

Cook County Sales Tax Revenue Forecast

Time Series Analysis and Forecast Final Project - May 27, 2023

Team 9

Annabella Liang

Juan Ramirez

Ujjwal Sehrawat

Overview

- **Business case and problem statement**
- **Modeling hypothesis and assumptions**
- **Data description and properties**
- **Data processing and EDA**
- **Feature Engineering**
- **Proposed modeling approaches**
- **Selected model results with justifications and tradeoffs**
- **Insights/Recommendations & Future work**

Business case and problem statement

Problem Statement:

- Forecast Sales Tax Revenues for cook county over a 3-year horizon.

Business Case:

- Stakeholders: Cook County board members, policy makers & local residents.
- Sales tax Forecasting gives stakeholders empirical basis for data-driven decision-making about:
 - future borrowing,
 - use of accumulated reserves,
 - specification of budget balancing measures.
- 3 -year planning especially important as expenditure typically lags by 3 years.

Modeling hypothesis and assumptions

Hypothesis:

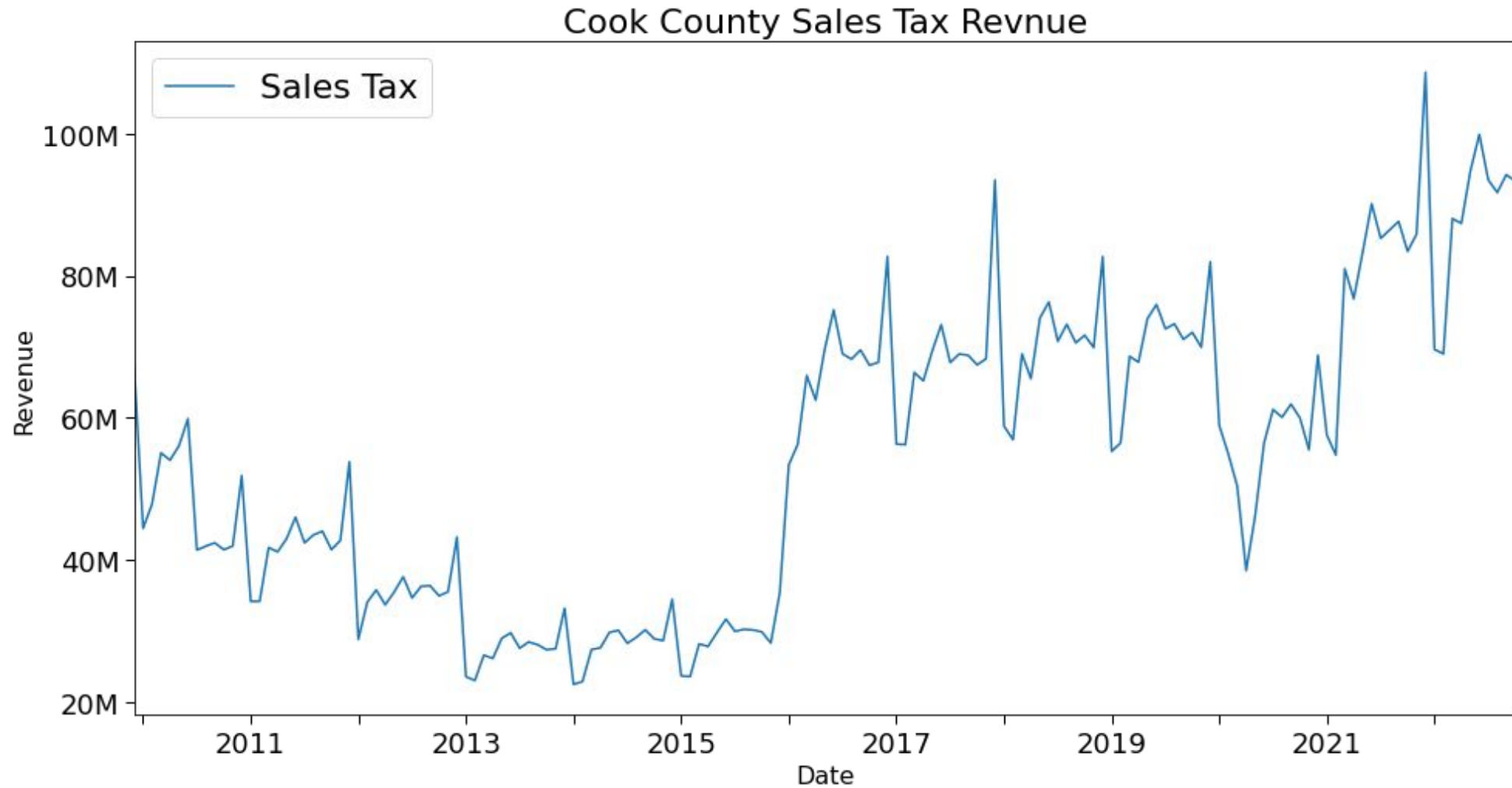
- We believe that sales tax revenue data from the past will be predictive of sales tax revenue data in the future.
- Exogenous variables (regional statistics) related to sales tax revenue will impact this aforementioned relationship & can help improve forecasting efficiency.

Assumptions:

- Publicly available Cook County's historical sales tax revenue data since its fiscal year 2010.
- The sales tax of the County has changed over the years, but currently stands at around 1.75%.
- Time series is stationary, i.e., statistical properties of time series does not change over time.
- Normal residuals.

Exploratory Data Analysis

Cook County - Sales Tax Revenue



Cook County - Changing Sales Tax Rates

Annual Pledged Sales Tax Revenues

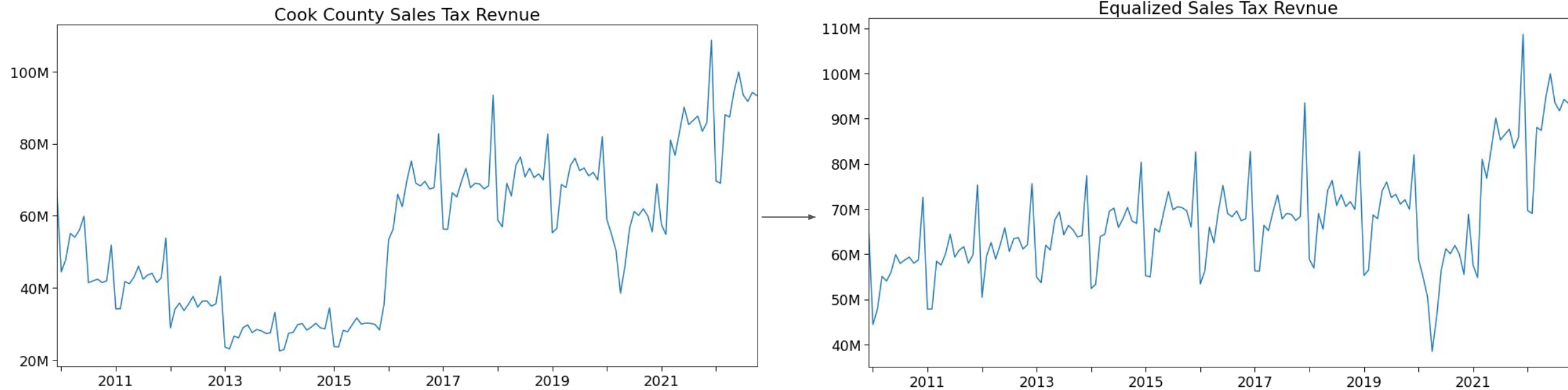
Fiscal Years 2012 to 2021

Fiscal Year Ended 11/30	Home Rule Sales Tax Rate	Effective Date	Home Rule Sales Tax Revenues	Percent Change Over Prior Year	Pro Forma Debt Service Coverage Ratio(1)
2021	1.75%		\$861.6	19.40%	10.2x
2020	1.75%		\$721.6	-13.96%	9.3x
2019	1.75%		838.7	-0.46%	10.8x
2018	1.75%		842.6	3.90%	10.8x
2017	1.75%		811.0	25.96%	14.0x
2016	1.75%	1/1/2016	643.8	85.65%	99.6x
2015	0.75%		346.8	3.99%	46x
2014	0.75%		333.5	-8.33%	44.5x
2013	0.75%	1/1/2013	363.8	-20.60%	47.9x
2012	1.00%	1/1/2012	458.2	-9.02%	60.8x

