PHASE III: Implementation

Data visualization project: creating App analysis dashboard forgoogle play store dataset

Team Members: SHRINA PATEL

CHANDANA KANDARI

Presented to – ASMAA ELBADRAWY

Table of Contents

[Section 1: The Dashboard 2](#_Toc39347684)

[Final Figure of the Dashboard: 2](#_Toc39347685)

[Explanation for what it is used for: 2](#_Toc39347686)

[Section 2: The Dataset 2](#_Toc39347687)

[Google play store dataset Explanation 2](#_Toc39347688)

[Variable description in Goggle play store dataset 3](#_Toc39347689)

[Plots & Pre-attentive attributes 3](#_Toc39347690)

[Data Preparation & Preprocessing 4](#_Toc39347691)

[Section 3 – Dashboard Users 4](#_Toc39347693)

[Section 4 – Final Set Of Questions 5](#_Toc39347694)

[Section 5 – Plots 5](#_Toc39347695)

[Section 6 – Dashboard Interactivity 17](#_Toc39347696)

# Section 1: The Dashboard

## Final Figure of the Dashboard:

A screenshot of a computer

Description automatically generated

## Explanation for what it is used for:

The data from the Play Store applications has enormous potential to push successful app-making businesses. The developers will draw actionable lessons to work on and win the Android market. The dataset is extracted from Kaggle. It's 10k Play Store apps site scraped data for evaluating the Android market. It is used for analyzing large amount of data and visualizing through plots and make visualization easy.

**Application developers** can use this dashboard for visualizing the information that make them understand the users’ preferences

**Mobile Application Users** can use this dashboard for visualizing the information that help them in analyzing and comparing the applications of same category that serve the intended function.

The designed Dashboard answers the below 4 questions that are commonly arises for both the **Application developers and Mobile Application Users.**

1. Visualization of Number of people reviewed each app, as percent of number of installations? (percentage of reviews over the number of Installations for a selected application)
2. Visualization of Top N Applications from the user selected Categories based on Ratings, Reviews, and Installs as per the users’ choice.
3. Visualization of Number of Distinct Applications over Content Rating across the selected Category.
4. Visualization of Average Rating for a selected Priced Application (How do price affect average rating?)

Number of rows and columns: This is composed of 10841 rows and 13 columns in total.

# Section 2: The Dataset

## Google play store dataset Explanation

* The data from the Play Store apps has tremendous potential to drive successful app-making businesses. The developers will draw actionable insights to work on and win the Android market! The data from the Play Store applications has enormous potential to push successful app-making businesses.
* The developers will draw actionable lessons to work on and win the Android market. The dataset is extracted from Kaggle. It's 10k Play Store apps site scraped data for evaluating the Android market.
* The google play store dataset consists of 10841 rows and 13 columns. The target variable is “Installs” (number of installation), and explanatory variables include category, price, ratings, and reviews, etc.
* Specifically, ‘Rating’ is a continuous variable with a scale of 1–5. **Family** and **Game** apps have the highest market prevalence. Interestingly, **Tools, Business and Medical** apps are also catching up.

## Variable description in Goggle play store dataset

#### App: Application name

#### Category: Category the app belongs to

#### Rating: Overall user rating of the app (as when scraped)

#### Reviews: Number of user reviews for the app (as when scraped)

#### Size: Size of the app (as when scraped)

#### Installs: Number of user downloads/installs for the app (as when scraped)

#### Type: Paid or Free

#### Price: Price of the app (as when scraped)

#### Content Rating: Age group the app is targeted at - Children / Mature 21+ / Adult

#### Genres: An app can belong to multiple genres (apart from its main category). For e.g., a musical family game will belong to Music, Game, Family genres.

#### Last Updated: Date when the app was last updated on Play Store (as when scraped)

#### Current Ver: Current version of the app available on Play Store (as when scraped)

#### Android Ver: Min required Android version (as when scraped)

## Plots & Pre-attentive attributes

* Scatter Plot - Form – 2D position, Spatial grouping, Form- Shapes.
* Bar chart - Form- line length, line width, color (Intensity, Hue), Enclosure and Orientation.
* Line chart - Form – curvature, shows changing over time- trend, Spatial position, color (for visualizing difference between types).
* Stacked bar chart - Form- line length, line width, color (Intensity, Hue), Enclosure and Orientation.

