



Taking off the Training Wheels

Rebalancing Citibike, A Time Series Analysis

By: Mitchell Krieger

Bike share systems like Citibike in NYC face a unique problem.





When bike docks are full,
riders can't return bikes.

When bike docks are empty,
there are no bikes to take out



System Balance:

- When a the Citibike System is balanced it means that bike are distributed so that stations don't *get too full nor too empty*.
- Signs of an unbalanced system:

Bike Pools:

Stations where riders deposit bikes but don't take them out

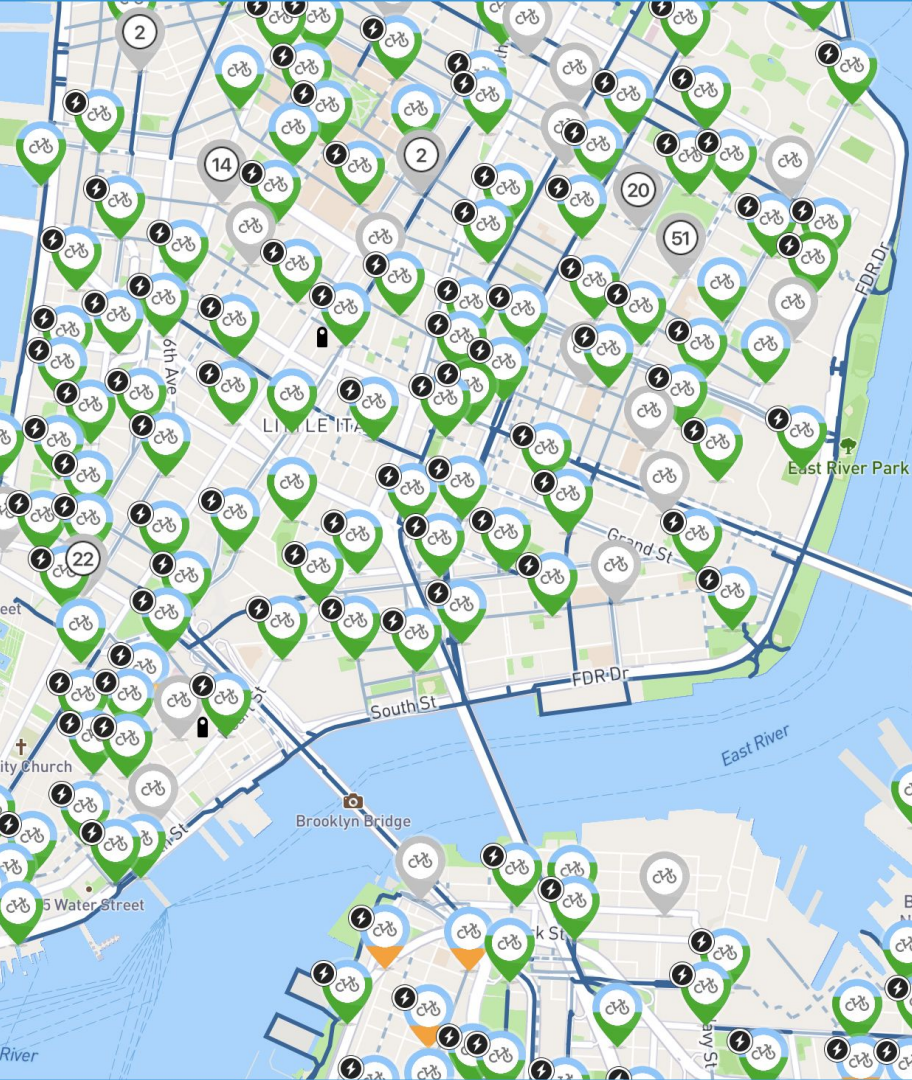
Causes too full stations



Bike Drains:

