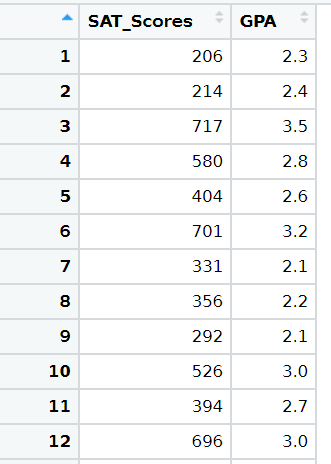
**Problem Statement: -**

A certain university wants to understand the relationship between students’ SAT scores and their GPA. Build a Simple Linear Regression model with GPA as the target variable and record the RMSE and correlation coefficient values for different models.



* **Business Objective:**

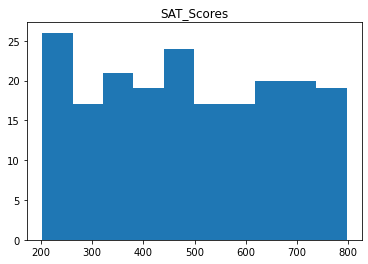
For today’s high school students, the good news is there are more A’s being given out than ever before on their report cards. The bad news is that SAT scores are continuing to fall. There’s a serious disconnect there that’s negatively affecting students without them realizing it.

**A recent study has shed light on the disturbing trend of high school grade inflation and its negative effect on SAT scores and college admissions. While high school GPAs may technically be higher than ever, their accuracy when measuring how much students are learning has come under question.**

* **Data Preprocessing:**

We have the dataset of university which have the GPA value and SAT\_Score of students. Using this data, we want to predict GPA value from the SAT\_Score of student.

So before doing prediction we need to preprocess the data. We check for null values or missing values in the data, but we got null values as zero.

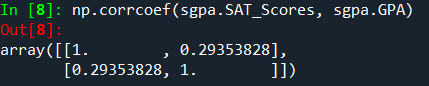
Chart, histogram

Description automatically generated

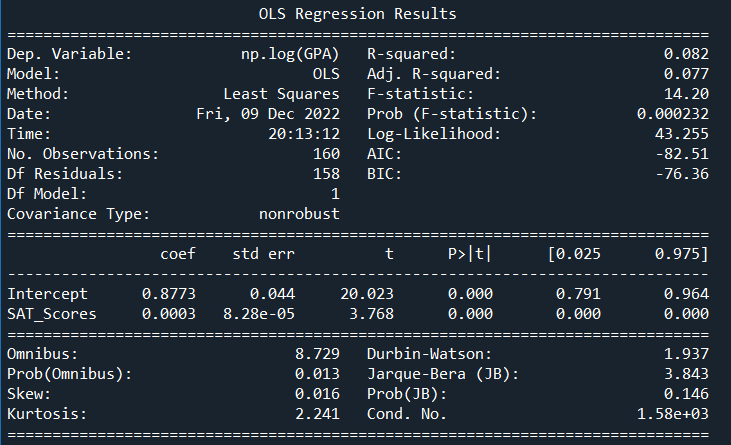
This are the Plots of the SAT\_Score and GPA which gives the insights of the data. The data is distributed here there is no skewness in the data.

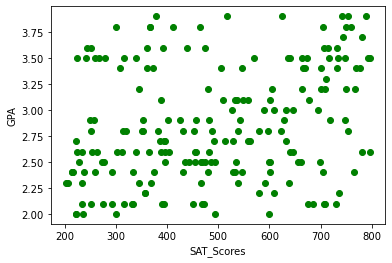
Now we find the correlation between the SAT\_Score and the GPA value and plotted the scatter plot.

Here we found that there is weak correlation between the GPA and SAT\_Score i.e. GPA is less dependent on SAT\_Score.

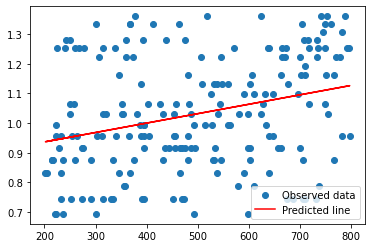


As the data was distributed we applied different transformations on the data such as log transformation, exponential transformation, square root transformation and polynomial transformation and calculated the errors of each model. After calculating errors we got the best fit model after exponential transformation as it gives the less error and probability of confidence is also less than 0.05. So we consider this model as best fit for our prediction model.





We have plotted the graph and regression line which gives graphical visualization of the prediction of our best fit final model.



Now we train and test our model. For training and testing we split our data into training and testing randomly. We use 80% data for training and 20% data for testing. And calculate error to check the accuracy of the model. The errors are very less so we can say that we have built the GPA prediction using SAT\_Score model with more accuracy.





**Conclusion:**

The SAT is a multiple-choice, pencil-and-paper test created and administered by the College Board.

The purpose of the SAT is to measure a high school student's readiness for college and provide colleges with one common data point that can be used to compare all applicants.

Your average GPA is a number that shows what you typically scored in your classes throughout the semester, term, and year.

As the SAT Score increases the GPA also increases. So, if a student is having a low SAT score then we can tell him to improve his/her skills to get a good GPA. We can analyze the student performance and will help them to improve their skills and get good scores.

But there is poor relation between the SAT score and GPA and we cannot conclude that if stude