Visualizing Singapore's Procurement Spending

Sheryl CHONG Man Er, KUNG Jung Wen, NGUYEN Dang Thanh Ha

Abstract— As part of the Smart Nation initiative in Singapore, public data of Singapore is made available by Singapore government's open data portal for the nation and decision makers to monitor and gain insights about Singapore's social and economic trends. However, a proper tool to visualize these data is lacking. To address this need, we designed and developed GeViz, an interactive visual analytics dashboard for the public and ministries of Singapore to uncover insights on Singapore government's procurement data, with the aim to identify spending patterns based on procurement categories, goods and services procured and ministry and suppliers relationships. We leveraged on machine learning to classify procurement categories based on the tender description and adopted various visualization techniques to explore and analyse Singapore government's procurement patterns.

Index Terms— . GeBiz, Procurement, Tenders, Suppliers, Visual Analytics

1 Introduction

GeBIZ is the Singapore Government's one-stop e-procurement portal which facilitates all tender activities between Singapore government and local and overseas suppliers since 2000. All the public sector's invitations for quotations and tenders are posted on GeBIZ hence suppliers can search for government procurement opportunities and submit their bids online. [1] There is more than S$10 billion worth of procurement opportunities posted annually to about 30,000 registered suppliers. [2]

For the past few years since its launch, GeBiz has collected a huge volume of procurement data which is a valuable source of data to understand the procurement needs and spending of the government.

2 Motivation & objectives

Our research and development efforts were motivated by the lack of available tool online to aid public and ministries to gain insights on GeBiz data. Through our project, we aim to provide public and ministries with an interactive visualization tool that allows the user to explore and discover the Singapore Government's procurement spending patterns. More specifically we aim to provide the user with the ability to:

1.Gain an overview of procurement spending made by each ministry and agency

2.Identify the relationships between ministries, agencies and suppliers

3.Identify what are the goods and services procured by ministries and agencies under each category

- *Sheryl CHONG Man Er is an undergraduate student at the School of Information Systems, Singapore Management University, E-Mail: mechong.2015@sis.smu.edu.sg*
- *KUNG Jung Wen is an undergraduate student at the School of Accountancy, Singapore Management University, E-Mail: jwkung.2015@accountancy.smu.edu.sg*
- *NGUYEN Dang Thanh Ha is an undergraduate student at the School of Information Systems, Singapore Management University, E-Mail: dthnguyen.2015@sis.smu.edu.sg*

3 The approach

3.1 Data Collection

1. Government Procurement via GeBIZ dataset that is available in www.data.gov.sg

2. Ministry – Agency data from Singapore Government Directory

3. Past procurement data on GeBIZ website

3.2 Exploratory Data Analytics (EDA)

Before performing any visualizations on the procurement dataset, exploratory data analysis (EDA) is applied to gain a summarization on the dataset, its features and whether any necessary assumptions are required. In this project, EDA is performed using Tableau and Python. There are some problems in the original dataset that we spotted out during the EDA process that are mentioned in more details below.

3.2.1 Missing Ministry Feature

There are Agency and Supplier features; however, Ministry feature is missing. This means that we cannot perform analysis on Ministry level.

3.2.2 Missing Procurement Category

The procurement category is not available, which means we cannot perform analysis on the Procurement Category level and conclude in which areas the government spend on.

3.2.3 Inconsistent Agency Naming Convention

Some Agency names include the corresponding Ministry name inside. This is not a major data issue, but it affects the naming consistency and may confuse the users.
The solutions to these three problems are mentioned in the following two sections on data cleaning and feature creation.

3.3 Text Classification

Since we are given a training dataset that includes the tender description and the procurement category, this is a supervised multi-class text classification problem and the goal is to find out the most suitable classification algorithm for this dataset.

Given a new tender description, the classifier categorizes the procurement into one and only one procurement category.
We perform the classification using 4 different classification algorithms that are:

1. Linear Support Vector Machine
2. Logistic Regression
3. Multinomial Naïve Bayes
4. Random Forrest

3.4 Data Transformation On Training Dataset

3.4.1 Encoding Categorical Data

The Category column is encoded into integers because only a few algorithms can work with categorical data directly, the others can only accept numerical input and produce numerical output.

3.4.2 Handling Imbalanced Categories

The distribution of procurements per category is imbalanced as some categories have significantly more procurements, such as Services, compared to others. Even though this will lead to prediction bias towards the majority classes, we do not need to do re-sampling in this specific case because we are more interested in categories with more procurements. Re-sampling will be more suitable for situations such as cancer detection where the minority or outliers are of greater interest.

