```
from matplotlib import pyplot as plt
import pandas as pd
import numpy as np

df=pd.read_csv('C:/Users/LOCHAN MANI GAVEL/Desktop/ap(1).csv')

df
```

```
    Serial No.  GRE Score  TOEFL Score  University Rating  SOP  LOR
CGPA  \
0            1        337          118                  4  4.5   4.5
9.65
1            2        324          107                  4  4.0   4.5
8.87
2            3        316          104                  3  3.0   3.5
8.00
3            4        322          110                  3  3.5   2.5
8.67
4            5        314          103                  2  2.0   3.0
8.21
..         ...        ...          ...                ...  ...   ...
...
95          96        304          100                  4  1.5   2.5
7.84
96          97        306          100                  2  3.0   3.0
8.00
97          98        331          120                  3  4.0   4.0
8.96
98          99        332          119                  4  5.0   4.5
9.24
99         100        323          113                  3  4.0   4.0
8.88

    Research  Chance of Admit
0          1              0.92
1          1              0.76
2          1              0.72
3          1              0.80
4          0              0.65
..       ...               ...
95         0              0.42
96         0              0.48
97         1              0.86
98         1              0.90
99         1              0.79

[100 rows x 9 columns]
```

```
df=df.loc[::,['Serial No.','GRE Score','TOEFL Score','CGPA']]
df
```

```
    Serial No.  GRE Score  TOEFL Score  CGPA
0            1        337          118  9.65
1            2        324          107  8.87
2            3        316          104  8.00
3            4        322          110  8.67
4            5        314          103  8.21
..         ...        ...          ...   ...
95          96        304          100  7.84
96          97        306          100  8.00
97          98        331          120  8.96
98          99        332          119  9.24
99         100        323          113  8.88

[100 rows x 4 columns]
```

```
x=df.iloc[:,0]
```

```
x.shape
```

```
(100,)
```

```
x=df.iloc[:,0].values.reshape(-1,1)
x.shape
```

```
(100, 1)
```

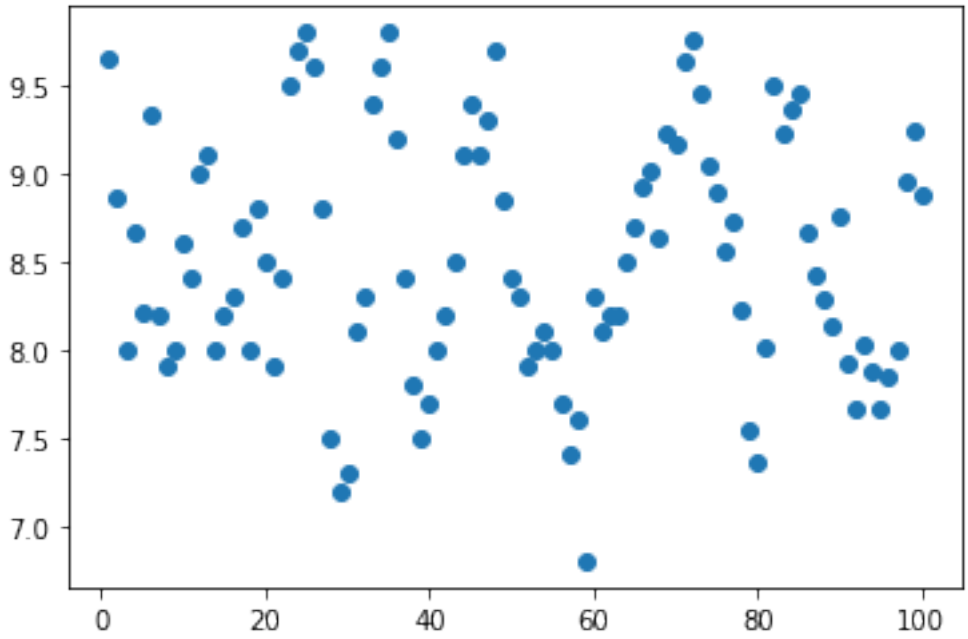
```
y=df.iloc[:,-1].values.reshape(-1,1)
y.shape
```

```
(100, 1)
```

```
import matplotlib.pyplot as plt
%matplotlib inline
```

```
plt.scatter(x,y)
plt.show
```

```
<function matplotlib.pyplot.show(close=None, block=None)>
```




```
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split

x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,rando
m_state=0)

lr=LinearRegression()
lr.fit(x_train,y_train)

LinearRegression()

y_pred=lr.predict(x_test)
y_pred

array([[8.55959806],
       [8.5609052 ],
       [8.5590752 ],
       [8.56022984],
       [8.56066556],
       [8.5610577 ],
       [8.5593802 ],
       [8.56062199],
       [8.56020806],
       [8.56110127],
       [8.56018627],
       [8.56103592],
       [8.56073092],
       [8.55931484],
       [8.55918413],
       [8.5596852 ],
       [8.55951092],
```

```
       [8.55955449],
       [8.55975056],
       [8.55920592]])
```

```
plt.scatter(x,y,color='blue')
plt.plot(x_test,y_pred,color='cyan')
```

```
[<matplotlib.lines.Line2D at 0x1f03d4f8fd0>]
```