CITI BIKE USAGE

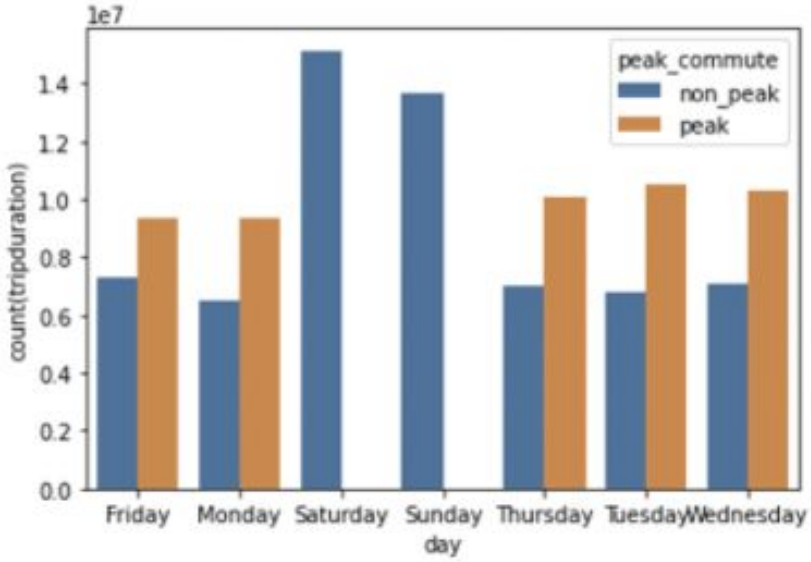
Abigail Bernhardt, Nick Daniello, David Fuentes,
Diana McSpadden, Eric Sarani

Introduction To Citi Bike

- A **shared service** business, Citi Bike must understand how conditions change bike usage.
  - **Rebalancing**
  - Public Transit **Interconnectivity**
- Variables inform Citi Bike
- 2020!!

Data Summary

- August 2018 - April 2021
  - **49,380,335** Citi Bike trips
  - Hourly weather data
  - Monthly real estate data
  - **== 38 variables per trip**



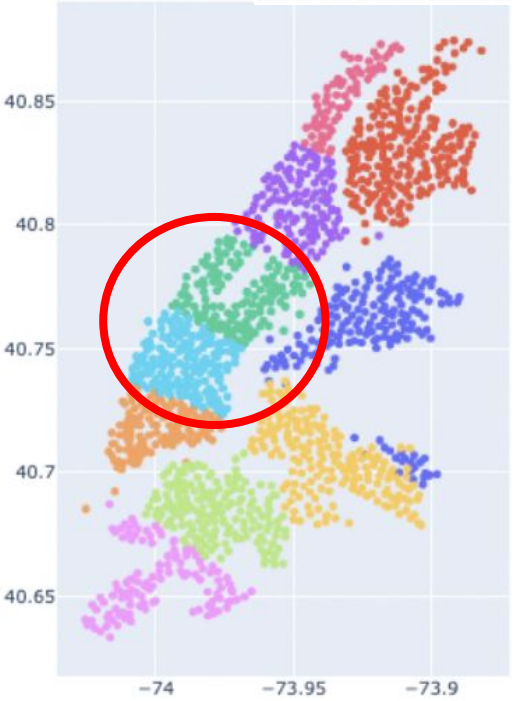
- Engineered data
  - Crow-flies distance
  - Good/Bad weather label per trip
  - Neighborhood and Bike Behavior Classifications

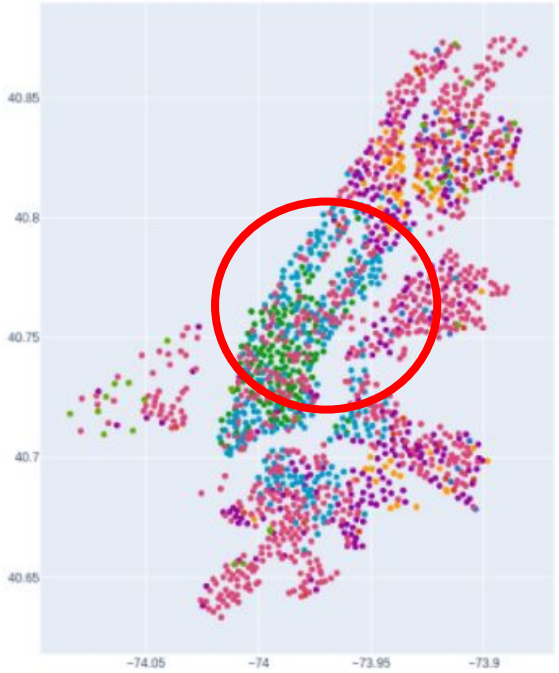
Feature Engineering

Borough	# Stations	% Total
Queens	179	11.72
Brooklyn	471	30.84
Manhattan	651	42.63
Bronx	226	14.8

