



NYC CITIBIKE ANALYSIS

Scarlett Zuo

March 21, 2016

OUTLINE

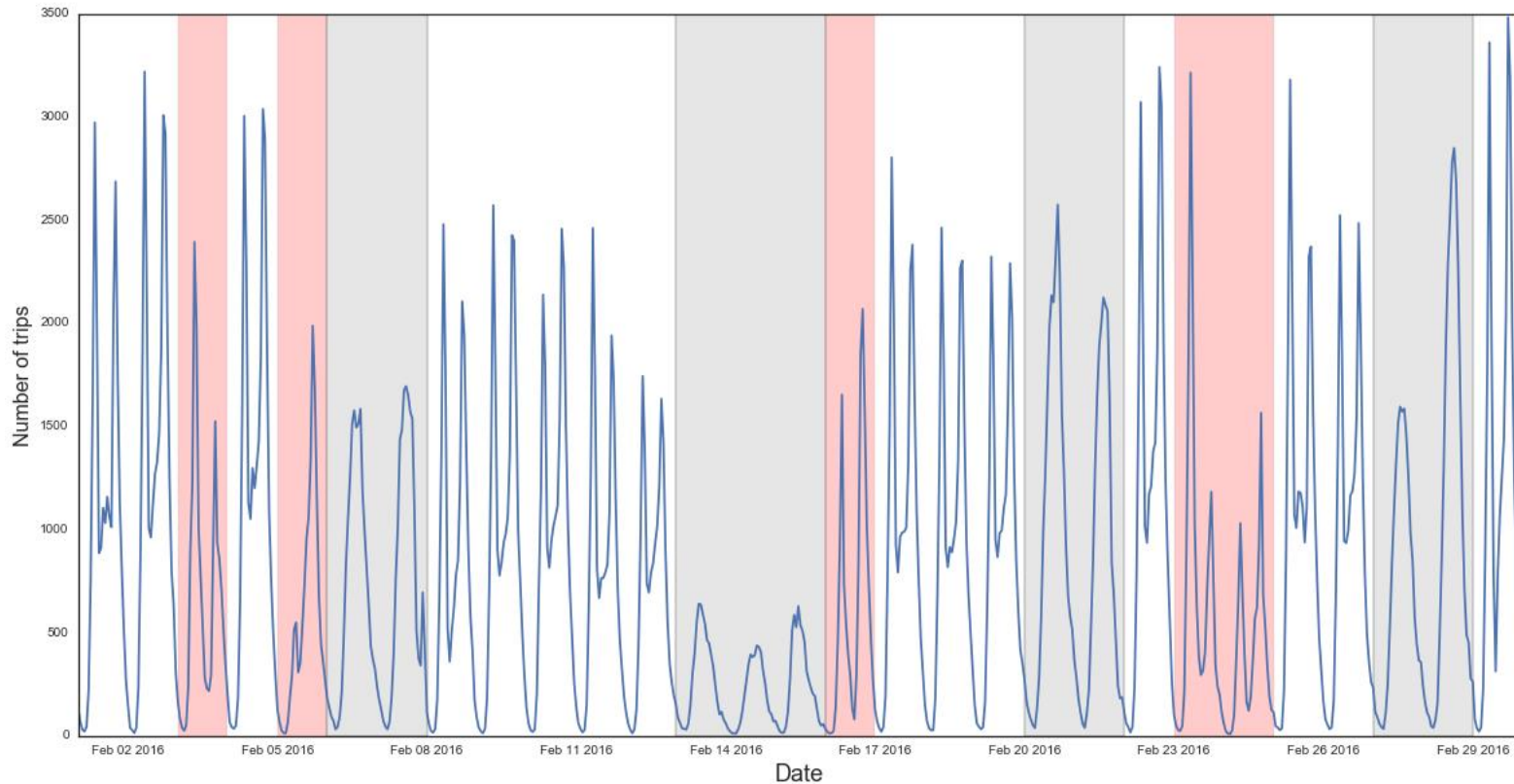
- Highlight of Findings
- Explore the Overall Trend
- A Closer Look at the Weekday and Weekend Patterns
- Station Clustering
 - Method
 - Results
- User Behavior Exploration
- Prediction and Precision
 - Stations require attendance
 - Other options

