POLITECNICO DI TORINO

Corso Di Laurea Magistrale in Ingegneria Informatica

Politecnico
di Torino

Tesi di Laurea Magistrale

Analisi di dati di micromobilità mediante tecniche di data mining

Supervisori

Prof. Silvia CHIUSANO

Dr. Elena DARAIO

Dr. Luca CAGLIERO

Candidato

Andrea VARA

Aprile 2022

Sommario

Al giorno d'oggi è una pratica comune noleggiare biciclette e monopattini elettrici, tramite i servizi di bike sharing e e-scooter sharing. Queste tipologie di mezzi sono caratterizzati da una notevole fruibilità e accessibilità da parte di un'ampia varietà di persone. L'utilizzo di questi mezzi sta cambiando le nostre abitudini. Questo cambiamento è dovuto alla facilità di utilizzo di questi mezzi, ma anche alla loro significativa ricaduta nell'ambito della mobilità green, ovvero una mobilità sostenibile in grado di ridurre l'impatto ambientale. L'utilizzo di questi mezzi ricade nella micro-mobilità, ovvero una mobilità relativa a spostamenti brevi in città. I mezzi di trasporto associati ai servizi di bike sharing e e-scooter non sono vincolati ad una stazione per il deposito o per il noleggio. L'utente può pertanto utilizzare il mezzo per raggiungere una destinazione di interesse, lasciando il mezzo nei pressi della destinazione stessa. In questa tesi viene presentata una metodologia per l'analisi dei dati di micro-mobilità. L'approccio utilizzato si basa sull'estrazione di pattern sequenziali che permettano di modellare i possibili spostamenti degli utenti. La metodologia utilizzata copre le diverse fasi del processo di Knowledge Discovery in Database (KDD). Partendo da una collezione di dati di mobilità viene prima eseguita una analisi preliminare esplorativa valutare la loro qualità ed integrità. Successivamente la collezione di dati di mobilità è arricchita integrando le informazioni relative ai luoghi di interesse raggiunti dall'utente. Infine la collezione dati risultante è analizzata per estrarne pattern di mobilità mediante l'algoritmo cSPADE. L'approccio utilizzato in questa tesi è generale è può essere utilizzato per l'analisi di mobilità relativi a diversi mezzi di trasporto.

Keywords: micro-mobility, dockless vehicles, mobility data analysis, POI, Data integration, sequential pattern mining, cSpade

Ringraziamenti

Prima di procedere con la trattazione, vorrei dedicare qualche riga a tutti coloro che mi sono stati vicini in questo percorso di crescita personale e professionale.

Ringrazio la Professoressa Chiusano, la Dottoressa Elena Daraio e il Dottor Luca Cagliero per la disponibilità ed il supporto tecnico.

Grazie ai miei amici e colleghi per essere stati sempre presenti anche durante questa ultima fase del mio percorso di studi. Grazie per aver ascoltato i miei sfoghi, grazie per tutti i momenti di spensieratezza.

Infine, un profondo ringraziamento va a mia madre, mio padre e mio fratello che mi sono stati accanto per ogni piccola e grande difficoltà di questi anni.

Indice

Elenco delle tabelle

Elenco delle figure

Acronimi

POI

Point of interest

OSM

OpenStreetMap

Capitolo 1

Introduzione

1.1 Contesto

Al giorno d'oggi è una pratica comune noleggiare biciclette e monopattini elettrici, tramite i servizi di bike sharing e e-scooter sharing. L'utilizzo di questi mezzi ricade nella micro-mobilità, questa è caratterizzata dall'impiego di mezzi di trasporto leggeri (*per es., monopattini elettrici , bici elettriche, skateboard ...*) ed è relativa a percorsi e distanze brevi in città. I mezzi messi a disposizione dai servizi di bike sharing e e-scooter sharing giocano un ruolo fondamentale nell'ambito della micro-mobilità urbana. I servizi legati a queste tipologie di mezzi si dividono in due categorie che si differenziano per la tipologia di noleggio: *Docked* e *Dockless*. La prima tipologia (*docked*), richiede il deposito dei mezzi in una delle stazioni di deposito/noleggio della città, al contrario la seconda(*dockless*) permette di lasciare i mezzi in un qualsiasi luogo pubblico. I provider dei servizi dockless forniscono delle piattaforme tramite le quali è possibile conoscere la posizione dei mezzi più vicini per il noleggio. Sfruttando queste piattaforme è possibile spostarsi autonomamente verso una destinazione di interesse, lasciando il mezzo nei pressi della destinazione stessa.

I dati di mobilità associati a questi mezzi sono solitamente caratterizzati da: un punto di partenza, un punto di arrivo, da un timestamp che indica l'istante di partenza e un timestamp che indica l'istante di arrivo. L'estrazione dei pattern di mobilità non è un lavoro semplice in quanto si devono tenere in considerazione diversi fattori come, per esempio, quali sono i luoghi di interesse da considerare per queste tipologie di mezzi ed il fatto che i pattern possono essere influenzati dal periodo preso in considerazione. Infatti, per quanto riguarda il secondo fattore, nel periodo invernale si tende ad utilizzare meno questa tipologia di mezzi per gli spostamenti.

Un punto di interesse [1] (in inglese point of interest - POI) è un luogo specifico

che qualcuno può trovare utile o interessante. Esempi di punto di interesse possono essere: monumenti, esercizi commerciali, hotel, stazioni di servizio o qualsiasi altra categoria utilizzata nei moderni sistemi di navigazione per autoveicoli. Un punto di interesse GPS specifica, come minimo, latitudine e longitudine del POI, assumendo un determinato dato cartografico. Di solito è incluso un nome o una descrizione per il POI.
L'individuazione dei POI può avvenire in diversi modi, ovvero: tramite i social network [2] come Facebook [3], Instagram [4], Foursquare [5] oppure possono essere ricavati dalle mappe come OpenStreetMap [6]. *OpenStreetMap (OSM)* [7] è un progetto collaborativo finalizzato a creare mappe del mondo a contenuto libero.

1.2 Lavori correlati

L'analisi dei dati di micromobilità è importante perché permette di capire come vengono utilizzati i mezzi di trasporto. In particolar modo attraverso l'analisi di questi dati è possibile individuare i periodi in cui vengono utilizzati maggiormente i mezzi e le aree della città in cui si spostano maggiormente gli utenti [8] [9]. Un'altra ragione per cui si vogliono analizzare questi dati è per capire per quali tipi di spostamenti vengono usati questi mezzi, ovvero modellare gli spostamenti degli utenti [2] [10]. Queste analisi sono importanti per i fornitori dei servizi di bike sharing e e-scooter sharing per distribuire meglio i loro mezzi nelle città.

Attualmente sono pochi gli studi su questa tipologia di mezzi, qui di seguito è riportato un elenco delle problematiche più rilevanti, ognuna con riferimenti ad alcuni contributi:

- *Analisi dei dati di micro mobilità:* l'analisi di questi dati è importante, perché permette di individuare i periodi in cui vengono utilizzati maggiormente i mezzi e le aree maggiormente frequentate dagli utenti. In letteratura sono presenti diversi lavori legati ai dati di micro-mobilità, ma questi studi sono prevalentemente legati ai servizi di bike sharing [11] [12]. Tra i lavori non legati al bike sharing troviamo [9] [8], dove sono stati presi in esame i dati relativi e-scooter. In questi lavori viene fatta prima un'analisi geo-spaziale per determinare le zone della città maggiormente frequentate. Successivamente viene costruito un modello di regressione binomiale negativo per stabilire delle relazioni tra l'utilizzo degli e-scooter e determinati luoghi della città [13] . Dai risultati emerge che questi mezzi vengono utilizzati prevalentemente nelle zone centrali della città e nei pressi dei campi universitari, mentre dal punto di vista temporale sono presenti delle differenze legate alla città esaminate.
- *Modellazione degli spostamenti:* In letteratura sono presenti degli studi per modellare gli spostamenti [10] [2]. Nel lavoro [2] viene fatto un confronto sulla

mobilità di due diverse città prendendo in considerazione i dati associai a due social media distinti. Per fare questo confronto sono stati creati dei grafi di transizione, tramite i quali è stato possibile modellare gli spostamenti comuni degli utenti e le aree della città di maggiore interesse.

Al giorno d'oggi esistono diversi provider per il servizio di bike sharing e e-scooter sharing. I mezzi che sono stati presi in considerazione in questo studio ricadono nella categoria *dockless*. In questa tesi viene presentata una metodologia per l'analisi dei dati di micro-mobilità, relativi alla città di Louisville(KY).
Partendo da una collezione di dati di mobilità viene prima eseguita una analisi preliminare esplorativa per valutare la loro qualità ed integrità. Successivamente la collezione di dati di mobilità è arricchita integrando le informazioni relative ai luoghi di interesse raggiunti dall'utente. Infine la collezione dati risultante è analizzata per estrarne pattern di mobilità mediante l'algoritmo cSPADE. Per poter integrare i luoghi di interesse è stato scelto OpenStreetMap utilizzando OverpassTurbo [14]. OverpassTurbo è un tool che permette tramite delle API di estrarre informazioni da OpenStreetMap.

La tesi è strutturata nel seguente modo:

Capitolo 2: Descrizione dei concetti teorici legati all'analisi associativa e all'estrazione di pattern sequenziali; **Capitolo 3**: Breve introduzione al framework; **Capitolo 4**: Descrizione della metodologia applicata; **Capitolo 5**: Descrizione dei dati presi in esame e i tool utilizzati per la metodologia presentata; **Capitolo 6**: Discussione dei risultati ottenuti applicando la metodologia presentata; **Capitolo 7**: Conclusione e lavori futuri;

Capitolo 2

Concetti generali

In questo capitolo verranno presentati i concetti teorici alla base di questa Tesi. I concetti esposti fanno riferimento al Data Mining.

2.1 Data Mining

Il *data mining* [15] è parte integrante del processo di Knowledge Discovery in Database (*KDD*). I task di data mining sono generalmente divisi in due categorie principali:

- *Task predittivi*, dove l'obiettivo è prevedere il valore di un particolare attributo basato sui valori di altri attributi.
- *Task descrittivi*, dove l'obiettivo è derivare pattern che riassumono le relazioni presenti nei dati. I task descrittivi sono spesso di natura esplorativa e richiedono spesso tecniche di post-elaborazione per convalidare e spiegare i risultati.

Uno dei task descrittivi è l'analisi descrittiva e viene utilizzata per scoprire i pattern che descrivono caratteristiche fortemente associate nei dati. I pattern scoperti vengono tipicamente rappresentati mediante regole di associazione.

2.1.1 Regole di associazione

Le regole di associazione ci permettono di estrarre legami di co-occorrenza tra gli item. Considerando un dataset contenente una collezione di transazioni (Figura 2.1), dove per transazione si intende un insieme di Item, si definisce regola associativa:

$$X \rightarrow Y$$

Dove: "$\rightarrow$" indica un'implicazione mentre X e Y sono degli itemset disgiunti. X

fa parte dell'antecedente, o corpo della regola mentre Y è il conseguente, o testa della regola.

tid	Items
1	{Milk, Bread}
2	{Beer, Diapers}
3	{Bread, Milk, Eggs}
4	{Bread, Milk, Diapers, Beer}

Figura 2.1: Esempio di transazioni

Definizioni

- **Item**: Considerando un dataset contenente una collezione di transazioni (Figura 2.1), si definiscono *item* gli oggetti presenti in tale transazione.
- **Item set**: è una collezione di Item
- **k-Itemset**: se un itemset contiene k item questo viene chiamato k-itemset
- **Frequenza**: Matematicamente può essere definita la frequenza di un itemset X come: $\sigma(X) = |\{t_i | X \subseteq t_i, t_i \in T\}|$
  Dove |.| rappresenta il numero di elementi in un set e T è l'insieme delle transazioni.

La forza di una regola di associazione può essere misurata in termini di supporto e confidenza.

- **Supporto**: il supporto di un itemset è ugale alla frequenza dell'intemset rispetto al numero totale di transazioni

$$Sup(X) = \frac{\sigma(X)}{(\#Transazioni)}$$

- **Confidenza**: è la frequenza del conseguente rispetto alle transazioni che contengono l'antecedente. Quindi non è altro che la probabilità condizionata di Y dato X.

$$Conf = P(Y|X) = \frac{P(X,Y)}{P(X)} = \frac{Sup(X,Y)}{Sup(X)}$$

- **Lift**: utilizzando la confidenza per l'estrazione delle regole si possono verificare dei problemi se la testa della regola è frequente. Infatti nella confidenza si tiene conto solo del supporto dell'antecedente e non del conseguente. Per risolvere questo problema è stato introdotto il lift.

$$Lift = \frac{Sup(X,Y)}{Sup(X)Sup(Y)}$$

Estrazione regole

L'estrazione delle regole di associazione può essere formulato nel seguente modo: Dato un insieme di transazioni T, si devono trovare tutte le regole che hanno un supporto $\geq minsup$ e una confidenza $\geq minconf$, dove $minsup$ e $minconf$ sono delle soglie definite dall'utente.

2.1.2 Tecniche di estrazione

Brute force

Consiste nell'enumerare tutte le possibili permutazioni degli item e per ciascuna permutazione si calcola supporto e confidenza. Il risultato sarà composto da quelle permutazioni che soddisfano le soglie di supporto e confidenza. Questo approccio è irrealizzabile a causa delle complessità computazionale.

Tecniche alternative

Invece di generare tutti i possibili candidati, si possono disaccoppiare i requisiti di supporto da quelli di confidenza. Così facendo si scompone il problema in due parti:

1. *Generazione degli itemset frequenti*: si trovano tutti gli itemset che soddisfano la soglia di supporto. Questi itemset vengono chiamati $itemsetfrequenti$

2. *Generazione delle regole*: si estraggono tutte le regole che hanno un valore alto di confidenza dagli itemset frequenti trovati nel passo precedenti.

Per la generazione degli itemset frequenti esistono diverse tecniche come *Apriori* e *Frequent-pattern growth*.

Apriori[16]: questo approccio si basa sulla seguente idea:
Se un itemset è frequente tutti i suoi sottoinsiemi di qualsiasi lunghezza sono anch'essi frequenti.
Da questo si evince che se un itemset non è frequente allora non lo saranno tutti i sui soprainsiemi. Quindi vale la proprietà antimonotona del supporto:
Considerando due itemset A e B se $A \subseteq B$ allora $Sup(A) \geq Sup(B)$

Questo approccio basato su dei livelli, ovvero ad ogni iterazione vengono estratti degli itemset di lunghezza k. Per ogni livello vengono effettuati due step:

- *Generazione dei candidati*: questo step si suddivide in altri due step
  - *Join*: vengono generati i candidati di lunghezza k +1 facendo il join degli itemset frequenti di lunghezza k che hanno in comune il prefisso di lunghezza k-1.
  - *Pruning*: viene applicato il principio Apriori, ovvero vengono esclusi i candidati di lunghezza k+1 ottenuti nello step precedente se almeno uno dei sottoinsiemi di lunghezza k non è contenuto nella lista degli itemset frequenti di lunghezza k.
- *Generazione degli itemset Frequenti*: vengono presi solo in considerazione solo i candidati che hanno un supporto maggiore della soglia minima. Questo viene fatto scansionando tutto il dataset per il calcolo del supporto.

Frequent pattern growth [16]: al contratio di Apriori, non vengono generati dei candidati ma viene creata una struttura compatta in memoria che contiene una rappresentazione compressa del dataset. Successivamente questa struttura chiamata FP-tree viene letta ricorsivamente per l'estrazione dei itemset frequenti.

2.1.3 Sequential pattern mining

L'estrazione delle regole di associazione viene spesso usata nella market basket analysis, tuttavia l'estrazione di questi pattern enfatizza solo legami di co-occorrenza tra gli item senza tenere in considerazione l'informazione sequenziale dei dati, ovvero le informazioni temporali non vengono prese in considerazioni. Sia $I = \{i_1, i_2, ..., i_m\}$ un set di m item distinti. Un evento è una collezione non vuota e non ordinata di item. Un evento è associato ad un itemset. Una sequenza è una lista ordinata di eventi. Una sequenza è del tipo :

$$\alpha_1 \rightarrow \alpha_2 \rightarrow ... \rightarrow \alpha_q$$

dove α_i è un evento.
Una sequenza può essere caratterizzata dalla sua lunghezza e dal numero di item in essa contenuti. La lunghezza di una sequenza corrisponde al numero di eventi presenti nella sequenza, mentre una k-sequence è una sequenza che contiene k item. Una sequenza α è una sottosequenza di un'altra sequenza s se ogni elemento(evento) ordinato in α è un sottoinsieme di un elemento ordinato in s.
La sequenza $\alpha = (\alpha_1, \alpha_2, ..., \alpha_m)$ è una sottosequenza di $s = (s_1, s_2, ..., s_n)$ se esistono degli interi $1 \leq j_1 \leq j_2 \leq ... \leq j_m \leq n$ tale per cui $\alpha_1 \subseteq s_{j_1}, \alpha_2 \subseteq s_{j_2}, ..., \alpha_m \subseteq s_{j_m}$. Se α è una sottosequenza di s, diremo che α è contenuta in s. Sia D un dataset che contiene uno o più dati sequenziali. Il termine dati sequenziali si riferisce a un elenco ordinato di eventi associati alla sequenza corrispettiva(vedi fig 2.2).
Il supporto di una sequenza s è la frazione di tutte le sequenze che contengono s.
Il problema dell'estrazione dei pattern sequenziali può essere formulato nel seguente modo:
Dato un sequence dataset D e data una soglia di supporto minimo, l'estrazione dei pattern sequenziali consiste nel trovate tutte le sequenze che hanno un supporto $\geq$ *minSup.*

Esistono diversi algoritmi per l'estrazione di pattern sequenziali come SPADE [17], GSP [18] e PrefixSpan [19]. Per i nostri esperimenti è stato usato cSPADE [20].

Sequenze		
sid	tid	Eventi
1	1	A, B, C
1	2	A,C,F
1	3	B,G
2	1	A,C
2	2	C,E
3	1	A,B
3	2	B,C
3	3	A,C,E

Figura 2.2: Sequence Dataset

Capitolo 3

Framework

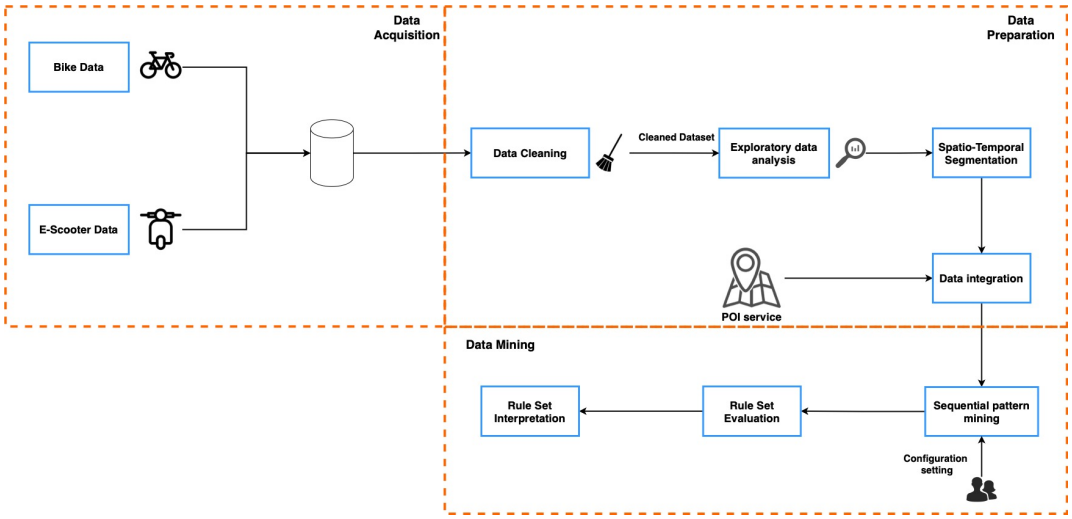


Figura 3.1: Framework proposto

Il Framework proposto è rappresentato in Fig 3.1 ed è composto dalle seguenti fasi:

- **Data acquisition**: Questa fase è dedicata all 'acquisizione dei dati di mobilità.
- **Data preparation**: In questa fase viene generato un cleaned dataset tramite la rimozione dei dati rumorosi e dei dati mancanti. Il dataset prodotto viene analizzato da un punto di vista spaziale e temporale per selezionare i dati utili. In fine vengono integrati i dati relativi ai POI presenti nell'area di interesse, individuata precedentemente, per la generazione dei dataset da dare in input alla fase successiva.

- **Data Mining**: In questa fase vengono estratti tramite un algoritmo di sequetial pattern minig dei pattern di mobilità. L'analisi effettuata ha esplorato diverse configurazioni di parametri. I risultati ottenuti sono stati interpretati prendendo in considerazione determinate metriche.

Capitolo 4

Metodologia

In questo capitolo vengono esaminate nel dettaglio le varie fasi viste nel capitolo precedente.

4.1 Data Acquisition

La Data Acquisition è la prima fase della metodologia proposta ed è un processo tramite il quale vengono acquisiti i dati provenienti da più sorgenti dati e integrati. Nel nostro caso di studio sono presenti due fonti di dati, una per i dati di mobilità e una per i luoghi di interesse. Per la prima parte della metodologia, verranno presi in considerazione i dati di mobilità.

4.2 Data Preparation

La seconda fase è Data preparation il cui scopo è trasformare i dati di input in un formato appropriato per l'analisi successiva. I passaggi solitamente coinvolti in questa fase sono: l'integrazione dei dati provenienti da più sorgenti dati, la pulizia dei dati per rimuovere il rumore e le osservazioni duplicate e la selezione di record e feature rilevanti per l'attività di data mining in corso.

Data Quality

I dati spesso non sono perfetti. Potrebbero esserci problemi dovuti a errori umani, limitazioni dei dispositivi di misurazione o difetti nel processo di raccolta dei dati. Potrebbero mancare valori o anche interi dati. In altri casi, potrebbero esserci dati duplicati. Sebbene la maggior parte delle tecniche di data mining possa tollerare un certo livello di imperfezione nei dati, focalizzarsi sulla comprensione

e al miglioramento della qualità dei dati in genere migliora la qualità dell'analisi risultante. Alcuni problemi legati alla *Data Quality* [15] sono i seguenti:

- *Rumore*: Si riferisce alla modifica dei valori originali.

