Jonathan Jackson

Data Science 330

Final Project Report

For this project, we were tasked with examining the activities of employees of the fictitious company GAStech through the geographic data we were provided. Our employees resided in the town of Abila, which was located in the country of Kronos. We were able to utilize visualization tools in Tableau to properly display the movement data of specific employees as they went through their daily activities. Tableau is a visual analytics tool focused on business analytics and general analysis. The software is focused on making data and visualizations interactive, rather than solely in code. From examining specific peoples' daily routines and unusual events, we were able to collaborate on what was the best way to approach this deep dive. We were tasked with utilizing four datasets that would help us visualize daily routines, an anomalous but benign event, and an anomalous and part of plot event. Our visualizations and data analysis serves as validation to events such as getting coffee before going to work and surveillance shifts at 3 am. In this report, we will be explaining our findings and thought processes as we worked through this data. To the left, we have a map depicting the city of Abila, in the country of Kronos.

Methods

Through the use of Tableau, a visualization software that helps us see and understand data, we were able to create multiple visualizations that validated the three events we were tasked with. Firstly, our data needed to be joined to form relationships that allowed for plotting across datasets. We decided to join the GPS data with the car assignments where "id" in the GPS data is equal to "CarID" in the car assignments data, as shown in Figure 1 and Figure 2. To visualize daily routines, location data from the GPS dataset was combined with car ID data from the car assignments dataset. This allowed us to see which car was where at any given time or date in the dataset. The timestamp in that visualization was used as a filter and was set to January 6, 2014, at 1:20 am to January 7, 2014, at midnight providing a view of one single day's routine. Another method used was filtering by exact dates, which was used to visualize the anomalous part of the plotting task. We were able to select which dates should and should not be included in our visualization through a single value selection list as displayed in Figure 3.

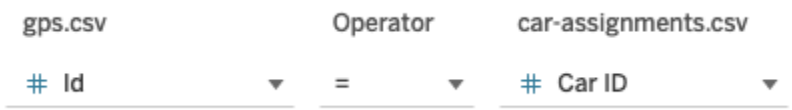


Figure 1: Joining GPS data with Car Assignments data on car ID

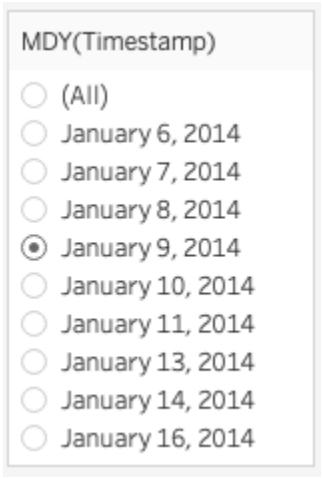
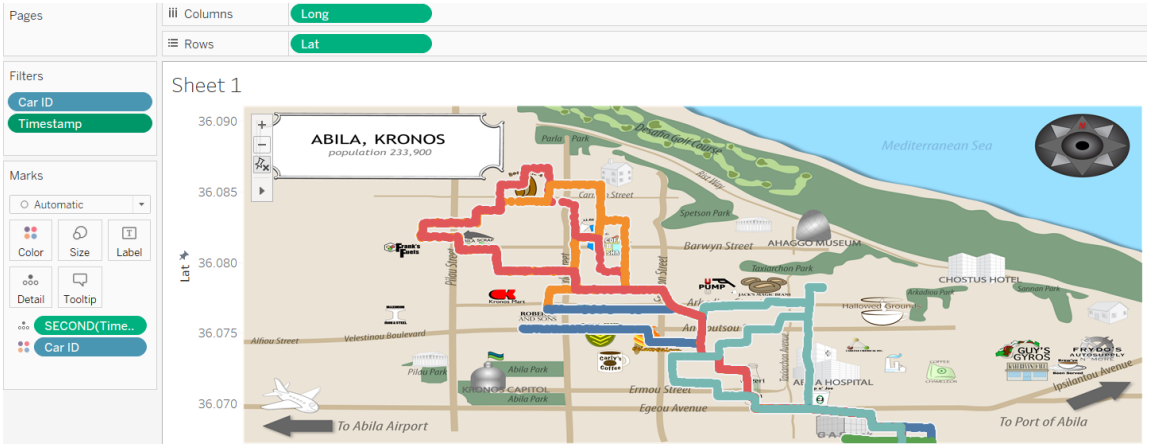


