Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:
Name - Gaurav Kumar Email - gauravrs195@gmail.com Contribution - Complete Project (Individual Project)
Please paste the GitHub Repo link.
Github Link:- https://github.com/gauravrs195/Capstone-1-Play-Store-App-Review-Analysis
Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)

Google Play, formerly known as Android Market, is the official distribution storefront for Android applications and other digital media, such as music, movies, and books, from Google.

In this capstone project we have compared thousands of applications across various categories. We have analyzed the data to discover key factors responsible for app engagement and success helping the developers to work and capture the android market.

This is my first EDA project. I made this project individually. In this project there are two csv files: the first csv file contains apps data, and the second csv file contains data of user review. I started this project from importing libraries then importing play store app csv file. First, I performed data cleaning on play store data. Handling null values in columns. We began with 'Android Ver' there are 3 null values. We remove all null values from this column. Then we moved to 'Current Ver' columns there are 8 null values. We remove all null values from the column. In Type column 1 null value. We convert this null value to 'Free' Type. In Rating columns 1470 null value we replace with median of all values. After Handling null value, the next step is Handling Duplicates. In 'App' columns we remove all duplicate values. Changing the data type of last updated column from string to date time. In String column, first drop the $ symbol from all the values. Then we can assign float datatype to those values. In the Installs column, we remove '+' and ','symbol from all the entities and convert into Int datatype. In the Size column contains data with different units. 'M' for MB and 'k' for KB. Convert all the values to a single unit (MB). Change Data type of reviews column from string to integer.

Import User Review csv file. Handling the NaN values in the User reviews dataframe. A total of 26868 rows contains NaN values in Translated Review column. Those apps which do not have a review (NaN value instead) also NaN values in the columns Sentiment, Sentiment Polarity, and Sentiment Subjectivity in most of the cases.

In the initial phase, we focused more on the problem statements and data cleaning, in order to ensure that we give them the best results out of our analysis.

- In the Sports category FIFA Soccer and 3D Bowling has the highest number of installs.
- Percentage of free apps = 92.20%
- Maximum apps in the play store are from Family category
- Category with the highest number of installs: Game
- Most popular app in the Play Store based on the number of reviews: Facebook
- I am Rich Premium app is the most expensive app in the play_store.
- Overall percentage of review sentiment in which Positive sentiment count is 64%, Negative 22% and Neutral 14%.
- Helix Jump has the highest number of positive reviews
- Angry Birds Classic has the highest number of negative reviews.