3.4.3 Term Frequency – Inverse Document Frequency

Many machine learning algorithms can only process a fixed-length input; therefore, it is necessary to do feature encoding on the Tender Description to convert the texts to vectors of numbers. In our case, we will use Term Frequency – Inverse Document Frequency (tf-idf) for feature encoding. The output tf-idf weight of each Tender Description represents the importance of words in a corpus and ignores the order in which they appear.

3.4.4 Classifier Training

After we have the tf-idf vector representations of the Tender Description, we will start training the four abovementioned classifiers using Scikit-learn library in Python. Below is the benchmarking result for each classifier:

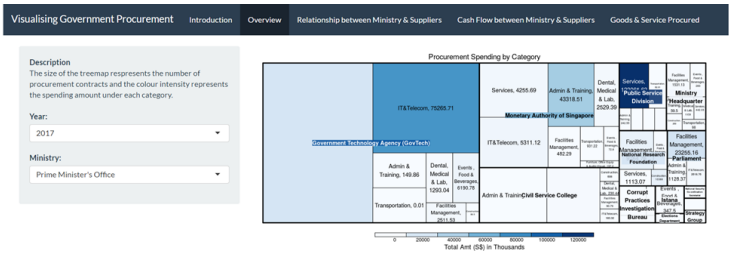
```
model_name
LinearSVC                0.898813
LogisticRegression       0.687043
MultinomialNB            0.591986
RandomForestClassifier   0.272587
```

Linear Support Vector Machine has the best performance with nearly 90% accuracy compared to the rest. The reason why linear SVM is chosen over non-linear SVM is due to the fact that when number of features is large, non-linear data mapping to higher dimensional space does not improve the performance; therefore, linear kernel is good enough, according to A Practical Guide to Support Vector Classification [3].

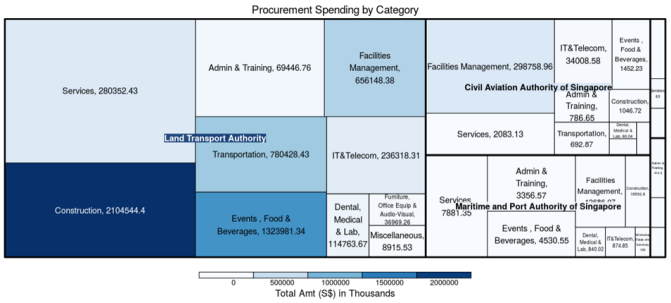
4 User Interface Design

Before we started on developing our application, we first researched various visualization libraries to find out the desired visualization charts to present the data. Eventually, we decided to use R libraries due to its ease and flexibility in creating visualizations. The following descriptions show the design considerations in the visualization tool:

a) Overview of Visualization
Our visualization tool created by R Shiny library is designed to fit any browser size. By removing the need to scroll, this enhances readability and user experience.

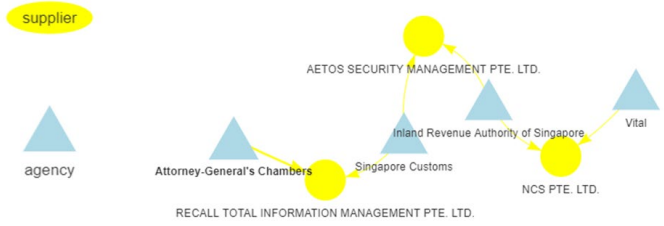


b) Treemap Diagram



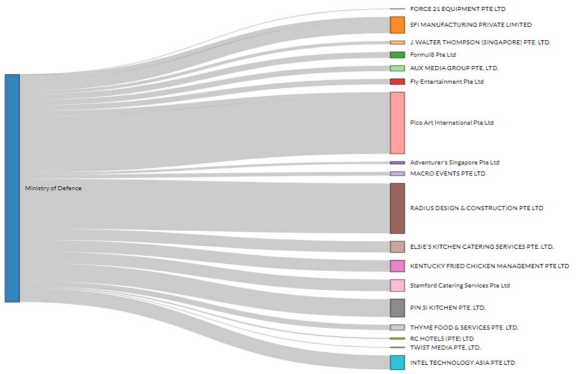
The treemap created by treemap library [4] to show the spending breakdown for each category of all agencies under the selected ministry. The size of the box represents the total amount of procurement contracts while the colour intensity represents the total amount of good and services procured.

c) Network Diagram



Network diagram created by VisNetwork library to show the relationship of agencies and suppliers of the selected ministry [5]. The triangle icon represents the agency while the circle icon represents the supplier.

d) Sankey Diagram



Sankey diagram created by NetworkD3 library to show the cash flow between the selected agency and suppliers for the selected category. The thickness of the path represented the total dollar amount of the goods and services procured from the supplier.

e) Word Cloud

Word cloud created by wordcloud2 library to show an overview of the tender description for the selected agency and selected category. The higher of the frequency of the words appeared in the procurement description, the bigger the word will appear in the word cloud.

f) Description Table


Show entries Search:

	ministry	agency	tender_description	awarded_amt
14	Prime Minister's Office	Civil Service College	Invitation to Tender for Design and Administration of Perception Surveys 2017/2018 for Civil Service College	138800
1	Prime Minister's Office	Civil Service College	Invitation to Tender for Provision of Full Design & Production Services for the Journal "Ethos" for Civil Service College	74410
2	Prime Minister's Office	Civil Service College	Invitation to Tender for Research Study on Expectations of Public Service's Delivery for Civil Service College	47016

Showing 1 to 3 of 14 entries Previous 1 2 3 4 5 Next

The description table is placed under the word cloud to allow the user to search using the keywords or phase for the exact description, as well as, to view to the top tenders by awarded amount by clicking on the awarded amount column name.

g) Filters

We have included dropbox filters which allow the user to filter the visualization based on the year, ministry, agency and category of their choice.

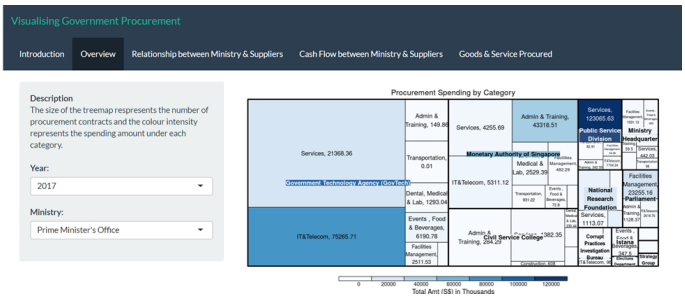
In addition, there is awarded amount range filter to allow the user to filter the visualization according to the awarded amount.

5 Web Application Architecture

This visualization tool is designed and developed using R programming language due to its ease and flexibility in creating and customising charts and integrating them. Our code uses R Shiny library to build the dashboard, dplyr and tidyverse libraries for data manipulation, ggplot2 libraries for visualization and lubridate library to manipulate the date related data. Besides that, we also used NetworkD3 library to build our Sankey diagram [6], WordCloud2 library to build the word cloud [7] and VisNetwork library to build our network diagram[5]. Our application is then deployed to Shinyapp.io which is a shiny server which allows users to deploy their shiny application online. Therefore, users are able to view our visualization tools online using their preferred browser.

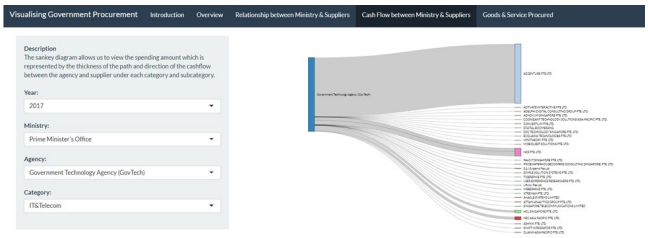
6 Case Study

To demonstrate how our application can be used, we will be examining the procurement spending and needs of Prime Minister's Office (PMO) in 2017.

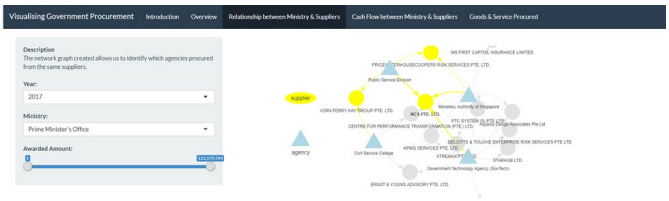


The treemap overview allows easy comparison of each agency spending within the PMO on each category by the quantity of procurement and total dollar amount of goods and services procured. We are able to gain insights on what each agency in the PMO has procured for. From this, taking Government Technology Agency (GovTech) as an example, GovTech has procured from categories which are Services, IT & Telecommunication, Administration &Training, Transportation, Dental, Medical & Lab, Events, Food & Beverages and Facilities Management. Within GovTech, they have procured most for Service category and the highest total dollar amount of goods and services is IT & Telecommunication.

