In [24]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
sns.set()
from sklearn.linear_model import LinearRegression
```

In [25]:
```
data = pd.read_csv('1.01. Simple linear regression.csv')
data.head() ##visaulize the 5 first data
```

Out[25]:

| | SAT | GPA |
|---|---|---|
| **0** | 1714 | 2.40 |
| **1** | 1664 | 2.52 |
| **2** | 1760 | 2.54 |
| **3** | 1685 | 2.74 |
| **4** | 1693 | 2.83 |

In [3]:
```
## Regression
```

In [26]:
```
x = data['SAT'] ## feature variable - input
y = data['GPA'] ## target variable - output
```

In [27]:
```
x.shape
```

Out[27]: (84,)

In [28]:
```
y.shape
```

Out[28]: (84,)

In [29]:
```
##convert into 2d array
x_matrix = x.values.reshape(-1,1)
x_matrix.shape
```

Out[29]: (84, 1)

In [32]:
```
reg = LinearRegression()
```

In [35]:
```
reg.fit(x_matrix,y)
```

Out[35]: LinearRegression()

In [38]:
```
#R-squared
reg.score(x_matrix,y)
```

Out[38]:
```
0.40600391479679765
```

In [40]:
```
reg.coef_
```

Out[40]:
```
array([0.00165569])
```

In [42]:
```
reg.intercept_
```

Out[42]:
```
0.2750402996602803
```

In [45]:
```
#making predictions
reg.predict([[1740]])
```

Out[45]:
```
array([3.15593751])
```

In [48]:
```
new_data = pd.DataFrame(data=[1740,1760],columns=['SAT'])
new_data
```

Out[48]:

| | SAT |
|---|---|
| **0** | 1740 |
| **1** | 1760 |

In [49]:
```
reg.predict(new_data)
```

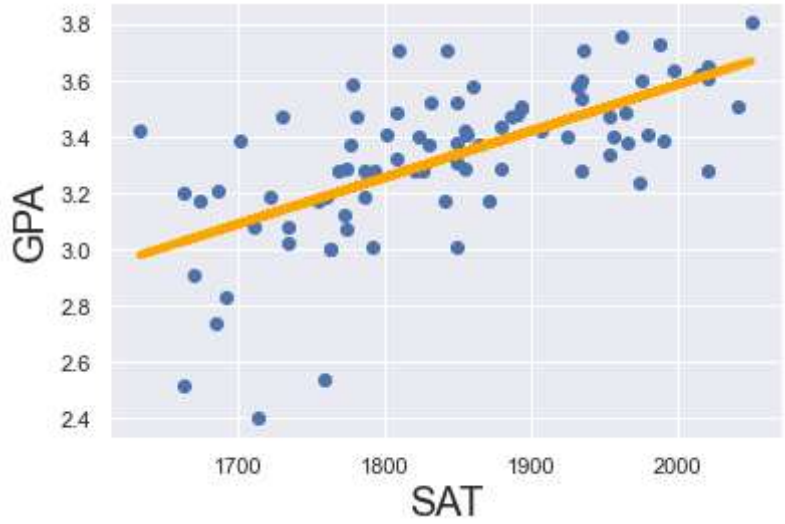
Out[49]:
```
array([3.15593751, 3.18905127])
```

In [51]:
```
new_data['Predicted_GPA'] = reg.predict(new_data)
new_data
```

Out[51]:

| | SAT | Predicted_GPA |
|---|---|---|
| **0** | 1740 | 3.155938 |
| **1** | 1760 | 3.189051 |

In [58]:
```
plt.scatter(x,y)
yhat = reg.coef_*x_matrix + reg.intercept_
fig = plt.plot(x,yhat, lw=4, c='orange', label='regression line')
plt.xlabel('SAT', fontsize=20)
plt.ylabel('GPA', fontsize=20)
plt.show()
```




In [ ]:
```
```