POLITECNICO DI TORINO

Corso di Laurea Magistrale

in

Ingegneria Informatica (Computer Engineering)

Tesi di Laurea Magistrale

Collecting and Analyzing User Feedback
to Guide Automated Data Exploration

Relatore
prof. Tania Cerquitelli

Candidato
Vincenzo Torcasio

Anno Accademico 2020/2021

Sommario

Introduzione

Al giorno d'oggi sono molti i singoli utenti, le piccole e grosse aziende e i ricercatori che puntano all'analisi del dato per poterne estrapolare conoscenza e quindi ottenere dei vantaggi nel loro campo.

La generazione di un dato è un evento molto importante, in quanto quel singolo frammento di dato può essere utile se messo in correlazione ad altri simili ad esso per poter estrarre numerose informazioni. C'è una distinzione netta da fare in merito alla terminologia adottata, in quanto un *dato* è formato da una struttura, è un oggetto presente e salvato su una sezione del disco di un calcolatore. Questo può confondersi nella miriade di dati che sono stati generati nello stesso istante o in quelli successivi e precedenti. Quello che conta, in ogni caso, è che esso può essere fonte di *informazione* per un certo individuo.

I dati risultano essere di fondamentale importanza se visti sotto una certa ottica, se possono essere analizzati e sottoposti a processi di analytics.
In altri termini possiamo affermare che un dato può essere una fonte inesauribile di informazioni.

Con l'avvento di *internet* e di tutti gli strumenti che mette a disposizione, negli anni il processo di analisi del dato è stato completamente stravolto. In molti si sono avvicinati a questa disciplina detta appunto *Data Science*.
Questa risulta essere un'area veramente molto vasta e molti sono i ricercatori e le aziende interessate ai vari processi, in quanto l'analisi dei dati porta svariati vantaggi sia da un punto di vista prettamente scientifico (sono infatti molte le discipline che hanno investito in queste ricerche) ma anche da un punto di vista economico, in quanto è possibile correggere bilanci o effettuare manovre in borsa.
Qualsiasi sia quindi l'obiettivo finale, i vari processi che puntano all'estrapolazione delle informazioni sono sempre e costantemente in ascesa.

Numerosi sono gli strumenti e le applicazioni sviluppate dalle varie aziende leader nel settore per fornire strumenti di supporto alle aziende più piccole, università o singoli utenti impegnati in progetti personali.
La maggior parte di tali applicazioni possono semplicemente trovarsi in rete e alcune di esse forniscono sempre dei manuali di supporto per una comprensione alquanto esaustiva di tutte le features che esse possiedono. Ma questo tipo di applicazioni possiedono veramente delle grandi limitazioni, il costo computazionale.

L'analisi di un dataset può risultare molto oneroso come processo se si fa riferimento alle risorse hardware e alle tempistiche. Proprio per questo motivo è stato necessario trovare delle soluzioni per poter fornire la possibilità di caricare dei dataset anche molto grandi da poter analizzare.
Il web mette a disposizione una moltitudine di queste applicazioni a cui affidarsi, per evitare di addentrarsi nello sviluppo di nuove applicazioni completamente da zero.

L'obiettivo di questo lavoro è però dedito allo sviluppo di un nuovo strumento che prende il nome di **ADESCA-(Automated Data Exploration with Storytelling CApability)** e che si basa sullo sviluppo di un *tool* per l'esplorazione automatica dei dati.

Questo è un progetto di ricerca già avviato e che nel tempo riesce a prendere sempre maggiormente forma. La sua versione base è descritta in questo paper [1] e consta al momento di una numerosa serie di sezioni già implementate per l'analisi dei dataset, che vengono direttamente scelti e caricati dall'utente.

Una delle tante sezioni è quella chiamata **StoryTelling**, che mira ad un'analisi completamente personalizzata del dataset caricato dall'utente per potergli spiegare e presentare i dati.

L'implementazione di questa sezione non prevede una semplice visualizzazione di grafici, in quanto il concetto citato poco fa è molto più ampio e complesso. Questo, infatti, mira a *presentare* i dati in modo che l'utente ne possa trarre il massimo beneficio e che possa comprenderli nel modo più semplice possibile. Questo porta numerose problematiche proprio perché gli utenti possono avere background diversi e soprattutto una diversa sensibilità.

L'obiettivo principale è quindi quello di riuscire a schematizzare per quanto possibile delle categorie di utenti, andando ad effettuare una suddivisione in base al loro livello di esperienza. Per fare ciò si sono rivelati necessari l'attuazione di processi per la raccolta di informazioni, andando ad implementare un form per raccogliere dati e vari applicativi da sviluppare in parallelo per l'attuazione delle analisi di tali risultati.

Una raccolta dati preliminare però non basta, in quanto per avere un quadro completo della situazione è necessario poter ottenere informazioni dagli utenti anche in un *post utilizzo* dell'applicazione, in modo tale che questi possano rilasciare dei *feedback* che possano aiutare a migliorare il tutto sotto ogni aspetto.

Facendo ciò è stato possibile implementare un sistema capace di migliorare tale sezione basandosi su dati concreti e grazie alle varie applicazioni che consentono di estrarre una maggiore conoscenza dai dati raccolti sarà possibile implementare delle soluzioni *ad hoc* per **ADESCA** e la sua sezione di **StoryTelling**.

Concetto di StoryTelling

Spiegare il concetto di **StoryTelling** risulta essere per certi versi abbastanza complesso.
In generale è possibile affermare che questo è visto come l'atto che sfrutta i principi dell'arte oratoria per poter narrare dei fatti o più in generale delle *storie*.
Questa iniziativa di narrazione si evolve e si applica nelle forme più disparate in quanto si riferisce sia ad opere letterarie (e quindi su supporti cartacei) ma anche ad opere artistiche (cinema, danza e musica). Questo importante concetto però si estende ai giorni nostri anche in maniere differenti in quanto con l'evolversi dei tempi cambiano gli strumenti che le persone possono adoperare ma nonostante ciò l'obiettivo è sempre comune: riuscire a far comprendere il messaggio agli utenti in ascolto.

Alcuni esempi che è possibile citare sono sicuramente quelli legati al lavoro aziendale, in quanto un *Project Manager* deve essere capace di comunicare nel modo più chiaro possibile quelle che sono le scelte organizzative e tecniche che il suo team deve svolgere. O ancora meglio questo processo potrebbe essere utilizzato come biglietto da visita per presentazione dell'azienda o di una singola persona. In sostanza, ciò che deve essere raggiunto è il "*far recepire il messaggio*".

Questo processo trova però spazio anche in molti altri ambiti come quello del mondo digitale, e un'evoluzione così drastica necessita di un buon adattamento. È quindi possibile affermare che:
*Lo **StoryTelling** è una forma di produzione dei media digitali, usata dalle persone per poter condividere la propria storia (Figura 1). I supporti usati a tale scopo spaziano nei vari ambiti del digitale, al di fuori dei supporti fisici.*[2]

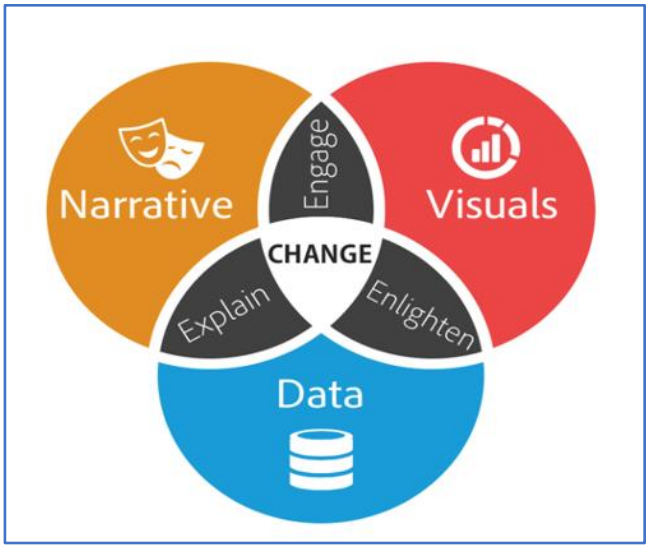


Figura 1: Grafico riassuntivo che cerca di spiegare il concetto di **StoryTelling** definendo come fondamentale la componente narrativa.

Storytelling in Data Science

Come introdotto già nel paragrafo precedente, il concetto di **StoryTelling** viene esteso anche nel settore digitale e si ramifica in tutti i suoi ambiti.
Questo processo viene talvolta adottato soprattutto dalle grandi multinazionali in diversi settori che mirano principalmente alla vendita e ad incrementare la loro produttività.
Tali indici di produttività vengono però tenuti a bada da una scienza ben più che esatta, ovvero l'analisi dei dati degli oggetti prodotti e soprattutto della loro vendita.

Questo consiste quindi nel riuscire ad ottenere dei risultati quanto più possibile corretti dall'analisi dei dati per poter successivamente investire (il riscontro in questo caso è puramente economico).
L'analisi dei dati viene però limitata sempre dallo stesso concetto, *è così semplice capire quelli che sono i risultati ottenuti?* Soprattutto nei casi in cui si faccia riferimento all'analisi di larghi dataset e il contenuto informativo è veramente ampio.

Ed è proprio qui che subentra il concetto di **StoryTelling**, che mira a far sì che tutti i risultati ottenuti vengano riportati come si deve, soddisfacendo così la necessità dei vari utenti che necessitano di "*risposte*".[3]

Dopo aver effettuato una serie di analisi è necessario poter riportare dei risultati, che possano giustificare gli esperimenti condotti.
Tali risultati necessitano a loro volta di essere presentati nella forma migliore possibile (rispetto al tipo di dato analizzato e al tipo di analisi effettuata), per poter essere compresi. Ed è qui che si ritiene necessaria l'azione di curare la forma di espressività, in quanto l'audience potrebbe risultare non esperta e quindi inadeguata alla comprensione dei risultati, che dovrebbero essere presentati nella forma più appropriata possibile.
L'utente finale, utilizzatore del programma, non sempre possiede la convinzione e le capacità di poter interpretare i dati risultanti e proprio per questo il concetto di **StoryTelling** risulta importante, perché ci consente di presentare i risultati ottenuti all'utente in una forma più semplice e concreta.

Infatti, un approccio allo **StoryTelling** in ambito Data Science risulta ancora più complesso, in quanto la necessità di poter apprezzare i dati per compiere successivamente delle scelte "mirate", pongono i progettisti di sistemi a mettersi nei panni degli utenti/utilizzatori che possono essere quanto più diversificati a livello di esperienza nell'interpretazione di risultati.

Un risultato non deve essere ottenuto e semplicemente visualizzato (e magari neanche giustificato).
Il focus è quello di andare a dimostrare il come e il perché quel risultato è stato ottenuto, presentandone i vari tratti che hanno portato al suo ottenimento e le caratteristiche salienti. Tutto ciò deve essere fatto tramite una serie di tecniche di comunicazione avanzate e adatte al tipo di utente.
Lo **StoryTelling** in un'applicazione generica (utilizzata per un qualche fine) mira a migliorare l'esperienza dell'utilizzatore, rendendola più semplice e gradevole, il suo compito è quello di andare a migliorare la **Human Experience**.

Non tutte le figure possiedono le capacità di comprendere il linguaggio tecnico, in quanto ci sono persone con sensibilità diverse e quindi possono maggiormente comprendere e preferire un metodo di esposizione diverso rispetto ad un altro.
Un concetto di fondamentale importanza è quello della distinzione tra visualizzazione e presentazione dei dati.

Un **Data Scientist**, può possedere svariate skill (progettazione, programmazione ecc…) e in un ambito così vario e vasto è difficilmente apprezzabile un utente che possa eccellere in tutto. In genere il compito di un analista è quello di riuscire ad estrapolare conoscenza a partire dai dati che ha a disposizione, ad oggi vi è la possibilità di sfruttare varie tecniche e strumenti per raggiungere tale goal. Ma una delle skill che ognuno dovrebbe poter padroneggiare è proprio quella dello **StoryTelling**, che però risulta molto difficile da acquisire proprio perché non bisogna guardare ai risultati ottenuti con gli occhi di chi ha progettato il sistema di analisi (quindi nei panni dell'ingegnere) ma con gli occhi di chi ne sarà l'utilizzatore (*Figura 2*).

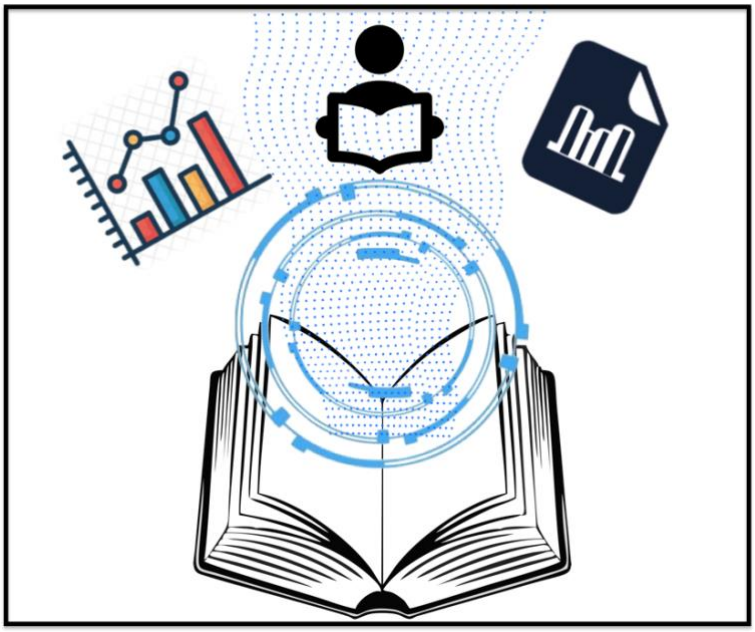


Figura 2: immagine rappresentativa del concetto di **StoryTelling** applicato all'ambio della **DataScience**.

Ogni dato "*possiede la propria storia*" a partire da quando esso è stato generato, e questa deve necessariamente essere affiancata alla visualizzazione del risultato ottenuto in seguito all'applicazione dei vari processi di analisi.
Il connubio tra queste due operazioni porta a quello che è il concetto di **presentazione**.

Un utente non deve necessariamente riuscire a capire i tecnicismi o a come dover settare un'applicazione, se esso è solamente interessato ad un risultato o a carpire determinate informazioni piuttosto che altre. Infatti, un utente che si ritroverebbe bombardato con troppi risultati sotto forma di grafici e numeri all'interno dell'interfaccia dei risultati, potrebbe essere confuso quanto più seccato di dover "impegnarsi a capire" i propri risultati, per questo motivo si ha anche la necessità di dover fornire solo ed esclusivamente quello che viene chiesto.

Lo **StoryTelling** introduce la possibilità di creare un'esperienza quanto più possibile personalizzata e inerente alle analisi di quello specifico Dataset e delle competenze dell'utenza, evitando così lo sforzo di utilizzare i tecnicismi. Inoltre, come già accennato in precedenza, una buona *data stories* contiene solo ciò che di fondamentale importanza è per l'utente, fornendo tutt'al più dei suggerimenti o consigli per le azioni da intraprendere successivamente.

Con il termine "***The Data StoryTelling***" si specifica il concetto di andare a prendere qualcosa che di per sé non è "umano" e trasformarlo in qualcosa di facilmente interpretabile e condivisibile con gli altri utenti. Questo è un concetto che può essere esteso in diversi ambiti e possiede diverse ramificazioni, ma in ogni caso quello che si cerca di fare è sempre quello di andare ad introdurre una **componente narrativa** che possa migliorare la nostra capacità espressiva.

Riassumendo:

- Data Visualization ≠ Data Presentation, in quanto è necessaria la componente narrativa
- Uso di termini facilmente comprensibili

"**Stories Bring Data To Life**", ed è questo il motivo per cui aggiungere una componente narrativa può essere di fondamentale importanza per la comprensione di essi, in quanto le storie colpiscono in maniera unica l'utente che in tal modo riesce ad immergersi nel contesto. *"Le storie sono la forma di contenuto più potente, perché gli umani sono programmati per le storie".*

Ci sono varie modalità di presentazione dei dati, dalla visualizzazione di immagini, valori numerici e rappresentazioni di risultati tramite **Dashboard**.

Uno dei principali esempi, spiegato in un video trovato online e dedicato allo **StoryTelling**, è la creazione di una locandina per l'associazione Save The Children, il cui obbiettivo era quello di presentare dei risultati per riuscire a coinvolgere più persone nella causa.
L'esperimento (*Figura 3*) è stato condotto creando ben due locandine, in cui sono stati **presentati** i dati in maniera diversa (in una tramite dati numerici e grafici, l'altra tramite immagini). Il risultato finale è stato quello che gli utenti sono risultati più sensibili alla seconda locandina (con immagini), eppure lo studio condotto è identico, ciò che è cambiato è stata solo la presentazione dei dati risultanti.

Da qui si evince che la sensibilità (e l'interpretabilità) delle persone varia da individuo a individuo, e il processo di presentazione risulta importante per una corretta comprensione dei dati.[4]

Figura 3: Immagine estrapolata dallo stesso video presentazione discusso sopra, questa tende sottolineare la differenza di potenza espressiva tra le due presentazioni dei dati. A sinistra una *Dashboard*, a destra una *Story*.

Nelle varie documentazioni/video spiegazioni è risultato che una buona prassi per la costruzione di una storia risiede in semplici e piccoli passi che possano catturare completamente l'attenzione di un utente.

Nel processo di **Data StoryTelling**, è possibile distinguere tre fasi principali (vedi *Figura 4*):

1) Setup – "Before state of data"
2) Conflict – "How the data changes"
3) Resolution – "After state"

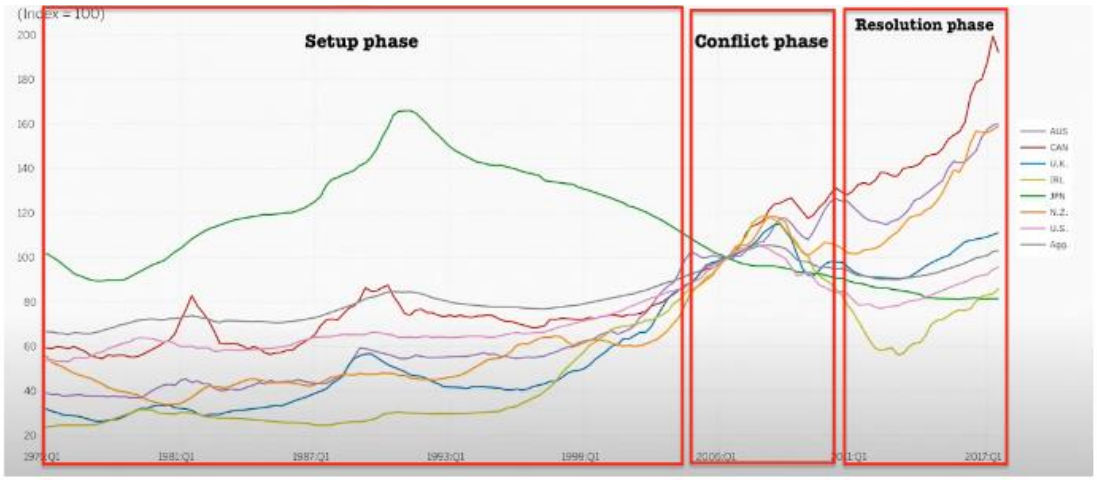


Figura 4: Screenshot prelevato dal video spiegazione, il quale mostra dei dati che vengono analizzati e quelle che sono le tre principali fasi.

In questi punti sono riassunte le tre principali fasi, quella di *Setup* consiste nel visionare i dati e capirne il significato, sottolinearne le caratteristiche principali e soprattutto specificare una condizione/definirne uno stato. [5]

La seconda fase, quella di *Conflict*, consiste invece nel trovare un qualsiasi evento che possa "alterare" il normale comportamento dei dati (ad esempio un evento imprevisto).

La terza fase, quella di *Resolution*, invece consiste nel valutare come i dati siano arrivati ad un nuovo stato, dovuto al cambiamento che c'è stato nella seconda fase. Questa fase richiede nuovamente interpretazione e valutazione dei nuovi dati.

In generale questa modalità di erogazione delle "*stories*", fa sì che un utente riesca a comprendere l'evento modificante, e quindi riesca a carpire il significato dei dati prima e post evento. In quanto il cambiamento riesce a far distinguere ogni fase e ciò che deve essere riportato sono solo gli eventi più importanti.

Intelligence and Automated Process of StoryTelling

Usare delle tecniche che possano andare ad automatizzare il processo di **StoryTelling** risulta essere molto più complesso di quanto in realtà si possa immaginare. Dopo aver analizzato una serie di dati e aver ottenuto una serie di risultati, è possibile andare ad applicare delle tecniche particolari per la selezione del risultato migliore, considerando sempre il fatto che siano state condotte più analisi, per poterle visualizzare all'utente.

Da qui sorge però il problema della presentazione, che si pone già nelle prime fasi quando l'utente sceglie il dataset da cui estrapolare conoscenza. "*Che tipo di Dataset sarà?*", questa è la principale domanda che la nostra applicazione dovrebbe chiedersi per poter costruire una componente narrativa finale (a parte il fatto che in base al tipo di dataset possono o meno essere condotte delle analisi piuttosto che altre).
Ad oggi si sono studiate varie soluzioni che però rimangono prototipi, e consistono nell'utilizzo di tecniche di **AI** (Artificial Intelligence).

Il primo passo verso la costruzione di un'applicazione che possa autonomamente creare delle stories adatte ai propri utenti (in base al dataset e alle analisi possibili) parte dal capire cosa vuole l'utente, raccogliendo quante più informazioni possibili anche in modo interattivo.

Questo tipo di processo non si estende solamente all'ambito dell'analisi dei dati, ma trova fondamenti anche in un'accezione più generale come la possibilità di far apprendere ad un calcolatore la capacità di raccontare delle *stories*.

Un **AI** di questo tipo viene completamente forzata ad apprendere dei pattern, tramite vari tipi di dati come immagini e video, che possano ricondurre a stati emozionali piuttosto che altri. Sempre considerando il concetto già espresso in precedenza che consiste nel sottolineare quelli che sono gli eventi più importanti e che possano fornire una distinzione netto dal resto della storia, potremmo denotare questi eventi come "*punto di rottura*".
Questo tipo di processo farà sì che una *story* possa essere "*raccontata*" con enfasi ed instillare negli ascoltatori determinate emozioni.

E questo è proprio lo stesso concetto di fondo che dovrebbe essere utilizzato nell'analisi di dati con *Automated StoryTelling Process*. Un calcolatore durante l'analisi di un dataset dovrebbe poter capire quali sono le *features* più importanti e che possono essere evidenziate rispetto ad altre, il tipo di algoritmi che sono stati utilizzati e le varie operazioni fornite come input dell'utente dovranno essere automaticamente processate per poter ottenere uno **StoryTelling** personalizzato.

L'uso dell'intelligenza artificiale al giorno d'oggi ha preso piede con una velocità esorbitante e i campi di applicazione risultano essere vari.
Queste sezioni di codice vanno infatti ad essere inserite ormai in svariate applicazioni e che hanno come scopo quello di aiutare gli utenti a migliorare i propri processi di generazione dei dati e quindi automaticamente nella loro presentazione.

La vera difficoltà rimane comunque il riuscire a capire quale possa essere un corretto modo di presentazione dei risultati, tanto da poter essere compreso dalla molteplicità di utenti con background differenti.

Recommender System Survey

Per la creazione di un'applicazione orientata all'analisi dei dati, per far sì che possa fornire risultati quanto più possibile appropriati in relazione all'esperienza dell'utente utilizzatore del sistema, vi è la necessità di creare un sistema di raccolta dati per la schematizzazione o profilazione degli utenti.

Da qui la necessità di introdurre un **Sistema di Raccomandazione** che possa essere utilizzato per consigliare agli utenti (in base al proprio profilo costruito) determinate operazioni e fornire i risultati in modo appropriato.

I **Recommender Systems** (RS) vengono utilizzati per fornire consiglio agli utenti nelle più svariate situazioni e ad oggi questo tipo di sistemi vengono usati nelle principali applicazioni che riguardano store online o intrattenimento (Amazon, Netflix ecc…).

