Bikeshare *Wizard*

Anna Cheng, Tiffany Cheng, Tina Fang, Giulia Olsson, Tres Pimentel

Fall 2022

Bikeshare is a popular form of transportation

Bikeshare: publicly-available bikes that are docked at stations, rented out, and returned to a station within the same metropolitan area.

Annual Members (2021)

NYC
147K[1]

Boston
23K[2]

Trips Taken (2021)

NYC
26M[3]

Boston
3M[2]

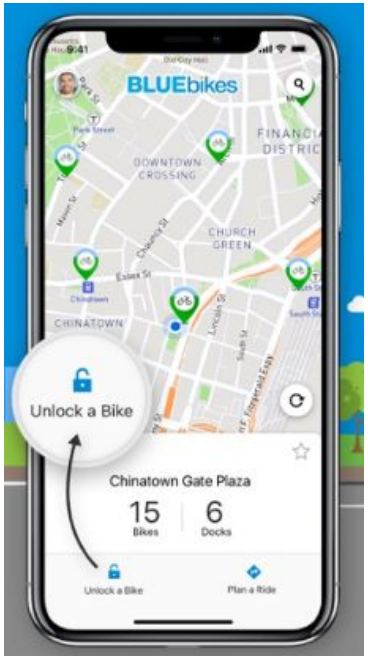
Worldwide Users: Up to 10MM[4]

[1]https://mot-marketing-whitelabel-prod.s3.amazonaws.com/nyc/December-2021-Citi-Bike-Monthly-Report.pdf
[2]https://www.bluebikes.com/system-data
[3]https://w42st.com/post/new-york-pedal-power-why-2021-was-the-year-of-the-citi-bike/
[4]PBSC Urban Development

User interviews uncovered areas of opportunity

Current State

- Users rely on bikeshare apps, which only provide real-time station status
- No availability forecasts



Desired State

- Users want to know bike and dock availability ahead of time to plan commute
- Ideal forecast timeframe: 10 mins to 1+ days into the future

Image: https://apps.apple.com/us/app/bluebikes/id1250203009

We designed our solution with the user's needs in mind

Problem
Users do not know whether there will be bikes or docks available when they need them

Our Solution
Provide bike and dock availability predictions by station in 15-minute increments for up to 3 future days

Our solution has positive impacts on multiple fronts

User Experience

- Ability to plan commute ahead of time
- More reliable and efficient commutes
- Better UX leads to more users ($)

Climate Impact

- More users mean less cars on the road
- 1 mile biked (instead of driven) reduces CO2 emissions by 1 pound[1]
- Citi Bike in NYC saved ~7K tons of CO2 from being emitted in 2021[2]

[1]https://www.scientificamerican.com/article/is-bike-sharing-really-climate-friendly/
[2]https://w42st.com/post/new-york-pedal-power-why-2021-was-the-year-of-the-citi-bike/

We chose Boston Bluebikes system for our MVP

BLUEbikes℠

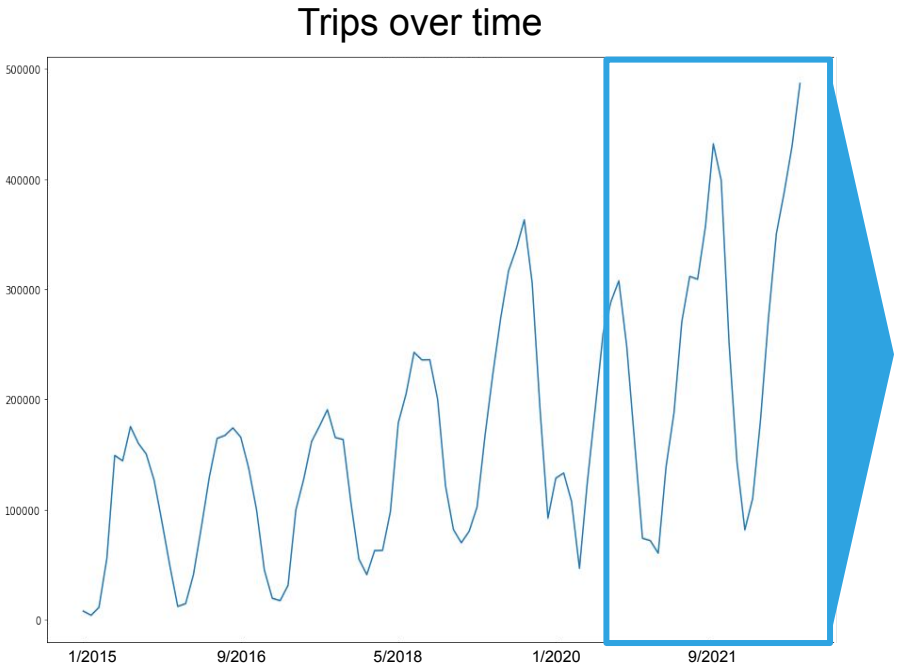
- Historical trips data from 2015 to present
- ~15 million trips
- ~500 stations