Stations where riders take bikes out but don't return them

Causes empty stations

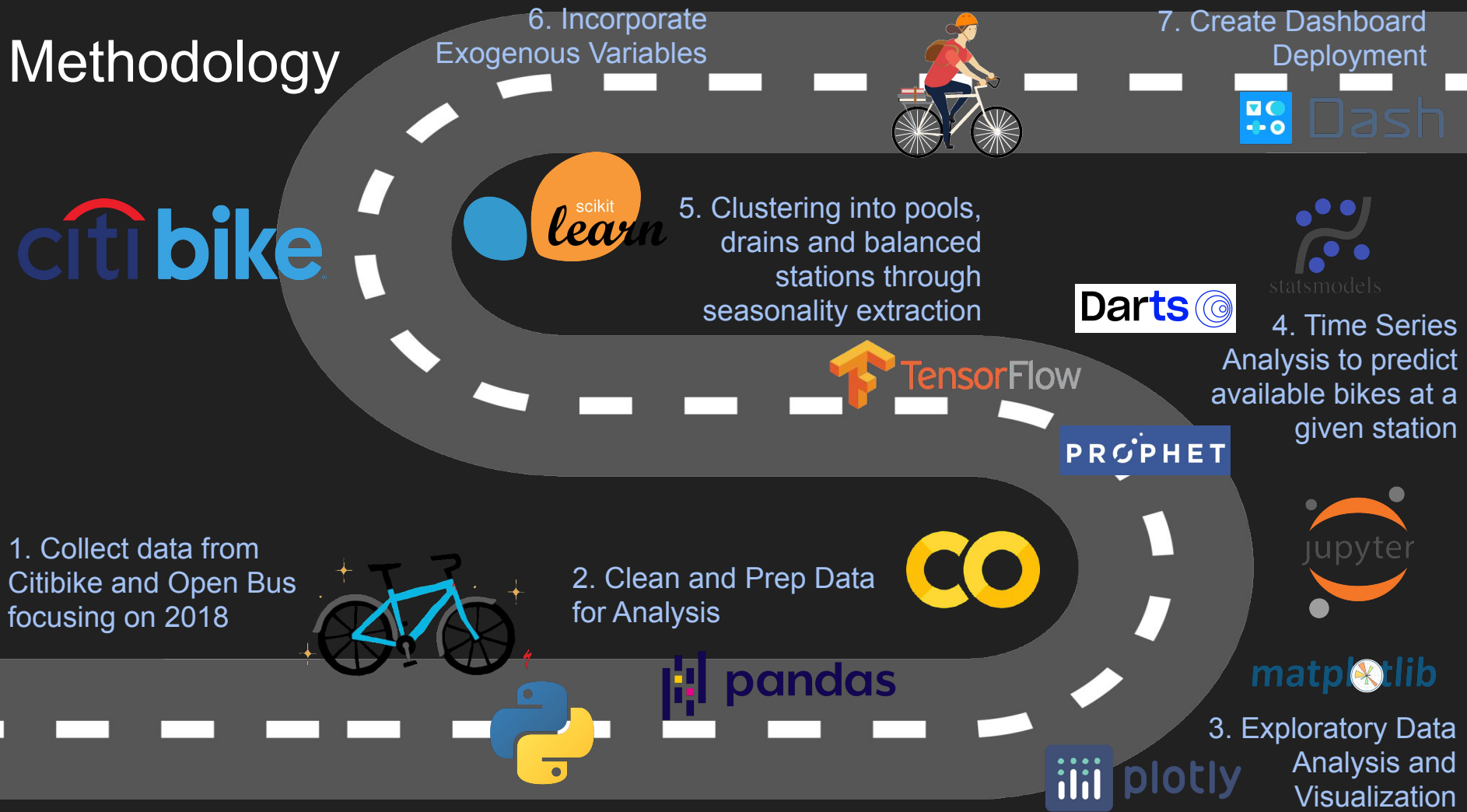


How can we ensure a balanced system?

To do so we'll need to be able to understand:

- System Traffic: How Users Ride
 - When and Where to Rebalance?
- Number of bikes at a given station
 - How much to re-balance?
- Station, Rider and other behavior that leads to bike drainage & poolage
 - Why did it become unbalanced?
 - Possible exogenous variables such as weather, elevation and holidays

Methodology

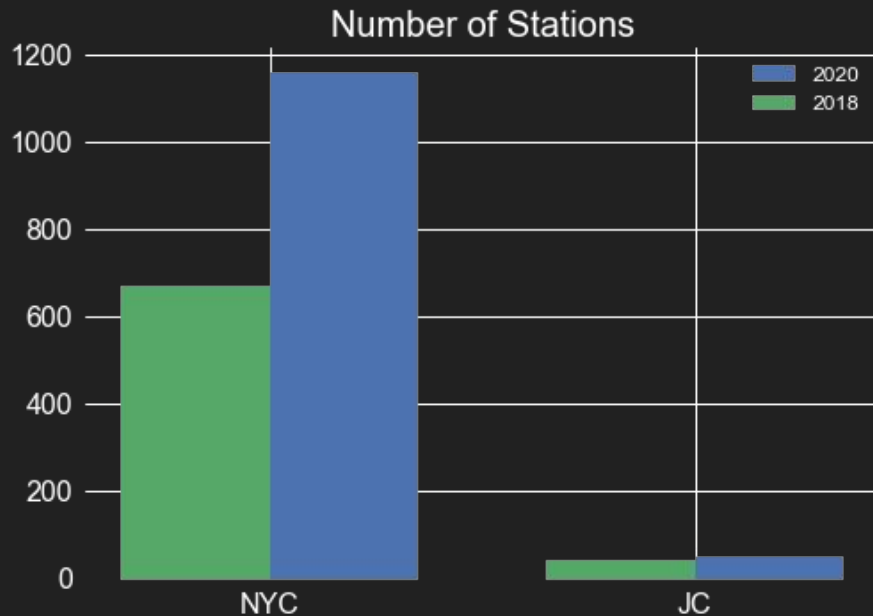


The Data:

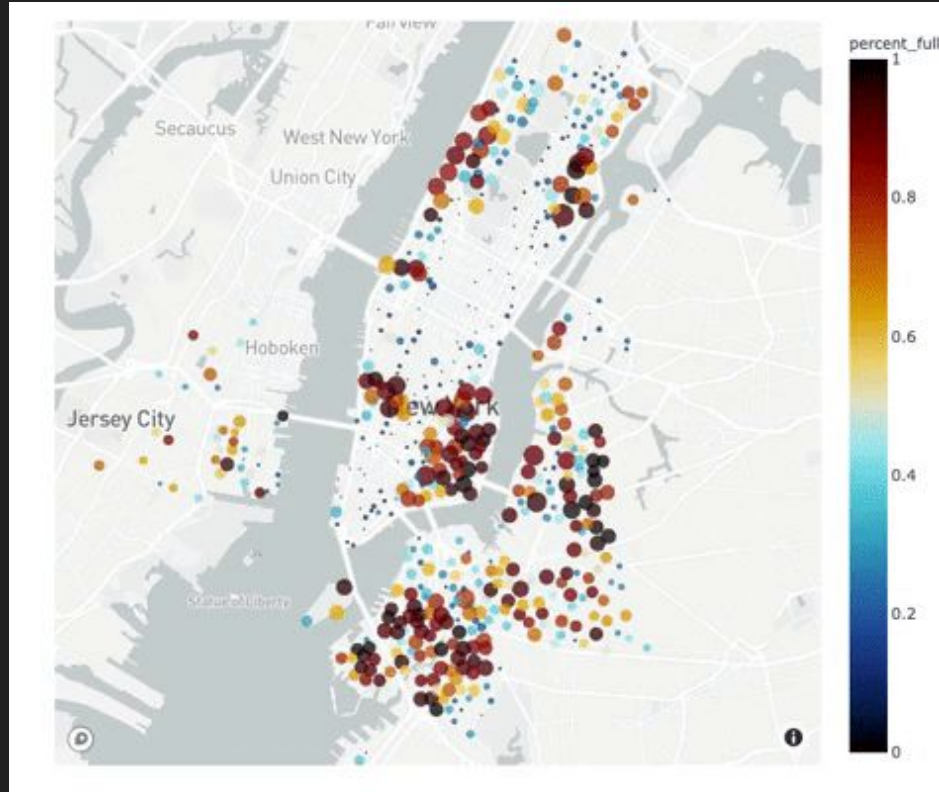
- Data Collected From Citbike Live Feed and Open Bus Project
- Focused Analysis on 2018
- Operated by Motivate, a subsidiary of Lyft

System Stats:

- Launched in 2013
- 1200+ stations between NYC & JC
 - Huge growth from 650+ stations in 2018
- 1.75 million trips in 2018
- Majority of trips are <15 minutes



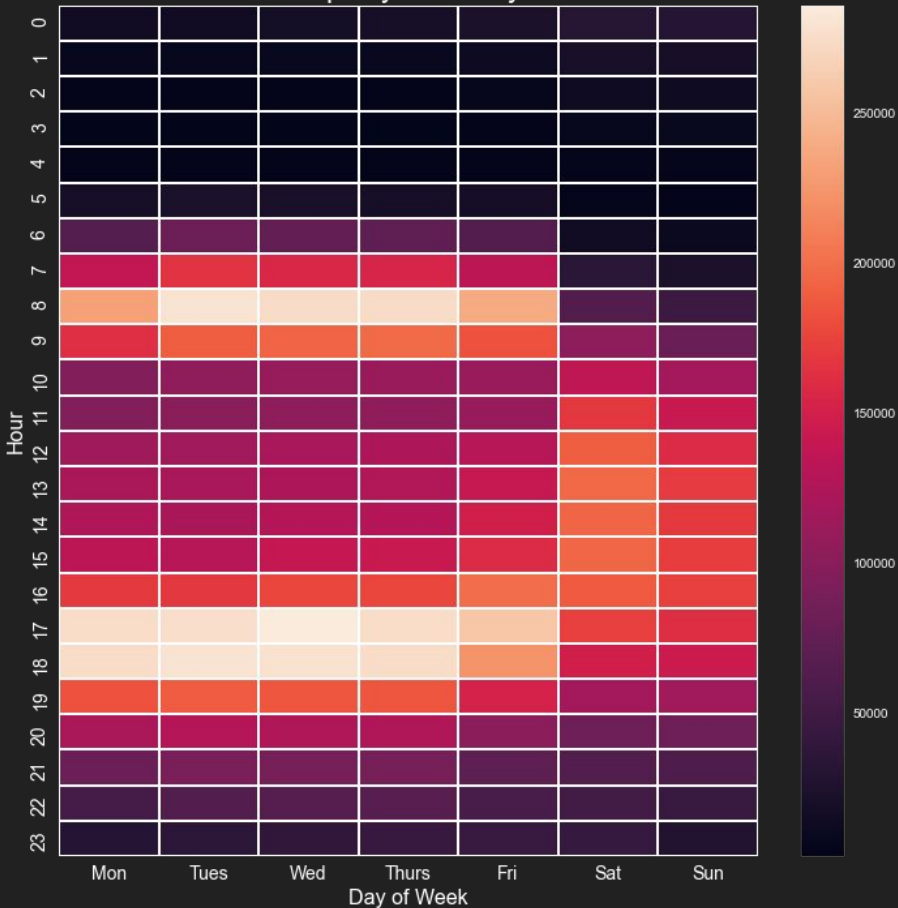
Visualizing the system in a typical weekday...



