Project Report

for

VAST Challenge 2014

Mini Challenge 2

Group 3

Abhishek Ganta, Sarvar Khamidov,

Jonathan Marquis, Jay Patel

Professor Zhang

DS 330

11 December 2021

Introduction

This project in its entirety was one that required the mining, processing, and visualization of data in conjunction with synthesizing processes of transforming data into insightful information that could be used to draw connections and conclusions in order for us to figure out a solution to a problem. The problem within Mini Challenge #2 required us to process and visualize data relative to employee movement and activity to gather more detailed information for the investigations on missing people within a company of interest, where the data indicated activity two weeks leading up to the disappearances. In this regard, there were many routines and other variables, such as credit card and mobile activity that were both benign and suspicious that supported and created evidence towards the idea that within the company, there are people of interests whose activities make them appear to be guilty. The use of Tableau as a visualization tool together with the imagination of the data analysts in this project allowed for the portrayal of a story line that matched that of the investigation, allowing us to make vital connections and therefore conclusions regarding the company's involvement.

Data

‘**car-assignments.csv**’

This csv file is used in the workbook datasource. It contains 5 features, all related to giving information about the employee and their car id. This csv is the most crucial one in identifying the exact employees we are filtering by their car ids and their names. The features included in order are ‘LastName’, ‘FirstName’, ‘Car ID’, ‘Current Employment Type’, and ‘Current Employment Title’. An example case is [Flecha, Sven, 17, Information Technology, IT Technician]. This file connects to the ‘cc_data.csv’ file by the variable ‘LastName’=’LastName’ and the ‘gps.csv’ file by the variables ‘Car ID’ = ‘Id’.

‘cc_data.csv’

This csv file is used in the workbook datasource. It contains 5 features, all related to the spending information of each employee, and this includes their exact location and date they spent their money at, but does not include the exact time. This is more so just the dataset to use to track the base route and base spending of each employee. The features included in order are ‘Timestamp’ in M/D/Y H:M:S format, ‘Location’, ‘Price’, ‘FirstName’, and ‘LastName’. An example case is [1/6/2014 7:36:00 AM, Hallowed Grounds, 16.72, Birgitta, Frente]. This file connects to the ‘car-assignments.csv’ file by the variables ‘LastName’ = ‘LastName’.

‘**gps.csv’**

This csv file is used in the workbook datasource. It contains 4 features, all related to the exact location of an employee’s car at an exact time. What this does is, if mapped correctly using certain types of plots, such as a scatter plot, and the latitude and longitude features, it can accurately portray the daily routes of each employee. The features included in order are ‘Timestamp’ in M/D/Y H:M:S format, ‘Id’, ‘Lat', and ‘Long’. An example case is [1/7/2014 8:05:44 AM, 2, 36.0827832, 24.8518351]. This file connects to the ‘car-assignments.csv’ file by the variables ‘Car ID’ = ‘Id’.

‘loyalty_data.csv’

This csv file is used in the workbook datasource. It contains 5 features, all related to the spending information of each employee, and this includes their exact location, date, and time they spent their money at. This dataset, although not used as much as the ‘car-assignments.csv’ file, is the source for all connections to the workbook used for mini-challenge 2. This dataset helps in verifying locations and spending of employees between each other. The features included in order are ‘Timestamp’, ‘Location’, ‘Price’, ‘First Name’, and ‘Last Name. An example case is [1/6/2014, Brewed Awakenings, 4.35, Ada, Campo-Corrente]. This file connects to the ‘car-assignments.csv’ file by the variables ‘Last Name’ = ‘LastName’.

‘mc2-tourist.jpg’

This image file is particularly important in the visualization for our project as without it, we would have to worry about the coordinates given in the ‘gps.csv’ file displaying our routes in the middle of the ocean on a live map. With this image file, we can layer it as a placeholder for the live map and a background for which we can easily match and refer to the locations in our datasets. Albeit, the locations, employees, and the actual image map is fake, but are still very helpful to match the data accordingly as it will easily help us plot whatever we want to measure. The place is called Abila, Kronos off the coast of the Mediterranean Sea.

Analyses

Common daily routine analysis

Routine to analyze:

Isande Borrasco #7 and Calixto Nils #8 represented in this dataset are very much creatures of habit and, in general, have a fairly predictable routine. On weekdays, they get up in the morning to drive to work while stopping off for coffee along the way. At lunch they will go to lunch then return to work. In the evening they may go out for dinner then back home. On weekends, they may go to lunch, go shopping, and go out to dinner in the evening.

