

POLITECNICO DI TORINO

Collegio di Ingegneria Gestionale

**Corso di Laurea Magistrale
in Ingegneria Gestionale**

Tesi di Laurea Magistrale

Progettazione e sviluppo di una dashboard per l'analisi della qualità dell'aria nel contesto urbano. Caso di studio: la città di Milano.



Relatori

prof. Silvia Chiusano
prof. Tania Cerquitelli

Candidato

Francesco Cagnazzo

Aprile 2018

Indice

1. Introduzione.....	1
2. Aspetti teorici della progettazione e analisi basi di dati.....	3
2.1 Basi di dati.....	3
2.2 Data Warehouse.....	3
2.2.1 Architettura del Data Warehouse.....	4
2.2.1.1 Strumenti ETL.....	6
2.2.1.2 Strumenti di Reportistica e OLAP.....	6
2.2.2 Fasi di Progettazione di un Data Warehouse.....	7
2.2.2.1 Analisi e Riconciliazione delle fonti dati.....	7
2.2.2.2 Analisi dei Requisiti.....	7
2.2.2.3 Progettazione Concettuale.....	8
2.2.2.4 Progettazione Logica.....	8
2.2.2.5 Progettazione Fisica.....	9
2.2.2.6 Progettazione dell'alimentazione.....	9
2.2.3 Dimensional Fact Model.....	9
2.2.4 Key Performance Indicator.....	10
2.3 Business Intelligence.....	11
2.3.1 Analisi What-If.....	12
2.3.2 Data Mining.....	12
2.3.3 Business Performance Management.....	12
3. Inquinamento atmosferico.....	13
3.1 Contesto.....	13
3.1.1 Fonti di inquinamento.....	13
3.2 Principali inquinanti.....	14
4. Stato dell'arte.....	17
4.1 Dashboard esistenti.....	17
4.2 Tecnologie disponibili.....	19
4.2.1 MySQL & PHP.....	19
4.2.2 Google Data Studio 360.....	20
4.2.3 Javascript.....	20
4.2.4 Angular JS.....	20
4.2.5 Analisi e scelta delle tecnologie.....	21
5. Framework Proposto	23
5.1 Fonti dei dati.....	23
5.2 Progettazione del Dimensional Fact Model.....	24
5.3 Progettazione dei Key Performance Indicator.....	26
5.4 Preparazione dei dati e creazione del Data Warehouse.....	27

5.4.1	Arpa Lombardia.....	27
5.4.2	Portale Comune di Milano.....	33
5.4.3	Wheater Underground.....	35
5.4.4	Viste.....	38
5.5	Estrazione e calcolo dei Key Performance Indicator.....	39
5.6	Sviluppo della Dashboard.....	41
5.6.1	Tecnologie Utilizzate.....	42
5.6.2	Dashboard.....	44
5.6.2.1	Monitoring Stations.....	45
5.6.2.2	Pollutant Concentration.....	47
5.6.2.3	Metereological Report.....	60
5.6.2.4	Vehicle Report.....	68
5.7	Confronto delle tecnologie per lo sviluppo della Dashboard.....	71
6.	Osservazioni sui dati.....	74
7.	Conclusioni e sviluppi futuri.....	78
8.	Bibliografia e Sitografia.....	79

1. Introduzione

Il fenomeno dell'inquinamento atmosferico, negli ultimi anni, è diventato un problema molto serio per molte città italiane, soprattutto per quelle come Milano con un'elevata numerosità demografica. I problemi conseguenza del fenomeno dell'inquinamento sono vari, sia per l'ambiente sia per la salute dei cittadini stessi. È stato così necessario l'intervento delle amministrazioni locali per prendere provvedimenti e decisioni, proprio atte a contenere l'inquinamento urbano. Per procedere alla risoluzione o quanto meno alla limitazione di questo grosso problema, è bene conoscere dettagliatamente questo fenomeno, le cause principali dell'emissione di sostanze inquinanti, e le sostanze stesse.

Una volta chiaro il problema e quali sono le informazioni più importanti da conoscere legate ad esso, è necessario raccogliere questi dati e fornire analisi e indicatori di facile comprensione per poter appunto arrivare a prendere determinate decisioni.

In quasi tutti gli ambiti (lavorativi e non), l'informazione è diventata il punto di riferimento per la pianificazione e il controllo delle attività. Di conseguenza, la raccolta di informazioni sotto forma di dati e la loro gestione, diventa un problema quando le classiche basi di dati non forniscono soluzioni efficienti e adeguate. Proprio per ovviare a tale necessità, è nato il fenomeno del Data Warehousing, che permette un accumulo sempre maggiore di dati e da essi estrae in tempi rapidi, informazioni di sintesi permettendo così la valutazione di un particolare fenomeno o evento.

L'obiettivo di questa tesi, quindi è proprio quello di apportare i vantaggi del data warehousing all'ambito dell'inquinamento atmosferico urbano nella città di Milano. Più precisamente, si vorrà progettare un DW che permetta a due differenti tipi di utenti: uno più esperto (come ad esempio un amministratore comunale) e a uno meno esperto (il comune cittadino) di avere report di analisi riguardo ai livelli di inquinamento della città e trovare possibili correlazioni con altri fattori; per permettere a tali utenti di poter prendere decisioni e fare valutazioni in maniera più informata. Tali utenti, avendo necessità differenti potranno consultare report e analisi diverse tramite due distinti strumenti: un'applicazione mobile per il comune cittadino interessato a valutare localmente la qualità dell'aria urbana e una dashboard per l'amministratore comunale interessato a comprendere le principali cause di inquinamento atmosferico che interessano la città.

La tesi seguirà un approccio graduale per raggiungere l'obiettivo di realizzare il DW e la dashboard appena descritti: nel presente capitolo, è stata data una descrizione generale degli strumenti che si vogliono sviluppare e per quali scopi questi sviluppi sono effettuati. Nel capitolo 2, si andrà nel dettaglio per chiarire alcuni aspetti teorici delle basi di dati, soprattutto dei DW, con la descrizione dei principi sui cui essi si basano, delle relative architetture e caratteristiche, della modellazione DFM, fino a vedere le singole fasi che compongono la progettazione e realizzazione di un Data Warehouse. Inoltre, verranno studiate le caratteristiche principali dei così detti KPI (Key Performance Indicator), che sono degli indicatori utili per la misurazione empirica di informazioni di interesse, tramite i dati estratti dal DW. Nel capitolo 3 si descriverà brevemente l'ambito di sviluppo di tale tesi, cioè l'inquinamento atmosferico, andando a vedere dal punto di vista teorico i principali inquinanti, le cause della produzione di essi e su quali aspetti vanno ad incidere tali inquinanti, sia dal punto di vista ambientale che della salute delle persone. Dopodiché, si passerà ad occuparsi dell'effettivo sviluppo di tale progetto, andando a vedere, nel capitolo 4, quali sono gli strumenti di visualizzazione già esistenti per mostrare report di questo genere e facendo una descrizione anche sulle varie tecnologie a disposizione che possono essere usate per tale sviluppo, andando inoltre a fare un confronto tra esse basandosi sui relativi vantaggi e svantaggi.

In seguito, nel capitolo 5, sarà la volta di progettare, modellare e realizzare il DW in questione, andando a descrivere nel dettaglio la sua composizione. Sarà necessario capire come e cosa voler mostrare all'utente finale e quale deve essere la granularità delle varie informazioni e i KPI fondamentali. Si è quindi prima passati da una progettazione concettuale con la definizione del Dimensional Fact Model, e poi ad una progettazione logica in cui le fonti dati a disposizione sono state pulite in base a quelle che sono le necessità individuate nella progettazione concettuale, e poi inserite nel database, pronte per essere poi consultate dagli applicativi creati. Si è poi ritenuto opportuno fare ricorso all'uso di una vista, la quale permette di rendere più immediato e performante l'utilizzo del data warehouse. Dopo aver progettato il data warehouse in maniera completa, quindi anche con la definizione delle tabelle riguardanti meteo e traffico oltre a quella degli inquinanti, è stato sviluppato lo strumento di visualizzazione delle informazioni raccolte nel DW, cioè la dashboard.^[1]

Solo dopo aver realizzato il DW ci si focalizzerà su un metodo di visualizzazione di tali informazioni, cioè si passerà a realizzare una vera e propria dashboard, dedicata principalmente agli utenti più esperti e più interessati a valutare possibili legami tra inquinamento e altri fattori. Il personale dell'amministrazione locale è principalmente interessato a comprendere le principali cause di inquinamento che interessano la città cercando poi di garantire un buon livello di qualità dell'aria. Questi cercano di capire in quali punti della città l'inquinamento risulta aver raggiunto condizioni critiche. Sono anche interessati a valutare gli aspetti contestuali che possono contribuire all'inquinamento atmosferico come le condizioni climatiche e il traffico di veicoli nelle aree urbane. Il personale dell'amministrazione locale quindi deve analizzare i flussi completi dei dati raccolti (anche a diverse granularità spaziali e temporali), per osservare e comprendere il fenomeno osservato, valutare le diverse componenti e identificare le possibili cause.

Successivamente, tale tesi si concluderà con un aspetto più pratico, nel capitolo 6 si dimostrerà l'utilità della dashboard e soprattutto delle informazioni raccolte nel DW progettato; per fare ciò si effettueranno delle osservazioni sui dati, mostrando come questi possano risultare utili per delle decisioni amministrative.

Infine, nel capitolo 7, si descriveranno brevemente, quali aspetti di tale progetto potrebbero subire dei miglioramenti e come tale progetto potrebbe espandersi, come ad esempio tramite la realizzazione di un'apposita applicazione mobile.

¹ La progettazione e realizzazione del Data Warehouse è in comune con la Tesi Magistrale "Progettazione e sviluppo di una applicazione mobile per il monitoraggio della qualità dell'aria nel contesto urbano. Caso di studio: la città di Milano" di Nunzio Caggiano

2. Aspetti teorici della progettazione e analisi di basi di dati

In questo capitolo vengono presentati gli aspetti teorici relativi a sistemi e strumenti utilizzati nello sviluppo del progetto, in particolare quindi le basi di dati, il Data Warehouse e gli strumenti di Business Intelligence.

2.1 Basi di Dati

La base di dati rappresenta una collezione di dati riguardanti uno o più argomenti correlati tra loro, ed è strutturata in maniera tale da consentire ai dati di essere utilizzati per diverse applicazioni.

Il DBMS o Data Base Management System è un sistema software progettato per assistere, mantenere e utilizzare una grande collezione di dati, assicurando condivisione, persistenza e affidabilità. Tale collezione di dati è gestita da un DBMS. Le caratteristiche generali di un DBMS sono le seguenti:

- possibilità di creare nuovi database e di specificare i rispettivi schemi mediante un linguaggio specializzato chiamato data definition language;
- permettere agli utenti di eseguire query ai dati e di modificarli, facendo uso di un apposito linguaggio detto query language o data-manipulation language;
- consente di salvare una enorme quantità di dati per un lungo periodo di tempo, rendendoli sicuri;
- garantisce l'integrità evitando accessi simultanei allo stesso dato, controllando l'accesso da più utenti alla volta. ^[1]

2.2 Data Warehouse

Un data warehouse (DW) è un archivio informatico contenente i dati di un'organizzazione, necessari per eventuali analisi di dati con lo scopo dell'attuazione di processi decisionali e del miglioramento del patrimonio informativo.

Alla base del Data warehousing ci sono alcuni strumenti tra cui quelli per localizzare, estrarre, trasformare e caricare i dati. Altri componenti importanti in un sistema di data warehouse sono quelli di business intelligence e quelli per gestire e recuperare i metadati.

Nel 1996, fu data da William H. Inmon per la prima volta la definizione di data warehouse:

“Un Data Warehouse è una collezione di dati di supporto per il processo decisionale che presenta le seguenti caratteristiche:

- *è integrata e consistente;*
- *è orientata ai soggetti di interesse;*
- *è rappresentativa dell'evoluzione temporale e non volatile.”*

Andando nel dettaglio della definizione di DW di Inmon, si analizzano di seguito le caratteristiche:

- Integrata e consistente: in un DW possono confluire dati provenienti da varie fonti eterogenee: dati estratti dall'ambiente di produzione e dati provenienti da sistemi informativi esterni all'azienda. Tramite l'utilizzo di metodi di codifica uniformi, il perseguimento di una omogeneità semantica di tutte le variabili, l'utilizzo delle stesse unità di misura, può essere raggiunta l'integrazione.
- Orientata al soggetto: il DW è incentrato sui concetti di interesse dell'azienda, quali clienti, prodotti, vendite e ordini. Si passa dalla progettazione per funzioni ad una modellazione dei dati che consenta una visione multidimensionale degli stessi.
- Variabile nel tempo: I dati contenuti in un DW sono aggiornati fino ad una certa data che di solito è antecedente a quella in cui l'utente interroga il sistema. Questa è una caratteristica che lo differenzia da un sistema transazionale, il quale invece è incapace di fornire un quadro storico di ciò che viene analizzato e in cui i dati corrispondono sempre ad una situazione aggiornata.
- Non volatile: in confronto alla progettazione di un'applicazione transazionale il DW ha una facilità di progettazione maggiore. Questa consente degli accessi in sola lettura per cui i dati contenuti nel DW non sono modificabili. Questo fa sì che non ci siano più le possibili anomalie dovute agli aggiornamenti, né si ricorre a strumenti complessi per gestire l'integrità referenziale o per bloccare record a cui possono accedere altri utenti in fase di aggiornamento.

Il data warehouse viene usato come supporto ai decision maker andando a descrivere quello che è il processo di acquisizione, trasformazione e distribuzione di informazioni presenti all'interno o all'esterno delle aziende. Questo lo differenzia dai normali sistemi gestionali i quali, al contrario, hanno il compito di automatizzare le operazioni di routine. Un data warehouse può essere costruito secondo diverse modalità che spaziano da una logica completamente accentrata ad una logica completamente distribuita, questo quando il focus viene posto sulla capacità di supportare il processo decisionale

Un concetto fondamentale del DW, da analizzare prima di entrare nel dettaglio delle possibili architetture, è quella di multidimensionalità. Dato che gli eventi che si susseguono in un'azienda sono molteplici per poterli raggruppare e selezionare si immagina di collocarli in uno spazio n-dimensionale i cui assi (dimensioni di analisi) definiscono le diverse prospettive per la loro identificazione. Dal concetto di dimensione nasce la diffusissima metafora del cubo per la rappresentazione dei dati multidimensionali: si immaginano gli eventi come celle di un cubo i cui spigoli rappresentano le dimensioni di analisi. Queste celle contengono valori per ciascuna misura, per cui un cubo multidimensionale è incentrato su un fatto di interesse per il processo decisionale e che indica un insieme di eventi, descritti in maniera quantitativa da misure numeriche. Gli assi di tale cubo rappresentano una possibile dimensione di analisi; ed ogni dimensione può avere più livelli di dettaglio individuati da attributi strutturati in gerarchie. [2]

2.2.1 Architettura del Data Warehouse

In un sistema di data warehousing le caratteristiche fondamentali sono le seguenti:

- Separazione: mantenere separate l'elaborazione analitica e quella transazionale.
- Scalabilità: con la crescita dei volumi di dati da gestire ed elaborare e del numero di utenti da soddisfare, l'architettura hardware e quella software devono poter essere facilmente ridimensionate.

- Estendibilità: deve essere in grado di supportare nuove applicazioni e tecnologie senza riprogettare integralmente il sistema.
- Sicurezza: i dati memorizzati hanno natura strategica per cui il controllo sugli accessi è essenziale.
- Amministrabilità: la complessità dell'attività di amministrazione non deve risultare eccessiva.

Esistono diverse architetture per la progettazione di un sistema data warehousing; tuttavia, quella più utilizzata risulta essere l'architettura a due livelli che evidenzia la separazione tra il livello delle sorgenti e quello del DW. In realtà, questo si estende su quattro livelli distinti che descrivono stadi successivi del flusso di dati:

1. *Livello delle sorgenti.* Le fonti di dati utilizzate dal DW sono eterogenee e possono essere estratte dall'ambiente di produzione (originariamente archiviati in database aziendali relazionali) oppure provenire da sistemi informativi non interni all'azienda.

2. *Livello dell'alimentazione.* Per eliminare le inconsistenze e completare eventuali parti mancanti, i dati memorizzati nelle sorgenti devono essere estratti e ripuliti. Si parla così di strumenti ETL (Extraction, Transformation and Loading) i quali permettono di integrare schemi eterogenei, nonché di estrarre, trasformare, pulire, validare, filtrare e caricare i dati dalle sorgenti nel DW.

3. *Livello del warehouse.* Tutte le informazioni raccolte dalla varie sorgenti vanno a caratterizzare il DW. Esso può essere consultato e usato come sorgente per costruire data mart ossia un sottoinsieme o un'aggregazione dei dati presenti nel DW primario.

4. *Livello di analisi.* Permette la consultazione efficiente e flessibile dei dati integrati a fini di stesura di report, di analisi, di simulazione.

Le principali motivazioni a sostegno dell'utilizzo di questa architettura sono:

- E' sempre disponibile informazione di buona qualità anche quando è precluso l'accesso alle sorgenti a livello del warehouse.
- L'interrogazione analitica effettuata sul DW non interferisce con la gestione delle transazioni a livello operativo
- L'organizzazione logica del DW si basa sul modello descritto in precedenza ovvero quello multidimensionale
- Tra i sistemi OLTP (trattano dati correnti al massimo livello di dettaglio), e sistemi OLAP (operano su dati storici e di sintesi), c'è una importante discordanza temporale e di granularità.
- Per ottimizzare le prestazioni per applicazioni di analisi e reportistica si possono impiegare tecniche specifiche.

Nella Figura 2.1 è riportata la struttura dell'architettura a due livelli appena descritta.

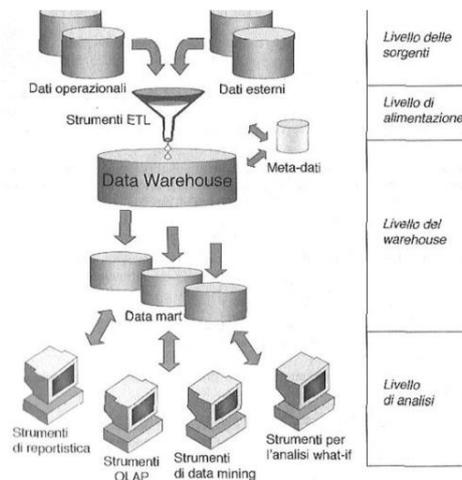


Figura 2.1: Architettura data warehouse a due livelli. [2]

Nei successivi paragrafi sono analizzati nel dettaglio alcune delle componenti presenti in questa architettura. [2]

2.2.1.1 Strumenti ETL

Componenti fondamentali per alimentare una sorgente dati singola, esauriente e di qualità in grado di alimentare il DW risultano essere gli strumenti ETL (Extraction, Transformation and Loading) in una architettura a due livelli.

Questa operazione si compone di quattro fasi:

1. *Estrazione.* Fase in cui i dati rilevanti vengono estratti dalle sorgenti. Se il DW viene popolato per la prima volta si parla di estrazione statica, se invece c'è un aggiornamento periodico del DW che cattura solamente i cambiamenti avvenuti nelle sorgenti dall'ultima estrazione, si parla di estrazione incrementale.
2. *Pulitura.* È una fase importantissima nel processo di data warehousing, in quanto si occupa di migliorare la qualità dei dati. I dati possono essere sporchi a causa di alcuni fattori: dati duplicati, dati mancanti, valori errati, valori inconsistenti.
3. *Trasformazione.* Fase in cui i dati vengono convertiti dal formato operativo sorgente a quello del DW. La presenza di più fonti distinte eterogenee rende complicata la corrispondenza con il livello sorgente, e questo richiede durante la progettazione una complessa fase di integrazione.
4. *Caricamento.* E' l'ultima fase da eseguire e si tratta di caricare i dati nel DW. Questo può avvenire secondo due modalità:
 - Refresh. I dati del DW vengono riscritti integralmente, sostituendo quelli precedenti.
 - Update. I soli cambiamenti occorsi nei dati sorgente vengono aggiunti nel DW, tipicamente senza distruggere o alterare i dati esistenti.

2.2.1.2 Strumenti di Reportistica e OLAP

Nell'architettura a due livelli prima descritta, a livello di analisi vengono utilizzati differenti strumenti. In particolare, per quanto riguarda quelli di data mining e what-if si rimanda al paragrafo riguardante BI; invece di seguito sono descritti quelli di reportistica e OLAP.

Gli strumenti di reportistica sono principalmente orientati agli utenti con necessità di accedere a informazioni strutturate in modo invariabile e a intervalli di tempo stabiliti. Un

rapporto (report) è invece caratterizzato da un'interrogazione in grado di generare la selezione e l'aggregazione dei dati multidimensionali e da una presentazione che può essere in forma tabellare o grafica.

Tuttavia, la principale modalità di fruizione delle informazioni contenute in un DW è rappresentata dagli strumenti OLAP: essi consentono agli utenti con difficoltà nell'effettuare le analisi da fare a priori, di esplorare in maniera interattiva i dati sulla base del modello multidimensionale. Quindi, gli utenti degli strumenti di reportistica svolgono un ruolo passivo, gli utenti OLAP sono in grado di costruire attivamente una sessione di analisi complessa in cui ogni passo effettuato è conseguenza dei risultati ottenuti al passo precedente.

Le funzioni di base di uno strumento OLAP sono:

- *Slicing*: permette di ridurre la dimensionalità del cubo andando a fissare un valore per una delle dimensioni;
- *Dicing*: mediante un determinato criterio di selezione viene ridotto l'insieme dei dati oggetto di analisi;
- *Roll-up*: tramite l'eliminazione di un livello di dettaglio da una gerarchia, c'è un aumento nell'aggregazione dei dati
- *Drill-down*: riduce l'aggregazione dei dati introducendo un altro livello di dettaglio in una gerarchia;
- *Drill-across*: permette un collegamento tra due o più cubi correlati al fine di compararne i dati;
- *Drill-through*: consiste nel passaggio dai dati aggregati multidimensionali del DW ai dati operazionali presenti nelle sorgenti;
- *Pivoting*: ha l'obiettivo di analizzare le stesse informazioni sotto diversi punti di vista con un cambiamento nella modalità di presentazione.

2.2.2 Fasi di progettazione di un Data Warehouse

2.2.2.1 Analisi e riconciliazione delle fonti dati

I dati presenti in un DW sono ricavati da varie sorgenti diverse sia per la tecnologia che le gestisce e sia per il modello con cui viene rappresentata la realtà aziendale. Per cui, durante la prima fase di progettazione è necessario andare a definire lo schema dei dati operazionali a partire dal quale verrà alimentato il data mart. Per fare questo si analizzano gli schemi delle sorgenti disponibili, le correlazioni tra le varie sorgenti e si valuta quella che è la qualità dei dati. Ad essere coinvolti in questa fase risultano essere il progettista e gli amministratori dei database operazionali capaci di dare un significato a schemi e tracciati record non facilmente comprensibili ad altri. Un aspetto molto importante del data warehousing è il concetto di dato integrato che permette di derivare informazioni consistenti e prive di errori, dalle quali poi è necessario un processo di riconciliazione che comporta integrazione, pulizia e trasformazione dei dati.

2.2.2.2 Analisi dei requisiti

Fase adibita alla raccolta, filtraggio e analisi dei requisiti degli utenti finali da parte del progettista, con l'obiettivo di capire quali sono le informazioni di interesse strategico. Da questa fase vengono poi individuati i fatti da modellare e alcune indicazioni sul carico di lavoro. La scelta dei fatti dipende dall'utente finale e per ognuno di essi è fondamentale definire l'intervallo di storicizzazione, ossia l'arco temporale che gli eventi memorizzati

dovranno abbracciare. La definizione del carico di lavoro ha lo scopo di consentire al progettista di identificare dimensioni e misure per la progettazione concettuale.

Dopo aver determinato il carico di lavoro viene definita la granularità (dipende da velocità di risposta richiesta e livello di dettaglio delle interrogazioni) nella rappresentazione dei fatti, cosa molto importante per la riuscita dell'intero progetto in quanto va a determinare la flessibilità di interrogazione del data mart.

Ci sono varie tecniche per l'analisi dei requisiti utente, tra le quali si distingue l'approccio basato sull'utilizzo del formalismo di Tropos. Quindi nel contesto specifico dei data warehouse, è necessario introdurre alcuni nuovi concetti:

- Fatti. Un fatto modella un insieme di eventi che si verificano quando un obiettivo viene raggiunto.
- Attributi. Sono dei campi la cui valorizzazione si accompagna alla registrazione di un fatto ad opera di un obiettivo.
- Dimensioni. Una dimensione è una proprietà di un fatto che ne descrive una possibile coordinata di analisi.
- Misure. Una misura è una proprietà numerica di un fatto che ne descrive un aspetto quantitativo.

2.2.2.3 Progettazione concettuale

E' la fase in cui viene disegnato uno schema concettuale per il data mart. Questo schema è il così detto Dimensional Fact Model (DFM), in cui per ogni fatto di interesse evidenziato dall'utente deve essere creato uno schema che permetta di descrivere in maniera grafica tutti i concetti del modello multidimensionale: fatti, misure, dimensione e gerarchie. Ci sono tre fasi per quanto riguarda la progettazione concettuale:

1. *Mappatura dei requisiti.* Lo scopo principale di questa fase è quello di creare una corrispondenza tra fatti, dimensioni e misure individuate precedentemente e le relazioni e attributi presenti nello schema operativo.
2. *Costruzione dello schema di fatto.* Ciascuna dimensione e misura mappati con successo da un diagramma di ragionamento esteso allo schema operativo vengono inclusi nello schema di fatto. Successivamente vengono aggiunti anche gli attributi, eliminando quelli inutilizzati.
3. *Raffinamento.* Avviene una affinazione dello schema di fatto per renderlo sempre più vicino ai fabbisogni dell'utente.

Nel paragrafo 2.2.3 verrà analizzato più nel dettaglio il DFM

2.2.2.4 Progettazione logica

E' la fase in cui vengono determinate le procedure necessarie per andare a determinare lo schema logico del data mart, partendo dallo schema concettuale precedentemente estratto.

Il principale obiettivo di questa fase è la massimizzazione della velocità di reperimento dati, autorizzando il ripetuto utilizzo di dati ridondanti e denormalizzati. Essa può essere riassunta in tre fasi:

- Traduzione degli schemi di fatto in schemi logici. Se schema logico a stella, la tabella dei fatti presenta tutti gli attributi e tutte le misure collegate al fatto in maniera diretta ed inoltre per le gerarchie vengono create tabelle delle dimensioni contenenti tutti gli attributi.

- Materializzazione delle viste. Processo con cui vengono selezionate un insieme di viste con lo scopo di esaltare gli obiettivi di un progetto.
- Frammentazione delle viste. La tabella viene suddivisa in più tabelle chiamate frammenti in maniera tale da aumentare le prestazioni del sistema

2.2.2.5 Progettazione Fisica

La fase di progettazione fisica è una fase in cui vengono individuati gli indici che ottimizzano le prestazioni. In questa fase un ruolo importante lo svolgono il carico di lavoro e il volume dati e inoltre occorre anche riferirsi ad uno specifico DBMS.

2.2.2.6 Progettazione dell'alimentazione

E' una fase in cui si prendono le decisioni sul processo di alimentazione del livello riconciliato e del data mart.

Fondamentale è la scelta, a partire dalle sue sorgenti, dell'intervallo di aggiornamento periodico del data mart.

2.2.3 Dimensional Fact Model

Il Dimensional Fact Model (DFM) è una vera e propria specializzazione del modello multidimensionale discusso in precedenza per le applicazioni di data warehousing. E' di tipo grafico e altro non è che un modello concettuale creato per fare da supporto alla progettazione di data mart. Il DFM si pone i seguenti obiettivi:

- dare supporto efficace alla progettazione concettuale;
- creare un ambiente in cui le interrogazioni utente possano essere formulate in modo intuitivo;
- rendere possibile la comunicazione tra il progettista e l'utente finale con l'obiettivo di raffinare le specifiche dei requisiti;
- costituire una piattaforma stabile per la progettazione logica;
- fornire una documentazione di progetto espressiva e non ambigua.

Il DFM genera una rappresentazione concettuale costituita da un insieme di schemi di fatto. Gli elementi di base modellati da questi schemi sono:

- Fatto: permette di modellare un insieme di eventi che avvengono nell'impresa ed esprime il concetto di interesse per il processo decisionale. Il fatto deve evolvere nel tempo quindi è fondamentale che abbia aspetti dinamici.
- Misura: descrive un aspetto di interesse quantitativo per l'analisi ed è quindi una vera e propria proprietà numerica di un fatto.
- Dimensione: va a descrivere le coordinate di analisi e indica la proprietà con dominio finito di un fatto. Ogni fatto di solito ha più dimensioni che ne vanno a determinare la minima granularità di rappresentazione.
- Evento primario: è l'occorrenza di un fatto, realizzata da una ennupla formata da un valore per ogni dimensione. Ad ogni evento primario viene associato un valore per ciascuna misura.
- Attributo dimensionale: sono gli attributi e le dimensioni, a valori discreti che le descrivono.

- Gerarchia: albero in cui i nodi sono attributi dimensionali e gli archi vanno a modellare associazioni di tipo molti a uno tra coppie di attributi dimensionali. Essa comprende una dimensione indicata sulla radice dell'albero oltre agli attributi dimensionali che la descrivono.
- Evento secondario: considerando un blocco di attributi dimensionali, ogni ennupla di loro valori porta ad un evento secondario che aggrega tutti gli eventi primari ad essa associato.

Ad ognuno di questi eventi è anche associato un valore per misura, che sintetizza tutti i valori della stessa misura negli eventi primari corrispondenti. Vengono ora introdotti alcuni concetti di modellazione avanzata del DFM:

- Attributi descrittivi: sono attributi che non sono usati per l'aggregazione in quanto hanno spesso domini con valori continui. Questi attributi specificano una proprietà di un attributo dimensionale e di una gerarchia ed è da quest'ultimo determinato mediante una dipendenza funzionale.
- Attributi cross-dimensionali: attributi di tipo dimensionale o descrittivo, il cui valore è determinato dalla combinazione di due o più attributi dimensionali, che possono anche appartenere a gerarchie distinte.
- Convergenza: avviene quando all'interno di una gerarchia due attributi dimensionali risultano essere collegati da due o più percorsi alternativi di associazioni molti a uno.
- Archi multipli: sono una connessione tra due attributi dimensionali che modellano un'associazione molti a molti
- Archi opzionali: vengono usati nel caso in cui un'associazione rappresentata in uno schema di fatto non viene definita per un sottoinsieme di eventi. ^[2]

2.2.4 Key Processor Indicator

I Key Process Indicator (KPI), sono indicatori numerici che consentono di valutare diversi aspetti dei processi aziendali. Tramite il loro utilizzo i manager capiscono come i vari processi gestiti dal sistema informativo stanno procedendo.

I KPI devono possedere cinque caratteristiche (SMART) ^[3]:

- *Specific*: il KPI deve essere specifico per ogni processo, e chiaramente dipende dall'azienda in questione e dal modo di lavorare. Non è possibile definire a priori un indicatore numerico buono in assoluto, ma dipendono dal singolo processo, ad esempio le soglie di minimo e di massimo che non devono essere superate non possono essere definite.
- *Measurable*: il KPI deve essere facile da misurare e la sua misurazione deve essere fattibile sia a livello economico sia a livello pratico.
- *Achievable*: per ogni indicatore le soglie stabilite devono sempre essere realisticamente raggiungibili oltre che essere sotto il controllo di chi valuta l'indicatore.
- *Relevant*: un indicatore deve essere rilevante. Questo vuol dire che si deve essere in grado di interpretare l'indicatore quindi capire da esso cosa non funziona, fornendo indicazioni per intervenire, correggendo, sul processo. Per cui, in generale, risulta essere utile misurare gli indicatori solo se le informazioni che ricavo sono utili per le esigenze informative. E' importante definire quelli che sono gli stakeholders e gli utenti a cui interessa il KPI ed è necessario definirli in maniera tale che essi possano valutare il processo.
- *Timely*: il KPI deve essere disponibile nei tempi prestabiliti: infatti se un indicatore non risulta essere disponibile nel momento in cui serve risulta essere inutile, anche se è un

indicatore fondamentale. Qualsiasi indicatore ha un intervallo di tempo in cui è utile e rilevante.

Per fornire una visione completa e multidimensionale è di grande importanza andare a scegliere indicatori che abbraccino diverse categorie e che possano bilanciarsi a vicenda. Ci sono 4 principali categorie di KPI:

1. *KPI Generali o di Volume*: sono la base per nuovi indicatori e consentono di comprendere il contesto in cui avviene il processo.
2. *KPI di Efficienza*: ci dà informazioni sulla bontà o meno del processo rispetto alle risorse utilizzate e alla quantità di output prodotto.
3. *KPI di Qualità*: verificano la qualità del prodotto come conseguenza del processo andandosi a focalizzare sui prodotti del processo finito.
4. *KPI di Servizio*: cercano di capire la relazione che c'è tra una richiesta e la relativa risposta ottenuta in termini di prodotto, di tempestività etc. Per cui questo tipo di KPI sono legati agli input e agli output del processo.

2.3 Business Intelligence

Tra gli strumenti informatici di supporto alle attività decisionali, i sistemi di data warehousing, prima descritti, sono sicuramente i più diffusi. Tuttavia esistono insieme più ampi ed eterogenei di soluzioni, i quali prendono il nome di sistemi di Business Intelligence (BI). La Business Intelligence è una disciplina che consente ai decisori aziendali, di capire attraverso soluzioni software, i fattori chiave del business e in seguito di prendere le migliori decisioni in quel momento.