Figure 3: Our date filter

Figure 2: The relationships between our data

Daily Routines

Our first task was to make a visual representation of a chosen daily routine. This scenario asked us to track the routine of employees on a weekday. For example, a routine could show an employee getting coffee before work, heading to work then after work would take a drive to the golf course. We picked 5 random CarIDs as the 5 employees we wanted to track. By doing that we filtered out CarIDs from the car-assignments.csv file and picked 5 random IDs in that bunch. To plot the routes that these cars took, we plotted the longitude and the latitude that comes from the gps.csv. We also filtered out the times that these routes took place. We wanted to visualize the routine of one day and it has to be a weekday. So we chose January 6th and verified that it is a weekday. After that, we wanted to color code the employees so that we can keep track of the routes being taken throughout the day. Then, we marked "second(timestamps)" so that whenever we hover our mouse on a certain CarID path, we can see at what time that route took place. Therefore, we were able to see when these employees did certain tasks. For example, from this visualization, we were able to see that CarID(light blue) arrived at work near 8 After and didn't get out until 6:25 PM. After that he headed to Roberts and Sons at 7 PM then he headed back home at around 7:30.



We wanted to track the morning routine of some employees. We filtered out the hours of the

morning to see what the employees do before work on a weekday. The day January 8th was chosen because it is a weekday. We also choose 4 random CarIDs to track. After that, we found out that CarID #14 went and bought coffee before they went to work. CarID #5 went to the supermarket before they arrived at work. CarID #7 went straight to work instead of doing anything else. The last random employee CarID #17 went to the auto supply store before they got to work. There is nothing unusual about these routines but it's interesting to see if these are consistent routes since this is one day out of a year.

Anomalous and Part of Plot Event

To validate that surveillance was taking place at odd hours of the night by certain GasTech employees, we visualized the data on the given dates of January 6, 7, 9, 10, 11, 13, and 14. We were able to plot the locations of specific employees and aggregate them by Car ID because our GPS data was joined to the Car Assignments data. Upon analyzing the visualization,

we found multiple dates and times in which surveillance was taking place and was able to infer that the routes being taken were going to and from executive homes. At 3:20 am on January 9, 2014, Car 15 left their home in the southeast part of Abila and drove to the northwest. As shown in Table 1, ten minutes later, at 3:30 am, Car 24 is documented leaving the northwest part of town to go back home in the southeast using the same route as Car 15. After further analysis, we inferred that these two cars were changing surveillance shifts at approximately 3:30 am. Our given information states that the shifts are three hours long and so we can infer that on January 9, Car 24 was on duty from 12:30 am - 3:30 am, and Car 15's shift was from 3:30 am - 6:30 am.

View Data: Surveillance

Show aliases | Copy | Export All

Car ID	Lat	Long	First Name	Second of Timestamp
15	36.0721241700	24.8745984300	Loreto	1/9/2014 3:32:42 AM
24	36.0721457200	24.8746267000	Minke	1/9/2014 3:30:01 AM

Table 1: Car 15 arriving to switch surveillance shifts with Car 24

Figure 2: Car 21 arriving to switch surveillance shifts with Car 16

On January 11, 2014, as illustrated in Figure 2, we saw similar behavior with Car 16 leaving the executive home at 3:23 am and Car 21 arriving at the executive home at 3:32 am. These cars take a slightly different route to the executive homes than the cars on January 9, but end up in a similar location. Overall, we find that the suspected surveillance is indeed happening by specific

cars at odd hours of the night.

Anomalous but Benign Event

To build a visualization tool to validate our "anomalous but benign" event, we chronicled Employee #29 Bertrand Ovan cruising around town on January 11. We utilized Tableau to filter data corresponding to Ovan's identifiers, which include his Car ID, name, and the date he traveled from his home. Through this data, we were able to determine his path and location around Abila, Kronos through the variables of Latitude and Longitude. An additional piece of data that was included was the time he was positioned in each coordinate, and we observed that he cruised around town from approximately 6:30 pm to midnight. Furthermore, the amount of money spent was included in the dataset, and based on Ovan's event, we observed that he made his only transaction at Frydo's Auto Supply, which occurred around 8 pm. Based on the GPS, we

infer that he made this particular purchase to fix his car, as he went back home immediately after. In addition, based on the timestamps featured in the data, we can observe how Ovan left his house at about 10 pm and cruised around again until midnight. We also see that his route

is similar to a loop and how he passes by various spots, including a coffee shop, the park, and a hospital.

