Minnesota Department of Transportation Traffic Project

Nathan Wodarz

September 2021

Problem Statement

Outline

Problem Statement

Background: MnDOT

- **Minnesota Department of Transportation** (MnDOT)
- Minnesota ranks 4th of 50
  - Centerline mileage
  - Lane mileage
- Road volume
  - Automatic Traffic Recorders (ATR) and Weigh-in-Motion (WIM)
  - Over 155 total
    - 75+ in Minneapolis-St. Paul metro area (7 counties)
    - 80+ outstate (80 counties)
  - Available 2002-present

Problem Statement

Background: MnDOT

- **Minnesota Department of Transportation** (MnDOT)
- Minnesota ranks 4th of 50
  - Centerline mileage
  - Lane mileage
- Road volume
  - Automatic Traffic Recorders (ATR) and Weigh-in-Motion (WIM)
  - Over 155 total
    - 75+ in Minneapolis-St. Paul metro area (7 counties)
    - 80+ outstate (80 counties)
  - Available 2002-present

Problem Statement

Background: MnDOT

- **Minnesota Department of Transportation** (MnDOT)
- Minnesota ranks 4th of 50
  - Centerline mileage
  - Lane mileage
- Road volume
  - Automatic Traffic Recorders (ATR) and Weigh-in-Motion (WIM)
  - Over 155 total
    - 75+ in Minneapolis-St. Paul metro area (7 counties)
    - 80+ outstate (80 counties)
  - Available 2002-present

Problem Statement

Background: MnDOT

- **Minnesota Department of Transportation** (MnDOT)
- Minnesota ranks 4th of 50
  - Centerline mileage
  - Lane mileage
- Road volume
  - Automatic Traffic Recorders (ATR) and Weigh-in-Motion (WIM)
  - Over 155 total
    - 72+ in Minneapolis-St. Paul metro area (7 counties)
    - 80+ outstate (80 counties)
  - Available 2002-present

Problem Statement

Background: MnDOT

- **Minnesota Department of Transportation** (MnDOT)
- Minnesota ranks 4th of 50
  - Centerline mileage
  - Lane mileage
- Road volume
  - Automatic Traffic Recorders (ATR) and Weigh-in-Motion (WIM)
  - Over 155 total
    - 75+ in Minneapolis-St. Paul metro area (7 counties)
    - 80+ outstate (80 counties)
  - Available 2002-present

Problem Statement

Background: MnDOT

- **Minnesota Department of Transportation** (MnDOT)
- Minnesota ranks 4th of 50
  - Centerline mileage
  - Lane mileage
- Road volume
  - Automatic Traffic Recorders (ATR) and Weigh-in-Motion (WIM)
  - Over 155 total
    - 75+ in Minneapolis-St. Paul metro area (7 counties)
    - 80+ outstate (80 counties)
  - Available 2002-present

Problem Statement

Background: MnDOT

- **Minnesota Department of Transportation** (MnDOT)
- Minnesota ranks 4th of 50
  - Centerline mileage
  - Lane mileage
- Road volume
  - Automatic Traffic Recorders (ATR) and Weigh-in-Motion (WIM)
  - Over 155 total
    - 75+ in Minneapolis-St. Paul metro area (7 counties)
    - 80+ outstate (80 counties)
  - Available 2002-present

Problem Statement

Background: MnDOT

- **Minnesota Department of Transportation** (MnDOT)
- Minnesota ranks 4th of 50
  - Centerline mileage
  - Lane mileage
- Road volume
  - Automatic Traffic Recorders (ATR) and Weigh-in-Motion (WIM)
  - Over 155 total
    - 75+ in Minneapolis-St. Paul metro area (7 counties)
    - 80+ outstate (80 counties)
  - Available 2002-present

Problem Statement

Background: MnDOT

- **Minnesota Department of Transportation** (MnDOT)
- Minnesota ranks 4th of 50
  - Centerline mileage
  - Lane mileage
- Road volume
  - Automatic Traffic Recorders (ATR) and Weigh-in-Motion (WIM)
  - Over 155 total
    - 75+ in Minneapolis-St. Paul metro area (7 counties)
    - 80+ outstate (80 counties)
  - Available 2002-present

Problem Statement

Background: MnDOT

- **Minnesota Department of Transportation** (MnDOT)
- Minnesota ranks 4th of 50
  - Centerline mileage
  - Lane mileage
- Road volume
  - Automatic Traffic Recorders (ATR) and Weigh-in-Motion (WIM)
  - Over 155 total
    - 75+ in Minneapolis-St. Paul metro area (7 counties)
    - 80+ outstate (80 counties)
  - Available 2002-present

Outline

Problem Statement
Motivation

- Interest in geography/road network
- Interest in time series
- Availability of data

Problem Statement

Motivation

- Interest in geography/road network
- Interest in time series
- Availability of data

Problem Statement

Motivation

- Interest in geography/road network
- Interest in time series
- Availability of data

Outline

Problem Statement

Problem Statement

- Fit model to existing data
- Predict future traffic levels

Problem Statement

Problem Statement

- Fit model to existing data
- Predict future traffic levels

Data Wrangling

Outline

Data Wrangling
Raw Data: MnDOT Data Products

- MnDOT Data Products
  - .csv format (2017-)
  - .txt format (2002-2017)
- Hourly values
  - One row per station per direction per day
  - More recently, also per lane
  - 24 hourly totals per row
  - Some values are estimated

Data Wrangling
Raw Data: MnDOT Data Products

- MnDOT Data Products
  - .csv format (2017-)
  - .txt format (2002-2017)
- Hourly values
  - One row per station per direction per day
  - More recently, also per lane
  - 24 hourly totals per row
  - Some values are estimated

Data Wrangling

Raw Data: MnDOT Data Products

- MnDOT Data Products
  - .csv format (2017-)
  - .txt format (2002-2017)
- Hourly values
  - One row per station per direction per day
  - More recently, also per lane
  - 24 hourly totals per row
  - Some values are estimated

Data Wrangling

Raw Data: MnDOT Data Products

- MnDOT Data Products
  - .csv format (2017-)
  - .txt format (2002-2017)
- Hourly values
  - One row per station per direction per day
  - More recently, also per lane
  - 24 hourly totals per row
  - Some values are estimated

Data Wrangling

Raw Data: MnDOT Data Products

- MnDOT Data Products
  - .csv format (2017-)
  - .txt format (2002-2017)
- Hourly values
  - One row per station per direction per day
  - More recently, also per lane
  - 24 hourly totals per row
  - Some values are estimated

Data Wrangling

Raw Data: MnDOT Data Products

- MnDOT Data Products
  - .csv format (2017-)
  - .txt format (2002-2017)
- Hourly values
  - One row per station per direction per day
  - More recently, also per lane
  - 24 hourly totals per row
  - Some values are estimated

Data Wrangling

Raw Data: MnDOT Data Products

- MnDOT Data Products
  - .csv format (2017-)
  - .txt format (2002-2017)
- Hourly values
  - One row per station per direction per day
  - More recently, also per lane
  - 24 hourly totals per row
  - Some values are estimated

Data Wrangling

Raw Data: MnDOT Data Products

- MnDOT Data Products
  - .csv format (2017-)
  - .txt format (2002-2017)
- Hourly values
  - One row per station per direction per day
  - More recently, also per lane
  - 24 hourly totals per row
  - Some values are estimated

Data Wrangling
Raw Data: ATR Stations

- Location
  - Rural vs Urban
- Functional Class
  - Interstates
  - Principal Arterial - Other Freeways and Expressways
  - Principal Arterial - Other
  - Minor Arterial
  - Major Collector
  - Local

Data Wrangling
Raw Data: ATR Stations

- Location
  - Rural vs Urban
- Functional Class
  - Interstates
  - Principal Arterial - Other Freeways and Expressways
  - Principal Arterial - Other
  - Minor Arterial
  - Major Collector
  - Local

Data Wrangling
Raw Data: ATR Stations

- Location
  - Rural vs Urban
- Functional Class
  - Interstates
  - Principal Arterial - Other Freeways and Expressways
  - Principal Arterial - Other
  - Minor Arterial
  - Major Collector
  - Local

Data Wrangling
Raw Data: ATR Stations

- Location
  - Rural vs Urban
- Functional Class
  - Interstates
  - Principal Arterial - Other Freeways and Expressways
  - Principal Arterial - Other
  - Minor Arterial
  - Major Collector
  - Local

Data Wrangling
Raw Data: ATR Stations

- Location
  - Rural vs Urban
- Functional Class
  - Interstates
  - Principal Arterial - Other Freeways and Expressways
  - Principal Arterial - Other
  - Minor Arterial
  - Major Collector
  - Local

Data Wrangling
Raw Data: ATR Stations

- Location
  - Rural vs Urban
- Functional Class
  - Interstates
  - Principal Arterial - Other Freeways and Expressways
  - Principal Arterial - Other
  - Minor Arterial
  - Major Collector
  - Local

Data Wrangling
Raw Data: ATR Stations

- Location
  - Rural vs Urban
- Functional Class
  - Interstates
  - Principal Arterial - Other Freeways and Expressways
  - Principal Arterial - Other
  - Minor Arterial
  - Major Collector
  - Local

Data Wrangling

Raw Data: ATR Stations

- Location
  - Rural vs Urban
- Functional Class
  - Interstates
  - Principal Arterial - Other Freeways and Expressways
  - Principal Arterial - Other
  - Minor Arterial
  - Major Collector
  - Local

Data Wrangling
Raw Data: ATR Stations

- Location
  - Rural vs Urban
- Functional Class
  - Interstates
  - Principal Arterial - Other Freeways and Expressways
  - Principal Arterial - Other
  - Minor Arterial
  - Major Collector
  - Local

Data Cleaning

Data Cleaning

- Remove duplication
- Remove inactive stations
- Remove stations with no data in last year
- Remove stations missing more than 80% of all months
  - January 2002-July 2021
  - Missing: no entries for month

Data Cleaning

- Remove duplication
- Remove inactive stations
- Remove stations with no data in last year
- Remove stations missing more than 80% of all months
  - January 2002-July 2021
  - Missing: no entries for month

Data Cleaning

- Remove duplication
- Remove inactive stations
- Remove stations with no data in last year
- Remove stations missing more than 80% of all months
  - January 2002-July 2021
  - Missing: no entries for month

Data Cleaning

- Remove duplication
- Remove inactive stations
- Remove stations with no data in last year
- Remove stations missing more than 80% of all months
  - January 2002-July 2021
  - Missing: no entries for month

Data Cleaning

- Remove duplication
- Remove inactive stations
- Remove stations with no data in last year
- Remove stations missing more than 80% of all months
  - January 2002-July 2021
  - Missing: no entries for month

Data Cleaning

- Remove duplication
- Remove inactive stations
- Remove stations with no data in last year
- Remove stations missing more than 80% of all months
  - January 2002-July 2021
  - Missing: no entries for month

Transformation

Transformation

: MADT

- Traffic counts given by **Annual Average Daily Traffic** (AADT)
  - Weighted mean of **Monthly Average Daily Traffic** (MADT)
  - Weights: number of days in month

  - m: month; j: day of week; h: hour of day
  - w_{jm}: ocurrances of jth day of week in month m
  - n_{hjm}: non-missing values for hth hour of jth day of week in month m
  - VOL_{ihjm}: ith data point for hth hour of jth day of week in month m
  - $MADT_m$: monthly average daily traffic for month m

Transformation

: MADT

- Traffic counts given by **Annual Average Daily Traffic** (AADT)
  - Weighted mean of **Monthly Average Daily Traffic** (MADT)
  - Weights: number of days in month

- m: month; j: day of week; h: hour of day
- w_{jm}: ocurrances of jth day of week in month m
- n_{hjm}: non-missing values for hth hour of jth day of week in month m
- VOL_{ihjm}: ith data point for hth hour of jth day of week in month m
- $MADT_m$: monthly average daily traffic for month m

Transformation

: MADT

- Traffic counts given by **Annual Average Daily Traffic** (AADT)
  - Weighted mean of **Monthly Average Daily Traffic** (MADT)
  - Weights: number of days in month

- m: month; j: day of week; h: hour of day
- w_{jm}: ocurrances of jth day of week in month m
- n_{hjm}: non-missing values for hth hour of jth day of week in month m
- VOL_{ihjm}: ith data point for hth hour of jth day of week in month m
- $MADT_m$: monthly average daily traffic for month m

Transformation

: MADT

- Traffic counts given by **Annual Average Daily Traffic** (AADT)
  - Weighted mean of **Monthly Average Daily Traffic** (MADT)
  - Weights: number of days in month

$$MADT_m = \frac{\sum_{j=1}^{7} w_{jm} \sum_{h=1}^{24} \left[\frac{1}{n_{hjm}} \sum_{i=1}^{n_{hjm}} VOL_{ihjm} \right]}{\sum_{j=1}^{7} w_{jm}}$$

- m: month; j: day of week; h: hour of day
- w_{jm}: ocurrances of jth day of week in month m
- n_{hjm}: non-missing values for hth hour of jth day of week in month m
- VOL_{ihjm}: ith data point for hth hour of jth day of week in month m
- $MADT_m$: monthly average daily traffic for month m

