DS-GA 3001.001 Final Project

Time Series Analysis on Transportation Usage and COVID-19

Cora Hyun Jung
Emmy Phung
Sujeong Cha

Agenda

I. Problem Formulation
II. Data Overview
III. Solution Approach
IV. Result Summary
V. Conclusion

I. Problem Formulation

Motivation

- Steep changes in the mobility trends in NYC have been observed during COVID-19 pandemic.
- Since our daily lives are heavily affected by means of transportation, it's essential to investigate the relationship between the mobility pattern and COVID-19

Goal

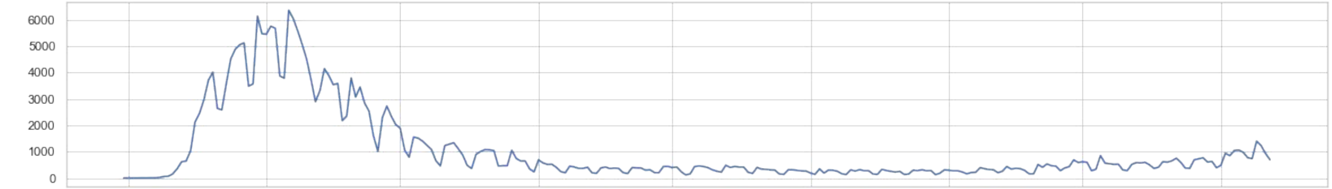
- To investigate the severity of the mobility pattern change across three means of transportation in NYC during COVID-19 pandemic
  - Subway
  - Taxi/For-hire vehicles
  - Citibike

II. Data Overview

COVID-19

(03/01/20 - 06/30/20)

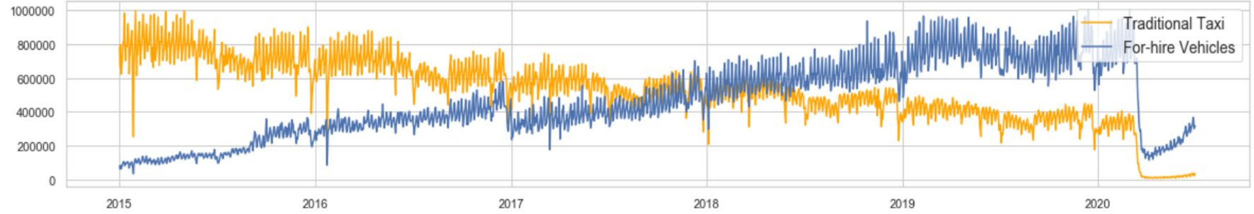
**NYC Health COVID data*



Taxi

(01/01/15 - 06/30/20)

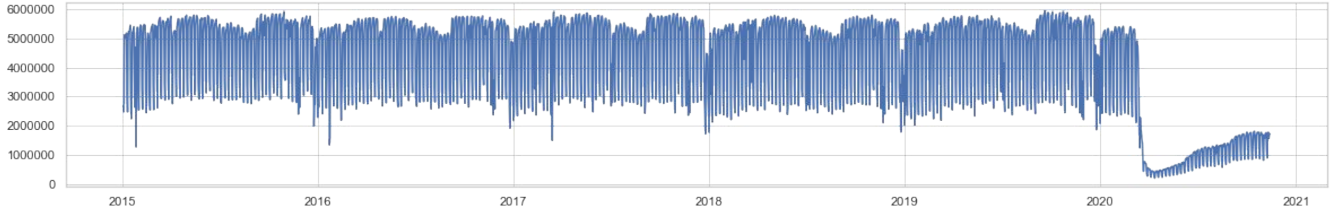
**NYC TLC Trip Record*



Subway

(01/01/15 - 06/30/20)

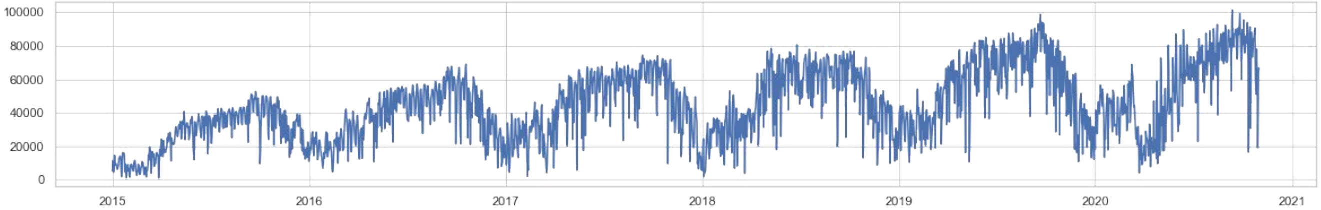
**NYC MTA Turnstile*



Citibike

(01/01/15 - 06/30/20)

