Capital Bikeshare Fleet Modeling

A toolkit for station agnostic fleet rebalancing operations

Introduction: The Business Case

Capital Bikeshare

Capital Bikeshare is the largest provider of rented bikes in DC.

Recently purchased by Lyft.

Context

Traditional docked bikes must be left at one of around 600 stations throughout the region.

"Dockless" bikes introduced in July 2020.

Problem

Maximize availability of bikes and parking without depending on dock-based models.
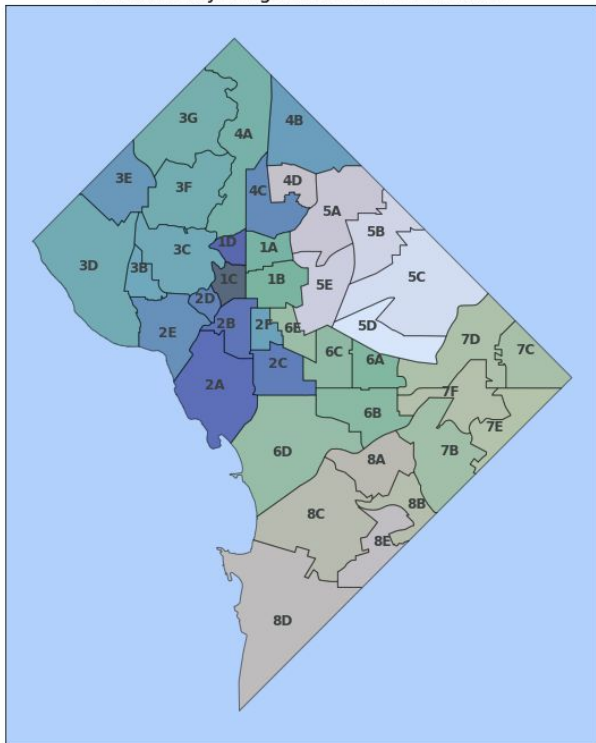
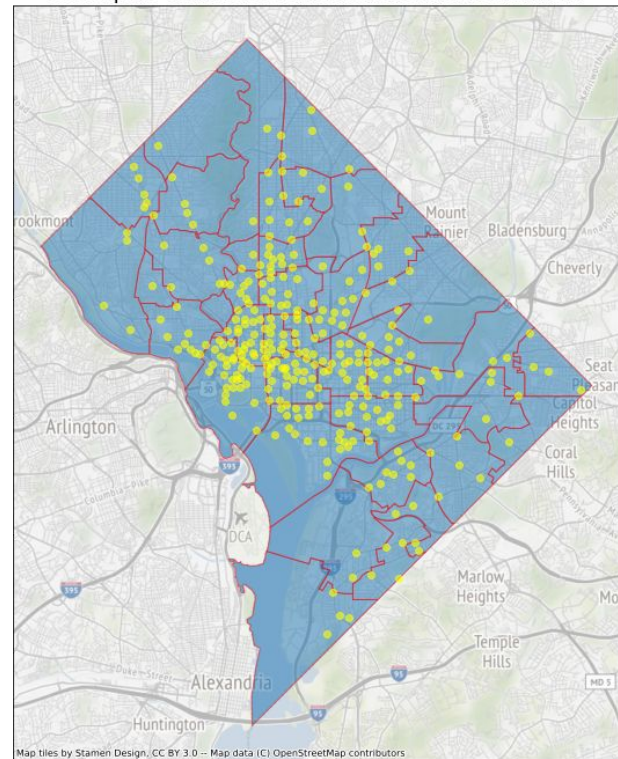We propose models by ANC instead.

