Distribution Plots

For Contineous Plots that is for int or float data,there are three plots:

- Displot
- Joinplot
- Pairplot

```
# Let f1,f2,f3,f4 are four features of a dataset
# If we analyze f1 and f2 is called Bivarient Analysis  --> Joinplot
# For more than two features --> Pairplot
# For one feature --> Distplot
```

```
import seaborn as sns
import matplotlib.pyplot as plt
```

```
df = sns.load_dataset("tips")                  #load_dataset() is an inbuilt fn inside seaborn
```

```
# tips they give over the given bill and other related parameters
# Tip --> dependent Feature
# Total bill,sex,smoker,day,etc  --> Independent Features
print(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 x 7 columns]
```

```
df.head()                      # Returns first five rows
```

	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

```
df.dtypes                # data type of different columns
```

```
total_bill     float64
tip            float64
sex           category
smoker        category
day           category
time          category
size             int64
dtype: object
```

Correlation with Heatmap

A correlation heatmap uses colored cells, typically in a monochromatic scale, to show a 2d correlation matrix(table) between two discrete dimensions or event types. It is very important in Feature Selection.

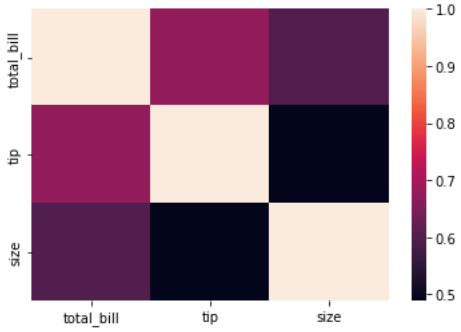
```
# Correlation --> Interrelationship b/w all the features (Features are either int or float)
df.corr()
```

```
# 100% corr, 67% corr, ... represent:
# When total bill increses and tip always increases and vice-varsa that is 100% correlation
```

	total_bill	tip	size

```
sns.heatmap(df.corr())
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7ffb08809d60>
```



Note:

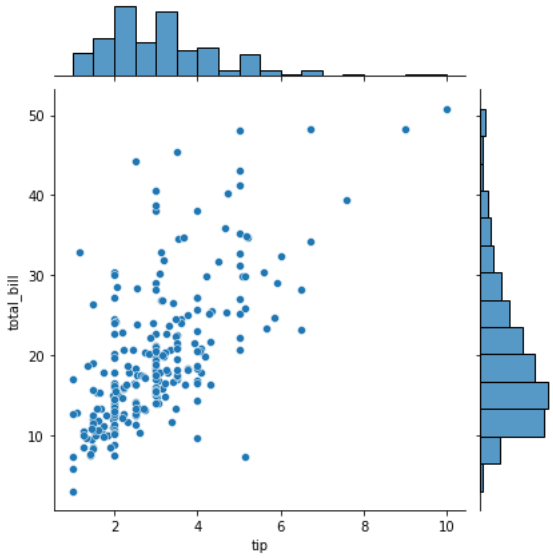
Heatmap is used in EDA when we do Feature selection, if there are two independent features and their corr is same or higher 0.9 or 1 , it is possible that we can use one feature instead of two.

JoinPlot

A join plot allows to study the relationship between 2 numeric variables. The central chart dsiplays that correlation. It is usually a scatterplot, a haxbin plot, a 2D histogram or a 2D density plot.

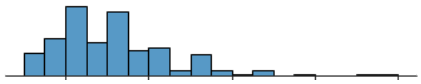
```
# Univariate Analysis or Bivarient Analysis --> have two Features
sns.jointplot(x='tip',y='total_bill',data=df)
```

```
<seaborn.axisgrid.JointGrid at 0x7f0146893790>
```



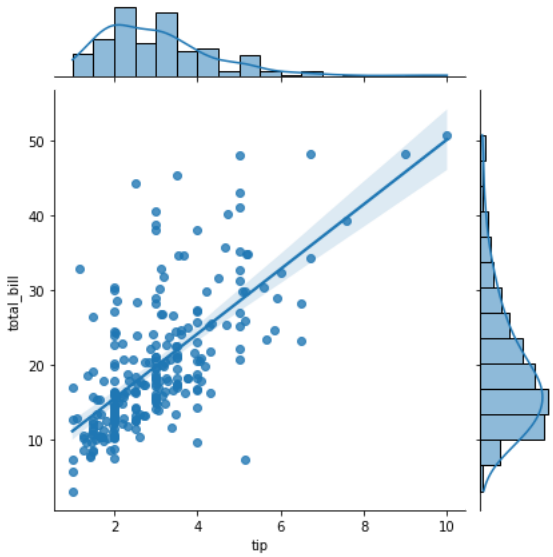

```
sns.jointplot(x='tip',y='total_bill',data=df,kind='hex')          #kind='hex' --> Gives hexagonal structure instead of point
```

```
<seaborn.axisgrid.JointGrid at 0x7f014672c5b0>
```



```
# Kind= 'reg' it will draw a regression line based on probabilty density function
sns.jointplot(x='tip',y='total_bill',data=df,kind='reg')
```

```
<seaborn.axisgrid.JointGrid at 0x7f01465b7be0>
```

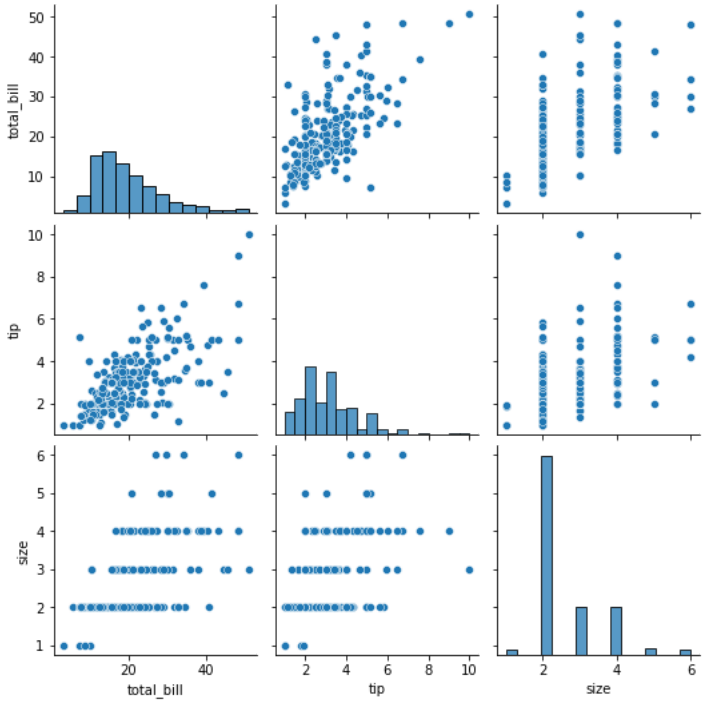


Pair Plot

A 'pair plot' is also known as a scatterplot, in which one variable in the same data row is matched with another variable's value,like this: Pairs plot are just elaborations on this,showing all variables paired with all other variables

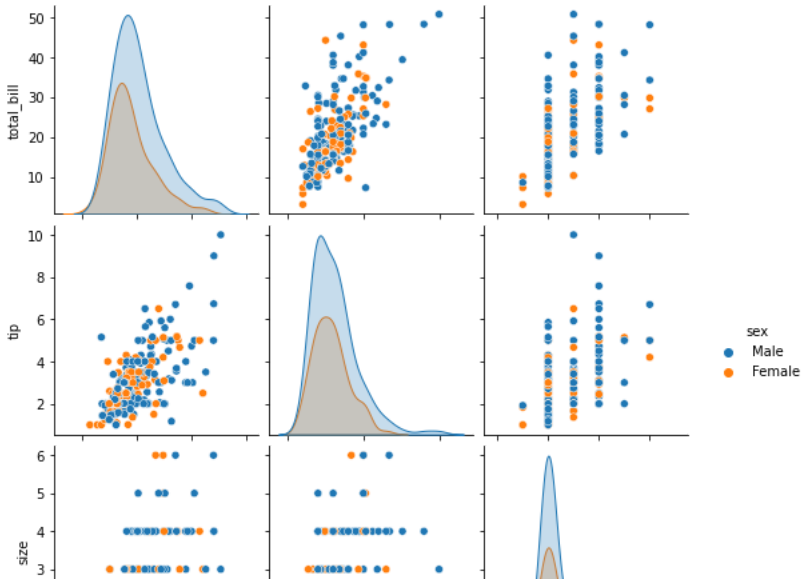
```
#For more than two independent Features
# For pair plot feature needs to have a int value or floating value
sns.pairplot(df)
```

```
<seaborn.axisgrid.PairGrid at 0x7f01464e0970>
```




```
# Scatterplot based on Categorical Value i.e,'Sex', we use the fn hue.
sns.pairplot(df,hue='sex')
```

```
<seaborn.axisgrid.PairGrid at 0x7f01460fa970>
```

