Project Report
Visualization of Google Play store App Data

Name of the Student	Santosh Kumar Medisetty
Banner ID	B00836673
Name of the Course	CSCI 6406 - Visualization

Abstract:

Google Play Store a variety of applications. No product provides insight into how well an application will perform or provides information on how to improve the application that needs to be developed. So, visualizations which depicts the information of top applications will be helpful to a user who is in search of best applications.

Introduction:

Android, an open-source software, has created a shift in the technology sector, and especially the smartphone industry. Play store is a one-stop-shop for Android users to download millions of apps from every country. There are apps in the play store that relate to various categories such as news, gaming, personalization, athletics, music, etc. Several applications have a huge number of installs relative to others, due to different factors that determine the application's popularity. Consumers struggle to choose an application in the group they like among the thousands of applications. So, a visualization that shows top applications based on the specified rating in a particular category would help the user to choose an application.

Motivation and benefits:

The main aim of this project is to help the user get an idea of what application can be downloaded from a set of applications from the chosen category based on user ratings. A visualization that displays the top 3 applications based on the category be it paid or free application is shown. The Sunburst chart and Hierarchical Bar chart gives insights into the dataset used in this project. Another visualization i.e., a bar chart that displays the average number of ratings based on the genre in games category is shown.

Dataset Used:

The dataset used for this project is taken from the Kaggle website [1]. The dataset contains around 32K applications available in google play store. The dataset contains the following attributes.

Column	Description
App Name	Name of the application
Category	Type of the application
Rating	Rating of the application
Price	Cost of the application
Installs	Number of Installs of the application
Reviews	Number of reviews of the application
Size	Size of the application
Last Updated	Last updated date
Minimum version	Minimum version of android required
Latest version	Latest version of android required

Description of data:

Category: There are different categories of applications like Communication, Education, Entertainment, Finance, Food, Music, Photography, Social, Shopping, Sports, and Games. Games are again divided into categories like Action, Adventure, Arcade, Card, Board, Racing, Puzzle, Word.

Ratings: The applications are rated from 1 to 5 where 5 is the highest and 1 is the lowest.

Price: There are applications which are free to download, and some are paid applications.

Preparatory Tasks:

Data Pre-processing: The dataset with 32K applications is cleaned to remove duplicates, null values and some rows with special characters. The columns “Last Updated”, “Minimum version” and “Latest version” are deleted from the dataset as they are not helpful for our analysis. The final dataset contains approximately 13K records.

JSON file creation: A JSON file is created hierarchically to draw a sunburst chart and Hierarchical bar chart with application category, price, and rating of the application. The order of the hierarchy is given below.

Application Data → Application Category → Price (Paid/Unpaid)→Application Rating

CSV file creation: A CSV file is created with an average number of ratings for each genre of game in the games category and the paid/unpaid category.

Visualizations:

Visualization of information refers to methods used by visual representation to convey data knowledge. Its main objective is to condense large data sets into visual graphics so that complex relationships within the data can be easily understood. Three different types of visualizations are created in this project to help users get the insights of Google play store data.

1. Sunburst Chart:

Sunburst chart helps one to visualize data hierarchically [2]. The attributes used for this chart are application type, target audience, and category[5]. By selecting the required category, paid/Unpaid and rating of application, one will get details about the top three applications in that selection. In the data-preprocessing step, the below steps are followed.

- Few records have only special characters, those records were deleted.
- Application type was of two classes: Paid and Free. All the records were checked for errors.

As seen in the Figure 1, the sunburst chart has many levels [10]. These levels can be filtered out based on the selection made by the user.

Figure-1 Sunburst Chart

For instance, if the user selects category as “Finance”, price of the app as “Unpaid” and rating as “4”, the top 3 applications will be displayed beside the chart as shown in Figure 2.

