Internship Report

Name: Bhargavi K

Project Title: Google Play Store Analytics

Internship Duration: [11.06.2025] – [11.07.2025]

Company: Null Class Edtech Private Limited

Email: bargaveek@gmail.com

Tools Used: Python, Pandas, NumPy, Plotly, Word Cloud, NLTK,

Jupyter Notebook, HTML

1. Introduction

The internship project focuses on analyzing applications and user reviews from the Google Play Store. The objective was to extract insights about app popularity, user sentiments, and trends in installations and categories. The final deliverable was a dynamic and interactive dashboard created using Plotly and HTML, integrating multiple Python modules.

2. Background

With millions of apps hosted on the Play Store, it becomes important to understand app performance and user behavior. This project uses real-world datasets to explore categories, ratings, installs, revenue, and review sentiments to support product decisions and trend identification.

3. Learning Objectives

- Understand data preprocessing techniques for large-scale datasets
- Perform exploratory data analysis using Pandas and NumPy
- Conduct sentiment analysis using NLTK VADER
- Create interactive charts using Plotly
- Build a responsive dashboard integrating all visualizations
- Apply time-based logic for conditional chart rendering

4. Activities

I worked on cleaning and transforming the Play Store data, conducting sentiment analysis on reviews, and building multiple visualization charts. Each task involved generating specific plots and insights, later merged into a unified dashboard.

5. Tasks Completed

-Task 1: Generated 10 different Plotly visualizations for key metrics like installs, revenue, ratings, categories, and updates.
-Task 2: Created a word cloud based on positive reviews for Health & Fitness apps.
-Task 3: Built a dual-axis chart comparing installs and revenue for Free vs Paid apps viewable between 1 PM and 2 PM IST.
-Task 4: Created a time-series trend chart for app installs highlighting MoM growth over 20%, shown only between 6 PM and 9 PM IST.

6. Skills Developed

- Python programming and data manipulation with Pandas
- Data cleaning and preprocessing techniques
- Interactive data visualization with Plotly
- Natural Language Processing with NLTK
- HTML integration and dashboard layout styling
- Timezone-aware scripting in Python

7. Competencies Gained

- Independent project handling
- Real-world data analysis
- Insight communication through dashboards
- Coding best practices and documentation
- Time-based scripting and conditional display logic

8. Feedback

The project received positive feedback for being well-structured and visually informative. My mentors were very kind and helpful. They supported me in every step of the project and always answered my questions. Their guidance made it easier for me to understand the tasks and complete them well. I am thankful for their help.

9. Evidence of Work

- Jupyter Notebook files containing all Python code
- The integrated HTML dashboard
- Links to hosted notebook or dashboard (GitHub)

10. Challenges Faced

- Handling null, inconsistent or duplicate values in datasets
- Processing reviews for sentiment while avoiding noise
- Implementing time-based chart conditions accurately
- Combining multiple plots and insights into one dashboard

11. Solutions Applied

- Used `fillna`, typecasting, and filtering to clean data
- Applied NLTK's VADER for sentiment scores
- Used Python's `datetime` with `pytz` for IST time filtering
- Created modular code to generate and save charts using Plotly and WordCloud

12. Outcomes and Impact

The final dashboard allows stakeholders to explore key Play Store metrics in one place. The project improved my confidence in building complete data solutions, starting from raw CSV files to visual storytelling and dashboard development.

13. Conclusion

This internship project provided me with end-to-end experience in data science workflow like data cleaning, analysis, visualization, and presentation. It helped me strengthen technical skills, problem-solving abilities, and the capacity to build interactive dashboards independently.