**CSE - 5334 - 001 : Data Mining**

**Assignment 1 - Exploratory Data Analysis**

**Instructor: Dr. Elizabeth D Diaz**

**Team 18: Urmi Manish Sheth – 1002064934**

**Harsh Navinbhai Shah – 1002057387**

**Sai Swetha Tadivaka – 1002112726**



A screenshot of a computer

Description automatically generated

**Task 1A:**

Display all the details about the dataset.

A close-up of a red line

Description automatically generated

A close-up of a computer screen

Description automatically generated

**Task 1B:**

Display all the details about the dataset.

1. Check for null values within each column.
2. Fill the values for more than 100 missing columns with certain labels such as No Director, Country Unavailable, etc. Drop the rows for the rest of the missing columns.

A screenshot of a computer

Description automatically generated

A screenshot of a computer program

Description automatically generated

**Task 1C:**

Display a two horizontal bar chart side by side for Top 10 countries with total number of movies and TV shows.

A screenshot of a computer program

Description automatically generated

A comparison of a graph

Description automatically generated

**Task 1D:**

Print the first row based on the longest duration time of a movie from each country column with its director, date\_added, release\_year, duration of the movie and its description.

A close-up of a computer code

Description automatically generated

A screenshot of a document

Description automatically generated

**Task 1E:**

Display the title of the movie, its director, the date it was added to the platform and the date it was officially released where the official release date and the date it was added to the plafform has same year.

A screenshot of a computer program

Description automatically generated

A screenshot of a computer

Description automatically generated

**Task 1F:**

Display Director, the release year and number of movies and TV Shows directed by that director within a year and sort the results from highest.

A screenshot of a computer

Description automatically generated

**Task 1G:**

Display the title of the movie/TV shows, the Director, the date it was added to netflix and the category it was listed in from the data which belongs to Documentary/Docuseries category.



A screenshot of a computer

Description automatically generated

**Task 1H:**

Display title, the date it was added to the plaform, type of category it was listed in and its description for Family Dramas.

Hint: Use Description to look for type of Drama.



A screenshot of a computer

Description automatically generated

**Task 1I:**

Plot the Distribution of TV shows based on their number of seasons. (horizontal bar chart)

1. Less than 3 seasons
2. 3 Seasons
3. 4 Seasons
4. 5 to less than 10 seasons.
5. 10 or more seasons.

A screenshot of a computer code

Description automatically generated

A graph with blue squares

Description automatically generated

**Task J:**

Display a side by side pie chart where it shows different rating a movie and a TV show belongs to.

Movie:

1. Uncut/Not rated
2. Restricted
3. Parentel guidance
4. General audience
5. Adults only

TV Show:

1. All Children
2. Older Children
3. Parentel Presence
4. General audience
5. Mature

A screenshot of a computer code

Description automatically generated

A comparison of pie charts

Description automatically generated

**References:**

<https://pandas.pydata.org/docs/user_guide/index.html#user-guide>

[Pandas groupby() and count() with Examples - Spark By {Examples} (sparkbyexamples.com)](https://sparkbyexamples.com/pandas/pandas-groupby-count-examples/)

[Using pandas and Python to Explore Your Dataset – Real Python](https://realpython.com/pandas-python-explore-dataset/)