**IMDb Movie Dataset Analysis**

This Python code encompasses the exploration and analysis of an IMDb movie dataset, including data preparation, visualization, and statistical hypothesis testing.

**1. Data Loading and Preprocessing**

**Libraries Used:**

pandas: for data manipulation and analysis.

matplotlib.pyplot, seaborn, plotly.express: for data visualization.

datetime: for date and time manipulation.

**Data Loading:**

Imported the dataset from the file NetflixOriginals.csv using pd.read\_csv() from Pandas.

**Data Exploration:**

Overview of the dataset using methods like head(), describe(), info(), and isna().sum() to understand its structure, statistical summary, and missing values.

**Data Preprocessing:**

Extracted and transformed 'Premiere' column to obtain the release date and year. Handled string manipulations and converted data to appropriate date formats for analysis.

**2. Exploratory Data Analysis (EDA)**

**Genre and Language Analysis:**

Counted and visualized the distribution of movie genres and languages using bar plots, identifying popular genres and languages.

**Runtime Analysis:**

Explored movie runtimes by visualizing their distribution using a density plot and conducted statistical analysis on runtimes using one-sample and independent t-tests.

**3. Statistical Analysis of Runtimes**

**Hypothesis Testing:**

Utilized scipy.stats to perform hypothesis tests on movie runtimes.

Conducted one-sample t-tests with null hypotheses set at specific values (e.g., 91.3, 93) to test for differences in runtime means.

Executed independent t-tests comparing movie runtimes against various sets of dummy data.

**Visualization of Statistical Analysis:**

Visualized the runtime distributions and outcomes of t-tests using density plots and legends to compare different distributions.

**4. Conclusion**

The analysis provides insights into the IMDb movie dataset, highlighting popular genres, languages, and statistical inferences related to movie runtimes.

This code represents an end-to-end process, from data loading and cleaning to exploratory analysis and hypothesis testing. It offers a comprehensive