

The background of the slide features a large teal triangle on the left side, pointing towards the center. The rest of the background is white. The title text is centered in the white area.

# **Minnesota Department of Transportation Traffic Project**

Nathan Wodarz

September 2021

The background of the slide is composed of two large, diagonal, triangular sections. The top-left section is a solid teal color, and the bottom-right section is a solid light gray color. These two sections meet at a diagonal line that runs from the top-left towards the bottom-right, leaving a large white triangular area in the center where the text is located.

# Problem Statement

# Outline

## Problem Statement

- Background

- Motivation

- Problem Statement

## Data Wrangling

- Raw Data

## Data Cleaning

## Transformation

## Exploratory Data Analysis

- Data Distribution

- Autocorrelation

## Imputation

- Methods

- Results

## Modeling

- Methods

- Results

## Future Directions

# 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
    - ▶ 104 in Minnesota's 35 road network (7 counties)
    - ▶ 50+ out of state (20 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
    - ▶ 104 in Minnesota, 51 Paid metro area (7 counties)
    - ▶ 51 outside (120 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
    - ▶ 104 in Minneapolis-St. Paul metro area (7 counties)
    - ▶ 51 on other 120 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
    - ▶ 100 in Minnesota, 55 Paid measures (7 counties)
    - ▶ 55 unaided (100 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

Background

**Motivation**

Problem Statement

## Data Wrangling

Raw Data

## Data Cleaning

## Transformation

## Exploratory Data Analysis

Data Distribution

Autocorrelation

## Imputation

Methods

Results

## Modeling

Methods

Results

## Future Directions

# 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

Background

Motivation

**Problem Statement**

## Data Wrangling

Raw Data

## Data Cleaning

## Transformation

## Exploratory Data Analysis

Data Distribution

Autocorrelation

## Imputation

Methods

Results

## Modeling

Methods

Results

## Future Directions

# 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

The background of the slide is composed of two large, diagonal geometric sections. The top-left section is a solid teal color, and the bottom-right section is a solid light gray color. These two sections meet at a diagonal line that runs from the top-left towards the bottom-right, leaving a large white triangular area in the center where the text is located.

# Data Wrangling

# Outline

Problem Statement

Background

Motivation

Problem Statement

**Data Wrangling**

**Raw Data**

Data Cleaning

Transformation

Exploratory Data Analysis

Data Distribution

Autocorrelation

Imputation

Methods

Results

Modeling

Methods

Results

Future Directions

# 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

The background of the slide is composed of three geometric sections. A teal-colored triangle is in the top-left corner. A light gray triangle is in the bottom-left corner. The remaining area is a white trapezoid. The text "Data Cleaning" is centered within the white area.

# 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

The background of the slide is composed of three geometric sections. A teal-colored triangle is in the top-left corner. A light gray triangle is in the bottom-left corner. The remaining area is a white trapezoid. The word "Transformation" is centered within the white area.

# 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
- ▶  $w_{jm}$ : occurrences of  $j$ th day of week in month  $m$
- ▶  $n_{hjm}$ : non-missing values for  $h$ th hour of  $j$ th day of week in month  $m$
- ▶  $VOL_{ihjm}$ :  $i$ th data point for  $h$ th hour of  $j$ th 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
- ▶  $w_{jm}$ : occurrences of  $j$ th day of week in month  $m$
- ▶  $n_{hjm}$ : non-missing values for  $h$ th hour of  $j$ th day of week in month  $m$
- ▶  $VOL_{ihjm}$ :  $i$ th data point for  $h$ th hour of  $j$ th 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
- ▶  $w_{jm}$ : occurrences of  $j$ th day of week in month  $m$
- ▶  $n_{hjm}$ : non-missing values for  $h$ th hour of  $j$ th day of week in month  $m$
- ▶  $VOL_{ihjm}$ :  $i$ th data point for  $h$ th hour of  $j$ th 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
- ▶  $w_{jm}$ : occurrences of  $j$ th day of week in month  $m$
- ▶  $n_{hjm}$ : non-missing values for  $h$ th hour of  $j$ th day of week in month  $m$
- ▶  $VOL_{ihjm}$ :  $i$ th data point for  $h$ th hour of  $j$ th 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
- ▶  $w_{jm}$ : occurrences of  $j$ th day of week in month  $m$
- ▶  $n_{hjm}$ : non-missing values for  $h$ th hour of  $j$ th day of week in month  $m$
- ▶  $VOL_{ihjm}$ :  $i$ th data point for  $h$ th hour of  $j$ th 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
- ▶  $w_{jm}$ : occurrences of  $j$ th day of week in month  $m$
- ▶  $n_{hjm}$ : non-missing values for  $h$ th hour of  $j$ th day of week in month  $m$
- ▶  $VOL_{ihjm}$ :  $i$ th data point for  $h$ th hour of  $j$ th 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
- ▶  $w_{jm}$ : occurrences of  $j$ th day of week in month  $m$
- ▶  $n_{hjm}$ : non-missing values for  $h$ th hour of  $j$ th day of week in month  $m$
- ▶  $VOL_{ihjm}$ :  $i$ th data point for  $h$ th hour of  $j$ th 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
- ▶  $w_{jm}$ : occurrences of  $j$ th day of week in month  $m$
- ▶  $n_{hjm}$ : non-missing values for  $h$ th hour of  $j$ th day of week in month  $m$
- ▶  $VOL_{ihjm}$ :  $i$ th data point for  $h$ th hour of  $j$ th 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
- ▶  $w_{jm}$ : occurrences of  $j$ th day of week in month  $m$
- ▶  $n_{hjm}$ : non-missing values for  $h$ th hour of  $j$ th day of week in month  $m$
- ▶  $VOL_{ihjm}$ :  $i$ th data point for  $h$ th hour of  $j$ th day of week in month  $m$
- ▶  $MADT_m$ : monthly average daily traffic for month  $m$



