Data Analysis

AN INTERNSHIP REPORT

Real-Time Google Play Store Data Analysis

KHUSHI GUPTA
NULLCLASS

INTRODUCTION

This report outlines my experience as an intern at NullClass, where I worked on real-time Google Play Store data analytics using Python. The project involved data visualization, sentiment analysis, and interactive filtering techniques to extract meaningful insights from user reviews and app performance data. The internship at NullClass was an excellent opportunity to gain hands-on experience in real-time data analytics using Python. The project focused on analysing Google Play Store data to extract meaningful insights related to app performance, user sentiment, and market trends. The primary objective was to develop interactive visualizations and analytics dashboards to better understand app categories, user reviews, and ratings using data science techniques.

This report provides an in-depth overview of my learning journey, tasks undertaken, challenges faced, and the skills developed throughout the internship.

This report provides an overview of my internship experience at NullClass, where I worked on real-time data analytics for Google Play Store applications using Python. The primary objective was to extract, clean, analyse, and visualize app data to generate meaningful insights.

BACKGROUND

The rapid growth of mobile applications has made app store analytics an essential aspect of digital business strategy. Understanding user sentiment, app popularity, and category-based trends is crucial for developers and businesses. This internship provided hands-on experience in analysing large-scale datasets and creating visualizations to assist in data-driven decision-making.

With the increasing dependency on **mobile applications**, analysing data from platforms like the Google Play Store is critical for developers, businesses, and researchers. The **availability of user-generated data** (ratings, reviews, download counts) allows organizations to assess app performance, identify trends, and optimize future development strategies.

This internship aimed to **build analytical models** that process, clean, and visualize Play Store data to uncover insights. The main areas of focus included:

- Understanding **user sentiment** and how it affects app success.
- Comparing **app categories** based on reviews and ratings.
- Implementing **data filtering and transformation** to extract meaningful patterns.
- Utilizing **visualization techniques** to present findings effectively.

Through **data analytics and machine learning concepts**, this internship bridged the gap between **theoretical knowledge and practical implementation** in a professional setting.

LEARNING OBJECTIVE

Develop expertise in Python for data analytics and visualization. Gain proficiency in handling large datasets using Pandas and NumPy. Implement real-time data filtering and graphical representation techniques. Understand user sentiment distribution and its impact on app ratings.

The primary learning objectives of this internship included:

Technical Learning Objectives:

1. **Data Handling & Preprocessing:**
    - Learn **data cleaning techniques** to handle missing, incorrect, or inconsistent values.
    - Convert **raw Play Store data** into structured formats suitable for analysis.
    - Master **data transformation** using Pandas and NumPy to extract meaningful insights.
2. **Data Visualization & Interpretation:**
    - Develop **interactive visualizations** using Matplotlib and Seaborn.
    - Understand different **graphing techniques** (bar charts, violin plots, scatter plots, histograms).
    - Interpret patterns in **app ratings, reviews, and user sentiments** effectively.
3. **Filtering & Data Extraction:**
    - Implement **dynamic filtering techniques** to select relevant categories.
    - Use **conditional logic and numerical thresholds** to refine datasets.
    - Optimize performance by applying **efficient data querying** methods.

4. **Sentiment Analysis & Market Trends:**
    - Categorize user reviews into **positive, neutral, and negative** sentiments.
    - Identify trends in user ratings across different app categories.
    - Compare app performance based on **install counts, review scores, and sentiment distribution**.
5. **Real-Time Data Handling & Conditional Visualization:**
    - Implement **time-based conditions** to modify graphs dynamically.
    - Automate updates in visualizations based on **date and time intervals**.
6. **Version Control & Project Management:**
    - Set up and manage a **GitHub repository** for collaboration.
    - Use **Git commands (init, add, commit, push, pull)** to maintain version control.
    - Write effective **documentation (README.md)** for project clarity.

Professional & Personal Development Learning Objectives:

1. **Problem-Solving & Critical Thinking:**
    - Learn to **identify and resolve dataset inconsistencies**.
    - Adapt to new **data processing techniques and troubleshooting methods**.
    - Apply **logical reasoning** to develop optimized solutions.
2. **Time Management & Productivity:**

- Plan and execute tasks **within assigned deadlines**.
- Balance **multiple project components** without compromising quality.

3. **Effective Communication & Reporting:**
    - Document findings in a structured and **concise manner**.
    - Create **clear and informative reports** using insights from data analysis.
    - Present analytical results in a **visually engaging format**.
4. **Collaboration & Adaptability:**
    - Work within an **internship team structure**.
    - Adapt to **new tools, coding practices, and methodologies**.

By the end of this internship, I aimed to become **confident in handling large-scale datasets**, drawing **actionable insights**, and presenting them in an industry-standard format.

ACTIVITIES AND TASKS

During this internship, I successfully completed several **data analytics tasks**, each aimed at extracting useful insights from the **Google Play Store dataset**.

Task 1: Sentiment Distribution Analysis

- **Objective:** Analyse the **distribution of user sentiments** (positive, neutral, negative) across different rating groups.
- **Implementation:**
    - Cleaned and preprocessed the review dataset.
    - Categorized reviews into sentiment groups.
    - Created a **stacked bar chart** using Matplotlib to visualize sentiment trends.

Task 2: Category Comparison & Filtering

- **Objective:** Compare the **average rating and total review count** for the **top 10 app categories** based on installs.
- **Implementation:**
    - Filtered categories that had **more than 50 apps** to ensure meaningful comparisons.
    - Created a **grouped bar chart** using Seaborn to compare ratings and review counts.
    - Applied **multiple filtering conditions** to refine the dataset.