Source: Cook County Comptroller's Office

- Sales Tax Rates have changed several times over the span of the data
- Changing Tax Rates account for the jumps in the data
- Our next step is to scale the data

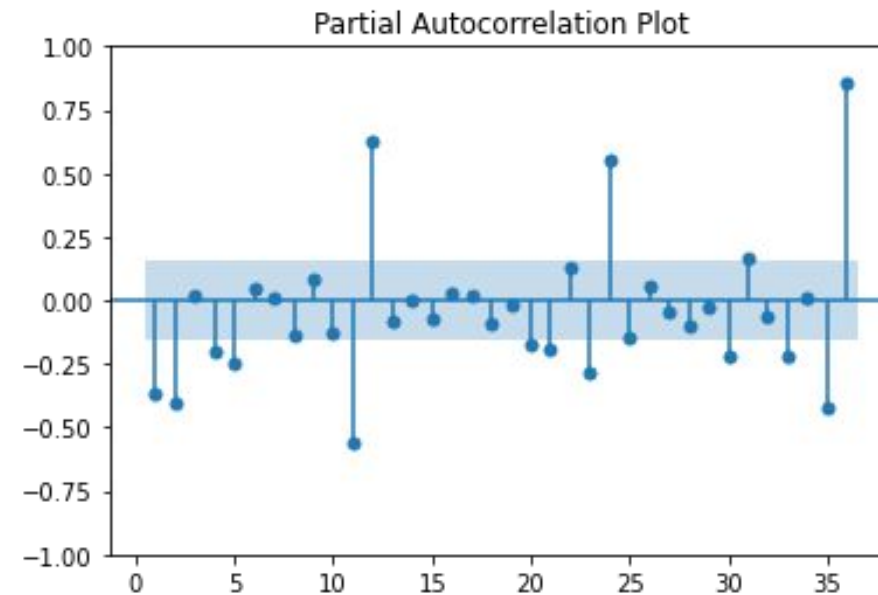
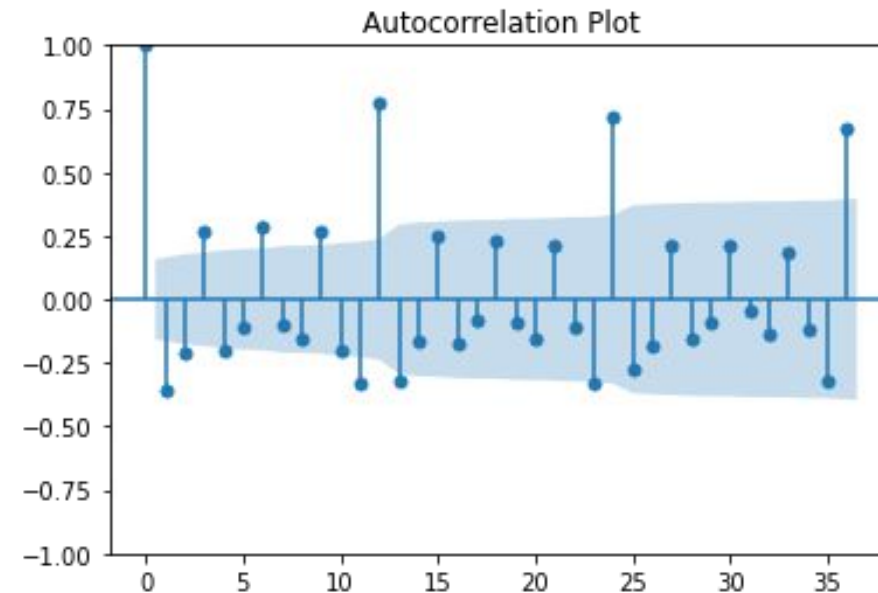
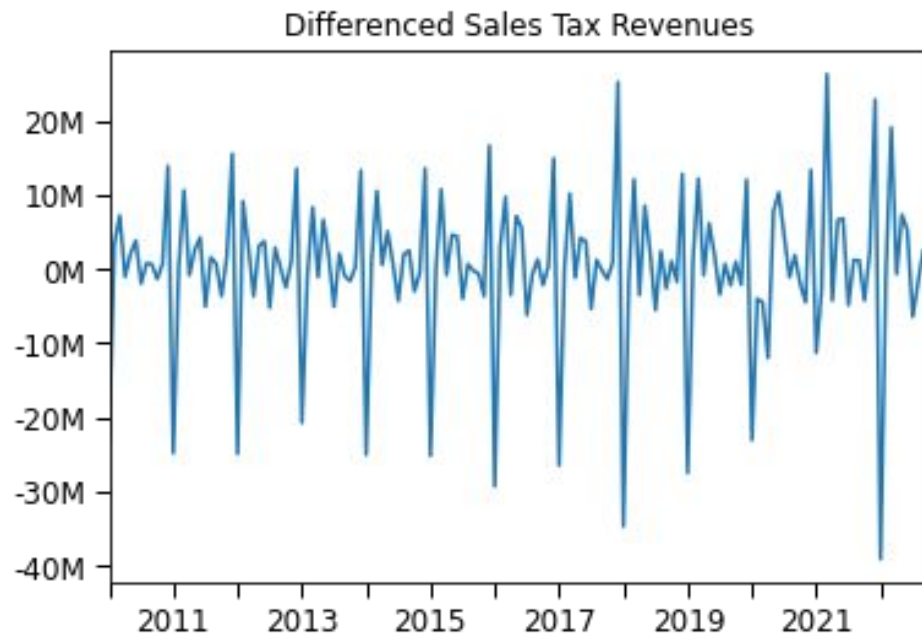
Cook County - Equalized Sales Tax Rates



