

```
In [20]: #TASK ONE :- EXPLORATORY DATA ANALYSIS(EDA)
import pandas as pd
import numpy as np
```

```
In [13]: #Load dataset
df=pd.read_csv(r'D:\CardioGoodFitness.csv')
```

```
In [14]: df
```

```
Out[14]:
```

	Product	Age	Gender	Education	MaritalStatus	Usage	Fitness	Income	Miles
0	TM195	18	Male	14	Single	3	4	29562	112
1	TM195	19	Male	15	Single	2	3	31836	75
2	TM195	19	Female	14	Partnered	4	3	30699	66
3	TM195	19	Male	12	Single	3	3	32973	85
4	TM195	20	Male	13	Partnered	4	2	35247	47
...	...	...	...	...	...	...	...	...	...
175	TM798	40	Male	21	Single	6	5	83416	200
176	TM798	42	Male	18	Single	5	4	89641	200
177	TM798	45	Male	16	Single	5	5	90886	160
178	TM798	47	Male	18	Partnered	4	5	104581	120
179	TM798	48	Male	18	Partnered	4	5	95508	180

180 rows × 9 columns

```
In [15]: data=np.genfromtxt("D:\CardioGoodFitness.csv",delimiter=",",skip_header=1)
```

```
In [16]: data
```

```
Out[16]: array([[ nan, 1.80000e+01, nan, ..., 4.00000e+00,
        2.95620e+04, 1.12000e+02],
       [ nan, 1.90000e+01, nan, ..., 3.00000e+00,
        3.18360e+04, 7.50000e+01],
       [ nan, 1.90000e+01, nan, ..., 3.00000e+00,
        3.06990e+04, 6.60000e+01],
       ...,
       [ nan, 4.50000e+01, nan, ..., 5.00000e+00,
        9.08860e+04, 1.60000e+02],
       [ nan, 4.70000e+01, nan, ..., 5.00000e+00,
        1.04581e+05, 1.20000e+02],
       [ nan, 4.80000e+01, nan, ..., 5.00000e+00,
        9.55080e+04, 1.80000e+02]])
```

```
In [21]: # data Characteristics
df.head()
```

Out[21]:

	Product	Age	Gender	Education	MaritalStatus	Usage	Fitness	Income	Miles
0	TM195	18	Male	14	Single	3	4	29562	112
1	TM195	19	Male	15	Single	2	3	31836	75
2	TM195	19	Female	14	Partnered	4	3	30699	66
3	TM195	19	Male	12	Single	3	3	32973	85
4	TM195	20	Male	13	Partnered	4	2	35247	47

```
In [22]: df.tail()
```

Out[22]:

	Product	Age	Gender	Education	MaritalStatus	Usage	Fitness	Income	Miles
175	TM798	40	Male	21	Single	6	5	83416	200
176	TM798	42	Male	18	Single	5	4	89641	200
177	TM798	45	Male	16	Single	5	5	90886	160
178	TM798	47	Male	18	Partnered	4	5	104581	120
179	TM798	48	Male	18	Partnered	4	5	95508	180

```
In [23]: df.describe()
```

Out[23]:

	Age	Education	Usage	Fitness	Income	Miles
count	180.000000	180.000000	180.000000	180.000000	180.000000	180.000000
mean	28.788889	15.572222	3.455556	3.311111	53719.577778	103.194444
std	6.943498	1.617055	1.084797	0.958869	16506.684226	51.863605
min	18.000000	12.000000	2.000000	1.000000	29562.000000	21.000000
25%	24.000000	14.000000	3.000000	3.000000	44058.750000	66.000000
50%	26.000000	16.000000	3.000000	3.000000	50596.500000	94.000000
75%	33.000000	16.000000	4.000000	4.000000	58668.000000	114.750000
max	50.000000	21.000000	7.000000	5.000000	104581.000000	360.000000

```
In [24]: df.isnull().sum()
```

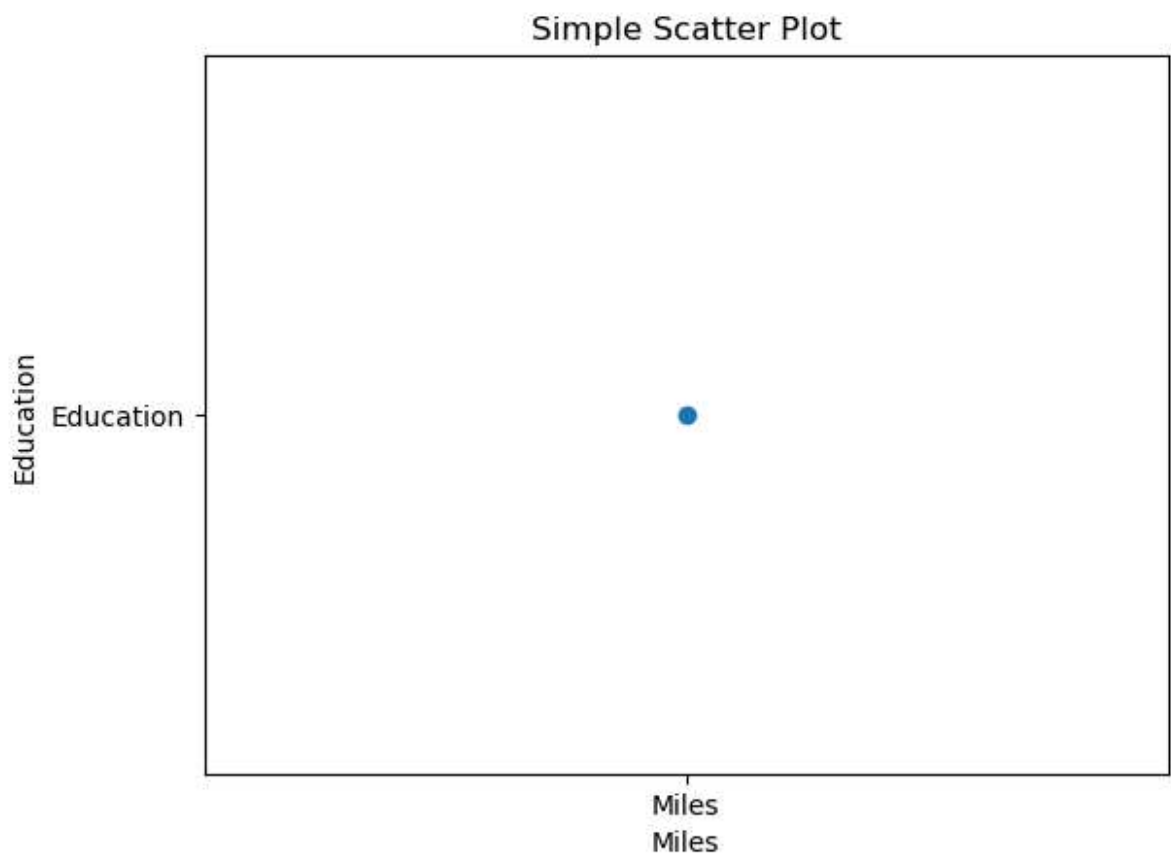
```
Out[24]: Product      0
Age      0
Gender    0
Education 0
MaritalStatus 0
Usage     0
Fitness   0
Income    0
Miles     0
dtype: int64
```

```
In [28]: df.info()
```

