

# New York Citi Bike Trip Duration Prediction

## CSE 258 Assignment 2

Zhuo Cheng  
A53233235  
UC San Diego  
San Diego, CA  
zhc262@eng.ucsd.edu

Tianran Zhang  
A53237053  
UC San Diego  
San Diego, CA  
tiz217@eng.ucsd.edu

Jiamin He  
A53243479  
UC San Diego  
San Diego, CA  
jih426@eng.ucsd.edu

### ABSTRACT

In this project, a model that predicts the ride duration of a share bike trip in New York City is built. The patterns related to the geographical locations and start and end time with passenger businesses are observed and evaluated. We utilize the dataset released by the Citi Bike Share System in New York City<sup>1</sup>, which provides the necessary information to accomplish our duration prediction goal. We manipulated and mined the dataset by multiple models we learned from CSE 258 course taught by Prof. Julian McAuley and the further reading materials. We used linear regression, ridge regression, random forest, XGBoost and the ensemble of random forest and XGBoost to build different regression models. And we finally get a best model by utilizing XGBoost, which helps to predict the duration of a ride and thus helps users decide a time-saving transportation means.

### KEYWORDS

Share Bike, Duration Prediction, Linear Regression, Random Forest Regression, XGBoost

### 1 INTRODUCTION

Bike shares are riding a wave of popularity in the intermodal transit planning community. Through bike sharing systems in a city, people are able to rent a bike from a one location and return it to a different place on an as-needed basis. The number of bike share systems, defined as publicly-available systems with at least 10 stations and 100 bikes, has steadily increased year-over-year, from four systems in 2010 to 55 systems in 2016 across U.S, with over 42,000 bikes available in cities of all sizes[2]. In addition, 80% of systems that have been in operation for more than a year have expanded since they launched. The number of bikes in the nation also increased substantially, up 30%, as existing large systems have continued to grow. Figure 1 in page 1 shows the bike share growth in the US from 2010 to 2016<sup>2</sup>.

The growth of bike share shows no signs of stopping. A number of U.S. cities, such as Detroit, New Haven, and New Orleans, have either selected vendors or are planning to launch systems, and

many existing systems are also rolling out major expansions: New York's Citi Bike Program [3] is adding another 2,000 bikes, for a total of 12,000; Houston is more than tripling in size to over 100 stations; and the San Francisco Bay Area is expanding from a 700 to a 7,000 bike system[5]. Here in our assignment, we conduct a case study based on the bike sharing data on the Citi Bike Share System in New York City in order to predict the duration for a user in one bike trip at a certain time. We try to forecast the duration of a share-bike trip to help improve and expand the bike sharing system in NYC. Predicting duration of a bike ride can also help users to make wise decisions on hailing a cab or riding a bike, for example.

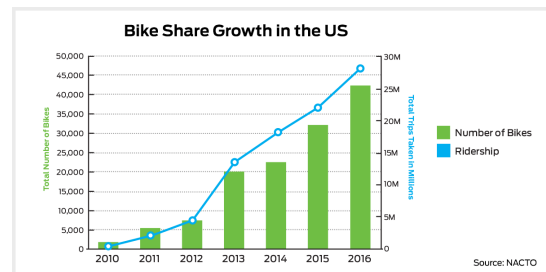


Figure 1: Bike Share Growth in the US

In this project, we obtained the primary dataset provided by Citibike System Data. The data generated by these bike share systems are consisted of duration of travel, start time and date, stop time and date, start station name, stop station name, station ID, station latitude and longitude, bike ID, user type (Customer = 24-hour pass or 3-day pass user; Subscriber = Annual Member), user gender and user's year of birth. Given adequate data from Jul 2013 to Sep 2017, we used data in 2016 year round. In our analysis, we attempt to evaluate the importance of different features and extract the non-trivial ones to create a prediction model for this problem. We then compare performances of different models and discuss their effectiveness and shortcomings. In this assignment, We mainly use the method of multiple linear regression analysis, Random Forest and XGBoost to forecast the bike share trip duration in New York City. An ensemble of Random Forest and XGBoost is also applied in this report. We evaluate these models based on the Fraction of Variance Unexplained (FVU).

<sup>1</sup>Citi Bike Share System: <https://www.citibikenyc.com>

<sup>2</sup>Figure taken from <https://nacto.org/bike-share-statistics-2016/>, December 02, 2017

## 2 DATA

### 2.1 Dataset Variables

We obtain the data from the data system of Citi Bike, which is the national largest bike share program, with 10,000 bikes and 600 stations across Manhattan, Brooklyn, Queens and Jersey City[3]. It consists of data from the program's launch date, from 07/2013 to 09/2017. The dataset provided information<sup>34</sup> is as shown in Table 1 in page 2.

**Table 1: Dataset Variable and Type**

Variate	Format
Trip Duration	in seconds format
Start Time and Date	Timestamp
Stop Time and Date	Timestamp
Start Station Name	String
End Station Name	String
Station ID	Number
Station Lat/Long	Number
Bike ID	Number
User Type	Customer or Subscriber
Gender	Number
Year of Birth	Number

### 2.2 Data Cleaning

Basically in 2016, there were 11231 bike stations and the total number of rides in dataset is 12224313. The total number of customers' rides is 1543944 and subscribers' rides is 12301711.

