NYC Public Transport Analysis

Studying the relationship between Citi Bikes and the Taxi and Subway Systems

Vaibhav A Gadodia
Richa Pandey
Kristin Liu

Why?

The Citi Bike System was deployed in NYC in June 2013 and has since expanded to over 15,000 bikes and 1,000 bike stations

Citi Bikes offer a better user experience. Compared to a taxi or subway ride, they are:

- greener
- faster (sometimes)
- more fun to use
- cheaper

Are New Yorkers using Citi Bikes as a taxi and subway alternative?

Citi Bike App Splash Screen

Datasets

Citi Bike System Data

Trip histories (June 2013 - December 2019)

Rich information about trip start/end dates/times, locations and user demographics

TLC Trip Record Data

Taxi trip information (January 2012 - December 2019)

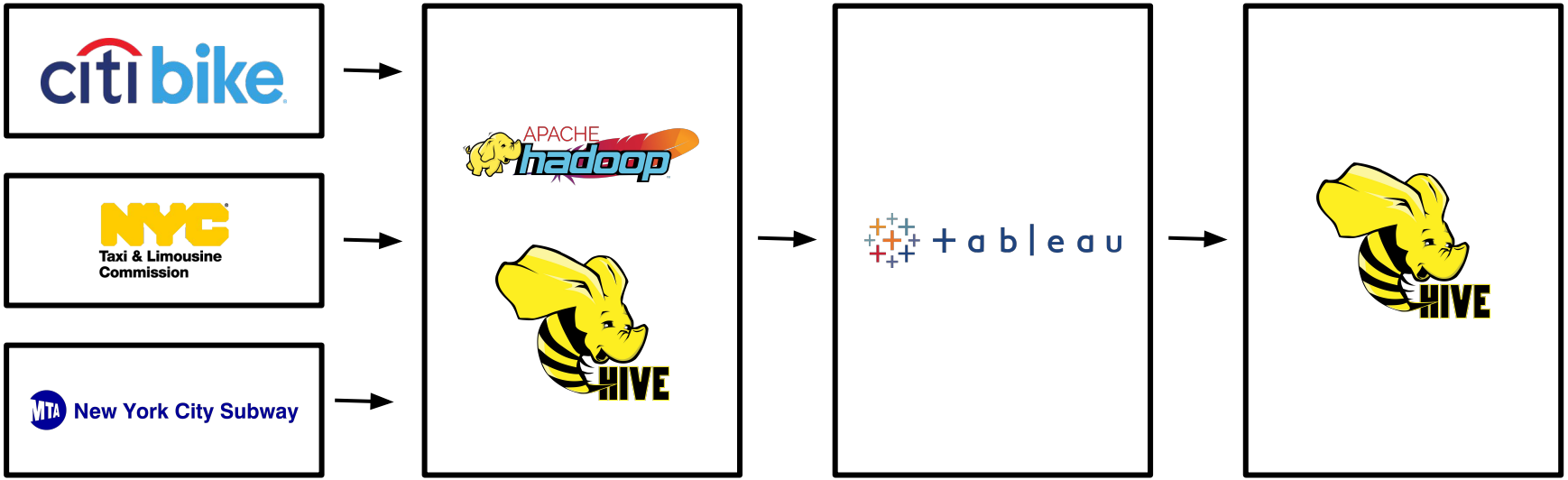
Fields such as pick-up/drop-off dates/times, locations, fare, distance, and passenger count

MTA Turnstile Data

Cumulative station entry/exit counts(January 2012 - December 2019)

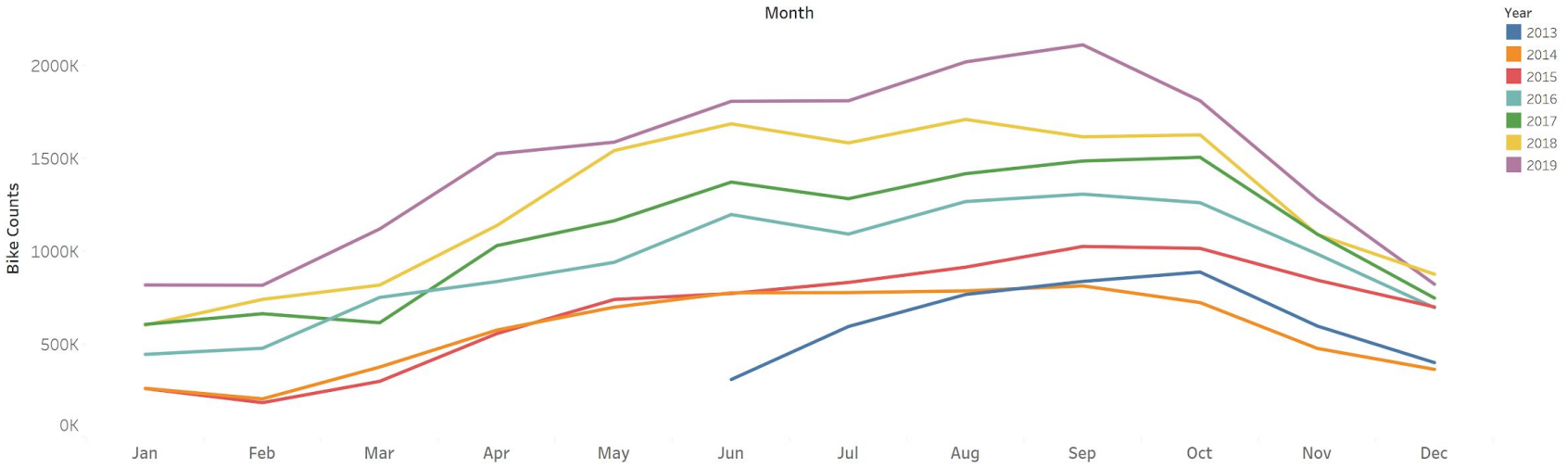
Other information like entry/exit date/time, station unit, subunit and linename

Pipeline


citi bike
NYC
Taxi & Limousine Commission
MTA New York City Subway
APACHE hadoop
HIVE
+ableau
HIVE
HDFS
ETL
Visualizations
Analysis

