Predicting Dockless Bike and Scooter Availability

Elisabeth Ericson

MUSA Capstone 2022

The topic: Dockless micromobility

- New transportation mode; emerged in last several years
- Potential to address first- and last-mile problem for urban neighborhoods poorly served by transit and traditional bikeshare
- Controversy over growth and management has led to regulation, onerous restrictions

I considered a range of very serious questions...

- ➔ Does dockless micromobility improve accessibility for neighborhoods poorly served by transit and traditional bikeshare?
- ➔ Can cluster analysis of origin-destination data tell us anything interesting about where people travel using dockless modes?
- ➔ What neighborhood characteristics predict high rates of dockless bikeshare adoption relative to traditional bikeshare?
- ➔ How do micromobility providers compare on equity? What companies are best and worst at expanding access to low-income or non-white neighborhoods?
- ➔ Is it possible to infer precise trip start and end coordinates by combining anonymized trip history data with real-time inactive bike location data?
- ➔ How does deep learning for origin-destination demand prediction work, and how could I adapt it to this data?

…but decided to go with something more lighthearted

addressing a problem that, if you've used scooters or bikeshare, you might be able to relate to

What are the chances that a bike or scooter listed as “available” in the app will be taken before I can get to it?

The premise

- ➔ Build a model to predict how much longer an inactive bike or scooter is likely to remain available
- ➔ Build a simple web app where a user can enter their current location, select a nearby bike or scooter, and see how likely it is to still be available when they get there

Bike & scooter location data

- Real-time API publishes coordinates for all inactive vehicles
- Standardized format: General Bikeshare Feed Specification (GBFS)
- Wrote demo Python scraper and collected location data every 60 seconds for 24 hours

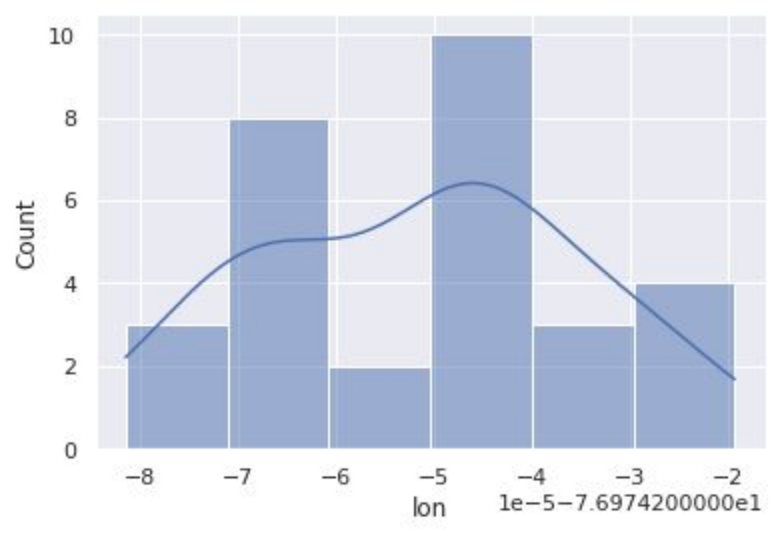
Complications

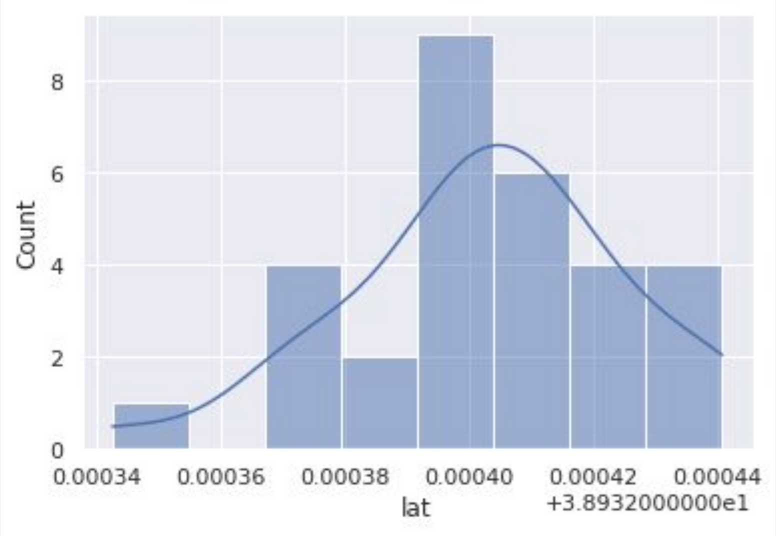
- ➔ **Dynamic vehicle IDs:** Bicycle IDs reset every 30 minutes
- ➔ **GPS error:** Coordinates not always consistent between records, even when the bicycle hasn't moved

bike_id	is_reserved	is_disabled	type	lon	lat	timestamp
002604d3123025e6e2fa8384ee72d2a6	0	0	electric_bike	-76.974229	38.932343	09:30:07
002604d3123025e6e2fa8384ee72d2a6	0	0	electric_bike	-76.974219	38.932373	09:31:09
002604d3123025e6e2fa8384ee72d2a6	0	0	electric_bike	-76.974228	38.932403	09:32:11
002604d3123025e6e2fa8384ee72d2a6	0	0	electric_bike	-76.974228	38.932403	09:33:13
002604d3123025e6e2fa8384ee72d2a6	0	0	electric_bike	-76.974240	38.932374	09:34:14

GPS readings for one bike over half an hour

- Range of about 0.0001 degree, or about 11 meters





Longitude (left) is around -76.97 and latitude (right) around 38.93;
if anyone knows why Seaborn does this to the axis labels, please let me know

Other data to incorporate

- **Spatial:** Population and employment density; distance to major streets or intersections; distance to transit
- **Temporal:** Day of week and time of day
- **Weather:** Temperature, precipitation
- **Access:** Other bikes/scooters nearby
- Others?

Other data to incorporate

- ➔ **Spatial:** Population and employment density; distance to major streets or intersections; distance to transit and bikeshare
- ➔ **Temporal:** Day of week and time of day
- ➔ **Weather:** Temperature, precipitation
- ➔ **Access:** Other bikes or scooters nearby
- ➔ Others?

Important next steps

- ➔ Figure out how to compute time to vehicle activation (dependent variable)
- ➔ Write a more polished scraper for larger-scale data collection
- ➔ Figure out best way to set up data pipeline
- ➔ Decide a) whether to include both bikes and scooters, or just bikes; b) if just bikes, whether to look at SF instead of D.C.

Questions?