NAME:MUKESH ELANGO

4. Netflix Titles

In [2]:
```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as  sns
```

In [3]:
```python
a=pd.read_csv("netflix_titles.csv")
a.head()
```

Out[3]:

	show_id	type	title	director	cast	country	date_added	release_year	rati
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	T N
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	T N
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	T N
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	T N

In [4]:
```python
a.isnull()
```

Out[4]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	dur
0	False	False	False	False	True	False	False	False	False	
1	False	False	False	True	False	False	False	False	False	
2	False	False	False	False	False	True	False	False	False	
3	False	False	False	True	True	True	False	False	False	
4	False	False	False	True	False	False	False	False	False	
...	...	...	...	...	...	...	...	...	...	
8802	False	False	False	False	False	False	False	False	False	
8803	False	False	False	True	True	True	False	False	False	
8804	False	False	False	False	False	False	False	False	False	
8805	False	False	False	False	False	False	False	False	False	
8806	False	False	False	False	False	False	False	False	False	

8807 rows × 12 columns

In [5]: `a.drop_duplicates()`

Out[5]:

	show_id	type	title	director	cast	country	date_added	release_year
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021
...	...	...	...	...	...	...	...	...
8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2019	2007
8803	s8804	TV Show	Zombie Dumb	NaN	NaN	NaN	July 1, 2019	2018

	show_id	type	title	director	cast	country	date_added	release_year
8804	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	November 1, 2019	2009
8805	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	2006
8806	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015

8807 rows × 12 columns

In [6]: `a.bfill()`

Out[6]:

	show_id	type	title	director	cast	country	date_added	release_year
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	United States	September 25, 2021	2020
1	s2	TV Show	Blood & Water	Julien Leclercq	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	India	September 24, 2021	2021
3	s4	TV Show	Jailbirds New Orleans	Mike Flanagan	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021
4	s5	TV Show	Kota Factory	Mike Flanagan	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021
...	...	...	...	...	...	...	...	...
8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2019	2007
8803	s8804	TV Show	Zombie Dumb	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson,	United States	July 1, 2019	2018

	show_id	type	title	director	cast	country	date_added	release_year
8804	s8805	Movie	Zombieland	Ruben Fleischer	Emma Stone, … Jesse Eisenberg, Woody Harrelson, Emma Stone, …	United States	November 1, 2019	2009
8805	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma…	United States	January 11, 2020	2006
8806	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan…	India	March 2, 2019	2015

8807 rows × 12 columns

In [13]:
```python
a.ffill()
```

Out[13]:

	show_id	type	title	director	cast	country	date_added	release_year
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020
1	s2	TV Show	Blood & Water	Kirsten Johnson	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	South Africa	September 24, 2021	2021
3	s4	TV Show	Jailbirds New Orleans	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	South Africa	September 24, 2021	2021
4	s5	TV Show	Kota Factory	Julien Leclercq	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021
...	...	...	...	...	...	...	...	...
8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	United States	November 20, 2019	2007
8803	s8804	TV Show	Zombie Dumb	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey	United States	July 1, 2019	2018

	show_id	type	title	director	cast	country	date_added	release_year
					J...			
8804	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	United States	November 1, 2019	2009
8805	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	United States	January 11, 2020	2006
8806	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	India	March 2, 2019	2015

8807 rows × 12 columns

```
In [32]: a['country']=a['country'].str.upper()
         a
```

Out[32]:

	show_id	type	title	director	cast	country	date_added	release_year
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	UNITED STATES	September 25, 2021	2020
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	SOUTH AFRICA	September 24, 2021	2021
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	INDIA	September 24, 2021	2021
...	...	...	...	...	...	...	...	...
8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	UNITED STATES	November 20, 2019	2007
8803	s8804	TV Show	Zombie Dumb	NaN	NaN	NaN	July 1, 2019	2018

	show_id	type	title	director	cast	country	date_added	release_year
8804	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	UNITED STATES	November 1, 2019	2009
8805	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	UNITED STATES	January 11, 2020	2006
8806	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	INDIA	March 2, 2019	2015

