



---

## VAST CHALLENGE 2014 MC2

---

FRANCESCO FALLENÌ (559939)

PROGETTO VISUAL ANALYTICS

DATA SCIENCE & BUSINESS INFORMATICS

Anno Accademico 2020/2021

# Indice

<b>1</b>	<b>Introduzione</b>	<b>2</b>
<b>2</b>	<b>Preparazione dei dati e presentazione del modello</b>	<b>2</b>
2.1	Preparazione dei dati . . . . .	2
2.2	Presentazione del modello . . . . .	3
<b>3</b>	<b>Descrizione dell'interfaccia e delle Interazioni</b>	<b>3</b>
3.1	Pannello di Selezione e Filtraggio dei dati . . . . .	4
3.2	Mappa . . . . .	5
3.3	Timeline Transazioni . . . . .	6
<b>4</b>	<b>Vast 2014 Mini Challenge 2</b>	<b>7</b>
4.1	Mini Challenge 2.1 . . . . .	7
4.2	Mini Challenge 2.2 . . . . .	8
4.3	Mini Challenge 2.3 . . . . .	9
<b>5</b>	<b>Dashboard finale</b>	<b>10</b>

# 1 Introduzione

L'obiettivo del progetto è quello di implementare uno strumento che consenta di rispondere alle domande della nella Mini-Challenge 2 di VAST 2014 [2]. Nello specifico, viene richiesto di individuare le routine e i pattern inusuali degli impiegati dell'azienda GASTech a partire dai dati relativi agli spostamenti e alle transazioni effettuate. Infine, è richiesto anche di individuare eventuali anomalie o dati mancanti.

## 2 Preparazione dei dati e presentazione del modello

### 2.1 Preparazione dei dati

In questa sezione verrà descritta la fase di preparazione dei dati effettuata per semplificare la realizzazione della dashboard finale. Alcune fasi della preparazione sono state realizzate in Python mentre altre vengono eseguite durante l'utilizzo dello strumento.

Il dataset fornito è composto dai seguenti file:

- *gps.csv*: 685169 record relativi agli spostamenti degli impiegati, per ogni record sono presenti gli attributi id, Timestamp, latitudine e longitudine.
- *car-assignments.csv*: 44 record relativi agli assegnamenti delle auto ai dipendenti, per ogni record sono presenti gli attributi CarId, Nome, Cognome, Dipartimento e Titolo del dipendente.
- *cc\_data.csv*: 1491 record relativi alle transazione effettuate dai dipendenti, per ogni record sono presenti gli attributi Nome, Cognome, Luogo dell'acquisto, Timestamp e cifra spesa.
- *loyalty\_data.csv*: 1393 record, dati simili a quelli di *cc\_data* ma con Timestamp ridotto (solo la data).

In aggiunta, sono stati forniti anche i dati necessari per poter disegnare la mappa della città di Abila e un'immagine dei luoghi di interesse. Da tali file è stato ricavato un unico file in formato geojson che rappresenta le strade della città su cui mostrare gli spostamenti.

La mappa disegnata utilizzando il file *Abila.geojson* è priva di informazioni sui luoghi di interesse e quindi non permette di dare un contesto ai movimenti. A tale proposito si è tentato di allineare le strade con l'immagine della mappa, ottenendo scarsi risultati. Per risolvere il problema è stato deciso di creare un ulteriore file *location.geojson* contenente le coordinate indicative dei luoghi di interesse ricavate da soluzioni presenti in rete [3].

Successivamente sono stati individuati 9 valori mancanti dell'attributo *CarId* nel file *car-assignments.csv*, la caratteristica che accomuna questi record è il Dipartimento e il Titolo dei dipendenti rispettivamente *Facilities* e *Truck Driver*.

Per poter visualizzare gli spostamenti sulla mappa è stato necessario raggruppare i record di *gps.csv* per un determinato criterio (Ad esempio *CarId* e *Data*) per poi successivamente ordinarli sull'attributo *Timestamp*. Si è proseguito separando ulteriormente la lista di coordinate (ad esempio se la differenza tra due timestamp di elementi adiacenti è maggiore di 10 minuti) per numerare tutti i singoli spostamenti di ogni dipendente nell'arco della giornata. queste operazioni sono state implementate direttamente all'interno della dashboard in JavaScript.

I dati dei file *gps.csv* e *cc\_data.csv* sono stati arricchiti effettuando una join rispettivamente sugli attributi *CarId* e *Nome/Cognome* con i dati del file *car-assignments.csv*. In questo modo per ogni record *gps* e transazione sono disponibili tutti gli attributi di un dipendente.