Task 3: Violin Plot Visualization

- **Objective:** Represent **rating distribution** for different app categories.
- **Implementation:**

    - Selected **categories based on criteria** such as app count and total reviews.
    - Plotted **violin charts** to visualize the density and spread of ratings.
    - Incorporated **data filtering techniques** to enhance visualization accuracy.

Task 4: Real-Time Data Handling

- **Objective:** Implement time-based conditions for displaying different graphs.
- **Implementation:**
    - Used Python's datetime module to set **time-based display conditions**.
    - Dynamically updated visualizations based on time intervals.

Task 5: GitHub Repository Management

- **Objective:** Maintain version control and documentation for project files.
- **Implementation:**
    - Set up a **GitHub repository** for tracking project progress.
    - Used Git commands (git init, git add, git commit, git push) for updates.
    - Created a README.md file to document project details and instructions.

SKILLS AND COMPETENCIES

- Data analysis and manipulation using Pandas and NumPy.
- Data visualization with Matplotlib and Seaborn.
- Real-time data filtering and time-based graphical conditions.
- Handling large datasets efficiently.
- Git and GitHub for version control and collaboration.
- Problem-solving and independent research skills.

Technical Skills:

- **Data Analytics:** Proficient in **data cleaning, transformation, and visualization**.
- **Python Programming:** Extensive use of **Pandas, NumPy, Matplotlib, and Seaborn**.
- **Data Filtering:** Applied **conditional filters** to extract relevant insights.
- **Version Control:** Hands-on experience with **Git and GitHub** for project management.

Soft Skills:

- **Problem-Solving:** Effectively addressed dataset inconsistencies and visualization challenges.
- **Project Management:** Maintained structured workflows and completed tasks within deadlines.
- **Communication:** Documented findings and explained complex concepts in a simple manner

---

FEEDBACK AND EVIDENCE

The project provided valuable insights into handling real-world data challenges. Successfully implementing time-based conditions for dashboard visualization demonstrated practical expertise in Python. The completion of all tasks within the given timeframe highlights strong analytical and programming skills.

One of the most significant takeaways from this internship was the **ability to work independently on complex data-driven tasks**. Without direct mentor support, I was encouraged to **rely on self-learning, research, and troubleshooting skills**, which significantly boosted my **confidence and adaptability**. This approach mirrored real-world industry expectations, where professionals must **analyse problems, develop solutions, and optimize workflows** efficiently.

Additionally, working with **real-time data and dynamic visualizations** introduced a new level of complexity that required **critical thinking and logical reasoning**. Understanding how to apply **time-based conditions for dynamic dashboard visualization** was a crucial challenge that pushed me to explore innovative coding techniques and **efficient data filtering** strategies.

Overall, the feedback I received from my **self-assessments, task evaluations, and internship guidelines** helped me refine my work, improve efficiency, and adopt a **more structured approach to problem-solving**. The internship was not just about completing assigned tasks but about **gaining hands-on expertise, improving independent research skills, and building a strong analytical mindset**—all of which will be invaluable in my future career endeavours.

CHALLNGES AND SOLUTIONS

- Challenge: Missing or inconsistent data in the dataset.
  - Solution: Used data preprocessing techniques, such as handling null values and converting data types appropriately.

- Challenge: Time-based conditions for visualization.
  - Solution: Implemented real-time filtering using Python's datetime module.

- Challenge: Initial issues with GitHub integration.
  - Solution: Configured Git properly and ensured seamless synchronization between VS Code and GitHub.

OUTCOMES AND IMPACT

- Successfully developed a functional real-time analytics dashboard for Google Play Store data.
- Improved proficiency in data science tools and techniques.
- Gained confidence in independent research and problem-solving.
- Strengthened technical documentation and reporting skills.
- The project contributes to better understanding app store analytics for developers and businesses.
- Developed an **interactive analytics dashboard** for Google Play Store data.
- Gained proficiency in **Python data analytics tools and techniques**.
- Successfully met all project requirements, making me eligible for an **internship certificate and stipend**.
- Enhanced **coding efficiency** and **data visualization expertise**.
- Improved **ability to handle large-scale datasets** for real-world applications.

CONTRIBUTION AND FUTURE SCOPE

- The visualizations created can aid.
- Businesses in app market analysis.
- Future enhancements may include real-time API integration for live data updates.
- Implementing machine learning models for predictive analytics on app performance.
- Provided **data-driven insights** that can help app developers understand user sentiment and market trends.
- Developed **optimized visualization techniques** for analysing Play Store data.
- Ensured project documentation and code clarity for **future reference and scalability**.

CONCLUSION

This internship was an enriching experience that provided hands-on exposure to data analytics and visualization. The project not only strengthened my technical abilities but also improved my ability to work independently and manage a structured workflow. The insights gained will significantly contribute to my professional growth in the field of data analytics and software development.

This internship at **NullClass** provided a **valuable opportunity to apply data science techniques** in a **real-world scenario**. The hands-on approach helped refine my **technical, analytical, and problem-solving skills**, while also improving my ability to work with **large datasets and visualization tools**.

By working on **real-time data analytics, visualization, and GitHub project management**, I gained **practical exposure to industry-level data science applications**. This experience has strengthened my confidence and prepared me for future **data analytics roles**.

Submitted by: Khushi Gupta