Short Paper

Alice Anonymous
Some Institute of Technology
alice@example.com

Bob Security
Another University
bob@example.com

ABSTRACT
XXXXXXXXXXXXXXXXXXX

1.   1.INTRODUCTION

2.   2.MOTIVATION OF THE APPLICATION

Nullam semper imperdiet orci, at lacinia est aliquet et. Sed justo nibh, aliquet et velit at, pharetra consequat velit. Nullam nec ligula sagittis, adipiscing nisl sed, varius massa. Mauris quam ante, aliquet a nunc et, faucibus imperdiet libero. Suspendisse odio tortor, bibendum vel semper sit amet, euismod ac ante. Nunc nec dignissim turpis, ac blandit massa. Donec auctor massa ac vestibulum aliquam. Fusce auctor dictum lobortis. Vivamus tortor augue, convallis quis augue sit amet, laoreet tristique quam. Donec id volutpat orci. Suspendisse at mi vel elit accumsan porta ac ut diam. Nulla ut dapibus quam.

Sed est odio, ornare in rutrum et, dapibus in urna. Suspendisse varius massa in ipsum placerat, quis tristique magna consequat. Suspendisse non convallis augue. Quisque fermentum justo et lorem volutpat euismod. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Morbi sagittis interdum justo, eu consequat nisi convallis in. Sed tincidunt risus id lacinia ultrices. Phasellus ac ligula sed mi mattis lacinia ac non felis. Etiam at dui tellus.

3.   3.REVIEW AND CRITIC ON PAST WORKS

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.

4.   4.DESIGN FRAMEWORK

4.1   4.1 Data

The data we used is from the VAST 2021, which contains four main datasets, one tourist map and a geospatial folder: * car-assignments.csv: A list of vehicle assignments by employee, which including Employee Last Name, Employee First Name, Car ID, Current Employment Type and Current Employment Title * gps.csv: A file of vehicle tracking data, which including Timestamp, Car ID, Latitude and Longitude * loyalty_data.csv: A file containing loyalty card transaction data, which including Timestamp. Location, Price and Loyalty Number * cc_data.csv: A file containing credit and debit card transaction data, which including Timestamp, Location, Price and Last 4 digits of the credit or debit card number * MC2-Tourist.jpg: A tourist map of Abila with lobations of interest identified * Geospatial folder: ESRI shape files of Abila and Kronos The typical data analysis workflow process of data preparation, data visualization, insights discovery, Shiny architecture and build the Shiny application. The application includes visual analysis methods and geo-spatial analytical methods to uncover the anomalies we found. For the data preparation, we first changed the timestamp into useful format like DD/MM/YY HH/MM, and then separated it into date, day and hour in order to visualize it easily. We also inner-joined some data files (e.g., cc_data and loyalty_data) to find some anomalies.

4.2   4.2 Shiny Architecture

The Shiny application was built by using a free and open-source software —- R language. The method selected refers to the papers from City University London, Central South University as well as the submission conducted by Kevin Griffin of University of California, Davis. We used Shiny to webify our results and providing options and visualisations for interactivity. It consists of 4 major views: [XXXXXXXXXXXXXXX and 6 components (Bar charts, heatmap, boxplot, bipartite graph, map and data table) were used to visualize the results.

4.3   4.3 Analysis technology

4.3.1   4.3.1 Bar Chart

The bar chart showed the popularity of each location by the number of times people visit it. We separated the location into 5 categories (Business, Dining, Living, Unknown and NA), it's easily to show which kind of location is the most popular. In addition, user can click "Show the data" on the

sidebar to see details or just put their mouse on the bar because the chart is interactive.

4.3.2 4.3.2 Heatmap

There were 3 types of heatmap showed the popularity of the locations by day, hour and people. An interactive sidebar with radio button is provided on the left for user to select.

4.3.3 4.3.3 Boxplot

To explore the insights about transaction, we prefer to use boxplot, which graphically depicting groups of numerical data through their quartiles. Box plots may also have lines extending from the boxes (whiskers) indicating variability outside the upper and lower quartiles, in this case, these outliers would help us to identify the suspicious transactions.

4.3.4 4.3.4 Bipartite graph

When modelling relations between two different classes of objects, bipartite graphs very often arise naturally. Hence, we used bipartite graph to check whether there is cross-used between credit card and loyalty card.

4.3.5 Map

The fifth visual component which is also the topmost of the application is the map. We used maps extensively to uncover all suspicious relationships and movements. In this view, some movements data records will be represented as point symbols and others will be shown as route.

5. CASE STUDY

When analyzing the popularity of locations, the first thing to do is to look at the number of times people visit it (Fig.1). "Katerina's Cafe," "Hippokampos," "Guy's Gyros," "Hallowed Grounds" and "Brew've Been Served" are the TOP 5 popular places among employees. Then we performed heatmaps to show the visit frequency both by date and by hour (Fig 2). Some obvious patterns appear which is F&B services busy during mornings, lunch times and evenings. Katerina's Café and Hippokampos were popular almost every day while people went to Guy's Gyros, Hallowed Grounds and Brew've Been Served only on weekdays. When combining the data of credit card and loyalty card, we found there is a gap, credit card data has total 1490 records while loyalty card data includes 1392 records. Hence, we suspected that some loyalty cards were cross used.

5.1 5.1 Most popular Location in Kronos

5.2 5.2

5.3 5.3

5.4 5.4

5.5 5.5

6. DISCUSSION

The application was developed using the Shiny architecture on R, an interactive web-based application to promote visualization analytics on VAST 2021 mini challenge-2 dataset. The interactivity and functionality of our application provides evidence on the robustness of R shiny as a framework to develop web application, along with the variety of available R packages that serve as building blocks for each module in our application. This project is mainly focus on visual analytics. The various components of application were developed with interactive visualisations. Compared with other visualization tool such as Tableau, the R shiny application shows it application can provide the right balance between friendly user-interface and depth of analysis.

7. FUTURE WORK

Project Shiny App was built in relation to VAST 2021 mini challenge-2 dataset as a usecase. The Shiny App enable users to perform exploratory using several visualization methods without users needing extensive programming or statistical knowledge. The application could be further enhanced by including a data load and wrangling function to accommodate different datasets. The current types of charts are limited to only 4 types of charts. Other charts, such as network and bar charts, can be incorporated further. The user case scenario set in the fictional region. Hence, further interactivity choices of geographic area such as county, neighbourhood can be added if using the real gps dataset. Lastly, the shiny app only focuses on visualization analytics, further analysis method such as clustering analysis can be incorporated further.

8. ACKNOWLEDGEMENT

References

[1] Fenner, M. 2012. One-click science marketing. Nature Materials. 11, 4 (Mar. 2012), 261–263.

[2] Meier, R. 2012. Professinal Android 4 Application Development. John Wiley & Sons, Inc.