**Internship Report: Real-Time Google Play Store Data Analytics**

Intern: Anjana

Internship Provider: NullClass

Project: Real-Time Google Play Store Data Analytics Dashboard

**1. Introduction**

This report summarizes the work completed during the internship with NullClass, which focused on developing a real-time data analytics dashboard for Google Play Store applications. The project utilized Python, Streamlit, and Plotly to create an interactive and dynamic dashboard. The primary goal was to apply complex data cleaning, filtering, and visualization techniques, including the implementation of strict time-based chart display restrictions, to analyze Google Play Store data.

**2. Background**

The Google Play Store contains a vast number of mobile applications, each associated with key metrics such as ratings, reviews, installs, and size. Analyzing this data can reveal critical insights into app performance, user preferences, and market trends. The central challenge of this internship was to build a functional and interactive dashboard that adhered to a strict set of data filters, transformations, and display rules.

**3. Learning Objectives**

The internship was designed to achieve the following learning objectives:

* Gain hands-on experience in **data preprocessing and cleaning** using Python libraries like Pandas.
* Implement **interactive visualizations** with Plotly and Streamlit.
* Apply complex, **conditional filtering** to large datasets.
* Integrate **time-based data visibility** rules, specifically handling the Indian Standard Time (IST) zone.
* Understand and implement **translation integration** for multilingual category displays.
* Work with a **real-world dataset**, specifically the Kaggle Google Play Store Apps dataset.
* Adhere to professional **coding and documentation practices**.

**4. Activities and Tasks**

The internship was structured around five distinct tasks, each requiring the creation of a specific chart with unique data filters and display schedules.

**Task 1: Grouped Bar Chart**

* **Objective:** To compare the average rating and total reviews for the top 10 categories by installs.
* **Filters:** Applied a minimum rating of 4.0, a minimum size of 10MB, and filtered for apps last updated in January.
* **Time Restriction:** Chart was displayed only between 3 PM and 5 PM IST.

**Task 2: Choropleth Map**

* **Objective:** To visualize global app installs by category.
* **Filters:** Restricted the view to the top five categories with over one million installs and excluded any categories starting with the letters A, C, G, or S.
* **Time Restriction:** Chart was displayed only between 6 PM and 8 PM IST.

**Task 3: Dual-Axis Chart**

* **Objective:** To compare average installs and revenue for free versus paid apps.
* **Filters:** A multi-condition filter was applied, including a minimum of 10,000 installs, a minimum of $10,000 in revenue, an Android version greater than 4.0, a size larger than 15MB, a content rating of "Everyone," and an app name length of 30 characters or less.
* **Time Restriction:** Chart was displayed only between 1 PM and 2 PM IST.

**Task 4: Time Series Trend**

* **Objective:** To display total installs over time by category, highlighting periods of significant growth.
* **Filters:** Apps not starting with X, Y, or Z; categories starting with E, C, or B; over 500 reviews; and no "S" in the app name.
* **Translations:** The categories "Beauty," "Business," and "Dating" were translated into Hindi, Tamil, and German, respectively.
* **Time Restriction:** Chart was displayed only between 6 PM and 9 PM IST.

**Task 5: Bubble Chart**

* **Objective:** To show the relationship between app size and average rating, with bubble size representing total installs.
* **Filters:** Rating greater than 3.5, a specific list of categories, over 500 reviews, no "S" in the app name, a subjectivity score greater than 0.5, and over 50,000 installs. The "Game" category was specifically highlighted in pink.
* **Translations:** The same category translations from Task 4 were applied.
* **Time Restriction:** Chart was displayed only between 5 PM and 7 PM IST.

**5. Skills and Competencies**

This internship provided the opportunity to apply and develop a range of technical and professional skills:

* **Python Programming:** Utilized for data manipulation and dashboard logic.
* **Data Cleaning & Transformation:** Proficiently used Pandas and NumPy for preprocessing data.
* **Interactive Dashboards:** Created a dynamic user interface using Streamlit.
* **Data Visualization:** Generated complex and informative charts with Plotly Express.
* **Time Zone Handling:** Implemented the pytz library to handle IST for time-based conditional rendering.
* **Multilingual Support:** Implemented category translations for internationalization.
* **Version Control:** Used Git and GitHub to manage the project's code and track changes.
* **Professional Documentation & Reporting:** Maintained clear project documentation and prepared this report.

**6. Challenges and Solutions**

* **Challenge 1: Insufficient Data:** Some initial data filters resulted in "No data" being returned for certain charts.
  + **Solution:** The dataset was replaced with a richer Kaggle dataset that contained a broader range of free and paid apps across all required categories.
* **Challenge 2: Inconsistent Data Formatting:** App categories had inconsistent casing (e.g., "GAME" vs. "Game").
  + **Solution:** A case-insensitive matching logic was implemented for all category-based filters to ensure accuracy.
* **Challenge 3: Overly Strict Filtering:** The combination of multiple strict filters sometimes left very few rows of data for visualization.
  + **Solution:** The filters were carefully verified against the dataset to ensure they were correctly applied. Category mappings were also used to address any inconsistencies in the data.

**7. Outcomes and Impact**

The internship successfully culminated in the delivery of a fully functional, real-time data analytics dashboard. All project requirements, including complex filters, category translations, and time-based display restrictions, were met with precision. This project is now a significant portfolio piece, demonstrating the ability to build professional-grade dashboards with real-world constraints. The successful completion of all tasks also satisfied the requirements for stipend eligibility.

**8. Conclusion**

This internship provided valuable, hands-on experience in building a professional data analytics project from concept to completion. The project simulated a real work environment where strict requirements and deadlines had to be met with accuracy. The experience significantly enhanced my skills in Python, data analysis, visualization, and project management, preparing me for future roles in the data science and analytics field.