Are the Major National Parks stealing tourists from surrounding parks?

Casey Ellis

Meenu Ravi

Kentaro Hama

Introduction

Yosemite National Park is one of the most famous tourist destinations in the world. In 2019, it saw nearly four and a half million visitors during the year. However, there are other great parks around Yosemite, such as Kings Canyon National Park and Sequoia National Park (hereinafter referred to as "surrounding parks".) If you were to visit Yosemite on your next vacation, would you want to stop by any of the surrounding parks? Or would you rather enjoy nature in Yosemite and not have to visit the surrounding parks? In other words, do you consider Yosemite and its surrounding parks to be complementary or substitute? Since both cases are theoretically possible, an empirical approach is needed to see which relationship is stronger.

Therefore, our group will build models to explain the demand for visits to the surrounding parks and examine the impact of increases or decreases in the number of visitors to Yosemite on the number of visitors to the surrounding parks. Specifically, we fit SARIMA, Auto ARIMA, and VAR models to the training set and predict the test set. Then, we compare the error metrics of each model and finally propose a best model.

Data

The data used in our analysis is the number of visitors by park and by month from 1979 to 2019 (we dropped the data for 2020 as an outlier to account for the impact of COVID-19.) The data is obtained from Visitor Use Statistics provided by the U.S. National Park Service. We also use monthly gasoline prices and monthly US population as external regressors.

We defined "surrounding parks" as any national parks or national monument within a 100-mile radius of Yosemite. The parks that fall into this category are Devils Postpile National Monument, Kings Canyon National Park, Sequoia National Park.

Data Preprocessing

Process for cleaning up data

Exploratory Data Analysis

- Quantitative statistics
- Qualitative statistics
- Plots of visitors by parks (Yosemite, surrounding parks)
- Correlation between visitors to Yosemite and surrounding parks (Section 2.6 of the textbook may be useful.)
- Stationarity check (ADF and KPSS tests), transformation/differencing, ACF/PACF plots

Models

(1) ARIMA with external regressors

- Fit the model
- Examine the residuals (plots, histograms ACF, Ljung Box results)
- Calculate the error metrics

(2) Prophet with external regressors

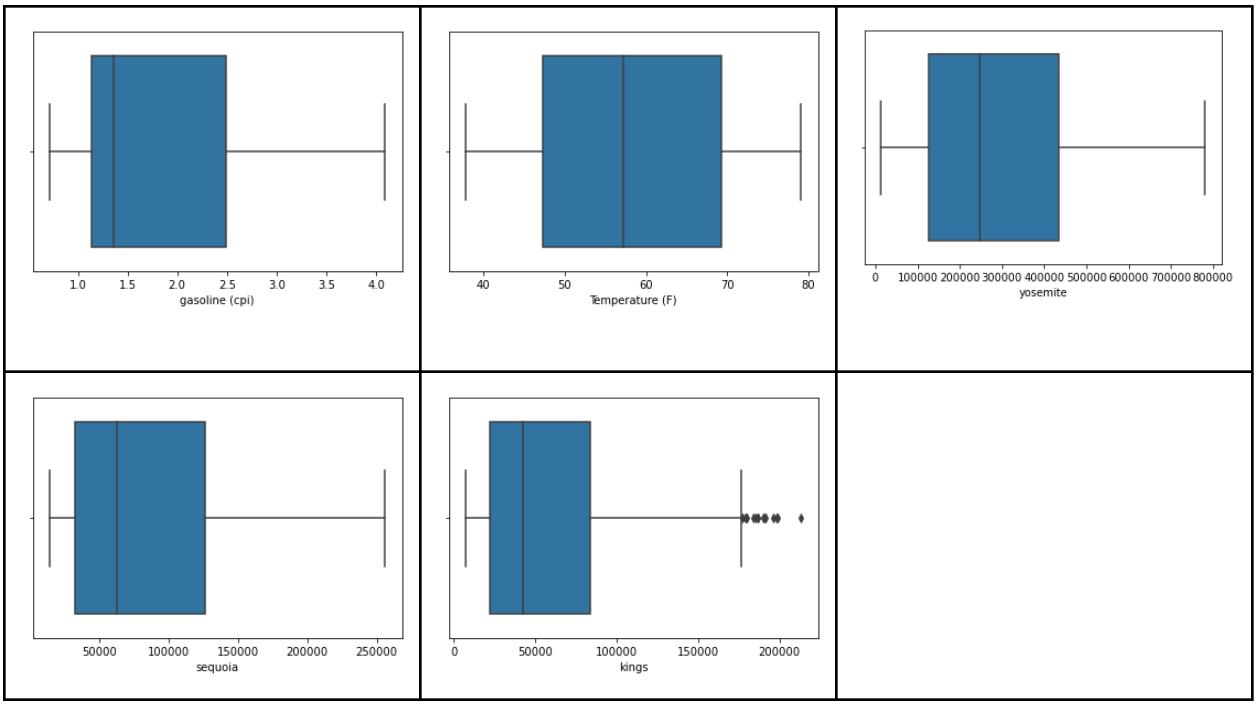
- Fit the model
- Examine the residuals (plots, histograms ACF, Ljung Box results)
- Calculate the error metrics

