

# ADS Project 5: CitiBike Explorer

Team: #5

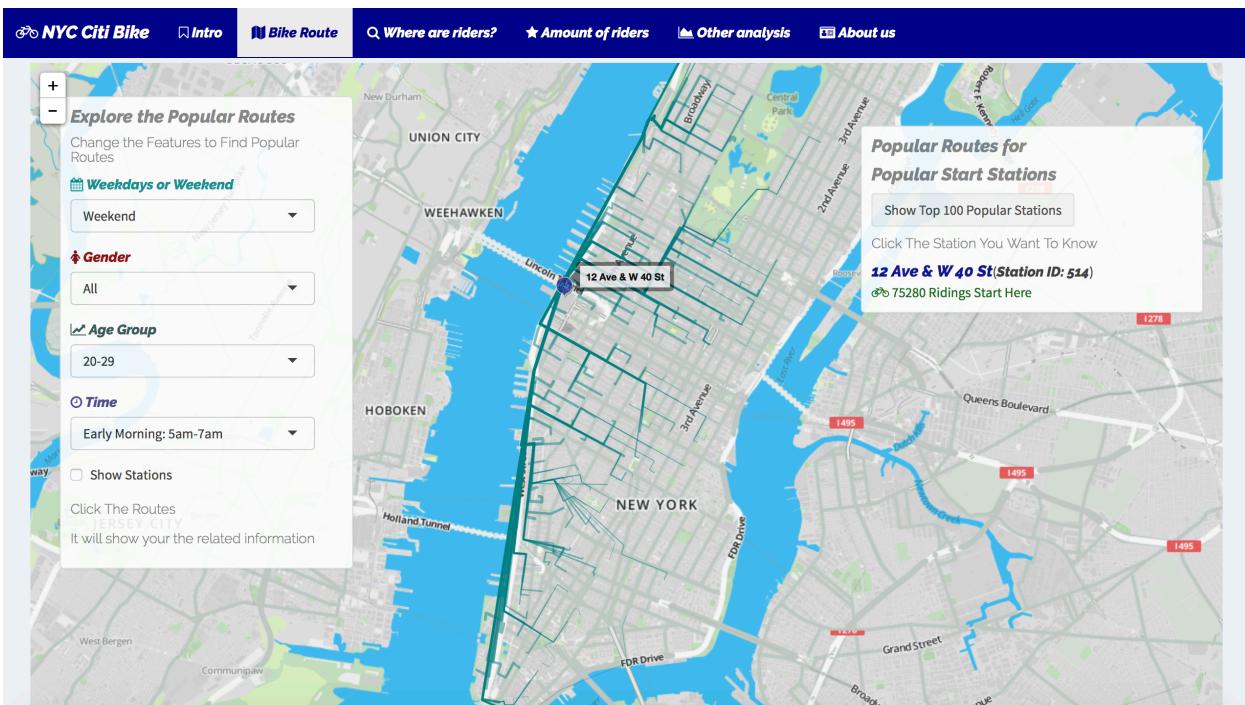
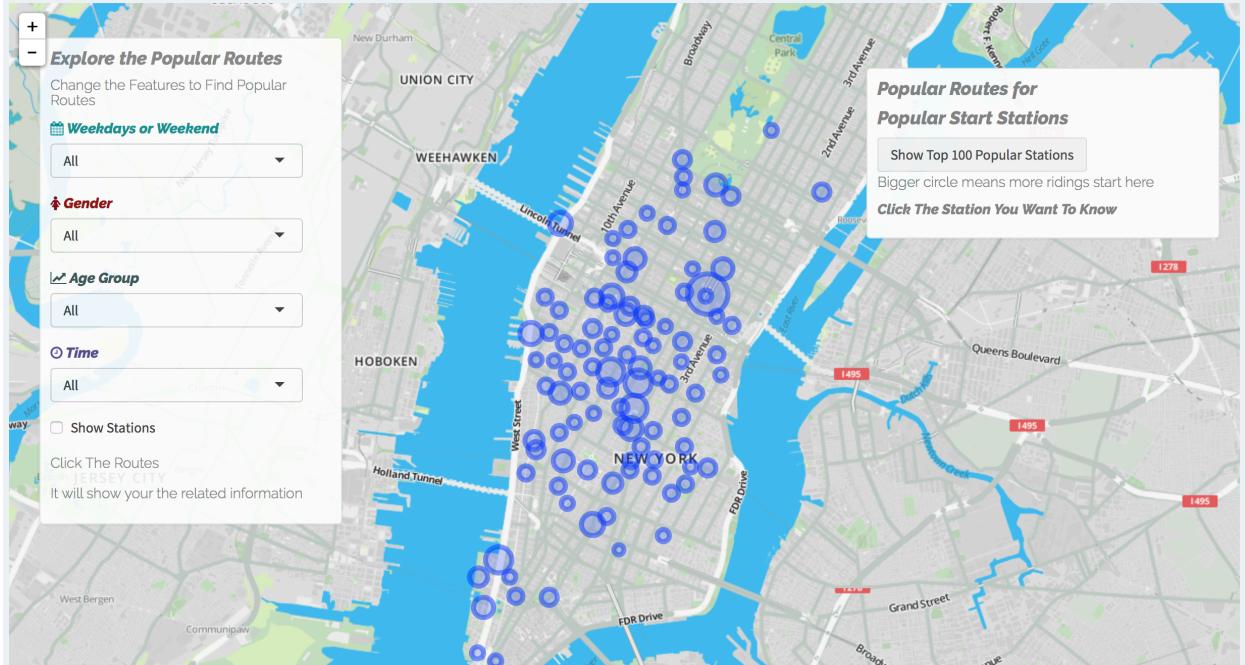
Team members: Hongyang Yang (hy2500), Pinren Chen (pc2751), Siyi Tao (st3036), Xin Gao (xg2249), Xin Luo (xl2614)

