**NETFLIX DATA ANALYSIS REPORT**

***BY SAURABH SINGH***

***AF0417029***

**1. Title**

**Netflix Data Analysis**

**2. Introduction**

Netflix's extensive content catalog spans various genres, categories, and formats that attract millions of viewers worldwide. This project will analyze Netflix's dataset, focusing on factors like category, country, release date, ratings, and type. By analyzing these attributes, we aim to uncover trends and preferences, providing insights that could inform Netflix's content strategies.

**3. Objectives**

The primary objectives of this project are:

* To understand the diversity in Netflix’s content by analyzing categories, genres, and types.
* To explore content ratings and their distribution across countries and categories.
* To analyze release trends and seasonality patterns in Netflix’s content.
* To provide visual insights to better understand the popularity and types of available content.

**4. Scope of Work**

The project will include the following tasks:

* **Data Exploration**: Understanding the dataset structure, including each feature's role and characteristics.
* **Data Preprocessing**: Handling missing values in fields like Director and Cast, and standardizing date formats.
* **Feature Analysis**: Analyzing categories, ratings, release dates, and duration to understand trends.
* **Data Visualization**: Using graphs to visualize insights on genre popularity, release trends, and country-based distributions.
* **Reporting**: Documenting and presenting findings in a structured report

**5. Methodology**

The project will be structured as follows:

1. **Data Collection**: Obtained the Netflix dataset from Kaggle.
2. **Data Preprocessing**:
   * Handle missing values, especially in columns like Director and Cast.
   * Parse Release\_Date and convert to a standard date format.
   * Remove duplicate entries if any and standardize the Duration format.
3. **Exploratory Data Analysis (EDA)**:
   * Summarize categories, ratings, and types.
   * Create visualizations to analyze distributions and trends in features like Category, Rating, and Country.
4. **Feature Analysis**:
   * Analyze relationships between Rating and Category to see which types of content have different ratings.
   * Identify seasonal trends in Release\_Date.
5. **Visualization**:
   * Generate plots to showcase trends and findings, including bar charts for category and country distributions, line plots for release trends, and pie charts for rating distributions.
6. **Reporting**:
   * Document analysis and findings in a comprehensive report.

**6. Tools and Technologies**

* **Programming Language**: Python
* **Libraries**: Pandas, NumPy, Matplotlib, Seaborn
* **IDE**: Jupyter Notebook
* **Data Source**: Kaggle (Netflix Dataset)

**7. Expected Outcomes**

* Insights into the distribution of Netflix content by category, type, and rating.
* Identification of the most common content types, genres, and popular ratings.
* Visualizations to understand release trends and content patterns.
* A final report with recommendations for potential content strategy.

**8. Timeline**

* Data Collection and Preprocessing
* Exploratory Data Analysis and Feature Analysis
* Visualization and Interpretation
* Report Preparation and Final Submission

**9. Conclusion**

This project offers insights into the types of content available on Netflix, focusing on categories, ratings, and release trends. These findings can provide useful guidance for content strategy and viewer engagement enhancement.