- *Outliers*: Gli outlier possono essere sample che hanno caratteristiche diverse dalla maggior parte degli altri sample del dataset, oppure valori di un attributo che sono insoliti rispetto ai valori tipici per quello attributo. Inoltre, è importante fare una distinzione tra rumore e outlier. Gli outlier possono essere sample o valori legittimi. Pertanto, a differenza del rumore, gli outlier a volte possono essere interessanti.

- *Dati mancanti*: Mancanza di uno o più valori di attributo in un'oggetto. In alcuni casi, le informazioni non sono state raccolte. In altri casi, alcuni attributi non sono applicabili a tutti i sample. Indipendentemente da ciò, i valori mancanti dovrebbero essere presi in considerazione durante l'analisi dei dati. Esistono diverse strategie per trattare i dati mancanti, ognuna delle quali può essere appropriata in determinate circostanze.

Le strategie per la gestione dati mancanti sono le seguenti:

- *Eliminare i sample o gli attributi*: Una strategia semplice ed efficace consiste nell'eliminare i sample con valori mancanti. Tuttavia, un sample con dati mancanti contiene sempre delle informazioni e se molti sample hanno valori mancanti, può essere difficile o impossibile fare un'analisi affidabile. Tuttavia, se in un dataset sono presenti solo pochi sample con valori mancanti, potrebbe essere opportuno eliminarli. Una strategia correlata consiste nell'eliminare gli attributi con valori mancanti. Ciò dovrebbe essere fatto con cautela poiché gli attributi eliminati potrebbero essere quelli critici ai fini dell'analisi.

- *Stimare i valori dei dati mancanti*: In alcuni casi i valori mancanti possono essere stimati (interpolati) utilizzando i valori degli altri sample. Per esempio, se i dati mancanti fanno riferimento ad un attributo continuo, allora può essere utilizzato il valore medio dell'attributo dei k più vicini; se l'attributo è categorico, può essere preso il valore dell'attributo più comune.

- *Ignorare i dati mancanti durante l'analisi*: Molti approcci di data mining possono essere modificati per ignorare i valori mancanti.

Il rilevamento e la correzione dei problemi legati alla qualità dei dati viene chiamato Data Cleaning.

4.2.1 Data Cleaning

Prima di iniziare l'analisi, è buona norma verificare la qualità dei dati e pulirli. Poiché i dati di mobilità sono caratterizzati da: un punto di partenza, un punto di arrivo, un timestamp di partenza e un timestamp di arrivo; nel caso in cui vi siano dati rumorosi o dati mancanti relativi a questi campi caratterizzanti devono essere gestiti. Per questo motivo sono stati applicati i seguenti filtri:

- Rimozione dei sample con dati mancanti.
- Rimozione dei sample con coordinate sbagliate.

Per l'individuazione delle coordinate sbagliate, si deve vedere quali sono le coordinate che appartengono alla città/zona presa in esame. L'output di questa fase produce il cleaned dataset che può essere passato alla fase di analisi successiva.

4.2.2 Analisi esplorativa dei dati

I dati di mobilità contengono informazioni spazio-temporali, che possono essere utilizzati per fare delle analisi che servono per identificare i dati da prendere in considerazione per il processo di *Data Minig*. Tramite un'analisi temporale si possono individuare i periodi in cui vengono utilizzati maggiormente i mezzi presi in esame, mentre con un'analisi spaziale si possono identificare le aree maggiormente frequentate dagli utenti.

4.2.3 Spatio-Temporal Segmentation

Data Selection

La data selection è il processo mediante il quale vengono presi in considerazione solo i dati rilevanti. Alla luce delle analisi della fase precedente, sono stati preso in considerazione solo i trip effettuati nell'anno in cui si sono verificati il maggior numero di trip. Su questi dati sono state applicate delle segmentazioni temporali e spaziali.

Data Segmentation

Sui dati selezioni è stata applicata una segmentazione spaziale in modo tale da considerare solo le zone della città più frequentate. In fine sono state applicate delle segmentazioni temporali per vedere il cambiamento dei pattern.
Sono stati considerati due metodi per la segmentazione temporale:

- *Mesi più attivi*: si selezionano i mesi col maggior numero di trip effettuati.
- *Stagioni*: il dataset viene diviso in base alle stagioni

4.2.4 Data Integration

Partendo dai dati selezionati nella fase precedente, questi sono stati arricchiti integrando i dati rativi ai POIs della zona di interesse tramite *OpenStreetMap* [6]. OpenStreetMap contiene una raccolta di dati geografici che è possibile utilizzare per trovare i luoghi di interesse (POIs). OpenStreetMap usa tre tipologie di dato: *Nodes*, *Ways*, *Relations*.

I *nodes* servono per rappresentare un punto geografico, quindi queste tipologie di dato sono identificate da delle coordinate geografiche.
Le *ways* sono rappresentate da due o più nodes e possono essere utilizzate per rappresentare un'area come per esempio un parco.
Le *relations* vengono usate per organizzare più nodes o ways, quindi possono essere usati per rappresentare un aeroporto, uno stadio o un particolare tipo di edificio.
Tutte queste tipologie di dato sono caratterizzate da tag, ovvero delle coppie chiave-valore, che servono a descrivere il dato. Su OpenStreetMap sono presenti molti *tag*, nel nostro caso di studio sono stati presi in considerazioni solo i tag associati alle seguenti chiavi che permettono l'identificazione dei POIs:

- *Amenity*: identifica quelle strutture utilizzate da visitatori e residenti.
- *Shop*: Identifica quei luoghi legati ad un business, ovvero alla vendita di determinate merci
- *Tourism*: Identifica quei luoghi di interesse per i turisti.
- *Building*: identifica singoli edifici o gruppi di edifici collegati
- *Public_transport*: Identifica le stazioni/fermate dei mezzi pubblici

Poiché esistono diverse tipologie di edifici sono state prese in considerazioni solo le seguenti: *stadium*, *university*, *hospital*.

I POIs estratti sono stati suddivisi in 19 categorie riportate nella Tabella 4.1.

Categoria	
Bank	Luoghi in cui è possibile ritirare o depositare denaro
Bar/Cafe	Luoghi in cui è possibile mangiare ed intrattenersi con altre persone.
Vehicles_related_places	Luoghi legati ai veicoli motorizzati
Bicycle_related_places	Luoghi legati alle bici
Diplomacy/Services	Luoghi legati ai servizi pubblici
Entertainment	Luoghi legati all'intrattenimento
Food	Luoghi in cui è possibile prendere del cibo da asporto
Health	Luoghi legati alla cura della persona
Market	Luoghi legati alla vendita di generi alimentari
Post_office	Luoghi legati ai servizi postali
Public_transport	Luoghi legati ai mezzi pubblici
Religious_place	Luoghi religiosi
Restaurant	Ristoranti
School/College	Luoghi legati all'apprendimento
Store/Shop	Negozi
Theater	Teatri
Hotel	Luoghi in cui è possibile soggiornare
Public_places	Luoghi pubblici
Other	Utilizzato quando vicino al punto di arrivo non sono presenti POIs

Tabella 4.1: Categorie di POI considerate

4.2.5 Creazione input sequence dataset

Per eseguire le tecniche di data mining, i dati devono essere trasformati in un formato idoneo alla tecnica utilizzata. Per l'estrazione di pattern sequenziali, si deve generare un *input sequence dataset*. Un input sequence dataset è un dataset che contiene uno o più dati sequenziali. Il termine dati sequenziali si riferisce a un elenco ordinato di eventi associati alla sequenza corrispettiva. Avendo i dati di micromobilità e i dati relativi ai POI, ogni viaggio corrisponde ad una sequenza (Capitolo 2.1.3) che ha due eventi che contengono rispettivamente le categorie associate ai POI che ricadono in un intorno del punto di partenza e del punto di arrivo. I POI vicini ad un punto possono essere pensati come i POI che ricadono all'interno di una circonferenza di raggio r. Ogni singola riga di questo dataset deve essere del formato :

cid tid numItems itemlist

Dove :

- *cid* : indica la sequenza di riferimento

- *tid* : indica l'identificativo dell'evento/un'informazione temporale
- *numItems*: il numero di item associati all'evento
- *itemlist*: lista ordinata di item per l'evento

cid, *tid* e *numItems* sono dei numeri interi mentre *itemList* è una lista di numeri interi, dove ogni numero indica uno specifico item. Avendo i dati di mobilità e i POIs, il dataset sarà generato nel seguente modo:

1. Ad ogni singolo trip è associato un numero intero distinto che verrà usato come *cid* (riga 8 Algoritmo 1)
2. Come *tid* verrà considerato lo slot temporale di inizio trip e fine trip ottenuto tramite discretizzazione temporale. La discretizzazione è stata fatta considerando slot temporali di 15 minuti, in quanto per una questione di privacy i timestamp sono stati approssimati ai 15 minuti più vicini. La discretizzazione è stata fatta suddividendo un giorno (24h) in slot di 15 minuti ottenendo così 96 slot ((24 * 60) /15 = 96). Gli slot temporali sono stati ottenuti considerando l'orario di inizio noleggio e fine noleggio (riga 9 Algoritmo 1).
3. Essendo i dati di mobilità caratterizzati da un punto di partenza e un punto di arrivo, è possibile calcolare i POI vicini al punto di partenza e quelli vicini al punto di arrivo (riga 10 Algoritmo 1). I POIs vicini ad un punto geo-referenziato possono essere pensati come i POI che ricadono all'interno di una circonferenza di raggio r. Il valore del raggio deve essere pensato come la distanza massima che un utente percorre per raggiungere il mezzo da noleggiare oppure come la distanza massima che percorre un utente dopo il noleggio del mezzo. Calcolando i POI vicini al punto di partenza e al punto di arrivo, verranno considerati come *numItems* il numero di categorie distinte associate ai POI vicini al punto di arrivo e al punto di partenza (riga 13 e 14 Algoritmo 1). Come *itemList* verranno considerate le categorie distinte di POI che ricadono vicino il punto di partenza e il punto di arrivo (riga 15 e 16 Algoritmo 1).
4. Avendo calcolato tutti i dati sopracitati, è possibile creare per ogni singolo trip due righe che hanno lo stesso *sid*, come *tid* lo slot temporale associati rispettivamente al tempo di inizio noleggio e fine, come *numItems* il numero di item associati al punto di partenza e al punto di arrivo e come *itemList* gli item associati al punto di partenza e al punto di arrivo. (riga 17 e 18 Algoritmo 1).

Quanto detto si può riassumere nel seguente pseudocodice:

Algorithm 1 Generazione input sequence dataset

procedure GENERAZIONE_INPUT_SEQUENCE_DATASET($trips, pois, r_d, r_a$)
    ▷ *trips* sono i viaggi
    ▷ *pois* sono i punti di interesse
    ▷ r_d è il raggio associato al punto i partenza
    ▷ r_a è il raggio associato al punto di arrivo
    $rows \leftarrow []$
    for (*index*,*trip*) in *trips* **do**
        $sid \leftarrow index$ ▷ calcolo sid
        $tidDeparture, tidArrival \leftarrow getSlots(trip)$ ▷ calcolo tid associati rispettivamente alla partenza e all'arrivo
        $startingPois, arrivalPois \leftarrow getPois(trip, pois, r_d, r_a)$ ▷ Calcolo dei POIs vicini al punto di partenza e al punto di arrivo basato sui raggi
        $startingCategories \leftarrow getCategories(startingPois)$
        $arrivalCategories \leftarrow getCategories(arrivalPois)$
        $lenStartingItems \leftarrow len(startingCategories)$
        $lenArrivalItems \leftarrow len(arrivalCategories)$
        $startingR \leftarrow row(sid, tidDeparture, lenStartingItems, startingCategories)$
        $arrivalR \leftarrow row(sid, tidArrival, lenArrivalItems, arrivalCategories)$
        $rows \leftarrow append(startingR)$
        $rows \leftarrow append(arrivalR)$
    end for
    return *rows*
end procedure

4.2.6 Considerazioni

Questa metodologia permette di analizzare diverse tipologie di situazioni tramite la configurazione di parametri caratterizzanti (Figura 4.1). Tra questi, si deve considerare la distanza tra il punto di partenza e il punto di arrivo, intesa non come distanza percorsa ma come distanza relativa. Poichè molti viaggi coprono brevi spostamenti, nel momento in cui viene generato il sequence dataset, potrebbero essere presenti trip a cui sono associati gli stessi POIs vicini al punto di partenza e al punto di arrivo. Questa uguaglianza è dovuta ai valori impostati per il raggio di partenza e il raggio di arrivo, ma anche alla distanza relativa tra i due punti di partenza e arrivo. Minore è la distanza relativa tra i due punti, maggiore è la probabilità che i punti di interesse vicini al punto di partenza e al punto di arrivo coincidano all'aumentare dei raggi di partenza e arrivo. Per questo motivo, sono stati generati diversi sequence dataset al variare dei due raggi e filtrando i viaggi in base alla distanza relativa tra il punto di partenza e il punto di arrivo. Per i valori dei raggi sono state considerate le seguenti ipotesi:

- *simmetrica*

- *asimmetrica*

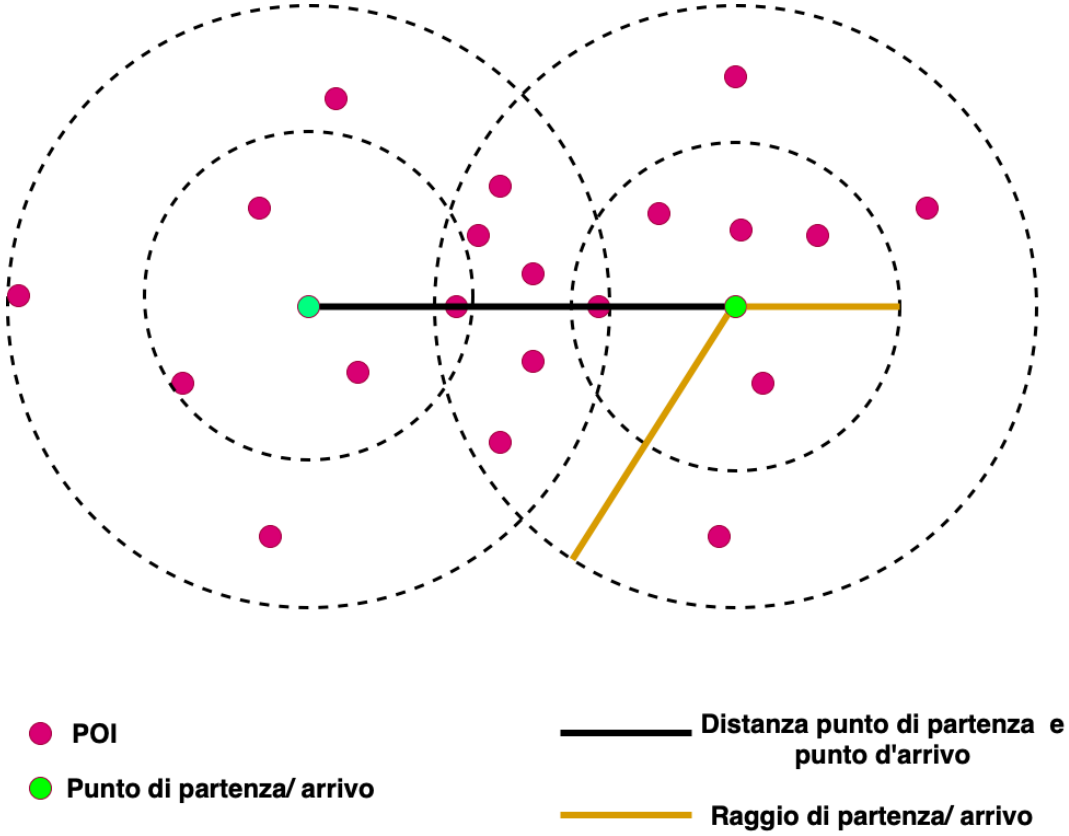


Figura 4.1: Considerazioni parametri

Nell'ipotesi simmetrica viene considerato lo stesso valore per il raggio di partenza e il raggio di arrivo, quindi si suppone che la distanza massima percorsa da un utente per raggiungere il mezzo da noleggiare sia la stessa di quella fatta dopo il noleggio. Nell'ipotesi asimmetrica si assume un valore per il raggio di partenza maggiore di quello di arrivo, quindi si assume che la distanza massima percorsa da un utente per raggiungere il mezzo da noleggiare sia maggiore della distanza percorsa dopo il noleggio. Nell'ipotesi asimmetrica si assume che la distanza percorsa per noleggiare un mezzo sia maggiore perché non è detto che vi sia disponibilità di mezzi nelle vicinanze e quindi si è costretti a camminare di più.
Un'altra considerazione è legata al numero di POI vicini al punto di partenza e al punto di arrivo. Variando il raggio di partenza e quello di arrivo, si modificano i POI che ricadono nell'intorno del punto di partenza e del punto di arrivo. Per la generazione dell'input sequence dataset, si è deciso di non considerare quei viaggi che

non hanno POI nell'intorno del punto di partenze, al contrario saranno considerati quei viaggi che non hanno POI nell'intorno del punto di arrivo, a condizione che abbiano almeno un POI nell'intorno del punto di partenza. Nell'ultimo caso verrà considerato per il punto di arrivo un POI appartenente alla categoria "Other" (da riga 13 a 20 Algoritmo 2). Con queste considerazioni lo pseudocodice può essere riformulato nel seguente modo:

Algorithm 2 Generazione input sequence dataset

1: **procedure** GENERAZIONE_INPUT_SEQUENCE_DATASET($trips, pois, r_d, r_a, d$)
2: ▷ *trips* sono i viaggi
3: ▷ *pois* sono i punti di interesse
4: ▷ r_d è il raggio associato al punto i partenza
5: ▷ r_a è il raggio associato al punto di arrivo
6: ▷ d distanza minima tra il punto di partenza e arrivo
7: $rows \leftarrow []$
8: $trips \leftarrow min\_distance(trips, d)$ ▷ si filtrano i viaggi considerando quelli che hanno una distanza tra il punto di partenza e il punto di arrivo maggiore della distanza minima
9: **for** (*index*,*trip*) in *trips* **do**
10: $sid \leftarrow index$ ▷ calcolo sid
11: $tidDeparture, tidArrival \leftarrow getSlots(trip)$ ▷ calcolo tid associati rispettivamente alla partenza e all'arrivo
12: $startingPois, arrivalPois \leftarrow getPois(trip, pois, r_d, r_a)$ ▷ Calcolo dei POIs vicini al punto di partenza e al punto di arrivo basato sui raggi
13: **if** $len(startingPois) == 0$ **then**
14: *continue*
15: **end if**
16: $startingCategories \leftarrow getCategories(startingPois)$
17: $arrivalCategories \leftarrow getCategories(arrivalPois)$
18: **if** $len(arrivalCategories) == 0$ **then**
19: $arrivalCategories \leftarrow$ ["*Other*"]
20: **end if**
21: $lenStartingItems \leftarrow len(startingCategories)$
22: $lenArrivalItems \leftarrow len(arrivalCategories)$
23: $startingR \leftarrow row(sid, tidDeparture, lenStartingItems, startingCategories)$
24: $arrivalR \leftarrow row(sid, tidArrival, lenArrivalItems, arrivalCategories)$
25: $rows \leftarrow append(startingR)$
26: $rows \leftarrow append(arrivalR)$
27: **end for**
28: **return** *rows*
29: **end procedure**

4.3 Data Mining

4.3.1 CSPADE

CSPADE [20] è un algoritmo che permette l'estrazione di pattern sequenziali, considerando una varietà di vincoli. Infatti tramite questo algoritmo è possibile: limitare la lunghezza o la larghezza di una sequenza, imporre vincoli sul gap minimo e massimo tra gli eventi, applicare finestre temporali su sequenze ammissibili, incorporare vincoli degli item, e trovare sequenze predittive di una o più classi. Questi vincoli sono stati introdotti perchè ad esempio, si potrebbe essere interessati solo a quelle sequenze ravvicinate nel tempo, quelle che si verificano molto distanti nel tempo, quelle che si verificano entro un determinato intervallo di tempo, quelle che contengono elementi specifici o quelle che predicono un determinato attributo. Questo algoritmo si basa su SPADE(Sequential pattern discovery using equivalent classes)[17], che utilizza un vertical data layout, ovvero per ogni item vengono riportate la coppie id-sequenza e id-evento (sid, eid). Lo spazio di ricerca è rappresentato da una struttura reticolare(lattice). Questo spazio di ricerca viene partizionato in tanti sotto-reticoli attraverso la nozione di classe di equivalenza, ciascuna partizione può essere processata applicando una delle seguenti strategie di ricerca: BFS (Breadth-First Search) o DFS (Depth-First Search). L'estrazione delle sequenze frequenti avviene tramite questi passi:

1. Calcolo *1-sequences* frequenti

2. Calcolo *2-sequences* frequenti

3. Decomposizione basata sui prefissi delle classi di equivalenza

4. Per ogni classe di equivalenza si generano tutte le sequenze tramite un JOIN sui prefissi di tutti gli elementi(atomi) (ovvero le sequenze di una determinata lunghezza che sono frequenti) e si controlla che il supporto ti tali sequenze sia maggiore o uguale al supporto minimo. Questo passo viene eseguito ricorsivamente applicando due strategie di ricerca BFS (Breadth-First Search) o DFS (Depth-First Search).

Tra i vincoli che è possibile impostare tramite cSPADE si trovano:

- Limite sulla lunghezza(*MaxLen*): è possibile impostare una lunghezza massima per le sequenze estratte. La lunghezza di una sequenza equivale al numero di eventi in essa presenti. La limitazione sulla lunghezza massima consentita per l'estrazione di pattern, viene fatta solitamente per rendere il task trattabile, poiché altrimenti si ottiene un aumento nel numero di sequenze frequenti.

- Limite sulla larghezza(*MaxSize*): è possibile impostare una larghezza massima per le sequenze estratte. La larghezza di una sequenza equivale al numero massimo di elementi in un evento. Anche in questo caso, la limitazione sulla larghezza massima consentita per l'estrazione di pattern, viene fatta solitamente per rendere il task trattabile.

- Gap minimo(*MinGap*): In alcuni casi potremmo essere interessati alle sequenze che si verificano dopo un determinato intervallo. Imponendo un gap minimo quindi è possibile impostare un vincolo temporale nelle sequenze, ovvero è possibile impostare un gap minimo per tenere in considerazioni solo quelle sequenze in cui il gap tra gli eventi e maggiore delle soglia minima.

- Gap massimo(*MaxGap*): Segue lo stesso concetto del gap minimo, solo che in questo caso viene impostata una soglia massima.

