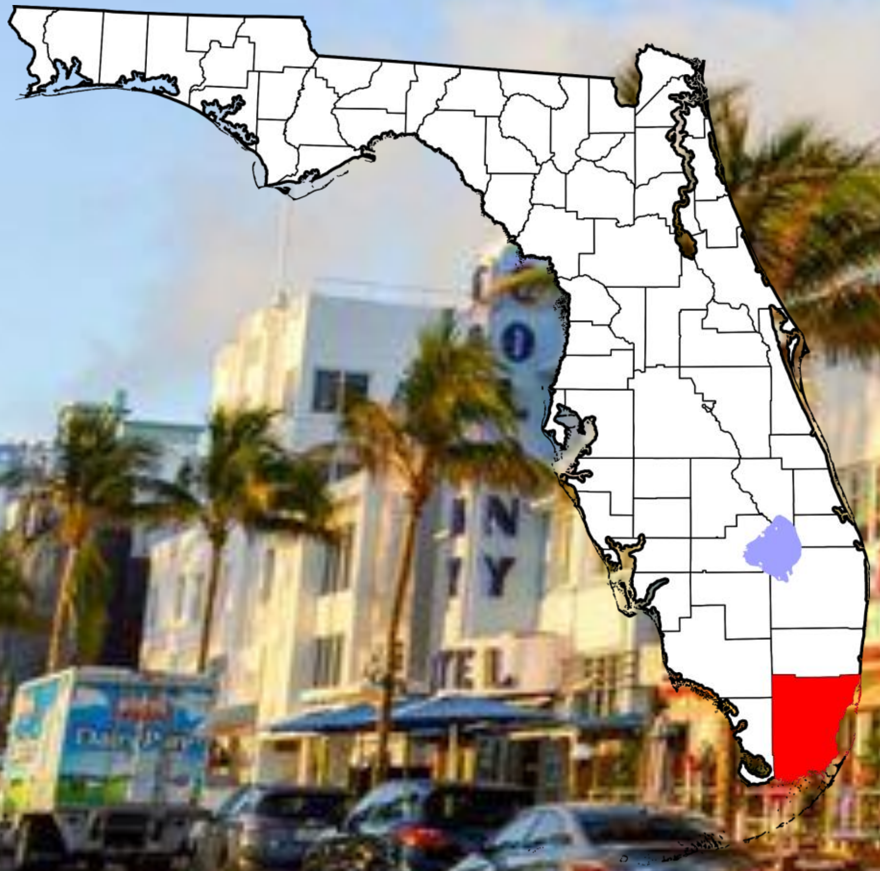
Miami-Dade Public Transit Analysis

Forecasting ridership and a brief look into ridership trends



Public transportation can be a ride out of poverty

The relationship between transportation and social mobility is stronger than that between mobility and several other factors, like crime, elementary-school test scores or the percentage of two-parent families in a community

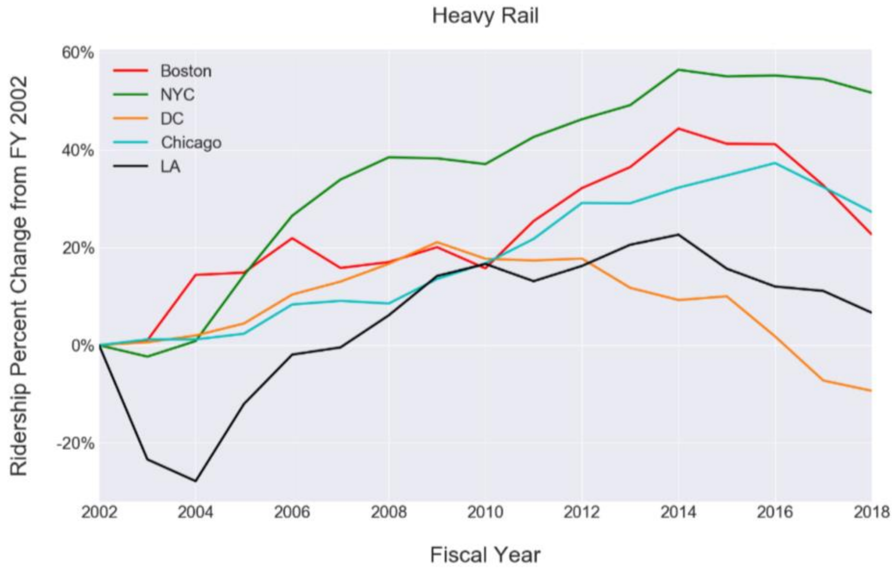
commuting time has emerged as the single strongest factor in the odds of escaping poverty

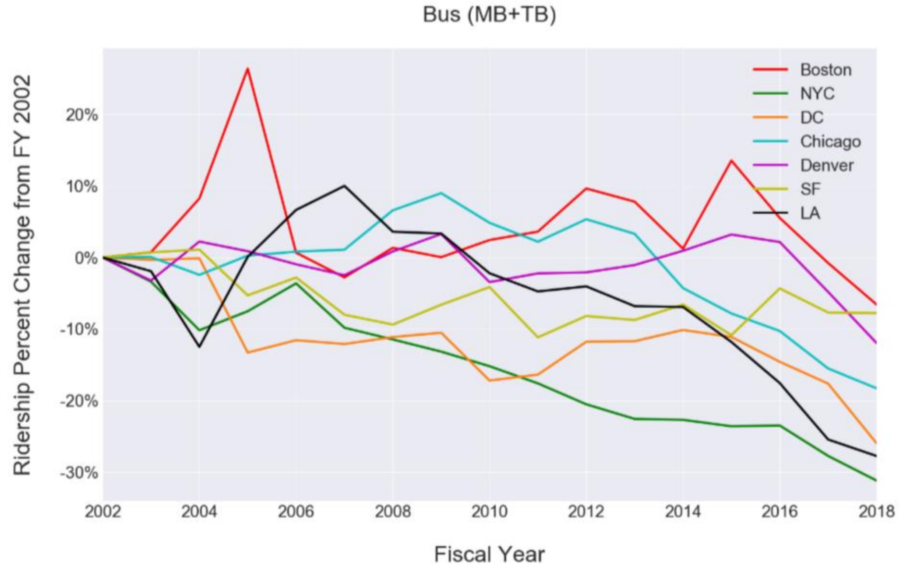
$1 invested → $4 economic return

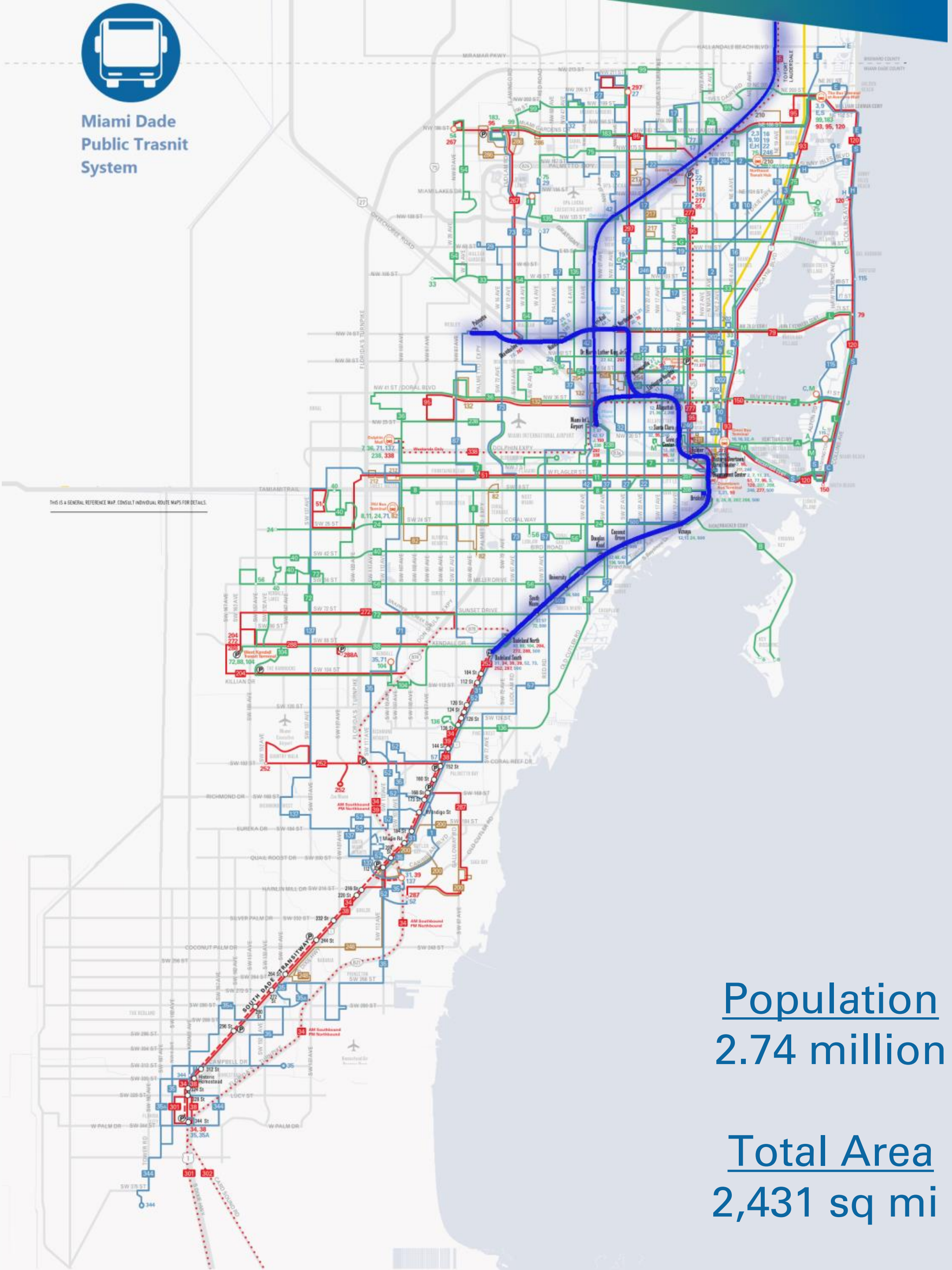
Ridership decreasing across US

<u>Why, tho?</u>

- decrease in gas prices
- increased car ownership
- expansion of ridesharing
- expansion of bike and scooter sharing
- reliability issues with public transit
- service cuts