## Data Preparation & Preprocessing

Data Cleaning done before loading the dataset into the tableau:

1. Removed '+' from 'Number of Installs' to make it numeric
2. One of the Rows had mismatch data entries, so shifted the cells towards right to match the data entries.
3. In the ‘Rating’ attribute, replaced NaN Values to -1 to make it Numeric instead of String.

After successful preprocessing of dataset, data is finally prepared for visualization.

# 

# Section 3 – Dashboard Users

Dashboard is developed for Application developers and Mobile Application Users.

**Application developers** can use this dashboard for visualizing the information that make them understand the users’ preferences like:

* The Applications that have the highest Installs, Rating, Reviews in a particular Category or among Categories
* The type(free/paid) of application of a particular category affect Rating.
* Price affect the average rating of applications.
* Number of Distinct Applications over Content Rating across the selected Category.
* Comparison of number of paid and free applications across categories
* Number of people reviewed each app, as percent of number of installations.

**Mobile Application Users** can use this dashboard for visualizing the information that help them in analyzing and comparing the applications of same category that serve the intended function:

* The applications that has the highest installs, Reviews, Rating in a particular Category or among Categories.
* Distinct Applications over Content Rating across the selected Category.
* Number of people reviewed each app, as percent of number of installations.
* Price affect the average rating of applications.

# Section 4 – Final Set Of Questions

1. Visualization of Number of people reviewed each app, as percent of number of installations? (percentage of reviews over the number of Installations for a selected application)
2. Visualization of Top N Applications from the user selected Categories based on Ratings, Reviews, and Installs as per the users’ choice.
3. Visualization of Number of Distinct Applications over Content Rating across the selected Category.
4. Visualization of Average Rating for a selected Priced Application (How do price affect average rating?)
5. What is the rating across the categories based on types(free/paid)?
6. Comparison of number of paid and free applications across categories?

# Section 5 – Plots

1. Visualization of Number of people reviewed each app, as percent of number of installations? (percentage of reviews over the number of Installations for a selected application)

* **Scatter plot** is used for this type of visualization where the bar is apps and ATTR (installs) and the value is Reviews%(Installs).
* List of Pre-attentive attributes: Form – 2D position, Spatial grouping, Form- Shapes

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

A screenshot of a social media post

Description automatically generated

A screenshot of a social media post

Description automatically generated

1. Visualization of Top N Applications from the user selected Categories based on Ratings, Reviews, and Installs as per the users’ choice.
   1. **Bar chart** is used to visualize the data where the bar is application and the values are taken as average number of Reviews/Ratings/Installs as per the user choice and sorted in descending order.
   2. It determines the list of application that has highest average number of Reviews/Ratings/Installs as per the user choice.
   3. Sliding bar as a parameter control is created as the user can select based on their interest how many applications average number of Reviews/Ratings/Installs they want to visualize.

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

In this Visualization,

Column is Application which has only relevant values based on the category/categories selected by the user. Category attribute (Multiple values list) is added to the context based on this App Attribute relevant values are shown.

Row is Average(row attribute), here the row attribute can take the values (Reviews/Ratings/Installs) from ‘Select Row Attribute’ Parameter linked to a ‘Row Attribute’ calculated Field.

‘Select Row Attribute’ Parameter

A screenshot of a social media post

Description automatically generated

‘Select Row Attribute’ Parameter A screenshot of a cell phone

Description automatically generated

List of Pre-attentive attributes: Form- line length, line width, color (Intensity, Hue), Enclosure and Orientation.

**Top N Applications Vs Reviews**

A screenshot of a cell phone

Description automatically generated

**Top N Applications Vs Rating**

A screenshot of a cell phone

Description automatically generated

**Top N Applications Vs Installs**

A screenshot of a social media post

Description automatically generated

1. Visualization of Number of Distinct Applications over Content Rating across the selected Category.

* **Bar chart** is used for visualization where column is content rating and rows is CNTD(Apps).
* Parameter control is created as selected Category along with calculated field to represent different application over content rating and category is selected as per user’s choice for visualization.
* Calculated field is created for selecting the category for user’s choice.
* List of pre-attentive attributes: form- length, width, color, Enclosure and orientation.

