Low Level Design

Google Play Store Data Analysis

**Overview:**

The google play store is one of the largest and the most popular android app stores. It has an enormous amount of data that can be used to make the various prediction and Analytical models. This dataset consists of 13 different features which is used to for data analysis.

**Introduction:**

Mobile application are one of the fastest growing segments of downloadable software application market. Out of all of the markets I choose Google Play Store due to its increasing popularity and recent fast growth. One of the main reasons of its popularity is the fact the about 81% of the apps are free of cost. Developers and users play key roles in determining the impact that market interactions have on future technology. However, the lack of a clear understanding of the inner working and dynamic of popular app markets impacts both the developers and users.

**Analysis Methodology:**

Our analysis approach is divided into three phases:

1. Data extraction (From the provided dataset)
2. Data cleansing
3. Visualization
4. Dashboard / Report Making

**Data Extraction** 🡪 This process deals with data collection in this we gathered the dataset provided by Ineuron.

**Data Cleansing** 🡪 This process deals with data manipulation, data which is collected in the previous step consist of NaN values, unwanted Column, outliers, handling these error is also known as Exploratory Data Analysis (EDA).

**Visualization** 🡪 This step is also one of the part of EDA, Generally after data cleansing. Various observations are performed over the data in order to find some relationships amongst the (feature – feature) or (feature – labels). And during this process we use different types of chart, graph in order to make better visualizations.

**Creation dashboard** 🡪 Creation dashboard come with lots of challenges like manipulation of data obtained from python EDA still some changes have been done through Excel, power Query. For converting our numerical data analysis / observations into visuals we used power BI for dashboard report creation.

**Technology Used:**

1. Excel -> Minor manipulation / original dataset reading.
2. Python
   * Pandas -> data frame manipulation
   * NumPy -> array
   * Seaborn -> visualization
   * Matplotlib -> visualization
   * Statistics -> math
3. Power BI -> creating Dashboard visuals.
4. Power Query -> manipulation over excel retrieved table.
   * M language