(3) X model

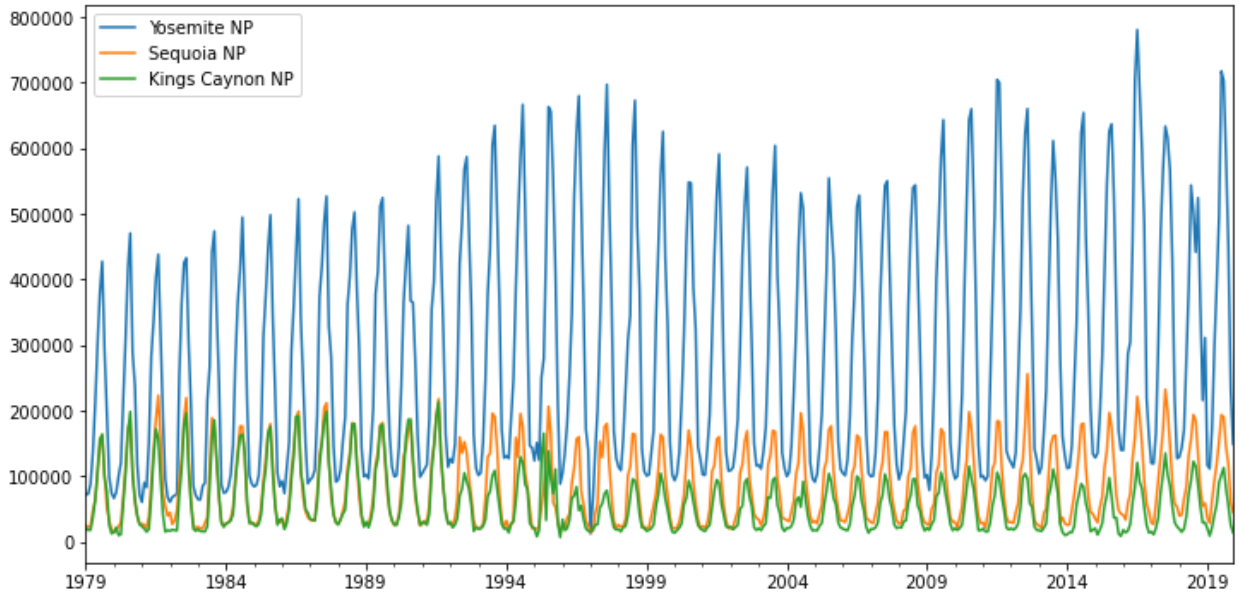
- Fit the model
- Examine the residuals (plots, histograms ACF, Ljung Box results)
- Calculate the error metrics