A screenshot of a social media post

Description automatically generated

A screenshot of a cell phone

Description automatically generated

A screenshot of a social media post

Description automatically generated

1. How do price affect average rating?

* Visualization of Average Rating for a selected Priced Application.
* **Scatter plot** is used for doing the visualization where the bar is price (a multiple values list with excluded values) and the value is taken as average rating.
* List of Pre-attentive attributes: Form – 2D position, Spatial grouping, Form- Shapes

Filters:

* Type – only paid Apps are visible
* Price – added to the context based on its values the relevant app attribute values are displayed.
* Rating - -1 Rating is excluded as it is NaN that is unknown Rating

Average Rating for all Applications over a selected price

A screenshot of a cell phone

Description automatically generated

Average Rating for a selected Priced Application over a selected price

A screenshot of a cell phone

Description automatically generated

1. What is the rating across the categories based on types?

* **Line chart** is used to visualize the data where the bar is rating, and the values is taken as number of records. This graph visualizes based on the created calculated field and parameter control we can select category for which we want to visualize rating that are differentiated based on types (Free/ paid apps)
* List of Pre-attentive attributes: Form – curvature, added marks, shows changing over time- trend, Spatial position, color (for visualizing difference between types).

A screenshot of a social media post

Description automatically generated

A screenshot of a cell phone

Description automatically generated

A screenshot of a social media post

Description automatically generated

6. Comparison of number of paid and free applications across categories?

**Stacked bar chart** is used for visualization where bar is Category and value is taken as Number of records. Stacked bar charts are extremely useful as we want to visualize and compare category wise data in one visualization. It lets us accommodate a lot of values and detail into one chart. In our visualization, the chart is visualized based on Type.

List of Pre-attentive attributes: Form- line length, line width, color (Intensity, Hue), Enclosure and Orientation.

A screenshot of a cell phone

Description automatically generated

# Section 6 – Dashboard Interactivity

**List of Interactive Controls:**

When you want to add interactivity and versatility to a study, or play with what-if scenarios, parameters are useful. Suppose you're unsure which fields will work better for your viewers to include in your view or which style. You should include parameters in your view to allow viewers to choose how they wish to look at the data.

1. **How to price affect rating- Drop down menu parameter control for apps and checklist for Price.**

Range for price- 0 to $ 400

Range for Apps – As a drop-down menu

A screenshot of a cell phone

Description automatically generated A screenshot of a cell phone

Description automatically generated

1. **Create parameter control for Top N categories by field- Visualization of Top N Applications from the user selected Categories based on Ratings, Reviews, and Installs as per the users’ choice.**

* Checklist created for selecting category for visualization and slider bar for Top N Categories.
* Multiple value list for selecting the category based on which the app are displayed for a selected row attribute i.e.. Visualization based on Reviews, Rating and Installs
* Select Row Attribute’ Parameter – Drop down menu
* Range for selecting Top N Categories – 100 maximum and 1 minimum

A screenshot of a cell phone

Description automatically generated

1. **Selected category parameter control-** Drop down menu that selects name of category(category)\_name) calculated field as per user’s choice for visualization of content rating VS no of apps.

**Range of values :** ART\_AND\_DESIGN,AUTO\_AND\_VEHICLES, BEAUTY, BOOKS\_AND\_REFERENCE, BUSINESS, COMICS, COMMUNICATION, DATING, EDUCATION, ENTERTAINMENT, EVENTS, FAMILY, FINANCE, FOOD\_AND\_DRINK, GAME, HEALTH\_AND\_FITNESS, HOUSE\_AND\_HOME, LIBRARIES\_AND\_DEMO,LIFESTYLE , MAPS\_AND\_NAVIGATION, MEDICAL , NEWS\_AND\_MAGAZINES, PARENTING, PERSONALIZATION, PHOTOGRAPHY, PRODUCTIVITY, SHOPPING, SOCIAL, SPORTS, TOOLS, TRAVEL\_AND\_LOCAL, VIDEO\_PLAYERS, WEATHER

A screenshot of a cell phone

Description automatically generated

1. **App\_Name –** Used for Visualization of Number of people reviewed each app, as percent of number of installations

A screenshot of a social media post

Description automatically generated