Lab3 Data Visualization

1. About the Google Play Store Apps dataset

Columns in the Google Play Store Apps dataset:

```
Index(['App', 'Category', 'Rating', 'Reviews', 'Size', 'Installs', 'Type',
       'Price', 'Content Rating', 'Genres', 'Last Updated', 'Current Ver',
       'Android Ver'],
      dtype='object')
```

Select a few columns that contain valuable information for consideration:

- **Category**: There are a total of 33 distinct categories. It would be worthwhile to display the proportion of apps belonging to each category.
- **Rating**: This column consists of floating-point numbers ranging from 1 to 5, representing the percentage distribution of different ratings. Consider displaying highly rated apps as recommended content for users.
- **Installs**: The number of installations varies significantly, indicating the popularity of each app. It would be useful to showcase some popular apps to users.
- **Reviews**: This column reflects the level of discussion and popularity surrounding each app, potentially having a strong correlation with the number of installations.
- **Size**: Out of the total apps, 278 end with 'M' and 182 end with 'K'. The size of the remaining apps will vary depending on the specific device. Displaying the distribution of these three categories could be beneficial.
- **Type** and **Price**: The 'Type' column indicates whether the app is free or paid. If an app is free, the price is listed as 0; otherwise, it ranges from $0.99 to $400.
- **Content Rating**: This column denotes the age limit for each app.
- **Last Updated**: The 'Last Updated' column spans from 2010 to 2018, allowing for an analysis of the number of app updates per year across different categories.

2. My design for the dashboard

The layout of the page is shown as depicted in the screenshot.

2.1 Pie Chart

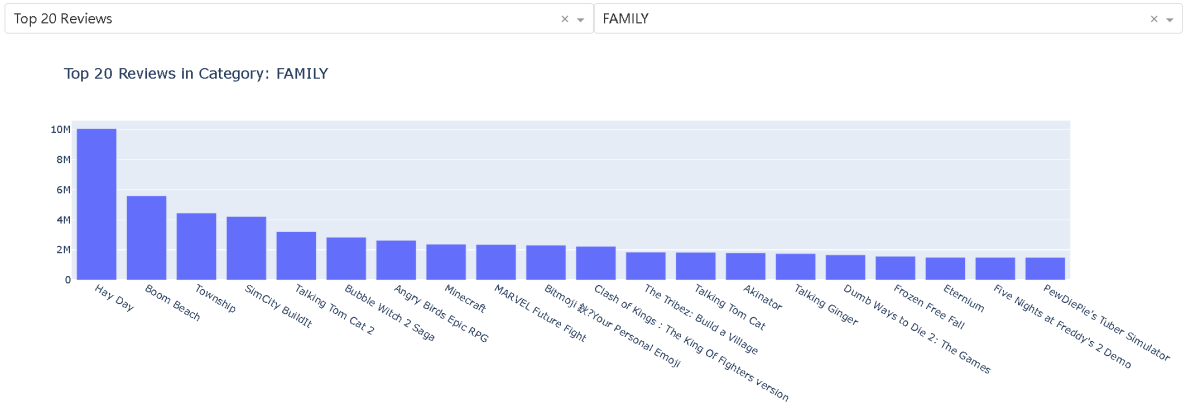
Users can click on the dropdown menu on the left to select one of the following options: Category, Type, Size, Rating, or Content Rating. Upon selection, the pie chart will be updated to display the proportion of each app category. For instance, among all the apps, the category with the highest proportion is FAMILY.

2.2 Line Chart

Users can click on the dropdown menu on the right to select a specific category or choose all categories directly. Upon selection, the line chart will be updated to display the number of updates for that category in each year. It can be observed that the majority of apps were updated between 2016 and 2018.

2.3 Bar Chart

Users can click on the two adjacent dropdown menus to select the ranking criterion(Reivew, Installs, Ratings or Price) and the app category. For example, in the graph below, the bar chart showcases the top 20 apps in the FAMILY category based on the installation count. The highest-ranked app in this category has nearly 10 million installations.



2.4 Scatter Plot

Lastly, we have a 3D scatter plot. Users can drag and adjust the view to explore the distribution of variables and visualize the relationships between different variables.

In this scatter plot, the x, y, and z axes represent Last Updated, Rating, and Category, respectively. The size of the data points indicates the number of installations, while the color ranges from red (indicating more reviews) to blue (indicating fewer reviews), reflecting the number of reviews for each data point.

3D Scatter Plot of Apps(size for Installs, color for Reviews)