# Topic(s): Simple Linear Regression

**Instructions**

Please share your answers filled inline in the word document. Submit Python code and R code files wh

A food delivery service recorded the data of delivery time taken and the time taken for the deliveries to be sorted by the restaurants in order to improve their delivery services. Approach – A Simple Linear regression model needs to be built with target variable ‘Delivery.Time’. Apply necessary transformations and record the RMSE values, Correlation coefficient values for different transformation models.



**Business Problem**

**Objective :-** Calories.Consumed’

**Python code details :**

Data Frame name is df. It has 21 rows and 2 columns.

**Work on each feature of the dataset to create a data dictionary as displayed in the below image:**

Then we create a data frame that’s contain details of each columns ,like- description ,data types ,and save the details named as data\_details .all of them are important .

**Data Pre-processing**

**Data Cleaning and Data Mining.**

Now we check info and describe for df .Check for data types ,unique value and variance .

Then we check for unique value in each columns

:-

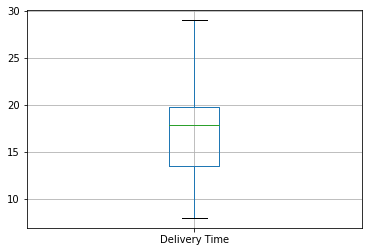
Delivery Time 21

Sorting Time 9

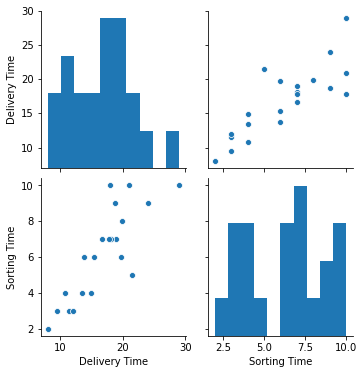
Dataframe has no missing values in columns .

We have done EDA for each columns and save the details as EDA. covariance for data set save as covariance . historgam and scatter plot for each column all data are normally distributed as well as we check for boxplot .there is no outliers present.

Boxplot:-



Histogram and Scatter plot:-



1. **Model Building:**
   1. **Perform Simple Linear Regression on the given datasets**
   2. **Apply different transformations such as exponential, log, polynomial transformations and calculate RMSE values, R-Squared values, Correlation**
   3. **Coefficient for each model**
   4. **Build the models and choose the best fit model**
   5. **Briefly explain the model output in the documentation**

We build ML models that predict delivery time gained using sorted time. here y is continuous .so we tried simple linear regression with different transformation.

**Result = MODEL RMSE R-squared Adj. R-squared Correlation coefficient**

**SLR 2.79 0.682 0.665 0.82**

**Log model 2.73 0.695 0.679 0.84**

**Exp model 2.94 0.710 0.695 0.83**

**Poly model 2.79 0.764 0.737 0.82**

**Model details saved Model\_details.**

Polynomial transformation is increasing Multiple R Squared Value. So we build our final mode on poly model.

Model reg is best fit model= Poly model ,root mean square error 2.79 , R\_squared 0.764 , . . Adj. R-squared 0.737 .

We split the data 50-50 in train and test.

So , we build our final model on SLR build on train data and test on test data .

Now we split our data in X\_train, X\_test, Y\_train, Y\_test 50% data on train and 50% test . Preparing a Simple linear regression on training data set (Polynomial with 2 degrees )

R-squared: 0.88

Adj. R-squared: 0.85

,then test on test data , Evaluation on Test Data as result root mean square error=5.01

Evaluation on Train Data also as result root mean square error=1.8

* used library :- pandas, numpy, sk learn, matplotlib, statsmodels