Transformation

: MADT

- Traffic counts given by **Annual Average Daily Traffic** (AADT)
  - Weighted mean of **Monthly Average Daily Traffic** (MADT)
  - Weights: number of days in month

$$MADT_m = \frac{\sum_{j=1}^{7} w_{jm} \sum_{h=1}^{24} \left[\frac{1}{n_{hjm}} \sum_{i=1}^{n_{hjm}} VOL_{ihjm} \right]}{\sum_{j=1}^{7} w_{jm}}$$

- m: month; j: day of week; h: hour of day
- w_{jm}: ocurrances of jth day of week in month m
- n_{hjm}: non-missing values for hth hour of jth day of week in month m
- VOL_{ihjm}: ith data point for hth hour of jth day of week in month m
- $MADT_m$: monthly average daily traffic for month m

Transformation

: MADT

- Traffic counts given by **Annual Average Daily Traffic** (AADT)
  - Weighted mean of **Monthly Average Daily Traffic** (MADT)
  - Weights: number of days in month

$$MADT_m = \frac{\sum_{j=1}^{7} w_{jm} \sum_{h=1}^{24} \left[\frac{1}{n_{hjm}} \sum_{i=1}^{n_{hjm}} VOL_{ihjm} \right]}{\sum_{j=1}^{7} w_{jm}}$$

- m: month; j: day of week; h: hour of day
- w_{jm}: ocurrances of jth day of week in month m
- n_{hjm}: non-missing values for hth hour of jth day of week in month m
- VOL_{ihjm}: ith data point for hth hour of jth day of week in month m
- $MADT_m$: monthly average daily traffic for month m

Transformation

: MADT

- Traffic counts given by **Annual Average Daily Traffic** (AADT)
  - Weighted mean of **Monthly Average Daily Traffic** (MADT)
  - Weights: number of days in month

$$MADT_m = \frac{\sum_{j=1}^{7} w_{jm} \sum_{h=1}^{24} \left[\frac{1}{n_{hjm}} \sum_{i=1}^{n_{hjm}} VOL_{ihjm} \right]}{\sum_{j=1}^{7} w_{jm}}$$

- m: month; j: day of week; h: hour of day
- w_{jm}: ocurrances of jth day of week in month m
- n_{hjm}: non-missing values for hth hour of jth day of week in month m
- VOL_{ihjm}: ith data point for hth hour of jth day of week in month m
- $MADT_m$: monthly average daily traffic for month m

Transformation

: MADT

- Traffic counts given by **Annual Average Daily Traffic** (AADT)
  - Weighted mean of **Monthly Average Daily Traffic** (MADT)
  - Weights: number of days in month

$$MADT_m = \frac{\sum_{j=1}^{7} w_{jm} \sum_{h=1}^{24} \left[\frac{1}{n_{hjm}} \sum_{i=1}^{n_{hjm}} VOL_{ihjm} \right]}{\sum_{j=1}^{7} w_{jm}}$$

- m: month; j: day of week; h: hour of day
- w_{jm}: ocurrances of jth day of week in month m
- n_{hjm}: non-missing values for hth hour of jth day of week in month m
- VOL_{ihjm}: ith data point for hth hour of jth day of week in month m
- $MADT_m$: monthly average daily traffic for month m

Transformation

: MADT

- Traffic counts given by **Annual Average Daily Traffic** (AADT)
  - Weighted mean of **Monthly Average Daily Traffic** (MADT)
  - Weights: number of days in month

$$MADT_m = \frac{\sum_{j=1}^{7} w_{jm} \sum_{h=1}^{24} \left[\frac{1}{n_{hjm}} \sum_{i=1}^{n_{hjm}} VOL_{ihjm} \right]}{\sum_{j=1}^{7} w_{jm}}$$

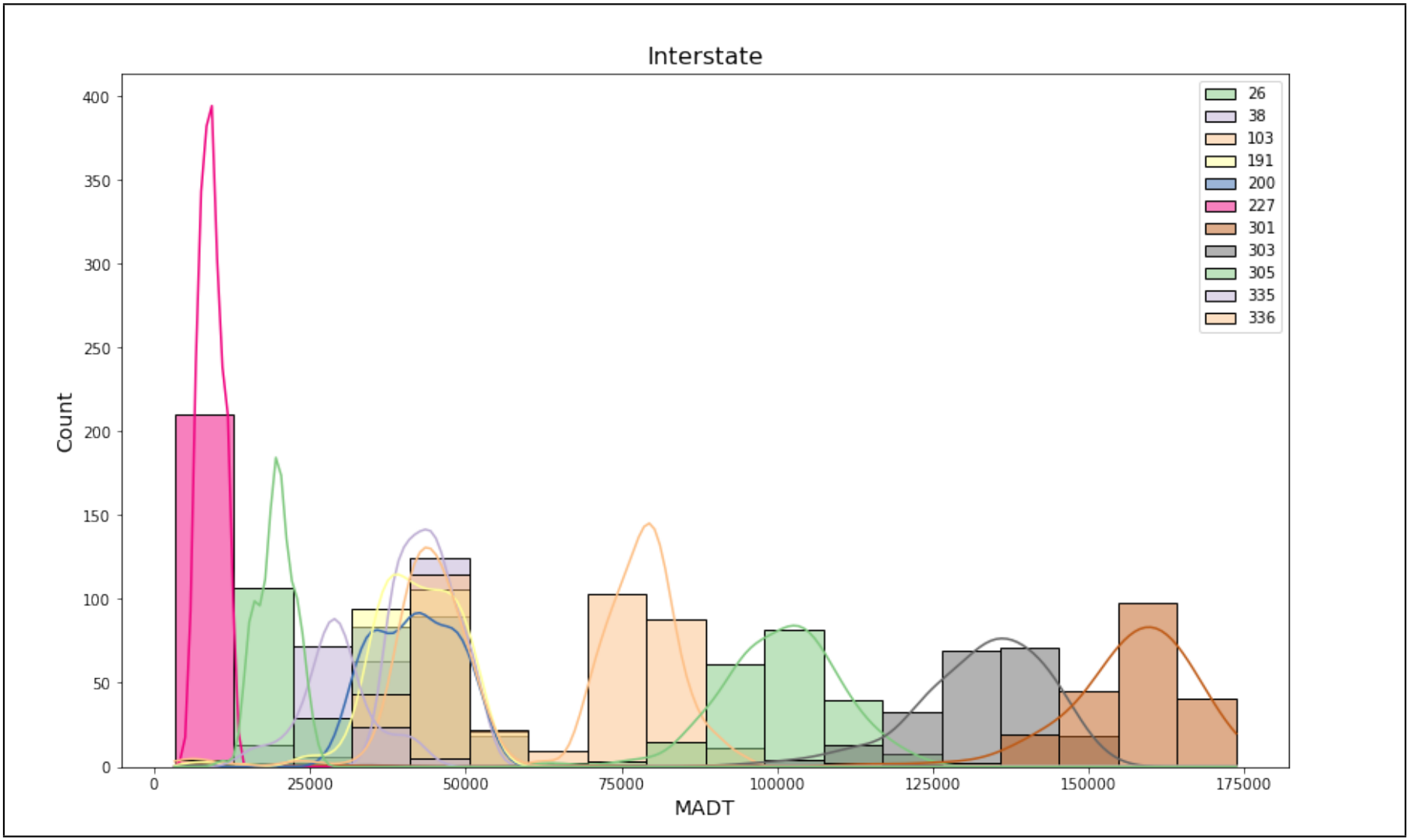
- m: month; j: day of week; h: hour of day
- w_{jm}: ocurrances of jth day of week in month m
- n_{hjm}: non-missing values for hth hour of jth day of week in month m
- VOL_{ihjm}: ith data point for hth hour of jth day of week in month m
- $MADT_m$: monthly average daily traffic for month m