Infine, sono state assegnate le categorie dei luoghi di interesse e aggiunte ai record di *cc\_data.csv*.

## 2.2 Presentazione del modello

Il modello proposto permette di visualizzare i dati descritti nella mappa di Abila sulla base di vari criteri come il dipendente, il dipartimento del dipendente, il titolo del dipendente, la data e ora attraverso l'interazione con gli elementi grafici.

# 3 Descrizione dell'interfaccia e delle Interazioni

La dashboard è composta principalmente da 3 sezioni:

1. Componenti per la selezione, aggregazione e filtraggio dei dati.
2. Componente mappa dove vengono mostrati i luoghi di interesse e gli spostamenti dei dipendenti.
3. Componente timeline dove vengono mostrate le transazioni nel tempo dei vari dipendenti.

Tutte le componenti sono collegate e mostrano i dati interattivamente sulla base dei criteri di selezione applicati.

## 3.1 Pannello di Selezione e Filtraggio dei dati

Le componenti di selezione permettono di filtrare i dati sulla base di più criteri contemporaneamente, tra questi:

- Dipendenti: è possibile filtrare i dati sulla base di uno o più dipendenti attraverso la tabella con righe selezionabili, per ogni dipendente sono presenti gli attributi CarId, Nome, Cognome, Dipartimento e Titolo.

Inoltre, è possibile assegnare un colore agli spostamenti utilizzando la palette di colori associata. I colori selezionabili sono 13 (Figura 3.1), di cui 12 scelti con lo strumento Color Brewer [1] e uno di default (nero).

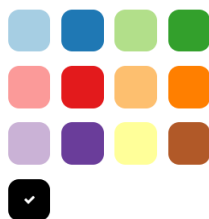


Figura 3.1: Colori dipendenti

- Dipartimento: è possibile selezionare tutti i dipendenti di uno o più dipartimenti (Engineering, Executive, Facilities, Information Technology, Security).

Scegliendo un dipartimento, anche le righe della tabella interessate vengono automaticamente selezionate.

- Data: è possibile filtrare gli spostamenti e le transazioni relative ad uno o più date (yyyy-mm-dd) utilizzando il selettore per la data, il range selezionabile è ristretto alle sole presenti nei record.
- Minuti e ore: è possibile filtrare i dati in un determinato intervallo orario selezionando l'area interessata sul grafico. Il grafico mostra la quantità di record gps per ogni slot orario, questo facilita la ricerca di eventuali spostamenti effettuati dai dipendenti nell'arco della giornata (Figura 3.2).

I dati del grafico si aggiornano dinamicamente al variare dei criteri selezionati.

Infine, è possibile cambiare l'intervallo orario trascinando quello già esistente oppure rimuoverlo con un doppio click in un qualsiasi punto del grafico.

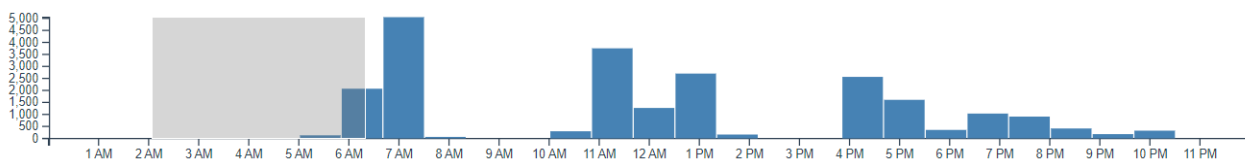






Figura 3.2: Istogramma con selettore

In Figura 3.3 viene mostrato il pannello per la selezione dei dipendenti, dipartimenti e date.

Car ID	First Name	Last Name	Current Employment Type	Current Employment Title	Color
35	Willem	Vasco-Pais	Executive	Environmental Safety Advisor	
4	Ingrid	Barranco	Executive	SVP/CFO	
19	Vira	Frente	Engineering	Hydraulic Technician	
10	Ada	Campo-Corrente	Executive	SVP/CIO	

Select all
Clear selected

Select Employment Type

Engineering
Executive
Facilities
Information Technology
Security

Select dates




Figura 3.3: Selettori per dipendente, data e dipartimento

## 3.2 Mappa

La mappa sulla quale vengono rappresentati gli spostamenti è quella mostrata in Figura 3.4. Le informazioni sui luoghi di interesse sono rappresentate da punti, selezionandoli è possibile vedere il nome del luogo e il numero di transazioni sulla base dei criteri scelti. È presente una legenda che identifica la categoria del luogo con 8 colori qualitativi [1].