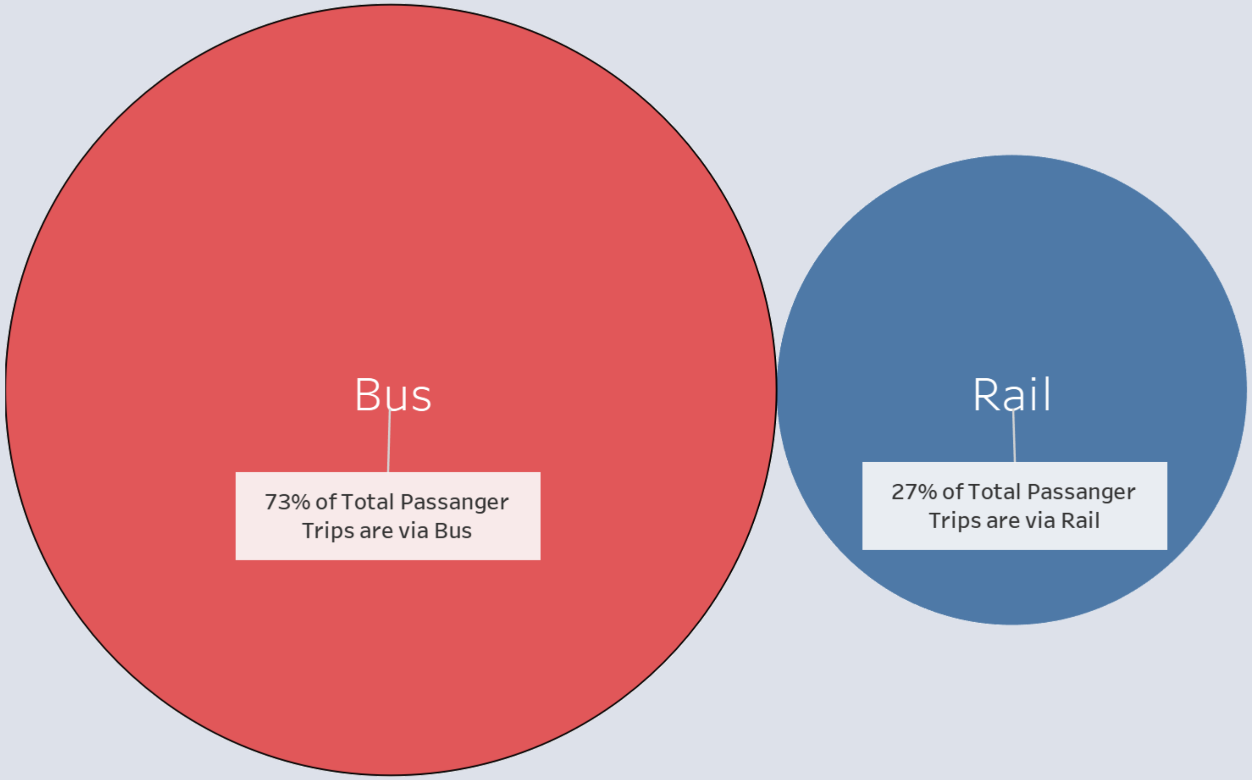






Population
2.74 million

Total Area
2,431 sq mi



Forecasting Analysis

Method

Time series analysis independently for Bus and Rail

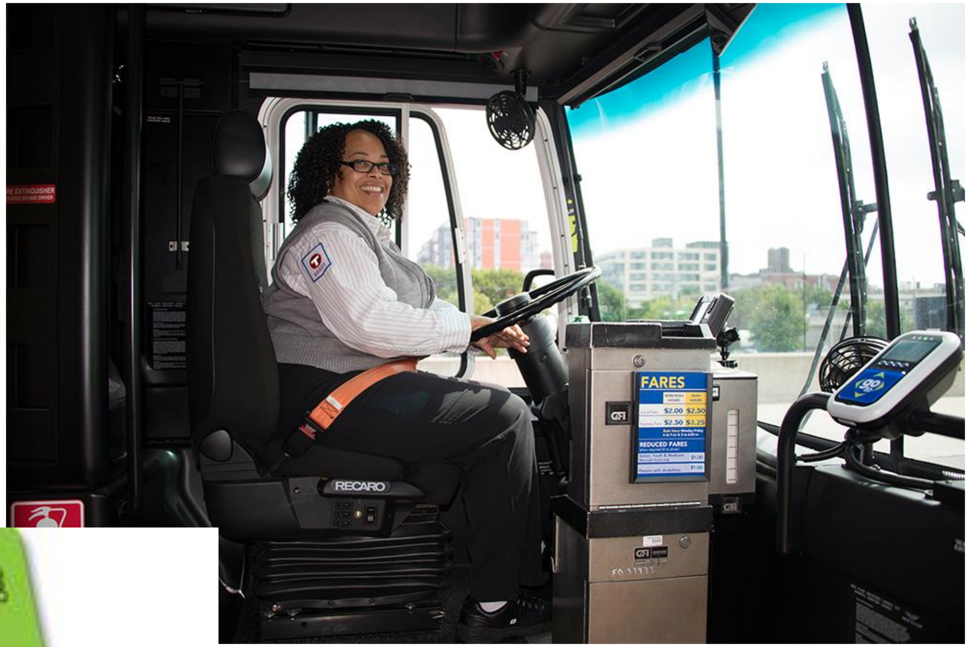
Data

Number of Unlinked Passenger Trips
per month from January 2002 until January 2020
sourced from the National Transit Database
filtered for Miami-Dade County