HIGHLIGHT OF THE FINDINGS

- Weekdays and weekends have different patterns
- 3 types of stations: Am-Peak-Only, Am-Pm-Peaks, Pm-Peak-Only
- Subscribers and non-subscribers have different behaviors
 - Subscribers: short trips <10 min, daily commute
 - Non-subscribers: 10-20 min trips

EXPLORE THE OVERALL TREND

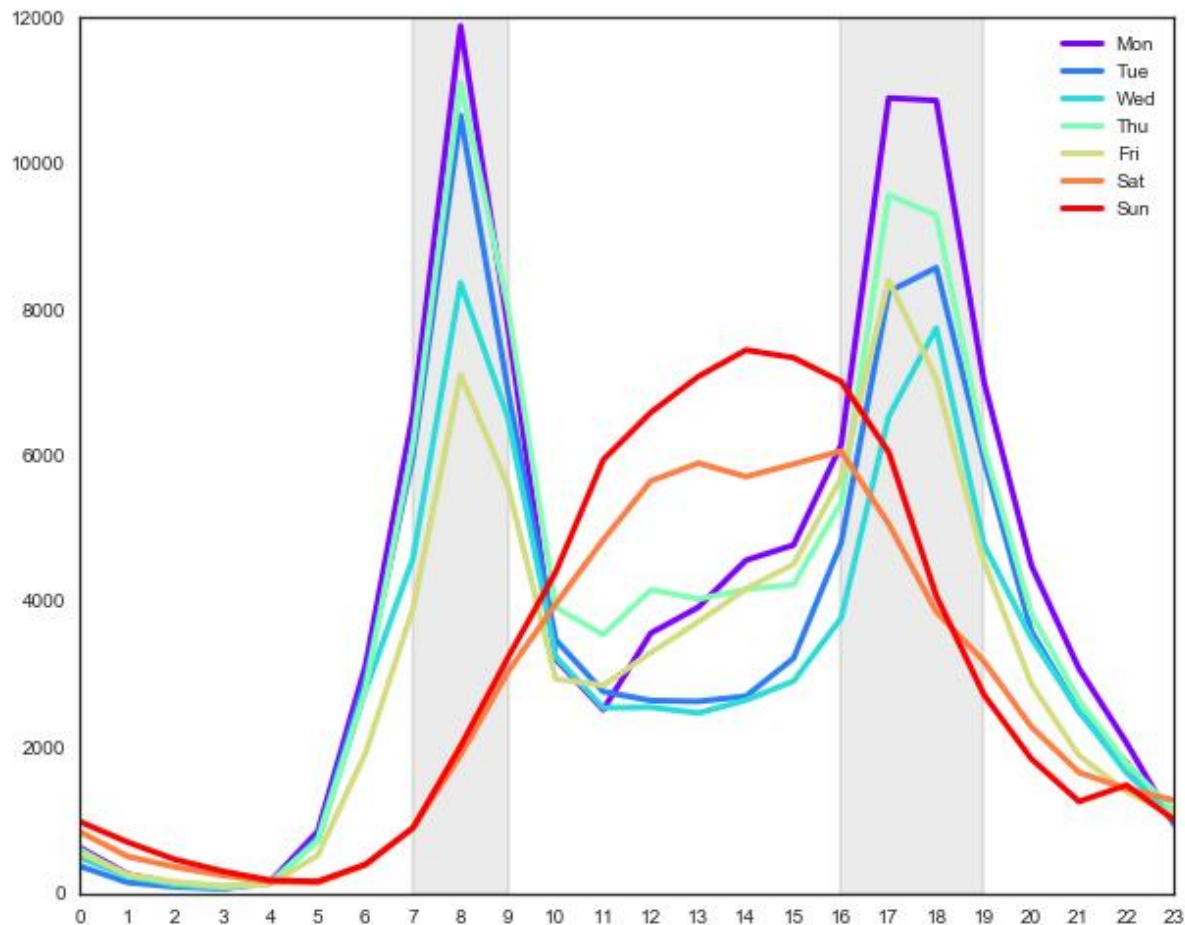
Total Bikeshare Trend



- Weekday patterns are different from weekend & holiday patterns.
- Bad weather lowers number of bike rental.

