Project: PlayStore Data Analysis Bootcamp

Project Title: Analyzing Google Play Store App Trends using Python & Power BI.

Project Description:
This project focuses on performing an end-to-end data analysis of the Google Play Store dataset.
You'll learn how to clean and preprocess real-world data using Python (Pandas, Matplotlib, Seaborn), explore insights through visual analysis, and finally create an interactive dashboard in Power BI that summarizes business insights such as app popularity, ratings, and pricing trends.

What is This Project About?
This project aims to help you understand the complete data analytics workflow — from raw data to actionable insights. You'll develop core data-handling, analysis, and visualization skills used in the analytics industry.
You will learn to:

- 🧹 Clean and preprocess real-world datasets using Python and Pandas.
- 📊 Explore and visualize patterns using Seaborn and Matplotlib.
- 💡 Derive business insights and design a professional Power BI dashboard.
- 🧠 Think like a data analyst — interpret trends, correlations, and KPIs.

What You Have to Do (Step-by-Step):

Day 1 – Data Loading & Cleaning

1. Use Google Colab and kagglehub to import the Google Play Store dataset from Kaggle.
2. Inspect the dataset using .info(), .describe(), and check for missing or inconsistent values.
3. Handle missing data, remove duplicates, and clean text-based numeric columns (Price, Installs, Reviews).
4. Convert data types appropriately and filter invalid entries (e.g., ratings > 5).
5. Add a derived column isPaid (1 = Paid, 0 = Free).
6. Save the cleaned dataset as cleaned_playstore.csv.

Day 2 – Exploratory Data Analysis (EDA)

1. Load your cleaned dataset into Colab.
2. Explore top app categories and their distributions.
3. Visualize rating trends using bar plots, histograms, and scatter plots.
4. Perform correlation analysis between Rating, Installs, Reviews, and Price.

5. Interpret insights such as:
    - Which categories are most popular?
    - Do paid apps have higher ratings?
    - What factors influence app installs?

Day 3 – Dashboard Creation (Power BI Web)

1. Go to https://app.powerbi.com and sign in with a Microsoft account.
2. Upload your cleaned_playstore.csv dataset (Create → Upload File → Local File).
3. Create visuals:
    - 📊 Bar Chart – Top 10 app categories by count
    - 🍩 Donut Chart – Free vs Paid app distribution
    - 📈 Line Chart – Average rating by category
    - 💳 Cards – KPIs (Average Rating, Total Installs)

What to Submit:
Students must submit following to be eligible for certification

- Python Notebook (Cleaning + EDA)
- PPT & VIDEO RECODING

Tips for Students: To make your project stand out:

- Always comment your code in Colab and include short markdown explanations.
- In Power BI, keep visuals clean and consistent (color theme, font size, labels).
- Add meaningful titles to each chart (e.g., "Top Categories by App Count").
- Use insight captions: one short takeaway under each visual.
- Verify all numeric columns (especially Price and Installs) are in correct format before uploading.
- Save your work frequently in Power BI Web (autosave can lag on poor internet).
- Before submission, ensure all visuals update dynamically when using slicers.

Deadline:
Submit your project by 25 October 2025, 11:59pm in
Submission link - (https://forms.gle/MDmxuQAk5ZRBhZSa6)

.