8807 rows × 12 columns

In [56]:
```python
a['date_added']=pd.to_datetime(a['date_added'],dayfirst=True,errors='coerce')
a['date_added']=a['date_added'].dt.strftime('%d-%m-%Y')
a
```

Out[56]:

	show_id	type	title	director	cast	country	date_added	release_year
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	UNITED STATES	25-09-2021	2020
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	SOUTH AFRICA	24-09-2021	2021
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	24-09-2021	2021
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	24-09-2021	2021
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	INDIA	24-09-2021	2021
...	...	...	...	...	...	...	...	...
8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	UNITED STATES	20-11-2019	2007
8803	s8804	TV Show	Zombie Dumb	NaN	NaN	NaN	01-07-2019	2018

	show_id	type	title	director	cast	country	date_added	release_year
8804	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	UNITED STATES	01-11-2019	2009
8805	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	UNITED STATES	11-01-2020	2006
8806	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	INDIA	02-03-2019	2015

8807 rows × 12 columns

In [60]:
```python
a.columns = (a.columns.str.strip().str.upper())
a
```

Out[60]:

	SHOW_ID	TYPE	TITLE	DIRECTOR	CAST	COUNTRY	DATE_ADDED	RELE
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	UNITED STATES	25-09-2021	
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	SOUTH AFRICA	24-09-2021	
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	24-09-2021	
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	24-09-2021	
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	INDIA	24-09-2021	
...	...	...	...	...	...	...	...	
8802	s8803	Movie	Zodiac	David Fincher	Mark Ruffalo, Jake Gyllenhaal, Robert Downey J...	UNITED STATES	20-11-2019	
8803	s8804	TV Show	Zombie Dumb	NaN	NaN	NaN	01-07-2019	

	SHOW_ID	TYPE	TITLE	DIRECTOR	CAST	COUNTRY	DATE_ADDED	RELE
8804	s8805	Movie	Zombieland	Ruben Fleischer	Jesse Eisenberg, Woody Harrelson, Emma Stone, ...	UNITED STATES	01-11-2019	
8805	s8806	Movie	Zoom	Peter Hewitt	Tim Allen, Courteney Cox, Chevy Chase, Kate Ma...	UNITED STATES	11-01-2020	
8806	s8807	Movie	Zubaan	Mozez Singh	Vicky Kaushal, Sarah-Jane Dias, Raaghav Chanan...	INDIA	02-03-2019	

8807 rows × 12 columns

In [78]: `a.dtypes`

Out[78]:
```
SHOW_ID          object
TYPE             object
TITLE            object
DIRECTOR         object
CAST             object
COUNTRY          object
DATE_ADDED       object
RELEASE_YEAR      int64
RATING           object
DURATION         object
LISTED_IN        object
DESCRIPTION      object
dtype: object
```

a['RELEASE_YEAR'] = pd.to_datetime(a['RELEASE_YEAR'], errors='coerce') a.dtypes

5.Sales Data

In [172…
```
a=pd.read_csv("Auto Sales data.csv")
a.head()
```

Out[172...

	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	ORDER
0	10107	30	95.70	2	2871.00	24/02
1	10121	34	81.35	5	2765.90	07/05
2	10134	41	94.74	2	3884.34	01/07
3	10145	45	83.26	6	3746.70	25/08
4	10168	36	96.66	1	3479.76	28/10

In [97]: `a.isnull()`

Out[97]:

	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	ORD
0	False	False	False	False	False	
1	False	False	False	False	False	
2	False	False	False	False	False	
3	False	False	False	False	False	
4	False	False	False	False	False	
...	...	...	...	...	...	
2742	False	False	False	False	False	
2743	False	False	False	False	False	
2744	False	False	False	False	False	
2745	False	False	False	False	False	
2746	False	False	False	False	False	

2747 rows × 20 columns

In [99]: `a.bfill()`

Out[99]:

	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	OR
0	10107	30	95.70	2	2871.00	2
1	10121	34	81.35	5	2765.90	0
2	10134	41	94.74	2	3884.34	0
3	10145	45	83.26	6	3746.70	2
4	10168	36	96.66	1	3479.76	2
...	...	...	...	...	...	
2742	10350	20	112.22	15	2244.40	0
2743	10373	29	137.19	1	3978.51	3
2744	10386	43	125.99	4	5417.57	0
2745	10397	34	62.24	1	2116.16	2
2746	10414	47	65.52	9	3079.44	0

2747 rows × 20 columns

In [101...

```python
a.ffill()
```

Out[101…	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	OR
0	10107	30	95.70	2	2871.00	2
1	10121	34	81.35	5	2765.90	0
2	10134	41	94.74	2	3884.34	0
3	10145	45	83.26	6	3746.70	2
4	10168	36	96.66	1	3479.76	2
...	...	...	...	...	...	
2742	10350	20	112.22	15	2244.40	0
2743	10373	29	137.19	1	3978.51	3
2744	10386	43	125.99	4	5417.57	0
2745	10397	34	62.24	1	2116.16	2
2746	10414	47	65.52	9	3079.44	0

2747 rows × 20 columns

In [103…
```python
a.drop_duplicates()
```

Out[103...

	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	OR
0	10107	30	95.70	2	2871.00	2
1	10121	34	81.35	5	2765.90	0
2	10134	41	94.74	2	3884.34	0
3	10145	45	83.26	6	3746.70	2
4	10168	36	96.66	1	3479.76	2
...	...	...	...	...	...	
2742	10350	20	112.22	15	2244.40	0
2743	10373	29	137.19	1	3978.51	3
2744	10386	43	125.99	4	5417.57	0
2745	10397	34	62.24	1	2116.16	2
2746	10414	47	65.52	9	3079.44	0

2747 rows × 20 columns

In [105...
```python
a['PRODUCTLINE']=a['PRODUCTLINE'].str.upper()
a
```

Out[105…

	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	OR
0	10107	30	95.70	2	2871.00	2
1	10121	34	81.35	5	2765.90	0
2	10134	41	94.74	2	3884.34	0
3	10145	45	83.26	6	3746.70	2
4	10168	36	96.66	1	3479.76	2
...	...	...	...	...	...	
2742	10350	20	112.22	15	2244.40	0
2743	10373	29	137.19	1	3978.51	3
2744	10386	43	125.99	4	5417.57	0
2745	10397	34	62.24	1	2116.16	2
2746	10414	47	65.52	9	3079.44	0

2747 rows × 20 columns

In [107…

```
a['ORDERDATE']=pd.to_datetime(a['ORDERDATE'],dayfirst=True,errors='coerce')
a['ORDERDATE']=a['ORDERDATE'].dt.strftime('%d-%m-%Y')
a
```

Out[107…

	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	OR
0	10107	30	95.70	2	2871.00	2
1	10121	34	81.35	5	2765.90	0
2	10134	41	94.74	2	3884.34	0
3	10145	45	83.26	6	3746.70	2
4	10168	36	96.66	1	3479.76	2
...	...	...	...	...	...	
2742	10350	20	112.22	15	2244.40	0
2743	10373	29	137.19	1	3978.51	3
2744	10386	43	125.99	4	5417.57	0
2745	10397	34	62.24	1	2116.16	2
2746	10414	47	65.52	9	3079.44	0

2747 rows × 20 columns

In [111…

```python
a.columns = (a.columns.str.strip().str.lower())
a
```

Out[111…

	ordernumber	quantityordered	priceeach	orderlinenumber	sales	orderdate	day
0	10107	30	95.70	2	2871.00	24-02-2018	
1	10121	34	81.35	5	2765.90	07-05-2018	
2	10134	41	94.74	2	3884.34	01-07-2018	
3	10145	45	83.26	6	3746.70	25-08-2018	
4	10168	36	96.66	1	3479.76	28-10-2018	
...	...	...	...	...	...	...	
2742	10350	20	112.22	15	2244.40	02-12-2019	
2743	10373	29	137.19	1	3978.51	31-01-2020	
2744	10386	43	125.99	4	5417.57	01-03-2020	
2745	10397	34	62.24	1	2116.16	28-03-2020	
2746	10414	47	65.52	9	3079.44	06-05-2020	