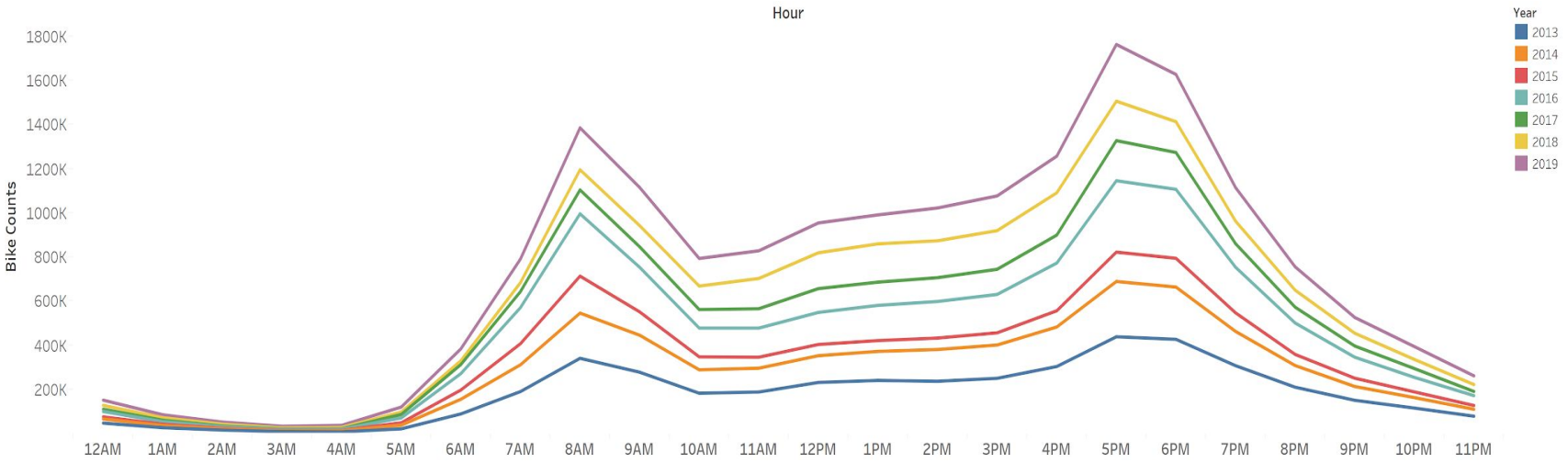
Trends

Citi Bike Usage - Monthly



Since introduction in June 2013, monthly Citi Bike usage has consistently increased

Citi Bike Usage - Hourly

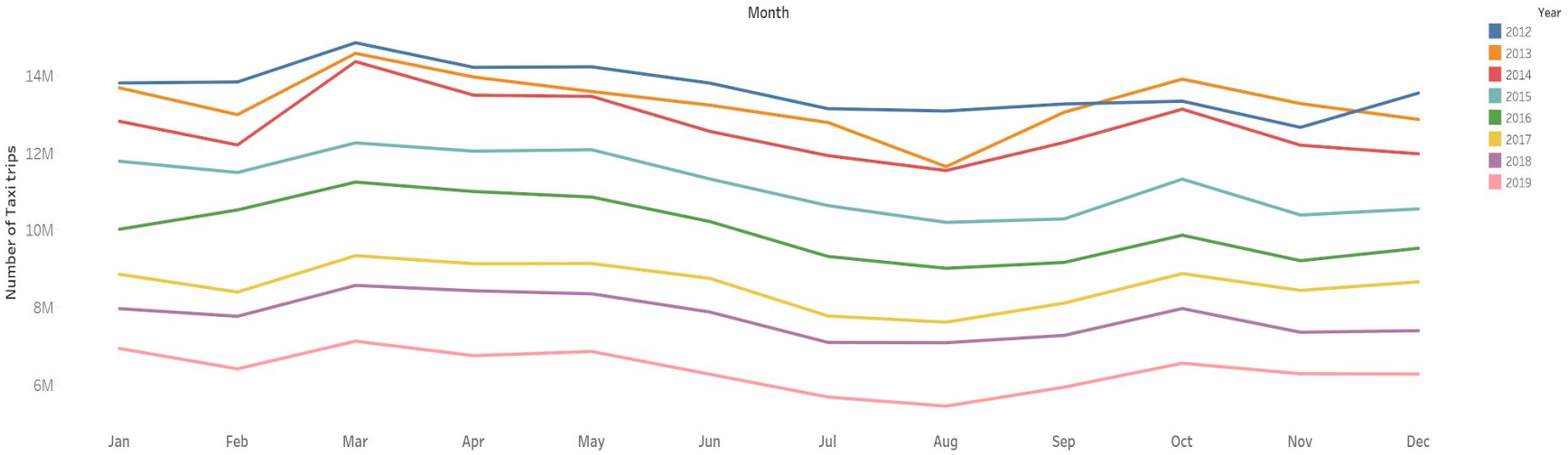


Citi Bike System's hourly usage has also shown consistent growth over the past seven years

298.52%

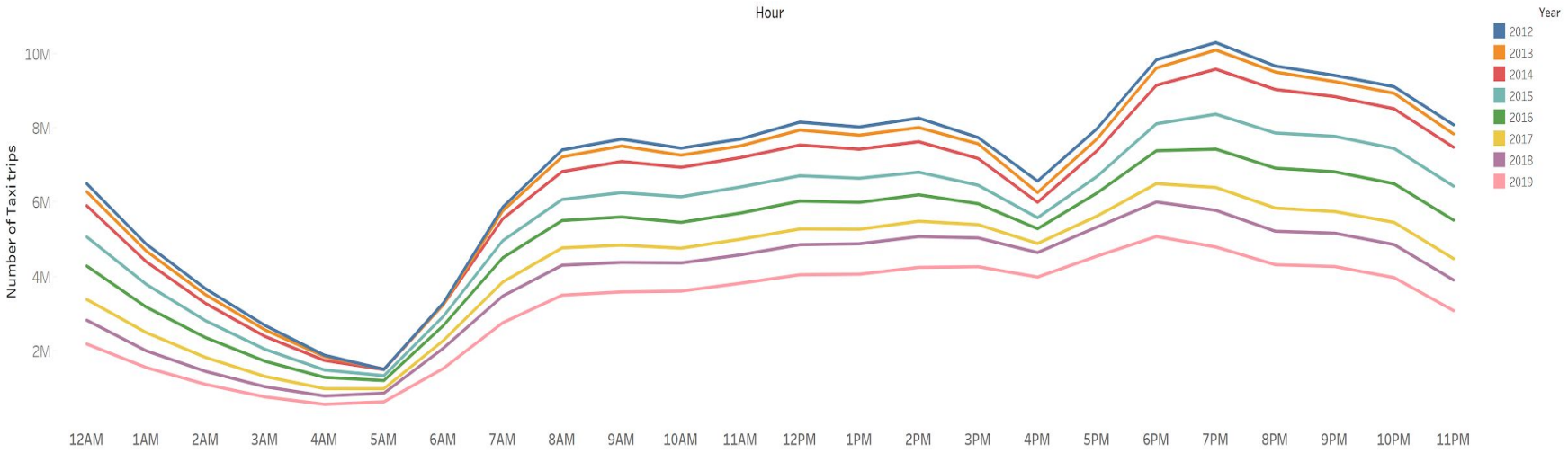
Change in yearly Citi Bike usage over the last seven years

