**. Introduction:** The objective of this project is to analyze the TMDB (The Movie Database) dataset containing information about movies, including details such as titles, genres, keywords, cast, and crew. The dataset provides a rich source of information for exploring trends and patterns in the movie industry.

**2. Data Collection:**

* The dataset comprises two CSV files: "tmdb\_5000\_movies.csv" and "tmdb\_5000\_credits.csv".
* The "tmdb\_5000\_movies.csv" file contains details about movies, including titles, overviews, genres, and keywords.
* The "tmdb\_5000\_credits.csv" file contains information about the cast and crew involved in each movie.

**3. Data Cleaning and Preparation:**

* Loaded the datasets using pandas library.
* Merged the two datasets based on the 'id' and 'movie\_id' columns.
* Removed unnecessary columns and handled missing values.
* Transformed JSON strings into readable formats to extract relevant information.

**4. Exploratory Data Analysis (EDA):**

* Conducted EDA to understand the distribution of variables such as genres, keywords, and cast.
* Generated summary statistics and visualizations to gain insights into the dataset.

**5. Feature Engineering:**

* Extracted features from JSON strings such as genres, keywords, cast, and crew.
* Created new variables based on these features to enhance the dataset for analysis.

**6. Conclusion:**

* The analysis of the TMDB movie dataset provided valuable insights into the movie industry.
* Key findings include trends in movie genres, popular keywords, and the involvement of cast and crew in different movies.
* Further exploration could involve predictive modeling to forecast movie success based on various factors.

**7. References:**

* Python documentation: <https://docs.python.org/3/>
* Pandas documentation: https://pandas.pydata.org/docs/
* Matplotlib documentation: https://matplotlib.org/stable/contents.html

**8. Appendix:**

* Code snippets and scripts used for data cleaning, analysis, and visualization are available upon request.

Top of Form