**Capstone Project Submission**

Play Store App Review Analysis

Smartphone applications are widely available. They are easy to design and may be lucrative, because of these two characteristics, more and more applications are being developed. In this notebook, we'll compare more than 10,000 apps from different categories in the Google Play store to conduct a thorough analysis of the Android app industry.

There is a vast amount of data in the Google Play store app dataset that may be leveraged to provide insightful analyses. The success and user engagement are greatly influenced by a number of critical aspects. Comparing our issue statement to the current App market on the Google Play store, it is essentially inevitable. In order to identify all the critical elements influencing app engagement and success, two datasets—one containing basic information and the other user reviews—are offered for exploration and analysis.

After doing extensive study, it was discovered that the Google Play store library is growing by more than 3000 applications every day. Hence, conclusions for company improvements may be drawn from vast datasets & a range of insights.

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.

They are:

* App :- Name of the App
* Category :- Category under which the App falls.
* Rating: - Application's rating on play store
* Reviews :- Number of reviews of the App.
* Size :- Size of the App.
* Install :- Number of Installs of the App
* Type :- Whether the App is free/paid
* Price :- Price of the app (0 if it is Free)
* Content Rating :- Appropriate Target Audience of the App.
* Genres:- Genre under which the App falls.
* Last Updated :- Date when the App was last updated
* Current Ver :- Current Version of the Application
* Android Ver :- Minimum Android Version required to run the App

1. 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
2. Sentiments (Positive, Negative, Neutral)
3. Sentiment Polarity
4. Sentiment Subjectivity

At first, we break down the datasets

* By importing necessary library classes
* Then by basic inspection of dataset.
* Then data cleaning and clean that attributes which have not to be useful and Replace all the Null values with the Average of their Columns or with not null values.
* Then followed by Data manipulation and handling duplicate data 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.
* As the data become ready for analyze we went into the visualization steps.

Some insights on which we worked are as follows:

1. Checking Correlation
2. Getting the average Rating of the Apps
3. Total number of applications in each category
4. Check the number of installs in each Category
5. Getting App Size Distribution
6. Check the number of installs on the basis of their size
7. Lets check the app on the basis of price in playstore.(Free or Paid)
8. let’s compare Category according to there Installs
9. Category and Rating
10. Category and Reviews
11. Review sentiments in all the app dataset
12. Let's see a more depth understanding of sentiment Polarity and Subjectivity.
13. Does sentiments Polarity is proportional to sentiments subjectivity.
14. Sentiment analysis of user reviews
15. Word Cloud

**Contributor Role**

**Nagendra Pratap Singh (nnnagendrasingh@gmail.com)**

* Upload dataset to Google colab.
* Analyze null values and filter them.
* Data cleaning.
* Correction of data types
* Data wrangling and manipulation
* Data Visualizations
* Technical documentation
* PowerPoint presentation
* Project summary

**Please paste the GitHub Repo link:** [**https://github.com/Nagendra-hash**](https://github.com/Nagendra-hash)