From a preliminary analysis on these data, we need to detect and deal with some abnormal data. For some cases too short duration or even zero or negative is deleted; cases that take too long duration in a trip ( $> 1$  million seconds, roughly around a month) are also removed; for some cases the users' information is missing, age and sex, for example; for some the ratio of distance by duration is unreasonably high (30 mile/hour) or low (1 mile/hour). A lower bound (2.2369 mile/hour) and an upper bound (19.0137 mile/hour) is used in this analysis to protect our models from being skewed by extremities and make them immune to outliers, which is a reasonable range from walking to driving[17].

After all these data cleaning, we randomly choose 30k trips from each month in 2016, which creates a dataset with a size of 360k in total. And we split it into three sets: training, validation and test, each with 120,000 rides.

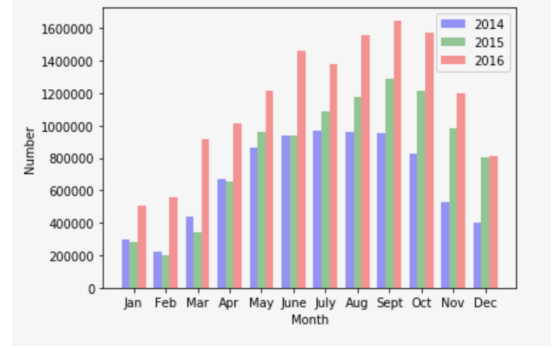
### 2.3 Exploratory Analysis

To better understand the problem and extract non-trivial features, we perform an exploratory analysis on the data.

**2.3.1 Prevalance.** Figure 2 in page 2 illustrates an increase in the number of trips starting from 2014, when the citi bike program just launched, to 2016. This shows that in 2016, a great number of people start to choose citi bike share program.

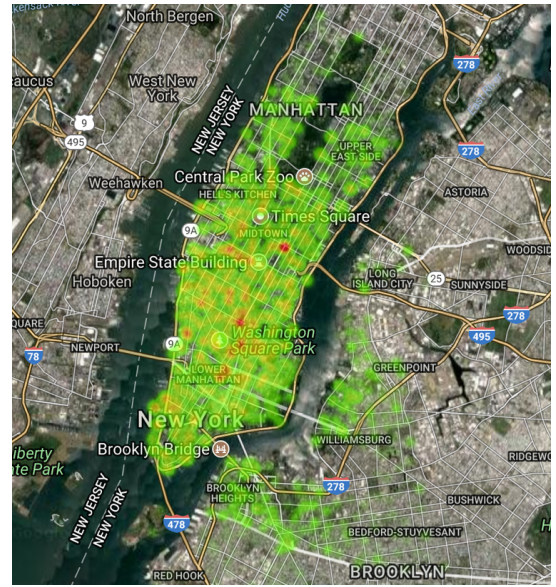
<sup>3</sup>Customer = 24-hour pass or 3-day pass user; Subscriber = Annual Member

<sup>4</sup>0=unknown; 1=male; 2=female



**Figure 2: The number of user from 2014 to 2016**

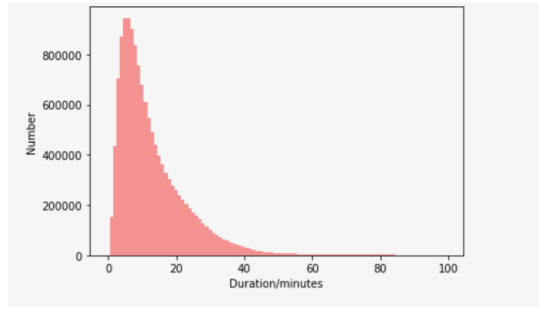
**2.3.2 Docking location.** To investigate how the bike docking location is distributed and figure out the active area in New York City, we plotted the start and end location distribution of each bike trip (Figure 3 in page 2). Red color represents the most active area, followed by yeallow, and then green.



**Figure 3: The distribution of start and end location in 2016**

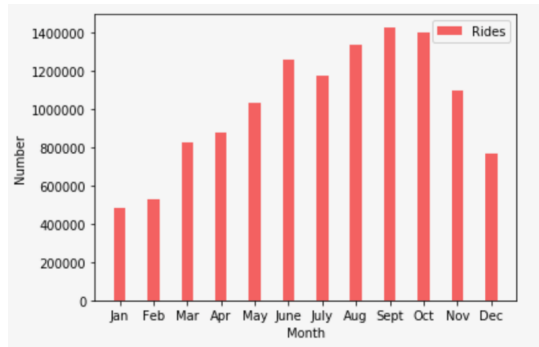
**2.3.3 Distribution of duration.** Figure 4 in page 3 shows the overall distribution of bike duration in 2016. We can see that most of the duration are in the range (0, 20), some are in (20, 40) and few of them are larger than 40.

