

INTERNSHIP PROJECT REPORT ON

**Google Play Store Apps Analysis**

Submitted By:

KENNETH AMAN (3PD21AI012)

BACHELOR OF ENGINEERING IN

ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

POOJYA DODDAPPA APPA COLLEGE OF ENGINEERING KALABURAGI

Unified Mentors Internship :

Data Analytics 16 weeks 10th Oct to 10th Feb(online) Unified Mentors is a professional training platform specializing in data analytics, AI, ML, DATA Science and more. The organization focuses on equipping aspiring interns with real-world experience, hands on projects, and industry relevant skills under expert mentorship

Address:

Plot No 247B/7,Dhola Kuva Road Hansi City, Hansi Hisar, Haryana 125033 India

Contact Information:

• Phone: +91 6283 800330

• Email: hello@unifiedmentor.com

**Google Play Store Apps Analysis Report**

**Project Overview**

This project focuses on analysing the Google Play Store Apps dataset to extract meaningful insights that can help developers and businesses optimize their app strategies. The dataset consists of various app-related features, including category, rating, size, number of installs, and more.

**Problem Statement**

The primary goal of this project is to conduct an exploratory data analysis (EDA) of Google Play Store apps to:

* Identify trends in app categories, ratings, and installations.
* Understand the correlation between ratings, reviews, and installations.
* Detect missing values, duplicates, and anomalies in the dataset.
* Provide recommendations for app developers based on the findings.