**NYC Citi Bike Data*



III. Solution Approach

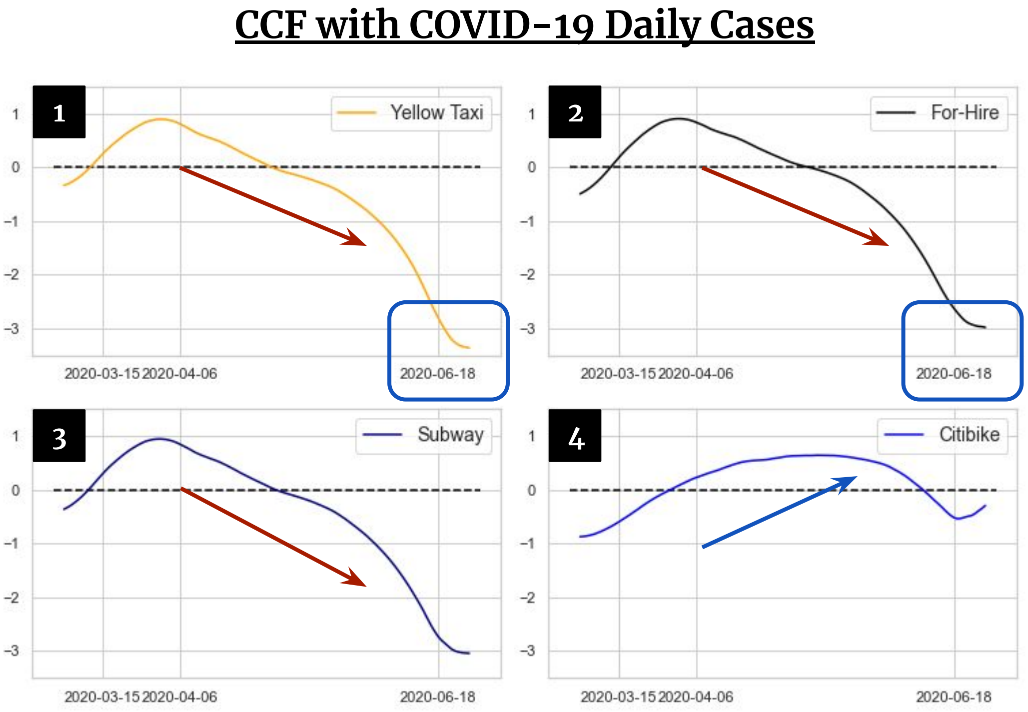
Approach 1:

- Cross correlation (CCF) between COVID data and ridership during post-COVID

Approach 2:

- Build a pre-COVID model that captures the trend and seasonality of ridership and use it to predict ridership during post-COVID to see the difference (metrics: **MSE**)

IV. Results – Approach 1 (Cross Correlation Function)



1) Highly Similar CCFs for 2 & 3:
For-Hire Vehicles and Subway are affected by COVID-19 in a very similar fashion.

2) Higher tail for 2 compared to 1:
For-Hire Vehicles shows a sign of quicker recovery than Yellow Cabs (regions in blue boxes)

3) Upward Sloping CCF for 4:
After the outbreak peak at April 6, Citibike gained more popularity.

IV. Results – Data Split for Approach 2

Train Data: 2015.01~2019.10 (4.75 years)

Test Data: 2019.11~2020.2 (4 months)

first COVID-19 case

2015.01.01 | 2019.10.31 | 2020.02.29 | 2020.06.30

Pre-COVID Train (4 years and 9 months)

Pre-COVID Test (4 Months)