Il DW fino ad oggi è stato lo strumento principale per la BI, grazie ai vantaggi e le potenzialità descritti nei paragrafi precedenti. Tuttavia, andare ad utilizzare i DW comporta alcuni limiti, sono qui di seguito elencati quelli principali:

- L'intervallo di aggiornamento dei dati difficilmente è minore della settimana/giorno
- Se utilizzato per formulare interrogazioni complesse che non fanno parte del modello multidimensionale, il DW è poco efficiente.
- Non permette di formulare scenari di previsione in quanto consente di registrare solo il passato

Con il passare degli anni le aziende sono cambiate e ciò ha reso necessaria l'adozione di nuove soluzioni che consentano di rispondere alle sempre più crescenti richieste informative, che si possono riassumere nei seguenti punti:

- Tecniche di analisi più potenti e non basate sul modello multidimensionale;
- Possibilità di effettuare analisi sui dati provenienti da sorgenti informative eterogenee e con aggiornamenti rapidi
- Possibilità di prevedere il futuro e non solo di analizzare il passato.

Tre possibili proposte per il soddisfacimento dei punti precedenti sono: l'analisi what-if, il data mining e il Business Performance Management. ^[2]

2.3.1 Analisi What-If

L'analisi what-if supera uno dei limiti degli strumenti di reportistica e OLAP ossia quello di andare a tenere traccia solo del passato non consentendo di analizzare scenari futuri. Il what if è l'insieme di tecniche di analisi previsionale che consentono di valutare il comportamento di un sistema reale assumendo un particolare insieme di condizioni iniziali.

Si possono classificare le tecniche di analisi what-if in base all'approccio utilizzato quando si elabora il modello:

- Tecniche induttive: in base al comportamento che il sistema ha avuto durante un certo intervallo temporale si cerca di ricavarne il modello.
- Tecniche deduttive: puntano ad identificare e caratterizzare i rapporti di tipo causa-effetto tra i vari componenti del sistema.

2.3.2 Data Mining

Il Data Mining è quel processo che permette di estrarre conoscenza da banche dati di dimensioni molto elevate mediante l'applicazione di algoritmi che individuano le associazioni nascoste tra le informazioni e le rendono visibili. L'utente non riesce ad individuare tutti i pattern significativi quando ci sono moli di dati molto alte. Per pattern si intende la rappresentazione sintetica e ricca di semantica di un insieme di dati. Il Data Mining ha come obiettivo quello di aiutare l'utente nella ricerca del pattern e per farlo raccoglie un insieme di tecniche e metodologie dalle aree dell'intelligenza artificiale e del pattern recognition. Questo fa sì che non occorre più effettuare manualmente l'analisi dei dati, ma basta indicare con una certa approssimazione cosa e dove si vuole ricercare e lasciare che uno strumento automatico si accoli il peso della computazione.

Le tipologie di pattern più usate ed estraibili dai dati con tecniche di data mining sono:

- Regole Associative: consentono di individuare i gruppi di affinità tra oggetti, quindi determina le regole di implicazione logica presente nelle basi di dati.
- Clustering: è un insieme di tecniche di analisi multivariata dei dati con lo scopo di selezionare e raggruppare elementi omogenei in un insieme di dati.
- Alberi Decisionali: rappresentano un particolare tipo di classificatore, vengono usati per la comprensione di un particolare fenomeno poiché permettono di classificare in ordine di importanza, le cause che portano al verificarsi di un evento;
- Serie Temporal: a partire da sequenze di dati complesse, individuano i pattern ricorrenti o atipici.

2.3.3 Business Performance Management

Il Business Performance Management (BPM) rappresenta tutte le attività che permettono di misurare le proprie prestazioni andando ad incoraggiare l'uso efficiente delle risorse umane e l'efficacia dei processi aziendali

3. Inquinamento atmosferico

Questo capitolo è una introduzione teorica di quello che è il fenomeno dell'inquinamento atmosferico, le sue fonti e i suoi principali inquinanti. Inquinamento atmosferico che è l'argomento di fondo su cui vengono applicati i vantaggi del Data Warehouse.

3.1 Contesto

L'aria è una miscela di gas costituita da azoto (circa 78%), ossigeno (circa 21%) e da altri gas (argon, anidride carbonica, metano, ecc.) presenti in concentrazioni molto inferiori.

A causa di fenomeni di origine naturale (ad esempio eruzioni vulcaniche) e di origine antropica (attività umane), una sostanza presente nell'aria può avere concentrazioni diverse rispetto a quelle naturali. Quando si presentano queste alterazioni si parla di **inquinamento atmosferico**.

Le sorgenti artificiali di inquinamento atmosferico sono: le emissioni degli inquinanti industriali, delle centrali termiche, degli impianti di riscaldamento e dei mezzi di trasporto.

Gli inquinanti possono essere divisi in primari e secondari. I primi si hanno quando le sostanze contenute nelle emissioni si ritrovano direttamente nell'ambiente, le seconde invece si hanno quando queste sostanze subiscono in atmosfera dei processi di trasformazione. La presenza di tali inquinanti nell'aria va ad influire in maniera negativa su di essa, colpendo di conseguenza la salute delle persone. Per questo, molte città si adoperano per monitorare la qualità dell'aria urbana con l'obiettivo di prendere le dovute precauzioni e nello stesso tempo informare i cittadini. ^[4]

3.1.1 Fonti di inquinamento

Le cause dell'inquinamento atmosferico, come detto in precedenza, possono essere distinte in: fonti naturali (vulcani (SO₂), incendi (PM₁₀), ghiaioni (amianto), processi biologici (allergeni)) e fonti antropiche (traffico veicolare, riscaldamento domestico, industrie e attività artigianali, agricoltura e altre attività).

In particolare, per quanto riguarda il traffico veicolare, le emissioni dipendono dal tipo di combustibile, dal tipo di veicolo e da quando è stato prodotto. I veicoli alimentati a combustibili fossili provocano principalmente la produzione di anidride carbonica (CO₂), particolato (PM₁₀ e inferiori), idrocarburi (HC), ossidi di azoto (NO_x), mentre i veicoli a metano e GPL emettono NO_x, particolato ultrafine e scarsi idrocarburi. Negli ultimi anni si cerca di diminuire le emissioni da parte dei veicoli con l'introduzione delle fasce euro 1,2,3 ecc..

Un'altra fonte importante di sostanze inquinanti nell'aria è il riscaldamento domestico in cui gli inquinanti emessi dipendono dal combustibile utilizzato, dalla tipologia di riscaldamento, dalla tipologia, dalla vetustà e dalla manutenzione dello stesso.

La terza fonte più importante di inquinamento atmosferico riguarda l'industria e l'artigianato. Sono tantissime le lavorazioni in campo industriale che determinano inquinanti molto diversi tra loro in base alla lavorazione eseguita, e vanno da solventi, nebbie acide a metalli polveri ecc.. Tuttavia nel settore industriale le emissioni sono fortemente regolamentate per cui le industrie utilizzano diversi sistemi di abbattimento degli inquinanti.

Tra i principali contribuenti alle emissioni di sostanze inquinanti ci sono le industrie e la produzione di energia. Tuttavia, queste emissioni si verificano principalmente a notevoli

altezze. Le emissioni sono mescolate con grandi volumi d'aria e il loro contributo alle concentrazioni ambientali è spesso piccolo. Le emissioni da traffico invece avvengono a basse altezze, in strada, dove la dispersione e la diluizione è lenta. Quindi questo, potrebbe causare un gravissimo inquinamento dell'aria, anche quando le emissioni sono di piccole dimensioni.
[4]

3.2 Principali inquinanti

Le sostanze emesse da industrie, veicoli, centrali elettriche ecc vanno a contaminare l'aria che si respira. Queste sostanze sono un grosso problema poiché portano ad effetti dannosi sulla salute o sull'ambiente. Ovviamente in base alla quantità, alla durata e alla pericolosità dell'inquinante, gli effetti e l'impatto possono essere diversi. Ad esempio, sulla salute possono essere di piccola entità e reversibili, come può essere ad esempio una irritazione agli occhi, oppure debilitanti se non addirittura fatali, ad esempio il cancro.

Di seguito sono analizzati i principali inquinanti dell'aria. [5][6]

- *Particolato Atmosferico (PM10/PM2.5)*

A causa delle loro piccolissime dimensioni (<10 micron, <2,5 micron), il PM10 e il PM2.5 sono di grande interesse poiché entrano facilmente in profondità nei polmoni e quindi potenzialmente pongono rischi significativi per la salute.

Le particelle in sospensione nell'aria sono variabili sia nella forma fisica sia nella composizione chimica.

Le cause principale dell'emissione di PM10 e PM2.5 sono: le lavorazioni industriali, gli impianti di riscaldamento, l'usura dell'asfalto, degli pneumatici e dei freni e le emissioni di scarico degli autoveicoli, in particolare quelli con motore diesel.

Il PM10 viene trasportato a distanze molto elevate e costituisce uno dei principali mezzi di diffusione di composti tossici. Il PM10 ha anche una componente secondaria, formata direttamente in atmosfera a partire da altri inquinanti gassosi già presenti.

Dal punto di vista della salute, queste particelle possono arrivare in profondità nei polmoni provocando infiammazioni e peggioramenti nelle persone con malattie cardiache e polmonari e possono portare in superficie composti cancerogeni.

Inoltre le polveri sottili possono agire pesantemente su statue, monumenti ed edifici sporchi, aumentandone costi di pulizia e manutenzione.

Molto alta risulta essere la concentrazione di tale inquinante nelle zone maggiormente urbanizzate dove si verificano numerosi superamenti del limite giornaliero di 50 µg/m³. Un grosso problema è che la componente secondaria di tale inquinante è un "fondo" su cui risulta difficile incidere in maniera significativa.

- *Ossidi di Azoto (NOx)*

Con il termine NOX si descrive una miscela di ossido nitrico (NO) e biossido di azoto (NO₂). NO è prodotto in quantità molto maggiori di NO₂, ma si ossida a NO₂ nell'atmosfera. L'emissione dell'NOx avviene quando il carburante viene bruciato ad esempio nel settore dei trasporti, nei processi industriali e nella produzione di energia.

Nel caso di un elevato irraggiamento solare porta allo smog fotochimico il quale a sua volta crea altre sostanze inquinanti. Per questo è da considerarsi uno degli inquinanti atmosferici più pericolosi.

I gas di scarico delle auto risultano essere una delle principali fonti di produzione di NOx. La quantità delle emissioni dipende dalle caratteristiche del motore e dalla modalità del suo utilizzo, questo poiché se il motore lavora ad un elevato numero di giri la produzione di NOx

aumenta.

Esso dipende anche dagli impianti di riscaldamento infatti durante la stagione invernale a causa degli impianti di riscaldamento, le concentrazioni di tale inquinante aumentano,

Dal punto di vista della salute, il biossido di azoto è in grado di irritare i polmoni e dare minore resistenza alle infezioni respiratorie come l'influenza. Ad essere molto soggetti ad eventuali problemi causati da questo inquinante sono i bambini.

Inoltre, gli ossidi di azoto contribuiscono in maniera significativa alla formazione di piogge acide e all'aumento della concentrazione di nitrati nel suolo.

- *Ozono (O₃)*

L'ozono troposferico (O₃) è un inquinante che non viene emesso direttamente in atmosfera, ma è prodotto dalla reazione tra il biossido di azoto (NO₂), idrocarburi e luce solare ed è quindi di tipo secondario. La luce solare fornisce l'energia per iniziare la formazione di ozono, per cui elevati livelli di ozono sono osservati durante il periodo estivo.

I livelli di ozono sono molto elevati nelle aree urbane (dove alti livelli di NO sono emessi dai veicoli), più che nelle zone rurali.

Dal punto di vista della salute, l'ozono è causa di irritazione alle vie respiratorie dei polmoni, aumentando i sintomi di chi soffre di asma e malattie polmonari.

Nell'ambiente invece, l'ozono provoca danni ad una serie di colture commerciali ed alla vegetazione naturale. Esso è un potente agente ossidante nel deterioramento della qualità della gomma, dei tessuti.

- *Biossido di Zolfo (SO₂)*

L'SO₂ è causata da combustibili fossili bruciati che contengono tracce di composti di zolfo ed è dovuto alla produzione di energia e a fonti di trasporto di piccole dimensioni.

Le emissioni di SO₂ sono molto variabili a seconda della stagione, Infatti presenta valori massimi nella stagione invernale, laddove sono in funzione impianti di riscaldamento domestici.

L'esposizione all'SO₂ può danneggiare la salute causando una diminuzione della funzione polmonare negli asmatici. L'inquinamento dovuto all'anidride solforosa è considerato più dannoso quando sono elevati le concentrazioni da inquinamento di particolato e da altri composti.

Dal punto di vista ambientale, il principale costituente delle piogge acide è proprio l'acido solforico generato da reazioni atmosferiche di SO₂. Inoltre le particelle di solfato di ammonio sono quelle più abbondanti che si trovano in aria. Questo inquinante risulta anche essere una minaccia per determinati tipi di pietra utilizzati negli edifici.

- *Monossido di Carbonio (CO)*

L'incompleta combustione dei materiali contenenti carbonio (tra cui quelli usati nei mezzi di trasporto) producono il CO, che è un gas inodore, insapore e incolore. Questo inquinante ha concentrazioni più elevate laddove c'è un elevato traffico urbano intenso e rallentato, infatti i valori risultano più elevati con motore al minimo e in fase di decelerazione. Tuttavia, un aspetto positivo, è quello che tali concentrazioni raramente superano gli standard relativi alla salute, anche nei centri urbani più affollati.

Il CO va a ridurre la capacità di trasporto dell'ossigeno nel sangue, agendo mediante reazione con emoglobina.

Inoltre il monossido di carbonio ha anche un importante ruolo nella formazione di ozono ed è trasformato in anidride carbonica.

- *Benzene (C₆H₆)*

Il benzene (C₆H₆), emesso dall'uso del petrolio e da oli minerali, altro non è che un idrocarburo aromatico incolore, liquido e infiammabile.

Il benzene è una sostanza cancerogena per l'uomo. Con esposizione a concentrazioni elevate, si verificano danni al midollo osseo, mentre con una esposizione cronica professionale sono frequenti i casi di leucemia (molto frequenti nei lavoratori dell'industria manifatturiera, dell'industria della gomma e dell'industria petrolifera).

4. Stato dell'arte

Il fenomeno dell'inquinamento atmosferico è un problema che ad oggi attira molto interesse a livello mondiale.

Tale interesse proviene sia dalle amministrazioni, sia dal singolo cittadino. L'attenzione delle amministrazioni riguardo la raccolta, analisi e rappresentazione dei dati relativi a questo fenomeno è legata alla volontà di capirne le cause e di conseguenza prendere le dovute decisioni per risolvere o quanto meno limitare questo problema. D'altra parte, il singolo cittadino che ha interesse per la propria salute ha necessità di conoscere la qualità dell'aria che respira giornalmente per prendere eventualmente le dovute precauzioni.

Per ovviare a queste necessità, varie città mondiali hanno fatto ricorso all'utilizzo di sistemi di monitoraggio dell'inquinamento atmosferico per la raccolta di informazioni riguardanti i vari inquinanti, i quali insieme ai dati meteorologici e quelli sulla numerosità e tipologia di veicoli in circolazione, hanno permesso delle analisi più approfondite in merito al fenomeno dell'inquinamento.

Negli ultimi decenni, lo sviluppo di Internet e delle varie tecnologie informatiche ha consentito di realizzare differenti dashboard/applicazioni di facile consultazione che rappresentassero in maniera grafica e numerica i risultati delle analisi dei dati raccolti.

Quindi, prima di realizzare la dashboard riguardante il Comune di Milano, è stato utile consultare un certo numero di siti web che mostrassero, come prima descritto, le informazioni relative all'inquinamento atmosferico per capire quali possono essere gli indicatori più importanti e come questi vengono solitamente rappresentati.

Inoltre, è stata fatta una analisi sulle tecnologie a disposizione per lo sviluppo Web con l'obiettivo di individuare quella più idonea per la realizzazione della dashboard.

4.1 Dashboard esistenti

- *United States Environmental Protection Agency*

L'Agenzia per la protezione dell'ambiente statunitense fornisce un sito che dà all'utente l'opportunità di accedere ai dati riguardanti i livelli dei diversi inquinanti nelle differenti zone degli USA, con diverse granularità temporali (giornaliera e annuale). Tali dati possono essere scaricati oppure consultabili tramite reports che forniscono informazioni di sintesi in forma tabellare o tramite rappresentazione su grafici.

Tali reports permettono all'utente di selezionare l'anno e la zona di interesse e per ogni città della zona scelta, visualizzare:

- il valore medio e massimo dell'indice AQI con la possibilità di conoscere il numero di giorni in cui il valore di tale indice supera una certa soglia
- il valore medio di ogni inquinante e l'eventuale segnalazione nel caso questo valore abbia superato i limiti consentiti
- il valore giornaliero dell'indice AQI dell'inquinante selezionato

Per quanto riguarda la visualizzazione su grafici dei dati analizzati, l'utente ha la possibilità di selezionare l'inquinante e la zona di interesse e di visualizzare:

- un Tile Plot relativo agli anni scelti dall'utente, con una colorazione diversa in base alla soglia di pericolo in cui ricade il singolo valore dell'inquinante

- un Line Chart sui valori dell'inquinante nel periodo e nella stazione di monitoraggio scelti dall'utente
- uno Scatter Plot con i valori di AQI giornalieri durante uno specifico anno dell'inquinante di riferimento e di un altro sempre scelto dall'utente.

- *Air Pollution: Real-Time Air Quality Index*

Tramite il sito si ha la possibilità di avere informazioni sui valori di AQI di differenti città mondiali sia a livello generale sia per singolo inquinante, con un aggiornamento quotidiano dei dati.

Nella dashboard di questo sito, l'utente potrà consultare una mappa della città selezionata, visualizzandone per ogni stazione di monitoraggio il livello di AQI giornaliero.

Tramite, una serie di istogrammi relativi ai singoli inquinanti, vengono mostrati i trend dell'AQI nelle precedenti 48 ore e i rispettivi valori massimi e minimi e sempre con la stessa rappresentazione grafica e con lo stesso periodo temporale vengono monitorati i trend dei principali fenomeni atmosferici: Temperatura, Pressione, Umidità e Velocità del vento.

Inoltre, tramite le previsioni metereologiche, vengono visualizzate le previsioni settimanali sul valore di AQI della città scelta.

- *Ambience Data*

Questo sito è relativo ad una azienda privata che fornisce a singoli cittadini o ad aziende un prodotto che funziona da piccola stazione di monitoraggio privata. In questo modo, chi installa questo apparecchio, potrà accedere tramite una dashboard personale a reports di analisi dettagliati riguardanti l'inquinamento da polveri sottili, i valori di AQI e ai dati di umidità, temperatura e pressione.

- *Air Quality in Europe*

Tale sito è relativo alla qualità dell'aria delle principali città europee, per le quali viene calcolato e il valore di CAQI giornaliero. Questo valore viene in primis mostrato tramite la mappa dell'Europa, dove per ogni città è indicato oltre che il CAQI di fondo della città (qualità dell'aria esterna vissuta dal cittadino medio), anche il CAQI lungo la strada (i cittadini interessati sono coloro che vivono, lavorano e visitano queste strade, incluse le persone in auto o bus). Su questa mappa possono essere visualizzati anche i valori di CAQI del giorno precedente e quelli previsti per il giorno successivo.

Tutte le informazioni riportate su mappa, possono essere consultate anche in forma tabellare, permettendo così un confronto più veloce e immediato tra le varie città europee. Per ogni città, inoltre, si possono conoscere i valori di CAQI (stradale e di fondo) per ogni singolo inquinante.

Il sito, dà anche l'opportunità di effettuare un confronto con granularità annuale tra le diverse città in negli ultimi 3 anni, utilizzando l'Indice Annuale della Qualità dell'Aria.

- *Open Puglia*

Questo sito riguarda la città di Bari e mostra in maniera statica all'utente i valori giornalieri di AQI calcolati per ogni stazione di monitoraggio della città. Questi valori sono riportati all'interno di un grafico a linee per rappresentare il trend dei 60 gg precedenti dell'inquinamento atmosferico nelle varie stazioni. In questo grafico, l'utente è facilitato nella valutazione del trend tramite l'appartenenza dei valori di AQI a tre diverse soglie di pericolosità riconoscibili dalla diversa colorazione.

Inoltre, si ha dettaglio maggiore sull'ultimo giorno disponibile grazie a una rappresentazione grafica tramite tachimetro per mostrare il livello di AQI su ogni stazione di monitoraggio e la relativa appartenenza ad una determinata soglia.

- *Arpae Emilia-Romagna*

Tale sito riguardante la regione Emilia-Romagna, fornisce differenti informazioni in diverse modalità di rappresentazione, con la possibilità all'utente di inserire i filtri per indicare i dati di proprio interesse. In particolare, indicando la data e l'inquinante, verranno mostrati su mappa i valori giornalieri dell'inquinante selezionato per tutte le città più importanti della Regione. Per ogni città visualizzata sarà possibile un maggior dettaglio di informazioni riguardanti i valori misurati nelle singole stazioni di monitoraggio della città stessa.

Oltre alla rappresentazione su mappa, il sito mostra in forma tabellare i dati relativi al PM10 nelle ultime due settimane con una granularità per Provincia, dando risalto ai valori oltre la soglia consentita per legge tramite una diversa colorazione.

Inoltre, per ogni inquinante, in forma tabellare vengono riportati i valori medi giornalieri rilevati nelle singole stazioni e il numero di volte in cui tali valori hanno superato la soglia consentita per legge a partire dal 1 Gennaio.

L'utente ha anche la possibilità di scaricare dei reports con granularità temporali diverse riguardanti una singola città.

- *Applicazioni mobile*

Alcuni delle dashboard appena descritte sono direttamente legate ad applicazioni mobile che vengono messe a disposizione degli utenti per fornire funzionalità aggiuntive, per riportare informazioni già presenti sulla dashboard o per fornire informazioni real-time.

In particolare, queste app si focalizzano sul presentare all'utente una informazione molto riassuntiva e chiara, infatti vengono solitamente mostrati report specifici basati sull'AQI del luogo in cui ci si localizza o della città cercata, tale indice viene spesso mostrato anche in maniera grafica tramite mappe. Un'altra funzionalità fornita spesso dalle applicazioni mobile è l'aggiornamento in tempo reale delle condizioni atmosferiche e le previsioni future sia su tali condizioni, sia sui livelli di inquinamento.

4.2 Tecnologie disponibili

4.2.1 MySQL & PHP

Nel mondo dell'IT esistono differenti software di tipo server, detti Database management System(DBMS), utili per la gestione dei database. Tra questi, ha riscosso notevole successo, anche molto velocemente, MySQL: un RDBMS (Relational Database Management System) open source composto da un client a riga di comando e un server. I punti di forza che hanno permesso a MySQL di raggiungere il successo, sono stati:

- alta efficienza nonostante le moli di dati affidate;
- integrazione di tutte le funzionalità che offrono i migliori DBMS: indici, trigger e stored procedure, ecc.
- altissima capacità di integrazione con i principali linguaggi di programmazione, ambienti di sviluppo e suite di programmi da ufficio (ODBC, Java, Mono, .NET, PHP, Python).

Per utilizzare e amministrare i database MySQL sono stati ideati dei Manager appositi. Il più importante è phpMyAdmin; esso permette di creare database e tabelle e di eseguire operazioni di ottimizzazione. Per l'utilizzo di phpMyAdmin è necessario un server web come Apache HTTP Server ed il supporto del linguaggio PHP. MySQL, al momento della creazione delle tabelle fornisce un feedback per evitare eventuali errori. Inoltre, mette a disposizione altre funzionalità riguardanti l'inserimento dei dati (popolazione del database), per le query, per il backup dei dati, ecc. Come detto in precedenza, per utilizzare il DBMS

MySQL è necessario phpMyAdmin che a sua volta deve supportare il linguaggio server-side PHP. Esso risiede in un server in remoto e tramite la richiesta HTTP da parte del client, in fase di esecuzione interpreta le informazioni ricevute, le elabora e restituisce un risultato al client che ha formulato la richiesta. Il PHP è stato sviluppato principalmente per il Web, infatti consente di accedere alle richieste HTTP di tipo GET e POST; questo è solo un esempio delle caratteristiche di questo linguaggio implementate in funzione del Web. Un altro esempio importante è l'accesso in lettura/scrittura ai cookie del browser e il supporto alle sessioni sul server.

Infine, PHP mette a disposizione delle librerie per accedere ai database (MySQL, Postgres, SQLite, ecc), per interagire con i server Web (Apache), per manipolare le immagini (GD) e per effettuare connessioni remote (cUrl). ^{[7][8]}

4.2.2 Google Data Studio 360

Per la rappresentazione di report e dashboard dinamiche, uno strumento da utilizzare potrebbe essere Google Data Studio 360. Questo servizio trasforma i dati, e fornisce analisi facili da leggere e personalizzabili.

Data Studio ha principalmente 4 funzionalità:

- Connessione ai dati di rapporti provenienti da database o da Fogli Google, Google Analytics.
- Creazione di dimensioni e metriche e calcolo per ripulire e trasformare i dati senza aggiornare i dati non elaborati.
- Visualizzazione di differenti tipi di grafici come serie temporali, grafici a barre, grafici a torta, ecc.
- Personalizzazione grafica della dashboard
- Condivisione dell'accesso e possibilità di modifica dei rapporti con altri,

4.2.3 Javascript

Javascript è un linguaggio lato client orientato ad oggetti e ad eventi, che tramite funzioni di script permette di creare effetti dinamici interattivi su siti web e applicazioni web. Tali effetti scaturiscono da eventi innescati dall'utente stesso della pagina web. Tali funzioni di script, utilizzati dunque nella logica di presentazione, possono essere opportunamente inserite in file HTML, in pagine JSP o in appositi file separati con estensione .js poi richiamati nella logica di business.

L'accesso ad operazioni specifiche è fornito dal programma ospite (quello che ospita ed esegue lo script) tramite un'API ben definita implementata dal programma ospite e fornita allo script. Per uno script JavaScript, l'esempio più noto di programma ospite è quello del browser. Un browser tipicamente incorpora un interprete JavaScript; questo quando viene visitata una pagina web interpreta il codice JavaScript che è stato portato in memoria primaria, e lo esegue. JavaScript riesce a rapportarsi con un browser tramite delle interfacce chiamate DOM (Document Object Model in italiano Modello a Oggetti del Documento). In ambito Web, nelle pagine HTML, viene integrato codice JavaScript basato sulla scrittura di piccole funzioni che interagiscono con il DOM del browser permettendo azioni come controllare i valori nei campi di input, nascondere o visualizzare determinati elementi, ecc., non possibili con il semplice HTML statico ^[9]

4.2.4 AngularJS

Per semplificare lo sviluppo di applicazioni web, Google ha sviluppato un web framework open source denominato AngularJS. Tale framework legge prima la pagina HTML, che ha

incapsulato degli attributi personalizzati aggiuntivi (esempio: ng-controller), interpreta questi attributi come delle direttive (comandi) e lega le parti di ingresso e uscita della pagina al modello. Questo modello è rappresentato da variabili Javascript a cui si può impostare manualmente il valore, tramite codice o recuperandolo da risorse JSON statiche o dinamiche.

AngularJS in pratica fornisce strumenti per costruire un'architettura modulare e testabile della logica applicativa dell'applicazione, ma allo stesso tempo si basa sull'approccio dichiarativo dell'HTML nella definizione dell'interfaccia grafica.

AngularJS permette di creare le Single Page Application, applicazioni moderne che sfruttano le più recenti tecnologie, esse hanno la particolarità di avere risorse che vengono caricate dinamicamente su richiesta, senza necessità di ricaricare l'intera pagina.

Gran parte del codice da scrivere viene eliminato grazie all'associazione di dati in AngularJS e all'iniezione di dipendenze. In questo modo, si risparmia molto tempo e l'interpretazione del tutto avviene all'interno del browser. Questo framework strutturale raggiunge quindi l'obiettivo di aumentare la capacità MVC (Model View Controller) delle applicazioni web. ^[10]

4.2.5 Analisi e scelta delle tecnologie

Dopo aver descritto dal punto di vista teorico le principali tecnologie a disposizione per la realizzazione della dashboard, viene fatta ora un'analisi più dettagliata sui pro e contro dell'utilizzare queste tecnologie, con l'obiettivo di individuare quelle più idonee.

La prima scelta da effettuare è quella riguardante la costruzione dei reports della dashboard. Quindi, tra le tecnologie viste in precedenza, la scelta sarà tra il DBMS MySQL con PHP e il servizio Google Data Studio 360.

Le differenze principali individuate tra le due alternative riguardano:

- Possibilità di gestione dei dati provenienti da tabelle diverse
- Necessità di conoscenza del linguaggio SQL e di scrittura di codice lato server
- Possibilità di integrazione di mappe
- Possibilità di personalizzare le informazioni da richiedere all'utente come parametri di ricerca

In particolare, MySQL e PHP permettono la gestione delle relazioni tra tabelle diverse (Join) a differenza di Google Data Studio in cui la consultazione dei dati da parte dell'utente può avvenire solo tramite reports relativi ad una singola tabella.

Inoltre, a differenza di MySQL/PHP, con Google Data Studio è possibile realizzare reports senza dover scrivere righe di codice lato server ed è possibile interrogare le tabelle contenenti i dati senza scrivere query in SQL. Se questo da una parte è un vantaggio poiché permette di creare reports in maniera rapida e veloce senza la necessità di avere conoscenze di programmazione, d'altra parte questo risulta essere uno svantaggio se si vogliono effettuare operazioni sui dati prima di mostrarli all'utente (con Google Data Studio, ciò è impossibile).

Per quanto riguarda le mappe, entrambe le tecnologie permettono l'utilizzo di esse, con la differenza che con Google Data Studio il massimo focus della mappa è per città. Invece, utilizzando MySQL e PHP, si possono sfruttare le API di Google per mostrare informazioni dettagliate sulla singola città. In aggiunta, con MySQL e PHP rispetto a Google Data Studio, è possibile indirizzare maggiormente l'utente rispetto ai parametri da inserire in input per consultare i reports.

Riassumendo, nonostante Google Data Studio sia un ottimo strumento per la realizzazione di reports di facile consultazione, in base alle differenze sopra descritte e le funzionalità che la dashboard da realizzare dovrà avere, la scelta è ricaduta su l'utilizzo di MySQL e PHP.

In particolare, è stata ritenuta fondamentale la possibilità tramite codice di elaborare e gestire insieme i dati presi da diverse tabelle del database prima di mostrarli graficamente all'utente.

Nella scelta ha influito anche la possibilità di avere tramite la mappa un focus su una singola città (la dashboard riguarderà il Comune di Milano).

Infine, dopo aver scelto la tecnologia più idonea per la costruzione dei reports, è necessario capire se le funzionalità lato client della dashboard richiedano l'utilizzo del framework AngularJS o sia sufficiente l'utilizzo del tradizionale codice Javascript. Quindi, dato che tali funzionalità, non hanno quelle caratteristiche che permetterebbero di usufruire dei vantaggi forniti da AngularJS, la scelta è ricaduta sull'utilizzo del tradizionale Javascript, anche perché la mole di codice da scrivere lato client non risulta essere elevata.

5. Framework Proposto

Questo capitolo, legando i concetti teorici legati all'obiettivo prefissato, visti nei capitoli precedenti, andrà a descrivere dettagliatamente la progettazione e realizzazione del DW di riferimento e successivamente si vedrà come verrà realizzata la dashboard per la visualizzazione e l'analisi delle informazioni contenute nella base dati, da mostrare agli utenti.

Il primo passo è quello di definire le tipologie di dati su cui costruire il DW e di conseguenza gli indicatori di interesse (KPI). I dati presi in considerazione sono di tre tipi e riguardano la città di Milano per il periodo che va dal 2015 al 2017:

- Misurazioni sulle concentrazioni di inquinanti nelle aree urbane
- Condizioni climatiche nelle aree urbane
- Dati relativi al traffico di veicoli

Nei paragrafi successivi, verranno descritti nel dettaglio le fonti dei dati sopra citati, verrà effettuata la progettazione concettuale e logica del DW basata su questi dati e verranno indicati i KPI scelti per presentare le informazioni di interesse agli utenti che andranno ad usare la dashboard e l'applicazione mobile.

Nella figura 5.1 è riportato in maniera grafica il framework del progetto.

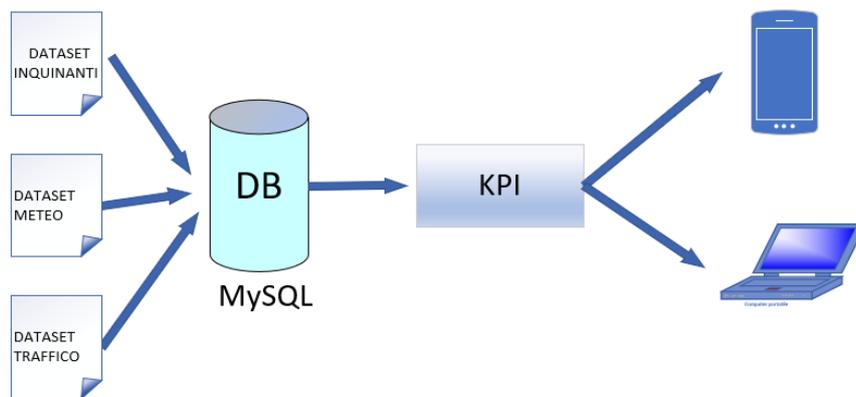


Figura 5.1: Framework

5.1 Fonti dei dati

Questo paragrafo presenta le tre fonti di dati considerate per il progetto, ognuna di esse fornisce una delle tipologie di dati prima citate. L'analisi è stata focalizzata sui dati della città di Milano.

1. *Misurazioni sulle concentrazioni di inquinanti:* questi dati sono estratti dal dataset di ARPA Lombardia e comprende i valori di concentrazione per un insieme di inquinanti (PM10, PM2.5, NO2, CO, O3, SO2, C6H6) raccolti attraverso alcune stazioni di monitoraggio situate nel Comune di Milano. Ogni stazione è dotata di un set di sensori, ognuno dei quali misura un determinato inquinante I dati forniti consistono in una serie di letture orarie o giornaliere, a seconda del tipo di inquinante (PM10 e PM2.5 solo

giornaliere). Ogni lettura è caratterizzata dall'identificatore della stazione di monitoraggio, dall'identificatore del sensore, dal nome dell'inquinante misurato, dalla concentrazione misurata dell'inquinante e dalla data e ora della lettura.