Il principio di funzionamento (e.g. Come rappresenta la *Figura 5*) si basa sul suggerire agli utenti oggetti (in caso di marketing online) o azioni da intraprendere, che possano essere simili agli interessi dell'utente stesso. Il concetto è quindi quello di andare a raccogliere più informazioni possibili da questo singolo utente per crearne una sorta di profilo.
Un secondo approccio invece, leggermente diverso, consiste nel costruire tale profilo utente basandosi anche su profili di altri utenti, che paragonati al singolo in questione, possono risultare simili a livello di interessi (ad esempio la recensione di un oggetto appartenente alla stessa categoria).
Quindi, oggetti e azioni verranno suggeriti dall'applicazione prendendo informazioni dal profilo costruito in questa modalità e facendo visualizzare all'utente utilizzatore solo questo tipo di pubblicità (se in relazione ad oggetti da acquistare) oppure in una sezione che potrebbe prendere il nome di "*consigliati*" i titoli di film o qualsiasi altro genere di forma di intrattenimento.

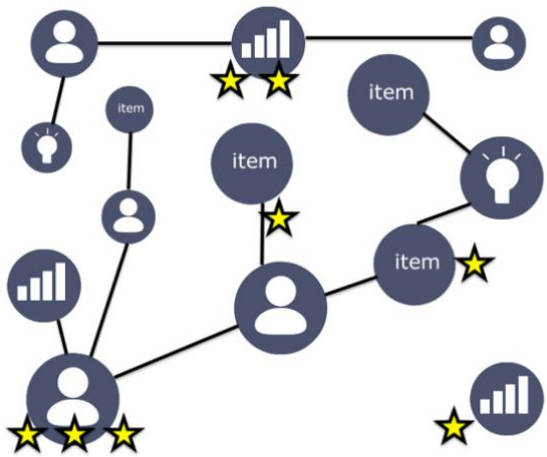


Figura 5: Grafico rappresentante la logica dietro un sistema di raccomandazione, che consiste nel creare dei link ed associare un ranking legato al livello di similitudine degli utenti.

Quando si parla di RS, si rende necessario il concetto di "**Information Filtering**", che si fonda sul principio che un'informazione classificata come rilevante deve essere recapitata, a discapito di altre info meno importanti. Il filtraggio delle informazioni deve essere fatto in base alle preferenze individuali e alla descrizione dell'utente, ecco appunto l'utilità di creare un profilo quanto più veritiero possibile.

Da quanto detto sopra, si possono distinguere i concetti di:

- **Collaborative Filtering**
- **Content-Based Filtering**:

Il primo consiste nel predirre/filtrare le preferenze dell'utente, basandosi su informazioni di altri utenti.
Il secondo invece, basa il suo approccio sul costruire le preferenze in relazione ai gusti dell'utente e consigliandone i top oggetti, appartenenti a quella categoria.

Un approccio invece più complesso, ma che potrebbe rendere il sistema funzionante e prestante e quello di andare ad inserire delle tecniche di **Machine Learning** (ML) per riuscire nella creazione di un sistema autonomo, che possa evolversi nel tempo e adattarsi sempre di più alle caratteristiche dell'utente. Un esempio lampante potrebbe essere la costruzione di un sistema di raccomandazione per un negozio online, allo stesso tempo però risultano esserci una serie di problematiche da dover risolvere, legate all'immenso numero di oggettistiche a cui l'utente può essere interessato, e che quindi possono mandare fuoristrada il sistema. Proprio per questo motivo le analisi che devono essere condotte necessitano di grande precisione e si limitano (sempre in base a quella che sarà la strategia di implementazione) a costruire i profili utenti basandosi principalmente sulle ricerche effettuate e sulla loro frequenza.

Gli **RS** possono comunque essere adottati in ogni ambito, in quanto la loro principale funzione è proprio quella di andare a raccogliere informazioni, capirne le relazioni tra esse e profilare gli utenti. Gli algoritmi in questi sistemi si basano fondamentalmente sulla costruzione di strutture dati (come matrici) che possano andare a mettere in relazione l'utente con i vari oggetti/azioni e calcolandone la percentuale di matching. Vari sono gli schemi che possono essere proposti per il computo di tali risultati, usando però di base un approccio probabilistico. Muore quindi quello che è un approccio statico e basato su metodi obsoleti in quanto una **AI** può automaticamente evolversi e quindi adattarsi a qualsiasi cambiamento dell'utente. [6]

Riflettendo sul concetto di RS, questi sono sistemi che possono portare a creare grosse opportunità sia a livello commerciale e sia a livello di miglioramento della user experience.

La raccolta delle informazioni non avviene solo tramite dei singoli sondaggi, ma avviene anche andando a valutare le azioni (precedenti e future) dell'utente. Un **RS** deve inoltre riuscire a discriminare le informazioni utili da quelle non utili, ed è proprio questa la sfida che ci si pone davanti. Sondaggi ed Online Evaluation riscontrano più consensi, rispetto a quelli offline, in quanto risultano avere un grado di interazione maggiore. Ci sono ovviamente pro e contro, molto spesso la raccolta di informazioni risulta essere onerosa in termini di tempo.

Come accennato nell'ultimo sotto paragrafo è possibile implementare un **RS** basandosi principalmente su due approcci diversi:

- **Registrazione utente**
- **Suddivisione categorica**

Queste due alternative portano allo stesso risultato ma una o l'altra possono essere preferite in base anche al tipo di progetto a cui si sta lavorando.

Nel primo caso questo prevede che l'utente per poter utilizzare l'applicazione deve necessariamente essere registrato al servizio, questo tipo di struttura è molto comune al giorno d'oggi ed è utilizzata principalmente dai colossi del mercato nell'ambito dello shopping online.
In questo modo ogni singolo utente potrà essere profilato e tale profilo potrà essere continuamente aggiornato nel tempo, in quanto a gestire il tutto sarà sempre un **AI** e quindi non vi sarà necessità di re-implementare il codice.
Agli utenti verranno quindi inviate comunicazioni riguardo le principali novità sull'applicazione e verranno costantemente suggeriti oggetti simili a quelli già ricercati in un modo totalmente automatizzato.

La seconda strategia invece riprende un concetto opposto, ovvero quello legato al fatto che siano gli utenti ad effettuare un'autovalutazione per poter poi effettuare una scelta ed utilizzare la modalità "*semplificata*" di un programma piuttosto che quella per utenti "*esperti*". Basta pensare semplicemente a quelle applicazioni che richiedono anche durante l'installazione all'utente di effettuare delle scelte.
Questo tipo di implementazione viene realizzata seguendo un approccio di raccolta moduli, compilati dagli utenti tramite un'opportuna applicazione e poi sottomessi.
In tal modo gli sviluppatori possono autonomamente analizzare i dati raccolti per poi creare in base alle varie risposte ottenute una serie di categorie in cui posizionare gli utenti.
Questo, quindi fa ricadere direttamente all'utente la scelta del poter utilizzare l'applicazione in una configurazione piuttosto che in un'altra.[7]

Feedback all'utente

L'altra faccia della medaglia di un sistema di raccomandazione è quella dei feedback.
Il principio di funzionamento è:

sistema di raccomandazione raccoglie dati ←→ *utente fornisce feedback al sistema*

Un sistema deve quanto più possibile riuscire ad acquisire feedback da parte degli utenti, in quanto questi ci consentono di migliorare la qualità dell'applicazione.
Infatti, gli utenti tenderanno a fornire il proprio pensiero riguardo l'uso dell'applicazione e questo tramite varie forme, come ad esempio un sistema di ranking (come in vari siti internet l'assegnare più o meno stelle), tramite mail, tramite messaggi diretti o short survey.

Un modo, infatti per poter raccogliere quante più informazioni possibili riguarda la possibilità di mettere a disposizione più canali di comunicazione agli utilizzatori dell'applicazione. Non solo un sistema di feedback ben orchestrato può andare a visionare se l'esperienza dell'utente è stata gradevole o meno, ma fornisce anche informazioni su come migliorare la *user experience* e il sistema, in quanto gli utenti potrebbero lamentarsi ad esempio di alcuni aspetti, ed elogiarne di altri.

Figura 6: Spesso le varie compagnie preferiscono ottenere informazioni sugli utenti analizzandone le azioni, tramite like o condivisioni.

Da vari studi è apparso che molte società non sottopongono a sondaggi i propri clienti, ma bensì raccolgono informazioni tramite vie alternative, come sui social network (*Figura 6*).

Alcuni dei punti cardine nella raccolta di informazioni consiste anche nel filtrare questi feedback, in quanto potrebbero esserci dei dati più importanti di altri e capire quali dati siano invece forvianti (ad esempio form completati a caso o risposte insensate).
Bisogna sempre considerare come importanti le recensioni dei clienti fidati (che ad esempio utilizzano la piattaforma da un lungo periodo di tempo), i feedback dei nuovi utenti (per capire l'impressione che l'applicazione ha avuto di loro), la quantità delle stesse problematiche e quante volte si siano ripetute. Alla base di un'analisi più precisa vi è la necessità di classificare i dati, in quanto i feedback possono essere raccolti da più sottosistemi diversi e quindi ognuno deve essere gestito nella maniera migliore possibile (si pensi ad esempio alla gestione di dati strutturati o non strutturati).

I feedback forniti dagli utenti a posteriori dell'uso di quella che è l'applicazione forniscono numerose informazioni congiuntamente a quelle già ottenute dall'implementazione di un **RS**, che sia esso stato strutturato come il primo o il secondo caso. Per poter ottenere quindi maggiori vantaggi è necessario integrare tutti questi dati per poi migliorare sia il sistema di raccomandazione che anche il poter implementare altre sezioni nella nostra applicazione o sistemarne di già esistenti.

L'utilità di fornire un *feedback* a fine utilizzo ci fa comprendere cosa un utente si aspettasse dall'applicazione e quali siano i punti critici questo è riuscito a trovare, motivo per cui il modo in cui si sottopongono i quesiti devono rispettare canoni di qualità e dell'aspettativa.

Rank delle analisi

Dopo aver utilizzato una determinata applicazione un utente si aspetterebbe di ottenere dei risultati, nell'ambito della Data Science questo corrisponde alla presentazione di grafici o risultati sotto forma di tabelle riassuntive.

In questo specifico caso di utilizzo è possibile visitare le varie sezioni di **ADESCA** ed ottenere dei risultati e poterli comparare tra di loro per poter far si che si possa avere una panoramica quanto più possibile ampia per una valutazione personale dei risultati, e quindi non lasciare che tutto sia svolto in automatico da quella che è l'applicazione che alla fine presenterà solo il risultato migliore.

L'implementazione di un'applicazione che consenta all'utente di scegliere se ottenere un unico risultato (quello ritenuto migliore) oppure avere la possibilità di scegliere tra la moltitudine di quelli ottenuti, porta un maggior coinvolgimento dell'utente che appunto possiede maggiori libertà di scelta.

Infatti, una distinzione può essere fatta se si pensa a quegli utenti *esperti* che vorrebbero prima di poter considerare un lavoro concluso poter valutare personalmente tutte le scelte che sono disponibili e poter scegliere la soluzione che secondo lui potrebbe essere più fruibile come soluzione e più adatta alla situazione.

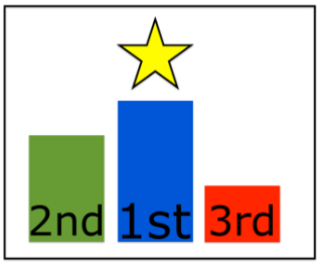


Figura 7: Sketch di un sistema di ranking.

Mentre, si pensi ad un utente che non possiede un grande background di base e che si affida totalmente ad una certa applicazione per poter ricavare le informazioni di cui necessita, questo utente *inesperto* non avrà di certo bisogno di ottenere molte scelte tra cui effettuare la sua selezione, il che potrebbe addirittura portarlo in confusione, ma semplicemente poter accedere ad un'unica soluzione definita come la "*migliore*".

Da qui si capisce l'importanza di possedere un sistema di **ranking delle analisi** (la *Figura 7* mostra una rappresentazione grafica del ranking) che possa fornire all'utente un vero e proprio supporto nella scelta.
Inoltre, questo tipo di implementazione consente anche di migliorare e ampliare gli orizzonti della nostra applicazione, poiché il fornire tali soluzioni porta necessariamente un pubblico più ampio perché si riescono a coinvolgere anche utenti non del settore.

Bisogna quindi costruire un *indice di affidabilità*, che possa riuscire a consigliare l'utente.
Un utente esperto, di certo potrebbe voler valutare tutti gli *indici qualitativi* e potrebbe anche essere interessato ai *parametri di input*, motivo per cui la nostra applicazione dovrà poter provvedere a fornire tali informazioni. Questo potrebbe evolversi anche nel concetto di viste differenti all'interno di un'applicazione e quindi possedere la sezioni che gestiscono i vari livelli di esperienza utente in concomitanza alle funzioni che possono utilizzare.

In ogni caso per fornire le basi ad un sistema di ranking si ha la necessità di implementare un metodo che possa confrontare i vari risultati ottenuti, mettendo in evidenza i vari parametri. Questo in un'applicazione orientata all'analisi dei dati potrebbe essere leggermente più complesso in quanto sarà l'utente a decidere il dataset da caricare e generalmente ogni dataset possiede i propri attributi, i quali possono essere completamente diversi tra loro. Da qui nasce la necessità di poter fornire supporto di base per ogni tipo di base dati caricata.

Uno dei modi principalmente utilizzato per poter presentare il risultato di un'analisi (ovviamente ciò varia sempre in base al dataset analizzato) consiste nel visualizzare i dati in una forma *matriciale*, assegnando un indice di affidabilità per ogni tipo di algoritmo utilizzato.

Questo tipo di forma può anche essere applicato agli attributi presenti per osservarne il loro grado di correlazione.
Questo tipo di soluzione permette all'utente di valutare le soluzioni e i punteggi (in relazione all'indice di affidabilità) ottenuti dall'applicazione dopo aver applicato i vari algoritmi al dataset. La forma matriciale, inoltre, si presenta meglio all'utente che può valutare il tutto direttamente da quello che sembra uno "*specchietto riassuntivo*".

Infine, questo tipo di sistemi possono essere applicate anche delle strategie di **AI**, il che andrebbe a semplificare le cose e anche a migliorarne l'efficacia.

Data visualization Tools: stato dell'arte

Il fine ultimo dell'andare ad analizzare un dataset consiste nel poter estrarre conoscenza, questa conoscenza deve poter essere quanto meno capita e sicuramente il metodo principale è quello di riuscire a presentare questi risultati in un modo consono.

Qui nascono i cosiddetti **Data Visualization Tools**.

Il principale vantaggio di questi strumenti è quello di poter essere utilizzati nella rappresentazione dei dati quando ad essere analizzati sono dataset troppo grandi. Questi mirano a creare una rappresentazione di centinaia di migliaia di dati andando totalmente ad automatizzare il processo ed ottenere quindi una visualizzazione precisa di tali dati.[8]

Generalmente questo tipo di applicazioni risultano essere molto complesse da un punto di visto progettuale in quanto sono da considerarsi una moltitudine di problematiche computazionali.
Poter estrarre dati e visualizzarli in uno spazio *2D* o *3D* impiega molte risorse e l'hardware utilizzabile potrebbe spesso non essere sufficientemente adatto, proprio per questo il lavoro di chi sviluppa questi strumenti deve essere quanto più ottimizzato possibile. In alcuni casi, infatti, è possibile che tali applicativi possano andare ad utilizzare l'hardware presente su macchine server e non direttamente su quella dell'utente, che invece avrà l'unica funzione di caricare i dati ed ottenere i risultati.

Da un punto di vista utente invece, questi strumenti forniscono dei tutorial di base che possono andare a fornire le basi per il proprio utilizzo. Infatti, le azioni che è possibile effettuare in un programma del genere risultano varie, come ad esempio la modifica di indici e parametri che possono poi influire sulla rappresentazione grafica del dataset. Esistono implementati grafici che variano moltissimo tra di loro e che in alcune applicazioni risultano essere talvolta innovativi e non banali da implementare ed essere capiti.

Esistono ad oggi una moltitudine di strumenti per la *data visualization*, i quali sono stati implementati seguendo svariate tecniche e progetti. Volendone effettuare una valutazione complessiva per eleggerne uno solo risulta infattibile in quanto ognuno mette a disposizione una serie di strumenti diversi e adatti ad una situazione piuttosto che un'altra.

Qui di seguito sono stati riportati alcuni dei principali programmi presenti al giorno d'oggi e che puntano alla visualizzazione di dati o alla gestione di risorse direttamente sfruttando interfacce grafiche o dashboard:

PENTAHO:

Pentaho (*Figura 8*) è una piattaforma *Enterprise di Business Intelligence* e *Big Data Analytics* che fornisce servizi *OLAP*, *ETL* e *Data Mining*. [9]
Questa comprende tutta una serie di strumenti capaci di poter analizzare i vari dati e necessari per poter effettuare analisi avanzate.
È presente la possibilità di andare a creare delle *Dashboard*, anche accessibili semplicemente tramite browser, in quanto sono disponibili online oppure si ha la possibilità di poter sviluppare queste *Dashboard* anche integrandole con soluzioni di terze parti, per poter controllare l'andamento in modo più globale e personalizzare la configurazione il più possibile. [10]

Le *Dashboard* implementabili con la versione base forniscono una serie di funzioni per l'appunto "*di base*", mentre per analisi più complesse ed avanzate sono necessari dei *Plug-Ins*.

Figura 8: Rappresentazione logo Pentaho.

Con **Pentaho** è possibile visualizzare i dati in vari modi sfruttando i vari grafici messi a disposizione.
Il modo più semplice per poter comprendere i dati è senz'altro quello di andare a visualizzarli nella maniera migliore possibile e con questo tool ci sono vari aspetti che possono essere considerati, in quanto i dati possono essere rappresentati sotto forma di numerosi grafici, ottenibili in base al tipo di analisi che viene effettuata.
Qui di seguito elencati una parte dei grafici:
Table, *Pivot Table*, *Bar Chart*, *Stacked bar chart*, *Column chart*, *Stacked column chart*, *Line chart*, *Area Chart*, *Pie chart*, *Scatter Chart*. [11]

In alcuni casi è possibile rendere il grafico in 2D o 3D e navigare dentro i vari grafici, per potersi concentrare su ciò che fondamentalmente è importante per l'utente/per l'analisi.
Qui di seguito una serie di link utili e che mostrano i vari casi d'uso:

GOOGLE DASHBOARD:

Google Dashboard è un servizio che sfrutta la raccolta dati degli utenti per poter poi visualizzare in una pagina dedicata un riassunto di tutte le attività svolte dall'utente.
Questo, inoltre, consente di visualizzare tutti quelli che sono i servizi attivi per un certo account **Google** e di verificare quali siano i dati che il servizio **Google** raccoglie dall'utente per migliorare le proprie piattaforme.

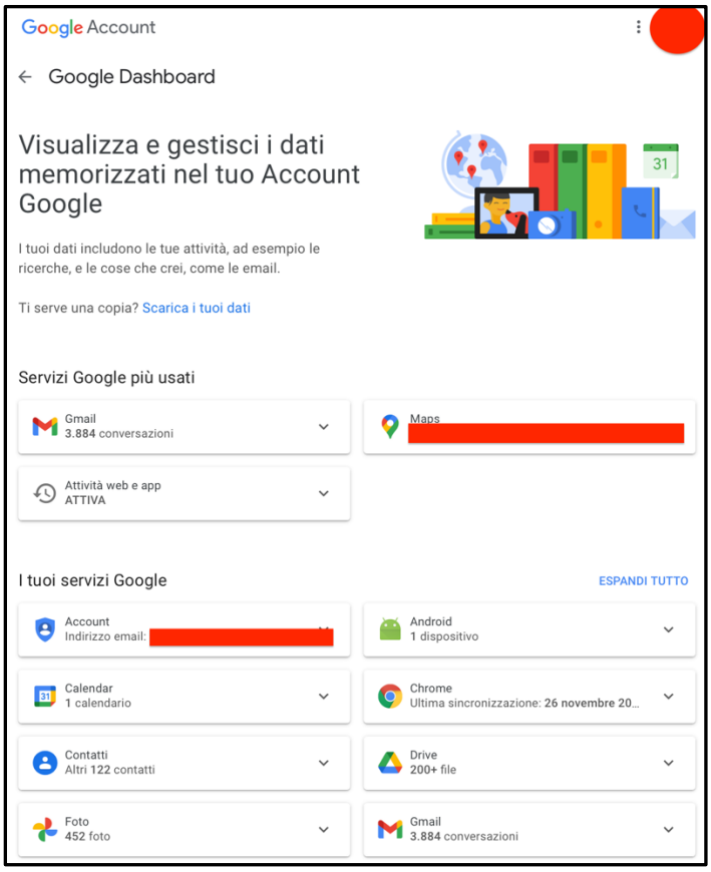


Figura 9:
Screenshot prelevato dalla pagina iniziale di Google Dashboard, in cui è possibile vedere tutti i dati di un determinato utente sotto forma di tabella riassuntiva, e da cui è possibile anche accedere ai vari servizi.

Una delle forme più semplici per la gestione di un account è senz'altro una sorta di pagina riepilogativa, proprio per questo una **Dashboard** (*Figura 9*) è divenuta l'idea migliore da adottare, in quanto con essa è possibile gestire tutto come se fosse una sorta di centro di comando che ci consente di visualizzare tutte le nostre azioni.

La **Dashboard** è formata da una serie di elementi in forma tabellare, suddivisi in:
"Servizi Google più usati" e *"I tuoi servizi Google"*.

Ogni singolo elemento risulta cliccabile e quindi c'è la possibilità di poterci interagire.
Una volta aperta una sezione sarà possibile vedere un riassunto ancora più dettagliato di tutte quelle informazioni, per cui il focus sulle informazioni cambierà da pagina a pagina e sarà appunto possibile andare a concentrarsi solo sui dati di interesse.[12]

GOOGLE DATA STUDIO:

Google Data Studio (*Figura 10*) è un'applicazione sviluppata da **Google** e che essa mette a disposizione degli utenti. Questa contiene una suite di strumenti che possono aiutare l'utente a tenere traccia dei propri dati e di poterli analizzare nel modo migliore possibile. Lo strumento è accessibile direttamente tramite *account Google* e si ha la possibilità
di poter importare in un foglio di lavoro i dati raccolti, presenti in locale oppure presenti online nel proprio spazio *Google Drive* (o altre applicazioni sviluppati sempre dalla stessa azienda)
o altre piattaforme di terze parti.

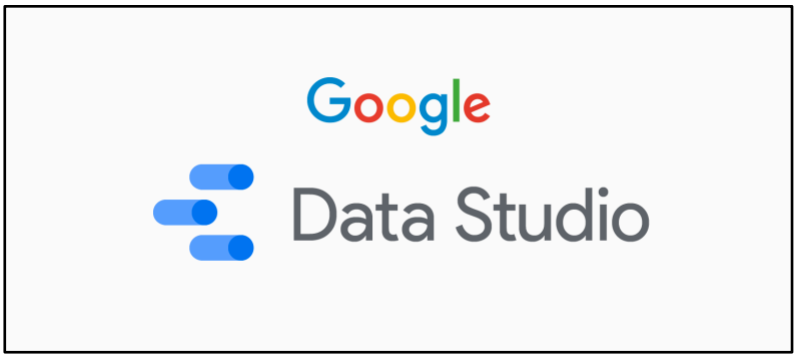


Figura 10: Logo di Google Data Studio.

Una volta effettuata tale operazione è possibile visualizzare i dati in tabelle, scegliendo gli attributi da voler visualizzare in aggiunta alla scelta delle metriche, possibilità di effettuare ordinamento secondo gli attributi delle colonne e aggiunta di filtri e stili.

Tramite una serie di grafici è infine possibile ottenere una visualizzazione più concreta dei dati, anche con grafici interattivi.
I tools di visualizzazione variano in base al tipo di dati che si stanno analizzando, è infatti possibile inserire dei grafici a tabella, a scheda, serie temporali, a barre, mappe, a torta ecc...

È possibile personalizzare i grafici in vari modi giocando sugli attributi presenti nel dataset caricato che può essere in vari formati.
Inoltre, questo strumento ci da la possibilità di poter creare dei fogli di lavoro quanto più personalizzati possibile, inserendo immagini, filtri e caselle di testo, aggiungere loghi e scrivere brevi report su ciò che si sta analizzando, con anche la possibilità di andare a linkare il foglio di lavoro all'interno di un altro website oppure di condividerlo con altri utenti in modo
da poterci lavorare a gruppi.