Aim

Forecast until December 2021

Unlinked Passenger Trip explained



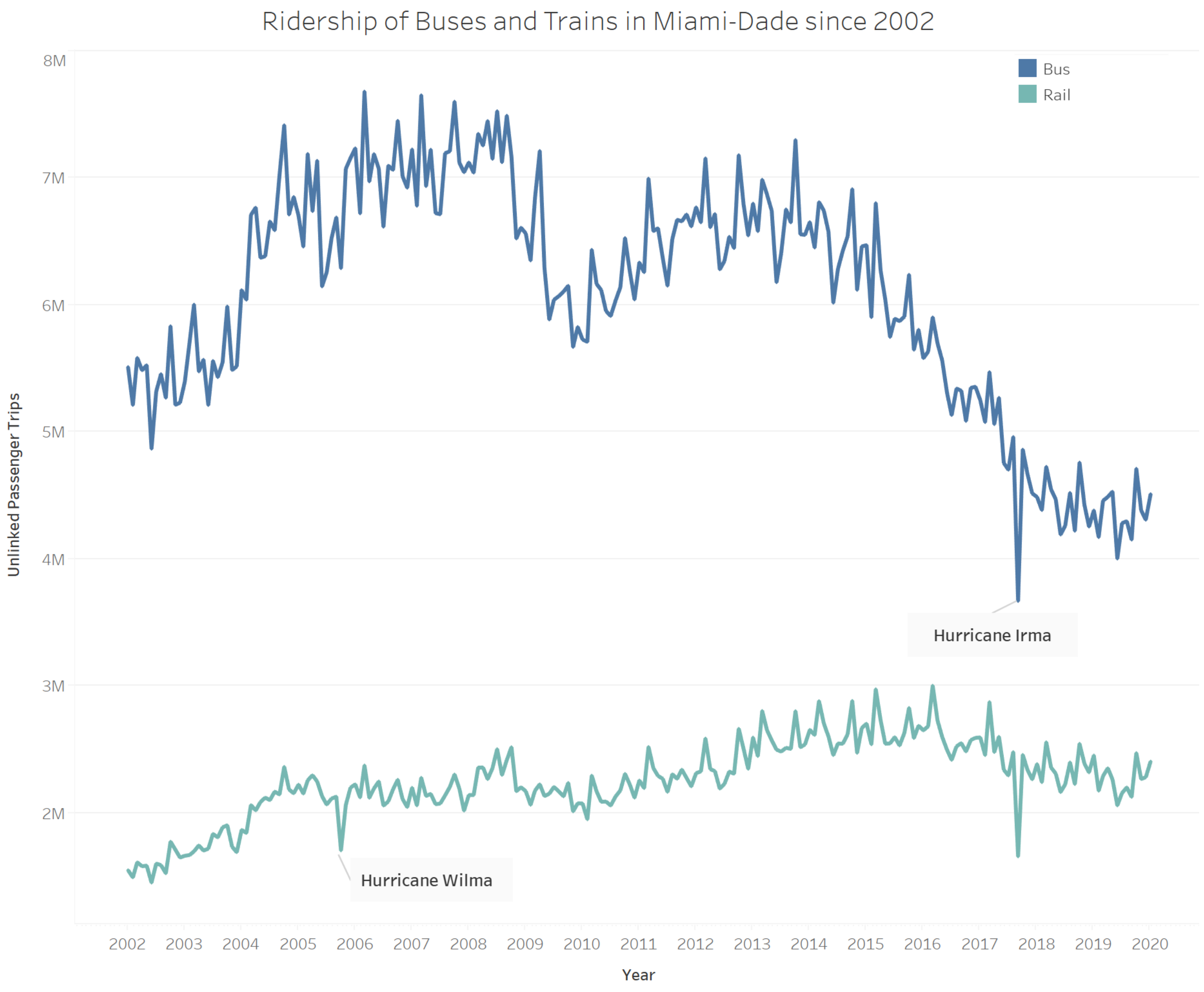
Key Insights

General drop for both Bus and Rail, but stronger for Bus

Effects of hurricanes strongly felt

Last year annual growth seen:
- 2013 for Bus
- 2015 for Rail

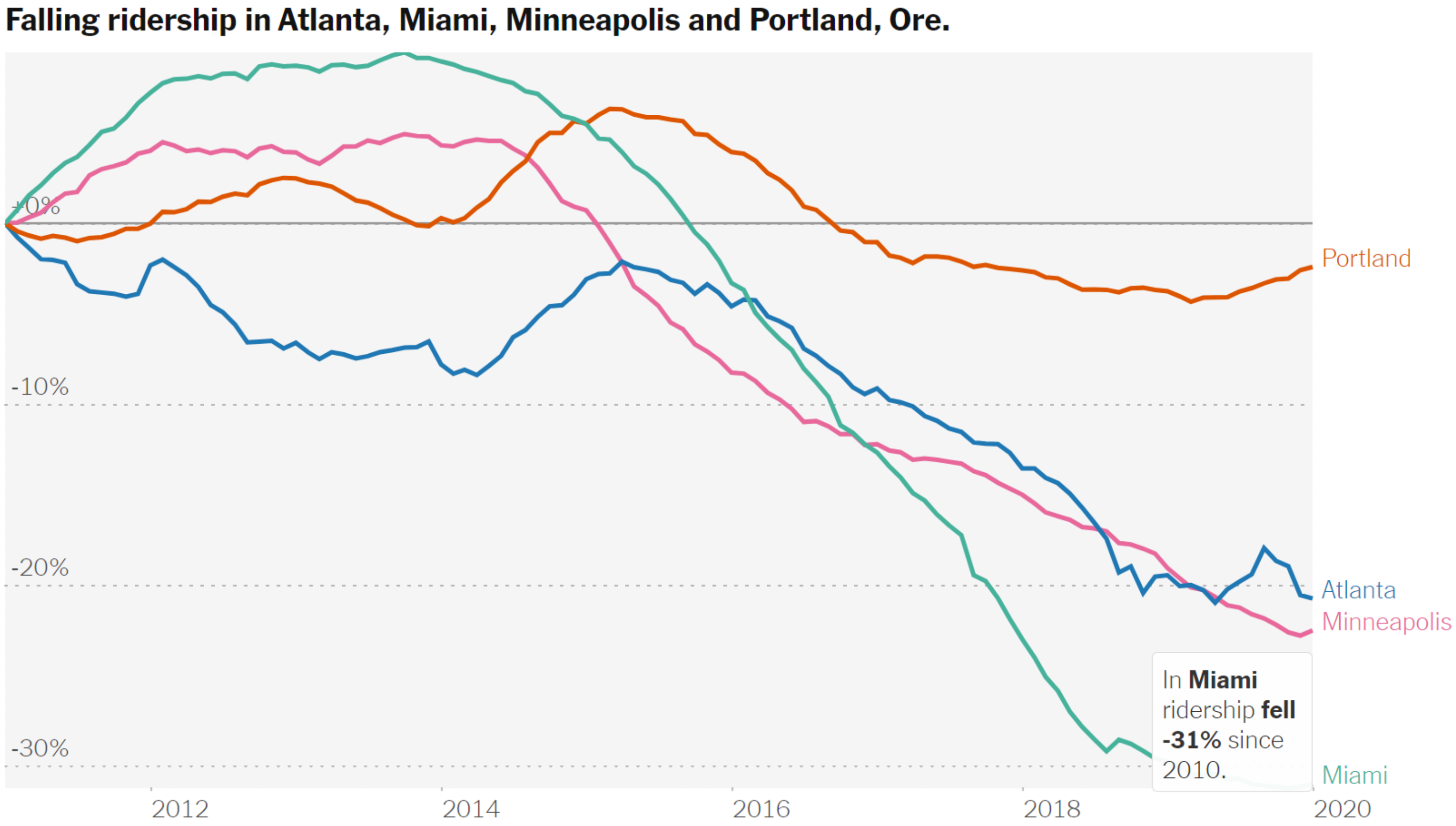
Scaled variance (coefficient of variation) relative for both:
-0.15 for Bus
-0.13 for Rail