**2.3.4 Month.** To see the influence of date and time, we firstly investigate the influence of months on the bike trip duration. This can be clearly seen in Figure 5 in page 3. It shows the distribution of months in bike share system in 2016. In January and February, not many people are using the bike sharing system and we suspect that it is due to the cold weather, which makes the riding time painful. Similarly, though the number of rides is increasing, there



**Figure 4: The overall distribution of bike duration in 2016**

is a subtle decrease in July, which is possibly due to a too high temperature during that period. Overall, in spring and fall, there are many people who are using bike share system, while in summer, this system becomes most popular and as winter comes, it becomes less attractive to people.



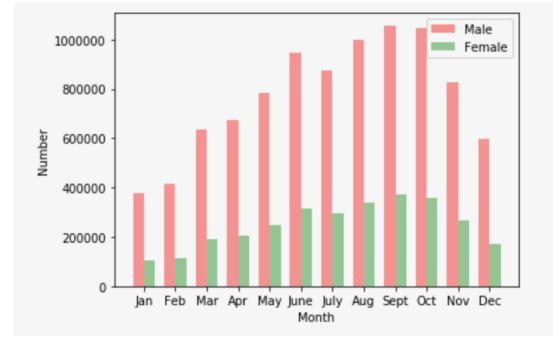
**Figure 5: Distribution of months in bike share system in 2016**

**2.3.5 Gender:** To understand the influence of the gender on the duration of each bike trip, we draw Figure 6 in page 3. It shows the distribution of gender in bike share system in 2016. It is obvious that males ride much more frequently than females. From Figure 7 in page 3, it generally takes more time for females during a ride trip because of the different velocity. It can also be deduced from Figure 8 in page 3.

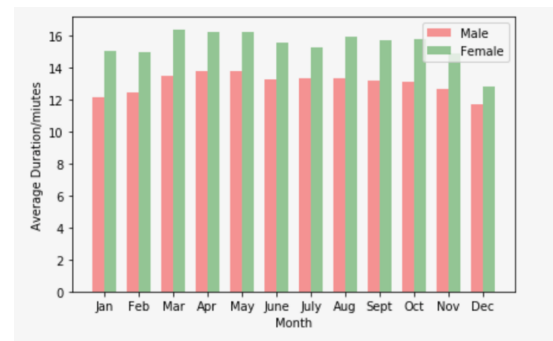
**2.3.6 User Subscription:** The users' subscribed plan also makes a difference in their corresponding bike trip duration. From Figure 9 in page 4, we can clearly see that subscribers ride much more frequently than ordinary customers. But for subscribers, they prefer riding even in a short duration trip. For customers, it is more economic to ride a longer duration trip as it costs them the same fee during one trip in a certain duration.

### 3 PREDICTIVE TASK

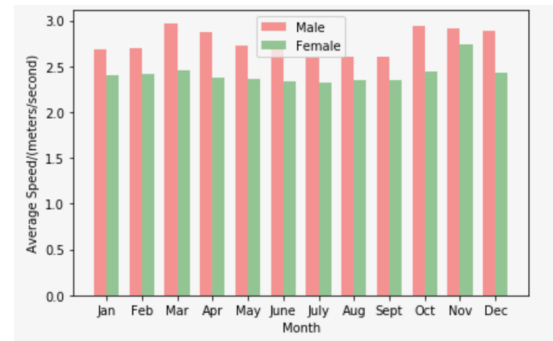
Based on the given dataset, what we want to do is to predict a single trip duration in the Citi Bike Share System in New York City.



**Figure 6: Numbers of rides of gender in bike share system in 2016**



**Figure 7: Durations of rides of gender in bike share system in 2016**



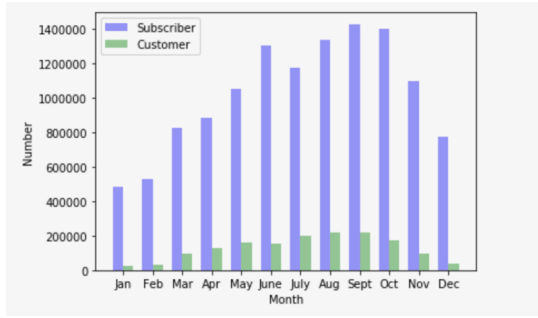
**Figure 8: Velocity of each ride of gender in bike share system in 2016**

### 3.1 Evaluation Method

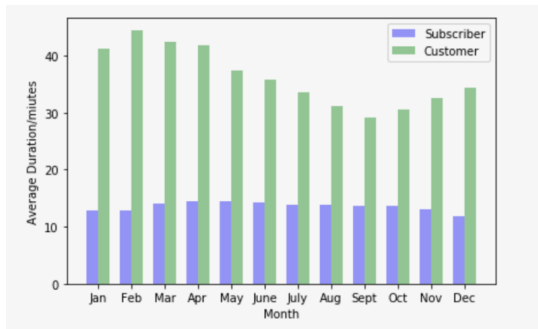
We evaluate the performance of different models on the basis of the fraction of variance unexplained (FVU). FVU is calculated as:

$$FVU(f) = 1 - R^2 = \frac{MSE(f)}{Var(y)}$$

FVU is suitable for our problem because FVU in statistics, is the fraction of variance of the regressand (dependent variable)  $Y$  which cannot be explained, i.e., which is not correctly predicted, by the