- Limitazione finestra temporale(*MaxWindow*): Una finestra temporale indica l'interesse verso quei pattern che si verificano all'interno di un intervallo temporale, cioè, invece dei vincoli di gap minimo e massimo che si applicano tra gli elementi della sequenza, la restrizione della finestra temporale si applica all'intera sequenza. Impostando la limitazione della finestra temporale viene specificata la differenza di tempo massima consentita tra l'ultima e la prima occorrenza di eventi in qualsiasi elemento di un pattern sequenziale.

4.3.2 Impostazione delle configurazioni

Per l'estrazione dei pattern sequenziali sono usate diverse configurazioni. Queste configurazioni riguardano sia i parametri usati per l'esecuzione di cSPADE ma anche i parametri utilizzati per la creazione dei dataset dai quali estrarre le sequenze.
Per i parametri relativi a cSPADE sono stati presi in considerazione solo i seguenti campi: *MinSup*, *MinGap*, *MaxGap*, *MaxSize*, *MaxLen*.

Mentre per gli input sequence dataset sono stai considerati i seguenti parametri: *Raggio partenza*, *Raggio arrivo*, *Distanza minima tra il punto di partenza e il punto di arrivo*, *Periodo temporale*.

Sono stati presi in considerazione questi parametri perchè variando i raggi di partenza e arrivo è possibile capire l'influenza del raggio sull'informazione estratta. Variando la distanza minima tra il punto di partenza e il punto di arrivo è possibile vedere l'impatto che ha la distanza sugli spostamenti. In fine applicando delle segmentazioni temporali è possibile ricercare i pattern ricorrenti nei diversi periodi temporali.

4.3.3 Valutazione delle sequenze

Per la valutazione delle sequenze sono stati presi in considerazioni le seguenti metriche (Capitolo 2.1.1):

- **Supporto**: Il supporto di una sequenza s è la frazione di tutte le sequenze che contengono s.
- **Confidenza**: è la probabilità condizionata di Y dato X.

$$Conf = P(Y|X) = \frac{P(X,Y)}{P(X)} = \frac{Sup(X,Y)}{Sup(X)}$$

- **Lift**: utilizzando la confidenza per l'estrazione delle regole si possono verificare dei problemi se la testa della regola è frequente. Infatti nella confidenza si tiene conto solo del supporto dell'antecedente e non del conseguente. Per risolvere questo problema è stato introdotto il lift.

$$Lift = \frac{Sup(X,Y)}{Sup(X)Sup(Y)}$$

Avendo estratto le sequenze, queste sono state ordinate per lift, confidenza e supporto per identificare le sequenze più significative.

Capitolo 5

Descrizione dei dati

Per il nostro caso di studio, sono stati presi in esame due tipologie di dato:

- Dati di mobilità
- Dati relativi ai POI

I dati di mobilità sono dei dati anonimizzati, relativi ad utenti che utilizzano il servizio di bike sharing e e-scooter sharing . Questi dati sono caratterizzati da: un punto di partenza, un punto di arrivo e dai timestamp di inizio e fine noleggio. Per l'estrazione dei pattern sequenziali, non bastano solamente i dati di mobilità ma servono anche i dati relativi ai POI per determinare quali potrebbero essere i possibili spostamenti fatti dagli utenti.

5.1 Dockless vehicles Louisville

I dati presi in esame sono degli open data relativi ai viaggi associati ai Dockless vehicles relativi a monopattini elettrici e bici della città di Louisville (KY). I dati fanno riferimento ai seguenti provider:

- Bird - lanciato nell'agosto 2018
- Lime - lanciato novembre 2018
- Bolt - lanciato a luglio 2019
- Spin - lanciato ad agosto 2019

I dati presi in esame sono composti dai seguenti campi:

- **TripID** - uid che identifica il trip

- **StartDate** – Data inizio trip, formato YYYY-MM-DD
- **StartTime** – Ora inizio trip, arrotondato ai 15 minuti più vicini in formato HH:MM
- **EndDate** – Data fine trip, formato YYYY-MM-DD
- **EndTime** – Ora fine trip, arrotondato ai 15 minuti più vicini in formato HH:MM
- **TripDuration** - durata del trip in minuti
- **TripDistance** - distanza del trip basata in miglia
- **StartLatitude** – Latitudine di partenza
- **StartLongitude** – Longitudine di partenza
- **EndLatitude** – Latitudine di arrivo
- **EndLongitude** – Longitudine di arrivo
- **DayOfWeek** – Giorno della settimana. Valori nel range 1-7, questi campo è derivato dalla data. Domenica = 1; Sabato = 7.
- **HourNum** - Fascia oraria da 0-24 di ottenuta da StartTime, utile per l'analisi

I dati presi in esame contengono sia campi che contengono informazioni spaziali che temporali. I campi che contengono informazioni spaziali sono: *StartLatitude*, *StartLongitude*, *EndLatitude* e *EndLongitude*; questi campi identificano rispettivamente le coordinate del punto di partenza e del punto di arrivo. I campi che contengono informazioni temporali sono: *StartDate*, *StartTime*, *EndDate* e *EndTime*; questi campi identificano rispettivamente il timestamp di inizio noleggio e fine noleggio.

Per il nostro caso di studio, sono stati ritenuti caratterizzanti per i viaggi i seguenti campi:

- **TripID**
- **StartDate**
- **StartTime**
- **EndDate**
- **EndTime**

- **StartLatitude**
- **StartLongitude**
- **EndLatitude**
- **EndLongitude**
- **DayOfWeek**
- **HourNum**

In più è stato considerato un campo derivato *MinDistance*, ottenuto calcolando la distanza relativa tra il punto di partenza e il punto di arrivo. Per calcolare questa distanza, data la latitudine e longitudine dei punti, è stata applicata la formula di Haversine [21]. La formula di Haversine calcola la distanza minima fra due punti posti su una superficie sferica (superficie terrestre). La distanza di Haversine viene calcolata nel seguente modo:

$$a = \sin^2 \frac{\Delta\varphi}{2} + \cos\varphi_1 \cdot \cos\varphi_2 \cdot \sin^2 \frac{\Delta\lambda}{2}$$

$$c = 2 \cdot atan2(\sqrt{a}, (\sqrt{(1-a)}))$$

$$d = R \cdot c$$

Dove φ è latitudine, λ è la longitudine, R è il raggio terrestre (raggio medio = 6371km);

5.2 OpenStreetMap dataset

OpenStreetMap contiene una raccolta di dati geografici che è possibile utilizzare per trovare i POI in una zona di interesse. Nel nostro caso di studio sono stati presi in considerazione i POI delle zone più frequentate dagli utenti che utilizzano queste tipologie di mezzi (bici e monopattini elettrici). I dati estratti sono caratterizzati dai seguenti campi:

- **ID**: è un numero intero che identifica il POI
- **type**: questo campo ci da un'informazione sul tipo di dato estratto da OpenStreetMap

- **lat**: identifica la latitudine del POI
- **lon**: identifica la longitudine del POI
- **name**: identifica il nome del POI
- **POI**: identifica la tipologia di POI
- **category**: identifica la categoria a cui è associato il POI

5.3 Strumenti

Per lo sviluppo di questa tesi è stato utilizzato Python come linguaggio di programmazione. Per il lavoro svolto sono stati usati i seguenti moduli:

- *pandas*: per la gestione dei file csv e tabelle
- *geopandas*: per la gestione di dati georeferenziati.
- *numpy*: per poter svolgere operazioni su vettori e matrici
- *matplotlib*: per la realizzazione dei plot
- *time e datetime*: per la gestione di dati temporali
- *overpay*: per usare le API di Overpass

Capitolo 6

Risultati e commenti

In questo capitolo sono riportati i risultati ottenuti seguendo la metodologia mostrata nel Capitolo 4.

6.1 Data acquisition

Per questo studio è stato preso in esame il dataset riportato nella Tabella 6.1 relativo ai dati di micromobilità della città di Louisville(KY). Tale dataset contiene 505993 samples relativi a viaggi effettuati nel periodo che va da ottobre 2018 a gennaio 2020.

Dataset	#samples	periodo
dataset1	505993	2018-08 al 2020-01

Tabella 6.1: Dataset analizzato

6.2 Data Preparation

Lo scopo di questa fase è trasformare i dati di input in un formato appropriato per l'analisi successiva. Per poter fare ciò i dati devono essere ripuliti e analizzati in modo tale da prendere in considerazione i dati rilevanti per l'attività di data mining.

Il dataset preso in esame è composto da 13 campi ma alcuni di questi presentano dei valori nulli : *EndDate*, *EndTime*.
EndDate e *EndTime* sono dei campi caratterizzanti per un viaggio, poiché rappresentano rispettivamente la data e l'ora di fine noleggio.

StartDate	StartTime	EndDate	EndTime	TripDuration	TripDistance	StartLatitude	StartLongitude	EndLatitude	EndLongitude
2019-07-10	22:45	NaN	NaN	0.0	0.0	38.225	-85.697	0.000	0.000
2019-07-10	22:45	NaN	NaN	0.0	0.0	38.224	-85.694	0.000	0.000
2019-10-28	14:15	NaN	NaN	0.0	0.0	38.262	-85.736	38.262	-85.736 2 14
2019-07-10	22:45	NaN	NaN	0.0	0.0	38.224	-85.694	0.000	0.000

Tabella 6.2: Dati mancanti

Nella Tabella 6.2 sono riportati i sample che presentano dati mancanti. Osservando questi sample si può notare la presenza di valori rumorosi come per esempio: EndLatitude, EndLongitude; per questo motivo questi sample sono stati rimossi.

Statistiche

	StartLatitude	StartLongitude	EndLatitude	EndLongitude	DayOfWeek	HourNum
count	505989.000000	505989.000000	505989.000000	505989.000000	505989.000000	505989.000000
mean	38.240815	-85.749592	38.238886	-85.746706	4.303007	14.104603
std	0.034990	0.081664	0.295808	0.602821	2.056213	4.620248
min	25.778000	-122.666000	-85.755000	-122.675000	1.000000	0.000000
25%	38.222000	-85.759000	38.222000	-85.759000	3.000000	11.000000
50%	38.251000	-85.755000	38.250000	-85.755000	4.000000	14.000000
75%	38.256000	-85.745000	38.256000	-85.744000	6.000000	18.000000
max	45.575000	-73.976000	48.863000	6.836000	7.000000	24.000000

Tabella 6.3: Statistiche attributi continui. Per ciascun attributo vengono riporti il numero di sample che hanno valori diversi dal volore nullo (count), il valore medio (mean), standard deviation (std), il valore minimo(min), il 25th percentile, il 50th percentile, il 75th percentile e il valore massimo (max)

	TripID	StartDate	StartTime	EndDate	EndTime
count	505989	505989	505989	505989	505989
unique	505989	540	97	541	97
top	0000045c-2677-3a7d-4b73-cad99a57	2019-09-07	15:45	2019-07-13	16:00
freq	1	3679	10506	3669	10412
max			24:00		24:00
min			00:00		00:00

Tabella 6.4: Statistiche attributi categorici. Per ciascun attributo vengono riportati il numero di sample che hanno valori diversi dal valore nullo (count), il numero di sample che hanno valori distinti(unique), il valore dell'attributo più frequente (top) e il numero di sample che hanno il volore più frequente (freq). Per i campi StartTime e EndTime sono riportati inoltre il valore massimo e minimo

Osservando le statistiche del dataset riportate in Tabella 6.3 e Tabella 6.4, emerge che alcuni sample presentano delle anomalie per alcuni campi:

- il campo HourNum assume valori in un range 0-24 invece di 0-23
- alcuni sample assumono valori per il campo StartTime uguale a 24:00
- alcuni sample assumono valori per il campo Endtime uguale a 24:00
- alcuni sample presentano del rumore nei campi relativi alle coordinate

Per il campo HourNum il rage 0-24 è dovuto al fatto che questo è derivato dal campo StartTime.
Osservando i valori minimo, massimo e medio delle coordinate relative al punto di partenza e al punto di arrivo (Tabella 6.3), si evince la presenza di coordinate anomale. Per l'individuazione di queste coordinate è stato preso in considerazione un file relativo ai confini e ai distretti della città di Louisville(KY). Tramite questo file è stato possibile individuare le coordinate rumorose. Nella Tabella 6.5 sono

Anomalie	#samples
Dati mancanti	4
Hournum == 24	771
StartTime == "24:00"	771
EndTime == "24:00"	841
Coordinate non appartenenti alla città	11377

Tabella 6.5: Anomalie riscontrate nel dataset

riportate le anomalie riscontrate. Osservando tale tabella possiamo notare che la prevalenza delle anomalie sono relative alle coordinate. Per poter analizzare i dati queste anomalie devono essere corrette attraverso il processo di *Data Cleaning*.

6.2.1 Data cleaning

Per la pulizia dei dati sono stati rimossi i sample con dati mancanti e i sample che hanno coordinate non appartenenti alla città di Louisville. Nella Tabella 6.6 sono riportati i filtri applicati, tra questi sono stati applicati i filtri per la rimozione dei sample che presentano dati mancanti e dei sample che presentano valori rumorosi per le coordinate. I sample con valori di HourNum, StartTime, EndTime uguali a 24 sono stati mappati col valore 0.

6.2.2 Analisi esplorativa

Per avere una visione più chiara sui dati sono state effettuate delle analisi da un punto di vista temporale e spaziale.

Filtro	#samples rimossi	Cardinalità post filtro
Dati nulli	4	505.985
Coordinate non appartenenti alla città	11.377	494.608

Tabella 6.6: Data cleaning

Distribuzione trip per ora di inizio

Sfruttando l'informazione relativa al campo *HourNum*, è possibile vedere la distribuzione dei viaggi rispetto ai vari timeslot, il primo timeslot è quello relativo a mezzanotte. Nella Figura 6.1 è riportata la distribuzione dei viaggi rispetto al campo *HourNum*. Da questa si evince che gli utenti utilizzano maggiormente i mezzi nel pomeriggio (15 - 17) e nell'orario di pranzo (12 - 14).

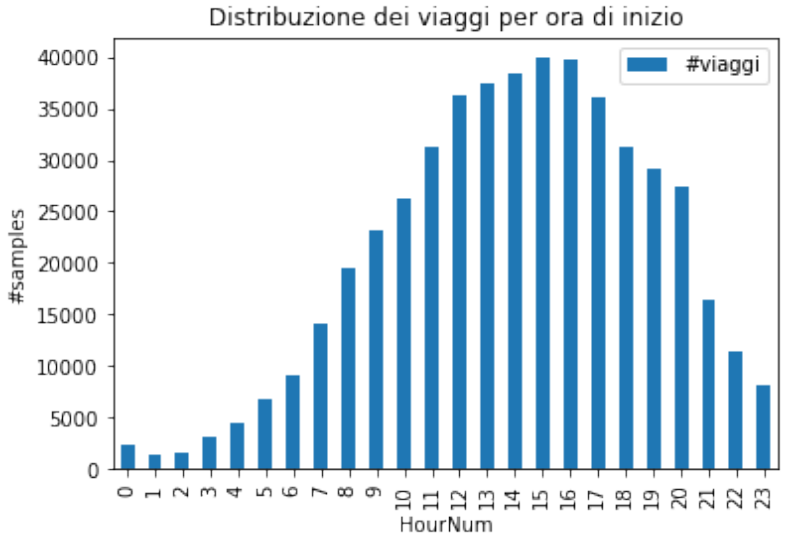


Figura 6.1: Distribuzione trip per ora inizio. Nell'asse delle ascisse sono riportati i timeslot, mentre nell'asse delle ordinate il numero di sample. Il timeslot *n* rappresenta sempre l'intervallo tra le ore *n* e *n+1* (ad esempio il timeslot 7 rappresenta le ore dalle 7:00 alle 8:00).

Distribuzione trip per giorno della settimana

Sfruttando l'informazione relativa al campo *DayOfWeek*, è possibile vedere la distribuzione dei viaggi rispetto ai giorni della settimana. Nella Figura 6.2 è riportata la distribuzione dei viaggi rispetto al campo *DayOfWeek*. Da tale distribuzione si evince che il maggior numero di trip sono avvenuti il venerdì e il sabato, inoltre i valori per la domenica non si discostano tanto dai giorni lavorativi.

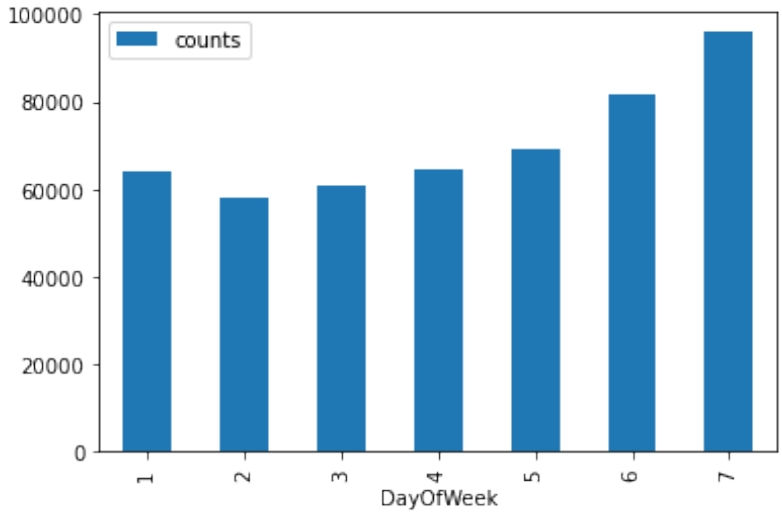


Figura 6.2: Distribuzione trip per giorno della settimana. Nell'asse delle ascisse sono riportati i giorni della settimana, mentre nell'asse delle ordinate il numero di sample. Valori ascisse: 1 = Domenica; 2 = Lunedì; 3 = Martedì; 4 = Mercoledì; 5 = Giovedì; 6 = Venerdì; 7 = Sabato;

Distribuzione trip annuali

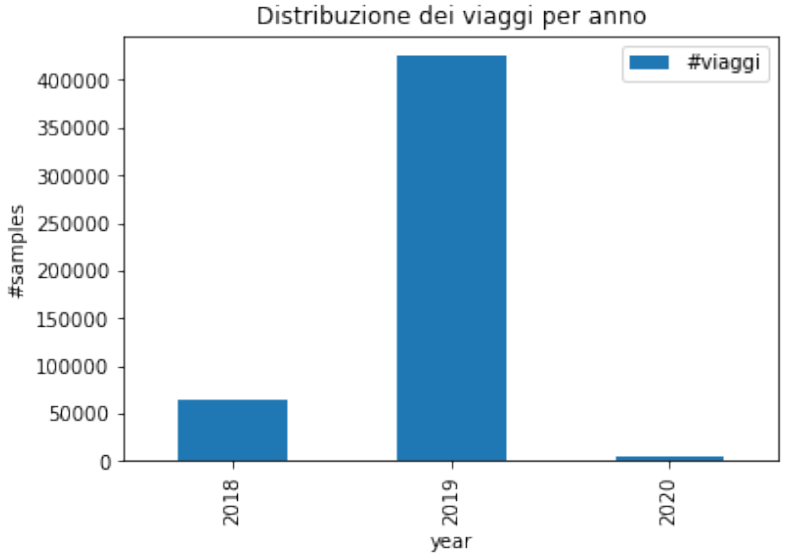


Figura 6.3: Distribuzione trip annuali. Nell'asse delle ascisse è riportato l'anno, mentre nell'asse delle ordinate il numero di sample.

Sfruttando l'informazione relativa all'anno, sono state calcolate le distribuzioni dei viaggi negli anni. I dati presi in esame coprono il periodo dal 2018-08 al 2020-01, infatti dalla Figura 6.3 si può notare che la maggior parte dei viaggi è relativo all'anno 2019.

Distribuzione trip mensili

Sfruttando l'informazione relativa ai mesi, è stata colcolata la distribuzione mensile dei viaggi (Figura 6.4). Da tale distribuzione si evince che la maggior parte dei viaggi sono relativi ai mesi di luglio, agosto e settembre del 2019, ovvero al periodo estivo. Confrontando i dati di agosto, settembre e Ottobre del 2018 con quelli del 2019 emerge che il numero di viaggi del 2018 sono molto più bassi di quelli del 2019. Questa differenza è dovuta al numero inferiore di provider per il servizio di noleggio, infatti alcuni servizi di noleggio sono stati lanciati nel 2019. I dati relativi all'anno 2019 presentano una netta differenza tra i trip fatti nella stagione estiva e quelli fatti nella stagione invernale. Tale differenza è riconducibile alle basse temperature raggiunte nei mesi invernali, infatti nel periodo invernale nella città di Louisville si raggiungono temperature sotto lo zero.

