**SYNOPSIS on**

**TITLE – IMDb Top 1000 Movies Analysis**

**VIKAS S HIREMATH(AF0378033)**

**ABOUT THE TOPIC (DATASET):**

**Data Structure**

The dataset is organized in a tabular format, where each row represents a movie record. The columns include:

1. **Title**: This column contains the names of the movies.
2. **Genre**: This column categorizes the movies by genre.
3. **Director**: This column lists the directors of the movies.
4. **Star1, Star2, Star3, Star4**: These columns list the main actors in the movies.
5. **Certificate**: This column indicates the movie ratings (e.g., PG, R).
6. **Runtime**: This column shows the duration of the movies in minutes.
7. **IMDB\_Rating**: This column provides the IMDb ratings of the movies.
8. **Overview**: This column contains a brief summary of the movies.
9. **Meta\_score**: This column provides the Metacritic score of the movies.
10. **No\_of\_Votes**: This column indicates the number of votes each movie received.
11. **Gross**: This column shows the gross earnings of the movies in USD.
12. **Released\_Year**: This column indicates the release year of the movies.

These columns collectively provide detailed information about each movie, including its genre, director, main actors, ratings, runtime, and financial performance. This dataset is useful for analyzing trends in movie genres, ratings, financial success, and more.

**Data Preprocessing**

1. **Data Cleaning**: Handle missing values and ensure data consistency.
2. **Normalization**: Normalize numeric features if required.
3. **Encoding Categorical Variables**: Encode categorical variables using one-hot encoding or label encoding.

**Implementation Steps**

1. **Data Ingestion**: Load the dataset into a data analysis environment (e.g., Python).
2. **Preprocessing**: Clean and prepare the data for analysis.
3. **Exploratory Data Analysis (EDA)**: Conduct EDA to understand data distribution and initial patterns.
4. **Visualization**: Create visualizations to communicate insights effectively (e.g., genre distribution, yearly trends, rating distributions).
5. **Reporting**: Summarize findings in reports or dashboards for stakeholders.

**Data Set: IMDb Top 1000 Movies**

**Technologies: Pandas, Matplotlib, Seaborn**

**Software Requirements:**

**Operating System**: Windows, Linux, macOS

**IDLE**: Jupyter Notebook

**Hardware Requirements:**

**RAM**: Minimum 4GB, recommended 8GB

**Processor**: Minimum Intel i3, recommended Intel i5