2747 rows × 20 columns

In [113…

```
a.dtypes
```

```
Out[113…    ordernumber                  int64
            quantityordered              int64
            priceeach                  float64
            orderlinenumber              int64
            sales                      float64
            orderdate                   object
            days_since_lastorder         int64
            status                      object
            productline                 object
            msrp                         int64
            productcode                 object
            customername                object
            phone                       object
            addressline1                object
            city                        object
            postalcode                  object
            country                     object
            contactlastname             object
            contactfirstname            object
            dealsize                    object
            dtype: object
```

In [117…
```python
a['msrp'] = pd.to_datetime(a['msrp'], errors='coerce')
a['days_since_lastorder'] = pd.to_datetime(a['days_since_lastorder'], errors='coerc
a['orderlinenumber'] = pd.to_datetime(a['orderlinenumber'], errors='coerce')
a.dtypes
```

```
Out[117…    ordernumber                       int64
            quantityordered                   int64
            priceeach                       float64
            orderlinenumber         datetime64[ns]
            sales                           float64
            orderdate                        object
            days_since_lastorder    datetime64[ns]
            status                           object
            productline                      object
            msrp                    datetime64[ns]
            productcode                      object
            customername                     object
            phone                            object
            addressline1                     object
            city                             object
            postalcode                       object
            country                          object
            contactlastname                  object
            contactfirstname                 object
            dealsize                         object
            dtype: object
```

3.Mall Customer Segmentation Data

In [120…
```python
a=pd.read_csv("Mall_Customers.csv")
a.head()
```

Out[120…

	CustomerID	Gender	Age	Annual Income (k$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40

In [122…

```
a.isnull()
```

Out[122…

	CustomerID	Gender	Age	Annual Income (k$)	Spending Score (1-100)
0	False	False	False	False	False
1	False	False	False	False	False
2	False	False	False	False	False
3	False	False	False	False	False
4	False	False	False	False	False
...	...	...	...	...	...
195	False	False	False	False	False
196	False	False	False	False	False
197	False	False	False	False	False
198	False	False	False	False	False
199	False	False	False	False	False

200 rows × 5 columns

In [124…

```
a.bfill()
```

Out[124…

	CustomerID	Gender	Age	Annual Income (k$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40
...	...	...	...	...	...
195	196	Female	35	120	79
196	197	Female	45	126	28
197	198	Male	32	126	74
198	199	Male	32	137	18
199	200	Male	30	137	83

200 rows × 5 columns

In [126…

```
a.ffill()
```

Out[126…

	CustomerID	Gender	Age	Annual Income (k$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40
...	...	...	...	...	...
195	196	Female	35	120	79
196	197	Female	45	126	28
197	198	Male	32	126	74
198	199	Male	32	137	18
199	200	Male	30	137	83

200 rows × 5 columns

In [128…

```
a.drop_duplicates()
```

Out[128…

	CustomerID	Gender	Age	Annual Income (k$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40
...	...	...	...	...	...
195	196	Female	35	120	79
196	197	Female	45	126	28
197	198	Male	32	126	74
198	199	Male	32	137	18
199	200	Male	30	137	83

200 rows × 5 columns

In [130…

```
a['Gender']=a['Gender'].str.upper()
a
```

Out[130…

	CustomerID	Gender	Age	Annual Income (k$)	Spending Score (1-100)
0	1	MALE	19	15	39
1	2	MALE	21	15	81
2	3	FEMALE	20	16	6
3	4	FEMALE	23	16	77
4	5	FEMALE	31	17	40
...	...	...	...	...	...
195	196	FEMALE	35	120	79
196	197	FEMALE	45	126	28
197	198	MALE	32	126	74
198	199	MALE	32	137	18
199	200	MALE	30	137	83

200 rows × 5 columns

In [132…

```
a.columns = (a.columns.str.strip().str.upper())
a
```

Out[132...