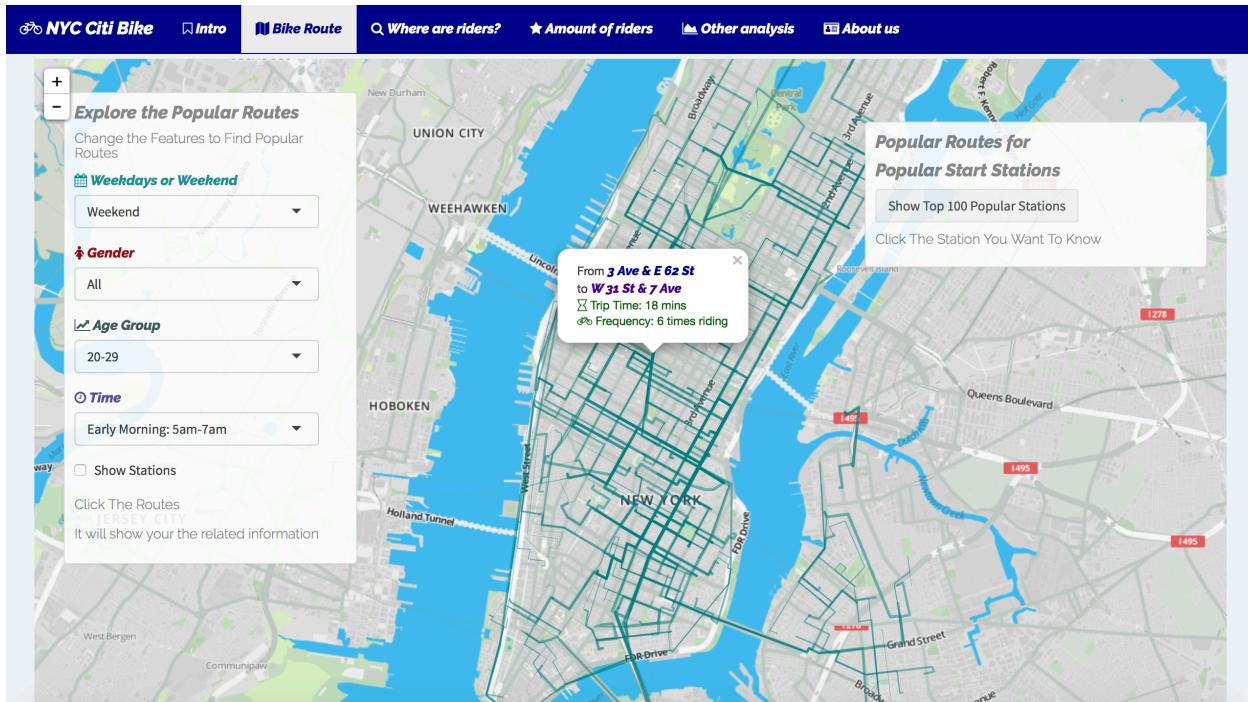
CitiBike, New York City's bike share system, is extremely popular among New Yorkers. We built an RShiny app to visualize the CitiBike's historical rent data, recommend routes for riders, simulate riding every day based on data in 2016. At the same time, we provide some insights between the weather and the number of rents per day according to CitiBike's historical rent data from 2013 to 2017. These information can provide insights about customer consumption habits for Citi Bike company and help to optimize company resources.

