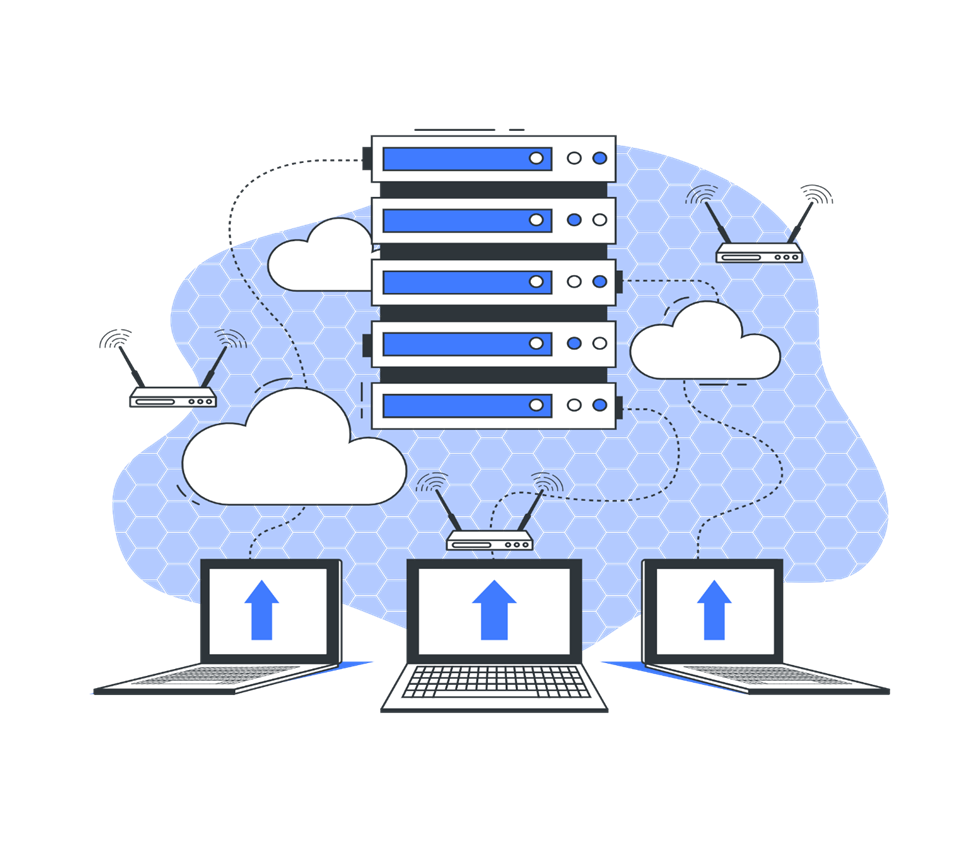
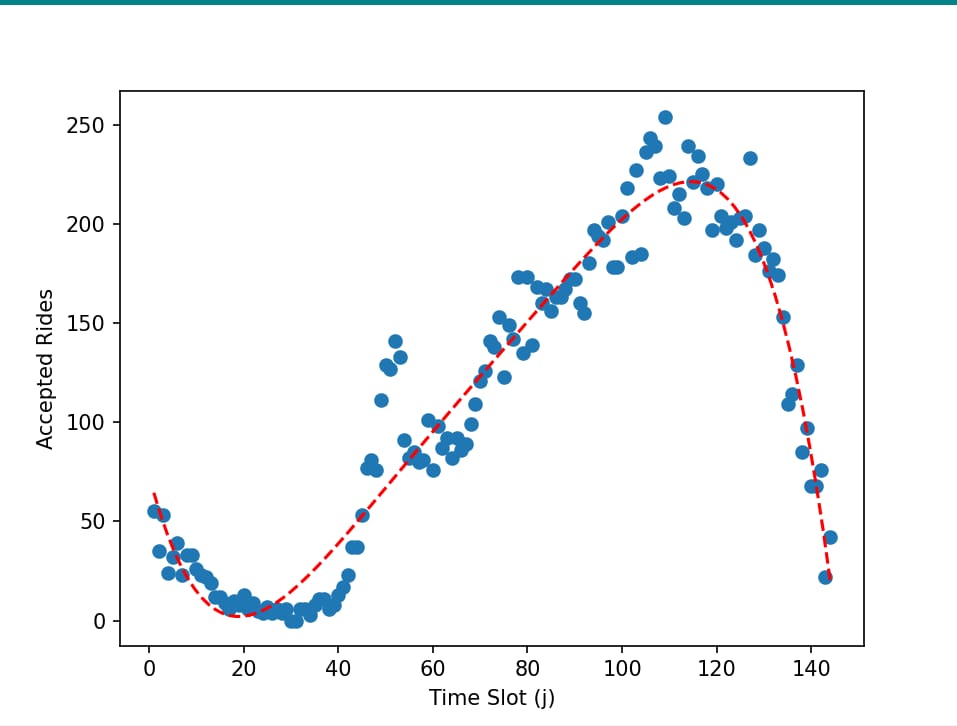
Applied Artificial Intelligence

