FAANG and Microsoft Employee Reviews

Semester 1, 2019

FIT5147 - Data Exploration and Visualisation

Tutor: Xue Sun

Student Name: Xue Chen

Student ID: 30100763

Introduction

In these years, Most of the software developers consider Facebook, Apple, Amazon, Netflix, Google and Microsoft as their dream companies to work. However, after you receive the offer, the size of the salary package is not the only thing you need to thinking about. This project tries to help the job seeker getting an insight of which tech company is the best to work for.

The intent audience will be the job seekers who is looking for a position in these 6 tech companies.

Design

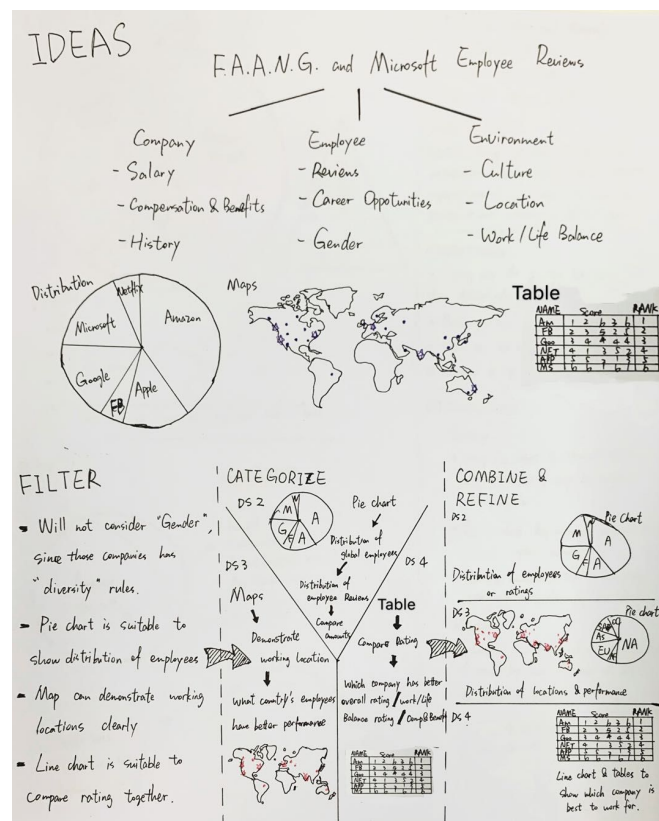The project will mainly focused on the following five design sheets.

Figure 1.1 Design Sheet - 1

The first sheet is considering what information is going to present and used in the analysis. For the brainstorming part, the idea came out into three aspects. The first part is "Company", I will not consider the "Salary" because there are different kinds of positions in these companies, e.g. there might some Amazon employees from fulfillment department which is included in the dataset. Also "History" will not be considered. The second part is "Employee", which will including "Reviews" and "Career Opportunities". "Gender" will not

be considered, since these tech companies normally has diversity rules. Finally, "Culture", "Location" and "Work/Life Balance" will be included in "Environment" part. The project will use pie chart for the distribution of employees, maps for the location, table for the rating and wordcloud for the textual reviews.
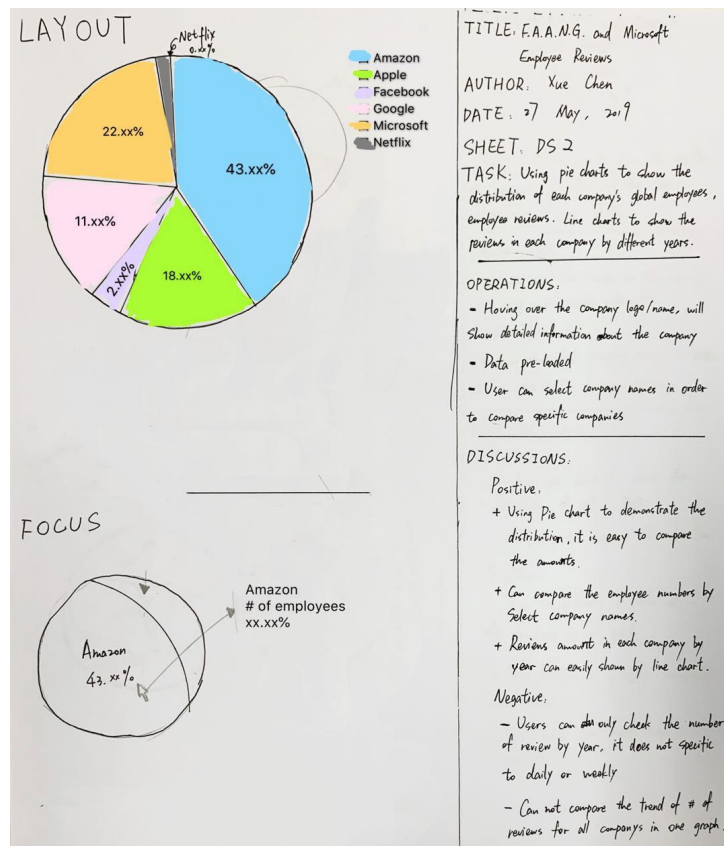