Taxi Usage - Monthly



Taxi usage over the past eight years has declined significantly

Taxi Usage - Hourly

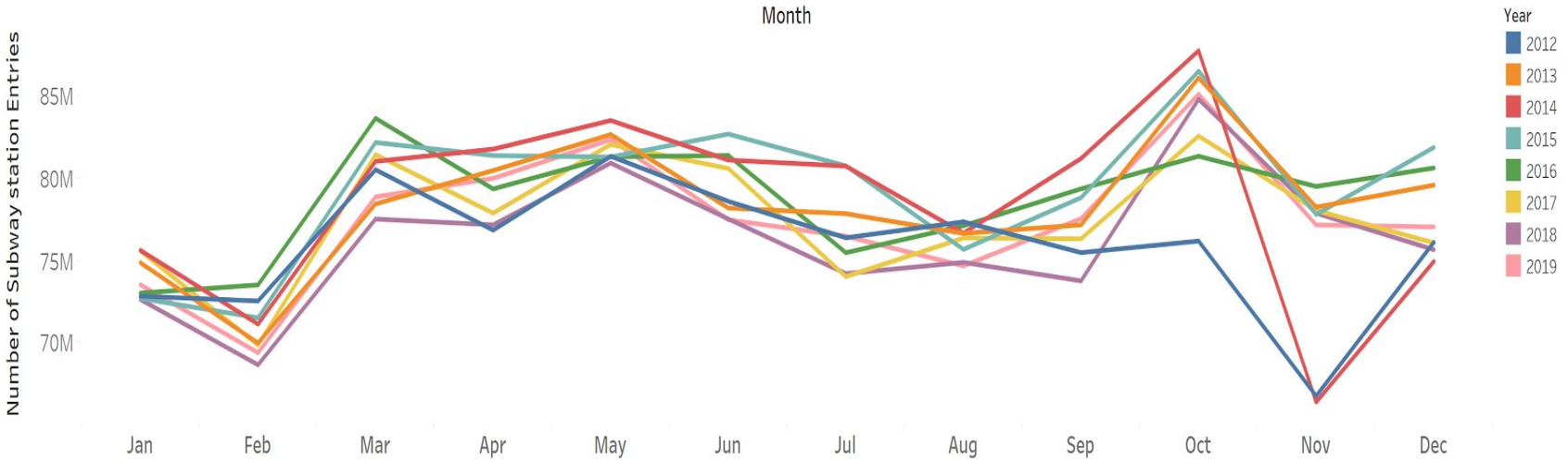


The number of hourly taxi rides have seen a consistent declining pattern

-53.28%

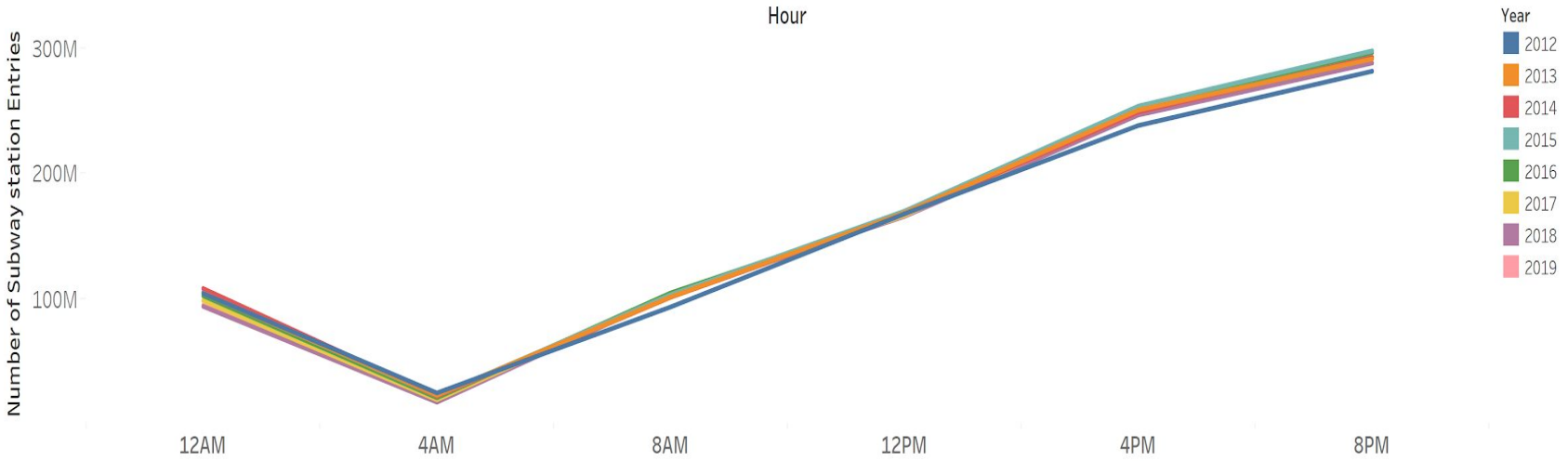
Change in yearly taxi usage over the last eight years

