# Hypothesis

I was interested in a season that would be popular for bike riding and then I was wondering if the data would show a peak before covid, drop during 2020, and if it would equate or exceed popularity after vaccines were made available.

# Data Source

Exported the below CSV files from <https://s3.amazonaws.com/tripdata/index.html>

### 2019

* 201903-citibike-tripdata.csv
* 201904-citibike-tripdata.csv
* 201905-citibike-tripdata.csv

### 2020

* 202003-citibike-tripdata.csv
* 202004-citibike-tripdata.csv
* 202005-citibike-tripdata.csv

### 2021

* 202103-citibike-tripdata.csv
* 202104-citibike-tripdata.csv
* 202105-citibike-tripdata.csv

### 2022

* 202203-citibike-tripdata.csv
* 202204-citibike-tripdata.csv
* 202205-citibike-tripdata.csv

# Data Cleaning

### 2019\_Spring\_Clean.ipynb, 2020\_Spring\_Clean.ipynb

* Removed columns 'starttime', 'stoptime', "start station id", "end station id", "bikeid" as they did not provide the information I cared to report on.
* Updated the ‘gender’ column to contain the string values opposed to integers
  + 0 = Other, 1 = Male, 2 = Female

### 2021\_Spring\_Clean.ipynb, 2022\_Spring\_Clean.ipynb

* Removed columns "ride\_id", "started\_at", "ended\_at", "start\_station\_id", "end\_station\_id" as they did not provide the information I cared to report on.
* A gender column was no longer available data.

### File merge.ipynb

* Added a ‘year’ column to each csv file and then merged similar data sets
  + 2019 & 2020
    - Renamed column usertype to member\_casual to match 2021 & 2022 df.
    - Replaced values in updated column member\_casual to match 2021 & 2022 df.
      * 'Subscriber' to 'member'
      * 'Customer' to 'casual'
    - Exported as new CSV
  + 2021 & 2022
    - Exported as new CSV

# Analysis

## '19 - '20 Avg Age of Gender

* Males predominantly use the bikes.
* Men and Women averaging between their 40s and 50s are using the bikes.
* “Other” genders in their early 50s use the bikes the least and are casual riders.
* Created calculated field “Age” for visual purpose, average age as integer instead of original float.
  + INT(AVG(2022-[Birth Year]))

### '19 - '20 Avg Age of Gender Detail

The average age range for men and women vary by about 10 years; 33-42 years of age.

“Other” have a smaller average age range of about 5 years; 47-52.

#### Males

* 91% of male users are members.
* Older males seem to sign up as members over casual users.
* Overall Gender detail: 66% of total users are men.

#### Females

* 84.6% of female users are members.
* Older females seem to sign up as members over casual users.
* Overall Gender detail: 25% of total users are female.

#### Other

* “Other” have typically been casual users at 83.4% in their early 50s.
* Overall Gender detail: Only 9% of total users are classified as “Other”.

## '21 - '22 Ride Type

* Over 70% of users are members
  + Most prefer the classic bike with electric being secondary