Le tecniche di visualizzazione dei dati che ci offre **Google Data Studio**, sono ampie e variegate e offrono la possibilità di visualizzare i dataset importati nel modo preferito dall'utente (ad esempio usare un attributo piuttosto che un altro).

Inoltre, è possibile collegare diversi campi di un foglio di lavoro per poter rendere il tutto ancora più interattivo (ad esempio spostare l'indicatore su un grafico fa cambiare il valore
in una cella) e dare stime delle analisi in tempo reale.
Infine, con questo tipo di tool è possibile creare o un foglio di analisi interattivo, e quindi più avanzato in termini di contenuti, oppure la possibilità di creare delle **Dashboard** e quindi avere una sorta di centro di comando con solo le informazioni principali e rilevanti (si prenda come esempio la già citata *Google Dashboard*).

In sostanza l'obbiettivo che si pone **Google Data Studio** è quello di fornire agli utenti un sistema potente e veloce per la creazione di report interattivi orientati all'analisi dei dati.[13]

TABLEAU:

Tableau (*Figura 11*) è un software interattivo di visualizzazione per i dati, questo ha sviluppato numerosi prodotti software tra versioni desktop, online e mobile.

Questo possiede numerose funzionalità come quella introdotta per poter analizzare dati geo-referenziati in vari formati. Il punto di forza di questo programma è che si pone a tutti gli effetti nei panni degli utenti, i quali possono essere variegati a livello di conoscenze e proprio per questo, Tableau ha creato un programma in cui non sarà necessaria alcune competenza di programmazione, ma basterà solamente interfacciarsi con le principali schermate utente e utilizzare i vari strumenti messi a disposizione, consentendo così a qualsiasi utente di poter esplorare ed analizzare i propri dati o di poter andare a creare *Dashboard* interattive.

Figura 11:
Tableau logo.

Questo software, inoltre, vanta di numerosi aspetti positivi come ad esempio un'elevata velocità di accesso ai propri dati e alla loro visualizzazione rispetto ai software concorrenti e consente di visualizzare dati in diversi formati (quali fogli di testo, file di calcolo, dump di database oppure da servizi cloud o applicazioni esterne), fornisce la possibilità di condividere il lavoro con altri utenti (per integrare il lavoro di coloro che possiedono qualifiche diverse all'interno di un azienda come ad esempio per l'addetto all'analisi o l'addetto alla presentazione del prodotto) e consente la creazione di *Dashboard* personalizzate da cui è possibile ottenere una visualizzazione dei dati risultanti più riassuntiva e dettagliata, combinando i vari elementi di visualizzazione messi a disposizione.

Sono anche disponibili strumenti vari per la *Data Exploration* e con dei *Built-in Tool* per la visualizzazione, i dati possono essere gestiti direttamente da interfaccia grafica, ottimizzando il lavoro.
Nella gestione dei dati vengono impiegate tecniche di **AI**/**ML** completamente integrate e impiegate anche nel processo di visualizzazione attraverso un processo di **StoryTelling**.

Uno dei punti forti dell'applicazione risulta essere quella di poter andare a visualizzare i dati combinando una serie di grafici, tra di loro connessi, in modo da poter capire come determinati indici possono incidere più di altri, in quanto ovviamente i grafici suggeriti variano in base al tipo di dataset che si analizza, e questo risulta di fondamentale importanza in quanto un'appropriata visualizzazione dei dati, fornisce risposte che una semplice analisi non può dare.

Prima di poter effettuare un plot dei dati in un grafico, è possibile andare a gestire i dati in tabelle e sfruttare strumenti che implementano funzioni come *split*, *join*, *in*, *funzioni di ordinamento* e *selezione* che consentono di avere il focus solo su determinati attributi.

È possibile poter visualizzare anche un'anteprima di **Data Visualization** per poter rifinire successivamente il lavoro, in quanto a livello di gestione dei grafici si ha possibilità di grande personalizzazione e interazione, come ad esempio la selezione di attributi nel grafico cambia gli

elementi dell'intera *Dashboard* (consentendo all'utente di concentrarsi direttamente su quelli selezionati).

Tableau è un software usato principalmente da *Data Analyst* e aziende che per la gestione di aspetti economici si affidano appunto all'analisi dei dati.
La potenza di tale applicazione risiede proprio nel fatto che sono stati implementati una serie di tool automatizzati che consentono all'utente di effettuare le operazioni semplicemente consentendogli di mettere a punto quello che è il processo di **Data StoryTelling**, che come abbiamo visto non si ferma alla semplice visualizzazione dei dati. [14]

QLIK:

Qlik (*Figura 12*) è un software creato per la *visualizzazione dei dati* e la *Business Intelligence*, creazione di report, analisi guidate o integrate.
Questo permette agli utenti di dare a vita a *Dashboard* personalizzabili capaci di fornire le informazioni più rilevanti e utili sui dati messi a disposizione. Esistono varie versioni dell'applicazione, da quella desktop scaricabile sul proprio pc a quella server, che invece consente una gestione più elastica dei propri file.

Qlik si pone come punto principale il processo di trasformazione dei dati in informazioni, per poter attuare nuove strategie per la propria azienda oppure aiutare il processo conoscitivo in ambito scientifico.
Il punto cardine è quindi quello di dover condividere i propri dati per ottenerne informazioni utili. L'idea che sta alla base dell'applicazione è definita: *democratizzazione dei dati*.

Questa sta ad indicare che con l'avvento dei *Big Data*, se non si è finemente formati (da un punto di vista lavorativo) risulta difficile poter analizzare dati complessi.
A seguito di tutto ciò, **Qlik** mette a disposizione degli utenti la possibilità di poter sottoporre i propri interrogativi direttamente alla piattaforma, utilizzando i fondamenti del linguaggio naturale per estrapolare informazioni sotto forma di grafici ottenuti grazie ad opportune tecniche di visualizzazione.

Figura 12:
Qlik logo.

Inoltre, viene introdotto il concetto di "*alfabetizzazione dei dati*", basato sul fatto di fornire spiegazioni di quelli che sono i dati in modo guidato, per assicurarsi che l'utente possa comprendere a pieno le informazioni (rimando al concetto di **StoryTelling**).

Si vanta una velocità nella lettura, importazione, estrapolazione e riaggregazione dei dati e un uso intenso di quelle che sono tecniche di intelligenza artificiale, su cui **Qlik** basa la tecnica dell'"*indicizzazione associativa*", un sistema capace di mettere in relazione tutti i dati, evidenziando quelle che potrebbero essere le possibili associazioni che potranno più o meno essere esplorate dall'utente.

È anche possibile andare a combinare una serie di sorgenti dati esterne ed indipendenti, dalle quali i vari strumenti estrapoleranno relazioni e che a loro volta verranno visualizzate in
grafici interattivi e capaci di essere compresi e utilizzati dall'utente. La vasta quantità di grafici utilizzabili, fornisce all'utente e soprattutto gli viene suggerito quello più adatto alla rappresentazione dei propri dati, alcuni dei grafici utilizzabili sono *Sankey Chart*, *Mekko Chart*, *Variance Waterfall*, *Radar Chart* e molti altri.

Ognuno di questi risulta essere interattivo e quindi fornisce in real-time come le informazioni cambiano all'arrivo di nuovi dati oppure come l'andamento può variare andando a simulare dei cambiamenti.
Anche in questa applicazione la componente descrittiva dei dati è molto forte, e rende piacevole la *User-Experience* in quanto l'utente viene guidato *step-by-step* nell'esplorazione dei dati.[15]

La raccolta delle informazioni

La raccolta delle informazioni è una delle operazioni preliminari e più importanti per quanto riguarda la buona riuscita del *processo di profilazione* degli utenti.

Quest'ultimo termine assume una grande importanza soprattutto in ambiti quali *Marketing* e *Finanziario*, poiché uno dei principali scopi è quello di andare a categorizzare i propri clienti in gruppi omogenei secondo le preferenze di ognuno, in modo tale che specifiche offerte commerciali o sponsorizzazione di determinati prodotti possano essere messe in atto.

Per fornire un esempio chiave che possa far ben capire il concetto si pensi alla raccolta punti di un supermercato: in tale processo ogni cliente possiede generalmente una propria card, che verrà utilizzata ad ogni suo acquisto come oggetto per il riconoscimento del cliente e come *token* per raccolta punti che la catena di supermercato mette a disposizione. Questo tipo di strategia mira a far si che il cliente venga a poco a poco *profilato*, ovvero: leggendo a mano a mano i resoconti di tutte le spese effettuate, si può facilmente capire quali tipi di prodotti quella persona è abituata ad acquistare e in tal modo la dirigenza può proporre delle offerte dedicate alla singola persona, oppure analizzando i dati nell'insieme è possibile investire in articoli più richiesti dai clienti.

Con questo semplice ma efficace esempio, si può facilmente intuire come una semplice operazione di raccolta dati possa avere grosse ripercussioni sulle attività in corso e di come questo processo offre la possibilità di migliorare i propri prodotti.

L'iniziativa di acquisire informazioni è senz'altro una grande opportunità, che però deve essere fatta nel modo corretto, acquisendo solo le informazioni necessarie e andando a trattare i dati personali (nei casi in cui vengano richiesti) nel rispetto delle leggi sulla privacy.

Nel nostro stato ci si rifà al regolamento Europeo il quale pronostica (in base alla finalità del trattamento) che il titolare debba fornire agli interessati (prima del trattamento) le informazioni che verranno richieste dalle norme (art. 12 GDPR). Il tutto viene espresso tramite l'informativa, la quale altro non è che una comunicazione rivolta all'interessato che ha lo scopo di mettere al corrente l'utente/il cittadino ancor prima che esso ne diventi l'interessato, sulle modalità di trattamento dei dati. Inoltre, l'informativa ha anche lo scopo di rendere valido un eventuale consenso dell'interessato.

Una raccolta dati può avvenire in svariati modi e al giorno d'oggi possiamo notare come numerose aziende sfruttano questo tipo di processo. Altri esempi che possono essere presi in esame sono quelli relativi alla navigazione in internet e al visitare determinati siti, i quali prima di potervi accedere forniscono all'utente una schermata con presenza dell'informativa dei dati personali o accettazione di cookie.

Per quanto riguarda la privacy policy questa deve necessariamente spiegare in modo chiaro quali siano i dati personali che il sito raccoglie, il perché della loro raccolta, il metodo di conservazione e se saranno ceduti a terze parti o meno.
In sostanza, devono essere spiegate chiaramente tutte le informazioni in modo tale che l'utente possa decidere consapevolmente se fornire il suo consenso.
Un altro metodo utilizzato per la raccolta dati online è quello di andare ad utilizzare dei *cookie*.[16]

Questi sono degli oggetti che contengono una piccola quantità di informazioni e offrono la possibilità di avere un'esperienza quanto più personalizzata possibile andando a mantenere delle piccole quantità di dati sull'utente. Ovviamente, questi cookie non sono pericolosi da un punto di vista prettamente informatico e tecnico, ma la preoccupazione sta nel fatto che il sito potrebbe violare la privacy dell'utente analizzandone i dati. In sostanza si può quindi affermare che i *cookie* non sono altro che dei frammenti di dati sugli utenti, che vengono salvati direttamente sul dispositivo della persona che sta utilizzando quel servizio.

Ci sono vari tipi di *cookie*: quelli di prima parte (salvati direttamente dal sito internet visitato) e quelli di terze parti (che trasmettono le informazioni a siti diversi da quello visitato). Qualsiasi sia quindi la tipologia, questi oggetti possono essere utilizzati principalmente per scopi tecnici o per *profilazione*.

Nella prima categoria ricadono i *cookie* che migliorano la navigazione in internet da un punto di vista della sicurezza e dell'efficienza, in sostanza sono elementi necessari per la fruizione del servizio. Mentre gli oggetti che ricadono nella seconda categoria sono prettamente orientati a costruire dei profili per gli utenti, acquisendo dati e condividendoli con i principali siti di *Analytics*.

La raccolta delle informazioni può però avvenire anche in altri modi, che per certi versi vengono definiti come "*metodologie dirette*". In tal caso parliamo quindi di pagine web o applicazioni che possono benissimo essere interfacciate dagli utenti, i quali esprimono formalmente la volontà di voler partecipare al processo di raccolta.

Cos'è un form per la raccolta dati?

Un form per la raccolta dati è uno strumento molto utile e versatile e viene utilizzato per vari scopi come ad esempio progetti di ricerca, nei quali sono necessari dei dati per poter condurre le proprie analisi. Un form è composto generalmente da una serie di campi e ognuno di essi fa riferimento ad una specifica domanda, ogni domanda a sua volta presenterà delle risposte che saranno appunto selezionabili dall'utente interfacciato.

Un foglio per la raccolta dati è un modulo strutturato e ben progettato espressamente per accumulare e conservare i dati e in un momento successivo poterli analizzare comodamente e facilmente. Questi fogli di raccolta sono molto utili e soprattutto sono uno strumento sorprendentemente versatile, in quanto partono con delle caratteristiche molto generiche per quanto riguarda il funzionamento (e quindi il processo di raccolta e gestione) e sarà semplicemente personalizzato dall'amministratore del form, che adatta la struttura alle varie domande che devono essere poste agli utenti.

Ci sono varie caratteristiche che uno strumento del genere può avere e proprio per questo si ha ampia scelta nel costruire il proprio form. Questi moduli, infatti, possono anche essere direttamente condivisi con siti di terze parti ed è quindi necessario riuscire ad ottenere un formato per i risultati che si possa prestare bene ad ulteriori elaborazioni, e il foglio elettronico necessario in questo caso è proprio uno di questi formati che bene o male siamo tutti portati a conoscere ovvero formati *.csv* o stile *.xlsx*.

Ricapitolando quindi, un form altro non è che un modulo per la raccolta dati che sarà composto principalmente da due parti fondamentali:

1. *Una pagina web o una schermata di un'applicazione che si interfacci con l'utente*
2. *Un foglio elettronico che gestisca i risultati*

Il primo assume la sua importanza in quanto consente all'utente di rispondere alle domande o i quiz preposti, gestendo tutti gli elementi necessari ad acquisire tutti gli input degli utenti.

Il secondo invece assume importanza nella conservazione e gestione dei dati che dovranno essere leggibili, ordinati e accessibili (e nel caso di applicazioni web anche scaricabili) dall'amministratore del form per poter condurre le proprie analisi.

La costruzione del form

Al giorno d'oggi le attività di ricerca e raccolta dati sono condotte da vari enti e aziende che coprono vari settori. Ma qualsiasi sia l'ambito il concetto di raccolta dati rimane invariato, e l'unica differenza è la modalità dei quiz/domande che vengono sottoposti agli utenti.

Questa differenza è data dalla possibilità dell'amministratore del form di gestire i campi di input (come meglio crede) da sottoporre al destinatario. Un esempio banale potrebbe essere quello di chiedere all'utente di rispondere ad una domanda tramite una casella di testo oppure semplicemente di rispondere effettuando il check di una casella. Il concetto di fondo è identico ma i dati da gestire sono completamente diversi.

Per la costruzione di un form ci sono diverse fasi da percorrere e non sempre questo lavoro di progettazione risulta essere semplice e immediato.

La prima fase è quella che consiste nel decidere se utilizzare delle applicazioni già preposte alla creazione di moduli oppure (nei casi più particolari) sviluppare e gestire il modulo direttamente tramite codice, costruendolo direttamente da zero.

Il secondo caso di certo lascia molto più spazio al programmatore che gestisce al meglio tutti gli aspetti che dovrà avere il modulo, sotto ogni singolo particolare. Di certo il lavoro di implementazione risulterebbe lungo e ricco di insidie in quanto ci si dovrà preoccupare sia dell'aspetto principale che si interfaccerà con l'utente e della gestione di tutti i campi (che potrebbero appunto essere diversi) ma anche della gestione lato server, ovvero di un programma applicativo messo in run su un'altra macchina e che comunicherà con il client e attenderà che i vari utenti facciamo il submit del modulo.

D'altro canto, sono molte le applicazioni presenti online che consentono agli utenti (che siano essi esperti o meno) di poter creare il proprio form da zero e sfruttando librerie grafiche.
Le caratteristiche di queste applicazioni variano da una all'altra e soprattutto possono fornire funzionalità diverse. L'utente potrà quindi scegliere l'applicazione da utilizzare semplicemente andando a valutare l'efficienza del prodotto in relazione al lavoro che dovrà affrontare.
Inoltre, in base alle funzionalità offerte (ad esempio creazione di moduli illimitati, possibilità di utilizzare specifiche funzioni, supporto tecnico ecc…) tali applicazioni potrebbero essere più o meno a pagamento o abbonamenti temporanei. Vi sono comunque grandi aziende che mettono a disposizione della comunità questi strumenti in modo totalmente gratuito, si pensi per esempio a *Google form*.

La scelta ricade quindi su queste due alternative, scegliere un prodotto già esistente tra quelli disponibili e che può in certi versi essere limitante oppure addentrarsi nella fase di sviluppo della propria applicazione.

Entriamo ora però nel vivo di un'applicazione del genere per riuscire a capire quali siano le modalità di utilizzo di questi fogli elettronici e di come è possibile utilizzare gli oggetti che mettono a disposizione.

Prendendo come esempio il già citato *Google form*, questo ci consente di creare dei moduli di base adatti all'esigenza dell'utente ed è veramente semplice da utilizzare. L'interfaccia risulta essere molto intuitiva ed è suddivisa in due principali sezioni: quella relativa alle domande e quella relativa alle risposte.

Nella prima pagina, quella delle domande, è l'applicazione stessa che suggerisce l'inserimento di una domanda alla quale si potrà rispondere in vari modi (scelta multipla, scelta singola, risposta breve ecc…) sempre ottenendo un suggerimento dall'applicazione stessa.
Vi sono poi delle sezioni totalmente personalizzabili, quali l'inserimento di un titolo, immagini di background e scelte di stili. La personalizzazione non termina qui, in quanto si estende anche alle domande e alle singole risposte come, ad esempio, il rendere una domanda obbligatoria prima di poter proseguire oppure inserire degli elementi grafici come risposte (una pura selezione di immagini).

Le risposte saranno poi amministrate direttamente da quello che un foglio di calcolo elettronico *Google* che archivierà tutte le risposte.

A valle di questa progettazione, la possibilità di poter rispondere al form sarà data dal link di condivisione che verrà automaticamente generato e basterà solo ed esclusivamente riuscire a condividere tale link ai vari utenti per poter inviare una risposta.

Per la costruzione di un form uno step fondamentale risulta essere quello della scrittura delle domande. Questa operazione non risulta essere semplice in quanto la stesura di domande ben formulate risulta complessa e soprattutto la metodologia di input fornita all'utente per poter rispondere deve essere compatibile con la domanda formulata.

Lo strumento usato

Il caso in questione risulta molto particolare in quanto si basa sempre su quelli che sono i principi di una raccolta dati, ma allo stesso tempo richiede qualche funzionalità in più nel form di presentazione.

Come verrà specificato anche in seguito, il principale lavoro sarà quello di riuscire a raccogliere i pareri dei vari utenti per riuscire a capire il loro livello di esperienza nell'ambito della Data Science e migliorare il processo di **Data StoryTelling**.

Una delle caratteristiche che l'applicazione deve avere è proprio quella di porre delle domande agli utenti e la possibilità di rispondere non solo selezionando delle checkbox ma anche scegliendo delle immagini. Tale caratteristica è presente anche in altre applicazioni (come visto in precedenza con *Google form*) ma non è sempre ottimizzata. [17]

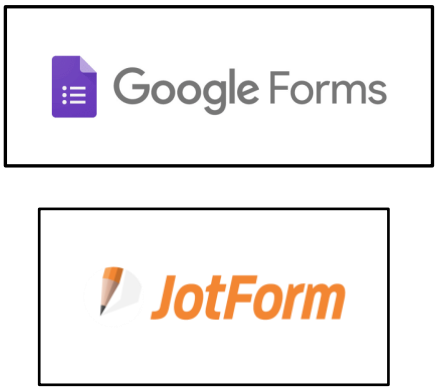


Figura 13:
Google Forms e JotForm logo.

Si è scelto quindi, in prima battuta, di affidarsi ad una delle applicazioni presenti e funzionali che il potente web mette a disposizione.
Una prima fase di sperimentazione è stata condotta con il già citato strumento di **Google** (*Figura 13*), il quale si fa forte delle proprie caratteristiche implementative e il quale rende semplice l'implementazione del modulo per qualsiasi tipologia di utente.

La principale limitazione è stata però quella di non riuscire ad ottenere una buona (per non dire ottima) qualità delle immagini caricate, facendo così decadere la possibilità di poter essere usato

come strumento principale in quanto la possibilità di gestire gli elementi grafici non è presente e la visualizzazione di un grafico (ricavato da vari dataset) se non dettagliato, non può assumere alcun valore.

Dopo una serie di ricerche tra le varie applicazioni e i loro "pro e contro", la scelta è ricaduta su "**JotForm**" (*Figura 13*). [18]

Quest'ultima è un'applicazione web formata da varie sezioni e che mette a disposizione degli utenti numerosi strumenti per la creazione non solo di form per la raccolta dati, ma fornisce anche la possibilità di poter creare vari moduli da inserire in pagine web e quindi avere una grande libertà di scelta nel creare il proprio documento virtuale.

Ogni singolo modulo possiede un layout di base che può essere modificato in base alle esigenze del programmatore, e quindi si può generare una pagina con uno stile classico oppure una pagina adatta solo ed esclusivamente per uno scopo specifico.
Per citare un caso d'uso riscontrato si potrebbe affermare che un form può avere un layout di base formato da pagine oppure essere compatto ed essere gestito da una singola pagina. Inoltre, la personalizzazione non possiede limiti, in quanto è possibile modificare a proprio piacimento le dimensioni e forme dei layout, personalizzandone addirittura i colori e caricando immagini di background.

L'implementazione del form richiede il proprio tempo e questa applicazione consente di gestire qualsiasi tipo di input che sia una checkbox o una selezione di immagini.
Infatti, la principale problematica relativa alla qualità drasticamente ridotta viene completamente superata con l'utilizzo di questo nuovo strumento.

Ogni campo è stato gestito singolarmente selezionando l'opportuno widget che l'applicazione mette a disposizione, consentendo di gestire i campi nella maniera più consona possibile. Sono per l'appunto selezionabili una quantità illimitata di widget per la creazione del modulo, e data questa possibilità l'utente può quindi decidere di strutturare il form secondo le proprie idee ed esigenze.

Un'altra alternativa di personalizzazione consiste nell'andare ad iniettare all'interno del layout principale (o di ogni singolo widget) del codice **CSS**, con il quale è possibile ottenere un livello di personalizzazione maggiore. Questa possibilità che viene messa a disposizione degli utenti è una feature veramente potente, in quanto la capacità di personalizzare a tal punto da poter scrivere del codice rende il processo di sviluppo del programmatore molto simile a quello che avrebbe incontrato se avesse implementato il programma da zero. Ovviamente, non è possibile agire sulle funzioni del form ma è possibile modificare la logica di base aggiungendo degli elementi condizionali tramite uno specchietto riassuntivo presente nelle impostazioni.

Un'ulteriore caratteristica è quella di poter effettuare delle traduzioni sia automatiche che manuali dei propri moduli in modo tale che questi possano essere compilati direttamente in più lingue, e quindi adatte ad un pubblico più ampio. Questa particolare funzione è accessibile direttamente da dalle impostazioni del modulo che consentirà l'apertura di un menù dal quale sarà possibile scegliere la lingua per la traduzione. Tali traduzioni non saranno fatte automaticamente dal programma (come si vede nella *Figura 14)*, ma dovranno essere svolte direttamente dallo sviluppatore e poste nelle relative caselle.

Figura 14: Screenshot del menù traduzioni direttamente riportato dalle impostazioni del form in questione. È possibile notare che l'unica opzione implementata in questo caso è relativa alla lingua inglese.

JotForm consente, inoltre, un'efficace gestione delle risposte grazie all'implementazione delle *Table* che riprendono il principale formato *.xlsx* o *.csv* e in tal modo è possibile accedere direttamente ai risultati e poterli valutare.

L'accesso ad essi può avvenire attraverso la pagina principale di gestione dei form che consente di avere una panoramica su tutti i form che sono stati creati ed effettuare operazioni su di essi come modifica/aggiornamento, pubblicazione e cancellazione.
Una volta entrati in tale sezione, dopo aver effettuato l'autenticazione con il proprio account, sarà possibile anche scaricare i risultati in uno dei formati già citati precedentemente, in modo da poter avere una copia in locale e poterci lavorare direttamente.

