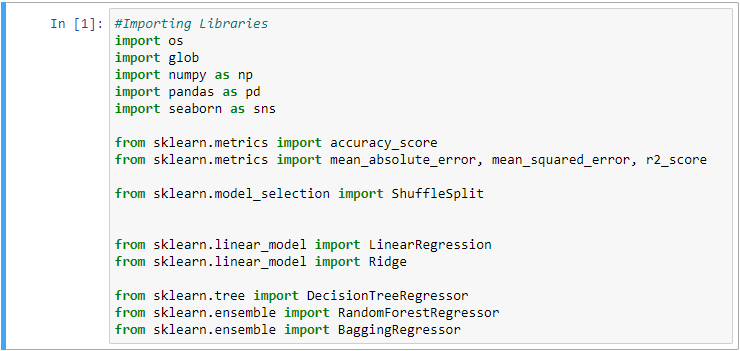
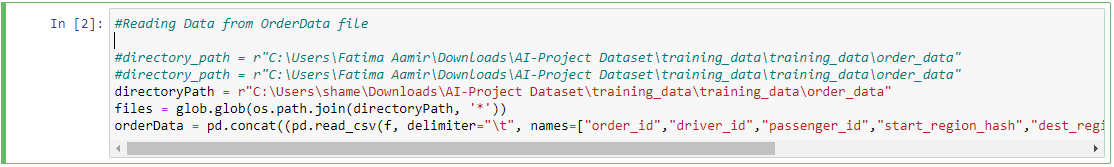
**AI PROJECT PHASE 1 REPORT**

Below is an explanation of the code we have used for this phase.

1. ***Libraries are imported***: The code begins by bringing in the important libraries like os, glob, NumPy, pandas, and seaborn, which are utilized for different data computation and analysis tasks. Where we used pandas and NumPy to the basic computation and handling of data. And seaborn for plotting of data.

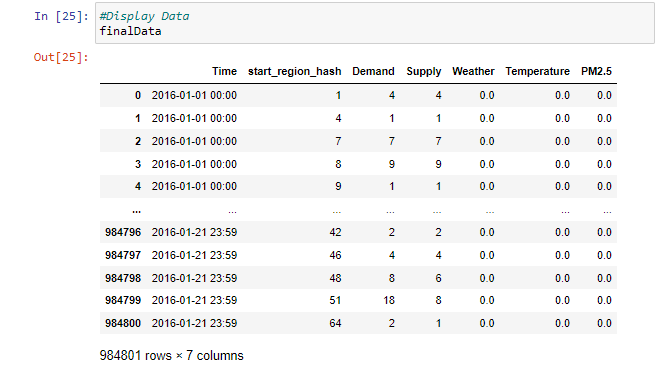


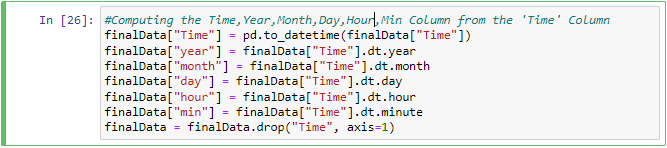
1. ***Reading Data from files***: The code reads data from the following files that were provided order Data, Weather Data, Poi Data and Cluster Map which are located in various directories with the help of glob library. The information is perused utilizing pandas read\_csv function with the suitable delimiter and feature names.
2. ***Data cleaning and pre-processing***: The code performs different data cleaning and pre-processing routines for example trimming time to exclude seconds, splitting and renaming features, combining datasets(merging), dropping duplicate rows, filling NULL values with 0, and mapping values starting with one dataset onto the other with the help of python (built-in dictionaries)



1. Calendar

   Description automatically generated***Data analysis and visualization***: The code utilizes seaborn library to make a pairplot of the dataset, which is a scatterplot framework that shows pairwise connections between factors. This aides in visualizing the connections between various factors in the dataset.
2. ***Data aggregation and computation:*** The code aggregates data with the help of group by function to calculate supply and demands based on different columns. The computed data is then merged with other datasets with the help of merge function to form a final dataset.



1. ***Data analysis and Feature Engineering***: The code further performs data analysis and feature engineering tasks, for example, extricating time, year, month, day, and hour data from the dataset utilizing datetime capabilities. This assists in making extra elements that with canning be utilized for additional examination and demonstrating.
2. ***Machine learning***: The code uses various machine learning algorithms such as Decision Tree Regressor, Random Forest Regressor, and Ada Boost Regressor for predicting the target variable based on the features. These models are imported from scikit-learn library and used for training and evaluation.
3. ***Model evaluation***: The code utilizes different assessment measurements, for example, accuracy\_score, mean\_absolute\_error, mean\_squared\_error, and r2\_score to assess the presentation of the prepared models. These measurements help in estimating the exactness and prescient force of the models.

Text

Description automatically generated

1. ***Model deployment:*** The trained models can be deployed in a production environment for making real-time predictions based on new data.
2. ***Model Optimization***: The model was optimized using a grid search for hyperparameter tuning, and the best-performing model was selected based on cross-validation results.
3. ***Results***: The best-performing Random Forest Regression model was selected based on the grid search results, and it was used to make predictions on the test dataset. The mean absolute error (MAE) between the predicted and actual values was calculated to be mean Abs Error. The accuracy of the model, measured by the R^2 score, was found to be [accuracy].