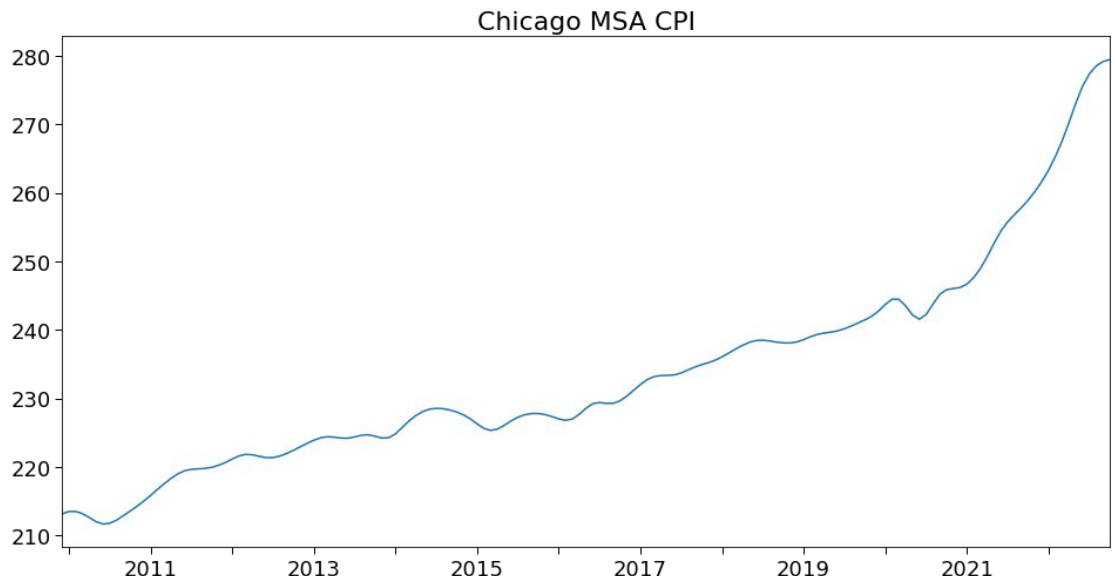
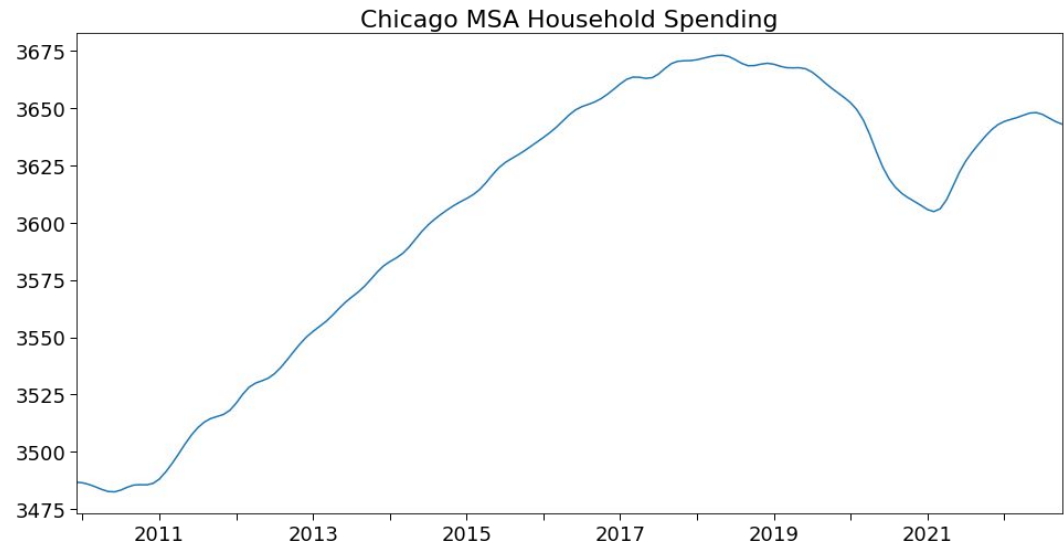
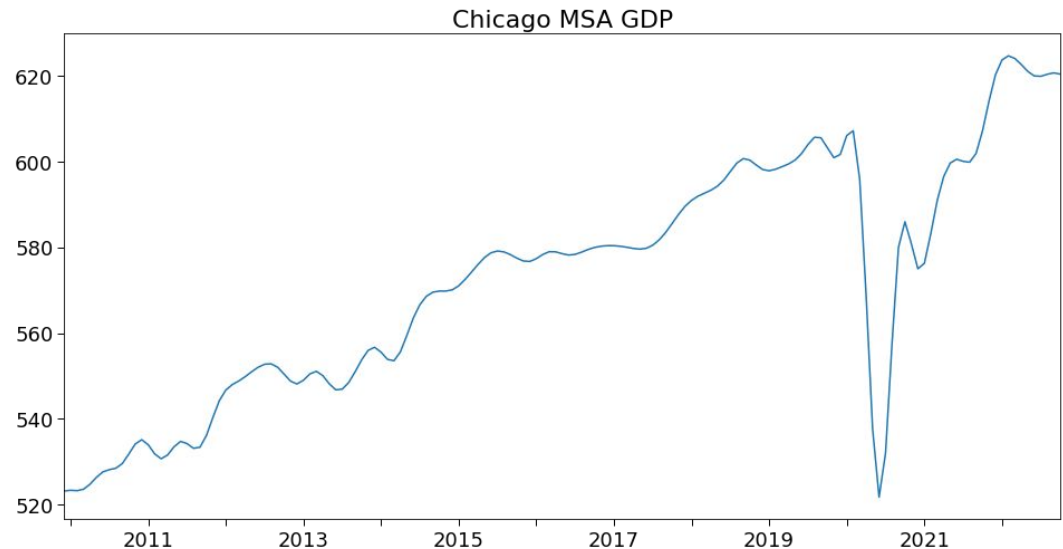
- The equalized Cook County Sales tax rate data was scaled so that all the tax rates looked the same
- We're trying to forecast using the patterns of the data, not possible tax rate changes which are legislatively decided
- The trend and seasonality in the data are much more clear
- We can clearly see the effects of the COVID 19 Pandemic

Stationarity, ACF and PACF

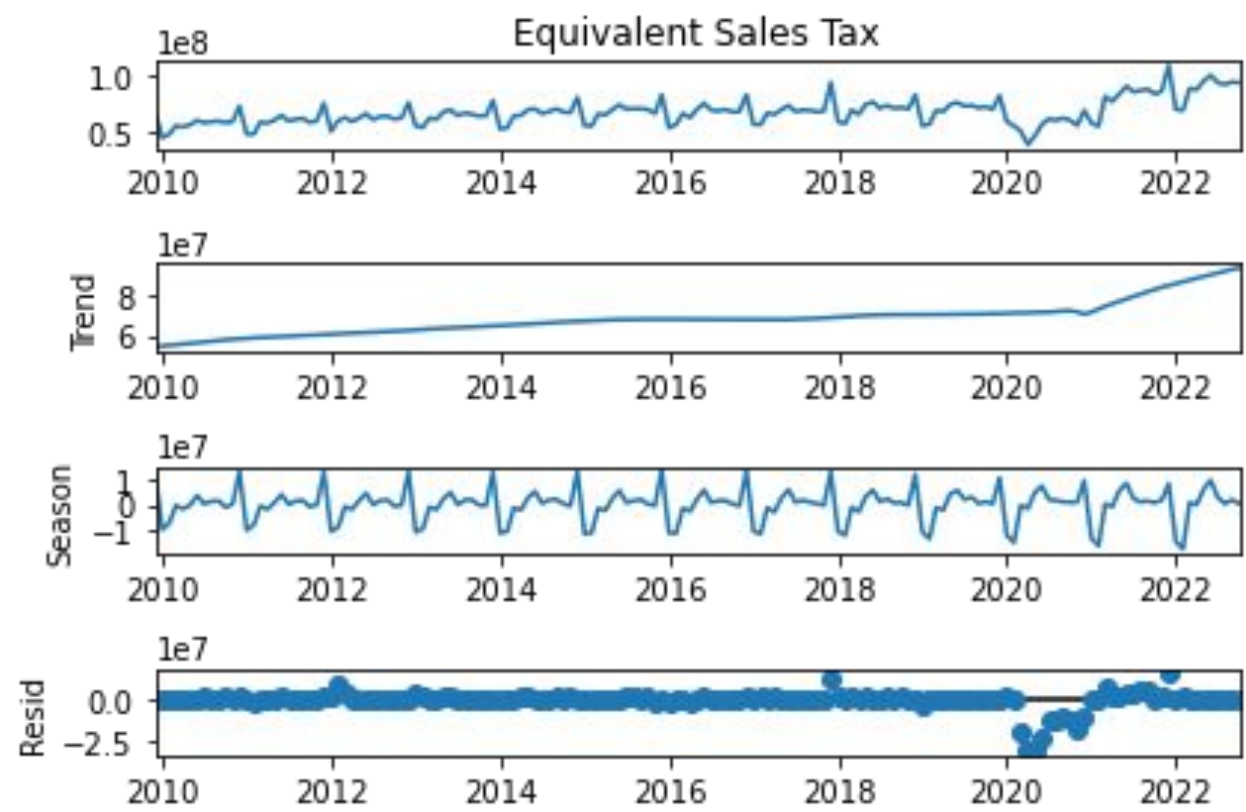
- Differenced sales tax revenue:
- ADF & KPSS test results: stationary time series