Exploratory Data Analysis

Outline

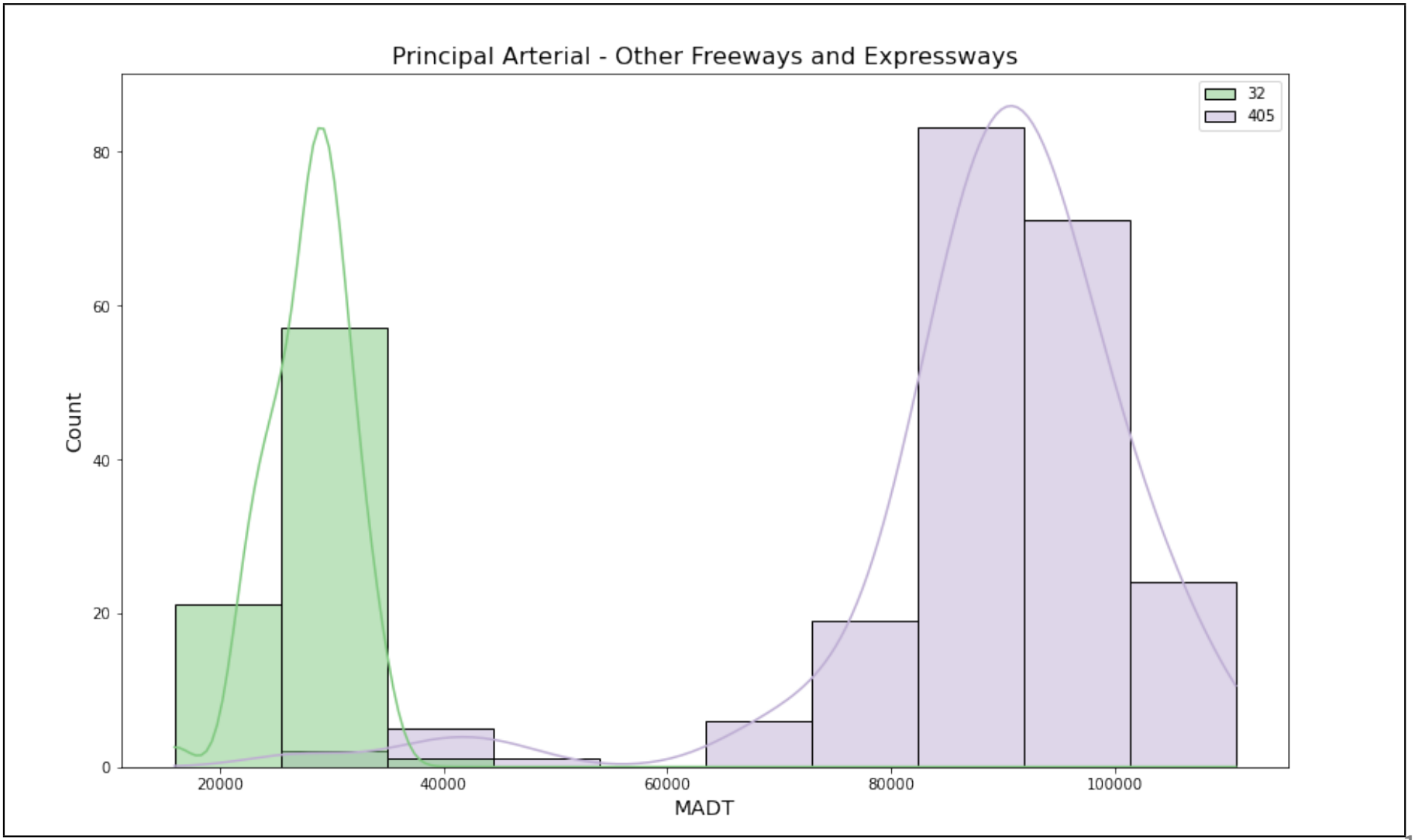
Exploratory Data Analysis

Data Distribution: Interstates



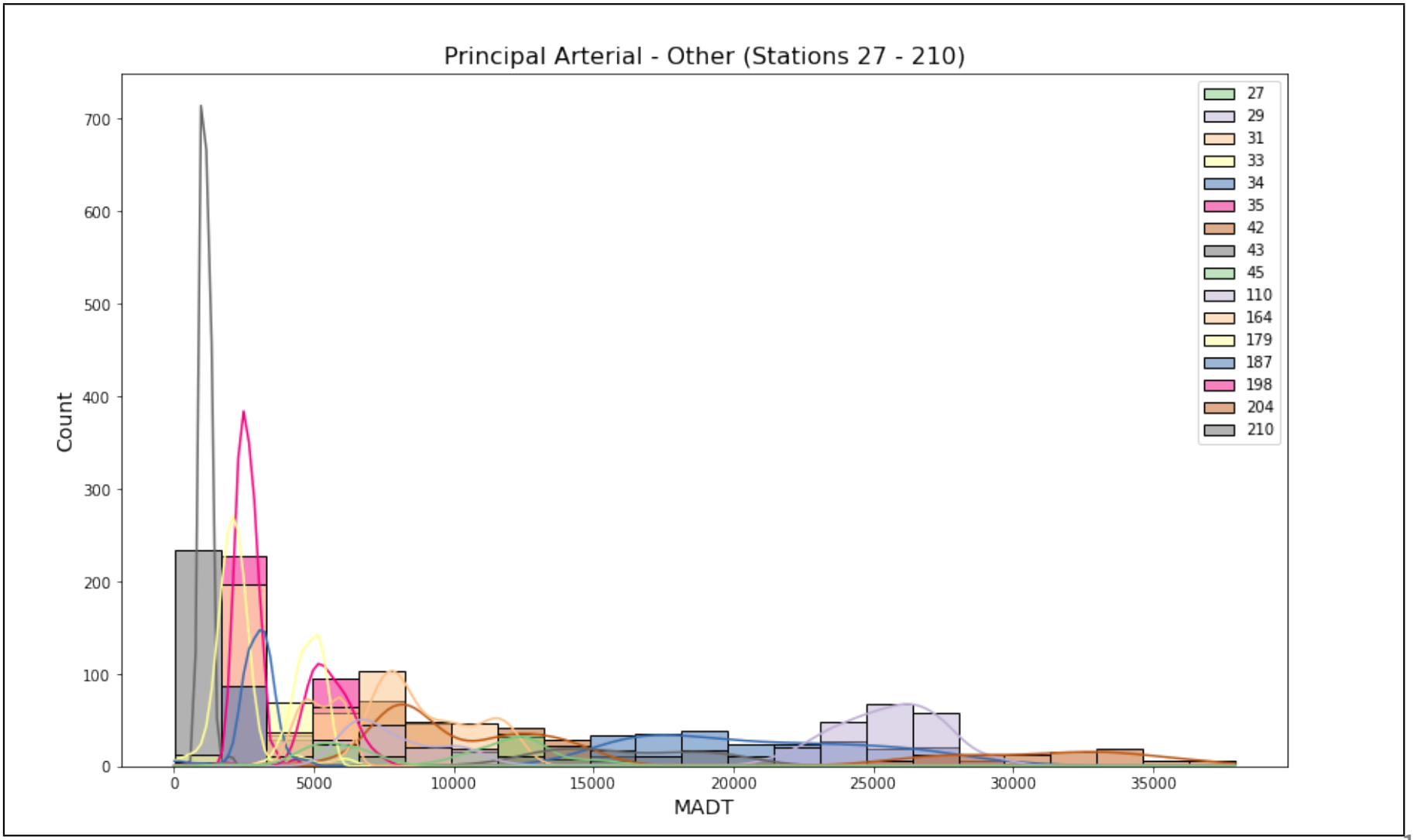
Exploratory Data Analysis

Data Distribution: Principal Arterial - Other Freeways



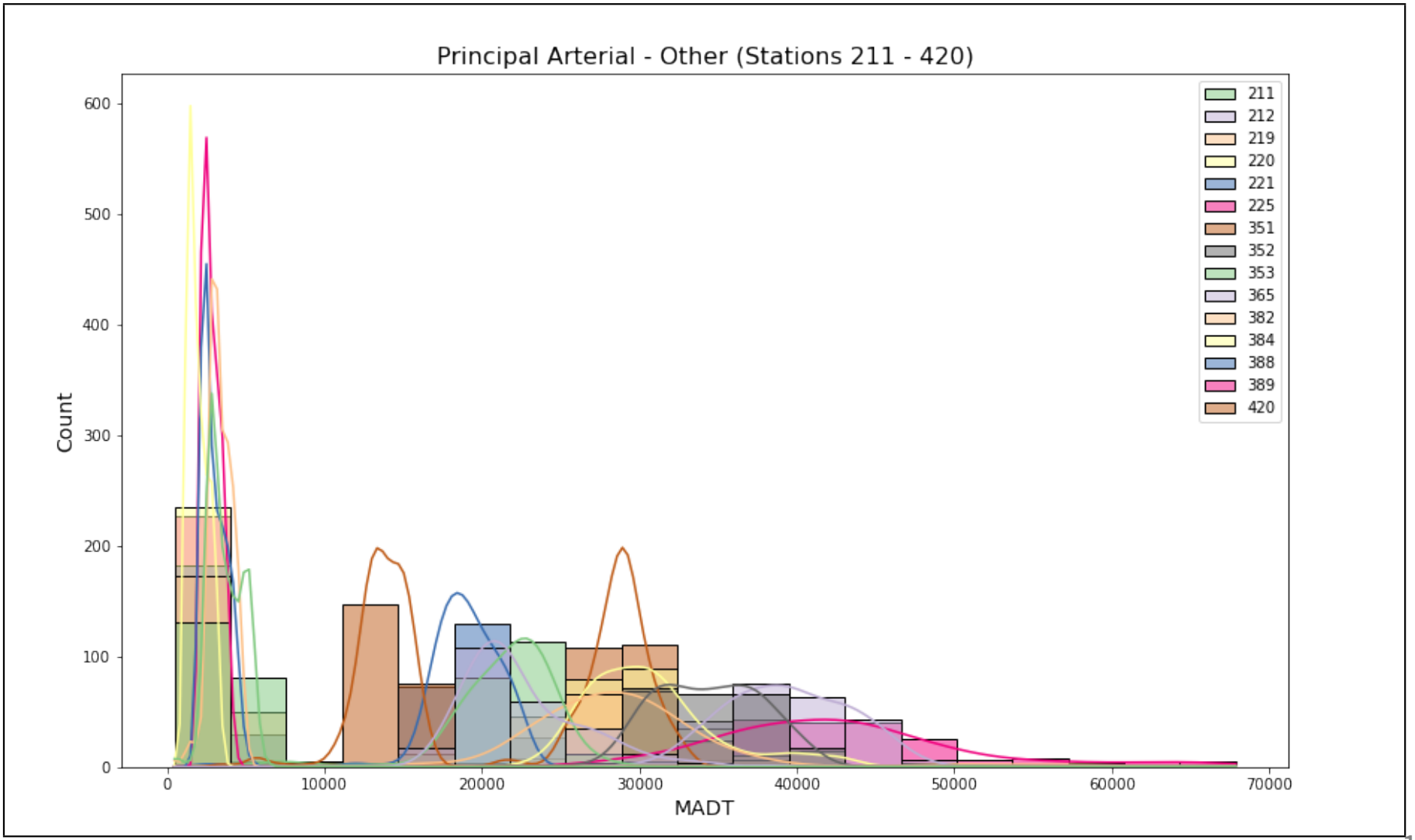
Exploratory Data Analysis

Data Distribution: Principal Arterial - Other



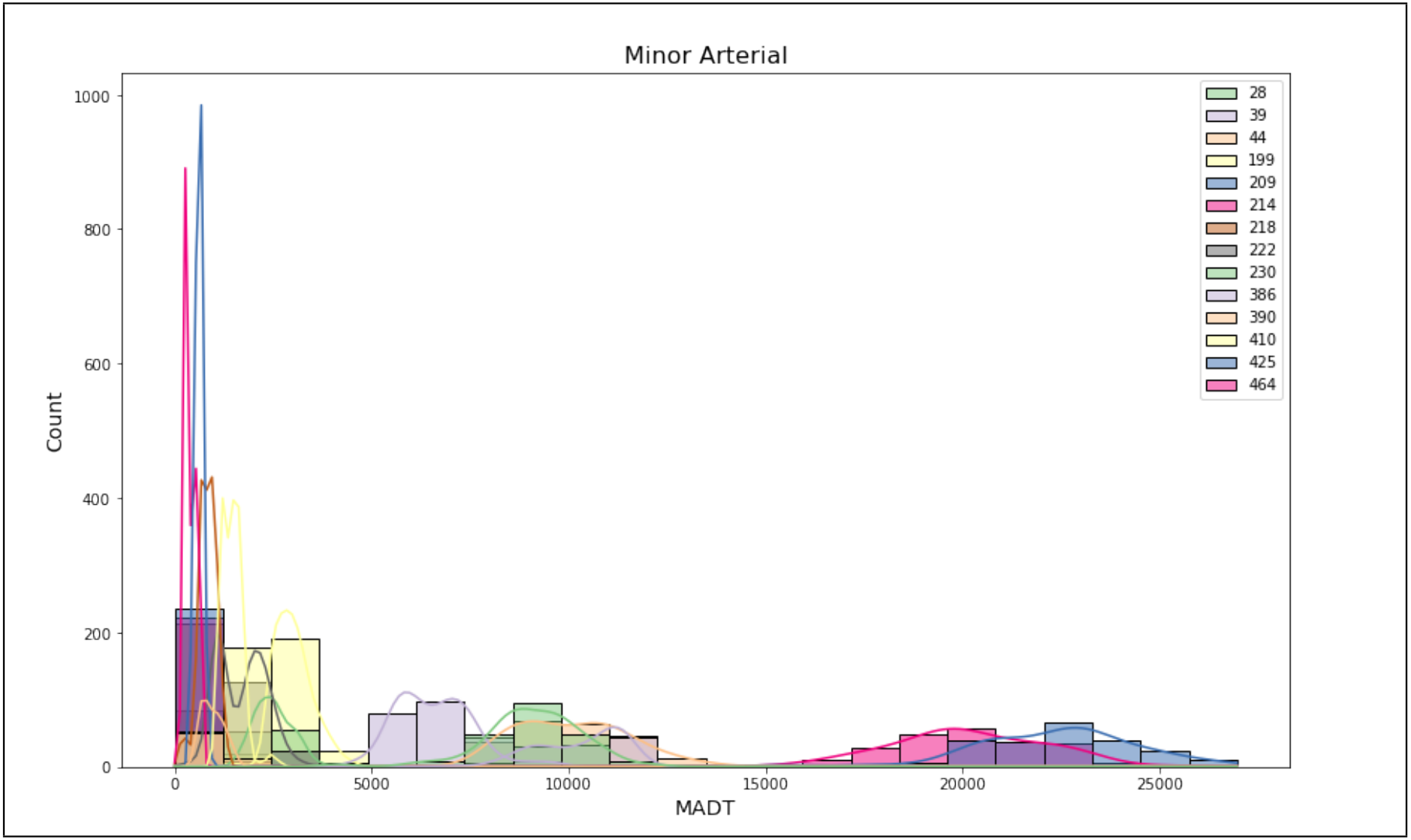
Exploratory Data Analysis

Data Distribution: Principal Arterial - Other



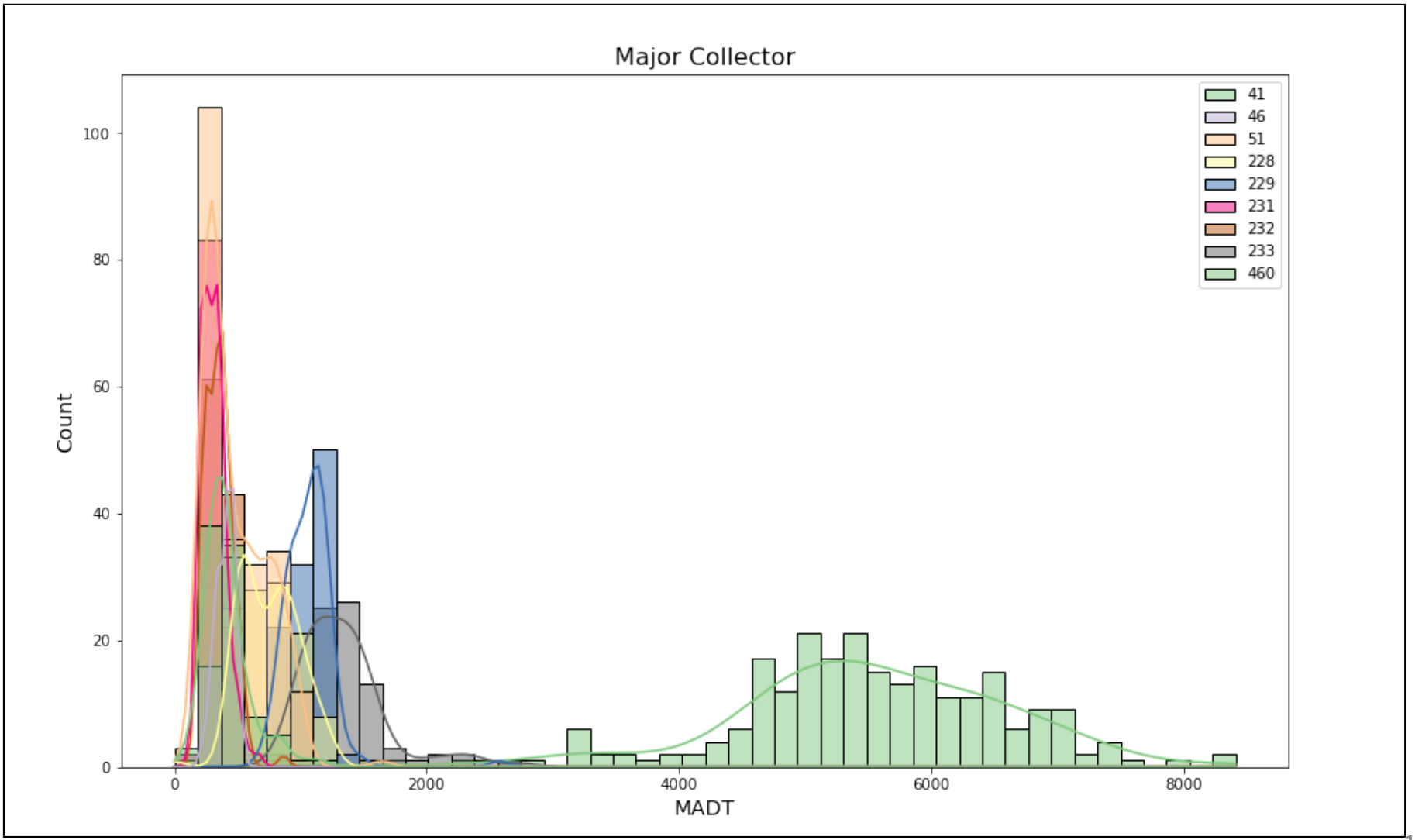
Exploratory Data Analysis

Data Distribution: Minor Arterial



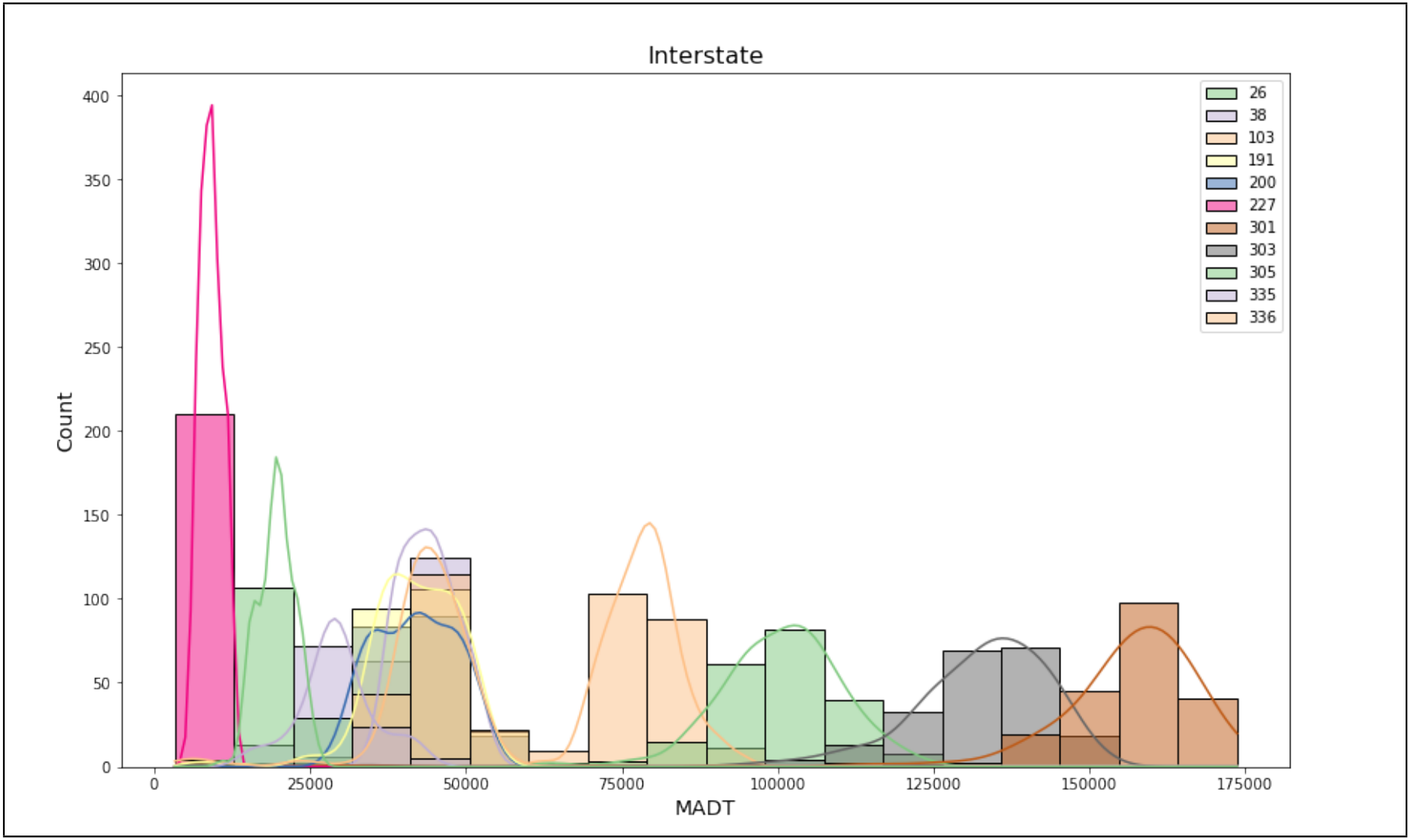
Exploratory Data Analysis

Data Distribution: Major Collector



Exploratory Data Analysis

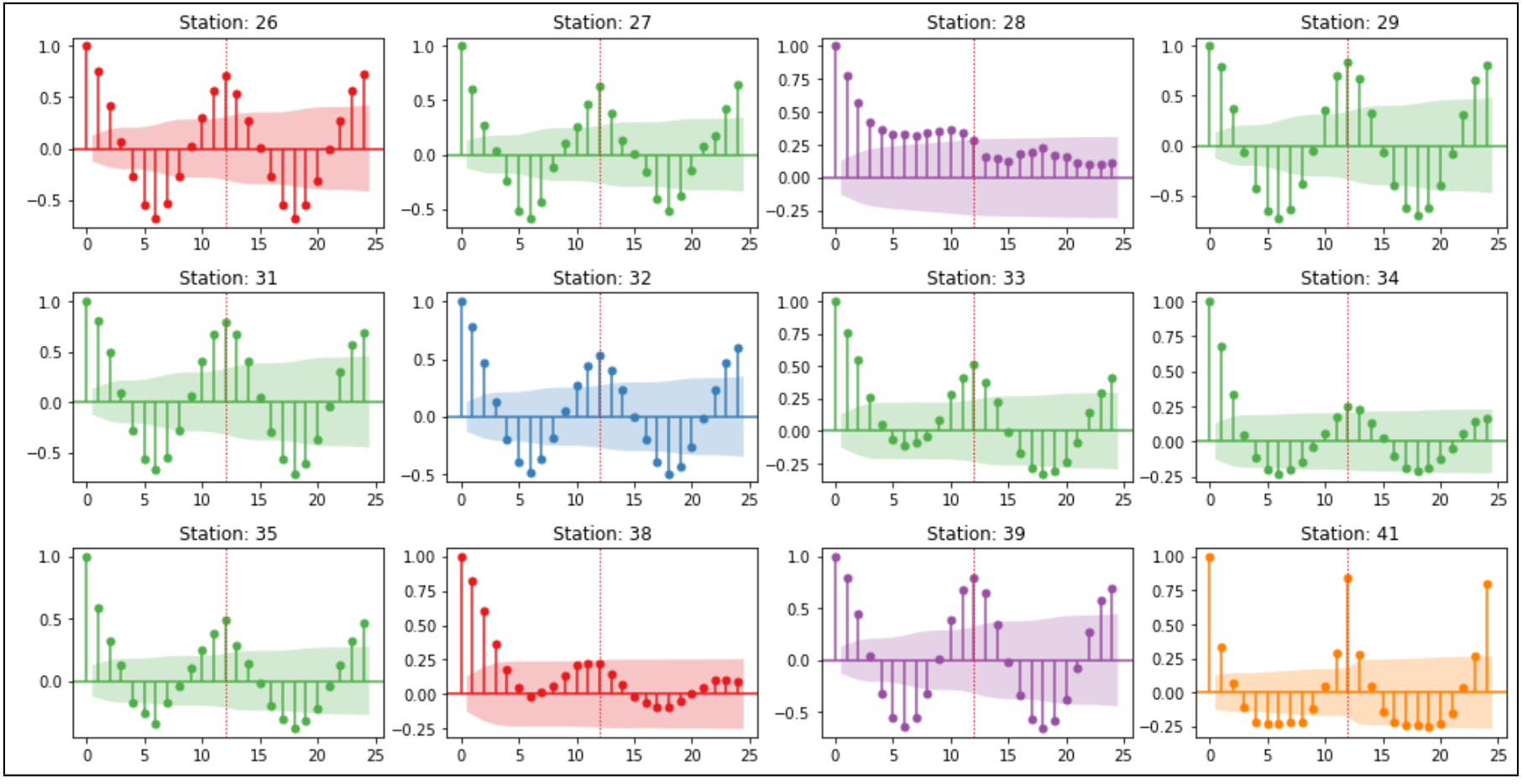
Data Distribution: Local



Outline

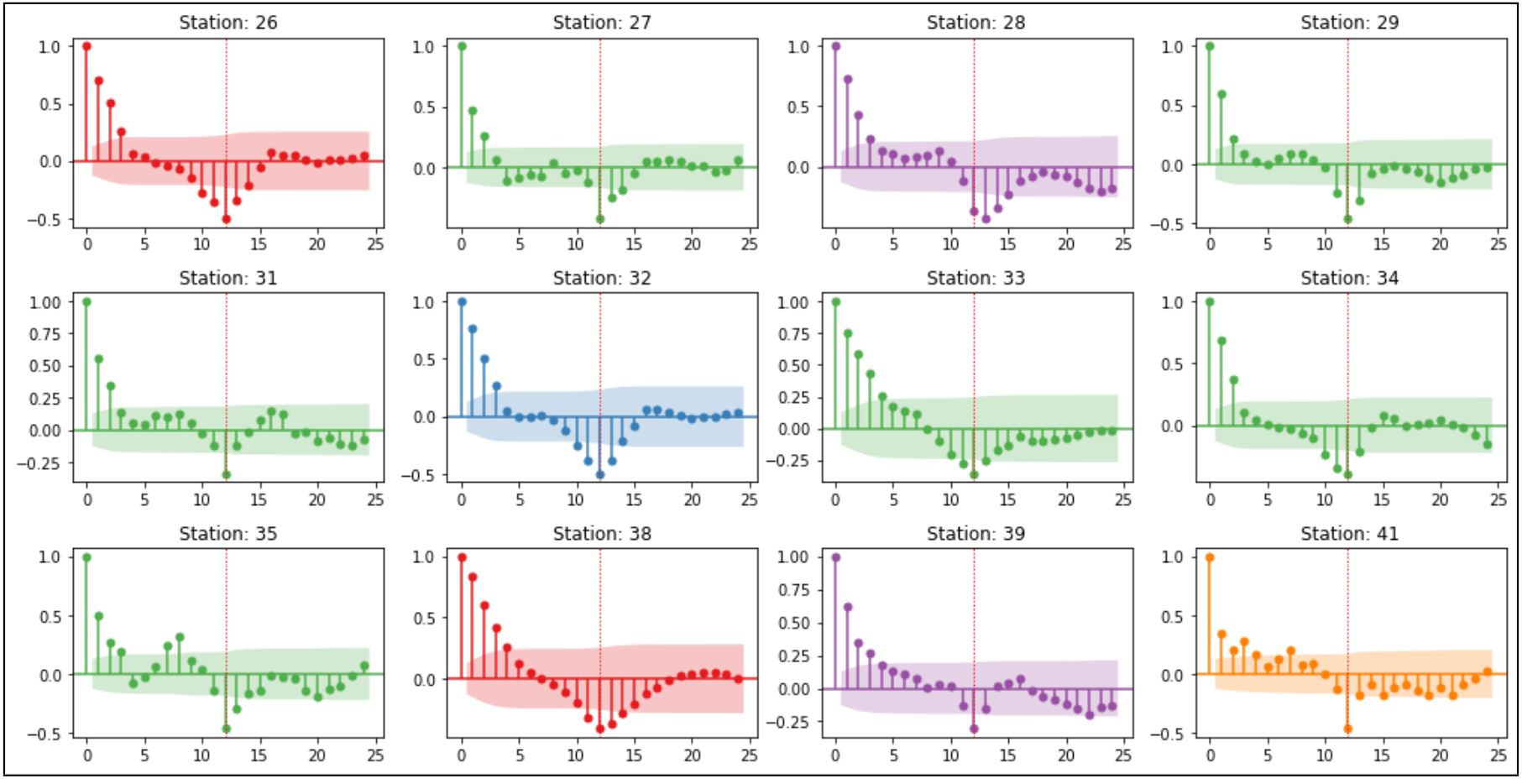
Exploratory Data Analysis

Autocorrelation: ACF



Exploratory Data Analysis

Autocorrelation: Differenced ACF



Imputation

Outline

Imputation
Methods

- Metric: Mean Square Error
- Methods
  - Mean
  - Seasonal Mean
  - CDRec (Centroid Decomposition)
  - Seasonal CDRec
  - Prophet
