INTERNSHIP REPORT

Project Title: *Real-Time Google Play Store Data Analytics Using Python*

Intern Name: *Vishal Tiwari*

Organization: *Null Class*

Internship Duration: *26 June 2025 – 26 September 2025*

Role: *Data Analytics Intern*

INTRODUCTION

As a Data Analytics Intern at Null Class, I had the opportunity to work on a real-time data visualization project titled *"Real-Time Google Play Store Data Analytics Using Python."* The objective was to analyze various aspects of mobile applications available on the Google Play Store using Python, data analysis libraries, and visualization tools. I built a fully functional dashboard that includes various graphs representing filtered data insights, which dynamically respond to time-based logic and complex filtering conditions. This report summarizes my experience, technical contributions, and personal growth over the three-month internship period.

BACKGROUND

With the exponential growth of mobile applications, app developers and marketers need tools to analyze user engagement, trends, and overall app performance. The Google Play Store is a rich source of publicly available data, offering insights into user ratings, downloads, reviews, and more. Through this internship, I was tasked with extracting actionable information from this data and transforming it into meaningful visualizations using Python and advanced data processing techniques. The focus was on real-time analytics, where graphs would appear or disappear based on specific time frames (Indian Standard Time), simulating real business dashboard behavior.

LEARNING OBJECTIVES

At the start of the internship, I set clear goals to achieve by the end of the three-month period. These included:

- Understanding the real-time processing of app data from the Google Play Store.
- Learning how to filter and transform large datasets using Python.
- Creating interactive visualizations using libraries like Plotly, Matplotlib, and Streamlit.
- Applying time-sensitive logic in dashboards to mimic live business intelligence systems.
- Improving my programming, analytical thinking, and problem-solving skills.
- Practicing version control and team collaboration using GitHub.

ACTIVITES AND TASKS

I was assigned six complex and highly specific tasks that involved designing different visualizations with multiple filters, including time-bound logic. Here's a breakdown of each:

Task 1: Word Cloud for 5-Star Reviews (Health & Fitness Apps)

- **Objective:** Generate a word cloud from 5-star reviews.
- **Filters Applied:**
    - Only 5-star reviews were included.
    - App names and common stopwords were removed.
    - Apps were filtered by the **Health & Fitness** category.
- **Tech Used:** Python (wordcloud, pandas, matplotlib)

Purpose: To identify the most common positive feedback terms, which help developers understand what users like most in health & fitness apps.

Task 2: Grouped Bar Chart (Top 10 App Categories by Installs)

- **Objective:** Compare the **average rating** and **review count** for the top 10 app categories.
- **Filters Applied:**
  - Rating > 4.0
  - Size > 10 MB
  - Last update should be in January
  - Time-based condition: Visible only between **3 PM and 5 PM IST**
- **Tech Used:** Plotly for grouped bar chart; datetime for time logic.

Purpose: To allow stakeholders to compare categories based on popularity and user satisfaction within certain constraints.

Task 3: Choropleth Map (Global Installs by Category)

- **Objective:** Show a world map visualizing install distribution.
- **Filters Applied:**
  - Top 5 app categories (excluding those starting with A, C, G, S)
  - Only categories with **>1 million installs**
  - Time-based condition: Active between **6 PM and 8 PM IST**
- **Tech Used:** Plotly Choropleth, GeoJSON mapping, pandas for data aggregation.

Purpose: This task visualizes how certain app categories perform globally while ignoring overly generic or dominating ones.

Task 4: Dual Axis Chart (Free vs Paid Apps)

- **Objective:** Compare **average installs** and **revenue** for free vs paid apps in the **top 3 categories**.

- **Filters Applied:**
  - Installs > 10,000
  - Revenue > $10,000
  - Android version > 4.0
  - Size > 15MB
  - Content Rating: Everyone
  - App name length < 30 characters
  - Time-based condition: Shown only **1 PM to 2 PM IST**
- **Tech Used:** Plotly (Secondary y-axis), pandas

Purpose: To understand how monetization strategies affect user reach and financial performance across categories.