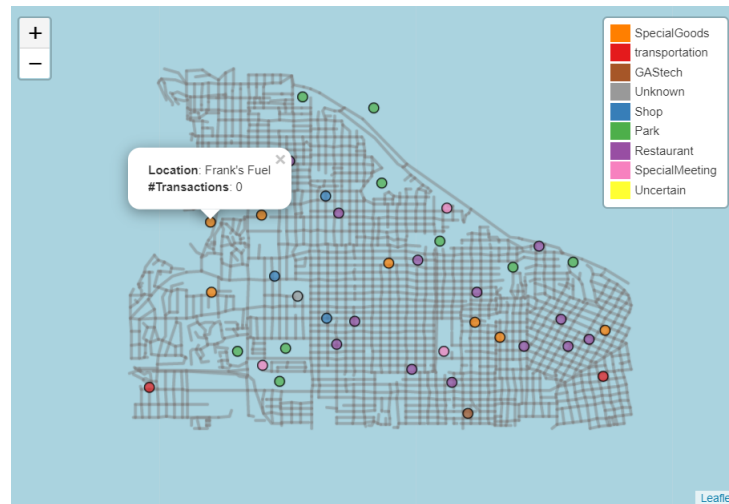


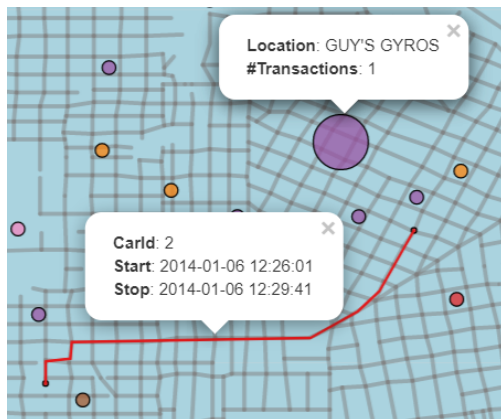
Figura 3.4: Mappa di Abila con luoghi di interesse

Gli spostamenti vengono mostrati sulla base dei criteri scelti, le linee sulla mappa sono interattive e se selezionate mostrano:

- Id della macchina che effettuato lo spostamento.
- Data e ora di inizio dello spostamento.
- Data e ora di fine dello spostamento.

I punti cambiano dimensione sulla base delle transazioni avvenute, questo evidenzia i luoghi dove c'è stata attività.

In Figura 3.5a e 3.5b vengono mostrati alcuni esempi di interazione con la mappa.



(a) Singolo spostamento



(b) Spostamenti di 4 dipendenti

### 3.3 Timeline Transazioni

Il componente Timeline permette di visualizzare le transazioni dei dipendenti nell'arco della giornata (Figura 3.6).

Sull'asse orizzontale si trovano le ore della giornata, sugli assi verticali si trovano gli Id delle auto e le date selezionate.

É presente anche una legenda con gli stessi colori utilizzati per codificare la categoria dei luoghi nella mappa, viene utilizzata per determinare la tipologia di una transazione (SpecialGoods, Transportation, Unknown, Shop, Restaurant, SpecialMeeting, Uncertain).

Per meglio visualizzare le transazioni è possibile effettuare uno zoom sull'intervallo orario interessato.

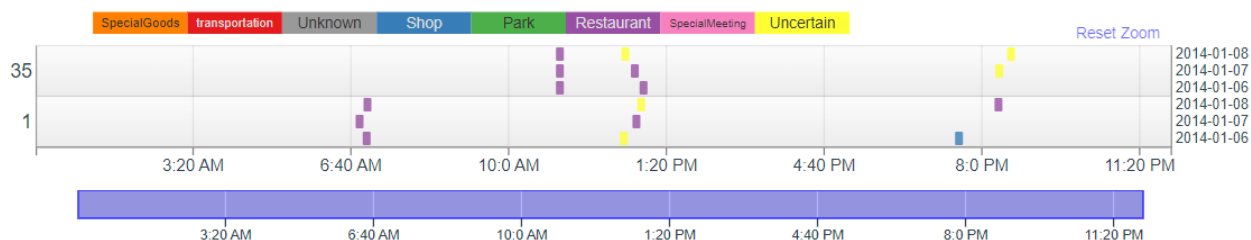


Figura 3.6: Timeline delle Transazioni

Passando con il puntatore su una transazione è possibile visualizzare i suoi dettagli (3.7), per ognuna viene mostrato il prezzo, l'ora esatta e il luogo dove avvenuta.