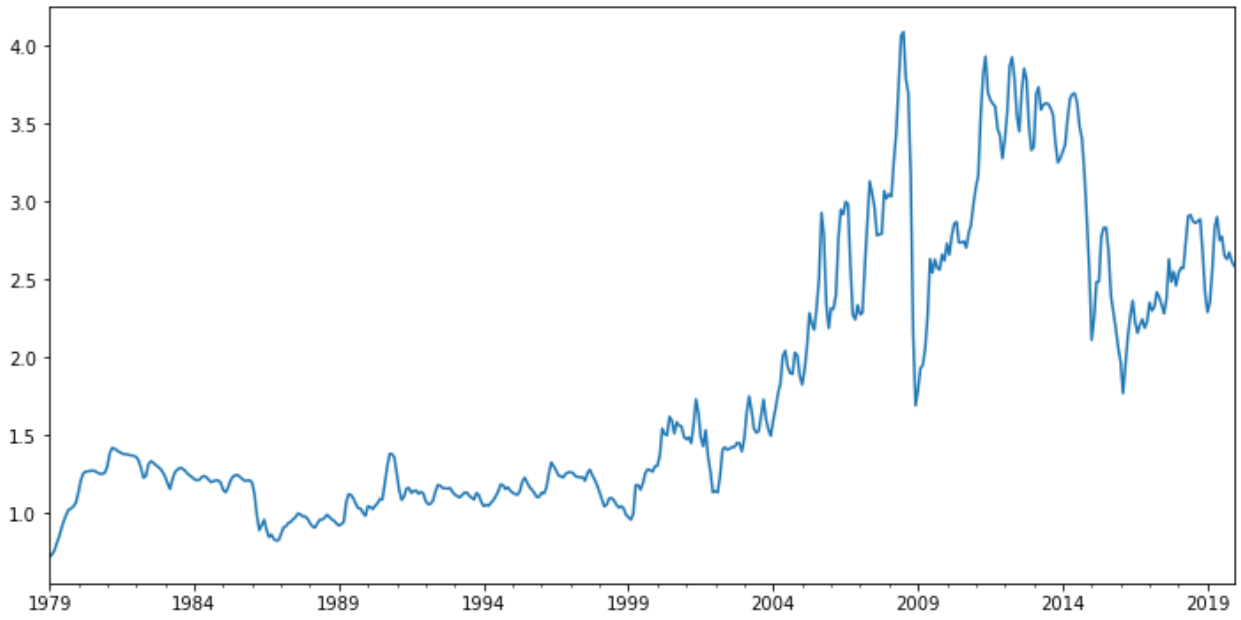
EDA Figures

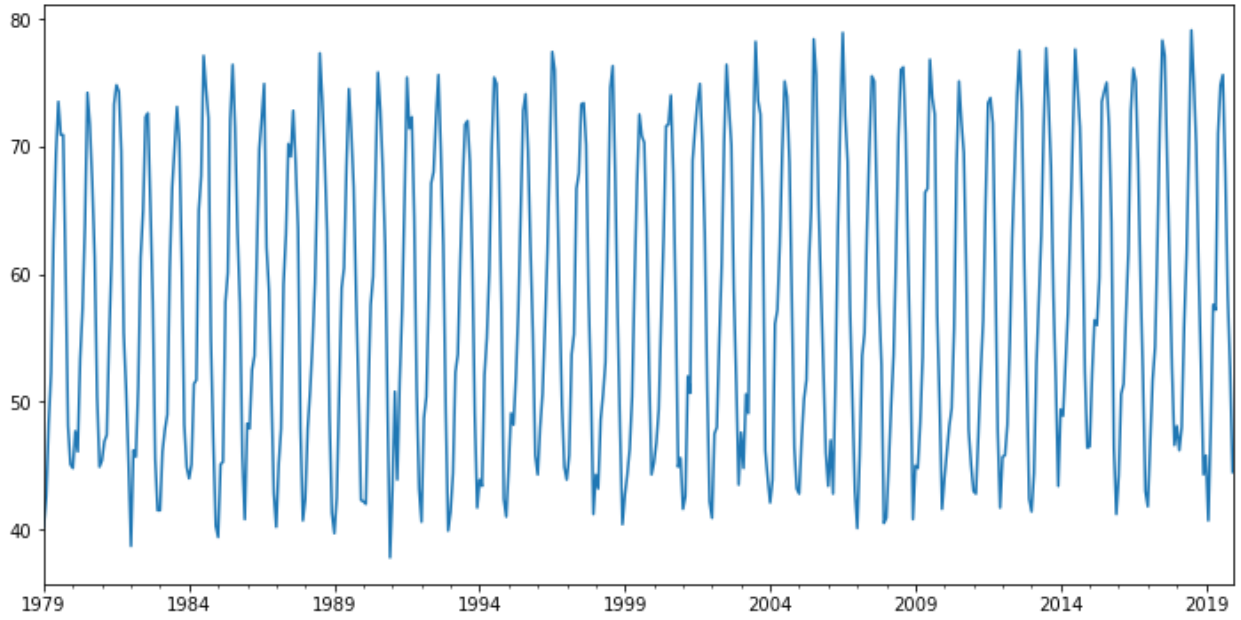
Outliers:


1.0 1.5 2.0 2.5 3.0 3.5 4.0
gasoline (cpi)
40 50 60 70 80
Temperature (F)
0 100000 200000 300000 400000 500000 600000 700000 800000
yosemite
50000 100000 150000 200000 250000
sequoia
0 50000 100000 150000 200000
kings

