

Shiny Investigative Tool into GASTech Personnel Disappearance

Aryah Umralkar Chopra
School of Computing and Information Systems,
Singapore Management University
aryahc.2020@mitb.smu.edu.sg

Rhoda Tong Min Ting
School of Computing and Information Systems,
Singapore Management University
rhoda.tong.2020@mitb.smu.edu.sg

Davmes Tan Chee Seng
School of Computing and Information Systems,
Singapore Management University
davmes.tan.2020@mitb.smu.edu.sg

ABSTRACT

A fictitious scenario was created as part of VAST Challenge 2021. A group of staff members from GASTech, an oil and gas company situated on an island known as Abila on Kronos, had gone missing mysteriously. A group known as Protectors of Kronos (POK) was the prime suspect into the disappearance. A series of unprocessed data were made available to the law enforcement agencies to investigate on. The data were split across three mini challenges, which our team had undertaken Mini Challenge 1 and 2.

Mini Challenge 1 consists of email correspondences, employee records and resumes, historical documents and news articles. The objective of this mini challenge is to identify the complex relationships among the people and organisations, and possibly infer the disappearance of GASTech to any individuals/group who might be involved. Mini Challenge 2 consists of gps tracking, transaction and loyalty card records, together with car assignment records - linked to the gps tracking data. The objective of this mini challenge is to discover anomalies and suspicious activities that may require additional investigating.

The objective of this research paper is to share on the methods and models used to develop an online investigative tool where a law enforcement at Kronos and Tethys could use, to piece the raw data into useable information and evidences.

1. INTRODUCTION

A fictitious scenario was created as part of VAST Challenge 2021. A group of staff members from GASTech, an oil and gas company situated on an island known as Abila on Kronos, had gone missing mysteriously. A group known as Protectors of Kronos (POK) was the prime suspect into the disappearance. While Mini Challenge 1 and 2 provided

a set of raw data that allow investigators to establish and identify complex relationships among the people and organisations, discover anomalies and suspicious activities, such investigative work may require humongous man hours and effort, without data analytics and visualisation.

We would be using Shiny R to develop an online investigative tool to aid in the analysis into the disappearance of GASTech Personnel, allowing investigators to explore information and inferential statistics derived from the unprocessed data available.

2. MOTIVATION

The motivation of this project would be two-fold. First, the data presented to the investigators were raw and unprocessed, and to link and derive insights from these data would require tremendous man hours and effort. Second, while insights could be derived and useful information could be formed, there would be a need to present the information in a visually appealing format to facilitate information dissemination and to allow quick collective appreciation of events among the investigators.

To this end, we would be looking to develop a R Shiny app based on three principles: (a) informative; (b) intuitive and; (c) interactive. Our 3Is principles would have the data undergoing baseline cleaning, making them into suitable formats for subsequent processing for information delivery. The user-interface would be made intuitive so that the investigator would be able to use the application without much references to our user guide. The online investigative tool would also be interactive, such that the investigator would be able to provide varied inputs into the formation towards the final visualisation report.

The R Shiny app would comprise of two main modules: (a) Exploratory Data Analysis allowing investigators to draw information such as transaction records, employee records, email correspondences and such; (b) Inferential Statistics allowing investigators to infer relationship linkages among user-selected employees, possible coded words within email correspondences within an identified group of personnel, their movements towards identified locations and possible anomalies at the locations and transaction analysis using both credit card and loyalty card data.

3. REVIEW AND CRITIC OF PAST WORKS

Taking reference on a submission by students from International Institute of Information Technology Hyderabad, they had derived an interface that uses geotools to show the paths of the cars, moving at a particular time or date. While it was a useful tool to visualise the movement of the vehicles, it was unable to plot the locations where the vehicles may had visited. This would be addressed in one of our module.

While credit card data and loyalty card data were provided for, to the investigators, there are no direct linkages between both set of data, except for the locations and price of item. The timestamp for both data are recorded differently with the credit card indicating the time of transaction while loyalty data indicating date of transactions only. Reviewing past work from members of University of Calgary, they attempted Parallel Coordinates plot by linking the locations, timestamp, price and employees, to establish possible linkages among the four variables. Inspired by their work and the sharing from Prof Kam on using Parallel Coordinates Plot on R, we would attempt to create parallel plots and allow the investigators to decide the variables that they are keen to establish linkages with.

4. DESIGN FRAME

Our R Shiny Investigative Tool's design would be based on our 3Is principle.