Figura 3.7: Dettagli di una Transazione

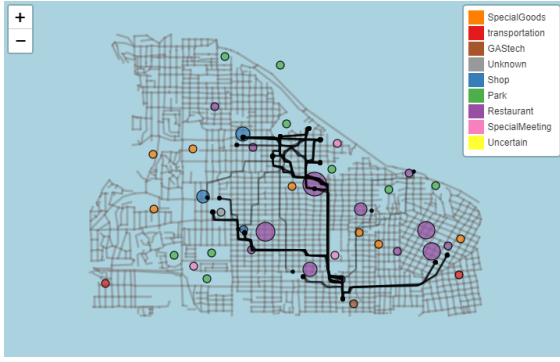
## 4 Vast 2014 Mini Challenge 2

In questo capitolo verranno fornite delle risposte semplificate alle domande della Mini Challenge 2 utilizzando la dashboard implementata.

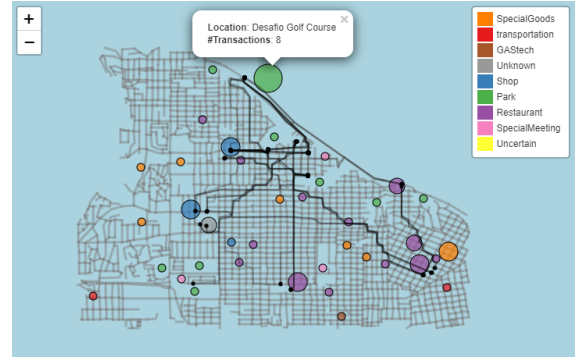
### 4.1 Mini Challenge 2.1

La prima domanda richiede di descrivere una giornata tipo di un dipendente della GASTech. Per rispondere si sono considerati tutti gli spostamenti dei dipendenti del dipartimento Executive nei giorni feriali (4.1a) e nei giorni festivi (4.1b).



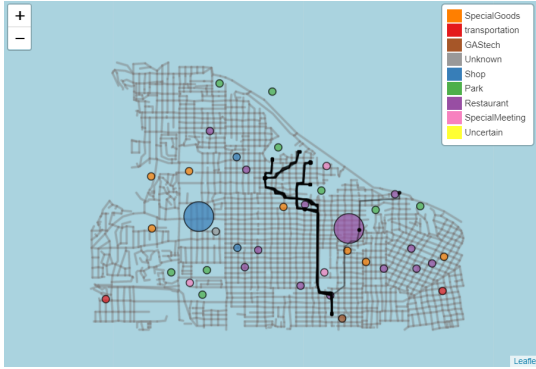


(a) Spostamenti dei dipendenti del dipartimento Executive nei giorni feriali

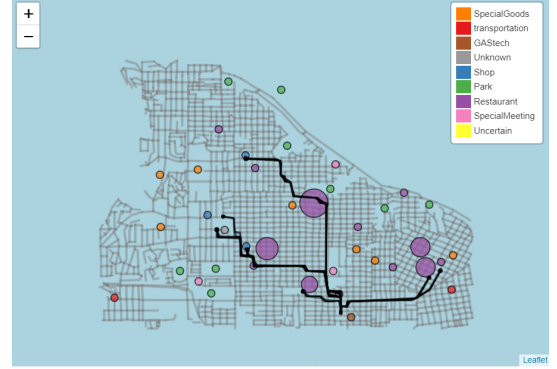


(b) Spostamenti dei dipendenti del dipartimento Executive nei giorni festivi

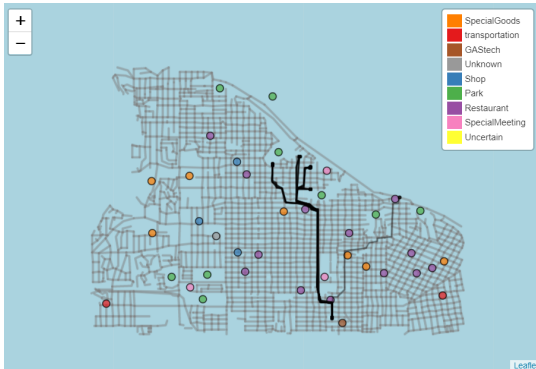
Sono stati considerati gli stessi spostamenti nei giorni feriali dei dipendenti del dipartimento Executive filtrandoli per intervalli orari differenti 4.2.