Most Popular Trips

Start Station	End Station	Number of Trips
MIT at Mass Ave / Amherst St	MIT Vassar St	23,572
MIT Pacific St at Purrington St	MIT Stata Center	22,934
MIT Vassar St	MIT Stata Center	22,146

Image: https://upload.wikimedia.org/wikipedia/commons/e/e1/Bluebikes_Logo.svg

EDA showed strong seasonality and impacts from the pandemic



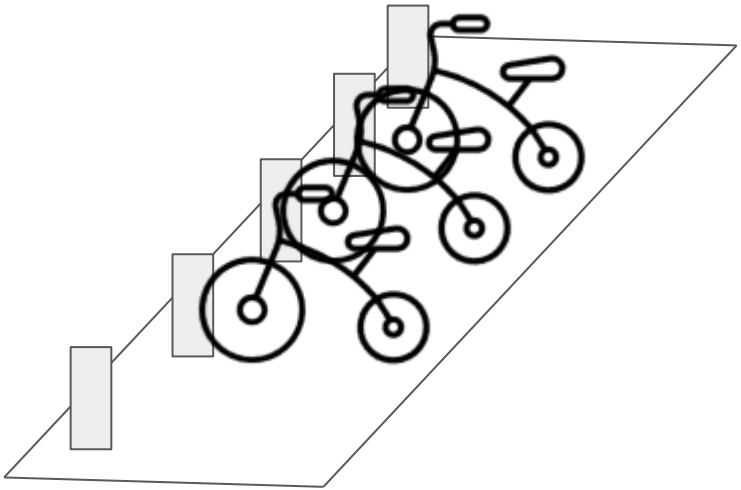
- Build the models with data from Aug 2020 - Aug 2022
- Changes in bikeshare usage due to COVID-19 pandemic (e.g. WFH, tourists, public transit avoidance)[1]

[1]https://bloustein.rutgers.edu/bikeshare-subway-ridership-changes-during-covid19-pandemic-nyc/

For our baseline, we assessed the predictive accuracy of the real-time data in the Bluebikes app

1. Real-time station data:

- 3 bikes
- 2 docks



2. Assumption for next 15 mins:

- 0 bikes leave (outbound)
- 0 bikes arrive (inbound)

3. Accuracy (RMSE):

- Inbound: 1.17
- Outbound: 1.20
-

Note: This is a hypothetical example for illustrative purposes only.

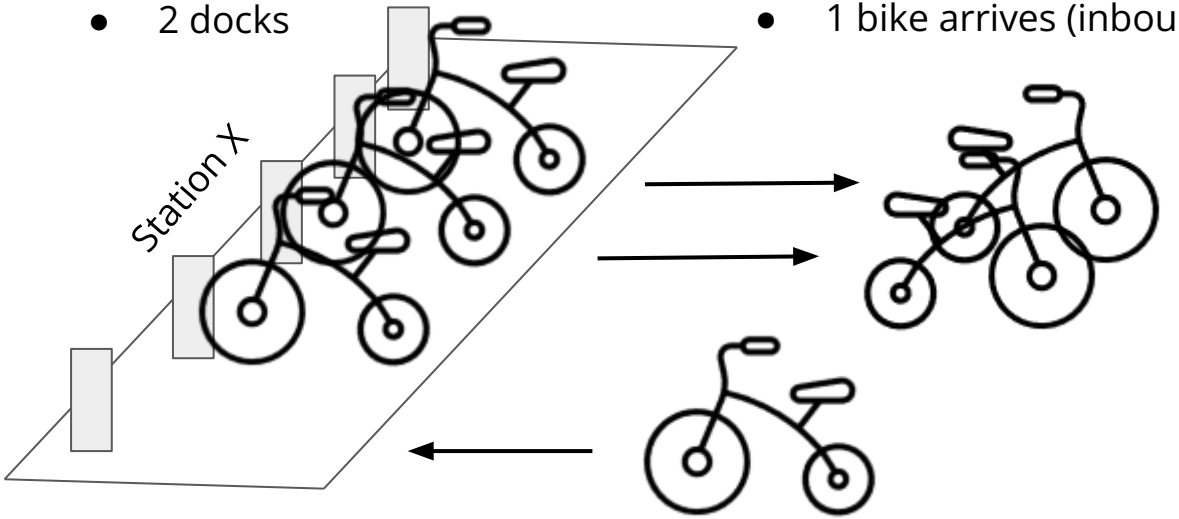
Our prediction approach combines real-time data with bike movement forecasts from two DeepAR models