The background of the slide is composed of two large, diagonal geometric shapes. The top-left shape is a solid teal color, and the bottom-right shape is a solid light gray color. These two shapes meet at a diagonal line that runs from the top-left towards the bottom-right, leaving a white triangular area in the center where the text is located.

# Exploratory Data Analysis

# Outline

Problem Statement

Background

Motivation

Problem Statement

Data Wrangling

Raw Data

Data Cleaning

Transformation

**Exploratory Data Analysis**

**Data Distribution**

Autocorrelation

Imputation

Methods

Results

Modeling

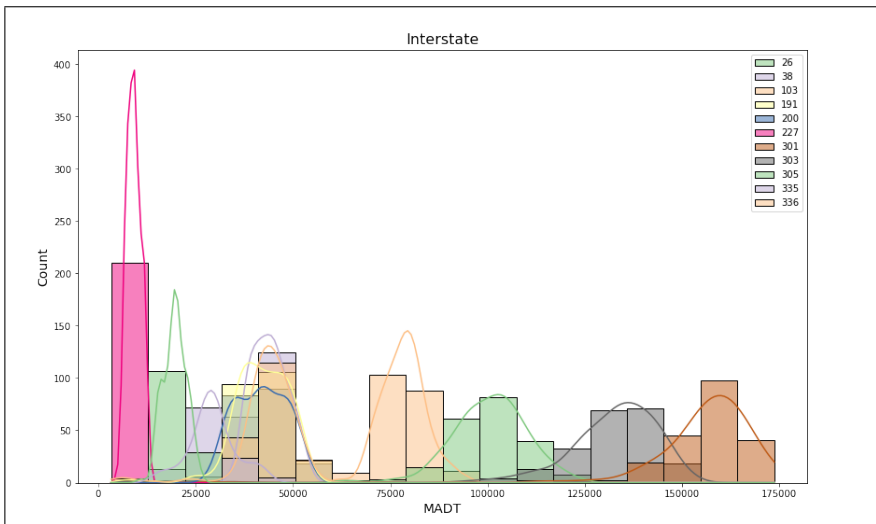
Methods

Results

Future Directions

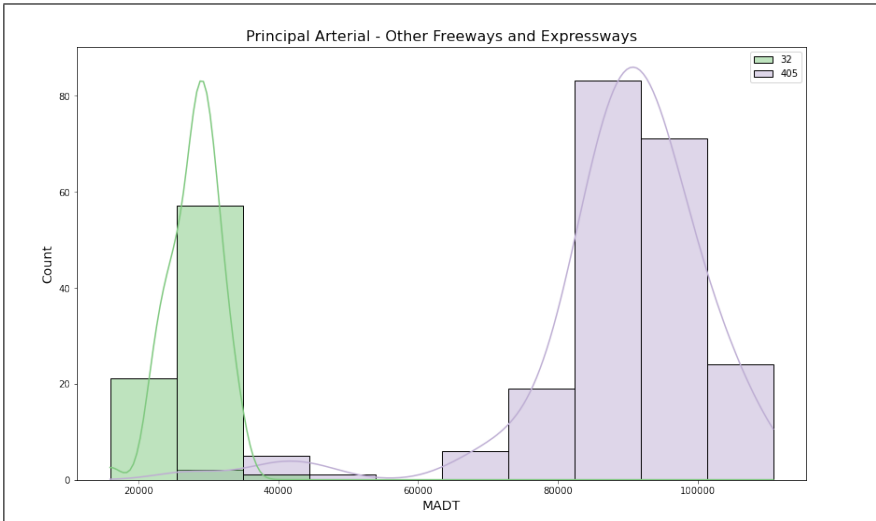
# Exploratory Data Analysis

## Data Distribution: Interstates



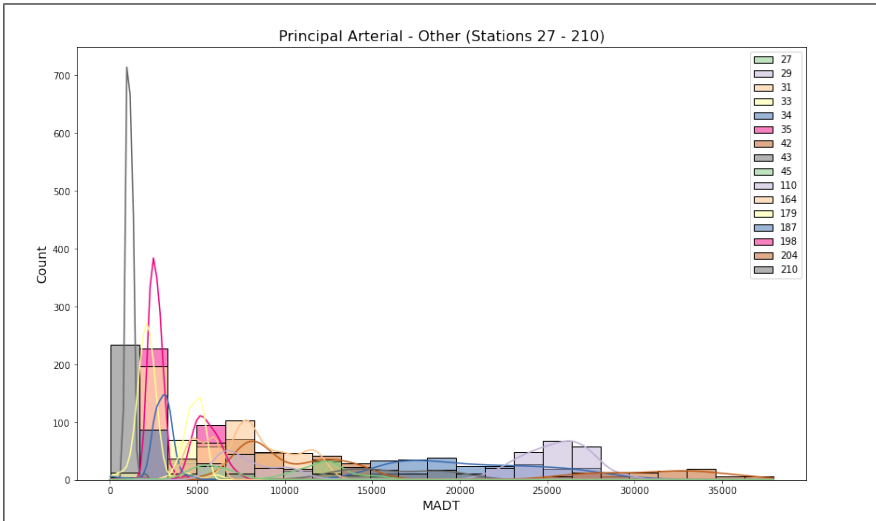
# Exploratory Data Analysis

## Data Distribution: Principal Arterial - Other Freeways



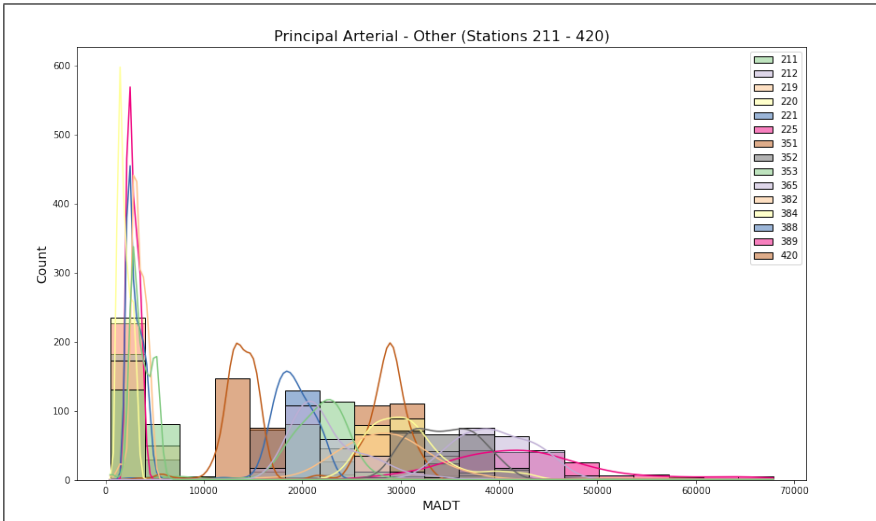
# Exploratory Data Analysis

## Data Distribution: Principal Arterial - Other



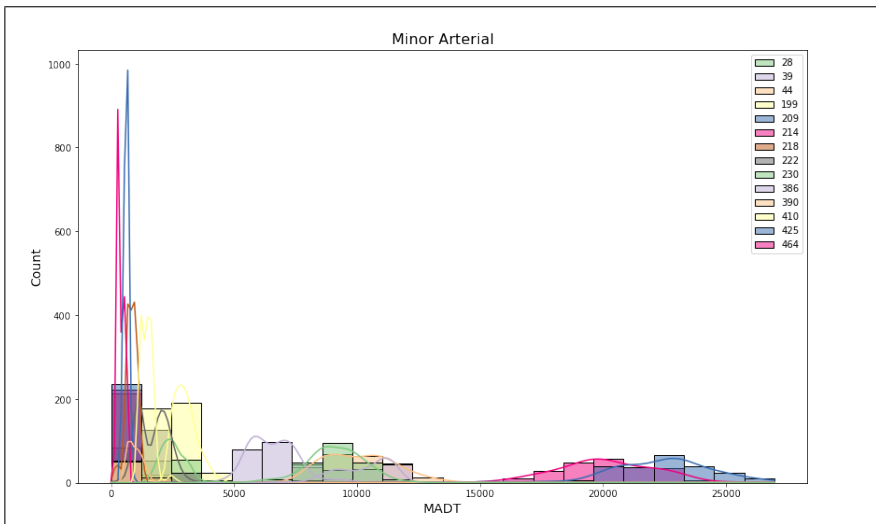
# Exploratory Data Analysis

## Data Distribution: Principal Arterial - Other



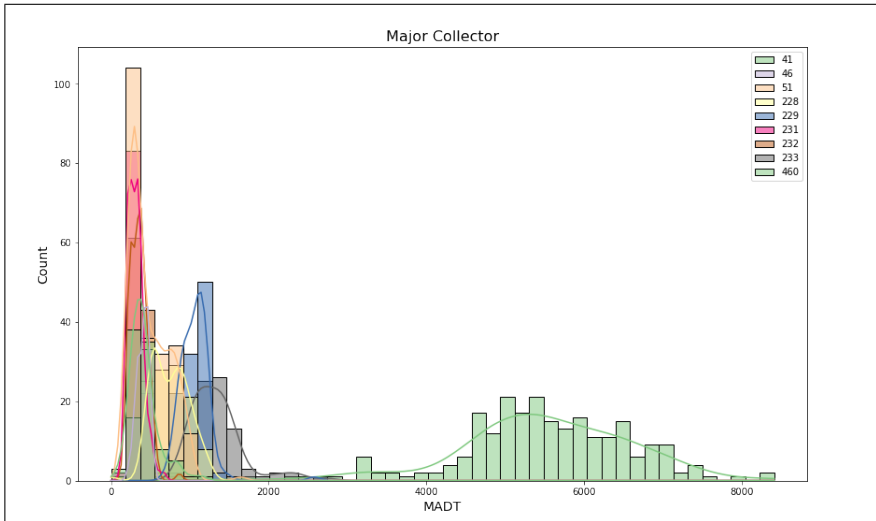
# Exploratory Data Analysis

## Data Distribution: Minor Arterial



# Exploratory Data Analysis

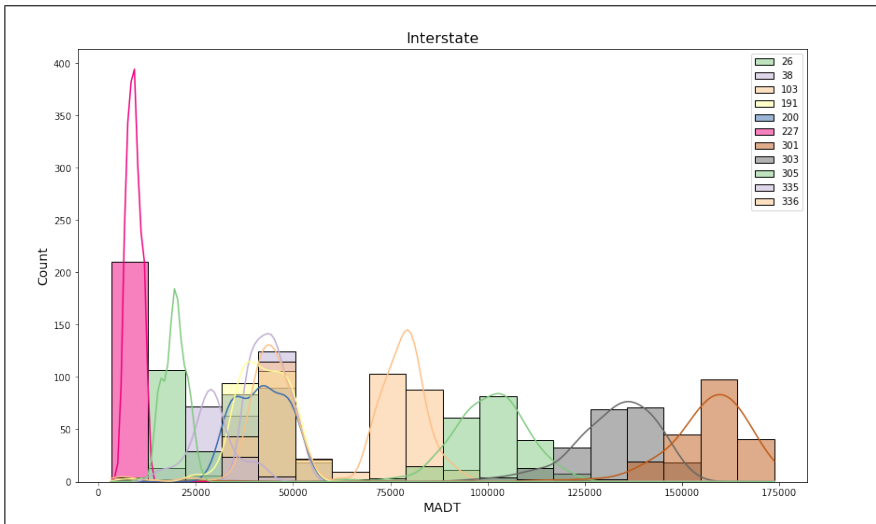
## Data Distribution: Major Collector





# Exploratory Data Analysis

## Data Distribution: Local



# Outline

Problem Statement

Background

Motivation

Problem Statement

Data Wrangling

Raw Data

Data Cleaning

Transformation

**Exploratory Data Analysis**

Data Distribution

**Autocorrelation**

Imputation

Methods

Results

Modeling

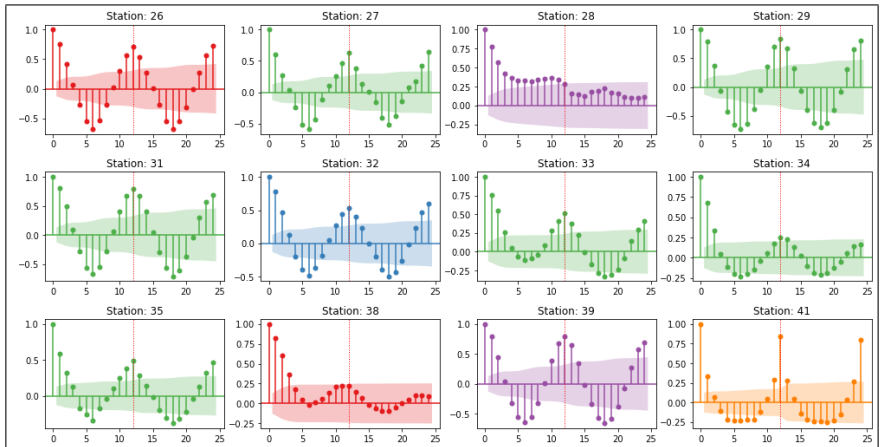
Methods

Results

Future Directions

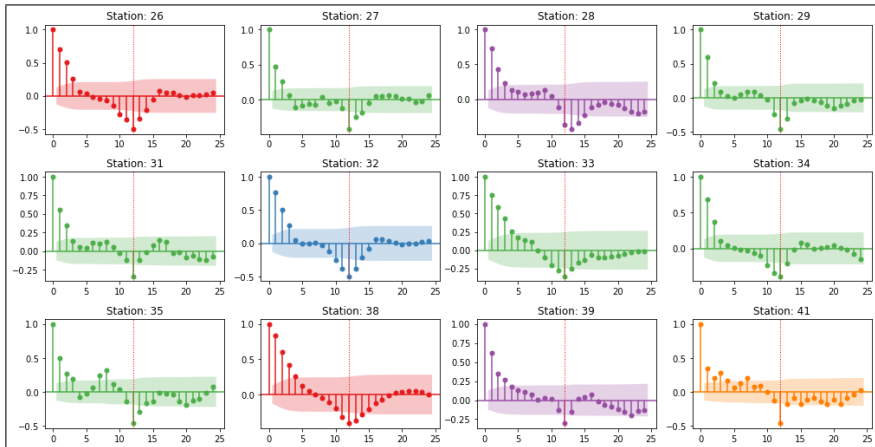
# Exploratory Data Analysis

## Autocorrelation: ACF



# Exploratory Data Analysis

## Autocorrelation: Differenced ACF



The background of the slide is composed of two large, solid-colored triangular areas. A teal triangle is located in the top-left corner, and a light gray triangle is in the bottom-left corner. These two triangles meet at a diagonal line that runs from the top-left towards the bottom-right, leaving a large white triangular area on the right side of the slide.

# Imputation

# Outline

Problem Statement

Background

Motivation

Problem Statement

Data Wrangling

Raw Data

Data Cleaning

Transformation

Exploratory Data Analysis

Data Distribution

Autocorrelation

**Imputation**

**Methods**

Results

Modeling

Methods

Results

Future Directions

# 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
  - ▶ CDRoc (Centroid Decomposition)
  - ▶ Seasonal CDRoc
  - ▶ Prophet
  - ▶ Prophet (Logistic Floor)
- ▶ All but CDRoc 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
  - ▶ CDRoc (Centroid Decomposition)
  - ▶ Seasonal CDRoc
  - ▶ Prophet
  - ▶ Prophet (Logistic Floor)
- ▶ All but CDRoc were column-wise