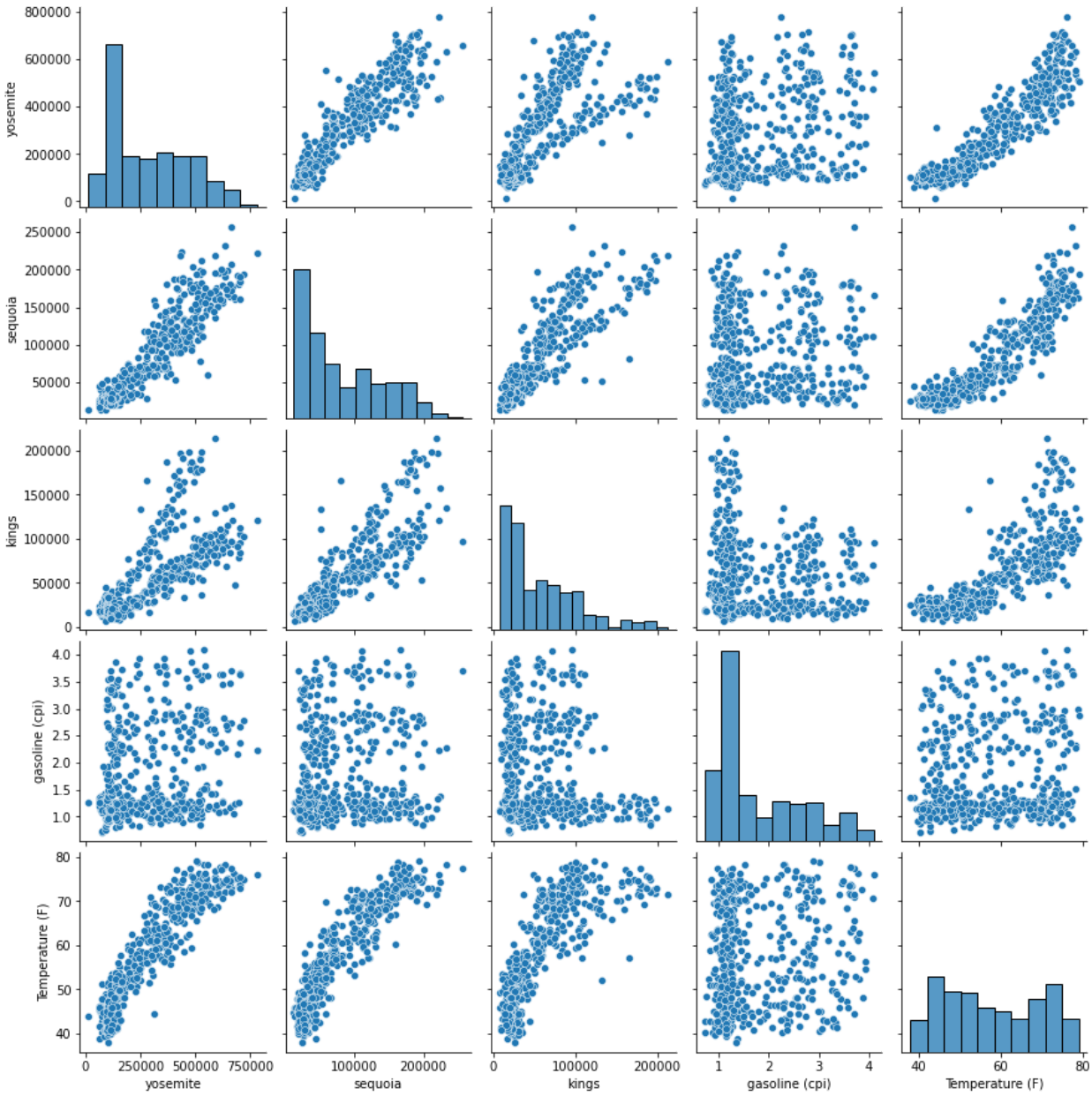
Graphs


Yosemite NP
Sequoia NP
Kings Caynon NP
800000
700000
600000
500000
400000
300000
200000
100000
0
1979
1984
1989
1994
1999
2004
2009
2014
2019