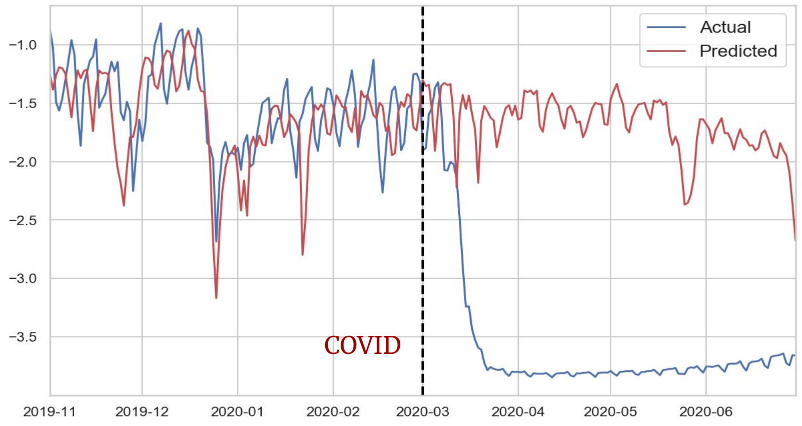
Post-COVID Test (4 Months)

*All data are standardized to make valid comparison across different datasets

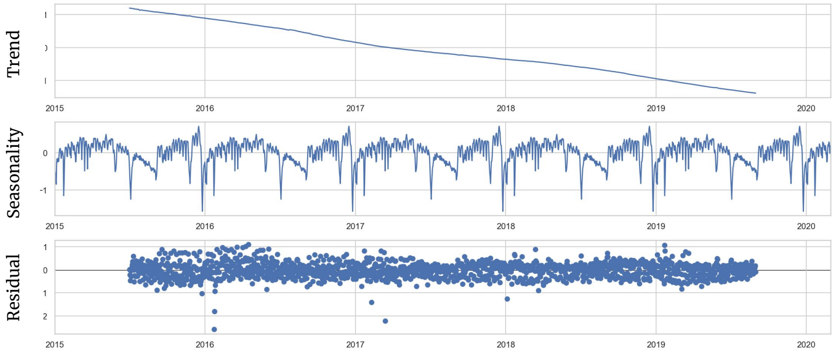
IV. Results – Approach 2 (ARMA)

- **Procedure**
  - Use Dickey Fuller Test to check stationarity
  - Fit on pre-COVID residual after detrend/de-seasonality
  - Use ACF, PACF & grid search (AIC) to determine p & q
- **For-hire vehicle is the most affected** by COVID.
  (Eg. Uber, Lyft)

Yellow Taxi with ARMA



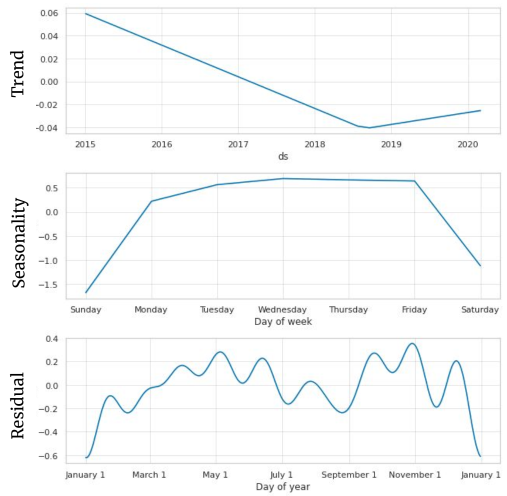
Detrend non-stationary Taxi data



	PreCOVID (MSE)	PostCOVID (MSE)	Diff (‱)
Yellow Taxi	0.1509	3.8689	▲24.64
For-Hire	0.2241	7.9733	▲34.58
Subway	1.6298	11.7228	▲6.19
CitiBike	2.468	3.525	▲0.43

*‱: basis point (percent per thousand)

IV. Results - Approach 2 (Facebook Prophet)

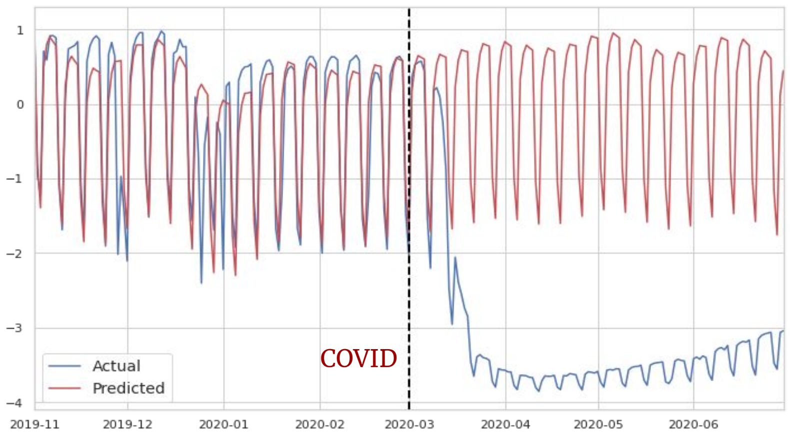


- Time series available for everyone (with many customizable options)

$$y(t) = g(t) + s(t) + h(t) + X(t)\beta + \epsilon_t$$

- **For-Hire Vehicles** is most severely affected by COVID-19
  (Eg. Uber, Lyft)