1. Real-time station data:

- 3 bikes
- 2 docks

2. Forecast for next 15 mins:

- 2 bikes leave (outbound)
- 1 bike arrives (inbound)



3. Station X forecast:

- 2 bikes
- 3 docks

Note: This is a hypothetical example for illustrative purposes only.

DeepAR is a forecasting algorithm built by Amazon and is well-suited for our MVP

Streamlined

- One model for all 450 stations
- LSTM and Prophet train one model per station

Flexible

- Time series can differ in length, which accommodates stations that are added at different times

Generalizable

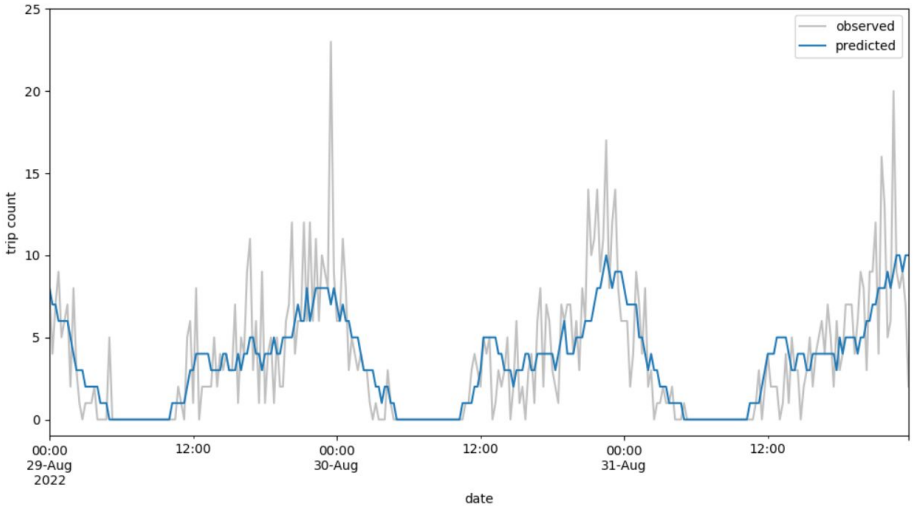
- Can generate forecasts for new stations that were not in the training data

After testing several iterations, our best models outperform the baseline

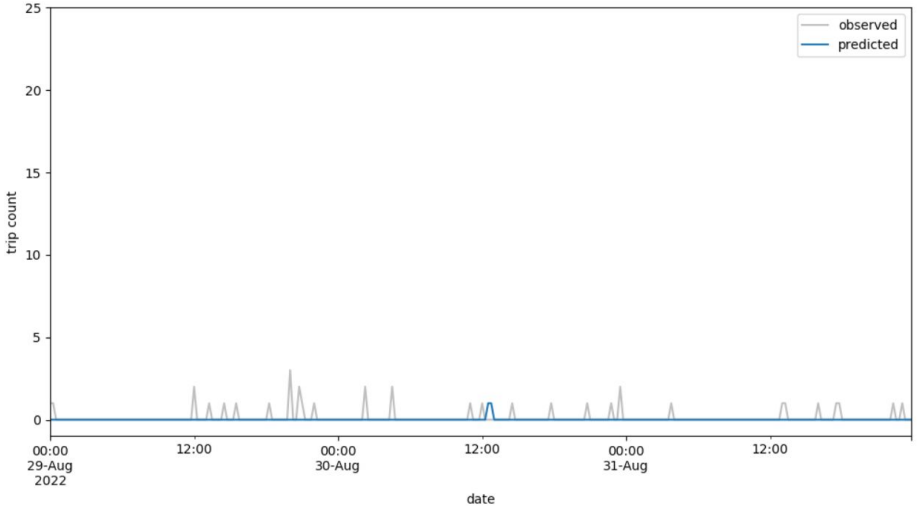
	Baseline	LSTM	Prophet	DeepAR			
Time Increment	15 min	15 min	15 min	15 min	**15 min**	15 min	15 min
Likelihood	✗	✗	✗	Student T	**Negative Binomial**	Student T	Negative Binomial
Station Clusters (k-means)	✗	✗	✗	✗	✗	✓	✓
Outbound Model RMSE	1.20	*Unsuitable for our MVP because both require training separate models for each station.*		1.00	**0.91**	1.23	0.93
Inbound Model RMSE	1.17			0.97	**0.87**	1.13	0.86

