

Internship Report

Internship Title: Learn to Build Real-time Google Play Store Data Analytics - Python

Internship Organization: NullClass

Internship Duration: 19-01-2025 to 19-04-2025 (3 Months)

Intern's Name: Monika Gaha

Introduction

This report provides an overview of my internship experience at NullClass, where I worked on developing advanced data visualizations and an interactive dashboard using Python, Plotly, and HTML. The internship focused on building a real-time data analytics dashboard using Python and Plotly, specifically targeting Google Play Store data. Background

Background

The internship provided an opportunity to engage with real-world datasets, perform exploratory data analysis (EDA), and build visual tools for insight generation. The scope included creating charts and dashboards that dynamically reflect business logic, time constraints, and conditional filtering.

Learning Objectives

- Apply data analysis techniques to large datasets.
- Create interactive visualizations using Plotly and Dash.
- Implement advanced filtering and logic-driven rendering of graphs.
- Develop modular, reusable code for dashboards.
- Generate visual insights that adapt to time-based conditions.

Activities and Tasks

The core tasks completed during the internship include:-

Task 1: Scatter Plot

Visualized the relationship between revenue and number of installs for paid apps. Color-coded by category with trendline.

Task 2: Grouped Bar Chart

Compared average rating and total reviews for top 10 categories with filters (rating ≥ 4.0 , size $\geq 10\text{MB}$, updated in Jan). This plot was shown only between 3 PM–5 PM IST.

Task 3: Choropleth Map

Created a map visualizing global installs by category for top 5 categories excluding ones starting with A, C, G, S. Displayed only between 6 PM–8 PM IST.

Task 4: Dual Axis Chart

Compared installs and revenue for free vs paid apps (top 3 categories), with various filters (install count, revenue, content rating, etc.). Displayed only between 1 PM–2 PM IST.

Task 5: Violin Plot

Distribution of ratings per category for apps containing letter “C” in name and with more than 10 reviews. Time-based visibility (4 PM–6 PM IST).

Task 6: Correlation Heatmap

Correlation between installs, reviews, and ratings for filtered apps. Displayed only between 2 PM–4 PM IST.

Task 7: Time Series Line Chart

Trends in total installs by category with shaded growth zones (20%+ MoM growth). Filtered by content rating, name pattern, and installs. Visible only between 6 PM–9 PM IST. Each visualization was saved separately as an HTML file, and only relevant ones were included in the dashboard based on the time filters. All plots were also saved independently as HTML files for external review, and a centralized dashboard was developed with logic to **exclude** time-restricted plots outside their valid visibility windows.

Skills and Competencies

- Python (Pandas, NumPy, datetime)
- Data Visualization (Plotly Express, Plotly Graph Objects)
- HTML and Web Integration
- Timezone-aware logic handling (IST)
- Data Cleaning and Feature Engineering
- Code modularization and reusable utility design

Feedback and Evidence

- Plots were successfully rendered and saved individually.
- Dashboard logic respected time-window constraints.
- Visualizations aligned with expected insights and business rules.
- Debugging and enhancements were continuously performed based on trial runs.

Challenges and Solutions

- **Challenge:** Managing timezone-aware logic for filtering time-bound plots.

- **Solution:** Developed a utility function to convert UTC to IST and validate current time ranges.
- **Challenge:** Preventing time-limited visualizations from appearing in the dashboard.
 - **Solution:** Separated HTML writing logic for dashboard and individual task visualizations.
- **Challenge:** Debugging module imports and runtime errors due to missing packages.
 - **Solution:** Ensured all required packages like statsmodels were installed and handled dependency errors proactively.

Outcomes and Impact

- Delivered a modular dashboard framework supporting dynamic, time-based visualization logic.
- Created seven meaningful visualizations with advanced filtering and UI interactions.
- Developed strong practical experience in Python data pipelines and frontend integration.

Conclusion

The internship at NullClass was a transformative learning experience. The tasks gave me deep insights into how real-world dashboards can be tailored for contextual, intelligent decision-making. This internship has been a valuable step in my data science journey. It enhanced both my technical and logical thinking capabilities. The experience of building a real-time, condition-aware dashboard was both challenging and rewarding.