  - Prophet (Logistic Floor)
- All but CDRec were column-wise

Imputation
Methods

- Metric: Mean Square Error
- Methods
  - Mean
  - Seasonal Mean
  - CDRec (Centroid Decomposition)
  - Seasonal CDRec
  - Prophet
  - Prophet (Logistic Floor)
- All but CDRec were column-wise

Imputation
Methods

- Metric: Mean Square Error
- Methods
  - Mean
  - Seasonal Mean
  - CDRec (Centroid Decomposition)
  - Seasonal CDRec
  - Prophet
  - Prophet (Logistic Floor)
- All but CDRec were column-wise

Imputation
Methods

- Metric: Mean Square Error
- Methods
  - Mean
  - Seasonal Mean
  - CDRec (Centroid Decomposition)
  - Seasonal CDRec
  - Prophet
  - Prophet (Logistic Floor)
- All but CDRec were column-wise

Imputation
Methods

- Metric: Mean Square Error
- Methods
  - Mean
  - Seasonal Mean
  - CDRec (Centroid Decomposition)
  - Seasonal CDRec
  - Prophet
  - Prophet (Logistic Floor)
- All but CDRec were column-wise

Imputation
Methods

- Metric: Mean Square Error
- Methods
  - Mean
  - Seasonal Mean
  - CDRec (Centroid Decomposition)
  - Seasonal CDRec
  - Prophet
  - Prophet (Logistic Floor)
- All but CDRec were column-wise