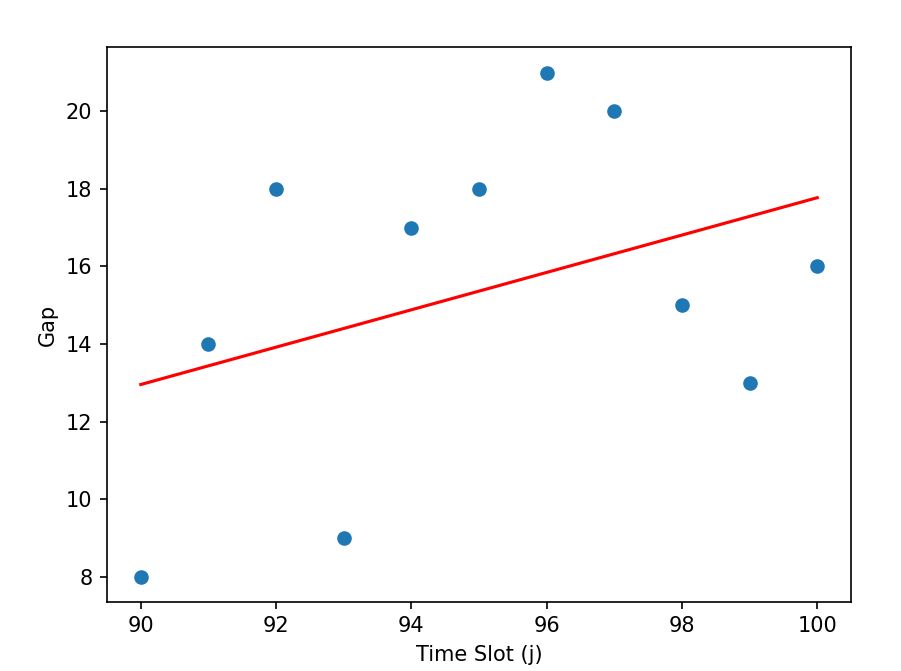


**Explanation**

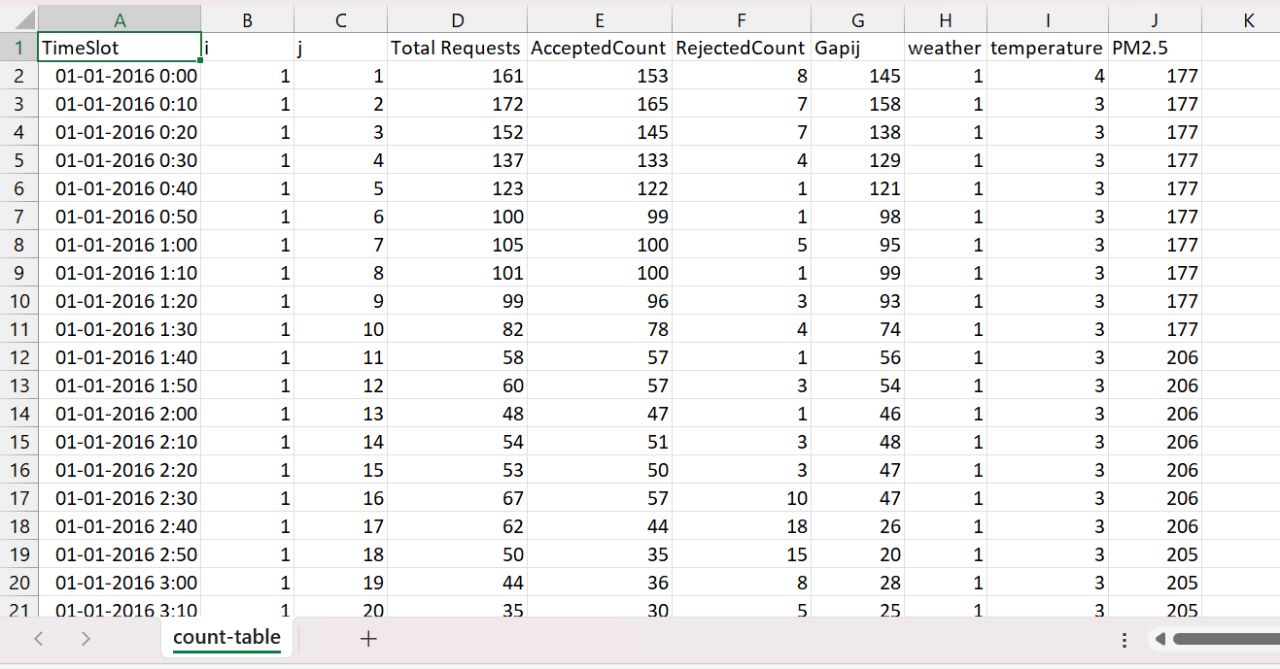
We started the project by reading the data files provided to us and trying to denormalize the data into a csv file, by first saving the data made in a data frame that is a data structure by pandas library and is a table like structure and then using join and merge commands to concat the data together corresponding to region id, hash id, order id and so on. After this made a new table that has a total 144 time slots 10 minutes each for a day and for each time slot we placed the count of accepted and rejected rides through which the gap is calculated by subtracting both. As the data is formulated of 144 time slots then weather data is added into the table for each time slot, added a column into that table for weather data.

