



Visual Analytics Project

VAST Challenge 2021 - Mini-Challenge 2

Li Yumeng | Liang Yingfang | Shen Xiaojing

INTRODUCTION

Background

- This project is to do the data visualization on the Vast-challenge 2021 mini-challenge 2 to give some insights on some abnormal activities. The background of this challenge is that some employees from company GASTech located in Abila is missing and we need to find out the abnormality by analysis the credit card purchase history data and gps tracking data.
- The datasets given are the 2 weeks data before some employees' missing date. Including employees' credit card and loyalty card transaction records, gps records, and car assignment. The company cars which are approved for both personal and business use, and the vehicles are tracked periodically as long as they are moving.



Motivation

- Our research and development efforts were motivated by the general lack of effective and easy to use web-enabled geospatial analytics tool for discovering patterns in the map with gps data. It aims to provide car movement track with a dedicated date and car ID on a specific time or a time period.
- By using the free application R-shiny to build a free open-source tool without requiring users to code or understand R.

DESIGN FRAMEWORK

1. Application Flowchart

1 Location

- Popular Locations - Bar chart
- Popular Locations - Heatmap
- Location Sales - Boxplot

2 GPS Record

- GPS Record Timeline -Scatter plot
- GPS Record Path
- GPS Record Dot plot

2. Exploratory Spatial Data Analysis

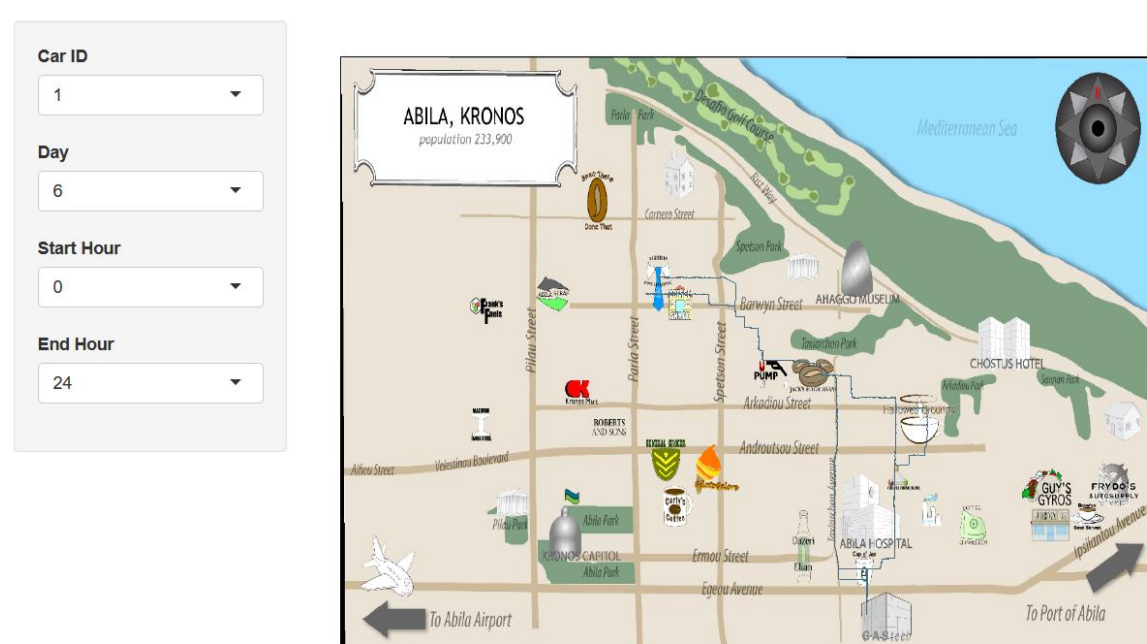


Figure 1.

GPS record path plot can work for visualization of path for different car ID, we can filter car ID and specific day, start hour and end hour. In this way we can see who hangout frequently and what specific time period and location they pass by.

3. Exploratory Data Analysis

Figure 2.

To investigate the most popular locations over different time period. Heat map by plotting locations at x-axis and hours of time stamp in the y-axis, coloured by count of transactions. Side bar is option to choose different date.