Imputation
Methods

- Metric: Mean Square Error
- Methods
  - Mean
  - Seasonal Mean
  - CDRec (Centroid Decomposition)
  - Seasonal CDRec
  - Prophet
  - Prophet (Logistic Floor)
- All but CDRec were column-wise

Imputation
Methods

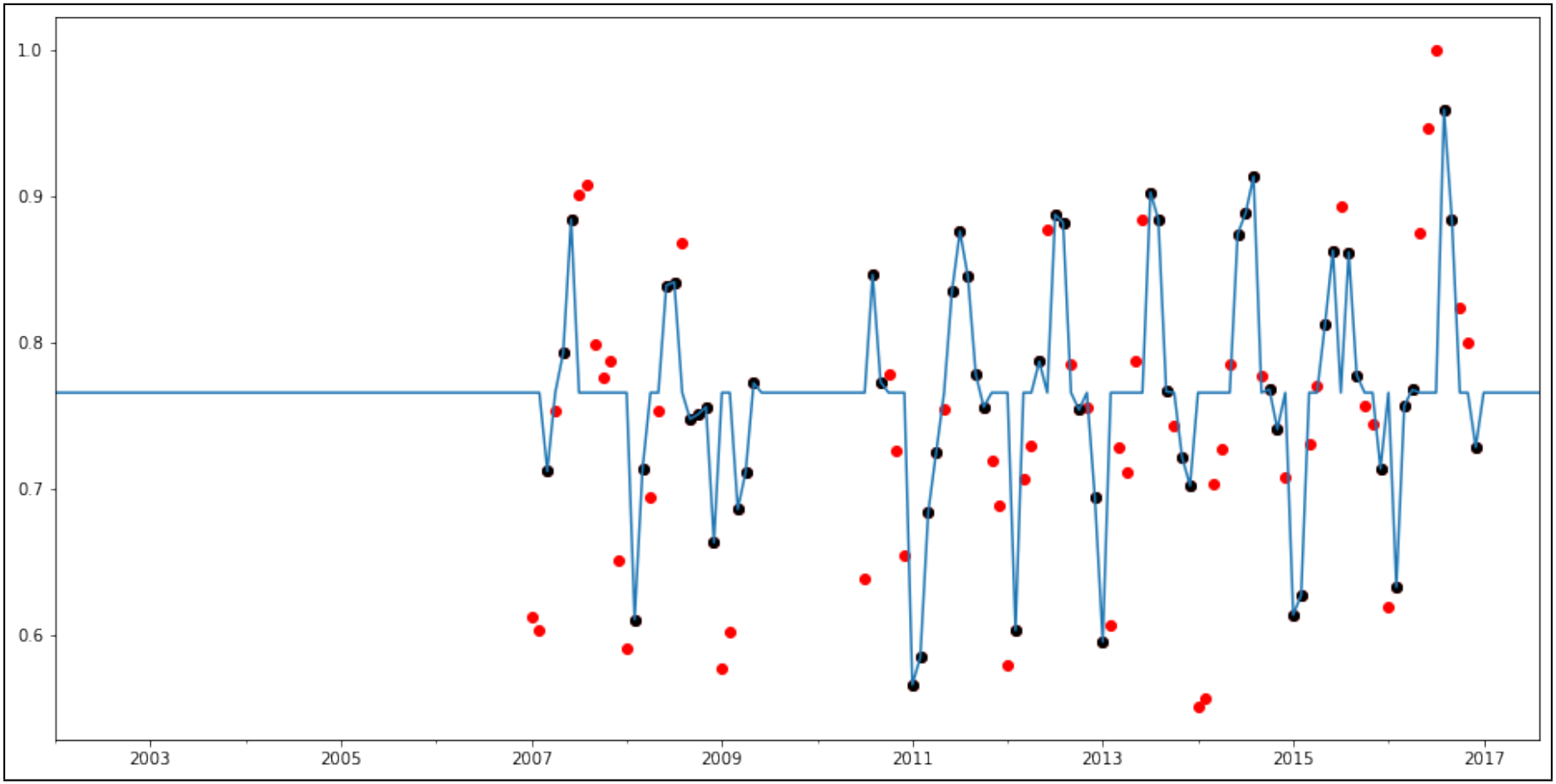
- Metric: Mean Square Error
- Methods
  - Mean
  - Seasonal Mean
  - CDRec (Centroid Decomposition)
  - Seasonal CDRec
  - Prophet
  - Prophet (Logistic Floor)
- All but CDRec were column-wise

Imputation
Methods

- Metric: Mean Square Error
- Methods
  - Mean
  - Seasonal Mean
  - CDRec (Centroid Decomposition)
  - Seasonal CDRec
  - Prophet
  - Prophet (Logistic Floor)
- All but CDRec were column-wise

