Assessing Citi Bike's Bike Speed

ROSS BOEHME

Overview

Background

- Rates of biking are **quickly increasing** in NYC post-COVID
- More biking -> **cleaner air, safety in numbers, healthier citizens**
- **Citi Bike**, NYC's bike share program, can accelerate this increase
- But I believe the **non-electric Citi Bike's speed** hurts the program
- **Speed improvements**: decrease weight (45 lbs), thinner tires, no basket, change from "cruiser" frame

The Goal

Evaluate Citi Bike data for 2022 to **determine if the non-electric Citi Bike's speed could be improved** to encourage more usage.

If it appears the non-electric Citi Bike's speed could be improved, **suggest performing a cost-benefit analysis**.

Approach

1. Determine **most common long Citi Bike trips** (10 min+)

2. **Compare trip times** between e-Citi Bike, non-e ("classic") Citi Bike, and Google Maps biking estimate

3. **Compare usage rates** between e-Citi Bike and classic. If the e-bike is much more popular, it could be due to its superior speed.

4. If the trip times are much slower for the classic Citi Bike than for the e-bike and Google Maps' estimates, and its usage rate is much lower, **speed is likely an issue -> suggest cost/benefit analysis**

Research Questions

1. How do electric and non-electric **Citi Bike travel times compare** to Google Maps' bike time estimates?

2. Despite the higher price ($0.26 more/minute) does it appear that Citi Bike **riders opt for the electric option** when given the choice between the two?

Data Overview

- Data from Citi Bike website and Google Maps API
- One observation per ride for 2022. 30.7M total observations.
- 14 Fields:

1. ride_id - Unique ride ID
2. rideable_type - Bike: "electric", "classic" (non-electric), "docked" (instant unlocking/relocking)
3. started_at - Time and date ride started
4. ended_at - Time and date ride ended
5. start_station_name - Typically in the format of cross streets e.g. "7 Ave & Central Park South"
6. start_station_id - 6 digit float giving unique ID to start station
7. end_station_name - formatted the same as start station name
8. end_station_id - 6 digit float giving unique ID to end station
9. start_lat - Trip starting latitude at station
10. start_lng - Trip starting longitude at station
11. end_lat - Trip starting latitude at station
12. end_lng - Trip starting longitude at station
13. member_casual - Binary of whether trip was taken by Citi Bike subscriber "member" or non-member "casual"
14. google_estimate_mins - Biking time est. from Google Maps based on start/end coordinates

Presentation Outline

1. Data Challenges, Acquisition, and Transformation
2. Exploratory Data Analysis
3. Research Question Analysis
4. Conclusions, Next Steps, and Limitations

Data Challenges, Acquisition, and Transformation

Data Challenges/Solutions

Challenges

- One row for each ride -> 30.7M rows -> Slow rendering and calculations
- Avg classic and electric Citi Bike trip times skewed by leisurely journeys and undocked bikes
- Short journeys may not reveal trip time disparities
- Google Maps API is only free for first 28,500 requests

Solutions

- Assess the most common trips (unique start/end combinations)
- Assess only long (10 min+ trips)
- Remove time outliers using IQR
- Double check queries beforehand to minimize test runs
- If needed, Leverage Google API free trial: $300.00 credit for first 90 days

Data Acquisition/Cleaning (1/3)

