

**Part C: Summary of the coding question**

- So, I begun with importing the data from its parent repo because I found the data is not an exact CSV file
- From the repo, I downloaded the data and inserted into the feature matrix and target matrix
- Then I checked whether or not the data is empty or corrupted, by printing some of them
- Checked them by using `X.info()`, which outputs the Non-Null Count and datatype of the features
- I plotted the histograms between the frequency and the value of the feature for better visualization.
- Now, from skleram model, I used the `train_test_split` which randomly takes out 20 percent of this data, which can be used for testing
- Then I imported the accuracy and `r2` score from `sklear` metrics to compare know how well the model is performing
- Now from `statsmodel.api` , I imported the ordinal regression and used to train the model
- Then, I used this model on the test data and stored all the values into a matrix `y_pred`
- I also used `predicted_bins` to round of the predicted data
- The same is done for linear regression also

**Appropriate metric for comparison:**

I chose the  $R^2$  score metric, also known as the coefficient of determination. It is a metric used to evaluate the performance of a regression model in machine learning. It provides a measure of how well the predictions of the model match the actual observed values.

**Conclusion:**

I have imported this in my code file and at the end, I printed the value for both ordinal and linear regression.

We can see that ordinal is slightly better than the linear regression for this dataset as it's  $R^2$  value says that it is a better fit to the data.