Chicago Metropolitan Area Statistics



Other EDA Features



Models Selection and Evaluation

Auto SARIMAX

Performing stepwise search to minimize aic

```
ARIMA(2,1,2)(1,1,1)[12]      : AIC=inf, Time=0.76 sec
ARIMA(0,1,0)(0,1,0)[12]      : AIC=3380.865, Time=0.02 sec
ARIMA(1,1,0)(1,1,0)[12]      : AIC=3375.851, Time=0.09 sec
ARIMA(0,1,1)(0,1,1)[12]      : AIC=3375.430, Time=0.16 sec
ARIMA(0,1,1)(0,1,0)[12]      : AIC=3382.207, Time=0.05 sec
ARIMA(0,1,1)(1,1,1)[12]      : AIC=3373.686, Time=0.28 sec
ARIMA(0,1,1)(1,1,0)[12]      : AIC=3376.622, Time=0.10 sec
ARIMA(0,1,1)(2,1,1)[12]      : AIC=3375.628, Time=0.66 sec
ARIMA(0,1,1)(1,1,2)[12]      : AIC=inf, Time=1.59 sec
ARIMA(0,1,1)(0,1,2)[12]      : AIC=3375.234, Time=0.34 sec
ARIMA(0,1,1)(2,1,0)[12]      : AIC=3377.582, Time=0.27 sec
ARIMA(0,1,1)(2,1,2)[12]      : AIC=inf, Time=1.58 sec
ARIMA(0,1,0)(1,1,1)[12]      : AIC=3376.772, Time=0.22 sec
ARIMA(1,1,1)(1,1,1)[12]      : AIC=3375.311, Time=0.41 sec
ARIMA(0,1,2)(1,1,1)[12]      : AIC=3376.106, Time=0.35 sec
ARIMA(1,1,0)(1,1,1)[12]      : AIC=3373.302, Time=0.25 sec
ARIMA(1,1,0)(0,1,1)[12]      : AIC=3374.742, Time=0.15 sec
ARIMA(1,1,0)(2,1,1)[12]      : AIC=3375.248, Time=0.81 sec
ARIMA(1,1,0)(1,1,2)[12]      : AIC=inf, Time=1.62 sec
ARIMA(1,1,0)(0,1,0)[12]      : AIC=3380.797, Time=0.18 sec
ARIMA(1,1,0)(0,1,2)[12]      : AIC=3374.718, Time=0.39 sec
ARIMA(1,1,0)(2,1,0)[12]      : AIC=3376.879, Time=0.39 sec
ARIMA(1,1,0)(2,1,2)[12]      : AIC=inf, Time=1.51 sec
ARIMA(2,1,0)(1,1,1)[12]      : AIC=3375.711, Time=0.32 sec
ARIMA(2,1,1)(1,1,1)[12]      : AIC=3377.176, Time=0.44 sec
ARIMA(1,1,0)(1,1,1)[12] intercept : AIC=3374.976, Time=0.29 sec
```

Best model: ARIMA(1,1,0)(1,1,1)[12]

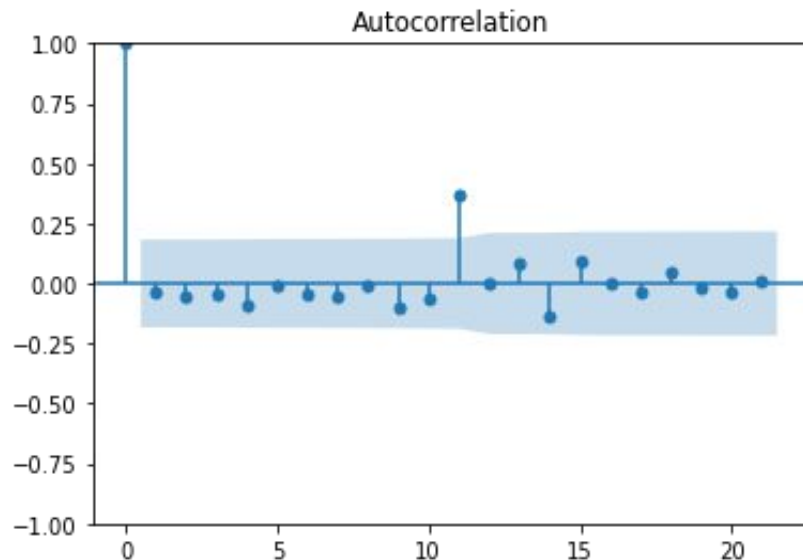
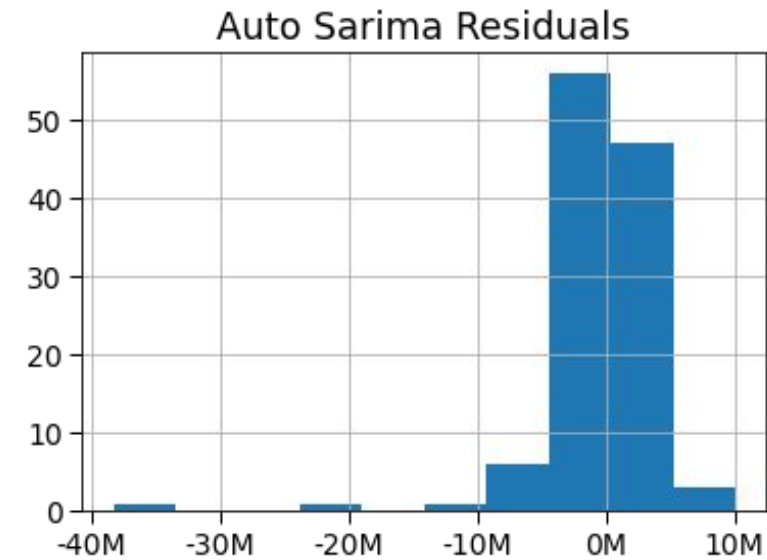
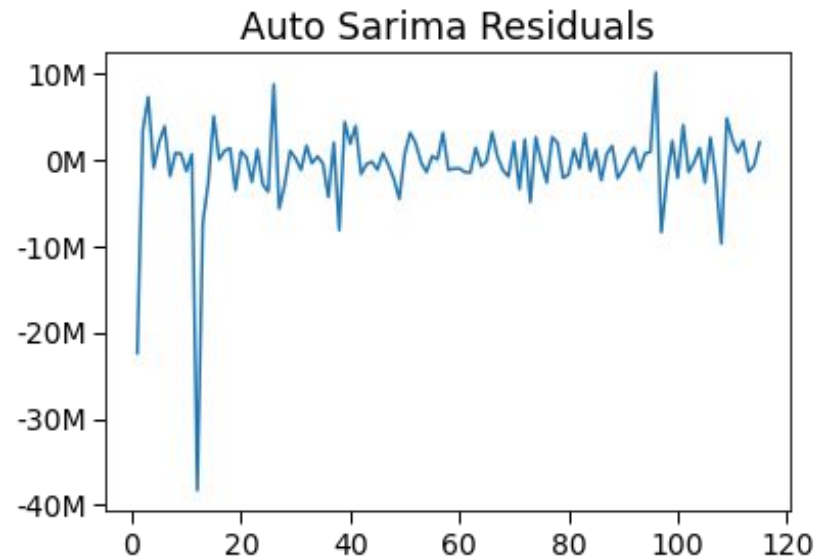
Total fit time: 13.286 seconds

SARIMAX Results

Dep. Variable:	y	No. Observations:	116			
Model:	SARIMAX(1, 1, 0)x(1, 1, [1], 12)	Log Likelihood	-1682.651			
Date:	Tue, 23 May 2023	AIC	3373.302			
Time:	21:52:33	BIC	3383.841			
Sample:	0	HQIC	3377.571			
	- 116					
Covariance Type:	opg					
	coef	std err	z	P> z	[0.025	0.975]
ar.L1	-0.0070	0.036	-0.194	0.846	-0.078	0.064
ar.S.L12	0.3455	0.135	2.557	0.011	0.081	0.610
ma.S.L12	-0.6193	0.136	-4.555	0.000	-0.886	-0.353
sigma2	8.915e+12	5.66e-15	1.58e+27	0.000	8.91e+12	8.91e+12
Ljung-Box (L1) (Q):	9.41	Jarque-Bera (JB):	19.51			
Prob(Q):	0.00	Prob(JB):	0.00			
Heteroskedasticity (H):	1.03	Skew:	-0.11			
Prob(H) (two-sided):	0.94	Kurtosis:	5.12			