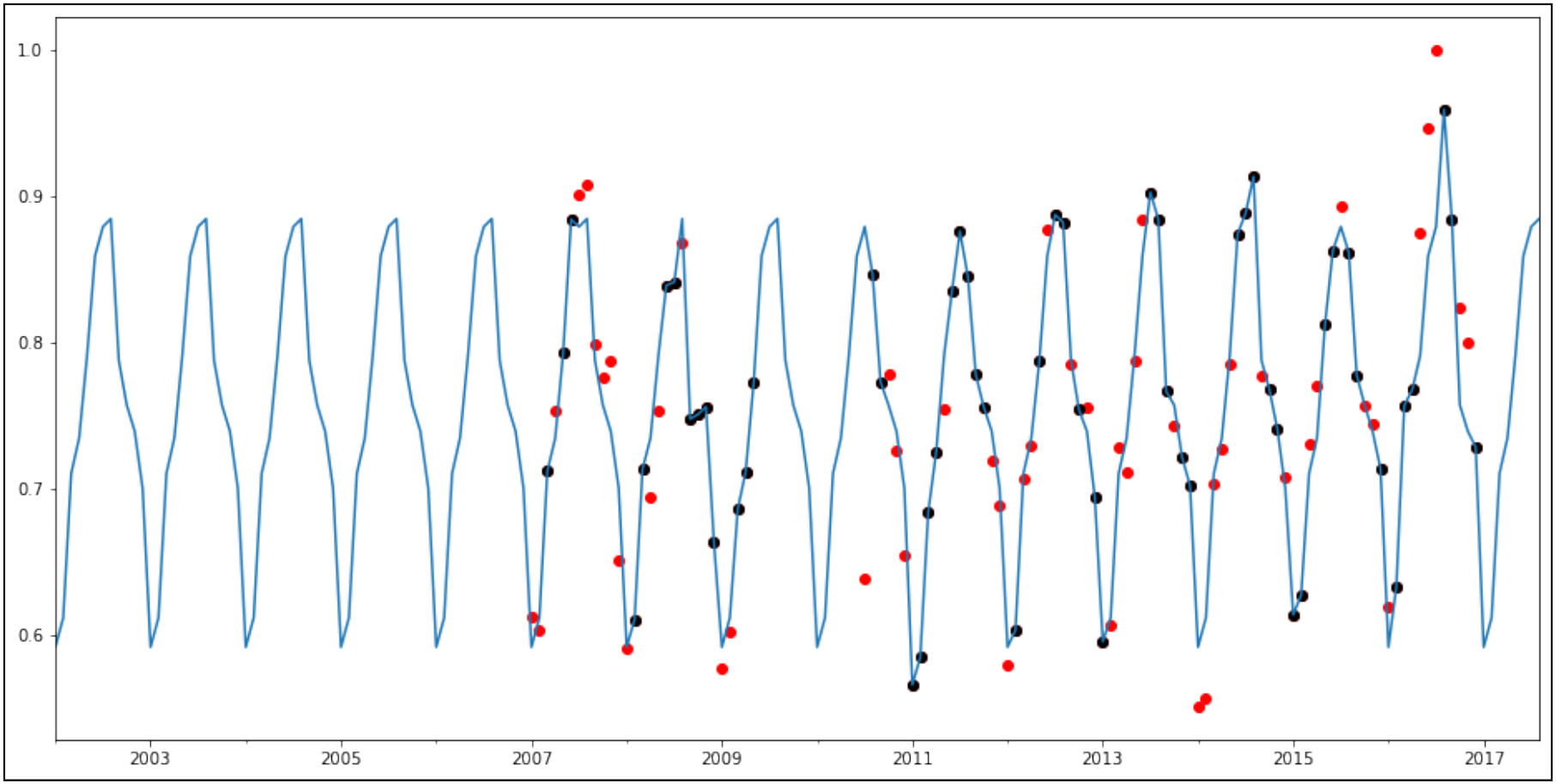
Imputation

Methods: Mean



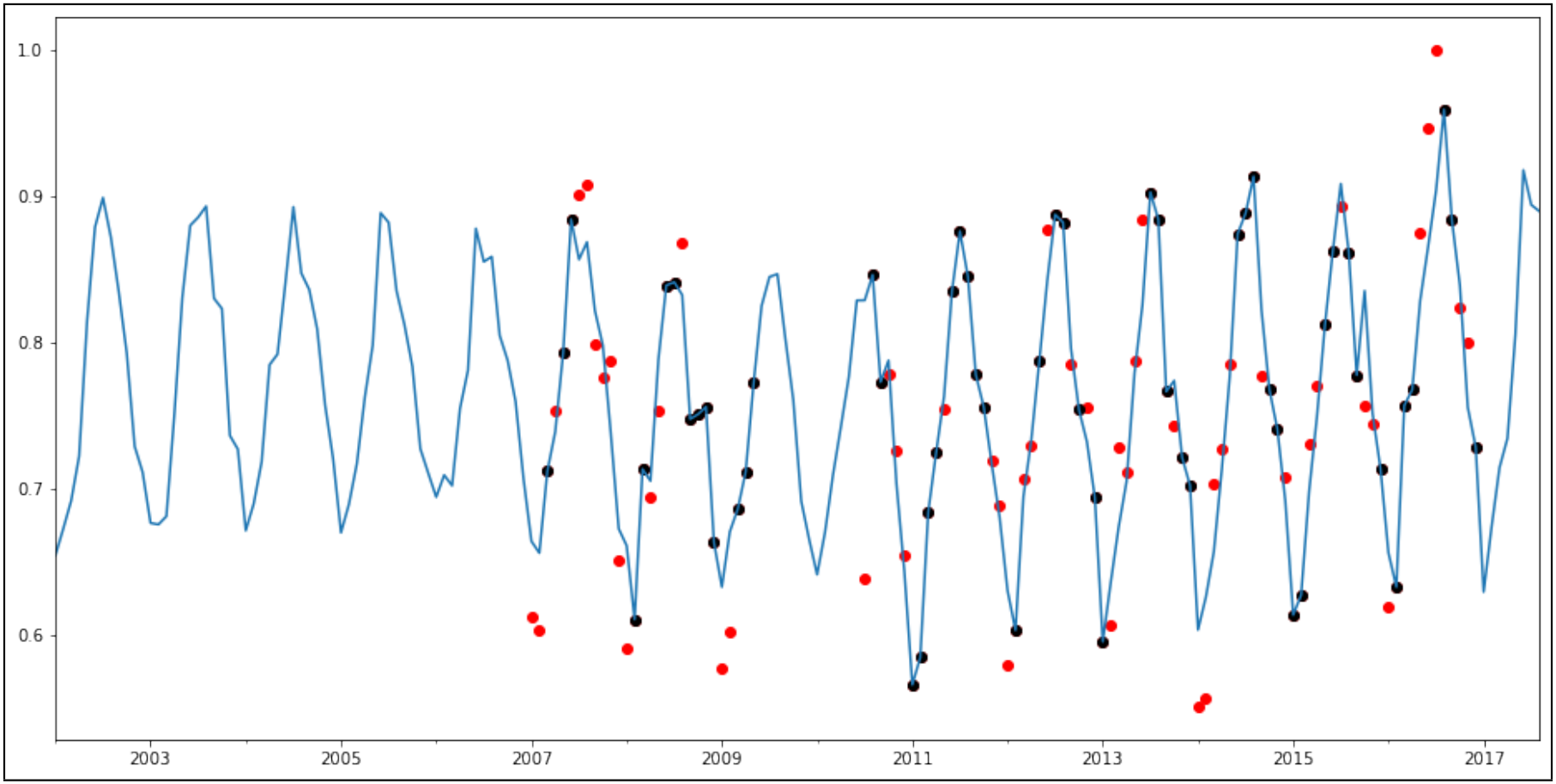
Imputation

Methods: Seasonal Mean



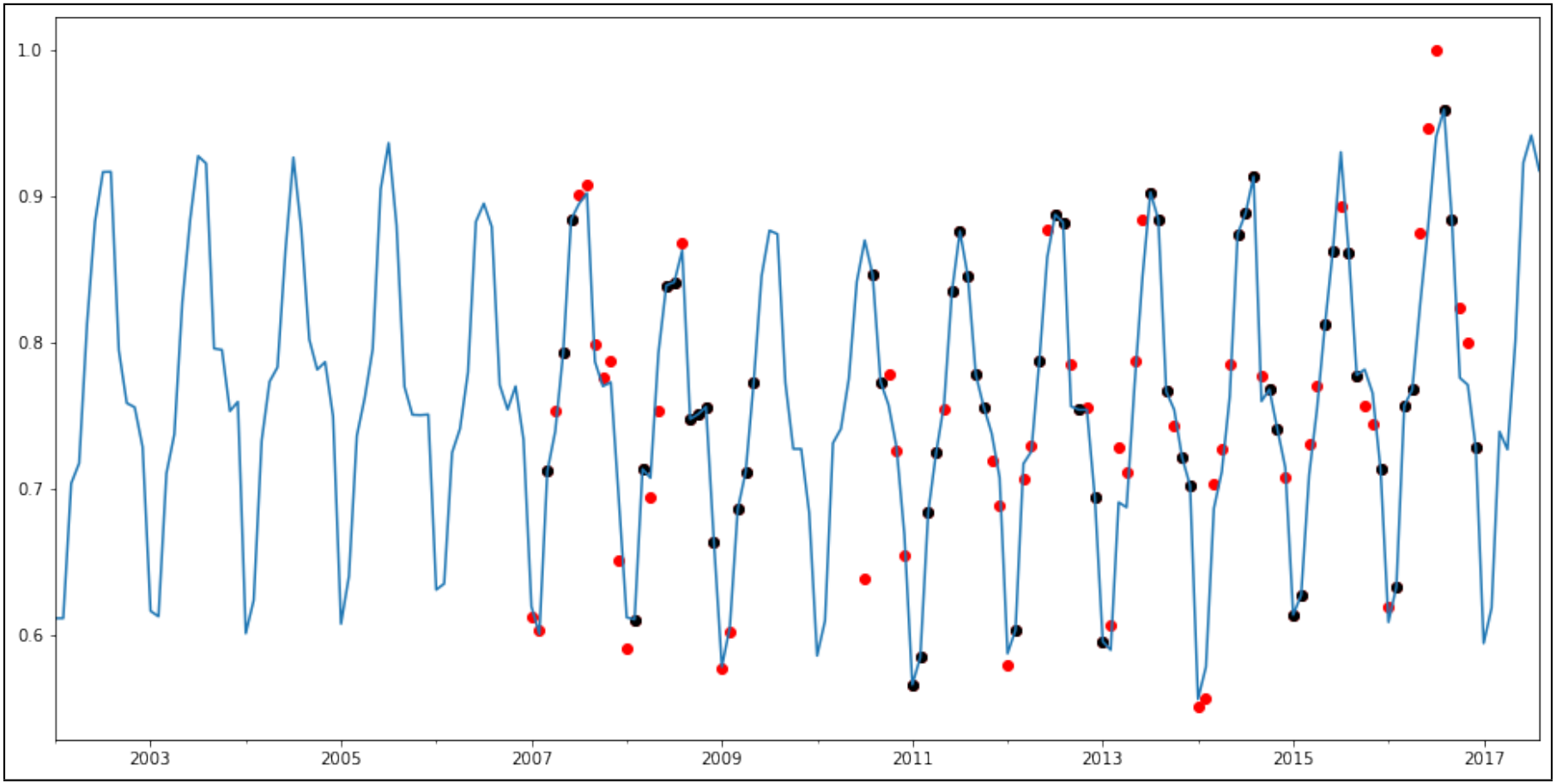
Imputation

Methods: CDRec



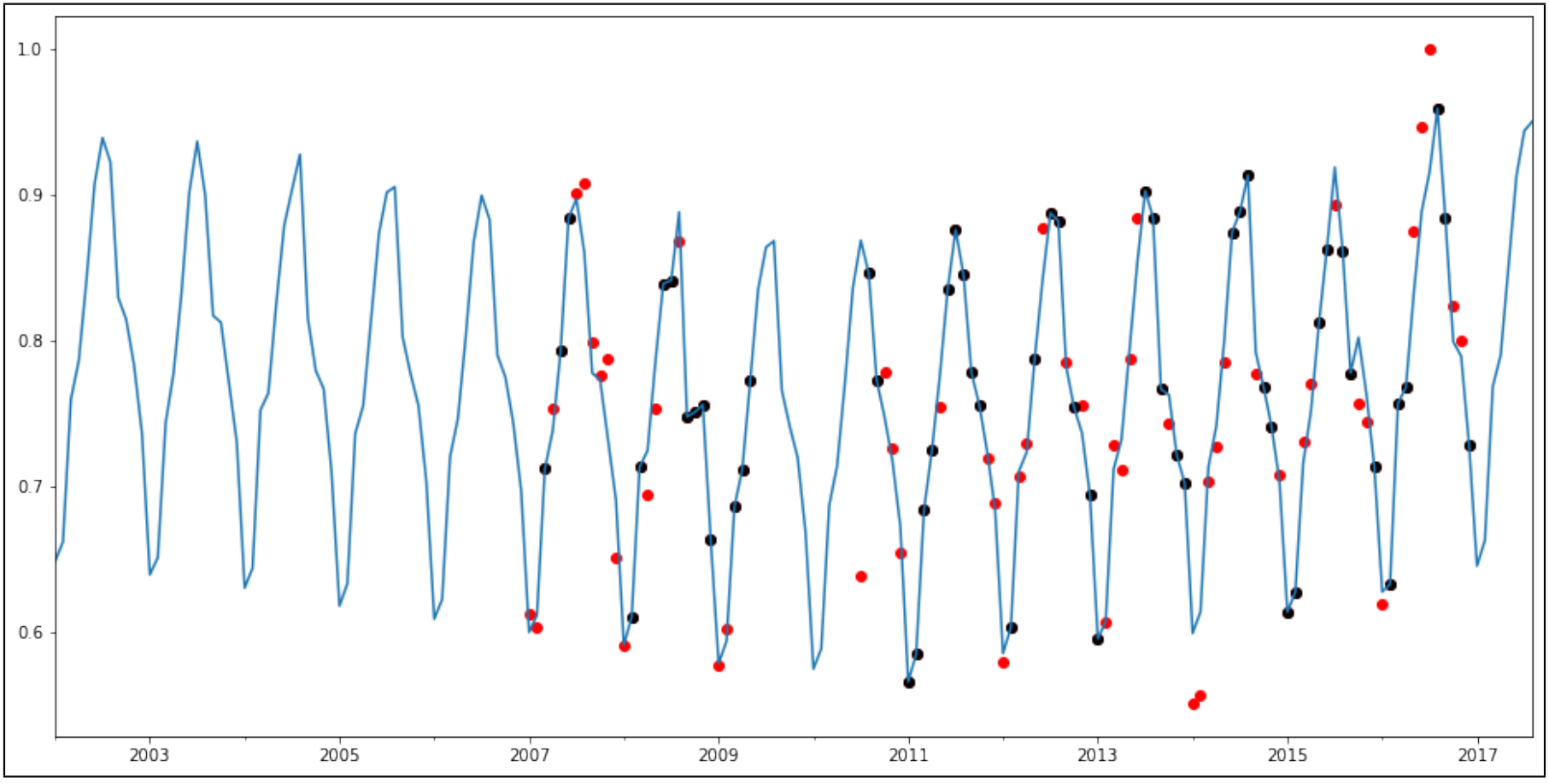
Imputation

Methods: Seasonal CDRec



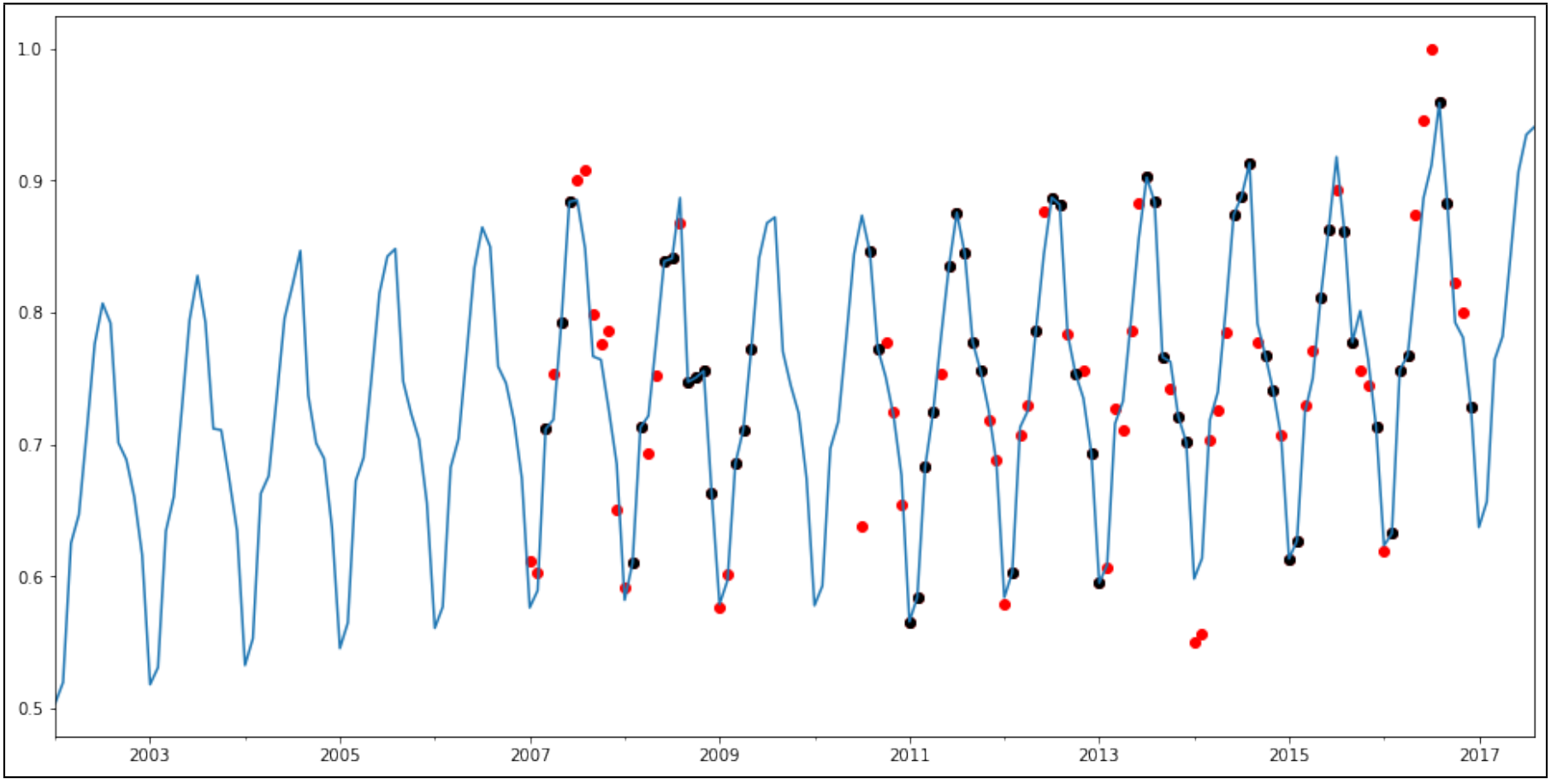
Imputation

Methods: Prophet



Imputation

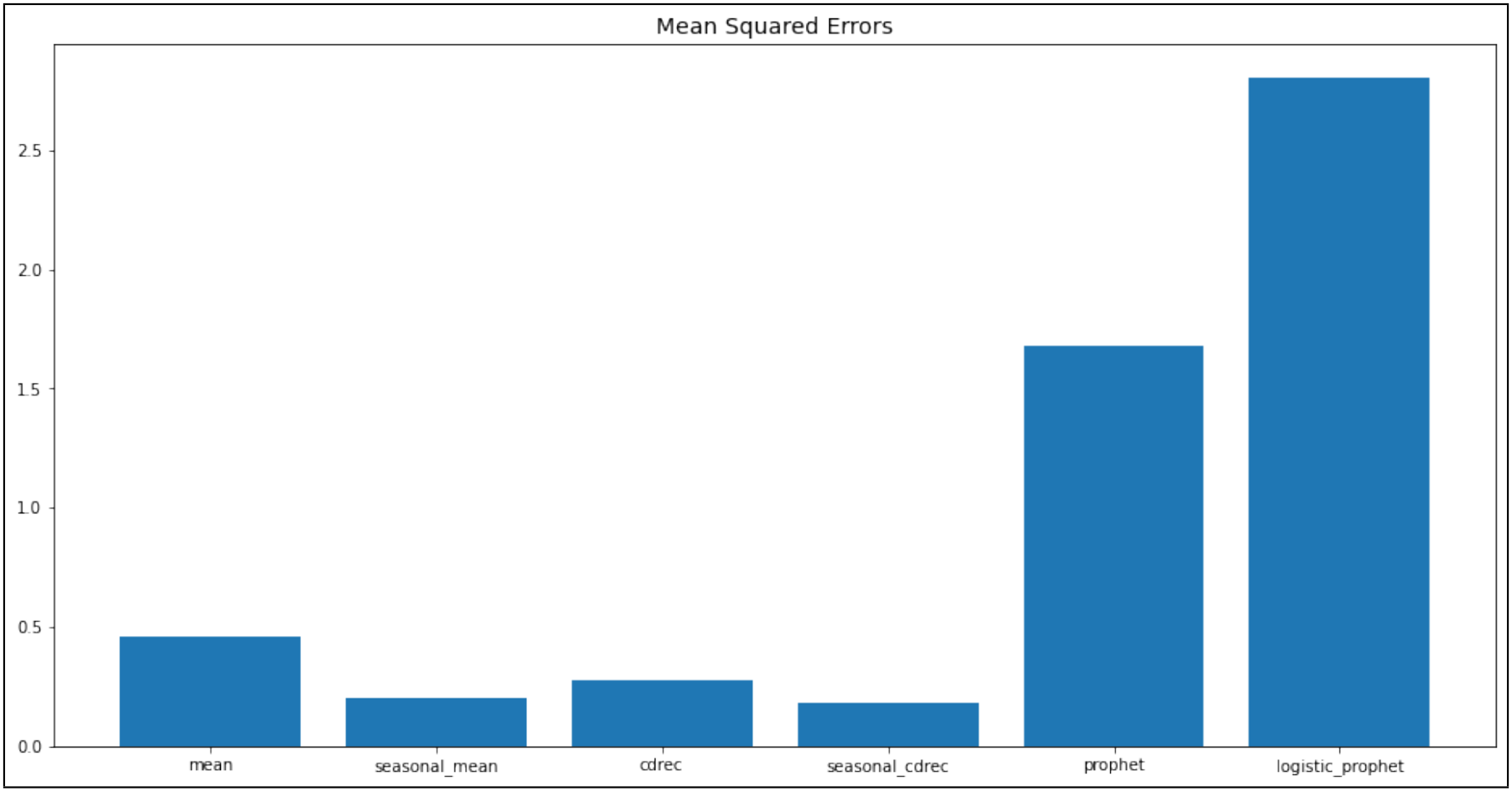
Methods: Logistic Prophet



Outline

Imputation

Results



Modeling

Outline

Modeling
Methods

- Metric: Mean Square Error
- Methods
  - Baseline: Lag 1M / 12M
  - Prophet
  - Exponential Smoothing
  - SARIMA
- All column-wise