This is the main page of our RShiny App, which shows a simple summary of our project and a general layout of our app.

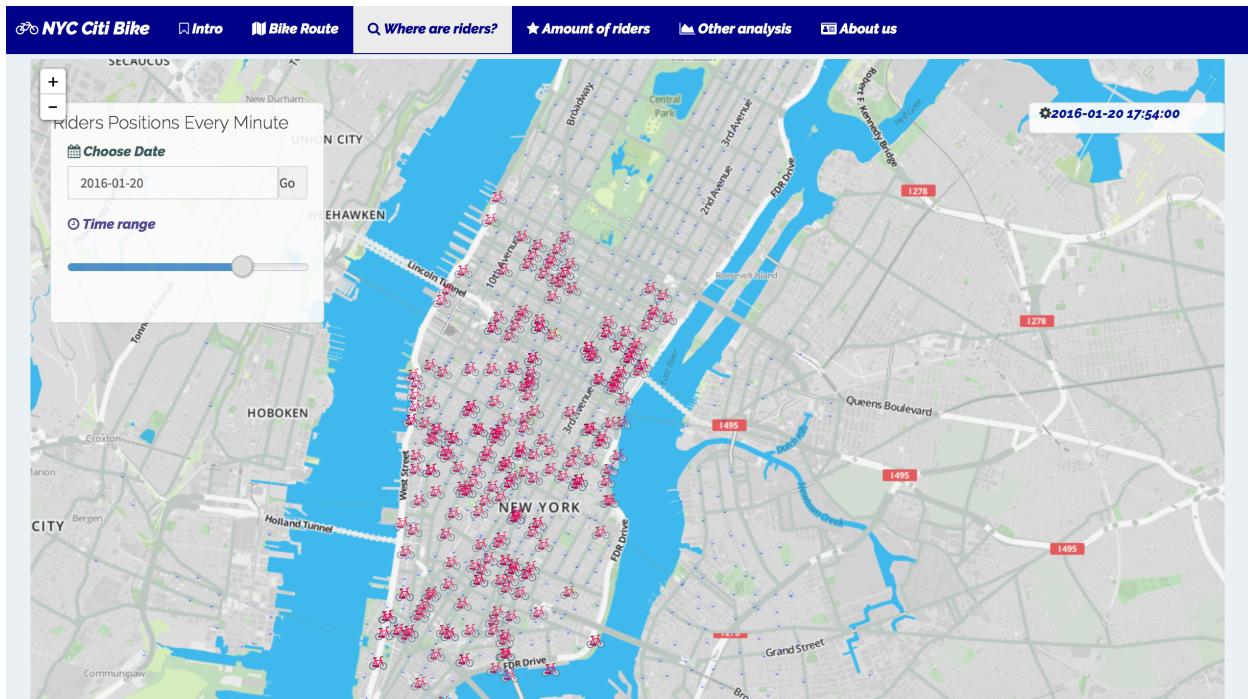
The screenshot shows the main interface of an RShiny application for NYC CitiBike. At the top, there is a dark blue header bar with white text and icons. From left to right, the menu items are: "NYC Citi Bike" (with a gear icon), "Intro" (with a person icon), "Bike Route" (with a map icon), "Where are riders?" (with a magnifying glass icon), "Amount of riders" (with a star icon), "Other analysis" (with a chart icon), and "About us" (with a document icon). Below the header, the main content area has a light gray background. In the top-left corner of this area, the text "Project 5: an RShiny app for NYC CitiBike" is displayed. To the right of this text, there is a large, semi-transparent watermark image of a person riding a CitiBike. On the left side of the content area, there is a sidebar with the title "Content" and a list of links: "Bike Route", "Where are riders?", "Amount of riders", "Other analysis", and "About us".

