# Google Play Store Data Analysis - Project Documentation

## 1. Project Overview

### Objective

The goal of this project is to analyze Google Play Store data to extract meaningful insights about app ratings, installs, revenue, and user engagement. The analysis aims to identify trends, compare free vs. paid apps, and determine factors influencing app success.

## 2. Dataset Details

### Dataset Name

Play Store Data Cleaned.csv

### Source

Google Play Store data

### Key Columns

- App Name  
- Category  
- Rating  
- Reviews  
- Installs  
- Type (Free/Paid)  
- Price  
- Content Rating  
- Revenue

## 3. Data Cleaning & Preprocessing

### Handling Missing Values

- Removed rows with missing critical values (Rating, Installs, Reviews).  
- Imputed missing values where necessary (e.g., replacing null ratings with category-wise averages).

### Data Type Conversions

- Converted Installs column from object to integer by removing `+` and `,` symbols.  
- Converted Price and Revenue columns to float for financial analysis.  
- Changed Last Updated column to Date format.

### Outlier Handling

- Removed extreme outliers in Revenue and Installs to avoid skewed analysis.  
- Filtered apps with unrealistic ratings (e.g., greater than 5).

### Feature Engineering

- Derived Revenue using `Installs \* Price` for paid apps.  
- Categorized apps based on installs (e.g., Low, Medium, High installs).  
- Extracted Year from the 'Last Updated' column to analyze trends over time.

## 5. Power BI Dashboard Implementation

### Visualizations & Interactivity

- Used Bar, Column, Pie, Scatter, and Line Charts.  
- Implemented Slicers for Category, Free/Paid, Installs Range, and Rating Range.  
- Enabled drill-through and cross-filtering for deeper analysis.

### DAX Measures Used

```DAX  
Avg\_Rating = AVERAGE(PlayStore[Rating])  
Total\_Revenue = SUM(PlayStore[Revenue])  
Apps\_Per\_Category = COUNT(PlayStore[App])  
Free\_Paid\_Percentage =   
 DIVIDE(  
 COUNTROWS(FILTER(PlayStore, PlayStore[Type] = "Free")),  
 COUNTROWS(PlayStore),  
 0  
 )  
```

## 6. Conclusion & Future Improvements

### Key Findings

✅ Education & Productivity apps perform well in ratings.  
✅ Social media & entertainment apps lead in installs.  
✅ Paid apps have lower installs but generate higher revenue per user.  
✅ Frequent updates contribute to app success.  
✅ There is no strong correlation between ratings and installs.

### Future Enhancements

- Include user sentiment analysis using NLP on app reviews.  
- Use machine learning to predict app success based on historical data.  
- Implement a real-time dashboard with API integration for live tracking.

## 7. Tools Used

- Python (Pandas, NumPy, Plotly) - Data Preprocessing & Analysis  
- Power BI - Dashboard & Visualization  
- Excel - Initial Data Cleaning

## 8. References & Credits

- Data Source: Google Play Store  
- Python Libraries: Pandas, NumPy, Matplotlib, Seaborn  
- Power BI Dashboard for visualization & insights

🚀 This project helped in understanding app success metrics, trends, and user preferences in the Google Play Store.