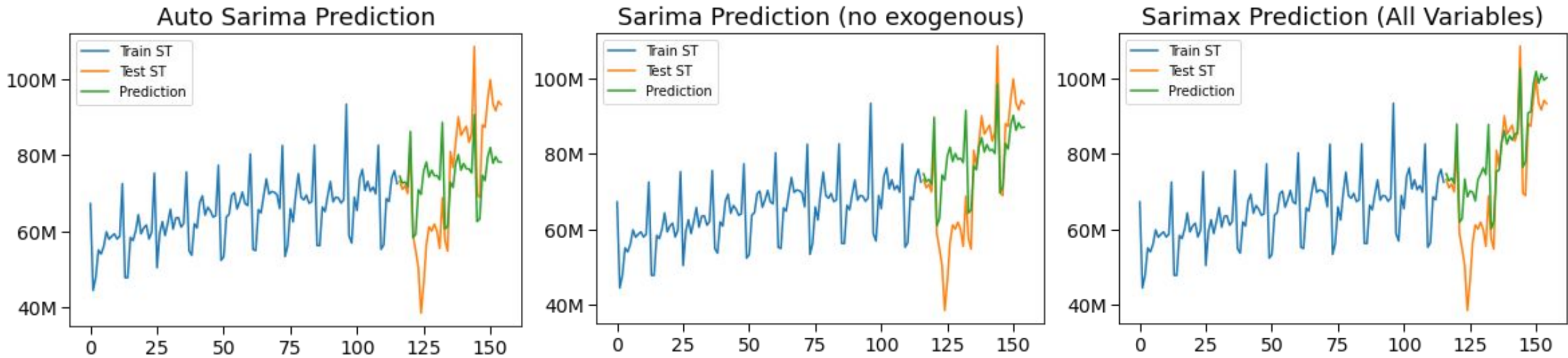


Auto SARIMA Residual Analysis and Error Metrics



```
lb_stat      21.143228  
lb_pvalue     0.048322  
Name: 12, dtype: float64
```


SARIMA Model Predictions

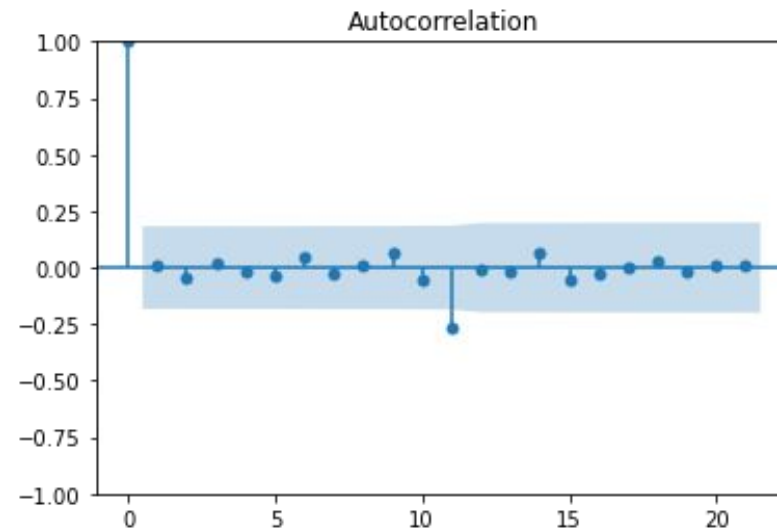
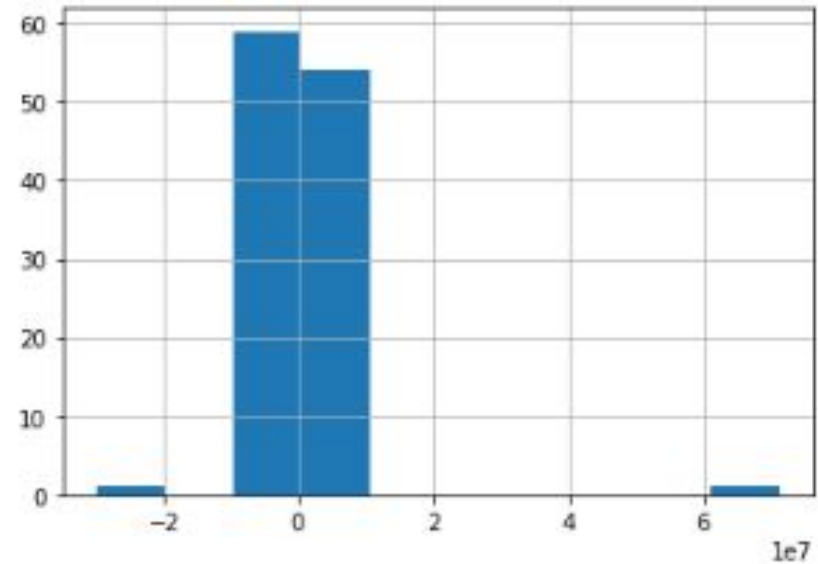
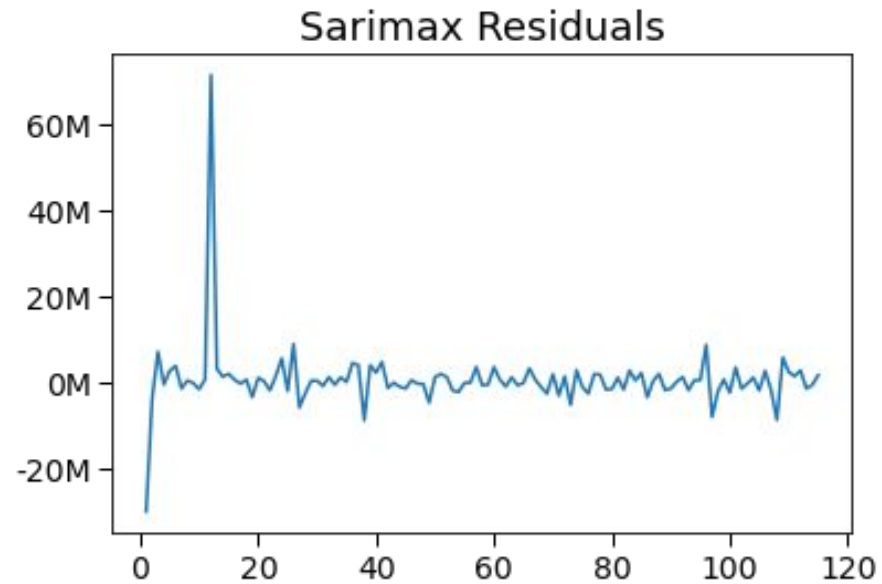


Errors	
MAE	11,469,496.42
MAPE	0.17
sMAPE	15.82
MASE	1.85

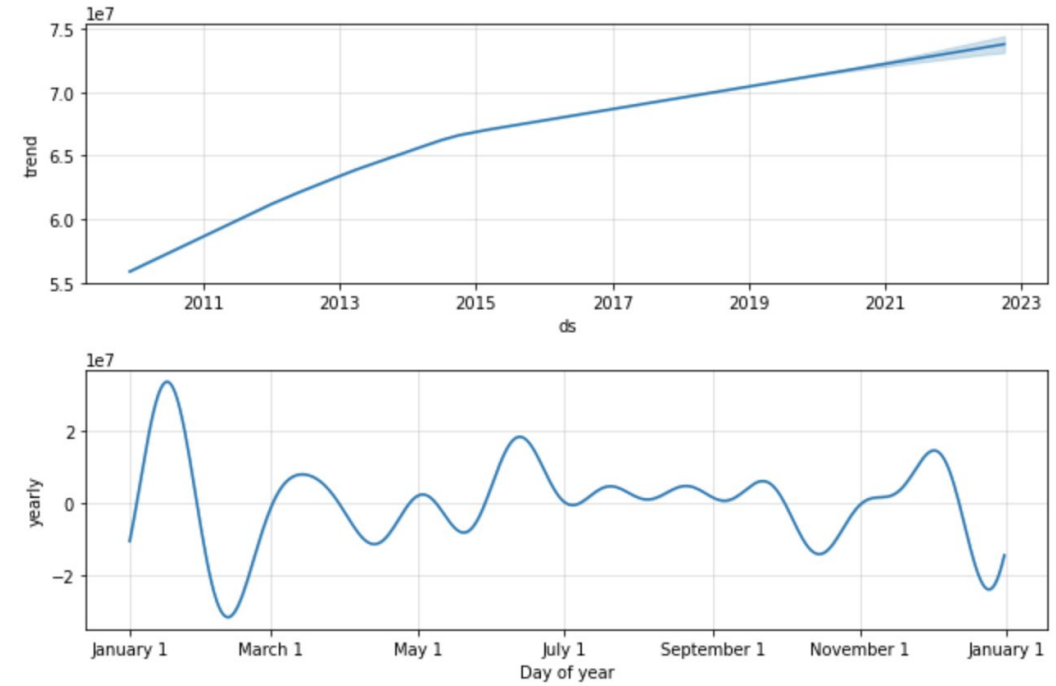
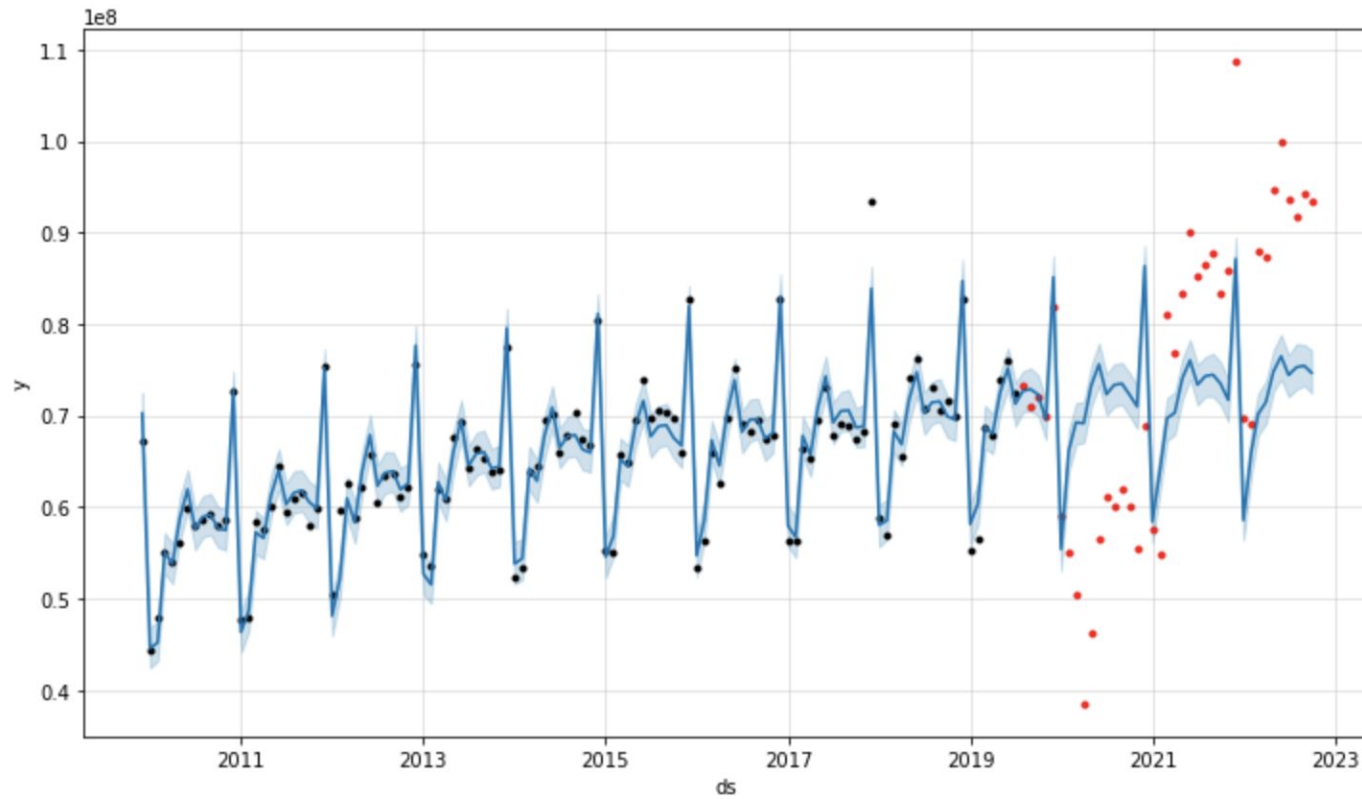
Errors	
MAE	10,758,156.75
MAPE	0.17
sMAPE	14.95
MASE	1.74

Errors	
MAE	7,541,089.67
MAPE	0.13
sMAPE	10.99
MASE	1.22

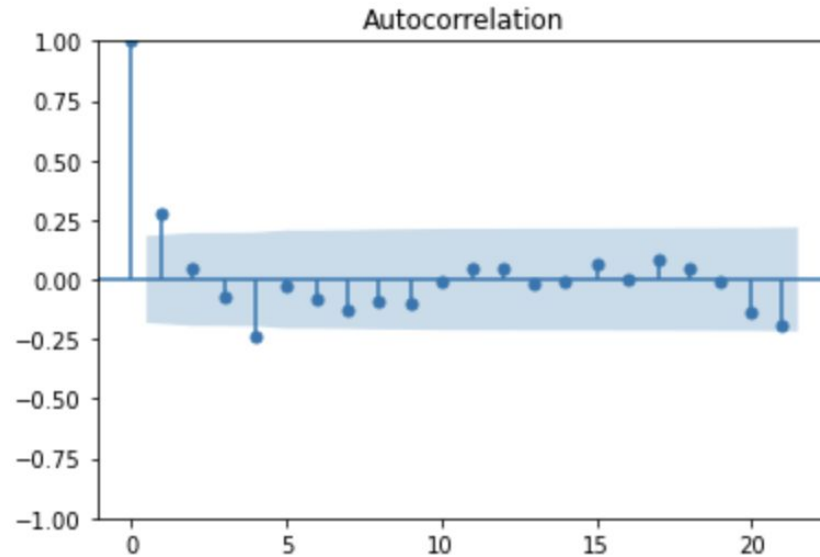
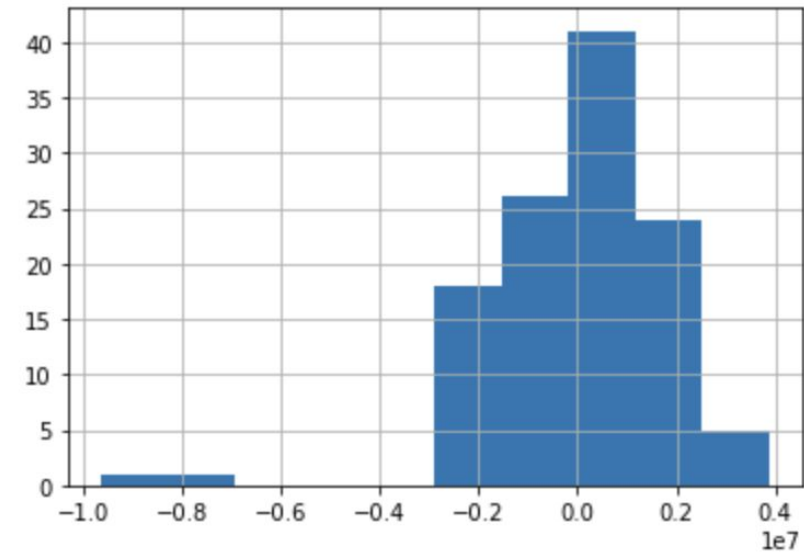
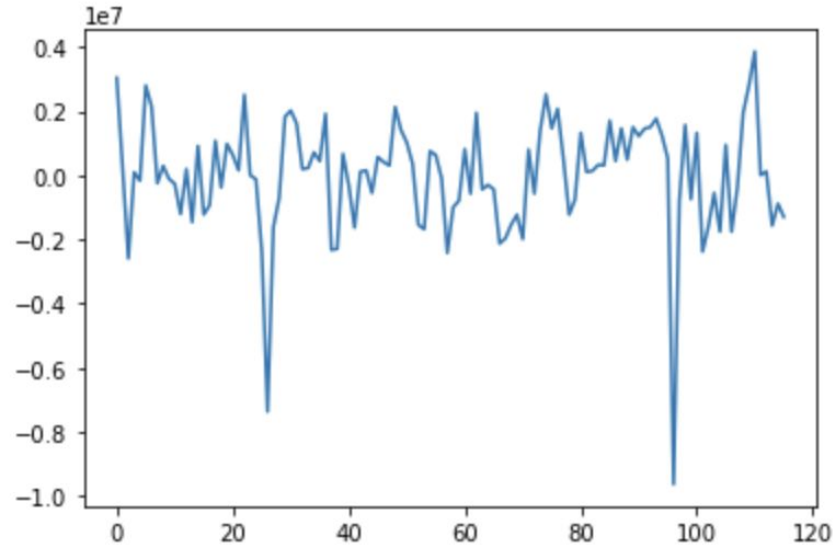
SARIMAX Residuals