The second part of our App is called Bike Route, this interactive map shows the popular route and popular station in NYC, we can adjust those features to explore the routes, we can also see the top 100 popular stations. And after clicking one specific station, you can find details about the most popular 300 routes starting from this specific station. The deeper, the more overlap and the thicker, the higher frequency.

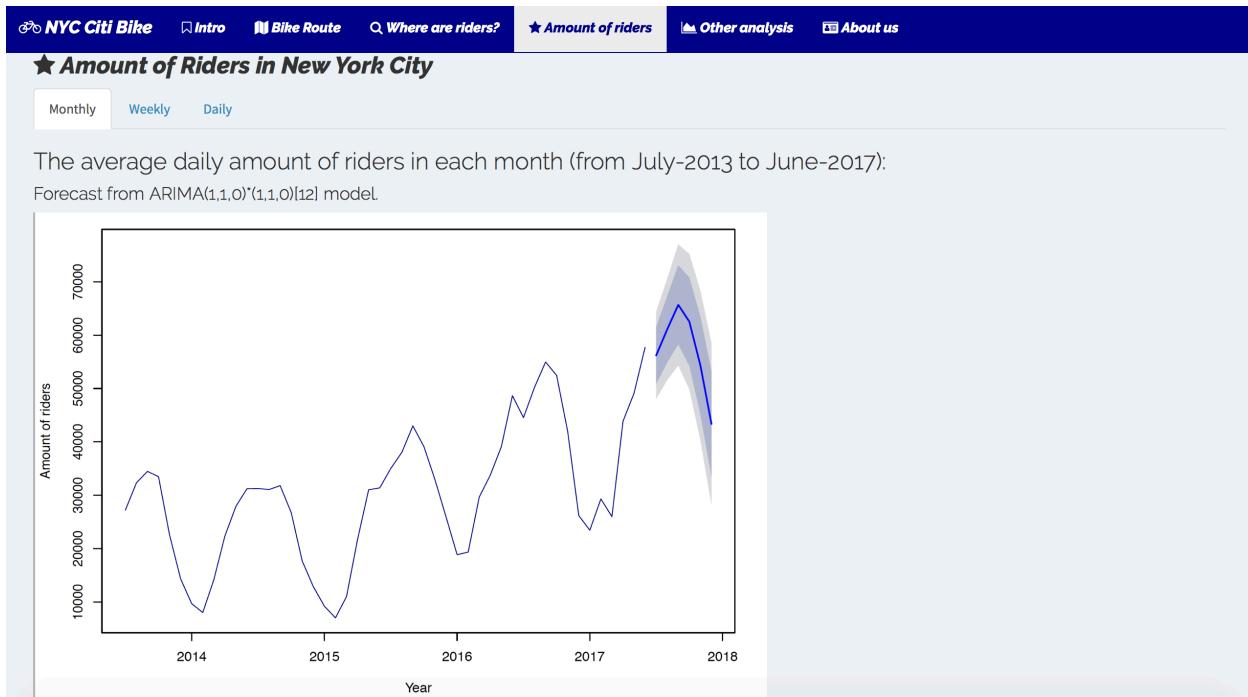




The third part of our App is called “Where are riders”. This map, every bike icon on the map represents a single rider, and the small blue dots represent all the bike stations scattered through the city, in this way we can simulate if the rider goes from station A to station B.