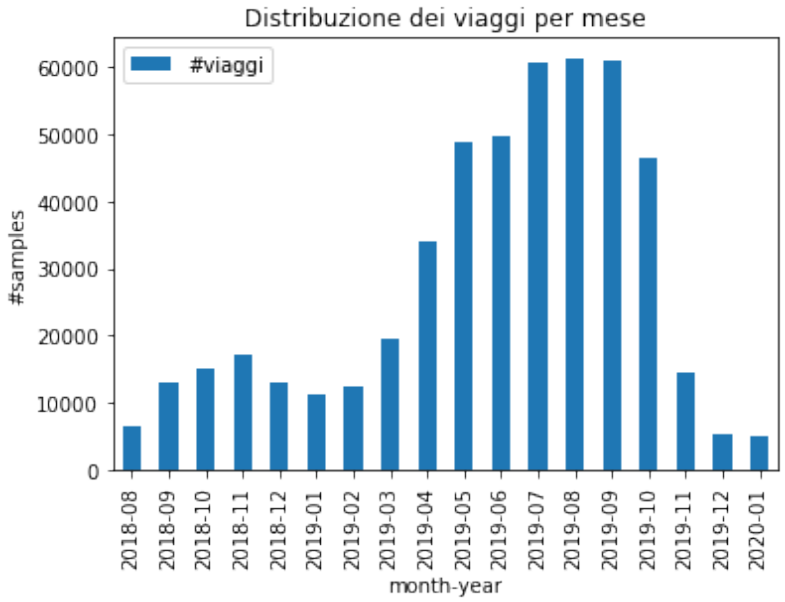


Figura 6.4: Distribuzione trip mensili. Nell'asse delle ascisse è riportato il mese e l'anno, mentre nell'asse delle ordinate il numero di sample

Distribuzione trip giornalieri

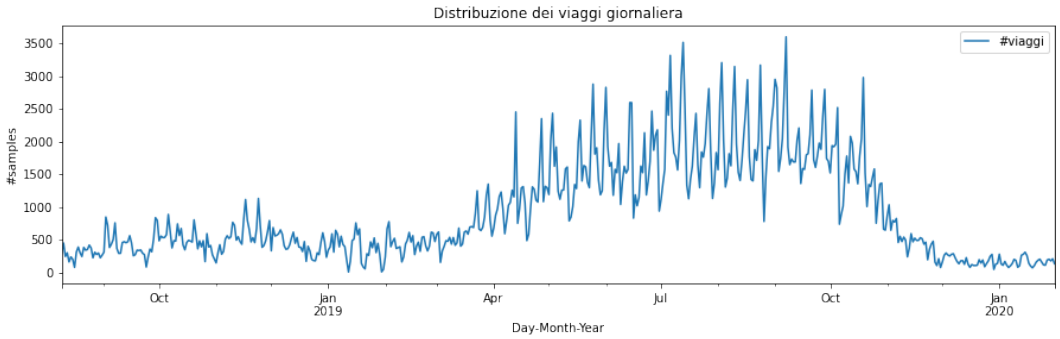


Figura 6.5: Distribuzione trip giornalieri. Nell'asse delle ascisse è riportato l'informazione relativa al giorno (dd-mm-yyyy), mentre nell'asse delle ordinate il numero di sample

Sfruttando l'informazione relativa ai giorni, è stata calcolata la distribuzione giornaliera dei viaggi. Dalla Figura 6.5 emerge che sono presenti picchi giornalieri relativi all'utilizzo dei monopattini e bici nelle giornate estive del 2019. Considerando le giornate del 2019 si ha un aumento giornaliero dei trip che inizia nel periodo primaverile del 2019 per poi diminuire drasticamente nel mese di novembre.

Confronto trips giorni lavorativi, festivi, weekend, festivi e weekend mensili

Prendendo in considerazione i mesi sono stati raggruppati i giorni in base a :

- **giorni lavorativi** : Come giorni lavorativi sono stati considerati tutti i giorni non festivi che vanno dal lunedì al venerdì.
- **giorni festivi**: Come giorni festivi sono stati considerati tutti i giorni festivi nel calendario americano più le domeniche.
- **weekend**: Come weekend sono stati considerati solo il sabato e la domenica.
- **giorni festivi e weekend**: sono stati presi in considerazione sia i giorni festivi e i weekend

Nella Figura 6.6 è riportato il confronto tra le distribuzione dei viaggi in base alla suddivisione dei giorni descritta in precedenza. Da tale confronto si evince un maggior numero di viaggi effettuati nelle giornate lavorative, inoltre il numero di viaggi fatti nei weekend dei mesi estivi è maggiore dei trip fatti nei giorni lavorativi dei mesi invernali.

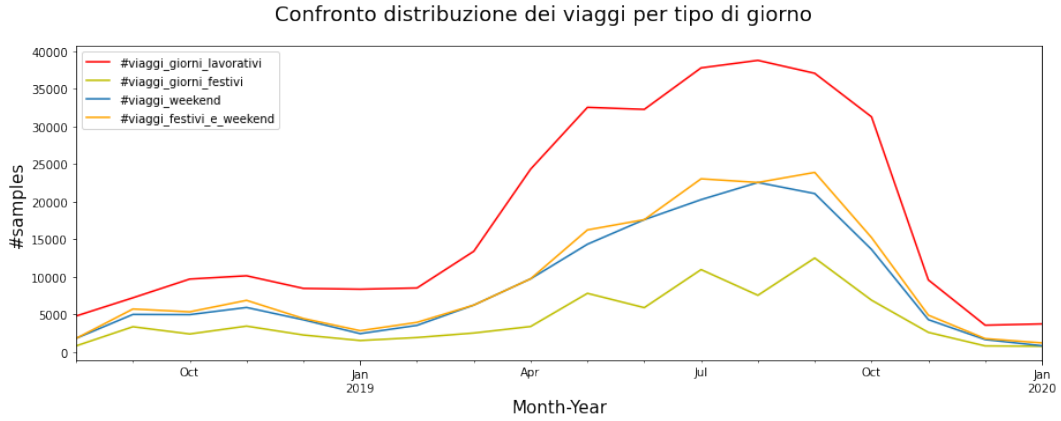


Figura 6.6: Confronto trips giorni lavorativi, festivi, weekend, festivi e weekend mensili. Nell'asse delle ascisse è riportato il mese e l'anno, mentre nell'asse delle ordinate il numero di sample.

Distribuzione geografica

Per determinare quali sono le aree maggiormente frequentate dagli utenti sono stati considerati i viaggi effettuati nel 2019 ed è stata considerata una mappa delle città di Louisville, contenente anche le informazioni sui distretti della città. Suddividendo tale mappe in aree di $1km^2$ sono stati conteggiati i trip effettuati in ciascuna delle seguenti aree. Avendo anche le informazioni sui distretti è stato fatto lo stesso lavoro per determinare i distretti maggiormente frequentati ed infine è stato fatto un confronto tra le aree e i distretti.

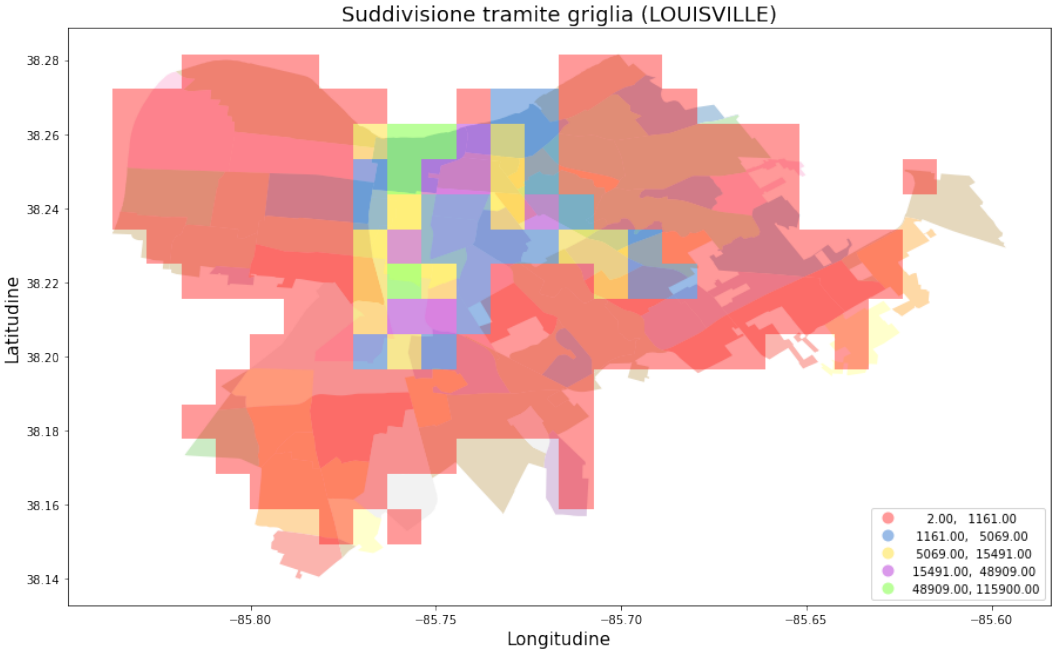


Figura 6.7: Suddivisione tramite griglia.

In Figura 6.7 è riportata la heat map delle aree frequentate dagli utenti, ottenuta suddividendo la città in una griglia e assegnando dei colori che indicano l'attività degli utenti nelle varie aree. I colori sono caratterizzati da un valore minimo e massimo di viaggi, in base al numero di viaggi effettuati in una zona viene assegnato il colore corrispondente.
Le aree delle heat map sono state classificate nel seguente modo:

- **zone verdi, viola e gialle**: zone maggiormente frequentate
- **zone blu**: mediamente frequentate

- **zone rosse**: scarsamente frequentate

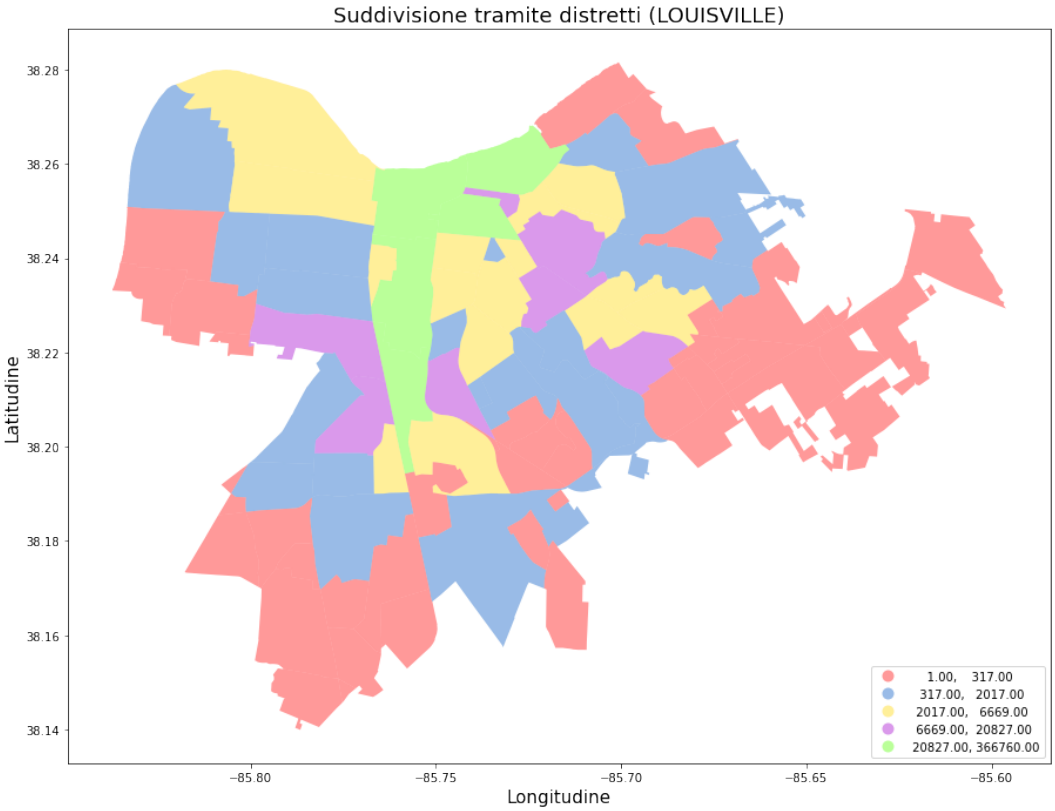


Figura 6.8: Suddivisione tramite distretti

Nella Figura 6.8 è riportata la heat map, ottenuta suddividendo la città nei distretti.
Le aree delle heat map sono state classificate nel seguente modo:

- **zone verdi e viola**: altamente frequentate
- **zone gialle e blu**: mediamente frequentate
- **zone rosse** : scarsamente frequentate

Confrontando le figure 6.7 e 6.8, si evince che le parti della città maggiormente frequentate sono le zone centrali. Se nella Figura 6.8 non vengono presi in considerazione i distretti rossi e blu, le zone della città maggiormente frequentate coincidono con le aree della Figura 6.7 non rosse.

6.2.3 Spatio-Temporal Segmentation

Data selection

Dei dati presi in esame sono stati considerati solo i viaggi relativi all'anno 2019, in quanto il 2019 è l'unico anno per il quale sono disponibili le informazioni relative a tutti i mesi.

Data Segmentation

Per le analisi successive è stata applicata una segmentazione spaziale, nella quale sono state prese in considerazione le aree della Figura 6.7 non rosse, ovvero le zone della città maggiormente frequentate. Guardando le distribuzioni temporali (Figura 6.4), sono state applicate due segmentazione temporali sui dati relativi all'anno 2019:

- Mesi più attivi: in questo caso sono stati considerati i 3 mesi in cui si sono verificati il maggior numero di viaggi. Dalla Figura 6.4, si evince che questi mesi sono quelli relativi a luglio, agosto e settembre. Questi mesi sono relativi all'inizio dell'estate , piena estate e fine estate.
- Stagioni: L'anno è stato partizionato in base alle stagioni, per capire come variano i pattern e per cercare i pattern significativi per le varie stagioni.

6.2.4 Data Integration

Considerando le aree non rosse della Figura 6.7, sono stati estratti i POIs che ricadono in tali zone. I POIs sono stati estratti da Openstreetmap tramite Overpass turbo. L'identificazione dei POI è stata fatta attraverso due step:

- Identificazione dei POI geo-referenziati da un punto (node)
- Identificazione dei POI descritti da un'area (way e relation)

Nel primo step sono stati considerati solo i POIs caratterizzati da un punto geografico. Per questa estrazione sono stati considerati solo le seguenti tipologie di dato: *Amenity*, *Shop*, *Tourism*, *Building*, *Public_transport*. Poichè esistono diverse tipologie di edifici sono stati considerati solo quelli relativi a *hospital* e *university*. Considerendo questi tipi di dato sono stati estratti in totale 1660 POIs.
Nel secondo step sono stati presi in considerazione anche POI descritti da un'area. Per mantenere una coerenza con i dati precedenti, è stato considerato un punto rappresentativo di tali aree, il centro. Per questa estrazione sono stati considerati solo le seguenti tipologie di dato: *Amenity*, *Shop*, *Tourism*, *Building*.

In questo caso per gli edifici è stato considerato oltre a *hospital* e *university* anche *stadium*.

Categoria	#POI
Public transport	633
Public Places	268
Restaurant	175
Store/Shop	115
Bar/Cafe	104
Vehicles related places	73
Bicycle related places	60
Entertainment	48
School/College	36
Diplomacy/Services	29
Health	29
Food	27
Religious place	23
Bank	17
Post Office	11
Market	9
Theater	2
Hotel	1

(a) Distribuzione dei POI geo-referenziati da un punto

Categoria	#POI
Vehicles related places	697
Public transport	635
Restaurant	290
Public Places	281
Diplomacy/Services	273
Store/Shop	175
Bar/Cafe	151
School/College	134
Religious place	107
Entertainment	88
Bicycle related places	72
Health	59
Bank	37
Food	35
Hotel	30
Market	24
Theater	17
Post Office	14

(b) Distribuzione dei POI geo-referenziati da un punto e da un'area

Tabella 6.7: Distribuzione POI nelle categorie

Considerando anche i dati che identificano un'area, il numero totale di punti di interesse è quasi raddoppiato infatti considerando anche questi dati sono stati estratti in totale 3120 POIs.
Guardando la distribuzione dei punti di interesse nelle varie categorie utilizzando solo i dati del primo step (Tabella 6.7a), alcune di queste hanno pochi elementi. Questo fenomeno è dovuto al diverso modo in cui possono essere rappresentati i dati su OpenStreetMap. Includendo i dati del secondo step (Tabella 6.7b), si verifica un aumento degli elementi che ricadono nelle varie categorie.

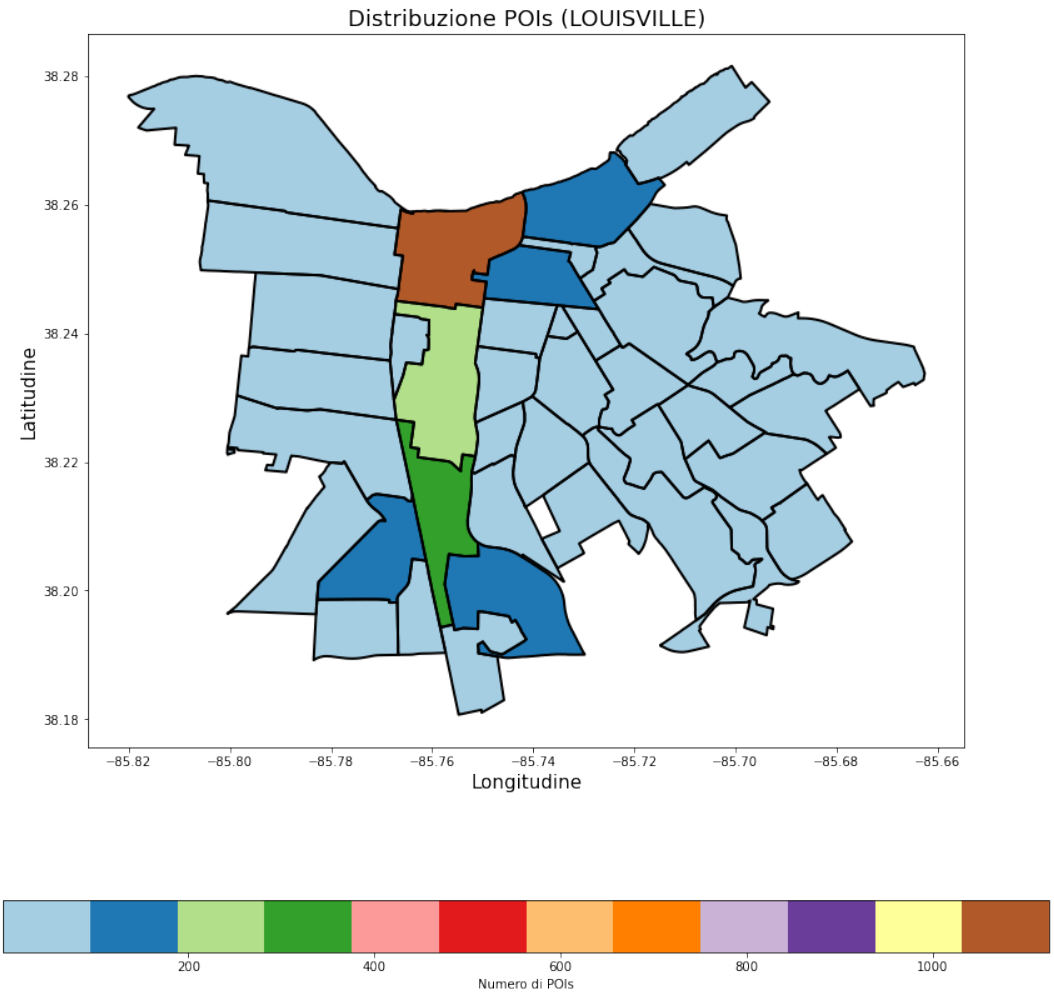


Figura 6.9: Distribuzione dei punti di interesse nei distretti integrando i nuovi dati

Avendo le informazioni sui distretti, è stato possibile vedere come sono distribuiti i punti di interesse nei vari distretti. In Figura 6.9 si evince che i distretti col maggior numero di POI sono quelli centrali. Quindi i trip si verificano maggiormente nella zona centrale della città che è anche quella zona con il maggior numero di POI.

Nella Tabella 6.8 sono riportati i distretti col maggior numero di luoghi di interesse. Tra questi distretti possiamo notare un distretto legato al business con 1125 luoghi di interesse e il distretto universitario frequentato dagli studenti.

Distretto	#POIs
CENTRAL BUSINESS DISTRICT	1125
UNIVERSITY	320
OLD LOUISVILLE	252

Tabella 6.8: Distretti col maggior numero di punti di interesse

Tassonomia

Avendo estratto i luoghi di interesse, questi sono stati divisi in 18 categoria in modo tale da poter determinare per ogni coppia punto di partenza e punto di arrivo le categorie di luoghi di interesse nelle vicinanze. I luoghi di interesse sono stati suddivisi considerando le categorie riportate nella Tabella 4.1. Nella tabella 6.9 sono riportati alcuni POI, individuati tramite OSM, suddivisi in alcune categorie.

Categoria	
Bank	bank, atm
Bar/Cafe	bar, cafe, pub
Bicycle_related_places	bicycle_repair_station, bicycle_parking
Diplomacy/Services	charity,community_centre, police, townhall
Entertainment	cinema, arts_centre, nightclub, museum
Health	hairdresser, hospital
Market	convenience shop, supermarket
Post_office	post_office, mailroom
Public_transport	bus_station
School/College	school,university
Store/Shop	boutique, clothes shop, jewelry
Hotel	hotel, motel

Tabella 6.9: Esempi di POI divisi nelle categorie

Tra questi esempi possiamo notare come nella categoria *Store/shop* ricadono diverse tipologie di negozio come per esempio negozi di vestiti e di gioielleria, mentre nella categoria *Hotel* ricadono luoghi in cui è possibile soggiornare come hotel e motel.

6.2.5 Configurazioni input sequence dataset

Avendo preso come rifermento una determinata area della città e avendo i punti di interesse associati a quell'area, sono stati creati diversi input sequence dataset tenendo in considerazione i seguenti parametri: *raggio di partenza*, *raggio di arrivo*, *distanza relativa tra il punto di partenza e il punto di arrivo*, *periodo temporale*.

Nella Tabella 6.10 sono stati riportati tutti i sequence dataset presi in considerazione negli esperimenti effettuati.

R.partenza(m)	R.arrivo(m)	Distanza minima(m)	Periodo	Esperimento
400	400		2019	1
400	200		2019	1
200	50		2019	1,2
100	50		2019	2
50	50		2019	2
200	50	100,200,300	2019	3
100	50	100,200,300	2019	3
50	50	100,200,300	2019	3
200,100,50	50	100	Mesi più attivi	4
200,100,50	50	100	Stagioni	4

Tabella 6.10: Sequence dataset considerati

6.2.6 Configurazioni cSpade

Avendo a disposizione gli input sequence dataset sono stati fatti degli esperimenti esplorando diverse configurazioni di cSpade. Nella Tabella 6.11 sono riportate alcune configurazioni prese in considerazione. Tra queste configurazioni si possono notare le configurazioni con $MinGap = 0$ che sono state utilizzate per prendere in considerazione quei viaggi in cui il timeslot di partenza coincide col timeslot di arrivo. Impostando come valore di supporto minimo = 1%, è stato possibile estrarre un buon numero di pattern. Impostando come valore di $MaxGap = 4$, verranno presi in considerazione i viaggi fatti in un arco temporale massimo di un'ora. In tutte le configurazioni esplorate è stato usato il valore $MaxLen = 2$, perchè ad ogni viaggio sono associati due eventi. Nella sezione dei risultati verrà usata la prima configurazione della Tabella 6.11.