L'applicazione mette a disposizione anche degli strumenti per poter agire e modificare le *tuple* della tabella che accoglie i risultati o ancor più semplicemente ne consente una semplice visualizzazione andando a fornire azioni di supporto come raggruppamenti, spostamenti di colonne o formati diversi da quelli tabellari.

Un'altra caratteristica importante è quello che ci consente di effettuare un collegamento diretto tra un foglio di lavoro *Google Sheets* e il nostro modulo. In tal modo sarà possibile ottenere una vera è propria sincronizzazione dei fogli elettronici ed averne una copia a portata di mano direttamente nel proprio account **Google**.

Come già stato citato più volte, il form sarà accessibile direttamente tramite un link generato automaticamente dall'applicazione e l'applicativo sarà posto in *run* direttamente sui loro server. Questo fornisce anche grandi vantaggi in quanto previene problematiche di gestione di *crash* dell'applicazione e fornisce un supporto anche alla *compilazione parallela* del modulo.

La costruzione delle domande e dei grafici

Una volta valutato lo strumento e fatte le proprie considerazioni, è iniziata la fase di sviluppo del form. Questa molto spesso è una fase sottovalutata in quanto si crede che basti buttar giù qualche domanda ed aver automaticamente fatto tutto il grosso del lavoro.

La prima fase invece, risulta essere quella più intensa e quella sempre più sottoposta a revisione, poiché è proprio in questo stadio che verranno formulate le domande e una domanda non deve essere né troppo generica e né troppo specifica altrimenti non si andrebbero ad acquisire la totalità delle informazioni che una raccolta dati mette a disposizione.
Spiegandone meglio il concetto: una buona varietà di domande consente di ottenere determinate informazioni (anche più di una) e quindi questa deve essere formulata nella maniera più corretta possibile. [19]

Le domande che sono state generate in questa occasione si rifanno direttamente alla capacità e preparazione degli utenti nell'ambito della Data Science e si è scelto di suddividere il form in più sezioni, precisamente tre:

- *La **prima** sezione fa riferimento a delle domande di carattere generale*
- *La **seconda** sezione fa riferimento a delle domande basate su grafici generati a partire da un Dataset strutturato*
- *La **terza** sezione fa invece riferimento a delle domande basate su grafici generati a partire da un Dataset time-series*

Nella **prima** sezione la maggior parte delle domande sono state basate su una scelta singola e sulla compilazione di questi campi che forniscono informazioni solamente sullo stato della persona che compila il modulo e sul suo livello di conoscenza di base di alcuni dei principali concetti relativi all'ambito della Data Science.
Informazioni generiche come sesso, età e professione ci consentono di inquadrare direttamente la persona e capire quanto potrebbe essere familiare a questo mondo e dalle risposte fornite successivamente, le quali sono state accuratamente poste in una scala da *uno* (valore minimo) a *cinque* (valore massimo), ci consentono di andare ad effettuare una prematura valutazione sul singolo utente e quindi di poter schematizzare delle categorie in cui questi possono essere inseriti.
L'idea della scala graduata porta numerosi vantaggi sia lato utente che lato sviluppatore, in quanto l'utente riceve maggiore libertà nella risposta mentre lo sviluppatore avrà già dei range di valori ben suddivisi e pronti per poter essere utilizzati (senza dover andare ad effettuare normalizzazione su dati e parametri).

Nella **seconda** sezione invece, sono stati generati dei grafici a partire da dataset diversi. La prima cosa con cui l'utente a che fare è proprio una breve descrizione del dataset in questione, che risulta necessaria per poter fornire le informazioni di base all'utente, che potrà essere più o meno esperto e quindi vorrebbe conoscere qualcosa in più sui dati che sta analizzando e fornisce anche una buona dose di **StoryTelling**, in quanto i dati vengono meglio capiti se raccontati e/o spiegati.
I grafici che sono stati qui generati sono vari e riprendono solamente quei dati provenienti da dataset strutturati. Questi dati sono stati ricercati nei vari siti web che mettono a disposizione degli utenti dei dataset, per l'appunto pubblici e scaricabili.
Una volta trovato il dataset che si presta meglio per la rappresentazione sono stati generati i grafici che, ovviamente, rappresentano per ogni domanda la stessa informazione, ma in formato (parlando

di modalità di visualizzazione) diversa. Si è cercato di differenziare quanto più possibile i vari grafici, prendendone allo stesso tempo in considerazione molti, proprio per sottolineare un aspetto fondamentale e ovvero: *la visualizzazione*.

La visualizzazione di un grafico è automaticamente sinonimo di visualizzazione dei dati, e questi possono avere mille forme. Proprio per questo motivo esistono varie forme di rappresentazione e ognuno di esse può essere più o meno apprezzata da un utente piuttosto che da un altro e questo è dovuto al fatto che una persona può avere una sensibilità differente.
Il nostro lavoro è quindi proprio quello di capire tale sensibilità e migliorare la rappresentazione dei dati in modo che tutti i tipi di utenti (con più o meno esperienza) possano effettivamente essere in grado di capire il significato dei dati.

Un dataset strutturato è formato (come suggerisce anche il nome) da una struttura ben definita e ripetitiva, e che viene appunto utilizzata per conservare ed amministrare i dati.
I dataset che sono stati scelti sono diversi tra loro e rappresentano situazioni differenti, e anche questo è un concetto da non sottovalutare in quanto una base dati può essere più facilmente intuibile rispetto ad un'altra e può essere rappresentata in maniera diversa rispetto ad un'altra.

Proprio per questo la scelta è ricaduta su basi di dati diverse, per poter abituare l'utente a concetti tra di loro diversi e quindi a sua volta rappresentazioni diverse. Le basi di dati analizzate sono qui di seguito riportare: "*Gender Dataset*", "*Sales Dataset*" e "*Iris Dataset*".

Il primo tra questi fa riferimento a dati che rappresentano delle caratteristiche fisiche distinte per genere. Il numero di records che lo compongono è abbastanza consistente e le informazioni contenute permettono di conoscere gli attributi quali "*altezza*" e "*peso*" separatamente per genere.

Il "*Sales Dataset*" è stato rilasciato da un super-store e contiene delle informazioni relative alle vendite registrate in determinati slot temporali, precisamente tre anni, per tre categorie principali ("*Meat*", "*Snacks*" e "*Vegetables*").

L'ultimo dataset presente invece contiene una serie di attributi appartenenti a varie classi di fiori. Il valore di tali attributi ("*Sepal Length*", "*Sepal Width*", "*Petal Length*" e "*Petal Width*") vanno così a discriminare una classe piuttosto che un'altra.

Per ogni singola domanda è stato chiesto all'utente quale tra i seguenti grafici va a rappresentare meglio la distribuzione di determinati attributi oppure quali tra i grafici preposti rappresenta meglio i dati nella sua totalità, facendo dichiarare all'utente quale grafico risulta quindi essere più esaustivo e adatto alle sue esigenze. Una limitazione potrebbe essere la scelta singola di un grafico, e proprio per questo si è scelto di lasciare via libera all'utente di individuare e selezionare i grafici che più ritiene accettabili.
I tipi di domande sono state poi analizzati più volte per cercare di prendere più informazioni da un dataset anche per diverse analisi. Sono infatti state poste delle domande relative ai concetti di *clustering* e di *outliers detection* oltre che alle semplici *rappresentazioni delle distribuzioni*, presentando appunto opportuni grafici.[20] [21]

La **terza** e ultima sezione invece si rifà al concetto di dataset time-series, ovvero una base dati che possiede delle caratteristiche diverse da quelle viste nella sezione precedente in quanto riprende il processo di raccolta dei dati in circostanze diverse, come ad esempio l'ambito industriale. [22]

In questi casi vi saranno dei macchinari a svolgere numerose mansioni lavorative e ognuno di essi effettuerà un resoconto ogni preciso istante di tempo. Questi "*log*" generati dalle macchine/sensori dovranno essere in qualche modo analizzati per poter ottimizzare il lavoro oppure per riscontrare dei problemi. Da qui nascono quindi le basi di dati time-series che sono generalmente formati da valori numerici in serie e da un timestamp che ne identifica il momento della loro generazione.

I dataset utilizzati in questa sezione sono sempre stati scelti dalla moltitudine presente online nei vari siti e proprio come nel caso precedente sono stati scelti in base alle caratteristiche di rappresentazione. Il concetto esposto in questa sezione risulta essere per certi versi più complicato, in quanto i dati da analizzare hanno un formato diverso e non possono essere utilizzati i grafici già visti nella sezione precedente in quanto non sarebbero appunto adatti alla rappresentazione. Questi, infatti, dovranno avere una struttura quanto più possibile adatta all'analisi nel tempo, una struttura che possa far capire come un determinato attributo/valore possa essere simile o completamente diverso da quello successivo/precedente a cui è direttamente collegato. Grafici del genere richiedono quindi un'invettiva maggiore e soprattutto maggiore impegno per la loro generazione, che deve essere quanto più precisa possibile.

Tutti i grafici (riferimento alla *Figura 15*) sono stati generati utilizzando strumenti e librerie differenti. Per la generazione di alcuni sono state prima necessarie delle trasformazioni dei dati oppure l'applicazione di determinate funzioni, e quindi in questi casi è risultato necessario utilizzare come linguaggio di programmazione **Python**, che grazie a specifiche librerie (***Seaborn***, ***Pandas*** ecc…) ci consente di analizzare, gestire e trasformare i dati con conseguente raggiungimento dell'obiettivo preposto, ovvero la generazione del grafico (*Figura 16*). [23] [24] [25]

Gli strumenti utilizzati per questo scopo sono stati **Google Colab Online**, un vero e proprio applicativo che riprende la struttura di un notebook e che quindi consente l'esecuzione di pezzi di codice anche separatamente. [26]
Un'altra applicazione che si è resa di appoggio e come alternativa allo strumento sopra citato è **PyCharm**, un prodotto completamente diverso che fornisce un IDE al programmatore.

In altri casi, per concludere il discorso sulla trasformazione e gestione dei dati, si è rilevato di aiuto anche la speciale suite di grafici implementata per **Microsft Excel**.

Una libreria che invece si è resa molto utile è stata **Highcharts**, questa consente agli utenti di visualizzare in modo completamente gratuito alcuni dei grafici più interessanti per le diverse tipologie di Dataset. Fornisce inoltre, la possibilità di analizzare il grafico aprendo **JSFiddle**, e quindi automaticamente di poter accedere al codice **JavaScript** per poter usufruire del grafico. Grazie a questa libreria è stato possibile gestire anche i grafici più complessi e soprattutto personalizzarli sotto ogni aspetto.[27]

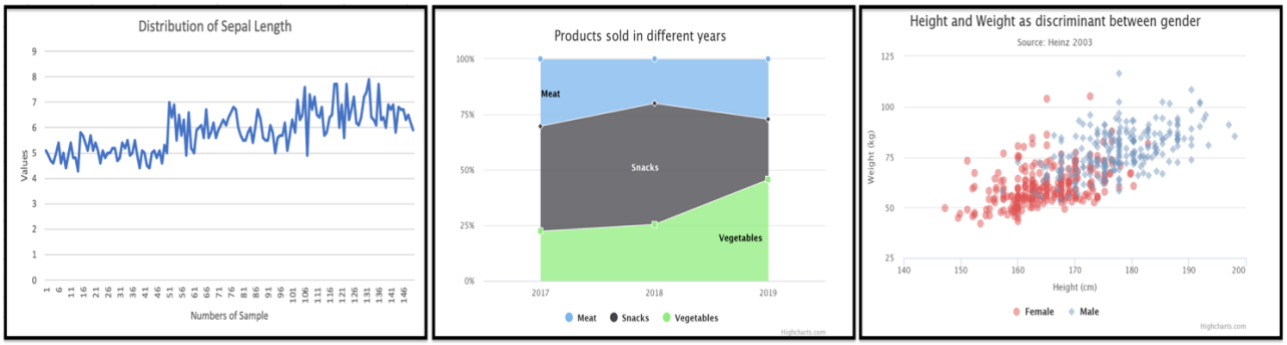


Figura 15: Alcuni dei grafici generati con i vari strumenti sopra citati e appartenenti alle varie domande della sezione numero due.
-Il primo a sinistra relativo al *Iris Dataset*
-Il centrale relativo al *Sales Dataset*
-A destra quello relativo al *Gender Dataset*

Una volta definite le tre sezioni il form ha quindi raggiunto il suo obiettivo, ovvero essere utilizzato per la raccolta dati. Ma per renderlo ulteriormente accattivante e fornirgli un aspetto più professionale sono stati inseriti alcuni elementi.

Il primo elemento ad essere stato inserito è quello usato per selezionare la lingua, che come anche già citato in precedenza, assume un'importanza fondamentale per far sì che il modulo possa essere condivisibile con più persone. Questo è presente come "*button*" nella prima pagina del form, in modo tale che prima di poter continuare l'utente possa selezionare la lingua preferita (che in questo caso sono *italiano* e *inglese*). In ogni caso il linguaggio potrà essere modificato in qualsiasi momento dall'utente durante la compilazione.

Il secondo elemento messo in aggiunta è una "*barra di progressione*", ovvero un widget che tiene conto delle risposte che sono state date dall'utente in modo che esso possa rendersi conto che si trova in un determinato punto della compilazione del form. In questo caso le domande poste agli utenti sono 24, per cui verrà aggiornata la percentuale di completamento ogni qual volta l'utente prosegue con la compilazione.

Infine, l'ultimo elemento che è stato aggiunto è un logo. Quest'ultimo è un elemento da non sottovalutare, nonostante questo sia solamente un oggetto puramente grafico e che non possiede funzioni, gioca comunque un ruolo importante in quanto fornisce all'utente un volto, un marchio riconoscibile per ciò che sta facendo.

Per poter caricare tutti i grafici che sono stati generati per ogni singola sezione, si è rivelato necessario utilizzare **Dropbox** come applicativo per poter caricare direttamente all'interno di una cartella condivisibile tutti i grafici generati. [28]
Da qui, è stato poi possibile ottenere un link per la condivisione delle immagini e il successivo collegamento nei campi **JotForm**. Un'alternativa a questo sarebbe stata possibile grazie allo spazio di archiviazione online di *Google Drive*. [29]

Figura 16: Immagine rappresentante tutti i loghi dei principali strumenti utilizzati per la generazione dei vari grafici.

La fase di revisione e di test

La fase di revisione risulta essere una delle fasi più importanti in quanto l'intero lavoro viene rivisto e sottoposto ad occhio critico. Si riprendono quindi le tutte le casistiche e si risolvono i vari problemi, oppure semplicemente si va a migliore piccoli aspetti.
L'aspetto sicuramente più importante del lavoro in questo form è stato quello di ottimizzare i grafici e renderli quanto più possibile leggibili e in alta qualità. Questo perché la generazione di immagini e il successivo caricamento ne comporta delle riduzioni sia dal punto di vista della qualità ma anche da quello delle dimensioni, in quanto per ogni grafico si è rivelato utile andare a ridimensionare l'immagine originale. Questa operazione è stata fatta per ogni singolo grafico, quindi è stato necessario andare a generarli nuovamente impostando come misure di *width* e *height* dei parametri che potessero essere adatti alla nuova struttura base del form.

Un altro punto posto a revisione è stata la struttura principale del layout, in modo che le tre sezioni possano essere suddivise in tre pagine differenti. Questo è stato possibile grazie anche alla possibilità offerta dall'applicativo di poter andare a modificare il layout di base e di renderlo adattabile anche ai vari schermi su dispositivi diversi. Ulteriore ragione per cui è stata effettuata tale scelta, la possibilità fornita agli utenti di poter rispondere al sondaggio direttamente da dispositivi mobili senza dover necessariamente accedere da pc per la compilazione (come si vede in *Figura 17)*.

Figura 17:
Tale immagine rappresenta uno screenshot proveniente da un dispositivo mobile (smartphone) su cui è stato aperto il link al form e mostra come è possibile anche compilare il modulo facilmente senza essere collegati da pc.

Ancor prima di iniziare la compilazione viene ora fornita la possibilità all'utente di poter scegliere la lingua, per poi arrivare ad una pagina di benvenuto in cui è presente una breve descrizione del progetto. Un'ultima pagina inserita è stata invece quella di ringraziamento, che compare a form completato. Da non sottovalutare sono infatti queste parti in cui si cerca di avere un rapporto diretto con l'utente, fornendogli spiegazioni e soprattutto andando a renderlo partecipe il più possibile.

Infine, come ultima fase dell'analisi critica sono state riprese le domande delle varie sezioni, che sono state anche riordinate in base a ciò che veniva richiesto per poter avere una logica di fondo nel completamento del form, ad esempio le domande su uno stesso dataset (si faccia riferimento alla sezione numero due) sono state poste una di seguito all'altra per rendere il tutto più compatto e complementare, oppure domande generiche ma raggruppabili sotto lo stesso macro-concetto.

Una volta effettuato il lavoro di revisione è stato avviato il *periodo di test* in cui il form è stato pubblicato e quindi accessibile tramite un apposito link.
Questa fase ci permette di verificare che tutto vada a buon fine e consente di gestire le criticità che vi si presentano in corso d'opera, prima ancora che il modulo possa essere largamente condiviso.
In questa fase sono infatti state riscontrate delle problematiche nella gestione delle risposte solo in alcuni casi, motivo per il quale è stato convenuto di scaricare periodicamente i dati dalle tabelle in fogli di lavoro elettronici locali.
Per ovviare a questa problematica è stato introdotto il collegamento a *Google Sheets* in modo tale da poter gestire le risposte direttamente dal proprio account **Google** in maniera più *sicura* ed *efficiente*. La possibilità di poter utilizzare i collegamenti e quindi poter ottenere delle copie di *backup* rassicura lo sviluppatore, che tal modo va a minimizzare i rischi di perdita dei dati.

Durante la fase di test sono state raccolta decisamente una buona parte delle risposte, che hanno già permesso di valutare l'applicativo per quello che è nella sua interezza, e questo è stato fatto grazie all'invio del link e alla collaborazione di amici e colleghi che si sono offerti di partecipare al sondaggio. Aggirandosi intorno ad un *quaranta* invii di modulo, solo in questa prima fase, è stato possibile testare non solo il funzionamento di questo ma anche le politiche di gestione delle risposte.

Tutti i dati che sono stati raccolti in questa fase non saranno scartati, ma andranno comunque a far parte del processo di ricerca a tutti gli effetti, in modo da poter avere anche un campione variegato.

La fase di pubblicazione

Superata la precedente fase di test si è arrivati alla fase di pubblicazione vera e propria, è quindi doveroso prepararsi alla possibilità di essere inondati da invii di moduli dai vari utenti.

La fase di pubblicazione risulta per certi versi la più contorta, in quanto si dovrà scegliere l'utenza a cui sottoporre tali domande e soprattutto ci si dovrà impegnare per cercare di invogliare tali utenti a rispondere e sottoporsi al test.

Come già citato altre volte, la raccolta dati spesso risulta essere tediosa e non sempre condivisa da tutti gli utenti che spesso cercano invece di evitare di "perdere tempo" nel compilare i form.
A tal proposito, il primo passo che è stato fatto è stato quello di scrivere una *bozza di e-mail*, in cui dopo aver spiegato in grandi linee il motivo che sta alla base della creazione di tale form e degli obiettivi prefissati, si invitano i gentili utenti a rispondere al modulo.
Anche tale processo è stato sottoposto più volte a revisione e soprattutto tradotto anche in lingua inglese, in modo tale che anche un pubblico più ampio vi possa partecipare. In tale bozza iniziale, poi raffinata, si è fatto riferimento alle principali motivazioni che hanno spinto a portare avanti il progetto di ricerca, sottolineandone i punti cardine e dando qualche piccola spiegazione al riguardo, in modo che gli utenti conoscendo quanto più possibile il contesto possono essere sensibili alla causa e parteciparvi in modo attivo.

Gli utenti target sono stati principalmente quelli frequentanti corsi molto attinenti al contesto di ricerca (come ad esempio *ingegneria informatica*, *matematica* ecc…) e che teoricamente dovrebbero far parte del mondo della *Data Science* o che comunque possiedono un background di base, in modo da poter avere un riscontro oggettivo e quanto più possibile preciso sulle domande poste. Ma non è soltanto la categoria degli studenti ad essere stata presa in considerazione, in quanto questi potrebbero fornire una sottosezione di un'utenza (sicuramente la più numerosa) ma il sondaggio è rivolto anche a professionisti del settore con le competenze più variegate, si passa infatti da utenti che svolgono mansioni lavorative all'interno di aziende a professori universitari che trattano proprio tali argomenti.

È quindi possibile affermare che una valutazione più ampia possibile sarebbe sicuramente la condizione migliore in cui ci potrebbe trovare, facendo sì che vari tipi di utenza possano effettivamente fornire il proprio contributo, rendendo la parte di categorizzazione da parte dello sviluppatore più precisa.

Figura 18: Screenshot della prima pagina del form in cui si chiede all'utente di andare a selezionare la lingua. In questa immagine è possibile vedere come il form si presenta subito dopo aver cliccato sul link.

La fase di pubblicazione risulta essere la fase più importante e soprattutto ci si dovrà dare delle scadenze temporali, legati al tempo in cui il form può appunto essere accessibile agli utenti. In questo caso si è deciso di tenere online il form per circa un mese, in modo che sia stato dato molto tempo agli utenti per rispondere e soprattutto si è avuto il tempo materiale per poter inviare a più gruppi di utenti il modulo (in *Figura 18 è stata riportata la prima schermata del form*).

Una volta trascorso questo range temporale, saranno ritenute utili e valide solo le risposte raccolte in questo arco di tempo. Questa fase non può essere prolungata troppo proprio come non dovrà essere troppo breve, generalmente questo può dipendere dalle tempistiche del progetto in corso ma può dipendere anche da fattori esterni in quanto gli utenti potrebbero compilare subito il form, fornendo immediatamente il loro contributo, oppure potrebbero rispondere in un secondo momento (in relazione ai vari impegni di studio/lavorativi) o nel caso peggiore potrebbero ignorare completamente l'invito.

Uno degli approcci che è stato utilizzato in questo progetto di ricerca è stato quello di poter invitare gli utenti a partecipare selezionando e suddividendo l'intero campione in più blocchi. Ad esempio, il primo blocco è stato quello relativo alla fase di test e quindi campioni molto variegati, mentre il secondo è stato principalmente diretto ai vari studenti e così via.

Analisi Statistica

Prima di poter parlare di *analisi statistica dei dati* sarebbe quantomeno necessario fare anche solo un cenno alla macroarea che gestisce ed ingloba questa disciplina, ovvero la statistica.

Quest'ultima ha infatti il fine di riassumere un insieme di dati sia che esso sia molto vasto o che sia un piccolo sottoinsieme. Tale processo si avvale di numerose metodologie soprattutto per quanto riguarda la presentazione delle informazioni che avviene tramite grafici opportunamente generati per poter evidenziare le principali caratteristiche oppure tramite delle tabelle riassuntive che espongono in maniera più diretta tutti i risultati ottenuti durante l'analisi. [30]

L'obiettivo primario di questa scienza risulta quindi quello di essere legato alla previsione, ovvero riuscire a stimare eventuali parametri all'evolversi della situazione ed al passare del tempo, una volta verificatasi una certa condizione. Questa disciplina è infatti adottata in vari ambiti come nel mondo dell'economia, nel settore aziendale (vari tipi di aziende), in medicina e più in generale negli ambiti scientifici.

Con il termine *analisi statistica dei dati* andremo quindi a sottintendere un processo ragionato e che mira con un'analisi preliminare all'ispezione, trasformazione e pulizia dei dati e in secondo luogo il contrassegnare quelle informazioni di spicco che possono risultare importanti.

Numerosi termini sono stati coniati nei vari anni ed è necessario citare quei processi come *Data Mining* e *Business Intelligence,* il primo risulta essere focalizzato nel processo di scoperta dei dati principalmente per scopi predittivi mentre con il secondo termine si esprime maggiormente il processo di aggregazione dei dati necessario per poi effettuare delle scelte strategiche.

Dunque, il concetto di *analisi statistica dei dati è colui* che si pone l'obiettivo di affrontare ed analizzare in modo efficace ed efficiente la sempre crescente quantità di dati che vengono generati istante per istante, consentendo quindi ad aziende (o team di scienziati) di progettare misure o prendere decisioni basate su tali risultati.

