Business Case: Netflix - Data Exploration and Visualisation

Mindset

1. Evaluation will be kept lenient, so make sure you attempt this case study.
2. Read the question carefully and try to understand what exactly is being asked.
3. Brainstorm a little. If you're getting an error, remember that Google is your best friend.
4. You can watch the lecture recordings or go through your lecture notes once again if you feel like you're getting confused over some specific topics.
5. Discuss your problems with your peers. Make use of the Slack channel and WhatsApp group.
6. Only if you think that there's a major issue, you can reach out to your Instructor via Slack or Email.
7. There is no right or wrong answer. We have to get used to dealing with uncertainty in business. This is exactly the skill we want to develop.

About NETFLIX

Netflix is one of the most popular media and video streaming platforms. They have over 10000 movies or tv shows available on their platform, as of mid-2021, they have over 222M Subscribers globally. This tabular dataset consists of listings of all the movies and tv shows available on Netflix, along with details such as - cast, directors, ratings, release year, duration, etc.

Business Problem

Analyze the data and generate insights that could help Netflix ijn deciding which type of shows/movies to produce and how they can grow the business in different countries

Dataset

Link: Dataset_link

(After clicking on the above link, you can download the files by right-clicking on the page and clicking on "Save As", then naming the file as per your wish, with .csv as the extension.)

The dataset provided to you consists of a list of all the TV shows/movies available on Netflix:

Show_id: Unique ID for every Movie / Tv Show
Type: Identifier - A Movie or TV Show
Title: Title of the Movie / Tv Show
Director: Director of the Movie
Cast: Actors involved in the movie/show
Country: Country where the movie/show was produced
Date_added: Date it was added on Netflix
Release_year: Actual Release year of the movie/show
Rating: TV Rating of the movie/show
Duration: Total Duration - in minutes or number of seasons
Listed_in: Genre
Description: The summary description

Hints

1. The exploration should have a goal. As you explore the data, keep in mind that you want to answer which type of shows to produce and how to grow the business.
2. Ensure each recommendation is backed by data. The company is looking for data-driven insights, not personal opinions or anecdotes.
3. Assume that you are presenting your findings to business executives who have only a basic understanding of data science. Avoid unnecessary technical jargon.
4. Start by exploring a few questions: What type of content is available in different countries?
    1. How has the number of movies released per year changed over the last 20-30 years?
    2. Comparison of tv shows vs. movies.
    3. What is the best time to launch a TV show?
    4. Analysis of actors/directors of different types of shows/movies.
    5. Does Netflix has more focus on TV Shows than movies in recent years
    6. Understanding what content is available in different countries

Evaluation Criteria (100 Points):

1. Defining Problem Statement and Analysing basic metrics **(10 Points)**

2. Observations on the shape of data, data types of all the attributes, conversion of categorical attributes to 'category' (If required), missing value detection, statistical summary **(10 Points)**

3. Non-Graphical Analysis: Value counts and unique attributes **(10 Points)**

4. Visual Analysis - Univariate, Bivariate after pre-processing of the data

Note: Pre-processing involves unnesting of the data in columns like Actor, Director, Country

4.1 For continuous variable(s): Distplot, countplot, histogram for univariate analysis **(10 Points)**

4.2 For categorical variable(s): Boxplot **(10 Points)**

4.3 For correlation: Heatmaps, Pairplots **(10 Points)**

5. Missing Value & Outlier check (Treatment optional) **(10 Points)**

6. Insights based on Non-Graphical and Visual Analysis **(10 Points)**

6.1 Comments on the range of attributes

6.2 Comments on the distribution of the variables and relationship between them

6.3 Comments for each univariate and bivariate plot

7. Business Insights **(10 Points)** - Should include patterns observed in the data along with what you can infer from it

8. Recommendations **(10 Points)** - Actionable items for business. No technical jargon. No complications. Simple action items that everyone can understand

Submission Process:

1. Type your insights and recommendations in the rich-text editor.
2. Convert your jupyter notebook into PDF (Save as PDF using Chrome browser's Print command), upload it in your Google Drive (set the permission to **allow public access**), and paste that link in the text editor.
3. Alternatively, you can directly submit your PDF on the portal.

4. Optionally, you may add images/graphs in the text editor by taking screenshots or saving matplotlib graphs using plt.savefig(...).
5. After submitting, you will not be allowed to edit your submission.