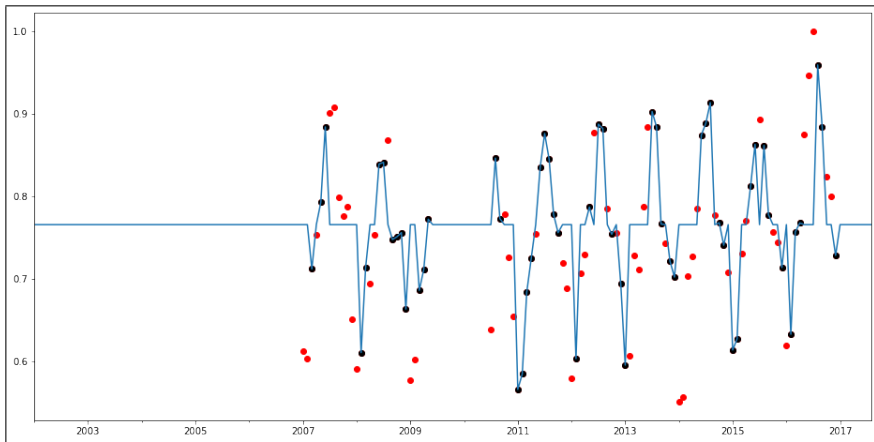
# Imputation

## 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

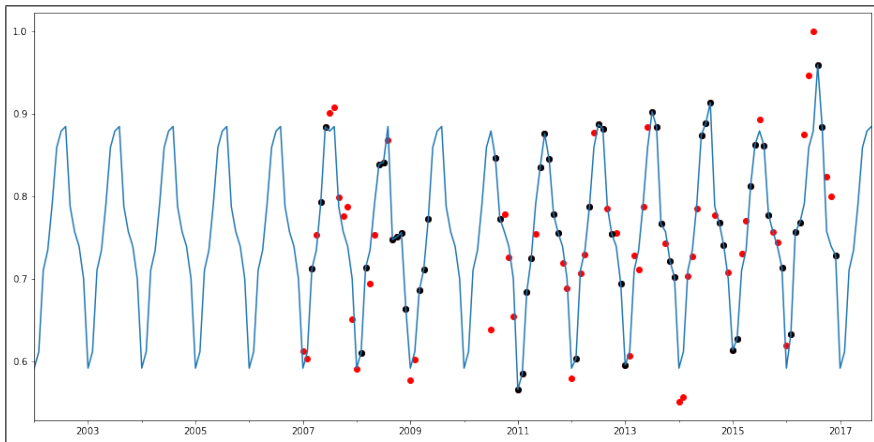
Methods: Mean





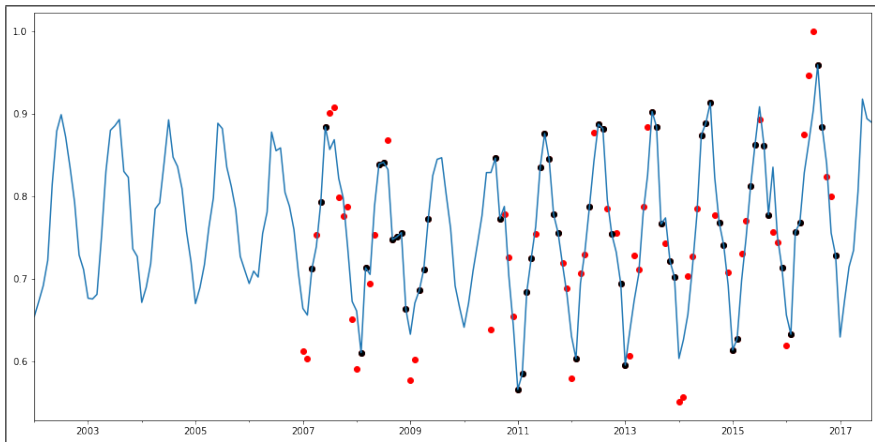
# Imputation

Methods: Seasonal Mean



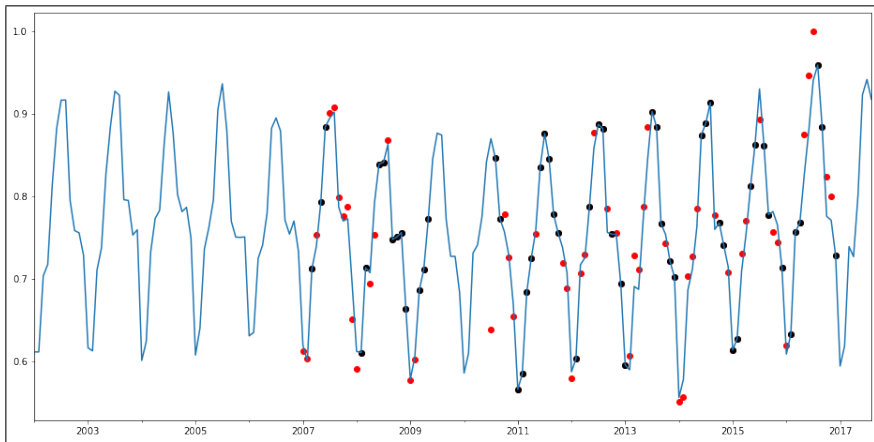
# Imputation

Methods: CDRec



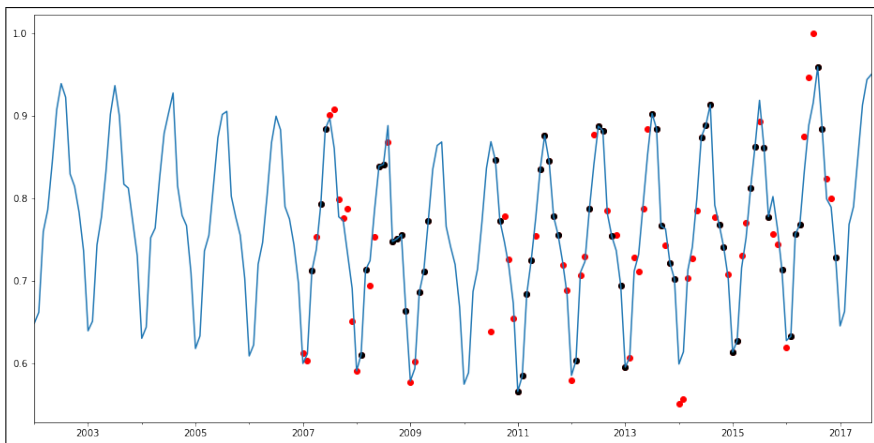
# Imputation

Methods: Seasonal CDRec



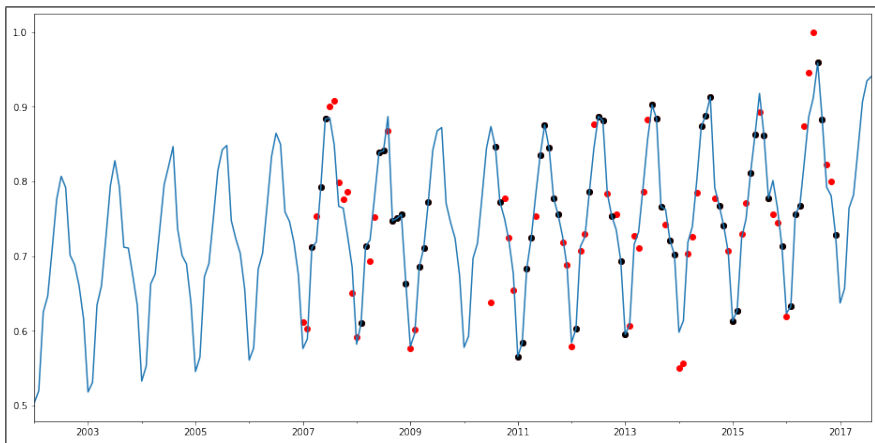
# Imputation

Methods: Prophet



# Imputation

Methods: Logistic Prophet



# Outline

Problem Statement

Background

Motivation