I campioni raccolti

Introduciamo ora un ulteriore termine finora non ancora citato: *Big Data Analytics*.
Questo è definito come il principale processo per la raccolta di grandi quantità di dati e per l'estrazione di informazioni da essi.

Un campione, termine con cui si intende l'insieme degli elementi su cui condurre le analisi, può essere molto variegato e soprattutto difficile da gestire quando sono molte le sue sfaccettature. Nel caso corrente l'intera popolazione dei campioni corrisponde al numero esatto di individui che hanno risposto correttamente al form, inviando quindi il modulo.

Una volta ottenute tutte le risposte, e quindi terminato il range temporale in cui i vari utenti possono accedere al modulo tramite opportuno link, il primo passo da affrontare è proprio quello di riuscire ad individuare i campioni non coerenti e non utili per lo scopo ultimo dell'analisi.
Questi campioni sono generalmente formati da campi vuoti (lasciati senza risposta dagli utenti) e anche da risposte fornite senza criterio (nel caso vi siano utenti che compilino male il form accidentalmente o appositamente).

I dati raccolti (la *Figura 19* mostra tali dati) riflettono in maniera precisa tutti i campi del modulo che è necessario gestire con grande precisione e questi possono essere organizzati suddividendo il tutto in base al contenuto informativo che possiedono e in base a quanto il processo conoscitivo può estrapolare da essi.

La prima suddivisione che può essere fatta è relativa ai primi *attributi* compilati dagli utenti e ovvero le informazioni personali di quest'ultimi:

- **Sesso**
- **Età**
- **Professione**

Da questi tre parametri è possibile estrapolare informazioni riguardanti il tipo di utenza coinvolta e di conseguenza anche la familiarità con il mondo della *Data Analysis*. Svolgendo un'analisi completa di quelli che sono tutti questi dati raccolti è possibile conoscere in percentuale i numeri di donne e uomini, la loro età e la professione che svolgono.

Questo tipo di informazione viene principalmente utilizzata per avere un riscontro diretto, con dati alla mano, di chi potrebbero essere gli utenti interessati ad utilizzare una applicazione per la *Data Analysis*. Infatti, analizzando meglio gli attributi relativi all'*età* e alla *professione*, è possibile notare come una determinata classe di utenti sia predominante rispetto ad altre, che possono essere più o meno ristrette se paragonate tra loro. In generale i dati raccolti mostrano come la maggioranza sia composta da "*studenti*" il cui percorso di studi coincide con ambiti scientifici. Ma analizzando anche l'ultimo attributo, possiamo notare come anche una buona sezione della totalità dei dati sia caratterizzata da professionisti in altri ambiti (ad esempio economico e giuridico o medico), il che fornisce numerose informazioni in quanto esprime l'intenzione di tali utenti ad avvicinarsi all'ambito della *Data Science*, nonostante in alcuni casi non si abbia l'esperienza necessaria.

	A	B	C	D	E
1	Submission Date	ProgressBar	Domanda 1 - Sesso?	Domanda 2 - Fascia di età?	Domanda 3 - Qual è la tua professione?
2	2021-01-24 13:58:09		Uomo	Tra 20 - 28 anni	Studente*
3	2021-01-24 16:33:32		Uomo	Tra 20 - 28 anni	Studente*
4	2021-01-25 09:13:27		Donna	Tra 20 - 28 anni	Data Scientist
5	2021-01-25 12:06:10		Donna	Tra 20 - 28 anni	Data Scientist
6	2021-01-25 15:37:55		Donna	Tra 20 - 28 anni	Data Scientist
7	2021-01-26 14:24:51		Uomo	Tra 20 - 28 anni	Data Scientist
8	2021-01-27 19:11:47		Uomo	Tra 20 - 28 anni	Domain Expert*
9	2021-01-27 23:17:25		Uomo	Tra 20 - 28 anni	Data ScientistStudente*
10	2021-02-05 15:04:55		Donna	Tra 20 - 28 anni	Inoccupato
11	2021-02-10 19:33:07		Uomo	Tra 20 - 28 anni	Studente*
12	2021-02-10 20:03:00		Uomo	Tra 20 - 28 anni	Studente*
13	2021-02-10 20:06:07		Uomo	Tra 20 - 28 anni	Studente*
14	2021-02-10 20:11:27		Uomo	Tra 20 - 28 anni	Studente*

Figura 19: Immagine che rappresenta la forma tabellare dei risultati raccolti dal form. Tale file possiede estensione *.xlsx* ed è possibile accedervi grazie a strumenti con *Microsoft Excel*, da cui sarà poi possibile modificarne l'estensione.

I parametri rimanenti invece fanno riferimento diretto all'esperienza acquisita degli utenti nei vari ambiti, a quanto essi utilizzino specifici strumenti e alle preferenze di grafici piuttosto che altri per la rappresentazione dei dati. Tutto questo è definito come un'autovalutazione da parte degli utenti, e quindi necessario poter fornire ad essi un modo per poter comunicare le loro intenzioni. A tal proposito quasi tutte le domande sono state disposte in un range di valori che va da "*uno*" a "*cinque*", che identificano rispettivamente il valore minimo e quello massimo, in modo tale che l'utente non si ritrovi a dover scegliere solo due alternative, in quanto molto spesso le alternative binarie non corrispondono alla vera esperienza dell'utente, ma che possa esprimere il suo livello all'interno di una scala graduata. Stesso principio è stato adottato per la selezione dei grafici, molteplici scelte sono state inserite in modo tale che l'utente possa esprimersi al meglio nella loro preferenza.

Tutte queste informazioni dovranno poi essere elaborate per poter ottenere dei risultati utili, in quanto queste a primo impatto non sono altro che una serie di numeri e stringhe concatenati tra loro.

Analisi preliminare

A valle di quanto affermato nel precedente paragrafo, viene quindi avviata l'analisi preliminare dei dati raccolti. Dopo aver fatto una breve panoramica sui campi del form, e su come questa struttura si riflette perfettamente su quelli che sono i risultati, bisogna far si che questi dati siano accessibili. Il formato prediletto per quanto concerne un dataset è quello di un foglio elettronico in formato .xlsx oppure un .csv in quanto i dati sono suddivisi grazie ad un carattere che si interpone tra un valore e l'altro, effettuando così una distinzione tra i vari campi.[31]

Il file ottenuto e generato dall'applicazione può essere scaricato in diversi formati, tra cui quelli già citati sopra, oppure vi si può accedere direttamente tramite browser e visualizzare i dati online, modificarli ed effettuare anche varie operazioni. Ovviamente si tratta di operazioni base legate alle funzionalità di un foglio di calcolo elettronico, non vengono minimamente introdotte tecniche avanzate.

Una volta scaricato il file è possibile accedere ad ogni singolo dato e la struttura che salta fuori corrisponde ad un numero di colonne pari al numero di domande e sotto domande (come si mostra nella *Figura 20)*, ad esempio per la domanda numero quattro:

"In una scala da 1 (valore minimo) a 5 (valore massimo), qual è il tuo livello di conoscenza dei seguenti termini?"

Le cui risposte sono molteplici:

"Machine Learning"
"Artificial Intelligence"
"Data Analytics"
Ecc…

Le colonne contenenti tali risposte sono ovviamente più di una, o ancora meglio una per ogni sotto risposta, ad esempio:

| "Q4-Machine Learning" | "Q4-Artificial Intelligence" | …

In tal modo è possibile quindi riuscire a generare un foglio elettronico ordinato e con ogni risposta messa in sequenza alla domanda principale (a cui possiamo riferirci come macrodomanda).

Quindi ricapitolando, il file sarà composto da una serie di attributi e ogni attributo viene rappresentato da una colonna che a sua volta comprenderà un insieme di colonne (raggruppate per macrodomanda) che corrispondono alla risposta dell'utente per quel particolare quesito.

A1 fx Q4MachineLearning

	A	B	C	D	E	F
1	Q4MachineLearning	Q4ArtificialIntelligence	Q4DataAnalytics	Q4TecnicheStatistiche	Q4DataManagement	Q4DeepLearning
2	4	4	4	3	4	4
3	1	1	1	4	1	1
4	3	3	4	5	4	1
5	4	4	4	3	4	3
6	5	5	5	5	3	5
7	3	2	2	2	2	3
8	4	4	4	3	3	4
9	5	4	5	4	5	4
10	4	4	4	2	4	4
11	3	3	3	2	3	3
12	4	3	2	2	4	3

Figura 20: Rappresentazione tabellare in formato *.csv* delle domande del form. L'immagine mostra come gli attributi (le colonne) siano stati rinominati per poter ottenere la combinazione *Domanda + Risposta*.

Una volta scaricato il file è anche possibile aprirlo, appunto in con un'applicazione per la visualizzazione e da qui è possibile anche gestire la struttura interna del file, come ad esempio il separatore tra i campi (che generalmente è impostato come una "*,*" oppure un "*;*") oppure è possibile modificare a mano i vari campi agendo direttamente sui dati.
Avendo tale possibilità di accesso ai risultati è possibile anche sistemare il file tramite spostamenti di colonne o comunque creare più copie del file per poter effettuare vari test, servendosi direttamente di quelle che sono le funzioni di base dei programmi come *Excel* o *Numbers*.

L'obiettivo da raggiungere

Raccolti tutti i dati necessari, aver scaricato il file che li contiene e dopo effettuato quella che è l'analisi preliminare si è arrivati al punto in cui inizia la vera e propria fase di ricerca.
Gli obiettivi (mostrati in *Figura 21*) che sono stati prefissati si estendono su due rami principali:

- **Analisi statistica**
- **Estrazione delle regole di associazione**

Cercando di procedere in parallelo su entrambi i fronti, questi due punti di studio e di ricerca possono fornire numerose informazioni sui dati raccolti.

L'obiettivo principale è quello di riuscire ad estrapolare informazioni che saranno poi utili per l'implementazione della sezione di **StoryTelling** in **ADESCA**. Questo corrisponde altresì al concetto che ci sono vari tipi di utenti utilizzatori e ciò corrisponde al fatto che ci siano vari livelli di esperienza da tenere in considerazione. Continuando a ragionarci su e seguendo questo filo conduttore, ciò che salta subito fuori è la possibilità di andare ad utilizzare i dati raccolti come base da porre per un futuro **Recommender System** e **sistema di profilazione utente**.

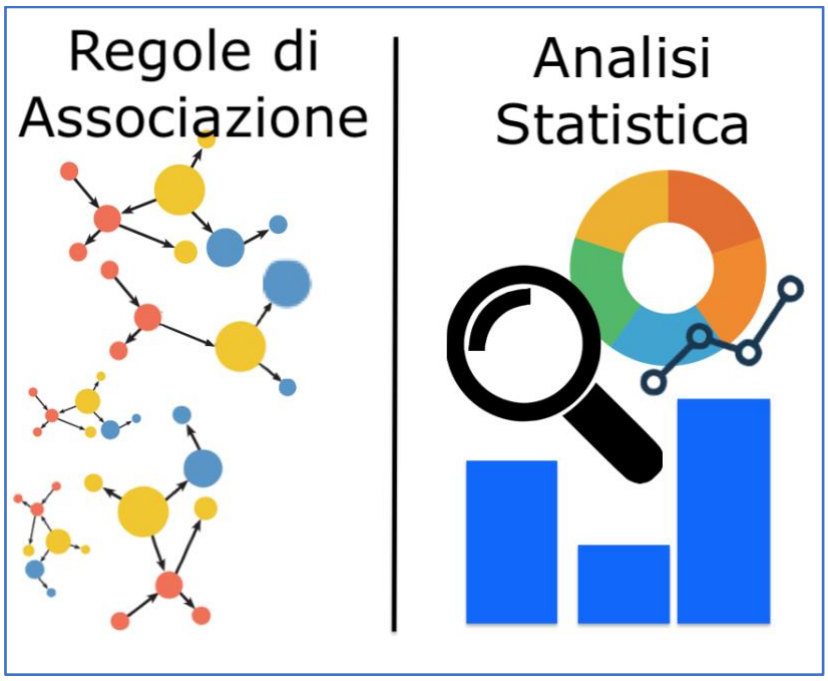


Figura 21: Immagine rappresentativa che mostra quali siano i principali obiettivi da raggiungere in seguito alla raccolta dati.

Con il concetto di profilazione degli utenti si intende la capacità e possibilità di poter definire delle categorie in cui gli utenti possano andare ad essere inseriti. In questo caso d'uso la categorizzazione degli utenti è direttamente riferita al loro livello di esperienza dell'ambito della **Data Science** e quindi ricadere in una piuttosto che in un'altra categoria corrisponde all'essere più o meno esperto.

L'implementazione dell'applicazione

Per la conduzione dell'*analisi statistica* ci si avvale dell'intero dataset, che come abbiamo visto può essere suddiviso in due sezioni principali, quella riguardante le informazioni generiche di un utente e quella relativa alla sua preparazione. Per poter condurre questo tipo di analisi sono necessari però alcuni strumenti molto più potenti di quelli che abbiamo in realtà già visto. In questo caso, il lavoro è stato proiettato verso un'ottica più professionale e quindi è stato necessario sviluppare una vera e propria **webapp** capace di fornire delle funzionalità all'utente utilizzatore e che ha la necessità di analizzare i dati.

Lo strumento utilizzato è **PyCharm** (*Figura 22*). Questo è un ambiente di sviluppo che è formato da un editor e che mette a disposizione vari tools come *python console* e *debugger*. Questo **IDE** (Integrated Develpment Environment) è nato ed è appunto utilizzato per lo sviluppi di applicazioni in *Python* e rende l'esperienza di programmazione più completa fornendo vari supporti allo sviluppatore.
Servendosi di questa applicazione per lo sviluppo, è stato possibile gestire tutto il materiale tramite i servizi offerti dall'**IDE** che consente la creazione di una directory ben organizzata e in cui sono presenti tutti file del progetto.
L'idea di base per questa app consiste nello sviluppare un'applicazione web, e quindi un applicativo che possa essere direttamente eseguito nel browser di un calcolatore collegandosi ad uno specifico indirizzo e porta, senza dover obbligare l'utente a doverla installare direttamente sul proprio pc.

Figura 22: Immagine rappresentante i loghi degli strumenti utilizzati.

Per fare ciò è stato necessario usufruire di un *micro-framework* che consentisse di accedere alle pagine web e riflettere in esse lo stato dell'applicazione. Questo prende il nome di **Flask** (*Figura 22*) e in un'accezione più generale può essere definito come un insieme di programmi necessari per la creazione di servizi web e che mette a disposizione un supporto per la costruzione di servizi *REST*.

Questo è stato direttamente installato nell'ambiente di lavoro grazie anche alla **python console** che mette a disposizione **PyCharm** e tramite una serie di comandi ci si ritrova direttamente l'ambiente di sviluppo totalmente configurato e pronto ad essere utilizzato.
Per la conduzione di questo tipo di analisi si è scelto di utilizzare un linguaggio come **Python** in quanto fornisce numerosi vantaggi da un punto di vista tecnico e risulta molto potente ed efficace se usato in ambiti come quello della *Data Analysis*.

Un ulteriore vantaggio che mette a disposizione è quello di lasciare allo sviluppatore una grande libertà nell'uso delle librerie da utilizzare, fornendo un supporto nativo o di terze parti per l'importazione e l'uso di queste. Una delle librerie che è stata appunto usata in questo progetto riguarda il concetto di visualizzazione dei dati.

Gli stessi dati possono essere rappresentati in varie forme e queste possono variare anche enormemente, questo comporta quindi che gli utenti possano comprendere o meno un tipo di rappresentazione o preferirne di altre. Una caratteristica che comunque gioca a favore dell'utente nel comprendere i dati potrebbe essere l'interazione con i grafici.

La libreria importata prende il nome di **Highcharts**. Questa è una libreria software che consente la creazione di grafici a partire da zero oppure sfruttando dei template di base, il tutto utilizzando come linguaggio un puro **JavaScript**.

Una volta trovati gli strumenti da adoperare per la creazione dell'applicazione e una volta settato l'ambiente di lavoro e la struttura del progetto (mostrata in *Figura 23*), l'ultimo passo consiste nell'importare il dataset all'interno del programma.

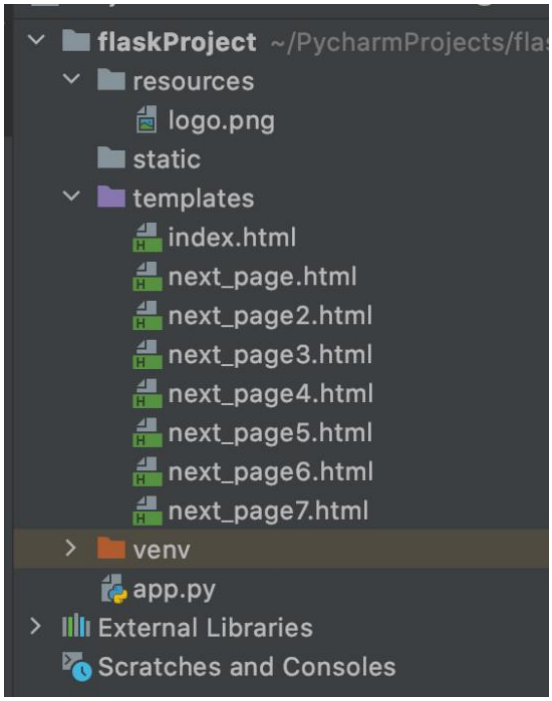


Figura 23:
Immagine rappresentante la struttura del progetto costruita in *PyCharm*. Questa mostra come vi sia la suddivisione tra la cartella *template*, che contiene tutti gli stili e le funzioni inseriti nelle pagine da visualizzare e poi distintamente il file *app.py*, che risulta essere il principale in quanto presenti al suo interno le principali funzioni per l'analisi e gestione dei dati raccolti.

Per poter gestire il documento sono state utilizzate delle funzioni base del linguaggio che consentono l'importazione di file con estensione *.csv* e una volta importato questo verrà tramutato in una struttura che prende il nome di *DataFrame*.
Quest'ultimo è una struttura dati di supporto per la gestione dei dataset importati ed è fornita da una libreria (a sua volta importata nel progetto) presente online che prende il nome di **Pandas**, che consente di manipolare i dati come tabelle numeriche o serie di valori.

Avendo già effettuato un'analisi preliminare del dataset nella prima fase, una buona parte dei dati sono stati sistemati però altre operazioni preliminari devono essere effettuate prima di poter dare in pasto alle funzioni il *DataFrame*.
Queste operazioni possono essere viste come un set obbligatorio da dover applicare e consistono nella rimozione automatica di valori *NULL* e nel rimuovere gli *spazi bianchi* presenti nelle righe del dataset. Tali funzioni possono essere combinate in modo tale da poter effettuare anche altre operazioni congiuntamente a quelle base.

Tutte queste operazioni qui descritte vengono effettuate nella pagina del progetto che prende il nome di *app.py*, e in cui sono definite tutte le funzioni dell'applicazione e che l'utente può utilizzare. Grazie a **Flask** è possibile sfruttare i *Decorators* che consentono di applicare a delle funzioni delle caratteristiche aggiuntive. Utilizzando *"@app.route()"* è possibile specificare l'endpoint su cui sarà presente tale funzione che a sua volta farà il render di un template.
Come è possibile osservare i *Decorators* possiedono una particolare notazione, ovvero quella di porre davanti il simbolo della chiocciola, inoltre questi vanno a marcare direttamente il metodo (N.b. un metodo può avere anche più di un *Decorators*).

La struttura dell'applicazione in fase di sviluppo (mostrata in Figura 24a) è stata definita in base a quelli che sono i dati raccolti, ed essendo questi stati suddivisi in informazioni generali ed informazioni specifiche la stessa suddivisione è stata adottata anche qui.
Per poter gestire le varie viste dell'applicazione sono stati creati una serie di template a cui sono stati poi aggiunti degli elementi in base al tipo di pagina e di informazioni da visualizzare.

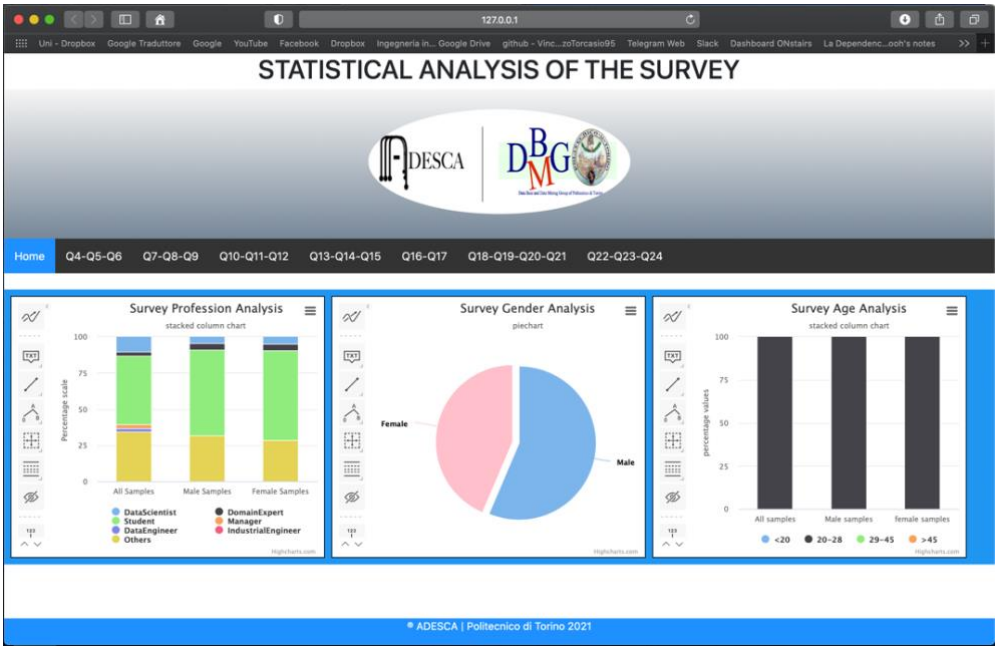


Figura 24a:
Screenshot dell'applicazione fatta girare in locale e a cui è possibile accedere direttamente tramite browser. Questa prima schermata risulta essere una pagina di riepilogo in cui sono presenti dati sugli attributi che forniscono informazioni sul numero di persone che hanno risposto al sondaggio.

Il primo template (in *Figura 24b*) corrisponde alla visione d'insieme che deve essere fornita non appena un utente si collega all'applicazione, e che quindi necessita di conoscere in prima battuta quelli che sono stati gli utenti ad aver risposto alle domande. Grafici come *Istogrammi in percentuale* e *PieChart* sono stati utilizzati per la visualizzazione di tali dati.
I dati vengono direttamente estrapolati nella funzione scritta in *app.py, che accede tramite delle funzioni di iterazione all'interno del DataFrame,* e gestiti in modo da poter essere direttamente visualizzati nel template senza alcun'ulteriore operazione.

La forza di tali librerie e strumenti sta proprio nel fatto che è possibile combinare una serie di linguaggi per poter ottenere servizi e funzionalità aggiuntive. Infatti, i dati estrapolati dalla funzione verranno poi iniettati nella vista dell'applicazione e presentati tramite il template che è stato scritto con una combinazione di linguaggi quali **HTML**, **CSS** e **JavaScript**.
Utilizzando quest'ultimo è stato possibile anche scrivere delle mini-funzioni che possano fornire informazioni aggiuntive nel grafico e che lavorano non su una serie statica di valori inseriti dal programmatore o dall'utente, ma bensì lavora direttamente sulla variabile renderizzata dall'*app.py* e che contiene i dati estratti dal *DataFrame*.