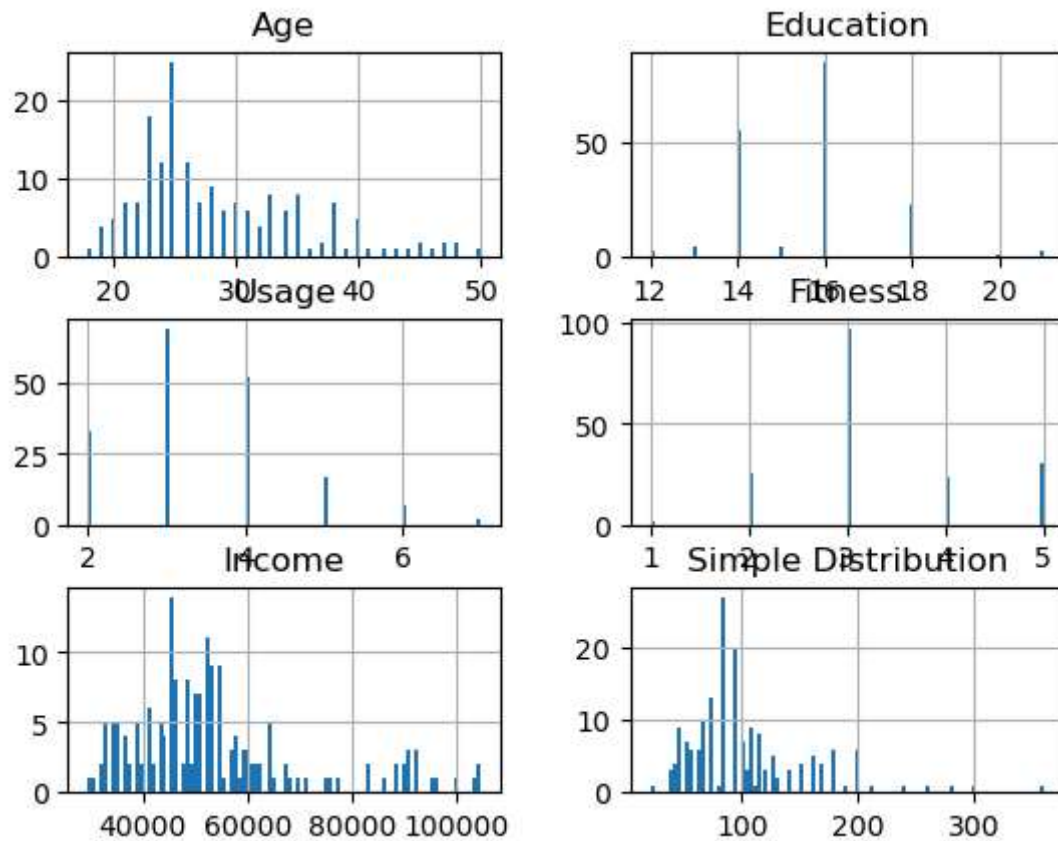
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 180 entries, 0 to 179
Data columns (total 9 columns):
 #   Column          Non-Null Count  Dtype
---  -
 0   Product         180 non-null   object
 1   Age             180 non-null   int64
 2   Gender          180 non-null   object
 3   Education       180 non-null   int64
 4   MaritalStatus   180 non-null   object
 5   Usage          180 non-null   int64
 6   Fitness         180 non-null   int64
 7   Income          180 non-null   int64
 8   Miles           180 non-null   int64
dtypes: int64(6), object(3)
memory usage: 12.8+ KB
```