2. *Condizioni climatiche*: Il set di dati riguardanti le misurazioni metereologiche sono estratti da Weather Underground e sono raccolti dagli utenti attraverso tutti i registri di Personal Weather Station (PWS) più vicini alla città di Milano. Per ciascuna PWS selezionata, le misurazioni meteorologiche considerate (Temperatura, Pressione, Umidità, Precipitazioni, Velocità del vento) sono relative all'anno 2015. Ogni registrazione include l'identificatore della PWS, il timestamp della misurazione e per ogni misura meteorologica il corrispondente valore meteorologico misurato. I dati meteorologici raccolti sono stati poi aggregati attraverso diverse PWS per ottenere un valore rappresentativo alla granularità giornaliera associata a ciascuna stazione di monitoraggio degli inquinanti.
3. *Dati relativi al traffico di veicoli*: questi dati sono forniti dal comune di Milano sotto forma di informazioni aggregate sul numero di veicoli che entrano nella zona centrale di Milano giornalmente. Il set di dati contiene il numero totale giornaliero di veicoli in entrata e il numero giornaliero di veicoli separatamente per ciascuna categoria di veicolo. I veicoli sono classificati in base al tipo di carburante (Benzina, Diesel, Elettrico, GPL-Metano, Ibrido, GPL, Miscela, Ibrido Gasolio, Metano, Ibrido Benzina), al tipo di veicolo (Bus, Merci, Privati, Altro) e alla categoria di Euro. I dati sul traffico qui descritti sono relativi al periodo 2015-2016.

5.2 Progettazione del Dimensional Fact Model

I dati sulla qualità dell'aria, delle condizioni metereologiche e del traffico vengono quindi archiviati in un unico data warehouse, la cui rappresentazione concettuale, secondo il Dimensional Fact Model, è riportata in figura 5.2.

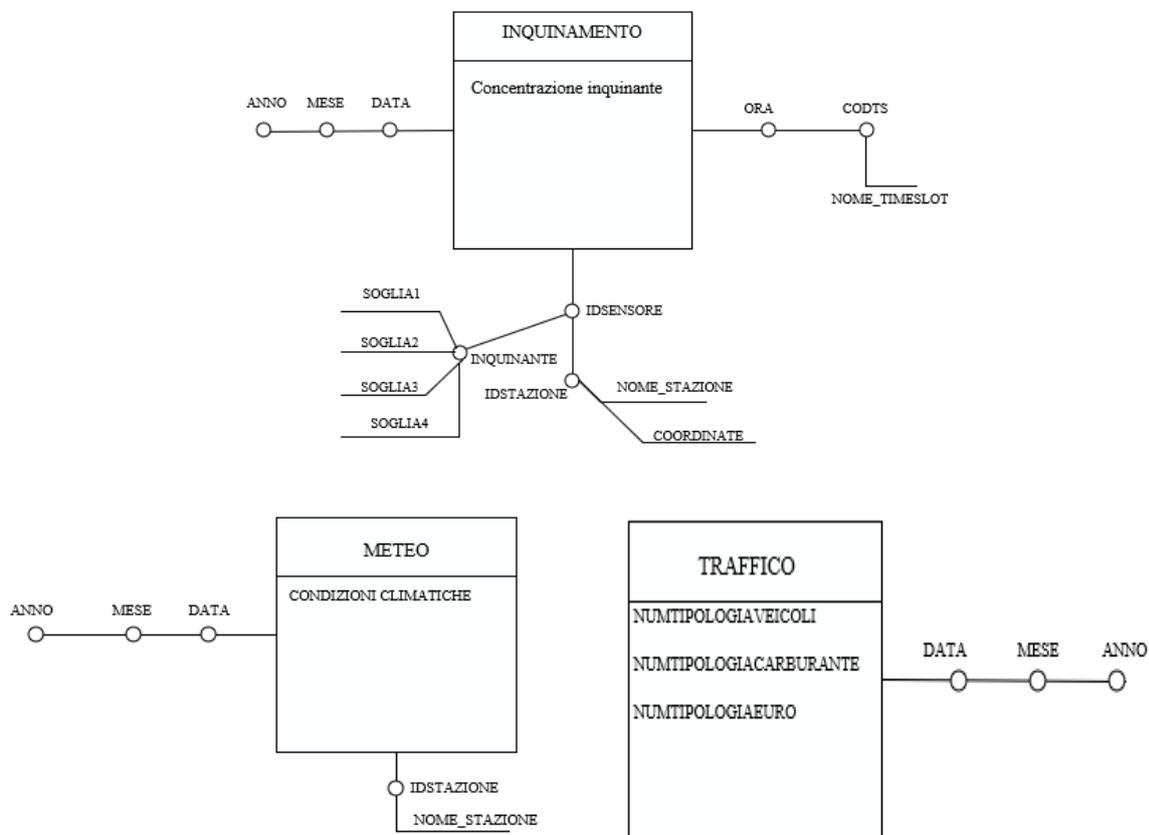


Figura 5.2: Dimensional Fact Model

Come si può notare, il DW progettato è composto da tre tabelle dei fatti, una riguardante l'inquinamento atmosferico, una riguardante le condizioni climatiche e una relativa al traffico.

Per quanto riguarda la tabella dei fatti relativa all'inquinamento, essa fornisce una misura sulla concentrazione dell'inquinante; inoltre, definisce tre gerarchie di dimensioni per analizzare la distribuzione spaziale e temporale della qualità dell'aria. Nello specifico, sono state definite una gerarchia delle dimensioni legate alla posizione collegata al dispositivo sensore che monitora la concentrazione di un determinato inquinante e due gerarchie di dimensioni temporali: una relativa alla data della misurazione e una relativa all'ora.

La tabella dei fatti corrispondente al meteo, ha come misura la condizione climatica e due gerarchie di dimensioni, una spaziale e una temporale. La prima indica la stazione in cui avviene il rilevamento climatico, la seconda indica la data in cui questa rilevazione viene effettuata.

Infine, la tabella dei fatti relativa al traffico urbano permette la misurazione del traffico da tre punti di vista: rispetto al tipo di veicolo, al tipo di carburante e rispetto alla categoria Euro. Questa tabella definisce una sola dimensione temporale con granularità giornaliera.

Successivamente al Dimensional Fact Model, viene mostrata la relativa Progettazione Logica, per avere maggiore facilità e intuitività per la costruzione effettiva delle tabelle che dovranno essere realizzate per lo sviluppo del progetto.

TEMPO-DATA (IdData, Data, Mese, Anno)

TEMPO-ORA (IdOra, Ora, TimeSlot, Nome_TS)

SENSORE (IdSensore, IdStazione, Nome_Stazione, Latitudine, Longitudine, Inquinante, Soglia1, Soglia2, Soglia3, Soglia4)

INQUINAMENTO (IdData, IdOra, IdSensore, Concentrazione_Inquinante)

TRAFFICO (IdData, NumTipoVeicolo, NumTipoAlimentazione, NumTipoEuro)

STAZIONE (IdStazione, Nome_Stazione)

METEO (IdStazione, IdData, Condizione_Climatica)

Le gerarchie relative alle dimensioni temporali forniscono diverse combinazioni di intervalli di tempo. In una gerarchia, il tempo si estende da una data all'altra, con l'indicazione del mese e dell'anno di appartenenza. Nell'altra gerarchia, il tempo si estende da ore a fasce orarie (mattina, pomeriggio, sera, notte).

La gerarchia delle dimensioni basata sulla posizione, prende come riferimento la posizione del sensore fisico usato nella misurazione e si estende fino all'indirizzo della stazione di appartenenza. In questa dimensione, ogni sensore misura uno specifico inquinante con le relative soglie di pericolosità.

Per i dati riguardanti le misurazioni delle condizioni climatiche, la gerarchia di dimensione basata sulla posizione ha una granularità diversa dalla precedente, in particolare prende come riferimento non più il sensore fisico ma la stazione di appartenenza del sensore che ha effettuato la rilevazione.

In realtà, quanto appena descritto, differisce dall'effettiva realizzazione delle tabelle utilizzate per il progetto. Per alcune dimensioni non sono state create delle tabelle dedicate ma si è preferito integrarle con quelle relative ai fatti, per rendere più semplice e veloce l'interrogazione della base dati. Quanto appena detto, si vedrà in maniera approfondita nel capitolo 5.4.

5.3 Progettazione dei Key Performance Indicator

I dati che compongono il Data Warehouse vengono analizzati per ottenere informazioni sulla condizione di inquinamento dell'aria nell'area urbana di Milano. Lo scopo dell'analisi è di produrre feedback utili agli utenti finali utilizzando Key Performance Indicator(KPI). Nel contesto del progetto sviluppato, i KPI sono indicatori quantitativi della condizione di inquinamento atmosferico monitorata. Inoltre, vengono utilizzati ulteriori KPI per quantificare le condizioni climatiche e di traffico dei veicoli. I KPI individuati sono:

- A. *KPI di concentrazione di inquinanti spazio-temporali*: segnalano la concentrazione di inquinanti per stazione, a diversi livelli di granularità nel tempo, in una data specificata dall'utente:
 - A1. Concentrazione media giornaliera degli inquinanti per stazione
 - A2. Concentrazione media degli inquinanti per fascia oraria giornaliera per stazione
 - A3. Concentrazione media oraria degli inquinanti per stazione
 - A4. Concentrazione media mensile degli inquinanti per stazione
- B. *KPI sul clima*: segnalano i valori medi climatici per stazione, a diversi livelli di granularità nel tempo, in un intervallo temporale specificato dall'utente
 - B1. Valori medi giornalieri per i diversi fenomeni metereologici
 - B2. Valori medi mensili per i diversi fenomeni metereologici
- C. *KPI di traffico veicolare*: riportano il numero totale di veicoli giornalieri suddivisi per diverse categorie, entro un periodo temporale specificato dall'utente, nell'area del centro di Milano.
 - C1. Valori totali giornalieri per tipologia di veicolo
 - C2. Valori totali giornalieri per tipologia di carburante
 - C3. Valori totali giornalieri per categoria Euro
- D. *KPI sull'analisi statistica di Correlazione*: le concentrazioni di inquinanti possono essere influenzate dalle condizioni meteorologiche. Per approfondire la valutazione dell'impatto delle condizioni climatiche sull'inquinamento atmosferico, è stata analizzata la correlazione statistica tra concentrazione di inquinanti e dati climatici. Il coefficiente di correlazione di Pearson stabilito è stato attualmente adottato per il calcolo del KPI sulla sequenza di misurazioni raccolte in un dato intervallo di tempo (giornalieri o mensili). Il coefficiente di correlazione di Pearson assume valori nell'intervallo [-1, +1]. Valori superiori a 0 rappresentano correlazioni positive, mentre valori inferiori a 0 rappresentano correlazioni negative. Più alto è il valore del coefficiente assoluto più forte è la correlazione. Il calcolo del coefficiente di correlazione di Pearson avviene come mostrato nella formula 5.3.1.

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \quad (5.3.1)$$

Dove il numeratore indica la Covarianza tra le due serie numeriche (formula 5.3.2) e il denominatore rappresenta il prodotto delle due Deviazioni Standard (formula 5.3.3).

$$\sigma_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n} \quad (5.3.2)$$

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} \quad (5.3.3)$$

5.4 Preparazione dei dati e creazione del Data Warehouse

In questa fase sono stati utilizzati gli strumenti ETL (Extraction, Transformation, Loading), utili per alimentare una sorgente dati singola, dettagliata, esauriente e di alta qualità che possa a sua volta alimentare il DW. Durante il processo di alimentazione del DW, la riconciliazione avviene in due occasioni: quando il DW viene popolato per la prima volta e periodicamente quando il DW viene aggiornato. Questa fase di riconciliazione consiste in quattro processi distinti:

- *Estrazione*: consiste nell'estrarre i dati rilevanti dalle sorgenti
- *Pulitura*: consiste nel migliorare la qualità dei dati (normalmente piuttosto scarsa nelle sorgenti)
- *Trasformazione*: consiste nella conversione dei dati dal formato operativo sorgente a quello del DW
- *Caricamento*: consiste nel caricamento dei dati puliti e trasformati nel DW

Nella realizzazione di questo progetto, la fase di Trasformazione dei dati viene effettuata sia prima che dopo la fase di Caricamento, per una maggiore praticità nella gestione e conversione dei dati nel formato desiderato.

Si descrivono ora nel dettaglio, per ogni differente fonte dati descritta nei paragrafi precedenti (ARPA Lombardia, Wheater Underground e il portale del Comune di Milano), l'applicazione degli strumenti ETL.

5.4.1 ARPA Lombardia

Per quanto riguarda i dati estratti da ARPA Lombardia, è stato necessario l'utilizzo del relativo sito per effettuare le richieste per lo scaricamento delle informazioni riguardanti i livelli di inquinamento. Tali richieste sono state fatte singolarmente per ogni stazione di monitoraggio presente nella città di interesse indicata (Milano) e per il periodo temporale indicato (2015-2017); le risposte di ARPA Lombardia alle richieste ricevute, sono state fornite tramite email in formato csv, suddividendo le misurazioni effettuate dalla stazione di monitoraggio in base ai sensori che la compongono (ogni sensore misura un differente inquinante).

Per facilitare la comprensione della struttura dei dati forniti da ARPA Lombardia, viene ora analizzata la risposta alla richiesta delle misurazioni dell'anno 2016 relative alla stazione di monitoraggio sita in Viale Marche.

Nella figura 4.3 è mostrata una parte del file csv ricevuto da ARPA Lombardia contenente le rilevazioni.

	A	B	C
1	5504,2016/01/01 01:00,	61.0	
2	5504,2016/01/01 02:00,	60.4	
3	5504,2016/01/01 03:00,	64.4	
4	5504,2016/01/01 04:00,	63.9	
5	5504,2016/01/01 05:00,	62.4	
6	5504,2016/01/01 06:00,	61.7	

Figura 5.3: file csv ARPA Lombardia

Come si può notare, sono presenti tre informazioni: le prime due indicano il codice del sensore che ha effettuato la rilevazione e quando essa è stata effettuata (data e ora), invece il terzo valore indica l'effettiva misurazione della concentrazione dell'inquinante. In allegato a tali rilevazioni, ARPA Lombardia fornisce una legenda composta come in figura 5.4.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Legenda											
2	-999 dato non disponibile											
3												
4	Riepilogo estrazione											
5	Id_Stazione, Nome_Stazione, Id_Sensore, Nome_Sensore, Unita_Misura, Periodo_Dal, Periodo_Al, Operatore											
6	501, Milano - viale Marche, 5504, Milano - viale Marche Biossido di Azoto, $\mu\text{g}/\text{m}^3$, 2016/01/01 01:00, 2017/01/01 00:00, Media											
7	501, Milano - viale Marche, 5827, Milano - viale Marche Monossido di Carbonio, mg/m^3 , 2016/01/01 01:00, 2017/01/01 00:00, Media											
8	501, Milano - viale Marche, 6328, Milano - viale Marche Ossidi di Azoto, $\mu\text{g}/\text{m}^3$, 2016/01/01 01:00, 2017/01/01 00:00, Media											
9	501, Milano - viale Marche, 17127, Milano - viale Marche Benzene, $\mu\text{g}/\text{m}^3$, 2016/01/01 01:00, 2017/01/01 00:00, Media											
10												

Figura 5.4: legenda dati ARPA Lombardia

Da tale legenda è possibile ricavare l'identificativo e il nome della stazione, l'identificativo e i nomi dei sensori ad essa associati, l'unità di misura in cui sono effettuate le misurazioni, il periodo temporale di riferimento e la tipologia di misurazione effettuata. Inoltre, tale legenda indica che valori di misurazioni pari a -999 corrispondono all'assenza del dato stesso.

Per quanto riguarda il file mostrato in figura 5.3, contenente le misurazioni, le tre tipologie di informazioni sono state suddivise in altrettante colonne; in questo modo, è stato possibile lavorare singolarmente sulla colonna contenente le informazioni temporali, suddividendola ulteriormente in due colonne distinte: una contenente la data e l'altra contenente l'ora della misurazione. Con la singola Data, è stato possibile creare ulteriori due colonne contenenti rispettivamente le informazioni sul mese e l'anno della misurazione; queste informazioni saranno utili quando si vorranno estrarre i KPI con una granularità diversa da quella giornaliera.

Dopo aver effettuato tali trasformazioni, nella prima riga del file sono stati inseriti i nomi di quelli che saranno gli attributi della tabella che verrà successivamente usata nel DW.

In figura 5.5 è riportata la stessa parte del file precedente, aggiornata e pronta per essere caricata.

	A	B	C	D	E
1	Idsensore,Data,Ora,Valore,Mese,Anno				
2	5504,01/01/2016,01:00:00,61.0,01-2016,2016				
3	5504,01/01/2016,02:00:00,60.4,01-2016,2016				
4	5504,01/01/2016,03:00:00,64.4,01-2016,2016				
5	5504,01/01/2016,04:00:00,63.9,01-2016,2016				
6	5504,01/01/2016,05:00:00,62.4,01-2016,2016				
7	5504,01/01/2016,06:00:00,61.7,01-2016,2016				

Figura 5.5: file csv ARPA Lombardia da caricare

Raccogliendo le legende delle stazioni di monitoraggio, sono state applicate delle trasformazioni sui dati con l'obiettivo di creare un nuovo file contenente informazioni sulle diverse stazioni di monitoraggio di Milano, i loro rispettivi sensori e gli inquinanti da quest'ultimi misurati. Tale file sarà caricato anch'esso nel DW e utilizzato per creare tramite l'identificativo del sensore una corrispondenza tra la stazione, l'inquinante e la misurazione effettuata. In figura 5.6 è mostrato una parte del file appena descritto.

	A	B	C	D	E
1	Id_Stazione,Nome_Stazione,Id_Sensore,Inquinante				
2	501,Milano - viale Marche,5504,NO2				
3	501,Milano - viale Marche,5827,CO				
4	501,Milano - viale Marche,6328,NO				
5	501,Milano - viale Marche,17127,C6H6				
6	549,Milano - P.zza Abbiategrasso,5552,NO2				
7	549,Milano - P.zza Abbiategrasso,6344,NO				

Figura 5.6: legenda ARPA Lombardia da caricare

Come si può vedere da un facile confronto, rispetto alla struttura delle legende ricevute da ARPA Lombardia, alcune informazioni sono state rimosse perché ritenute inutili. In particolare, sono stati eliminati i dati relativi al periodo temporale (perché già contenuto nel file riguardante le misurazioni), all'unità di misura (perché tutti gli inquinanti sono misurati in $\mu\text{g}/\text{m}^3$ ad eccezione del CO misurato in mg/m^3) e alla tipologia di misurazione effettuata (perché sempre corrispondente alla Media). Inoltre, per una maggiore facilità di lettura, si è preferito sostituire al nome del sensore (contenente il nome dell'inquinante misurato) la formula chimica dell'inquinante.

A questo punto, sia il file delle misurazioni, sia quello riguardante le stazioni sono pronti per essere caricati nel DW in formato csv.

Per quanto riguarda il primo file, ossia quello delle misurazioni, esso è stato caricato e convertito in una tabella del database con il nome "inquinanti". Al momento della creazione della tabella, alcuni campi riguardanti le informazioni temporali (Data e Mese) hanno subito una conversione di formato diversa da quella desiderata (varchar anziché date).

Di conseguenza, è stato necessario convertire questi due attributi nel formato corretto e per fare questo sono stati seguiti i seguenti passaggi:

- Aggiunta di due colonne temporanee ('DataTemp' e 'MeseTemp') con formato date

- Esecuzione delle istruzioni sql:

```
UPDATE 'inquinanti' SET 'DataTemp'=(SELECT STR_TO_DATE('Data',
%d/%m/%Y));
```

```
UPDATE 'inquinanti' SET 'MeseTemp'=(SELECT STR_TO_DATE('Mese',
%m/%Y));
```

- Cancellazione delle colonne 'Data' e 'Mese'
- Aggiornamento del nome delle colonne 'DataTemp' e 'MeseTemp' rispettivamente in 'Data' e 'Mese'

Tali conversioni sono state necessarie per permettere la futura estrazione dei KPI descritti nei paragrafi precedenti. Infatti, senza le informazioni temporali nel formato date non sarebbe stato possibile consultare il DW per avere informazioni con granularità giornaliera o mensile (come indicato nei KPI). Inoltre, sempre per la possibilità di estrarre i KPI prima descritti, è stato ritenuto utile inserire due ulteriori colonne nella tabella "inquinanti", con l'obiettivo di raggruppare i dati in fasce orarie. In particolare, sono stati aggiunti i campi 'CodTS' e 'TimeSlot': il primo è un valore da 1 a 4 che identifica la fascia oraria di appartenenza della misurazione, il secondo è una descrizione letterale della fascia oraria stessa. Usando le seguenti istruzioni sql è stato possibile assegnare i valori corrispondenti a tali campi:

- **UPDATE** 'inquinanti' **SET** 'CodTS'=1, 'TimeSlot'=Morning (0-11) **WHERE** 'Ora' >=00:00:00 **AND** 'Ora' <=11:00:00;
- **UPDATE** 'inquinanti' **SET** 'CodTS'=2, 'TimeSlot'=Afternoon (12-16) **WHERE** 'Ora' >=12:00:00 **AND** 'Ora' <=16:00:00;
- **UPDATE** 'inquinanti' **SET** 'CodTS'=3, 'TimeSlot'=Evening (17-20) **WHERE** 'Ora' >=17:00:00 **AND** 'Ora' <=20:00:00
- **UPDATE** 'inquinanti' **SET** 'CodTS'=4, 'TimeSlot'=Night (21-23) **WHERE** 'Ora' >=21:00:00 **AND** 'Ora' <=23:00:00

In seguito a queste trasformazioni degli attributi della tabella "inquinanti", sono state effettuate delle operazioni su alcuni dati presenti nella tabella. In particolare, nella colonna 'Valore', in cui sono presenti le misurazioni effettuate, sono presenti dei valori pari a -999.0 che indicano (come detto prima) l'assenza della misurazione. Per evitare che tale valore possa influire in qualche modo nelle analisi successive basate sui KPI da estrarre, si è deciso di trasformare tramite l'operazione sql di UPDATE, il valore -999.0 in NULL.

Nella figura 5.7 è riportata la struttura della tabella "inquinanti", mentre nella figura 5.8 è rappresentato un esempio di come essa è stata popolata.

#	Nome	Tipo	Codifica caratteri	Attributi	Null	Predefinito
<input type="checkbox"/>	1	Idsensore	int(5)		Sì	NULL
<input type="checkbox"/>	2	Data	date		No	Nessuno
<input type="checkbox"/>	3	Mese	date		No	Nessuno
<input type="checkbox"/>	4	Ora	varchar(8)	utf8_general_ci	Sì	NULL
<input type="checkbox"/>	5	CodTS	int(11)		No	Nessuno
<input type="checkbox"/>	6	TimeSlot	varchar(50)	utf8_general_ci	No	Nessuno
<input type="checkbox"/>	7	Valore	decimal(6,1)		Sì	NULL
<input type="checkbox"/>	8	Anno	int(4)		Sì	NULL

Figura 5.7: Struttura tabella “inquinanti”

Idsensore	Data	Mese	Ora	CodTS	TimeSlot	Valore	Anno
5722	2017-10-01	2017-10-00	00:00:00	1	Morning (0-11)	26.3	2017
5550	2017-10-01	2017-10-00	00:00:00	1	Morning (0-11)	36.0	2017
5506	2017-10-01	2017-10-00	00:00:00	1	Morning (0-11)	31.6	2017
10280	2017-10-01	2017-10-00	00:00:00	1	Morning (0-11)	0.9	2017
10282	2017-10-01	2017-10-00	00:00:00	1	Morning (0-11)	36.8	2017
20004	2017-10-01	2017-10-00	00:00:00	1	Morning (0-11)	2.7	2017
10279	2017-10-01	2017-10-00	00:00:00	1	Morning (0-11)	45.8	2017
6372	2017-10-01	2017-10-00	00:00:00	1	Morning (0-11)	39.3	2017
5841	2017-10-01	2017-10-00	00:00:00	1	Morning (0-11)	0.6	2017

Figura 5.8: Popolazione tabella “inquinanti”

Prima di concludere la descrizione di tale tabella è importante sottolineare un aspetto concettuale riguardante i sensori relativi agli inquinanti PM10 e PM2.5. A differenza degli altri sensori, questi effettuano rilevazioni con frequenza giornaliera e non oraria; di conseguenza, i campi ‘Ora’, ‘CodTS’ e ‘TimeSlot’ avranno rispettivamente valori pari a NULL, 0 e stringa vuota. Per cui le analisi sui KPI basati sui dati raggruppati per fasce orarie, non prenderanno in considerazione i valori delle misurazioni relative al PM10 e al PM2.5.

L’altro file ancora da analizzare è quello prima descritto riguardante le stazioni di monitoraggio con i relativi sensori. A differenza del caso precedente, dopo il caricamento del file non risultano necessarie ulteriori trasformazioni sui campi della tabella creata (“stazioni”).

Vengono riportate rispettivamente in figura 5.9 e in figura 5.10, la struttura ed una parte della popolazione di tale tabella.

#	Nome	Tipo	Codifica caratteri	Attributi	Null	Predefinito
<input type="checkbox"/>	1	Id_Stazione	int(3)		Sì	NULL
<input type="checkbox"/>	2	Nome_Stazione	varchar(28)	utf8_general_ci	Sì	NULL
<input type="checkbox"/>	3	Id_Sensore	int(5)		Sì	NULL
<input type="checkbox"/>	4	Inquinante	varchar(5)	utf8_general_ci	Sì	NULL

Figura 5.9: Struttura tabella “stazioni”

Id_Stazione	Nome_Stazione	Id_Sensore	Inquinante
501	Milano - viale Marche	5504	NO2
501	Milano - viale Marche	5827	CO
501	Milano - viale Marche	6328	NO
501	Milano - viale Marche	17127	C6H6
549	Milano - P.zza Abbiategrasso	5552	NO2
549	Milano - P.zza Abbiategrasso	6344	NO
547	Milano - Parco Lambro	5550	NO2
547	Milano - Parco Lambro	5722	O3

Figura 5.10: Popolazione tabella “stazioni”

Successivamente alla creazione di queste due tabelle, si è ritenuta necessaria l’implementazione di ulteriori due tabelle che fornissero informazioni aggiuntive riguardo alle stazioni di monitoraggio e riguardo agli inquinanti misurati. In particolare, per le stazioni di monitoraggio è stata creata la tabella “latlngstazioni” che per ogni stazione indica le relative coordinate geografiche (latitudine e longitudine); questo perché nello sviluppo della dashboard e dell’applicazione mobile facente parte di questo progetto, risultano necessarie le informazioni sulla posizione delle stazioni di monitoraggio.

Questa tabella è legata a “stazioni” tramite l’identificativo della stazione; nelle figure 5.11 e 5.12 sono mostrate la struttura e i dati contenuti.

#	Nome	Tipo	Codifica caratteri	Attributi	Null
<input type="checkbox"/> 1	Id_Stazione	int(11)			No
<input type="checkbox"/> 2	Lat	double			No
<input type="checkbox"/> 3	Lng	double			No
<input type="checkbox"/> 4	Nome_Stazione	varchar(100)	latin1_swedish_ci		No

Figura 5.11: Struttura tabella “latlngstazioni”

Id_Stazione	Lat	Lng	Nome_Stazione
501	45.4954866	9.1921222	Viale Marche
549	45.4304475	9.1741345	P.zza Abbiategrasso
547	45.4957003	9.2465629	Parco Lambro
539	45.4437788	9.1695722	Viale Liguria
705	45.4785456	9.2326961	Via Pascal
548	45.4704374	9.1952129	Via Senato
528	45.4631078	9.1938893	Via Verziere
503	45.4759561	9.1408294	P.zza Zavattari

Figura 5.12: Popolazione tabella “latlngstazioni”

La seconda tabella che è risultato utile creare è stata “soglie”, essa identifica per ogni inquinante le soglie di pericolosità da prendere come riferimento. Questa tabella è stata necessaria per fornire agli utenti della dashboard un indicatore cromatico differente a seconda della soglia in cui ricade il valore dell’inquinante. Essa è legata con la tabella “stazioni” tramite il nome univoco dell’inquinante.

Nelle figure 5.13 e 5.14 sono mostrate rispettivamente la struttura e i dati in essa contenuti.

#	Nome	Tipo	Codifica caratteri	Attributi	Null	Predefinito
<input type="checkbox"/> 1	Nome_Inquinante	varchar(5)	utf8_general_ci		Si	NULL
<input type="checkbox"/> 2	Soglia1	int(3)			Si	NULL
<input type="checkbox"/> 3	Soglia2	int(3)			Si	NULL
<input type="checkbox"/> 4	Soglia3	int(3)			Si	NULL
<input type="checkbox"/> 5	Soglia4	int(3)			Si	NULL

Figura 5.13: Struttura tabella “soglie”

Nome_Inquinante	Soglia1	Soglia2	Soglia3	Soglia4
PM10	25	50	75	100
PM2.5	20	40	60	80
NO2	100	200	300	400
O3	90	180	240	300
CO	5	10	15	20
SO2	62	125	187	250
C6H6	4	8	12	16

Figura 5.14: Popolazione tabella “soglie”

5.4.2 Portale della città di Milano

Per quanto riguarda i dati del traffico, tramite il portale del Comune di Milano è stato possibile scaricare le informazioni relative agli ingressi dei veicoli nell’Area C di Milano (corrispondente al centro cittadino). Questi dati riguardano gli ultimi due anni più recenti (2015 e 2016) e forniscono i valori giornalieri degli accessi in quest’area.

Al momento dello scaricamento dei dati, dal portale sarà possibile specificare la classificazione per cui i valori devono essere suddivisi; in particolare, per le necessità del progetto, quelle ritenute utili sono: per categoria veicolo, per tipo di carburante e per categoria Euro. In figura 5.15 sono mostrati una parte dei file (uno per ogni classificazione) in formato csv, contenenti i dati appena descritti.

	A	B	C	D
1	timeIndex,NC,ALTRO,BUS,MERCI,PERSONE			
2	19/08/2016,2155,235,506,1712,13667			
3	20/08/2016,2423,164,513,816,12923			
4	21/08/2016,2027,79,460,485,14661			
5	22/08/2016,2505,332,574,1963,13328			
6	23/08/2016,2640,419,517,2514,16355			
7	24/08/2016,2858,359,452,2176,14026			

	A	B	C	D	E	F	G
1	timeIndex,EURO0,EURO1,EURO2,EURO3,EURO4,EURO5,EURO6,NC						
2	19/08/2016,18,41,438,1157,5247,6713,2452,2209						
3	20/08/2016,34,57,365,1354,4109,5751,2739,2430						
4	21/08/2016,26,41,369,1073,4735,6902,2531,2035						
5	22/08/2016,24,13,392,1117,4540,7289,2756,2571						
6	23/08/2016,18,24,334,980,5325,10337,2755,2672						
7	24/08/2016,27,23,277,1024,4818,8030,2746,2926						

	A	B	C	D	E	F	G	H	I	J	K	L
1	timeIndex,BENZINA,DIESEL,ELETTRICO,GPL_METANO,IBRIDO,NC,GPL,MISCELA,IBRIDO GASOLIO,METANO,IBRIDO BENZINA											
2	19/08/2016,5640,7718,56,,,2159,1115,,6,582,999											
3	20/08/2016,5358,7024,8,,,2428,944,,17,278,782											
4	21/08/2016,5600,7725,10,,,2029,1050,,7,237,1054											
5	22/08/2016,4752,8114,74,,,2509,1395,,20,594,1244											
6	23/08/2016,5248,10767,41,,,2641,1691,,10,667,1380											
7	24/08/2016,5572,8189,70,,,2859,1460,,19,551,1151											
8	25/08/2016,5524,9863,53,,,2026,1643,,18,790,2374											

Figura 5.15: file csv Portale città di Milano

Come si può notare, i tre file hanno una struttura molto simile, infatti tutti e tre hanno una informazione temporale indicante la data della rilevazione e la numerosità dei veicoli entranti nell'area C, suddivisi per categoria a seconda della classificazione scelta.

Tutti questi file csv sono stati suddivisi per colonne, per distinguere al meglio i valori di ogni categoria e soprattutto per aggiungere ulteriori due colonne, contenenti informazioni temporali riguardo al mese e all'anno; informazioni che risulteranno utili in seguito, durante le analisi, per l'estrazione di alcuni KPI. In figura 5.16 sono riportate le stesse parti di file precedenti aggiornati e pronti per essere caricati.