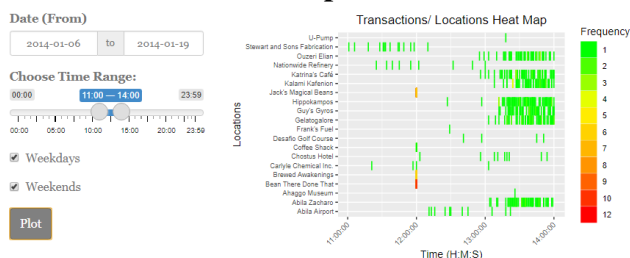
Informative - A series of informative modules would be made available under the Exploratory Data Analysis (EDA) tab, comprising of: (a) Locations Exploration; (b) Transactions Exploration; (c) Cards Exploration; (d) GASTech Employees Information; and (e) Email Correspondences. Details of each modules would be covered under the sub-sections later.

Intuitive - The entire UI design would be simple and intuitive, allowing investigators to use it without frequent references to our user guide. This would be done by ensuring that appropriate input methods were used and that the UI would be kept as clean and simple as possible.

Interactive - The respective modules would require specific inputs from the investigators before a representative visualisation could be produced. It would be made interactive such that the investigators would have the flexibility to select various variables or inputs to create the desired visualisation. Similarly, the investigators would have the option to save the produced visualisation separately.

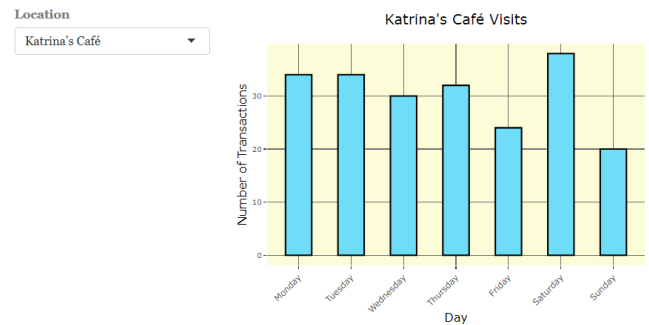
5. EXPLORATORY DATA ANALYSIS

5.1 Locations - For Exploration!



The intention of this module would be to allow the inves-

tigators to explore the popularity of the locations through the analysis of transaction records. It would be conducted in a manner by aggregating the transaction records based on the time of transactions. The transaction counts would be plotted against the aggregated time of transactions and locations using a heatmap to infer the peak periods for the locations. The selected records would be filtered based on the range of date inputs, time range of transactions, with the ability to toggle between weekdays, weekends or both.



The function of this module would be to identify the operating days of the locations through the analysis of credit card transactions, through the aggregation of records based on the day of transactions. The output of this visualisation would suggest the days where the locations would be having transactions, thus inferring their operating days of the week.

5.2 Transactions - For Exploration!

Duis nec purus sed neque porttitor tincidunt vitae quis augue. Donec porttitor aliquam ante, nec convallis nisl ornare eu. Morbi ut purus et justo commodo dignissim et nec nisl. Donec imperdiet tellus dolor, vel dignissim risus venenatis eu. Aliquam tempor imperdiet massa, nec fermentum tellus sollicitudin vulputate. Integer posuere porttitor pharetra. Praesent vehicula elementum diam a suscipit. Morbi viverra velit eget placerat pellentesque. Nunc congue augue non nisi ultrices tempor.

5.3 Individual Card Transactions

Duis nec purus sed neque porttitor tincidunt vitae quis augue. Donec porttitor aliquam ante, nec convallis nisl ornare eu. Morbi ut purus et justo commodo dignissim et nec nisl. Donec imperdiet tellus dolor, vel dignissim risus venenatis eu. Aliquam tempor imperdiet massa, nec fermentum tellus sollicitudin vulputate. Integer posuere porttitor pharetra. Praesent vehicula elementum diam a suscipit. Morbi viverra velit eget placerat pellentesque. Nunc congue augue non nisi ultrices tempor.

5.4 Employees of GasTech

Duis nec purus sed neque porttitor tincidunt vitae quis augue. Donec porttitor aliquam ante, nec convallis nisl ornare eu. Morbi ut purus et justo commodo dignissim et nec nisl. Donec imperdiet tellus dolor, vel dignissim risus venenatis eu. Aliquam tempor imperdiet massa, nec fermentum tellus sollicitudin vulputate. Integer posuere porttitor pharetra. Praesent vehicula elementum diam a suscipit. Morbi viverra velit eget placerat pellentesque. Nunc congue augue non nisi ultrices tempor.

5.5 Email Correspondence

Duis nec purus sed neque porttitor tincidunt vitae quis augue. Donec porttitor aliquam ante, nec convallis nisl ornare eu. Morbi ut purus et justo commodo dignissim et nec nisl. Donec imperdiet tellus dolor, vel dignissim risus venenatis eu. Aliquam tempor imperdiet massa, nec fermentum tellus sollicitudin vulputate. Integer posuere porttitor pharetra. Praesent vehicula elementum diam a suscipit. Morbi viverra velit eget placerat pellentesque. Nunc congue augue non nisi ultrices tempor.