```
In [25]: #Data Visualization
import matplotlib.pyplot as plt
import seaborn as sns
```

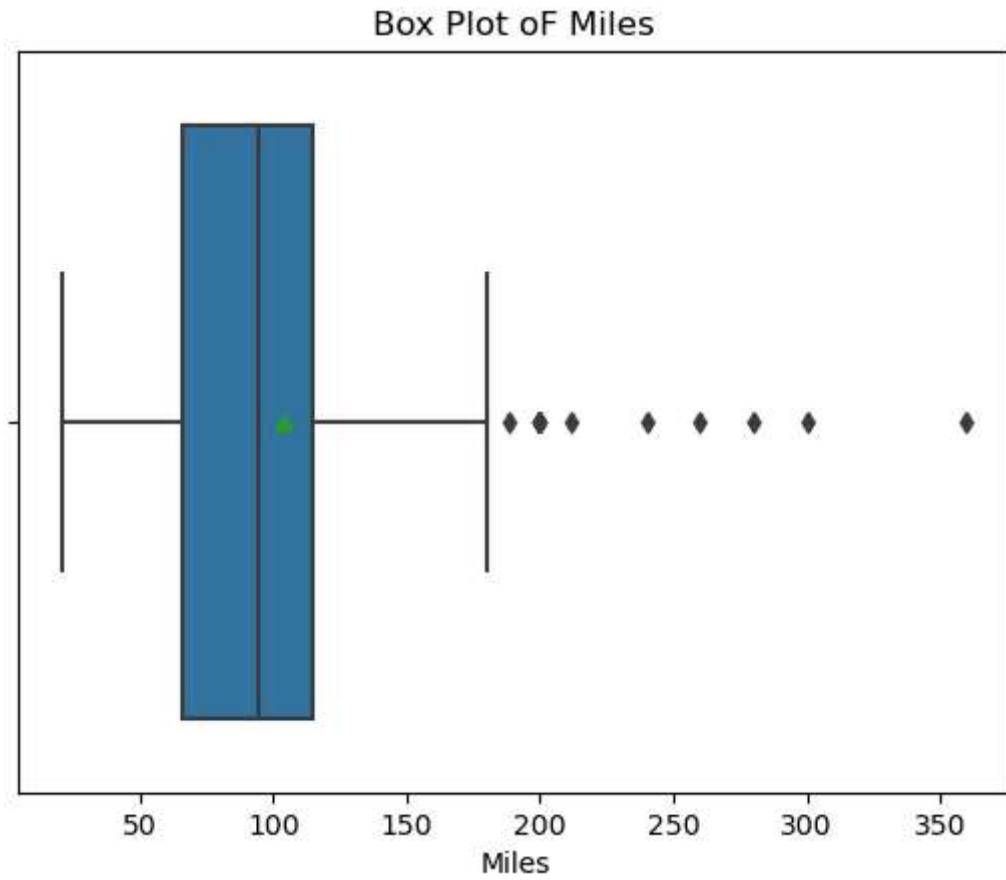
```
In [32]: plt.scatter('Miles', 'Education')  
plt.xlabel('Miles')  
plt.ylabel('Education')  
plt.title('Simple Scatter Plot')  
plt.show()
```



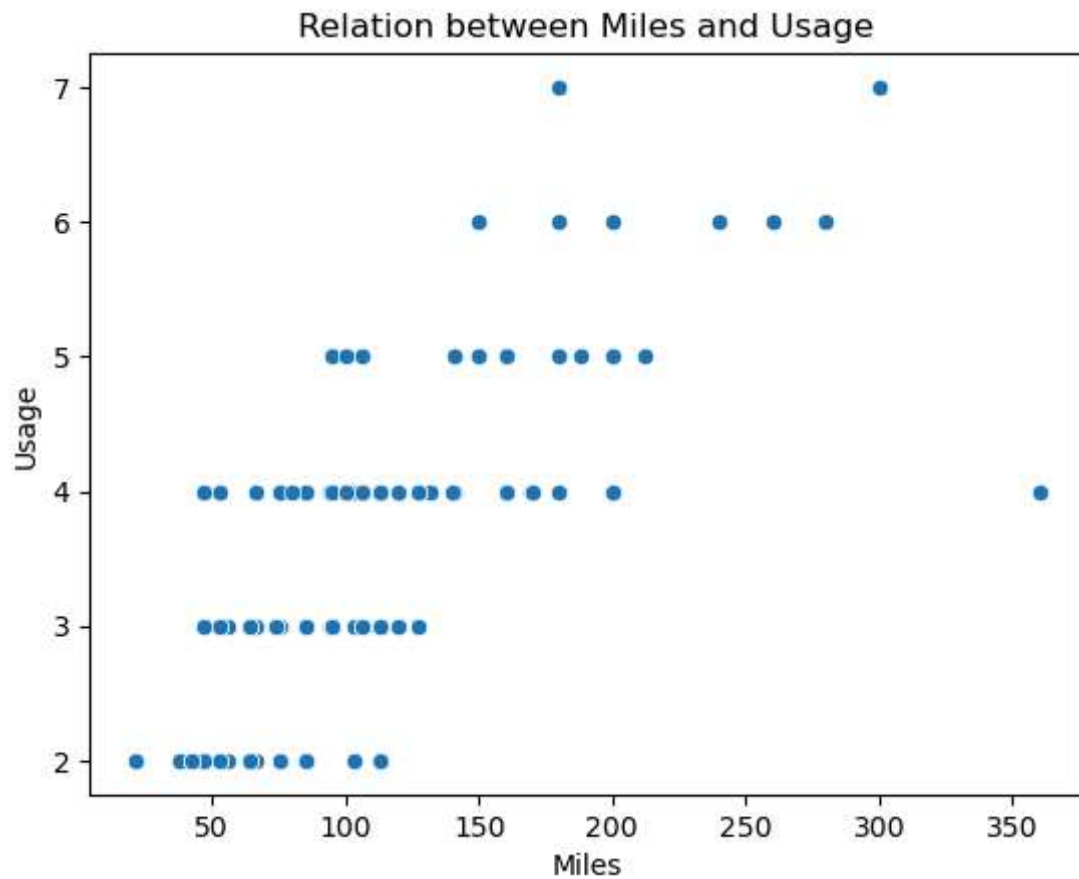