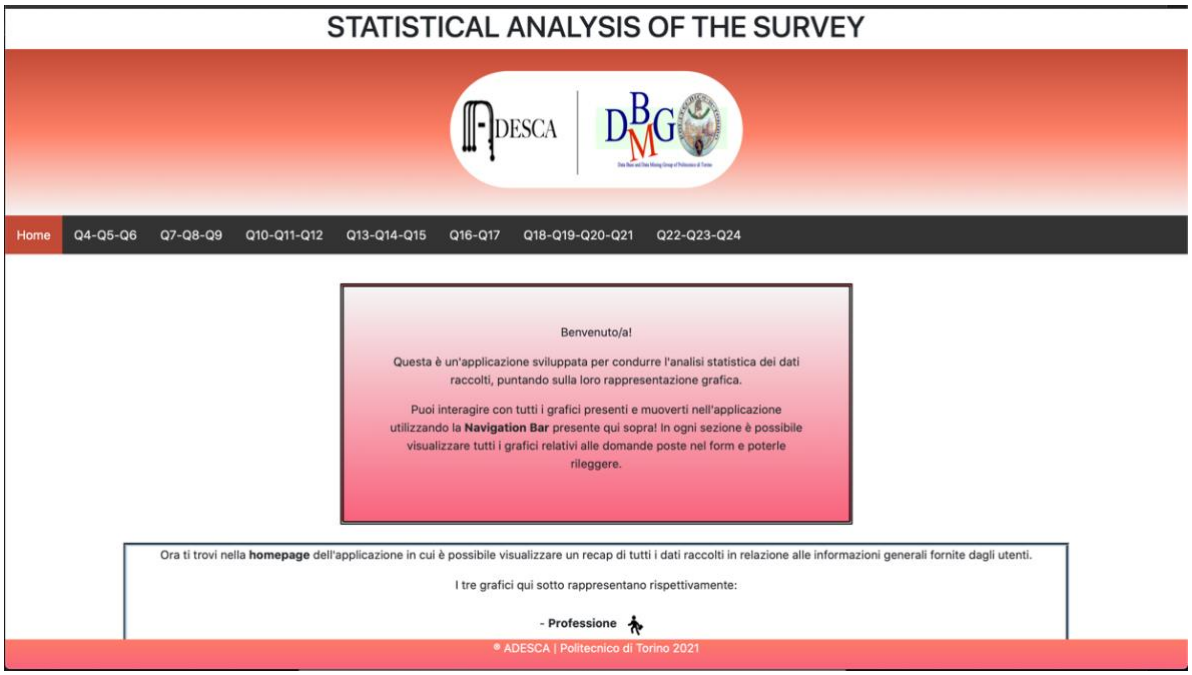


Figura 24b:
Prima parte della schermata principale dell'applicazione definitiva, nella quale viene mostrata un riquadro riassuntivo.

Le altre pagine della **webapp** sfruttano similmente la stessa struttura già descritta ma con delle modifiche che rendono migliori i dati da dover visualizzare. La struttura di base prevede che ci sia una singola funziona per ogni singolo template, il che corrisponde ad ogni singola pagina. Una suddivisione necessaria è stata quella delle varie domande che sono state accorpate in base all'ambito, ad esempio le domande relative alla visualizzazione degli attributi di un certo dataset descritto sono state aggregate in un'unica pagina. E questo corrisponde anche a far si che se tutti gli attributi di una determinata domanda necessitano di un particolare trattamento (che sia standardizzazione, sostituzione ecc…) possono essere gestiti non come casi particolari ma come casistiche generali, in quanto una suddivisione è stata già effettuata a monte dell'analisi.

Nella sezione principale sono stati poi aggiunti ulteriori strumenti che si rendono necessari nel caso l'utente decida di effettuare delle operazioni sui grafici mostrati (*Figura 24c*), quali disegnare delle linee rette o inserire annotazioni. **Highcharts** comunque risulta essere una delle librerie open-source migliori in circolazione in quanto consente anche una vera e propria interazione con l'utente. Si passa da animazioni legate alla generazione del grafico (e quindi ogni volta che si ricarica la pagina) a quelle interazioni che cambiano i dati da visualizzare o disabilitano un certo valore.

Ogni singola pagina è stata poi riempita con un sommario in cui è possibile rileggere le domande che sono state proposte nel form ed una *navigation bar* è stata aggiunta per far sì che l'utente possa spostarsi all'interno dell'applicazione come e quando vuole.

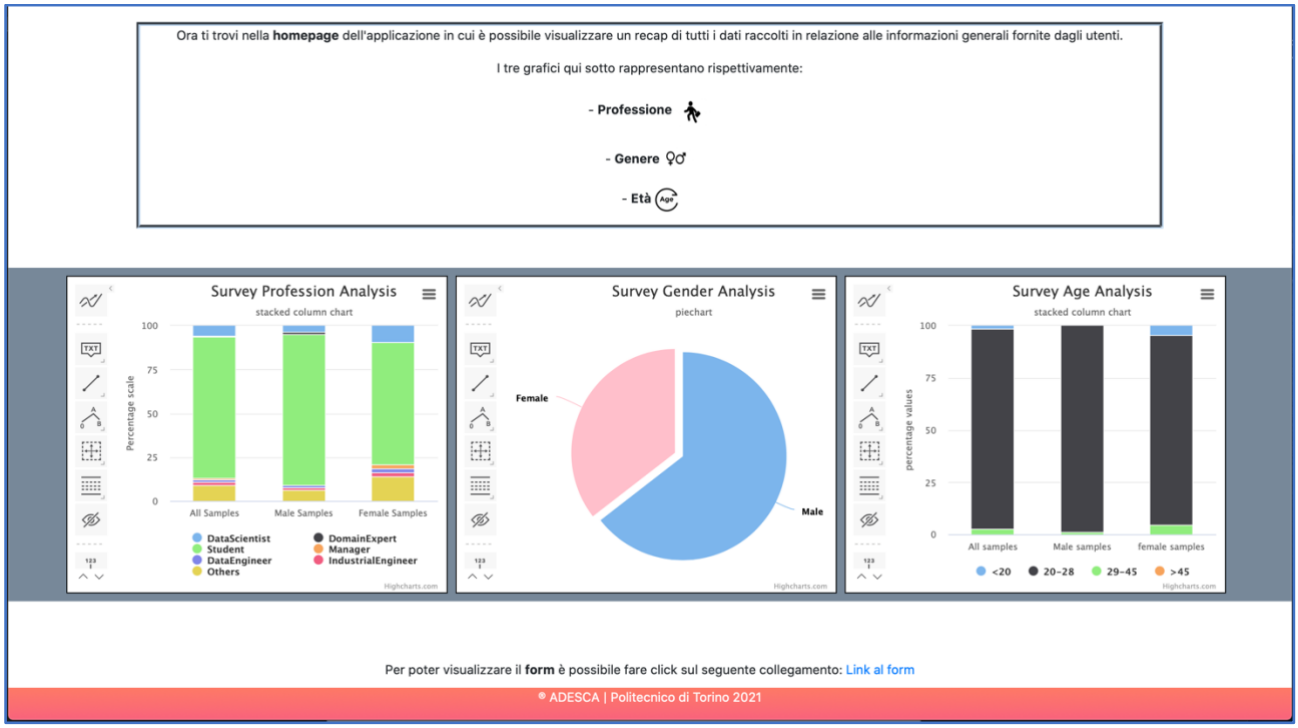


Figura 24c:
Seconda parte della pagina principale dell'applicazione, nella quale vengono mostrati i grafici.

Implementazione delle Association Rules

Un altro obiettivo da raggiungere è contrassegnato da un'analisi diretta e più approfondita delle risposte fornite dal singolo utente, che messe a confronto con tutte le altre forniscono delle informazioni molto importanti. Ciò che è stato appena descritto con parole molto spicciole va ad introdurre il concetto di **Regole di Associazione**.

Nell'ambito della **Data Science**, o più nello specifico in quello del *Data Mining*, le **Association Rules** descrivono uno dei modi per poter estrarre, a partire da un dataset, delle relazioni che risultano nascoste all'utente.
L'esempio generico che spiega in maniera molto breve ed esaustiva questo concetto è quello del "*supermercato*", nel quale si mette alla luce il fatto che è possibile analizzare le transazioni fatte dagli utenti per poter capire quali siano gli oggetti che un determinato cliente ha acquistato insieme.

Questo può fornire numerose informazioni in quanto mette in correlazione quelli che sono i vari oggetti e che in un'analisi iniziale potrebbero sembrare non correlati.
Quindi, possiamo affermare che le **Regole di Associazione** forniscono un grande supporto per la scoperta di queste relazioni che vengono estratte direttamente da immensi dataset. Infine, l'obiettivo principale non è quello di andare a trovare delle *preferenze* di un utente (considerandone il singolo) ma bensì andare a trovare delle relazioni tra item di ogni *tupla*.

Avendone introdotto il concetto principale è ora necessario fare un breve recap di come fondamentalmente funzionano le **Association Rules** e come vengono applicate nella pratica.

Vengono introdotti due parametri fondamentali:

- **Support**
- **Confidence**

Ognuno di questi è considerato come un indice che fornirà un singolo valore calcolato a partire dai dati del dataset.

Il **Support** (*Formula 1*) va ad indicare la frequenza con cui determinati item si trovano all'interno di una tupla, ed è definito matematicamente come:

$$\text{Support} = \frac{\text{Freq (A)}}{\text{N}}$$

Formula 1: Support.

La **Confidence** (*Formula 2*) è invece utilizzata per spiegare come un determinato item si trovi in relazione con un altro item:

$$\text{Conf (A} \longrightarrow \text{B)} = \frac{\text{Support (A U B)}}{\text{Support (A)}}$$

Formula 2: Confidence.

Dopo aver introdotto i primi due indici è possibile parlare del terzo parametro che sfrutta, per poter essere calcolato, la combinazione dei primi due. Questo parametro prende il nome di **Lift** (*Formula 3*) ed è definito da un punto di vista matematico come:

$$\text{Lift} = \frac{\text{Support}\ (\ A\ \ U\ \ B)}{\text{Support}\ (A)\ x\ \text{Support}\ (B)}$$

Formula 3: Lift.

Questo identifica quanto un item è in relazione con un secondo item, e quanto il primo risulta essere frequente all'interno del dataset.

Una volta definiti i vari parametri su cui è possibile basarsi per ottenere dei buoni risultati, è necessario andare ad introdurre il concetto di **Algoritmo Apriori**, che si pone l'obiettivo di trovare i set di elementi più frequenti all'interno di una base dati. Tale algoritmo tenta un approccio di tipo iterativo per poter trovare gli itemset più frequenti (un itemset è definito come un'insieme di item che vengono per l'appunto associati) e che richiede quindi una scansione intera del dataset.

Questo tipo di algoritmo risulta essere uno strumento molto potente e che viene utilizzato in vari settori come ad esempio nel campo medio, universitario ma anche nelle grandi industrie ed aziende. Ovviamente l'utilizzo di questo tipo di algoritmo possiede dei *pro* e dei *contro*, e di sicuro il principale aspetto negativo è quello legato all'aspetto computazionale, in quanto risulta essere veramente costoso da un punto di vista delle risorse utilizzate. Proprio per questo motivo è necessario riuscire a settare con criterio i vari i parametri che vengono utilizzati per l'esecuzione e che sono stati precedentemente citati.

Valutati tutti gli aspetti che un algoritmo del genere pone davanti è stato necessario trovare anche degli strumenti capaci di poter gestire una gran mole di lavoro (dal punto di vista del calcolo) e di poter fornire un risultato accettabile.
Lo strumento principalmente utilizzato è **Google Colab Online**, uno strumento messo a disposizione appunto da una delle aziende leader per quando riguarda l'informatica, e che consiste nel far collegare l'utente (a cui è associato un account Google) ad un notebook online in cui è possibile andare ad eseguire del codice **Python** sfruttando gli hardware messi a disposizione dall'azienda. Altri vantaggi messi a disposizione dell'utente sono quelli legati alla condivisione tra utenti molto semplificata e la possibilità di iniziare a lavorare sui propri progetti direttamente senza effettuare alcuna configurazione.

Su questo tipo di piattaforma è possibile sfruttare tutti i vantaggi di un ambiente di lavoro **Python** e quindi importare semplicemente le librerie esterne o installare pacchetti di terze parti. Il tutto effettuando un semplice collegamento ai server che fungono da host per l'hardware utilizzato (*GPU* e *RAM*).

Questo strumento fornisce anche la possibilità di salvare varie copie del codice utilizzato nel proprio spazio personale su Google Drive oppure in alternativa la possibilità di poterlo scaricare direttamente sul proprio computer in un formato *.ipynb* o *.py*.

Per prima cosa sono stati importati quelli che sono i moduli principali per la gestione del dataset e per l'implementazione dell'algoritmo che ci consentirà di estrarre le regole di associazione.
L'operazione immediatamente successiva prevede quella che è l'importazione del dataset tramite una serie di funzioni capaci di far selezionare il file direttamente dal *file system* e che verrà caricato nello spazio di lavoro all'interno dell'applicazione. Ad operazione completata ci saranno delle operazioni che consentono di ricostruire i dati caricati e gestirli grazie alla libreria **Pandas**.

Prendendo in considerazione il dataset ottenuto dai dati raccolti, è possibile notare il grande numero di colonne che formato tale base di dati e che quindi porta ad una complessità computazionale non indifferente e potrebbe portare al superamento di quelli che sono i limiti forniti dallo strumento (che risultano essere già sopra la media).
La prima constatazione ad essere stata effettuata si riferisce all'obiettivo principale che ci si era posto, ovvero quello di estrapolare informazioni utili per poter implementare una sezione in cui gli utenti vengono divisi in base alla loro esperienza nel settore.

Per poter far ciò è stato necessario servirsi di quelle che sono state le impostazioni delle domande nel form, e cioè servirsi di quelle che sono le scale graduate delle risposte fornite dagli utenti.
Il primo passo è quindi stato quello di suddividere gli utenti con questa operazione preliminare in tre categorie principali (come mostrato in *Figura 25*):

- **Rookie**
- **Middle**
- **Advanced**

Suddividendo il range di valori in queste tre categorie è stato possibile snellire il costo computazionale rendendo possibile l'ottenimento dei risultati.

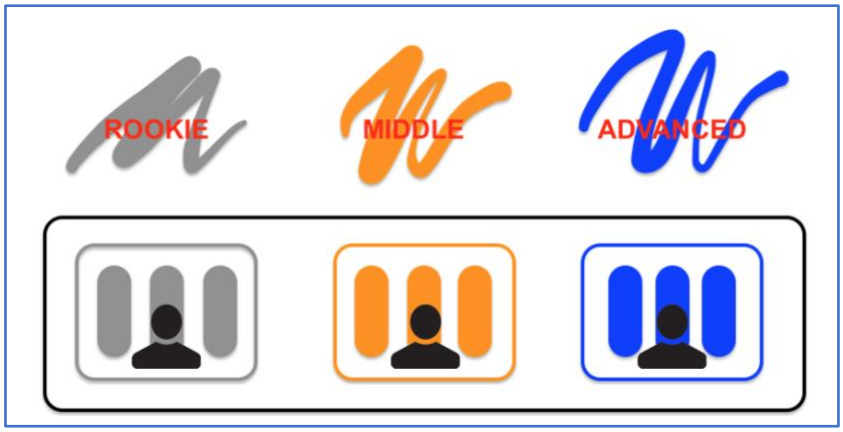


Figura 25:
Immagine rappresentativa e schematica di quella che è la suddivisione degli utenti in tre categorie in base a quelle sono le loro conoscenze di base e che consentono un approccio diverso al programma.

Una seconda operazione da effettuare fa riferimento a quelle che sono le colonne del dataset relative alla scelta dei grafici, in quanto gli utenti sono stati abilitati a poter effettuare anche più di una scelta e questo comporta che all'interno di una certa *row* è possibile avere anche più di un parametro che viene separato da un carattere speciale che in questo caso risulta essere "\n".
In questo passaggio è quindi necessario andare a gestire questo aspetto altrimenti si potrebbe andare a rendere dubbia la correttezza dell'analisi in quando le singole risposte non verrebbero conteggiate e si farebbe riferimento solo alle possibili combinazioni.
Ad essere implementato (*Figura 26*) è stato quindi un codice capace di scandire tutte le colonne in cui sono presenti questi parametri e con un costrutto iterativo è stato possibile generare altre colonne a partire da quella originale. Questa operazione corrisponde ad una vera e propria modifica del dataset e si conclude quindi con la cancellazione delle colonne originali per evitare che l'informazione possa essere non corretta.

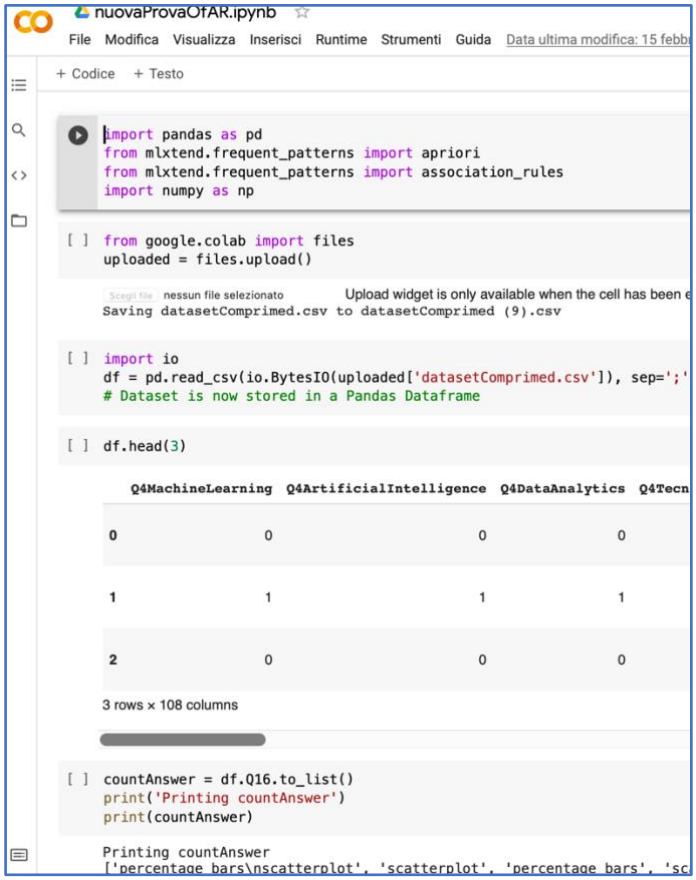


Figura 26:
Questo screenshot prelevato dalla console di *Google Colab Online*, ci consente di poter visualizzare la struttura a notebook dell'applicazione e mostra la prima parte del codice che è stato utilizzato per l'estrazione delle *Association Rules*.

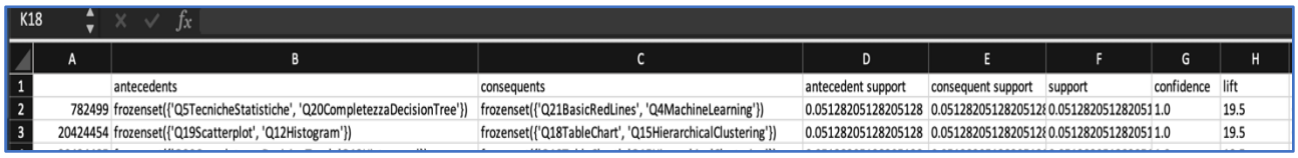
Una volta effettuate le operazioni sopra descritte sono state definite le righe di codice che implementano il concetto di **Association Rules** e che per poter essere utilizzati necessitano di quella che l'operazione definita come tuning dei parametri, nella quale si testano una serie di valori associati ai vari parametri per poter ottimizzare il processo (già descritto ormai nei paragrafi precedenti).

Questo algoritmo accoglie come parametro per ottenere i dati su cui lavorare quella che una struttura che abbiamo già visto, ovvero quella di un *DataFrame*.
Una volta terminato il processo questo non fa altro che ritornare i risultati sempre sotto la stessa forma, andando così a fornire all'utente quelli che sono dei metodi da poter chiamare direttamente

su queste strutture dati, come ad esempio le funzioni di ordinamento secondo un determinato criterio (in questo caso effettuato tramite il parametro **LIFT**) oppure la visualizzazione diretta nel *notebook* di quelle che sono le *n* righe del *DataFrame*.

Una volta completata l'analisi i risultati devono essere presentabili e leggibili anche da quelli che sono utenti esterni e che potrebbero servirsi anche di queste informazioni, ragion per cui i sono stati associati agli item i nomi delle colonne (che rappresentano appunto gli attributi) in modo che itemset possano essere direttamente riconosciuti senza dover effettuare alcun altro tipo di trasformazione sui dati.

Ma tali strutture sono nuovamente esportate dal metodo sotto forma di *DataFrame* ed è proprio per questo motivo che risulta necessario aggiungere ulteriori righe di codice per poter ottenere un file scaricabile sotto forma di documento. Per far ciò è stato utile sfruttare le caratteristiche dello strumento che consente appunto di collegarsi al proprio spazio virtuale online e di poter esportare grazie alle caratteristiche del linguaggio base (riferimento a **Python**) i dati sotto forma di file *.csv*.
Dopo aver condotto l'intera analisi si potrà accedere ai risultati direttamente tramite i file scaricati (e.g. Come evidenziato in *Figura 27*) che saranno appunto suddivisi in tre parti, ognuno relativo ad una categoria di utente: *Rookie*, *Middle* e *Advanced*.



	A	B	C	D	E	F	G	H
1		antecedents	consequents	antecedent support	consequent support	support	confidence	lift
2	782499	frozenset({'Q5TecnicheStatistiche', 'Q20CompletezzaDecisionTree'})	frozenset({'Q21BasicRedLines', 'Q4MachineLearning'})	0.05128205128205128	0.0512820512820512(	0.0512820512820513	1.0	19.5
3	20424454	frozenset({'Q19Scatterplot', 'Q12Histogram'})	frozenset({'Q18TableChart', 'Q15HierarchicalClustering'})	0.05128205128205128	0.0512820512820512(	0.0512820512820513	1.0	19.5

Figura 27:
Immagine rappresentante il file *.csv* ottenuto una volta terminato il run dello script per l'estrazione delle regole di associazione. Questo file viene direttamente scaricato ed è visualizzabile grazie a strumenti come *Microsft Excel*. È possibile da qui notare gli *itemset* coinvolti e affianco quelli che sono i punteggi ottenuti e calcolati dai vari indici.

Classi utente e preferenze

L'estrazione delle **Association Rules** e l'implementazione dell'applicazione per l'**Analisi Statistica** mettono alla luce una serie di risultati, i quali possono essere sfruttati per studi futuri.

Nell'immediato è invece possibile capire dall'applicazione sviluppata quali siano state le principali risposte fornite dagli utenti, che come è stato già appurato possiedono expertise diverse.
I grafici generati a partire da questi dati mettono in evidenza le risposte in percentuale, in modo tale che si possa capire quali siano state le più gettonate. Il tutto è stato implementato senza andare però a considerare nessun tipo di suddivisione utenti in classi.

La suddivisione delle classi porta numerosi vantaggi in quanto fornisce agli sviluppatori la possibilità di costruire *ad hoc* le sezioni del programma, in modo tale che ogni tipologia di utenza possa ritenersi soddisfatta.

Ciò è stato implementato principalmente nell'estrazione delle **Regole di Associazione**, dalle quali è possibile capire le preferenze degli utenti, che essi siano *Rookie*, *Middle* o *Advanced*.
L'estrazione di tali informazioni è stata fatta con lo scopo di avere una maggiore precisione nelle scelte effettuate dagli utenti, tenendo conto quindi anche di quelli che sono i vari grafici proposti e delle combinazioni possibili con le domande della prima sezione.

Ricapitolando, i file ottenuti come risultato da questo processo forniscono informazioni necessarie per quanto riguarda le preferenze utente e quindi quali siano le risposte principalmente correlate tra loro. Seguendo tali combinazioni sarà possibile implementare e porre basi per un programma con delle categorie adatte agli utenti, in quanto ad essere implementate saranno delle funzionalità coerenti con quanto raccolto.

Un ulteriore analisi che è stata condotta per poter ulteriormente andare ad ottenere informazioni circa la costruzione di queste categorie, è quella relativa alle preferenze dei soli grafici.
A partire dai dati raccolti è stato possibile utilizzare la scala graduata della prima sezione del form per poter ottenere le tre categorie utenti sopracitate.
Il primo step prevede che la suddivisione in classi venga effettuata grazie al calcolo della media relativa alle risposte fornite dagli utenti.
La suddivisione è stata effettuata utilizzando tali metri di giudizio:

- $0 < \textit{Mean Value} < 1.5 \rightarrow$ ROOKIE
- $1.5 < \textit{Mean Value} < 3.5 \rightarrow$ MIDDLE
- $3.5 < \textit{Mean Value} < 5 \rightarrow$ ADVANCED

Questa particolare suddivisione consente di ottenere una maggioranza all'interno della classe *Middle*, la quale sarà a tutti gli effetti quella più corposa anche a livello di contenuti. E soprattutto consente di discriminare al meglio quelli che sono gli utenti *Advanced* e *Rookie*, che necessitano di una maggiore attenzione considerata la loro posizione specifica.

