# Exploratory Data Analysis Report

## Dataset – Netflix

| Task | Method Used |
|---|---|
| To show top 5 records of the dataset | data.head(5) |
| To show bottom 10 records of the dataset | data.tail(10) |
| To show no. of rows and columns | data.shape |
| To show no. of total values(elements) in the dataset | data.size |
| To show the column names of the dataset | data.columns |
| To show the data type of the columns of the dataset | data.dtypes |
| To show the indexes, column's datatypes and memory usage | data.info() |
| **TASK 1** | |
| To check duplicate records in the dataset | data.duplicated() |
| To Remove duplicates permanently | data.drop_duplicates(inplace = True) |
| **TASK 2** | |
| To show null values | data.isnull() |
| To show count of null values | data.isnull().sum() |
| To create heatmap for null value | import seaborn as sns<br>sns.heatmap(data.isnull() |
| **TASK 3** | |
| To display the show id and director name for "House of Cards" | data.loc[data.title == "House of cards", ['Show_id', 'Director']] |
| **TASK 4** | |
| To change the date datatype in a dataset | data['Date_N']=pd.to_datetime(data['Release_Date'], format ="mixed") |
| To find the year the highest no. of Tv shows and Movies were Released | data['Date_N'].dt.year.value_counts().head(1) |

| | |
|---|---|
| To show the count for individual | data['Date_N'].dt.year.value_counts() |
| To display the bar graph | lata['Date_N'].dt.year.value_counts().plot(kind = 'bar', color = red') |
| **TASK 5** | |
| To show the no of movies and Tv show<br>In the dataset | data.groupby('Category').Category.count() |
| To show the count plot | sns.countplot(x = 'Category', data=data ) |
| **TASK 6** | |
| To separate the year in the date column | data['year'] = data['Date_N'].dt.year |
| To show the movies released in 2000 | data[(data['Category'] == 'Movie') & (data['year'] == 2000)]<br>**BY USING FILTERING METHOD** |
| To show the title of TV shows were released in India | data[(data['Category'] == 'TV Show') & (data['Country'] == 'India')] ['Title']<br>**BY USING FILTERING METHOD** |
| To show the top 10 directors who gave the highest no. of Tv shows and Movies | data['Director'].value_counts().head(10) |
| Show all the records where "category is movie and type is comedies" or "county is united kingdom" | data[(data['Category']=='Movie')&(data['Type']=='Comedies') \| (data['Country']=='United Kingdom')]<br>**BY USING FILTERING METHOD** |
| To delete null values | new_data = data.dropna()<br>new_data.reset_index(inplace = True) |
| To show the no. of Movies/Tv shows , tom cruise were cast | new_data[new_data['Cast'].str.contains('Tom Cruise')] |
| **TASK 7** | |
| To show the unique rating in the dataset | data['Rating'].nunique()<br>data['Rating'].unique()<br>**BY USING UNIQUE METHOD** |
| To show distribution of content ratings in a graph | piechart = px.pie(x, values= 'counts', names = 'Rating', title = 'Distribution of content ratings')<br>piechart.show() |
| To check no of movies got "TV-14"rating in canada | data[(data['Category'] == 'Movie') & (data['Rating'] == 'TV-14') & (data['Country'] == 'Canada')].shape<br>**BY USING FILTERING METHOD** |
| To check no. of TV shows got "R" after 2018 | data[(data['Category'] == 'TV Show') & (data['Rating'] == 'R') & (data['year'] > 2018)] **BY USING FILTERING METHOD** |

| TASK 8 | |
|---|---|
| To show max duration of a Movie/TV show | data['Minutes'].max()<br>**BY USING MAX METHOD** |
| To show min duration of a Movie/TV show | data['Minutes'].min()<br>**BY USING MIN METHOD** |
| **TASK 9** | |
| To show the country has highest no. of Tv shows | data_tvshow.Country.value_counts().head(1)<br>**BY USING COUNT METHOD** |
| To show all the countries and no. of Tv shows in a graph | **BY USING COUNT PLOT IN SEABORN** |
| To compare each countries TV shows and Movies count in graph | data.groupby(['Category']).size().reset_index(name='counts')<br>**BY USING GROUPBY AND SIZE METHOD** |
| **TASK 10** | |
| To sort the dataset by year | data.sort_values(by = 'year')<br>**BY USING SORT METHOD** |
| To Show the count of movies/tv shows released in a year in line chart | **BY USING LINE CHART IN MATPLOTLIB** |
| To Show the Top 5 directors | data['Director'].value_counts().head(5)<br>**BY USING VALUE_COUNTS METHOD** |