...we can see both seasonality and pools/drains.

Diving Deeper into Seasonality:

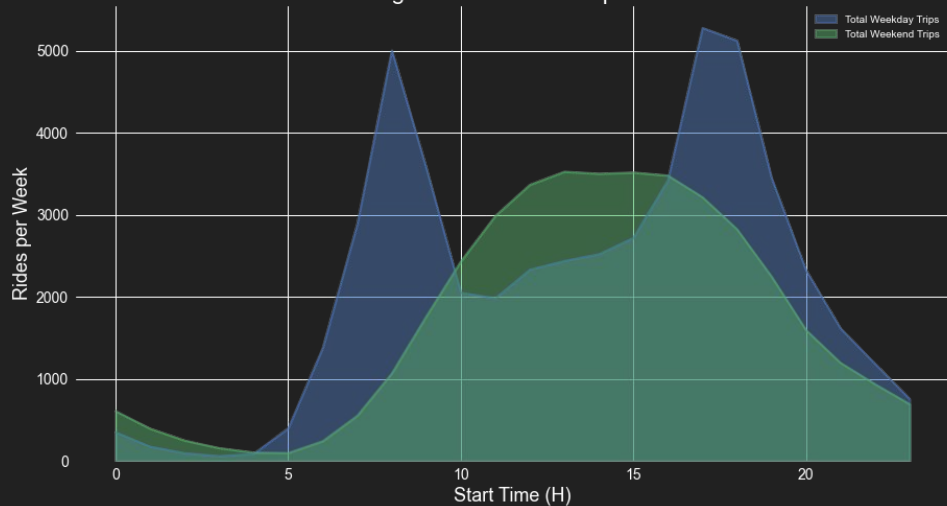
Trips by Weekday



There is both Weekly and Daily Seasonality present

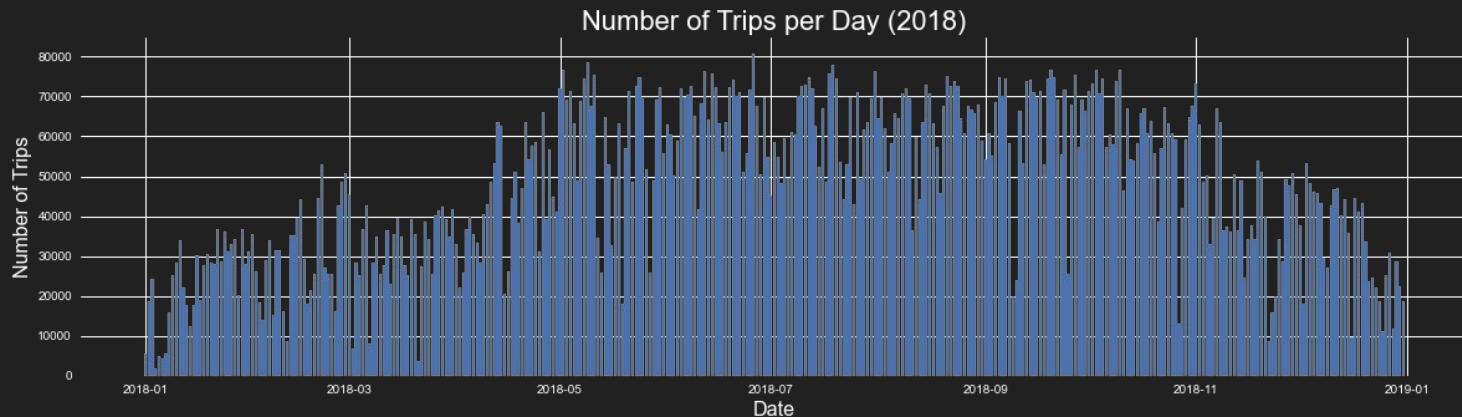
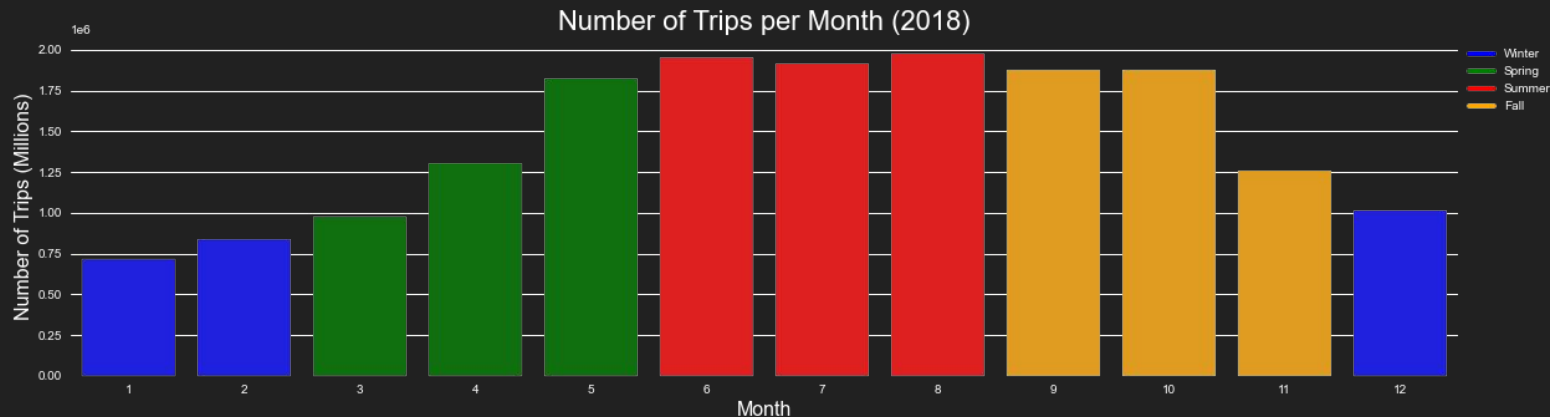
- Weekly:
 - Weekdays behave differently than weekends
 - Tues/Weds/Thurs are heaviest days
- Daily:
 - Peaks occur during rush hour commuting time