	CUSTOMERID	GENDER	AGE	ANNUAL INCOME (K$)	SPENDING SCORE (1-100)
0	1	MALE	19	15	39
1	2	MALE	21	15	81
2	3	FEMALE	20	16	6
3	4	FEMALE	23	16	77
4	5	FEMALE	31	17	40
...	...	...	...	...	...
195	196	FEMALE	35	120	79
196	197	FEMALE	45	126	28
197	198	MALE	32	126	74
198	199	MALE	32	137	18
199	200	MALE	30	137	83

200 rows × 5 columns

In [134...
```
a.dtypes
```

Out[134...
```
CUSTOMERID                int64
GENDER                    object
AGE                       int64
ANNUAL INCOME (K$)        int64
SPENDING SCORE (1-100)    int64
dtype: object
```

In [138...
```
a['AGE'] = pd.to_datetime(a['AGE'], errors='coerce')
a.dtypes
```

Out[138...
```
CUSTOMERID                     int64
GENDER                    datetime64[ns]
AGE                       datetime64[ns]
ANNUAL INCOME (K$)             int64
SPENDING SCORE (1-100)         int64
dtype: object
```

2.Medical Appointment No Shows

In [141...
```
a=pd.read_csv("Medical Appointment No Shows.csv")
a.head()
```

Out[141…

	PatientId	AppointmentID	Gender	ScheduledDay	AppointmentDay	Age
0	2.987250e+13	5642903	F	2016-04-29T18:38:08Z	2016-04-29T00:00:00Z	62
1	5.589978e+14	5642503	M	2016-04-29T16:08:27Z	2016-04-29T00:00:00Z	56
2	4.262962e+12	5642549	F	2016-04-29T16:19:04Z	2016-04-29T00:00:00Z	62
3	8.679512e+11	5642828	F	2016-04-29T17:29:31Z	2016-04-29T00:00:00Z	8
4	8.841186e+12	5642494	F	2016-04-29T16:07:23Z	2016-04-29T00:00:00Z	56

In [143…

```
a.isnull()
```

Out[143…

	PatientId	AppointmentID	Gender	ScheduledDay	AppointmentDay	Age	Neighb
0	False	False	False	False	False	False	
1	False	False	False	False	False	False	
2	False	False	False	False	False	False	
3	False	False	False	False	False	False	
4	False	False	False	False	False	False	
...	...	...	...	...	...	...	
110522	False	False	False	False	False	False	
110523	False	False	False	False	False	False	
110524	False	False	False	False	False	False	
110525	False	False	False	False	False	False	
110526	False	False	False	False	False	False	

110527 rows × 14 columns

In [145…

```
a.bfill()
```

Out[145…

	PatientId	AppointmentID	Gender	ScheduledDay	AppointmentDay
0	2.987250e+13	5642903	F	2016-04-29T18:38:08Z	2016-04-29T00:00:00Z
1	5.589978e+14	5642503	M	2016-04-29T16:08:27Z	2016-04-29T00:00:00Z
2	4.262962e+12	5642549	F	2016-04-29T16:19:04Z	2016-04-29T00:00:00Z
3	8.679512e+11	5642828	F	2016-04-29T17:29:31Z	2016-04-29T00:00:00Z
4	8.841186e+12	5642494	F	2016-04-29T16:07:23Z	2016-04-29T00:00:00Z
...	...	...	...	...	...
110522	2.572134e+12	5651768	F	2016-05-03T09:15:35Z	2016-06-07T00:00:00Z
110523	3.596266e+12	5650093	F	2016-05-03T07:27:33Z	2016-06-07T00:00:00Z
110524	1.557663e+13	5630692	F	2016-04-27T16:03:52Z	2016-06-07T00:00:00Z
110525	9.213493e+13	5630323	F	2016-04-27T15:09:23Z	2016-06-07T00:00:00Z
110526	3.775115e+14	5629448	F	2016-04-27T13:30:56Z	2016-06-07T00:00:00Z