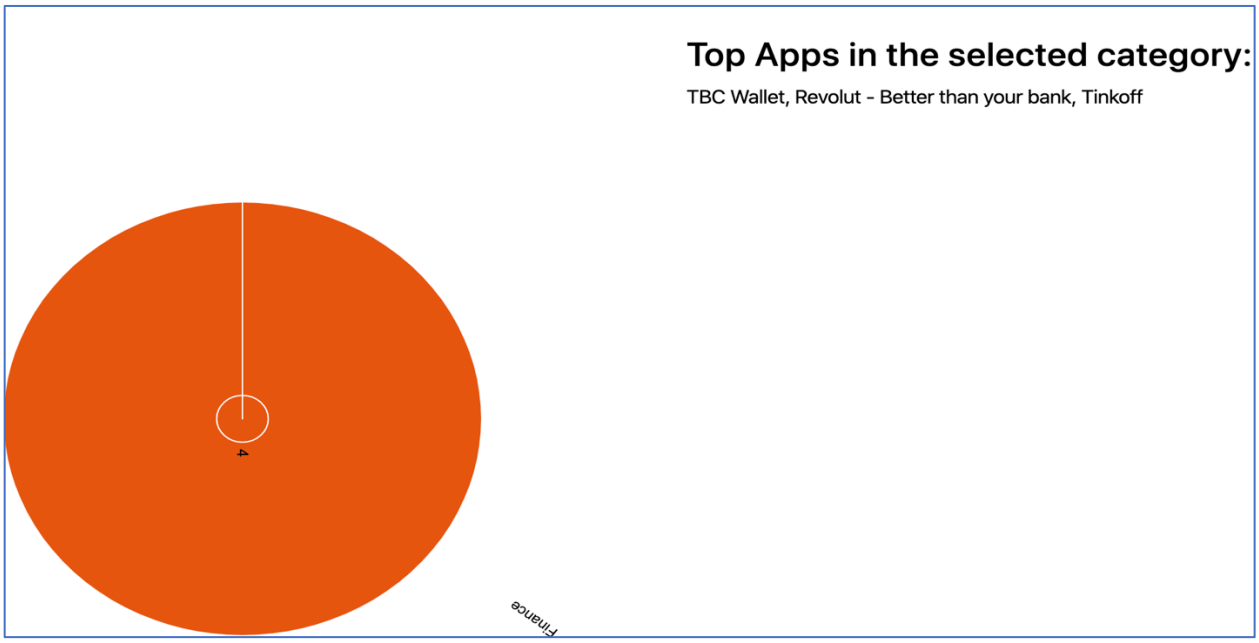

Figure-2 Sunburst chart shows the top 3 apps beside the chart.

2. Hierarchical Bar Chart:

An interactive Hierarchical Bar Chart helps to visualize the data hierarchically [3]. The data used for this chart is the same as the Sunburst chart above. The data is visualized in the form of bars extending towards the horizontal axis and base on the vertical axis. “App categories” are displayed along the vertical axis and the “Total number of apps” are displayed along the horizontal axis. When clicked on a bar, the bar is divided into its sub-category. The Hierarchical bar chart is shown in Figure-3.

3. Bar Chart:

An interactive bar chart [4] that visualises the average number of ratings to each genre in games category is displayed. The average number of ratings is displayed along vertical axis and each genre in games sub-category are displayed along horizontal axis [7]. This bar chart is categorised into 2 types based on the cost of application (Paid/Unpaid). The use of this chart is to let the user know which genre of games received a greater number of ratings which implies the user to know which genre of games does people use more. This is displayed in Figure-4.

