VIEWERSHIP ANALYSIS

Exploratory Data Analysis using Python, they include checking for missing values, understanding data types and getting summary statistics.

1. Importing libraries

```
[15] # for data manipulation & mathematical calculations
     import pandas as pd
     import numpy as np
```

18s

2 Data ingestion

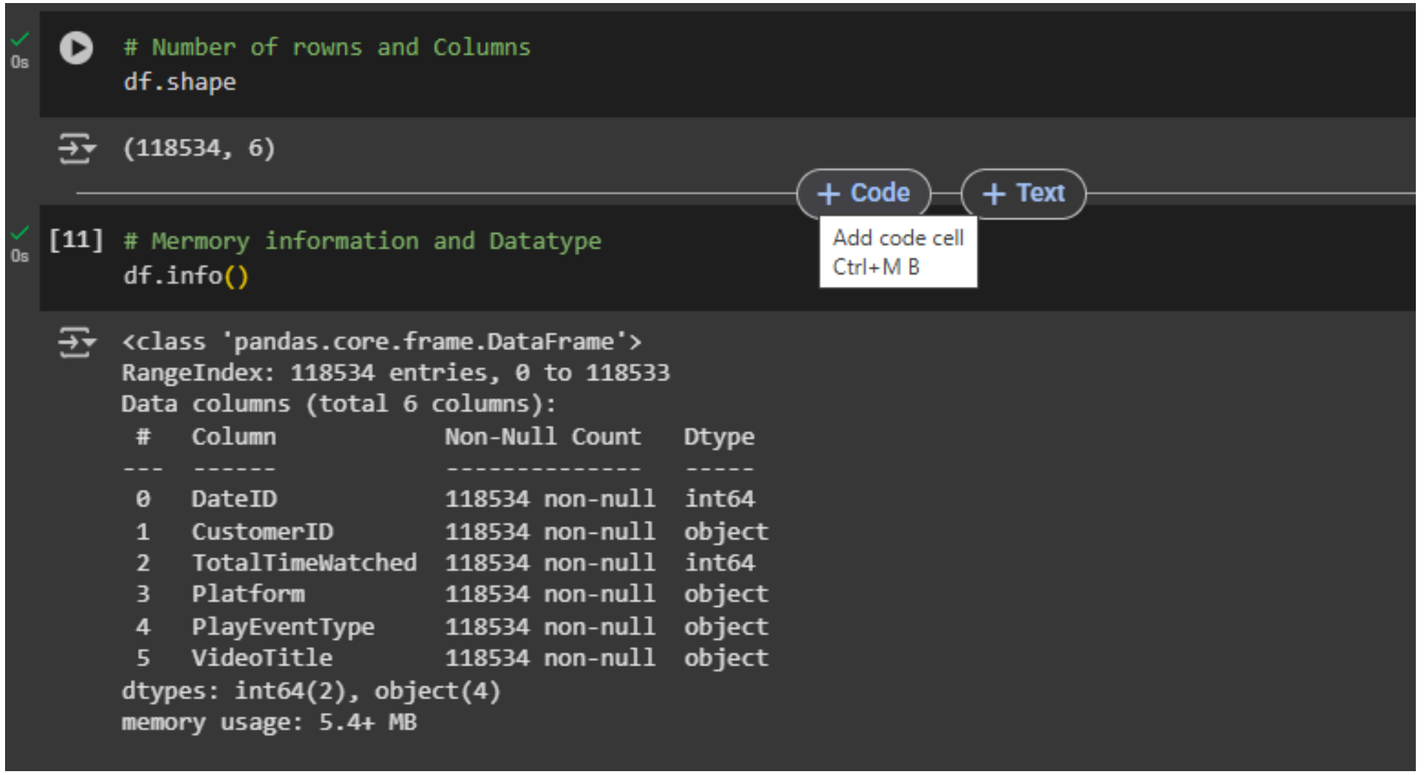
```
[7]  data_path='/content/Viewership Analysis .xlsx - Data.csv'
```

0s

```
# From a csv file
df=pd.read_csv(data_path)
```

```
[9]  # Summary Stats for numerical Columns
     display(df)
```

	DateID	CustomerID	TotalTimeWatched	Platform	PlayEventType	VideoTitle
0	20201101	EW1DENH0EC1J3M9WAOZF9LSV004O	300	Leanback	LiveTV	F1 '20: Emilia Romagna GP
1	20201101	EW1DENH0EC1J3M9WAOZF9LSV004O	300	Leanback	LiveTV	F1 '20: Emilia Romagna GP
2	20201101	6TS2LLY0L3G66FVY86Q0JEZE000K	360	Leanback	Other	Chasing The Sun
3	20201101	6TS2LLY0L3G66FVY86Q0JEZE000K	360	Leanback	Other	Chasing The Sun
4	20201101	6PMV67PLJ2S47S68J0Y30XFK003C	120	Leanback	LiveTV	Sonic The Hedgehog
...	...	...	...	...	...	...
118529	20210416	DGXM5TJ18W1LANOLKJJN5MFT000G	600	Leanback	LiveTV	New Day
118530	20210416	DMMVDBZFDT1F7WL4BGWQ5KOJ000G	5194	Web	LiveTV	Buying And Selling With Property...
118531	20210416	1JKBIAY7MEUH6BWX0SCU93GJ000I	47660	Web	LiveTV	Expresso


