CAPSTONE PROJECT-1 PLAY STORE APP REVIEW ANALYSIS

BY-Team Achievers(Prashant Singh
Sachin
Jhaanvi khandelwal
pankaj kumar pandit)

AI

Content

- Why analyze the google play store?
- Introduction
- Agenda
- Dataset Preparation
- Attributes in Google Play store Data
- Attributes in User reviews
- Overview of analysis
- Exploring app categories
- Installs according to genres
- Correlation Heatmap
- Rating fluctuation
- Count of application in each category diffrentiated by their types.
- Distribution of subjectivity
- Distribution of reviews
- Relation between sentiment subjectivity and sentiment polarity
- Percentage of review sentiments
- Conclusion

AI

Why analyze the google play store?

Mobile app market is set to grow 20% by 2023.

Android Apps comprise 90% of the Mobile App Market

What makes an App popular? Can we predict how popular it's going to be?

What are some interesting patterns in user behavior related to app usage & feedback

Introduction

•Android is the most popular operating system in the world, with over 2.5 billion active users spanning over 190 countries.

•Google Play was launched on March 6, 2012, bringing together Android Market marking a shift in Google's digital distribution strategy .

• Android is the dominant mobile operating system today more than 85% of all mobile devices running Google's OS. The Google Play Store is the largest and most popular Android app store.

•There are more than 3.04 million apps found on Google Play Store.

• The Play Store apps data has enormous potential to drive app-making businesses to success.

•Actionable insights can be drawn for developers to work on and capture the Android market. The main goal of our project is-

1)The purpose of our project is to gather and analyze detailed information on apps in the Google Play Store in order to provide insights on app features and the current state of the Android app market.

2) The Objective of the project to Explore and analyze the data to discover key factors responsible for app engagement and success.

Problem Statement

❏ Two datasets are provided, one with **basic information** and the other with **user reviews** for the respective app.

❏ We must examine and evaluate the data in both datasets in order to identify the important characteristics that influence app engagement and success.

So, what factors influence an app's success?

An app is said to be successful if it has:

❏ A high average user rating

❏ A good number of positive reviews

❏ A good number of monthly average users

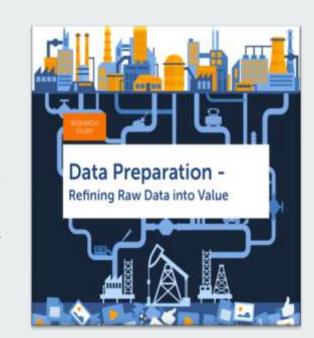❏ High revenue per customer and so on.

Agenda

- Exploring app categories.
- Installs according to genres.
- Heatmap of the Dataset.
- Rating Fluctuation.
- Count of application in each category differentiated by their types.
- Distribution of subjectivity.
- What is the distribution of type of reviews in the dataset?
- Does sentiment subjectivity is proportion to sentiment polarity
- Percentage of review sentiments.

AI

Dataset Preparation

▪**Loading the data sets:** Two datasets, First Play store app dataset and User Reviews dataset.

▪**Import Libraries:** NumPy, Pandas, Seaborn and Matplotlib

▪**Data cleaning:** Null values, Finding and removing Outliers, Removing duplicate data.

▪**Data Imputation:** Filling the missing categorical values with mode and numerical values with median. Conversion of price, installs, reviews into numerical values.

▪**Exploratory Data Analysis:** Analyzing the data sets to summarize their main characteristics using statistical graphics and data visualizations method.

Data Preparation -
Refining Raw Data into Value

Attributes in Google Play store Data

1.App : This column Contains the name of the app for each observation.
2.Category : This column Contains Category to which the app belongs.
3.Rating : This column contains the average rating for the app.
4.Reviews : This column contains the number of reviews that the app has received on the play store.
5.Size : This column contains the amount of memory the app occupies on the device.
6.Installs : This column contains the number of times that the app has been downloaded and installed from the play store.
7.Type : This column contains the information whether the app is free or paid.
8.Price: If the app is a paid app, this column contains the data about its price.
9.Content Rating: This column contains the maturity rating of the app i.e. the age group of the audience for which it is suitable.
10.Genres: This column contains the data about to which genre the app belongs. Genres can be considered as a further division of the group of Category.
11.Last Updated: Contains the date on which the latest update of the app was released.
12.Current Version: Contains information on the current version of the app available on the play store.
13.Android Version: Contains information about the android versions on which the app is supported.