	A	B	C	D	E
1	Data,NC,ALTRO,BUS,MERCI,PERSONE,Mese,Anno				
2	01/01/2015,9312,350,675,1352,63652,01-2015,2015				
3	02/01/2015,10793,1085,891,6520,69847,01-2015,2015				
4	03/01/2015,12293,890,729,3967,81270,01-2015,2015				
5	04/01/2015,12532,358,709,1911,78579,01-2015,2015				
6	05/01/2015,13239,1384,852,7981,87271,01-2015,2015				
7	06/01/2015,12381,520,780,2540,87908,01-2015,2015				

	A	B	C	D	E	F	G
1	Data,EURO0,EURO1,EURO2,EURO3,EURO4,EURO5,EURO6,NC,Mese,Anno						
2	01/01/2015,67,202,2691,7516,26085,28086,1377,9317,01-2015,2015						
3	02/01/2015,71,186,2598,7596,30179,35784,1884,10838,01-2015,2015						
4	03/01/2015,93,207,2939,8911,33425,39125,2127,12322,01-2015,2015						
5	04/01/2015,73,162,2680,8261,31452,36857,2047,12557,01-2015,2015						
6	05/01/2015,65,147,2748,8261,36255,47320,2639,13292,01-2015,2015						
7	06/01/2015,92,181,2830,8504,34252,43288,2583,12399,01-2015,2015						

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Data,BENZINA,DIESEL,ELETTRICO,GPL_METANO,IBRIDO,NC,GPL,MISCELA,IBRIDO GASOLIO,METANO,IBRIDO BENZINA,Mese,Anno												
2	01/01/2015,27461,29968,12,5515,3073,9312,0,0,0,0,0,0,01-2015,2015												
3	02/01/2015,29222,38132,47,7622,3320,10793,0,0,0,0,0,0,01-2015,2015												
4	03/01/2015,35091,41227,42,7216,3280,12293,0,0,0,0,0,0,01-2015,2015												
5	04/01/2015,33605,38512,29,6346,3066,12531,0,0,0,0,0,0,01-2015,2015												
6	05/01/2015,35797,47802,58,9269,4562,13239,0,0,0,0,0,0,01-2015,2015												
7	06/01/2015,36840,43655,28,7213,4013,12380,0,0,0,0,0,0,01-2015,2015												

Figura 5.16: file csv Portale città di Milano da caricare

A questo punto, i tre file così composti sono pronti per essere caricati nel DW in formato csv; alla conclusione del loro caricamento verranno create tre tabelle così definite. “tipalimentazione”, “catveicolo” e “euro”. Anche in questo caso, come in quello della tabella “inquinanti”, al momento della creazione, alcuni campi riguardanti le informazioni temporali (Data e Mese) subiscono una conversione di formato diversa da quella desiderata (varchar anziché date). Di conseguenza, è stato necessario convertire questi due attributi nel formato corretto nelle stesse modalità e medesimi passaggi descritti in precedenza. Dopo tali conversioni sarà possibile estrarre in una futura analisi alcuni KPI descritti nei paragrafi precedenti. Infatti, senza le informazioni temporali nel formato date non sarebbe stato possibile consultare il DW per avere informazioni con granularità giornaliera o mensile (come

indicato nei KPI). Vengono ora riportate, rispettivamente nelle figure 5.17 e 5.18 la struttura di una delle tre tabelle ed un esempio di come essa è stata popolata (non vengono riportate le altre perché molto simili a quella mostrata).

#	Nome	Tipo	Codifica caratteri	Attributi	Null	Predefinito
<input type="checkbox"/>	1	Data	date		No	Nessuno
<input type="checkbox"/>	2	NC	int(5)		Sì	NULL
<input type="checkbox"/>	3	ALTRO	int(4)		Sì	NULL
<input type="checkbox"/>	4	BUS	int(4)		Sì	NULL
<input type="checkbox"/>	5	MERCI	int(5)		Sì	NULL
<input type="checkbox"/>	6	PERSONE	int(6)		Sì	NULL
<input type="checkbox"/>	7	Mese	date		No	Nessuno
<input type="checkbox"/>	8	Anno	int(4)		Sì	NULL

Figura 5.17: Struttura tabella “catveicolo”

Data	NC	ALTRO	BUS	MERCI	PERSONE	Mese	Anno
2016-01-07	1133	154	171	655	5057	2016-01-00	2016
2016-01-28	1043	182	212	1051	7829	2016-01-00	2016
2016-08-15	1283	71	384	365	10316	2016-08-00	2016
2016-08-17	1698	271	392	1804	11184	2016-08-00	2016
2016-08-14	2773	162	380	367	11515	2016-08-00	2016
2015-10-24	3941	119	154	509	11645	2015-10-00	2015
2016-08-16	2225	246	315	1566	12251	2016-08-00	2016
2016-08-18	2046	262	490	1601	12857	2016-08-00	2016
2016-08-13	2246	172	426	844	12891	2016-08-00	2016

Figura 5.18: Popolazione tabella “catveicolo”

5.4.3 Wheater Underground

Per quanto riguarda le informazioni metereologiche, queste sono state reperite tramite il portale Wheater Underground. Questo sito si occupa di fornire previsioni dettagliate del meteo di qualsiasi città a livello mondiale e mettendo a disposizione degli utenti la possibilità di scaricare dati o report sulle condizioni climatiche di una determinata città in uno specifico periodo temporale. In particolare, questa ultima funzione è risultata utile per avere a disposizione i dati del meteo della città di Milano dell’anno 2015. Il file scaricato contenente queste informazioni si presenta come nella figura 5.19:

1	ID,"Date","HOUR","HH","SENSOR_ID","Area","MEASUREMENT","temperature","humidity","precipitations","windspeed","pressure","dew_point","wind_gust","wind_direction","wind_chill","heat_index","precipitations rat
2	1,"2015/01/01","00:00","0","5531","1","55.2","-0.534185666432","83.4446305473","0","7.01376328808","1028.82414379","-3.36597352739","8.81934483581","227.773409674","-3.24858288986","-9999","0","0","0","Milai
3	2,"2015/01/01","00:00","0","6366","1","76.3","-0.534185666432","83.4446305473","0","7.01376328808","1028.82414379","-3.36597352739","8.81934483581","227.773409674","-3.24858288986","-9999","0","0","0","Milai
4	3,"2015/01/01","00:00","0","5725","1","15.3","-0.534185666432","83.4446305473","0","7.01376328808","1028.82414379","-3.36597352739","8.81934483581","227.773409674","-3.24858288986","-9999","0","0","0","Milai
5	4,"2015/01/01","00:00","0","5551","1","54.2","-0.534185666432","83.4446305473","0","7.01376328808","1028.82414379","-3.36597352739","8.81934483581","227.773409674","-3.24858288986","-9999","0","0","0","Milai
6	5,"2015/01/01","00:00","0","5834","1","0.9","-0.534185666432","83.4446305473","0","7.01376328808","1028.82414379","-3.36597352739","8.81934483581","227.773409674","-3.24858288986","-9999","0","0","0","Milai

Figura 5.19: file csv Wheater Underground

Le informazioni contenute sono risultate essere maggiori rispetto a quelle necessarie per le analisi da effettuare in questo progetto; per questo dopo aver suddiviso i dati per colonna, si è deciso di prendere in considerazione solo le condizioni climatiche ritenute correlabili con i livelli di inquinamento (temperatura, umidità, velocità del vento, pressione, precipitazioni).

La granularità dei dati a livello temporale è oraria, a livello spaziale riguarda i singoli sensori di ogni stazione presente nella città di Milano. Dopo aver osservato i valori meteorologici misurati da ogni sensore della stessa stazione di monitoraggio, si è notato che questi fossero pressoché identici tra loro e di conseguenza si è deciso di portare la granularità ad un livello minore di dettaglio, cioè andando a considerare per ogni stazione nella medesima ora del medesimo giorno, solo un singolo valore rappresentativo dell'intera stazione di monitoraggio; andando quindi a rimuovere tutte le informazioni relative ai sensori.

Quindi riassumendo, le informazioni ritenute importanti da inserire nel DW del progetto, sono la data, l'ora, le condizioni climatiche prima citate e il nome della stazione che ha effettuato le rilevazioni. Inoltre, come visto anche in precedenza, sono state aggiunte altre due colonne: Mese e Anno, sempre con l'obiettivo di avere la possibilità di effettuare analisi con granularità differenti. Il file csv pronto per il caricamento nel DW è mostrato in figura 5.20.

	A	B	C	D	E	F	G	H	I	J	K
1	Date,Mese,Anno,Ora,temperature,humidity,wind_speed,pressure,precipitation_rate,Nome_Stazione										
2	01/01/2015,01-2015,2015,00:00,1.265838858,79.26020396,3.830803563,1030.76711,0,Milano - via Verziere										
3	01/01/2015,01-2015,2015,00:00,1.243883747,79.30366993,3.929370982,1030.369198,0,Milano - via Senato										
4	01/01/2015,01-2015,2015,00:00,1.200327454,78.85514945,4.694661182,1027.008018,0,Milano - Parco Lambro										
5	01/01/2015,01-2015,2015,00:00,1.226741512,78.86289216,4.512110389,1027.775475,0,Milano -via Carlo Pascal										

Figura 5.20: file csv Wheater Underground da caricare

Dopo essere stato caricato sono state necessarie altre operazioni per rendere questi dati utilizzabili. Per prima cosa, anche in questo caso, come in quello delle tabelle precedentemente descritte, al momento della creazione alcuni campi riguardanti le informazioni temporali (Data e Mese) subiscono una conversione di formato diversa da quella desiderata (varchar anziché date). Di conseguenza, è stato necessario convertire questi due attributi nel formato corretto nelle stesse modalità e medesimi passaggi descritti in precedenza.

Dopodiché, si è notato come alcuni valori delle misurazioni dei fenomeni meteorologici fossero concettualmente errati: sono stati rilevati valori al di fuori dei range possibili per questi fenomeni (es. temperatura -300°C); in questi casi, si è provveduto alla sostituzione di tali valori con il valore di default NULL, per evitare che questi influissero sul valore dei KPI che saranno estratti tramite questa tabella.

Infine, dal punto di vista temporale, come già detto in precedenza nel paragrafo relativo al DFM, l'interesse era quello di avere dati meteorologici con granularità giornaliera; quindi, in virtù del fatto che le informazioni fornite dal portale hanno una granularità maggiore (oraria), la scelta è stata quella di effettuare una media giornaliera per ogni stazione per ogni condizione meteorologica misurata; questo è stato possibile tramite un'interrogazione sql. In base al risultato è stata aggiornata l'intera tabella "meteo", passando così da una granularità oraria ad una giornaliera. Nelle figure 5.21 e 5.22 sono riportate rispettivamente la struttura di tale tabella e una parte della popolazione di essa, successive alle operazioni prima descritte.

#	Nome	Tipo	Codifica caratteri	Attributi	Null	Predefinito
<input type="checkbox"/>	1	Data	date		No	Nessuno
<input type="checkbox"/>	2	Mese	date		No	Nessuno
<input type="checkbox"/>	3	Anno	varchar(50)	latin1_swedish_ci	No	Nessuno
<input type="checkbox"/>	4	temperature	double		Sì	NULL
<input type="checkbox"/>	5	humidity	double		Sì	NULL
<input type="checkbox"/>	6	wind_speed	double		Sì	NULL
<input type="checkbox"/>	7	pressure	double		Sì	NULL
<input type="checkbox"/>	8	precipitation_rate	double		Sì	NULL
<input type="checkbox"/>	9	Nome_Stazione	varchar(50)	latin1_swedish_ci	No	Nessuno

Figura 5.21: Struttura tabella “meteo”

Data	Mese	Anno	temperature	humidity	wind_speed	pressure	precipitation_rate	Nome_Stazione
2015-01-01	2015-01-00	2015	1.265838858	79.26020396	3.830803563	1030.76711	0	Milano - via Verziere
2015-01-01	2015-01-00	2015	1.243883747	79.30366993	3.929370982	1030.369198	0	Milano - via Senato
2015-01-01	2015-01-00	2015	1.200327454	78.85514945	4.694661182	1027.008018	0	Milano - Parco Lambro
2015-01-01	2015-01-00	2015	1.226741512	78.86289216	4.512110389	1027.775475	0	Milano - via Carlo Pascal
2015-01-01	2015-01-00	2015	1.323304423	79.35583162	3.349713785	1032.815653	0	Milano - P.zza Abbiategrasso
2015-01-01	2015-01-00	2015	1.290933842	79.66544131	3.233451539	1033.409764	0	Milano - viale Liguria
2015-01-01	2015-01-00	2015	1.199877418	80.29196117	3.166798345	1033.905607	0	Milano - P.zza Zavattari
2015-01-01	2015-01-00	2015	1.19730829	79.46449143	4.065368521	1029.855049	0	Milano - viale Marche

Figura 5.22: Popolazione tabella “meteo”

Le unità di misura utilizzate per i valori delle condizioni climatiche della tabella “meteo” sono: °C per la temperatura, % per l’umidità, Km/h per la velocità del vento, hPa per la pressione e % per i livelli di precipitazione.

5.4.4 Viste

Dopo aver effettuato la descrizione dettagliata delle tabelle che compongono il DW del progetto, sono necessarie alcune considerazioni sull’utilizzo di tali tabelle. Analizzando i KPI precedentemente descritti, si nota che molte delle interrogazioni che verranno effettuate sul DW, per estrarre dati utili agli utenti, riguardano principalmente tre tabelle: “inquinanti”, “stazioni” e “latlngstazioni”.

In particolare, molte delle informazioni e analisi che verranno mostrate all’utente tramite la dashboard e l’applicazione mobile, richiedono l’interrogazione della tabella “inquinanti” per l’estrazione dei dati necessari per il calcolo dei KPI. La maggior parte di quest’ultimi richiedono una granularità giornaliera ma la tabella in questione, ha una granularità oraria; per cui, ogni qualvolta un utente accede a report o analisi giornalieri, sarebbe necessaria una interrogazione al DW che comporti una elaborazione per raggruppare i dati orari presenti nella tabella e calcolare i rispettivi valori giornalieri. Naturalmente, affinché questi calcoli siano disponibili all’utilizzo su dashboard o app, è necessario un certo tempo di elaborazione, dovuto all’elevato numero di record presenti nella tabella “inquinanti” (oltre 700 mila righe).

Considerando la frequenza con il quale si interrogherebbe tale tabella e l’elevata numerosità di essa, la soluzione trovata è quella di creare e utilizzare una così detta Vista.

Le viste sono un elemento utilizzato dalla maggior parte dei DBMS. Una vista è rappresentata da una query (SELECT), il cui risultato può essere utilizzato come se fosse una tabella. Generalmente i DBMS rielaborano le query sulle viste in modo che agiscano sulle tabelle che fanno parte della vista stessa. Le viste generalmente vengono utilizzate per semplificare le query. Leggere un insieme di dati avente un significato potrebbe essere complesso, perché potrebbe richiedere eccessive JOIN fra tabelle; con una vista è possibile semplificare molto la stesura di query che leggono le informazioni.

Un altro scopo delle viste potrebbe essere semplificare o potenziare la gestione dei

permessi. Ad esempio, si potrebbe creare una query che legge solo alcuni dati da una tabella (tramite la clausola WHERE), per poi assegnare il permesso in lettura ad un certo utente sulla vista, ma non sulla tabella di base. In questo modo l'utente non vedrà i dati che non vengono estratti dalla vista.

Le viste possono essere aggiornabili, cioè è possibile eseguire su di esse comandi DML come INSERT, UPDATE e DELETE.

Quindi, dopo aver capito cosa è e a cosa serve una vista, si è passati a crearne una. Come già detto, la vista da dover creare dovrà permettere il raggruppamento dei dati contenuti nella tabella “inquinanti”, passando ad una granularità giornaliera e di conseguenza diminuendo il numero di righe. Inoltre, con lo scopo di facilitare la futura stesura di query relative alle informazioni su inquinanti e stazioni di monitoraggio, all’interno della vista sono stati inseriti i dati provenienti dalle tabelle “latlngstazioni” e “stazioni”, al fine di evitare il frequente uso dell’operatore JOIN per accedere a dati legati tra loro ma contenuti in differenti tabelle.

Per cui, dopo la creazione della vista, ogni qualvolta tramite dashboard o applicazione mobile sarà necessario l’accesso ai dati relativi alle misurazioni giornaliere degli inquinanti e alle stazioni (con relativa posizione geografica) che hanno effettuato tali misurazioni, sarà sufficiente interrogare questa vista con inoltre l’utilizzo di query molto più semplici da scrivere.

Quindi, la vista che si è deciso di creare prenderà il nome di “viewdaily” e sarà possibile realizzarla tramite la seguente query sql:

CREATE VIEW viewdaily **AS**

(SELECT latlngstazioni.Lat, latlngstazioni.Lng, stazioni.Nome_Stazione, inquinanti.Data, stazioni.Inquinante, AVG(inquinanti.Valore) AS Media_Giornaliera, inquinanti.Mese, inquinanti.Anno

FROM inquinanti, stazioni, latlngstazioni

WHERE stazioni.Id_Sensore = inquinanti.Idsensore AND stazioni.Id_Stazione = latlngstazioni.Id_Stazione

GROUP BY stazioni.Nome_Stazione, inquinanti.Data, stazioni.Inquinante, inquinanti.Mese, inquinanti.Anno)

Il risultato di questa query è quindi una vista composta come mostrato in figura 5.23.

#	Nome	Tipo	Codifica caratteri	Attributi	Null	Predefinito
<input type="checkbox"/> 1	Lat	double			No	Nessuno
<input type="checkbox"/> 2	Lng	double			No	Nessuno
<input type="checkbox"/> 3	Nome_Stazione	varchar(28)	latin1_swedish_ci		Si	NULL
<input type="checkbox"/> 4	Data	date			No	Nessuno
<input type="checkbox"/> 5	Inquinante	varchar(5)	latin1_swedish_ci		Si	NULL
<input type="checkbox"/> 6	Media_Giornaliera	decimal(10,5)			Si	NULL
<input type="checkbox"/> 7	Mese	date			No	Nessuno
<input type="checkbox"/> 8	Anno	int(4)			Si	NULL

Figura 5.23: Struttura vista “viewdaily”

Tale vista è formata da circa 35 mila record, numerosità molto minore rispetto alle oltre 700 mila righe della sola tabella “inquinanti”. Un esempio di tali record è riportato nella figura 5.24.

Lat	Lng	Nome_Stazione	Data	Inquinante	Media_Giornaliera	Mese	Anno
45.4759561	9.1408294	Milano - P.zza Zavattari	2015-01-01	C6H6	NULL	2015-01-00	2015
45.4759561	9.1408294	Milano - P.zza Zavattari	2015-01-01	CO	0.97391	2015-01-00	2015
45.4759561	9.1408294	Milano - P.zza Zavattari	2015-01-01	NO	97.04783	2015-01-00	2015
45.4759561	9.1408294	Milano - P.zza Zavattari	2015-01-01	NO2	56.52609	2015-01-00	2015
45.4759561	9.1408294	Milano - P.zza Zavattari	2015-01-02	C6H6	NULL	2015-01-00	2015
45.4759561	9.1408294	Milano - P.zza Zavattari	2015-01-02	CO	1.25000	2015-01-00	2015
45.4759561	9.1408294	Milano - P.zza Zavattari	2015-01-02	NO	146.09167	2015-01-00	2015

Figura 5.24: Popolazione vista “viewdaily”

Con la creazione e l’utilizzo della vista “viewdaily”, uno dei problemi che si aveva necessità di risolvere, cioè diminuire il tempo di estrazione dei dati utili alle analisi, non ha trovato soluzione. La causa di ciò è che questo tipo di vista non viene scritta fisicamente su disco da parte del DBMS, ma viene creata ogni qualvolta essa viene richiamata in una interrogazione sql. Di conseguenza, permette di ridurre il numero di righe da consultare e diminuisce la complessità delle query da scrivere ma non influisce positivamente sul tempo di esecuzione della query stessa, andando a volte ad aumentarne il tempo di attesa.

Per ovviare anche a questo problema, si passa all’utilizzo delle viste materializzate: esse sono viste scritte fisicamente su disco per consentirne una lettura più rapida. In pratica, essa non è altro che la copia fisica della vista originale; in questo modo, richiamando nelle interrogazioni sql la vista materializzata anziché quella originale, si diminuisce drasticamente il tempo di esecuzione perché non sarà necessario al DBMS ricreare ogni volta la vista. Nel caso in questione, la vista materializzata creata sulla base di “viewdaily” prende il nome di “vistagiornaliera”.

Uno degli aspetti da prendere in considerazione quando si parla di viste non materializzate (“viewdaily”), riguarda il loro aggiornamento: infatti, ogni qualvolta vengono inseriti nuovi dati all’interno di tabelle utilizzate per la creazione della vista, quest’ultima si aggiorna automaticamente inserendo i nuovi record.

Per le viste materializzate (“vistagiornaliera”) invece, l’aggiornamento non avviene nello stesso momento di quello della vista originale, ma sarà il DBMS a intervalli regolari ad aggiornarla inserendo i nuovi record della vista originale.

5.5 Estrazione e calcolo dei Key Performance Indicator

Come visto nel paragrafo 2.4, l’inserimento dei dati all’interno del DW aveva come obiettivo quello di estrarre determinati KPI. Dopo aver descritto nel dettaglio la struttura delle tabelle del DW, si possono ora scrivere le query sql per l’estrazione di tali KPI.

In alcuni casi, è necessario una elaborazione del risultato della relativa query per permettere l’effettivo calcolo del KPI. Tale elaborazione è stata sviluppata in linguaggio PHP.

A1. Concentrazione media giornaliera degli inquinanti per stazione

```
SELECT 'Nome_Stazione', 'Data', 'Inquinante', 'MediaGiornaliera'
FROM 'vistagiornaliera'
GROUP BY 'Nome_Stazione', 'Data', 'Inquinante'
```

A2. Concentrazione media degli inquinanti per fascia oraria giornaliera per stazione

```
SELECT S. 'Nome_Stazione', I. 'Data', S. 'Inquinante', I. 'TimeSlot', AVG(I. 'Valore')  
FROM 'stazioni' as S, 'inquinanti' as I  
WHERE S. 'Id_Sensore'=I. 'Idsensore'  
GROUP BY S. 'Nome_Stazione', I. 'Data', S. 'Inquinante', I. 'TimeSlot'
```

A3. Concentrazione media oraria degli inquinanti per stazione

```
SELECT S. 'Nome_Stazione', I. 'Data', S. 'Inquinante', I. 'Ora', I. 'Valore'  
FROM 'stazioni' as S, 'inquinanti' as I  
WHERE S. 'Id_Sensore'=I. 'Idsensore'  
GROUP BY S. 'Nome_Stazione', I. 'Data', S. 'Inquinante', I. 'Ora'
```

A4. Concentrazione media mensile degli inquinanti per stazione

```
SELECT 'Nome_Stazione', 'Mese', 'Inquinante', AVG('MediaGiornaliera')  
FROM 'vistagiornaliera'  
GROUP BY 'Nome_Stazione', 'Mese', 'Inquinante'
```

B1. Valori medi giornalieri per i diversi fenomeni metereologici per stazione

```
SELECT 'Nome_Stazione', 'Data', 'temperature', 'humidity', 'wind_speed', 'pressure',  
      'precipitation_rate'  
FROM 'meteo'
```

B2. Valori medi mensili per i diversi fenomeni metereologici per stazione

```
SELECT 'Nome_Stazione', 'Mese', 'Anno', AVG('temperature'), AVG('humidity'),  
      AVG('wind_speed'), AVG('pressure'), AVG('precipitation_rate')  
FROM 'meteo'  
GROUP BY 'Nome_Stazione', 'Mese', 'Anno'
```

C. Valori totali giornalieri per tipologia di veicolo/tipologia carburante/categoria Euro

```
SELECT *  
FROM 'catveicolo'/'tipalimentazione', 'euro'
```

D. KPI sull'analisi statistica di Correlazione(Pearson): per calcolare questo KPI, sono stati dapprima estratti tramite query sql i dati necessari e poi questi sono stati elaborati in PHP tramite delle apposite istruzioni. Nel dettaglio, l'obiettivo è quello di calcolare la correlazione di Pearson tra un particolare fenomeno meteorologico e la concentrazione di un inquinante scelto, in un intervallo temporale.

Di seguito sono mostrate la query e il blocco di codice php per il calcolo del suddetto KPI; è da sottolineare che l'utente deve indicare il fenomeno meteorologico (\$m) e l'inquinante da confrontare (\$inquinante) e il periodo temporale in cui misurare tale correlazione (\$inizio, \$fine).

```

SELECT V.`Data`, V.`Media_Giornaliera`, M.`temperature` AS Media_T, M.`humidity`
        AS Media_H, M.`wind_speed` AS Media_WS, M.`pressure` AS Media_P ,
        M.`precipitation_rate` AS Media_PR
FROM `vistagiornaliera` as V, `meteo` as M
WHERE V.`Inquinante`="`.Sinquinante.`" AND M.`Data` BETWEEN "`.Sinizio.`" AND
        "`.Sfine.`" AND M.`Data`=V.`Data` AND M.`Nome_Stazione`='Milano - via
        Senato'
GROUP BY V.`Data`

```

Sql fornirà i dati di tutti i fenomeni metereologici e le concentrazioni dell'inquinante scelto, solo per il periodo temporale indicato.

A questo punto, tramite il blocco di codice php sarà possibile calcolare il KPI riguardante la correlazione di Pearson tra l'inquinante e il fenomeno meteorologico scelto dall'utente:

```

$mediaX = array_sum($X)/$count; // media delle x dove $x è un vettore contenente i livelli di concentrazione dell'inquinante
$mediaY = array_sum($Y)/$count; // media delle y dove $y è un vettore contenente le rilevazioni del fenomeno metereologico
$SS = 0;$SX = 0;$SY = 0;
for($i=0;$i<$count;$i++){
    $SS += ($X[$i] - $mediaX) * ($Y[$i] - $mediaY);
    $SX += pow(($X[$i] - $mediaX),2);
    $SY += pow(($Y[$i] - $mediaY),2);
}
$RR = $SS / (sqrt($SX) * sqrt($SY));

```

5.6 Sviluppo della Dashboard

Nei capitoli precedenti si è visto come modellare i dati e progettare il DW necessari al progetto in questa tesi descritto. E' bene ricordare che l'obiettivo che ci si era prefissati era quello di progettare un DW per la gestione di dati riguardanti l'inquinamento e alcuni fattori come meteo e traffico, per fornire le relative informazioni e analizzare possibili correlazioni tra queste tramite l'estrazione di particolari KPI. Questi KPI, come detto anche in precedenza, servono per mostrare report e analisi riassuntive a differenti utenti, sia quelli facenti parte di una amministrazione, sia il comune cittadino che vuole tenersi aggiornato riguardo la qualità dell'aria e le possibili cause che potrebbero peggiorare questa qualità. Si è detto che per queste due tipologie di utenti, si sono individuati due differenti strumenti di visualizzazione di questi report: una dashboard per mostrare analisi più dettagliate e più incentrata sulle possibili correlazioni tra inquinamento urbano e sue possibili cause (meteo e traffico) e un'applicazione mobile che fornisca le informazioni più importanti per un comune cittadino, cioè i livelli più recenti di inquinamento nelle differenti zone urbane.

Essendo lo sviluppo di tali strumenti molto ampio e complesso, si è deciso per quanto riguarda la Tesi in questione, di occuparsi della realizzazione della dashboard sopra descritta.

In particolare, per individuare quali devono essere le sezioni contenute all'intero del cruscotto, cioè quali informazioni mostrare e quali analisi e confronti effettuare, bisogna basarsi sui dati e sui KPI a disposizione.

Nel capitolo 5.4 si sono ben descritte le tabelle che compongono il DW a disposizione e quali KPI da queste, poter estrarre tramite query sql e poter calcolare tramite codice php. Con l'utilizzo di tali KPI, verranno mostrati report di facile lettura, costituiti da mappe, istogrammi, grafici a linea e grafici plot, contenenti le informazioni necessarie ad un utente

più esperto come appunto un utente dell'amministrazione comunale, riguardo al contesto descritto (inquinamento e possibili cause); tali report possono risultare molto utili per prendere decisioni e fare delle scelte per migliorare la qualità dell'aria urbana.

Prima di procedere con lo sviluppo di questo cruscotto, deve essere chiaro, che i reports a disposizione dell'utente non saranno così detti report statici, cioè non saranno delle semplici informazioni prese dal DW e mostrate sempre allo stesso modo all'utente; si avrà la possibilità di scegliere periodi temporali con differenti granularità, gli inquinanti di interesse e per quanto riguarda il meteo, le differenti condizioni meteorologiche da confrontare con i dati sull'inquinamento. Queste scelte porteranno di conseguenza alla realizzazione di report dinamici e basati sulle sole informazioni richieste dall'utente, risultando più dettagliati e quindi più utili.

Nei paragrafi successivi, verranno dapprima descritti nel dettaglio gli strumenti e linguaggi di programmazione a disposizione per la realizzazione della dashboard che si vuole sviluppare e successivamente si andrà a vedere sezione per sezione, la composizione dei report contenuti nel cruscotto. Infine, verrà fatto un confronto con un report realizzato con un differente strumento da quelli usati, andando ad analizzare i vantaggi e svantaggi di utilizzare quest'ultimo per i report della dashboard. Per concludere, verranno mostrate alcune osservazioni effettuate sui dati stessi, proprio con l'utilizzo dei report forniti dalla dashboard, con l'obiettivo di dimostrare l'efficienza e l'utilità di quanto realizzato.

5.6.1 Tecnologie Utilizzate

Prima di descrivere le singole sezioni che comporranno la dashboard che si sta andando a realizzare, si è ritenuto necessario descrivere i linguaggi di programmazione e gli strumenti usati. Per realizzare una dashboard consultabile dagli utenti, essa deve essere raggiungibile tramite qualsiasi browser, collegandosi ad un sito web. Di conseguenza, affinché il browser possa mostrare la dashboard e possa permettere l'interazione all'utente, bisogna utilizzare quei linguaggi di programmazione nati proprio con l'obiettivo di permettere lo sviluppo di applicazioni Web. I linguaggi in questione sono: HTML, CSS, Javascript, PHP e SQL. Ognuno di questi linguaggi ha una funzione specifica per lo sviluppo della dashboard, si descrivono ora singolarmente nel dettaglio per cosa sono usati e quali sono le loro principali caratteristiche.

- **HTML:** è un linguaggio di formattazione che descrive le modalità di impaginazione o visualizzazione grafica (layout) del contenuto, testuale e non, di una pagina web attraverso tag di formattazione. Sebbene l'HTML supporti l'inserimento di script e oggetti esterni quali immagini o filmati, non prevede alcuna definizione di variabili, strutture dati, funzioni o strutture di controllo che possano realizzare programmi, il suo codice è in grado soltanto di strutturare e decorare dati testuali. Ogni tag specifica un diverso ruolo dei contenuti che esso contrassegna e descrivono caratteristiche come la funzione, il colore, le dimensioni, la posizione relativa all'interno della pagina. I browser che leggono il codice mostrano all'utente formattazioni predefinite per ogni tag che incontrano. Tuttavia, questa formattazione è completamente sotto il controllo dell'utente, che può modificarla nelle Impostazioni del suo browser.
- **CSS:** è un linguaggio usato per separare i contenuti delle pagine HTML dalla loro formattazione e permettere una programmazione più chiara e facile da utilizzare, sia per gli autori delle pagine stesse sia per gli utenti, garantendo contemporaneamente anche il riutilizzo di codice ed una sua più facile manutenzione.
- **JAVASCRIPT:** come già visto nel paragrafo 4.2, è un linguaggio di scripting orientato agli oggetti e agli eventi, comunemente utilizzato nella programmazione Web lato client per la creazione, in siti web e applicazioni web, di effetti dinamici interattivi tramite

funzioni di script invocate da eventi innescati a loro volta in vari modi dall'utente sulla pagina web in uso (mouse, tastiera, caricamento della pagina ecc...).

- **PHP:** Anche questo linguaggio è stato già in parte descritto in precedenza nel paragrafo 4.2. Come già in parte detto quindi, è un linguaggio di scripting interpretato, originariamente concepito per la programmazione di pagine web dinamiche. Attualmente è principalmente utilizzato per sviluppare applicazioni web lato server, ma può essere usato anche per scrivere script a riga di comando o applicazioni stand-alone con interfaccia grafica.
- **SQL:** è un linguaggio standardizzato per database basati sul modello relazionale (RDBMS) progettato per: creare e modificare schemi di database (DDL - Data Definition Language); inserire, modificare e gestire dati memorizzati (DML - Data Manipulation Language); interrogare i dati memorizzati (DQL - Data Query Language) e creare e gestire strumenti di controllo ed accesso ai dati (DCL - Data Control Language).

Quindi come si può intuire dalle descrizioni appena date, per realizzare il layout e la struttura della dashboard verrà utilizzato l'HTML e per tutto ciò che riguarderà la parte grafica di questo cruscotto si riterrà opportuno usare CSS. Javascript essendo linguaggio lato client si occuperà soprattutto di presentare le analisi all'utente facendo le dovute elaborazioni sulle informazioni ricevute dal server. Infine, PHP svolgerà un ruolo fondamentale nel funzionamento della dashboard perché essendo un linguaggio lato server si occuperà di comunicare con il server in cui è contenuto il DW, validare i parametri inseriti dall'utente, interrogare le tabelle del DW, tramite query scritte in SQL, per ricevere gli opportuni dati e infine elaborare tali dati per passarli a Javascript che si occuperà come detto poco fa, di mostrare le opportune analisi all'utente.

Oltre ai linguaggi di programmazione descritti nel paragrafo precedente, si illustrano ora gli altri strumenti e piattaforme usate per la realizzazione e gestione della dashboard del quale si sta trattando in questa Tesi. Di seguito sono analizzate singolarmente le caratteristiche di questi strumenti e poi successivamente verrà spiegato come questi sono stati utilizzati nel caso pratico dello sviluppo della dashboard.