Problem Statement

Data Wrangling

Raw Data

Data Cleaning

Transformation

Exploratory Data Analysis

Data Distribution

Autocorrelation

**Imputation**

Methods

**Results**

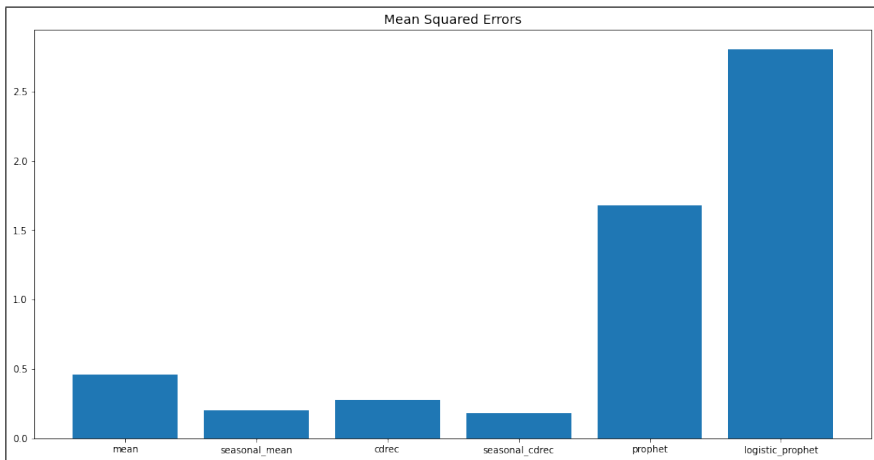
Modeling

Methods

Results

Future Directions

# Imputation Results



The background of the slide is composed of three geometric regions. A teal-colored triangle is located in the top-left corner. A light gray triangle is located in the bottom-left corner. The remaining area of the slide is white. The word "Modeling" is centered in the white region.