Task 5: Time Series Chart (Monthly Installs by Category)

- **Objective:** Plot a time series showing installs over months segmented by app category.
- **Filters Applied:**
  - App names must not start with X, Y, or Z
  - Category must start with E, C, or B
  - Reviews > 500
  - App name must not contain letter "S"
  - Translations:
    - "Beauty" ⟶ Hindi
    - "Business" ⟶ Tamil
    - "Dating" ⟶ German

- Time-bound: **6 PM to 9 PM IST**
- Highlight 20%+ MoM growth with shaded areas

- **Tech Used:** Plotly line chart, dictionary mapping for translations, time series smoothing.

Purpose: Show category-specific growth while also managing multilingual label presentation and pattern analysis.

Task 6: Bubble Chart (Size vs Rating, Bubble = Installs)

- **Objective:** Visualize relationship between app size and average rating.
- **Filters Applied:**
  - Rating > 3.5
  - Categories: Game, Beauty, Business, Comics, Communication, Dating, Entertainment, Social, Event
  - Reviews > 500
  - App name must not contain “S”
  - Sentiment subjectivity > 0.5
  - Installs > 50k
  - Translations:
    - Beauty ⟶ Hindi
    - Business ⟶ Tamil
    - Dating ⟶ German
  - Highlight Game category bubbles in pink
  - Time-based: Visible only **5 PM to 7 PM IST**
- **Tech Used:** Plotly Bubble chart, sentiment analysis using TextBlob

Purpose: To analyze whether size influences app ratings and popularity, with detailed filtering and visualization features.

SKILLS AND COMPENTENCIES DEVELOPED

Throughout this internship, I improved in several key areas:

Skill	Description
Python	Used libraries like pandas, numpy, plotly, matplotlib, datetime, wordcloud, and TextBlob.
Data Cleaning	Learned how to preprocess messy data, handle missing values, and apply complex filters.
Data Visualization	Built aesthetically pleasing and functional graphs using Plotly and Streamlit.
Time-Based Logic	Implemented time filtering so charts appear/disappear based on IST.
Dashboard Development	Used Streamlit to build a live dashboard integrating all graphs.
Multilingual Presentation	Translated app category labels using custom mapping.
Version Control	Used Git and GitHub to manage and submit code efficiently.

FEEDBACK AND EVIDENCE

- **GitHub Repository**: [Uploaded]
- **Dashboard Demo (if hosted)**: [Optional - Netlify or Vercel]
- **Daily Work Reports** submitted consistently on: https://dailyreport.nullclass.com/

- **Screenshots of all visualizations** have been added in the appendix (if you're submitting a Word/PDF file with images)

CHALLENGES AND SOLUTION

Challenge	Solution
Dealing with complex filtering	Broke filters into manageable logical expressions and chained them in pandas.
Time-based chart rendering	Used datetime.now(pytz.timezone("Asia/Kolkata")) to control rendering in Streamlit.
Translating category names dynamically	Used a dictionary and replace() function in pandas to convert specific values.
Filtering based on character count	Used len() and apply() to enforce app name limits.
Highlighting growth in time series	Calculated percent change month-over-month and used Plotly's fill for shaded areas.

OUTCOMES AND IMPACT

By the end of this internship, I had created a fully functional, filter-heavy, real-time analytics dashboard. I:

- Developed a strong command over real-world data manipulation and visual representation.
- Learned how to mimic business intelligence systems with real-time behavior.
- Produced a professional portfolio project that I can showcase to potential employers.

- Became more confident in tackling open-ended, independent assignments with limited guidance.

DATASET

https://www.kaggle.com/datasets/lava18/google-play-store-apps?resource=download&select=googleplaystore.csv

---

CONCLUSION

My internship with Null Class was a significant step in my career development. I not only enhanced my Python and data visualization skills but also learned how to structure and execute a real-world analytics project independently. The exposure to complex filtering, time-bound logic, and multilingual requirements challenged me to think critically and implement efficient solutions. I look forward to applying these skills in future roles in the data analytics and tech industry.