# **Google Play Store Project**

Hagar Abdallah

Nerdeen Muhammed

Nadeen Ahmed

➤ **Objectives**

To gain insights into various aspects of app performance, user engagement, and market trends.

➤ **Story**

we divided our analysis into 3 parts:  
-Category wise analysis: Top categories, Content rating distribution, Average rating of categories, Revenue, Average app size per category, Sentiment distribution across all categories.  
-Applications wise analysis: An overview of all applications with different parameters, Bar plots for Installs, Ratings, Reviews and Size.  
-Sentiment analysis of user reviews: Sentiment count by category, Top 5 most positively and negatively reviewed categories and Subjectivity distribution across categories:  
                0-0.25 ~ Factual opinion  
                0.25-0.5 ~ Moderately factual  
                0.5-0.75 ~ Moderately personal  
                0.75-1 ~ Personal opinion  
  
**➤ Tools Used**  
• Microsoft Excel  
• Jupter Notebook (Python libraries - Numpy, Pandas, sklearn)  
• SQL Server  
• PowerBI  
  
**➤ Data Source**  
The datasets were taken from Kaggle.  
• playstore\_apps.csv - It contains the basic details of the app with the following columns: App, Category, Rating, Reviews, Size, Installs, Type, Price, Content rating, Genres, Last Updated, Current version.  
• playstore\_reviews.csv - It contains the user reviews for the respective app.  
App: It contains the name of the app  
Translated\_Review: It contains the English translation of the review dropped by the user of the app Sentiment: Sentiment\_Polarity: Sentiment\_Subjectivity:

**➤ Steps Involved**  
• Data Cleaning and Transforming-  
To prepare the data for analysis, we performed several data cleaning steps. First, we checked for null values to ensure that there were no missing pieces of data. We also identified and removed outliers, especially in cases where ratings had a specific range. If null values existed and needed to be replaced with some value, we inserted values using the sklearn library . We modified data values, for size column we removed symbols like K and M and replaced them with their respective values. We changed data types of columns to ensure that the data was in a format that was easy to analyze. Finally, we dropped duplicates and checked unique values to ensure the integrity of the data and that it was compliant with the intended datatype. A new column 'Revenue' was added by multiplying the prices and installs value for each record.  
  
• Sentiment analysis   
polarity, subjectivity, and sentiment. The polarity column indicates the sentiment of each review as positive, negative, or neutral. The subjectivity column measures the degree of personal opinion in the review on a scale of 0 to 1, where 0 represents a factual statement and 1 represents a highly opinionated statement. The sentiment column is an aggregate score between -1 and 1 that takes into account both polarity and subjectivity. A score of -1 indicates an extremely negative review, 1 indicates an extremely positive review, and 0 indicates a neutral review. These insights can be useful for improving product quality and customer satisfaction.

• Insights using SQL Server  
[GooglePlayStore](http://googleplaystore.sql)  
  
**➤ Data Visualization**

Microsoft power Bi

**➤ Conclusion**

There are about 33 categories and 0ver eight thousand applications in the given dataset with an overall average rating of about 4 and an average application size of about 19 MB.

For content rating, majority of the applications fall under ‘Everyone’ followed by ‘Teen’ and ‘Mature17+’. The least being ‘Adults only 18+’ and Unrated.

Majority of the apps, about 92% are listed as free while about only 7 % being of paid type. ‘Family’ followed by ‘Game’ have the highest number of paid apps.

‘Family’ is the top category with the highest number of applications followed by ‘Game’ while ‘Beauty’ and ‘Comics’ being the lowest.

While ‘Family’ is the top category, it lags behind in revenue. ‘Lifestyle’ made the highest revenue followed by ‘Game’ and ‘Family’ while ‘Comics’ and ‘House and home’ made the least revenue.

‘Events’ is the top category with the highest average rating of 4.45 followed by ‘Education’ and ‘Art and design’.

‘Google news’ and ‘Subway surfers’ are the most popular apps in terms of installs with about 1000M installs.

‘I am rich’ is the highest earning app followed by it’s premium version and ‘Hitman Sniper’.

Clash of Clans, Subway Surfers and Clash Royale are the most reviewed apps with about 44 million+ reviews.

The largest app on the Play Store is ‘Hungry Shark Evolution’ belonging to ‘Game’ category with about 95 MB.

The total number of reviews are 37,427 out of which 19,222 are of Free apps and 225 reviews are of Paid apps.

Majority of the reviews have a positive sentiment followed by Negative and Neutral.

‘Game’ has the highest number of reviews when it comes to Positive and Negative Sentiments followed by ‘Family’.

Sentiment Subjectivity is directly proportional to sentiment polarity in maximum reviews.

40.02% of the Total reviews are Moderately Personal, followed by Moderately Factual at 27.54% and then by Factual Opinion at 16.53% and 15.90% are Personal Opinion.

Bivariate analysis of the following presented some deeper insights as depicted -

Price vs Rating

Rating and Last updated

Type and Last updated