# Modeling



# Outline

Problem Statement

Background

Motivation

Problem Statement

Data Wrangling

Raw Data

Data Cleaning

Transformation

Exploratory Data Analysis

Data Distribution

Autocorrelation

Imputation

Methods

Results

**Modeling**

**Methods**

Results

Future Directions

# 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

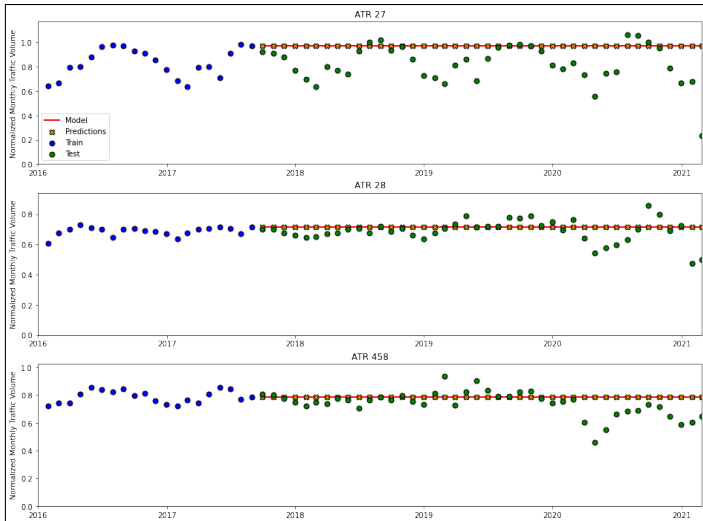
## Methods

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



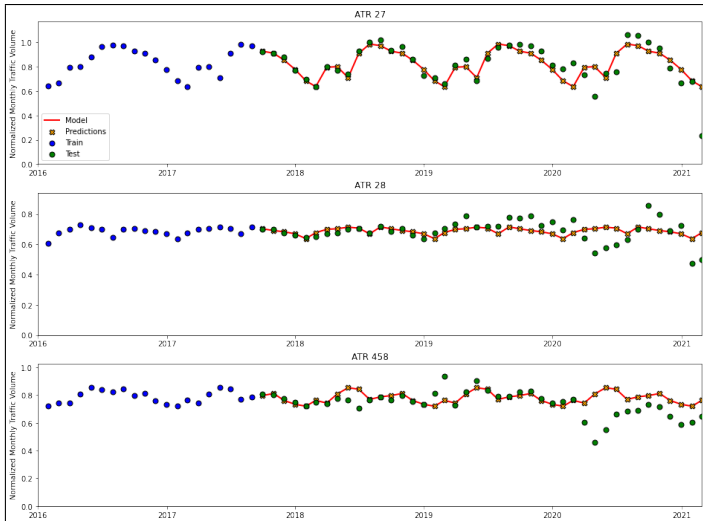
# Modeling

## Methods: 1 Month Lag



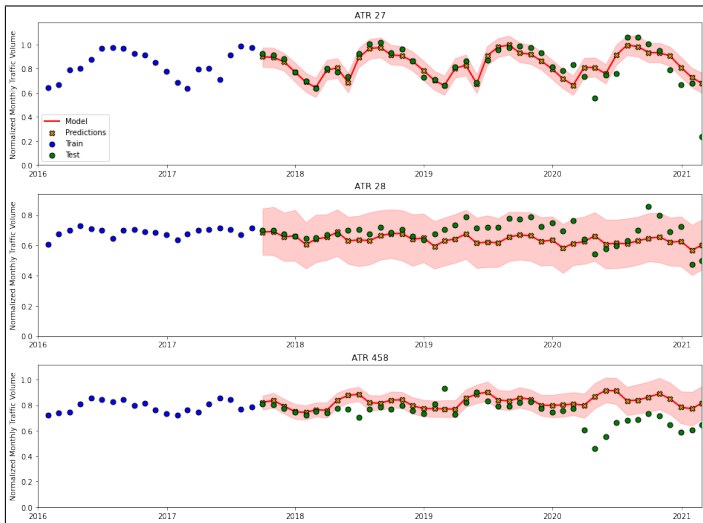
# Modeling

## Methods: 12 Month Lag



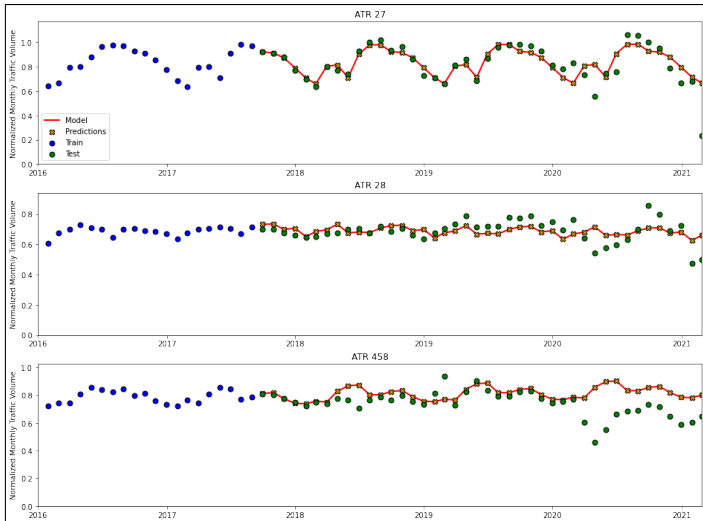
# Modeling

## Methods: Prophet



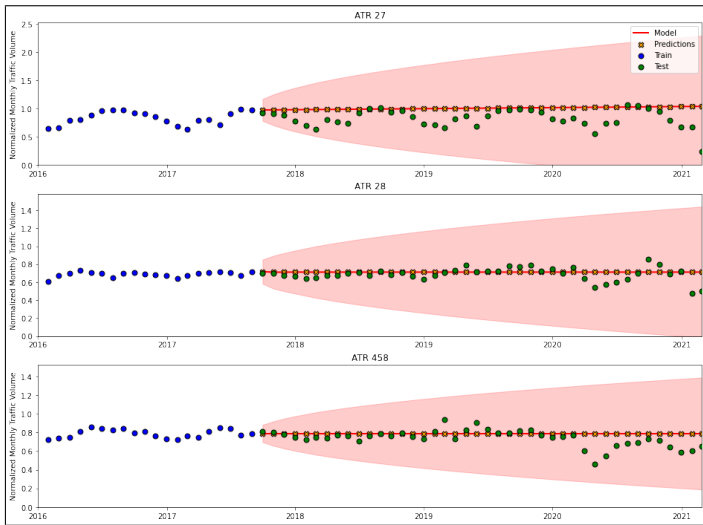