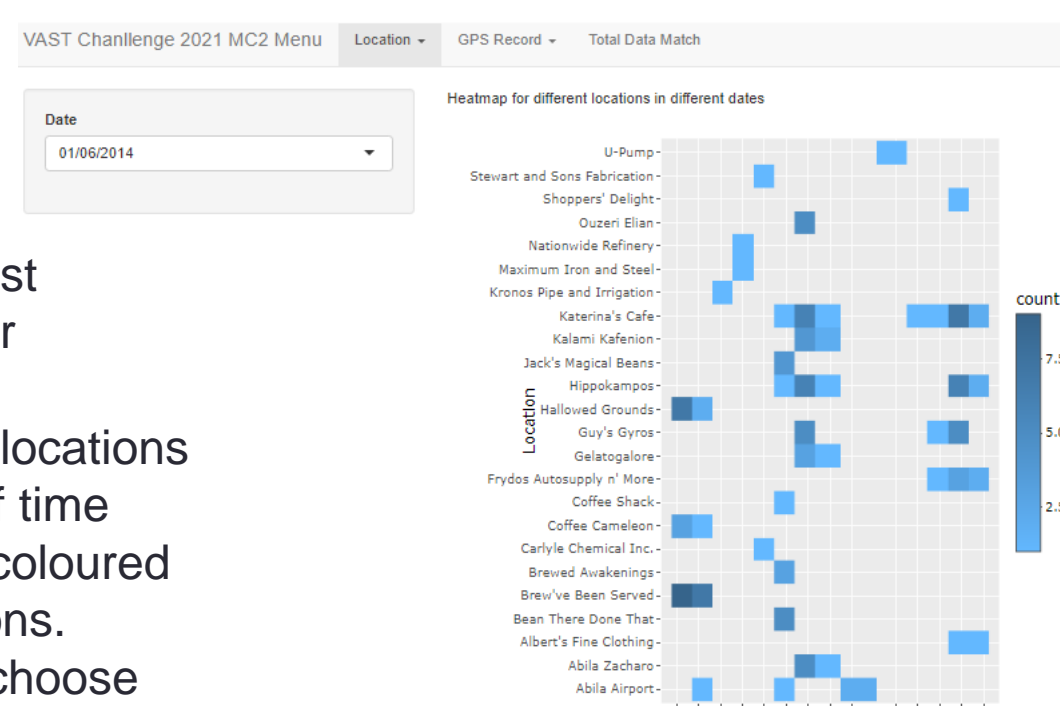


Figure 3.

In order to sort the location name by their sales respectively, we write a re-order function to sort the y-axis. The total number of transactions at different locations are displayed by a bar chart.

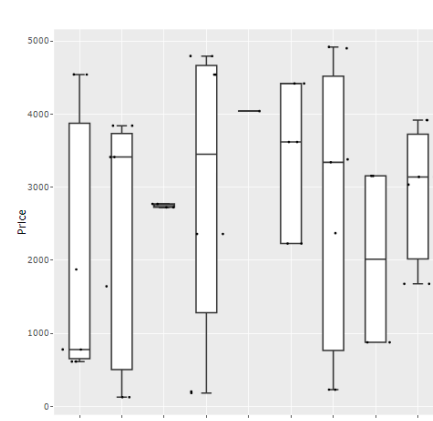
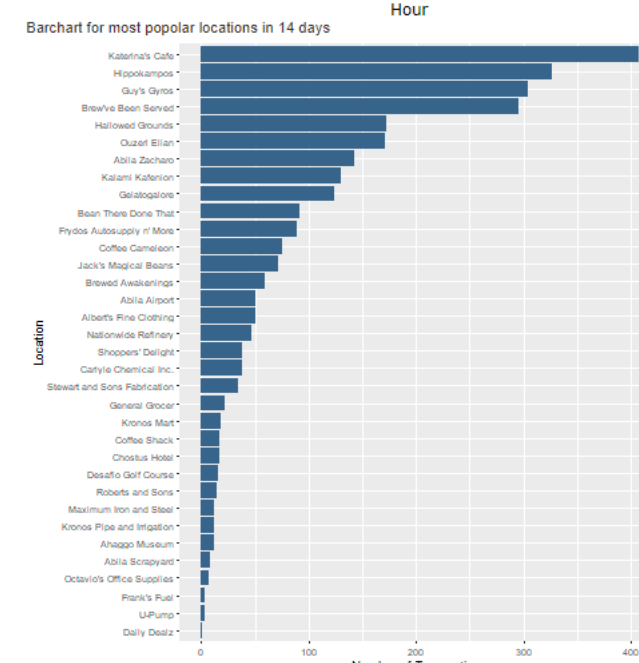


Figure 4.

A box plot is designed to show the distribution of price of each order.