```
In [56]: #Data Distribution
df.hist(bins=100)
plt.title('Simple Distribution')
plt.show()
```



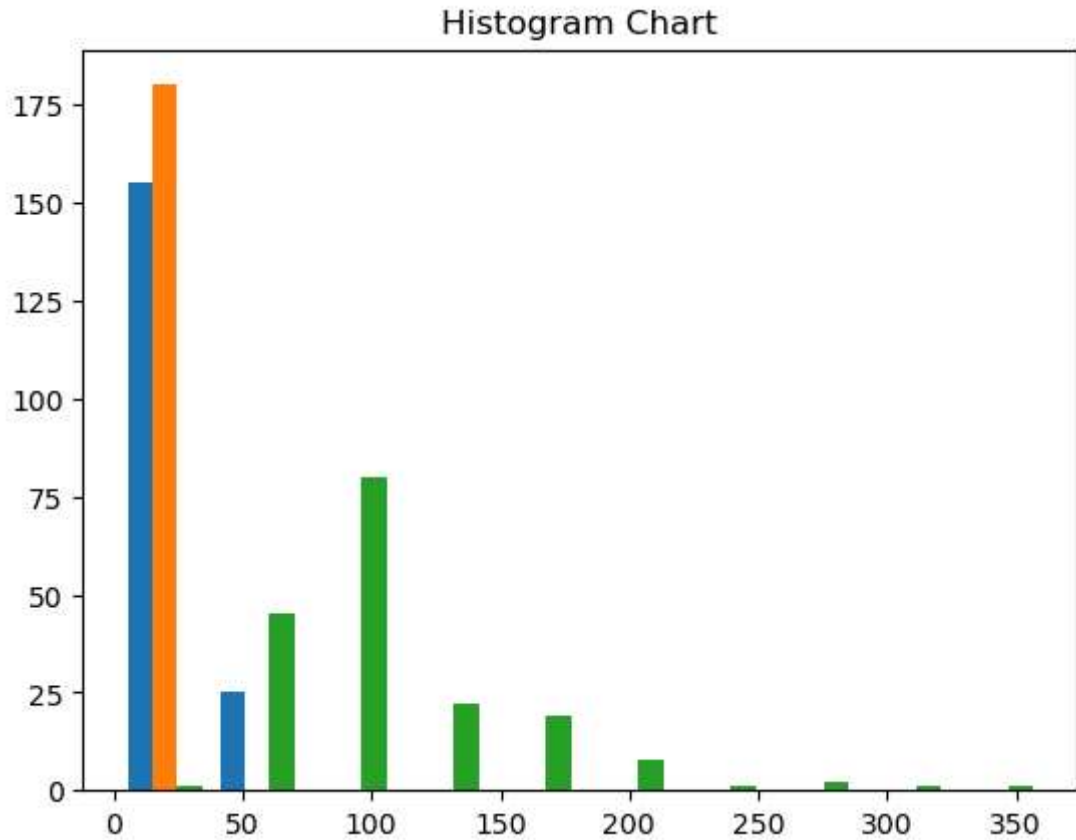
```
In [34]: # Find Outliers  
#using Box Plot  
sns.boxplot(data=df,x='Miles',showmeans=True)  
plt.title('Box Plot of Miles')  
plt.show()
```



