EXPLORATORY DATA ANALYSIS

REPORT – by NIVEDHA R V

DATASET – Netflix.csv

TASK	METHOD USED
To show top 5 records of the dataset	nfdata.head(5)
To show bottom 10 records of the dataset	nfdata.tail(10)
To show no. of rows and columns	nfdata.shape
To show no of total values(elements) in the dataset	nfdata.size
To show the column names of the dataset	nfdata.columns
To show the data type of the columns of the dataset	nfdata.dtypes
To show the indexes, column's datatypes and memory usage	nfdata.info
TASK 1	
To check and show duplicate records in the dataset	nfdata.duplicated()
To remove duplicate records permanently	nfdata.drop_duplicates(inplace=True) nfdata
TASK 2	
To show null values present in dataset	nfdata.isnull()
To show the count of null values in each column	nfdata.isna().sum()
To create heatmap and show null value count	sb.heatmap(nfdata.isna())
TASK 3	
To show the show id and director name, for the show 'House of Cards'	nfdata.loc[nfdata["Title"]=="House of Cards"][["Show_Id","Director"]]
TASK 4	
To show the year that released highest number of tv shows and movies	nfdata[['Date','Year']] = nfdata.Release_Date.str.split(",",expand=True) year=nfdata.Year.value_counts() year[0:1]
To show the count for individual	nfdata.Year.value_counts()
To show them in bar graph	sb.barplot(x="Year",y="Category",data=nfdata) plt.gca().invert_yaxis()
TASK 5	
To show no of movies and tv shows in the dataset	nfdata.groupby("Category").Category.count()
To show the count of movies and tv shows separately	nfdata.Category.value_counts()
To show them using count plot as well.	sb.countplot(x="Category",data=nfdata)
TASK 6	
Show the movies that were released in the year 2000	nfdata[(nfdata.Year==2000) & (nfdata.Category=="Movie")][["Title","Category"]]

Show the title of tv shows that were released only in india	nfdata[nfdata.Country=="India"][["Title"]]
Show the Top 10 directors who gave the highest number of TV shows and movies to Netflix.	director=nfdata.Director.value_counts() director.head(10)
Show all the records where "category is movie and type is comedies" or "county is united kingdom"	nfdata[((nfdata.Type.str.contains("comedies"))& (nfdata.Category=="Movie"))| (nfdata.Country.str.contains("United Kingdom"))]
In how many movies/shows, Tom Cruise were cast? Show the records as well	nfdata.dropna(inplace=True) nfdata[(nfdata.Cast.str.contains("Tom Cruise"))][["Category"]].count() nfdata[(nfdata.Cast.str.contains("Tom Cruise"))]
TASK 7	
To show all the Ratings defined by Netflix?	nfdata.Rating.values
To show the distribution of content ratings in a graph	sb.displot(nfdata.Rating,aspect=3) plt.title("Distribution of Rating",fontsize=17)
To show the ratings and no. of unique ratings	no=nfdata.Rating.unique() no print("Num of Unique ratings=",len(no))
To show the no of movies that got "TV-14" rating in Canada	res=nfdata.query("Category.str.contains('ovie') & Rating.str.contains('TV-14') & Country=='Canada'") len(res)
To show the no of Tv Shows got rating "R" after year 2018	Ratings=nfdata[(nfdata["Rating"]=="R") & (nfdata["Year"]>=2018)][["Category","Year"]] len(Ratings)
TASK 8	
To show the maximum duration of a movie/tv show in Netflix	nfdata[["Minutes","Unit"]]=nfdata["Duration"].str.split(" ",expand=True) maxi=nfdata[nfdata.Unit.str.contains("eas")] maxi.max()
To Show the lowest duration of a movie/tv show in Netflix	nfdata[(nfdata.Unit.str.contains("min"))&(nfdata.Minutes.min())].min()
TASK 9	
To show the country that has the highest no. of .Tv Shows	nfdata[nfdata["Category"]=="TV Show"][["Country"]].value_counts().head(1)
To show all the countries and no. of tv shows in a graph To compare each countries tv show and movie count and show them in a graph	nfdata[nfdata["Category"]=="TV Show"][["Country"]].value_counts().plot(kind="barh")
TASK 10	
To Sort the dataset by year and show them	nfdata.sort_values("Year")
To show the count of movies/tv shows released in a year in line chart	nfdata[(nfdata["Category"]=="TV Show") | (nfdata["Category"]=="Movies")][["Yearp"]].value_counts() .plot(kind="line",fontsize=6)
To show the Top 5 directors in the Netflix based on the count of movies/tv shows directed by them.	nfdata["Director"].value_counts().head(5)