NULLCLASS

Internship Report

Intern Name: Abdulsattar Modin Inamdar

Internship Title: Learn To Build Real Time Google Play store data analytics - python

Company Name: NULLCLASS EDTECH PRIVATE LIMITED.

Duration: Example: 16-06-2025 to 16-09-2025

NULLCLASS

Declaration

I, Abdulsattar Modin Inamdar, hereby declare that this internship report titled **"Google Play store data analytics - python"** is the result of my own work carried out during my internship with NullClass from 16th June 2025 to 16th September 2025.

This report has not been submitted to any other institution for academic purposes. All data used in the report is for educational use only and the dashboards, notebooks, and visualizations were independently created as part of the assigned internship tasks.

Date: 15th July 2025

Place: Ichalkaranji.

NULTCLASS

Table Content

NULLCLASS

1. Introduction

During my online internship at NullClass, I worked on a project focused on data analysis of Google Play Store apps using Python, Pandas, and Matplotlib,plotly. The goal was to clean, analyze, and visualize insights from the app dataset to understand category trends, user engagement, pricing models, and more.

2. Background

Null Class is an ed-tech platform that enables hands-on learning through real-world internships. For this internship, I was assigned a project involving the analysis of the Google Play Store dataset using Python. The objective was to extract meaningful insights from app data like installs, reviews, ratings, content ratings, and price, and present them via Jupyter Notebook with clean and understandable visualizations.

NULLCLASS

3. Learning Objectives

- Clean and preprocess real-world datasets using Pandas.
- Handle missing and inconsistent data effectively.
- Create interactive visualizations using Matplotlib and Seaborn.
- Perform group-wise aggregations and comparative analysis.
- Derive insights from user reviews, installs, ratings, and app categories

NULLCLASS

4. Activities and Tasks

During this internship, I completed a set of advanced data analysis and visualization tasks using Python, Pandas, Plotly, and Matplotlib, with a strong emphasis on condition-based filtering and interactivity. Below are the major tasks:

Task 1 – Revenue vs Installs Scatter Plot (Paid Apps Only)

Created a scatter plot to visualize the relationship between revenue and number of installs for paid apps only.

- Trendline added to show correlation
- Points color-coded based on app categories

Task 2 – Dual-Axis Chart for Free vs Paid Apps in Top Categories

Created a dual-axis chart to compare average installs and revenue for free vs paid apps in the top 3 categories.

Applied filters:

- Installs ≥ 10,000
- Revenue ≥ $10,000
- Android Version > 4.0
- App Size > 15MB
- Content Rating = "Everyone"
- App name ≤ 30 characters (including spaces & special characters)
  Visibility Condition:
- Only visible between 1 PM to 2 PM IST

Task 3 – Grouped Bar Chart for Ratings and Reviews by Category

Plotted a grouped bar chart comparing average rating and total review count for the top 10 app categories by installs.

Filters applied:

- Average Rating ≥ 4.0
- App Size ≥ 10MB
- Last Update = January
  Visibility Condition:
- Only visible between 3 PM to 5 PM IST

Task 4 – Global Installs Choropleth Map by Category

Designed a Choropleth map using Plotly to visualize global installs by app category.
Filters applied:

- Top 5 app categories by install count
- Installs > 1 million
- Exclude categories starting with "A", "C", "G", or "S"
  Visibility Condition:
- Only visible between 6 PM to 8 PM IST

Task 5 – Time Series Line Chart of Installs Over Time by Category

Built a time series line chart to track total installs over time, segmented by app category.
Logic and Filters:

- Highlight 20%+ MoM growth with shaded areas
- Category must begin with "E", "C", or "B"
- App name must NOT begin with "x", "y", or "z"
- App name must NOT contain the letter "S"
- Reviews > 500
  Localization for categories:
- "Beauty" → Hindi

- "Business" → Tamil
- "Dating" → German
  Visibility Condition:
- Only visible between 6 PM to 9 PM IST

Task 6 – Bubble Chart: App Size vs Rating with Installs

Created a bubble chart to analyze relationship between app size and average rating. Bubble size represents install count.

Filters applied:

- Rating > 3.5
- Categories: Game, Beauty, Business, Comics, Communication, Dating, Entertainment, Social, Event
- Reviews > 500
- App name must NOT contain the letter "S"
- Sentiment subjectivity > 0.5
- Installs > 50,000
- Game category highlighted in pink
  Localization for categories:
- "Beauty" → Hindi
- "Business" → Tamil
- "Dating" → German
  Visibility Condition:
- Only visible between 5 PM to 7 PM IST

NULLCLASS

5. Skills and Competencies Gained

Technical Skills:

- Data cleaning with Pandas
- Data visualization using Matplotlib and Seaborn
- GroupBy and aggregation techniques
- Data type conversions and handling outliers
- Jupyter Notebook documentation

Soft Skills:

- Time management during remote work
- Debugging and logical reasoning
- Communicating findings through clear visuals
- Report writing and project documentation

NULLCLASS

6. Feedback and Evidence

As mentor guidance was not part of the internship, I took an independent learning approach. I referred to trusted online sources such as documentation, tutorials, and forums including Stack Overflow to solve challenges and improve task outcomes. The results—Jupyter notebooks, Python scripts, and visualizations—are recorded as evidence of the skills developed and tasks completed. Screenshots and logs were also maintained to document my progress throughout the internship.

NULLCLASS

7. Challenges and Solutions

Challenge: Parsing and converting the 'Installs' column which included symbols like '+' and ','

Solution: Used str.replace and astype(int) after cleaning string characters.

Challenge: Dealing with missing ratings and app types.

Solution: Imputed missing values using mode or removed null rows if imputation wasn’t appropriate.

Challenge: Visual clutter in bar charts with many categories.

Solution: Filtered and visualized only the top 10 or most relevant entries.

NULLCLASS

8. Outcomes and Impact

- Gained hands-on experience in real-world data analysis.
- Understood market trends in the mobile app ecosystem.
- Improved my proficiency in Pandas, data wrangling, and visual storytelling.
- Delivered a well-commented and structured Jupyter Notebook as a portfolio project.

9. Conclusion

The Google Play Store analysis project enabled me to explore the mobile application landscape using Python. This internship not only enhanced my technical and analytical skills but also introduced me to best practices in data storytelling. The experience has motivated me to pursue further projects in data science and analytics.