Figure 1.2 Design Sheet - 2

The design sheet 2 is focused on the distribution of the employees from different company. Pie chart will be used to visualize the data.
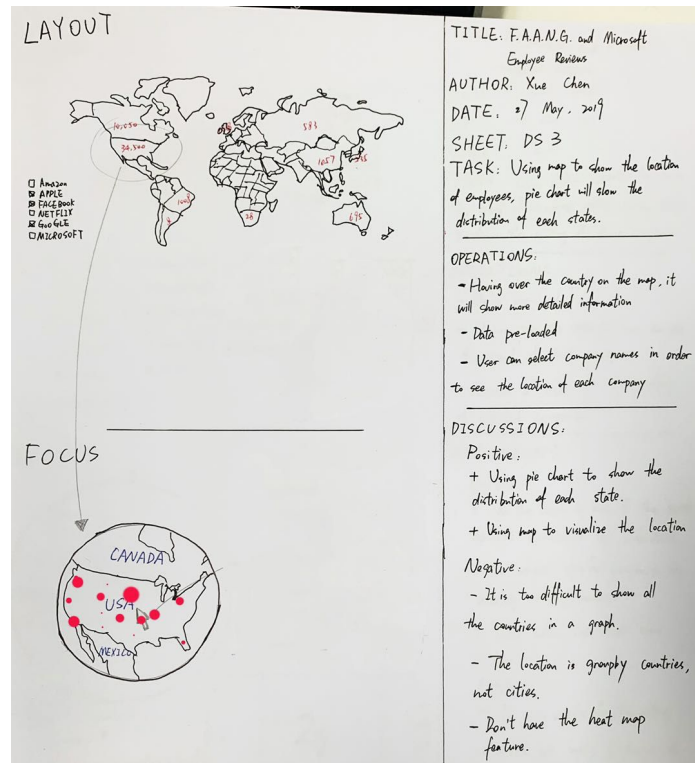
Figure 1.3 Design Sheet - 3

Maps can show the working location of employees. The radius of the circles on the map refers to the amount of employees in this location.
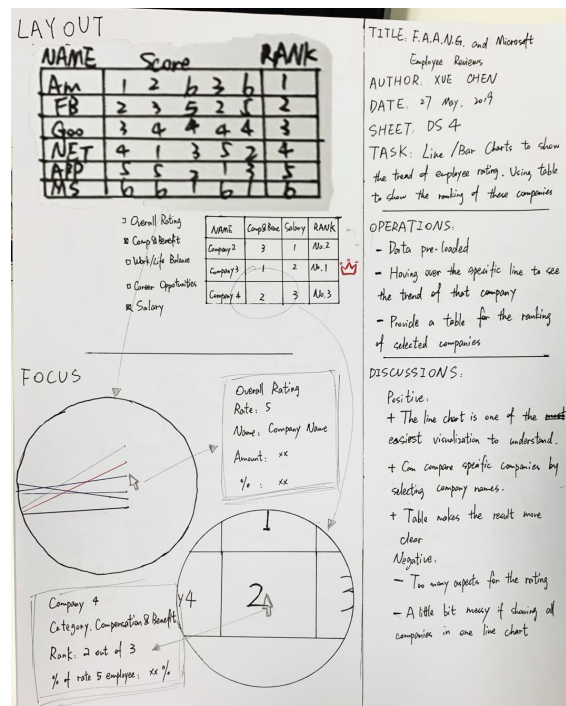

