**Predict Delivery Time Based on Sorting Time**

Project Overview

The project’s ambition is to predict the ‘Delivery Time’ based on ‘Sorting Time’.

Steps Of Project

Importing the Packages and Data

The essential packages are imported from Python like Seaborn, Pandas, matplotlib in the Jupyter environment.

Data is also imported, checking for null values are done and are recovered.

EDA (Exploratory Data Analysis)

Heat map Analysis: This gives the correlation between ‘Delivery Time’ and ‘Sorting Time’. The correlation between them is 0.83 and is a positive linear regression.

Histogram Analysis: It interprets the skewness of the ‘Delivery Time’ which calculated is 0.33 and is slightly right skewed.

Box plot Analysis: There were no outliers found in the plot and no imputing was done.

Model Building

The three models were built based on splitting the data for training and testing follows as:

80:20

75:25

70:30

Models were built using scikit learn.

Results

The outcome of 80:20 model was best compared to the other two models. The R2\_Score with 84.01% is selected with RMSE 1.58.

This model can be applied in advertising to predict the money spent and revenue as response variable. Also like in Agriculture to predict the yield of crops based on fertilizers and water supply.