Modeling
Methods

- Metric: Mean Square Error
- Methods
  - Baseline: Lag 1M / 12M
  - Prophet
  - Exponential Smoothing
  - SARIMA
- All column-wise

Modeling
Methods

- Metric: Mean Square Error
- Methods
  - Baseline: Lag 1M / 12M
  - Prophet
  - Exponential Smoothing
  - SARIMA
- All column-wise

Modeling
Methods

- Metric: Mean Square Error
- Methods
  - Baseline: Lag 1M / 12M
  - Prophet
  - Exponential Smoothing
  - SARIMA
- All column-wise

Modeling
Methods

- Metric: Mean Square Error
- Methods
  - Baseline: Lag 1M / 12M
  - Prophet
  - Exponential Smoothing
  - SARIMA
- All column-wise

Modeling
Methods

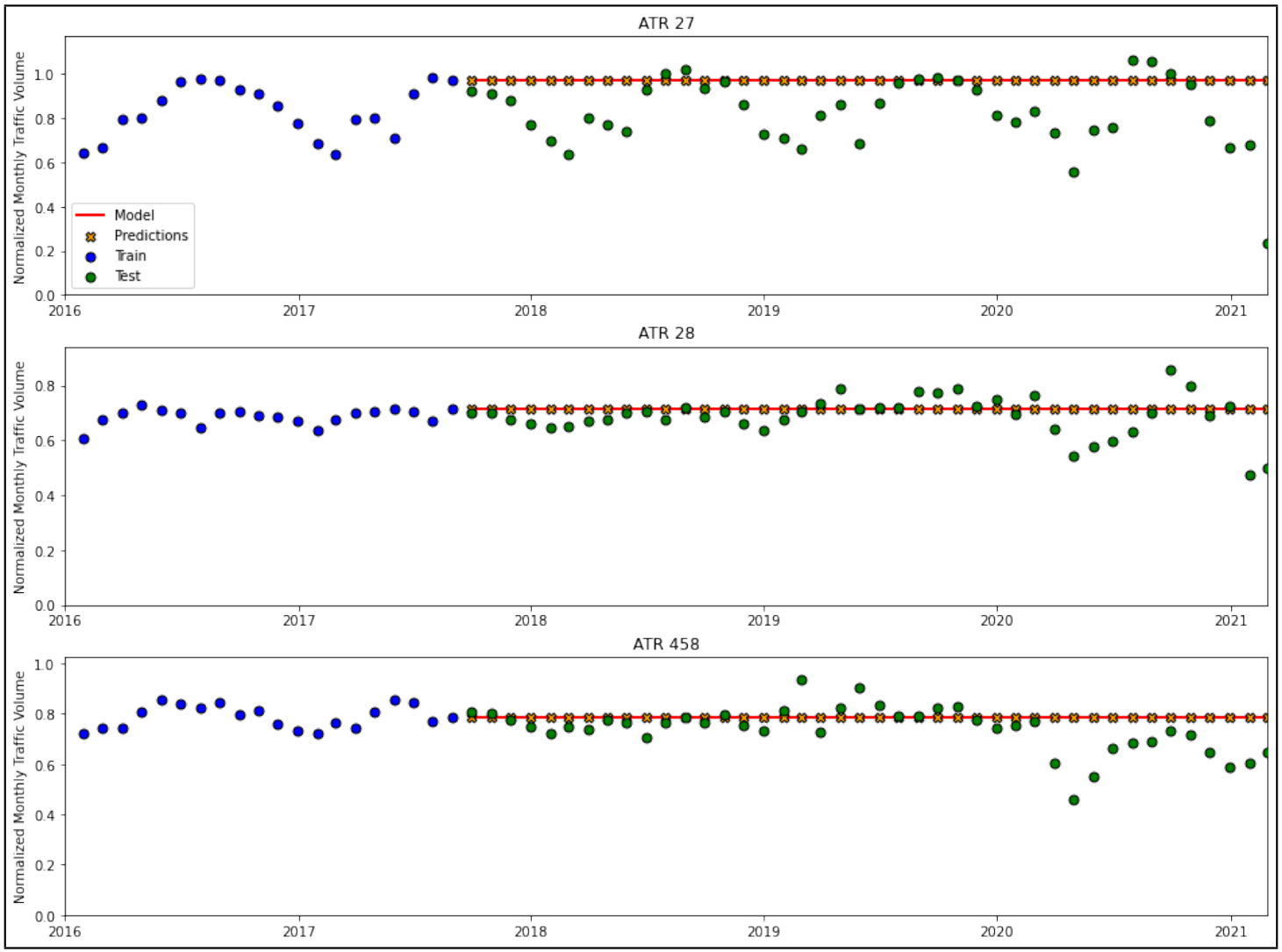
- Metric: Mean Square Error
- Methods
  - Baseline: Lag 1M / 12M
  - Prophet
  - Exponential Smoothing
  - SARIMA
- All column-wise

Modeling
Methods

- Metric: Mean Square Error
- Methods
  - Baseline: Lag 1M / 12M
  - Prophet
  - Exponential Smoothing
  - SARIMA
- All column-wise