**Figure 9: Numbers of rides of different user type: subscriber and customer in bike share system in 2016**



**Figure 10: Durations of rides of different user type: subscriber and customer in bike share system in 2016**

explanatory variables  $X$ . When trying to predict  $Y$ , the most naive regression function that we can think of is the constant function predicting the mean of  $Y$ , i.e.,  $f(x_i) = \bar{y}$ . It follows that the MSE of this function equals the variance of  $Y$  and the FVU then reaches its maximum value of 1. So if the explanatory variables  $X$  tell us nothing about  $Y$ ,  $FVU(f) = 1$ . But as prediction gets better, MSE can then be reduced and ideally the best model will have  $FVU(f) = 0$ .

### 3.2 Data Preprocessing

In this dataset, 30k observations is randomly extracted from data in each month in 2016. And this aggregated dataset now contains 360k observations and is well distributed in all year round. It is then shuffled and divided into three splits, individually 1/3 for training, validation and test. Each training sample consists of 15 entries – the first 11 contains the basic information for a certain bike trip. In addition, there are also 4 entries that represent the rider’s information. To make a better use of these training samples, we preprocess and convert them into a more intuitive form as described below.

(1) Day of month:

From the given timestamp of the start date and end date we can easily compute the day of the month, by using one-hot encoding, we use 31 features to represent different days in a month.

(2) Day of week:

Using the timestamp provided in our dataset, by using one-hot encoding, we construct 7 features to represent different days in a week in order to provide more explanation on week-days and weekends.

(3) Year:

As we focused mainly on data in 2016, so the year feature here is set all 2016, and in later discussion be left behind.

(4) Month:

With the same technique, we use 12 features to represent the months from January to December.

(5) Hours:

Given the specific start time and stop time, by one-hot encoding we can use 24 features to represent the 24 hours in a day.

(6) Minutes:

From the timestamp value present in the data set, we divide 0~60 minutes into four features, each of them covers a range in 15 minutes.

(7) Distance:

In order to represent the latitude and longitude information of the start and end station in some useful forms, we consider several different ways, such as mapping them into zip code and use one-hot encoding to represent them into categorical features. Due to the largeness of our data set, it takes too long to convert each pair of latitude and longitude into a zip code representation. We choose to use the distance between the start station and end station as the feature. The conversion is simply by using the vincenty distance of the Geopy library in Python[7].

(8) Gender:

In the original dataset, 1 for males and 2 for females. After processing, we convert it into a boolean feature, 1 for females and 0 for males.

(9) Age:

In order to represent the age in a useful form, we set 30 year-old as a watershed and use two categories to represent it. If the given age is smaller than or equal to 30, we set the negative value of given age to the first category and 0 to the second category. If the given age is greater than 30, we set 30 to the first category and the difference between the given age and 30 to the second category. For example, if the given age is 15, then we set -15 to the first category and 0 as the second. If the given age is 40, then we set 30 to the first category and 10 to the second. The intuition to represent the feature age in this way is that we consider the age 30 and below is negatively related to the duration of the ride, which means that the closer to the 30, the shorter duration it takes. And we consider the age above 30 is positively related to the duration of the ride, which means that the further to the 30, the longer duration it takes.

### 3.3 Features Selection

The previous exploratory analysis on different features illustrated a few important features that we should include. We firstly incorporated the latitude and longitude of the start and end station, month, week, hour, minutes, gender and age as our feature vectors.



To analyze the importance of each feature, we remove each of these features from our feature set to see how the performance changes. Figure 11 in page 5 shows the importance of some chosen features.

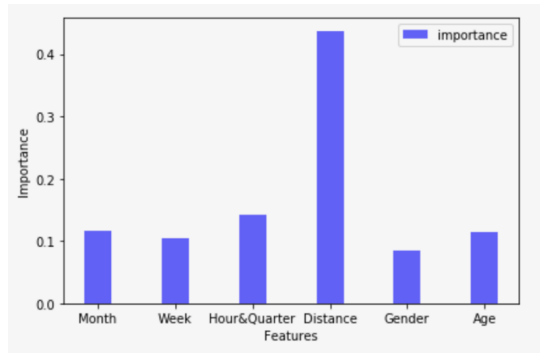


Figure 11: Importance of features

When we consider if selecting the day or week as one of the features and if adding time into the feature vector, there are different opinions in our group, so we draw a bar graph (Fig. 12 in page 12) to compare four cases: week, week and time, day, day and time. We add one of these combinations into our feature set each time and run our linear regression model to see which one works best. As shown in Figure 12 in page 5, by adding the features week and time, the performance is the best.