0s
Number of rowns and Columns
df.shape
(118534, 6)
+ Code
+ Text
Add code cell
Ctrl+M B
[11]
0s
# Mermory information and Datatype
df.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 118534 entries, 0 to 118533
Data columns (total 6 columns):
Column Non-Null Count Dtype
--- ------ -------------- -----
0 DateID 118534 non-null int64
1 CustomerID 118534 non-null object
2 TotalTimeWatched 118534 non-null int64
3 Platform 118534 non-null object
4 PlayEventType 118534 non-null object
5 VideoTitle 118534 non-null object
dtypes: int64(2), object(4)
memory usage: 5.4+ MB

```
[12] # Views the unique Values and Counts
     df.value_counts(dropna=False)
```

DateID	CustomerID	TotalTimeWatched	Platform	PlayEventType	VideoTitle	count
20201222	KN42BQKA57DYKED96YC6FB3E002E	60	Android	LiveTV	Unknown	20
20210203	JAJ82KH7E2I02BRM9VQJHDJM004H	1140	Leanback	Catch Up	Man V. Food	20
20210313	14WMFY74CBBN2CETEUP15O78001Q	60	Leanback	LiveTV	Unknown	16
20210117	1KFN67JSCELVDIBF4QZN391X002C	60	Leanback	LiveTV	Unknown	14
20210317	F2CI4KOVM23ILX8X0PHIGNEI006P	60	Leanback	LiveTV	Unknown	13
...	...	...	...	...	...	...
20210416	MDB1B8CG7EBC81XV199E2Y39006D	480	Android	LiveTV	Unknown	1
	M2KHDKDI8JIVAQ5HJDM2EAJM004F	17388	Web	LiveTV	Ontbyt Grotes	1
	M2EACDFMI3UO6C5N4FM297ZW004F	1234	Android	LiveTV	Unknown	1
20201101	6PMV67PLJ2S47S68J0Y30XFK003C	120	Leanback	LiveTV	Sonic The Hedgehog	1

Frozen II	1

108661 rows × 1 columns

dtype: int64

```
[13] # Detect and count missing values
     df.isnull().sum()
```

	0
DateID	0
CustomerID	0
TotalTimeWatched	0
Platform	0
PlayEventType	0
VideoTitle	0

dtype: int64

+ Code + Text

[14]
```
# Summary statistics for numerical columns
df.describe()
```

	DateID	TotalTimeWatched
count	1.185340e+05	118534.000000
mean	2.020945e+07	2046.869860
std	2.583117e+03	3735.512071
min	2.020110e+07	1.000000
25%	2.021012e+07	240.000000
50%	2.021022e+07	1020.000000
75%	2.021032e+07	2400.000000
max	2.021042e+07	88500.000000

```
[16] # Viewing the first 5 rows
     df.head(5)
```

	DateID	CustomerID	TotalTimeWatched	Platform	PlayEventType	VideoTitle
0	20201101	EW1DENH0EC1J3M9WAOZF9LSV004O	300	Leanback	LiveTV	F1 '20: Emilia Romagna GP
1	20201101	EW1DENH0EC1J3M9WAOZF9LSV004O	300	Leanback	LiveTV	F1 '20: Emilia Romagna GP
2	20201101	6TS2LLY0L3G66FVY86Q0JEZE000K	360	Leanback	Other	Chasing The Sun
3	20201101	6TS2LLY0L3G66FVY86Q0JEZE000K	360	Leanback	Other	Chasing The Sun
4	20201101	6PMV67PLJ2S47S68J0Y30XFK003C	120	Leanback	LiveTV	Sonic The Hedgehog

```
#Viewing the last 5 rows
df.tail(5)
```

	DateID	CustomerID	TotalTimeWatched	Platform	PlayEventType	VideoTitle
118529	20210416	DGXM5TJ18W1LANOLKJJN5MFT000G	600	Leanback	LiveTV	New Day
118530	20210416	DMMVDBZFDT1F7WL4BGWQ5KOJ000G	5194	Web	LiveTV	Buying And Selling With Property...
118531	20210416	1JKBIAY7MEUH6BWX0SCU93GJ000I	47660	Web	LiveTV	Expresso
118532	20210416	1ADX18R3B33REZ2T0NK911E4002W	1335	Leanback	LiveTV	Unknown
118533	20210416	BO0REPHA6FP7C15B0K9N1GVO0050	360	iOS	LiveTV	Imbewu

```
# Returns the number of non-null values in each dataframe
df.count()
```

	0
DateID	118534
CustomerID	118534
TotalTimeWatched	118534
Platform	118534
PlayEventType	118534
VideoTitle	118534

dtype: int64

```python
# Returns the highest value in the columns
df.max()
```

	0
DateID	20210416
CustomerID	MGSP66WAG6HL0U4C8TCMJU5L002C
TotalTimeWatched	88500
Platform	iOS
PlayEventType	Other
VideoTitle	zou

dtype: object

```
[20] # Returns the lowest value in the columns
     df.min()
```

	0
DateID	20201101
CustomerID	00ICB6499OJXJ380F7HIKZBJ006Y
TotalTimeWatched	1
Platform	Android
PlayEventType	Catch Up
VideoTitle	#10 Yanos

dtype: object