

Day 1 - SQL - Measure of Central Tendency_by_Virat Tiwari

October 7, 2023

1 MEASURE OF CENTRAL TENDENCY -:

2 IN THIS TOPIC WE HAVE TO CALCULATE THE MEAN , MODE , MEDIAN IN PYTHON

3 A - Mean (Average)

```
[19]: age = [45,2,5,33,6,44,51,26,84,13,20,54,91,36,47,97,83,64,28]
```

Note - In python we have a NUMPY library for getting the MEAN or AVERAGE of Dataset , so we simply import the NUMPY library for finding the MEAN of any Dataset

```
[20]: import numpy as np
      np.mean(age)
```

```
[20]: 43.63157894736842
```

```
[21]: weights=[48,54,69,75,84,35,75]
```

```
[22]: np.mean(weights)
```

```
[22]: 62.857142857142854
```

Note - In python we have another library called as “SEABORN” that provide the built in datasets so we simply used that dataset by importing this library

```
[23]: import seaborn as sns
      df=sns.load_dataset('tips')
```

```
[24]: df.head()
```

```
[24]:
```

	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4

```
[25]: np.mean(df['total_bill'])
```

```
[25]: 19.78594262295082
```

```
[26]: np.mean(df["tip"])
```

```
[26]: 2.99827868852459
```

4 B - Median

```
[27]: np.median(age)
```

```
[27]: 44.0
```

```
[28]: np.median(df["total_bill"])
```

```
[28]: 17.795
```

```
[29]: np.median(df["tip"])
```

```
[29]: 2.9
```

```
[30]: # In this dataset last value is bigger than other values so that value is ^  
      ↪ outlier of dataset
```

```
age = [45, 2, 5, 33, 6, 44, 51, 26, 84, 13, 20, 54, 91, 36, 47, 97, 83, 64, 300]
```

```
[31]: np.median(age)
```

```
[31]: 45.0
```

Note - In case of OUTLIERS we have to use Median for better result, OUTLIERS are the one or two bigger value in a dataset comparatively other values

5 C - Mode

Note - In python for MODE we have to import a specific library called as SCIPY

```
[32]: from scipy import stats
```

```
[33]: stats.mode(age)
```

```
/tmp/ipykernel_70/2474845003.py:1: FutureWarning: Unlike other reduction  
functions (e.g. `skew`, `kurtosis`), the default behavior of `mode` typically  
preserves the axis it acts along. In SciPy 1.11.0, this behavior will change:  
the default value of `keepdims` will become False, the `axis` over which the  
statistic is taken will be eliminated, and the value None will no longer be
```

```
accepted. Set `keepdims` to True or False to avoid this warning.  
stats.mode(age)
```

```
[33]: ModeResult(mode=array([2]), count=array([1]))
```

THANK YOU SO MUCH !!

YOURS VIRAT TIWARI :)