The Problem

Geography Approach: Areas Instead of Stations

DC Advisory Neighborhood Commissions

DC Capital Bikeshare stations with ANC boundaries

Trips Approach:
2020 Is Not A "Normal" Year

Daily Capital Bikeshare Ride Counts, 2018-2020
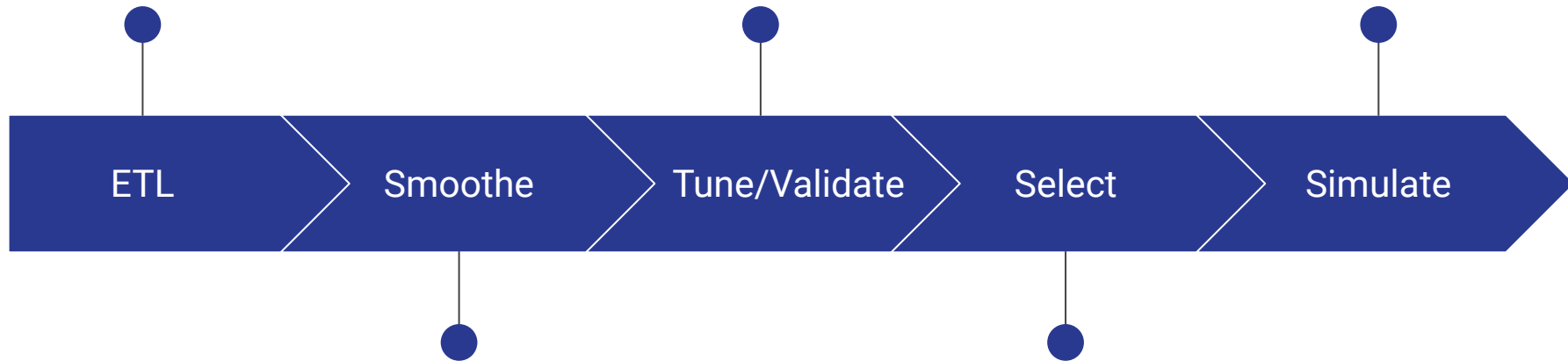Green and Red Lines are 15 Day Moving Averages

Data Science Process

Trip start/end points assigned to an ANC, loaded into postgis by time

Find best components for each model. Validate over different time periods.
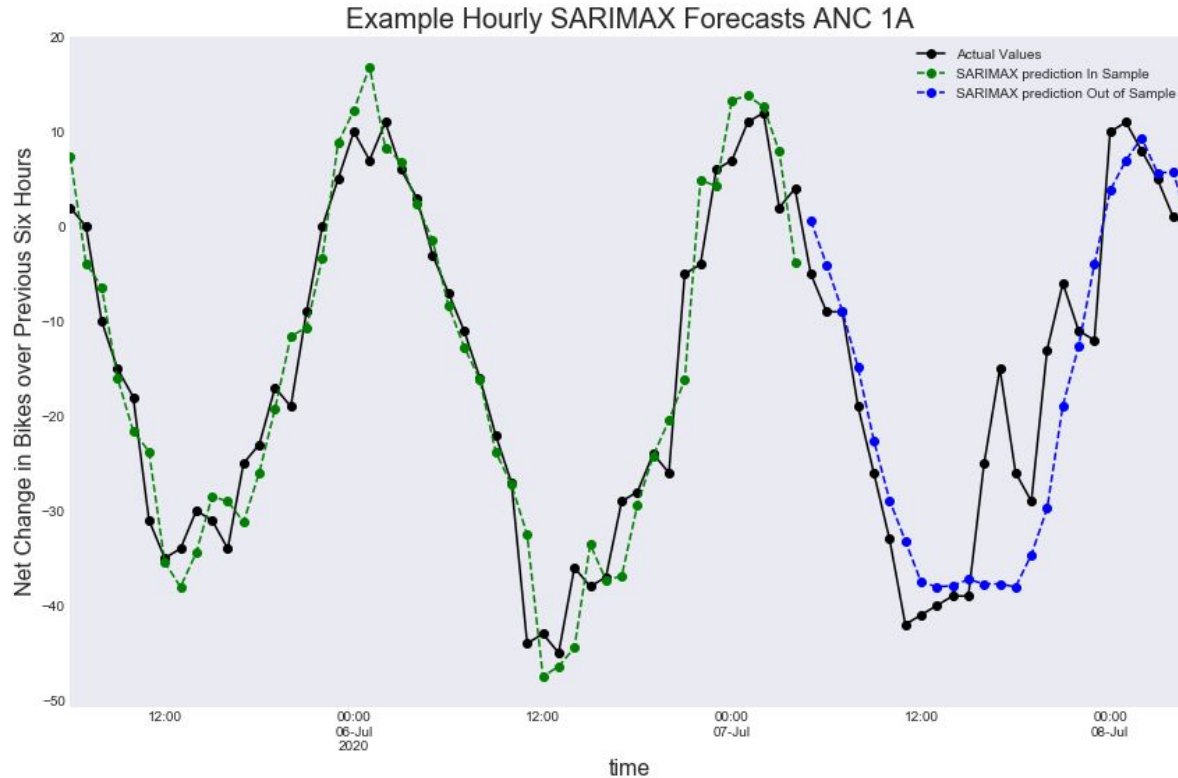
Simulate business application to be tested on August data soon to be released.

| ETL | Smoothe | Tune/Validate | Select | Simulate |

Select the right frequency for each model.

Compare model performance. Select best models for each task.