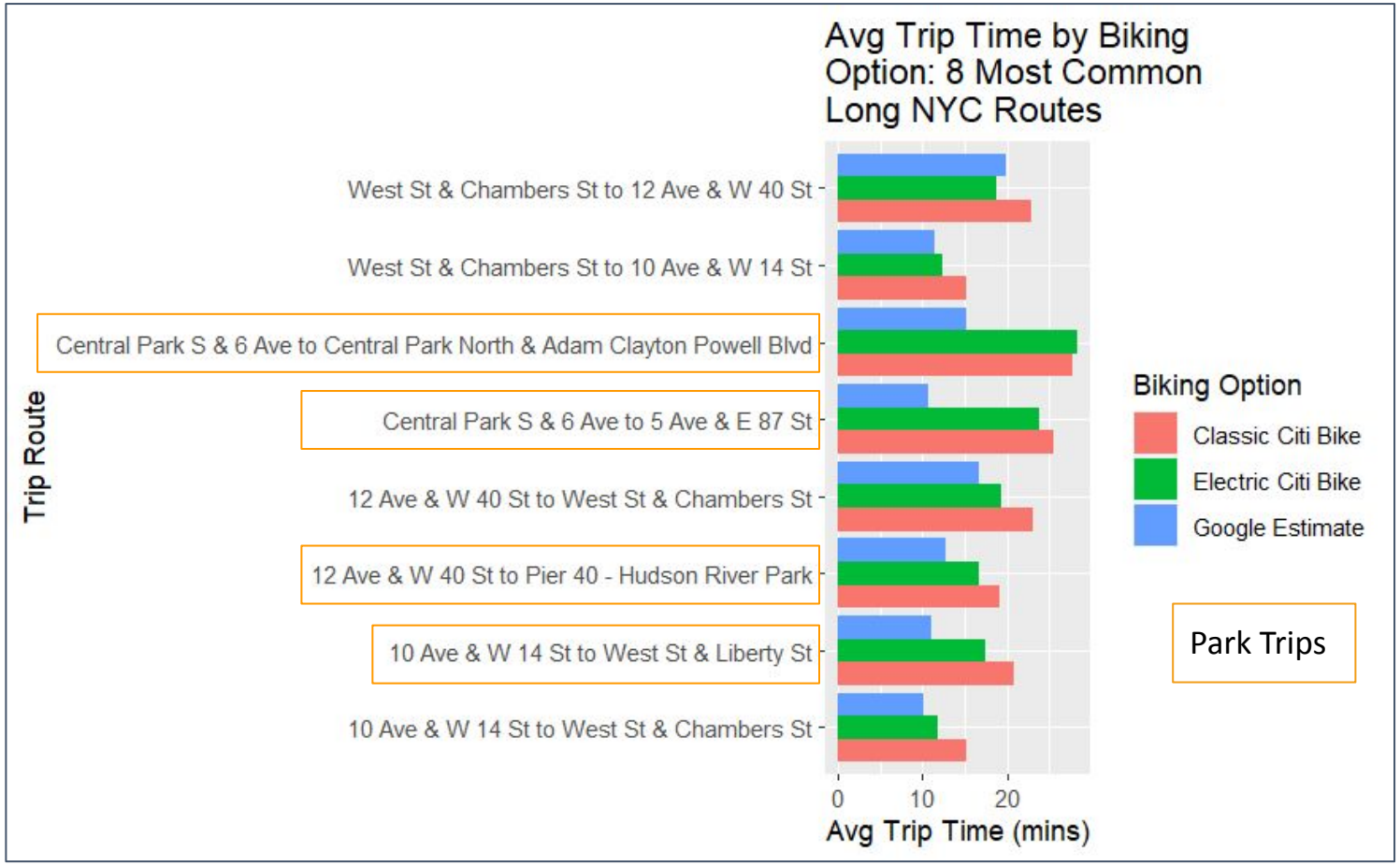
1. Use base R to acquire CSV from Citi Bike website
2. Bind 12 monthly dataframes (30.7M rows, 1 per ride)
3. Light cleaning: Fix station names, remove docked trips
4. Assess most common trips of 2022 which had duration of 10 min+ via Google Maps API

```
trip_count <- cb %>%  dplyr::filter(start_station_id != end_station_id ,trip_time >= 10

trip_count <- plyr::ddply(trip_count , c( 'start_station_name' , 'start_full_coord',
'start_station_id', 'end_station_name', 'end_full_coord', 'end_station_id' ), nrow)

top_trips <- trip_count %>% dplyr::arrange(desc(V1)) %>% top_n(8)

for (i in 1:nrow(top_trips)){
google_time[i] <- gmapsdistance(origin = top_trips$start_full_coord[i],
                                destination = top_trips$end_full_coord[i],
                                mode = "bicycling",key)$Time
}
```

Data Acquisition/Cleaning (2/3)