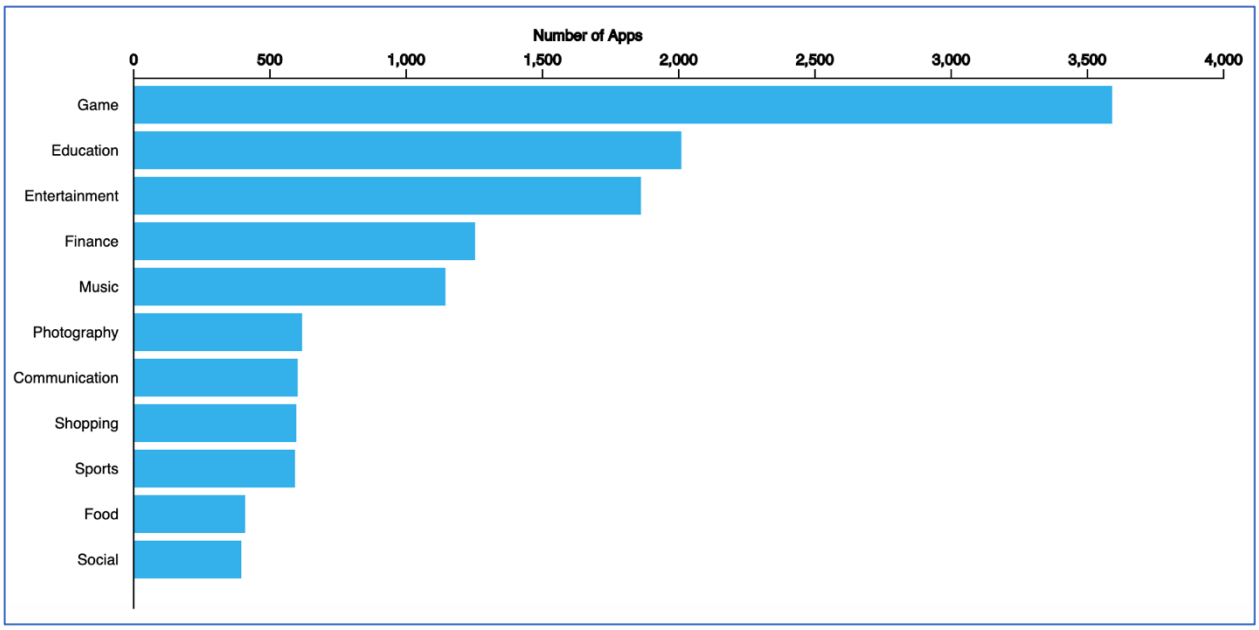

Figure-3 Hierarchical Bar Chart that displays the number of apps in each category.

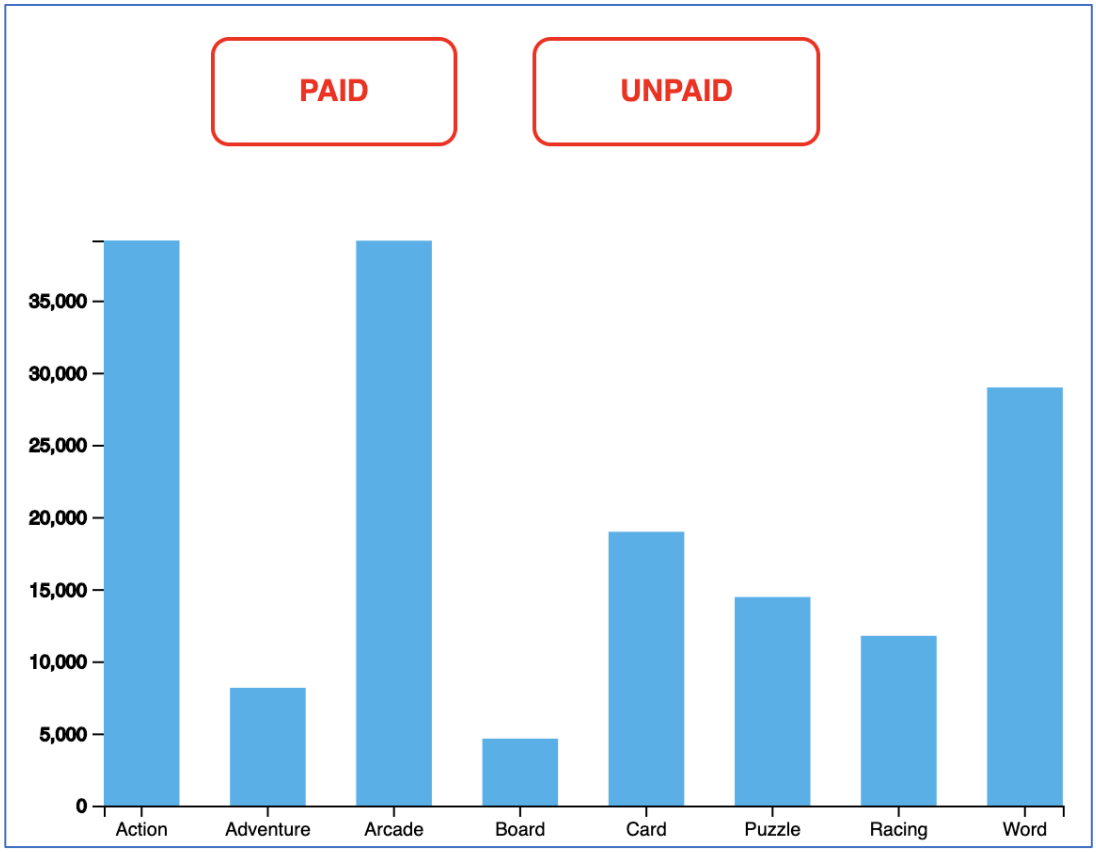

Figure-4 Bar Chart that displays average number of ratings in each genre of Games in Paid category.