```
In [36]: #finding Outliers
sns.scatterplot(x='Miles',y='Usage',data=df)
plt.xlabel('Miles')
plt.ylabel('Usage')
plt.title('Relation between Miles and Usage')
plt.show()
```

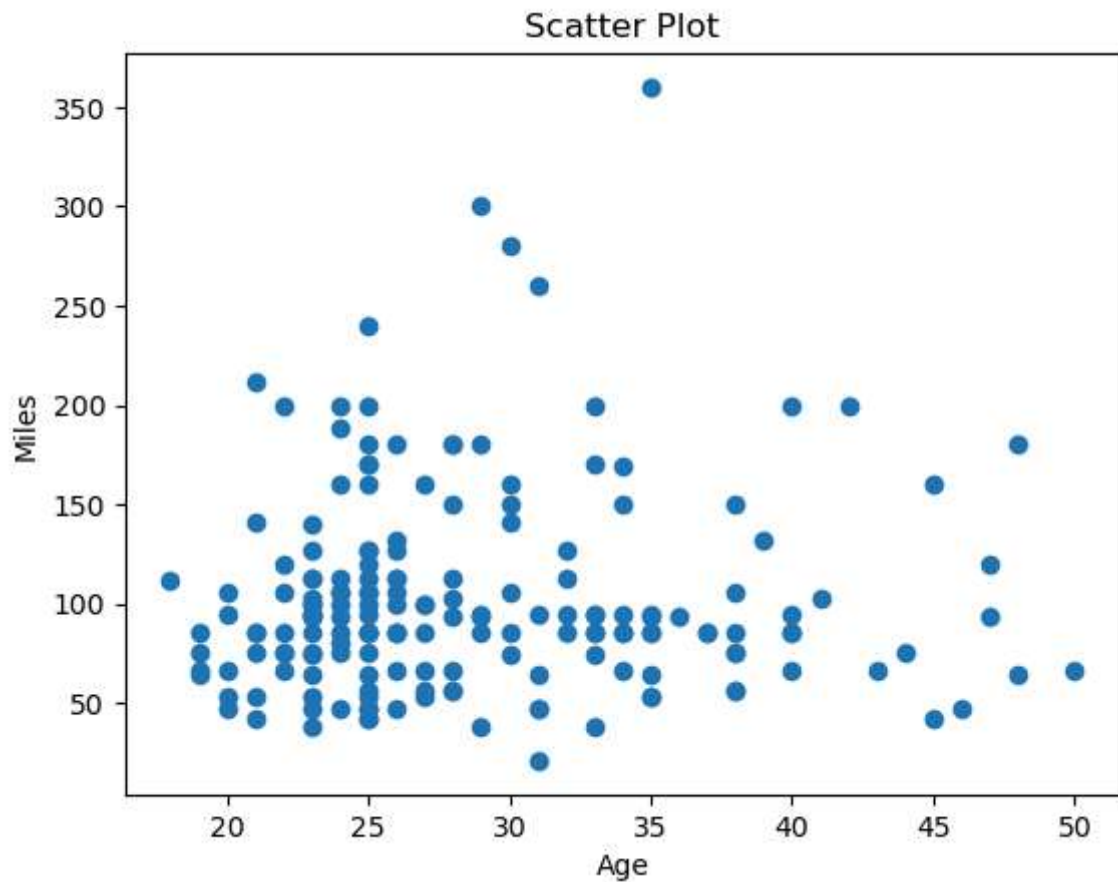


