

## Contribution Sheet for Big Data Course Work

- **Makoto Ono-** Data downloading & uploading, Data cleaning, Feature engineering and exploratory data analysis (weekends holidays, peak hours, demand ratio, hourly/daily analysis, Correlation Matrix generation from data.gov), Code refactoring and optimization, Debugging, Size-up/Speed-up/Scale-up experiments for linear and nonlinear models. Report writing for data preparation, cleaning, feature engineering.
- **Raonak Shukla-** Data Cleaning, Feature Engineering- Temperature data from NOAA and Traffic Data from New York city departments. Feature engineering using embedded method using local approach and the using voting to select features from different partitions. Data Preparation for time series, Build Local and Global Regression model, Compared the performance of local models with global model. Report writing for size-up, speed-up and scale-up experiment, global and local models, plots for ensemble effect in distributed learning and heatmap for demand visualization.
- **Helin Lei-** File concatenation (column pruning technique), Data cleaning (null values, date format), Data Visualization (Stacked Bar chart etc.), Research for different models for time series and Regression, build Local Linear Regression model and Global Linear Regression model, Literature Review, Report Writing (Introduction, Problem Description, Related work, Speed-up, cross referencing).
- **Hande Simay Telatar-** Data cleaning for time series, data visualisation for time series, fitting different models like ETS, ARIMA, SARIMA for time series, evaluation of fitted models using cross validation, optimisation of final model by hyper parameter tuning, demand forecasting for next 12 weeks based on final model. Report writing for time series section.
- **Palak Bajaj-** Initial research ideas and collection of information for the project, applied different feature selection techniques, analyzed the statistical importance of features.
- **Kumar Harsh –** Heatmap based on location and the number of trips as per the zones and boroughs of the New York city, Feature engineering – incorporated holiday data frame alongside with the weekend days as a new feature. Initial research idea discussion for the project.