- **GOOGLE CHARTS:** è un servizio che offre un modo per visualizzare i dati su di un sito web. Tale servizio offre un gran numero di tipi di grafici pronti all'uso: dai semplici grafici a linee a complesse mappe ad albero gerarchiche. Il modo più comune per utilizzare Google Charts è con un semplice JavaScript da incorporare nella propria pagina web. Tramite alcune librerie di Google Chart, si ha la possibilità di indicare dati da tracciare, selezionare le opzioni per personalizzare il proprio grafico e infine creare un oggetto grafico con uno specifico ID. Con la creazione di un <div> all'interno della propria pagina html, con l'indicazione dell'id specificato si potrà visualizzare il grafico di Google. I grafici sono altamente interattivi e permettono la connessione per creare dashboard complessi e sono resi utilizzando la tecnologia HTML5 / SVG per fornire compatibilità cross-browser e portabilità cross-platform per iPhone, iPad e Android, cioè gli utenti non dovranno mai utilizzare plugin o software per visualizzare correttamente i grafici. Tutti i tipi di grafico vengono popolati con i dati utilizzando la classe DataTable; questa fornisce metodi per l'ordinamento, la modifica e il filtraggio dei dati e può essere popolato direttamente dalla pagina Web, da un database o da qualsiasi provider di dati che supporta il protocollo DataSource.
- **MAPPE DI GOOGLE:** sono uno strumento utile per georeferenziare attività e punti di interesse. Sono state adottate su moltissimi siti web come sostituto o integrazione per la sezione 'Dove siamo' ma le possibilità di utilizzo vanno ben oltre la semplice visualizzazione aerea di un'area geografica. Fino ad ora un utilizzo più professionale e personalizzato di questo strumento era possibile solo utilizzando in modo consistente le API messe a disposizione da Google e quindi mettendo pesantemente mano al codice

JavaScript all'interno delle proprie pagine. Da quando è stata attivata la sezione Le mie mappe su Google Maps è possibile una personalizzazione e una integrazione delle mappe più completa e alla portata di tutti, anche degli utenti meno esperti. Per utilizzare le mappe di Google Maps occorre avere un account Gmail. Per integrare una mappa sul proprio sito prima di tutto bisogna possedere una chiave di registrazione (API key). Il risultato dell'invio del form è una pagina dove viene indicata la chiave (API key) che è stata generata: tenetela da parte, la utilizzeremo tra poco. Più in basso, nella stessa pagina, viene proposto il codice da cui partire per includere la mappa all'interno della propria pagina HTML. Utilizzando il codice HTML proposto durante la creazione della chiave come punto di partenza per integrare la mappa all'interno del proprio sito si è già fatto parte del lavoro. Nel dettaglio, il codice fornito è costituito da un primo tag script che utilizza come parametro, tra gli altri, la chiave generata per validare la mappa relativamente alla url del sito. Il secondo tag script contiene una funzione JavaScript che viene lanciata al caricamento della pagina HTML (tramite l'istruzione onload presente all'interno del tag body) e che genera la mappa all'interno dell'elemento che ha id="map", un div che si trova all'interno del corpo della pagina al quale si può dare le dimensioni che più interessano impostando i valori di width e height. In questo script è presente anche una riga che posiziona geograficamente la mappa centrandola ad un valore di latitudine e imposta lo zoom. La mappa verrà visualizzata all'interno del body della pagina HTML dove dovrà essere presente un div vuoto che abbia come id il valore map e che abbia delle dimensioni definite a piacimento. A questo div si potranno assegnare anche altri valori di stile. Infine, tramite la chiamata a getElementById(map) passata come parametro, lo script utilizzerà questo div come contenitore per la mappa generata.

- **GOOGLE FUSION TABLES:** è un servizio web fornito da Google per la gestione dei dati. Le tabelle Fusion possono essere utilizzate per raccogliere, visualizzare e condividere tabelle di dati. I dati sono memorizzati in più tabelle che gli utenti di Internet possono visualizzare e scaricare. Il servizio web fornisce i mezzi per visualizzare i dati con grafici a torta, grafici a barre, grafici a linee, grafici a dispersione, linee temporali e mappe geografiche. I dati vengono esportati in un formato di file con valori separati da virgola.

Nella dashboard che si vuole realizzare in questo lavoro di Tesi, gli strumenti appena descritti sono stati utilizzati per fornire all'utente delle mappe contenenti informazioni di facile lettura, andando a indicare su tali mappe le stazioni di monitoraggio nel comune di Milano e per ognuna di esse, i dati più significativi, resi più chiari anche tramite una diversità cromatica dei vari marker. Su queste mappe tramite un file xml contenente i contorni delle diverse zone di Milano, si sono potuti disegnare tali contorni trasformando questo file xml in una vera e propria tabella e poi integrandola con l'utilizzo del servizio di Google delle Fusion Tables.

Infine, lo strumento più usato per fornire analisi e report immediati, chiari e ricchi di informazioni, è stato quello dei grafici di Google che prendendo i dati estratti dal DW, hanno permesso di mostrare tali informazioni in maniera grafica, permettendo anche personalizzazioni utili per chiarificare ancor di più queste analisi.

Nei paragrafi successivi, si descriverà nel dettaglio ogni sezione facente parte della dashboard che si vuole realizzare.

5.6.2 Dashboard

La dashboard che si andrà a sviluppare, come ben descritto in tutta la Tesi, ha l'obiettivo di fornire report e analisi di facile lettura e ricche di informazioni ad utenti più esperti o più interessati a informazioni dettagliate riguardo l'inquinamento atmosferico urbano e la correlazione con le condizioni climatiche e il traffico cittadino. Tali informazioni, verranno mostrate sulla base di parametri inseriti dall'utente stesso, che potrà richiedere differenti KPI

da estrarre dal DW, tramite le diverse sezioni di questa dashboard.

In figura 5.25 è mostrata la pagina iniziale della dashboard, da questa si può notare come ci siano quattro differenti sezioni: la prima corrispondente alla pagina iniziale e le successive riguardanti i vari KPI estratti dal DW secondo le query descritte nel paragrafo 5.5.

Di seguito verranno descritte tali sezioni, quali KPI sono in esse mostrate e in quale forma grafica sono questi mostrati all'utente.

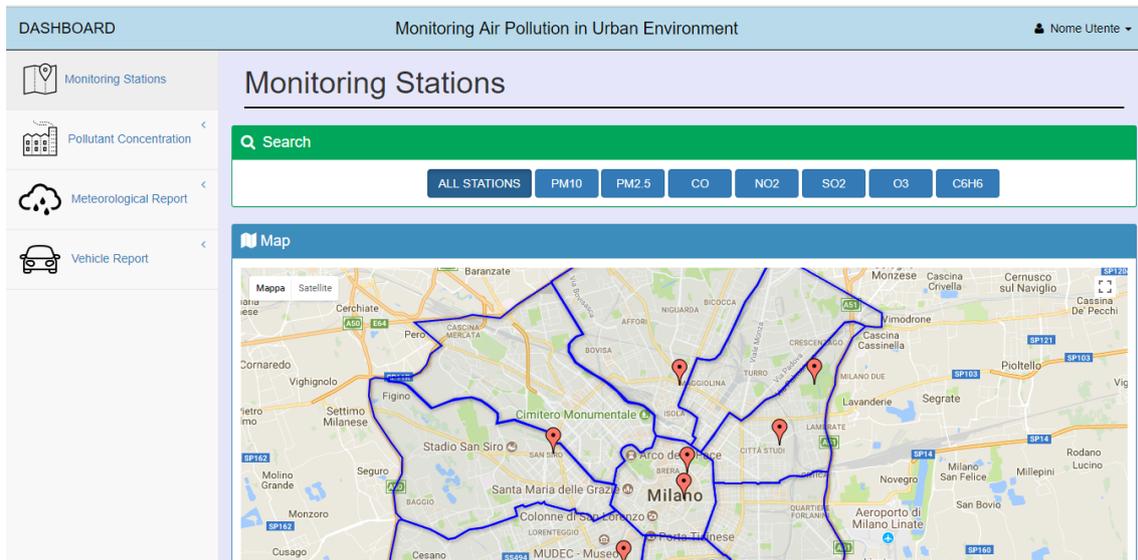


Figura 5.25: schermata iniziale della dashboard

5.6.2.1 Monitoring Stations

La sezione Monitoring Stations permette all'utente di avere una visione periferica sull'intera città di Milano, infatti qui vengono mostrate su mappa, le differenti zone cittadine e la posizione delle stazioni di monitoraggio che forniscono i dati presenti nel DW riguardanti le concentrazioni d'inquinamento dei diversi inquinanti. La mappa mostrata in questa sezione, fornisce all'utente una descrizione generale riguardo le stazioni di monitoraggio, infatti cliccando sui differenti marker, come mostrato dall'esempio in figura 5.26, l'utente potrà conoscere la via dove la stazione è posizionata e i differenti inquinanti che essa controlla.

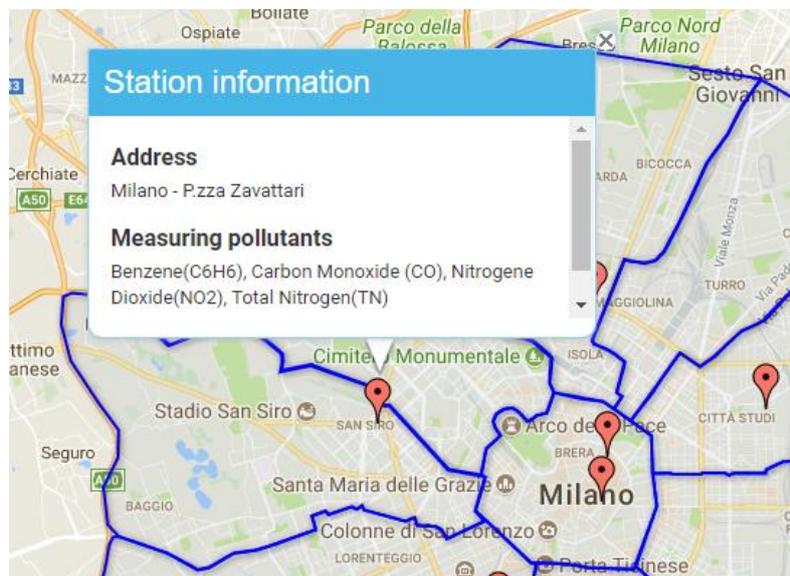


Figura 5.26: Finestra di informazioni della stazione

Proprio la conoscenza dei singoli inquinanti misurati può essere utile ad un utente che ha interesse di vedere la posizione delle sole stazioni di monitoraggio che rilevano le concentrazioni di un particolare inquinante. Per evitare che l'utente debba cliccare ogni singolo marker sulla mappa per capire quali stazioni sono di suo interesse, è stata data la possibilità all'utente stesso di richiedere la visualizzazione sulla mappa dei soli marker riguardanti le stazioni che monitorano un particolare inquinante. Tale scelta è possibile tramite la sezione "Search", mostrata in figura 5.27, in cui l'utente può indicare l'inquinante di proprio interesse.



Figura 5.27: Barra di ricerca di Monitoring Stations

Naturalmente al momento dell'accesso nella sezione Monitoring Stations, la selezione attivata sul tasto All Stations, infatti come visto prima, vengono mostrate tutte le stazioni di monitoraggio della città di Milano. Se ad esempio l'utente decide di visualizzare le sole stazioni che misurano l'inquinante PM10, il report che verrà mostrato sarà quello in figura 5.28

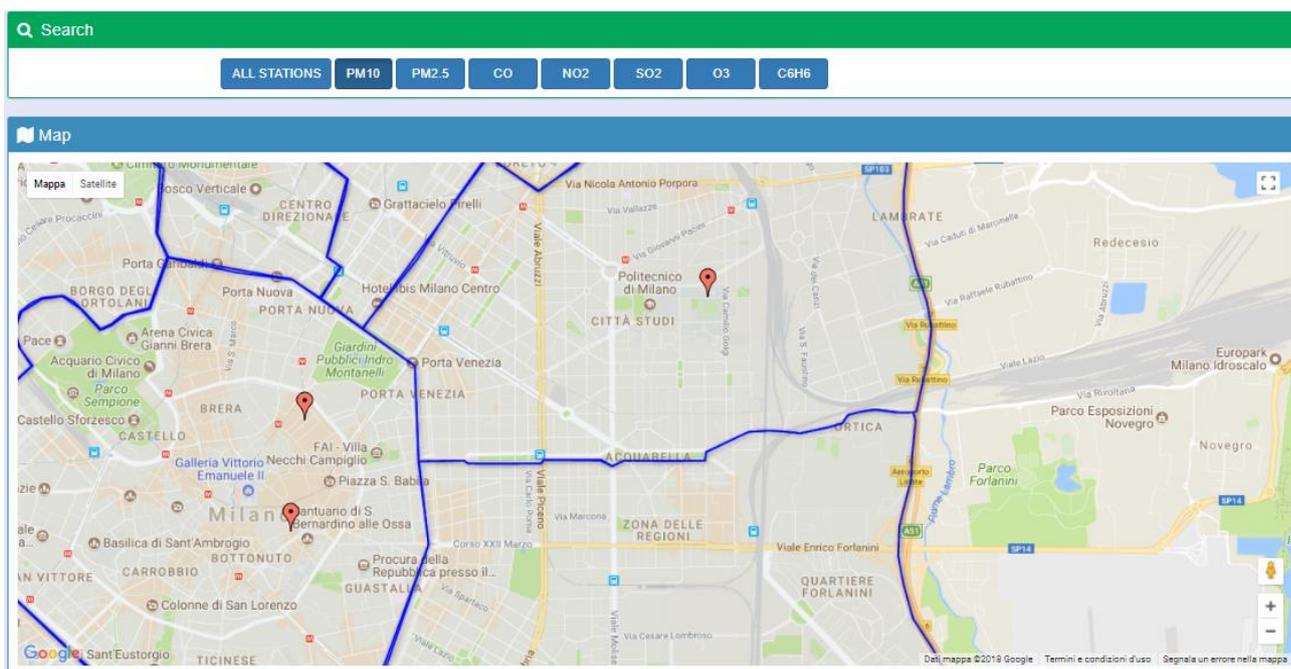


Figura 5.28: Monitoring Stations per inquinante PM10

Come si può notare, dopo la selezione PM10, le stazioni di monitoraggio verranno filtrate per mostrare come detto, solo quelle che controllano le concentrazioni di PM10.

Molto importante per la descrizione dello sviluppo di tale dashboard, non è solamente quale sia il funzionamento di ogni singola sezione, ma anche come è costituito tutto il flusso di informazioni.

Infatti, l'accesso alla sezione Monitoring Stations, farà partire direttamente una richiesta al server per recuperare le informazioni geografiche, riguardanti le stazioni di monitoraggio; per

fare questo, tramite codice php, verrà fatta una richiesta al DW, contenente una interrogazione sql così costituita:

```
SELECT latlngstazioni.`Lat`,latlngstazioni.`Lng`,latlngstazioni.`Nome_Stazione`  
FROM `latlngstazioni`,`stazioni`  
WHERE stazioni.Id_Stazione=latlngstazioni.Id_Stazione
```

La risposta del DW a questa interrogazione sarà la lista dei nomi di tutte le stazioni di monitoraggio contenute nella tabella “stazioni”, con l’indicazione per ognuna di esse delle relative coordinate geografiche contenute nella tabella “latlngstazioni”.

Tale richiesta è molto simile a quella fatta al server nel momento della selezione dell’inquinante da parte dell’utente, in questo caso infatti php memorizza la scelta fatta (\$inquinante) e effettua la seguente interrogazione al DW:

```
SELECT latlngstazioni.`Lat`,latlngstazioni.`Lng`,latlngstazioni.Nome_Stazione  
FROM `latlngstazioni`,`stazioni`  
WHERE stazioni.Id_Stazione=latlngstazioni.Id_Stazione AND  
stazioni.Inquinante=$inquinante
```

L’unica differenza tra le due query è che nella seconda verranno prese le stazioni dalla tabella “stazioni” che tra gli inquinanti monitorati hanno l’inquinante indicato dall’utente. In generale però, la struttura del risultato ricevuto dal DW all’interrogazione sarà lo stesso; a questo punto, tramite il codice php, questo risultato viene elaborato e fornito al codice JavaScript che tramite l’utilizzo delle mappe di Google, posizionerà dei marker sulla mappa nelle posizioni geografiche indicate dal server, andando per ogni marker, ad inserire le informazioni da visualizzare al momento del click sul marker stesso. Una volta creata la mappa, tramite il linguaggio HTML e lo stile impostato con CSS, essa verrà mostrata all’utente, che in questo modo avrà la possibilità di avere la visione periferica sul monitoraggio dell’inquinamento effettuato nella città di Milano.

5.6.2.2 Pollutant Concentration

Tramite il menù laterale, come mostrato in figura 5.29, si può accedere alla seconda sezione della dashboard, quella denominata Pollutant Concentration. La figura 5.29 mostra come tale sezione sia composta a sua volta da 3 sotto-sezioni: Daily, Time Slot e Hourly.

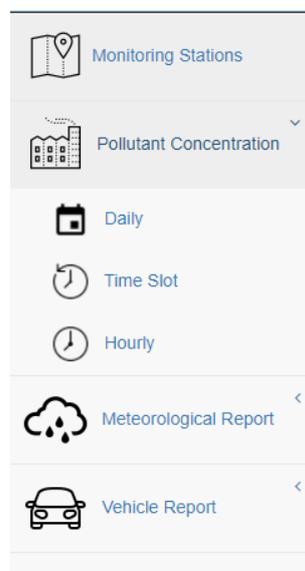


Figura 5.29: Sotto-sezioni di Pollutant Concentration nel menù laterale

Queste tre sotto-sezioni, come si può già intuire dai loro nomi, differiscono tra loro perché forniscono informazioni riguardanti le concentrazioni di inquinamento, rilevate con granularità diversa, rispettivamente: giornaliera, per fascia oraria e oraria.

Dato questa sezione, riguarda le concentrazioni di inquinamento si può subito immaginare quali siano gli indicatori di riferimento; infatti, dal paragrafo 5.3 dove sono stati descritti i KPI su cui ci si sarebbe basati per lo sviluppo della dashboard, si può notare che i KPI nella categoria A. riguardano proprio le rilevazioni delle concentrazioni di inquinamento.

Di conseguenza, è facile intuire che saranno proprio tali KPI il punto di partenza dei report e analisi mostrate in queste tre sezioni; con la sola precisazione che il KPI A4. che forniva le concentrazioni con granularità mensile, non verrà utilizzato nella sezione in questione ma nella sezione Meteorological Report; si vedrà meglio nei paragrafi successivi il perché di tale scelta.

Si torna ora a descrivere dettagliatamente la composizione dei report presenti nelle tre sottosezioni, come l'utente può interagire con tali report, quali informazioni sono mostrate e come queste sono estratte dal DW.

1. Daily Pollutant Concentration

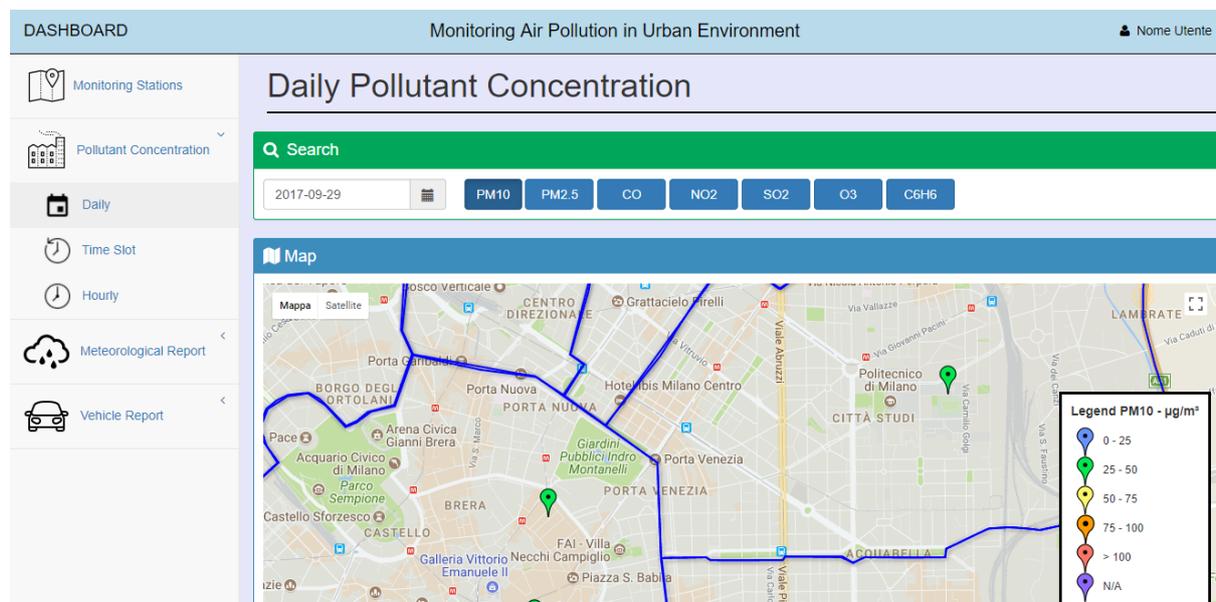


Figura 4.30: Daily Pollutant Concentration

La sezione Daily Pollutant Concentration permette all'utente di avere un report dettagliato e allo stesso tempo chiaro sulle concentrazioni di inquinanti misurate in una specifica data.

Tali informazioni verranno mostrate in due forme: tramite una mappa e tramite visualizzazione grafica con istogrammi. Quindi, come si può notare dalla figura 5.31 sottostante, l'utente ha due parametri da scegliere per impostare l'analisi e di conseguenza per indicare quali dati estrarre dal DW. Questi due parametri riguardano appunto, il periodo temporale di riferimento del report (la data), sia l'inquinante del quale mostrare i livelli di inquinamento.

La figura 5.31 mostra come si presenta la barra di Search al momento del caricamento della pagina, cioè con la selezione della data impostata sul giorno più recente presente nel DW (in questo caso è il 29 settembre 2017) e sull'inquinante più importante e significativo per quanto riguarda l'inquinamento urbano (PM10).



Figura 5.31: Barra di ricerca di Daily Pollutant Concentration

Nel caso in cui l'utente volesse consultare le informazioni riguardanti le concentrazioni di inquinamento dovute al PM10 ma di un giorno differente da quello più recente, sarà possibile farlo tramite il widget Calendar, come mostrato in figura 5.32.

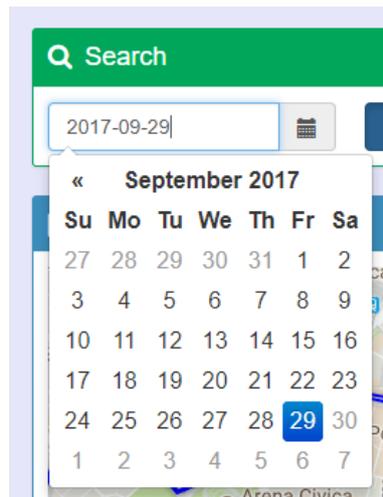


Figura 5.32: Widget Calendar

Una volta aver compreso quali sono i parametri scelti dall'utente, passiamo a descrivere le varie visualizzazioni mostrate nel report e quali dati vengono estratti dal DW e in che modo, per tali visualizzazioni.

Come si può notare dalla figura 5.33, la mappa contenuta in questo report mostra le stazioni di monitoraggio del comune di Milano che misurano l'inquinante selezionato e per ognuno dei marker posizionati sulla mappa è stato utilizzato un diverso colore a seconda della fascia di inquinamento in cui ricade la misurazione fatta di quel determinato inquinante in quella specifica data da quella specifica stazione.

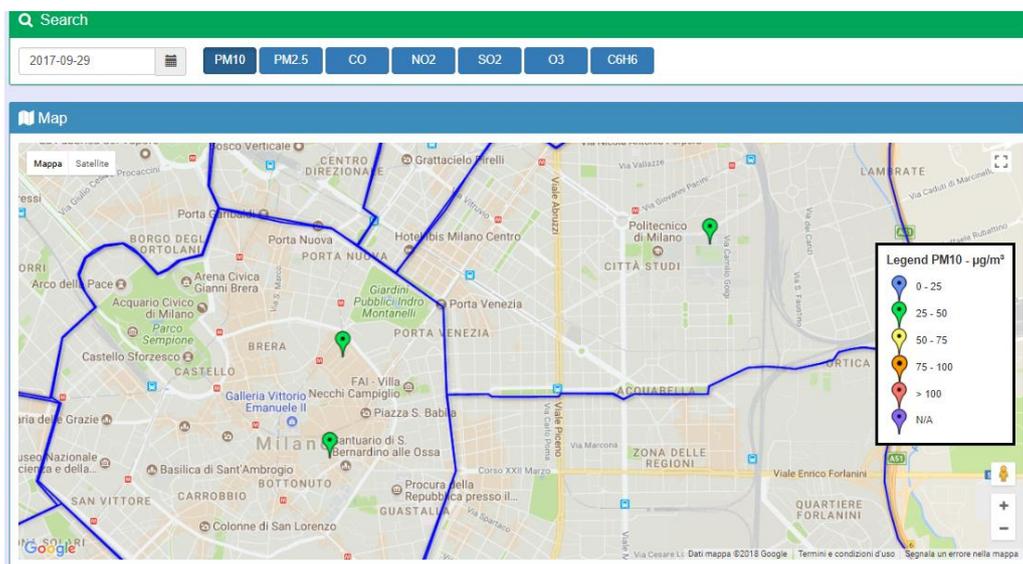


Figura 5.33: Mappa in Daily Pollutant Concentration

Nella figura 5.33 sono appunto mostrate le tre stazioni di monitoraggio dell'inquinante PM10 e tutte sono colorate di verde, tale colore è legato con la legenda mostrata a lato in cui sono mostrate le associazioni tra colori dei marker e fasce di valori delle concentrazioni dell'inquinante scelto. Nel caso in questione risulta che in data 29 Settembre 2017, le rilevazioni effettuate sulla concentrazioni giornaliera del PM10 hanno fornito tutte valori compresi nella fascia 25-50 $\mu\text{g}/\text{m}^3$.

Per far sì di conoscere il preciso valore di concentrazione giornaliera dell'inquinante si è fornita all'utente la possibilità di cliccare sui singoli marker per visualizzare tale valore, come mostrato in figura 5.34

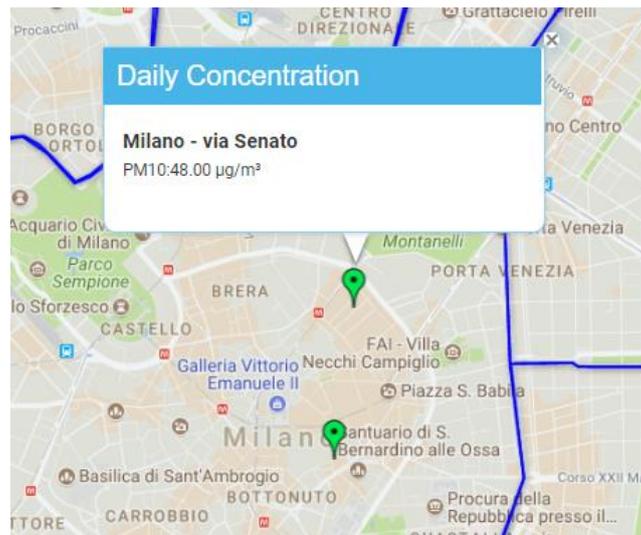


Figura 5.34: Informazioni su mappa in Daily Pollutant Concentration

Si è visto quindi che l'informazione principale gestita fin'ora da questa sezione Daily Pollutant Concentration è stata, come dice il nome stesso, la concentrazione media giornaliera dell'inquinante.

Come descritto anche per la sezione Monitoring Stations e come verrà fatto anche per le successive sezioni, si va a dettagliare ora quali dati vengono estratti dal DW per visualizzare le informazioni sulla mappa e come tali dati vengano elaborati prima dell'effettiva visualizzazione. Al momento della selezione dell'inquinante ($\$inquinante$) e della data ($\day) da parte dell'utente, verrà inviata una richiesta al server contenente questi due parametri, questa richiesta permetterà di estrarre i dati necessari dal DW, in particolare dalla vista creata durante il progetto "vistagiornaliera".

Come già visto, tale vista permette di estrarre informazioni con granularità giornaliera anziché oraria (come è nella tabella "inquinanti") riguardanti le concentrazioni misurate da ogni stazione per ogni inquinante. Nel caso in questione, da tale vista verranno estratte le coordinate geografiche, il nome delle differenti stazioni e le misurazioni dell'inquinante indicato fatte da ognuna di tali stazioni. Quanto detto è possibile tramite l'utilizzo della query scritta in linguaggio sql, qui di seguito riportata:

```
SELECT vistagiornaliera.Lat,vistagiornaliera.Lng,vistagiornaliera.Nome_Stazione,
        vistagiornaliera.Media_Giornaliera
FROM vistagiornaliera
WHERE vistagiornaliera.Data = $day AND vistagiornaliera.Inquinante = $inquinante
```

Da questa query è possibile intuire che l'indicatore che viene estratto coincide con quello indicato nel paragrafo 5.3, con la numerazione A1, cioè *Concentrazione d'inquinamento media giornaliera per stazione*.

Dopo aver estratto questi dati, essi vengono restituiti dal server e sarà JavaScript tramite lo strumento delle mappe di Google a posizionare i marker sulle rispettive posizioni geografiche e inserire all'interno della finestra informativa di ogni marker, il nome della stazione e il valore di inquinamento individuato in quella data per quello specifico inquinante. La mappa però, così non risulta ancora pronta per essere mostrata perché i marker devono essere colorati a seconda delle soglie di appartenenza mostrate nella legenda. Quindi, quando vengono estratti i dati sulle concentrazioni medie giornaliere, è necessario interrogare il DW un'ulteriore volta, tramite una diversa query, che acceda alla tabella "soglie" ed estragga solo quelle relative all'inquinante indicato dall'utente.

La query appena descritta è così composta:

```
SELECT `Soglia1`, `Soglia2`, `Soglia3`, `Soglia4`
FROM `soglie`
WHERE Nome_Inquinante=$inquinante
```

Le soglie verranno anch'esse gestite da Javascript che assegnerà un colore ad ogni soglia (come mostrato nella legenda della mappa) e di conseguenza si controllerà in quale soglia ricade la concentrazione d'inquinante misurata dalla stazione e si colorerà il relativo marker del colore di quella soglia individuata.

Dopo aver descritto la visualizzazione su mappa, si passa a mostrare delle informazioni su grafici tramite i Google Charts. In particolare, nella sezione Daily Pollutant Concentration verrà visualizzato un grafico che mostra i livelli di concentrazione dell'inquinante scelto dall'utente per ogni stazione, sia nella data selezionata, sia nei 10 giorni precedenti. Questo permette di avere un trend degli andamenti di quell'inquinante in un preciso lasso di tempo.

Anche qui, si può intuire che il KPI usato è sempre quello indicato nel capitolo 5.3 con A1, a differenza della mappa però non si estraggono solo i valori della giornata selezionata, ma anche quelli dei giorni precedenti.

Quanto appena descritto a parole, viene mostrato tramite la figura 5.35 in cui è rappresentato proprio l'istogramma in questione.

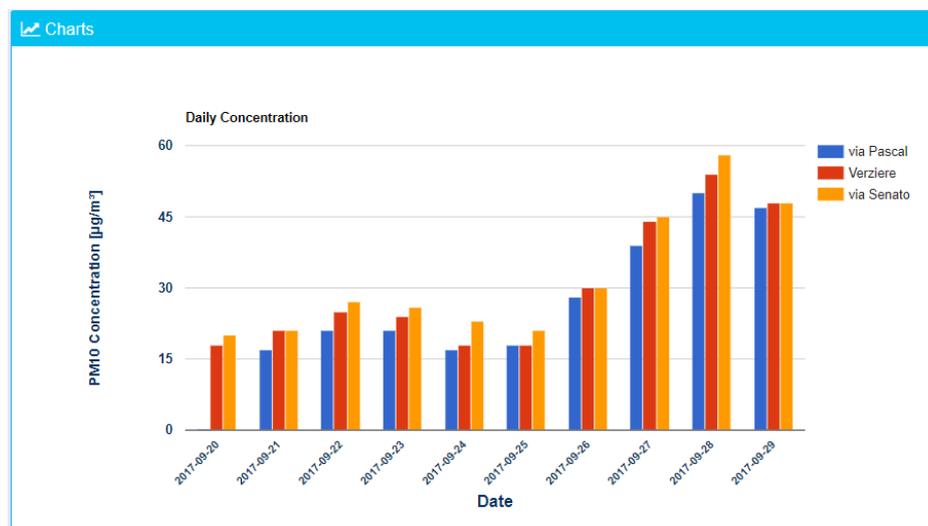


Figura 5.35: Daily Concentration

Questo grafico sarà quindi costituito da una serie di istogrammi (uno per ogni giorno), ognuno di essi sarà costituito da tante barre verticali, quante sono le stazioni di monitoraggio che misurano l'inquinante indicato dall'utente. L'appartenenza di una determinata rilevazione ad una stazione è facilmente riconoscibile tramite la differente colorazione delle barre in funzione proprio della stazione di monitoraggio. Questa visualizzazione grafica è possibile grazie all'elaborazione di Javascript dei dati ricevuti dal server in seguito alla richiesta dell'intero report.

Tale richiesta fatta ala server grazie a php, contiene un'interrogazione sql per estrarre i dati dal DW; tale query è così composta:

```
SELECT vistagiornaliera.Nome_Stazione,vistagiornaliera.Data,  
vistagiornaliera.Media_Giornaliera  
FROM vistagiornaliera  
WHERE vistagiornaliera.Data BETWEEN adddate($day,-9) and $day AND  
vistagiornaliera.Inquinante=$inquinante
```

Questa query estrae le informazioni così come sono necessarie all'istogramma prima descritto, cioè per ogni data nell'intervallo indicato, estrae il nome della stazione di monitoraggio che ha effettuato la rilevazione e la relativa concentrazione media giornaliera. Anche in questo caso è da notare che l'interrogazione è fatta solo sulla vista "vistagiornaliera" che come già detto permette di avere le concentrazioni medie giornaliere già calcolate.

Infine, per quanto riguarda la pagina Daily Pollutant Concentration, c'è un altro grafico a istogrammi da analizzare, esso va nel dettaglio della data indicata dall'utente e mostra le differenti rilevazioni delle stazioni di monitoraggio effettuate durante la giornata per l'inquinante selezionato ma con un livello di granularità maggiore: non più giornaliera ma per fascia oraria.