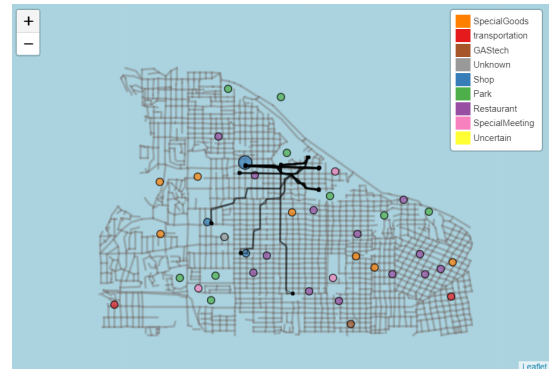
(a) Lavoro



(b) Pausa pranzo



(c) Ritorno dal lavoro



(d) Cena

Figura 4.2: Spostamenti dipendenti per intervalli orari

## 4.2 Mini Challenge 2.2

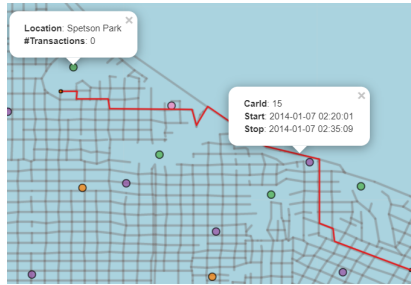
La seconda domanda richiede di identificare eventuali pattern inusuali nei dati e per ognuno di descriverne la tipologia, chi è coinvolto, quali luoghi sono coinvolti e quando è avvenuto.

Il pattern individuato vede coinvolti i dipendenti Isia Vann (16) e Loreto Bodrogi (15) nei giorni 6 e 7 Gennaio. Isia Vann si dirige verso Spetson Park alle ore 22:01 e arriva alle ore 22:09 del giorno 6 Gennaio (4.3a). Successivamente, Loreto Bodrogi raggiunge Isia

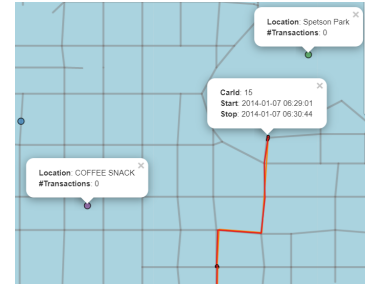
Vann a Spetson Park partendo alle ore 2:20 e arrivando alle ore 2:35 del giorno 7 Gennaio (4.3b). Entrambi rimangono a Spetson Park fino a quando, alle ore 7:06, si dirigono verso Coffee Snack per fare colazione, per poi proseguire alla GASTech (4.3c). Durante lo stop a Coffee Snack i due dipendenti non pagano il conto in quanto non si verificano transazioni in quell'intervallo orario.



(a) Isia Vann si dirige verso Spetson Park



(b) Loreto Bodrogi si dirige verso Spetson Park



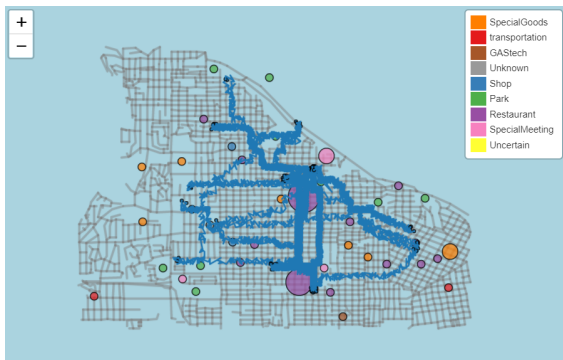
(c) Entrambi si dirigono verso GASTech

Figura 4.3: Incontro nella notte a Spetson Park

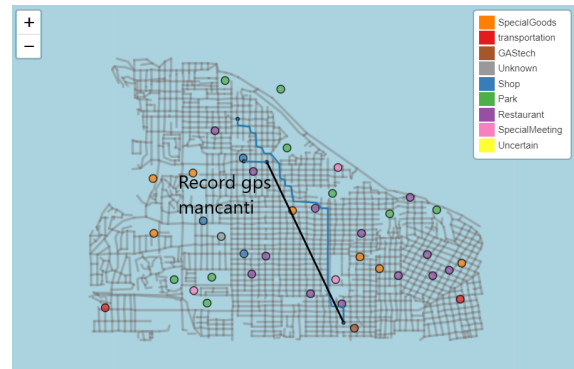
## 4.3 Mini Challenge 2.3

La terza domanda richiede di individuare e di gestire eventuali dati mancanti o in conflitto presenti nel dataset.