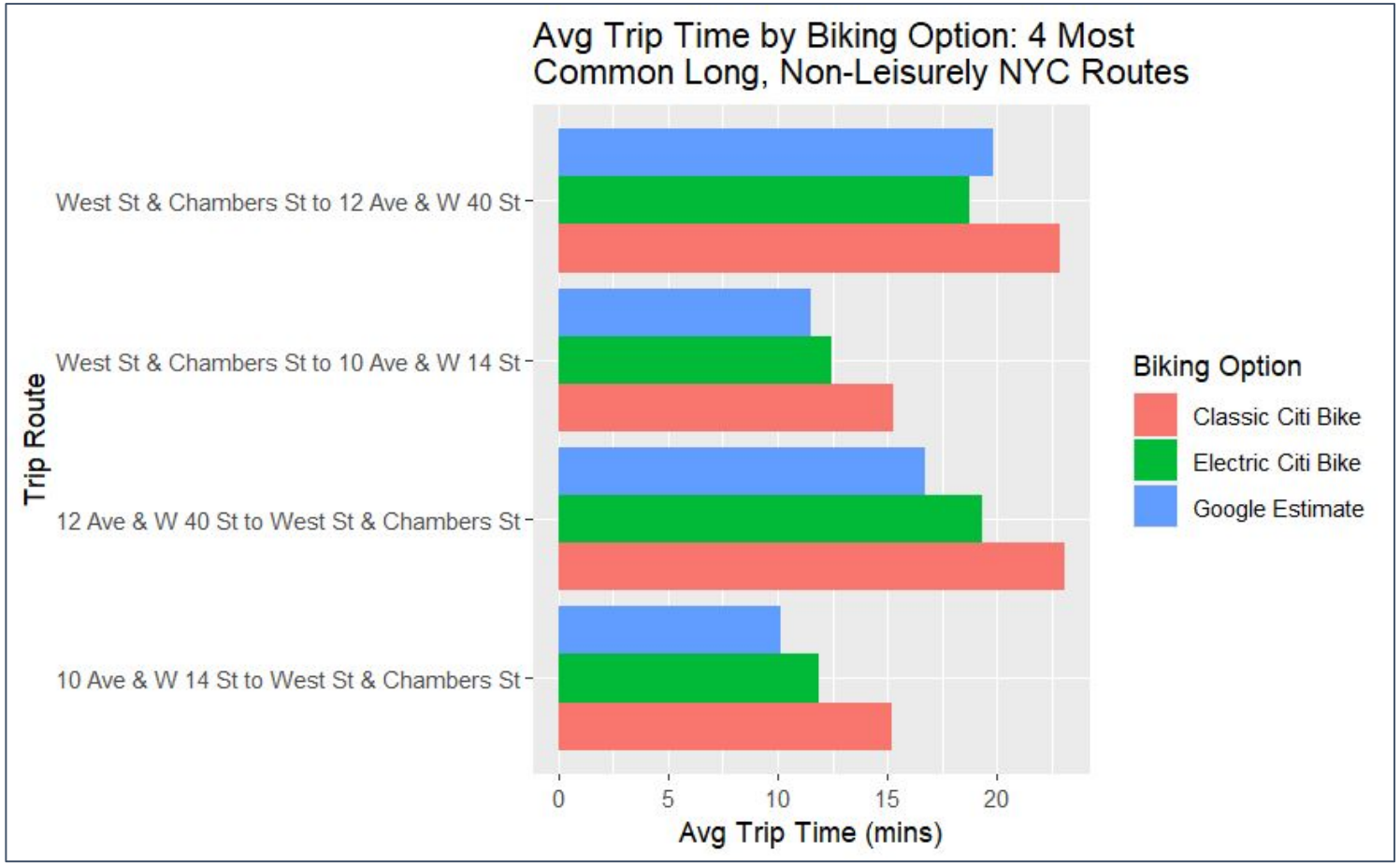
5. With Google Maps estimates calculated for 8 most common long trips, calculate average trip time (with outliers removed) per electric and classic Citi Bike

6. To account for trips where the avg time is skewed by leisure rides, remove trips where:
   a. Electric bike time > 1.3x Google estimate and
   b. Start or end at parks (Central, Hudson River)