AI

Attributes in User reviews

1. **App-** Application name
2. **Translated Review-** User review
3. **Sentiment-** Positive/Negative/Neutral
4. **Sentiment Polarity-** Sentiment polarity score
5. **Sentiment Subjectivity-** Sentiment subjectivity score

Overview of analysis

Data Cleaning

Understand the structure of the dataset and clean data before analysis.

Data Exploration

Uncover initial patterns, characteristics, and points of interest using visual exploration.
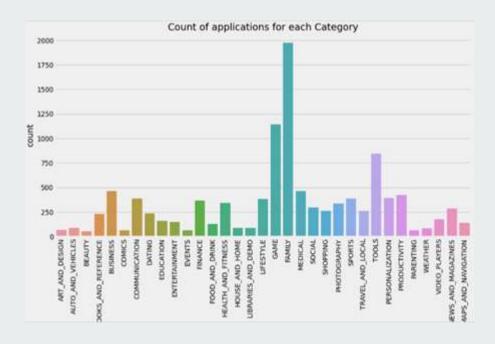
Predictive Modeling

Formulate a statistical model to forecast an outcome using relevant predictors.

Exploring app categories

- With more than 1 billion active users in 190 countries around the world, Google Play continues to be an important distribution platform to build a global audience. For businesses to get their apps in front of users, it's important to make them more quickly and easily discoverable on Google Play. To improve the overall search experience, Google has introduced the concept of grouping apps into categories. Which category has the highest share of (active) apps in the market? Is any specific category dominating the market? Which categories have the fewest number of apps?
- We will see that there are 33 unique app categories present in our dataset. Family and Game apps have the highest market prevalence. Interestingly, Tools, Business and Medical apps are also at the to*p*.
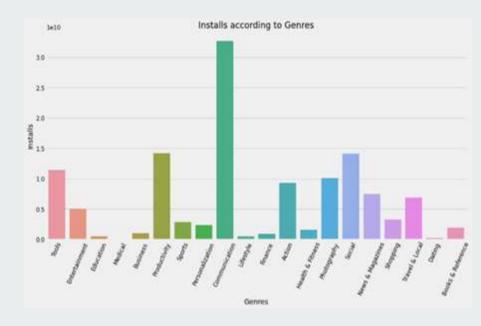

Count of applications for each Category

Installs according to genres

- As we can see communication genre has the highest number of installs in this genre social media apps like whatsapp and Gmail are included which has a huge number of customer base as everyone in this day and age uses social media apps.
- If a developer wants to develop an app then the apps which falls into these genres then it will have a huge customer base.
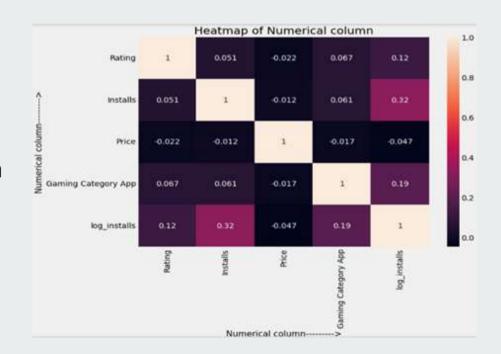

Installs according to Genres

Correlation Heatmap

- There is a strong **positive** correlation between the **Reviews** and **Installs**.
- The Price is slightly **negatively** correlated with **rating,reviews and installs.**
- The **Rating** is slightly **positively** correlated with the **Installs** and **Reviews**.

Heatmap of Numerical column

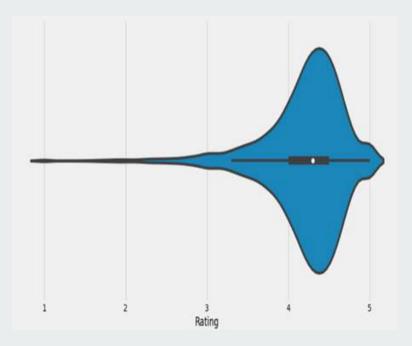| Numerical column ---->	Rating	Installs	Price	Gaming Category App	log_installs
Rating	1	0.051	-0.022	0.067	0.12
Installs	0.051	1	-0.012	0.061	0.32
Price	-0.022	-0.012	1	-0.017	-0.047
Gaming Category App	0.067	0.061	-0.017	1	0.19
log_installs	0.12	0.32	-0.047	0.19	1

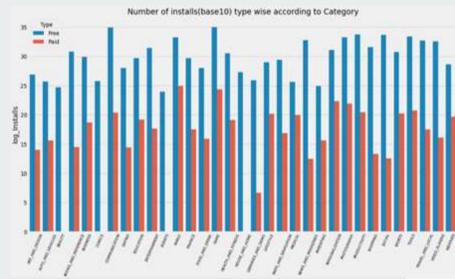Numerical column ------->

Rating fluctuation

- After having witnessed the market share for each category of apps, let's see how all these apps perform on an average. App ratings (on a scale of 1 to 5) impact the discoverability, conversion of apps as well as the company's overall brand image. Ratings are a key performance indicator of an app.
- From our research, we found that the average volume of ratings across all app categories is 4.17. The violin plot plot is broad between 4 and 5 indicating that the majority of the apps are highly rated with only a few exceptions in the low-rated apps.