The models are more useful for busy, high volume stations

Busy: Central Square at Mass Ave / Essex St

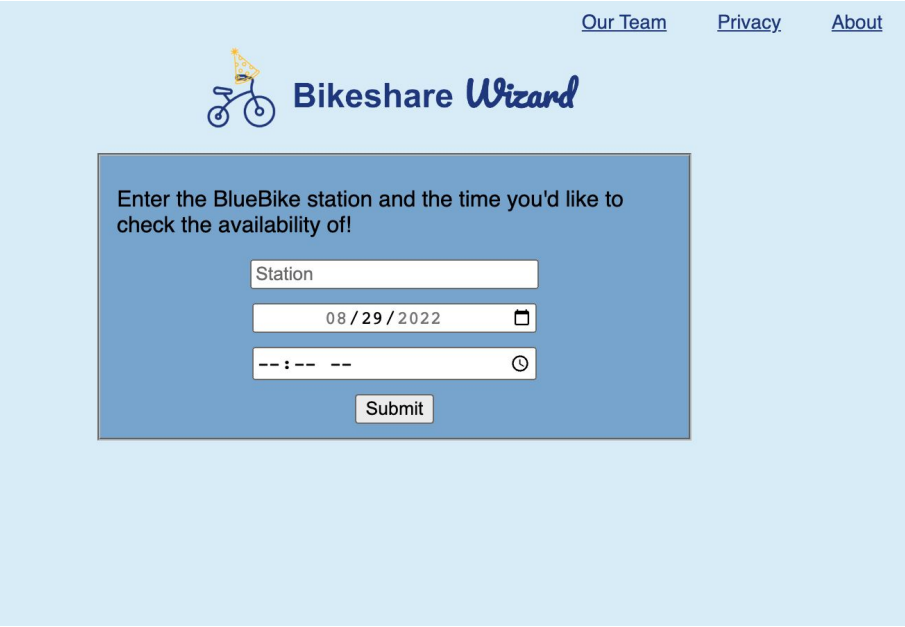


Not busy: Maverick St at Massport Path



Note: Charts show results for outbound models, and results are similar for inbound models.

Our web app is easy-to-use and provides quick forecasts

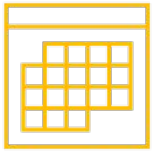


- App page is clean with intuitive UI
- User must enter three things (drop-down and autocomplete available to prompt user)
  - Station
  - Day
  - Time
- App page also has links to pages about the project, our team, and privacy

Our MVP would ideally be owned by and incorporated into the Bluebikes app

Data

- Monthly data uploaded by Bluebikes
- Delayed by 1-2 weeks



Model

- DeepAR forecasts max. 3 days in our case
- Needs to be retrained regularly with new data

User Interface

- Incorporate forecasts into existing Bluebikes app
- Users only need one app

We believe our MVP will help increase ridership, which will benefit the climate

Our mission is to **make bikeshare more reliable** and to ensure that it remains **an attractive and climate-friendly form of transportation**.

We approached this problem by providing **bike and dock availability predictions** for each station.

Thank you!

Questions?

Acknowledgements

We would like to thank our Capstone instructors, **Joyce Shen and Zona Kostic**, for their expert guidance and invaluable feedback as we made progress on our project. We would also like to thank **Robert Turnage** for helping us set up our web application. Last but not least, we would like to thank our fellow **classmates in Section 1** for their feedback during our weekly check-ins and presentations.

Thank you to **Mina Iskarous, Steffen, Maddie DiLullo Byrne, Jessica Sparacino, and Luis Octavios** for answering our user interview questions, which shaped how we built our MVP.