The two categories Paid and Unpaid are linked interactively to make the transition when clicked. Figure-4 displays the average ratings vs genre for the paid category. Whereas Figure-5 displays average ratings vs genre for the unpaid category.

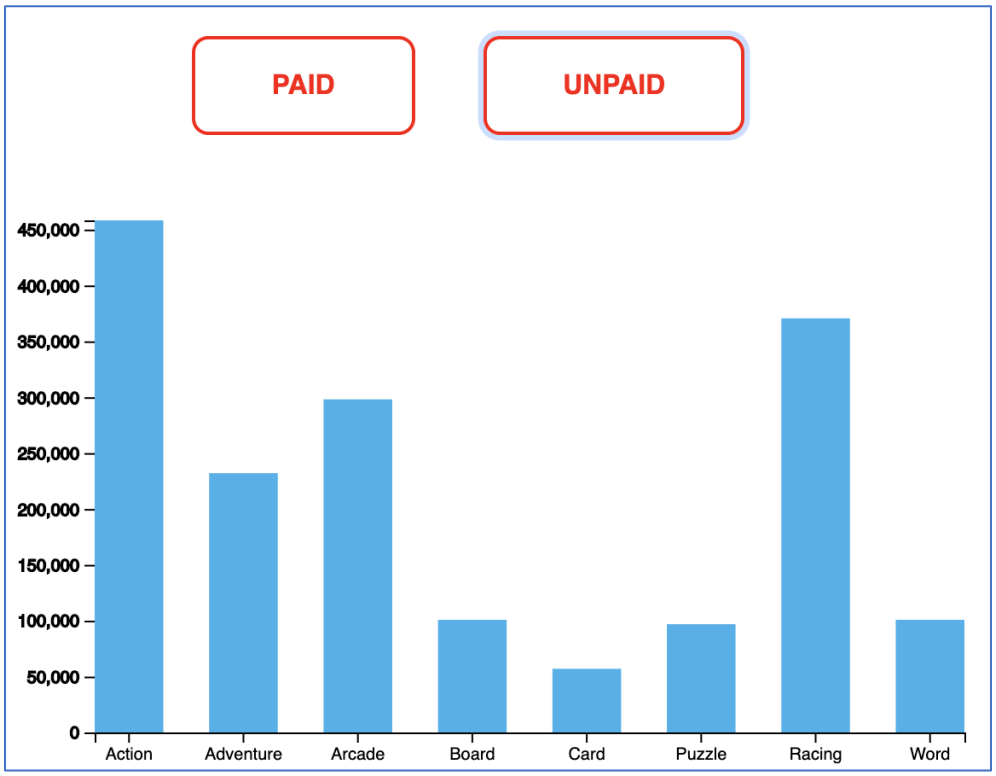

Figure-5 Bar Chart that displays average number of ratings in each genre of Games in Unpaid category.

To view the various charts, a web layout is designed using HTML, CSS, and bootstrap. A navigation bar is added in order to navigate through the charts. Separate pages are dedicated to each of the charts. Since sunburst and hierarchical bar chart both belong to hierarchical charts, they are placed under the same navigation item “Hierarchical charts”. These two charts are again linked to display interactively with the help of buttons. Bar Chart is displayed under another navigation item. Bonus web pages “Home” that gives the information about the project and “info” page that gives the basic overview of the play store application are added. A snapshot of the navigation bar is shown in Figure-6. Finally, to improve the view of the webpage small enhancements are done and a background image with the theme of the application is used. A screenshot of the application is shown below in Figure-7.