Subway Usage - Monthly



The average Subway usage has been mostly stable in the last few years

Subway Usage - Hourly

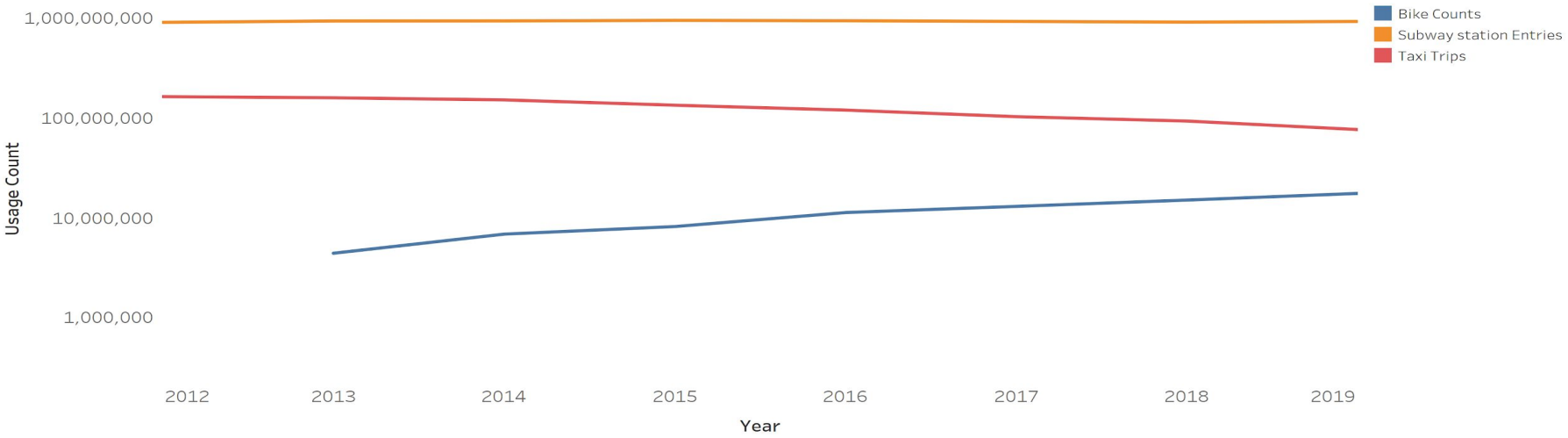


The hourly Subway usage has also been at a consistent level over the last eight years

2.05%

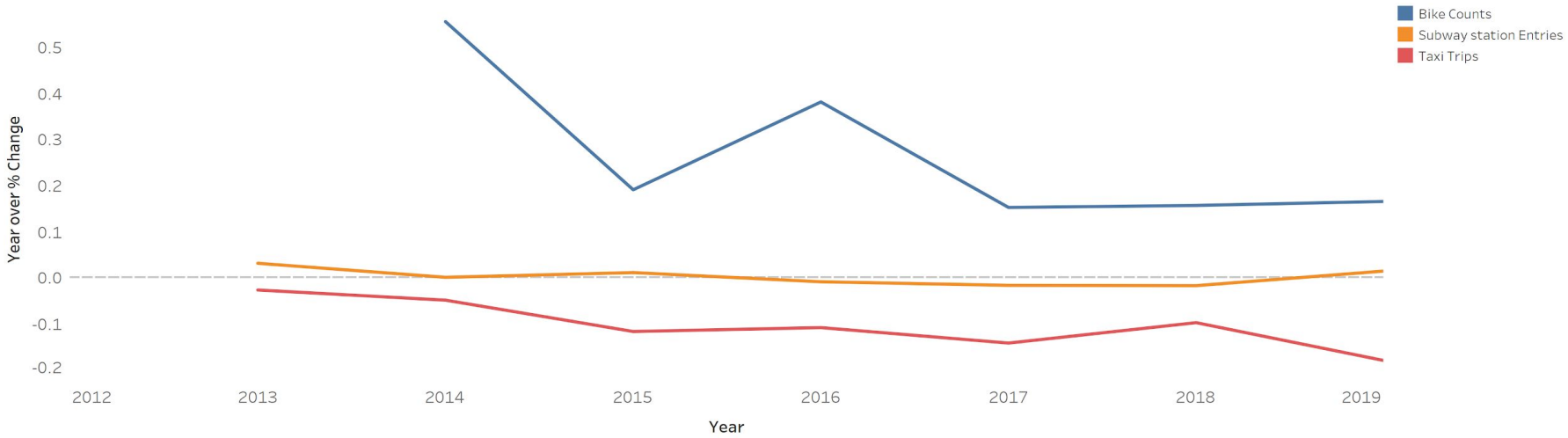
Change in yearly subway usage over the last eight years

Total Yearly Trips



Citi Bike trips have increased while taxi trips have gone down. Subway trips have remained at the same level

Yearly Growth Rate



Since Citi Bike's introduction, taxis have displayed negative growth as bike usage has consistently grown year over year. The subway system has mostly held its own

Since June 2013,

- Citi Bike usage increased
- Taxi usage declined
- Subway usage remained stable

Bikes and Taxis

Methodology

Pearson Correlation Coefficient

Also known as Pearson's r, this statistic measures linear correlation between two variables

Rubin Causal Model

Is an approach to the statistical analysis of cause and effect based on the framework of potential outcomes

Pearson Correlation Coefficient

The coefficient varies from -1 to 1

- -1 shows perfect negative correlation
- 0 shows no correlation
- +1 shows perfect positive correlation

$$r = \frac{\sum \left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{\sqrt{\sum \left(x_i - \bar{x}\right)^2 \sum \left(y_i - \bar{y}\right)^2}}$$

The coefficient for bikes and taxis is -0.975

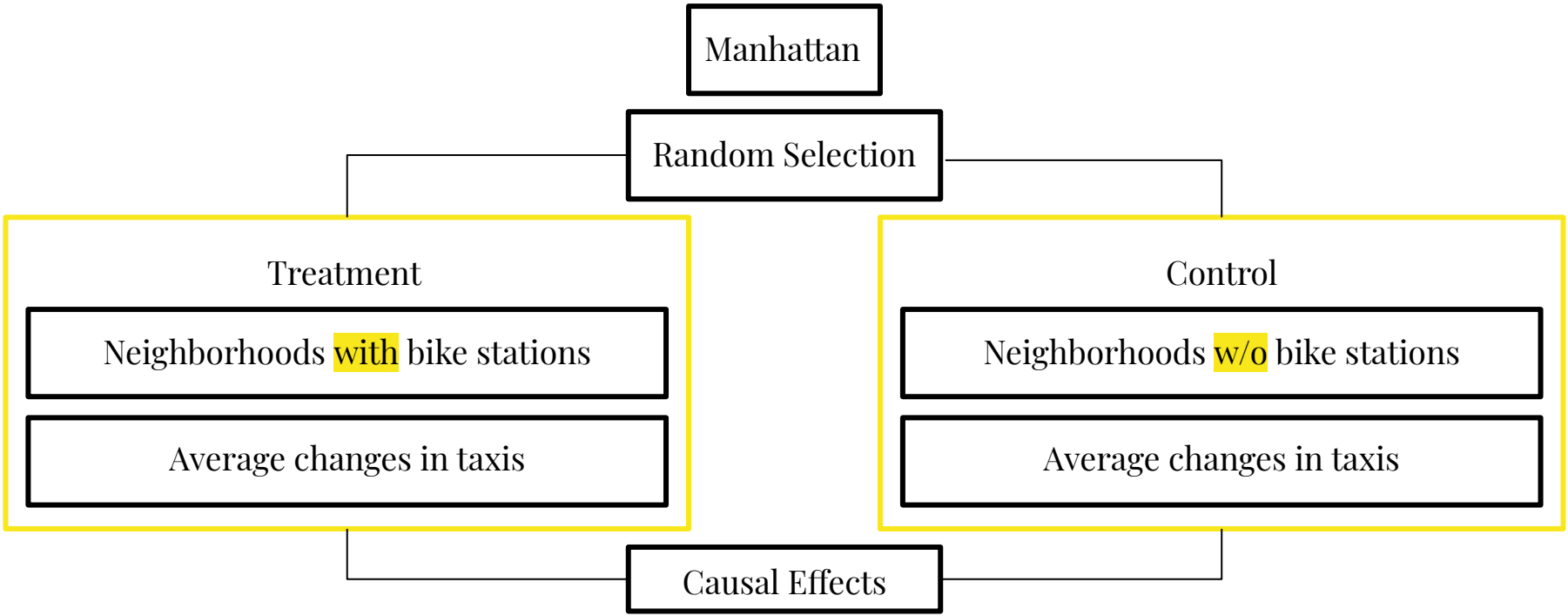
This is an indication of very high negative correlation