6. INFERENCE STATISTICS

6.1 Email Network Analysis

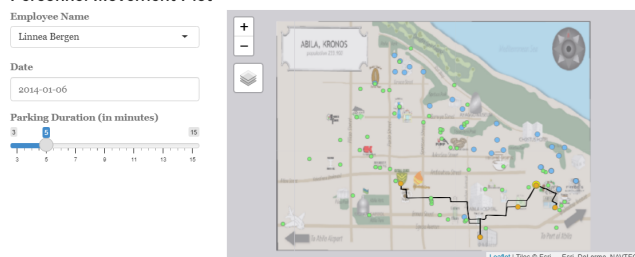
Duis nec purus sed neque porttitor tincidunt vitae quis augue. Donec porttitor aliquam ante, nec convallis nisl ornare eu. Morbi ut purus et justo commodo dignissim et nec nisl. Donec imperdiet tellus dolor, vel dignissim risus venenatis eu. Aliquam tempor imperdiet massa, nec fermentum tellus sollicitudin vulputate. Integer posuere porttitor pharetra. Praesent vehicula elementum diam a suscipit. Morbi viverra velit eget placerat pellentesque. Nunc congue augue non nisi ultrices tempor.

6.2 Networks

Duis nec purus sed neque porttitor tincidunt vitae quis augue. Donec porttitor aliquam ante, nec convallis nisl ornare eu. Morbi ut purus et justo commodo dignissim et nec nisl. Donec imperdiet tellus dolor, vel dignissim risus venenatis eu. Aliquam tempor imperdiet massa, nec fermentum tellus sollicitudin vulputate. Integer posuere porttitor pharetra. Praesent vehicula elementum diam a suscipit. Morbi viverra velit eget placerat pellentesque. Nunc congue augue non nisi ultrices tempor.

6.3 Employee Movement Plot

Personnel Movement Plot



Under this module, the investigators would be able to select a specific employee and to choose the date to determine the route of the employees. The list of locations and route would be plotted onto the map, and the duration of the parking could be refined based on the tolerance of the investigators to define how a location should be identified.



Under this sub-module, the locations that were visited by the employee would be made known, and listed in the order of sequence where he/she had visited. Possible location names may appear, if they had been identified prior, and the box plot of the duration spent at the said location, by employees, would be featured to determine whether there were any anomalies.

6.4 Transaction Amount Analysis

Duis nec purus sed neque porttitor tincidunt vitae quis augue. Donec porttitor aliquam ante, nec convallis nisl ornare eu. Morbi ut purus et justo commodo dignissim et nec nisl. Donec imperdiet tellus dolor, vel dignissim risus venenatis eu. Aliquam tempor imperdiet massa, nec fermentum tellus sollicitudin vulputate. Integer posuere porttitor pharetra. Praesent vehicula elementum diam a suscipit. Morbi viverra velit eget placerat pellentesque. Nunc congue augue non nisi ultrices tempor.

7. FUTURE WORKS

(Additional point from Nikki on how the modules could be improved.)

(Additional point from Rhoda on how the modules could be improved.)

An improvement towards the Employee Movement Plot could include further analysis on the movement route, in terms of the distance/ speed, which may suggest the vehicle were kept in idle mode, which is currently unable to detected under the current module. In addition, the streets of Abila could be identified based on the movement routes, and thus, and analysis could be conducted whether the vehicle was travelling in a suspicious manner.

8. ACKNOWLEDGEMENTS

The authors wish to thank Professor Kam Tin Seong of Singapore Management University for his extensive guidance and support during this project.

References

- [1] Kam, Tin Seong 2020. Hands-On Exercise 8: Creating Parallel Coordinate Plot. <https://rpubs.com/tskam/PCP>.

- [2] Pochampally, Yashaswi; Yarrabelly, Navya; Chikka, Veera Raghavendra and Karlapalem, Karlapalem 2014. VAST Challenge 2014 Mini Challenge 2. <http://visualdata.wustl.edu/varepository/VAST%20Challenge%202014/challenges/MC2%20-%20Patterns%20of%20Life%20Analysis/entries/International%20Institute%20of%20Information%20Technology%20Hyderabad/>.
- [3] Sahaf, Zahra et al. 2014. VAST Challenge 2014 Mini Challenge 2. <http://visualdata.wustl.edu/varepository/VAST%20Challenge%202014/challenges/MC2%20-%20Patterns%20of%20Life%20Analysis/entries/University%20of%20Calgary/>.