

Capstone Project Submission

Play Store App Review Analysis

Mobile apps are everywhere. They are easy to create and can be lucrative, because of these two factors, more and more apps are being developed. In this notebook, we will do a comprehensive analysis of the Android app market by comparing over ten thousand apps in Google play store across different categories

The Google play store app dataset consists of enormous data that can be used to create effective insights. There are various key factors that play a major role in the success & engagement from the user's end. Our problem statement is quite inevitable in comparison with the present Google Play store App market. Upon doing several pieces of research it can be seen that every day around 3000+ apps is being added to the play store library. Therefore enormous datasets & variety of insights can be concluded for business improvements.

In this EDA project we were provided with two datasets

1. Playstore.csv -> contains all the details of the applications of Google Play. There are 13 features that describe a given app.
2. User_reviews.csv -> contains 100 reviews for each app, most helpful first. The text in each review has been pre-processed and attributed with three new features
 - Sentiments (Positive, Negative, Neutral)
 - Sentiment Polarity
 - Sentiment Subjectivity

At first, we break down the datasets by importing necessary library classes, followed by checking unique values, converting the data types to similar objects, removing special characters as the analysis demands & making the entire dataset ready for analyzing & plotting actionable insights.

After examining null & missing values from the dataset we directly went deep into the visualization steps.

Some insights on which we worked are as follows:

- Top 10 categories which have the most number of application installations.

- Correlations check among 'Install' label and other labels of the datasets.
- Visualization using violin & density plot.
- Visualization using Histograms and Bar plots
- Insights over the content rating with the number of apps.
- Word cloud for better understanding of repetitive words
- Analyzing using a pie chart to understand the significant difference in percentage distributions of sentiments in the Data.
- Category Line plot for positive as well as negative sentiments opinion.

Contributor Roles

1. Nikhil Roewal (nikhilroewal89@gmail.com)

- Upload dataset to Google colab and explain dataset to team members.
- Analyze null values and filter them.
- Data cleaning.
- Correction of data types
- Data wrangling
- Data Visualizations
- Technical Write up
- PowerPoint presentation
- Project summary

2. Zayeem Ahmed (ahmedzayeem@gmail.com)

- Upload dataset to Google colab and explain dataset to team members.
- Analyze null values and filter them.
- Data cleaning.
- Correction of data types
- PowerPoint presentation
- Project summary

3. Sumanta Banerjee (sumantabanerjee77@gmail.com)

- Data wrangling
- Data Visualizations
- Technical Write up
- PowerPoint presentation
- Project summary

4. Ajit Padole (ajitpadole1998@gmail.com)

- Upload dataset to Google colab and explain dataset to team members.
- Analyze null values and filter them.
- Data cleaning.
- Correction of data types
- Data wrangling

5. Manoj Sridala (manojchidurala1@gmail.com)

- Upload dataset to Google colab and explain dataset to team members.
- Analyze null values and filter them.
- PowerPoint presentation
- Project summary

Please paste the Github Repo link:
https://github.com/ajitpadole/Playstore_Data_Analysis.git