Rubin Causal Model - Overview

This model is based on the idea of potential outcomes

- Unit-level causal effect cannot be directly observed
- Randomized experiments allow the observation population-level causal effects
- An estimate of average causal effects can be obtained by computing the difference in means of treatment and control samples

Rubin Causal Model - Operational Workflow



Rubin Causal Model - Findings

	Avg. % change in taxi trips over 2 years
Warmer Months	-13.12%
Cooler Months	-12.53%

The causal effects for bike and taxi, in both warm and cold months are negative

Negative causal effects indicate that introduction of Citi Bikes in a neighborhood leads to more decline in taxi usage than neighborhoods without Citi Bike stations.

Takeaways

- Pearson's r of -0.975 indicates high negative correlation between Citi Bike and Taxi usage

- Along with correlation, we also find indications of a causal relation between Citi Bikes and Taxi usage

- Rubin Causal Model findings show that Citi Bikes accelerate Taxi usage decline

Citi Bike System has caused NYC Taxi usage to decline

Bikes and Subway

Methodology

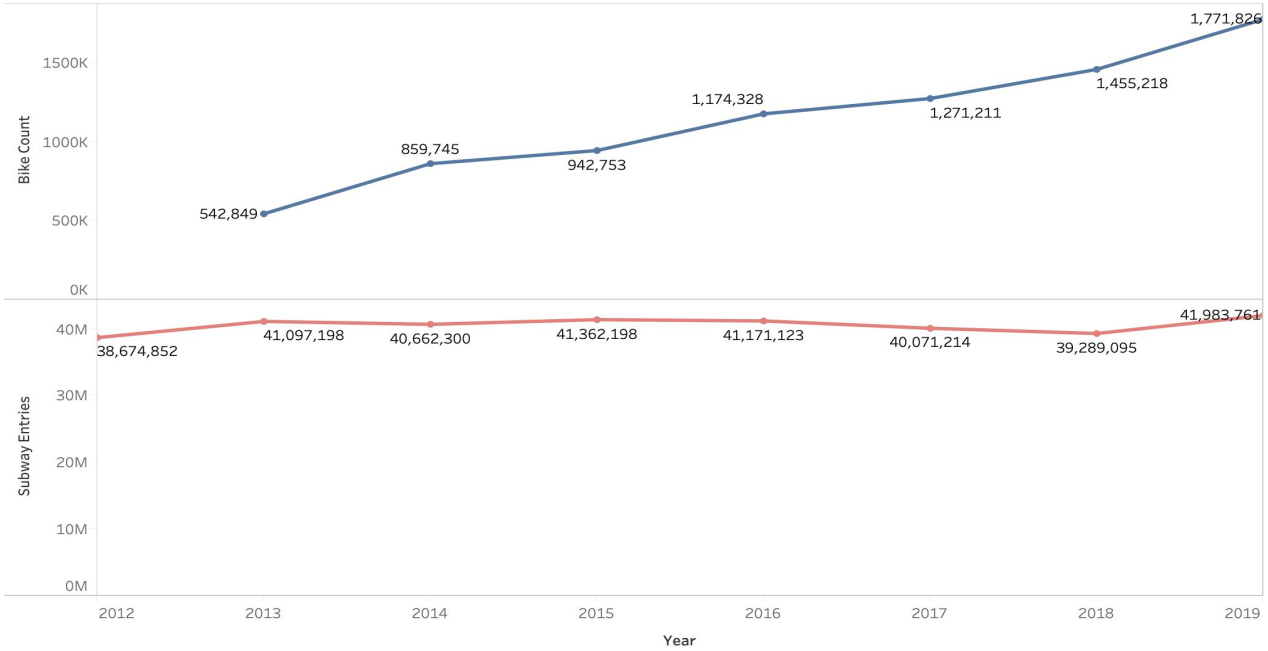
Correlation study for areas with low Subway availability

- Three neighborhoods with large average distance to subway stations were considered

- Yearly trip counts were visualized and yearly percentage change computed to study both substitute and complementary relation between Citi Bikes and Subway

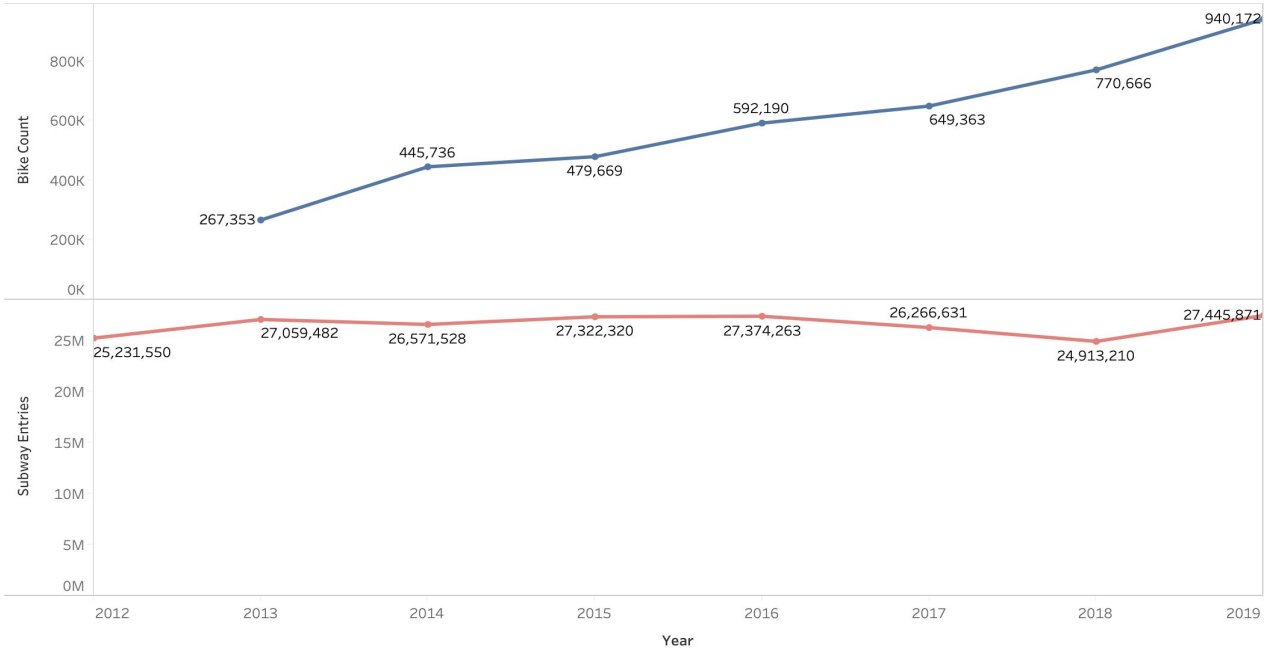
- Pearson Correlation Coefficient was calculated

