```
import seaborn as sns
```

```
df=sns.load_dataset("tips")
df
```

	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
...	...	...	...	...	...	...	...
239	29.03	5.92	Male	No	Sat	Dinner	3
240	27.18	2.00	Female	Yes	Sat	Dinner	2
241	22.67	2.00	Male	Yes	Sat	Dinner	2
242	17.82	1.75	Male	No	Sat	Dinner	2
243	18.78	3.00	Female	No	Thur	Dinner	2

244 rows × 7 columns

Check information about data

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 244 entries, 0 to 243
Data columns (total 7 columns):
 #   Column      Non-Null Count  Dtype   
---  ------      --------------  -----   
 0   total_bill  244 non-null    float64 
 1   tip         244 non-null    float64 
 2   sex         244 non-null    category
 3   smoker      244 non-null    category
 4   day         244 non-null    category
 5   time        244 non-null    category
 6   size        244 non-null    int64   
dtypes: category(4), float64(2), int64(1)
memory usage: 7.4 KB
```

Checking first five entries

```
df.head()
```

	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

Checking last five entries

```
df.tail()
```

	total_bill	tip	sex	smoker	day	time	size
239	29.03	5.92	Male	No	Sat	Dinner	3
240	27.18	2.00	Female	Yes	Sat	Dinner	2
241	22.67	2.00	Male	Yes	Sat	Dinner	2
242	17.82	1.75	Male	No	Sat	Dinner	2
243	18.78	3.00	Female	No	Thur	Dinner	2

Summary statistics

```
df.describe()
```

	total_bill	tip	size
count	244.000000	244.000000	244.000000
mean	19.785943	2.998279	2.569672
std	8.902412	1.383638	0.951100
min	3.070000	1.000000	1.000000
25%	13.347500	2.000000	2.000000
50%	17.795000	2.900000	2.000000
75%	24.127500	3.562500	3.000000
max	50.810000	10.000000	6.000000

Checking number of rows and columns

```
df.shape
```

```
(244, 7)
```

```
df.shape[0]
```

```
244
```

```
df.shape[1]
```

```
7
```

```
name="the number of rows are", df.shape[0]
print(name)
```

```
('the number of rows are', 244)
```

```
name="the number of columns are", df.shape[1]
print(name)
```

```
('the number of columns are', 7)
```

Checking columns name

```
df.columns
```

```
Index(['total_bill', 'tip', 'sex', 'smoker', 'day', 'time', 'size'], dtype='object')
```

Checking row headings

```
df.index
```

```
RangeIndex(start=0, stop=244, step=1)
```

Removing specific columns

```
df1=df.drop(["day","size"], axis=1)
df1
```

	total_bill	tip	sex	smoker	time
0	16.99	1.01	Female	No	Dinner
1	10.34	1.66	Male	No	Dinner
2	21.01	3.50	Male	No	Dinner
3	23.68	3.31	Male	No	Dinner
4	24.59	3.61	Female	No	Dinner
...	...	...	...	...	...
239	29.03	5.92	Male	No	Dinner
240	27.18	2.00	Female	Yes	Dinner
241	22.67	2.00	Male	Yes	Dinner
242	17.82	1.75	Male	No	Dinner
243	18.78	3.00	Female	No	Dinner

244 rows × 5 columns

Checking missing values

```
df.isnull().sum()
```

```
total_bill    0
tip           0
sex           0
smoker        0
day           0
time          0
size          0
dtype: int64
```

Checking unique values

```
df.time.unique()
```

```
['Dinner', 'Lunch']
Categories (2, object): ['Lunch', 'Dinner']
```

Groupby

```
df.groupby(["tip"]).mean()
```

```
<ipython-input-28-30ab68981819>:1: FutureWarning: The default value of numeric_only in Dat
  df.groupby(["tip"]).mean()
```

```
       total_bill  size
  tip
 1.00    7.167500   1.5
 1.01   16.990000   2.0
 1.10   12.900000   2.0
 1.17   32.830000   2.0
 1.25    9.696667   2.0
  ...         ...   ...
 6.70   34.300000   6.0
 6.73   48.270000   4.0
 7.58   39.420000   4.0
 9.00   48.330000   4.0
10.00   50.810000   3.0

123 rows × 2 columns
```