Subway with Prophet



*Reference: Forecasting at Scale (Taylor, et.al., 2017)

	PreCOVID (MSE)	PostCOVID (MSE)	Diff (‰)
Yellow Taxi	0.0806	3.8099	▲46.26
For-Hire	0.1403	7.7647	▲54.34
Subway	0.2700	12.0022	▲43.45
CitiBike	0.4695	2.3388	▲3.98

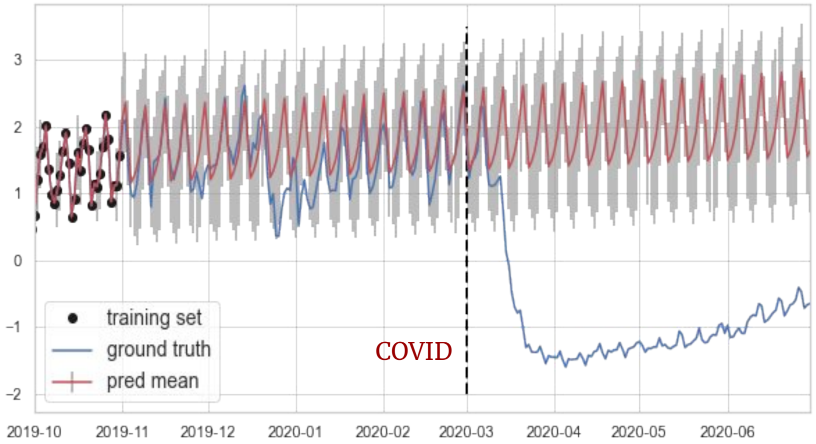
*‰: basis point (percent per thousand)

IV. Results – Approach 2 (Gaussian Process)

- **Kernel: RBF() + ExpSineSquared(*) x DotProduct() + ConstantKernel()**
  - For Yellow Taxi, For-Hire, and Subway, Periodicity = 7 reflects weekday seasonality
  - For Citibike, Periodicity = 365 reflects yearly seasonality
  - Multiplying **DotProduct Kernel** by **PeriodicKernel** captures linear trend with cycles

- Based on GP modelling, **For-Hire Vehicles** were most severely affected by COVID-19
  (Eg. Uber, Lyft)

For-Hire Vehicles with GP



	PreCOVID (MSE)	PostCOVID (MSE)	Diff (‰)
Yellow Taxi	0.1029	3.3475	▲31.54
For-Hire	0.1976	8.7102	▲43.09
Subway	0.3500	11.1570	▲30.88
CitiBike	0.3636	2.6036	▲6.16

*‰: basis point (percent per thousand)

IV. Results – Approach 2 (Kalman Filter/Smoother)

- **Least appropriate model** for our dataset (relatively large MSE values)

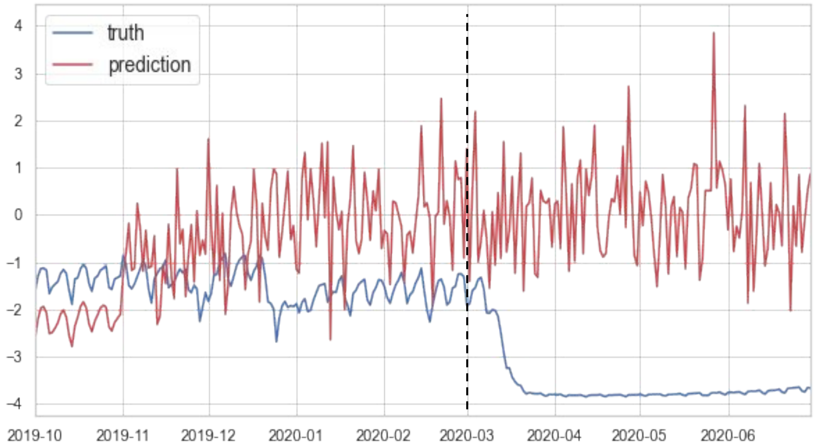
  [Training] Main goal of Kalman Filter/Smoother: find a sequence of "latent space"

  → Tends to mitigate the fluctuation coming from seasonality effect

  [Prediction] Random draw ~ N(μ, Σ,) produces volatile predictions (particularly for long-term)

