DATA ANALYSIS USING PYTHON COURSE BY

ACCENTURE & ANUDIP FOUNDATION

DATASET EXPLORATORY DATA ANALYSIS USING PYTHON PROJECT

BALAKRISHNA H S

STUDENT ID: AF0468107

ANUDIP FOUNDATION

Project GitHub Link: NETFLIX

PROJECT SYNOPSIS

1. Title

NETFLIX DATA ANALYSIS USING PYTHON

2. Introduction

Netflix is one of the world's largest streaming platforms, providing a wide range of movies and TV shows from various genres and regions. Understanding the content distribution, trends, and audience preferences can help improve decision-making for content recommendations and acquisitions. This project aims to analyze a Netflix dataset using Exploratory Data Analysis (EDA) techniques to identify trends, patterns, and insights about the platform's content. The insights from this analysis can help Netflix improve its content strategy and give viewers a better understanding of what to expect from the platform.

3. Objectives

The primary objectives of this project are:

- To explore and understand the key features of the Netflix dataset.
- To perform data preprocessing, including handling missing values and duplicates.
- To analyze factors affecting content trends, such as genre, country, and ratings.
- To analyze trends in content release over the years and how they relate to viewer preferences
- To visualize key insights and provide recommendations for better content strategies.

4. Analysis Process

- Data Exploration: Understanding the dataset, including the features such as genres, ratings, release year, and production country.
- Data Preprocessing: Cleaning the dataset by handling missing values, removing duplicates, and formatting data for analysis

- Feature Selection: Identifying the most significant features that influence Netflix content popularity, such as genre, rating, and release year.
- Data Visualization: Using plots and graphs to visualize trends and patterns in the dataset, such as genre distribution, content ratings, and release patterns.
- Trend Analysis: Studying release patterns, country-wise contributions, and genre preferences.
- Reporting: Documenting the findings and preparing a final report that presents the results and provides actionable recommendations.

5. Methodology

1. Data Collection

- The dataset is sourced from a public repository such as Kaggle, containing details about Netflix content.

2. Data Preprocessing

- Handle missing values by filling or removing them as necessary.
- Replace missing values in important columns like Directors, Cast, and Country with 'Not Specified'.
- Drop records with missing values in key fields like Date Added and Rating.
- Remove duplicate records to ensure clean data.

3. Exploratory Data Analysis (EDA)

- Use descriptive statistics to summarize dataset features.
- Create visualizations like bar charts, histograms, and heatmaps to identify trends.
- Analyze distribution of content by type, release year, and country.

4. Feature Selection

- Identify key features influencing content trends through correlation analysis.

5. Data Analysis & Visualization

- Count the number of movies vs. TV shows available on Netflix.
- Analyze trends in content release over the years and identify peak periods.
- Identify the top 10 countries contributing the highest number of Netflix shows.
- List the most common genres and their distribution.
- Study content trends based on director, cast, country, and ratings.

6. Evaluation and Interpretation

- Compare identified trends and patterns.
- Interpret results to understand their impact on content distribution.

7. Reporting

- Compile findings and insights into a detailed report.
- Provide recommendations based on the analysis.

6. Data Analysis & Visualizations

1. Content Type Distribution

✓ Netflix has more movies than TV shows, indicating higher demand for films.

2. Release Year Distribution

✓ Most content was released between 2000-2020, with a peak in 2018.

3. Top 10 Countries with Highest Content Contribution

✓ The United States has the most content, followed by India, highlighting regional production trends.

4. TV Shows & Movies from India

✓ Extracted a list of Indian movies and TV shows available on Netflix.

5. Top 10 Directors with Most Content on Netflix

✓ Identified the most frequent directors contributing to Netflix’s library.

6. Actor-Based Analysis

✓ Filtered content featuring specific actors like Mohanlal.

7. Rating-Based Analysis

✓ Identified the number of Indian movies rated 'TV-14'.

8. Country-wise TV Show Analysis

✓ Determined the country with the highest number of TV shows.

9. Top 10 Genres with Most Content

- ✓ Extracted and visualized the most common genres on Netflix.

6. Tools and Technologies

The project will utilize the following tools and technologies:

- ❖ **Programming Language:** Python
- ❖ **Libraries:** Pandas, NumPy, Matplotlib
- ❖ **IDE:** Jupyter Notebook or any Python-compatible Integrated Development Environment (IDE)
- ❖ **Data Source:** The dataset will be sourced from a public repository -Kaggle web platform

7. Expected Outcomes

- Identification of significant factors influencing Netflix content trends.
- Discovery of content distribution patterns based on genre, country, and ratings.
- Visual representation of key insights to help in decision-making.
- A final report summarizing findings and recommendations.

8. Timeline

The project is expected to be completed within a [specific timeframe, e.g., 4 weeks], with the following milestones:

- Week 1: Data Collection and Preprocessing
- Week 2: Exploratory Data Analysis and Data Visualization
- Week 3: Reporting, and Final Submission

9. Conclusion

This project provides useful insights into Netflix's content distribution. It helps us understand which movies and TV shows are popular in different regions and genres. The findings can help content creators and Netflix make better decisions to improve audience engagement and platform growth.