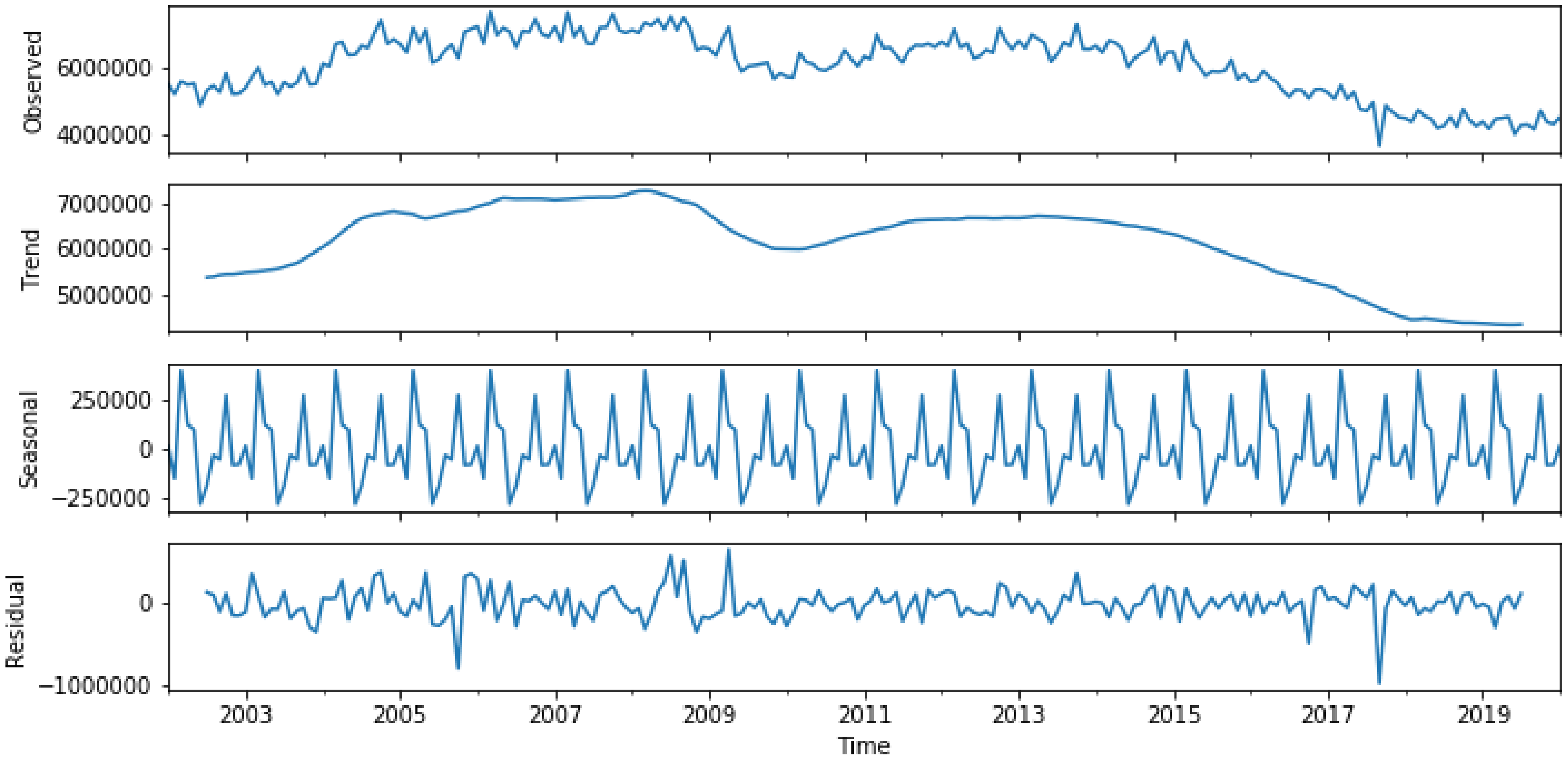
Bus Ridership Forecasting

Reality

Dream

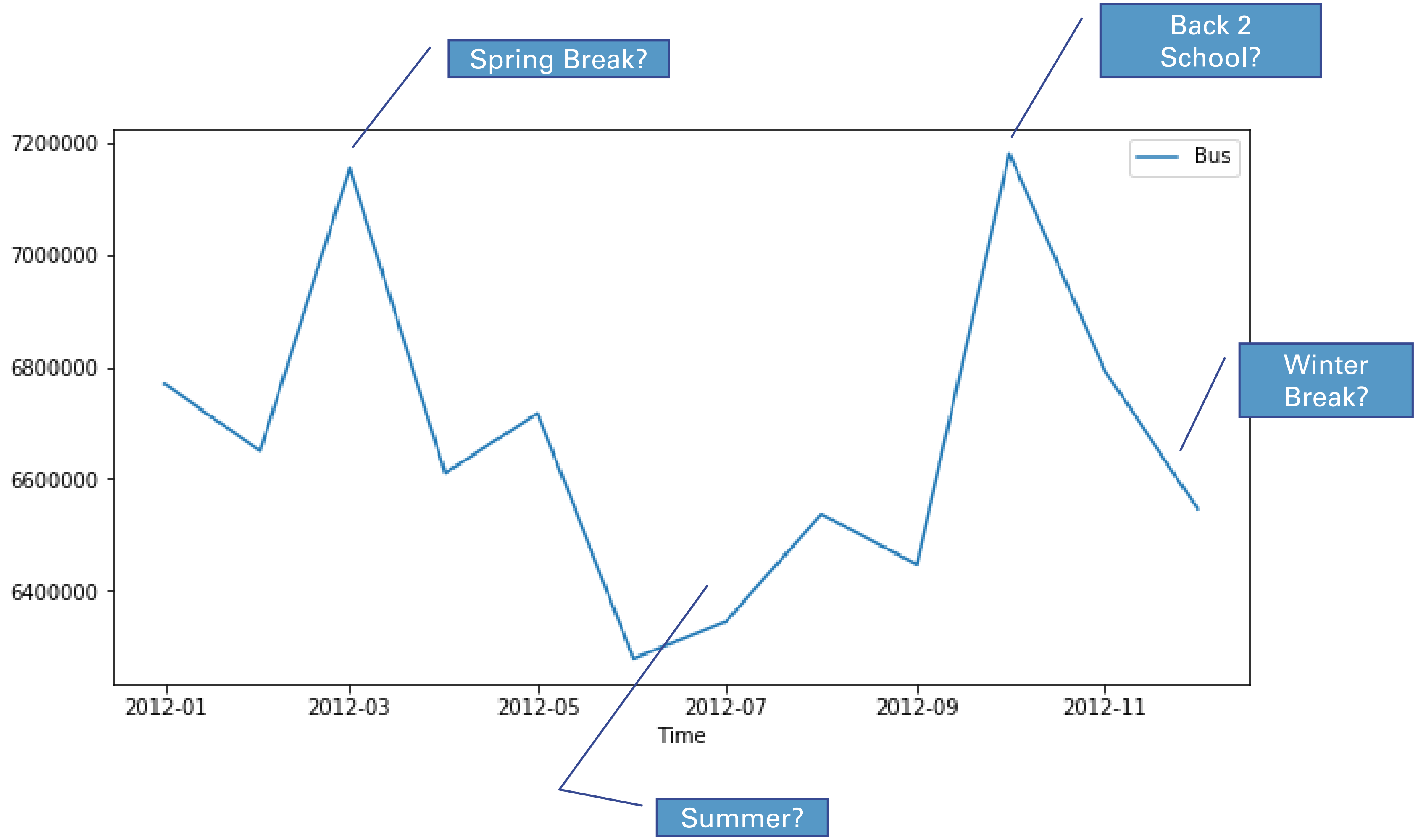


Time Series Forecasting
Seasonal Decomposition



Seasonality in Focus

Bus



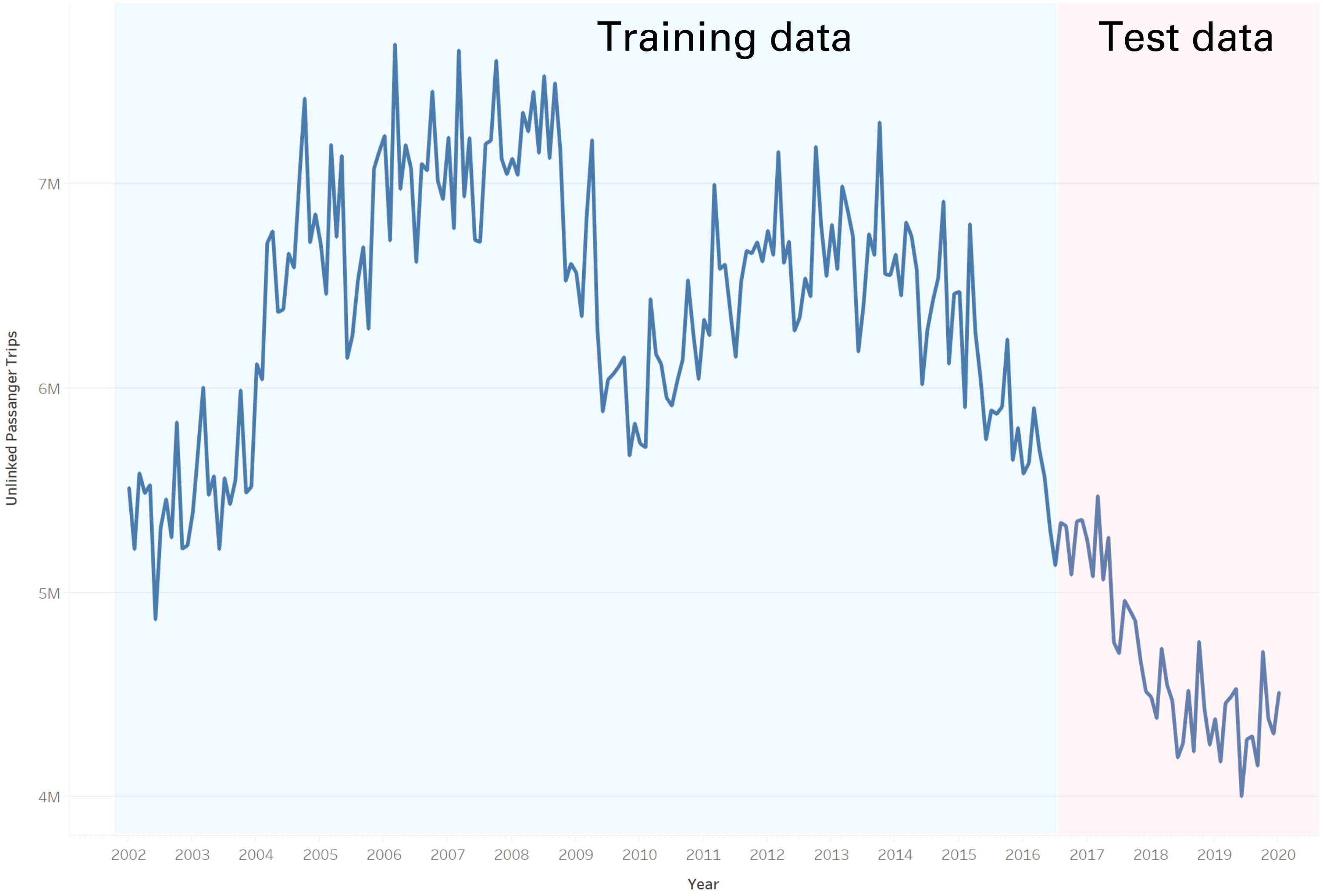
Methodology

- Averaged out hurricane data points

- Split data with 80% - 20%

- Grid search to find the best parameters for my model based on the lowest value of AIC, a measure of accuracy