<u>Yellow Taxi with KF</u>

COVID



	PreCOVID (MSE)	PostCOVID (MSE)	Diff (‱)
Yellow Taxi	2.6997	14.5950	▲4.41
For-Hire	2.3488	2.7412	▲0.17
Subway	*Not Performed*		
CitiBike			

*‱: basis point (percent per thousand)

IV. Results – Approach 2 (RNN)

Data preprocessing

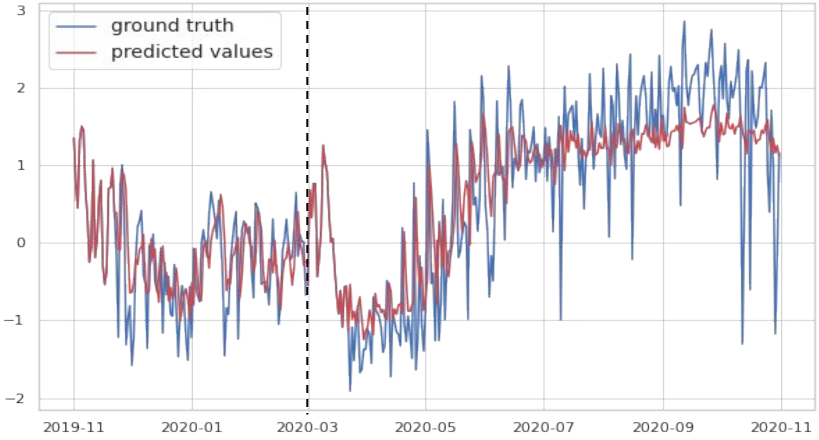
- Use sliding window to create data sequences and labels
- Window size is determined by Dickey Fuller test result on number of lags (e.g. 23 days)

Yellow Taxi is the most affected by COVID-19.

Model architecture & training

- 1 LSTM layer
- 1 Dropout layer (p=0.1)
- 1 Linear layer
- Adam Optimizer (lr=0.001)
- 2 Prediction models:
  point-wise (PW) & long-term (LR)

Citibike with RNN (PW)



	PreCOVID (MSE)	PostCOVID (MSE)	Diff (‰)
Yellow Taxi	0.0328	1.2395	▲36.76
For-Hire	0.0685	0.5086	▲6.42
Subway	0.2480	7.7452	▲30.23
CitiBike	0.2371	0.5186	▲1.19

COVID

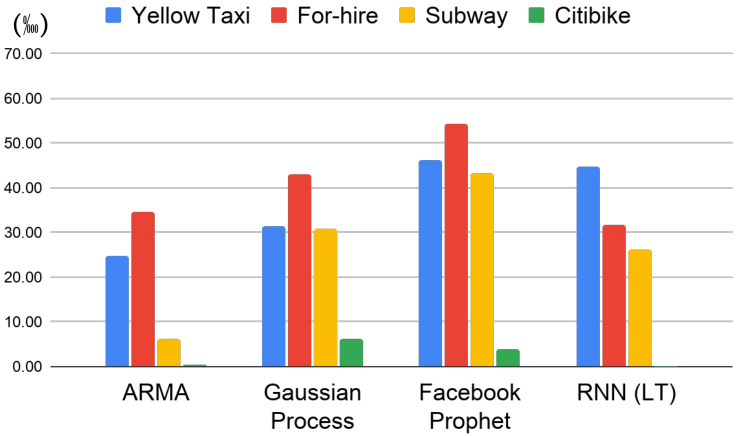
*Reference: Deep and Confident Prediction for Time Series at Uber (Zhu et. al.,2017)

V. Conclusion

Final Summary:

- Overall the level of impact that COVID outbreak has on transportation is as follows:
  Citibike < Subway < Yellow Taxi < For-hire Vehicles
- Unlike others, the demand for Citibike increases during the pandemic
- The results are expected because biking is an individualistic activity (more suitable during COVID time) while sharing the same vehicles (taxi, subway) is more dangerous and should be avoided.

Impacts of COVID on means of transportation
(results across all models)



Possible Next Steps:

- Forecasting the ridership demand for the unseen future in regression format
  ("If COVID daily cases increase by X%, the total ridership for OO will decrease by Y%.")
- Incorporating the effect of holidays and events that do not follow a periodic pattern into the model (e.g. Thanksgiving, the Super Bowl)

Thank you! :)

Any Questions?