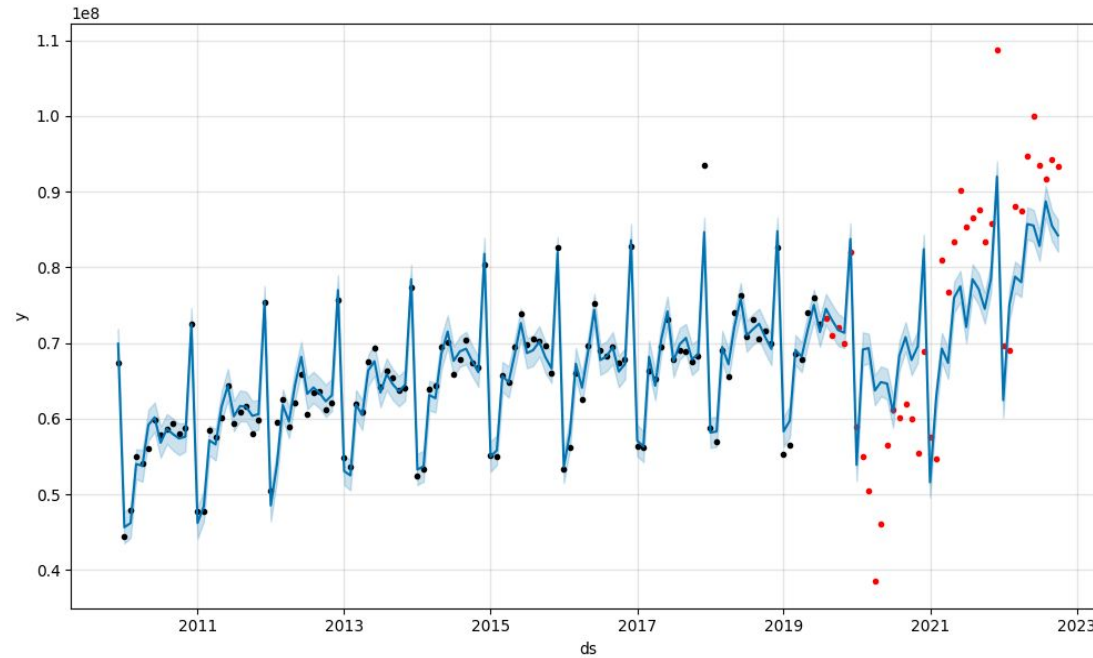
PROPHET



PROPHET RESIDUAL ANALYSIS



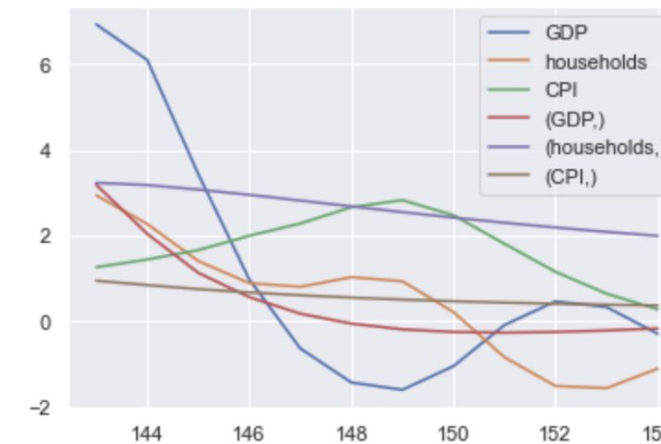
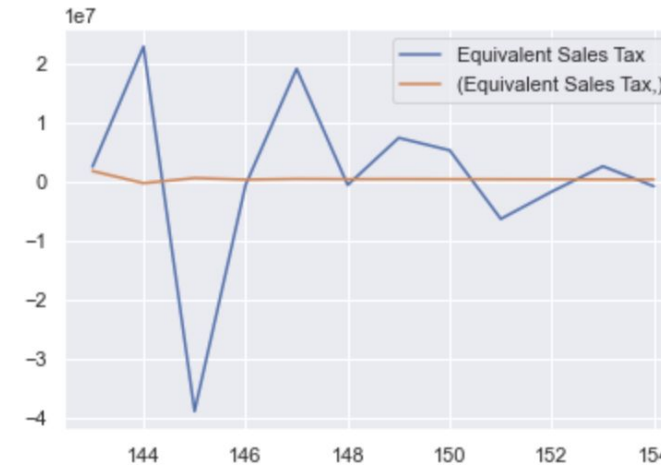
PROPHET with all 3 regressors



ERRORS	
MAE	18694668.759
MAPE	0.2362
sMAPE	45.62
MASE	1.92

Error covariance matrix

	coef	std err	z	P> z	[0.025	0.975]
sqrt.var.Equivalent Sales Tax	8.937e+06	4.76e-10	1.88e+16	0.000	8.94e+06	8.94e+06
sqrt.cov.Equivalent Sales Tax.GDP	0.5227	0.001	593.960	0.000	0.521	0.524
sqrt.var.GDP	3.5644	0.002	1439.778	0.000	3.560	3.569
sqrt.cov.Equivalent Sales Tax.households	0.1346	0.009	14.584	0.000	0.117	0.153
sqrt.cov.GDP.households	0.1951	0.007	29.953	0.000	0.182	0.208
sqrt.var.households	0.6316	0.035	17.958	0.000	0.563	0.700
sqrt.cov.Equivalent Sales Tax.CPI	0.0278	0.015	1.919	0.055	-0.001	0.056
sqrt.cov.GDP.CPI	0.1758	0.018	9.689	0.000	0.140	0.211
sqrt.cov.households.CPI	0.0980	0.019	5.073	0.000	0.060	0.136
sqrt.var.CPI	0.1719	0.010	16.810	0.000	0.152	0.192



ERRORS	
MAE	9004335.26
MAPE	1.04
sMAPE	178.86
MASE	1.21

VAR

Summary of Regression Results

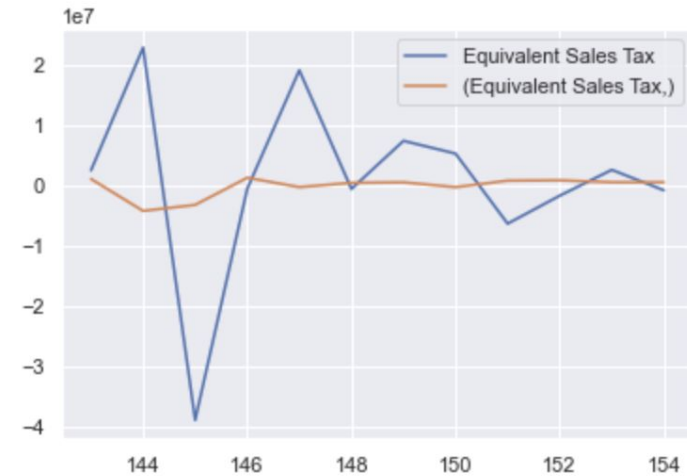
Model:	VAR		
Method:	OLS		
Date:	Thu, 25, May, 2023		
Time:	10:37:48		
No. of Equations:	4.00000	BIC:	24.6627
Nobs:	138.000	HQIC:	23.8065
Log likelihood:	-2317.45	FPE:	1.22077e+10
AIC:	23.2203	Det(0mega_mle):	7.67049e+09

Correlation matrix of residuals

	Equivalent Sales Tax	GDP	households	CPI
Equivalent Sales Tax	1.000000	0.179586	0.057303	0.101240
GDP	0.179586	1.000000	0.278093	0.673668
households	0.057303	0.278093	1.000000	0.477868
CPI	0.101240	0.673668	0.477868	1.000000