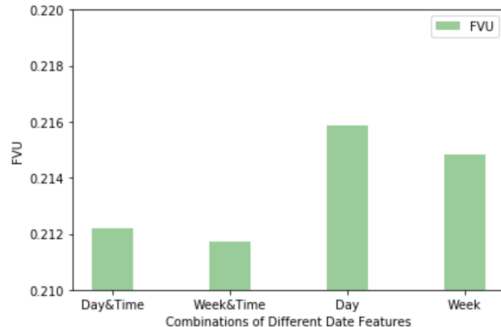


Figure 12: Date and Time Feature Selection

## 4 MODELS AND METHODOLOGY

### 4.1 Baseline

We define the baseline predictor by calculating the mean value of our label duration in the train set. However, the FVU value for the baseline model on the test set is 1, which proves to be a bad predictor. But it makes sense since no variation in the label can be accounted for, the variable  $X$  not providing any information about the label.

### 4.2 Linear Regression

Then we try a basic model we learned in class: linear regression, which has the form  $Y = \theta X$ . The feature we chose is from the

important features that we have discussed in the data exploration part. The FVU value we got is 0.21509, which performs much better than the baseline.

### 4.3 Ridge Regression

In order to avoid overfitting to the training set, we introduced the regularization into our linear regression model, where we penalize the model complexity during the training process.

$$\operatorname{argmin}_{\theta} = \frac{1}{N} \|y - X\theta\|_2^2 + \lambda \|\theta\|_2^2$$

We use the ridge regressor provided by the Sklearn library. To optimize the lambda term, we plot the function of validation FVU against different values of lambda ranging from 1 to 200. The FVU corresponding to different values of lambda can be seen in Figure 13 in page 5.

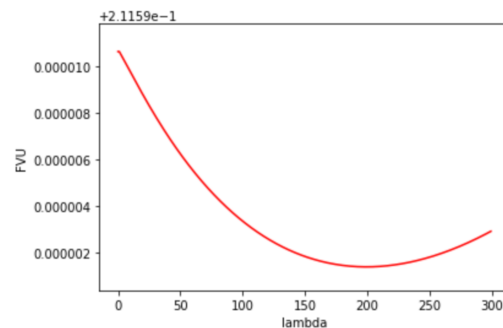


Figure 13: The relationship of FVU and lambda

As shown in the figure, the optimal value for lambda is  $\lambda = 199$ , which we select in our prediction. The FVU we got for the ridge regression model is 0.211591.

### 4.4 Random Forest Regressor

In our feature vectors, some features are in the numerical form such as the distance and age; others are categorical binary features such as the month, week, hour and minute. This is the reason how we come up with the idea to use the random forest regression since the random forest model works well with a mixture of numerical and categorical features. A random forest is a way of averaging multiple deep decision trees[9], trained on different parts of the same training set, with the goal of reducing the variance[8]. To implement the random forest model, we use the *RandomForestRegressor* of the library sklearn in Python.

To initialize the *RandomForestRegressor*, we need to pass in two parameters: the maximum depth and the random state. To figure out the optimal value of the max-depth, we plot the validation FVU value with different values of max-depth and from the figure that we can know the best value is 20. It is true that the performance of the model could be improved by increasing the value of max-depth. However, it is easy to see from figure 14 in page 6 that the improvement is not significant as the max-depth increases further from 20. Also the larger the value of the max-depth, the longer time the model takes to run.

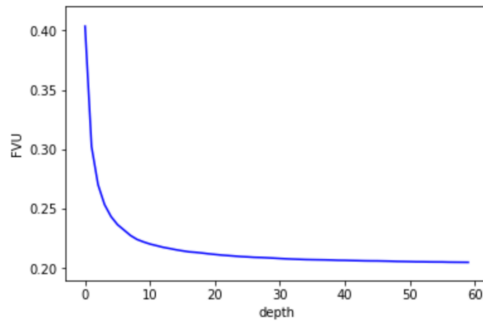


Figure 14: The FVU of random forests under different depth

The FVU value we got is around 0.21226, which is better than both the linear regression and ridge regression. It is because the random forest model is able to capture the subtle non-linear interaction effects between variables. If all variables of interest has the linear dependency from the predictors, with both algorithms we might get almost the same result. The difference between these two models is mainly due to some non-linear variables and this detection is an advantage of the tree-based models because these models are equipped to do this. However, this improvement is at the cost of more time for fitting, which also encourages us to continue to work on finding better models (It is discussed in the next part: XGBoost Regressor).

#### 4.5 XGBoost Regressor

XGBoost is a shorthand for eXtreme Gradient Boosting, which is an algorithm that has recently been dominating applied machine learning and Kaggle competitions for structured or tabular data[4]. After performing the random forest regression model, as we increased the max-depth value, the performance gets improved, but it takes longer and longer to run the model, which brings us to try the xgboost model since one of the most significant feature of the xgboost is its fast execution speed[1].