Figure 1.4 Design Sheet - 4

According to design sheet 4, a table including the rank of the companies will be generated.
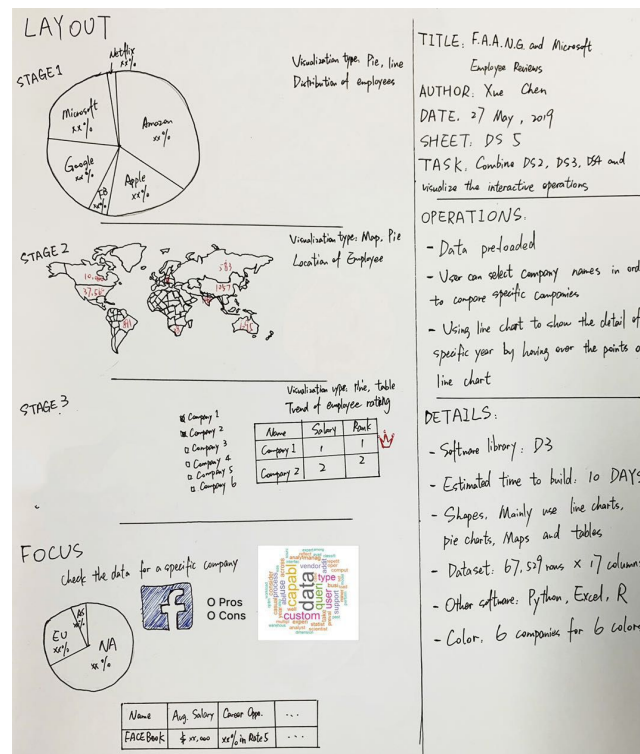
Figure 1.5 Design Sheet - 5

As the final design sheet, the project will mainly use pie chart, map and table to do the analysis. More importantly, user can check the result for a specific company with a wordcloud plot for the textual reviews. Such as the example shown on the sheet.

Compare to the original 5 design sheets, I did some justification, such as adding the wordcloud, change the world map to the US region only.

Implementation

The visualization is implemented by R-Shiny.

Library: shiny, tm, SnowballC, RColorBrewer, wordcloud, markdown, memoise, leaflet, dplyr.

The analysis will become very understandable if the plots can do interaction with users. The list of libraries used above most of them are for wordcloud, some are for the map, some are for the interactive plots.