Evidences to support the analysis:

We found that two individuals Isande Borrasco #7 and Calixto Nils #8 have same routine everyday except for the weekends. We can see that during the mornings of weekdays they usually visit ‘Hallowed Grounds’ for coffee most likely, during lunch time ‘Brewed Awakenings’ and ‘Chostus Hotel’, then for the later after work ‘Frydos Auto Supply N more’ and ‘Hippokampos’. However during the weekends both of them dinner time they visit ‘Katerna’s cafe’, during lunch time ‘Abila Zacharo’, ‘Hippokampos’ and ‘Kalami Kafenion’.

Data involved:

In order to make two visual analyses for this data analysis we used two datasets ‘gps.csv’ and ‘cc_data.csv’. First mapping visualization used ‘Long’ and ‘Lat’ columns from gps.csv table with filter of ID from gps.csv data of 7 & 8, Hour: 7 & 12, Date: January 15, 2014. Also the other table used data from cc_data.csv with columns SUM(Price) and rows Day, Location, Last Name. Then we filtered to individuals with id 7 & 8 again, then to different times of the day morning, lunch and dinner.

Map routes for Isande and Calixto

Rows and Columns

Long, Lat

Filters:

Id: #7 (Isande Barrasco), #8 (Calixto Nils)

Hour: 7 & 12

MDY: January 15, 2014

Weekday: Wednesday

Spending weekdays and weekends for Isande and Calixto

Columns: SUM(Price)

Rows: Day, Location, Last Name

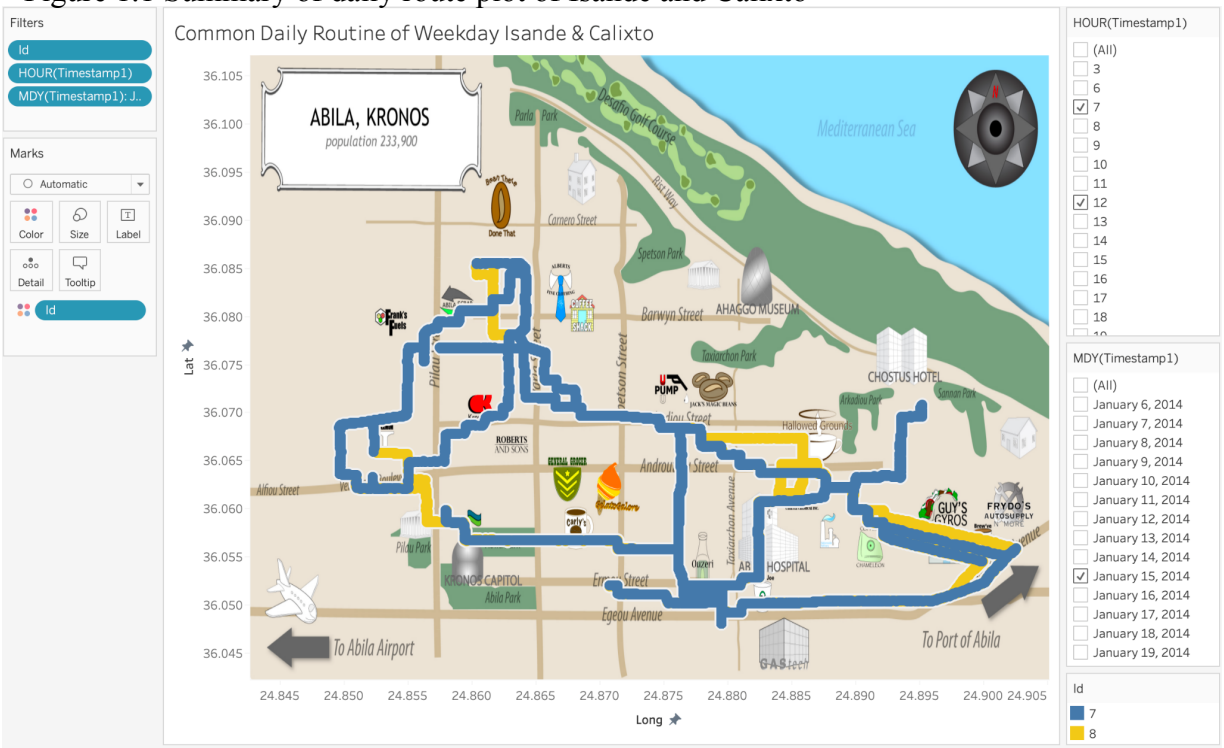
Filters:

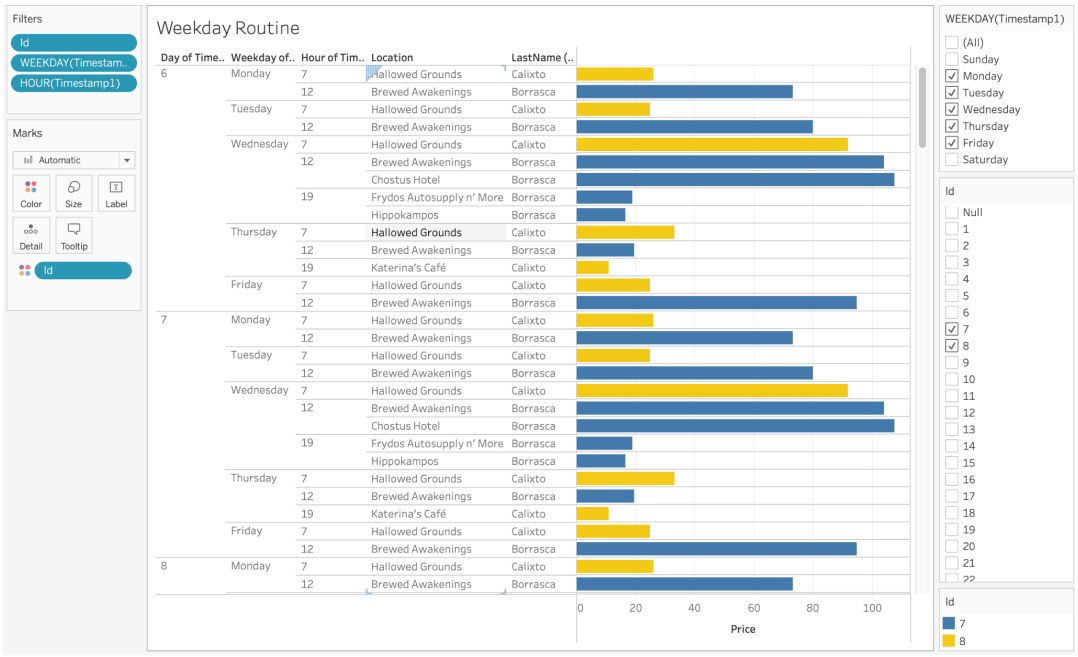
Id: #7 (Isande Barrasco), #8 (Calixto Nils)

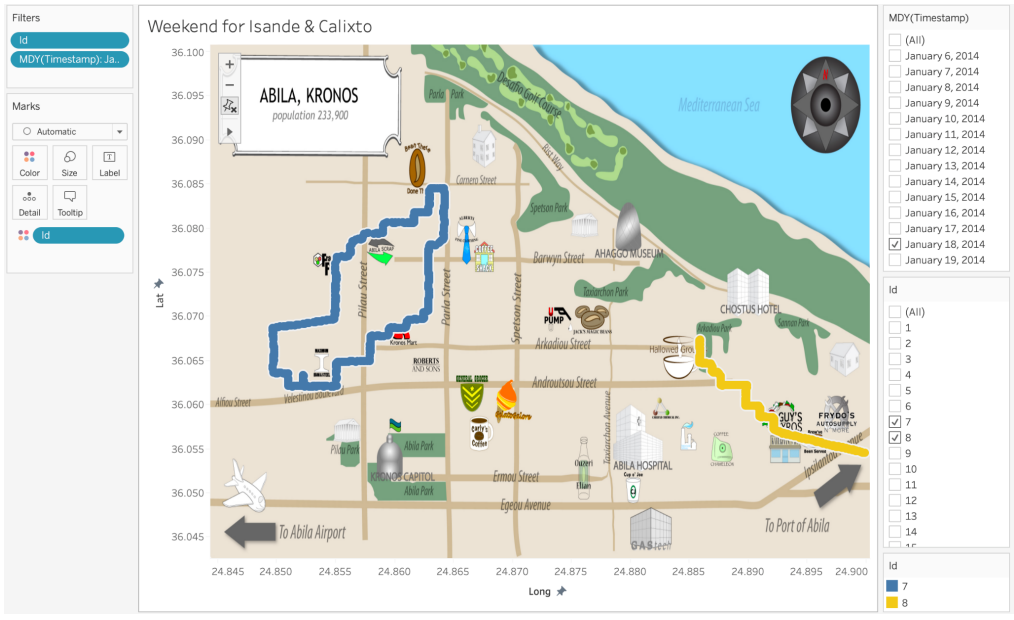
Hour: Morning, **Lunch, Dinner**

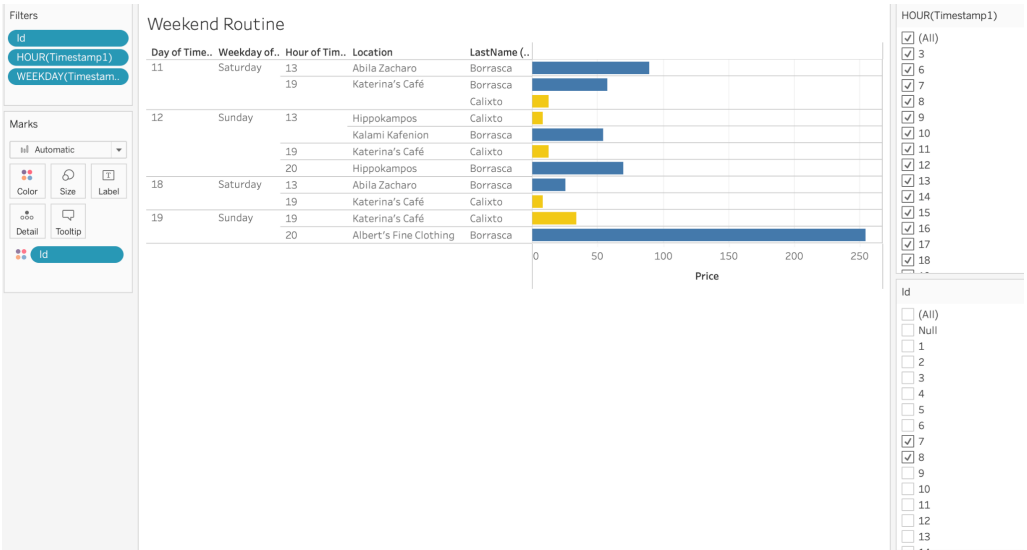
Visualization results (screenshots):

Figure 1.1 Summary of daily route plot of Isande and Calixto




Filters
Id
WEEKDAY(Timestam..
HOUR(Timestamp1)
Marks
Automatic
Color
Size
Label
Detail
Tooltip
Id
Weekday Routine
Day of Time..
Weekday of..
Hour of Tim..
Location
LastName (..
6
Monday
7
Hallowed Grounds
Calixto
12
Brewed Awakenings
Borrasca
Tuesday
7
Hallowed Grounds
Calixto
12
Brewed Awakenings
Borrasca
Wednesday
7
Hallowed Grounds
Calixto
12
Brewed Awakenings
Borrasca
Chostus Hotel
Borrasca
19
Frydos Autosupply n' More
Borrasca
Hippokampos
Borrasca
Thursday
7
Hallowed Grounds
Calixto
12
Brewed Awakenings
Borrasca
19
Katerina's Café
Calixto
Friday
7
Hallowed Grounds
Calixto
12
Brewed Awakenings
Borrasca
7
Monday
7
Hallowed Grounds
Calixto
12
Brewed Awakenings
Borrasca
Tuesday
7
Hallowed Grounds
Calixto
12
Brewed Awakenings
Borrasca
Wednesday
7
Hallowed Grounds
Calixto
12
Brewed Awakenings
Borrasca
Chostus Hotel
Borrasca
19
Frydos Autosupply n' More
Borrasca
Hippokampos
Borrasca
Thursday
7
Hallowed Grounds
Calixto
12
Brewed Awakenings
Borrasca
19
Katerina's Café
Calixto
Friday
7
Hallowed Grounds
Calixto
12
Brewed Awakenings
Borrasca
8
Monday
7
Hallowed Grounds
Calixto
12
Brewed Awakenings
Borrasca
0
20
40
60
80
100
Price
WEEKDAY(Timestamp1)
(All)
Sunday
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Id
Null
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
Id
7
8


Filters
Id
MDY(Timestamp): Ja..
Marks
Automatic
Color
Size
Label
Detail
Tooltip
Id
Weekend for Isande & Calixto
ABILA, KRONOS
population 233,900
Mediterranean Sea
Desmos Golf Course
Parda Park
Carnem Street
Spetson Park
Barwyn Street
AHAGGO MUSEUM
Taxiarchon Park
Pilaou Street
Parda Street
Spetsson Street
Arkadiou Street
Androutsou Street
CHOSTUS HOTEL
Hallowell Girls
ROBERTS AND SONS
Kronos Park
Abila Park
KRONOS CAPITOL
Ermou Street
Egeou Avenue
ABILA HOSPITAL
GUY'S AUTOSUPPLY
FRYDO'S AUTOSUPPLY
To Abila Airport
To Port of Abila
Lat
36.100
36.095
36.090
36.085
36.080
36.075
36.070
36.065
36.060
36.055
36.050
36.045
24.845
24.850
24.855
24.860
24.865
24.870
24.875
24.880
24.885
24.890
24.895
24.900
Long
MDY(Timestamp)
(All)
January 6, 2014
January 7, 2014
January 8, 2014
January 9, 2014
January 10, 2014
January 11, 2014
January 12, 2014
January 13, 2014
January 14, 2014
January 15, 2014
January 16, 2014
January 17, 2014
January 18, 2014
January 19, 2014
Id
(All)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
Id
7
8


Filters
Id
HOUR(Timestamp1)
WEEKDAY(Timestam..
Marks
Automatic
Color
Size
Label
Detail
Tooltip
Id
Weekend Routine
Day of Time..
Weekday of..
Hour of Tim..
Location
LastName (..
11
Saturday
13
Abila Zacharo
Borrasca
19
Katerina's Café
Borrasca
Calixto
12
Sunday
13
Hippokampos
Calixto
Kalami Kafenion
Borrasca
19
Katerina's Café
Calixto
20
Hippokampos
Borrasca
18
Saturday
13
Abila Zacharo
Borrasca
19
Katerina's Café
Calixto
19
Sunday
19
Katerina's Café
Calixto
20
Albert's Fine Clothing
Borrasca
0
50
100
150
200
250
Price
HOUR(Timestamp1)
(All)
3
6
7
8
9
10
11
12
13
14
15
16
17
18
Id
(All)
Null
1
2
3
4
5
6
7
8
9
10
11
12
13

Anomalous but Benign Activity Analysis

Activity to analyze:

Employees Isande Borrasca (#7) and Brand Tempesta (# 33) having an affair. Multiple times they go to a hotel over lunch, usually leaving and returning at times offset from each other. Happens on days: January 8, 10, 14, and 17.

Evidences to support the analysis:

Employees Isande and Brand can be seen hanging out at the same places spending approximately the same amount of amenities and services at the same exact times. Looking at the bar graph under visualization results, we can see that on January 8th, they both stayed at the Chostus Hotel in the afternoon. On Jan 10th, they both ate out at Katerina's Cafe at night. On Jan 14th, they both attended the Hippokampus at night, and on Jan 17th, they both ate at Guy's Gyros. The two maps below the bar graph show the route both Isande and Brand have taken on January 14th. Brand is colored in red while Isande is in blue, and you can clearly see that the routes shown on the map for both of them are almost exactly the same. Meaning, they've been going to places together.

Data involved:

‘cc_data.csv’:

Columns: SUM(Price CC Data)

Rows: DAY(Timestamp CC Data), Location(CC Data), LastName(CC Data)

Filters:

Day: 8, 10, 14, 17

LastNameCC

Marks:

Color:

Locations (suspected affair locations are highlighted in light blue)

FirstName (Brand (Tempestad) is dark orange, Isande (Borrasco) is blue)

Visualization results (screenshots):

Affair Bar Graph and Jan 14th Map Routes

Figure 2.1 Bar Graph of Isande and Brand Affair

Figure 2.1 Jan 14th Route of Isande

Figure 2.3 Jan 14th Route of Brand

Anomalous and Part of Plot Activity Analysis

Activity to analyze:

Loreto Bodrogi (#15), Isia Vann (#16), Hennie Osvaldo (#21) and Minke Mies (#21) carry out surveillance on executive houses in 3 hour shifts on January 6, 7, 9, 10, 11, 13, and 14 in the middle of the night when nobody else is driving.

Evidences to support the analysis:

We found the routes taken by Loreto Bodrogi (#15), Isia Vann (#16), Hennie Osvaldo (#21) and Minke Mies (#24) had recurrent stops and were seen taking the similar routes. The overlapping routes are evident from the fact that these employees were carrying out surveillance visits.

Data involved:

'gps.csv':

Columns: Long

Rows: Lat

Filters:

Id: #15 (Loreto Bodrogi), #16 (Isia Vann) , #21 (Hennie Osvaldo), #24 (Minke Mies)

MDY: January 6th, 7th, 9th, 10th, 11th, 13th, and 14th of 2014

Hour: 20, 21, 23

Visualization results (screenshots):

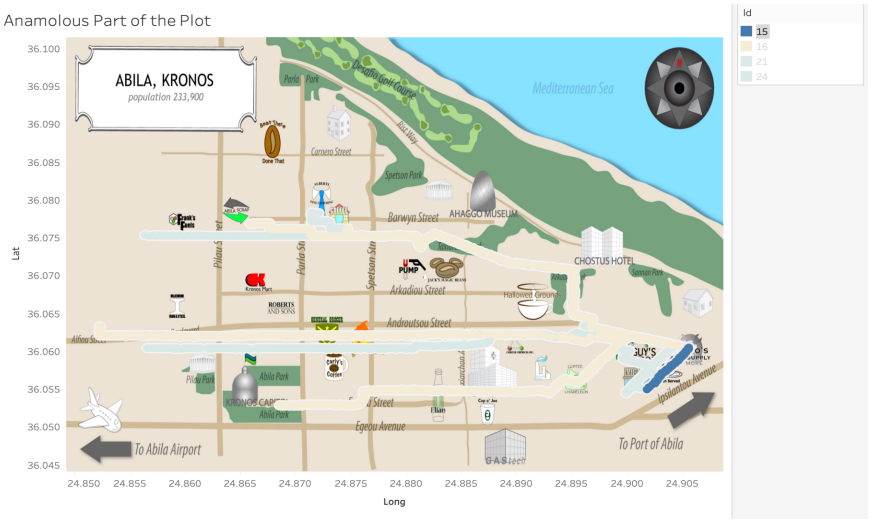


Figure 3.1 showing the route for Loreto Bodrogi (#15)

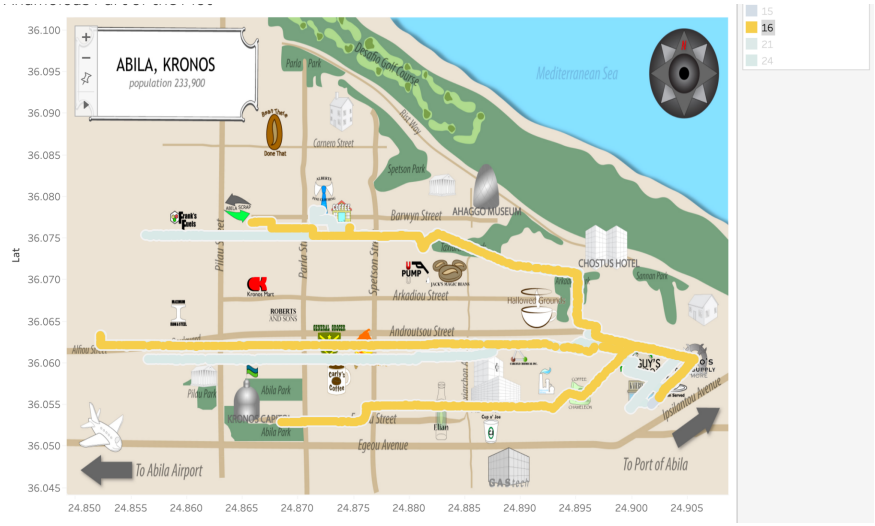


Figure 3.2 showing the route for Isia Vann (#16)

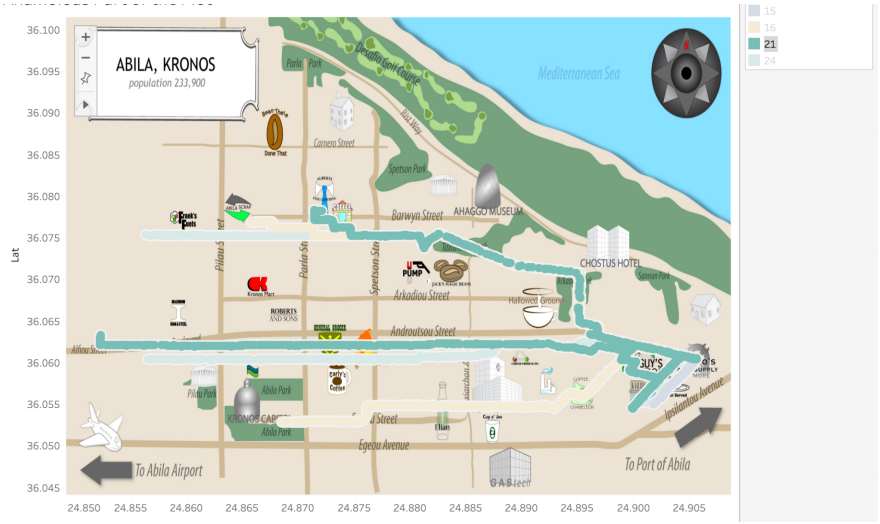


Figure 3.3 showing the route for Hennie Osvaldo (#21)

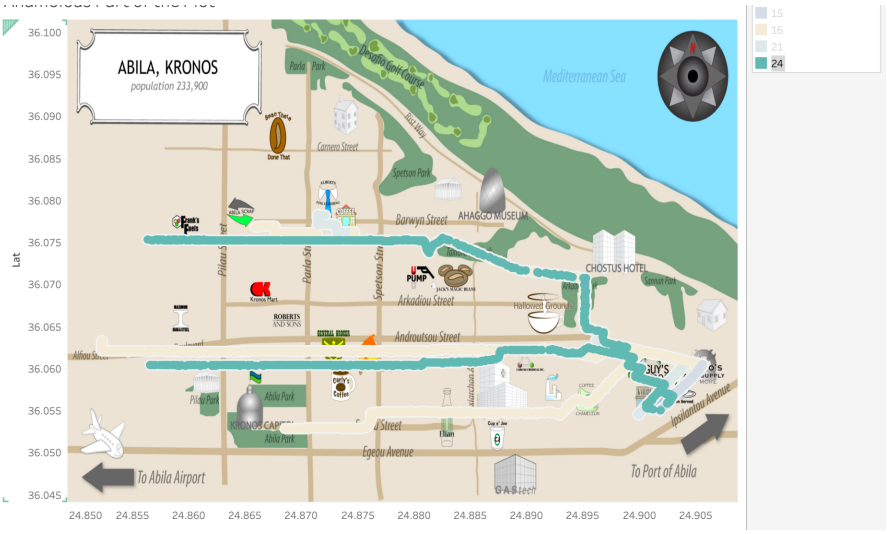


Figure 3.4 showing the route for Minke Mies (#21)

Figure 3.5 showing the overlapping paths for Loreto Bodrogi (#15), Isia Vann (#16), Hennie Osvaldo (#21) and Minke Mies (#21)

Conclusion

In conclusion there were many things that we learned during and after the completion of this project. We learned the importance of planning how one can use data to influence perception and decision-making. We also learned how data can be used to tell a story. In this project we were supposed to depict the daily routines of the employees on a geographical map, where local stores, shops, and places were labeled. Without the visual representation of the data with charts and the geospatial maps we would not have been able to magnify things such as overlapping routes and spending coincidences, in order to make sense of the company's involvement in the bigger investigation. Part of what goes into proper planning includes choosing the right graphics and visuals to portray exactly what it is that supports your claims. For example, figuring out what filters in the data gave us the most reasoning as to why person A and B may be accomplices and working together on something that can be considered suspicious. This also includes making sure that detailing is accurate. At first we wanted to make geospatial visuals using longitude and latitude but made none in order to have the plot without the live map. We also came across alignment issues with the scatter plot and background image we added. While filtering we had to make sure that when the plot was filtered with one feature, that the other features such as IDs do not disappear and hinder the plot. Even though it was difficult for us to get the mapping background and trace routes to align, the closer we were to proper alignment, the easier it was to deduce and conclude our finding with certainty, whereas misalignment caused more error in analysis, swaying the integrity and accuracy of our connections and conclusions. If we had more time on this project we could integrate different technologies and techniques to make our visual and analytical components easier to interact with. For example, working on cumulative score calculation, which would assist us in processing the transactions data on employees' credit cards better. In addition, using things such as Javascript and Unfolding to create depictive maps and geo visualizations that are more interactive and depictive with features such as zooming, panning multitouch gestures, would have been a great addition as well.