Visualizations - Substitution



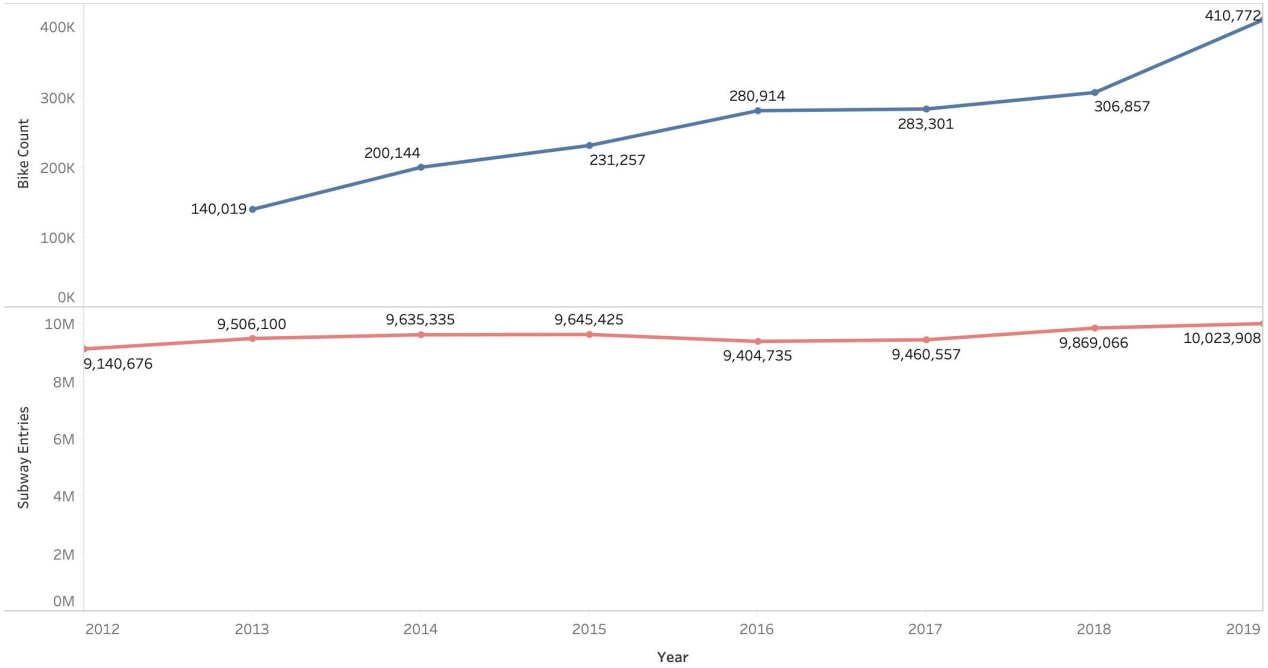
Subway usage in the three areas was mostly stable while Citi Bike usage increased

Visualizations - Substitution



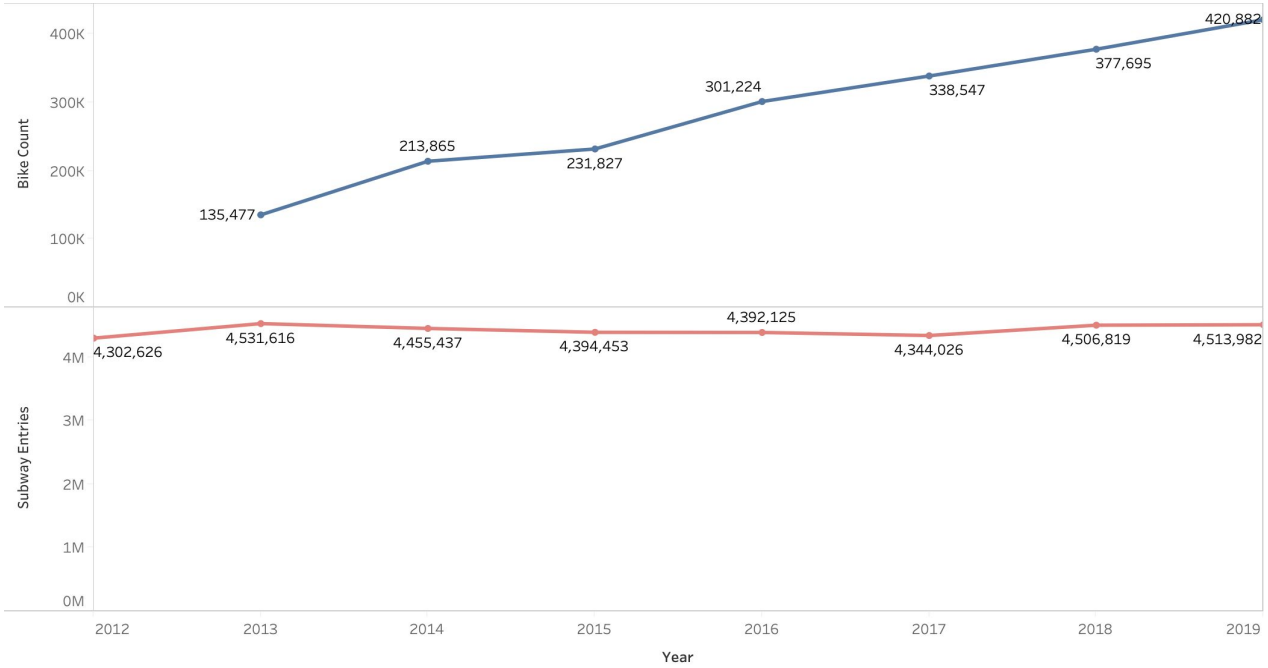
Subway usage in this area was mostly stable while Citi Bike usage grew consistently