To implement the xgboost model, we use the *XGBRegressor* of the library xgboost in Python. Similar to *RandomForestRegressor*, we also need to tune the max-depth parameter, however, we find that the xgboost model could achieve a very good performance with the max-depth only to be 7, which is way smaller than that of the Random Forest Regressor. The process about tuning the max-depth and the learning-rate can be seen in Figure 15 in page 6.

The FVU value we got for the xgboost model is 0.19888, which improves around 0.01 from the random forest regression model.

#### 4.6 Ensemble of Random Forest and XGBoost

In order to achieve a better performance, we came up with the method of ensemble of models[6], which is a very popular and powerful technique on a variety of predicting tasks[11]. The ensemble between different models is a process to select two or more models which are related but not same and combine their prediction results into a single value as our final prediction.

In our model, we create an ensemble between the random forest regression and xgboost. There are many ensemble ways such as

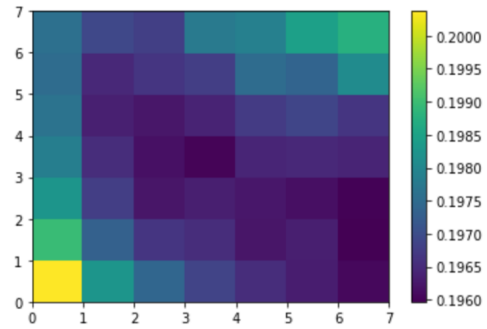


Figure 15: tuning parameter: maxdepth and learning rate

stacking and averaging. We adopt the averaging method, taking the mean of the predictions from each model as our final predictions. The FVU we got for the combination model is around 0.20058. We can see that the ensemble model works worse than the single xgboost, which is not usually the case.

To figure out the reason behind, we read some researches[11] and make some analysis on both the random forest regression model and xgboost[12]. We conclude that it is because the xgboost dominates in this ensemble models. It possesses its unique ability to capture the complexity of big data, so the single xgboost model works better than the ensemble model.

### 5 LITERATURE

#### 5.1 Relevant Dataset

Our dataset is provided by New York Citi Bike Share System and most of the bike share system provide opening dataset in similar formula to public. Similar dataset can also be visited and gained through Hubway Bike Share System in Boston, MA<sup>5</sup>; DecoBike in Miami, FL<sup>6</sup>; Relay Bike Share System in Atlanta, GA<sup>7</sup>; Reddy Bike Share System in Buffalo, NY<sup>8</sup>; Divvy Bike Share System in Chicago, IL<sup>9</sup>; Capital Bike Share System in Washington, DC<sup>10</sup>, etc. These systems generate a great deal of data relating to various ride details, including trip duration, start and end location, etc.

The dataset we used here is of great importance for us to study public transportation and city mobility. Leveraging the historical data provided can benefit us to providing more accurate prediction for bike trip duration at a time.

#### 5.2 Related Work

Many previous research and investigation have been conducted in bike trip prediction. But most of them are forecasting the share bike demand at a certain time or at a certain place. Also, there has been a share bike demand prediction Competition running on Kaggle three years ago. But few of relevant projects are delved into bike trip duration prediction as what we did in this project.

<sup>5</sup>Hubway Bike Share: <http://www.cambridgema.gov/CDD/Transportation/gettingaroundcambridge/bikeshare>

<sup>6</sup>Deco Bike, Citi Bike: <http://citibikemiami.com>

<sup>7</sup>Relay Bike Share: <http://relaybikeshare.com/system-data/>

<sup>8</sup>Relay Bike Share: <https://reddybikeshare.socialbicycles.com>

<sup>9</sup>Divvy Bike Share: <https://www.divvybikes.com>

<sup>10</sup>Relay Bike Share: <https://www.capitalbikeshare.com/system-data>

For those who are predicting the share bike demand, some techniques and key features that they have used and worth mentioning are as follows:

**5.2.1 Geographic and Distance Information.** Javier etc.[14] develops a rapid response ridership forecast model, based on the combined use of Geographic Information Systems (GIS), distance-decay functions and multiple regression models. Analyses carried out show that weighting the variables according to the distance-decay functions provides systematically better results. The choice of distance threshold also significantly improves outcomes. Osvaldo etc. [10] develops models based on GIS and proves its considerable advantages over the traditional four-step model, including simplicity of use, easy interpretation of results, immediate response and low cost. This study also uses geographically weighted regression (GWR) to deal with the station prediction problem. Patrick etc.[16] refrains from building a spatial model to assess the bike share repositioning services.

**5.2.2 Bike Docking Station Locations.** Yexin Li etc.[15] proposed a bipartite prediction algorithm based on hierarchical clustering to cluster bike stations into groups and therefore get different levels of hierarchical bike stations.

**5.2.3 User Subscription Plans and User Habits.** Elliot etc.[13] investigates several significant predictors of membership including reactions to mandatory helmet legislation, riding activity over the previous month, and the degree to which convenience motivated private bike riding. These results provide insight as to the relative influence of various factors impacting on bike share membership in Australia. The findings may assist bike share operators to maximize membership potential and help achieve the primary goal of bike share to increase the sustainability of the transport system.