A CLOSER LOOK AT THE WEEKDAY AND WEEKEND PATTERNS

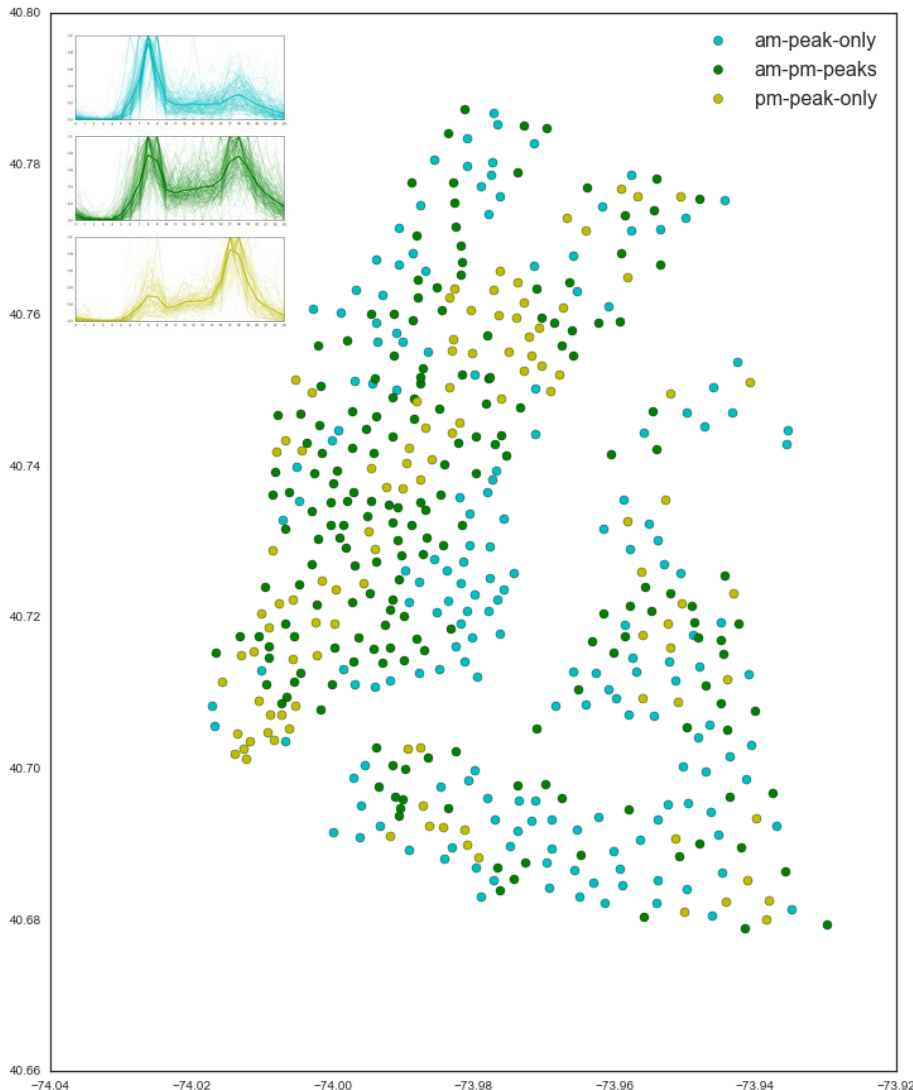
Weekday and Weekend Patterns



STATION CLUSTERING - METHOD

- Cathal Coffey – Capital Bikeshare
 - Am-Peak-Only
 - Am-Pm-Peaks
 - Pm-Peak-Only
- How about NYC CitiBike?
 - Method:
 - Separate weekdays and weekends
 - Aggregate trips taken by hour for each station, normalize counts
 - Use Euclidean distance and K-mean for clustering
 - Train on Feb 1 – Feb 20, and test set: Feb 21 – Feb 29
 - Metric: Normalized RMSE