110527 rows × 14 columns

In [147…

```python
a.ffill()
```

Out[147…

	PatientId	AppointmentID	Gender	ScheduledDay	AppointmentDay
0	2.987250e+13	5642903	F	2016-04-29T18:38:08Z	2016-04-29T00:00:00Z
1	5.589978e+14	5642503	M	2016-04-29T16:08:27Z	2016-04-29T00:00:00Z
2	4.262962e+12	5642549	F	2016-04-29T16:19:04Z	2016-04-29T00:00:00Z
3	8.679512e+11	5642828	F	2016-04-29T17:29:31Z	2016-04-29T00:00:00Z
4	8.841186e+12	5642494	F	2016-04-29T16:07:23Z	2016-04-29T00:00:00Z
...	...	...	...	...	...
110522	2.572134e+12	5651768	F	2016-05-03T09:15:35Z	2016-06-07T00:00:00Z
110523	3.596266e+12	5650093	F	2016-05-03T07:27:33Z	2016-06-07T00:00:00Z
110524	1.557663e+13	5630692	F	2016-04-27T16:03:52Z	2016-06-07T00:00:00Z
110525	9.213493e+13	5630323	F	2016-04-27T15:09:23Z	2016-06-07T00:00:00Z
110526	3.775115e+14	5629448	F	2016-04-27T13:30:56Z	2016-06-07T00:00:00Z

110527 rows × 14 columns

In [149…

```python
a.drop_duplicates()
```

Out[149…

	PatientId	AppointmentID	Gender	ScheduledDay	AppointmentDay
0	2.987250e+13	5642903	F	2016-04-29T18:38:08Z	2016-04-29T00:00:00Z
1	5.589978e+14	5642503	M	2016-04-29T16:08:27Z	2016-04-29T00:00:00Z
2	4.262962e+12	5642549	F	2016-04-29T16:19:04Z	2016-04-29T00:00:00Z
3	8.679512e+11	5642828	F	2016-04-29T17:29:31Z	2016-04-29T00:00:00Z
4	8.841186e+12	5642494	F	2016-04-29T16:07:23Z	2016-04-29T00:00:00Z
...	...	...	...	...	...
110522	2.572134e+12	5651768	F	2016-05-03T09:15:35Z	2016-06-07T00:00:00Z
110523	3.596266e+12	5650093	F	2016-05-03T07:27:33Z	2016-06-07T00:00:00Z
110524	1.557663e+13	5630692	F	2016-04-27T16:03:52Z	2016-06-07T00:00:00Z
110525	9.213493e+13	5630323	F	2016-04-27T15:09:23Z	2016-06-07T00:00:00Z
110526	3.775115e+14	5629448	F	2016-04-27T13:30:56Z	2016-06-07T00:00:00Z

110527 rows × 14 columns

In [151…

```python
a['Neighbourhood']=a['Neighbourhood'].str.lower()
a
```

Out[151...

	PatientId	AppointmentID	Gender	ScheduledDay	AppointmentDay
0	2.987250e+13	5642903	F	2016-04-29T18:38:08Z	2016-04-29T00:00:00Z
1	5.589978e+14	5642503	M	2016-04-29T16:08:27Z	2016-04-29T00:00:00Z
2	4.262962e+12	5642549	F	2016-04-29T16:19:04Z	2016-04-29T00:00:00Z
3	8.679512e+11	5642828	F	2016-04-29T17:29:31Z	2016-04-29T00:00:00Z
4	8.841186e+12	5642494	F	2016-04-29T16:07:23Z	2016-04-29T00:00:00Z
...	...	...	...	...	...
110522	2.572134e+12	5651768	F	2016-05-03T09:15:35Z	2016-06-07T00:00:00Z
110523	3.596266e+12	5650093	F	2016-05-03T07:27:33Z	2016-06-07T00:00:00Z
110524	1.557663e+13	5630692	F	2016-04-27T16:03:52Z	2016-06-07T00:00:00Z
110525	9.213493e+13	5630323	F	2016-04-27T15:09:23Z	2016-06-07T00:00:00Z
110526	3.775115e+14	5629448	F	2016-04-27T13:30:56Z	2016-06-07T00:00:00Z