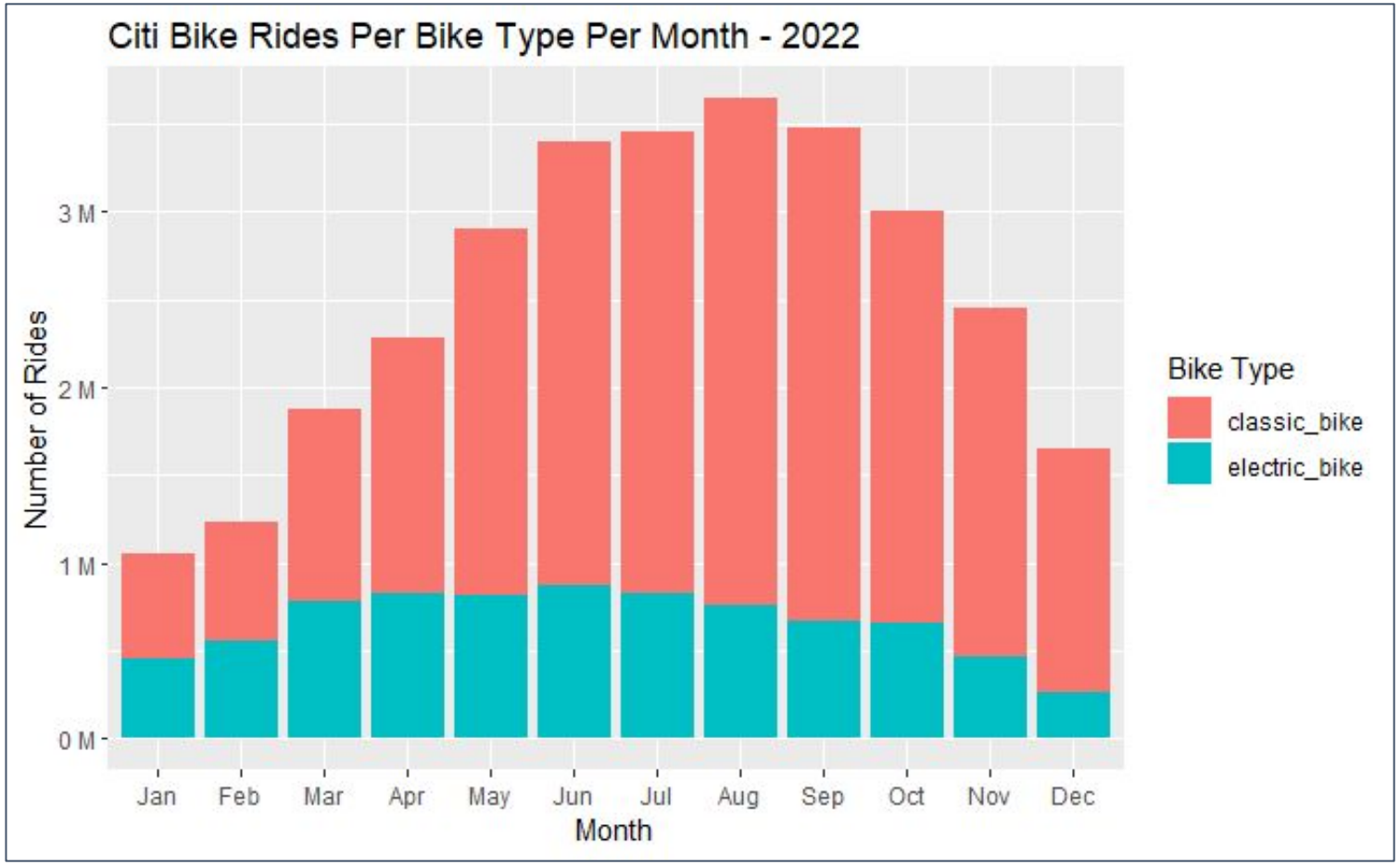


Data Acquisition/Cleaning (3/3)

7. Final time comparison df: 4 most common long, **non-leisurely** NYC Citi Bike trips for 2022, with avg trip times per electric and classic bike compared to Google Maps time estimate