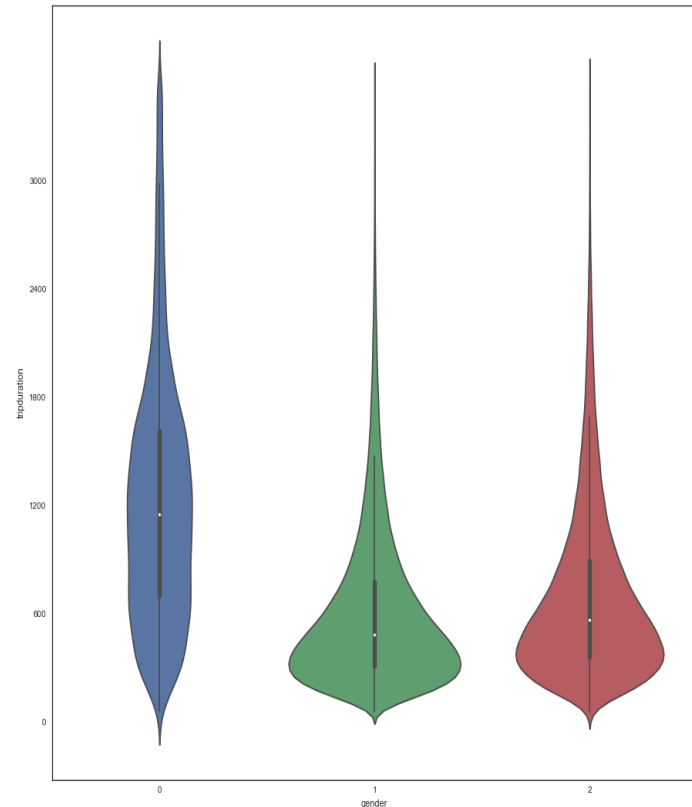
STATION CLUSTERING - RESULTS



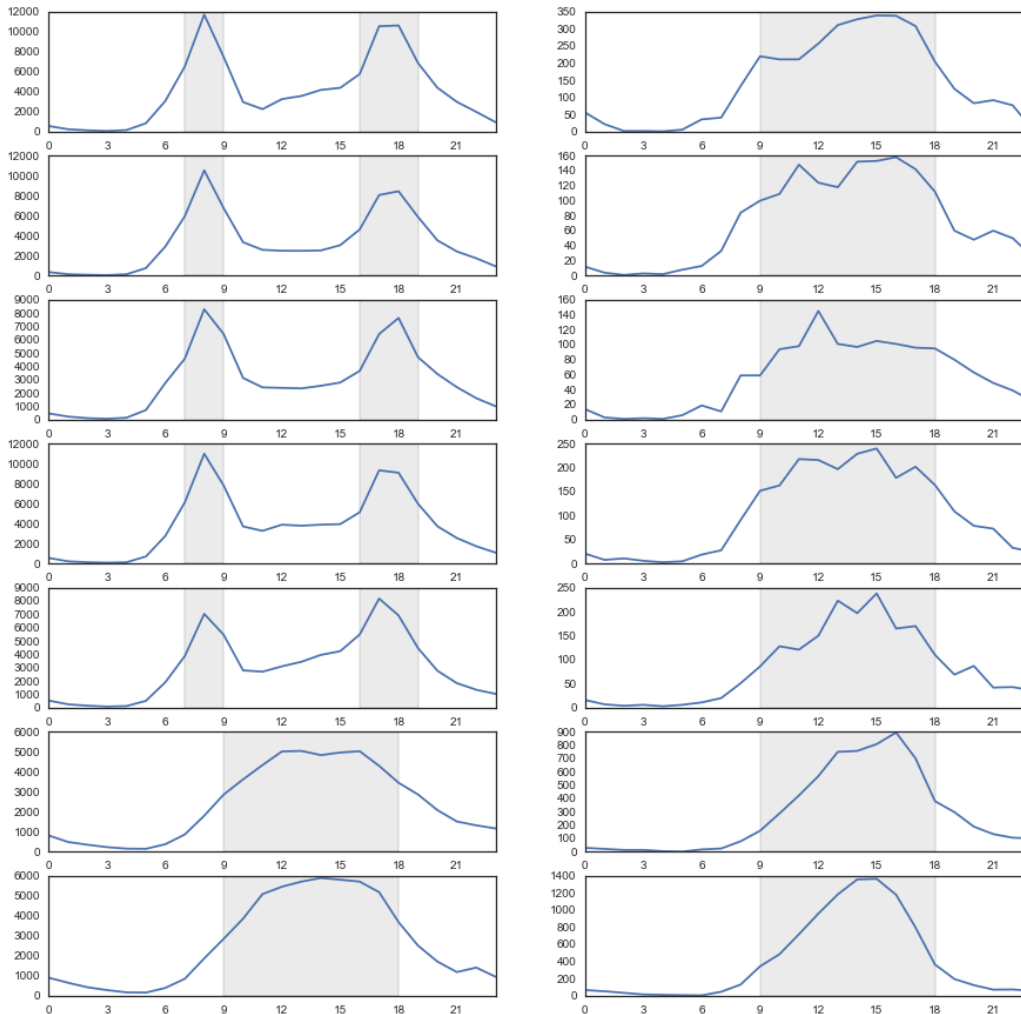
- Am-Peak-Only: near residential areas; they are used to commute to Pm-Peak-Only stations
- Pm-Peak-Only: work locations
- Am-Pm-Peaks: popular POIs / Metro stations
- Implications:
- Morning: bike and dock supply guarantee - transfer enough bikes to Am-Peak-Only stations from Pm-Peak-Only stations
- Evening: reverse

USER BEHAVIOR EXPLORATION

- Non-subscribers tend to take longer trips than subscribers: 75% trips are done in half an hour VS 15 minutes.
- For subscribers, most trips are done within 10 minutes, whereas non-subscribers make roughly the same amount of trips that last for 10 - 20 minutes.
- Little gender difference.



USER BEHAVIOR EXPLORATION (CONT.)



- During the weekdays, subscribers use bikes for daily commute.
- Non-subscribers' weekday pattern is similar to weekends.
- Non-subscribers have more activities during weekends than during the week.

PREDICTION AND PRECISION

- Prediction
 - Use the cluster centroids as baseline
 - Extrapolate for each station by the max count for that station.
- Precision – Normalized RMSE

| | Training Set | Test Set |
|----------|--------------|----------|
| Weekdays | 0.045 | 0.019 |
| Weekends | 0.056 | 0.033 |

- Results look good!