User guide

For the "Distribution" page, user can select company names in order to do the comparetions.
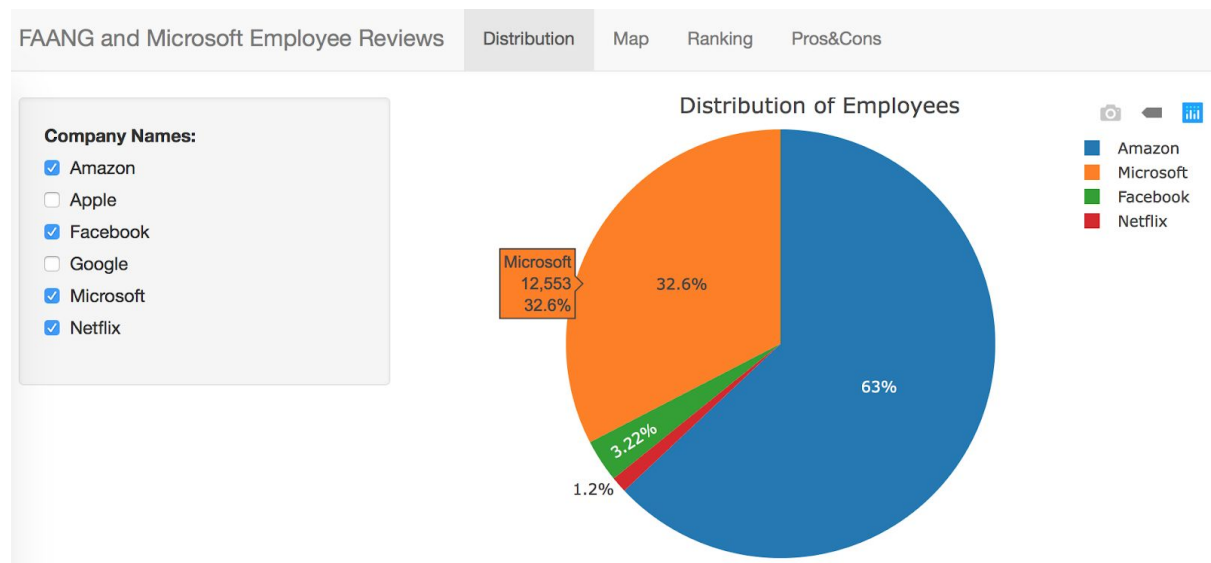

Figure 2.1 Distribution of Employees

By hovering over on the pie chart, detailed information will be shown(Figure 2.1).

The next page(Figure 2.2) shows the working location of each company. The radius of red circles on the map represent the amount of the employees in each location. User can select the company name on the left and click the circles for the names of cities.
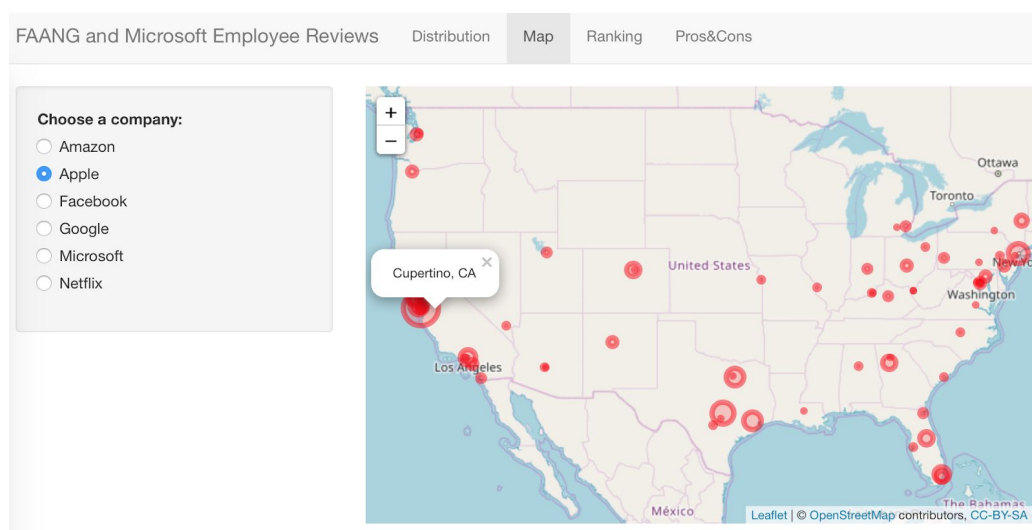