A partire da questi valori è stato poi possibile mappare le tre categorie in modo tale che il dataset possa essere appunto suddiviso in tre parti, o meglio ancora tre file.

Per l'implementazione effettiva di questa parte si è fatto riferimento al progetto già presente e che riguarda appunto **l'Analisi Statistica**. Essendo questa sviluppata in **Python** e attraverso l'uso di *PyCharm* come *IDE*, questa nuova analisi è stata direttamente inglobata al suo interno andandone a sviluppare una vera e propria sezione dedicata.

È stata appunto creata una pagina a cui è possibile accedere dall'applicazione esistente tramite un apposito tasto all'interno del menu (*Figura 28*). Questa si differenzia dalla precedente che risulta essere molto più generica, ma sfruttando lo stesso principio di funzionamento, è stato possibile integrare direttamente il tutto all'interno di essa, evitando quindi di creare un nuovo progetto *stand alone*.

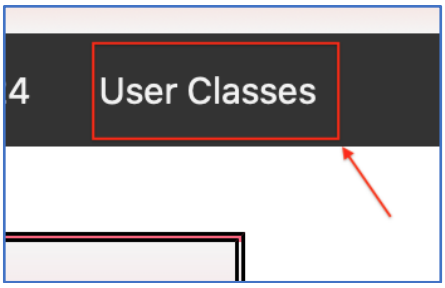


Figura 28:
Aggiunta della sezione relativa alla categorizzazione degli utenti all'interno del menu principale dell'applicazione.

Tale pagina è stata pensata per fornire le principali informazioni ricavate appunto dal nuovo processo messo in atto, e anche per poter effettuare dei confronti con quelli già conclusi.
Le principali informazioni messe in evidenza solo il numero di utenti che hanno partecipato al sondaggio e quindi che sono stati utilizzati per condurre l'esperimento (mostrato in *Figura 29*).

Ottenute le tre categorie, è stato poi possibile conteggiare il numero di utenti che sono ricaduti in una piuttosto che in un'altra e proprio da questa pagina è possibile accedere alle varie categorie cliccando sulla sezione corrispondente, messa in evidenza.

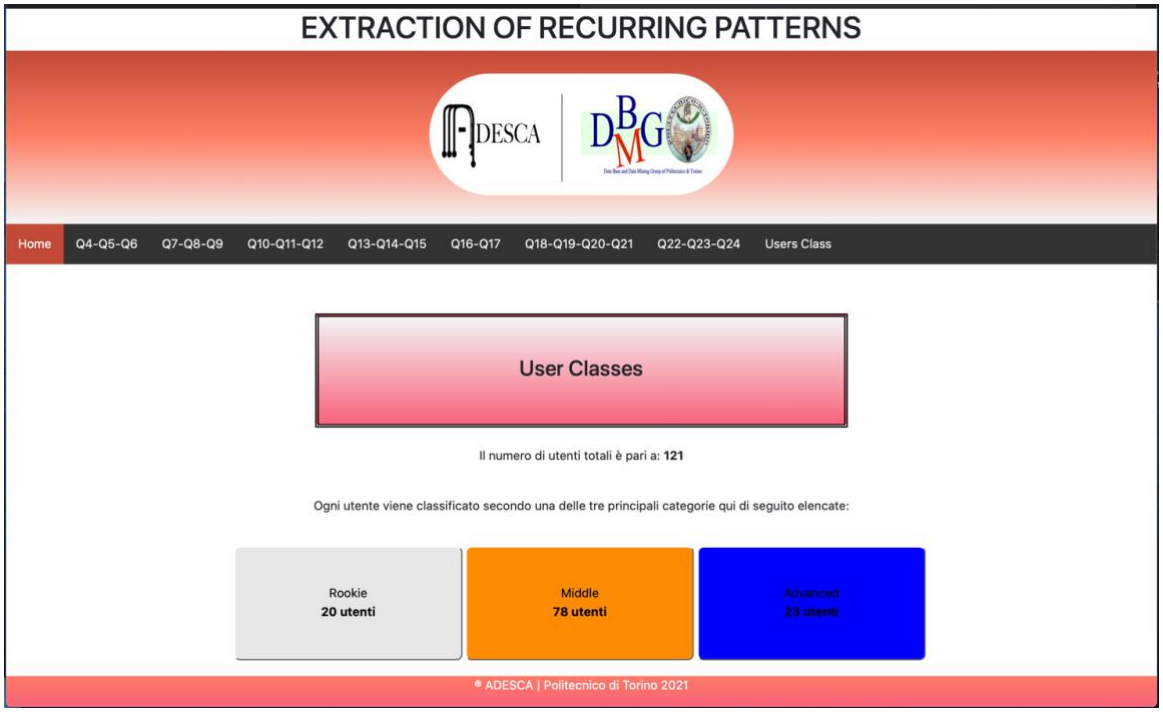
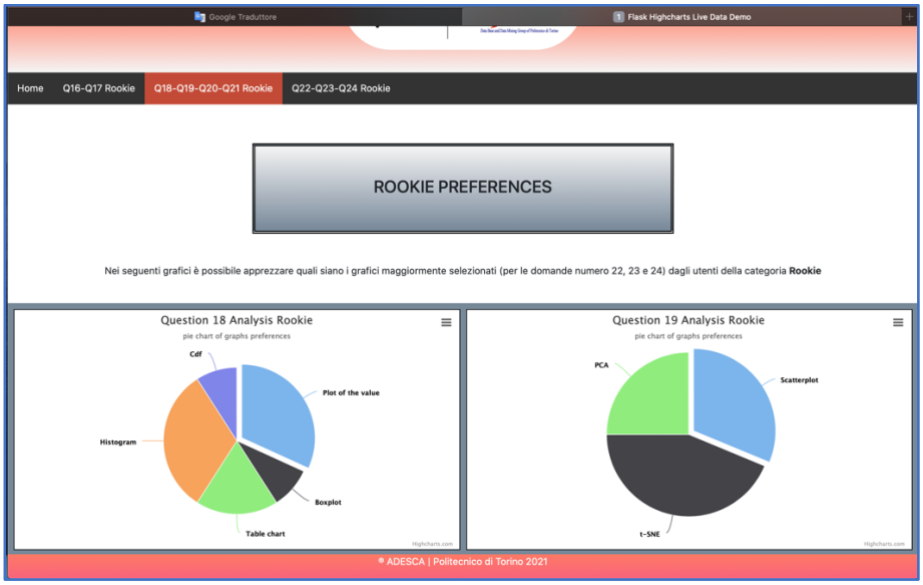


Figura 29:
Screenshot dell'applicazione che mostra come è stata implementata la pagina relativa alla categorizzazione degli utenti. Sono presenti le tre principali categorie, a cui è possibile accedere, con informazioni riguardo il numero di utenti.

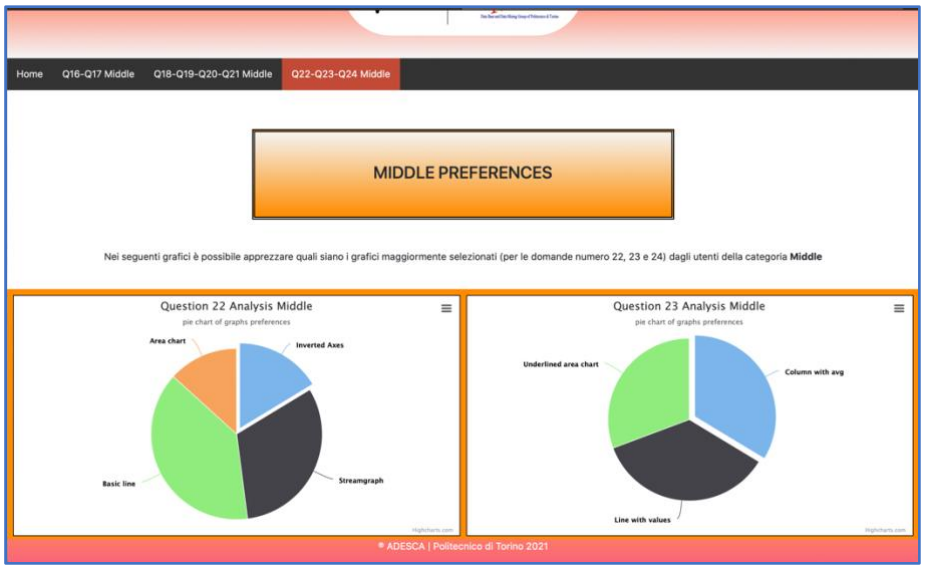
Accedendo ad una categoria piuttosto che ad un'altra (*Figura 30*), è possibile valutare quelle che sono le risposte fornite dagli utenti nel form di raccolta dati. Tali risposte prevedono quali siano i grafici maggiormente preferiti dagli utenti sulle varie domande poste all'interno del sondaggio.

Ogni grafico possiede le proprie caratteristiche e soprattutto è relativo ad una specifica domanda, proprio per questo motivo i grafici sono stati costruiti (proprio come nel caso precedente) considerando la suddivisione per domanda.

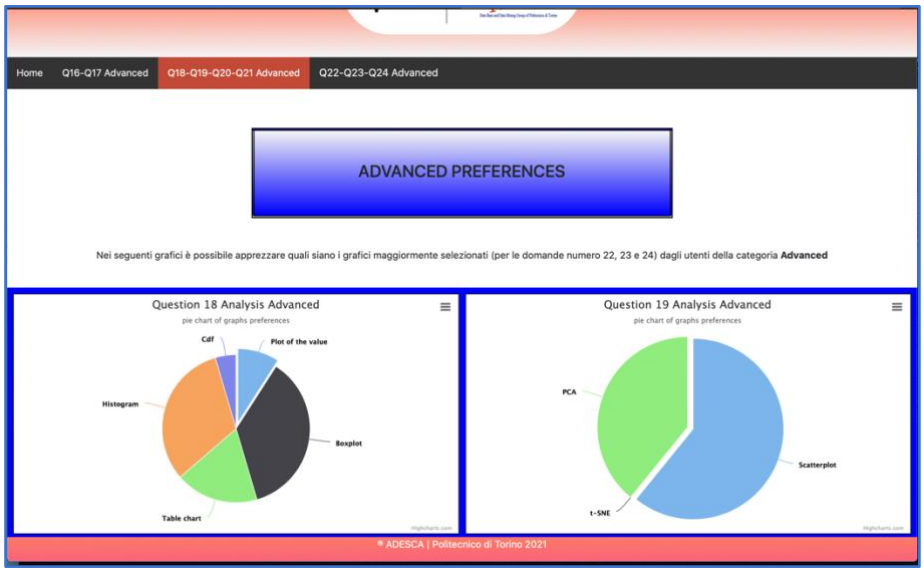
All'interno di ogni sezione è poi presente una *navigation bar* che fornisce la possibilità di navigare all'interno dell'applicazione e visualizzare tutti i grafici relativi alle varie domande.



a) Pagina di preferenza categoria Rookie.



b) Pagina di preferenza categoria Middle.



c) Pagina di preferenza categoria

Figura 30: Immagine composta da tre schermate dell'applicazione (a,b,c), ognuna relative alle varie sezioni Rookie, Middle e Advanced. È possibile notale i grafici anche in questo caso interattivi e la navigation bar tramite cui è possibile spostarsi nell'applicazione.

Questo tipo analisi si differenzia da quelle precedenti in quanto mette solamente a confronto, considerando le tre principali classi, quali siano le preferenze da un punto di vista prettamente grafico di rappresentazione dei dati. Quindi fornisce un ulteriore supporto appunto a quelle che sono le informazioni già ricavate dai processi precedenti.

Andando a basare la costruzione delle nuove sezioni su questi dati e su quelli già precedentemente raccolti, sarà possibile andare ad implementare con coerenza ed efficacia un sistema capace di fornire all'utente ciò di cui necessita.

Feedback su ADESCA

L'operazione preliminare di raccolta dati tramite form mira a migliorare quella che è l'applicazione già esistente e di per sé consente di raggiungere degli ottimi risultati, ma per poter completare il lavoro si ritiene necessario anche un secondo sondaggio, che possa contribuire a migliorare il lavoro già esistente.

Non è infatti una novità potersi scontrare con una pagina di riepilogo per fornire *feedback* riguardo una certa applicazione. Programmi che possiedono tale struttura se ne trovano a bizzeffe soprattutto online, una volta terminato il periodo di utilizzo è possibile caricare una pagina di cortesia in cui si chiede all'utente stesso di compilare gentilmente un breve form sull'esperienza appena avuta.

In genere le varie aziende avanzano con due approcci differenti nel sottoporre i questionari agli utenti, in relazione al fatto che i tipi di prodotti possono essere diversi.
Ad esempio, all'interno di applicazioni (soprattutto per dispositivi mobili) si chiede periodicamente, se ciò non è stato già fatto, di fornire una valutazione all'applicazione tramite un *pop-up* riassuntivo e in cui si esprime il giudizio in scala graduata (si pensi alle "*stelle*" assegnate ad una certa applicazione) oppure utilizzare la tecnica già citata che consiste nel ricondurre l'utente in una pagina di cortesia.

L'idea di base da implementare in **ADESCA** consiste del verificare tutti i passaggi che l'utente compie all'interno dell'applicazione e tenere traccia delle aree che esso visualizza.

Infatti, al momento l'applicazione tiene conto di tre sezioni principali che si dividono in:

- **DATA**
- **USER EXPLORATION**
- **STORYTELLING**

In queste sezioni l'utente può navigare liberamente senza alcuna distinzione e può usufruire di tutti gli strumenti messi a disposizione per la generazione di risultati o di grafici che possano mostrare determinate informazioni piuttosto che altre.
Da notare anche che la sezione *USER EXPLORATION* è al momento composta da ben due sottosezioni e che sono **INFO** e **CLUSTERING.**

Una volta che l'utente ha terminato l'analisi da condurre sarà necessario andare a visitare una pagina di cortesia in cui saranno presenti dei link esterni collegati ad un form in cui sarà possibile rispondere alle domande basate sul tipo di esperienza che l'utente ha avuto nell'utilizzare l'applicazione.

Il form in questo caso sarà di tipo interattivo in quanto è possibile che l'utente non abbia visitato tutte le sezioni e quindi potrebbe non sapersi esprimere riguardo le domande di una certa parte.

In questo caso la pagina web di cortesia che viene generata sarà direttamente costruita a partite dai pattern precedenti in modo tale che si possa mantenere la stessa scelta stilistica fino alla fine.
I linguaggi utilizzati sono principalmente **HTML** e **CSS** per quanto riguarda la costruzione di essa mentre continuerà ad essere utilizzato **Python** in accoppiata a **Flask** per poter gestire la pagina, il suo rendering e la possibilità di navigarci tramite browser.

Implementazione del modulo per la raccolta dei feedback

Avendo già acquisito esperienza nella creazione di un form per la raccolta dati, è stato possibile riuscire a schematizzare l'intero lavoro da svolgere nella maniera più semplice e spedita possibile.

Le fasi da percorrere comunque risultano pressoché identiche a quelle già affrontate in precedenza nell'implementazione dell'altro form.

Il primo passo da svolgere è quello di andare a scrivere una bozza per le domande che dovranno comporre il modulo, e in questo caso bisognerà costruirle in base alla struttura dell'applicazione considerando sia il fatto che l'utente possa andare ad utilizzare l'applicazione in maniera più approfondita (e quindi si necessita di domande un po' più specifiche) o che l'utente vada ad utilizzarla semplicemente usufruendo delle principali caratteristiche e che quindi non vada a spulciare nei meandri del programma.

Considerando lo stato attuale dell'applicazione sono presenti quattro sezioni:

- **DATA**
- **USER EXPLORATION: INFO**
- **USER EXPLORATION: CLUSTERING**
- **STORYTELLING**

Più una quinta in fase di completamento e che fornisce il supporto ad un tipo di analisi innovativa in quanto tende ad analizzare un dataset caricato dall'utente e che fa riferimento alle *Time-Series*.

Quindi il form sarà suddiviso in cinque sezioni principali, ovvero quelle appena citate, e per ogni singola parte verranno implementate una serie di *tre domande* che mirano a far esprimere all'utente il suo giudizio su quella singola sezione visitata. La struttura delle domande è stata così definita:
le prime due di ogni sezione come domande specifiche sulla soddisfazione dell'utente per quella sezione mentre l'ultima lascia all'utente stesso la libertà di esprimersi tramite un breve commento su ciò che si sarebbe aspettato di vedere implementato in quella sezione, oppure esprimere il suo giudizio su ciò che vorrebbe fosse implementato.

Questo fornisce all'utente un mezzo più potente per potersi esprimere, rispetto alla scelta multipla tra alternative già proposte, e consente agli sviluppatori di tenere conto di tutti gli aspetti citati (positivi e negativi) dall'utente.

Considerando questa come struttura base del form l'utente non sarà obbligato a rispondere a tutte le domande, in quanto potrebbe non aver visitato una certa sezione, e soprattutto non gli sarà chiesto di dedicare troppo tempo nel lasciare il suo feedback.

Un'ultima sezione che però è stata aggiunta riguarda principalmente la valutazione generale dell'applicazione, che si discosta da quella del valutare singolarmente le sezioni. Sono state generate un numero identico di domande da poter sottoporre agli utenti, i quali vi potranno rispondere prima di poter inviare il modulo. Questo tipo di domande sono generiche e chiedono all'utente se la sua esperienza con l'applicazione è stata positiva e se questi la consiglierebbe o tornerebbe ad utilizzarla. Insomma, questo tipo di domanda non serve per poter capire come migliorare la singola sezione, ma bensì per capire il livello di soddisfazione degli utenti utilizzatori.

Una volta scritta la bozza delle domande, queste sono state poste più volte ad analisi critica per poter ottimizzare il processo di raccolta dati. Come ben sappiamo, bisogna molto curare la forma espressiva per poter ottenere il massimo delle informazioni dall'utente e quindi una domanda deve poter essere rivista più e più volte.
Arrivati alla forma definitiva, il prossimo passo consiste nell'implementazione del form.

In questo caso il contesto risulta essere molto diverso rispetto a quello del precedente in quanto si tratta di lasciare dei feedback e di per sè un modulo di questo tipo si basa esclusivamente sul rapporto diretto domanda-risposta, senza andare a gestire strutture dati un po' più complesse come nel caso del primo form in cui vi erano dei grafici da dover mostrare.
In tal caso, superato questo limite, è stato utilizzato come strumento per l'implementazione del modulo il già citato *Google Form*.

Questo, infatti, consente una gestione migliore delle risposte in quanto vengono automaticamente salvate nella cartella online del proprietario del form. Questo proprietario può inoltre condividere il modulo in fase di sviluppo e soprattutto può condividere i risultati direttamente utilizzando G*oogle Sheets*.

È possibile affermare quindi che la scelta vincolante in questo caso è ricaduta sul fatto di dover andare a scegliere lo strumento che fornisse il supporto migliore per la condivisione e gestione dei risultati, non avendo appunto la problematica di gestire domande con particolari strutture.
La prima fase nello sviluppo del form fa riferimento alla struttura base, quindi la scelta del layout e la suddivisione delle sezioni. Ognuna di esse fa riferimento diretto alle sezioni presenti all'interno di **ADESCA** e prendono appunto lo stesso nome, in modo che l'utente possa avere familiarità.

La schermata principale (come mostrato in *Figura 31*) fornisce all'utente una breve introduzione al modulo e fornisce dei messaggi di ringraziamento per la collaborazione. Proseguendo nel modulo ci sarà una seconda schermata che consente all'utente di scegliere la sezione che vuole valutare, in tal modo quest'ultimo verrà automaticamente reindirizzato alla sezione scelta.

Figura 31:
Screenshot dell'applicazione aperta direttamente nel browser utente. Questa mostra la prima schermata davanti a cui l'utente si ritroverà andando a cliccare sul link.

All’interno di ognuna di esse, come ultima opzione l’utente sarà sempre indirizzato a scegliere una sezione da valutare (*Figura 32*), tale scelta dipende appunto solo da lui, vi è infatti anche la possibilità di saltare direttamente alla pagina di invio del modulo.

Per gestire l’intero form, è stata aggiunta anche una *barra di progressione* che mostra all’utente il livello di compilazione del form. Alcune delle domande sono state tra l’altro rese obbligatorie in modo tale che si possano gestire i casi in cui gli utenti vogliano danneggiare o invalidare la raccolta dati.

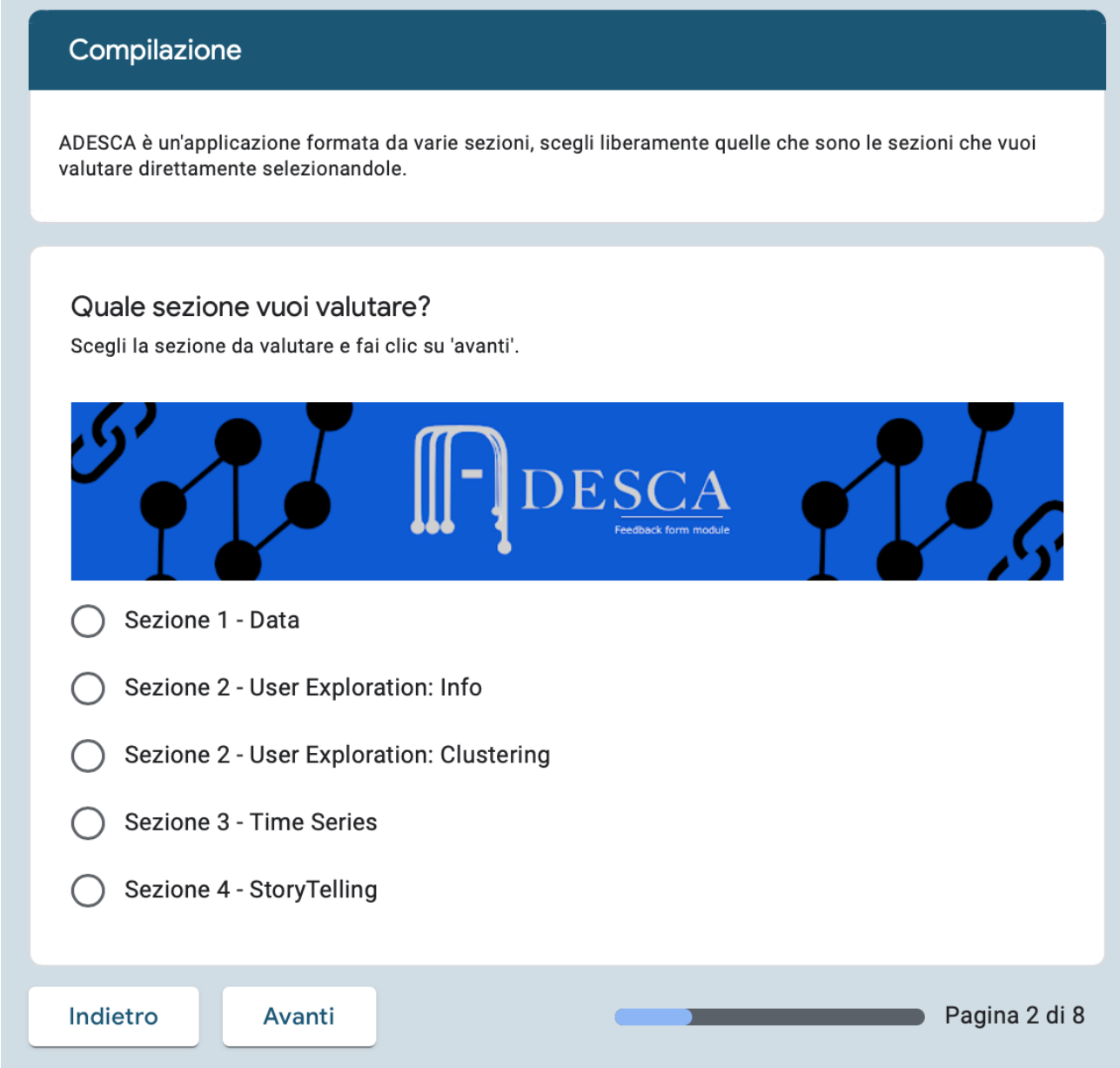


Figura 32:
Questa immagine mostra la schermata che fornisce all’utente la possibilità di navigare all’interno del form e compilare le varie sezioni.

