

```
In [1]: import pandas as pd
import numpy as np
from matplotlib import pyplot as plt
import statsmodels.api as sm
import seaborn as sns
sns.set()
```

```
In [2]: data=pd.read_csv('C:\\Users\\UP\\Downloads\\data science and ml\\Part_5_Advanced_Statis
```

```
In [3]: data
```

```
Out[3]:
```

	SAT	GPA
--	-----	-----

0	1714	2.40
1	1664	2.52
2	1760	2.54
3	1685	2.74
4	1693	2.83
...	...	...
79	1936	3.71
80	1810	3.71
81	1987	3.73
82	1962	3.76
83	2050	3.81

84 rows × 2 columns

```
In [4]: data.describe()
```

```
Out[4]:
```

	SAT	GPA
count	84.000000	84.000000
mean	1845.273810	3.330238
std	104.530661	0.271617
min	1634.000000	2.400000
25%	1772.000000	3.190000
50%	1846.000000	3.380000
75%	1934.000000	3.502500
max	2050.000000	3.810000

```
In [5]: y = data['GPA']
        x1 = data['SAT']
```

```
In [6]: plt.scatter(x1,y)
        plt.xlabel('SAT',fontsize=20)
        plt.ylabel('GPA',fontsize=20)
        plt.show()
```

```
-----
AttributeError                                Traceback (most recent call last)
<ipython-input-6-0a278f1dbdca> in <module>
      1 plt.scatter(x1,y)
----> 2 plt.xlabel('SAT',fontsize=20)
      3 plt.ylabel('GPA',fontsize=20)
      4 plt.show()
```

**AttributeError:** module 'matplotlib.pyplot' has no attribute 'xlabel'



```
In [ ]: x = sm.add_constant(x1)
        results = sm.OLS(y,x).fit()
        results.summary()
```

```
In [ ]: y = data['GPA']
        x1 = data['SAT']
```

```
In [8]: reg = LinearRegression()
```

```
-----
NameError                                Traceback (most recent call last)
<ipython-input-8-15df1378c066> in <module>
----> 1 rer = LinearRegression()
```

**NameError:** name 'LinearRegression' is not defined

```
In [1]: reg.fit(x,y)
```

```
-----
NameError                                Traceback (most recent call last)
<ipython-input-1-8fcf381088e5> in <module>
```

```
----> 1 reg.fit(x,y)
```

```
NameError: name 'reg' is not defined
```

In [ ]: