



# CTA Ridership:

**Forecasts & Predictions During the COVID-19 Pandemic**

Sam Waldner

---

# Problem Statement:

CTA (Chicago Transit Authority) ridership levels have dramatically changed in the last two years due in large part to the COVID-19 pandemic.

The goal of this project is to find the most suitable way to predict CTA total ridership at different datetime levels. Predicting at the daily level is more helpful for riders while a longer weekly and monthly predictions are more helpful for operational purposes. We will determine which Time Series Model performs the best via AIC and RMSE, and which Regression performs the best via  $R^2$  and RMSE.



## Background

- The Chicago Transit Authority (CTA) operates the nation's second largest public transportation system. On an average weekday, 1.6 million rides are taken on CTA. [\(source\)](#)
- Proposed 2022 operating budget of \$1.75 billion





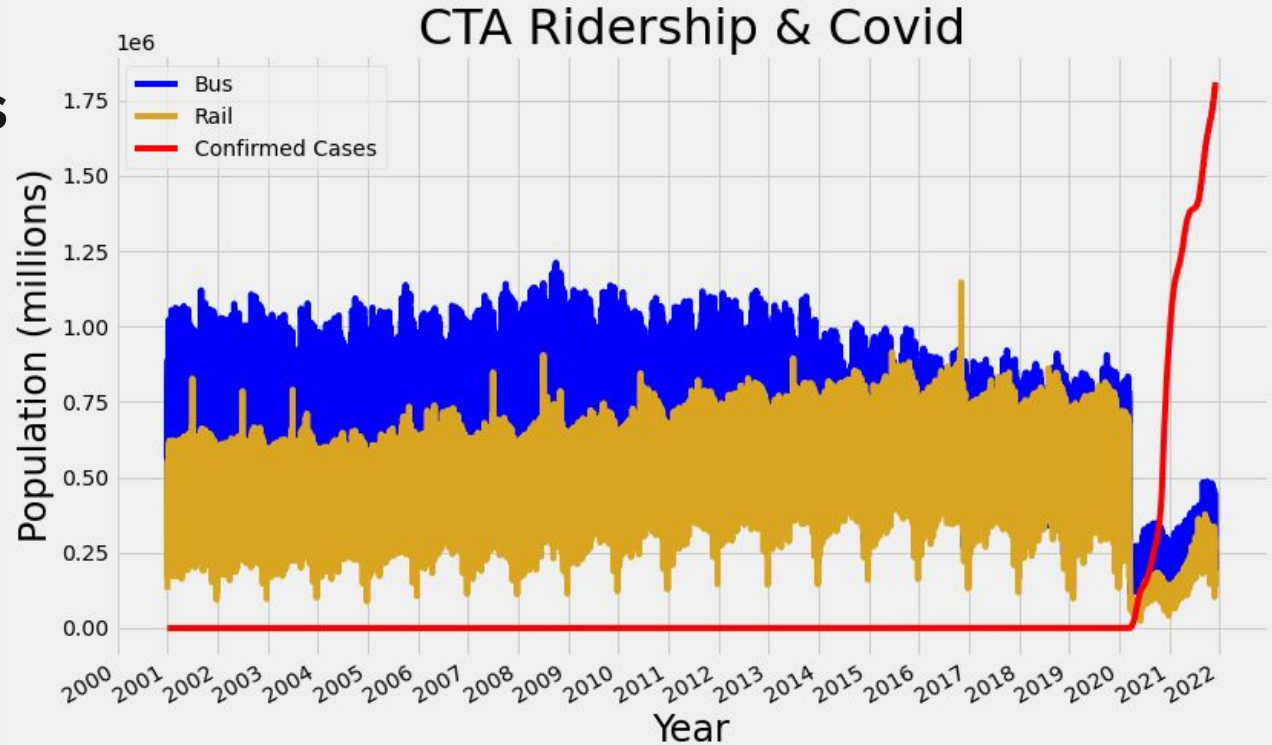
# Daily, Weekly, Monthly

- Primary data is CTA Ridership levels
  - Further data includes:
    - New Car Registrations
    - COVID Cases
    - Vaccine Rates
    - Variables representing levels of lockdown and restriction
    - Gas prices
    - Treasury Yield Curve
    - Weather

new_car_reg	0.6
20_yr	0.36
10_yr	0.34
7_yr	0.29
gas_open	0.26
gas_close	0.26
5_yr	0.25
3_yr	0.19
gas_price	0.18
2_yr	0.16
new_total_reg	0.16
1_yr	0.12
tmax	0.12
tmin	0.12

# Ridership Levels And Covid Cases

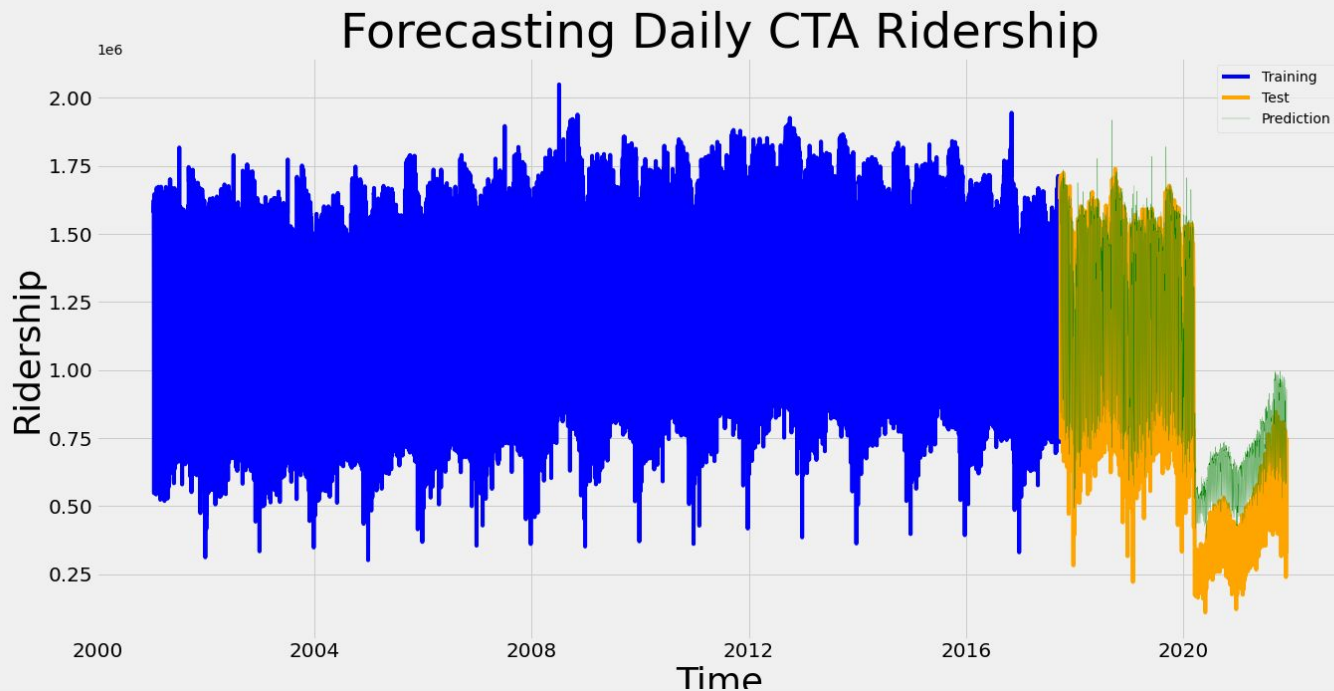
- Dense
- Inverse Relationship
- Bus vs Rail



No annual lag

Daily

- AIC:  
165,448.67
- RMSE:  
207,743.97



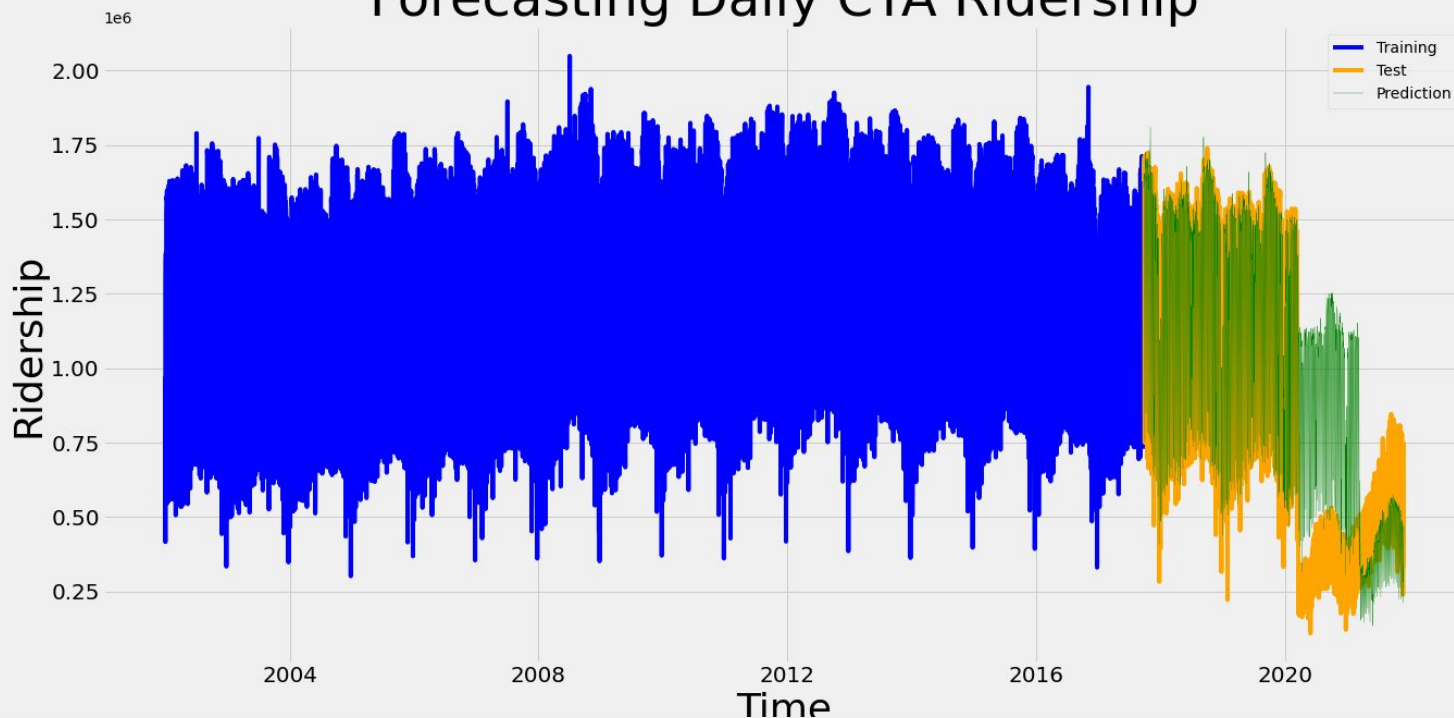
Dense, but catches the drop surprisingly well.

# Forecasting Daily CTA Ridership

Annual lag

Daily

- AIC:  
151,287.11
- RMSE:  
317,766.52

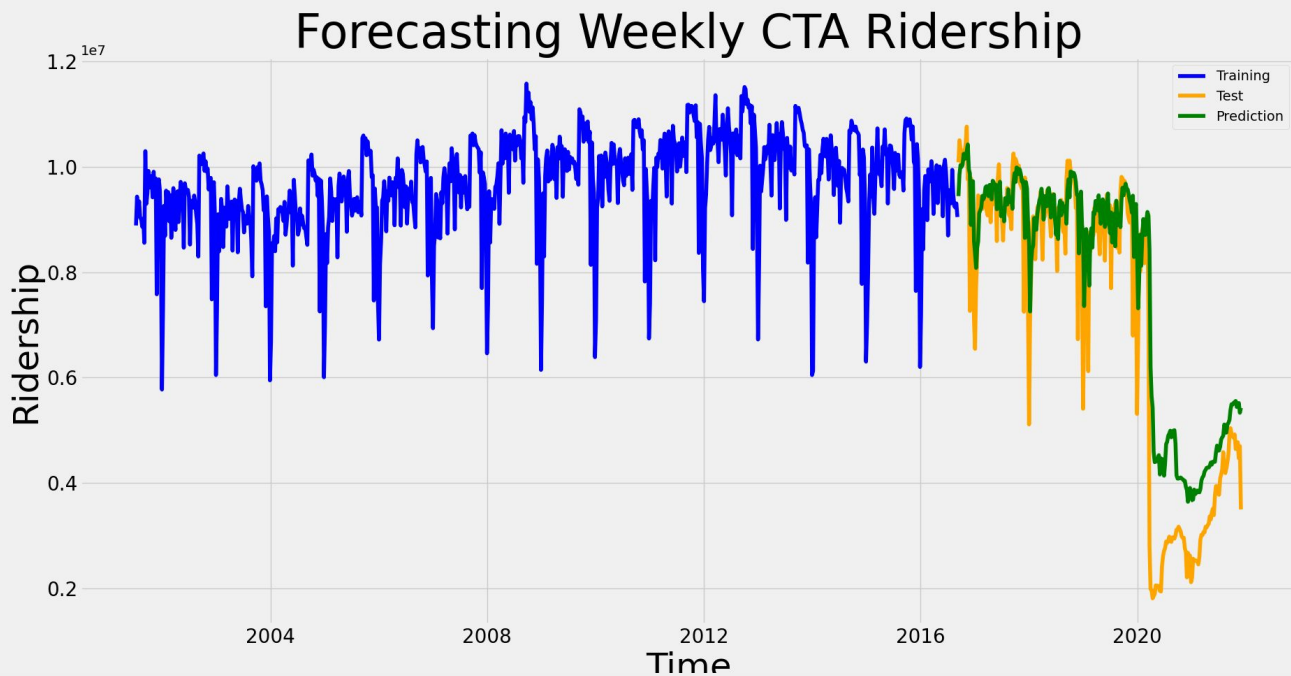


Better AIC worse RMSE, trouble adjusting

No annual lag

Weekly

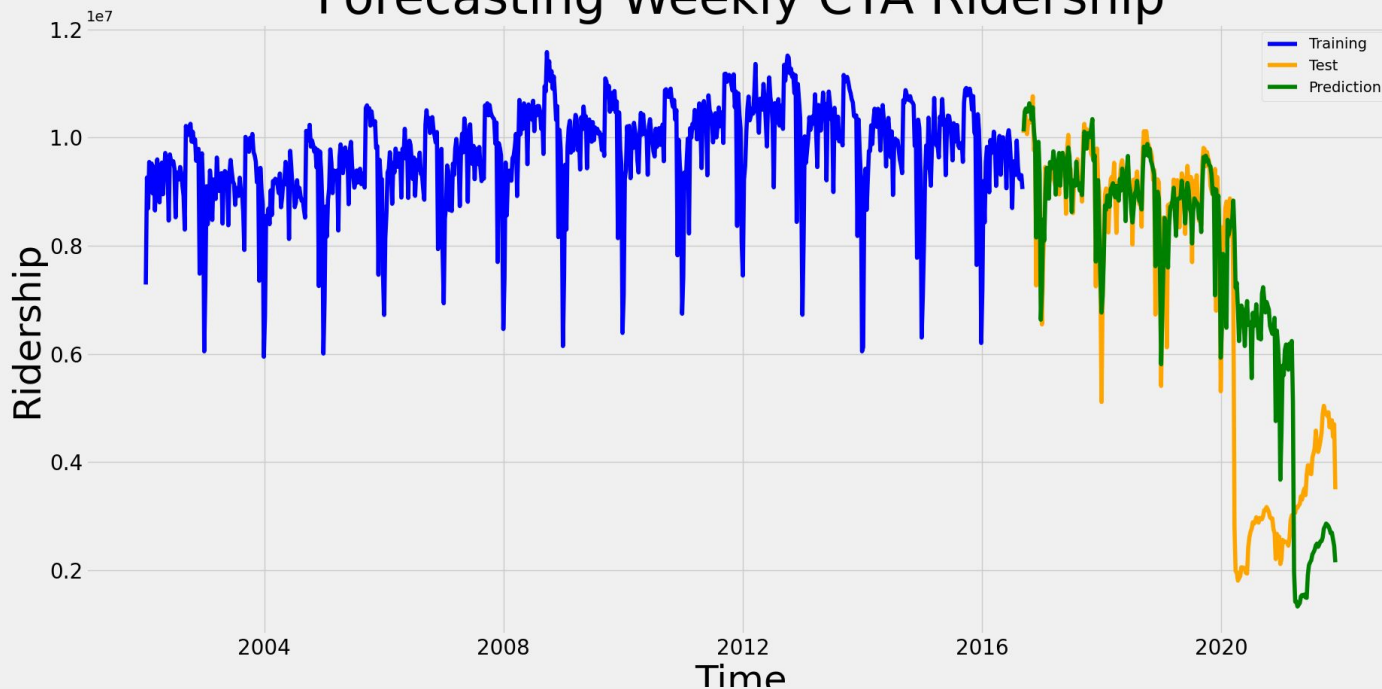
- AIC:  
23,562.97
- RMSE:  
1,141,954.78



Readable Seasonality, similar results



## Forecasting Weekly CTA Ridership



Annual lag

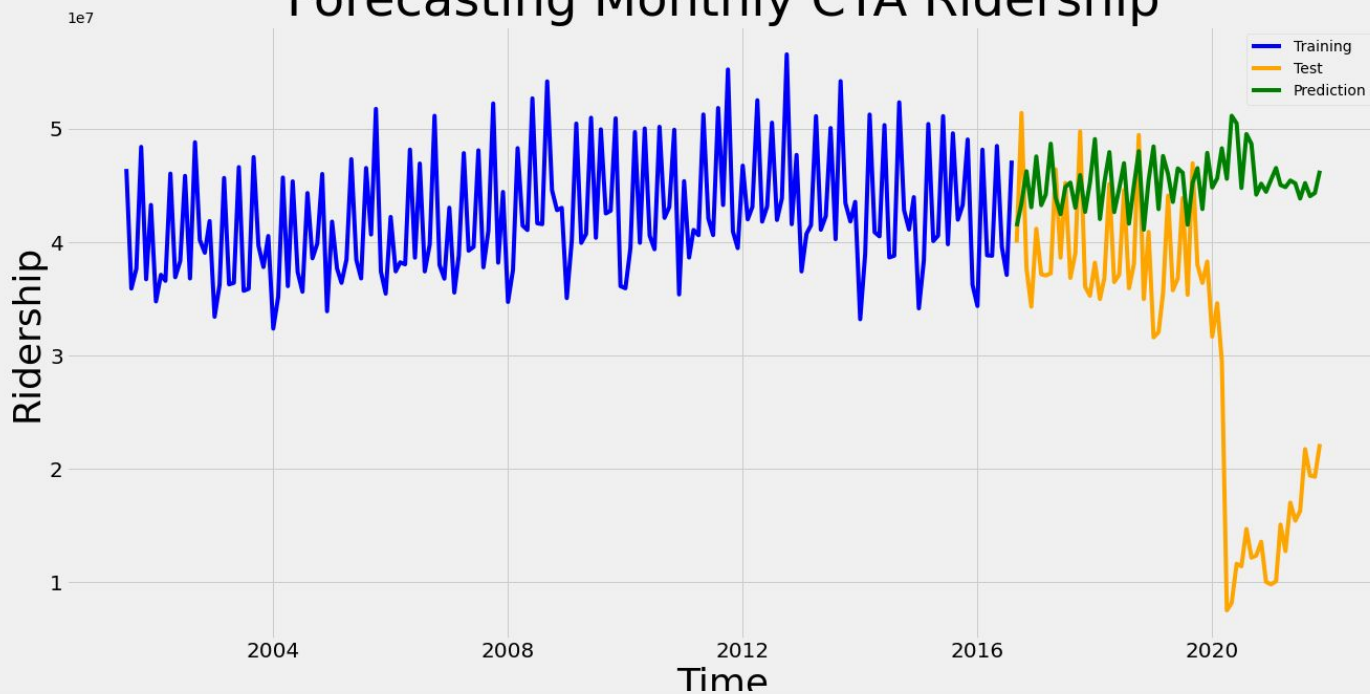
Weekly

- AIC:  
22,137.64
- RMSE:  
1,875,878.55

Better AIC worse RMSE, trouble adjusting

## Forecasting Monthly CTA Ridership

No annual lag

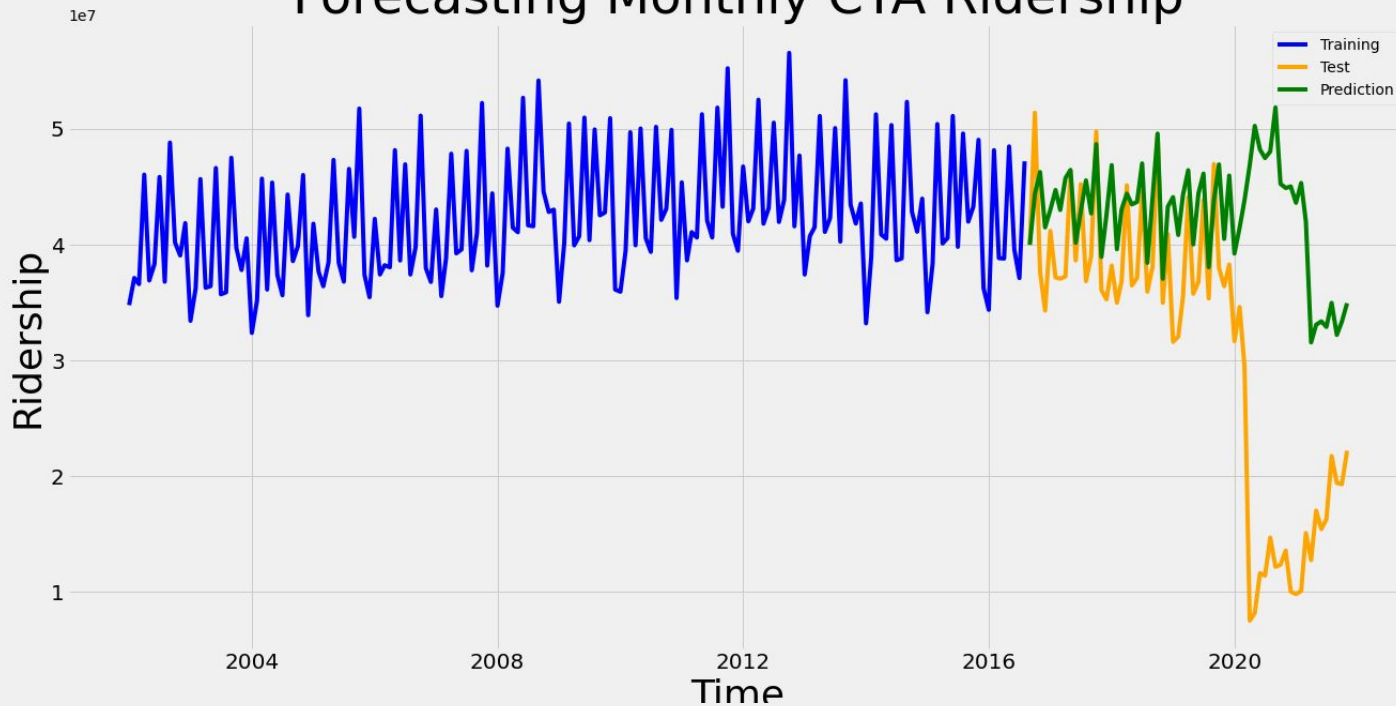


Monthly

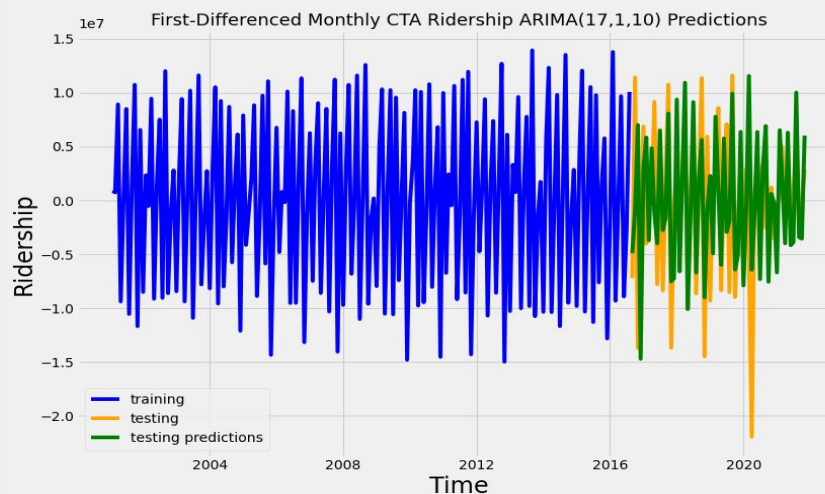
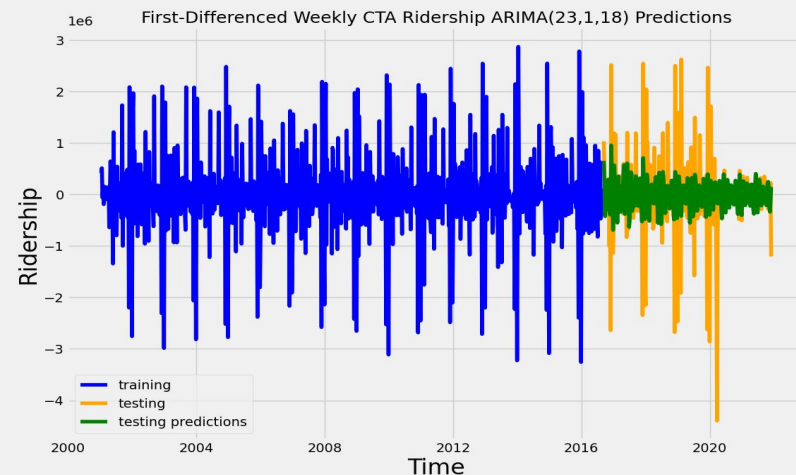
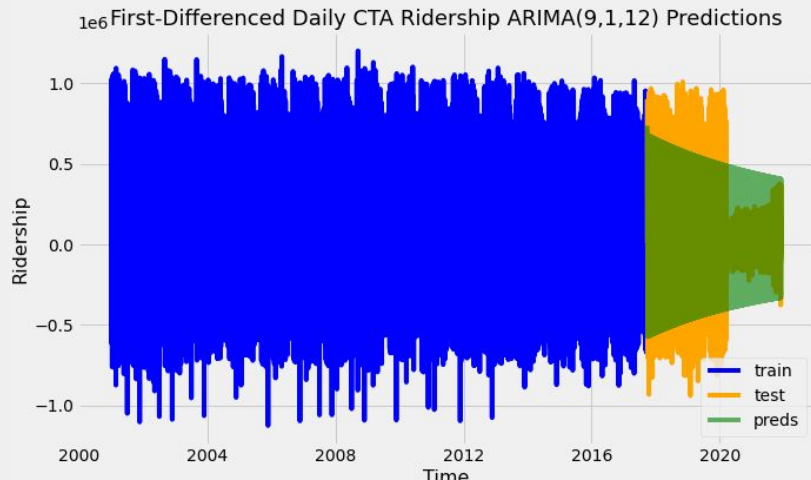
- AIC: 6,110.61
- RMSE: 19,463,350.10

- Loss of seasonality, poor performance

# Forecasting Monthly CTA Ridership



Better on both scores, shows some downward motion



- ARIMA strong bias, ignores drop

Daily:

AIC → 163,297.62

RMSE → 182,640.85

Weekly:

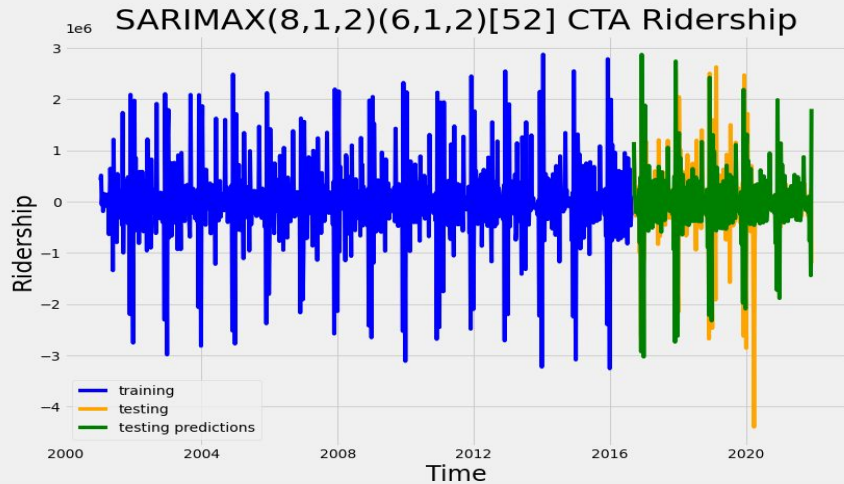
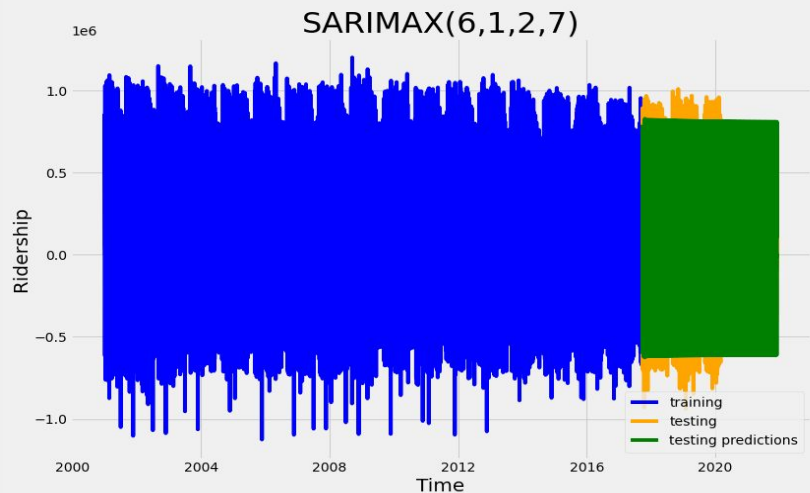
AIC → 24,116.00

RMSE → 736,057.29

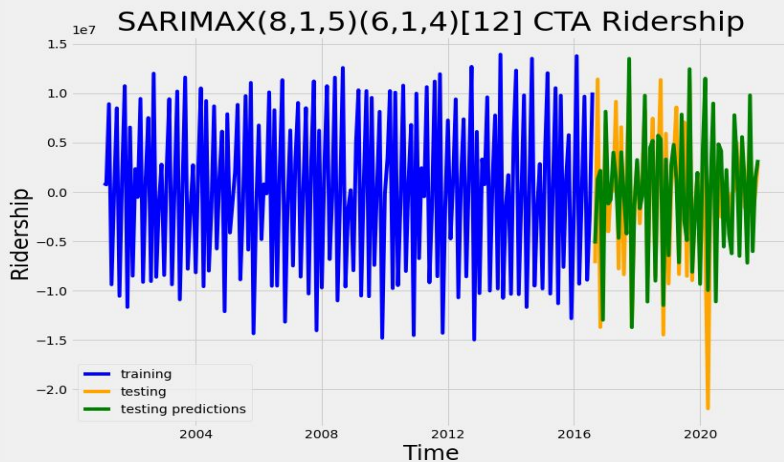
Monthly:

AIC → 6,139.96

RMSE → 6,515,672.95



- Except daily, SARIMA has best scores so far



Daily:

AIC → 164,875.68

RMSE → 729,083.44

Weekly:

AIC → 22,086.11

RMSE → 860,067.67

Monthly:

AIC → 5,781.00

RMSE → 5,461,922.37



# Vector Autoregression

Daily:

AIC/BIC

~18

RMSE

335,815.98

Weekly:

AIC/BIC

~16.5

RMSE

780,051.84

Monthly:

AIC/BIC

~44

RMSE

6,278,740.81

- Daily:
  - Total Rides
  - 20 year Yield Curve
  - Crash Damage indicator
  - Minimum temperature
- Weekly:
  - Total Rides
  - 20 year Yield Curve
  - Crash Damage indicator
  - Minimum temperature
  - Gas Prices
- Monthly:
  - Total Rides
  - 20 year Yield Curve
  - New Car Registrations
  - Minimum temperature
  - Gas Prices

## Mixed Results

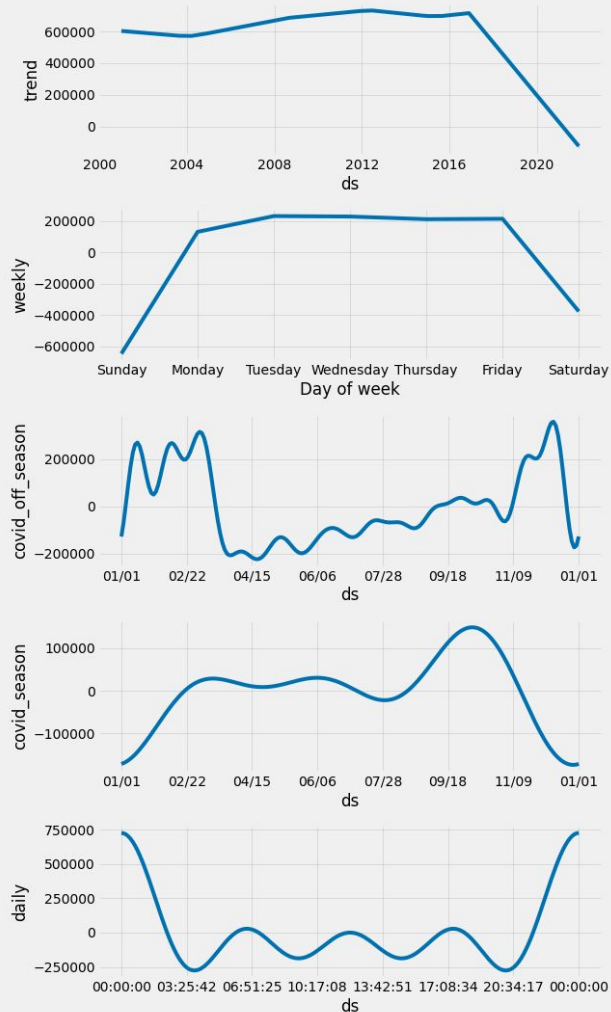
# Prophet

- Intuitive
- Flexible
  - Seasonality
  - Holidays
- Trend
- Measure

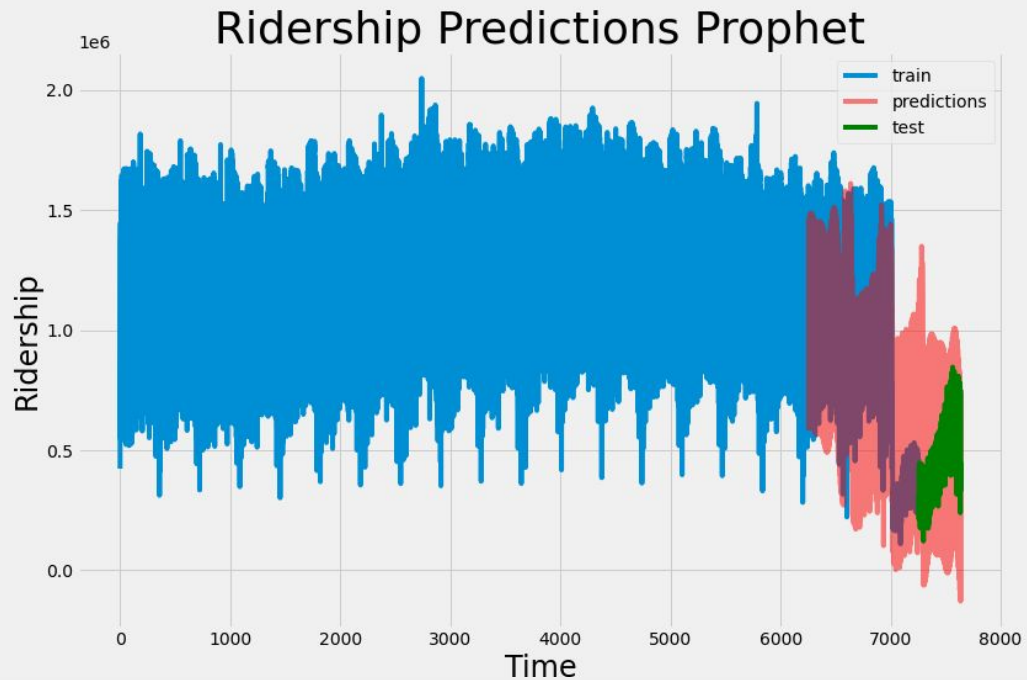
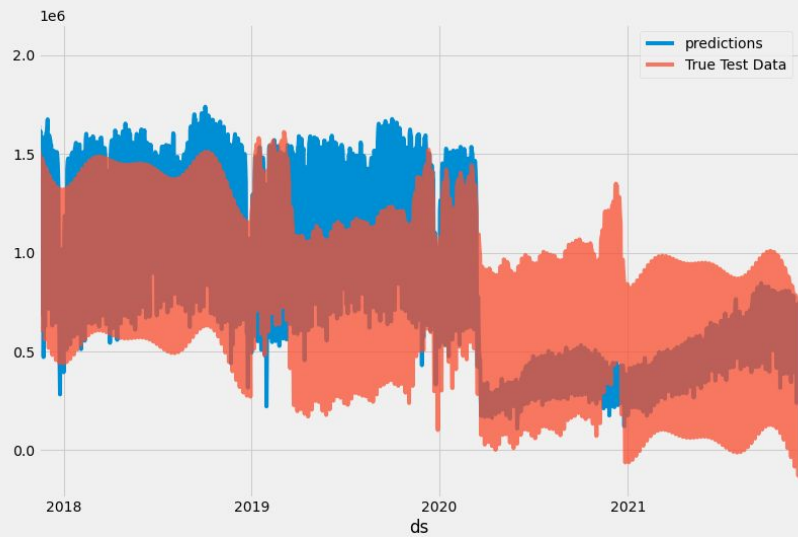
**Daily:**  
MDAPE: 0.042  
RMSE 373,906.89