Key Learnings

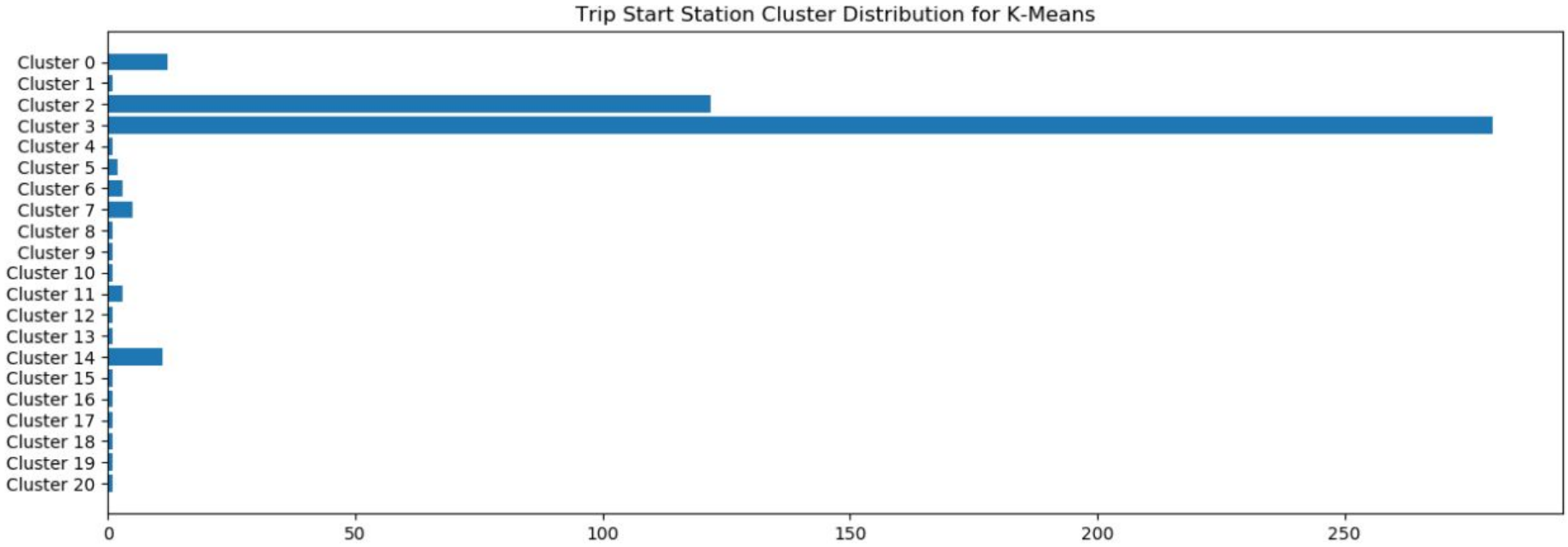
Data and EDA

- Changes in bikeshare usage due to the COVID-19 pandemic impacted timeframe chosen for modeling
- Engineer historical trips data to predict bike and dock availability