Visualizations - Substitution



Subway usage in this area remained almost constant as Citi Bike usage rose

Visualizations - Substitution



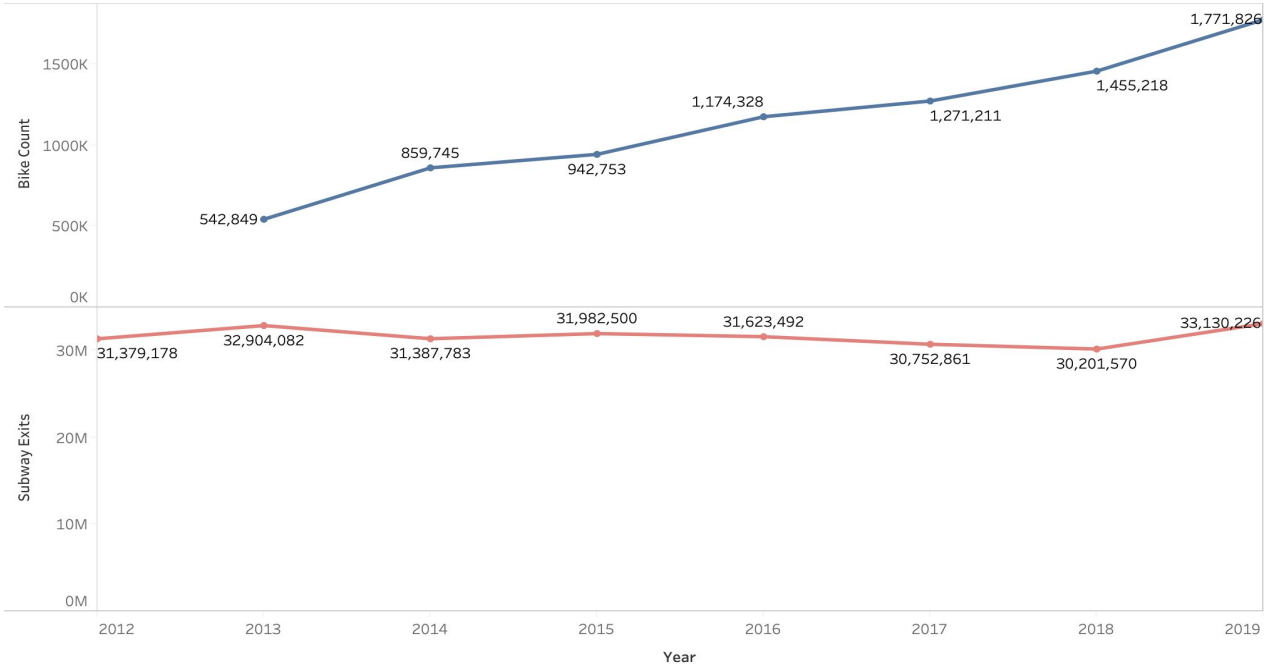
Subway usage in this area also remained stable while bike usage grew

Pearson Correlation Coefficient - Substitution

The coefficient for bikes and subway for the substitution study is 0.09

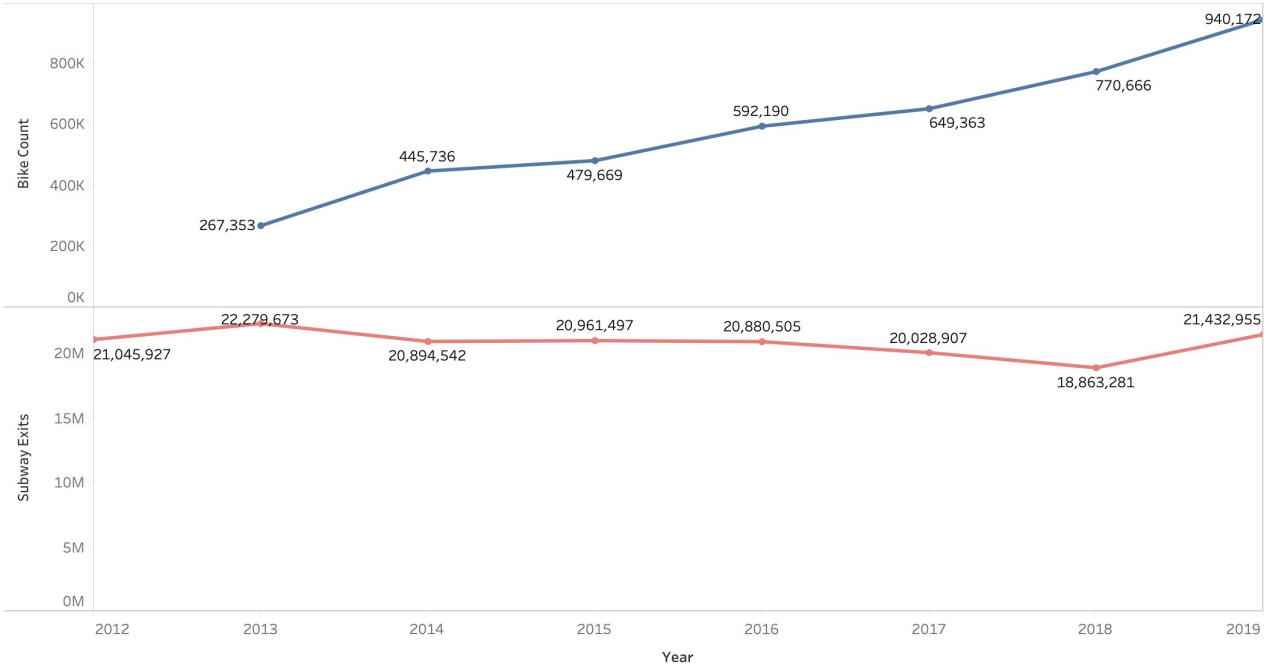
This indicates minimal to no correlation between Citi Bike usage and subway entries in the selected neighborhoods