 - Bikes nearly “freeze” in place overnight

Average Number of Rides per Hour



Yearly Seasonality?

- Exists but is weak on an individual station basis
- Only apparent when aggregated by month
- May not actually be useful in modeling



Modeling

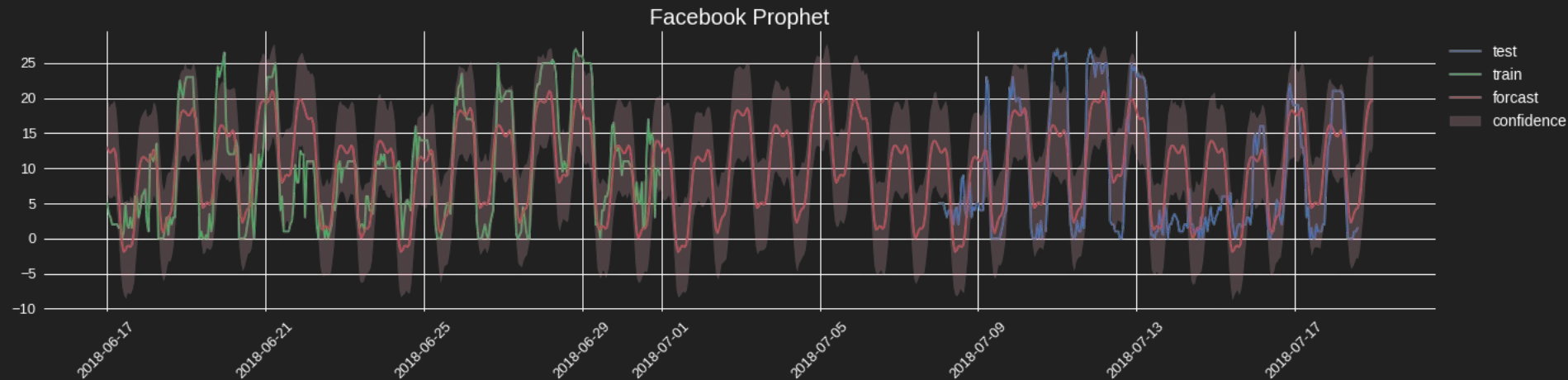
Goal: A model that captures the seasonality (weekly and daily) and closely predicts number of available bikes (low RMSE and MAE values).