Furthermore, after successfully creating the table containing all useful information we plotted the data with a regression line through the data to analyze overall data that is attached below:



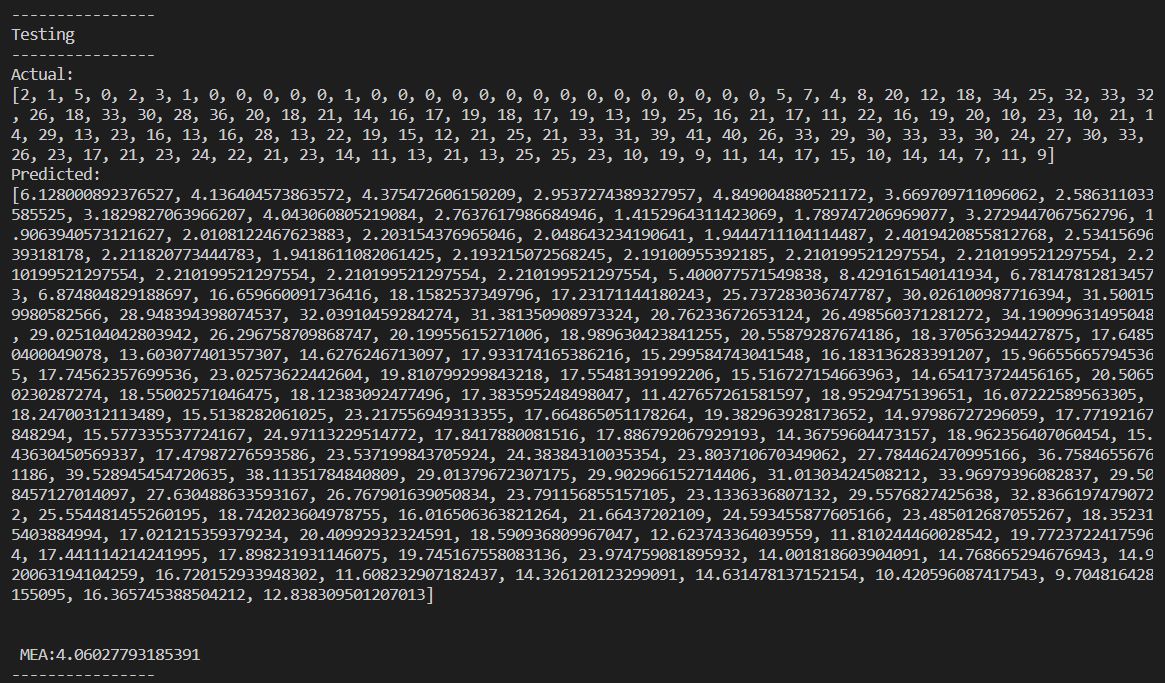


The table in csv format is attached below:



After successfully creating the data we created another data set having a rolling window of 10 which is our values of X and a single value of Y in each row for entire 144 time slots and then giving this new data to our regressor function from sklearn library which will train on our data set. Training the data will help us predict the gap on whatever time slot we need provided the rolling window is provided, that is data of the last 10 time slots is given in order to predict the value of Y through the regressor function.  
Moreover, we have divided our train data into 2 parts 1 for training and a small percent of that data to test our trained model and calculate the mean between the predicted gap and actual gap.

The following screenshot shows the output of arrays of predicted and actual gaps of a region on a specific day with MEA calculated



The provided training data was divided into days and first 15 days were kept to train out model, and next 6 days for testing of model. The following graph shows the Mean Absolute error of testing days

Chart, line chart

Description automatically generated

The program also supports the inputs by user. The user is required to input date and region from the training dataset and gaps of previous 10 time slots and it would output the prediction.

**<=THE END=>**