start_neighborhood	# Stations	% Total
Bronx	226	14.8
Uptown BK	183	11.98
Queens	179	11.72
Midtown BK	172	11.26
Midtown Manhattan	171	11.2
Uptown Manhattan	138	9.04
Downtown Manhattan	138	9.04
Central Park East...	137	8.97
Downtown BK	116	7.6
Harlem & Wash. He...	67	4.39

Neighborhood Grouping





Bike Behavior Grouping

Problem Statements and Approaches

1. **Dock Rebalancing:**
   - Approach: GNN
   - Important Variables:
     - *Good v. Bad Weather; COVID's Effect on Trips*

2. **End-Station Prediction:**
   - Approach: Log. Reg. and Rand. Forest
   - Important Variables:
     - *Starting neighborhood, time of day, DoW*

3. **Holiday Bike Behavior**
   - Approach: Log. Reg. and Rand. Forest
   - Important Variables:
     - *User type, ride duration, ride location*

4. **Trip Distance Prediction:**
   - Approach: Lin. Reg., GBT, Rand. Forest
   - Important Variables:
     - *Zip code, 2020, Subscribers, temperature*

5. **Real-Estate Price Prediction:**
   - Approach: Lin. Reg., Rand. Forest
   - Important Variables:
     - *Ride distances, ride times, peak/non-peak commute flag*

Modeling Results

- Informing Rebalancing:
  - Overall Importance: **Midtown and Central Park West**
  - Good v. Bad Weather: **Fringe stations non-importance**
  - Pre-COVID v. COVID: **Cycling new places during COVID**
  - Holiday Bike Behavior: **Less travel. Downsampling**
- Behavior Prediction:
  - End-Station Prediction: **Issues w/ inter-neighborhood**
  - Real-Estate: **Expensive neighborhoods had more stations**
  - Distance Prediction: **2020, user-type, and temperature**

ML Results - Distance

Feature	GBT Importance	RF Importance	Change in Estimated Miles (beta LinReg)
Night (time-bin category)	0.0736	0.0013	0.0262
Temperature	0.0558	0.1289	0.0016 per degree
Year 2020 (category)	0.0495	0.2406	0.0926
Subscriber vs. Customer (category)	0.0430	0.1578	-0.0739
11237 (Ridgewood)	0.0455	0.0097	-0.1756
11221 (Bushwick - BedStuy)	0.0430	0.0019	0.2181
10023 (Upper West Side)	0.0150	0.0465	0.4781
11222 (Bushwick - BedStuy)	0.0166	0.0421	0.5472
10460 (South Bronx)	0.0107	0.0367	0.3956

Low Explanatory Power: Sub 0.20 R^2 | Predicts some user behavior given available data | More robust user data will offer high returns | 2020 was powerful | Hipsters, Bronx Zoo, Central Park

Conclusion and Into the Future

- Citi Bike Transportation **worthwhile alternative in 2020**

- Areas **underserved by subways** or **near major attractions** (Central Park, Bronx Zoo, etc.) result in **longer bike trips**

- **User data is important**: Available data only got us so far in our analysis
  - User a triathlete or fitness instructor vs. office professional?
  - User bike every day?
  - User own a car?

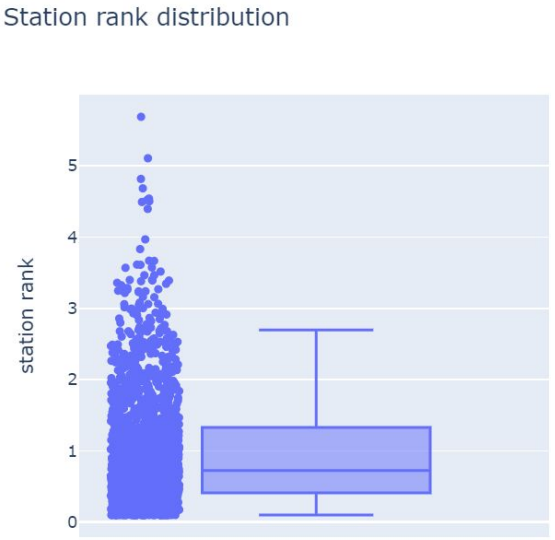
- Need **MORE DATA** to predict specific user behavior.