- Used an SARIMAX model to predict my test data, and forecast until 2021



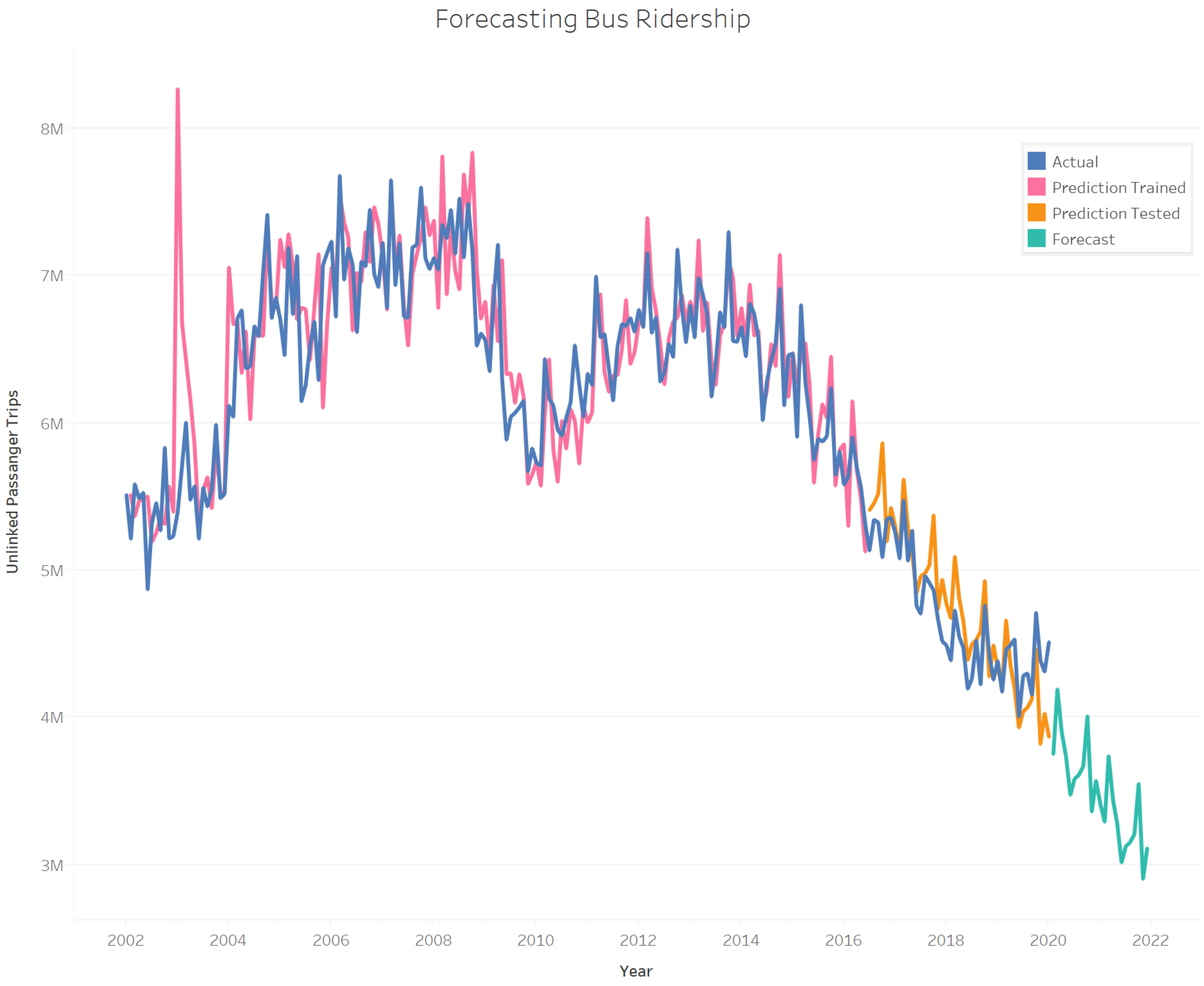
Time Series – Model

Grid Search -- SARIMA

```python
import itertools
#set parameter range
p = range(0,3)
q = range(0,3)
d = range(1,2)
s = range(12,13)
# list of all parameter combos
pdq = list(itertools.product(p, d, q))
seasonal_pdq = list(itertools.product(p, d, q, s))
# SARIMA model pipeline
for param in pdq:
    for param_seasonal in seasonal_pdq:
        try:
            mod = sm.tsa.statespace.SARIMAX(train,
                                            order=param,
                                            seasonal_order=param_seasonal)
            results = mod.fit(max_iter = 50, method = 'powell')
            print('SARIMA{},{} - AIC:{}'.format(param, param_seasonal, results.aic))
        except:
            continue
```

```python
model = SARIMAX(train, order=(0,1,1), seasonal_order=(1,1,0,12), enforce_stationarity=False, enforce_invertibility=False)
fitted = model.fit()
print(fitted.summary())
fig, ax = plt.subplots()
fig.set_size_inches(20, 10)
plt.plot(train)
fitted_df = fitted.fittedvalues[1:]
plt.plot(fitted_df, color='red')
plt.show()
```

```
                           Statespace Model Results
==========================================================================================
Dep. Variable:                               Bus   No. Observations:                  174
Model:             SARIMAX(0, 1, 1)x(1, 1, 0, 12)   Log Likelihood               -2092.309
Date:                           Thu, 19 Mar 2020   AIC                           4190.619
Time:                                   11:16:53   BIC                           4199.631
Sample:                               01-01-2002   HQIC                          4194.280
                                    - 06-01-2016
Covariance Type:                             opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -0.5093      0.065     -7.793      0.000      -0.637      -0.381
ar.S.L12      -0.3915      0.074     -5.261      0.000      -0.537      -0.246
sigma2      1.069e+11   8.02e-14   1.33e+24      0.000    1.07e+11    1.07e+11
===================================================================================
Ljung-Box (Q):                       69.41   Jarque-Bera (JB):                  1.01
Prob(Q):                              0.00   Prob(JB):                          0.60
Heteroskedasticity (H):               0.31   Skew:                             -0.01
Prob(H) (two-sided):                  0.00   Kurtosis:                          3.40
===================================================================================
```



