

EDA Analysis on Python

Promotes loose coupling between components by proving object

Install Dependencies

```
Oct 20, 2025 (4s) 2 Python Generate (Ctrl + I)  
#### To install and manage soft were packages  
!pip install openpyxl -q  
Note: you may need to restart the kernel using %restart_python or dbutils.library.restartPython() to use updated packages.  
💡
```

Assist offer specialized function

Import libraries

We have the data of the viewership analysis of the movies and tv shows.

The total time watched by the viewers. The platform is the platform where the movies and tv shows are watched. The play event type is the type of event that is played. The video title is the title of the movie or tv show.

meta data

```
Oct 20, 2025 (<1s) 5 Python  
#### importing panda help us process data  
import pandas as dp  
  
#### Importing numpy library help us to perform mathematical operations  
import numpy as np
```

```
Oct 20, 2025 (<1s) 6  
#### To access the file  
data_path = "/Workspace/Users/mavismbr@gmail.com/Viewership Analysis .xlsx"
```

```
Oct 20, 2025 (11s) 7  
#### Attaching the excel file  
  
survey_analysis = dp.read_excel(data_path)  
survey_analysis
```

```
Oct 20, 2025 (2s) 8  
#### display data given
```

Oct 20, 2025 (23)

```
### display data given

display(survey_analysis)
> See performance (1)

Table +
```

	DateID	CustomerID	TotalTimeWatched	Platform	PlayEventType	VideoTitle
1	20201101	EW1DENH0EC1J3M9WAQZF9LSV...	300	Leanback	LiveTV	F1 '20: Emilia Romagna GP
2	20201101	EW1DENH0EC1J3M9WAQZF9LSV...	300	Leanback	LiveTV	F1 '20: Emilia Romagna GP
3	20201101	6TS2LLY0L3G66FVV86Q0JEZE000K	360	Leanback	Other	Chasing The Sun
4	20201101	6TS2LLY0L3G66FVV86Q0JEZE000K	360	Leanback	Other	Chasing The Sun
5	20201101	6PMV67PLJ2S47S6B0Y30XFK003C	120	Leanback	LiveTV	Sonic The Hedgehog
6	20201101	6PMV67PLJ2S47S6B0Y30XFK003C	120	Leanback	LiveTV	Frozen II
7	20201101	6PMV67PLJ2S47S6B0Y30XFK003C	120	iOS	LiveTV	PL 20/21: Man Utd v Arsenal
8	20201101	6PMV67PLJ2S47S6B0Y30XFK003C	120	iOS	LiveTV	PL 20/21: Man Utd v Arsenal
9	20201101	6PMV67PLJ2S47S6B0Y30XFK003C	60	iOS	LiveTV	F1 '20: Emilia Romagna GP
10	20201101	6PMV67PLJ2S47S6B0Y30XFK003C	600	iOS	LiveTV	F1 '20: Emilia Romagna GP
11	20201101	JMCVLXSA196YJ0KC6ID0IEVB000...	1860	iOS	LiveTV	Betrayed
12	20201101	JMCVLXSA196YJ0KC6ID0IEVB000...	960	iOS	Catch Up	The Bradshaw Bunch
13	20201101	JMCVLXSA196YJ0KC6ID0IEVB000...	1080	iOS	Catch Up	Mbau Reloaded: Always Rise
14	20201101	JMCVLXSA196YJ0KC6ID0IEVB000...	240	iOS	Catch Up	The Bradshaw Bunch
15	20201101	JMCVLXSA196YJ0KC6ID0IEVB000...	120	iOS	LiveTV	PL 20/21: Newcastle v Everton

1000+ rows | Truncated data | 1.70s runtime

Refreshed 13 days ago

Oct 20, 2025 (<1s)

```
### display given number of rows and columns
survey_analysis.shape
```

(118534, 6)

Oct 20, 2025 (<1s)

```
### display Checking the data types for the columns

survey_analysis.dtypes
```

DateID	int64
CustomerID	object
TotalTimeWatched	int64
Platform	object
PlayEventType	object
VideoTitle	object
dtype:	object

Oct 20, 2025 (<1s)

```
### display the columns

survey_analysis.columns
```

```
Index(['DateID', 'CustomerID', 'TotalTimeWatched', 'Platform', 'PlayEventType',
       'VideoTitle'],
      dtype='object')
```

Oct 19, 2025 (<1s)

```
### display data information
survey_analysis.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
```

```

▶ ✓ Oct 19, 2025 (<1s) 13
    ### display unique values for the column 'Platform'

    survey_analysis['Platform'].unique()

array(['Leanback', 'iOS', 'Web', 'Android'], dtype=object)

▶ ✓ Oct 19, 2025 (<1s) 14
    ### display unique values for the column 'PlayEventType'

    survey_analysis['PlayEventType'].unique()

array(['LiveTV', 'Other', 'Catch Up', 'Downloads'], dtype=object)

▶ ✓ Oct 19, 2025 (<1s) 15
    ## display minimum time watched for the column 'TotalTimeWatched'

    survey_analysis['TotalTimeWatched'].min()

np.int64(1)

```

```

▶ ✓ Oct 19, 2025 (<1s) 16
    ### Check maximum time watched for the column 'TotalTimeWatched'

    survey_analysis['TotalTimeWatched'].max()