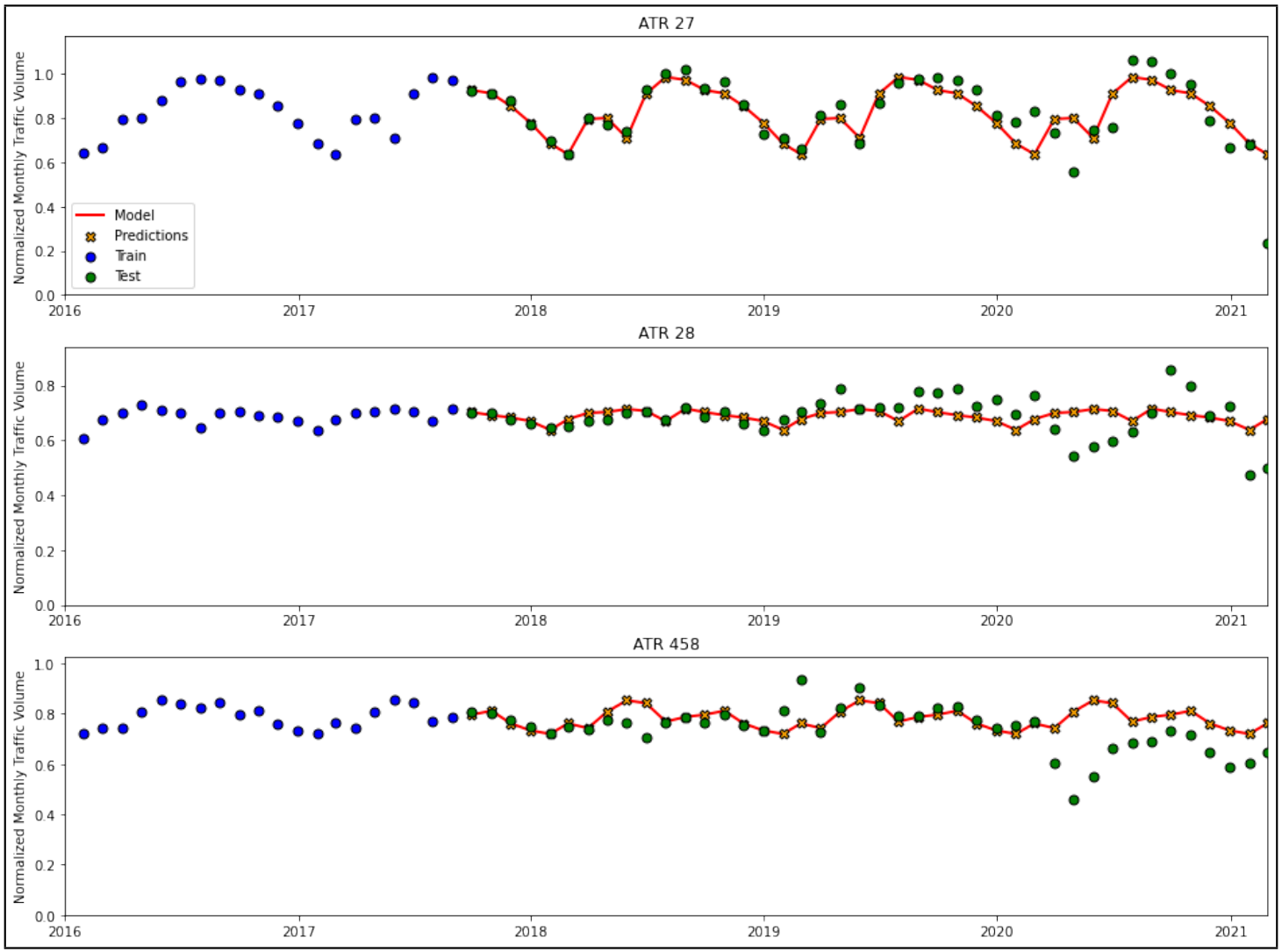
Modeling

Methods: 1 Month Lag



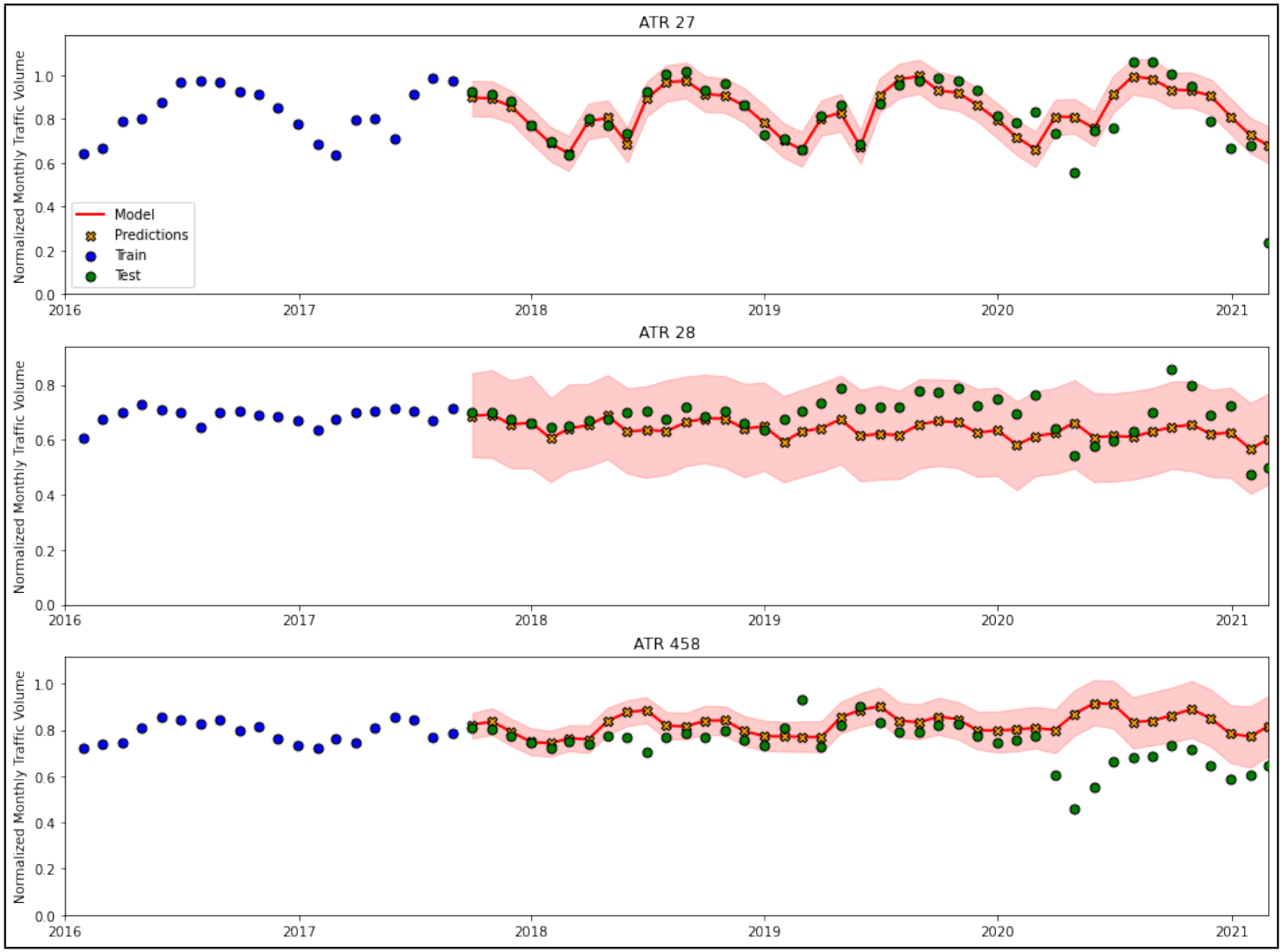
Modeling

Methods: 12 Month Lag



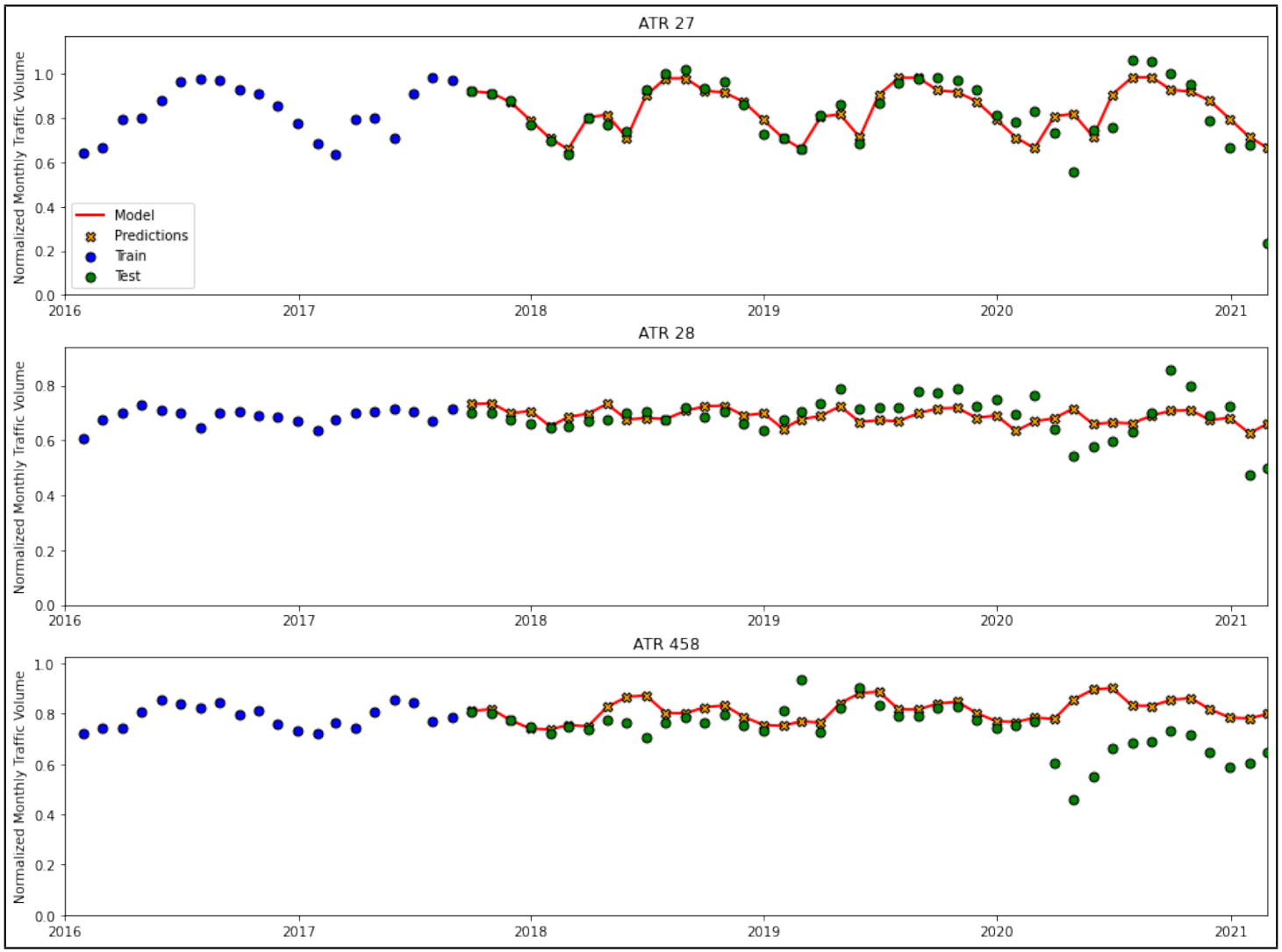
Modeling

Methods: Prophet



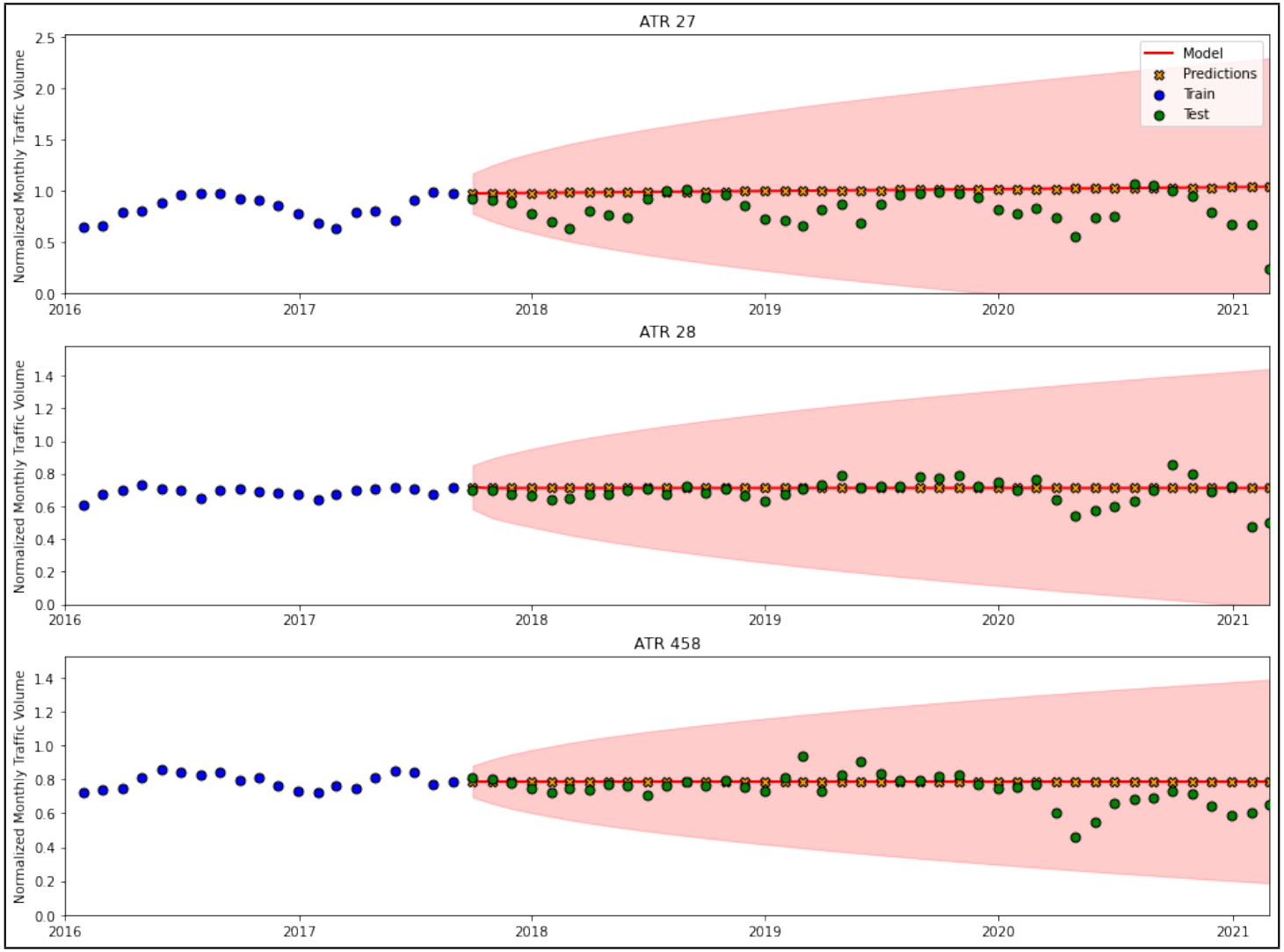
Modeling

Methods: Exponential Smoothing



Modeling

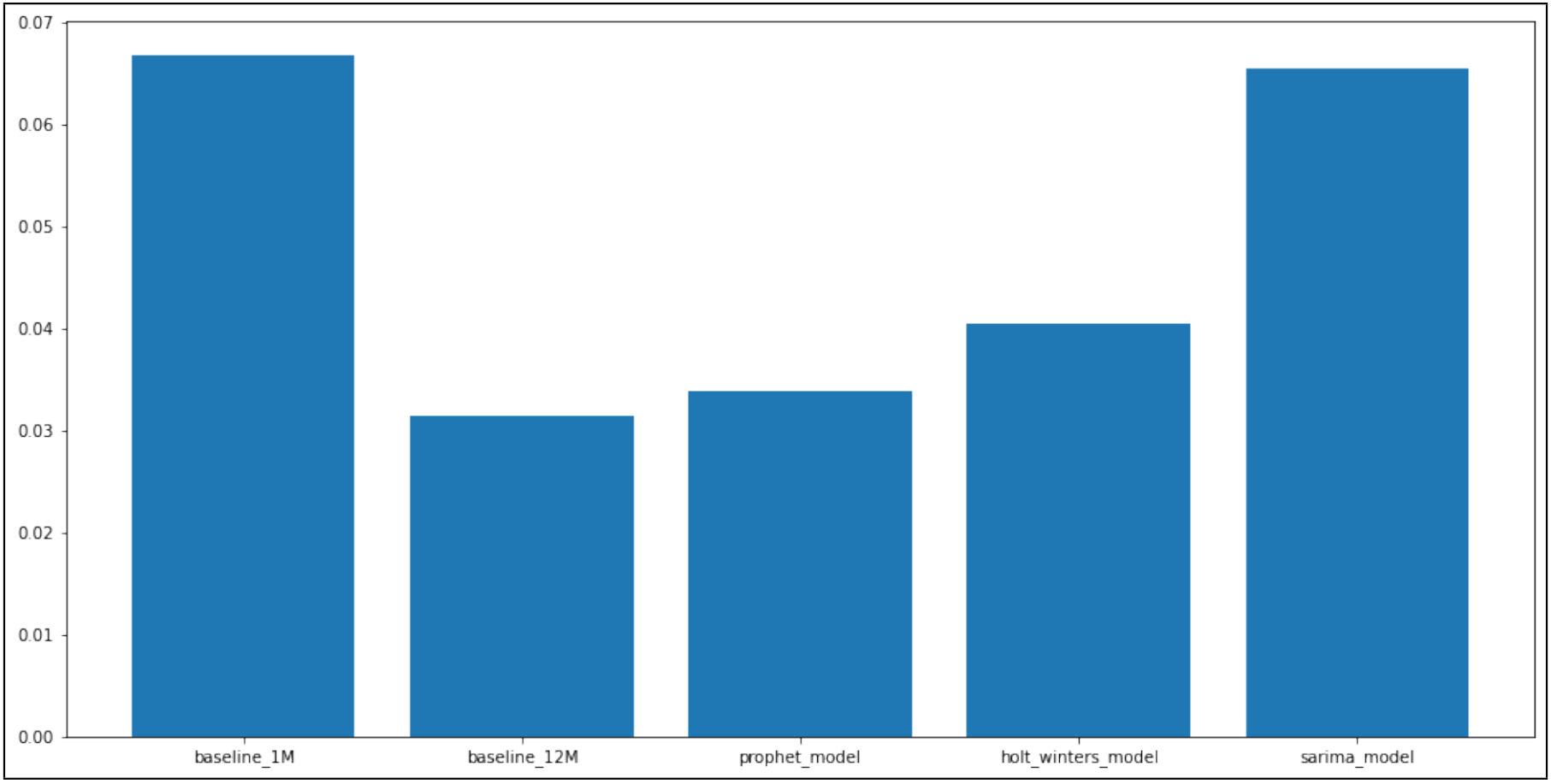
Methods: SARIMA



Outline

Modeling

Results



Future Directions

Future Directions

- Adjust for pandemic
- Deal with Interrelations
  - Deep Learning
  - STARIMA (Space-Time Autoregressive Integrated Moving Average)

Future Directions

- Adjust for pandemic
- Deal with Interrelations
  - Deep Learning
  - STARIMA (Space-Time Autoregressive Integrated Moving Average)

Future Directions

- Adjust for pandemic
- Deal with Interrelations
  - Deep Learning
  - STARIMA (Space-Time Autoregressive Integrated Moving Average)

Future Directions

- Adjust for pandemic
- Deal with Interrelations
  - Deep Learning
  - STARIMA (Space-Time Autoregressive Integrated Moving Average)