110527 rows × 14 columns

In [153...

```python
a['ScheduledDay']=pd.to_datetime(a['ScheduledDay'],dayfirst=True,errors='coerce')
a['ScheduledDay']=a['ScheduledDay'].dt.strftime('%d-%m-%Y')
a
```

```
C:\Users\umamu\AppData\Local\Temp\ipykernel_16184\3418324271.py:1: UserWarning: Pars
ing dates in %Y-%m-%dT%H:%M:%S%z format when dayfirst=True was specified. Pass `dayf
irst=False` or specify a format to silence this warning.
  a['ScheduledDay']=pd.to_datetime(a['ScheduledDay'],dayfirst=True,errors='coerce')
```

Out[153…

	PatientId	AppointmentID	Gender	ScheduledDay	AppointmentDay	Age
0	2.987250e+13	5642903	F	29-04-2016	2016-04-29T00:00:00Z	62
1	5.589978e+14	5642503	M	29-04-2016	2016-04-29T00:00:00Z	56
2	4.262962e+12	5642549	F	29-04-2016	2016-04-29T00:00:00Z	62
3	8.679512e+11	5642828	F	29-04-2016	2016-04-29T00:00:00Z	8
4	8.841186e+12	5642494	F	29-04-2016	2016-04-29T00:00:00Z	56
...	...	...	...	...	...	...
110522	2.572134e+12	5651768	F	03-05-2016	2016-06-07T00:00:00Z	56
110523	3.596266e+12	5650093	F	03-05-2016	2016-06-07T00:00:00Z	51
110524	1.557663e+13	5630692	F	27-04-2016	2016-06-07T00:00:00Z	21
110525	9.213493e+13	5630323	F	27-04-2016	2016-06-07T00:00:00Z	38
110526	3.775115e+14	5629448	F	27-04-2016	2016-06-07T00:00:00Z	54

110527 rows × 14 columns

In [155…
```python
a.columns = (a.columns.str.strip().str.upper())
a
```

Out[155…

	PATIENTID	APPOINTMENTID	GENDER	SCHEDULEDDAY	APPOINTMENTDAY	
0	2.987250e+13	5642903	F	29-04-2016	2016-04-29T00:00:00Z	
1	5.589978e+14	5642503	M	29-04-2016	2016-04-29T00:00:00Z	
2	4.262962e+12	5642549	F	29-04-2016	2016-04-29T00:00:00Z	
3	8.679512e+11	5642828	F	29-04-2016	2016-04-29T00:00:00Z	
4	8.841186e+12	5642494	F	29-04-2016	2016-04-29T00:00:00Z	
...	...	...	...	...	...	...
110522	2.572134e+12	5651768	F	03-05-2016	2016-06-07T00:00:00Z	
110523	3.596266e+12	5650093	F	03-05-2016	2016-06-07T00:00:00Z	
110524	1.557663e+13	5630692	F	27-04-2016	2016-06-07T00:00:00Z	
110525	9.213493e+13	5630323	F	27-04-2016	2016-06-07T00:00:00Z	
110526	3.775115e+14	5629448	F	27-04-2016	2016-06-07T00:00:00Z	

110527 rows × 14 columns

In [157…   ```python
            a.dtypes
            ```

Out[157…   ```
            PATIENTID          float64
            APPOINTMENTID        int64
            GENDER              object
            SCHEDULEDDAY        object
            APPOINTMENTDAY      object
            AGE                 int64
            NEIGHBOURHOOD       object
            SCHOLARSHIP         int64
            HIPERTENSION        int64
            DIABETES            int64
            ALCOHOLISM          int64
            HANDCAP             int64
            SMS_RECEIVED        int64
            NO-SHOW             object
            dtype: object
            ```

In [162…   ```python
            a['AGE'] = pd.to_datetime(a['AGE'], errors='coerce')
            a.dtypes
            ```

Out[162…   ```
            PATIENTID             float64
            APPOINTMENTID           int64
            GENDER                 object
            SCHEDULEDDAY           object
            APPOINTMENTDAY         object
            AGE            datetime64[ns]
            NEIGHBOURHOOD          object
            SCHOLARSHIP             int64
            HIPERTENSION            int64
            DIABETES                int64
            ALCOHOLISM              int64
            HANDCAP                 int64
            SMS_RECEIVED            int64
            NO-SHOW                object
            dtype: object
            ```

In [ ]: