Day 1 - SQL - Measure of Central Tendency_(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 1 - Mean (Average)

[4]:
```
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

[5]:
```
import numpy as np
np.mean(age)
```

[5]:
```
43.63157894736842
```

[6]:
```
weights=[48,54,69,75,84,35,75]
```

[7]:
```
np.mean(weights)
```

[7]:
```
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

[8]:
```
import seaborn as sns
df=sns.load_dataset('tips')
```

[9]:
```
df.head()
```

[9]:

	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

[10]:
```
np.mean(df['total_bill'])
```

[10]: 19.78594262295082

[11]:
```
np.mean(df["tip"])
```

[11]: 2.99827868852459

4 2 - Median

[12]:
```
np.median(age)
```

[12]: 44.0

[13]:
```
np.median(df["total_bill"])
```

[13]: 17.795

[14]:
```
np.median(df["tip"])
```

[14]: 2.9

[15]:
```
# In this dataset last value is bigger than other vakues 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]
```

[16]:
```
np.median(age)
```

[16]: 45.0

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

5 3 - Mode

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

[17]:
```
from scipy import stats
```

[18]:
```
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)
```

[18]:
```
ModeResult(mode=array([2]), count=array([1]))
```

THANK YOU SO MUCH !!

YOURS VIRAT TIWARI :)