Mini-project

Visualization in Tableau

CS 5346– Information Visualization

GENERAL INSTRUCTIONS

1. *Deadline:* **Monday 2 April, 10 AM.**
2. *Deliverable:* Submit your tableau workbook, all the datasets, and a short report (see template at Fig.1) in a zip file in IVLE folder **MP-Tableau**. Rename your zip file as {*your name*}_{*metric no.*}_{*partners name*}_{*partners metric no.*}, omitting the brackets.
3. *Assignment Type:* Pair
4. *Marks:* 15
5. *Assessment Criteria:* You shall get 60% for completing the mini-project successfully. That is, If it is timely, correct (truly represent the insight that you claim you have derived) and self-explanatory (As in the popular saying - A picture tells thousands words). You can expect to get at least 80% only if you are creative — meaning,
   - you made good use of supplementary dataset(s) from other sources.
   - you found novel analytic questions using those datasets.
   - you found novel insights using those datasets.

1 Overview

In this mini-project, you will use *Tableau* to analyze historical data of NEA (National Environment Agency) licensed eating establishments in Singapore. The data has been collected within the period of June 11, 2015 and September 6, 2016.

Tasks:

- **Tableau worksheets:** Prepare *worksheets* of visualizations with sufficient descriptive texts.
- **Demo:** Produce a *tableau story* using the series of visualizations (worksheets) with sufficient descriptive text to make a convincing argument. You will be asked to present your story in the class. Your story should be published in your Tableau public profile so that you can present it from the browser. Tentative date for presentation is **April 5** (subject to change depending on the availability of the tutors).
- **Short Report:** Submit a short report on your project following the template of Fig. 1.
- **Submission:** Finally, you are to submit your *workbook*, all the datasets and report in IVLE.

2 Link to the Dataset:

List of NEA Licensed Eating Establishments
This dataset is the modified version of this dataset. Visit here to learn more about the fields. You can also learn about the fields from the metadata shared in the google drive link.

3 Project Ideas

- Unlike the assignment on Tableau, in this project, you are strongly encouraged to supplement this dataset (NEA) with other publicly curated datasets (population data, for instance). However, the supplementary datasets must be used to facilitate additional insights from NEA, since NEA is the dataset of interest here.

Things to Include in your write-up	Details	Notes/Suggestions
Your Name, Metric No.		
Your Partners Name, Metric No.		
Introduction/Objective of this mini-project	Write a paragraph about the objectives of this mini-project in your own words. Mention individual contribution of each member.	Be precise and succinct.
Brief Description of the Supplementary Datasets (1-2 Paragraph/ dataset)	Each description must include the source of your data, description of the column fields, why you are using this dataset and how you integrate/combine this data with the original dataset.	Be precise and succinct.
Data Preprocessing/Cleaning:	Describe why and how you did preprocessing, cleaning of the datasets, if you had to do so	Be precise and succinct.
Analytic Questions Pursued and their outcomes (1 paragraph/question)	Describe the questions you answered. Describe what insight you gained from the visualization.	Do not describe the components of the visualization itself (e.g. In the row X, in the column Y). You might want to attach the picture of your visualization (if necessary). Be precise and succinct.

Figure 1: Template for the Short Report

- There is no limit on the number of worksheets or dashboard for the mini-project. However, since there will be only one story, we expect that you will create just enough worksheets required for the story.
- The specifics of the analytic question is open-ended. However, we expect you to cover most of the following categories -
  - Spatial aspects (involving map visualization), for instance, 'which regions of Singapore has Best/worst eating establishments? How the regions can be compared based on that criterion?'
  - Quantitative aspects, for instance, 'How the individual establishments compare in terms of their *suspension period* or *demerits points*'
  - Statistical aspects (e.g. correlation, clusters), for instance, 'Does *suspension period* has any correlation with *the number of suspensions*? Is there any correlation between the *population* in a region

with the *number of eating establishments* in that region?'

- Temporal aspects, for instance, When the establishments are typically suspended?

- Tips: **Be creative** with the data. Mine some **striking** insights. **Use** the flexibility of having supplementary datasets.

4 Queries

Send your queries to CS5346.tutor@gmail.com