```

17

```

    ##### display the summary of the data

    survey_analysis.describe()

<bound method NDFrame.describe of      DateID ...          VideoTitle
0   20201101 ...           F1 '20: Emilia Romagna GP
1   20201101 ...           F1 '20: Emilia Romagna GP
2   20201101 ...           Chasing The Sun
3   20201101 ...           Chasing The Sun
4   20201101 ...           Sonic The Hedgehog
...
118529 20210416 ...           New Day
118530 20210416 ...   Buying And Selling With Property...
118531 20210416 ...           Espresso
118532 20210416 ...           Unknown
118533 20210416 ...           Imbewu
[118534 rows x 6 columns]>

```

18

```

    ✓ Oct 19, 2025 (<1s)
    ### display the fist 5 rows

    survey_analysis.head (5)

      DateID  CustomerID  TotalTimeWatched  Platform  PlayEventType  VideoTitle
0  20201101  EW1DENH0EC1J3M9WA0ZF9LSV004O        300  Leanback     LiveTV  F1 '20: Emilia Romagna GP
1  20201101  EW1DENH0EC1J3M9WA0ZF9LSV004O        300  Leanback     LiveTV  F1 '20: Emilia Romagna GP
2  20201101  6TS2LLY0L3G66FVY86Q0JEZ0E000K        360  Leanback      Other  Chasing The Sun
3  20201101  6TS2LLY0L3G66FVY86Q0JEZ0E000K        360  Leanback      Other  Chasing The Sun
4  20201101  6PMV67PLJ2S47S68J0Y30XFK003C       120  Leanback     LiveTV  Sonic The Hedgehog

```

19

```

    ✓ Oct 19, 2025 (<1s)
    ### display the last 5 rows

    survey_analysis.tail (5)

      DateID  CustomerID  TotalTimeWatched  Platform  PlayEventType  VideoTitle
118529  20210416  DGXM5TJ18W1LANOLKJN5MFT000G        600  Leanback     LiveTV  New Day
118530  20210416  DMMVDBZFDT1F7YWL4BGWQ5KQJ000G        5194   Web     LiveTV  Buying And Selling With Property...
118531  20210416  1JKBIA7MEUH6BWX0SCU93GJ000I       47660   Web     LiveTV           Espresso
118532  20210416  1ADX18R3B33REZ2TONK911E4002W       1335  Leanback     LiveTV           Unknown
118533  20210416  BO0REPHA6FP7C15B0K9N1GV00050        360   iOS     LiveTV           Imbewu

```

Oct 19, 2025 (ε1)

20

Python Generate (Ctrl + I)

```
### display the duplicate row in the data

survey_analysis.duplicated()

<bound method DataFrame.duplicated of      DateID ...          VideoTitle
0      20201101 ...      F1 '20: Emilia Romagna GP
1      20201101 ...      F1 '20: Emilia Romagna GP
2      20201101 ...      Chasing The Sun
3      20201101 ...      Chasing The Sun
4      20201101 ...      Sonic The Hedgehog
...
...
118529 20210416 ...      New Day
118530 20210416 ...  Buying And Selling With Property...
118531 20210416 ...      Espresso
118532 20210416 ...      Unknown
118533 20210416 ...      Imbewu

[118534 rows × 6 columns]
```

Oct 19, 2025 (ε1)

21

Python

```
### display the duplicate row in the data if its true/false

survey_analysis.duplicated()

0    False
1    True
2   False
3   True
4   False
...
118529  False
118530  False
118531  False
118532  False
118533  False
Length: 118534, dtype: bool
```

Oct 19, 2025 (ε1)

22

Python

```
### drop all the duplicate row

survey_analysis.drop_duplicates()

  DateID CustomerID TotalTimeWatched Platform PlayEventType          VideoTitle
0  20201101 EW1DENH0EC1J3M9WA0ZP9LSV0040        300  Leanback     LiveTV  F1 '20: Emilia Romagna GP
1  20201101 6TS2LLY0L3G66FVY86Q0JEZE000K        360  Leanback     Other      Chasing The Sun
2  20201101 6PMV67PLJ2S47S68J0Y30XFK003C        120  Leanback     LiveTV      Sonic The Hedgehog
3  20201101 6PMV67PLJ2S47S68J0Y30XFK003C        120  Leanback     LiveTV      Frozen II
4  20201101 6PMV67PLJ2S47S68J0Y30XFK003C        120     iOS     LiveTV  PL 20/21: Man Utd v Arsenal
...
...
118529 20210416 DGXM5TJ18W1LANOLKJUN5MFT000G        600  Leanback     LiveTV      New Day
118530 20210416 DMMVDBZFDT1F7WL4BGWQ5KOJ000G       5194     Web     LiveTV  Buying And Selling With Property...
118531 20210416 1JKBIAY7MEUH6BWX0SCU93GJ000I       47660    Web     LiveTV      Espresso
118532 20210416 1ADX18R3B33REZ2T0NK911E4002W       1335  Leanback     LiveTV      Unknown
118533 20210416 B00REPHAA6FP7C15B0K9N1GV00050       360     iOS     LiveTV      Imbewu

108661 rows × 6 columns
```

Oct 20, 2025 (ε1)

23

Python

```
### Display the null value
survey_analysis.isnull()

  DateID CustomerID TotalTimeWatched Platform PlayEventType          VideoTitle
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
...
...
118529  False    False        False    False    False    False
118530  False    False        False    False    False    False
118531  False    False        False    False    False    False
118532  False    False        False    False    False    False
118533  False    False        False    False    False    False

118534 rows × 6 columns
```