Exploratory Data Analysis

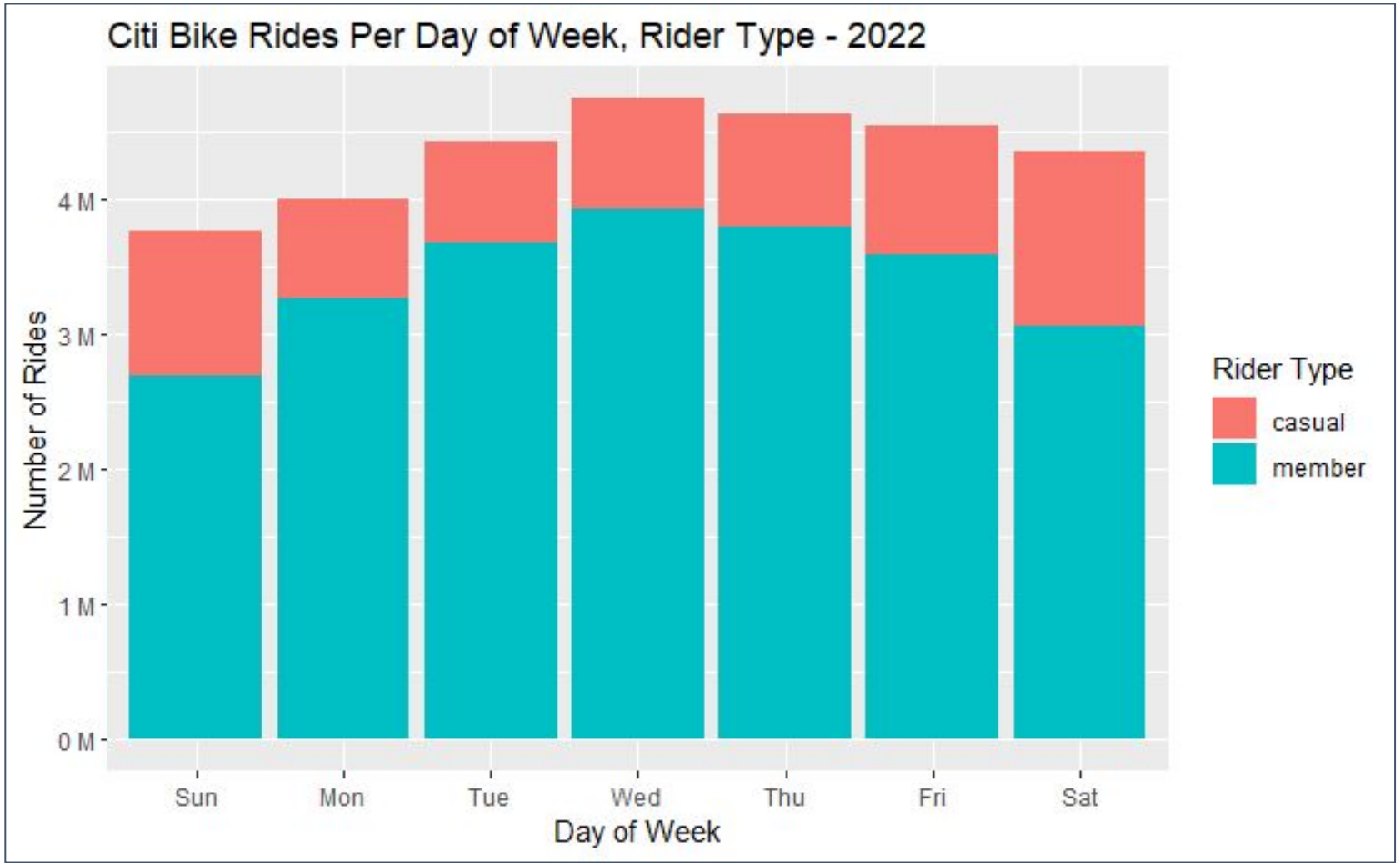


Rides peak in the summer

Trips correlated with temperature

e-Bikes are 1/5 of fleet but 1/3 of trips

More popular early in year (unsure why)



Weekdays > Weekends

Especially for members (subscribers)

Research Question Analysis

Research Questions

1. How do electric and non-electric **Citi Bike travel times compare** to Google Maps' bike time estimates?

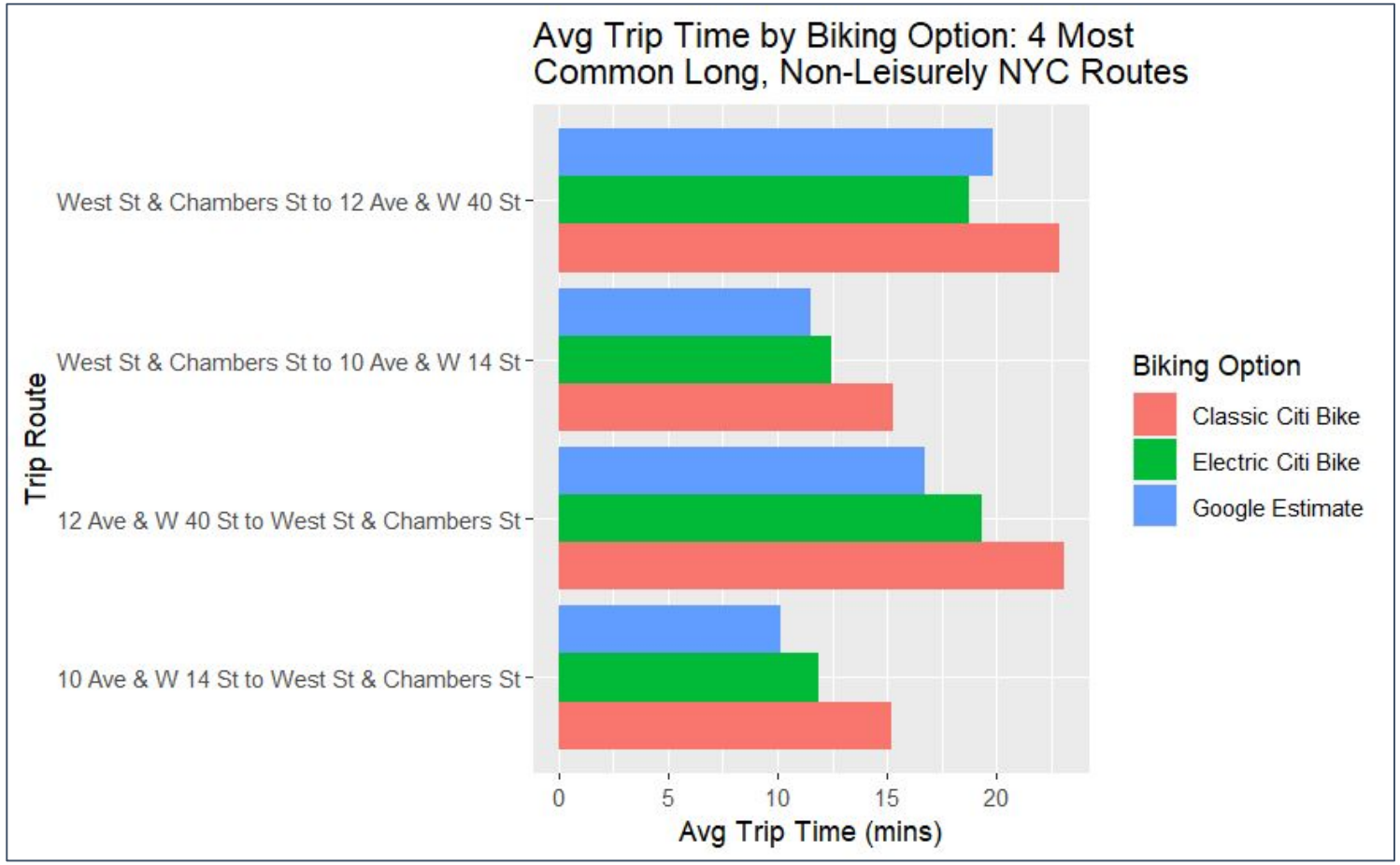
2. Despite the higher price ($0.26 more/minute) does it appear that Citi Bike **riders opt for the electric option** when given the choice between the two?

Travel Time Comparisons

The classic Citi Bike is much slower than Google Maps est. and the e-Bike

On avg, for the 4 most common trips:

- Classic bike 34% slower than Maps
- Classic bike 23% slower than e-Bike
- e-Bike 8% slower than Maps



Rider Bike Preference (1/3)

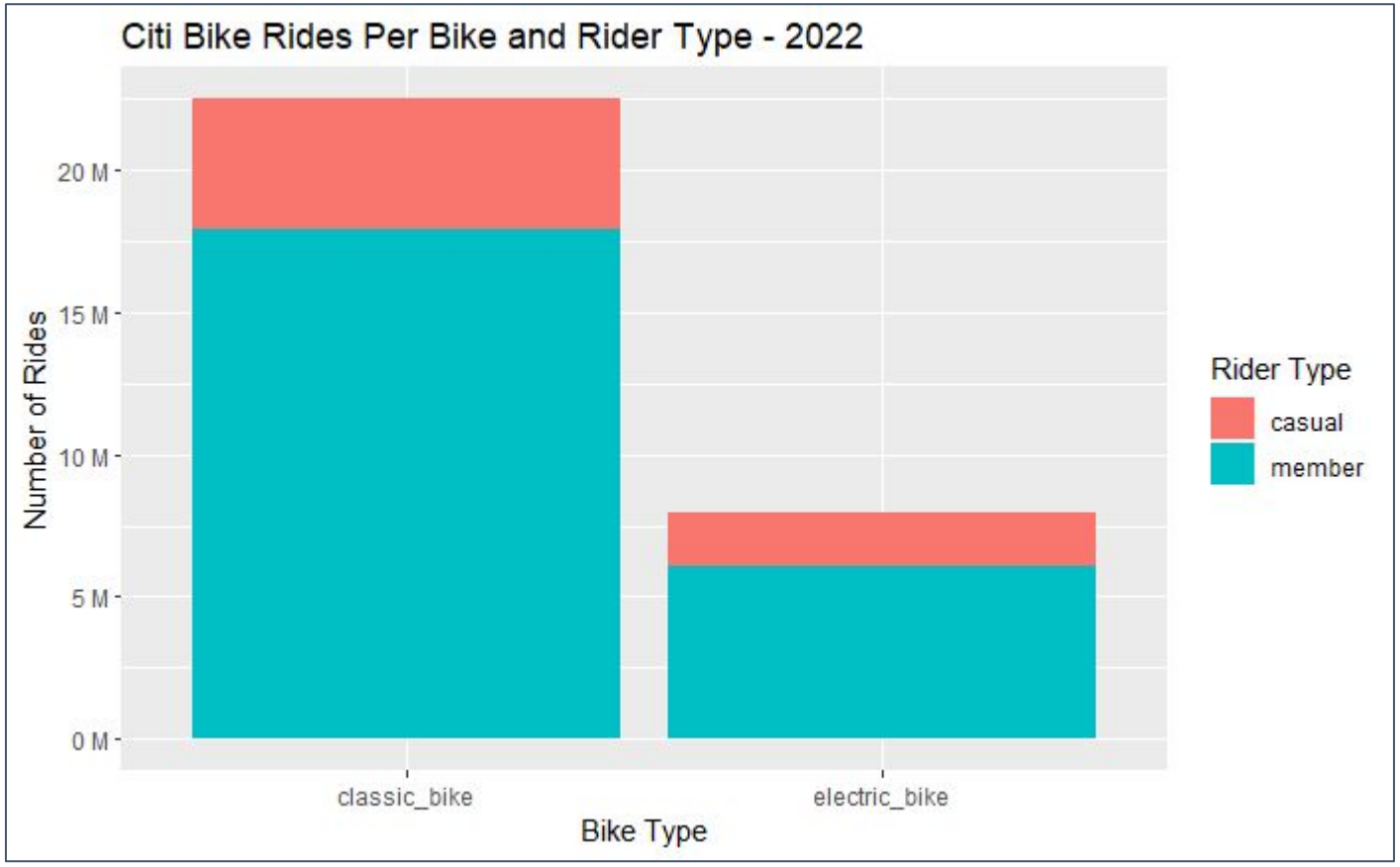
Lyft's 2022 Annual Report provides relevant background