MinSup	MaxSize	MaxLen	MaxGap	MinGap
1%	1	2	4	0
10%	1	2	4	0
1%	2	2	4	0
1%	3	2	4	0
1%	1	2	4	1

Tabella 6.11: Configurazioni cSpade

6.3 Risultati Sperimentali

In questa sezione sono riportati i risultati e le considerazioni, ottenute applicando la configurazione di riferimento di cSPADE ai sequence dataset della Tabella 6.10. Questa sezione sperimentale permette di evidenziare che i parametri relativi al raggio di partenza e al raggio di arrivo giocano un ruolo fondamentale per l'estrazione dei pattern sequenziali. Attraverso una prima fase di tuning è stato possibile individuare dei valori per questi parametri che permettono di estrarre sequenze con un valore alto di lift. Inoltre applicando delle segmentazioni temporali è stato possibile estrarre delle sequenze significative per il periodo preso in esame. Di seguito viene descritta una configurazione di riferimento per la generazione degli input sequence dataset, individuata mediante gli esperimenti e ritenuta ragionevole per affrontare l'analisi dei dati di micro-mobilità presi in esame. In particolar modo in questa sezione vengono descritti 4 esperimenti, di cui i primi 2 esperimenti riguardano il tuning dei parametri per la generazione dell'input sequence dataset, il terzo riguarda sia il tuning che il confronto con la configurazione di riferimento e il quarto esperimento riguarda solo il confronto con la configurazione di riferimento. Per ciascuno degli esperimenti riportati sono stati presi in considerazione le prime 30 sequenze estratte ordinate per lift, confidenza e supporto.

6.3.1 Configurazione di riferimento

La seguente configurazione è stata presa come riferimento per poter fare dei confronti con le altre configurazioni: Raggio partenza = 100 m; Raggio di arrivo = 50 m; Distanza minima = 100 m. In questa configurazione viene considerato un raggio di partenza non troppo grande, in modo tale da considerare lo spostamento degli utenti in un intorno di 100 m dal punto di partenza. Come raggio di arrivo viene considerato un valore entro il quale ricadono un numero accettabile di POI. Come distanza minima è stato considerato un valore non troppo grande in modo

tale da eliminare i round trip.

Periodo 2019

Nella Tabella 6.12 sono riportate le prime 30 sequenze estratte ordinate per lift, confidenza e supporto con questa configurazione.

Sequenze	rel. support	confidence	lift
Theater –> Theater	0.08	1.00	12.69
Post Office –> Post Office	0.08	1.00	12.42
Market –> Market	0.09	1.00	11.05
Food –> Food	0.12	1.00	8.29
Religious place –> Religious place	0.13	1.00	7.87
Health –> Health	0.15	1.00	6.73
School/College –> School/College	0.17	1.00	5.85
Bank –> Bank	0.18	1.00	5.53
Post Office –> Food	0.04	0.51	4.24
Food –> Post Office	0.04	0.34	4.24
Entertainment –> Entertainment	0.25	1.00	4.08
Hotel –> Hotel	0.25	1.00	4.00
Bank –> Post Office	0.05	0.27	3.31
Post Office –> Bank	0.05	0.59	3.26
Diplomacy/Services –> Diplomacy/Services	0.37	1.00	2.68
Bicycle related places –> Bicycle related places	0.37	1.00	2.67
Store/Shop –> Store/Shop	0.39	1.00	2.58
Public Places –> Public Places	0.46	1.00	2.19
Food –> Bank	0.05	0.39	2.17
Bank –> Food	0.05	0.26	2.17
Hotel –> Post Office	0.04	0.17	2.08
Post Office –> Hotel	0.04	0.51	2.03
Food –> School/College	0.04	0.35	2.02
School/College –> Food	0.04	0.24	2.02
Bar/Cafe –> Bar/Cafe	0.55	1.00	1.80
Diplomacy/Services –> Theater	0.05	0.14	1.79
Bank –> Store/Shop	0.12	0.69	1.78
Theater –> Diplomacy/Services	0.05	0.66	1.78
Store/Shop –> Bank	0.12	0.32	1.76
Store/Shop –> Food	0.08	0.21	1.74

Tabella 6.12: Sequenze configurazione di riferimento

Guardando le sequenze estratte nella Tabella 6.12, alcune di queste sequenze contengono la categoria *Bank*, queste sequenze evidenziano spostamenti da un luogo che potrebbe essere una banca o un atm verso luoghi legati alla vendita di cibo da asporto(*Food*), negozi(*Store/Shop*) o un ufficio postale(*Post_Office*). Tra le altre sequenze si trovano:

- *Market* → *Market*: questa sequenza evidenzia spostamenti da negozi di genere alimentare.

- *Food* → *Food*: la categoria food identifica quei luoghi dove si vende cibo da asporto, quindi questa evidenzia spostamenti tra stesse tipologie di luoghi.

- *Entertainment* → *Entertainment*: questa sequenze evidenzia spostamenti tra un luogo di intrattenimento ad un altro.

- *School/College* → *School/College*: questa sequenza evidenzia spostamenti da un plesso universitario/scuola ad un altro.

- *Store/Shop* → *Store/Shop*: nella categoria store/shop ricadono i negozi, quindi questa sequenza evidenzia spostamenti legati all'acquisto di beni di uso comune o per fare shopping.

- *School/Collage* → *Food* : questa sequenza evidenzia spostamenti da un luogo di apprendimento ad un luogo in cui è possibile acquistare cibo da asporto.

- *Bar/Cafe* → *Bar/Cafe*: nella categoria bar/cafe ricadono quei luoghi in cui le persone si riuniscono per conversare e nel frattempo mangiare o bere qualcosa. Questa sequenza evidenzia spostamenti da un bar/cafe ad un altro .

Mesi più attivi

Guardando le sequenze estratte con la configurazione di riferimento nei mesi più attivi (riportate nel dettaglio in Appendice A), alcune sequenze sono sempre presenti mentre altre sequenze sono presenti solo in alcuni mesi. Nella Tabella 6.13 sono riportate le sequenze presenti in tutti i mesi, tra queste è presente la sequenza *School/college* → *School/college*. La presenza di questa sequenza nel nel mese di agosto e settembre è riconducibile alla riaperture delle scuole, infatti le scuole americane riaprono nel periodo di agosto/settembre mentre nel periodo di luglio le scuole sono chiuse quindi la presenza di tale sequenza può essere ricondotta alle università, infatti nel periodo estivo le università sono aperte e solitamente nel periodo estivo avvengono le viste guidate nei campus universitari.

	Sequenze
0	School/College –> School/College
1	Bank –> Bank
2	Religious place –> Religious place
3	Food –> Bank
4	Hotel –> Post Office
5	Post Office –> Hotel
6	Public Places –> Public Places
7	Food –> Post Office
8	Hotel –> Hotel
9	Bank –> Food
10	Post Office –> Food
11	Bicycle related places –> Bicycle related places
12	Store/Shop –> Store/Shop
13	Post Office –> Post Office
14	Food –> Food
15	Entertainment –> Entertainment
16	Market –> Market
17	Theater –> Theater
18	Health –> Health
19	Diplomacy/Services –> Diplomacy/Services
20	Post Office –> Bank
21	Bank –> Post Office

Tabella 6.13: Sequenze uguali con raggio di partenza 100 m - mesi più attivi

Le seguenti sequenze sono presenti in alcuni mesi :

- *Hotel* → *Theater*: questa sequenza è presente nel mese di settembre. Questo vuol dire che il mese di settembre è importate dal punto di vista turistico.
- *Theater* → *Entertainment*: questa sequenza è presente nei mesi di luglio e agosto, questo evidenzia il fatto che il periodo estivo è più attivo dal punto di vista turistico.
- *School/college* → *Food*: questa sequenza è presente solo nel mese di settembre. La presenza di tale può essere ricondotta alla riapertura delle scuole.

Stagioni

Guardando le sequenze estratte con la configurazione di riferimento nelle stagioni, alcune sequenze sono sempre presenti mentre altre sequenze sono presenti solo in

alcune stagioni (riportate nel dettaglio in Appendice B). Nella Tabella 6.14 sono riportate le sequenze presenti in tutte le stagioni, anche in questo caso è presente la sequenza *School/College→School/College* in tutte le stagioni. Guardando la Tabella 6.14, 21 sequenze sono comuni a tutte le stagioni. La sequenza *School/College → Food* è presente in tutte le stagioni, questo evidenzia che una parte degli utenti sono studenti.

	Sequenze
0	School/College –> Food
1	School/College –> School/College
2	Bank –> Bank
3	Religious place –> Religious place
4	Hotel –> Post Office
5	Public Places –> Public Places
6	Food –> Post Office
7	Hotel –> Hotel
8	Post Office –> Food
9	Bicycle related places –> Bicycle related places
10	Store/Shop –> Store/Shop
11	Post Office –> Post Office
12	Food –> Food
13	Entertainment –> Entertainment
14	Market –> Market
15	Theater –> Theater
16	Food –> School/College
17	Health –> Health
18	Diplomacy/Services –> Diplomacy/Services
19	Post Office –> Bank
20	Bank –> Post Office

Tabella 6.14: Sequenze uguali con raggio di partenza 100 m - stagioni

Le seguenti sequenze sono presenti in alcune stagioni :

- *Theater → Entertainment*: questa sequenza è presente solo in primavera ed estate. Questo evidenzia il fatto che gli utenti preferiscono muoversi maggiormente versi in questi luoghi in questi periodi dell'anno.
- *Hotel → Health*: questa sequenza è presente solo nel periodo invernale, questo evidenzia il fatto che gli utenti preferiscono muoversi verso i centri legati alla cura delle persona nel periodo invernale.

- *Bar/cafe* → *Bar/cafe*: questa sequenza è presente in tutte le stagioni escluso l'estate, questo evidenzia il fatto che gli utenti preferiscono andare nei luoghi in cui è possibile conversare e riunirsi con amici in questi periodi.

- *Bank* → *Theater*: questa sequenza è presente solo nei periodi di autunno e inverno

- *Bank* → *Store/Shop*: questa sequenza è presente solo nei periodi di autunno e inverno

6.3.2 Esperimento 1

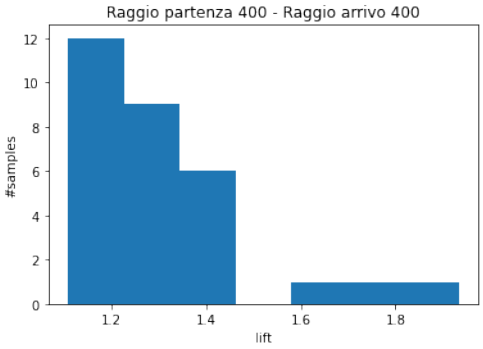
Per il primo esperimento sono stati considerati solo dei valori grandi per il raggio di partenza e raggio di arrivo, facendo un confronto tra l'ipotesi simmetrica e quella asimmetrica. Per questo esperimento sono stati considerati i sequence dataset ottenuti con le seguenti configurazioni:

- *Configurazione 1*: Raggio di partenza = 400m; Raggio di arrivo = 400m

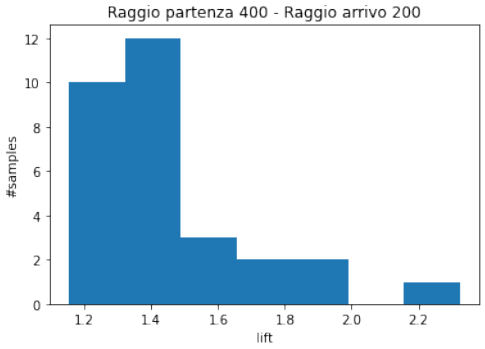
- *Configurazione 2*: Raggio di partenza = 400m; Raggio di arrivo = 200m

- *Configurazione 3*: Raggio di partenza = 200m; Raggio di arrivo = 50m

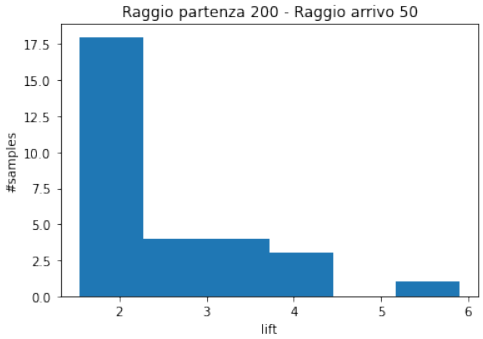
Per questo primo esperimento sono state considerate le prime 30 sequenze estratte ordinate per lift, confidenza e supporto. In Figura 6.10 sono riporti i valori di lift ottenuti dalla configurazioni presi in esame. Dai risultati di questo esperimento si evince che l'ipotesi asimmetrica porta a delle sequenze con valori di lift maggiore.



(a) Raggio di partenza 400 m - raggio di arrivo 400 m



(b) Raggio di partenza 400 m - raggio di arrivo 200 m



(c) Raggio di partenza 200 m - raggio di arrivo 50 m

Figura 6.10: Variazione del lift esperimento 1

Dal confronto delle sequenze della prima configurazione con quelle della seconda configurazione, si evince che 29 sequenze su 30 coincidono. Questo vuol dire che riducendo il raggio di arrivo l'informazione estratta non cambia molto. Confrontando le sequenze della seconda configurazione con le sequenze della terza configurazione, si evince che 20 sequenze su 30 coincidono, inoltre vi è un aumento significativo del lift (Figura 6.10).

6.3.3 Esperimento 2

Per questo secondo esperimento è stato ridotto ulteriormente il raggio di partenza per capire come varia l'informazione estratta rispetto al raggio. Nella Tabella 6.16 sono riportati il numero medio di POI che ricadono nell'intorno del punto di partenza al variare del raggio. Come si può notare da tale tabella riducendo il raggio fino ad un valore di 50 m ricadono un numero accettabile di POI. Il raggio

di arrivo non è stato ridotto in quanto, analizzando il numero di POI che ricadono nell'intorno del raggio di arrivo, si è visto che per valori inferiori a 50 m ricadono pochi POI negli intorni del punto di arrivo. Nella Tabella 6.15 sono riportati il numero medio di POI che ricadono nell'intorno del punto di arrivo e la percentuale di trip che ha almeno un POI in quell'intorno. Come si può vedere riducendo troppo il raggio di arrivo si riducono notevolmente i viaggi che hanno almeno un POI nell'intorno.
La riduzione del raggio di partenza è stata effettuate per vedere l'impatto che ha raggio rispetto all'informazione estratta. Per questo esperimento sono stati considerati i sequence dataset ottenuti con le seguenti configurazioni: Raggio di partenza = 200,100,50 m; Raggio di arrivo = 50m

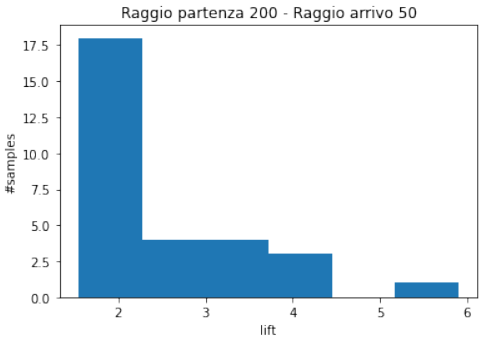
Raggio di arrivo(m)	Media POI	%Viaggi
50	4,4	71%
30	2,3	51%
10	1,3	7,6%

Tabella 6.15: Considerazioni sul raggio di arrivo

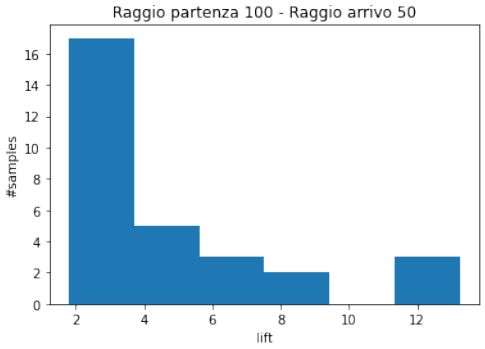
Raggio di partenza(m)	Media POI
50	4,6
100	13
200	43

Tabella 6.16: Considerazioni sul raggio di partenza

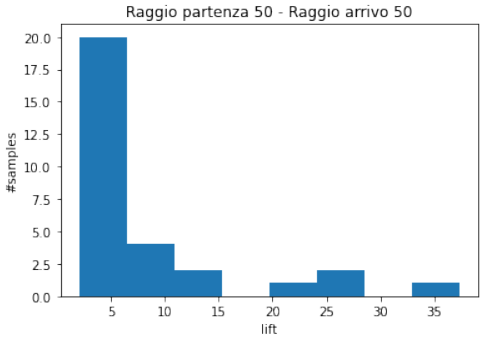
Facendo un confronto tra le sequenze in termini di lift, si evince un aumento del lift diminuendo il raggio di partenza. Dai risultati riportati in Figura 6.11, si evince che riducendo il raggio di partenza si ottengono dei risultati con valori di lift maggiori. Impostando come raggio di partenza 50 m il valore di lift aumenta notevolmente. Usare un raggio di partenza così piccolo implica brevi spostamenti degli utenti per il noleggio dei mezzi.



(a) Raggio di partenza 200 m - raggio di arrivo 50 m



(b) Raggio di partenza 100 m - raggio di arrivo 50 m



(c) Raggio di partenza 50 m - raggio di arrivo 50 m

Figura 6.11: Variazione del lift esperimento 2

Nella Tabella 6.17 sono riportate i confronti tra le diverse configurazioni prese in esame in termini di numero di sequenzi uguali. Dai confronti si evince una variazione di 10 sequenze diminuendo il raggio di partenza da 200 a 100 metri. Riducendo ulteriormente il raggio di partenza a 50 m, cambiano 11 sequenze rispetto alla configurazione con raggio di partenza 100 m, mentre rispetto alla configurazione con raggio di partenza 200 m cambiano 16 sequenze. Il cambiamento delle sequenze dovuto alla riduzione del raggio evidenzia come l'informazione estratta cambia nei diversi intorni.

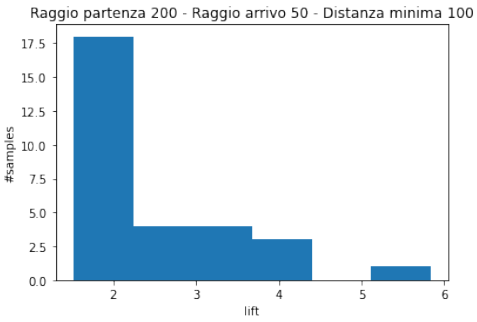
Raggio di partenza	Raggio di partenza	#sequenze uguali
200 m	100m	20
100 m	50m	19
200 m	50 m	14

Tabella 6.17: Variazione delle sequenze estratte al variare del raggio di partenza in termini di numero di sequenze uguali.

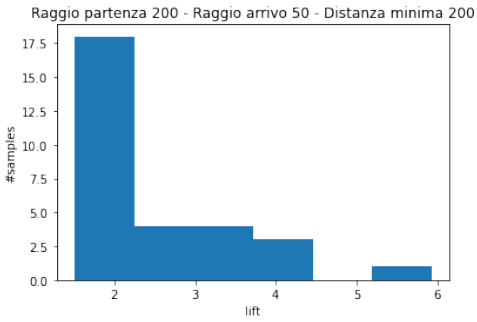
6.3.4 Esperimento 3

In questo esperimento è stata presa in considerazione la distanza tra il punto di partenza e il punto di arrivo. Tale considerazione nasce dal fatto che potrebbero essere presenti dei viaggi brevi o dei round trip tali per cui i POI nell'intorno del punto di partenza e del punto di arrivo coincidono. Per questo motivo, in questo esperimento sono stati posti dei vincoli relativi alla distanza tra il punto di partenza e il punto di arrivo, in modo tale da andare ad escludere quei viaggi tali per cui i POI che ricadono nell'intorno del punto di partenza e del punto di arrivo siano uguali. In questo esperimento per il raggio di arrivo è stato usato il valore 50 m, mentre sono stati posti dei vincoli per il raggio di partenza e la distanza minima tra il punto di partenza e il punto di arrivo. Per questo esperimento sono stati considerati i sequence dataset ottenuti con le seguenti configurazioni: Raggio partenza = 200,100,50 m; Raggio di arrivo= 50 m; Distanza minima = 100,200,300m.

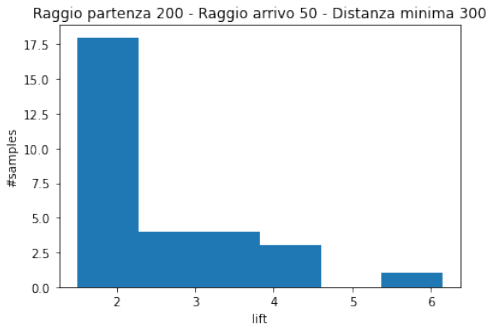
Da questo esperimento si evince un leggero aumento del lift aumentando la distanza minima tra il punto di partenza e il punto di arrivo. Questo aumento potrebbe essere dovuto alla riduzione delle sequenze di input che comporta la diminuzione del supporto di alcune sequenze
In Figura 6.12 sono riportati i valori di lift ottenuti impostando raggio di partenza 200 m, raggio di arrivo 50 m e variando la distanza minima. Osservando i risultati si evince un leggero aumento del lift.



(a) Raggio di partenza 200 m - raggio di arrivo 50 m - distanza minima 100m



(b) Raggio di partenza 200 m - raggio di arrivo 50 m distanza minima 200



(c) Raggio di partenza 200 m - raggio di arrivo 50 m - distanza minima 300

Figura 6.12: Variazione del lift esperimento 3

Confrontando le sequenze estratte a parità di raggio di partenza e di arrivo emerge una maggiore variabilità in termini di lift nella configurazione: Raggio di partenza = 50 m; Raggio di arrivo = 50 m.

Dalla Figura 6.13 si evince che impostando come raggio di partenza 200 m e variando la distanza minima, la variazione delle sequenze in termini di lift è minima, mentre nella configurazione con raggio di partenza 50 m si ha una variazione più accentuata. Questa variazione è dovuta al valore del raggio di partenza, infatti impostando come valore 50 m il numero di viaggi con almeno un POI nell'intorno del punto di partenza diminuisce, inoltre filtrando i viaggi per distanza vengono ridotti i viaggi da prendere in considerazione.

Distanza minima	Sequenza	Position	Lift
100 m	Bank → Entertainment	28	2,28
200 m	Bank → Entertainment	24	2,37
300 m	Bank → Entertainment	21	2,53
100 m	Public_places → Public_places	16	3,04
200 m	Public_places → Public_places	18	3,04
300 m	Public_places → Public_places	18	3,05
100 m	Hotel → Market	24	2,37
200 m	Hotel → Market	28	2,30
300 m	Hotel → Market	28	2,33

(a) Raggio di partenza 50 m - raggio di arrivo 50 m

Distanza minima	Sequenza	Position	Lift
100 m	Hotel → Theater	22	1,60
200 m	Hotel → Theater	23	1,61
300 m	Hotel → Theater	22	1,63
100 m	Hotel → Health	24	1,59
200 m	Hotel → Health	22	1,62
300 m	Hotel → Heath	23	1,63
100 m	Theater → Entertainment	31	1,51
200 m	Theater → Entertainment	28	1,52
300 m	Theater → Entertainment	28	1,52

(b) Raggio di partenza 200 m - raggio di arrivo 50 m

Figura 6.13: Variazione delle sequenze estratte (Esperimento 3)

Dalla Tabella 6.18 si evince che fissando il raggio di partenza e confrontando le sequenze astratte al variare della distanza minima, non si ha una grande variazione di sequenze. Da quanto detto in precedenza verrà utilizzato come valore di riferimento per i prossimi esperimenti il valore di distanza minima di 100 m, in modo tale da non filtrare troppi viaggi.

Nella Tabella 6.19 sono riportati il numero di sequenze uguali ottenute confrontando le diverse configurazioni al variare del raggio di partenza. Dal confronto si

Raggio di partenza	distanza minima	#sequenze uguali
200 m	300-200-100 m	28
100 m	300-200-100 m	29
50 m	300-200-100 m	29

Tabella 6.18: Variazione delle sequenze estratte al variare della distanza minima

evince una variazione di 10 sequenze, diminuendo il raggio di partenza da 200 a 100 metri. Riducendo ulteriormente il raggio di partenza a 50 m, cambiano 10 sequenze rispetto alla configurazione con raggio di partenza 100 m mentre rispetto alla configurazione con raggio di partenza 200 m cambiano 15 sequenze.

R.partenza	R.partenza	distanza minima	#sequenze uguali
200 m	100 m	100 m	20
100 m	50 m	100 m	20
200 m	50 m	100 m	15

Tabella 6.19: Variazione delle sequenze estratte al variare del raggio di partenza

Confronto con configurazione di riferimento

In questa sezione è stato fatto un confronto tra le sequenze estratte con la configurazione di riferimento e le configurazioni che hanno un raggio di partenza più piccolo o più grande. Tramite questo confronto si può vedere cosa accade aumentando o diminuendo il raggio di partenza.
Confrontando le sequenze estratte con la configurazione di riferimento (Tabella 6.12) con le sequenze estratte con raggio di partenza pari a 200 m (Tabella 6.20), si evince una differenza di 10 sequenze. Tra le sequenze presenti solo nella configurazione di riferimento sono presenti:

Bank –> Food, Bank –> Store/Shop, Bar/Cafe –> Bar/Cafe, Diplomacy/Services –> Theater, Food –> Bank, Food –> School/College, Public Places –> Public Places, School/College –> Food, Store/Shop –> Bank, Theater –> Diplomacy/Services.

Tra le sequenze presenti solo nella configurazione con raggio di partenza pari a 200 m sono presenti:

Bank –> Theater, Health –> Hotel, Hotel –> Health, Hotel –> Theater, Market –> Post Office, Post Office –> Market, Religious place –> Market, Theater –> Bank, Theater –> Entertainment, Theater –> Hotel

Confrontando i valori di lift delle sequenze comuni ad entrambe le configurazioni, emergono dei valori maggiori nella configurazione di riferimento.

Sequenze	rel. support	confidence	lift
Post Office –> Post Office	0.17	1.00	5.84
Theater –> Theater	0.24	1.00	4.17
Food –> Food	0.24	1.00	4.15
Market –> Market	0.25	1.00	4.03
Health –> Health	0.28	1.00	3.59
School/College –> School/College	0.29	1.00	3.40
Bank –> Bank	0.33	1.00	3.00
Religious place –> Religious place	0.37	1.00	2.69
Hotel –> Hotel	0.38	1.00	2.64
Post Office –> Food	0.11	0.62	2.58
Food –> Post Office	0.11	0.44	2.58
Theater –> Bank	0.16	0.68	2.04
Bank –> Theater	0.16	0.49	2.04
Entertainment –> Entertainment	0.49	1.00	2.03
Post Office –> Bank	0.11	0.64	1.91
Bank –> Post Office	0.11	0.33	1.91
Bicycle related places –> Bicycle related places	0.56	1.00	1.78
Diplomacy/Services –> Diplomacy/Services	0.58	1.00	1.73
Store/Shop –> Store/Shop	0.61	1.00	1.65
Hotel –> Theater	0.15	0.38	1.60
Hotel –> Post Office	0.10	0.27	1.60
Hotel –> Health	0.17	0.44	1.59
Theater –> Hotel	0.14	0.60	1.58
Post Office –> Hotel	0.10	0.59	1.57
Post Office –> Market	0.07	0.39	1.56
Health –> Hotel	0.16	0.59	1.55
Market –> Post Office	0.07	0.27	1.55
Theater –> Entertainment	0.18	0.75	1.52
Religious place –> Market	0.14	0.38	1.52
Market –> Religious place	0.14	0.56	1.51

Tabella 6.20: Sequenze conf: Raggio di partenza = 200 m - Raggio di arrivo = 50m - Distanza minima = 100m

Sequenze	rel. support	confidence	lift
Post Office –> Post Office	0.03	1.00	34.63
Food –> Food	0.04	1.00	25.93
Theater –> Theater	0.04	1.00	23.02
Market –> Market	0.05	1.00	19.16
Religious place –> Religious place	0.08	1.00	13.02
Health –> Health	0.10	1.00	10.05
School/College –> School/College	0.10	1.00	9.87
Bank –> Bank	0.10	1.00	9.66
Entertainment –> Entertainment	0.13	1.00	7.61
Hotel –> Hotel	0.14	1.00	6.90
Diplomacy/Services –> Diplomacy/Services	0.23	1.00	4.36
Bicycle related places –> Bicycle related places	0.26	1.00	3.89
Store/Shop –> Store/Shop	0.26	1.00	3.88
Public Places –> Public Places	0.33	1.00	3.04
Bank –> Store/Shop	0.08	0.78	3.01
Store/Shop –> Bank	0.08	0.31	3.01
Theater –> Entertainment	0.02	0.39	2.96
Entertainment –> Theater	0.02	0.13	2.95
Theater –> Diplomacy/Services	0.03	0.58	2.52
Diplomacy/Services –> Theater	0.02	0.11	2.49
Bar/Cafe –> Bar/Cafe	0.41	1.00	2.43
Market –> Hotel	0.02	0.34	2.37
Hotel –> Market	0.02	0.12	2.37
Diplomacy/Services –> Entertainment	0.07	0.31	2.34
Entertainment –> Diplomacy/Services	0.07	0.53	2.33
Bank –> Entertainment	0.03	0.30	2.28
Entertainment –> Bank	0.03	0.23	2.26
Bicycle related places –> Food	0.02	0.08	2.04
Food –> Bicycle related places	0.02	0.52	2.03
Public transport –> Public transport	0.50	1.00	2.01

Tabella 6.21: Sequenze conf: Raggio di partenza = 50 m - Raggio di arrivo = 50 m - Distanza minima = 100m

Confrontando le sequenze estratte con la configurazione di riferimento (Tabella 6.12) con le sequenze estratte con raggio di partenza 50 m (Tabella 6.21), si evince una differenza di 10 sequenze. Tra le sequenze presenti solo nella configurazione di riferimento sono presenti:

Bank –> Food, Bank –> Post Office, Food –> Bank, Food –> Post Office, Food –> School/College, Hotel –> Post Office, Post Office –> Bank, Post Office –> Food, Post Office –> Hotel, School/College –> Food

Tra le sequenze presenti solo nella configurazione con raggio di partenza pari a 50 m sono presenti:

Bank –> Entertainment, Bicycle related places –> Food, Diplomacy/Services –> Entertainment, Entertainment –> Bank, Entertainment –> Diplomacy/Services, Entertainment –> Theater, Food –> Bicycle related places, Hotel –> Market, Market –> Hotel, Theater –> Entertainment

Confrontando i valori di lift delle sequenze comuni ad entrambe le configurazioni, emergono dei valori maggiori nella configurazione con raggio di partenza pari a 50 m.

6.3.5 Esperimento 4

Per questo sono state prese in considerazione gli input sequence dataset ottenuti con le seguenti configurazioni: Raggio partenza = 200,100,50 m; Raggio di arrivo = 50 m; Distanza minima = 100 m; Periodo = mesi più attivi, stagioni.

Sono state prese in considerazione le sequenze estratte considerando i mesi più attivi(luglio, agosto e settembre) e le sequenze estratte al variare delle stagioni. In questo esperimento sono stati fatti dei confronti con le sequenze estratte con la configurazione di riferimento considerando gli stessi periodi.

Raggio di partenza 200 m - mesi più attivi

Guardando le sequenze estratte impostando come raggio di partenza 200 m (riportate nel dettaglio in Appendice A), alcune sequenze sono sempre presenti mentre altre sequenze sono presenti solo in alcuni mesi. Nella Tabella 6.22 sono riportate le sequenze presenti in tutti i mesi.

	Sequenze
0	School/College –> School/College
1	Bank –> Bank
2	Religious place –> Religious place
3	Hotel –> Post Office
4	Post Office –> Hotel
5	Theater –> Bank
6	Hotel –> Hotel
7	Food –> Post Office
8	Post Office –> Food
9	Bicycle related places –> Bicycle related places
10	Post Office –> Post Office
11	Bank –> Theater
12	Food –> Food
13	Entertainment –> Entertainment
14	Market –> Market
15	Theater –> Theater
16	Health –> Health
17	Diplomacy/Services –> Diplomacy/Services
18	Post Office –> Bank
19	Bank –> Post Office

Tabella 6.22: Sequenze uguali con raggio di partenza 200 m - mesi più attivi

Le seguenti sequenze sono presenti in alcuni mesi :

- *Hotel* → *Theater:* questa sequenza è presente nei mesi di luglio e agosto, questo evidenzia il fatto che questi mesi sono importanti dal punto di vista turistico.

- *Theater* → *Entertainment*: anche in questo caso questa sequenza è presente nel mese di luglio e agosto. Questo evidenzia maggiormente l'idea che questi periodi sono più attivi dal punto di vista turistico.

- *Hotel* → *Health*: questa sequenza è presente solo nel mese di settembre. Nella categoria Health ricadono quei luoghi legati alla cura della persona, quindi tale sequenza evidenzia il fatto che nel mese di settembre si sono verificati spostamenti verso questi luoghi.

- *School/college* → *Food*: questa sequenza è presente solo nel mese di settembre, questo è riconducibile alla riapertura delle scuole.

Confrontando tali sequenze con le sequenze ottenute con la configurazione di riferimento, si può notare una differenza nelle sequenza *Hotel* → *Theater*, infatti tale sequenza nella configurazione con raggio di partenza 200 m è presente solo nei mesi di luglio agosto mentre nella configurazione di riferimento tale sequenza è presente solo nel mese di settembre. Entrambe le sequenze *Theater* → *Entertainment* e *School/college* → *Food*, sono state estratte in entrambe le configurazioni negli stessi mesi, questo conferma la forza di tale sequenza. Nella configurazione di riferimento non è presente la sequenza *Hotel* → *Health*, quindi variando il raggio di partenza si ha una variazione delle informazioni estratte.

Raggio di partenza 50 m - mesi più attivi

Guardando le sequenze estratte impostando come raggio di partenza 50 m (riportate nel dettaglio in Appendice A), alcune sequenze sono sempre presenti mentre altre sequenze sono presenti solo in alcuni mesi. Nella Tabella 6.23 sono riportate le sequenze presenti in tutti i mesi, anche in questo caso è presente la sequenza *School/College→School/College* in tutti i mesi. Guardando la Tabella 6.23, 27 sequenze sono comuni a tutti i mesi presi in considerazione.

Le seguenti sequenze sono presenti in alcuni mesi :

- *Bank* → *Entertainment*: questa sequenza è presente nel mese di luglio e agosto. Tale sequenza non è presente nella configurazione di riferimento.

- *Bar/Cafe* → *Health*: questa sequenza è presente solo nel mese di settembre. Tale sequenza non è presente nella configurazione di riferimento.

Confrontando le sequenza uguali in tutti i mesi (Tabella 6.23) con le sequenze ottenute con la configurazione di riferimento (Tabella 6.13), si può notare un grande cambiamento infatti le due configurazioni hanno in comune le seguenti 14 sequenze:

Bank –> Bank, Bicycle related places –> Bicycle related places, Diplomacy/-Services –> Diplomacy/Services, Entertainment –> Entertainment, Food –> Food, Health –> Health, Hotel –> Hotel, Market –> Market, Post Office –> Post Office, Public Places –> Public Places, Religious place –> Religious place, School/College –> School/College, Store/Shop –> Store/Shop, Theater –> Theater

Inoltre nella la configurazione con raggio di partenza pari a 50 m, la sequenza *Theater –> Entertainment* è presente in tutti i mesi mentre nella configurazione di riferimento solo nel mese di settembre.

	Sequenze
0	Entertainment –> Diplomacy/Services
1	Bank –> Store/Shop
2	School/College –> School/College
3	Bank –> Bank
4	Religious place –> Religious place
5	Diplomacy/Services –> Entertainment
6	Food –> Bicycle related places
7	Public Places –> Public Places
8	Hotel –> Hotel
9	Store/Shop –> Store/Shop
10	Bicycle related places –> Bicycle related places
11	Store/Shop –> Bank
12	Theater –> Entertainment
13	Post Office –> Post Office
14	Food –> Food
15	Bicycle related places –> Food
16	Market –> Hotel
17	Entertainment –> Entertainment
18	Market –> Market
19	Theater –> Theater
20	Hotel –> Market
21	Bar/Cafe –> Bar/Cafe
22	Health –> Health
23	Diplomacy/Services –> Theater
24	Diplomacy/Services –> Diplomacy/Services
25	Entertainment –> Theater
26	Theater –> Diplomacy/Services

Tabella 6.23: Sequenze uguali con raggio di partenza 50 m - mesi più attivi

Raggio di partenza 200 m - Stagioni

Guardando le sequenze estratte impostando come raggio di partenza 200 m (riportate nel dettaglio in Appendice B), alcune sequenze sono sempre presenti mentre altre sequenze sono presenti solo in alcune stagioni. Nella Tabella 6.24 sono riportate le sequenze presenti in tutte le stagioni, anche in questo caso è presente la sequenza *School/College→School/College* in tutte le stagioni. Guardando la Tabella 6.24 si evince che 19 sequenze sono comuni a tutte le stagioni.

	Sequenze
0	School/College –> School/College
1	Bank –> Bank
2	Religious place –> Religious place
3	Theater –> Bank
4	Food –> Post Office
5	Hotel –> Hotel
6	Post Office –> Food
7	Bicycle related places –> Bicycle related places
8	Store/Shop –> Store/Shop
9	Post Office –> Post Office
10	Bank –> Theater
11	Food –> Food
12	Entertainment –> Entertainment
13	Market –> Market
14	Theater –> Theater
15	Health –> Health
16	Diplomacy/Services –> Diplomacy/Services
17	Post Office –> Bank
18	Bank –> Post Office

Tabella 6.24: Sequenze uguali con raggio di partenza 200 m - stagioni

Le seguenti sequenze sono presenti in alcune stagioni :

- *Hotel* → *Theater*: questa sequenza è presente in tutte le stagioni escluso in periodo autunnale. Questa sequenza evidenzia che tale spostamento è comune tutto l'anno tranne nel periodo autunnale. Tale sequenza non è presente nella configurazione di riferimento

- *Theater* → *Entertainment*: come nella configurazione di riferimento questa sequenza è presente solo in primavera ed estate.

- *Store/Shop* → *Hotel*: questa sequenza è presente solo nella stagione invernale. Questo sequenza evidenzia il fatto che in inverno questi mezzi vengono usati per l'acquisto di beni comuni. Tale sequenza non è presente nella configurazione di riferimento

- *Hotel* → *Health*: mentre nella configurazione di riferimento questa sequenza è presente solo nel periodo invernale, in questo caso questa sequenza è presente nelle stagioni di primavera ed autunno.

- *School/College* → *Food*: mentre nella configurazione di riferimento questa sequenza è presente in tutte le stagioni, in questo caso questa sequenza è presente solo nel periodo autunnale. Questa sequenza può essere ricondotta alla riapertura delle scuole.

- *Bar/Cafe* → *Bar/Cafe*: questa sequenza è presente solo nel periodo invernale, mentre nella configurazione di riferimento questa sequenza è presente in tute le stagioni tranne in estate.

- *Bank* → *Hotel*: questa sequenza è presente solo nel periodo invernale. Tale sequenza non è presente nella configurazione di riferimento

Raggio di partenza 50 m - Stagioni

Guardando le sequenze estratte impostando come raggio di partenza 50 m (riportate nel dettaglio in Appendice B), alcune sequenze sono sempre presenti mentre altre sequenze sono presenti solo in alcuni mesi. Nella Tabella 6.25 sono riportate le sequenze presenti in tutte le stagioni, anche in questo caso la sequenza *School/College→School/College* è presente in tutte le stagioni. Guardando la Tabella 6.24 si evince che 19 sequenze sono comuni a tutte le stagioni.

Le seguenti sequenze sono presenti in alcune stagioni :

- *Theater* → *Entertainment*: al contrario della configurazione di riferimento questa sequenza compare anche nel periodo autunnale.

- *School/College* → *Entertainment*: questa sequenza compare solo nel periodo primaverile (marzo-giugno). La presenza di questa sequenza può essere ricondotta al passaggio dal periodo invernale al periodo primaverile, ovvero ad un miglioramento delle condizioni climatiche. Tale sequenza non è presente nella configurazione di riferimento

- *Bicycle related place* → *Food*: questa sequenza compare solo nel periodo estivo. Tale sequenza non è presente nella configurazione di riferimento.

- *Bar/Cafe* → *Health*: questa sequenza compare solo nei periodi di inverno e autunno. Tale sequenza non è presente nella configurazione di riferimento

- *Bank* → *Market*: questa sequenza compare solo nel periodo invernale. Tale sequenza non è presente nella configurazione di riferimento

	Sequenze
0	Bank –> Store/Shop
1	School/College –> School/College
2	Bank –> Bank
3	Religious place –> Religious place
4	Public Places –> Public Places
5	Hotel –> Hotel
6	Bicycle related places –> Bicycle related places
7	Store/Shop –> Store/Shop
8	Store/Shop –> Bank
9	Post Office –> Post Office
10	Food –> Food
11	Entertainment –> Entertainment
12	Market –> Market
13	Theater –> Theater
14	Bar/Cafe –> Bar/Cafe
15	Health –> Health
16	Diplomacy/Services –> Theater
17	Diplomacy/Services –> Diplomacy/Services
18	Theater –> Diplomacy/Services

Tabella 6.25: Sequenze uguali con raggio di partenza 50 m - stagioni

Capitolo 7

Conclusione e Lavori Futuri

7.1 Conclusione

In questa tesi è stata introdotta una metodologia per l'analisi dei dati di micro-mobilità. Questa metodologia si basa su tre passi: acquisizione dati di mobilità, preparazione dei dati e data mining. La fase di preparazione è una fase cruciale di questa analisi, poichè in questa fase individuiamo le aree di maggiore interesse, i periodi più attivi, integriamo i luoghi di interesse e creiamo i dataset da usare per la fase di data mining. Nella Sezione 4.2.4 si è visto come i luoghi di interesse possono essere rappresentati in diversi modi su OSM. Nel nostro caso di studio sono stati presi in considerazione sia i luoghi di interesse geo-referenziati da un punto sia i luoghi di interesse identificati da un'area. Nella Sezione 4.2.6 sono state fatte diverse considerazioni per la generazione degli input sequence dataset, tra queste si trovano: la grandezza dell'intorno del punto di partenza, la grandezza dell'intorno del punto di arrivo, la distanza tra il punto di partenza e il punto di arrivo, e il periodo temporale. Dai risultati ottenuti dagli esperimenti effettuati si evince che tali parametri hanno una particolare importanza, infatti confrontando i risultati ottenuti al variare del raggio di partenza e del raggio di arrivo si è visto una variazione dell'informazione estratta. Tale variazione è dovuto ai luoghi di interesse che ricadono negli intorni del punto di partenza e di arrivo. Sebbene siano presenti delle variazioni al variare dei due parametri sopracitati, una parte dell'informazione rimane presente in tutte le configurazioni esaminate. Applicando inoltre delle segmentazioni temporali come i mesi più attivi o le stagioni è stato possibile estrarre pattern specifici per i periodi presi in esame. Come riportato nella Sezione 6.3.1, tramite la configurazione di riferimento è stato possibile estrarre dei pattern significativi per la micro-mobilità. In conclusione, la metodologia proposta è una metodologia molto flessibile che permette di considerare molti fattori con risultati interessanti.

Sarebbe possibile estendere questo lavoro di tesi in diversi modi. La metodologia proposta potrebbe essere applicata a diversi dataset, relativi a dati di mobilità, di diverse città per confrontare i pattern estratti e di conseguenza cercare le differenze relative alle abitudini degli utenti nelle diverse città.
Sarebbe anche possibile cambiare la fonte relativa ai luoghi di interesse. Nella Sezione 6.2.4, è stato introdotta una metodologia per integrare i luoghi di interesse tramite OSM, ma si potrebbero utilizzare altre fonti come per esempio *Foursquare* [5]. Tramite le API di Foursquare è possibile estrarre i luoghi di interesse in un intorno di un punto. I dati di Foursquare includono una tassonomia gerarchica di categorie che viene usata per classificare ogni record. Inoltre per ogni luogo di interesse è presente un punteggio, basato sui voti degli utenti, che potrebbe essere usato per l'individuazione dei luoghi più interessanti.

Appendice A

Mesi

In questa appendice sono riportate le sequenze ottenute esplorando diverse configurazioni e segmentando per mese.