The fourth part of our App is called “Amount of riders”. First we plot all the time series data, x is monthly data, y is amount of riders on that month, We found a pretty clear pattern that during winter time here the business starts to decease and reach to a minimum in that year and during summer time it reaches to a peak in that year. So we use a forecast model ARIMA to predict the latter half year of 2017 which we don’t have the data yet.



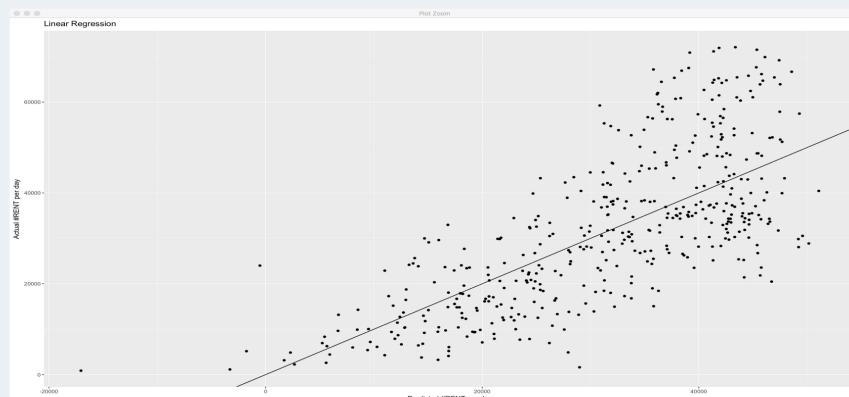
We answered some basic questions via a lot of fancy plots, like when do they ride, what days of the week are the most rides taken on, and so on. We still have a question that is there a strong relationship between the number of rents per day and the weather in that day? How strong is the influence? We run some machine learning models, linear regression, random forest, and GBM, those 3 are picture to show test y and predicted y, it looks Ok, we can basically say that the number of rent is related to the weather of that day. Next we want to see which factor in the weather data has the biggest influence. The random forest importance table and GBM influence table both tells us the TMAX-which is the maximum temperature of that day is the most important. We also want to see how strong is the influence. By running a regression tree, cross-validation, it can us that when the maximum temperature of that day is lower than 60 degree F, that's 15 degree C, rents start to decrease significantly. So the business owner can use the maximum temperature of the future days to adjust bikes.

## ★ Explore the Influence of Weather on the Amount of Riders

Step 1 Step 2 Step 3

Question 1: Is there any relationship between the weather and the amount of riders?

Linear Regression: adjusted R-square: 0.55



Random Forest

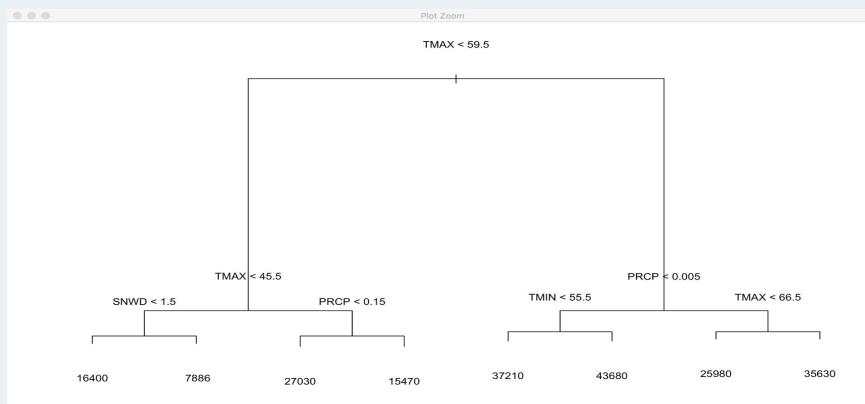


## ★ Explore the Influence of Weather on the Amount of Riders

Step 1 Step 2 Step 3

Question 3: How strong this factor influence the amount of riders?

Regression Tree



Regression Tree with CV

The fifth part of our App is called “Other analysis ”. It shows subscribers'/customers' trip duration distribution of each month (2016) and analysis of users by gender and age.