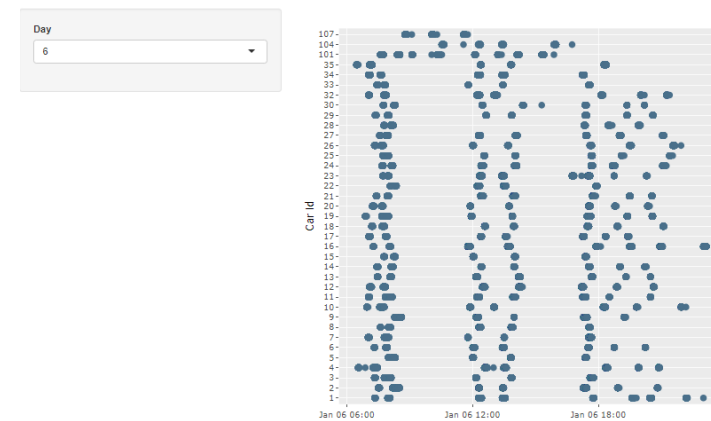
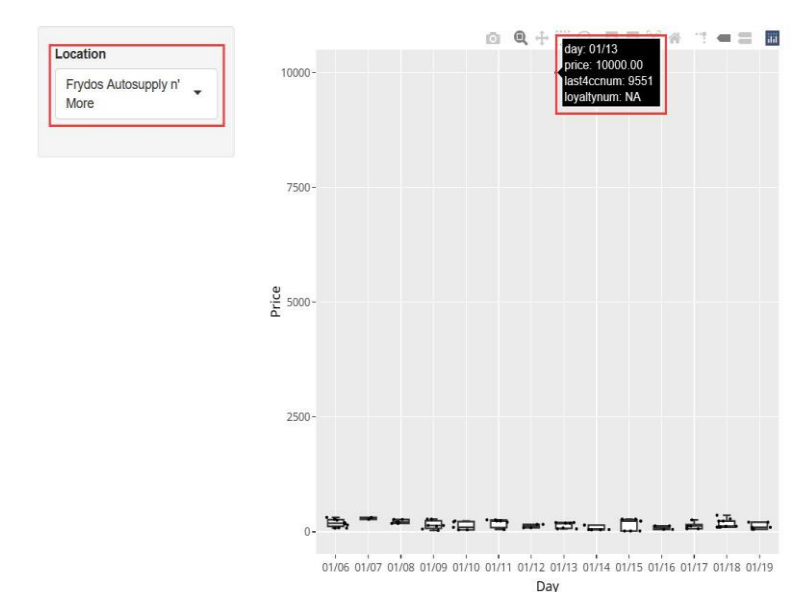


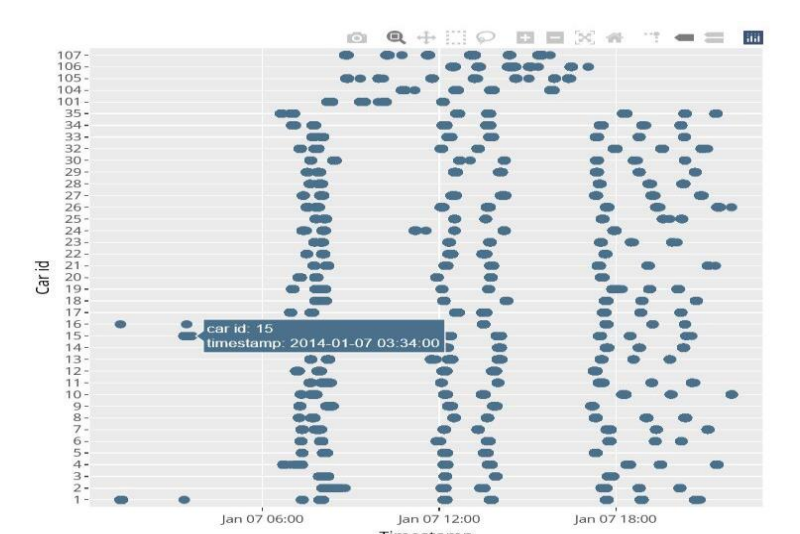
Figure 5.

A scatter plot is designed to show the car movement by different time daily.

4. Insights



Frydos Autosupply n' More has one very strange transaction in Jan 13. The average order price is around 200 to 300 in this location, but we can see the price of this suspicious transaction reached 10000, which is nearly 50 times of the average amount. If this is not caused by operating error, then that may involve money laundering for those kidnappers.



Nils Calixto (car id 1), Loreto Bodrogi (car id 15), Isia Vann (car id 16) and Minke Mies (car id 24) drives in the late night in Jan 7, 9, 11, 14, 16, nearly every other day. Coincidentally, Loreto, Isia and Minke are all work at the security department.



Elsa Orilla (car id 7) and Brand Tempestad (car id 33) may be in a relationship. They show up at Chostus Hotel together multiple times (in Jan 8, 10, 14, 17). According to the gps data, we can see that both of them drive their own car and most of the time they don't come from the same places, so it is more likely that they have an affair instead of being couples.

DISCUSSION

Discussion

RStudio is an Integrated Development Environment for R, a programming language for statistical computing and graphics. The application uses the free and open-source R language that offering a thriving programming environment for statistical and graphical analysis. In this data visualization challenge, we played as visual analytics experts to identify which GASTech employees made which purchases and identify suspicious patterns of behaviour. We learned to cope with uncertainties that result from missing, conflicting, and imperfect data to make recommendations for further investigation.



Future work

Regarding to the social relationship between employees, it would be easier to identify if we can visualize who are in the same department. Given the data sources provided, we should also identify potential informal or unofficial relationships among GASTech personnel if combined with timestamp because we can infer from their common activities during specific time periods.