# Abstract

How effective is the Earth Day marketing campaigns in New York? Earth Day is an internationally recognized as a day to promote green initiatives. Effective marketing and campaigns related to Earth Day should see an increased usage of the New York Transit system and other greener initiatives as to replace car usage. I used data from the [MTA](http://web.mta.info/developers/turnstile.html) and [Citi Bike](https://s3.amazonaws.com/tripdata/index.html) to model usages of both to analyze the effectiveness of the promoted greener transportation methods. The models suggest that current campaigns do not effectively promote greener transportation methods in New York.

# Design

This project could potentially be used by the New York City Government and other relevant Earth Day Organizers. By modeling this information we can hope to see how effective campaigns are in promoting Earth Day initiatives and further promote greener alternatives to driving. They will be able to more effectively allocate resources based on past turnouts to prior initiatives in order to better know what initiatives are effective.

# Data

There are 2 different datasets used. The MTA data looks at individual turnstiles for 379 stations. Each row of the data represents C/A (Control Area), Unit (Remote Unit for a Station), SCP (Subunit Channel), Station (Station Name), Linename, Division, Date, Time, DESc, Entries, and Exits. Entries were a cumulative entry at approximately 4 hour intervals. Daily Entries that I used in my modeling were found by subtracting the previous day’s entries from the current days entries. The information was then aggregated by stations and days to display the number of entries per day at a station.

The Citi Bike displays each individual ride as ride Id, rideable\_type, started\_at, ended\_at, start\_station\_name, start\_station\_id, end\_station\_name, end\_station\_id, start\_lat, start\_lng, end\_lat, end\_lng, and member\_casual. Each row in the dataset represents a bike ride that was started as such we can obtain daily ride counts by adding all the rows that started on a particular day.

# Algorithms

* MTA data was cleaned for any duplicates
* There were 37721 rows that had 0 entries and 0 exits logged that were removed.
* 77 rows had recovered audit tags. These rows happened in the middle of the day so they were ignored as we took the end of day cumulative entries.

# Tools

* Raw data was put into a SQL database and querying was used to preform initial analysis of the data and cleaning, such as removing the rows that had 0 entries and 0 exits.
* Exploratory data analysis in pandas
* Data visualized using matplotlib

# Communication

Findings were presented visually in slides.