Accuracy Measures

Akaike information criterion (AIC)
4179

Mean Absolute Percentage Error
0.047

Which translates to accounting for 95% of data

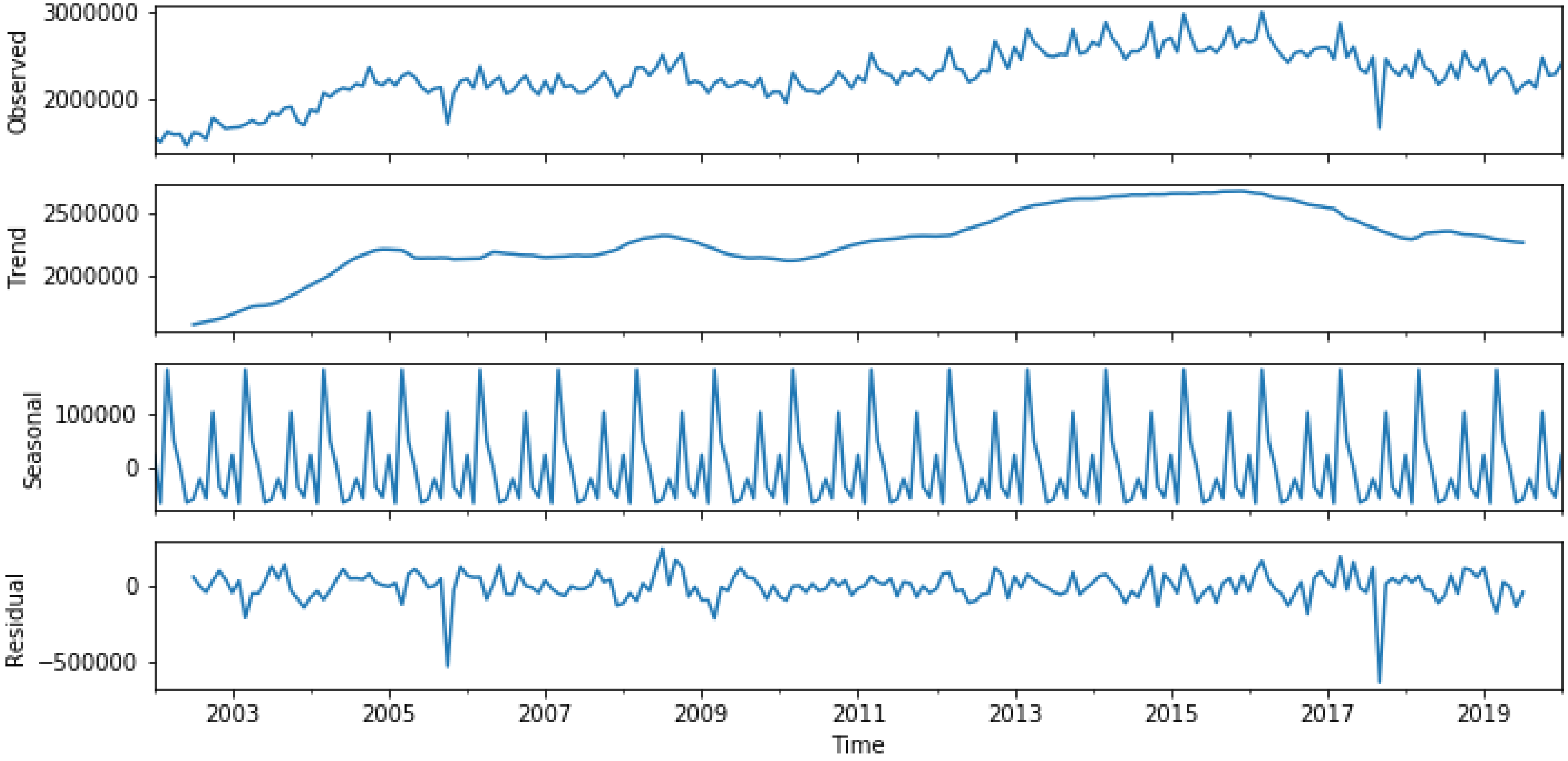
Mean Error
66,002

Mean Absolute Error
219,529

Root Mean Square Error
276,301

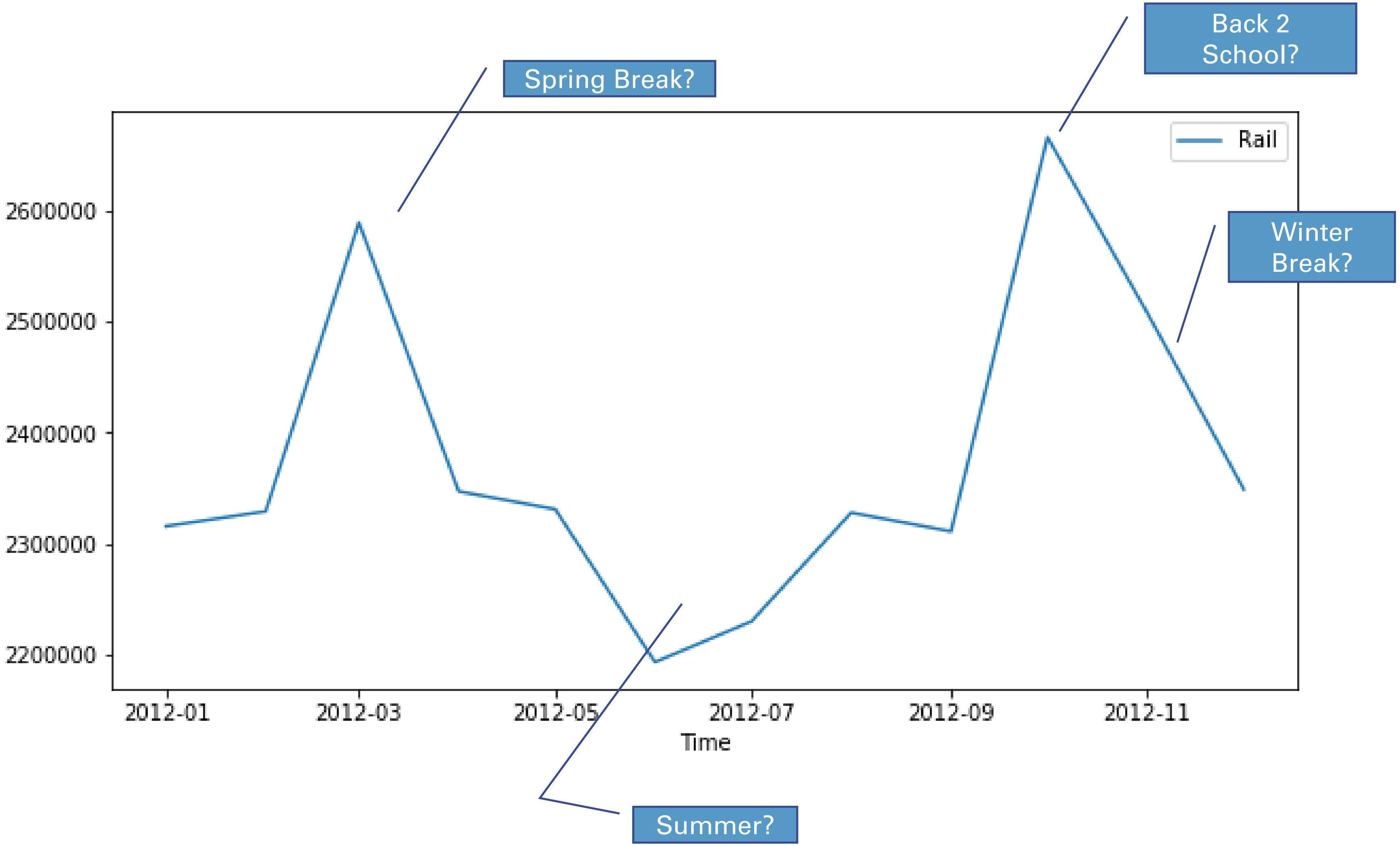
Rail

Time Series Forecasting
Seasonal Decomposition



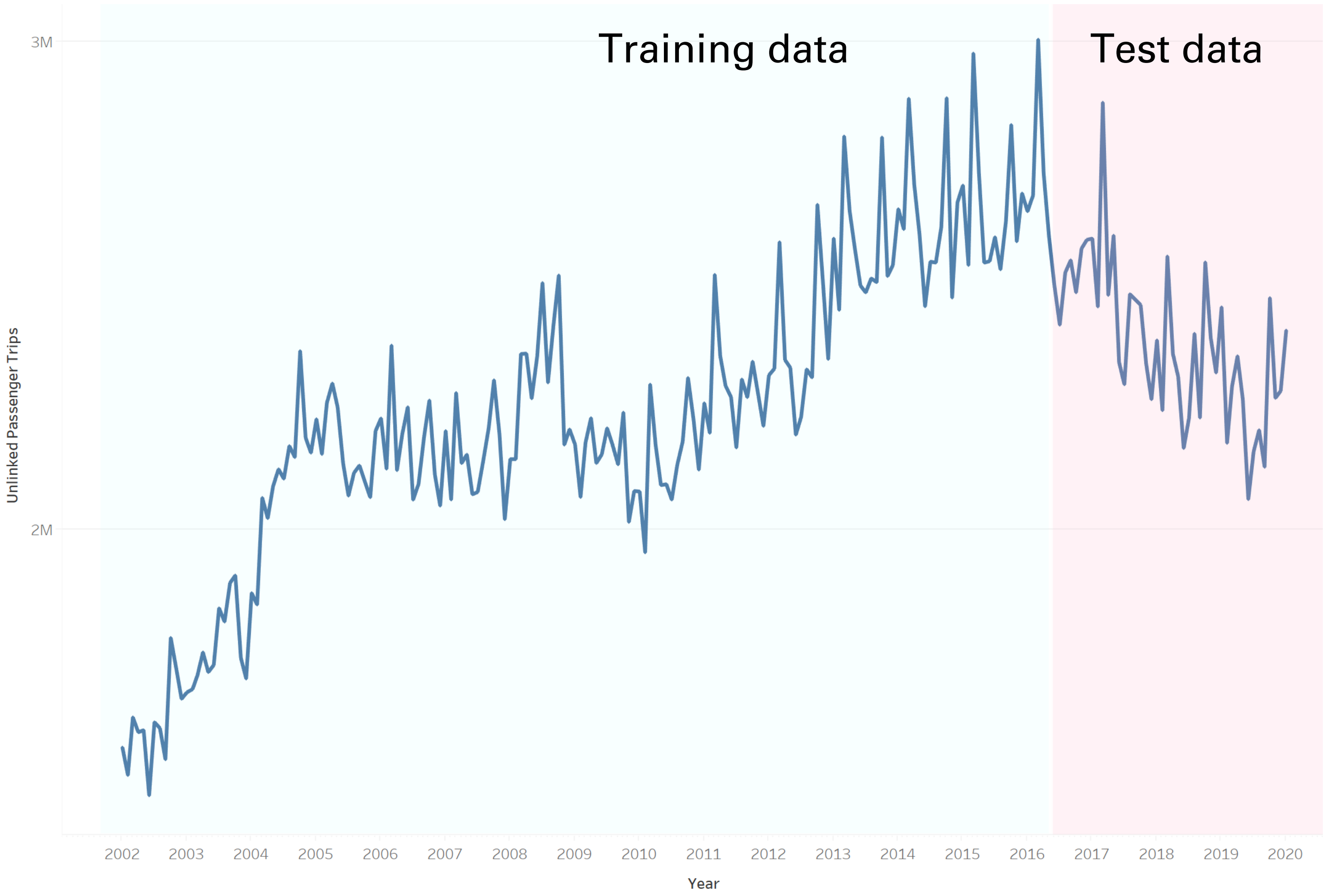
Seasonality in Focus

Rail



Methodology

- Averaged out hurricane data points
- Split data with 80% - 20%
- Grid search to find the best parameters for my model based on the lowest value of MAPE, a measure of accuracy
- Used an SARIMAX model to predict my test data, and forecast until 2021



