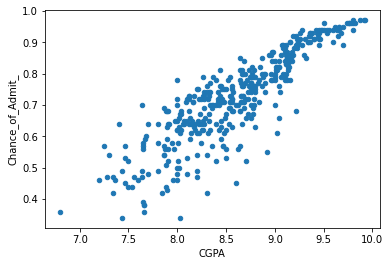
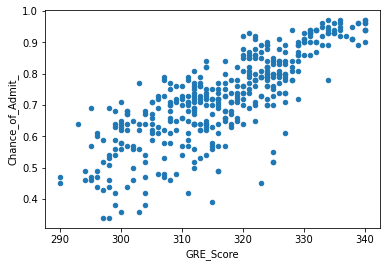
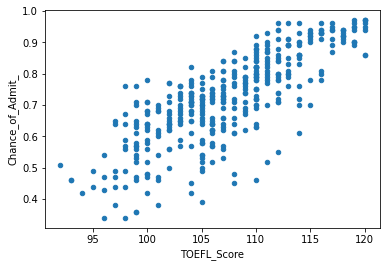
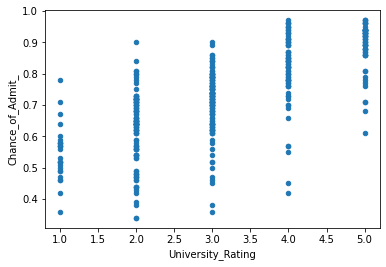
**Regression Task:**

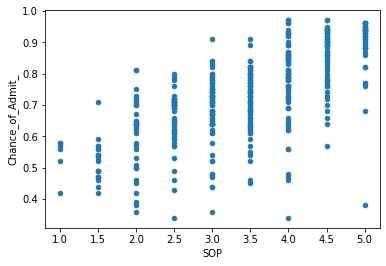
1. **Part 1(Predicting Chances of Admission from the CGPA acquired):**
2. **Coming up with a model:** We had to decide as to what model could fit the data best, so we plotted the *Chance of getting Admitted* vs *CGPA* acquired. We observed that chances of getting admission increase with increase in CGPA, and that it was possible to approximate the distribution as a linear distribution.

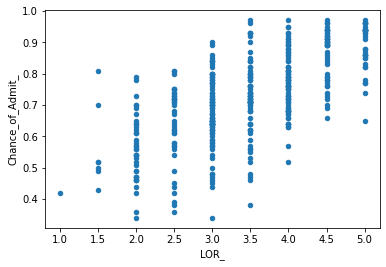














We also observe that plotting Chance of Admission v/s any other parameter does not yield much important insights, except for the fact that there is a general increase in chance of admission with the test scores. So we have an intuition that CGPA may be the most significant factor in getting admitted.

1. **Building the model:** We came up with a multilinear regression model to predict the chance of admission given other parameters.

(Insert abhi’s code here. He sent the multiple regression colab file on the group)

We observe that CGPA is indeed the most important factor in predicting chance of Admission

1. **Adding additional data from the other team:** We received 20 entries of the data from the other team. We added the data to our initial train data and tried to fit the initial multilinear model on the total data i.e. 420 entries.

**Observations-** We recieved additional 20 entries from the other team. We received an additional 20 entries from team Los Angeles. We included the entries in our train data and tried to fit another similar linear regression model on the net data.

Observations: Training on the original data led to 2 columns being insignificant, whereas training on the net data we found only one of the columns insignificant.

As we can see from the results the net data had a decreased error.

**Conclusions-** The additional data (la.csv) was a better fit and had a decreased spread. It’s distribution was represented by the model much better.

1. **2. Part 2(Predicting the TOEFL Score given the GRE Score):**
2. **Coming up with a model:** We had to decide as to what model could fit the data best so we plotted the *TOEFL Score* against the *GRE Score* . We observed that it was a fairly linear distribution, and the TOEFL Score generally increased with the increase in GRE Score.

[Plotting the graph](https://colab.research.google.com/drive/1XbpC2OopPIDwPHkuWOYn7qUK3EVvaPy2#scrollTo=_k415D3qRGEc&line=1&uniqifier=1)



1. **Building the model:** We came up with a simple linear regression to predict the TOEFL Score of a person given the GRE Score.

[Building model](https://colab.research.google.com/drive/1XbpC2OopPIDwPHkuWOYn7qUK3EVvaPy2#scrollTo=PwHVlbCEVLIk&line=1&uniqifier=1)

[Building model Continues](https://colab.research.google.com/drive/1XbpC2OopPIDwPHkuWOYn7qUK3EVvaPy2#scrollTo=S5jOSIjYV0Xq&line=4&uniqifier=1)

[And Continues](https://colab.research.google.com/drive/1XbpC2OopPIDwPHkuWOYn7qUK3EVvaPy2#scrollTo=47i24jHmWtNX&line=2&uniqifier=1)(Okay include all the cells in the heading train-test split, till before polynomial features)

1. Adding additional entries: We received an additional 20 entries from team Los Angeles. We included the entries in our train data and tried to fit another similar linear regression model on the net data.

Observations: (From the above given notebook link paste the complete code of Part 2 observing differences one, given at end)

i) The slope of line increased slightly

ii) The intercept of the model decreased slightly

iii) The mean squared error decreases

We now test the data on only the new data (la.csv)

i) The score of the model increases insanely and nearly touches 100%

ii) The mean squared error decreases to a negligible value

Conclusions: The data received is very very good and fits the model very well

Bugs Faced:

1. It was difficult to work with Columns having whitespaces in their names, and we recieved numerous errors: That’s when we decided to eliminate all the white spaces from column names.
2. For Task 2 we tried to normalize the