**Google Play Store Data Analytics with Python**

**Name:- Atharva Chougule**

**Duration:- 20th March – 21st May (2025)**

**Submission Date :- 21st May 2025**

**Introduction:**

Information on app performance, user behavior, and market trends is available through the Google Play Store Analytics initiative. It helps business analysts, marketers, and app developers make data-driven choices, enhance user engagement, and optimize app strategy.

**Background:**

Data cleansing, transformation, exploratory data analysis (EDA), sentiment analysis, and the development of an interactive dashboard are all part of the project's investigation of Google Play Store data. Finding valuable insights in the data to help guide company decisions is the aim.

**Learning Objectives:**

* Make use of the Python libraries Scikit-learn, Pandas, NumPy, and Plotly.
* Use NLTK to conduct sentiment analysis.
* Use Plotly to create interactive visuals.
* Use HTML to incorporate visuals into web apps.

**Activities and Tasks:**

**Task-1:** Create a scatter plot to visualize the relationship between revenue and the number of installs for paid apps only. Add a trendline to show the correlation and colour code the points based on app categories.

**Task-2:** Use a grouped bar chart to compare the average rating and total review count for the top 10 app categories by number of installs. Filter out any categories where the average rating is below 4.0 and size below 10 M and last update should be Jan month . this graph should work only between 3PM IST to 5 PM IST apart from that time we should not show this graph in dashboard itself.

**Task-3:** Create a dual-axis chart comparing the average installs and revenue for free vs. paid apps within the top 3 app categories. Apply filters to exclude apps with fewer than 10,000 installs and revenue below $10,000 and android version should be more than 4.0 as well as size should be more than 15M and content rating should be Everyone and app name should not have more than 30 characters including space and special character .this graph should work only between 1 PM IST to 2 PM IST apart from that time we should not show this graph in dashboard itself.

**Task-4:** Plot a bubble chart to analyse the relationship between app size (in MB) and average rating, with the bubble size representing the number of installs. Include a filter to show only apps with a rating higher than 3.5 and that belong to the Game, Beauty ,business , comics , communication , Dating , Entertainment , social and event categories. Reviews should be greater than 500 and sentiment subjectivity should be more than 0.5 and Installs should be more than 50k as well as this graph should work only between 5 PM IST to 7 PM IST apart from that time we should not show this graph in dashboard itself.

**Task-5:**Generate a heatmap to show the correlation matrix between installs, ratings, and review counts. Filter the data to include only apps that have been updated within the last year and have at least 100,000 installs and reviews count should be more than 1k and genres name should not be Starting with characters A , F , E , G , I , K . this graph should work only between 2 PM IST to 4 PM IST apart from that time we should not show this graph in dashboard itself.

**Data Loading and Cleaning:**

* Cleaned Google Play Store data using Pandas.
* Handled missing values, duplicates, and data type inconsistencies to ensure data reliability.

**Data Transformation:**

* Created new features like log-transformed install counts and categorized ratings.
* Calculated revenue metrics for deeper insights into app performance. **Exploratory Data Analysis (EDA):**
* Visualized trends in app categories, ratings, and reviews.
* Highlighted top-performing app categories.

**Sentiment Analysis:**

* Performed sentiment analysis on user reviews using NLTK.
* Determined sentiment polarity (positive, negative, neutral) and its impact on app metrics.

**Interactive Visualization and Dashboard Creation:**

* Built dynamic visualizations with Plotly.
* Integrated visualizations into web applications using HTML.
* Designed a user-friendly dashboard for key insights.

**Skills and Competencies:**

* Python Programming
* Data Analysis and Visualization
* Sentiment Analysis
* Dashboard Creation
* Web Application Integration

**Challenges and Solutions:**

* Data Quality Issues: By using cleaning procedures, missing and inconsistent data were addressed.
* Complex Visualizations: To improve visual representation, sophisticated Plotly capabilities were used.

**Outcomes and Impact:**

* Produced informative reports to aid marketers and app developers.
* A better comprehension of market trends for apps.
* Improved proficiency in sentiment analysis, data analysis, and visualization.

**Conclusion:**

The Google Play Store Analytics project offered a thorough education in interactive visualization, sentiment analysis, and data analysis. It improved technical proficiency while offering insightful information for app ecosystem decision-making.