**5.2.4 Time Series Data.** To predict the demand of bikes, a multi-similarity-based interference model based on gradient boosting regression tree is proposed to predict the rent proportion across clusters and the inter-cluster transition, based on which the number of bikes rent from/returned to each cluster can be easily inferred[15].

### 5.3 Analysis on this task

From the above analysis, we can see most of the prediction task are trying to solve the re-allocation problem or the arrangement of the bike docking station problem. So they pay more attention on the specific bike docking stations and try to deeply understand users' habits. Also, a majority of these investigations are made to predict the bike demand for an entire area and that is why they care more about geographic features. However, our predictive task is different from above mentioned work because we are trying to predict the duration of each bike trip. So we convert the specific location information such as longitude and latitude to a more concrete distance feature during a trip. Also, the user's habit and subscription would be helpful in our prediction and we also included gender, age and other user information in our prediction models.

## 6 RESULTS AND CONCLUSIONS

### 6.1 Results

FVU values with different models in this duration prediction task calculated with different models we built are as described in Figure 16 (baseline included), Figure 17 (baseline not included) and Table 2 in page 2:

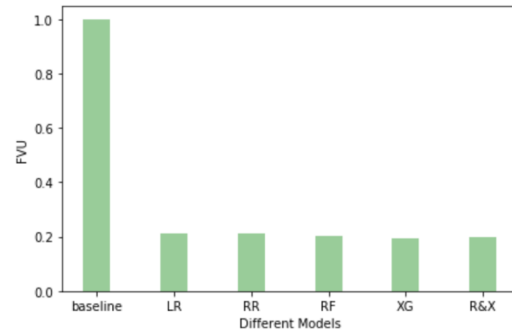


Figure 16: results of different models (baseline included)

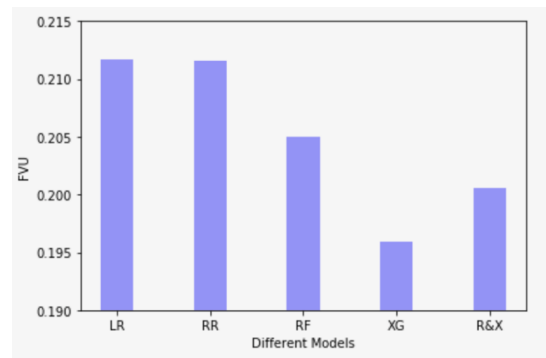


Figure 17: results of different models (baseline not included)

Table 2: Comparison of different models

Model	FVU
Baseline	1.000006
Linear Regression	0.211735
Ridge Regression	0.211591
Random Forest Regressor	0.205021
XGBoost Regressor	0.195970
Ensemble of Random Forest and XGBoost	0.200575

From comparison of different models we can get a lowest FVU of 0.195970 with XGBoosting Regressor as a bike trip duration prediction model, which can help provide efficient and accurate duration prediction for each bike trip. The predicted duration of each trip compared with the actual duration of each trip in test set

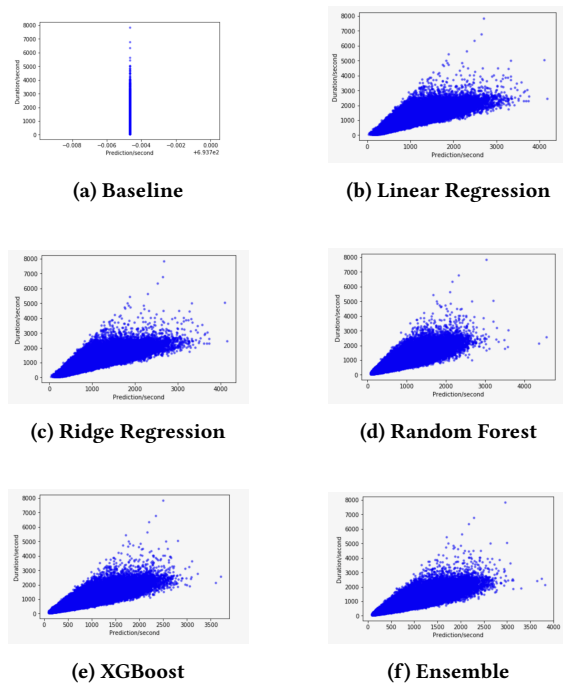


Figure 18: The predicted duration and actual duration

is drawn in Figure 18 in page 8. The XGBoost model works the best because the value of most predictions are close to the value of real data.

## 6.2 Conclusions

Linear Regression, Ridge Regression, Random Forest and XGBoost (and also an ensemble method which combines Random Forest and XGBoost, though it turns out to be not so good as predicted<sup>11</sup>) are the models we have trained to predict the duration of each share bike trip for a user.

From the results shown above, the XGBoost model works the best, both in running time and in prediction accuracy. We also choose different features according to their importance from our data exploration and data processing part. With these models and features, we successfully established the relationship between the chosen independent variables and the duration of a bike trip.