- Citi Bike fleet = 25,000 bikes
- 5,000 lectric, 20,000 classic
- e-Bikes 1/5th of fleet but 1/3 of rides
- e-Bikes used 3x more often per bike

Rider Bike Preference (2/3)

Members seem to prefer the e-Bike

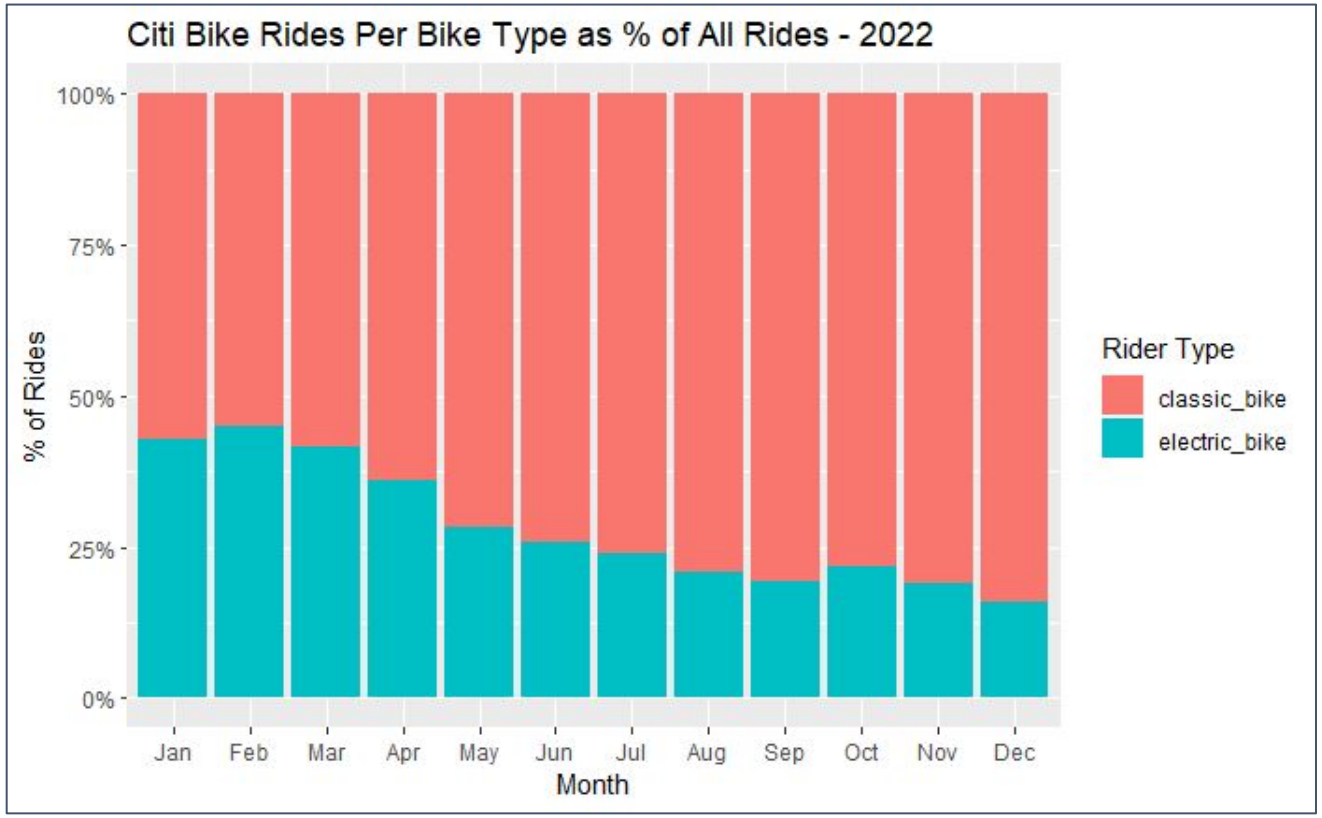
- Likely prefer it for the speed (most obvious difference)
- Perhaps trust members' opinions more as they're repeat customers
- If members are more profitable -> prioritize their opinions