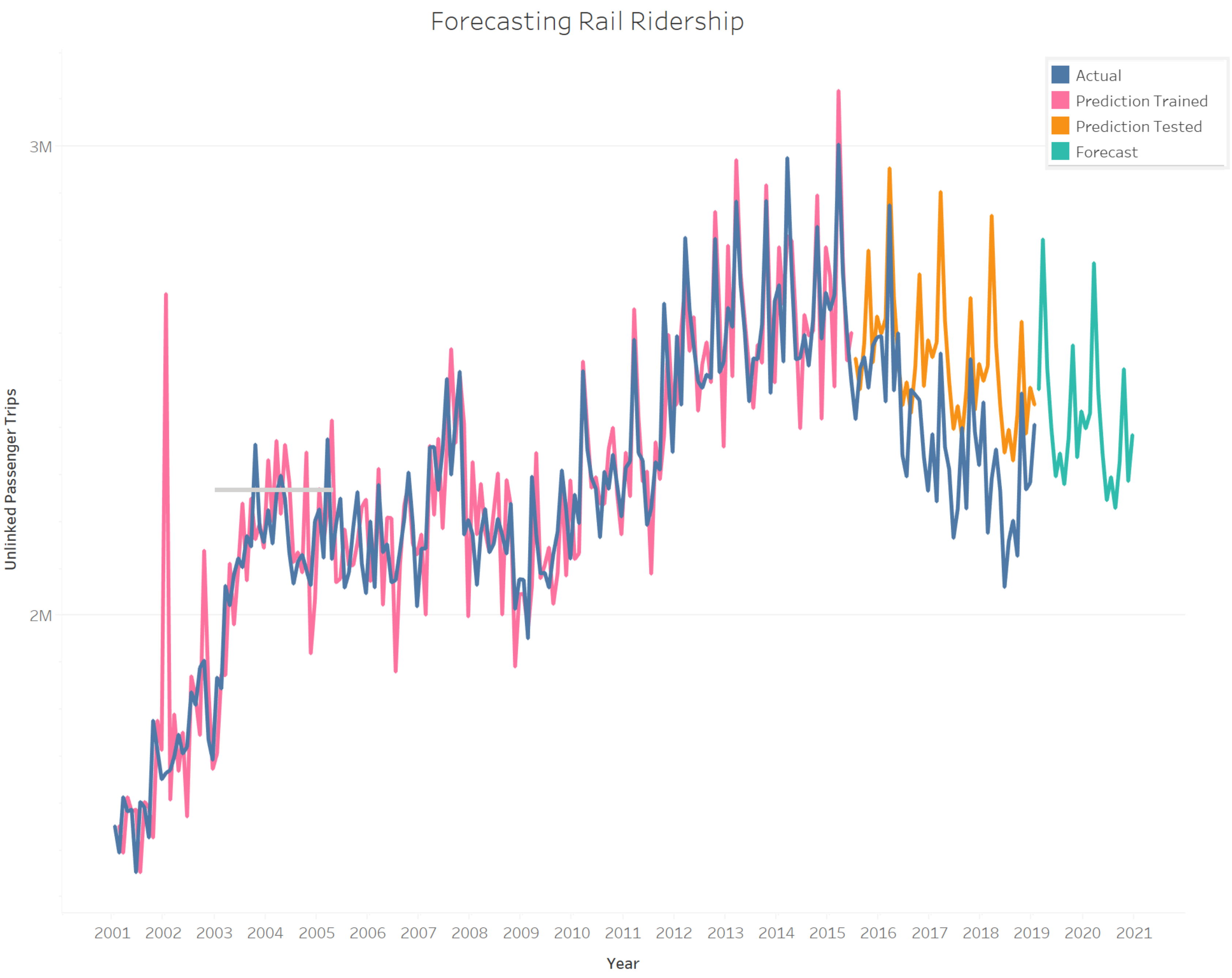
Time Series – Model

Grid Search -- SARIMA

```
import itertools
#set parameter range
p = range(0,3)
q = range(0,3)
d = range(1,2)
s = range(12,13)
# list of all parameter combos
pdq = list(itertools.product(p, d, q))
seasonal_pdq = list(itertools.product(p, d, q, s))
# SARIMA model pipeline
for param in pdq:
    for param_seasonal in seasonal_pdq:
        try:
            mod = SARIMAX(train,
                                order=param,
                                seasonal_order=param_seasonal)
            results = mod.fit(max_iter = 50, method = 'powell')
            forecast = results.predict(start=1, end=int(len(miami_test)))
            miami_test['Forecast'] = forecast
            start = datetime.datetime.strptime("2020-02-01", "%Y-%m-%d")
            date_list = pd.date_range('2020-02-01', freq='1M', periods=23)
            data = pd.DataFrame(index=date_list, columns= miami_test.columns)
            data.index = data.index.map(lambda t: t.replace(day=1))
            pred_uc = results.get_forecast(steps=66)
            data['Future'] = pred_uc.predicted_mean
            predictions = miami_test[174:]
            predictions = predictions.drop(['Future'], axis=1)
            mape = np.mean(np.abs(predictions["Forecast"] - predictions["Rail"])/np.abs(predictions["Rail"]))
            print('SARIMA{},{} - MAPE:{}'.format(param, param_seasonal, mape))
        except:
            continue
```

```
model = SARIMAX(train, order=(0,1,0), seasonal_order=(0,1,0,12), enforce_stationarity=False, enforce_invertibility=False)
fitted = model.fit()
print(fitted.summary())
fig, ax = plt.subplots()
fig.set_size_inches(20, 10)
plt.plot(train)
fitted_df = fitted.fittedvalues[1:]
plt.plot(fitted_df, color='red')
plt.show()
```

```
                           Statespace Model Results
==========================================================================================
Dep. Variable:                            Rail   No. Observations:                  174
Model:             SARIMAX(0, 1, 0)x(0, 1, 0, 12)   Log Likelihood               -2088.859
Date:                         Fri, 20 Mar 2020   AIC                           4179.717
Time:                                 00:07:09   BIC                           4182.792
Sample:                             01-01-2002   HQIC                          4180.966
                                  - 06-01-2016
Covariance Type:                           opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
sigma2      1.175e+10   1.39e+09      8.460      0.000    9.03e+09    1.45e+10
===================================================================================
Ljung-Box (Q):                      133.00   Jarque-Bera (JB):                 1.71
Prob(Q):                              0.00   Prob(JB):                         0.42
Heteroskedasticity (H):               0.73   Skew:                             0.08
Prob(H) (two-sided):                  0.25   Kurtosis:                         2.52
===================================================================================
```



Accuracy Measures

Akaike information criterion (AIC)
4179

Mean Absolute Percentage Error
0.075

Which translates to accounting for 92.5% of data

Mean Error
165,255

Mean Absolute Error
174,692

Root Mean Square Error
210,202

Correlations & Explanations (Graehler, Mucci & Erhardt, 2018)

Increase employment → Increase Ridership

Increase population → Increase Ridership

Increase gas prices → Increase Ridership

<u>Ridesharing Effects</u>:

Complimentary for Rail

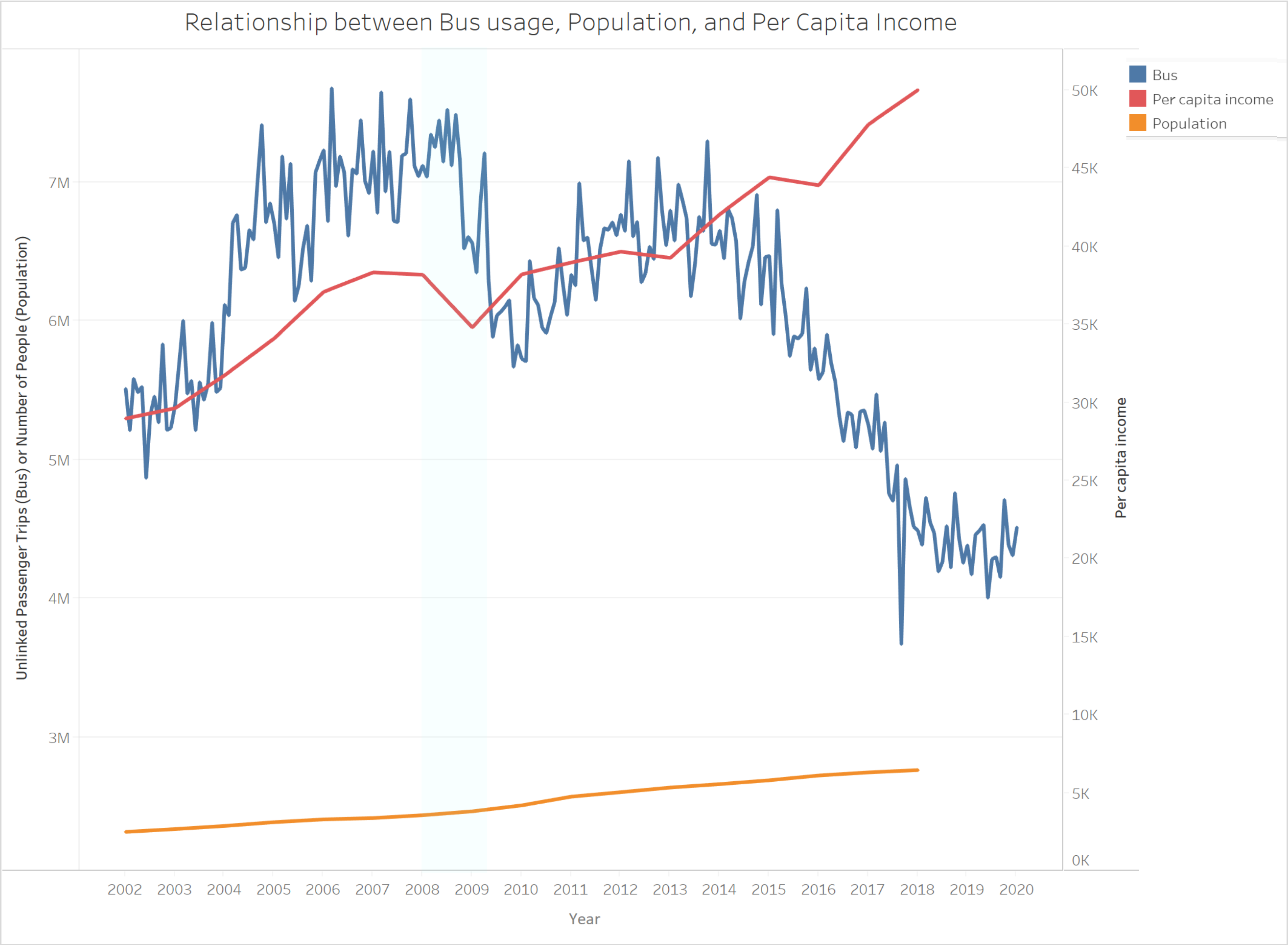
Competing for Bus (effect greater each year ridesharing in city by -1.70% per year)

Understanding the Recent Transit Ridership Decline in Major US Cities: Service Cuts or Emerging Modes?