Parlando di fascia oraria, è intuitivo capire che il KPI su cui ci si basa per tale grafico non è più quello indicato nel capitolo 5.3 con A1, ma sarà A2: *Concentrazione media dell'inquinante per fascia oraria per stazione*. Un'altra specificazione da dover fare prima di vedere il grafico e la relativa query per estrarre i dati necessari, è che sia in questo caso, sia negli altri casi in cui si parlerà di rilevazioni media per fascia oraria o direttamente rilevazioni orarie, ci saranno due inquinanti da non prendere in considerazione, cioè PM10 e PM2.5. Tali inquinanti vengono rilevati solo con granularità giornaliera, quindi analisi con granularità maggiori non sono possibili. Infatti, nel momento della selezione di uno di questi inquinanti da parte dell'utente, la pagina che verrà caricata conterrà solo la mappa e il grafico dei 10 giorni precedenti, prima descritto; invece, nel caso di selezione di un altro inquinante diverso da PM10 e PM2.5, la pagina caricata visualizzerà anche il grafico appena descritto, con il dettaglio sulle fasce orarie della data scelta. Tale grafico è mostarto in figura 5.36.

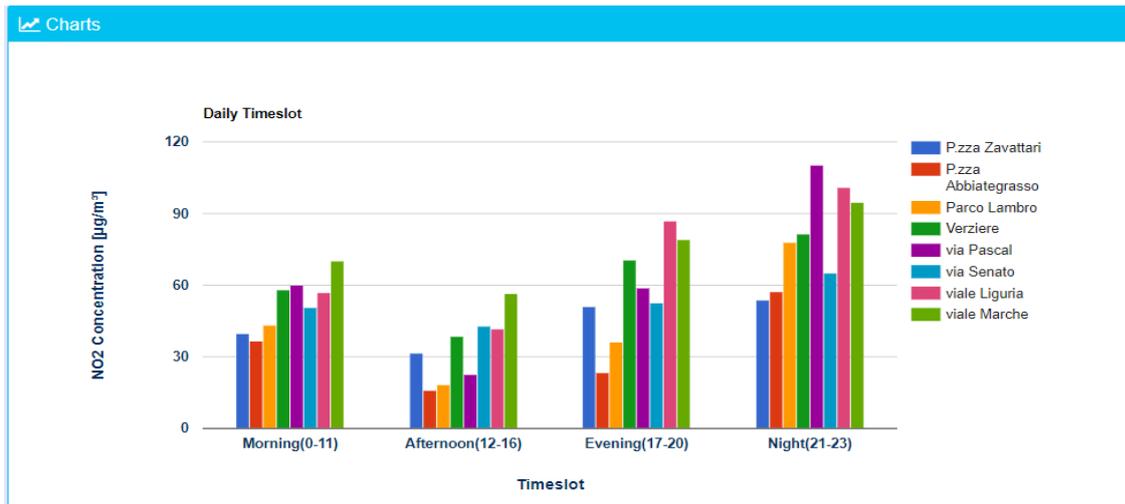


Figura 5.36: Daily Timeslot

In questo grafico, ci saranno quattro istogrammi, uno per ogni fascia oraria e ogni istogramma conterrà tante barre quante sono le stazioni di monitoraggio che rilevano i valori dell'inquinante scelto. In questo caso, i dati che gestirà Javascript per creare il grafico con lo strumento di Google Charts sono sì estratti dal DW tramite richiesta al server, ma effettuando una interrogazione sql per prendere i livelli di concentrazione, non più da “vistagiornaliera” (che contiene informazioni con granularità giornaliera) ma dalla tabella “inquinanti” (che ha un livello di granularità oraria). La query sarà così fatta:

```

SELECT stazioni.Nome_Stazione,inquinanti.TimeSlot,AVG (inquinanti.Valore)
AS Media_DailySlot
FROM inquinanti,stazioni
WHERE stazioni.Id_Sensore=inquinanti.Idsensore AND inquinanti.Data=$day AND
stazioni.Inquinante=$inquinante
GROUP BY stazioni.Nome_Stazione,inquinanti.CodTS

```

Questa query è un minimo più complessa rispetto a quelle viste precedentemente, questo perché per creare la vista giornaliera era già state effettuate alcune operazioni di join tra tabelle che non erano più necessarie; in questo caso invece, l'operazione di join tra le tabelle “inquinanti” e “stazioni” è stata necessaria per mettere in relazione il nome delle stazioni di monitoraggio e le relative misurazioni effettuate nella data e per l'inquinante specificato.

Naturalmente, avendo necessità di una granularità per fascia oraria, le concentrazioni orarie rilevate sono state raggruppate e da queste è stato estratto un singolo valore per ogni fascia oraria(CodTS).

2. TimeSlot Pollutant Concentration

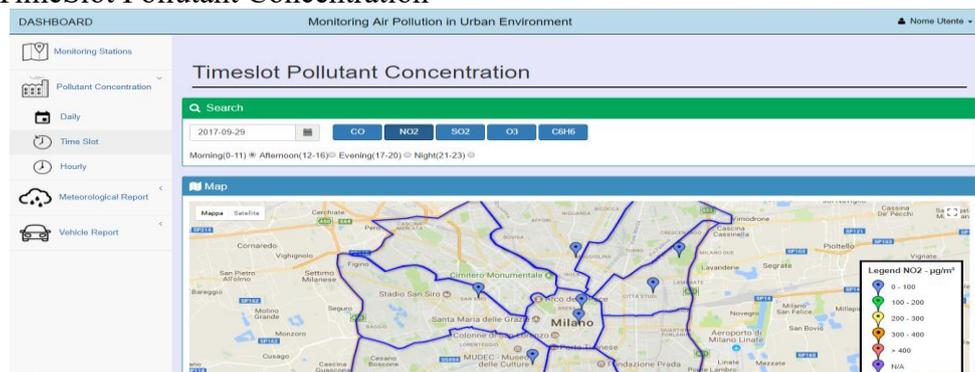


Figura 5.37: TimeSlot Pollutant Concentration

La sezione Time Slot Pollutant Concentration permette all'utente di avere un report dettagliato e allo stesso tempo chiaro sulle concentrazioni di inquinanti misurate in una specifica data, con un focus su una specifica fascia oraria(anch'essa scelta dall'utente).

Come per Daily Pollutant Concentration, anche per questo report verranno usate mappe e istogrammi per visualizzare le informazioni. Come si può notare dalla figura 5.38 sottostante, l'utente questa volta ha tre parametri da scegliere per impostare l'analisi e di conseguenza per indicare quali dati estrarre dal DW. Di questi tre parametri, due riguardano appunto, il periodo temporale di riferimento del report (la data e la fascia oraria), l'altro riguarderà l'inquinante del quale mostrare i livelli di inquinamento.

La figura 5.38 mostra come si presenta la barra di Search al momento del caricamento della pagina, cioè con la selezione della data impostata sul giorno più recente presente nel DW (in questo caso è il 29 settembre 2017), con la selezione della fascia orari impostata su quella mattutina (Morning) e sull'inquinante più importante e significativo per quanto riguarda l'inquinamento urbano (NO₂). E' da sottolineare l'assenza degli inquinanti PM₁₀ e PM_{2.5}, questo perché, come detto anche in precedenza, di questi sono disponibili solo rilevazioni giornaliera, quindi mostrare per essi informazioni di dettaglio per fascia oraria sarebbe impossibile.

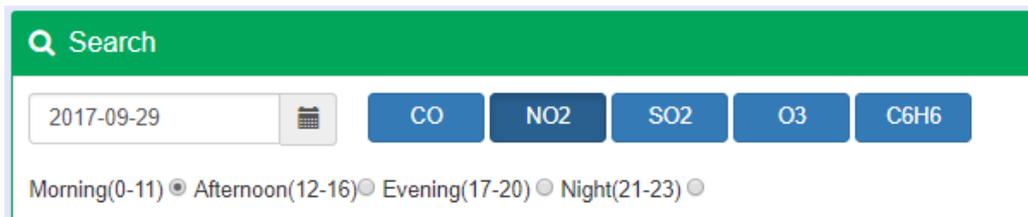


Figura 5.38: Barra di ricerca di Timeslot Pollutant Concentration

Una volta aver compreso quali sono i parametri scelti dall'utente, passiamo a descrivere le varie visualizzazioni mostrate nel report e quali dati vengono estratti dal DW e in che modo, per tali visualizzazioni.

Come si può notare dalla figura 5.39, la mappa contenuta in questo report mostra le stazioni di monitoraggio del comune di Milano che misurano l'inquinante selezionato e per ognuno dei marker posizionati sulla mappa è stato utilizzato un diverso colore a seconda della fascia di inquinamento in cui ricade la misurazione fatta di quel determinato inquinante in quella specifica data, in quella specifica fascia oraria e da quella specifica stazione.

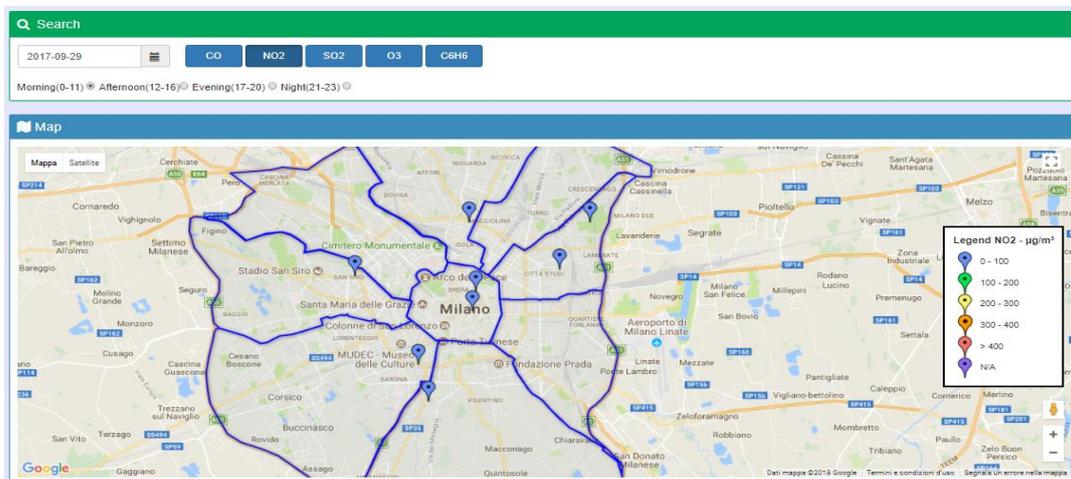


Figura 5.39: Mappa in Timeslot Pollutant Concentration

Il funzionamento della mappa in figura 5.39 sarà il medesimo dello stesso della sezione Daily Pollutant Concentration, cioè sono mostrare le sole stazioni di monitoraggio dell'inquinante scelto, per ognuna di esse verrà posizionato un marker, il colore di tale marker è legato con la legenda mostrata a lato in cui sono mostrate le associazioni tra colori dei marker e fasce di valori delle concentrazioni dell'inquinante scelto. La differenza in questo caso è che i valori che verranno considerati per l'assegnazione della soglia, ma anche che verranno visualizzati al momento del click sul marker non saranno quelli giornalieri ma quelli calcolati come media dei valori orari appartenenti alla sola fascia oraria indicata dall'utente.

L'informazione principale quindi, fino ad ora, è la concentrazione media dell'inquinante per fascia oraria. Come descritto per le sezioni precedenti e come verrà fatto anche per le quelle successive, si va a dettagliare ora quali dati vengono estratti dal DW per visualizzare le informazioni sulla mappa e come tali dati vengano elaborati prima dell'effettiva visualizzazione.

La richiesta al server conterrà i tre parametri inseriti dall'utente: inquinante (\$inquinante), data(\$day), fascia oraria(\$codts), questa richiesta permetterà di estrarre i dati necessari dal DW, in particolare, essendo necessari in questo caso dati con granularità maggiore di quella della vista, sarà necessari effettuare operazioni di join tra più tabelle per estrarre tutte le informazioni: le coordinate geografiche, il nome delle differenti stazioni e le misurazioni dell'inquinante indicato fatte da ognuna di tali stazioni nel periodo temporale specificato. Quanto detto è possibile tramite l'utilizzo della query scritta in linguaggio sql, qui di seguito riportata:

```
SELECT latlngstazioni.Lat,latlngstazioni.Lng,stazioni.Nome_Stazione,
        AVG(inquinanti.Valore) AS Media_TimeSlot
FROM inquinanti,stazioni,latlngstazioni
WHERE stazioni.Id_Sensore=inquinanti.Idsensore AND inquinanti.Data =$day AND
        stazioni.Inquinante=$inquinante AND
        stazioni.Id_Stazione=latlngstazioni.Id_Stazione AND inquinanti.CodTS= $codts
```

Da questa query è possibile intuire che l'indicatore che viene estratto coincide con quello indicato nel paragrafo 5.3, con la numerazione A2, cioè *Concentrazione d'inquinamento media per fascia oraria per stazione*.

Dopo aver estratto questi dati, essi vengono restituiti dal server e sarà JavaScript tramite lo strumento delle mappe di Google a posizionare i marker sulle rispettive posizioni geografiche e inserire all'interno della finestra informativa di ogni marker, il nome della stazione e il valore di inquinamento individuato in quella data, in quella fascia oraria e per quello specifico inquinante. La mappa però, così non risulta ancora pronta per essere mostrata perché i marker devono essere colorati a seconda delle soglie di appartenenza mostrate nella legenda. Quindi, come fatto per la mappa di Daily Pollutant Concentration, quando vengono estratti i dati sulle concentrazioni medie per fasce orarie, è necessario interrogare il DW un'ulteriore volta, tramite una diversa query, che acceda alla tabella "soglie" ed estraiga solo quelle relative all'inquinante indicato dall'utente. La query appena descritta è così composta:

```
SELECT `Soglia1`,`Soglia2`,`Soglia3`,`Soglia4`
FROM `soglie`
WHERE Nome_Inquinante=$inquinante
```

Le soglie verranno anch'esse gestite da Javascript che assegnerà un colore ad ogni soglia(come mostrato nella legenda della mappa) e di conseguenza si controllerà in quale

soglia ricade la concentrazione d'inquinante misurata dalla stazione e si colorerà il relativo marker del colore di quella soglia individuata.

Dopo aver descritto la visualizzazione su mappa, si passa a mostrare delle informazioni su grafici tramite i Google Charts. In particolare, nella sezione Time Slot Pollutant Concentration verrà visualizzato un grafico che mostra i livelli di concentrazione dell'inquinante scelto dall'utente per ogni stazione nella fascia oraria scelta ma con un dettaglio sui valori delle singole ore di quel time slot.

Si può intuire che il KPI usato in questo caso non sarà più quello della Concentrazione media dell'inquinante per stazione e per fascia oraria ma sarà la Concentrazione oraria dell'inquinante per stazione, cioè quello indicato nel capitolo 5.3 con A3. Quanto appena descritto a parole, viene mostrato tramite la figura 5.40 in cui è rappresentato proprio l'istogramma in questione.

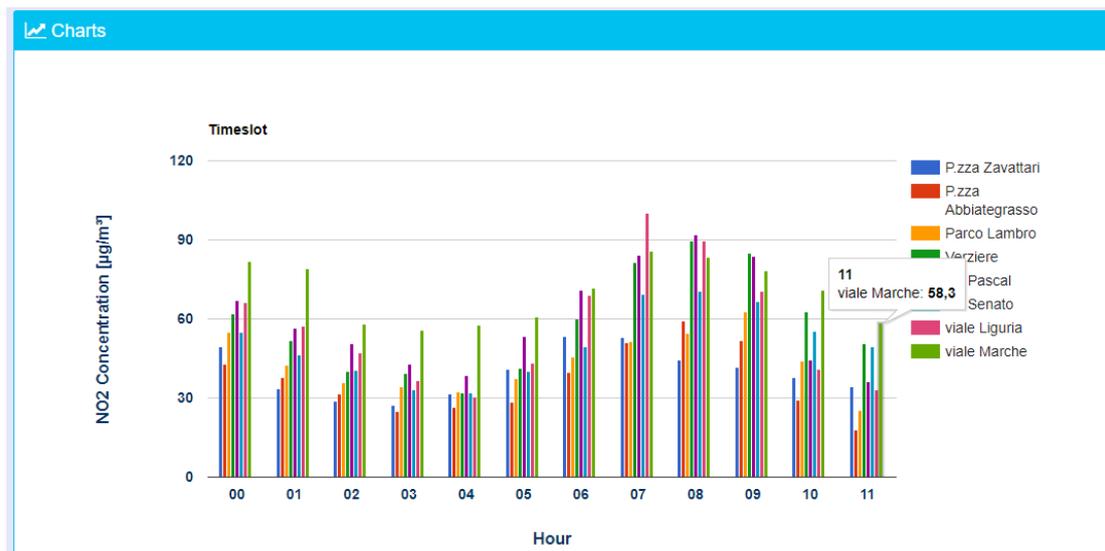


Figura 5.40: Timeslot

Questo grafico sarà quindi costituito da una serie di istogrammi (uno per ogni ora della fascia oraria selezionata), ognuno di essi sarà costituito da tante barre verticali, quante sono le stazioni di monitoraggio che misurano l'inquinante indicato dall'utente. L'appartenenza di una determinata rilevazione ad una stazione è facilmente riconoscibile tramite la differente colorazione delle barre in funzione proprio della stazione di monitoraggio.

Questa visualizzazione grafica è possibile grazie all'elaborazione di Javascript dei dati ricevuti dal server in seguito alla richiesta dell'intero report. Tale richiesta fatta al server grazie a php, contiene un'interrogazione sql per estrarre i dati dal DW; tale query è così composta:

```
SELECT stazioni.Nome_Stazione,inquinanti.Ora,inquinanti.Valore, inquinanti.CodTS
FROM inquinanti,stazioni
WHERE stazioni.Id_Sensore=inquinanti.Idsensore AND inquinanti.Data=$day AND
stazioni.Inquinante=$inquinante AND inquinanti.CodTS=$codts
GROUP BY stazioni.Nome_Stazione,inquinanti.Ora,inquinanti.Valore
```

Questa query estrae le informazioni così come sono necessarie all'istogramma prima descritto, cioè per ogni data nell'intervallo indicato, si individua il time slot indicato dall'utente e di conseguenza si estrae il nome della stazione di monitoraggio che ha effettuato la rilevazione e la relativa concentrazione oraria. Anche in questo caso è da notare che

l'interrogazione è fatta non usando la vista perché come già detto "vistagiornaliera" permette di avere solo una granularità giornaliera, invece in questo caso sono necessari i dati con massima granularità cioè senza calcolo di valori medi e tali dati hanno una granularità oraria.

3. Hourly Pollutant Concentration

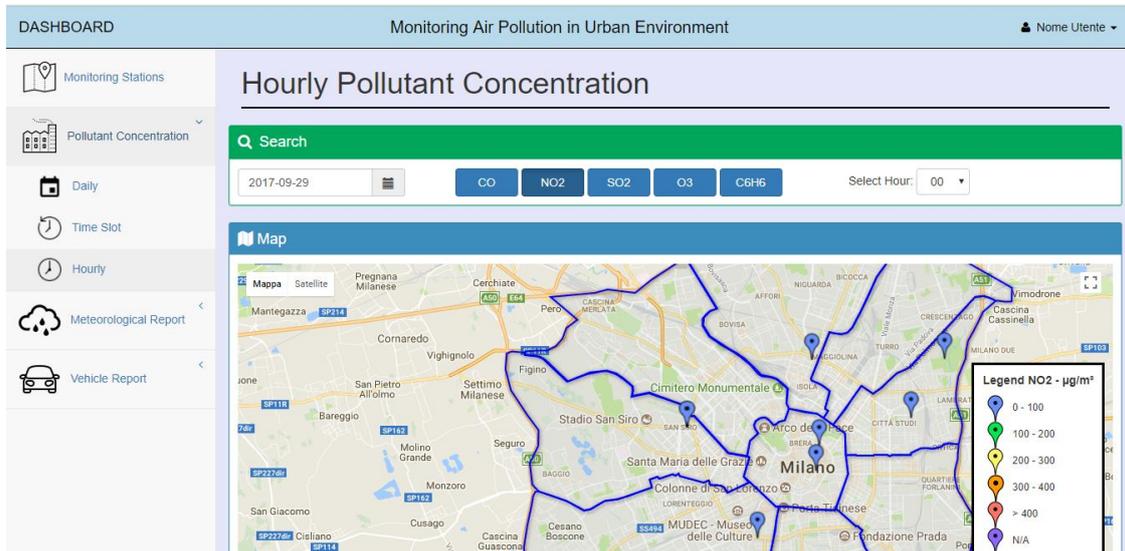


Figura 5.41: Hourly Pollutant Concentration

La sezione Hourly Pollutant Concentration permette all'utente di avere un report dettagliato e allo stesso tempo chiaro sulle concentrazioni di inquinanti misurate in una specifica data, con un focus su uno specifico orario (anch'esso scelto dall'utente).

Come per le sotto-sezioni di Pollutant Concentration descritte in precedenza, anche per questo report verranno usate mappe e istogrammi per visualizzare le informazioni. Come si può notare dalla figura 5.42 sottostante, l'utente ha tre parametri da scegliere per impostare l'analisi e di conseguenza per indicare quali dati estrarre dal DW. Di questi tre parametri, due riguardano appunto, il periodo temporale di riferimento del report (la data e l'ora), l'altro riguarderà l'inquinante del quale mostrare i livelli di inquinamento. La figura 5.42 mostra come si presenta la barra di Search al momento del caricamento della pagina, cioè con la selezione della data impostata sul giorno più recente presente nel DW (in questo caso è il 29 settembre 2017), con la selezione dell'ora impostata su 00 e sull'inquinante più importante e significativo per quanto riguarda l'inquinamento urbano (NO2).

Come per Time Slot Pollutant Concentration, anche in questa sezione si nota l'assenza degli inquinanti PM10 e PM2.5, sempre per lo stesso motivo, perché, come detto anche in precedenza, di questi sono disponibili solo rilevazioni giornaliera, quindi mostrare per essi informazioni di dettaglio oraria sarebbe impossibile.



Figura 5.42: Barra di ricerca in Hourly Pollutant Concentration

Una volta aver compreso quali sono i parametri scelti dall'utente, passiamo a descrivere le varie visualizzazioni mostrate nel report e quali dati vengono estratti dal DW e in che modo, per tali visualizzazioni.

Come già descritto per le sotto-sezioni precedenti, come si può notare dalla figura 5.43, la mappa contenuta in questo report mostra le stazioni di monitoraggio del comune di Milano che misurano l'inquinante selezionato e per ognuno dei marker posizionati sulla mappa è stato utilizzato un diverso colore a seconda della fascia di inquinamento in cui ricade la misurazione fatta di quel determinato inquinante in quella specifica data, in quella specifica ora e da quella specifica stazione.

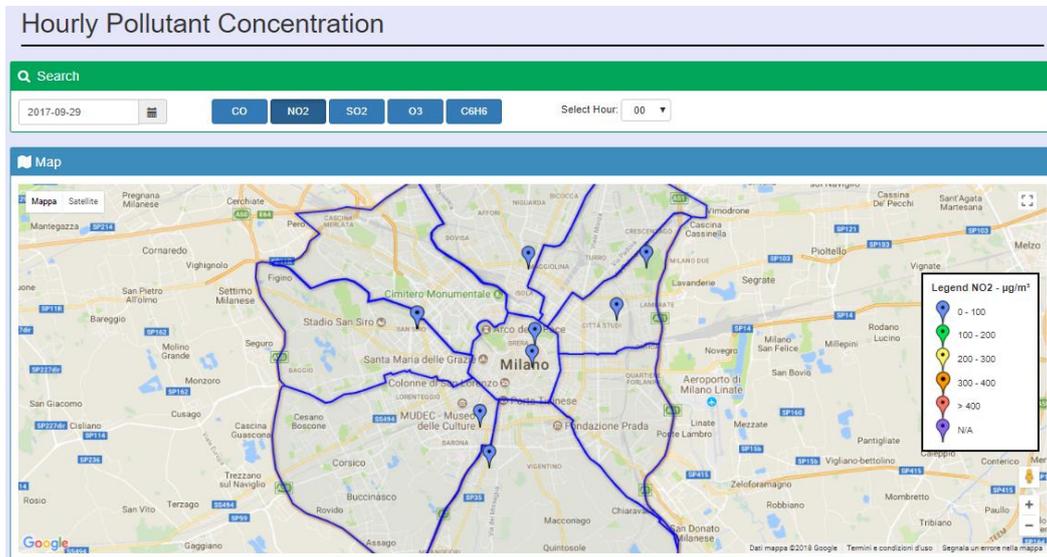


Figura 5.43: Mappa in Hourly Pollutant Concentration

Il funzionamento della mappa in figura 5.43 sarà il medesimo delle mappe delle sotto-sezioni precedenti, cioè sono mostrare le sole stazioni di monitoraggio dell'inquinante scelto, per ognuna di esse verrà posizionato un marker, il colore di tale marker è legato con la legenda mostrata a lato in cui sono mostrate le associazioni tra colori dei marker e fasce di valori delle concentrazioni dell'inquinante scelto. In questo caso, i valori che verranno considerati per l'assegnazione della soglia, e che verranno visualizzati al momento del click sul marker saranno quelli orari. Come descritto per le sezioni precedenti e come verrà fatto anche per le quelle successive, si va a dettagliare ora quali dati vengono estratti dal DW per visualizzare le informazioni sulla mappa e come tali dati vengano elaborati prima dell'effettiva visualizzazione.

La richiesta al server conterrà i tre parametri inseriti dall'utente: inquinante (\$inquinante), data(\$day), ora (\$mysql_time), questa richiesta permetterà di estrarre i dati necessari dal DW, in particolare, essendo necessari anche in questo caso dati con granularità maggiore di quella della vista, sarà necessari effettuare operazioni di join tra più tabelle per estrarre tutte le informazioni: le coordinate geografiche, il nome delle differenti stazioni e le misurazioni dell'inquinante indicato fatte da ognuna di tali stazioni nel periodo temporale specificato.

Quanto detto è possibile tramite l'utilizzo della query scritta in linguaggio sql, qui di seguito riportata:

```
SELECT latlngstazioni.Lat,latlngstazioni.Lng,stazioni.Nome_Stazione,inquinanti.Valore
FROM inquinanti,stazioni,latlngstazioni
WHERE stazioni.Id_Sensore=inquinanti.Idsensore AND inquinanti.Data =$day AND
stazioni.Inquinante=$inquinante AND
stazioni.Id_Stazione=latlngstazioni.Id_Stazione AND inquinanti.Ora=
$mysql_time
```

Da questa query è possibile intuire che l'indicatore che viene estratto coincide con quello indicato nel paragrafo 5.3, con la numerazione A3, cioè *Concentrazione d'inquinamento oraria per stazione*.

Dopo aver estratto questi dati, essi vengono restituiti dal server e sarà JavaScript tramite lo strumento delle mappe di Google a posizionare i marker sulle rispettive posizioni geografiche e inserire all'interno della finestra informativa di ogni marker, il nome della stazione e il valore di inquinamento individuato in quella data, in quell'orario e per quello specifico inquinante. La mappa però, così non risulta ancora pronta per essere mostrata perché i marker devono essere colorati a seconda delle soglie di appartenenza mostrate nella legenda.

Quindi, come fatto per le mappe precedenti, quando vengono estratti i dati sulle concentrazioni medie per fasce orarie, è necessario interrogare il DW un'ulteriore volta, tramite una diversa query, che acceda alla tabella "soglie" ed estragga solo quelle relative all'inquinante indicato dall'utente. La query appena descritta è così composta:

```
SELECT `Soglia1`, `Soglia2`, `Soglia3`, `Soglia4`
FROM `soglie`
WHERE Nome_Inquinante=$inquinante
```

Le soglie verranno anch'esse gestite da Javascript che assegnerà un colore ad ogni soglia(come mostrato nella legenda della mappa) e di conseguenza si controllerà in quale soglia ricade la concentrazione d'inquinante misurata dalla stazione e si colorerà il relativo marker del colore di quella soglia individuata.

Dopo aver descritto la visualizzazione su mappa, si passa a mostrare delle informazioni su grafici tramite i Google Charts. In particolare, nella sezione Hourly Pollutant Concentration verrà visualizzato un grafico che mostra i livelli di concentrazione dell'inquinante scelto dall'utente per ogni stazione per l'ora selezionata, sia nella data selezionata, sia nei 10 giorni precedenti. Questo permette di avere un trend degli andamenti di quell'inquinante in un preciso lasso di tempo di giorni mantenendo costante il confronto sullo stesso orario. Anche qui, si può intuire che il KPI usato è sempre quello indicato nel capitolo 5.3 con A3, a differenza della mappa però non si estraggono solo i valori della giornata selezionata, ma anche quelli dei giorni precedenti. Quanto appena descritto a parole, viene mostrato tramite la figura 5.44 in cui è rappresentato proprio l'istogramma in questione.

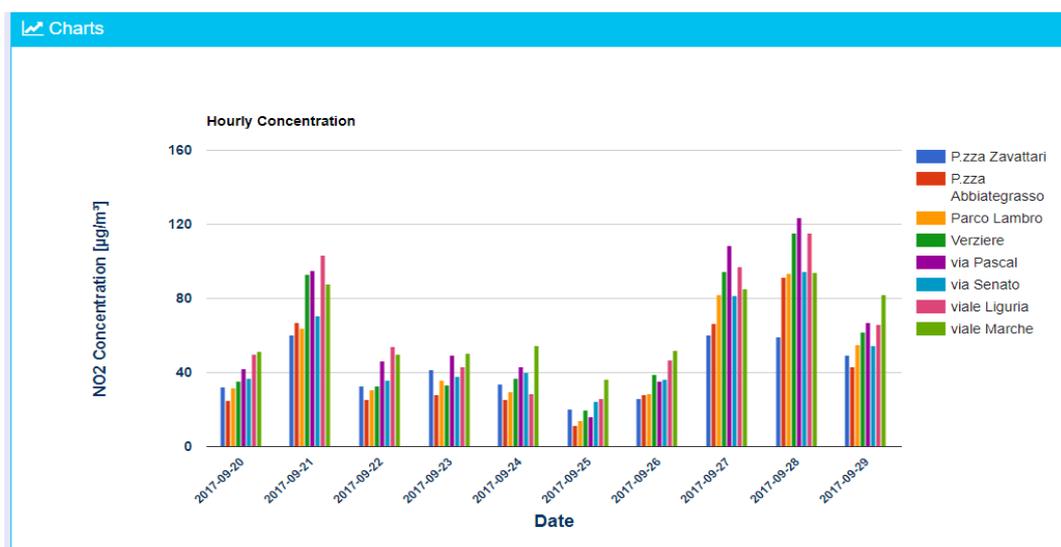


Figura 5.44: Hourly Concentration

Questo grafico sarà quindi costituito da una serie di istogrammi (uno per ogni giorno), ognuno di essi sarà costituito da tante barre verticali, quante sono le stazioni di monitoraggio che misurano l'inquinante indicato dall'utente nell'ora selezionata. L'appartenenza di una determinata rilevazione ad una stazione è facilmente riconoscibile tramite la differente colorazione delle barre in funzione proprio della stazione di monitoraggio.

Questa visualizzazione grafica è possibile grazie all'elaborazione di Javascript dei dati ricevuti dal server in seguito alla richiesta dell'intero report. Tale richiesta fatta al server grazie a php, contiene un'interrogazione sql per estrarre i dati dal DW; tale query è così composta:

```
SELECT stazioni.Nome_Stazione, inquinanti.Data, inquinanti.Valore
FROM inquinanti, stazioni
WHERE stazioni.Id_Sensore=inquinanti.Idsensore AND inquinanti.Data BETWEEN
    adddate($day,-9) and $day AND stazioni.Inquinante="'$inquinante.'" AND
    inquinanti.Ora=$mysql_time.
GROUP BY stazioni.Nome_Stazione, inquinanti.Data
```

Questa query estrae le informazioni così come sono necessarie all'istogramma prima descritto, cioè per ogni data nell'intervallo indicato, si individua l'ora indicata dall'utente e di conseguenza si estrae il nome della stazione di monitoraggio che ha effettuato la rilevazione e la relativa concentrazione oraria. Anche in questo caso è da notare che l'interrogazione è fatta non usando la vista perché come già detto "vistagiornaliera" permette di avere solo una granularità giornaliera, invece anche in questo caso come in quello di Time Slot Pollutant Concentration, sono necessari i dati con massima granularità cioè senza calcolo di valori medi e tali dati hanno una granularità oraria. Milano.

5.6.2.3 Meteorological Report

Questa sezione denominata Meteorological Report, si occupa di fornire un report in cui vengono confrontati con dapprima analisi separate e poi con una vera e propria correlazione, il legame tra i livelli di inquinamento urbano dovuti ai differenti inquinanti e i principali fenomeni meteorologici, come le precipitazioni, la temperatura, l'umidità, ecc.

Il Meteorological Report è composta a sua volta da due sotto-sezioni: Daily Range e Monthly Range. Come indicano già i due nomi, la differenza sostanziale di queste due sotto-sezioni è la granularità temporale sul quale sono effettuate le analisi e di conseguenza su cui si basa il report che verrà presentato. Di conseguenza, per evitare di ripetere le descrizioni delle varie parti che compongono questo report, si è deciso qui di descrivere questa differenza tra le due sotto-sezioni, per poi andare successivamente a descrivere nel dettaglio solo il report di Daily Range, ricordando ancora una volta che Monthly Range fornisce un report praticamente identico però dando la possibilità all'utente di selezionare un intervallo temporale diverso.

In particolare, in Daily Range si può scegliere un intervallo di giorni sul quale effettuare l'analisi; invece, in Monthly Range la selezione poteva essere fatta su un intervallo di mesi. Prima di passare all'effettiva descrizione di questo report è bene indicare quali indicatori sono stati usati: in Daily Range basandosi su intervalli giornalieri, verranno usati KPI che nel paragrafo 5.3 era identificati con A1 (Concentrazione media giornaliera dell'inquinante per stazione), B1 (Valori medi giornalieri per i diversi fenomeni meteorologici); invece, in Monthly Range basandosi su intervalli mensili, verranno usati KPI che nel paragrafo 5.3 era identificati con A4 (Concentrazione media mensile dell'inquinante per stazione), B2 (Valori medi mensili per i diversi fenomeni meteorologici). In entrambi sarà utilizzato il KPI che, sempre nel paragrafo , è identificato con D (KPI per l'analisi statistica della correlazione).