# Modeling

## Methods: Exponential Smoothing



# Modeling

## Methods: SARIMA



# Outline

Problem Statement

Background

Motivation

Problem Statement

Data Wrangling

Raw Data

Data Cleaning

Transformation

Exploratory Data Analysis

Data Distribution

Autocorrelation

Imputation

Methods

Results

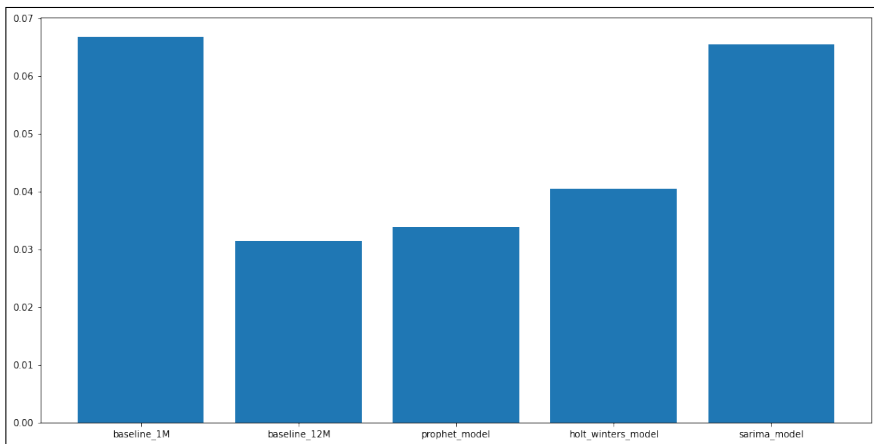
**Modeling**

Methods

**Results**

Future Directions

# Modeling Results



The background of the slide is composed of three geometric sections. A teal-colored triangle is in the top-left corner. A light gray triangle is in the bottom-left corner. The remaining area is a white trapezoid. The text "Future Directions" is centered within the white area.

## 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)