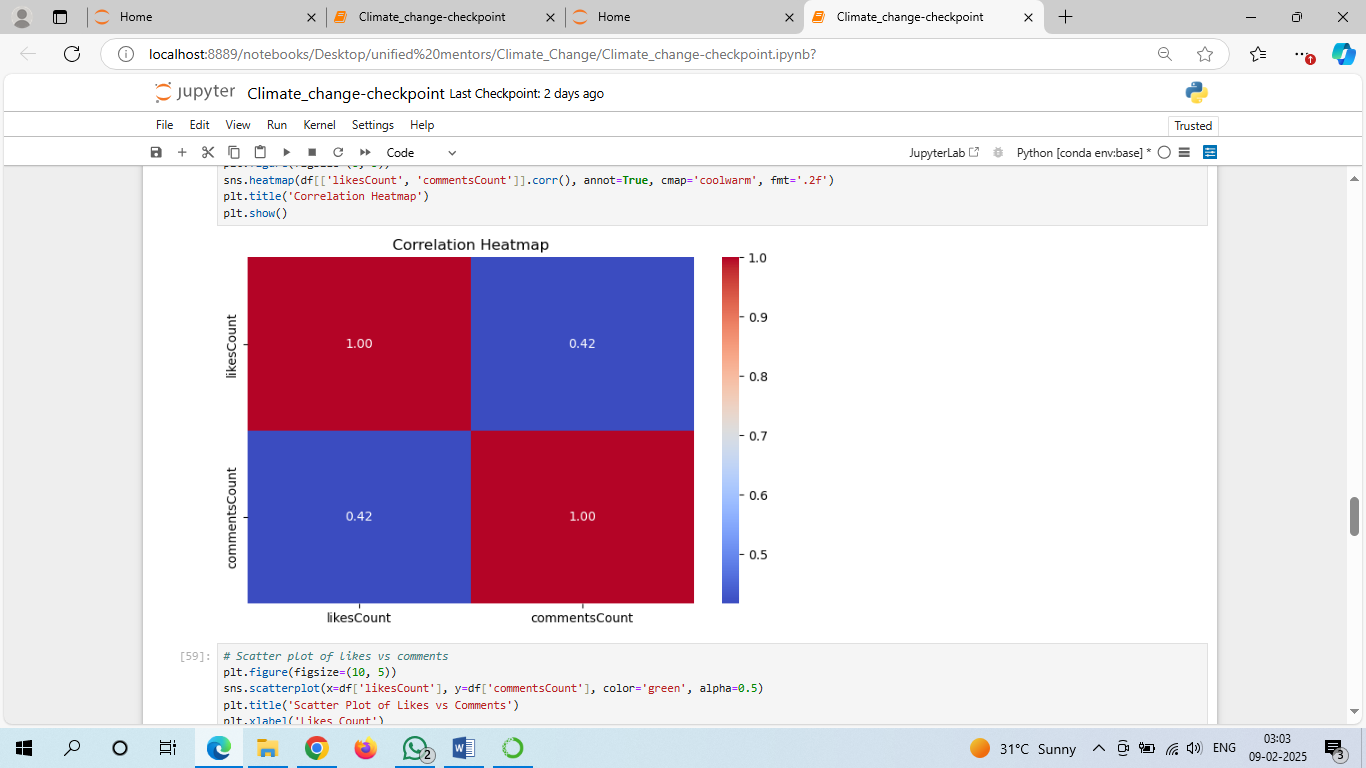
**Tools Used**

* **Python**: Data analysis and visualization.
* **Pandas**: Data manipulation and cleaning.
* **Matplotlib & Seaborn**: Data visualization.
* **SQL**: Querying and managing structured data.
* **Excel**: Preliminary data inspection and reporting.

**Data Sources**

The dataset was sourced from publicly available Google Play Store app data. The data includes attributes such as:

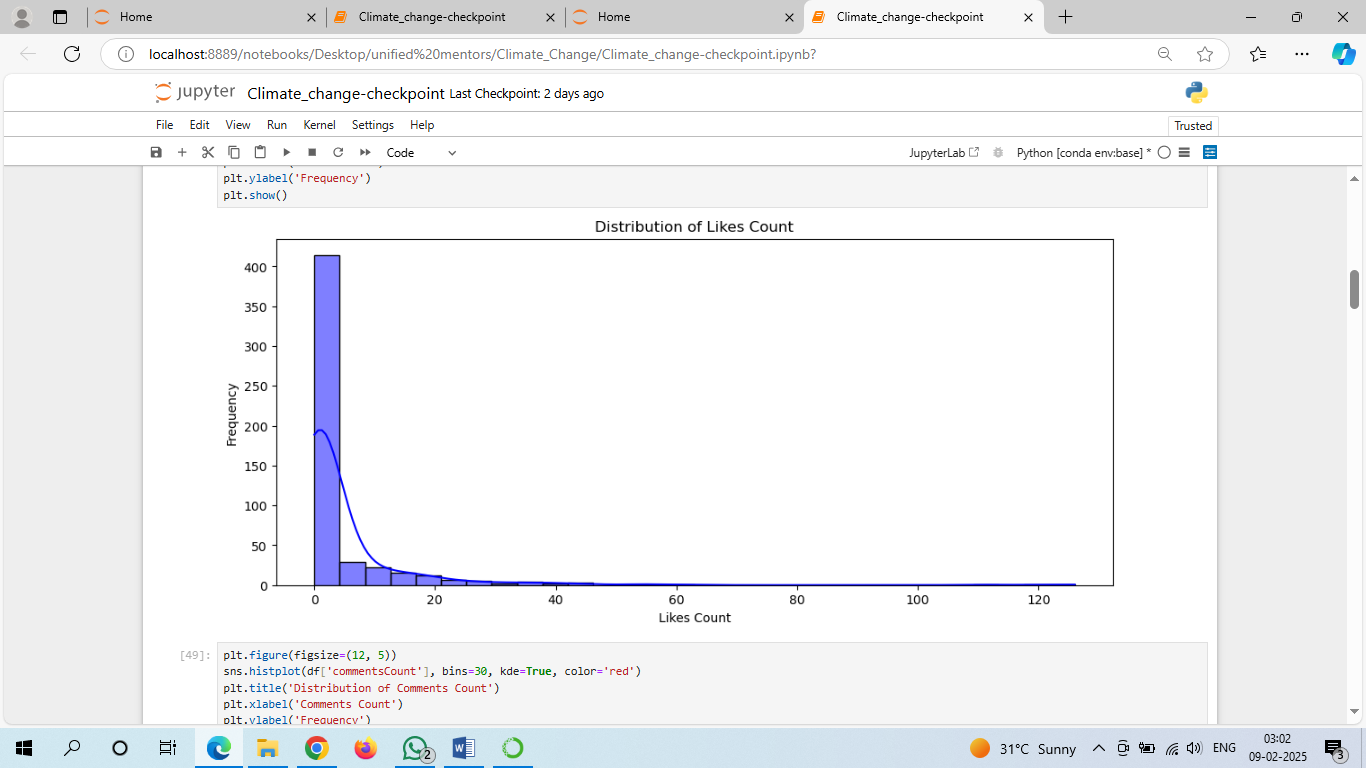
* App Name
* Category
* Rating
* Reviews
* Size
* Installs
* Type (Free/Paid)
* Price
* Content Rating
* Last Updated
* Current Version
* Android Version