STATIONS NEED ATTENDANCE

- If out > in, need bikes
- If in > out + docks, need space (take bikes away)

| id | 72 | 79 | 82 | 83 | 116 | 119 | 120 | 127 | |
|----|------|------|------|------|-----|-----|------|-----|---|
| 0 | NaN | 3.0 | NaN | NaN | 1.0 | NaN | NaN | 2.0 | 6 |
| 1 | NaN | 1.0 | NaN | NaN | NaN | NaN | NaN | NaN | 3 |
| 2 | NaN | 1.0 | NaN | NaN | NaN | NaN | NaN | 1.0 | 2 |
| 3 | NaN | 1.0 | NaN | NaN | NaN | NaN | NaN | NaN | 1 |
| 4 | NaN | NaN | NaN | NaN | 1.0 | NaN | NaN | NaN | 1 |
| 5 | 1.0 | 2.0 | 1.0 | 1.0 | 6.0 | 1.0 | 1.0 | 6.0 | 4 |
| 6 | 3.0 | NaN | 4.0 | 4.0 | 4.0 | 2.0 | 5.0 | 4.0 | 1 |
| 7 | 8.0 | 1.0 | 7.0 | 8.0 | 6.0 | 4.0 | 10.0 | 6.0 | 4 |
| 8 | 16.0 | NaN | 16.0 | 15.0 | 3.0 | 9.0 | 21.0 | 4.0 | 1 |
| 9 | NaN | NaN | 15.0 | 14.0 | NaN | 4.0 | 11.0 | NaN | 1 |
| 10 | NaN | 6.0 | 6.0 | 5.0 | NaN | 1.0 | 3.0 | NaN | 1 |
| 11 | NaN | 9.0 | 5.0 | 4.0 | NaN | NaN | NaN | NaN | 2 |
| 12 | NaN | 11.0 | 5.0 | 5.0 | NaN | NaN | NaN | NaN | 2 |

| id | 72 | 79 | 82 | 83 | 116 | 119 | 120 | 127 | 128 | 137 | ... | 3232 | 3233 | 3235 | 3236 | 3237 | 323 |
|----|------|-----|-----|-----|-----|-----|-----|-----|-----|------|-----|------|-------|------|-------|------|-----|
| 0 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN | NaN | NaN | 13.0 | NaN | NaN |
| 1 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN | NaN | NaN | 3.0 | NaN | NaN |
| 2 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN | NaN | NaN | NaN | NaN | NaN |
| 3 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN | NaN | NaN | NaN | NaN | NaN |
| 4 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN | NaN | NaN | NaN | NaN | NaN |
| 5 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN | NaN | NaN | NaN | NaN | NaN |
| 6 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN | NaN | NaN | NaN | NaN | NaN |
| 7 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | 3.0 | ... | NaN | 29.0 | NaN | NaN | NaN | NaN |
| 8 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | 61.0 | ... | NaN | 107.0 | NaN | NaN | NaN | NaN |
| 9 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | 50.0 | ... | NaN | 93.0 | NaN | NaN | NaN | NaN |
| 10 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN | NaN | NaN | NaN | NaN | NaN |
| 11 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN | NaN | NaN | NaN | NaN | NaN |
| 12 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN | NaN | NaN | NaN | NaN | NaN |
| 13 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN | NaN | NaN | 3.0 | NaN | NaN |
| 14 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN | NaN | NaN | 12.0 | NaN | NaN |
| 15 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN | NaN | NaN | 21.0 | NaN | NaN |
| 16 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN | NaN | NaN | 42.0 | NaN | NaN |
| 17 | 9.0 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN | NaN | NaN | 121.0 | NaN | NaN |
| 18 | 18.0 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN | NaN | NaN | 184.0 | NaN | NaN |
| 19 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN | NaN | NaN | 156.0 | NaN | NaN |
| 20 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN | NaN | NaN | 100.0 | NaN | NaN |
| 21 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN | NaN | NaN | 70.0 | NaN | NaN |
| 22 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN | NaN | NaN | 54.0 | NaN | NaN |
| 23 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN | NaN | NaN | 26.0 | NaN | NaN |

OTHER MODELING OPTIONS

- Separate bad-weather days to form another training set for bad-weather baseline
- Try Dynamic Time Warping instead of Euclidean distance for better non-linear alignment between time series.

Thank you!