Rider Bike Preference (3/3)

No revealed preference based on seasonal activity

- When there are fewer riders, preference may be more apparent, as the bike docks have more options to choose from.
- While in cold Q1, e-Bikes made up almost half of trips, they lagged in the busy summer -> Data towards e-Bikes being preferred
- However then in Q4, when it was cold and the docks were full of options, the trend was lost



Conclusions / Next Steps / Limitations

Conclusions

- The classic Citi Bike is **34% slower** than the typical bike (using Google Maps as proxy), and **23% slower** than the e-Citi Bike
- The classic bike appears to be **chosen less often** when available due to its slower speed
- Lyft: members use each e-Bike **3x more often**
- My data -> members (more informed users) **choose the e-Bike**

- The **Classic bike can be made easily faster** by:
  - Reducing the weight from its current 45 lbs
  - Making the tires less wide
  - Changing the frame so riders are more leaned forward

Next Steps

- My data are not definitive, but can be directionally useful. Therefore **I suggest Lyft perform a full cost benefit analysis** about whether making their classic bike faster would increase profits.

- **Cost benefit ideas**:
  - Survey customers about if they'd ride a faster classic bike more
  - Hold focus groups for different bike designs
  - Calculate cost estimates for a new bike design and compare to cost of replacing broken classic bikes over time

Limitations

- **2022 data may not be representative of all Citi Bike trips**. The city's infrastructure, such as bike lanes, are constantly changing and Google Maps may have outdated trip time estimates.

- **Lyft may prefer a slower Citi Bike** because it's less dangerous. However the fact that the company has increasingly introduced e-bikes to their fleet means speed is not a primary worry.

Appendix

Project Description

- NYC is at an inflection point: Due to COVID, many people are interested in trying biking in the city. If they enjoy it, New Yorkers could develop a life-long habit. That's why it's important to evaluate the Citi Bike program now.

- My project will evaluate Citi Bike data for 2022 to determine if the non-electric Citi Bike's speed could be improved to encourage more usage. I'm an avid NYC cyclist who occasionally uses Citi Bike when I don't have access to my own bike. The benefits of more Citi Bike usage are multifold: Less environmental pollution; Safety in numbers -- the more cyclists there are, the better/safer biking infrastructure becomes; Healthier citizens; Less space dedicated to car parking; and more. However it's my perception that the Citi Bike's speed holds it back: The non-electric options seem too slow compared to normal bikes. I'll compare trip data for the non-electric ("classic") Citi Bike, the electric Citi Bike, and Google Maps' bike estimates (based on the average bike) to test my hypothesis.

- Citi Bikes are primarily a fixed cost for their corporate owner Lyft. Therefore the more they can be used throughout the year, the more profitable the program will be. Improving the bike can be a win-win: More profit for the owning company, more biking in the city, perhaps creating a virtuous cycle.

RPubs and Downloads

- RPubs file including works cited: https://rpubs.com/rossboehme/1042065

- Github repository including data sources: https://github.com/rossboehme/DATA607/tree/main/finalproj

I’m Done!