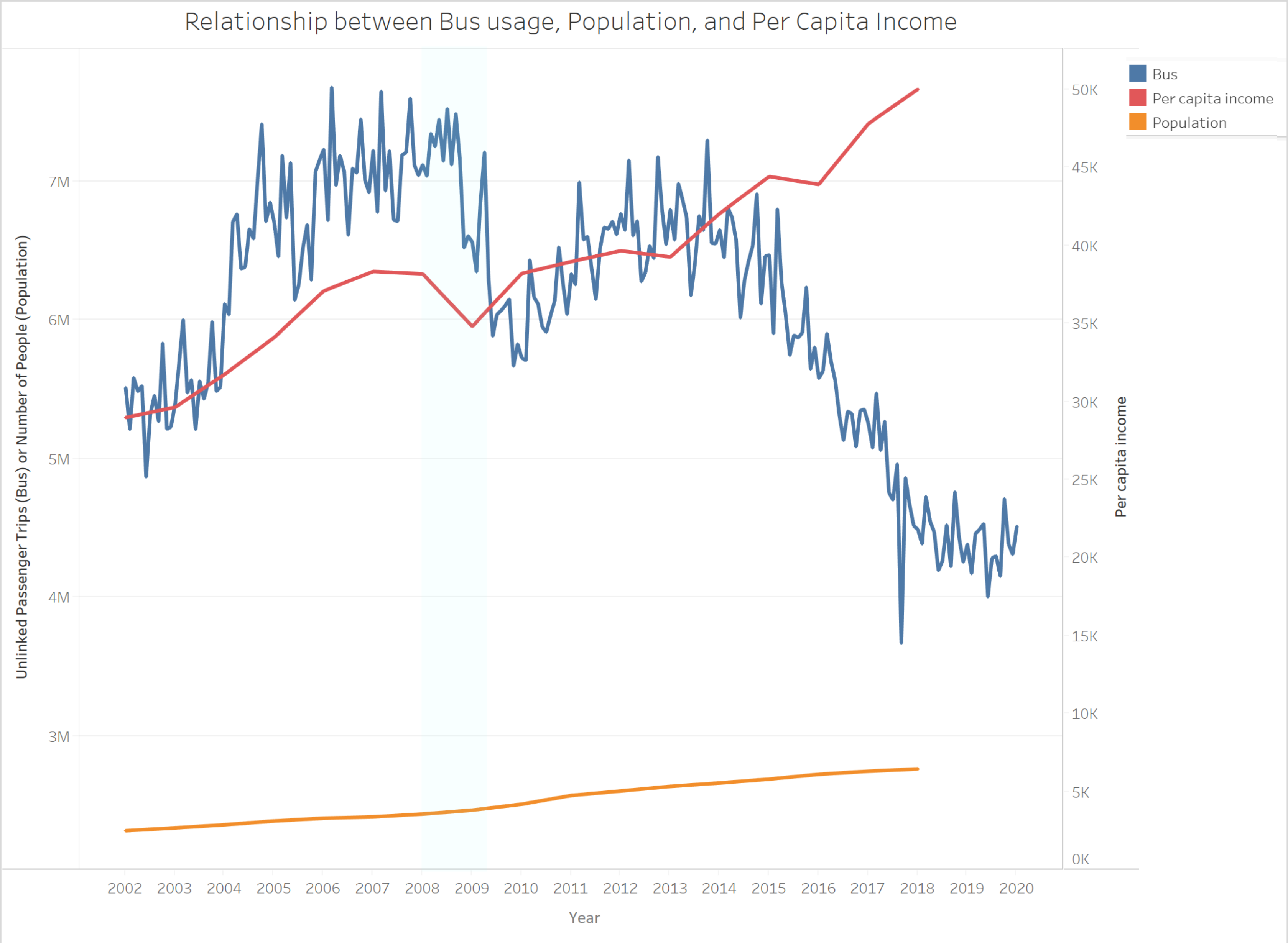
Michael Graehler, Jr.
Department of Civil Engineering, University of Kentucky
216 Oliver H. Raymond Bldg., Lexington, KY 40506
859-492-7535, michael.graehler@uky.edu

Richard Alexander Mucci
Department of Civil Engineering, University of Kentucky
216 Oliver H. Raymond Bldg., Lexington, KY 40506
859-257-4856, alex.mucci@uky.edu

Gregory D. Erhardt (corresponding author)
Department of Civil Engineering, University of Kentucky
261 Oliver H. Raymond Bldg., Lexington, KY 40506
859-323-4856, greg.erhardt@uky.edu



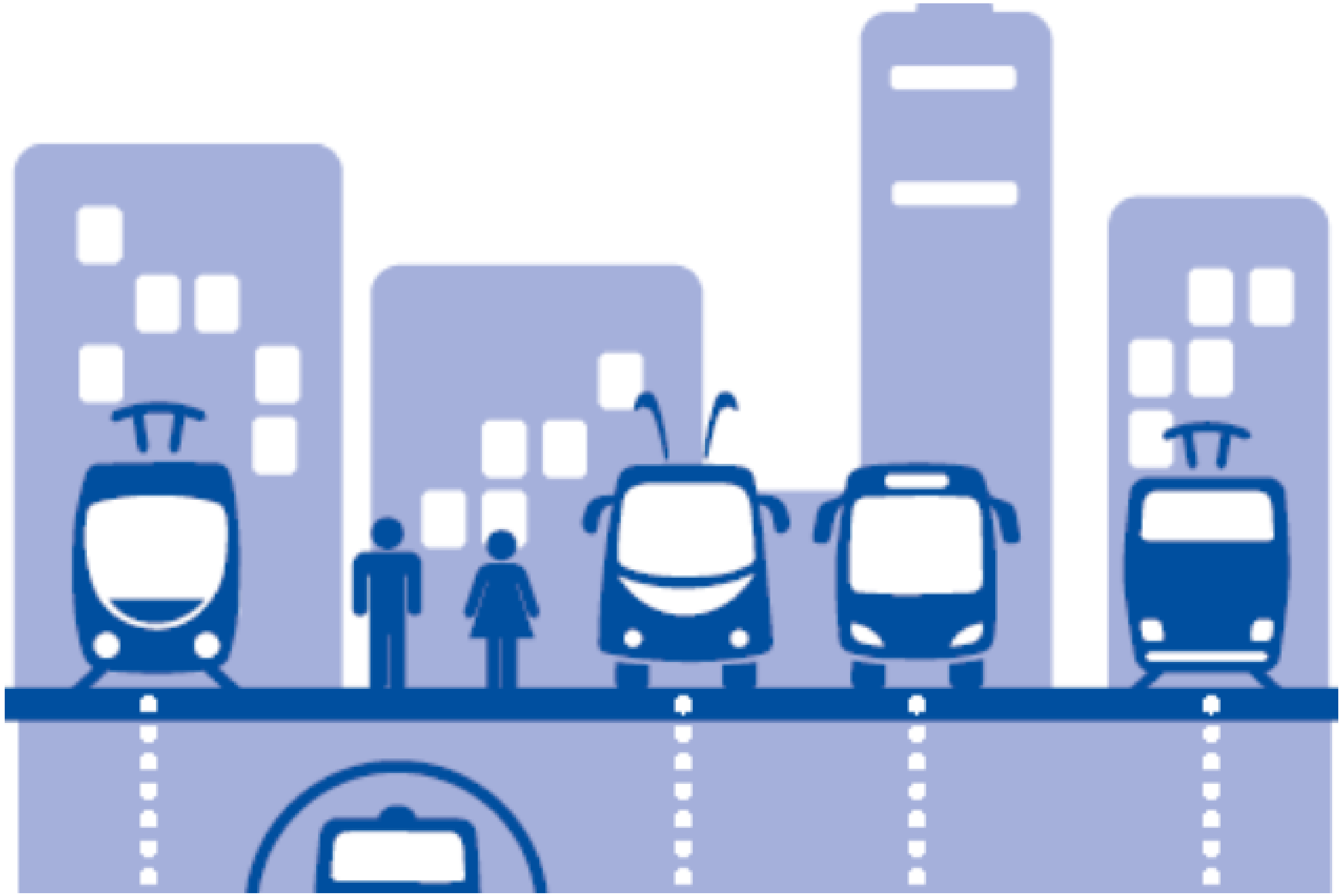
Future Ideas



Future Ideas

- Does Miami use public transit less compared to other major cities?

- Optimization Analysis looking at the 'flow' of the city, where people live compared to where they work



“One of my personal fears is the **suburbanization of poverty**. It’s hard to serve places that don’t have density. And it’s really hard to do it in a way that’s cost-effective. “ – Jacob Tzegaegbe, Transportation Advisor for Atlanta

Our* buses ineffective?

<u>commuting time has emerged as the single strongest factor in the odds of escaping poverty</u>