```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from plotnine.scales import *
from plotnine import *
df=pd.read_csv("C:\Users\jyoti\Desktop\Datasets\Mall_Customers.csv")
```

```
df.head()
```

	CustomerID	Gender	Age	Annual Income (k$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40

Univariate Analysis

```
df.describe()
```

	CustomerID	Age	Annual Income (k$)	Spending Score (1-100)
count	200.000000	200.000000	200.000000	200.000000
mean	100.500000	38.850000	60.560000	50.200000
std	57.879185	13.969007	26.264721	25.823522
min	1.000000	18.000000	15.000000	1.000000
25%	50.750000	28.750000	41.500000	34.750000
50%	100.500000	36.000000	61.500000	50.000000
75%	150.250000	49.000000	78.000000	73.000000
max	200.000000	70.000000	137.000000	99.000000

```
sns.distplot(df['Annual Income (k$)'])
```

```
df.columns
```

```
Index(['CustomerID', 'Gender', 'Age', 'Annual Income (k$)',
       'Spending Score (1-100)'],
      dtype="object")
```

```
columns = ['Age', 'Annual Income (k$)', 'Spending Score (1-100)']
for i in columns:
    plt.figure()
    sns.distplot(df[df["Gender"]=='Male'][i],label="Male",hist=False,kde=True)
    sns.distplot(df[df["Gender"]=='Female'][i],label="Female",hist=False,kde=True)
    plt.legend()
```

```
sns.kdeplot(df['Annual Income (k$)'],shade=True,hue=df['Gender'])
```

```
columns = ['Age', 'Annual Income (k$)', 'Spending Score (1-100)']
for i in columns:
    plt.figure()
    sns.boxplot(data=df,x='Gender',y=i)
```

```
df.shape()
```

```
(200, 5)
```

```
df.describe()
```

	Age	Annual Income (k$)	Spending Score (1-100)
count	200.000000	200.000000	200.000000
mean	38.850000	60.560000	50.200000
std	13.969007	26.264721	25.823522
min	18.000000	15.000000	1.000000
25%	28.750000	41.500000	34.750000
50%	36.000000	61.500000	50.000000
75%	49.000000	78.000000	73.000000
max	70.000000	137.000000	99.000000

```
df.corr()
```

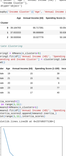
	Age	Annual Income (k$)	Spending Score (1-100)
Age	1.000000	-0.012398	-0.327227
Annual Income (k$)	-0.012398	1.000000	0.009903
Spending Score (1-100)	-0.327227	0.009903	1.000000

```
sns.heatmap(df.corr(),annot=True,cmap='rainbow')
```


Bivariate Analysis

```
sns.scatterplot(data=df, x='Annual Income (k$)', y='Spending Score (1-100)')
```

```
sns.pairplot(df[['Annual Income (k$)', 'Spending Score (1-100)']])
```



```
sns.pairplot(df,hue='Gender',diag_kind='kde')
```

```
df.groupby('Gender')[['Age', 'Annual Income (k$)', 'Spending Score (1-100)']].mean()
```

Gender	Age	Annual Income (k$)	Spending Score (1-100)
Female	38.098214	59.250000	51.526786
Male	39.806818	62.227273	48.511364

```
df.groupby('Gender')[['Age', 'Annual Income (k$)', 'Spending Score (1-100)']].median()
```

Gender	Age	Annual Income (k$)	Spending Score (1-100)
Female	35.0	60.0	50.0
Male	37.0	62.5	50.0

Clustering - Univariate, Bivariate, Multivariate

```
clustering1 = KMeans(n_clusters=3)
```

```
clustering1.fit(df[['Annual Income (k$)']])
```

```
KMeans(n_clusters=3)
```

```
clustering1.labels_
```

```
array([1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       
```