**Weekly:**  
MDAPE 0.028  
RMSE 1,570,733.39

**Monthly:**  
MDAPE: 0.10  
RMSE 13,110,986.04



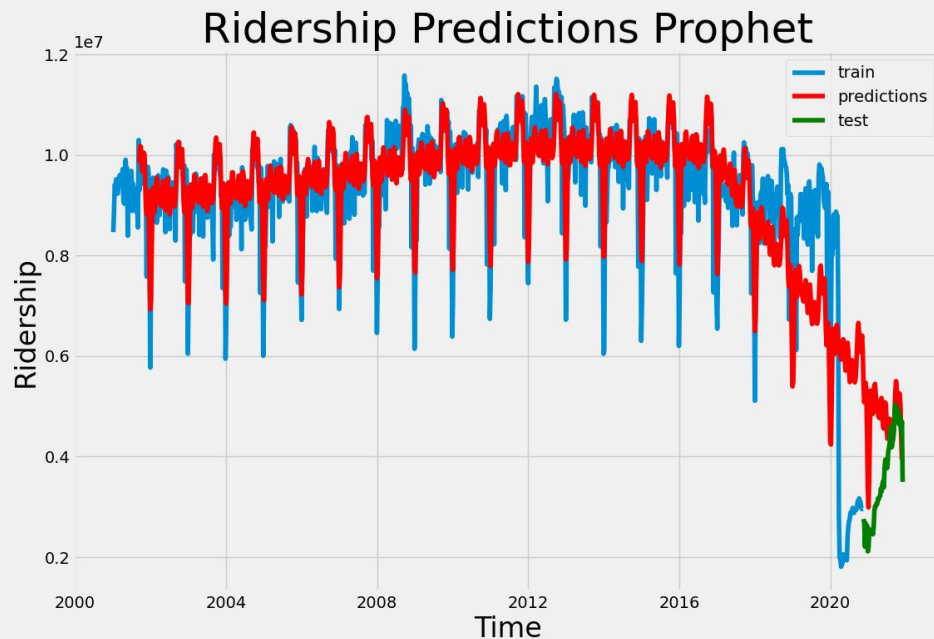
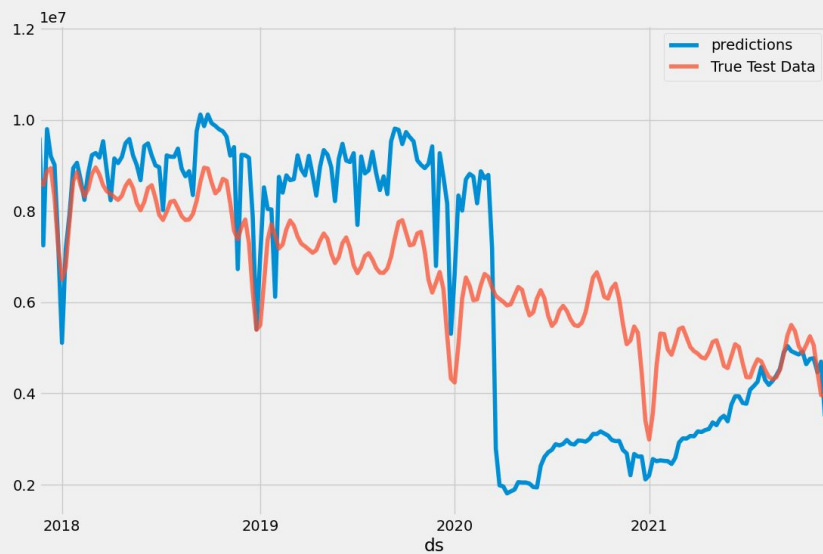
# Daily



# Prophet

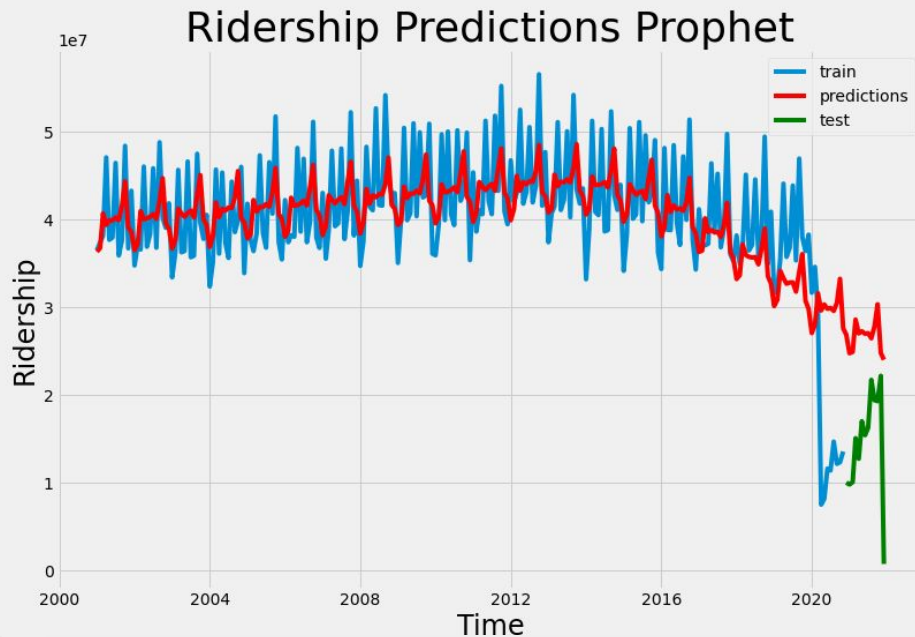
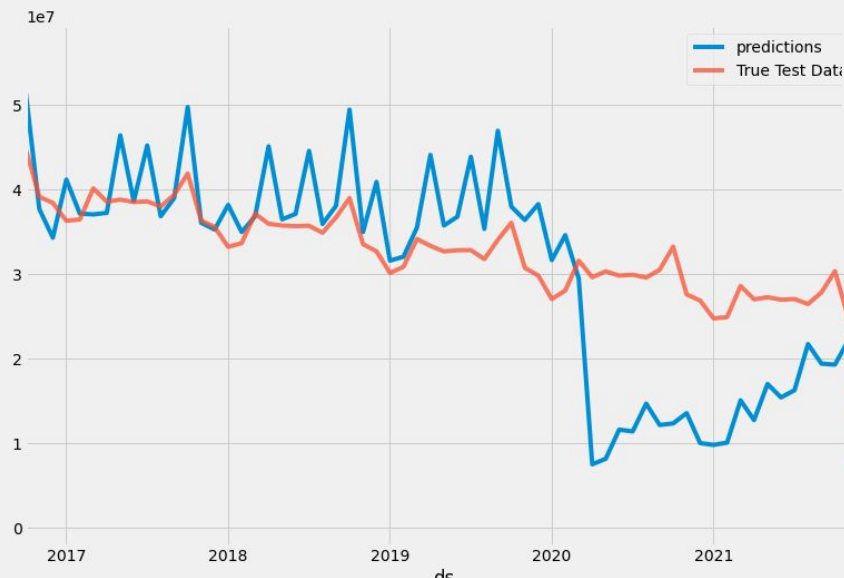


