Assessment -4

Perform an Exploratory Data Analysis (EDA) on a given dataset of movie ratings to extract insights and identify patterns. Use visualizations and descriptive statistics to support your findings.

Task 1

Data Understanding:

- Get familiar with the dataset. Check the size, columns, and basic summary statistics.
- Identify and report any missing values or duplicates in the dataset.

Task 2

Data Cleaning:

- Handle missing or incorrect data if present.
- Remove duplicates if applicable.
- Convert any relevant columns to appropriate data types (e.g., timestamp to datetime).

Task 3

Univariate Analysis:

- Analyze the distribution of ratings.
- Calculate the most common rating given by users.
- Analyze the genre distribution (if the dataset contains genres).

Task 4

Bivariate and Multivariate Analysis:

- Explore the relationship between user demographics (age, gender) and ratings. For example:
  - Are certain age groups more likely to rate higher or lower?
  - Do male and female users rate movies differently?
- Investigate how genre preferences vary by user demographics (age, gender).
- Explore correlations between movie ratings and genres.

Task 5

Popular Movies and Trends:

- Identify the top 10 highest-rated movies.
- Find the top 10 most-watched movies (by the number of ratings).
- Explore if certain movie genres have higher average ratings.

Task 6

Time-Based Analysis:

- Perform an analysis to see how ratings have changed over time (using the timestamp).
- Investigate if there are particular years or seasons where movies receive higher ratings.

Task 7

Visualizations:

- Provide relevant visualizations (histograms, bar plots, heatmaps, etc.) to support your findings.
- Visualize trends such as the distribution of ratings across different genres or user demographics.

Task 8

Conclusion:

- Summarize your key findings.
- Suggest areas for further analysis or recommendations for improving user engagement based on the insights.

Deliverables:

- A well-commented ipynb file (or any preferred tool) containing:
- Code used for analysis.
- Visualizations.
- Written explanations and insights based on the analysis.

Submission: The Entire assignment should be submitted by Friday (18/10/2024), You have to upload the dataset and 1 ipynb notebook file in Git Hub.