Best version of each type of Model:

Model	Seasonality Captured	RMSE	MAE
Dummy ARIMA	None	8.97	8.05
SARIMA	Daily	9.71	7.77
LSTM	Daily	10.44	8.80
CNN/LSTM	None	4.88	3.63
Facebook Prophet	Daily, Weekly	6.30	5.15

The Facebook Prophet Model

- Closely Models both the train (green) and test (blue) sets
 - Much of the data is within the confidence interval
- Captures both daily and weekly seasonality extremely well
- Very distant forecasts also perform quite well

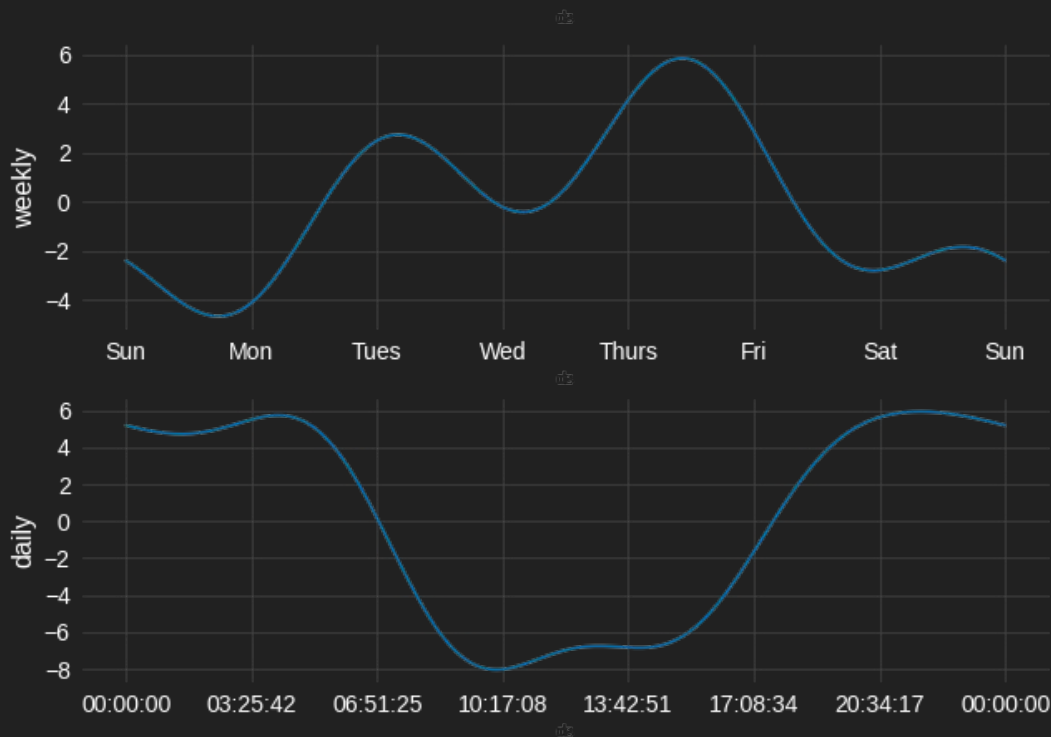


Seasonality Extraction

Captures patterns identified in the Exploratory Data Analysis:

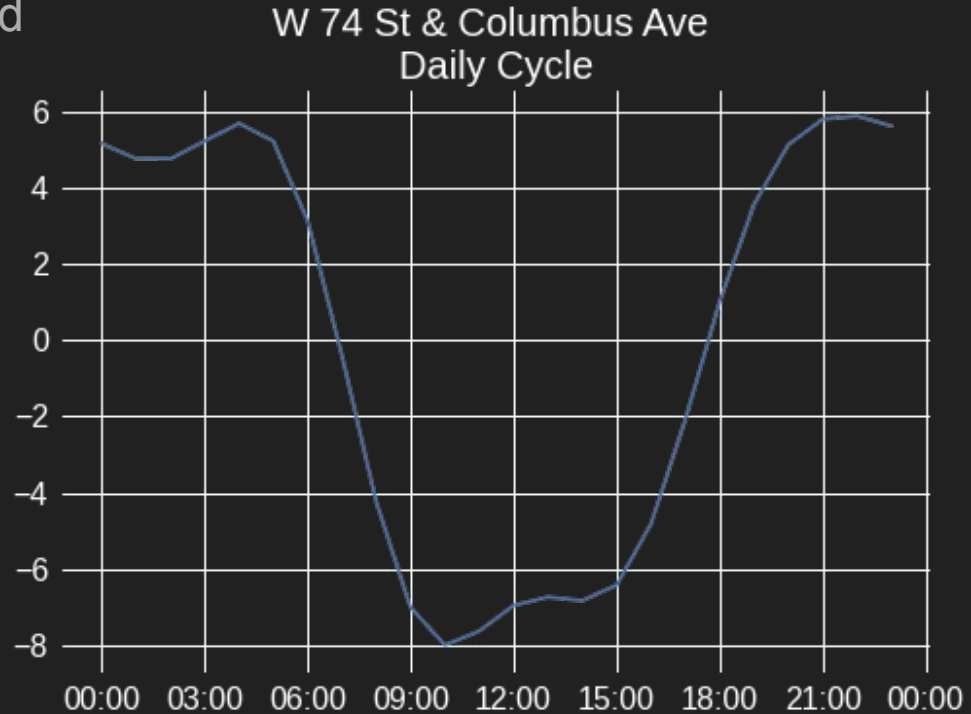
- **Weekly Season:**
 - Tues/Weds/Thurs are hot
 - Weekends behave differently
- **Daily Season:**
 - High activity during Rush Hour commuting times

We can use the daily season extraction to classify pools, drains and balanced stations.



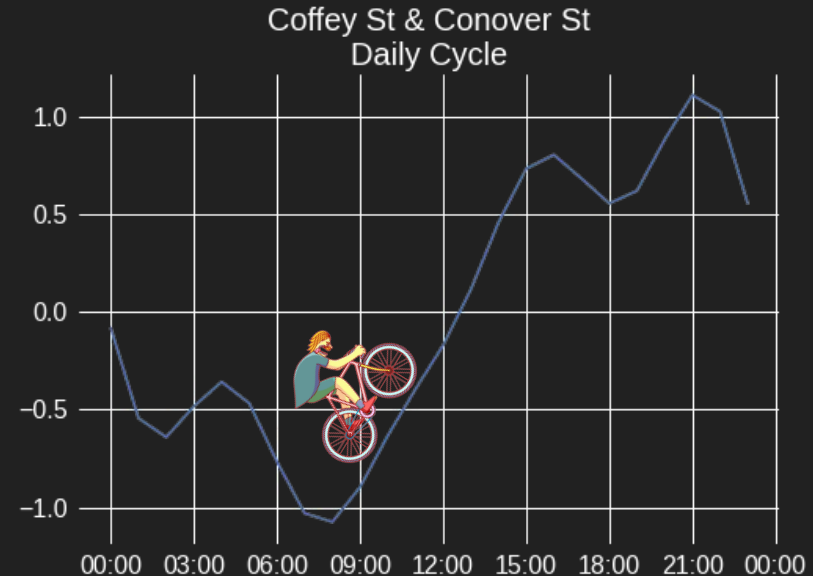
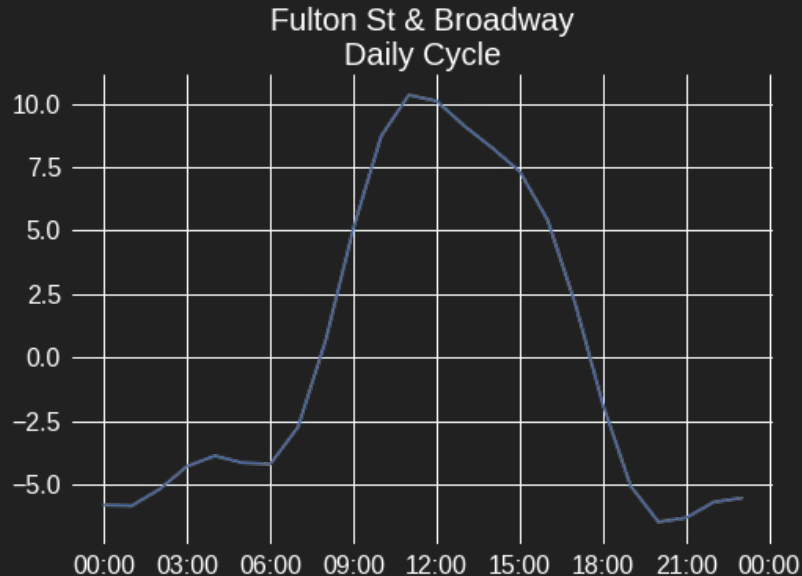
Balanced stations' extracted daily season should:

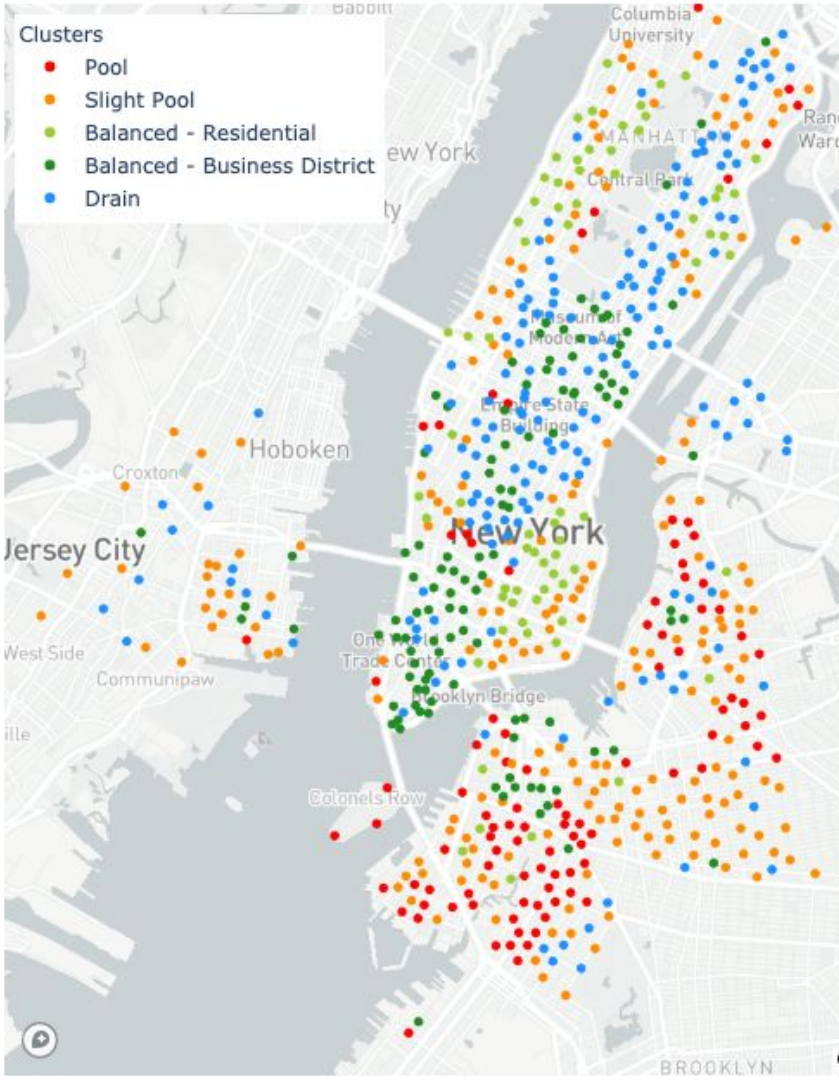
- Return to where they start after a cycle
- Have a curve centered mostly around zero in a:
 - U Shape (residential neighborhoods)
 - \cap Shape (business districts)



Unbalanced stations' extracted daily seasons could:

- Have a U or n curve centered far from zero
 - Center < 0 (drain)
 - Center > 0 (pool)
- Have a trend
 - Decreasing (drain)
 - Increasing (pool)
- Have a flat line
 - At Zero (station is unused)
 - Below Zero (drain)
 - Above Zero (pool)





Identifying stations as pools, drains, or balanced

- Implemented algorithm with 5 clusters to yield the most distinct groups of stations.
- Correctly identified key system trends:
 - Pooling in Brooklyn
 - Draining from midtown and the Upper East Side
 - Difference in seasonality between stations in residential districts and business districts
- Slight Pool cluster is not highly distinct, may be misclassification of pools & balanced

Conclusion

Now that stations are classified as pools, drains or balanced:

- Take advantage of seasonality to redistribute bikes strategic times:
 - During late night hours bikes are nearly frozen
 - The relative lull between the morning and evening commuter rush hour
 - Weekends
- Consider adding additional stations in areas with many pools/drains
- Distribute electric bikes strategically to battle elevation and long distances between stations
- Monitor changes in balancing as Citibike Continues to expand



Next Steps

- Refine Modeling/Clustering to better capture slight drains and slight pools
- Incorporate exogenous variables such as holidays, weather, electric bikes and elevation
- Collect data additional data and run analysis on 2020 as Citibike has expanded greatly since 2018
- Analyze the impact of COVID-19 on changes in rider behavior and station trends/clustering.



Thank you!

For more information:



- [GitHub Repository for this Project](#)
- [Citibike System Data](#)
- [Bloomberg 2014 profile on the mathematics of rebalancing](#)

Contact Information:

- Email: mitkrieger@gmail.com
- GitHub: @mitkrieg