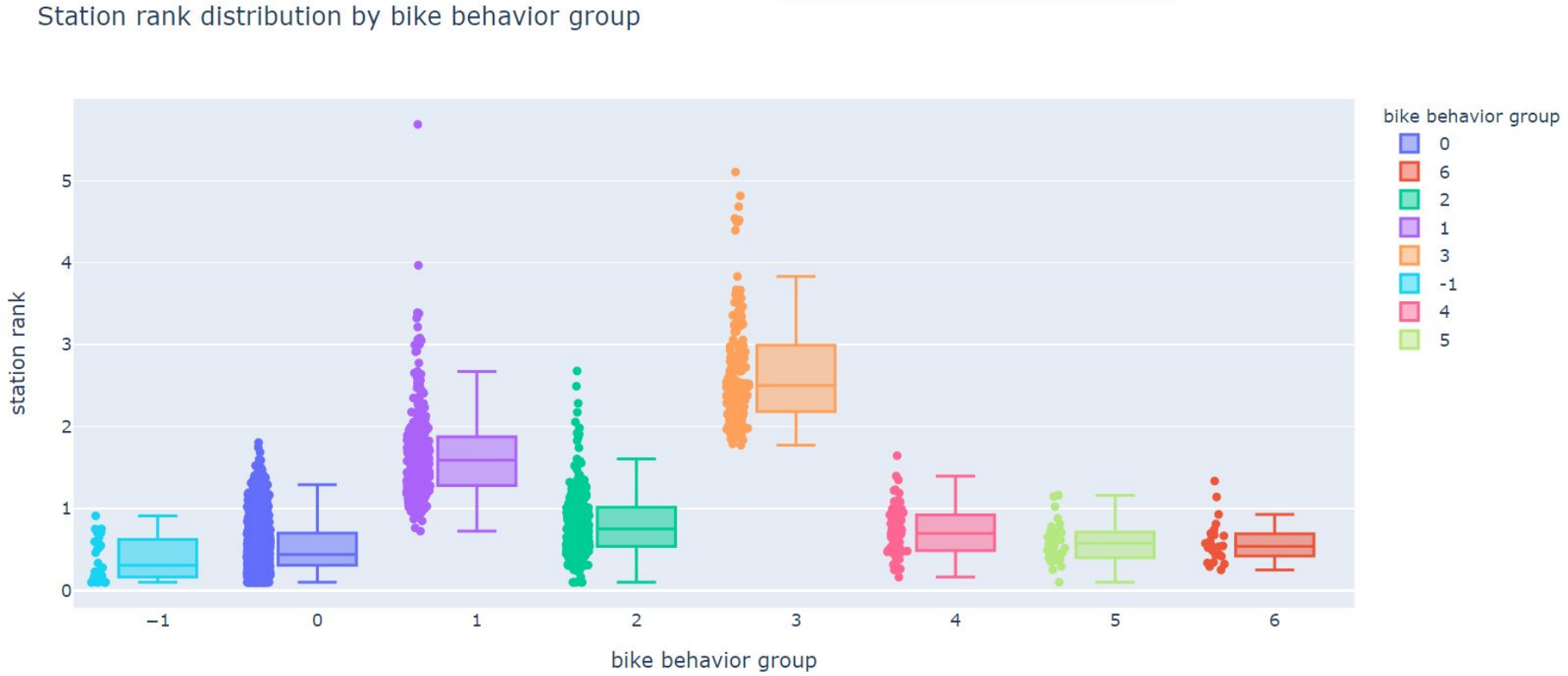
QUESTIONS

She who succeeds in gaining the mastery of the bicycle will gain the mastery of life. - Susan B. Anthony

BACKUPS

Station Rank





Neighborhood Trip Counts

Table 23: Rides by K-Means Neighborhood (left); Stations by K-Means Neighborhood (right)

start_neighborhood	# Rides	% Total
Midtown Manhattan	15825350	32.12
Downtown Manhattan	9991879	20.28
Central Park East...	8245275	16.73
Midtown BK	5233122	10.62
Uptown BK	4002562	8.12
Uptown Manhattan	3139583	6.37
Queens	1917315	3.89
Downtown BK	488525	0.99
Bronx	280608	0.57
Harlem & Wash. He...	146790	0.3

start_neighborhood	# Stations	% Total
Bronx	226	14.8
Uptown BK	183	11.98
Queens	179	11.72
Midtown BK	172	11.26
Midtown Manhattan	171	11.2
Uptown Manhattan	138	9.04
Downtown Manhattan	138	9.04
Central Park East...	137	8.97
Downtown BK	116	7.6
Harlem & Wash. He...	67	4.39