I dipendenti che hanno dati mancanti o in conflitto sono Elsa Orilla e Calzas Axel, il primo dipendente presenta delle rilevazioni gps imprecise (4.4a), mentre il secondo ha dei record mancanti (4.4b).



(a) Spostamenti di Elsa Orilla (28)



(b) Spostamenti di Calzas Axel (9)

## 5 Dashboard finale

La dashboard finale (Figura 5.1) comprende tutte le componenti descritte in precedenza.

La prima interazione dovrà avvenire con il pannello di selezione dipendenti, dipartimento e data in modo tale da aggiornare i dati l'istogramma.

Successivamente sarà possibile selezionare la fascia oraria di interesse sull'istogramma così da mostrare i soli spostamenti e le transazioni in quel dato intervallo.

Gli spostamenti sulla mappa e le transazioni nella timeline si aggiornano dinamicamente in base ai criteri scelti.

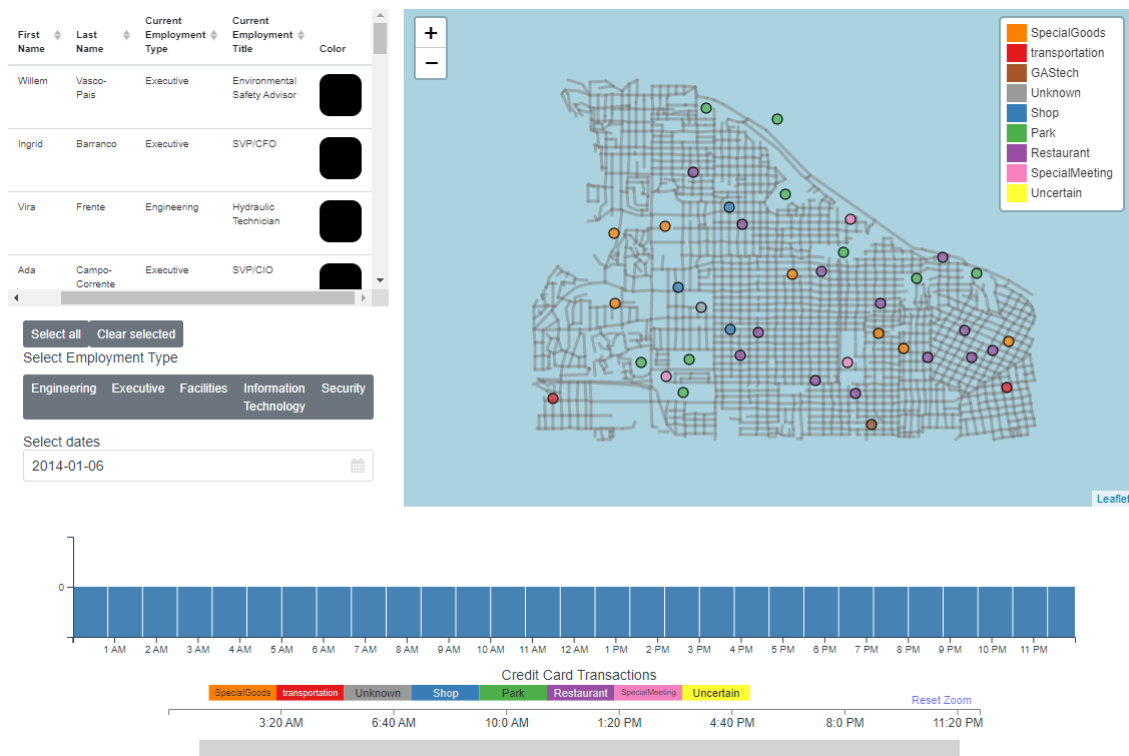


Figura 5.1: Dashboard

La dashboard realizzata prende ispirazione da uno strumento simile [3] proposto per la risoluzione dello stesso problema.

# Bibliografia

- [1] *Color Brewer 2*. URL: <https://colorbrewer2.org/>.
- [2] *Vast 2014*. URL: <https://vacommunity.org/VAST+Challenge+2014>.
- [3] *Visual Analytics Benchmark Repository, Vast 2014 MC2, PKU-Chen-2*. URL: <https://www.cs.umd.edu/hcil/varepository/index.php>.