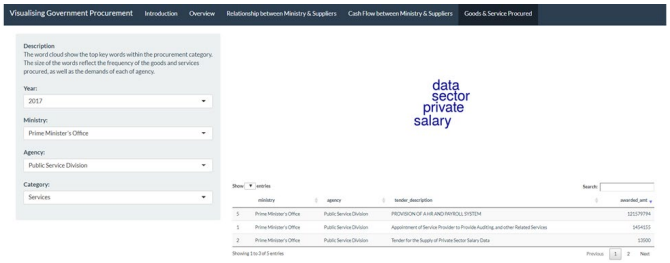
As a whole ministry, we also can infer that the highest total dollar amount of goods and services was procured by the Public Service Division for Service Category ($123 Million) while the highest number of procurement contracts are procured from GovTech.



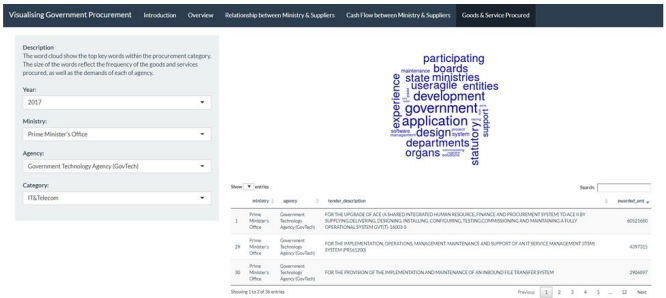
Under the Cash Flow between Ministry and Suppliers tab, the top suppliers of GovTech's IT& Telecommunication Services by dollar amount are by Accenture Pte Ltd and followed by NCS Pte Ltd.



NCS was also a common supplier of MAS, Public Service Division and GovTech in 2017, this finding corresponds to the fact that NCS is the primary ICT solution provider of Singapore Government.



With the identification of procurement categories and agency of interest, the word cloud under Goods and Service tab drills down further to identify the top keywords which appeared in the tender description. Private, sector, salary and data had been picked up by the word cloud as the agency procured for Supply of Private Sector Salary Data three times within 2017. With the sort function of the table, the top procurement item under Services is Provision of HR and Payroll System, amounting to S$ 121.5 million.



Whereas for top spending category of GovTech – IT&Telecommunication, the top goods and services procured by frequency are agile application development and user experience design services, software solutions and maintenance system for various statutory boards. The top procurement spending within this category was the upgrade of ACE (A shared integrated human resource, finance and procurement system) which amounted to S$60.5 million.

7 Conclusion

Through this project, we hope to provide a visualization tool for visualizing procurement spending made by Singapore Government. Our visualization tool enables public and ministries to explore and discover insights in the procurements made by respective agencies for different categories. Through our visualization, we hope that the public will be able to better understand government's procurement spending, while individual ministries will be able to detect useful procurement patterns to better guide them in making decisions for future purchases. This application mainly serves as an exploratory tool and will, therefore, requires additional qualitative research to view and understand the exact details and reason for each procurement transaction.

Acknowledgments

The authors wish to thank Professor Kam Tin Seong for his support and guidance throughout this project.

References

[1] GeBiz. (n.d.). About Us. About Us. Retrieved from https://www.gebiz.gov.sg/about-us.html

[2] Carol Lo, C. P. (n.d.). GeBiz - From Vision to Reality. Retrieved from https://www.dsta.gov.sg/docs/default-source/dsta-about/dh02200602-gebiz-from-vision-to-reality.pdf?sfvrsn=2

[3] Hsu, C.-W., Chang, C.-C., & Lin, C.-J. (2016, May 19). A Practical Guide to Support Vector Classification.

[4] Kopp, B. (n.d.). Creating a Treemap in R. Retrieved from https://rpubs.com/brandonkopp/creating-a-treemap-in-r

[5] Thieurmel, B. (n.d.). Introduction to visNetwork. Retrieved from https://cran.r-project.org/web/packages/visNetwork/vignettes/Introduction-to-visNetwork.html

[6] Bock, T. (n.d.). Creating Custom Sankey Diagrams Using R. Retrieved from https://www.displayr.com/sankey-diagrams-r/

[7] HOLTZ. (n.d.). THE WORDCLOUD2 LIBRARY. Retrieved from https://www.r-graph-gallery.com/the-wordcloud2-library/