```
In [64]: # Data Visualize  
# Histogram , Scatter , Heatmap  
  
# Histogram  
plt.hist(x=(df['Age'],df['Usage'],df['Miles']),bins=10)  
plt.title('Histogram Chart')  
plt.show()
```





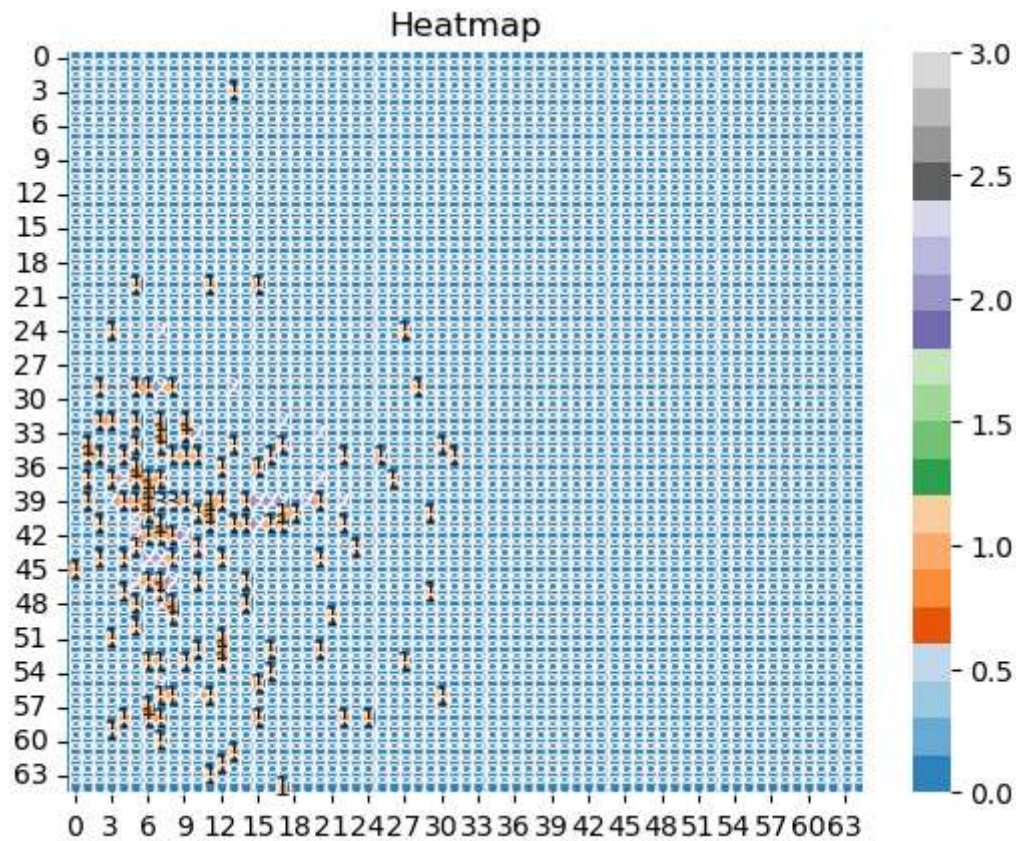
```
In [68]: # Scatter Plot
plt.scatter(x=df['Age'],y=df['Miles'])
plt.xlabel('Age')
plt.ylabel('Miles')
plt.title('Scatter Plot')
plt.show()
```



```
In [50]: # HeatMap
from sklearn.metrics import confusion_matrix as cn
```

```
In [70]: con=cn(df['Miles'],df['Age'])
```

```
In [73]: sns.heatmap(con,annot=True,cmap='tab20c')  
plt.title('Heatmap')  
plt.show()
```



```
In [ ]:
```