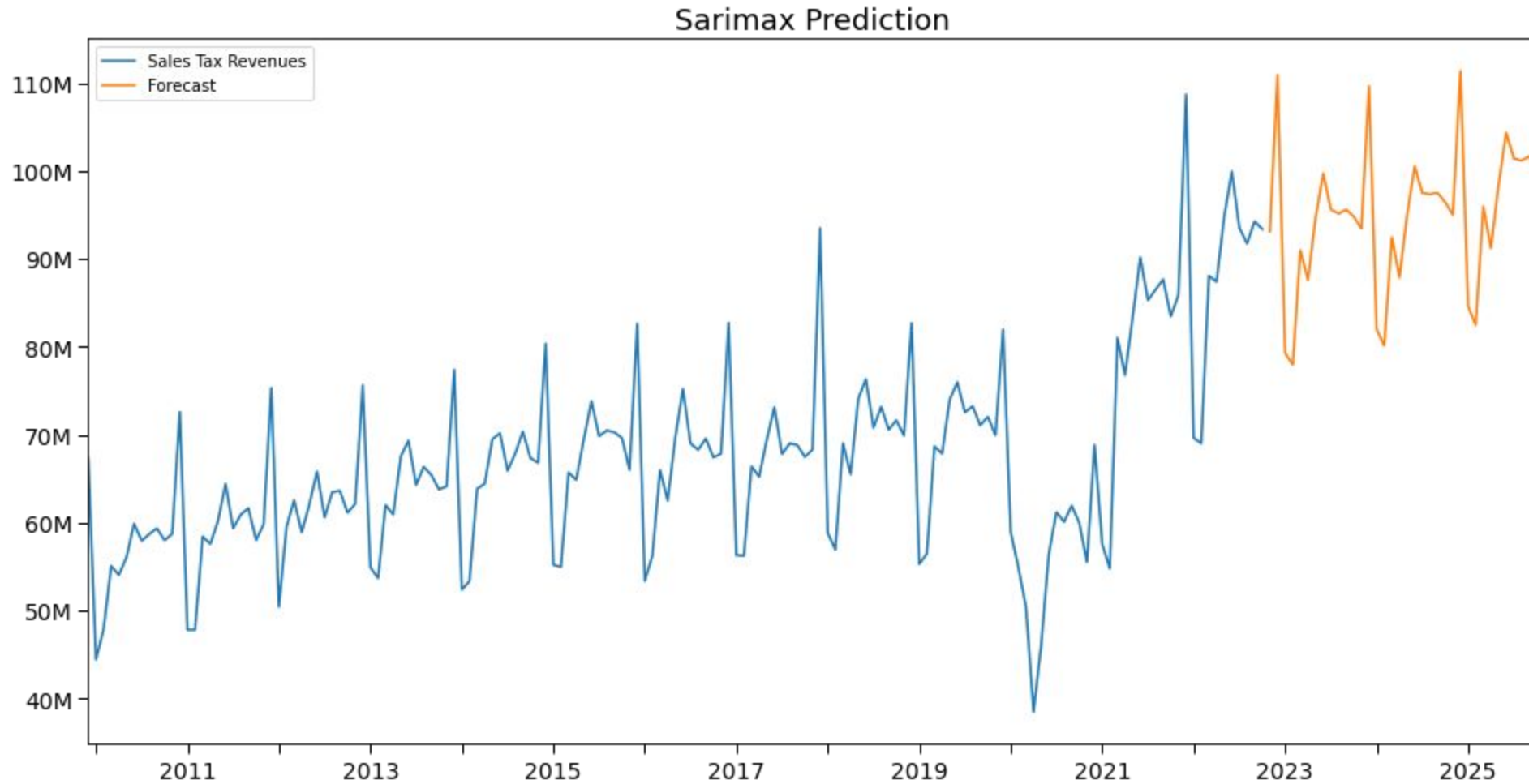
ERRORS

MAE	9341932.75
MAPE	1.25
sMAPE	180.58
MASE	0.84



Conclusion

Forecasting with the best Model

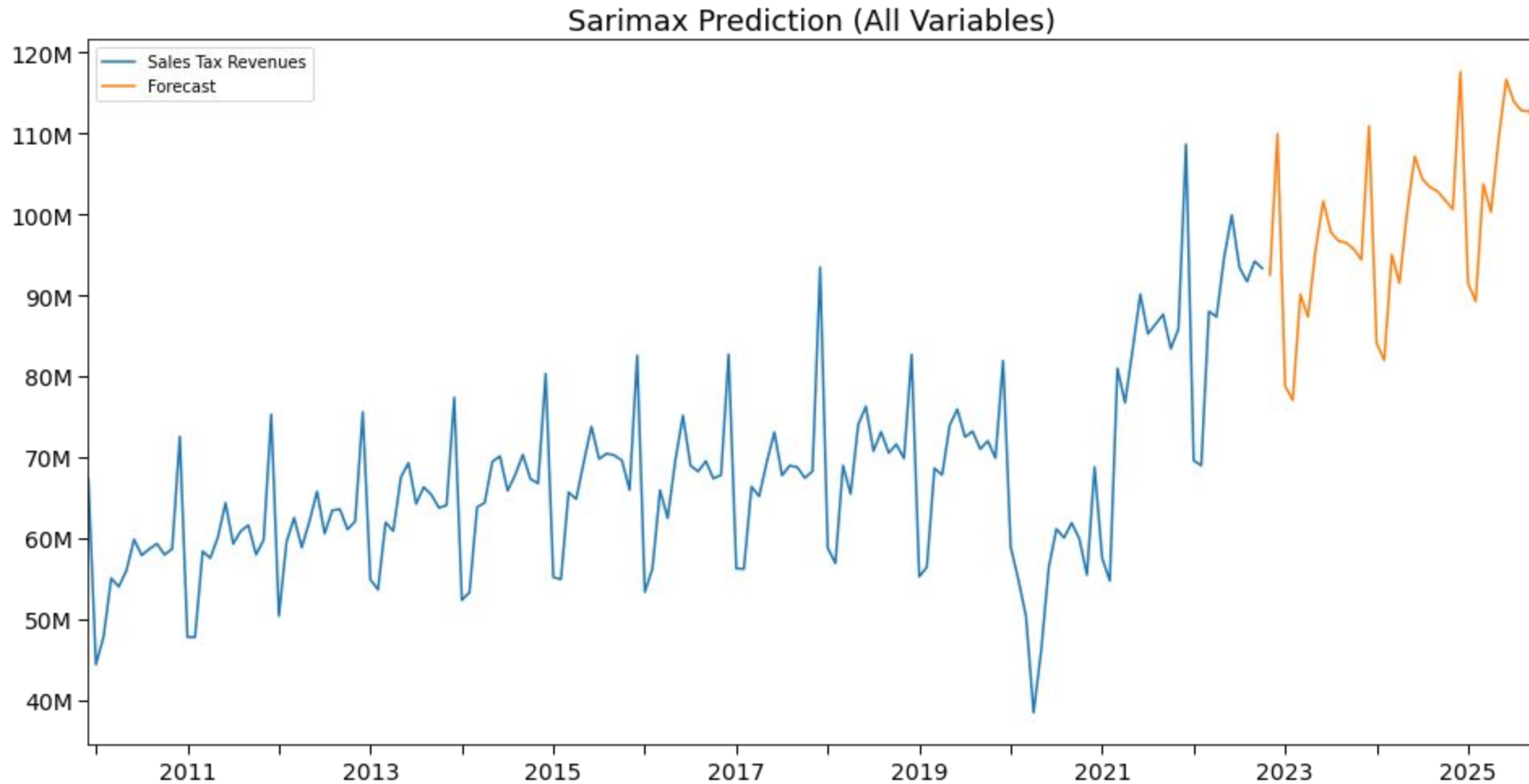


Next Steps and Future Recommendations

- Adjust ARIMA Orders
- Adjust Forecast time horizons
- Explore other economic variables (income statistics, population statistics, tourism etc.)
- Granger Causality Analysis
- Volatility Analysis (TBATs, GARCH, etc.)

Appendix

Forecasting with the Best Model



Predictions are based on Moody's analytics forecasts of the economic variables