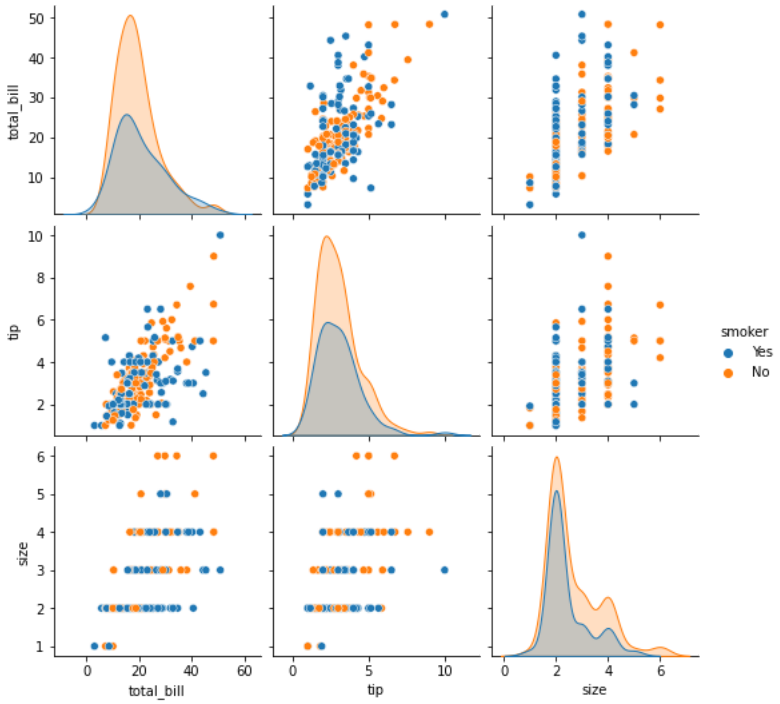



```
df['smoker'].value_counts()
```

```
No     151
Yes     93
Name: smoker, dtype: int64
```

```
sns.pairplot(df,hue='smoker')
```

```
<seaborn.axisgrid.PairGrid at 0x7f0145d7eb80>
```



Dist Plot

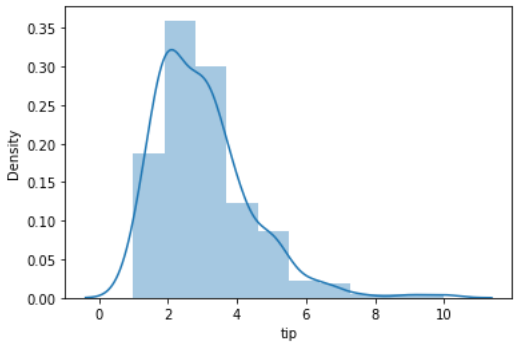
Dist plot helps us to check the distribution of the columns features

```
# For one feature
sns.distplot(df['tip'])
```

```
/usr/local/lib/python3.8/dist-packages/seaborn/distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be
  warnings.warn(msg, FutureWarning)
<matplotlib.axes._subplots.AxesSubplot at 0x7f0145a17040>
```

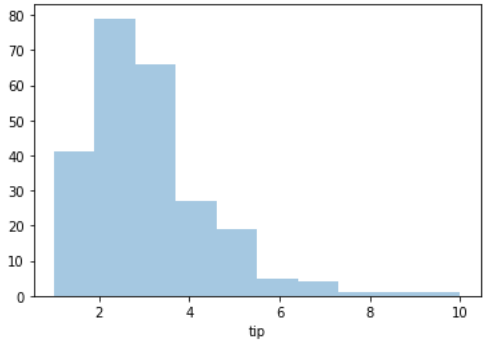
```
sns.distplot(df['tip'],kde=True,bins=10)          #Kde=true --> Density is given in terms of percentage
```

```
/usr/local/lib/python3.8/dist-packages/seaborn/distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be
  warnings.warn(msg, FutureWarning)
<matplotlib.axes._subplots.AxesSubplot at 0x7f0145a51dc0>
```




```
sns.distplot(df['tip'],kde=False,bins=10)
```

```
/usr/local/lib/python3.8/dist-packages/seaborn/distributions.py:2619: FutureWarning:
  warnings.warn(msg, FutureWarning)
<matplotlib.axes._subplots.AxesSubplot at 0x7f01458d65b0>
```



Categorical Plots

Seaborn also helps us in doing the analysis on categorical Data points.

- boxplot
- violinplot
- countplot
- bar plot

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

```
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

```
df['smoker'].value_counts()
```