Visualizations - Complementary



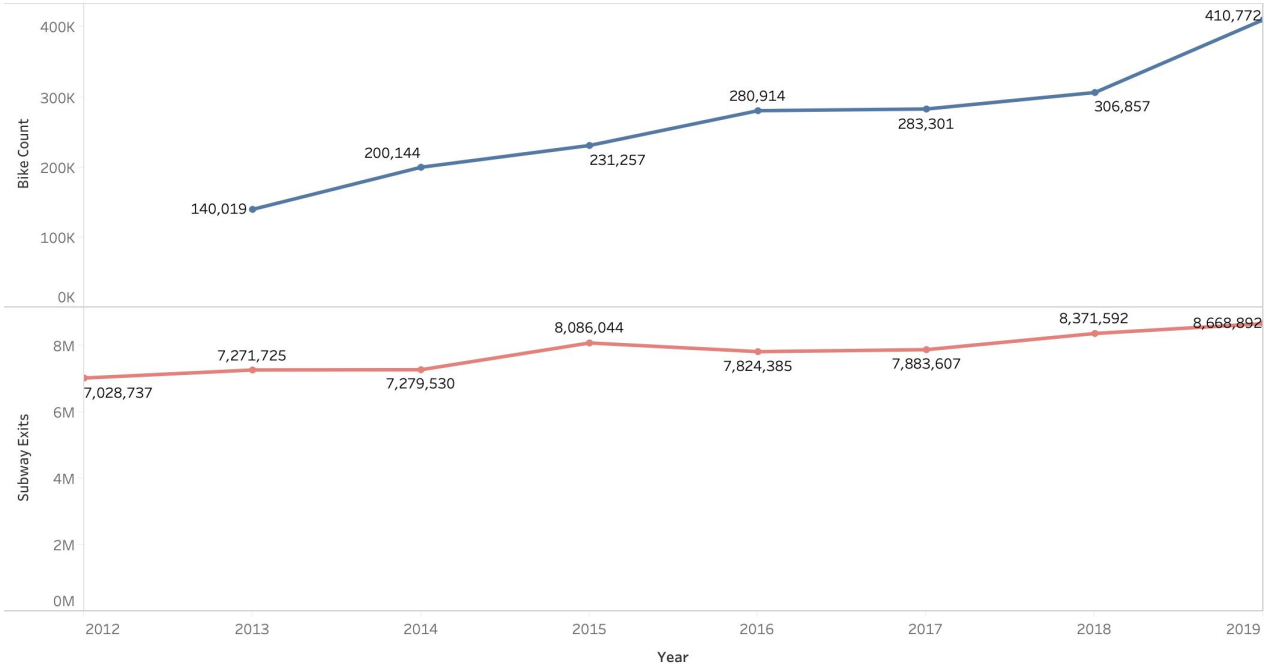
Subway usage in this area appears to be mostly stable while bike usage grew

Visualizations - Complementary



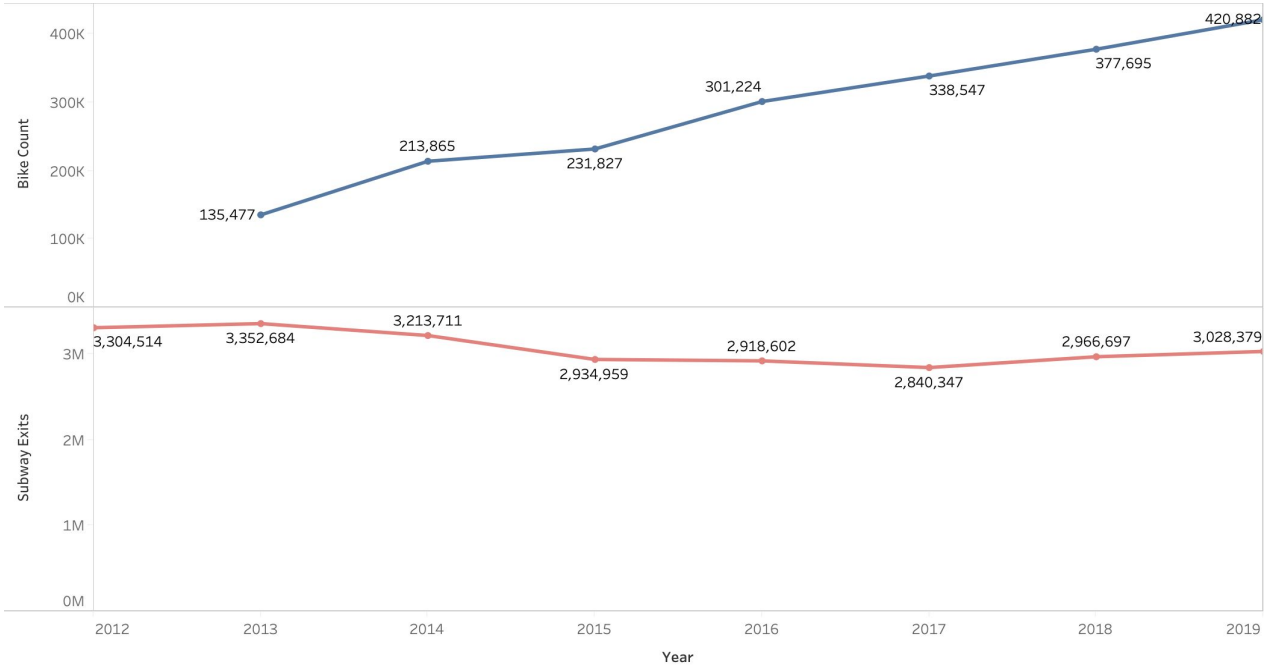
Subway usage in this area also displayed similar patterns of no significant growth

Visualizations - Complementary



Subway usage in this area grew at a minimal rate

Visualizations - Complementary



Subway usage in this area dropped minimally but mostly the trend remained stable

Pearson Correlation Coefficient - Complementary

The coefficient for bikes and subway for the complementary study is 0.26

This indicates a very weak positive correlation between Citi Bike usage and subway exits in the selected neighborhoods

This could be an indication that the Citi Bikes act as a complement to the Subway system in a limited capacity

Takeaways

- Pearson's r values of 0.09 and 0.26 indicate no correlation between Citi Bikes and Subway in the substitution case and a very weak positive correlation in the complementary case, respectively

- Overall, the Citi Bike system does not seem to impact the Subway in any appreciable manner

Citi Bikes have a minimal relation with the Subway system

Summary

- Citi Bike System has impacted the Taxi usage in NYC significantly
- Areas with Citi Bike stations see an increased decline in Taxi usage compared to areas with these stations
- The Subway system has mostly remained stable over the years and has not been influenced too much by Citi Bike usage both as an alternative and a complementary mode of transport

Thank You!