Una volta implementate tutte le sezioni e le varie domande si passa allo step successivo, che consiste nell'andare ad effettuare un test dell'applicazione per valutarne il funzionamento.

In tal caso questo non prevede come nel caso precedente l'invio del link di condivisione ad una moltitudine di utenti, ma è semplicemente bastato far compilare il form ad un sottogruppo abbastanza ristretto, in quanto la principale criticità potrebbe essere legata alla logica condizionale che vi è dietro e che consente all'utente di navigare nel modulo oppure all'archiviazione dei risultati nel foglio elettronico.
Tali risultati saranno poi messi a disposizione degli sviluppatori per una visualizzazione online oppure per lo scaricamento su computer, scegliendo tra vari formati. Da questa raccolta feedback sarà poi possibile estrapolare informazioni per poter migliorare l'intera applicazione.

Implementazione della sezione di valutazione

ADESCA è formato, come già citato molte volte, da una serie di sezioni. Queste sezioni implementano varie funzioni legate all'analisi dei dati e alla loro presentazione per l'ottenimento di risultati. Una volta che l'utente ha completato quello che il suo lavoro e la navigazione tra le sezioni sarà necessario trovare un modo per poterlo automaticamente ricondurre al modulo creato in precedenza e far si che vengano lasciati dei feedback.

Questa operazione prevede l'implementazione di una nuova sezione all'interno dell'applicazione che sarà sempre accessibile agli utenti proprio come risultano accessibili le altre.
Questa sezione sarà però completamente diversa rispetto alle precedenti, in quanto non implementerà alcuna funzionalità particolare riguardante l'analisi dei dati ma bensì dovrà mostrarsi all'utente come una pagina di cortesia in cui si invita quest'ultimo a lasciare dei feedback.

L'implementazione di questa sezione prevede l'utilizzo di strumenti e ambienti di lavoro già citati in precedenza e che risultano di fondamentale importanza per la riuscita dell'obiettivo e l'integrazione all'interno dell'applicazione principale.

Si riprende nuovamente **PyCharm**, già utilizzato per la creazione della webapp per l'analisi statistica dei risultati della prima survery. In questo caso è stato creato un nuovo progetto, il che ha reso necessario l'importazione di vari moduli e il setting del nuovo ambiente virtuale di lavoro.
La struttura base di questa applicazione riprende sostanzialmente la struttura su cui si basa **ADESCA**, e quindi prevede nuovamente l'utilizzo del framework **Flask**.

Grazie alla combinazione di questi strumenti è possibile sviluppare questa ultima sezione in modo spedito e coerente con quanto già presente. Il primo passo consiste nella creazione di un template di base che dovrà essere poi presentato all'utente (*Figura 33*). In questa prima schermata vi dovranno essere delle spiegazioni e dei ringraziamenti per l'utente, il quale dovrà essere appunto invogliato a lasciare dei feedback, evitando appunto la chiusura prematura dell'applicazione.

Figura 33:
Schermata dell'applicazione in fase si sviluppo in cui è presente un *message box* contenente le informazioni necessarie per l'utente e il collegamento per lasciare il feedback.

La struttura di questa sezione è stata pensata per renderla più lineare e semplice possibile, andando quindi a non complicare la situazione dell'utente.
Questa sezione è stata suddivisa in tre sezioni principali (come mostrato in *Figura 34*):

- **Home**
- **Recording Page**
- **Module Information**

Figura 34:
Immagine rappresentante la *navigation bar* in cui è possibile visionare le tre sezioni.

La prima sezione ha l'unico scopo di andare a fornire delle info utili all'utente e a ringraziarlo per aver utilizzato l'applicazione (come già citato in precedenza).
La seconda sezione invece è stata pensata per poter implementare una funzionalità innovativa in cui viene chiesto all'utente di andare a registrare un commento audio a caldo, nell'immediato post utilizzo di **ADESCA**, per fornire un'esperienza ancora più profonda e coinvolgente nell'uso dell'applicazione.
La terza e ultima sezione invece fornisce informazioni sul modulo implementato per la raccolta dei feedback e fornisce all'utente il collegamento diretto a quest'ultimo.

Una parte da non sottovalutare nell'implementazione di questa sezione è quella riferita alla lingua. Infatti, proprio come è accaduto per la fase iniziale e quindi il primo form, è stato necessario andare ad implementare anche in questo caso la funzionalità di poter cambiare linguaggio (*Figura 35)*.
Così, l'utente si troverà di fronte alla scelta di poter andare a selezionare la sua lingua madre e in tal caso le lingue corrispondono a *Italiano* ed *Inglese*.
Per la realizzazione della funzionalità è stato necessario sfruttare il linguaggio JavaScript.

L'applicazione, per come è stata strutturata, andrà a chiedere all'utente di selezionare la lingua non appena esso entra all'interno della sezione. Tale scelta una volta andati avanti non sarà più modificabile.

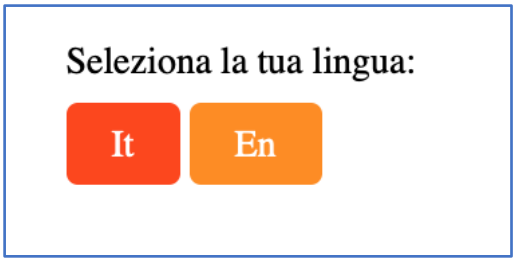


Figura 35:
Sezione della pagina iniziale in cui viene chiesto all'utente di selezionare la lingua.

Il modulo per la raccolta di commenti audio

L'implementazione di una funzionalità del genere fornisce all'utente un maggior coinvolgimento emotivo e soprattutto da un punto di vista dello sviluppo fornisce grande innovazione e tecniche più moderne per l'analisi e la conservazione dei dati.

Per poter implementare una funzionalità del genere ci si è serviti sempre del linguaggio **Python**, il quale fornisce delle librerie particolari per l'implementazione di questo tipo di features.
La libreria in questione è *sounddevice* e fornisce allo sviluppatore una serie di funzioni base da dover ottimizzare per la realizzazione di moduli che possano registrare suoni direttamente utilizzando il dispositivo dell'utente. [32]

La funzione è stata realizzata utilizzando le principali classi messe a disposizione, le quali sono state integrate con altri pezzi di codice per poter poi andare a soddisfare il principale requisito che è quello di renderlo accessibile agli utenti utilizzatori di **ADESCA**.

Il primo passo è stato quello di scrivere la funzione di registrazione che al momento di un clic su uno specifico *button*, inizia appunto a registrare i suoni intorno al dispositivo dell'utente, e quindi la sua voce.

La funzione necessita di andare a specificare alcuni parametri, come ad esempio il numero di secondi per la durata della registrazione, che in questo caso è stato limitato a trenta, in quanto risulta essere un compromesso più che giusto in termini di durata poiché c'è bisogno di estrapolare il maggior numero di informazioni cercando di usare meno risorse possibili. Un'altra necessità è quella di specificare il dispositivo che verrà utilizzato per la registrazione e che in tal caso corrisponde al microfono già integrato all'interno del dispositivo.

Ovviamente il processo di registrazione potrà essere manualmente interrotto dall'utente, che magari necessita di meno secondi per lasciare il commento audio. In ogni caso questa funzionalità si ripercuote su una seconda funzione di *stop* chiamata al click dell'apposito pulsante (*Figura 36*).

Tale registrazione sarà poi salvata in automatico direttamente nella directory specificata, oppure nel caso di successiva distribuzione sarà accessibile navigando nel file system del server.
Per poter salvare la registrazione e non commettere errori relativi alla sincronizzazione e al nome della registrazione, queste verranno salvate a partire dal loro *timestamp* di generazione e la loro estensione sarà *.wav* (un'alternativa al *.mp3*).

Una volta implementate le funzioni grazie alla libreria audio è necessario fornire all'utente la possibilità di utilizzare queste funzionalità, il che implica la costruzione di un template *ad hoc*. [33]
Utilizzando la combinazione di linguaggi quali **JavaScript** e **HTML**, è stato possibile andare a sviluppare dei template anche interattivi e in cui l'utente può svolgere le funzioni in maniera semplice e veloce.

Figura 36:
Schermata dell'applicazione in cui si mostra il principale template costruito per rendere possibile la registrazione audio da parte dell'utente.

Questo tipo di feedback lasciato dall'utente va a migliorare ulteriormente il processo di raccolta dati, in quanto essendo un metodo innovativo è possibile andare ad effettuare una serie di analisi specifiche, analizzando ad esempio la traccia audio tramite opportuni strumenti di terze parti oppure direttamente implementati tramite **Python**.

La pagina di collegamento al modulo di feedback

L'ultima pagina presente in questa sezione è quella relativa al collegamento al modulo. In questa pagina non sono state implementate delle funzioni particolari, ma bensì sono state fornite all'utente solo delle informazioni riguardanti il modulo e come esso è stato strutturato.

Fornire queste informazioni può risultare di fondamentale importanza in quanto l'utente deve poter decidere se intraprendere la compilazione o meno.
Uno schema riassuntivo potrebbe essere la scelta migliore poiché è il modo più diretto per poter comunicare con l'utente ed evitare che esso si infastidisca nell'andare leggere lunghi monologhi.

Il template su cui è stata costruita questa pagina è identico a quelle precedenti, in quanto è stato mantenuto lo stesso stile per poter fornire una continuità (*Figura 37*). Le info fornite sono presenti all'interno di un *box* ed informano l'utente sul tempo di compilazione del modulo e di come esso è strutturato, ovvero di come è possibile navigare al suo interno saltando le sezioni che non ha visitato e fornendo informazioni sul numero e tipo di domande.
Infine, l'accessibilità è stata semplificata al massimo fornendo un link diretto al modulo da questa pagina, in modo che l'utente non debba aprire ulteriori finestre nel browser.

Figura 37:
Schermata che rappresenta l'ultima pagina della sezione di valutazione di **ADESCA**.

Conclusioni e sviluppi futuri

ADESCA nasce come progetto di ricerca e mira al soddisfacimento di una serie di punti cardine, quali la possibilità di fornire supporto agli utenti che necessitano di analizzare dei dati nella maniera più semplice possibile e allo stesso tempo cerca di ottimizzare tutte le funzioni presenti al suo interno in modo che si possa usufruire di un'applicazione unica e performante.

Ad oggi l'applicazione consta di una serie di sezioni già implementate e funzionanti e non presenta troppe criticità impellenti da dover risolvere. Le funzionalità presenti in **ADESCA** sono il risultato di una serie di lavori portati avanti da diversi studenti che, a modo proprio, hanno dato una mano nello sviluppo di tale applicazione.

In questo lavoro condotto in vari mesi, è stato possibile porre le basi per l'implementazione di una sezione di **StoryTelling** più completa e complessa. Ricordando come questo sia un concetto molto importante e soprattutto uno strumento potente, in quanto le storie sono l'arma più potente per catturare l'attenzione delle persone:

"Stories are the most powerful form of content. Why? Because humans are hardwired for stories. Transformative hormones like dopamine and serotonin are released in the brain when we consume stories. That chemical reaction creates an emotional response within us, which helps us to remember a story's message, and can even compel us to take action." Cit.

Questo concetto abbraccia ormai diversi ambiti ed è oggetto di ricerca per molti, in quanto si cerca di trovare un modo per poter comunicare in maniera semplice e veloce anche il concetto più difficile. C'è da sottolineare il fatto che questo dipende anche dall'esperienza dell'utente che può avere più o meno determinate basi e quindi può riuscire a carpire prima determinati concetti.

La sezione di **StoryTelling** non mira alla sola visualizzazione dei dati ma, come già ripetuto diverse volte, punta ad una loro presentazione. Questo significa che grafici, immagini o altre strutture dovranno necessariamente avere una minima parte che possa giustificare il risultato ottenuto e che possa commentarlo, andando a massimizzare la potenza espressiva dell'intero processo.

L'idea di base è stata quella di avviare un vero e proprio processo di raccolta dati che consiste appunto nella creazione di un form. Una survey fornisce molte informazioni per chi deve sviluppare un sistema basato per migliorare l'esperienza utente, in quanto tutte le scelte effettuate dalle persone che hanno risposto al modulo, vanno a formare e rappresentare la popolazione di utenti che un domani potrebbe utilizzare **ADESCA**.
Il processo di creazione è stato abbastanza lungo se si fa riferimento allo sviluppo della struttura base del form e poi al processo di elaborazione e revisione delle domande, che sono state appunto poste ad occhio critico più volte per cercare di ottenere più informazioni dagli utenti.

Una prima fase che consta di questa messa online del modulo corrisponde ad un primo step, nel quale vengono collezionati tutti i dati e su cui ci si potrà basare per un'analisi successiva.

Il secondo passo da affrontare è stato quello di andare ad analizzare i vari dati ottenuti da questo processo, portando a termine una serie di obiettivi.
Il primo di questi è quello relativo allo sviluppo di un'applicazione web che possa visualizzare i dati raccolti e far si che quindi gli utenti e gli sviluppatori possano prendere nota di quello che è il risultato dell'analisi.

Questo processo è basato sulla creazione di un'applicazione per condurre un'*analisi statistica*, e quindi avere un'idea di quelli che sono stati i dati raccolti e di come poter iniziare a porre le basi per uno sviluppo futuro delle sezioni per i vari utenti di **ADESCA**.

Il secondo obiettivo è invece relativo all'estrazione delle *regole di associazione* a partire da quelli che sono i dati raccolti. Questo processo consiste nell'analizzare un dataset e poter estrarre della conoscenza inizialmente nascosta all'utente.
Ogni tupla viene presa in considerazione e viene confrontata rispetto alle altre e grazie ad opportune formule e funzioni fornisce come output dei risultati che mettono in correlazione i dati.

In questo modo è possibile capire quali siano state le scelte degli utenti e come una scelta possa aver condizionato un'altra, e così è anche possibile arrivare a capire gli ambiti in cui gli utenti sono più avanzati o meno e soprattutto la correlazione con quelli diversi.

Questo tipo di analisi è stata condotta in seguito ad una suddivisione per categoria di utente che può ricadere in *Rookie*, *Middle* o *Advanced*. E questo è stato fatto sia per avere un livello di dettaglio ancora maggiore e quindi una maggiore precisione nel processo di suddivisione (che possa tenere conto di tutte le singole scelte dell'utente) delle categorie ma anche per lo sfoltimento del dataset che avendo un grande quantitativo di attributi necessitava di potenti hardware alle spalle.

Analisi statistica e *regole di associazione* sono state le due strade intraprese nel post completamento del primo form e grazie a questi processi sarà possibile riuscire ad implementare una sezione di **StoryTelling** basandosi sulle preferenze degli utenti.

A valle di questi processi già completi, risulta necessario andare ad implementare anche un secondo form per la raccolta dei feedback degli utenti in seguito all'utilizzo di **ADESCA**. Questo si rende necessario in quanto il primo processo serve per poter buttare le basi per la sezione ed implementarla, mentre il secondo serve per poterlo migliorare.

Questo secondo modulo viene implementato direttamente seguendo le principali sezioni dell'applicazione in modo tale l'utente possa andare a fornire una valutazione quanto più precisa possibile e possa esprimere liberamente il suo giudizio.

Il processo è stato suddiviso in due parti: la creazione del modulo e lo sviluppo di una sezione da integrare in **ADESCA** per poter ricondurre l'utente a tale form e soprattutto l'implementazione di una innovativa funzione per lasciare appunto dei feedback.

Questa funzione consente all'utente di poter lasciare un commento audio che verrà poi analizzato successivamente con degli opportuni strumenti. Questo secondo modulo, quindi vede la combinazione di più tools per la raccolta dei feedback, che come già spiegato in precedenza, forniscono un grande supporto al miglioramento dell'applicazione.

In conclusione, il lavoro di sviluppo e di integrazione delle varie parti è risultato complesso e abbastanza lungo soprattutto considerando le tempistiche necessarie per la raccolta dei dati.
Il processo di sviluppo di **ADESCA** continuerà e questo lavoro ha gettato le basi per la costruzione vera e propria della sezione di **StoryTelling**.

I dati ottenuti aiuteranno i successivi sviluppatori nel processo di implementazione e gli forniranno una base su cui lavorare (considerando la possibilità di poter accedere ai risultati ottenuti dall'*analisi statistica* e a quelli estratti dalle *regole di associazione*) e delle linee guida nella fase più avanzata, in cui si dovrà tenere conto dei feedback degli utenti. Utilizzando i commenti audio sarà possibile anche tentare un nuovo approccio che consiste nell'analizzare direttamente le tracce ed estrarre il loro contenuto informativo.

Per finire, quindi, lo sviluppo di tale sezione sarà necessaria per poter fornire un supporto agli utenti qualsiasi sia il loro livello di esperienza e che necessitano di informazioni per poter orientare le loro scelte.

Bibliografia

[1] *Enhancing the friendliness of data analytics tasks: an automated methodology, Available at: http://ceur-ws.org/Vol-2841/DARLI-AP_15.pdf*

[2] *Data Storytelling: The Essential Data Science Skill Everyone Needs, www.forbes.com, Available at: https://www.forbes.com/sites/brentdykes/2016/03/31/data-storytelling-the-essential-data-science-skill-everyone-needs/*

[3] *Storytelling for Data Scientists, towardsdatascience.com, Available at: https://towardsdatascience.com/storytelling-for-data-scientists-317c2723aa31*

[4] *Data Storytelling: How to change statistics into stories that drive business action, www.youtube.it, Available at: https://www.youtube.com/watch?v=GFv9nHOpcS0&t=1373s*

[5] *Telling Stories with Data in 3 Steps (Quick Study), www.youtube.it, Available at: https://www.youtube.com/watch?v=r5_34YnCmMY*

[6] Machine Learning for Building Recommender System in Python, towardsdatascience.com, Available at: https://towardsdatascience.com/machine-learning-for-building-recommender-system-in-python-9e4922dd7e97

[7] *3 Approaches To Building A Recommendation System, towardsdatascience.com, Available at: https://towardsdatascience.com/3-approaches-to-build-a-recommendation-system-ce6a7a404576*

[8] *A Complete Overview of the Best Data Visualization Tools, www.toptotal.com, Available at: https://www.toptal.com/designers/data-visualization/data-visualization-tools*

[9] *Penthao data visualization 8 ways to better present data. www.extrasys.it, Available at: https://www.extrasys.it/en/smartblog/pentaho-data-visualization-8-ways-to-better-present-data*

[10] *Penthao platform plugin common ui. penthao.github.io, Available at:* https://pentaho.github.io/pentaho-platform-plugin-common-ui/platform/visual/

[11] *Visualization Types, help.penthao.com, Available at: https://help.pentaho.com/Documentation/8.1/Products/Data_Integration/Data_Integration_Perspective/Inspect_Your_Data/Visualization_Types*

[12] *Google Dashboard, google.com/dashboard, Available at: https://myaccount.google.com/dashboard*

[13] *Google Data Studio, support.google.com/datastudio, Available at: https://support.google.com/datastudio/answer/6283323?hl=it*

[14] *Tableau Web Site, tableau.com/trial, Available at: https://www.tableau.com/it*

[15] *Qlik Web Site, qlik.com, Available at: https://www.qlik.com/it-it/*

[16] Cookie e privacy policy, giuricivile.it, Available at: https://giuricivile.it/come-adeguare-proprio-sito-web-al-gdpr/#Tipo_di_cookies

[17] *Google Form Documentation, workspace.google.com, Available at: https://workspace.google.com/intl/en_ie/products/forms*

[18] *Jotform User Documentation, www.jotform.com, Available at: https://www.jotform.com/help/*

[19] *Top 50 Data Science Interview Questions and Answers for 2021, simplilearn.com, Available at: https://www.simplilearn.com/tutorials/data-science-tutorial/data-science-interview-questions*

[20] *An Indroduction to t-SNE with Python Example, towardsdatascience.com, Available at: https://towardsdatascience.com/an-introduction-to-t-sne-with-python-example-5a3a293108d1*

[21] *A Brief Overview of Outlier Detection Techniques, towardsdatascience.com, Available at: https://towardsdatascience.com/a-brief-overview-of-outlier-detection-techniques-1e0b2c19e561*

[22] *Time Series Data: serie storiche, andreaprovino.it, Available at: https://andreaprovino.it/time-series-data/*

[23] *Python Documentation, docs.python.org/3/, Available at: https://docs.python.org/3/*

[24] *Seaborn Documentation, seaborn.pydata.org, Available at: https://seaborn.pydata.org*

[25] *Pandas Documentation, pandas.pydata.org/docs/, Available at: https://pandas.pydata.org/docs/*

[26] *Google Colab Online Documentation, colab.research.google.com, Available at: https://colab.research.google.com/notebooks/intro.ipynb*

[27] *Highchart documentation and Demo, www.highcharts.com/demo, Available at: https://www.highcharts.com*

[28] *Q&A JotForm - Dropbox based image link, www.jotform.com/answers, Available at: https://www.jotform.com/answers/715086-Dropbox-based-image-link-is-not-working-as-a-product-image*

[29] *Q&A JotForm - How can I use images from Google Drive, www.jotform.com/answers, Available at: https://www.jotform.com/answers/1256214-How-can-I-use-images-from-Google-Drive-in-Image-Checkboxes*

[30] *Una semplice analisi statistica dei dati, vivalascuola.it, Available at: https://vivalascuola.studenti.it/come-fare-una-semplice-analisi-statistica-dei-dati-202332.html#steps_0*

[31] *Il tipo di analisi più semplice: L'Analisi Descittiva, lorenzogovoni.com, Available at: https://lorenzogovoni.com/il-tipo-di-analisi-dei-dati-piu-semplice-analisi-descrittiva/*

[32] *Playing and Recording Sound in Python, realpython.com, Available at: https://realpython.com/playing-and-recording-sound-python/#recording-audio*

[33] Official Documentation SoundDevice Python, python-sounddevice.readthedocs.io, Available at: https://python-sounddevice.readthedocs.io/en/0.4.1/

Ringraziamenti

Ripensando all'intero percorso sono stati tanti i dubbi e le incertezze che si sono presentate, le immense difficoltà legate alla vita accademica e anche agli aspetti personali, la curiosità e la voglia di vivere nuove esperienze e fare nuove conoscenze. Tutto questo, sinonimo di intraprendenza. Qualità da acquisire necessariamente, e una delle tante maturate durante questo bellissimo ed intenso viaggio.

Sono dovuti una lunga serie di ringraziamenti:

A partire da chi mi ha guidato e coinvolto in questo bellissimo lavoro di tesi, permettendomi di lavorare su un importante progetto. La Prof.ssa Tania Cerquitelli e Paolo Bethaz, sempre gentili e presenti per qualsiasi mio dubbio o richiesta.

Un ringraziamento particolarmente importante va ai miei compagni di avventura, gli amici e colleghi con cui abbiamo condiviso ansie, paure, sensazioni e gioie. Gli amici sono e resteranno un'ancora per la vita, specialmente se sono queste le persone con cui si sono condivise esperienze tali da legarci come una vera e propria famiglia. Ringrazio quindi Gianluca, Giacomo, Riccardo, Rebecca, Fabrizio, Nicola, Marianna, Leonardo e tutte le persone che hanno fatto parte di questo meraviglioso viaggio.

Un ringraziamento va a tutti i miei amici di sempre, coloro che non mi hanno mai abbandonato in nessuna circostanza e che mi hanno sempre appoggiato e sostenuto. Ringrazio Marcello, Giovanna, Maria, Matteo (Ciccio), Giuseppe.

Ringrazio i membri del mio gruppo della vita, Marco e Antonio, che hanno sempre rasserenato le mie giornate.

Ringrazio soprattutto il mio carissimo amico Mario per la sua costante presenza. Capace di portare gioia anche nei momenti più tristi che si sono presentati in questo percorso e con cui abbiamo condiviso momenti di immensa gioia.

Infine, il ringraziamento di gran lunga più importante va a tutta la mia famiglia, specialmente mio padre, mia madre e mio fratello, i quali mi hanno sempre supportato anche a distanza. Coloro che mi hanno sostenuto in tutti i momenti più difficili legati alle ansie e alle paure che sono però necessarie per poter crescere.
La mia famiglia, con cui ho condiviso tutte le gioie di questo grande traguardo.

Sono varie le cose che si imparano in un percorso lungo e articolato come questo, e per aggiungere degli ulteriori sentimentalismi, esprimo me stesso affermando il concetto che bisogna sempre provare a tirar fuori il meglio che si ha e ad adattarsi a tutte le situazioni che ci presenta la vita: be Brave!