Figure 2.2 Map for employee location

The ranking page helps the user get the detailed understanding of each company. User can choose the categories and do the sorting on the tables.
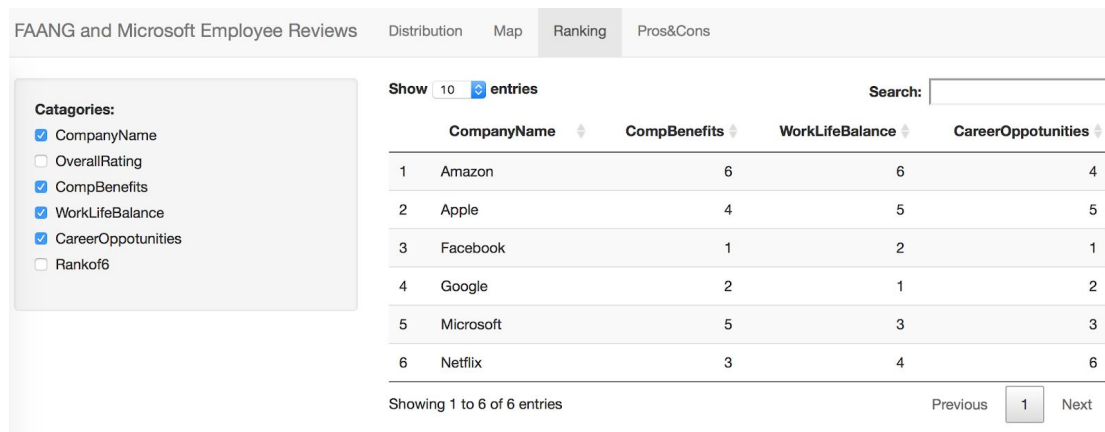

Figure 2.3 Table of ranking

The last page is the wordcould for the employee reviews. Users could choose the company names, pros/cons and maximum number of words. The default number of words is 20.
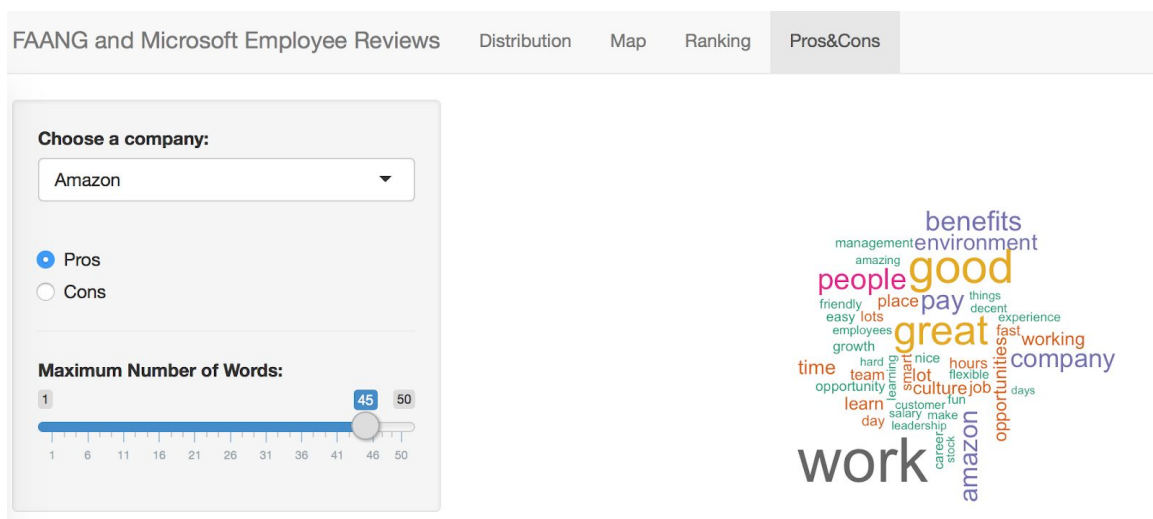

Figure 2.4 Wordcloud for pros&cons

Conclusion

The project helps to learn many of the skills of Shiny. I learned how to do the design by using five design sheets. If I can improve the project, I would like to add more interactions in it. For example, include year range into the project.

The reason why my project is difficult is:

1. Compare to normal pie chart, I used plotly. So there will be more interaction and has a better view of the chart.

2. For the wordcloud, there are lots of processes to do for the text before counting the word frequency. For example, remove punctuations, numbers, whitespace, useless words("the", "and", "but", etc.)
3. The marks in the map are in different size, it represents the amount of employees in a certain city.

Reference

1. Dataset of the project
   https://www.kaggle.com/petersunga/google-amazon-facebook-employee-reviews#employee_reviews.csv