A.1 Raggio di partenza = 200 m - Raggio di arrivo = 50 m - Distanza minima = 100 m

sequenze	rel. support	confidence	lift
Post Office –> Post Office	0.18	1.00	5.62
School/College –> School/College	0.20	1.00	5.06
Food –> Food	0.23	1.00	4.43
Theater –> Theater	0.23	1.00	4.41
Market –> Market	0.24	1.00	4.15
Bank –> Bank	0.33	1.00	2.99
Health –> Health	0.35	1.00	2.84
Food –> Post Office	0.11	0.48	2.68
Post Office –> Food	0.11	0.60	2.67
Religious place –> Religious place	0.41	1.00	2.42
Hotel –> Hotel	0.47	1.00	2.13
Theater –> Bank	0.16	0.70	2.10
Bank –> Theater	0.16	0.47	2.09
Entertainment –> Entertainment	0.49	1.00	2.06
Post Office –> Market	0.09	0.48	1.99
Market –> Post Office	0.09	0.35	1.99
Bank –> Post Office	0.12	0.35	1.96
Post Office –> Bank	0.12	0.65	1.95
Bicycle related places –> Bicycle related places	0.53	1.00	1.88
Diplomacy/Services –> Diplomacy/Services	0.60	1.00	1.67
Religious place –> Post Office	0.12	0.29	1.65
Post Office –> Religious place	0.12	0.68	1.64
Hotel –> Theater	0.17	0.37	1.64
Theater –> Hotel	0.17	0.76	1.61
Hotel –> Post Office	0.13	0.29	1.61
Theater –> Entertainment	0.18	0.78	1.60
Entertainment –> Theater	0.18	0.36	1.59
Religious place –> Food	0.15	0.36	1.59
Post Office –> Hotel	0.13	0.74	1.57
Food –> Religious place	0.15	0.65	1.57
Store/Shop –> Store/Shop	0.65	1.00	1.53

Tabella A.1: Raggio di partenza = 200 m - Raggio di arrivo = 50 m - Distanza minima = 100 m - Luglio

sequenze	rel. support	confidence	lift
Post Office –> Post Office	0.17	1.00	5.78
Food –> Food	0.23	1.00	4.35
Market –> Market	0.24	1.00	4.23
Theater –> Theater	0.24	1.00	4.22
School/College –> School/College	0.25	1.00	3.92
Health –> Health	0.33	1.00	3.06
Bank –> Bank	0.33	1.00	3.00
Post Office –> Food	0.11	0.61	2.67
Food –> Post Office	0.11	0.46	2.67
Religious place –> Religious place	0.39	1.00	2.54
Hotel –> Hotel	0.43	1.00	2.32
Theater –> Bank	0.16	0.68	2.04
Bank –> Theater	0.16	0.48	2.03
Entertainment –> Entertainment	0.50	1.00	2.01
Post Office –> Bank	0.11	0.64	1.93
Bank –> Post Office	0.11	0.33	1.93
Bicycle related places –> Bicycle related places	0.56	1.00	1.78
Post Office –> Market	0.07	0.42	1.78
Market –> Post Office	0.07	0.31	1.77
Diplomacy/Services –> Diplomacy/Services	0.58	1.00	1.71
Hotel –> Post Office	0.12	0.27	1.58
Store/Shop –> Store/Shop	0.64	1.00	1.56
Hotel –> Theater	0.16	0.37	1.56
Theater –> Entertainment	0.18	0.77	1.55
Post Office –> Hotel	0.12	0.67	1.55
Entertainment –> Theater	0.18	0.37	1.55
Theater –> Hotel	0.16	0.66	1.54
Post Office –> Religious place	0.10	0.60	1.53
Religious place –> Post Office	0.10	0.26	1.53
Market –> Religious place	0.14	0.59	1.50
Religious place –> Market	0.14	0.36	1.50

Tabella A.2: Raggio di partenza = 200 m - Raggio di arrivo = 50 m - Distanza minima = 100 m - Agosto

sequenze	rel. support	confidence	lift
Post Office –> Post Office	0.17	1.00	5.72
Theater –> Theater	0.25	1.00	4.02
Health –> Health	0.25	1.00	3.94
Market –> Market	0.26	1.00	3.78
Food –> Food	0.27	1.00	3.76
Bank –> Bank	0.32	1.00	3.11
Hotel –> Hotel	0.34	1.00	2.94
Religious place –> Religious place	0.34	1.00	2.91
School/College –> School/College	0.35	1.00	2.82
Post Office –> Food	0.12	0.67	2.53
Food –> Post Office	0.12	0.44	2.53
Theater –> Bank	0.16	0.65	2.04
Bank –> Theater	0.16	0.51	2.04
Post Office –> Bank	0.11	0.64	1.98
Bank –> Post Office	0.11	0.35	1.98
Entertainment –> Entertainment	0.52	1.00	1.93
Hotel –> Health	0.15	0.46	1.79
Health –> Hotel	0.15	0.60	1.76
Diplomacy/Services –> Diplomacy/Services	0.57	1.00	1.75
Bicycle related places –> Bicycle related places	0.59	1.00	1.68
Store/Shop –> Store/Shop	0.59	1.00	1.68
Hotel –> Post Office	0.10	0.28	1.63
Post Office –> Hotel	0.10	0.55	1.61
Market –> Bank	0.13	0.49	1.52
Bank –> Market	0.13	0.40	1.52
Hotel –> Bank	0.16	0.48	1.51
Bar/Cafe –> Bar/Cafe	0.66	1.00	1.50
Food –> School/College	0.14	0.53	1.50
School/College –> Food	0.14	0.40	1.50
Religious place –> Health	0.13	0.38	1.48
Religious place –> Post Office	0.09	0.26	1.48

Tabella A.3: Raggio di partenza = 200 m - Raggio di arrivo = 50 m - Distanza minima = 100 m - Settembre

A.2 Raggio di partenza = 100 m - Raggio di arrivo = 50 m - Distanza minima = 100 m

sequenze	rel. support	confidence	lift
Theater –> Theater	0.08	1.00	12.82
Market –> Market	0.08	1.00	11.92
School/College –> School/College	0.09	1.00	11.54
Post Office –> Post Office	0.09	1.00	10.89
Food –> Food	0.12	1.00	8.56
Religious place –> Religious place	0.15	1.00	6.78
Health –> Health	0.19	1.00	5.34
Bank –> Bank	0.21	1.00	4.83
Food –> Post Office	0.05	0.44	4.74
Post Office –> Food	0.05	0.55	4.72
Entertainment –> Entertainment	0.23	1.00	4.42
Hotel –> Hotel	0.31	1.00	3.20
Bank –> Post Office	0.06	0.29	3.14
Post Office –> Bank	0.06	0.64	3.08
Bicycle related places –> Bicycle related places	0.33	1.00	3.04
Diplomacy/Services –> Diplomacy/Services	0.40	1.00	2.49
Food –> Bank	0.06	0.48	2.33
Bank –> Food	0.06	0.27	2.32
Store/Shop –> Store/Shop	0.45	1.00	2.23
Public Places –> Public Places	0.46	1.00	2.16
Theater –> Entertainment	0.04	0.48	2.14
Entertainment –> Theater	0.04	0.16	2.11
Hotel –> Post Office	0.06	0.18	1.95
Diplomacy/Services –> Theater	0.06	0.15	1.94
Theater –> Diplomacy/Services	0.06	0.77	1.93
Post Office –> Hotel	0.05	0.59	1.90
Food –> Religious place	0.03	0.28	1.87
Religious place –> Food	0.03	0.21	1.82
Store/Shop –> Food	0.09	0.20	1.74
Food –> Store/Shop	0.09	0.76	1.71
Hotel –> Theater	0.04	0.13	1.70

Tabella A.4: Raggio di partenza = 100 m - Raggio di arrivo = 50 m - Distanza minima = 100 m - Luglio

sequenze	rel. support	confidence	lift
Post Office –> Post Office	0.08	1.00	12.29
Theater –> Theater	0.08	1.00	12.27
Market –> Market	0.08	1.00	11.79
Food –> Food	0.11	1.00	9.03
School/College –> School/College	0.14	1.00	7.32
Religious place –> Religious place	0.14	1.00	7.02
Health –> Health	0.17	1.00	5.78
Bank –> Bank	0.19	1.00	5.35
Post Office –> Food	0.04	0.51	4.57
Food –> Post Office	0.04	0.37	4.57
Entertainment –> Entertainment	0.24	1.00	4.19
Hotel –> Hotel	0.28	1.00	3.52
Bank –> Post Office	0.05	0.26	3.24
Post Office –> Bank	0.05	0.60	3.20
Bicycle related places –> Bicycle related places	0.37	1.00	2.73
Diplomacy/Services –> Diplomacy/Services	0.38	1.00	2.60
Store/Shop –> Store/Shop	0.42	1.00	2.37
Food –> Bank	0.05	0.42	2.25
Bank –> Food	0.05	0.25	2.25
Public Places –> Public Places	0.46	1.00	2.17
Hotel –> Post Office	0.05	0.17	2.04
Post Office –> Hotel	0.05	0.56	1.98
Food –> Religious place	0.03	0.27	1.93
Religious place –> Food	0.03	0.21	1.93
Theater –> Diplomacy/Services	0.06	0.70	1.83
Diplomacy/Services –> Theater	0.06	0.15	1.83
Theater –> Entertainment	0.03	0.43	1.78
Food –> School/College	0.03	0.24	1.78
Entertainment –> Theater	0.03	0.14	1.78
School/College –> Food	0.03	0.19	1.75
Store/Shop –> Food	0.08	0.19	1.74

Tabella A.5: Raggio di partenza = 100 m - Raggio di arrivo = 50 m - Distanza minima = 100 m - Agosto

sequenze	rel. support	confidence	lift
Post Office –> Post Office	0.08	1.00	12.84
Theater –> Theater	0.08	1.00	11.78
Market –> Market	0.10	1.00	10.10
Religious place –> Religious place	0.11	1.00	8.85
Food –> Food	0.13	1.00	7.88
Health –> Health	0.14	1.00	7.35
Bank –> Bank	0.16	1.00	6.11
Hotel –> Hotel	0.23	1.00	4.39
School/College –> School/College	0.23	1.00	4.33
Post Office –> Food	0.04	0.50	3.98
Food –> Post Office	0.04	0.31	3.97
Entertainment –> Entertainment	0.26	1.00	3.85
Bank –> Post Office	0.04	0.27	3.52
Post Office –> Bank	0.04	0.57	3.50
Diplomacy/Services –> Diplomacy/Services	0.36	1.00	2.78
Store/Shop –> Store/Shop	0.36	1.00	2.74
Bicycle related places –> Bicycle related places	0.41	1.00	2.44
Public Places –> Public Places	0.45	1.00	2.24
Food –> Bank	0.05	0.36	2.22
Bank –> Food	0.05	0.28	2.22
Hotel –> Post Office	0.04	0.17	2.18
Post Office –> Hotel	0.04	0.48	2.12
Food –> School/College	0.06	0.45	1.94
School/College –> Food	0.06	0.25	1.94
Bar/Cafe –> Bar/Cafe	0.52	1.00	1.91
Bank –> Store/Shop	0.11	0.67	1.83
Hotel –> Theater	0.04	0.16	1.83
Store/Shop –> Bank	0.11	0.30	1.81
Store/Shop –> Health	0.09	0.25	1.81
Health –> Store/Shop	0.09	0.63	1.74
Theater –> Diplomacy/Services	0.05	0.62	1.72

Tabella A.6: Raggio di partenza = 100 m - Raggio di arrivo = 50 m - Distanza minima = 100 m - Settembre

A.3 Raggio di partenza = 50 m - Raggio di arrivo = 50 m - Distanza minima 100 m

sequenze	rel. support	confidence	lift
Food –> Food	0.04	1.00	27.95
Post Office –> Post Office	0.04	1.00	27.32
Theater –> Theater	0.04	1.00	23.28
School/College –> School/College	0.05	1.00	22.02
Market –> Market	0.05	1.00	21.99
Religious place –> Religious place	0.09	1.00	10.88
Health –> Health	0.12	1.00	8.65
Entertainment –> Entertainment	0.12	1.00	8.20
Bank –> Bank	0.13	1.00	7.89
Hotel –> Hotel	0.18	1.00	5.71
Bicycle related places –> Bicycle related places	0.21	1.00	4.69
Theater –> Entertainment	0.02	0.51	4.15
Entertainment –> Theater	0.02	0.17	4.05
Diplomacy/Services –> Diplomacy/Services	0.25	1.00	4.02
Store/Shop –> Store/Shop	0.31	1.00	3.24
Public Places –> Public Places	0.32	1.00	3.08
Market –> Hotel	0.02	0.49	2.80
Hotel –> Market	0.02	0.13	2.77
Theater –> Diplomacy/Services	0.03	0.68	2.75
Diplomacy/Services –> Theater	0.03	0.12	2.70
Store/Shop –> Bank	0.10	0.33	2.60
Bank –> Store/Shop	0.10	0.80	2.59
Bicycle related places –> Food	0.02	0.09	2.49
Food –> Bicycle related places	0.02	0.52	2.46
Entertainment –> Bank	0.04	0.30	2.39
Bank –> Entertainment	0.04	0.29	2.39
Diplomacy/Services –> Entertainment	0.07	0.28	2.33
Entertainment –> Diplomacy/Services	0.07	0.58	2.32
Bar/Cafe –> Bar/Cafe	0.48	1.00	2.08
Public transport –> Public transport	0.54	1.00	1.84
Public transport –> Bank	0.12	0.21	1.70

Tabella A.7: Raggio di partenza = 50 m - Raggio di arrivo = 50 m - Distanza minima 100 m - Luglio

sequenze	rel. support	confidence	lift
Post Office –> Post Office	0.03	1.00	38.08
Food –> Food	0.03	1.00	31.86
Theater –> Theater	0.04	1.00	23.42
Market –> Market	0.05	1.00	21.48
School/College –> School/College	0.08	1.00	12.69
Religious place –> Religious place	0.08	1.00	11.83
Bank –> Bank	0.11	1.00	9.18
Health –> Health	0.11	1.00	8.85
Entertainment –> Entertainment	0.13	1.00	7.90
Hotel –> Hotel	0.16	1.00	6.34
Diplomacy/Services –> Diplomacy/Services	0.23	1.00	4.29
Bicycle related places –> Bicycle related places	0.25	1.00	4.01
Store/Shop –> Store/Shop	0.28	1.00	3.54
Theater –> Entertainment	0.02	0.42	3.31
Entertainment –> Theater	0.02	0.14	3.29
Public Places –> Public Places	0.33	1.00	3.05
Bank –> Store/Shop	0.08	0.78	2.75
Store/Shop –> Bank	0.08	0.30	2.74
Hotel –> Market	0.02	0.12	2.65
Theater –> Diplomacy/Services	0.03	0.61	2.63
Market –> Hotel	0.02	0.41	2.63
Diplomacy/Services –> Theater	0.03	0.11	2.55
Entertainment –> Bank	0.03	0.27	2.49
Bank –> Entertainment	0.03	0.31	2.48
Entertainment –> Diplomacy/Services	0.07	0.56	2.40
Diplomacy/Services –> Entertainment	0.07	0.30	2.39
Bar/Cafe –> Bar/Cafe	0.45	1.00	2.22
Bicycle related places –> Food	0.02	0.06	1.93
Food –> Bicycle related places	0.01	0.47	1.90
Public transport –> Public transport	0.53	1.00	1.88
Store/Shop –> Entertainment	0.06	0.22	1.77

Tabella A.8: Raggio di partenza = 50 m - Raggio di arrivo = 50 m - Distanza minima 100 m - Agosto

sequenze	rel. support	confidence	lift
Post Office –> Post Office	0.02	1.00	41.56
Food –> Food	0.04	1.00	24.82
Theater –> Theater	0.05	1.00	20.38
Market –> Market	0.06	1.00	15.82
Religious place –> Religious place	0.07	1.00	14.79
Bank –> Bank	0.09	1.00	10.75
Health –> Health	0.09	1.00	10.74
Hotel –> Hotel	0.13	1.00	7.52
Entertainment –> Entertainment	0.14	1.00	7.05
School/College –> School/College	0.15	1.00	6.83
Diplomacy/Services –> Diplomacy/Services	0.22	1.00	4.57
Store/Shop –> Store/Shop	0.24	1.00	4.19
Bicycle related places –> Bicycle related places	0.28	1.00	3.63
Bank –> Store/Shop	0.07	0.80	3.34
Store/Shop –> Bank	0.07	0.31	3.33
Public Places –> Public Places	0.32	1.00	3.11
Theater –> Entertainment	0.02	0.41	2.91
Entertainment –> Theater	0.02	0.14	2.90
Bar/Cafe –> Bar/Cafe	0.38	1.00	2.63
Theater –> Diplomacy/Services	0.03	0.55	2.51
Diplomacy/Services –> Theater	0.03	0.12	2.46
Hotel –> Market	0.02	0.15	2.43
Diplomacy/Services –> Entertainment	0.07	0.34	2.42
Market –> Hotel	0.02	0.32	2.41
Entertainment –> Diplomacy/Services	0.07	0.52	2.40
Public transport –> Public transport	0.47	1.00	2.10
Bicycle related places –> Food	0.02	0.08	2.07
Food –> Bicycle related places	0.02	0.57	2.06
Health –> Bar/Cafe	0.07	0.78	2.05
Bar/Cafe –> Health	0.07	0.19	2.05
Bank –> Entertainment	0.03	0.28	2.01

Tabella A.9: Raggio di partenza = 50 m - Raggio di arrivo = 50 m - Distanza minima 100 m - Settembre

Appendice B

Stagioni

In questa appendice sono riportate le sequenze ottenute esplorando diverse configurazioni e segmentando per stagione.

B.1 Raggio di partenza = 200 m - Raggio di arrivo = 50 m - Distanza minima = 100 m

sequenze	rel. support	confidence	lift
Post Office –> Post Office	0.15	1.00	6.47
Food –> Food	0.21	1.00	4.67
Theater –> Theater	0.23	1.00	4.44
Market –> Market	0.23	1.00	4.37
School/College –> School/College	0.26	1.00	3.87
Health –> Health	0.28	1.00	3.54
Bank –> Bank	0.33	1.00	3.05
Post Office –> Food	0.09	0.60	2.79
Food –> Post Office	0.09	0.43	2.79
Hotel –> Hotel	0.38	1.00	2.62
Religious place –> Religious place	0.39	1.00	2.58
Entertainment –> Entertainment	0.46	1.00	2.17
Theater –> Bank	0.15	0.69	2.09
Bank –> Theater	0.15	0.47	2.09
Bank –> Post Office	0.10	0.30	1.92
Post Office –> Bank	0.10	0.63	1.91
Bicycle related places –> Bicycle related places	0.53	1.00	1.90
Diplomacy/Services –> Diplomacy/Services	0.56	1.00	1.77
Hotel –> Theater	0.15	0.39	1.73
Post Office –> Market	0.06	0.39	1.72
Theater –> Hotel	0.15	0.65	1.71
Market –> Post Office	0.06	0.26	1.71
Store/Shop –> Store/Shop	0.59	1.00	1.68
Hotel –> Post Office	0.10	0.25	1.65
Post Office –> Hotel	0.09	0.61	1.61
Theater –> Entertainment	0.16	0.72	1.56
Hotel –> Health	0.17	0.44	1.56
Entertainment –> Theater	0.16	0.35	1.56
Hotel –> Bank	0.19	0.51	1.55
Religious place –> Market	0.14	0.36	1.55
Market –> Religious place	0.14	0.60	1.54

Tabella B.1: Raggio di partenza = 200 m - Raggio di arrivo = 50 m - Distanza minima = 100 m - Primavera

sequenze	rel. support	confidence	lift
Post Office –> Post Office	0.18	1.00	5.69
Theater –> Theater	0.24	1.00	4.25
Food –> Food	0.24	1.00	4.24
Market –> Market	0.25	1.00	4.06
School/College –> School/College	0.25	1.00	3.93
Health –> Health	0.32	1.00	3.14
Bank –> Bank	0.33	1.00	3.02
Post Office –> Food	0.11	0.63	2.65
Food –> Post Office	0.11	0.47	2.65
Religious place –> Religious place	0.39	1.00	2.57
Hotel –> Hotel	0.42	1.00	2.35
Theater –> Bank	0.16	0.68	2.05
Bank –> Theater	0.16	0.48	2.04
Entertainment –> Entertainment	0.50	1.00	2.02
Post Office –> Bank	0.11	0.65	1.96
Bank –> Post Office	0.11	0.34	1.96
Bicycle related places –> Bicycle related places	0.55	1.00	1.80
Post Office –> Market	0.08	0.43	1.76
Market –> Post Office	0.08	0.31	1.75
Diplomacy/Services –> Diplomacy/Services	0.59	1.00	1.70
Hotel –> Post Office	0.12	0.28	1.61
Religious place –> Post Office	0.11	0.28	1.59
Post Office –> Hotel	0.12	0.67	1.58
Post Office –> Religious place	0.11	0.61	1.58
Store/Shop –> Store/Shop	0.64	1.00	1.57
Hotel –> Theater	0.16	0.37	1.57
Theater –> Entertainment	0.18	0.77	1.56
Entertainment –> Theater	0.18	0.37	1.56
Theater –> Hotel	0.15	0.66	1.54
Religious place –> Market	0.14	0.37	1.50
Market –> Religious place	0.14	0.58	1.49

Tabella B.2: Raggio di partenza = 200 m - Raggio di arrivo = 50 m - Distanza minima = 100 m - Estate

sequenze	rel. support	confidence	lift
Post Office –> Post Office	0.19	1.00	5.39
Health –> Health	0.23	1.00	4.31
Theater –> Theater	0.27	1.00	3.73
Market –> Market	0.27	1.00	3.68
Food –> Food	0.28	1.00	3.54
Hotel –> Hotel	0.33	1.00	3.05
Religious place –> Religious place	0.33	1.00	3.03
Bank –> Bank	0.34	1.00	2.94
School/College –> School/College	0.37	1.00	2.69
Post Office –> Food	0.12	0.66	2.34
Food –> Post Office	0.12	0.43	2.34
Theater –> Bank	0.18	0.67	1.97
Bank –> Theater	0.18	0.53	1.97
Hotel –> Health	0.14	0.44	1.89
Entertainment –> Entertainment	0.54	1.00	1.85
Post Office –> Bank	0.12	0.63	1.84
Bank –> Post Office	0.12	0.34	1.84
Health –> Hotel	0.14	0.60	1.83
Store/Shop –> Store/Shop	0.59	1.00	1.71
Diplomacy/Services –> Diplomacy/Services	0.58	1.00	1.71
Bicycle related places –> Bicycle related places	0.61	1.00	1.63
Religious place –> Market	0.14	0.43	1.57
Market –> Religious place	0.14	0.52	1.56
Hotel –> Post Office	0.10	0.29	1.56
Food –> School/College	0.16	0.58	1.55
School/College –> Food	0.16	0.44	1.55
Religious place –> Health	0.12	0.36	1.55
Post Office –> Hotel	0.09	0.50	1.54
Health –> Religious place	0.12	0.50	1.52
Hotel –> Theater	0.13	0.41	1.51
Bar/Cafe –> Bar/Cafe	0.67	1.00	1.50

