Internship Report

Google Play Store Data Analysis

Introduction

This report outlines the data analytics tasks performed during my internship, focusing on sentiment analysis, comparative analytics, and time-series visualization for app data. The tasks involved filtering datasets, applying transformations, and creating insightful visualizations using Python and data visualization tools.

Background

The project aimed to analyze app performance through sentiment distribution, revenue comparisons, and install trends. The dataset included user reviews, ratings, app installs, revenue, and content ratings, which required preprocessing and filtering to meet specific criteria for each visualization.

Learning Objectives

1. To apply sentiment analysis techniques for user review insights.
2. To create advanced visualizations for comparing app performance.
3. To implement time-based filtering for dynamic dashboards.
4. To enhance Python and data visualization skills in a real-world scenario.

Activities and Tasks

Task 1: Sentiment Distribution Analysis

- Filtered apps with more than 1,000 reviews.
- Grouped apps by the top 5 categories.
- Categorized ratings into groups (1-2, 3-4, 4-5 stars).
- Created a stacked bar chart showing sentiment distribution (positive, neutral, negative).

Task 2: Free vs. Paid App Comparison

- Filtered apps with installs greater than 10,000 and revenue above $10,000.
- Included only apps with an Android version above 4.0, size over 15MB, content rating as "Everyone," and an app name of 30 characters or fewer.
- Created a dual-axis chart comparing average installs and revenue for free vs. paid apps in the top 3 categories.
- Ensured the graph was only visible between 1 PM IST to 2 PM IST in the dashboard.

Task 3: Install Trend Time Series

- Selected apps with more than 10,000 installs and content rating as "Teen."
- Included only apps whose names start with the letter 'E.'
- Created a time series line chart to show total installs over time, segmenting by category.
- Highlighted growth areas where installs increased by more than 20% month-over-month.
- Restricted the graph display to 6 PM IST to 9 PM IST in the dashboard.

Skills and Competencies

- **Data Cleaning & Preprocessing:** Applied filtering and transformations to ensure data quality.
- **Data Visualization:** Created stacked bar charts, dual-axis comparisons, and time-series graphs.
- **Python & Pandas:** Used Python for data manipulation and visualization.
- **Dashboard Optimization:** Implemented time-based visibility restrictions for graphs.
- **Sentiment Analysis:** Classified and visualized user sentiments effectively.

Feedback and Evidence

- Successfully visualized sentiment trends, app performance, and growth trends.
- Implemented effective filtering strategies, enhancing data-driven insights.
- Time-based dashboard filters ensured relevant information display.

Challenges and Solutions

- **Challenge:** Handling large datasets with complex filters.
    - **Solution:** Optimized data processing using Pandas and vectorized operations.
- **Challenge:** Implementing time-based visibility restrictions.

    - **Solution:** Used dashboard settings and conditional logic for time filters.
- **Challenge:** Identifying significant growth trends.
    - **Solution:** Applied month-over-month percentage calculations and conditional shading.

Outcomes and Impact

- **Improved Decision-Making:** Provided clear visual insights for app developers and stakeholders.
- **Efficient Data Processing:** Optimized filtering techniques reduced computational load.
- **Enhanced Dashboard Usability:** Time-based graphs ensured relevant information display at appropriate times.

Conclusion

This project provided hands-on experience in sentiment analysis, comparative analytics, and time-series visualization. It strengthened my Python skills, data preprocessing capabilities, and dashboard optimization techniques. The learnings from this internship will be valuable for future data-driven projects.

By: Khushi Sharma