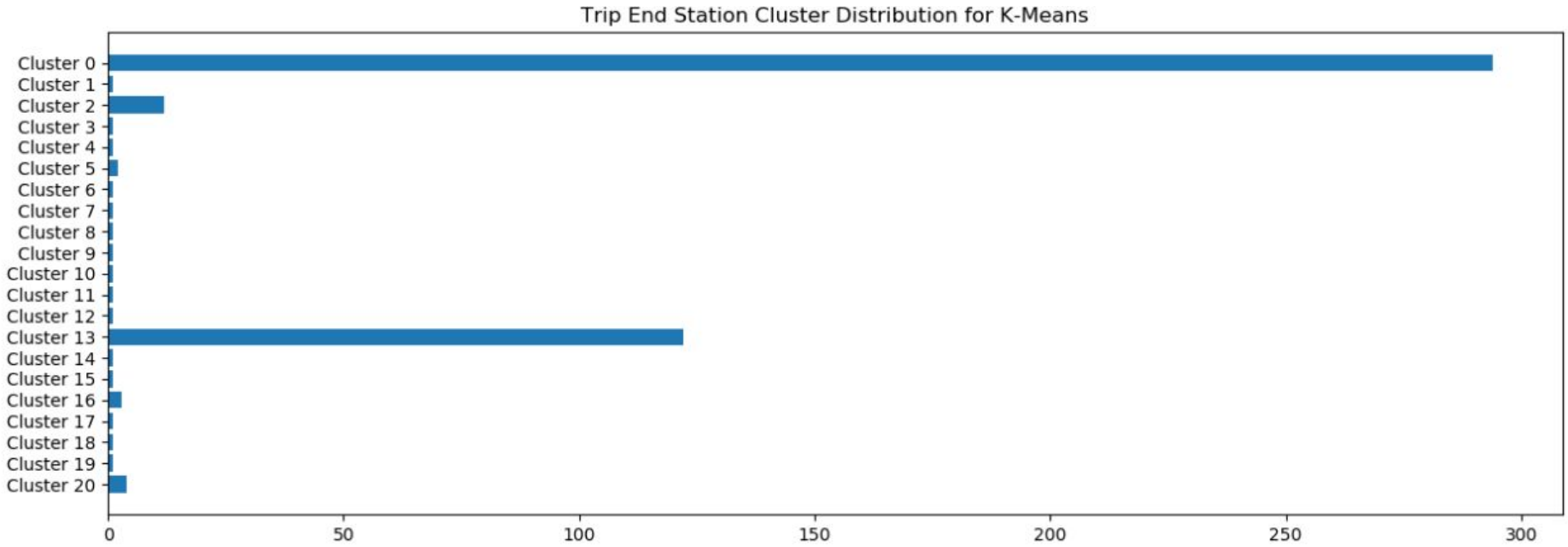
Modeling

- 1 model is preferred over 450 models
- Models need to be updated with new data and regularly trained for the long term

Outbound Model: Station Clusters (15-minute frequency)



Inbound Model: Station Clusters (15-minute frequency)



Additional Results for DeepAR (Student T likelihood)

	Baseline	**DeepAR**				
Time Increment	15 min	15 min	30 min	1 hour	15 min	15 min
Station Clusters (k-means)	✗	✗	✗	✗	✓ (15 min)	✓ (1 day)
Outbound Model RMSE	1.20	1.00	1.49	2.29	1.23	1.17
Inbound Model RMSE	1.17	0.97	1.43	2.13	1.13	1.11

Privacy concerns around the MVP have been addressed

Concern	Approach
Can a user's location be tracked? / Can someone see a user's route history?	User profiles/accounts are not used (data is not saved under any user)
What happens if someone gains access to our data?	All data used to build our models are publicly available & no user PII is used
What if there is a data breach at a bikeshare company?	We do not use bikeshare user IDs in our models

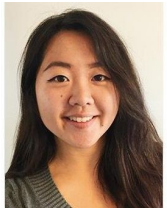
Additional valid security/privacy concerns have been raised but are considered out-of-scope for our project:

- Phone stolen from user
- Unauthorized access to outgoing queries from user

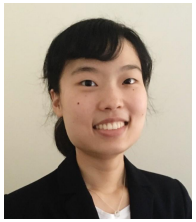
We believe no significant privacy concerns exist at this time.

Bikeshare Wizard

A brief introduction to the team



Anna Cheng
Data Scientist



Tiffany Cheng
Data Scientist

Tina Fang
Student

Giulia Olsson
Data Scientist

Tres Pimentel
Student

We want to make bikeshare a more reliable form of transport

Bikeshare: publicly-available bikes accessible through a mobile app that are docked at stations and rented out and returned to the same station or to a different station within the same metropolitan area.

Motivation
Bikeshare is a popular and fast form of transportation in many cities across the world

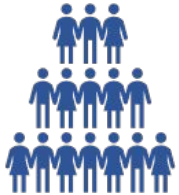
Problem
Users do not know whether there will be bikes or docks available when they need them

Our Solution
Provide bike and dock availability predictions by station

Our solution positively impacts bikeshare users and the climate

Market Opportunity

- $8B global ARR by 2022[1]
- 1 billion users worldwide by 2026[1]



User Experience

- Ability to plan commute ahead of time
- More reliable and efficient commutes

Climate Impact

- 1 mile biked (instead of driven) reduces CO2 emissions by 1 pound[2]

[1]https://www.statista.com/outlook/mmo/shared-mobility/shared-rides/bike-sharing/worldwide
[2]https://www.scientificamerican.com/article/is-bike-sharing-really-climate-friendly/