In figura 5.45, è mostrata la pagina del Meteorological Report e in particolare, quella di Daily Range.

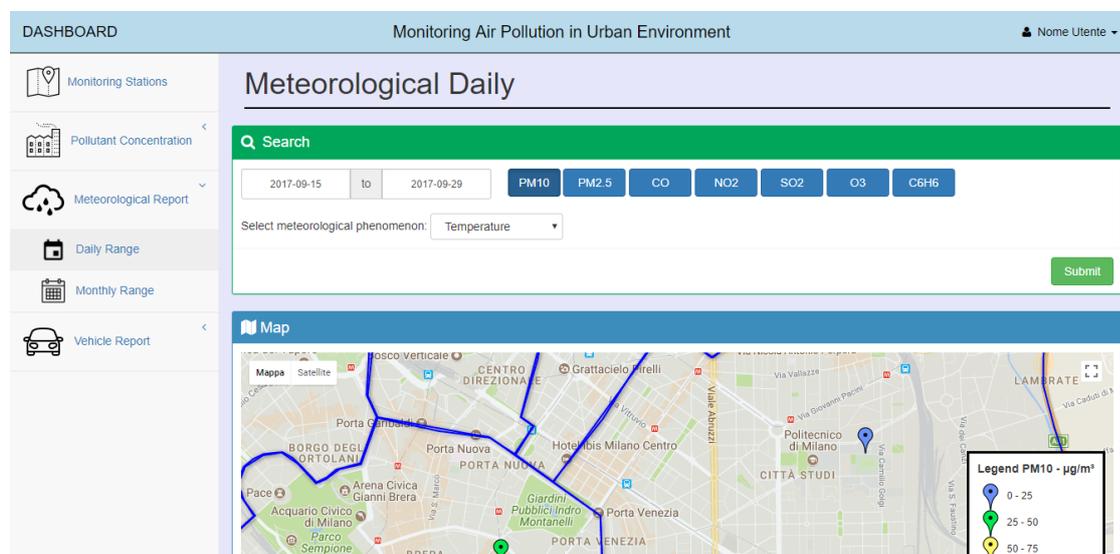


Figura 5.45: Meteorological Report

Come per la sezione Pollutant Concentration, anche in questa sezione le informazioni verranno mostrate in due forme: tramite una mappa e tramite visualizzazione grafica con istogrammi. Quindi, come si può notare dalla figura 5.46 sottostante, l'utente ha tre parametri da scegliere per impostare l'analisi e di conseguenza per indicare quali dati estrarre dal DW.

L'utente ha la possibilità di impostare il periodo temporale di riferimento del report cioè un intervallo temporale di giorni, l'inquinante del quale mostrare i livelli di inquinamento e soprattutto avrà la possibilità di scegliere tra una lista di fenomeni meteorologici, per il quale ha interesse ad effettuare un confronto con i livelli di inquinamento.

La figura 5.46 mostra come si presenta la barra di Search al momento del caricamento della pagina, cioè con la selezione dell'intervallo di giorni impostata sulla settimana più recente presente nel DW (in questo caso è dal 15 settembre 2017 al 29 settembre 2017), sull'inquinante più importante e significativo per quanto riguarda l'inquinamento urbano (PM10) e sul primo fenomeno meteorologico inserito nella lista.



Figura 5.46: Barra di ricerca in Meteorological Report

Nel caso in cui l'utente volesse consultare le informazioni riguardanti le concentrazioni di inquinamento dovute al PM10 ma di un intervallo di giorni differente da quello più recente, sarà possibile farlo tramite il widget Calendar, oppure se l'utente volesse cambiare il fenomeno meteorologico da analizzare, può farlo tramite il menu a tendina fornito. Entrambe queste funzionalità sono mostrate in figura 5.47

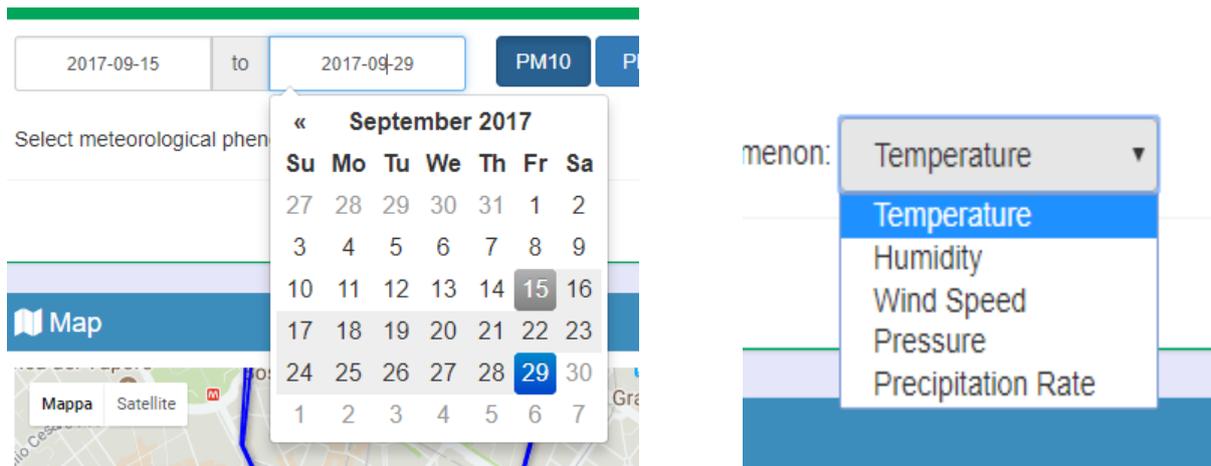


Figura 5.47: Widget Calendar & Menù

Una volta aver compreso quali sono i parametri scelti dall'utente, possiamo a descrivere le varie visualizzazioni mostrate nel report e quali dati vengono estratti dal DW e in che modo, per tali visualizzazioni.

Come si può notare dalla figura 5.48, la mappa contenuta in questo report mostra le stazioni di monitoraggio del comune di Milano che misurano l'inquinante selezionato e per ognuno dei marker posizionati sulla mappa è stato utilizzato un diverso colore a seconda della fascia di inquinamento in cui ricade la misurazione fatta di quel determinato inquinante in quell'intervallo di tempo da quella specifica stazione. Naturalmente, parlando di intervallo di giorni e non di singola data, il valore che ogni stazione prenderà come riferimento per l'inquinante in quel periodo temporale sarà la media dei valori giornalieri misurati dalla stazione nei giorni indicati.

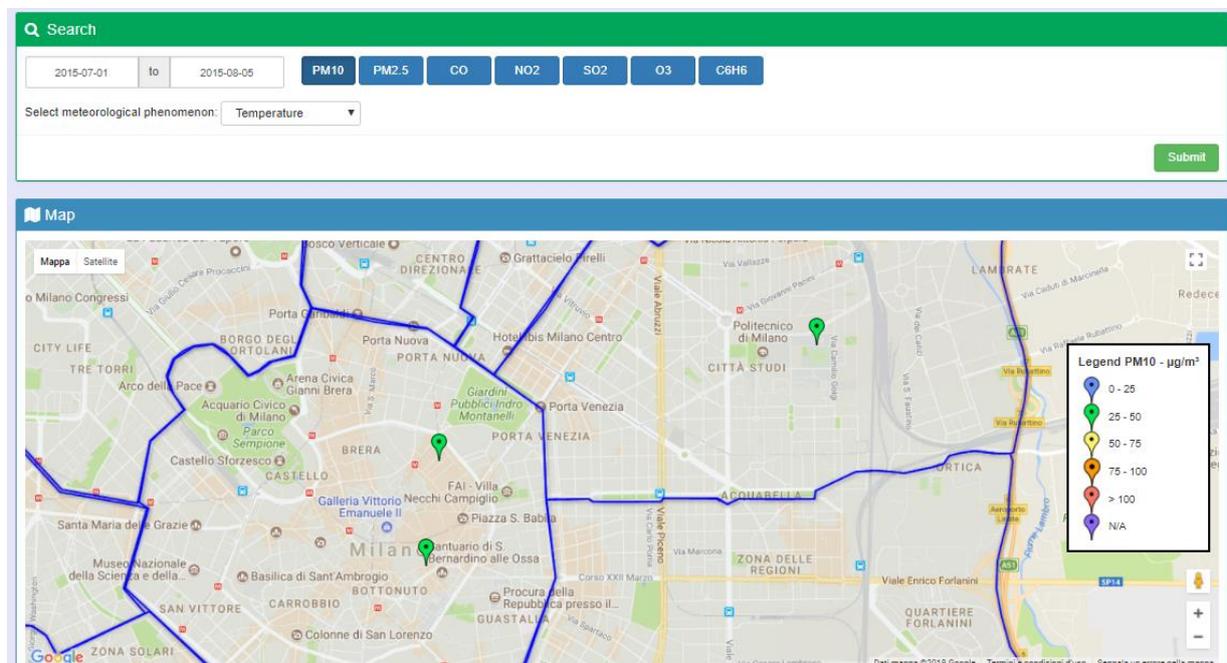


Figura 5.48: Mappa in Metereological Report

Nella figura 5.48 sono appunto mostrate le tre stazioni di monitoraggio dell'inquinante PM10 e tutte sono colorate di verde, tale colore è legato con la legenda mostrata a lato in cui sono mostrate le associazioni tra colori dei marker e fasce di valori delle concentrazioni

dell'inquinante scelto. Nel caso in questione risulta che le concentrazioni medie di PM10 per i giorni nell'intervallo indicato hanno fornito tutte valori compresi nella fascia 25-50 $\mu\text{g}/\text{m}^3$.

Per far sì di conoscere il preciso valore di concentrazione media dell'inquinante nel periodo specificato, si è fornita all'utente la possibilità di cliccare sui singoli marker per visualizzare tale valore, come mostrato in figura 5.49. Tale figura, mostra anche una informazione aggiunta riguardante il fenomeno meteorologico scelto, cioè il valore medio nell'intervallo scelto, delle rilevazioni fatte nella città di Milano sul fenomeno.

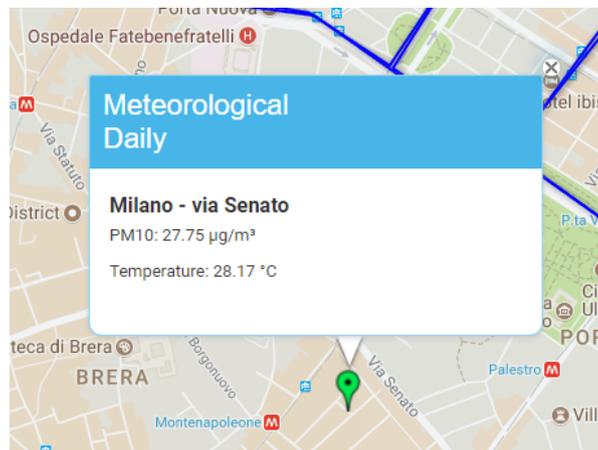


Figura 5.49: Informazioni su mappa in Meterological Report

Si è visto quindi che le informazioni principali gestite fin'ora da questa sezione Daily Range sono state la media per il numero di giorni scelto, della concentrazione media giornaliera dell'inquinante e dei valori medi del fenomeno meteorologico. Si va a dettagliare ora quali dati vengono estratti dal DW per visualizzare le informazioni sulla mappa e come tali dati vengano elaborati prima dell'effettiva visualizzazione.

Al momento della selezione dell'inquinante ($\$inquinante$), del periodo temporale data($\$inizio, \$fine$) e del fenomeno meteorologico ($\$meteo$) da parte dell'utente, verrà inviata una richiesta al server contenente questi parametri, questa richiesta permetterà di estrarre i dati necessari dal DW, in particolare dalla vista creata durante il progetto "vistagiornaliera" e dalla tabella "meteo". Come già visto, tale vista permette di estrarre informazioni con granularità giornaliera anziché oraria (come è nella tabella "inquinanti") riguardanti le concentrazioni misurate da ogni stazione per ogni inquinante. Nel caso in questione, da tale vista verranno estratte le coordinate geografiche, il nome delle differenti stazioni e le misurazioni dell'inquinante indicato fatte da ognuna di tali stazioni. Quanto detto è possibile tramite l'utilizzo di due query separate, scritte in linguaggio sql, qui di seguito riportate:

1. **SELECT** vistagiornaliera.Lat,vistagiornaliera.Lng,vistagiornaliera.Nome_Stazione,
AVG(vistagiornaliera.Media_Giornaliera) as Media_IntGiorni
FROM vistagiornaliera
WHERE vistagiornaliera.Data **BETWEEN** \$inizio **and** \$fine **AND**
vistagiornaliera.Inquinante=\$inquinante
GROUP BY vistagiornaliera.Nome_Stazione
2. **SELECT** AVG(meteo.`temperature`) AS Media_T, AVG(meteo.`humidity`) AS
Media_H, AVG(meteo.`wind_speed`) AS Media_WS,
AVG(meteo.`pressure`) AS Media_P, AVG(meteo.`precipitation_rate`) AS
Media_PR

```
FROM meteo
WHERE meteo.Data BETWEEN $inizio and $fine AND
meteo.`Nome_Stazione`='Milano - via Senato'
```

Dalla prima query quindi, verranno estratti i valori medi per stazione basati sulle concentrazioni medie dell'inquinante nel periodo selezionato. Nella seconda query invece, verranno estratti i valori medi delle rilevazioni giornaliere nel periodo selezionato, ma di ogni fenomeno meteorologico.

La scelta di prendere le misurazioni meteorologiche solo della stazione di monitoraggio di Via Senato (centro di Milano) è stata presa perché ci si è resi conto che tutti i valori giornalieri rilevati dalle differenti stazioni per un singolo inquinante, differivano di pochi centesimi, facendo diventare superfluo un ulteriore calcolo di media su tali rilevazioni, portano maggiormente ad optare per la scelta di una stazione di riferimento per i dati meteorologici.

Dopo aver estratto questi dati, essi vengono restituiti dal server e sarà JavaScript tramite lo strumento delle mappe di Google a posizionare i marker sulle rispettive posizioni geografiche e inserire all'interno della finestra informativa di ogni marker, il nome della stazione e il valore di inquinamento individuato in quella data per quello specifico inquinante. Tramite codice Javascript verrà anche individuato quale fenomeno meteorologico è di interesse, e tale valore verrà inserito nella finestra di informazioni di ogni singolo marker. Come fatto anche in passato, la mappa non risultando ancora pronta per essere mostrata dato che i marker devono essere colorati a seconda delle soglie di appartenenza mostrate nella legenda; si è deciso, di interrogare il DW un'ulteriore volta, tramite una diversa query, che acceda alla tabella "soglie" ed estraiga solo quelle relative all'inquinante indicato dall'utente. La query appena descritta è così composta:

```
SELECT `Soglia1`, `Soglia2`, `Soglia3`, `Soglia4`
FROM `soglie`
WHERE Nome_Inquinante=$inquinante
```

Le soglie verranno anch'esse gestite da Javascript che assegnerà un colore ad ogni soglia (come mostrato nella legenda della mappa) e di conseguenza si controllerà in quale soglia ricade la concentrazione d'inquinante misurata dalla stazione e si colorerà il relativo marker del colore di quella soglia individuata.

Dopo aver descritto la visualizzazione su mappa, si passa a mostrare delle informazioni su grafici tramite i Google Charts. In particolare, nella sezione Daily Range verranno visualizzati tre differenti grafici: il primo basato sui livelli di inquinamento dell'inquinante nel periodo temporale indicato, il secondo sulle rilevazioni meteo riguardanti il fenomeno meteorologico scelto nel medesimo periodo temporale e infine un ultimo grafico di correlazione basato sia sui dati dell'inquinamento sia su quelli del fenomeno meteorologico.

Nel primo grafico, quello sulle concentrazioni di inquinante, vengono mostrati nell'intervallo di giorni indicati dall'utente, i valori di concentrazione media giornaliera per l'inquinante selezionato, per ogni stazione. Questo permette di avere un trend degli andamenti di quell'inquinante in un preciso lasso di tempo. Infatti, il grafico è rappresentato da una linea per ogni stazione di monitoraggio dell'inquinante (le stesse mostrate sulla mappa) e per ogni giorno presente nell'intervallo, il relativo valore di concentrazione media. Si può intuire facilmente che il KPI usato è sempre quello indicato nel capitolo 5.3 con A1. In figura 5.50, il grafico a linee appena descritto.

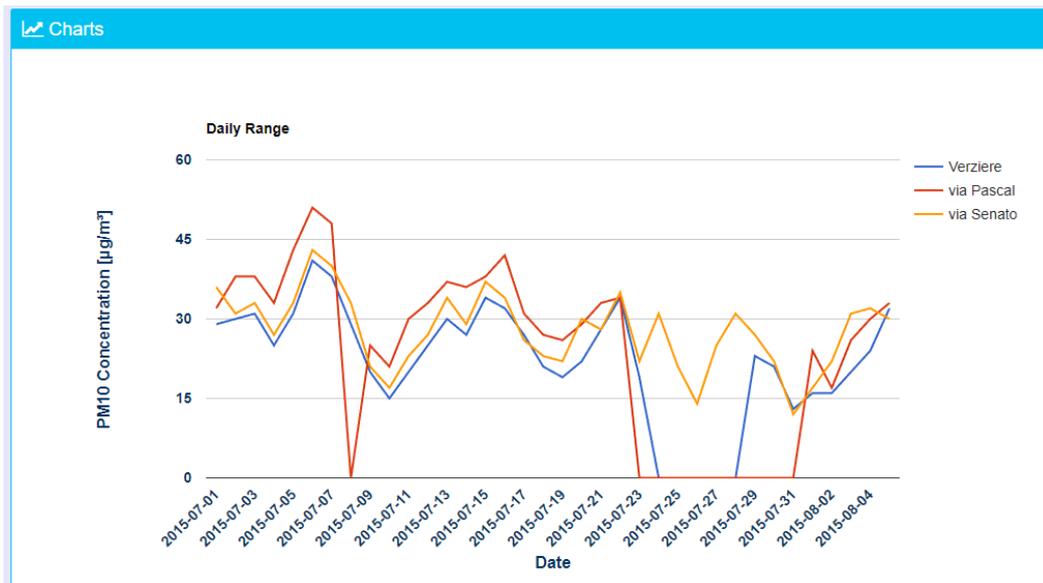


Figura 5.50: Daily Range

Questa visualizzazione grafica è possibile grazie all'elaborazione di Javascript dei dati ricevuti dal server in seguito alla richiesta dell'intero report. Tale richiesta fatta al server grazie a php, contiene un'interrogazione sql per estrarre i dati dal DW; tale query è così composta:

```
SELECT vistagiornaliera.Nome_Stazione,vistagiornaliera.Data,
        vistagiornaliera.Media_Giornaliera
FROM vistagiornaliera
WHERE vistagiornaliera.Data BETWEEN $inizio and $fine AND
        vistagiornaliera.Inquinante= $inquinante
GROUP BY vistagiornaliera.Nome_Stazione,vistagiornaliera.Data
```

Questa query estrae le informazioni così come sono necessarie al grafico a linee prima descritto, cioè per ogni data nell'intervallo indicato, estrae il nome della stazione di monitoraggio che ha effettuato la rilevazione e la relativa concentrazione media giornaliera. Anche in questo caso è da notare che l'interrogazione è fatta solo sulla vista "vistagiornaliera" che come già detto permette di avere le concentrazioni medie giornaliere già calcolate.

Il secondo grafico presente nel Meteorological Report è quello relativo ai dati del fenomeno meteorologico scelto, esso mostra le differenti rilevazioni giornaliere della stazione di monitoraggio del centro di Milano, nell'intervallo di giorni selezionato dall'utente. In questo caso, il KPI su cui ci si basa è quello indicato nel paragrafo 5.3 con B1: *Valori medi giornalieri per i diversi fenomeni meteorologici*. Tale grafico è mostrato in figura 5.51.

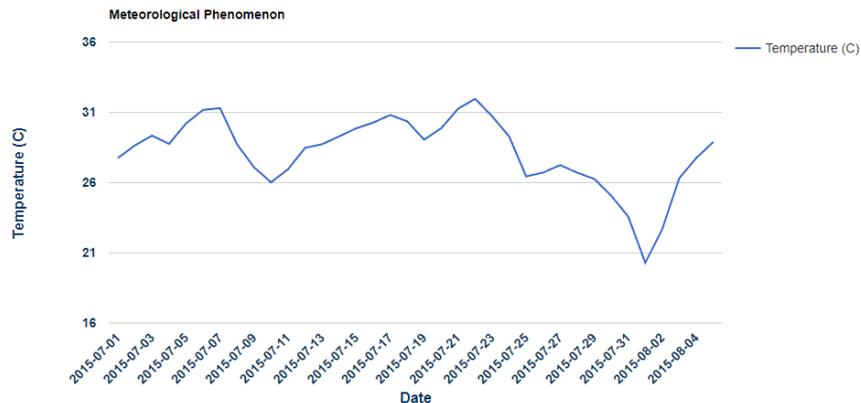


Figura 5.51: Meteorological Phenomenon

In questo grafico, ci sarà una sola linea, perché come detto in precedenza, le rilevazioni meteorologiche nella medesima giornata fatte dalle differenti stazioni, variavano di pochi centesimi, quindi si è valutata corretta la scelta di mostrare i dati di una singola stazione, quella nel centro di milano (Milano – Via Veneto). In questo caso, i dati che gestirà Javascript per creare il grafico con lo strumento di Google Charts sono estratti dal DW tramite richiesta al server, effettuando una interrogazione sql sulla tabella “meteo”. La query sarà così fatta:

```
SELECT meteo.Data,meteo.`temperature` AS Media_T, meteo.`humidity` AS Media_H,
        meteo.`wind_speed` AS Media_WS, meteo.`pressure` AS Media_P,
        meteo.`precipitation_rate` AS Media_PR
FROM meteo
WHERE meteo.Data BETWEEN $inizio and $fine AND
        meteo.`Nome_Stazione`='Milano - via Senato'
GROUP BY meteo.Data
```

Anche in questo caso, come per la mappa, dalla tabella sono state estratte le rilevazioni riguardanti tutti i fenomeni meteorologici analizzabili e solo successivamente tramite codice php, saranno presi i valori nell’intervallo temporale del solo fenomeno selezionato dall’utente.

Infine, per concludere l’analisi della sezione Meteorological Report, si descrive ora l’ultimo grafico presente nel report: quello di correlazione, basato sul calcolo del KPI denominato con D(KPI per l’analisi statistica della correlazione) nel paragrafo 5.3. Come visto, nel paragrafo appena citato, tale KPI a differenza degli altri indicatori utilizzati, non viene direttamente estratto dal DW tramite interrogazione sql ma viene calcolato con codice php partendo da dati già estratti. In particolare, questo grafico di correlazione, si occupa di trovare un indicatore numerico e di mostrare graficamente, il legame che c’è tra i livelli di concentrazione medi giornalieri dell’inquinante scelto dall’utente e i valori giornalieri del fenomeno meteorologico, anch’esso selezionato dall’utente.

Questa correlazione viene fatta nel periodo temporale indicato, quindi per l’intervallo di giorni scelto e serve per capire se c’è un legame nell’andamento temporale dei due grafici già mostrati (quello delle concentrazioni di inquinamento e quello del fenomeno meteorologico).

Come descritto nel paragrafo 5.3, l’indicatore di correlazione rappresenta la Correlazione di Pearson, che richiede l’utilizzo di due serie numeriche per calcolare il valore numerico di correlazione. Queste due serie numeriche, sono le stesse che verranno utilizzato in quest’ultimo grafico, che è diverso da tutti quelli precedentemente descritti, perché esso è un grafico plot.

Nel caso in questione, tale grafico si occupa di rappresentare sul piano cartesiano i punti che hanno come coordinate il livello di concentrazione medio giornaliero dell'inquinante calcolato come media tra tutte le stazioni e il valore giornaliero del fenomeno meteorologico.

Una volta riportati tutti i punti, per i differenti giorni nell'intervallo temporale scelto, verrà tracciata automaticamente la retta che può correla tali punti. Più tale retta interpola questi punti, più concentrazione d'inquinante e fenomeno meteorologico sono correlate. Il grafico appena descritto è mostrato in figura 5.52

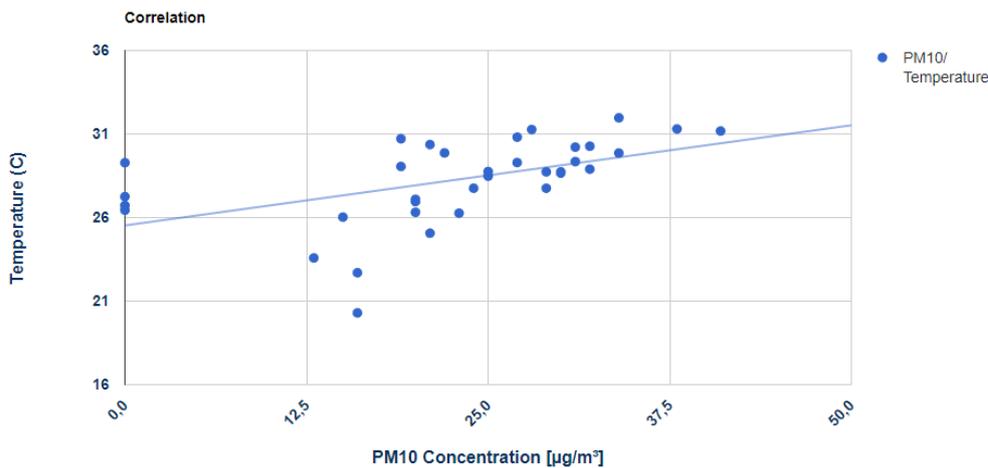


Figura 5.52: Correlation

Come detto, le due serie numeriche messe a confronto per la correlazione, riguardano i valori di inquinamento e fenomeno meteorologico rilevati giornalmente durante il periodo temporale scelto dall'utente. I valori del fenomeno meteorologico sono gli stessi estratti dal DW per il grafico di figura x.z, invece quelli della concentrazione d'inquinante, sono una media giornaliera tra i valori estratti per il grafico di figura x.w, in cui le rilevazioni giornaliere erano suddivise per stazione di monitoraggio. Di conseguenza per estrarre dal DW i valori d'inquinamento utili per il grafico 5.52 e per il calcolo numerico della correlazione, sarà necessario utilizzare una interrogazione sql scritta come nel seguente modo:

```
SELECT vistagiornaliera.Data,AVG(vistagiornaliera.Media_Giornaliera)
FROM vistagiornaliera
WHERE vistagiornaliera.Inquinante= $inquinante AND vistagiornaliera.Data
BETWEEN $inizio and $fine
GROUP BY vistagiornaliera.Data
```

Da questa query verranno estratti i valori medi giornalieri delle concentrazioni di inquinante di tutta la città di Milano. Fatto questo, Javascript potrà gestire tali dati e quelli dell'inquinamento per creare e visualizzare il grafico plot tramite lo strumento Google Charts. Come detto in precedenza però, l'indicatore KPI di quest'ultimo grafico, può essere quantificato numericamente tramite il calcolo del coefficiente di Pearson(il calcolo è mostrato nel paragrafo 5.5). Tale valore come già detto è compreso tra [-1,1], con gli estremi indici di massima correlazione (rispettivamente negativa e positiva), mentre con lo 0 si indica assenza completa di correlazione tra concentrazioni dell'inquinante e fenomeno atmosferico.

Sul report fornito all'utente, con i grafici e la mappa descritti in questa sezione, verrà indicato anche il valore di tale coefficiente di correlazione.

5.6.2.4 Vehicle Report

L'ultima sezione della dashboard messa a disposizione dell'utente, è denominata Vehicle Report, si occupa di fornire un report in cui vengono analizzati separatamente i livelli di inquinamento urbano dovuti ai differenti inquinanti e un'altra delle cause principali dell'inquinamento: il traffico urbano, entrambi valutati nel medesimo lasso temporale (intervallo di giorni). E' da specificare sin da subito, che per tale report, sono stati utilizzati i dati presenti nel DW riguardanti il traffico e come detto prima, questi sono relativi all'ingressi nel centro di Milano (la zona più trafficata).

Il Vehicle Report è composto a sua volta da tre sotto-sezioni: Type, Fuel e Euro. Come indicano già i nomi, la differenza sostanziale di queste sotto-sezioni è il criterio con il quale verranno classificati i vari veicolo nel centro di Milano. Di conseguenza, per evitare di ripetere le descrizioni delle varie parti che compongono questo report, si è deciso qui di descrivere questa differenza tra le tre sotto-sezioni, per poi andare successivamente a descrivere nel dettaglio solo il report che fornisce la classificazione per tipo di veicolo (Type) , ricordando ancora una volta che Fuel e Euro forniscono un report praticamente identico però classificando i veicoli rispettivamente, in base al tipo di carburante usato e in base alla categoria Euro di appartenenza del veicolo.

Quindi riassumendo, l'unica differenza tra le tre sottosezioni è che la classificazione dei veicoli è fatta secondo criteri differenti; le singole categorie facenti parte delle diverse classificazioni sono:

- TYPE → Persone, Merci, Bus, Altro, NC
- FUEL → Benzina, Diesel, Elettrico, GPL Metano, Ibrido, GPL, Miscela, Ibrido Gasolio, Metano, Ibrido Benzina, NC
- EURO → Euro 0, Euro 1, Euro 2, Euro 3, Euro 4, Euro 5, Euro 6, NC

Prima di passare all'effettiva descrizione di questo report è bene indicare quali indicatori sono stati usati: in Type, verrà usato il KPI che nel paragrafo 5.3 era identificato con C1 (Valori totali giornalieri per tipologia di veicolo), in Fuel verrà usato il KPI che nel paragrafo 5.3 era identificato con C2 (Valori totali giornalieri per tipologia di carburante) e in Euro verrà usato il KPI che nel paragrafo 5.3 era identificato con C3 (Valori totali giornalieri per categoria di Euro). In tutte e tre le sottosezioni sarà utilizzato il KPI che, sempre nel paragrafo 5.3, è identificato con A1 (Concentrazione media giornaliera degli inquinanti per stazione).

In figura 5.53, è mostrata la pagina del Vehicle Report e in particolare, come detto, quella di Type.

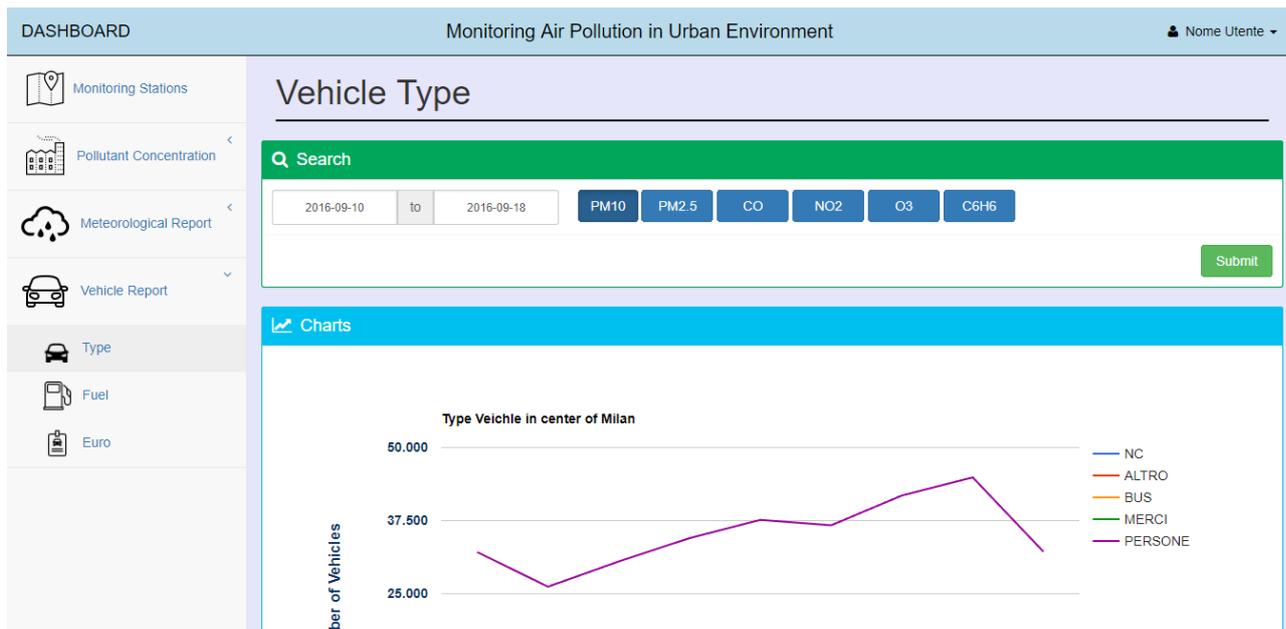


Figura 5.53: Vehicle Report

A differenza di tutte le sezioni descritte precedentemente per lo sviluppo della dashboard, in questa sezione l'intero report si baserà solo sulla visualizzazione grafica con grafici a linee, perché si è ritenuta sufficientemente chiara e completa un'analisi basata su dati mostrati su grafici e non su mappa. Come si può notare dalla figura 5.54 sottostante, l'utente ha due parametri da scegliere per impostare l'analisi e di conseguenza per indicare quali dati estrarre dal DW. L'utente ha la possibilità di impostare il periodo temporale di riferimento del report cioè un intervallo temporale di giorni (\$inizio, \$fine) e l'inquinante (\$inquinante) del quale mostrare i livelli di inquinamento (e quindi da voler confrontare con i dati del traffico).

La figura 5.54 mostra come si presenta la barra di Search al momento del caricamento della pagina, cioè con la selezione dell'intervallo di giorni impostata sulla settimana più recente presente nel DW (in questo caso è dal 10 settembre 2016 al 18 settembre 2016) e sull'inquinante più importante e significativo per quanto riguarda l'inquinamento urbano (PM10).