Tabella B.3: Raggio di partenza = 200 m - Raggio di arrivo = 50 m - Distanza minima = 100 m - Autunno

sequenze	rel. support	confidence	lift
Post Office –> Post Office	0.17	1.00	5.89
Health –> Health	0.17	1.00	5.79
Theater –> Theater	0.24	1.00	4.16
Hotel –> Hotel	0.24	1.00	4.16
Food –> Food	0.26	1.00	3.86
Market –> Market	0.27	1.00	3.72
Religious place –> Religious place	0.32	1.00	3.11
Bank –> Bank	0.34	1.00	2.96
Hotel –> Health	0.10	0.40	2.33
School/College –> School/College	0.44	1.00	2.30
Post Office –> Food	0.10	0.58	2.26
Health –> Hotel	0.09	0.54	2.26
Food –> Post Office	0.10	0.38	2.25
Entertainment –> Entertainment	0.48	1.00	2.08
Theater –> Bank	0.17	0.69	2.05
Bank –> Theater	0.17	0.49	2.05
Bank –> Post Office	0.11	0.32	1.86
Post Office –> Bank	0.11	0.62	1.85
Store/Shop –> Store/Shop	0.54	1.00	1.84
Diplomacy/Services –> Diplomacy/Services	0.55	1.00	1.81
Hotel –> Theater	0.10	0.43	1.79
Theater –> Hotel	0.10	0.43	1.77
Bicycle related places –> Bicycle related places	0.59	1.00	1.68
Hotel –> Bank	0.13	0.56	1.65
Bank –> Hotel	0.13	0.39	1.63
Bar/Cafe –> Bar/Cafe	0.62	1.00	1.61
Hotel –> Store/Shop	0.20	0.84	1.55
Store/Shop –> Hotel	0.20	0.37	1.55
Religious place –> Health	0.09	0.27	1.55
Hotel –> Post Office	0.06	0.26	1.53
Bank –> Market	0.14	0.41	1.52

Tabella B.4: Raggio di partenza = 200 m - Raggio di arrivo = 50 m - Distanza minima = 100 m - Inverno

B.2 Raggio di partenza = 100 m - Raggio di arrivo = 50 m - Distanza minima = 100 m

sequenze	rel. support	confidence	lift
Theater –> Theater	0.07	1.00	14.45
Post Office –> Post Office	0.07	1.00	13.40
Market –> Market	0.08	1.00	12.34
Food –> Food	0.11	1.00	8.85
Religious place –> Religious place	0.14	1.00	7.36
School/College –> School/College	0.14	1.00	7.31
Health –> Health	0.15	1.00	6.61
Bank –> Bank	0.18	1.00	5.46
Post Office –> Food	0.04	0.52	4.63
Food –> Post Office	0.04	0.34	4.61
Entertainment –> Entertainment	0.23	1.00	4.38
Hotel –> Hotel	0.25	1.00	4.06
Bank –> Post Office	0.05	0.25	3.29
Post Office –> Bank	0.04	0.59	3.24
Bicycle related places –> Bicycle related places	0.35	1.00	2.87
Diplomacy/Services –> Diplomacy/Services	0.37	1.00	2.69
Store/Shop –> Store/Shop	0.40	1.00	2.53
Public Places –> Public Places	0.45	1.00	2.21
Food –> Bank	0.05	0.40	2.19
Bank –> Food	0.05	0.25	2.18
Hotel –> Post Office	0.04	0.16	2.17
Food –> School/College	0.03	0.29	2.14
School/College –> Food	0.03	0.24	2.13
Post Office –> Hotel	0.04	0.52	2.11
Diplomacy/Services –> Theater	0.05	0.13	1.84
Entertainment –> Theater	0.03	0.13	1.84
Theater –> Diplomacy/Services	0.05	0.68	1.83
Bar/Cafe –> Bar/Cafe	0.55	1.00	1.81
Theater –> Entertainment	0.03	0.41	1.81
Store/Shop –> Food	0.08	0.20	1.79
Food –> Store/Shop	0.08	0.70	1.77

Tabella B.5: Raggio di partenza = 100 m - Raggio di arrivo = 50 m - Distanza minima = 100 m - Primavera

sequenze	rel. support	confidence	lift
Theater –> Theater	0.08	1.00	12.31
Post Office –> Post Office	0.09	1.00	11.74
Market –> Market	0.09	1.00	11.31
Food –> Food	0.12	1.00	8.55
Religious place –> Religious place	0.14	1.00	7.29
School/College –> School/College	0.14	1.00	7.26
Health –> Health	0.17	1.00	5.93
Bank –> Bank	0.19	1.00	5.27
Food –> Post Office	0.05	0.39	4.56
Post Office –> Food	0.05	0.53	4.55
Entertainment –> Entertainment	0.24	1.00	4.19
Hotel –> Hotel	0.28	1.00	3.52
Bank –> Post Office	0.05	0.28	3.29
Post Office –> Bank	0.05	0.62	3.25
Bicycle related places –> Bicycle related places	0.36	1.00	2.78
Diplomacy/Services –> Diplomacy/Services	0.38	1.00	2.60
Store/Shop –> Store/Shop	0.42	1.00	2.39
Food –> Bank	0.05	0.44	2.32
Bank –> Food	0.05	0.27	2.31
Public Places –> Public Places	0.46	1.00	2.19
Hotel –> Post Office	0.05	0.17	2.05
Post Office –> Hotel	0.05	0.57	1.99
Diplomacy/Services –> Theater	0.06	0.15	1.86
Theater –> Diplomacy/Services	0.06	0.71	1.85
Theater –> Entertainment	0.04	0.44	1.85
Entertainment –> Theater	0.04	0.15	1.83
Food –> Religious place	0.03	0.24	1.79
Food –> School/College	0.03	0.24	1.76
School/College –> Food	0.03	0.20	1.75
Religious place –> Food	0.03	0.20	1.75
Store/Shop –> Food	0.08	0.20	1.73

Tabella B.6: Raggio di partenza = 100 m - Raggio di arrivo = 50 m - Distanza minima = 100 m - Estate

sequenze	rel. support	confidence	lift
Post Office –> Post Office	0.08	1.00	12.30
Theater –> Theater	0.09	1.00	11.31
Market –> Market	0.10	1.00	9.88
Religious place –> Religious place	0.10	1.00	9.83
Health –> Health	0.12	1.00	8.08
Food –> Food	0.14	1.00	7.37
Bank –> Bank	0.17	1.00	5.94
Hotel –> Hotel	0.22	1.00	4.61
School/College –> School/College	0.25	1.00	4.05
Post Office –> Food	0.04	0.49	3.65
Food –> Post Office	0.04	0.30	3.64
Entertainment –> Entertainment	0.28	1.00	3.59
Bank –> Post Office	0.05	0.27	3.33
Post Office –> Bank	0.05	0.56	3.30
Store/Shop –> Store/Shop	0.35	1.00	2.87
Diplomacy/Services –> Diplomacy/Services	0.37	1.00	2.73
Bicycle related places –> Bicycle related places	0.43	1.00	2.34
Public Places –> Public Places	0.46	1.00	2.17
Food –> School/College	0.07	0.51	2.05
School/College –> Food	0.07	0.28	2.05
Food –> Bank	0.05	0.34	2.04
Bank –> Food	0.05	0.28	2.04
Hotel –> Post Office	0.04	0.16	2.03
Post Office –> Hotel	0.04	0.43	1.99
Bar/Cafe –> Bar/Cafe	0.51	1.00	1.95
Bank –> Store/Shop	0.11	0.67	1.91
Store/Shop –> Bank	0.11	0.32	1.89
Store/Shop –> Health	0.08	0.23	1.87
Bank –> Theater	0.03	0.16	1.84
Theater –> Bank	0.03	0.31	1.83
Hotel –> Theater	0.03	0.16	1.79

Tabella B.7: Raggio di partenza = 100 m - Raggio di arrivo = 50 m - Distanza minima = 100 m - Autunno

sequenze	rel. support	confidence	lift
Post Office –> Post Office	0.07	1.00	13.44
Theater –> Theater	0.07	1.00	13.42
Health –> Health	0.10	1.00	10.48
Religious place –> Religious place	0.10	1.00	9.57
Market –> Market	0.11	1.00	9.08
Food –> Food	0.13	1.00	7.74
Bank –> Bank	0.15	1.00	6.47
Hotel –> Hotel	0.16	1.00	6.20
Entertainment –> Entertainment	0.26	1.00	3.92
School/College –> School/College	0.27	1.00	3.64
Store/Shop –> Store/Shop	0.29	1.00	3.40
Bank –> Post Office	0.04	0.25	3.35
Post Office –> Bank	0.04	0.50	3.26
Post Office –> Food	0.03	0.40	3.11
Food –> Post Office	0.03	0.23	3.11
Diplomacy/Services –> Diplomacy/Services	0.33	1.00	3.06
Bicycle related places –> Bicycle related places	0.42	1.00	2.41
Bank –> Theater	0.03	0.17	2.33
Theater –> Bank	0.03	0.35	2.28
Public Places –> Public Places	0.45	1.00	2.24
Theater –> Post Office	0.01	0.16	2.19
Post Office –> Theater	0.01	0.16	2.17
Bar/Cafe –> Bar/Cafe	0.47	1.00	2.14
School/College –> Food	0.07	0.27	2.11
Food –> School/College	0.07	0.57	2.09
Store/Shop –> Health	0.06	0.20	2.06
Hotel –> Post Office	0.02	0.15	2.02
Bank –> Store/Shop	0.09	0.59	2.01
Hotel –> Health	0.03	0.19	2.00
Store/Shop –> Bank	0.09	0.31	1.98
Bank –> Hotel	0.05	0.32	1.97

Tabella B.8: Raggio di partenza = 100 m - Raggio di arrivo = 50 m - Distanza minima = 100 m - Inverno

B.3 Raggio di partenza = 50 m - Raggio di arrivo = 50 m - Distanza minima = 100 m

sequenze	rel. support	confidence	lift
Post Office –> Post Office	0.03	1.00	35.38
Theater –> Theater	0.04	1.00	26.96
Food –> Food	0.04	1.00	26.78
Market –> Market	0.04	1.00	23.72
School/College –> School/College	0.08	1.00	12.75
Religious place –> Religious place	0.08	1.00	12.40
Health –> Health	0.10	1.00	9.66
Bank –> Bank	0.11	1.00	9.50
Entertainment –> Entertainment	0.12	1.00	8.27
Hotel –> Hotel	0.14	1.00	7.03
Diplomacy/Services –> Diplomacy/Services	0.23	1.00	4.33
Bicycle related places –> Bicycle related places	0.25	1.00	4.05
Store/Shop –> Store/Shop	0.27	1.00	3.74
Public Places –> Public Places	0.33	1.00	3.06
Entertainment –> Theater	0.01	0.11	2.99
Theater –> Entertainment	0.01	0.36	2.96
Bank –> Store/Shop	0.08	0.77	2.87
Store/Shop –> Bank	0.08	0.30	2.86
Theater –> Diplomacy/Services	0.02	0.59	2.55
Diplomacy/Services –> Theater	0.02	0.09	2.53
Bank –> Entertainment	0.03	0.30	2.49
Entertainment –> Bank	0.03	0.26	2.45
Market –> Hotel	0.01	0.35	2.44
Hotel –> Market	0.01	0.10	2.44
Bar/Cafe –> Bar/Cafe	0.41	1.00	2.41
Diplomacy/Services –> Entertainment	0.07	0.29	2.38
Entertainment –> Diplomacy/Services	0.07	0.55	2.37
Entertainment –> School/College	0.02	0.16	2.00
School/College –> Entertainment	0.02	0.24	1.99
Public transport –> Public transport	0.51	1.00	1.97
Food –> Bicycle related places	0.02	0.49	1.97

Tabella B.9: Raggio di partenza = 50 m - Raggio di arrivo = 50 m - Distanza minima = 100 m - Primavera

sequenze	rel. support	confidence	lift
Post Office –> Post Office	0.03	1.00	33.24
Food –> Food	0.03	1.00	28.65
Theater –> Theater	0.04	1.00	22.44
Market –> Market	0.05	1.00	19.77
School/College –> School/College	0.08	1.00	12.45
Religious place –> Religious place	0.08	1.00	11.95
Health –> Health	0.11	1.00	9.22
Bank –> Bank	0.11	1.00	8.74
Entertainment –> Entertainment	0.13	1.00	7.80
Hotel –> Hotel	0.16	1.00	6.16
Diplomacy/Services –> Diplomacy/Services	0.23	1.00	4.27
Bicycle related places –> Bicycle related places	0.24	1.00	4.20
Theater –> Entertainment	0.02	0.47	3.65
Entertainment –> Theater	0.02	0.16	3.61
Store/Shop –> Store/Shop	0.28	1.00	3.53
Public Places –> Public Places	0.32	1.00	3.09
Bank –> Store/Shop	0.09	0.79	2.80
Store/Shop –> Bank	0.09	0.32	2.80
Theater –> Diplomacy/Services	0.03	0.63	2.70
Diplomacy/Services –> Theater	0.03	0.12	2.64
Market –> Hotel	0.02	0.43	2.63
Hotel –> Market	0.02	0.13	2.63
Entertainment –> Diplomacy/Services	0.07	0.56	2.40
Diplomacy/Services –> Entertainment	0.07	0.31	2.40
Bank –> Entertainment	0.03	0.30	2.31
Entertainment –> Bank	0.03	0.26	2.30
Bar/Cafe –> Bar/Cafe	0.45	1.00	2.24
Bicycle related places –> Food	0.02	0.08	2.16
Food –> Bicycle related places	0.02	0.51	2.15
Public transport –> Public transport	0.52	1.00	1.91
Public transport –> Bank	0.10	0.20	1.74

Tabella B.10: Raggio di partenza = 50 m - Raggio di arrivo = 50 m - Distanza minima = 100 m - Estate

sequenze	rel. support	confidence	lift
Post Office –> Post Office	0.03	1.00	36.93
Food –> Food	0.05	1.00	21.96
Theater –> Theater	0.05	1.00	19.66
Religious place –> Religious place	0.06	1.00	16.42
Market –> Market	0.07	1.00	15.34
Health –> Health	0.09	1.00	11.38
Bank –> Bank	0.09	1.00	11.10
Hotel –> Hotel	0.13	1.00	7.94
Entertainment –> Entertainment	0.15	1.00	6.64
School/College –> School/College	0.15	1.00	6.58
Store/Shop –> Store/Shop	0.22	1.00	4.47
Diplomacy/Services –> Diplomacy/Services	0.23	1.00	4.39
Bank –> Store/Shop	0.07	0.78	3.46
Store/Shop –> Bank	0.07	0.31	3.46
Bicycle related places –> Bicycle related places	0.29	1.00	3.45
Public Places –> Public Places	0.34	1.00	2.93
Bar/Cafe –> Bar/Cafe	0.36	1.00	2.75
Hotel –> Market	0.02	0.15	2.27
Market –> Hotel	0.02	0.28	2.26
Diplomacy/Services –> Entertainment	0.08	0.33	2.22
Entertainment –> Diplomacy/Services	0.08	0.50	2.21
Theater –> Entertainment	0.02	0.33	2.20
Entertainment –> Theater	0.02	0.11	2.19
Public transport –> Public transport	0.46	1.00	2.18
Theater –> Diplomacy/Services	0.03	0.50	2.18
Diplomacy/Services –> Theater	0.02	0.11	2.15
Bar/Cafe –> Health	0.07	0.19	2.13
Health –> Bar/Cafe	0.07	0.77	2.12
Bicycle related places –> Food	0.03	0.09	2.02
Food –> Bicycle related places	0.03	0.58	2.01
Bank –> Bar/Cafe	0.07	0.73	2.00

Tabella B.11: Raggio di partenza = 50 m - Raggio di arrivo = 50 m - Distanza minima = 100 m - Autunno

sequenze	rel. support	confidence	lift
Post Office –> Post Office	0.03	1.00	35.66
Theater –> Theater	0.04	1.00	24.66
Food –> Food	0.05	1.00	21.92
Religious place –> Religious place	0.06	1.00	15.60
Market –> Market	0.07	1.00	15.36
Health –> Health	0.07	1.00	15.06
Bank –> Bank	0.07	1.00	14.06
Hotel –> Hotel	0.10	1.00	9.64
Entertainment –> Entertainment	0.14	1.00	7.30
Store/Shop –> Store/Shop	0.17	1.00	5.92
School/College –> School/College	0.17	1.00	5.89
Diplomacy/Services –> Diplomacy/Services	0.20	1.00	4.88
Bank –> Store/Shop	0.05	0.68	4.04
Store/Shop –> Bank	0.05	0.29	4.01
Bicycle related places –> Bicycle related places	0.31	1.00	3.22
Bar/Cafe –> Bar/Cafe	0.31	1.00	3.21
Public Places –> Public Places	0.33	1.00	3.00
Bank –> Market	0.01	0.17	2.57
Market –> Bank	0.01	0.18	2.56
Hotel –> Bank	0.02	0.18	2.53
Public transport –> Public transport	0.41	1.00	2.46
Bank –> Hotel	0.02	0.26	2.46
Theater –> Diplomacy/Services	0.02	0.50	2.43
Diplomacy/Services –> Theater	0.02	0.10	2.40
Health –> Bar/Cafe	0.05	0.75	2.39
Bar/Cafe –> Health	0.05	0.16	2.39
Bank –> Entertainment	0.02	0.32	2.32
Entertainment –> Bank	0.02	0.16	2.31
Bar/Cafe –> Store/Shop	0.12	0.38	2.24
Store/Shop –> Bar/Cafe	0.12	0.70	2.23
Entertainment –> Diplomacy/Services	0.06	0.46	2.23

Tabella B.12: Raggio di partenza = 50 m - Raggio di arrivo = 50 m - Distanza minima = 100 m - Inverno

Bibliografia

[1] *POI*. URL: https://it.wikipedia.org/wiki/Punto_di_interesse (cit. a p. 1).

[2] Thiago Silva, Pedro Vaz de Melo, Juliana Salles e Antonio Loureiro. «A comparison of Foursquare and Instagram to the study of city dynamics and urban social behavior». In: ago. 2013. DOI: 10.1145/2505821.2505836 (cit. a p. 2).

[3] *Facebook*. URL: https://facebook.com (cit. a p. 2).

[4] *Instagram*. URL: https://instagram.com (cit. a p. 2).

[5] *Foursquare*. URL: https://foursquare.com (cit. alle pp. 2, 66).

[6] *OpenStreetMap*. URL: https://www.openstreetmap.org (cit. alle pp. 2, 15).

[7] *OSM*. URL: https://it.wikipedia.org/wiki/OpenStreetMap (cit. a p. 2).

[8] Junfeng Jiao e Shunhua Bai. «Understanding the Shared E-scooter Travels in Austin, TX». In: *ISPRS International Journal of Geo-Information* 9.2 (2020). ISSN: 2220-9964. URL: https://www.mdpi.com/2220-9964/9/2/135 (cit. a p. 2).

[9] Shunhua Bai e Junfeng Jiao. «Dockless E-scooter usage patterns and urban built Environments: A comparison study of Austin, TX, and Minneapolis, MN». In: *Travel Behaviour and Society* 20 (apr. 2020), pp. 264–272. DOI: 10.1016/j.tbs.2020.04.005 (cit. a p. 2).

[10] Cédric du Mouza e Philippe Rigaux. «Mobility Patterns». In: *GeoInformatica* 9 (gen. 2004). DOI: 10.1007/s10707-005-4574-9 (cit. a p. 2).

[11] Xiaohu Zhang, Yu Shen e Jinhua Zhao. «The mobility pattern of dockless bike sharing: A four-month study in Singapore». In: *Transportation Research Part D: Transport and Environment* 98 (2021), p. 102961. ISSN: 1361-9209. DOI: https://doi.org/10.1016/j.trd.2021.102961. URL: https://www.sciencedirect.com/science/article/pii/S1361920921002595 (cit. a p. 2).

[12] Zhifu Mi. «Environmental benefits of bike sharing: A big data-based analysis». In: *Applied Energy* 220 (giu. 2018), pp. 296–301. DOI: 10.1016/j.apenergy.2018.03.101 (cit. a p. 2).

[13] Reid Ewing e Robert Cervero. «Travel and the Built Environment: A Meta-Analysis». In: *Journal of the American Planning Association* 76 (giu. 2010), pp. 265–294. DOI: 10.1080/01944361003766766 (cit. a p. 2).

[14] *OverpassTurbo*. URL: https://overpass-turbo.eu (cit. a p. 3).

[15] PANG-NING TAN, MICHAEL STEINBACH e VIPIN KUMAR. *Introduction to Data Mining*. Addison-Wesley (cit. alle pp. 4, 13).

[16] Prof. Elena Baralis. *Association Rules Fundamentals*. https://dbdmg.polito.it/wordpress/wp-content/uploads/2019/10/DSL-3-MassRules.pdf (cit. a p. 7).

[17] Mohammed Zaki. «Zaki, M.J.: SPADE: An efficient algorithm for mining frequent sequences. Machine Learning 42(1), 31-60». In: *Machine Learning* 42 (gen. 2001), pp. 31–60. DOI: 10.1023/A:1007652502315 (cit. alle pp. 8, 21).

[18] Ramakrishnan Srikant e Rakesh Agrawal. «Mining Sequential Patterns: Generalizations And Performance Improvements». In: *EDBT, Lecture notes in computer science. vol. 1057* 1057 (mar. 1996). DOI: 10.1007/BFb0014140 (cit. a p. 8).

[19] Jian Pei, Jiawei Han, Behzad Mortazavi-Asl, Helen Pinto, Qiming Chen, Umeshwar Dayal e Mei-Chun Hsu. «PrefixSpan: Mining Sequential Patterns Efficiently by Prefix-Projected Pattern Growth». In: feb. 2001, pp. 215–224. ISBN: 0-7695-1001-9. DOI: 10.1109/ICDE.2001.914830 (cit. a p. 8).

[20] Mohammed Zaki. «Sequence Mining in Categorical Domains: Incorporating Constraints». In: gen. 2000, pp. 422–429. DOI: 10.1145/354756.354849 (cit. alle pp. 8, 21).

[21] R. Sinnott. «Virtues of the Haversine». In: *Sky and Telescope* 68 (nov. 1984), p. 158 (cit. a p. 26).