# Weekly



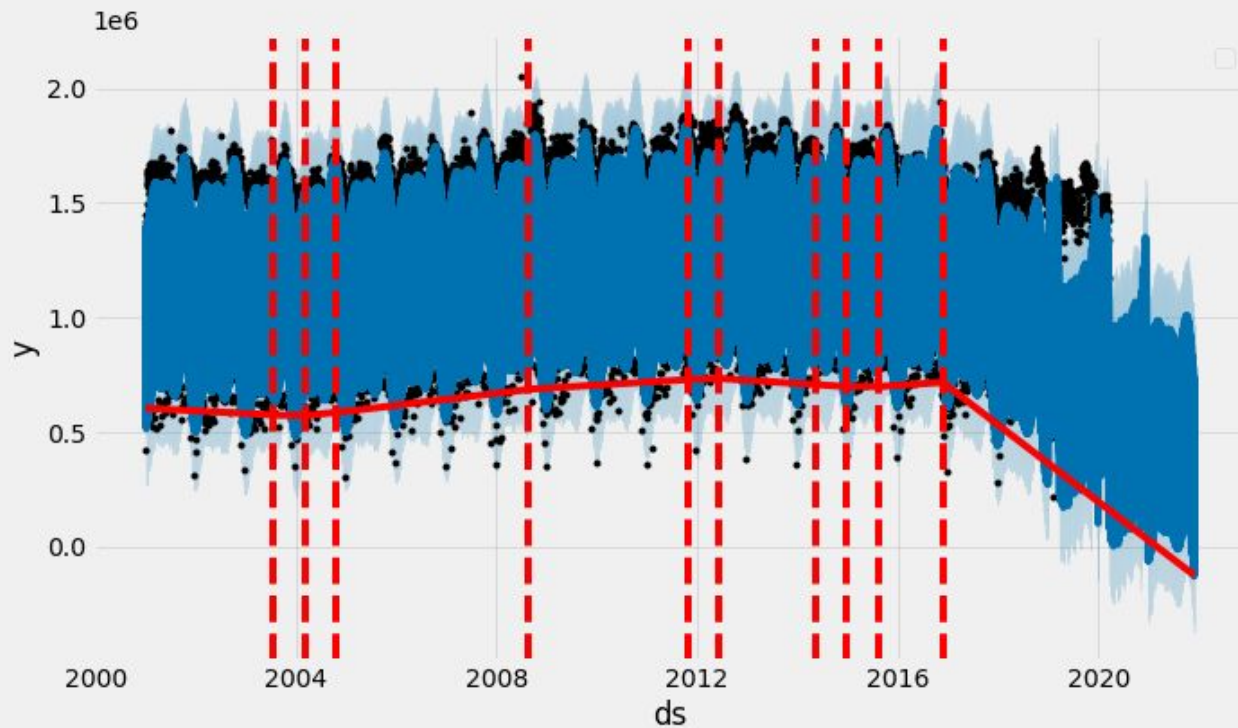
# Prophet

# Weekly



# Prophet

# Prophet's weakness



# Results

## Daily

Model	RMSE	Other
Linear w/ annual	317,766.52	AIC: 151,287.11
Linear w/o annual	207,743.97	AIC: 165,448.67
ARIMA(9,1,12)	182,640.85	AIC: 163,297.62
SARIMAX(6, 1, 2)x(5, 0, [1, 2, 3], 7)	729,083.44	AIC: 164,875.68
SARIMAX(6,1,2,7)	56,628,560,806	AIC: 164,823.30
VAR	335,815.98	AIC/BIC: 18.18
Prophet	373,906.89	MDAPE: .042
RandomForest	358,914.75	R2: 0.41, 0.35
LinearRegression	371,335.56	R2: 0.34 ,0.31
XGBoost	358,299.85	R2: 0.40, 0.36

## Weekly

Model	RMSE	Other
LinearTime w/ Annual	1,875,878.55	AIC: 22,137.64
LinearTime w/o Annual	1,141,954.78	AIC: 23,562.97
ARIMA(23,1,18)	736,057.29	AIC: 24,116.00
Auto-ARIMA(8,1,2)	780,236.69	AIC: 24,362.11
SARIMAX(8,1,2)(6,1,2)[52]	860,067.67	AIC: 22,086.11
VAR	780,051.84	AIC/BIC: ~16.5
Prophet	1,570,733.39	MDAPE: 0.028
LinearRegression	752,057.08	R2: 0.86, 0.84
RandomForest	674,141.51	R2: 0.96, 0.87
XGBoostRegressor	666,553.70	R2: 0.98, 0.87

## Monthly

Model	RMSE	Other
LinearTime w/ Annual	17,196,311.05	AIC: 5,879.04
LinearTime w/o Annual	19,463,350.10	AIC: 6,110.61
ARIMA(17,1,10)	6,515,672.95	AIC: 6,139.96
Auto-ARIMA(3,1,5)	6,155,633.02	AIC: 6,249.84
SARIMAX(8,1,5)(6,1,4)[12]	5,461,922.37	AIC: 5,781.00
VAR	6,278,740.81	AIC/BIC: ~44
Prophet	13,110,986.04	MAPE: 0.10
LinearRegression	5,239,220.40	R2:0.85, 0.68
RandomForest	5,612,426.81	R2: 0.89, 0.63
XGBoostRegressor(tree)	5,570,271.10	R2: 0.87, 0.64
XGBoostRegressor(linear)	5,574,917.50	R2: 0.83, 0.63

# Conclusion & Next Steps

- Weekly level best
  - Data lost in resample
  - SARIMAX
  - VAR
- XGBoost
  - Data available
- Prophet
  - Not the best scores
  - High potential
- CTA Data
- App integration or connected to Google