Model: Daily SARIMA

Example Hourly SARIMAX Forecasts ANC 1A

Results and Recommendations

Best Models had RMSE of around 6 over 24 hour forecast (off by around 6 bikes over next 6 hours for each hour of 24 hour cycle >> not bad)
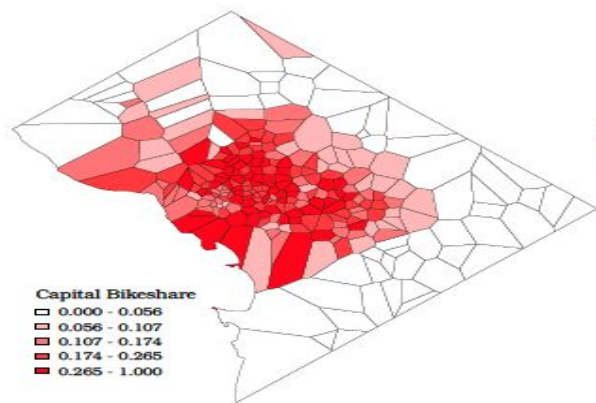
But the percentage error was more like 30% including when weekends (the models did not catch the "spikes" well (not good)

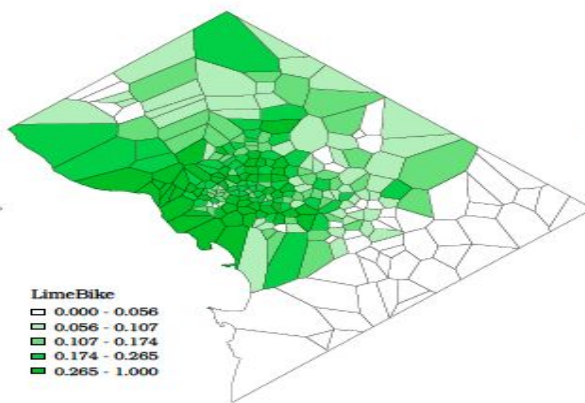Capital Bikeshare should use our model with an added "weekend effect"
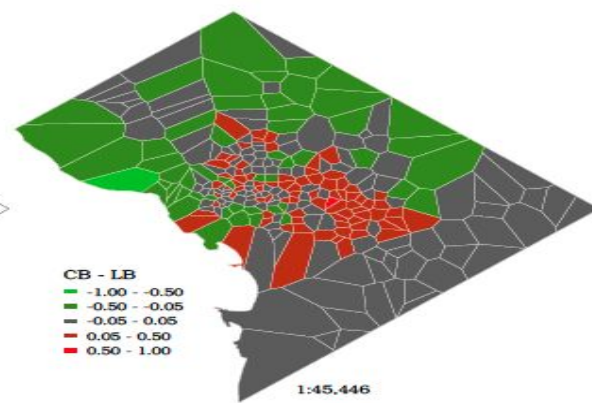
Future Improvements

- APIs make live predictions possible

- Time series predictions as an input feature in a larger model

- Clustering Gives More Actionable Results

(a) Capital Bikeshare

Capital Bikeshare
- 0.000 - 0.056
- 0.056 - 0.107
- 0.107 - 0.174
- 0.174 - 0.265
- 0.265 - 1.000

(b) LimeBike

LimeBike
- 0.000 - 0.056
- 0.056 - 0.107
- 0.107 - 0.174
- 0.174 - 0.265
- 0.265 - 1.000

(c) LB subtracted from CB

CB - LB
- -1.00 - -0.50
- -0.50 - -0.05
- -0.05 - 0.05
- 0.05 - 0.50
- 0.50 - 1.00

1:45,446

Figure 3 Normalized trip starts assigned to Capital Bikeshare station-based Voronoi polygons.

Sources

- Voronoi Polygons on previous slide are from Grant McKenzie's "Docked vs. Dockless Bike-sharing: Contrasting Spatiotemporal Patterns" (GIScience 2018)
- Analysis based on monthly trip data obtained from Capital Bikeshare (https://s3.amazonaws.com/capitalbikeshare-data/index.html)
- API address for real time info on available ebikes not at a dock: (https://gbfs.capitalbikeshare.com/gbfs/en/free_bike_status.json)
- API address for real time info on available bikes/ebikes at all CaBi docks (https://gbfs.capitalbikeshare.com/gbfs/en/station_status.json)
- TimeSeriesSplit Graphic from (https://scikit-learn.org/stable/auto_examples/model_selection/plot_cv_indices.html)
- Empty Dock Image from (https://www.popville.com/2014/09/problems-with-capital-bikeshare-for-commuting)

Thank You!

Contact

Github: https://github.com/jmillerbrooks/capital_bikeshare

Email: brooksjacobm@gmail.com

LinkedIn:
https://www.linkedin.com/in/jake-miller-brooks-a37a64106/