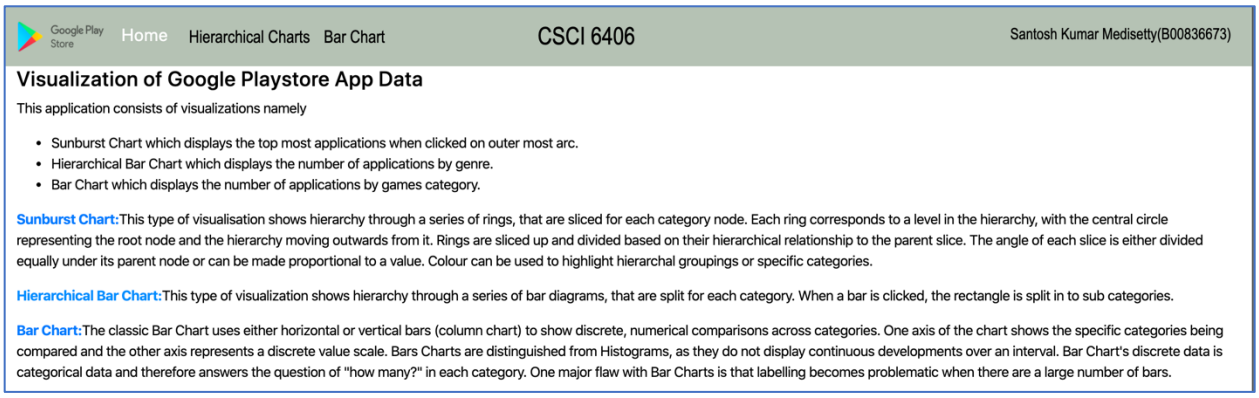

Figure-6 Navigation bar which displays the layout of all the charts.

Figure-7 Enhanced webpage

Tools Used:

Following is the description of the technologies used in the development of the project:

- HTML and CSS: HTML (Hypertext Markup Language) is used to construct the page's actual content, such as text, and CSS (Cascade Styling Sheets) is responsible for website design or style [12], including layout, background color, and visual effects.
- Python for data cleaning.
- D3.js to draw the charts.

Conclusion:

Based on the input selected by the user, the charts will be able to display the top applications in Sunburst Chart and Bar Chart helps the user to know which genre of applications are most used by android users.

References:

[1] “Google Play Store Apps", *Kaggle.com*, 2019. [Online]. Available: https://www.kaggle.com/gauthamp10/google-playstore-apps

[2] "Sunburst | Highcharts", Highcharts.com, 2019. [Online]. Available: https://www.highcharts.com/demo/sunburst

[3] “Hierarchical Bar Chart | bl.ocks ”, bl.ocks.org, 2019 [Online]. Available: https://bl.ocks.org/gallowayevan/f87b0911dcde85a5ccd3499a36e328e9

[4] “Bar Chart | d3-graph-gallery”, d3-graph-gallery.com, 2019 [Online]. Available: https://www.d3-graph-gallery.com/graph/barplot_button_data_csv.html

[5] Fan Bao and Jia Chen, "Visual framework for big data in d3.js," 2014 IEEE Workshop on Electronics, Computer and Applications, Ottawa, ON, 2014, pp. 47-50.

[6] L. Woodburn, Y. Yang and K. Marriott, "Interactive Visualisation of Hierarchical Quantitative Data: An Evaluation," 2019 IEEE Visualization Conference (VIS), Vancouver, BC, Canada, 2019, pp. 96-100.

[7] R. M. Amir Latif, M. Talha Abdullah, S. U. Aslam Shah, M. Farhan, F. Ijaz and A. Karim, "Data Scraping from Google Play Store and Visualization of its Content for Analytics," 2019 2nd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), Sukkur, Pakistan, 2019, pp. 1-8.

[8] T. Barlow and P. Neville. A comparison of 2-D visualizations of hierarchies. In IEEE Symposium on Information Visualization, 2001. INFOVIS 2001., pages 131–138, Oct. 2001

[9] R. Amar and J. Stasko. A Knowledge Task-Based Framework for Design and Evaluation of Information Visualizations. In IEEE Symposium on Information Visualization, pages 143–150, 2004

[10] J. Stasko and E. Zhang. Focus+context display and navigation techniques for enhancing radial, space-filling hierarchy visualizations. In IEEE Symposium on Information Visualization 2000. INFOVIS 2000. Proceedings, pages 57–65, 2000

[11] Briskman, J. (2019). Top Apps Worldwide for Q1 2019 by Downloads. Retrieved 5 October 2019, from https://sensortower.com/blog/top-apps-worldwide-q1-2019-downloads

[12] Coding Dojo Blog. (2019). The Must Read HTML vs CSS Infographic | Coding Dojo Blog. [online] Available at: https://www.codingdojo.com/blog/html-vs-css-inforgraphic