

```
✓ 23s [6] from google.colab import files
      uploaded = files.upload()
```

Choose Files archive.zip

- **archive.zip**(application/x-zip-compressed) - 491826 bytes, last modified: 4/24/2023 - 100% done

Saving archive.zip to archive (1).zip

READ DATASETS

```
✓ 1s 1 import pandas as pd
      pd.read_csv("archive.zip")
```



	id	Date	number of bedrooms	number of bathrooms	living area	lot area	number of floors	waterfront present	number of views	condition of the house	...	Built Year	Renovation Year	Postal Code	Latitude	Longitude	living_area_renov	lot_area_rer
0	6762810145	42491	5	2.50	3650	9050	2.0	0	4	5	...	1921	0	122003	52.8645	-114.557	2880	54
1	6762810635	42491	4	2.50	2920	4000	1.5	0	0	5	...	1909	0	122004	52.8878	-114.470	2470	40
2	6762810998	42491	5	2.75	2910	9480	1.5	0	0	3	...	1939	0	122004	52.8852	-114.468	2940	60
3	6762812605	42491	4	2.50	3310	42998	2.0	0	0	3	...	2001	0	122005	52.9532	-114.321	3350	42
4	6762812919	42491	3	2.00	2710	4500	1.5	0	0	4	...	1929	0	122006	52.9047	-114.485	2060	45
...

✓ 0s completed at 1:36 PM

UNIVARIATE ANALYSIS

```
[18] df=pd.DataFrame({'date':[42491,42491,42491,42491]})  
df.head()
```

	date
0	42491
1	42491
2	42491
3	42491

BI-VARIATE ANALYSIS

```
df=pd.DataFrame({'living area':[3650,2920,2910,3310,900], 'lot area':[9050,4000,6621,4770,9480]})
```

	living area	lot area
0	3650	9050
1	2920	4000
2	2910	6621
3	3310	4770
4	900	9480

MULTIVARIATE ANALYSIS

```
✓ [20] df=pd.DataFrame({'living area':[3650,2920,2910,3310,900], 'lot area':[9050,4000,6621,4770,9480], 'number of floors':[1.0,1.5,1.0,1.0,1.0]})  
      df.head()
```

	living area	lot area	number of floors
0	3650	9050	1.0
1	2920	4000	1.5
2	2910	6621	1.0
3	3310	4770	1.0
4	900	9480	1.0

DESCRIPTIVE STATISTICS[mean]

```
✓ [23] df=pd.DataFrame({'living area':[3650,2920,2910,3310,900], 'lot area':[9050,4000,6621,4770,9480]})  
      df.head()  
      df.mean()
```

```
living area    2738.0  
lot area       6784.2  
dtype: float64
```

MEDIAN

```
[27] df=pd.DataFrame({'living area':[3650,2920,2910,3310,900], 'lot area':[9050,4000,6621,4770,9480], 'number of floors':[1.0,1.5,1.0,1.0,1.0]})
df.head()
df.median(axis=0)
```

living area	2920.0
lot area	6621.0
number of floors	1.0
dtype:	float64

MODE

```
[31] df=pd.DataFrame({'date':[42491,42491,42491,42491]})
df.head()
df.mode()
```

MODE

```
✓ [31] df=pd.DataFrame({'date':[42491,42491,42491,42491]})  
02 df.head()  
df.mode()
```

	date
0	42491

VARIANCE

```
✓ [38] import numpy as np  
14 living_area=[3650,2920,2910,3310,900]  
  
print(np.var(living_area))  
  
919976.0
```

STANDARD DEVIATION

```
✓ [39] number_of_floors=[1.0,1.5,1.0,1.0,1.0]  
01 print(np.std(number_of_floors))  
  
0.19999999999999998
```

OneDrive



Screenshot saved
The screenshot was added to your
OneDrive.

MISSING VALUES

```
✓ [44] df=pd.DataFrame({'living area':[3650,2920,2910,3310,900],'lot area':[9050,4000,6621,4770,9480]})
da      df.head()
      print(df.shape)
      df.isna().sum()
      df["living area"].isnull()
```

```
(5, 2)
0    False
1    False
2    False
3    False
4    False
Name: living area, dtype: bool
```