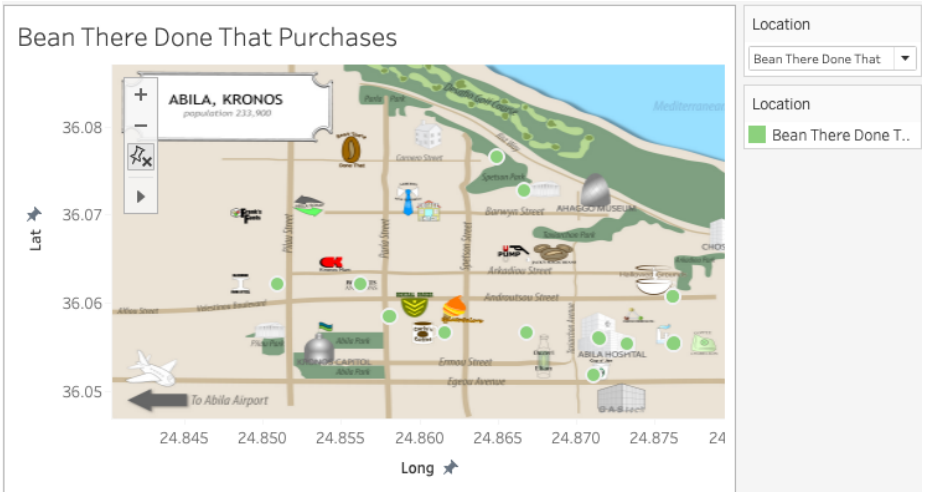
We built the visualization tool by filtering the employee(CarID #29) then filtering the days where that employee took that late-night drive. The columns are the average longitude and the average latitude are the rows. After that, we were able to see the employee's location for the entire day and what time he arrived through specific points on the map.

Imperfect Data

All five of our group members were brand new to the Tableau software, immediately presenting the first obstacle we had to overcome. Using sources such as eLearning and some other videos posted online, the group had a solid footing on the project ahead. Once we were feeling comfortable with the software, it was time to start breaking down the dataset provided to us. The first thing that the group noticed among the data was the extreme differences with each dataset. Some cases involved very simple to read and classified data that was able to be recognized in an instant. Only some portions of data left null values, namely the truck drivers in the employee dataset. The data on the GPS file contained over 600,000 rows of data that allowed the group to create points that would refer to dates and times throughout the project. Most of our project had us working through the GPS CSV. The portion that gave us trouble when rounding the table together was making comparisons with the data. Using filters, we were able to separate data and find the answers that we were looking for in each task. For some of our tasks, to make sure that our group was filtering things correctly, we would use tables/charts to count the number of points to see if we were calculating correctly. Seeing what went together and being able to make use of it is described in the following paragraphs.

Conclusion

In conclusion, for this project, we were able to utilize the visualization software Tableau to analyze the movement and activities of the employees of GAStech. By combining our credit card and car assignment datasets with our GPS data, we were able to create a data source that we continually referenced throughout the project. Through the use of specific features, such as filtering, we were able to find patterns specific to certain individuals on different days. The daily routine given to us was validated through our visualization which showed that employees made different stops such as to get coffee, go to the auto shop, and more before making their way to work at GAStech. Our visualization of Bertrand (#29) on January 11 validated that he was indeed cruising around town. Lastly, the suspected surveillance of executive homes was also validated by visualizing the data on January 9th and January 11th which showed cars switching shifts. In the future, we hope to have access to data that is more accurate to the map of Abila for our visualizations to look accurate and precise. As shown in the image above, the GPS data did not always align with the map which caused inaccuracies, however, despite our initial problems with our dataset and little experience with Tableau, we were able to visualize and carry out the tasks for this assignment.



References

"VAST Challenge 2014." VDAG, visualdata.wustl.edu/varepository/VAST%20Challenge%202014/challenges/MC2%20-%20Patterns%20of%20Life%20Analysis/entries/Central%20South%20University/.