**Data Analysis and Cleaning**

**Data Cleaning Steps**

1. **Handling Missing Values**:
   * The Rating column had missing values, which were filled based on the average ratings within each install category.
   * The Size column contained values such as "Varies with device," which were replaced with NaN values.
   * The Current Version and Android Version columns had minor missing values, which were dropped as they had a negligible impact on analysis.
2. **Data Type Conversions**:
   * Reviews and Installs were converted to integers.
   * Size was transformed into bytes for consistency.
   * Price values (e.g., '$4.99') were cleaned and converted to numeric values.
3. **Removing Duplicates**:
   * 483 duplicate rows were identified and removed to prevent data redundancy.



**Exploratory Data Analysis (EDA)**

**Visualizations and Insights**

1. **Distribution of Ratings**:
   * The average rating across all apps was **4.19**, with most apps receiving ratings between **4.0 and 4.5**.
   * Some anomalies were detected where apps had ratings above 5.0, which were removed.
2. **Top App Categories**:
   * The **Family** category had the highest number of apps, followed by **Games** and **Tools**.
   * The **Game** category had the highest total number of installations, indicating its popularity among users.
3. **Relationship Between Installs and Ratings**:
   * Apps with higher installs tend to have better ratings.
   * Free apps generally have higher downloads compared to paid apps.
4. **Most Popular Apps by Installs**:
   * The top categories by installs were **Games, Communication, and Social Media**.
   * Apps in the **Productivity and Business** categories also had significant user engagement
5. **Correlation Analysis**:
   * **Reviews and Installs** had a strong positive correlation (**0.64**), meaning that apps with more installs tend to have more reviews.
   * **Size and Installs** showed a weak correlation, indicating that app size is not a major factor in determining popularity.

**Results and Key Findings**

1. **The most popular app categories** in terms of installs are **Games, Communication, and Social**.
2. **Paid apps have significantly lower installs** compared to free apps, suggesting that users prefer free applications.
3. **Higher-rated apps tend to be in categories like Education, Events, and Books & Reference**.
4. **The number of installs is a strong indicator of user engagement, with apps that have more installs receiving more reviews**.
5. **Most apps have ratings between 4.0 and 4.5**, indicating general user satisfaction.

**Recommendations**

1. **For developers targeting high engagement**, focusing on **Games, Communication, and Social categories** is beneficial.
2. **Optimizing for user experience and gathering reviews** is crucial as reviews have a strong correlation with installs.
3. **Paid apps may need promotional strategies**, such as discounts or trial versions, to encourage more downloads.
4. **Ensuring apps are lightweight and optimized** can be beneficial, but size alone does not strongly impact downloads.
5. **Regular updates and user feedback implementation** can improve ratings and user retention.

**Conclusion**

The analysis of the Google Play Store dataset provides valuable insights into the app ecosystem. The trends observed suggest that user engagement, app categories, and pricing strategies play crucial roles in determining app success. Future work could involve **sentiment analysis of user reviews**, **time-series analysis of app trends**, and **predictive modelling to forecast app ratings and installs**.