## 6.3 Future Work

In future work, we want to add more important features in our prediction models, which needs more data provided to extract. For example, given the temperature or wind direction or humidity during a trip, these features can definitely added into the model and we can evaluate its effect on the final duration of a trip. Based on these hidden facts behind the data, we can provide more suggestions to the city share bike trip system and of course with more key features added into the prediction models, we can forecast the duration of a trip more accurately. Besides the environment and

weather data we have mentioned above, more covariates that we might want to consider in our models are as follows:

### (1) Traffic condition data.

It is of great possibility that when the traffic is busy, people might prefer riding bikes and riding for a longer duration. Thus we suspect that there could be a positive correlation between the traffic heaviness and the duration of a bike trip.

### (2) Holidays and Tourists Visiting data.

We suspect that there could be some relationships between holidays and the bike trip duration. But we cannot make assumption that it is a positive relationship or a negative relationship. During holidays, maybe people all drive to nearby cities to enjoy the nature and scenery, which results in less demand and less duration for each bike trip. Or it might leads to people's relaxing and exercising so there appears to be a longer duration for each bike trip.

Similarly, for a tourism city, during summer or the city's most beautiful and attractive season, the increasing number of tourists might result in longer duration of each trip as tourist like to enjoy the scenery and explore the city by riding for a whole day.

### (3) Taxi and subway data.

It is evident that bike, taxi and subway are all daily public transportation. And it is a reasonable suspect that an area that is short of taxi or is far away from subway station would have longer duration of a bike trip.

In a word, given more useful data, we can find more hidden relationship behind these big data with different effective models like gradient boosting, random forest, XGBoosting or light GBM. We can also try some neural network algorithms or consider applying hierarchical clustering in our models. And thus we can offer more efficient and accurate duration prediction for a bike trip.

## REFERENCES

- [1] [n. d.]. *Benchmarking random forest implementations*. <http://datascience.la/benchmarking-random-forest-implementations/>.
- [2] [n. d.]. *Bike Share in the US: 2010-2016*. <https://nacto.org/bike-share-statistics-2016/>.
- [3] [n. d.]. *Citi Bike Share system in New York City*. <https://www.citibikenyc.com>.
- [4] [n. d.]. *A Gentle Introduction to XGBoost for Applied Machine Learning*. <https://machinelearningmastery.com/gentle-introduction-xgboost-applied-machine-learning/>.
- [5] [n. d.]. *How bike share is changing American cities*. <https://www.curbed.com/2017/3/21/15006248/bike-share-ridership-transit-safety>.
- [6] [n. d.]. *KAGGLE ENSEMBLING GUIDE*. <https://mlwave.com/kaggle-ensembling-guide/>.
- [7] [n. d.]. *Python Geopy Toolbox*. <https://pypi.python.org/pypi/geopy>.
- [8] [n. d.]. *Random Forest Introduction*. [https://en.wikipedia.org/wiki/Random\\_forest](https://en.wikipedia.org/wiki/Random_forest).
- [9] Leo Breiman. 2001. Random forests. *Machine learning* 45, 1 (2001), 5–32.
- [10] Osvaldo Daniel Cardozo, Juan Carlos García-Palomares, and Javier Gutiérrez. 2012. Application of geographically weighted regression to the direct forecasting of transit ridership at station-level. *Applied Geography* 34 (2012), 548–558.
- [11] Corinna Cortes, Vitaly Kuznetsov, and Mehryar Mohri. 2014. Ensemble methods for structured prediction. In *International Conference on Machine Learning*. 1134–1142.
- [12] Thomas G Dietterich et al. 2000. Ensemble methods in machine learning. *Multiple classifier systems 1857* (2000), 1–15.
- [13] Elliot Fishman, Simon Washington, Narelle Haworth, and Angela Watson. 2015. Factors influencing bike share membership: An analysis of Melbourne and Brisbane. *Transportation research part A: policy and practice* 71 (2015), 17–30.
- [14] Javier Gutiérrez, Osvaldo Daniel Cardozo, and Juan Carlos García-Palomares. 2011. Transit ridership forecasting at station level: an approach based on distance-decay weighted regression. *Journal of Transport Geography* 19, 6 (2011), 1081–1092.

<sup>11</sup>The possible explanation for it have been given in the previous part.



- [15] Yexin Li, Yu Zheng, Huichu Zhang, and Lei Chen. 2015. Traffic Prediction in a Bike-sharing System. In *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems (SIGSPATIAL '15)*. ACM, New York, NY, USA, Article 33, 10 pages. <https://doi.org/10.1145/2820783.2820837>
- [16] Patrick Vogel and Dirk Christian Mattfeld. 2010. Modeling of repositioning activities in bike-sharing systems. In *World conference on transport research (WCTR)*.
- [17] Yu Zheng, Like Liu, Longhao Wang, and Xing Xie. 2008. Learning Transportation Mode from Raw Gps Data for Geographic Applications on the Web. In *Proceedings of the 17th International Conference on World Wide Web (WWW '08)*. ACM, New York, NY, USA, 247–256. <https://doi.org/10.1145/1367497.1367532>