Count of application in each category diffrentiated by their types.

- It looks like certain app categories have more free apps available for download than others. These majority of apps in the Family, Food & Drink, and Tools, as well as Social categories were free to install.
- At the same time Family, Sports, Tools, and medical categories had the biggest number of paid apps available for download.
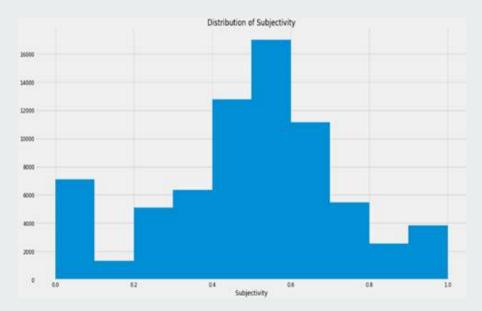

Number of installs(base10) type wise according to Category

Distribution of subjectivity

- It can be seen that maximum number of sentiment subjectivity lies between **0.4 to 0.7.** From this we can conclude that maximum number of users give reviews to the applications, according to their experience.
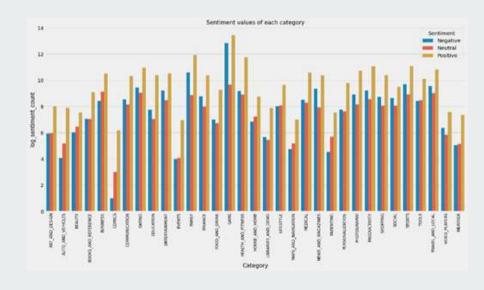

Distribution of Subjectivity

Distribution of reviews

- Sentiment basically determines the attitude or the emotion of the writer, i.e., whether it is positive or negative or neutral. Sentiment Polarity is float which lies in the range of [-1,1] where 1 means positive statement and -1 means a negative statement. Sentiment Subjectivity generally refer to personal opinion, emotion or judgment, which lies in the range of [0,1]
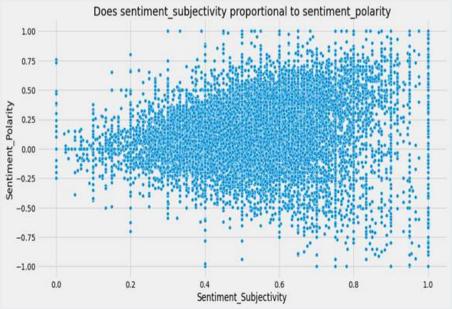
Relation between sentiment subjectivity and sentiment polarity

- Sentiment subjectivity is not always proportional to sentiment polarity but in maximum number of case, shows a proportional behavior, when variance is too high or low.
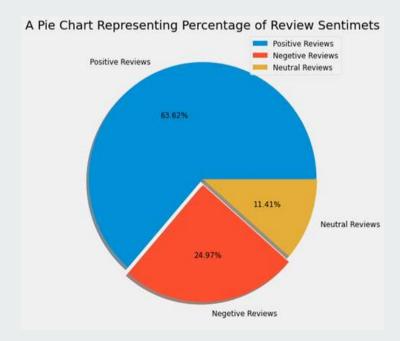

Does sentiment_subjectivity proportional to sentiment_polarity

Percentage of review sentiments

The number of **Unique** Apps from Play store and User reviews merged dataset are **816.**

From Sentiment column**, 63.62%** are **Positive, 24.97%** are **Negative** and **11%** are **Neutral** values.

A Pie Chart Representing Percentage of Review Sentimets

Conclusion

- Family category apps have the highest number of application.
- Communication gener has the most number of installs due to their huge user base
- Most of the apps are rated between 4 to five stars which shows that most of the apps are userfriendly.
- Free apps have high number of installs over paid apps.
- Maximum number of users give reviews to the applications, according to their experience.
- Sentiment subjectivity is not always proportional to sentiment polarity but in maximum number of case, shows a proportional behavior, when variance is too high or low.
- Most of the sentiments are positive with greatest number of percentage.