Figura 5.54: barra di ricerca in Vehicle Report

Nel caso in cui l'utente volesse consultare le informazioni riguardanti le concentrazioni di inquinamento dovute al PM10 ma di un intervallo di giorni differente da quello più recente, sarà possibile farlo tramite il widget Calendar. Come mostrato per la medesima funzionalità per il Meteorological Report in figura 5.45. Una volta aver compreso quali sono i parametri scelti dall'utente, passiamo a descrivere i due grafici visualizzati nel report tramite lo strumento Google Charts e quali dati vengono estratti dal DW e in che modo.

Il primo grafico, è un grafico a linee e mostra informazioni riguardanti la numerosità di veicoli entranti nel centro di Milano in un intervallo di tempo specificato dall'utente, distinguendo per tipologia di veicolo. Infatti, il grafico è rappresentato da una linea per tipologia e per ogni giorno presente nell'intervallo, il relativo numero di veicoli di quella

tipologia. Si può intuire facilmente che il KPI usato è sempre quello indicato nel capitolo 5.3 con C1. In figura 5.55, il grafico a linee appena descritto.

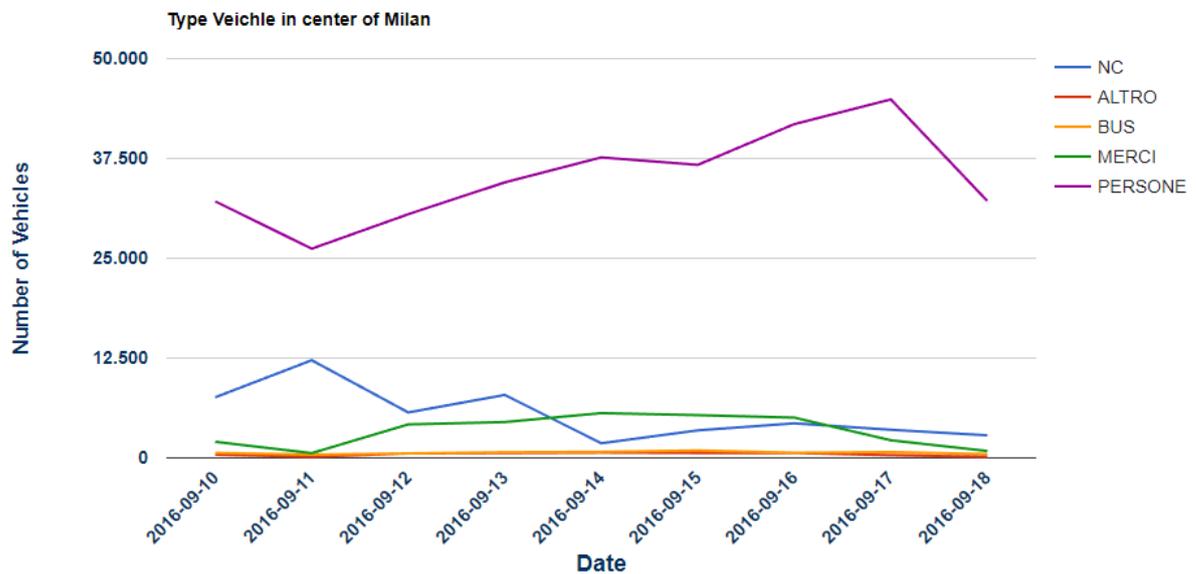


Figura 5.55: Type Vehicle in center of Milan

Questa visualizzazione grafica è possibile grazie all'elaborazione di Javascript dei dati ricevuti dal server in seguito alla richiesta dell'intero report. Tale richiesta fatta al server grazie a php, contiene un'interrogazione sql per estrarre i dati dal DW; tale query è molto semplice ed è così composta:

```
SELECT Data, 'NC', 'ALTRO', 'BUS', 'MERCI', 'PERSONE'
FROM catveicolo
WHERE Data BETWEEN $inizio and $fine
```

Questa query estrae le informazioni così come sono necessarie al grafico a linee prima descritto, cioè per ogni data nell'intervallo indicato, estrae i valori giornalieri già memorizzati nella tabella, suddivisi per tipologia.

Il secondo grafico presente nel Vehicle Report è quello relativo ai dati dell'inquinante; è da sottolineare che tale grafico è identico per tutte e tre le sotto-sezioni (Type, Fuel, Euro). Esso mostra le differenti rilevazioni giornaliere delle stazioni di monitoraggio del centro di Milano, nell'intervallo di giorni selezionato dall'utente. In questo caso, il KPI su cui ci si basa è quello indicato nel paragrafo 5.3 con A1: Concentrazione media giornaliera degli inquinanti per stazione. Tale grafico è mostrato in figura 5.56.

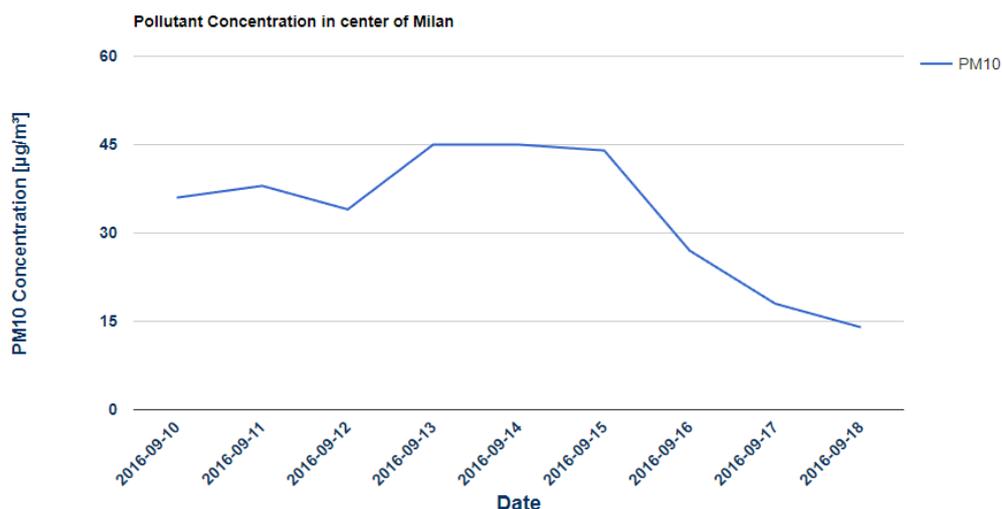


Figura 5.56: Pollutant Concentration in center of Milan

In questo grafico, ci sarà una sola linea, perché i dati riguardanti il traffico, come detto, solo relativi al centro di Milano, quindi per facilitare e soprattutto rendere più veritiero un confronto tra i due grafici, si è deciso di prendere i livelli giornalieri di inquinante dalle due stazioni nel centro di milano(Milano – Via Veneto oppure Milano – Via Verziere), scegliendo una o l'altra stazione a seconda dell'inquinante d'interesse selezionato dall'utente, questo perché queste due stazioni monitorano inquinanti differenti. In questo caso, i dati che gestirà Javascript per creare il grafico con lo strumento di Google Charts sono estratti dal DW tramite richiesta al server, effettuando una interrogazione sql sulla vista "vistagiornaliera". La query sarà così fatta:

```
SELECT vistagiornaliera.Inquinante,vistagiornaliera.Nome_Stazione,
vistagiornaliera.Data,vistagiornaliera.Media_Giornaliera
FROM vistagiornaliera
WHERE vistagiornaliera.Data BETWEEN $inizio and $fine AND
vistagiornaliera.Inquinante=$inquinante AND
vistagiornaliera.Nome_Stazione='Milano - Verziere'
```

5.7 Confronto delle tecnologie per lo sviluppo della Dashboard

Nel paragrafo 4.2, in occasione della descrizione delle tecnologie a disposizione per la realizzazione dei reports della dashboard, si è visto che per la gestione dei dati la scelta era tra MySQL/PHP e Google Data Studio. Senza descrivere nuovamente le due possibilità, si ricorda solamente il perché la scelta sia ricaduta sull'uso di MySQL,PHP: questi, a differenza di Google Data Studio, permettono la gestione delle relazioni tra tabelle diverse, permettono di effettuare operazioni sui dati estratti prima di mostrarli e hanno un focus maggiore per quanto riguarda le mappe. Detto questo, si effettua ora un confronto tra due grafici (un'istogramma e un grafico a linee) presenti nella dashboard appena sviluppata con MySQL, PHP e visualizzati con Google Charts e gli stessi grafici realizzati con il solo strumento Google Data Studio. Tale confronto verrà effettuato con gli stessi parametri inseriti dall'utente.

In figura 5.57 sono riportati i grafici presi dalla dashboard realizzata, riguardanti il trend degli ultimi 10 giorni della concentrazione di inquinante differenziato per stazione di monitoraggio e la numerosità di veicoli differenziati per tipologia, entrati nel centro di Milano in un certo intervallo di giorni. I parametri per questi grafici sono stati inseriti dall'utente

tramite le barre di ricerca descritte nelle relative sezioni dello sviluppo della dashboard.

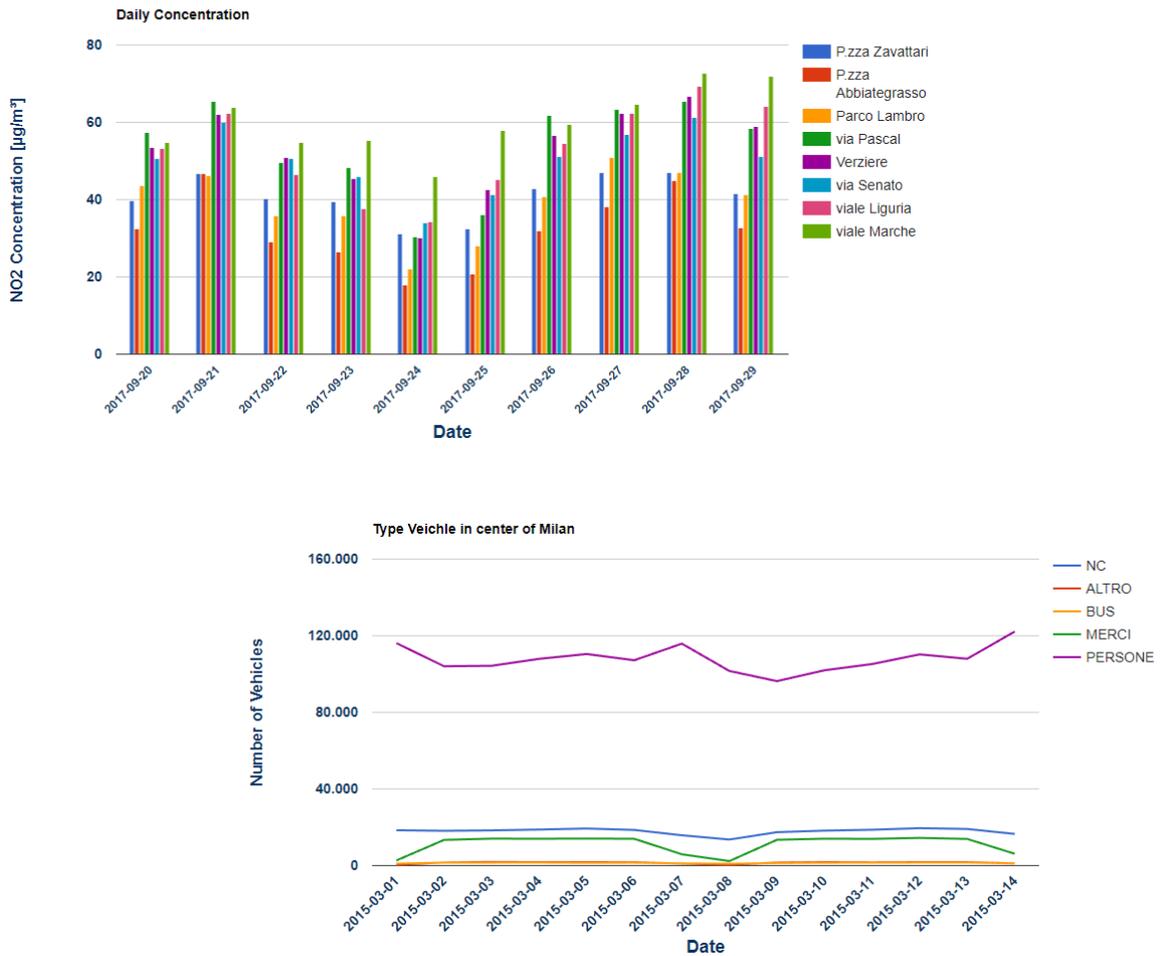


Figura 5.57: Istogramma e Grafico a Linee realizzati con MySQL/PHP

In figura 5.5 invece, sono mostrati gli stessi grafici realizzati con Google DataStudio, in questo caso vengono mostrati anche i parametri che l'utente può indicare per il proprio report.



Figura 5.58: Istogramma e Grafico a Linee realizzati con Google DataStudio

Oltre alle differenze di gestione dati citati in precedenza, DataStudio fornisce maggiore libertà all'utente in occasione dell'inserimento dei parametri. L'esempio più immediato è che l'istogramma era ideato per mostrare il trend dei 10 giorni precedenti ad una data indicata, invece con Data Studio sarà necessario chiedere all'utente di indicare l'intervallo di giorni e di conseguenza non si ha la certezza che questo trend riguardi un intervallo di 10 giorni. Questa maggiore libertà non è da considerarsi un vantaggio o uno svantaggio, dipende da cosa deve essere sviluppato e visualizzato.

6. Osservazioni sui dati

Questo capitolo precede la conclusione definitiva della trattazione della tesi qui sviluppata; si è valutato opportuno mostrare oltre che la parte di sviluppo del progetto descritto, anche l'utilizzo effettivo di ciò che è stato prodotto.

In particolare, si è deciso di analizzare alcune informazioni estratte dal DW e visualizzate tramite la dashboard, e da queste analisi effettuare delle osservazioni sui dati trattati. Tali osservazioni potrebbero ad esempio essere un punto di partenza per un utente esperto come un amministratore comunale di Milano, per prendere delle decisioni per migliorare la qualità dell'aria. Simulando il comportamento e l'interesse di un amministratore comunale che va ad utilizzare la dashboard messaggi a disposizione, probabilmente dopo aver analizzato in via generale la sezione di Pollutant Concentration e di conseguenza dopo aver capito quali sono gli inquinanti più dannosi per i cittadini e più presenti nell'aria; l'attenzione ricadrebbe sulle analisi riguardanti la relazione che c'è tra i livelli di inquinamento e altri fattori (meteo e traffico).

Con il Meteorological Report l'amministratore potrebbe notare facilmente quanto la presenza nell'aria di un inquinante molto pericoloso come il PM10 sia fortemente correlata con i principali fenomeni atmosferici. Infatti, la figura 6.1, in particolare il grafico della Correlazione di Pearson mostra l'elevata correlazione tra temperatura e PM10, in particolare, all'aumentare della temperatura, i livelli di PM10 diminuiscono.

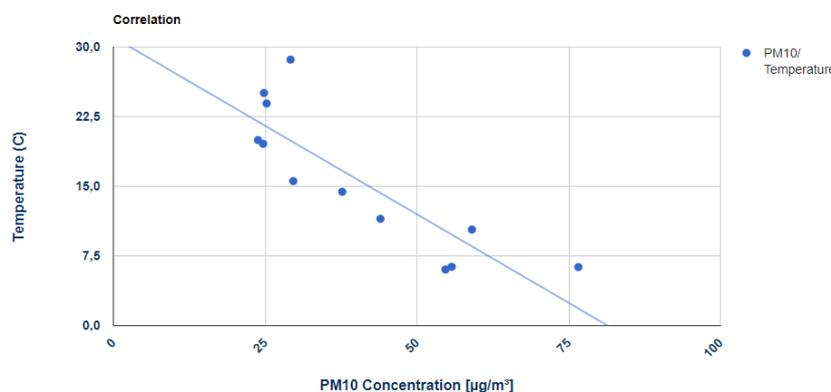


Figura 6.1: Correlazione di Pearson tra temperatura e PM10

Naturalmente, solo questo grafico dà una prima indicazione all'amministratore su quali periodi concentrare maggiormente il proprio interesse per future decisioni per migliorare la qualità dell'aria urbana, ma non è sufficiente per capire a pieno quali possano essere le cause di questa variazione di inquinamento.

Tramite la figura 6.2, che riporta uno dei grafici estratti sempre dalla sezione Meteorological Report (in particolare quella di Monthly Range), si può facilmente notare la forte variazione dei livelli di inquinante tra i mesi invernali e quelli estivi. Infatti, da tale grafico si può notare come le concentrazioni di inquinanti tra inverno ed estate siano praticamente dimezzate.

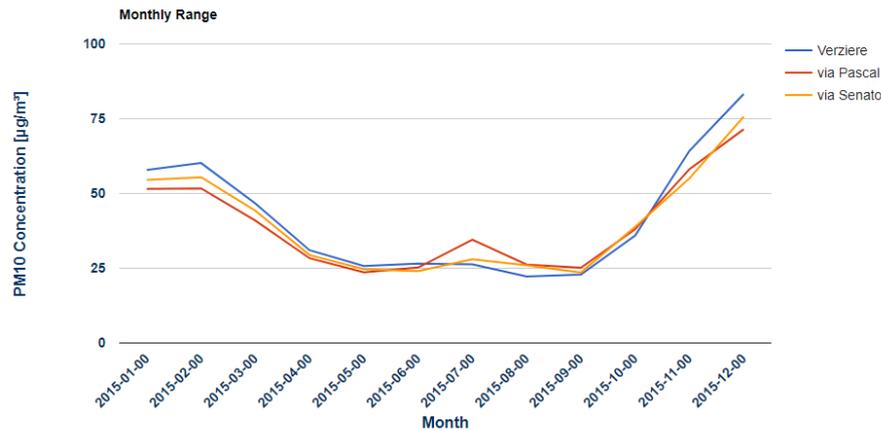


Figura 6.2: Concentrazione PM10 nell'anno 2015

Queste due prime osservazioni fatte, potrebbero portare l'amministratore a capire quanto influisca l'emissione nell'aria di sostanze inquinanti come il PM10, dovuta al riscaldamento domestico. Come ben si sa, questa è una delle cause dell'inquinamento atmosferico urbano, soprattutto nelle situazioni di abitazioni molto antiche e non modernizzate dal punto di vista energetico e queste analisi hanno potuto sottolineare la forte influenza che tale causa ha sul fenomeno dell'inquinamento. Da qui si intuisce che possibili decisioni devono essere prese per migliorare l'efficienza energetica delle abitazioni, proprio per evitare emissioni eccessive durante i mesi invernali.

Un'altra analisi interessante per un amministratore locale è mostrata in figura 6.3 e potrebbe essere quella riguardante l'altro principale causa inquinamento, cioè il traffico urbano. Tramite il Vehicle Report si può facilmente osservare come il numero di veicoli di quasi tutte le tipologie, tra il 2015 e il 2016 sia stato dimezzato (in conseguenza di alcune direttive comunali riguardanti la mobilità).

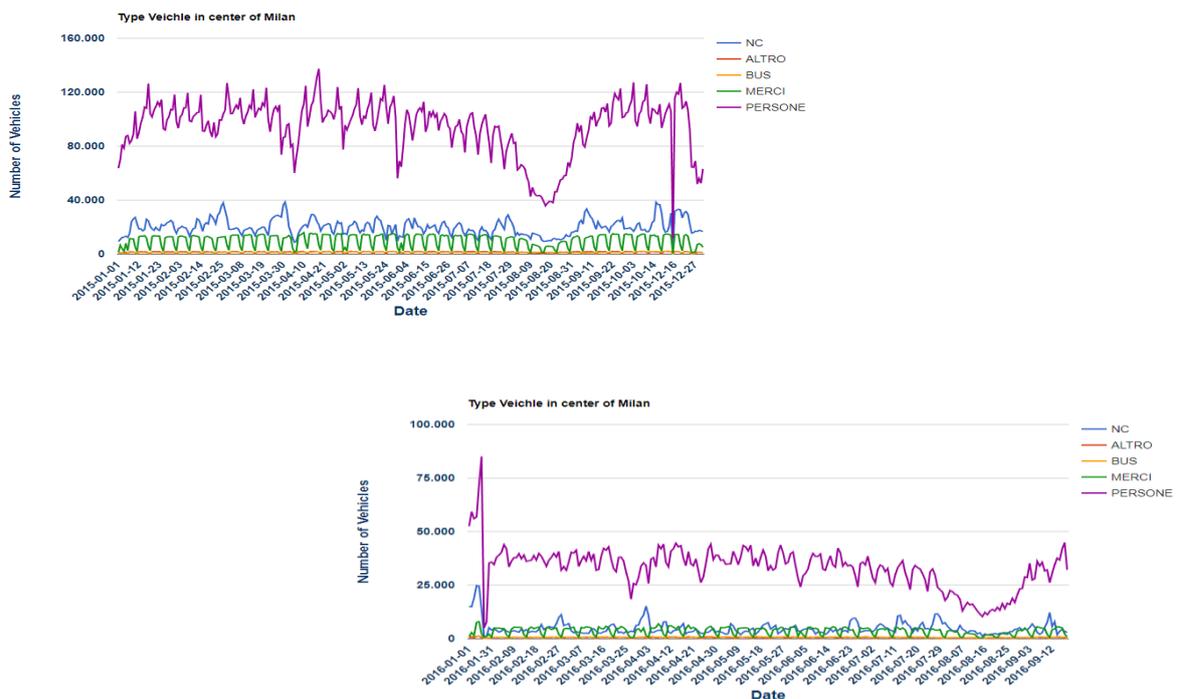
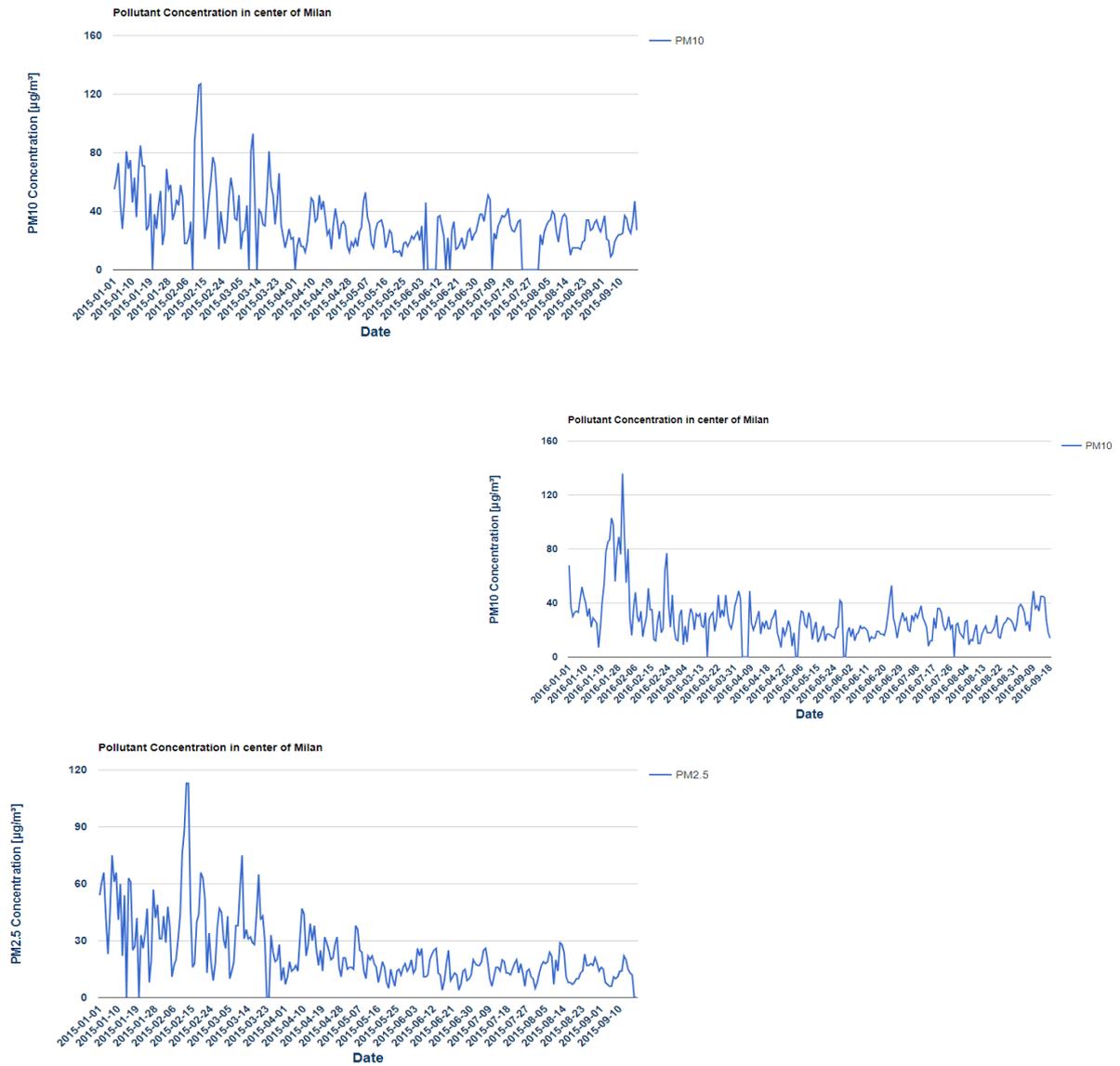


Figura 6.3: Numerosità di veicoli nel 2015 e nel 2016

Ciò che è interessante vedere, è come tale diminuzione abbia influito sulla qualità dell'aria urbana, andando a vedere nello stesso periodo, come sono variato le concentrazioni d'inquinamento delle principali sostanze emesse dai veicoli (PM10, PM2.5 e NO2). Dalla figura 6.4 si nota che il PM10 è rimasto praticamente invariato, il PM2.5 ha subito delle piccole variazioni e invece, il NO2 è diminuito notevolmente.



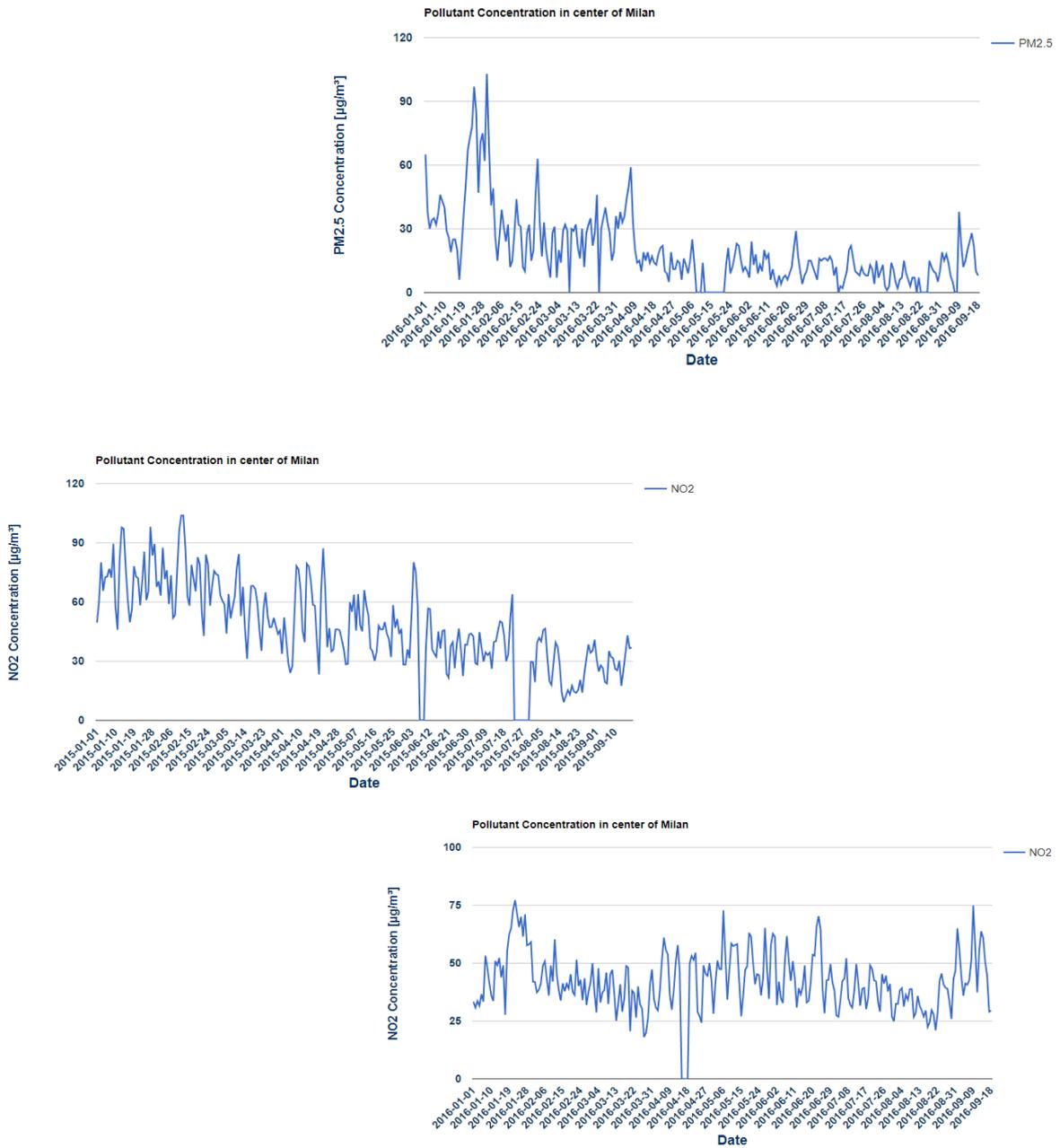


Figura 6.4: Concentrazioni inquinanti nel 2015 e nel 2016

Questa analisi porta facilmente a capire che se le politiche sulla mobilità portano sicuramente a dei miglioramenti, ma bisogna agire anche su altre fonti, affinché si possano diminuire anche i livelli di PM10 e PM2.5, che sono molto pericolosi per i cittadini.

7. Conclusioni e Sviluppi futuri

In conclusione, il progetto trattato in questa tesi aveva come obiettivo quello di progettare e realizzare un Data Warehouse che avrebbe messo a disposizione i dati di Milano riguardanti l'inquinamento atmosferico urbano e legare questo fenomeno con sue possibili cause (traffico cittadino) e con altri fattori che potrebbero influenzare i livelli di inquinamento (condizioni meteorologiche); estrarre indicatori utili per l'analisi di tutti questi aspetti e delle loro possibili correlazioni e sviluppare dei metodi di visualizzazione (dashboard e applicazione mobile) per utenti di vario tipo, interessati a informazioni diverse e più o meno dettagliate. Naturalmente, questo progetto è stato considerato fin troppo ampio per essere sviluppato nella sua interezza in una sola Tesi di Laurea. Quindi ci si è focalizzati sulla progettazione e modellazione del DW con tutti gli aspetti necessari al progetto, per poi sviluppare la dashboard utile principalmente ad un utente più esperto (amministratore comunale) rispetto a chi utilizzerebbe l'app. Tramite tale dashboard, come già detto, basandosi su particolari KPI ha permesso di fornire informazioni sui livelli di inquinamento, sul traffico e sulle condizioni climatiche principali, andando più nel dettaglio su particolari legami tra l'inquinamento e gli altri due tipi di dati.

Si passa ora, dopo essere riusciti a sviluppare tale dashboard a valutare quali possono essere possibili aspetti e di conseguenza sviluppi futuri da prendere in considerazione. In particolare, come già detto, di questo progetto fa parte anche un'applicazione mobile rivolta principalmente a comuni cittadini che hanno voglia di monitorare ed essere informati sulla qualità dell'aria del posto in cui vivono; naturalmente, quest'app dovrebbe contenere informazioni più immediate, più riassuntive, anche usando indici differenti da quelli usati nella dashboard e senza focalizzarsi sul legame con altri fattori. Per quanto riguarda la dashboard invece, si può pensare ad altre funzionalità e analisi da mettere a disposizione dell'utente: ad esempio, usando dati sulle emissioni dovute ad altre fonti di inquinamento (riscaldamento domestico, emissioni industriali) si possono effettuare dei report più dettagliati e più utili per un amministratore per avere una visione complessiva su quali fonti sono più causa dell'inquinamento atmosferico, e di conseguenza capire come agire e quali decisioni prendere per migliorare la qualità dell'aria urbana.

Dal punto di vista più teorico, dall'analisi delle altre dashboard, ma dando anche uno sguardo alle applicazioni mobili che riguardano questo tema, è da prendere in forte considerazione l'idea di utilizzare un indicatore denominato AQI. Tale indice viene spesso usato come indicatore generale della qualità dell'aria di una intera città o di una zona, di conseguenza, potrebbero essere incorporate nella dashboard ulteriori analisi basate su tale KPI.

8. Bibliografia e Sitografia

- [1] Ramakrishnan R., Gehrke J., “Sistemi di basi di dati”, McGraw-Hill, 2004
- [2] Golfarelli M., Rizzi S., “Data Warehouse – teoria e pratica della progettazione”, McGraw-Hill, 2006
- [3] Corno F., Torchiano M., “Sistemi Informativi Aziendali – Appunti per il corso”, 2018, <https://elite.polito.it/files/courses/02CIX/libro/cap10-standalone.pdf>
- [4] Inquinamento atmosferico, https://it.wikipedia.org/wiki/Inquinamento_atmosferico, consultato il 05/02/2018
- [5] Nozioni di base, https://www.airqualitynow.eu/it/pollution_home.php, consultato il 05/02/2018
- [6] Inquinanti, <http://www.arpa.piemonte.it/approfondimenti/temi-ambientali/aria/aria/cartella-qualita-inquinanti>, consultato il 05/02/2018
- [7] MySQL, <https://it.wikipedia.org/wiki/MySQL>, consultato il 06/02/2018
- [8] PHP, <http://www.html.it/guide/guida-php-di-base/>, consultato il 06/02/2018
- [9] Javascript, <http://www.html.it/guide/guida-javascript-di-base/>, consultato il 06/02/2018
- [10] AngularJS, <http://www.html.it/guide/guida-angularjs/>, consultato il 06/02/2018