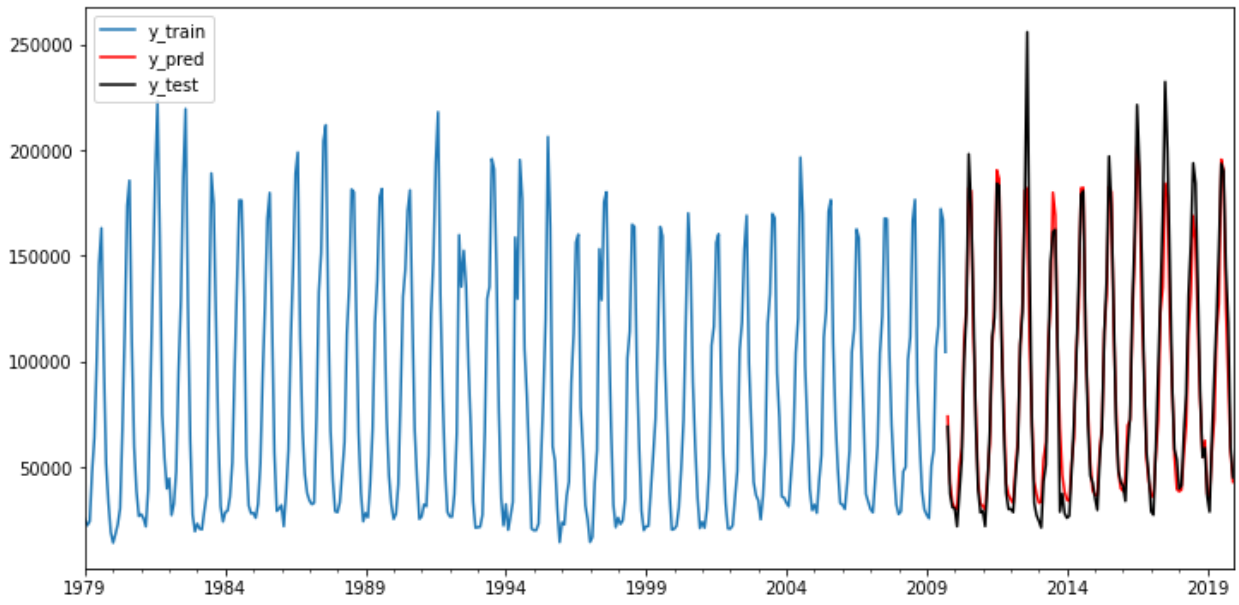




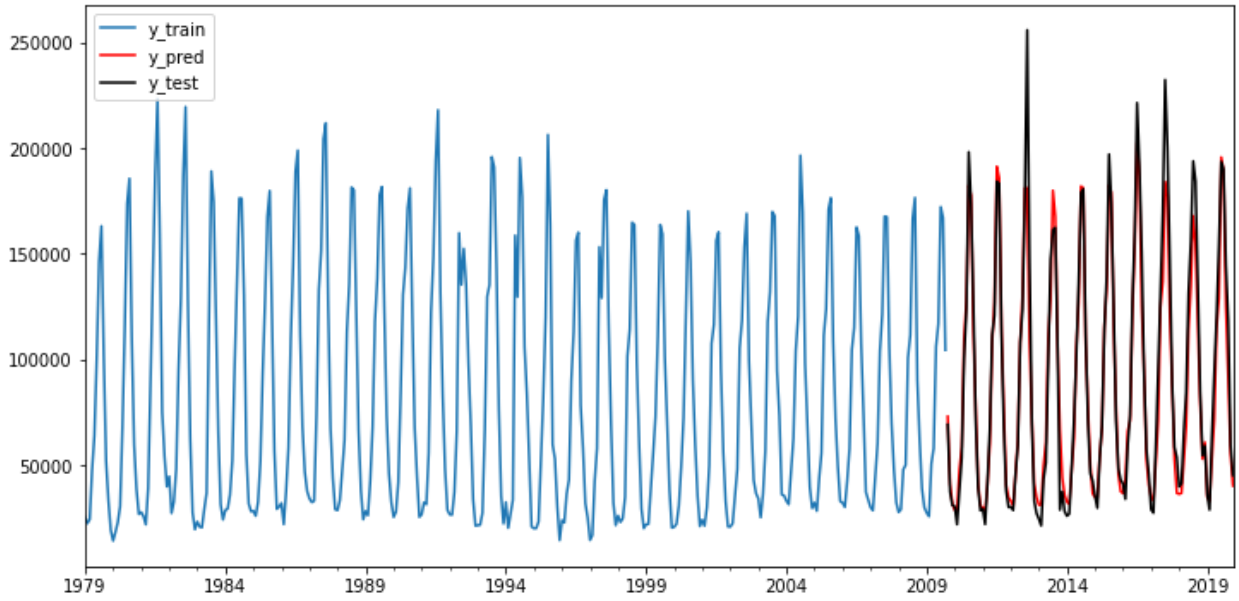
Correlation



Arima



Auto arima



Discussion

Initial Modeling Attempts:

Auto arima and sarima results:

Metrics	Sarima	Auto Sarima	Model
MSE	mse: 201469591.4395621	mse: 197891174.50464195	
MAE	mae: 9498.54714231295	mae: 9163.4322275975	
SMAPE	smape: 0.1223567793033211	smape: 0.11426232117844484	
MAPE	mape: 13.110400361633308	mape: 11.848661357599065	
MASE	mase: 0.34993384161301383	mase: 0.33758794831678496	

According to the error metrics of the three models, **X** is the best model. Looking at the results of the best model, the coefficient on visitors to Yosemite is positive and statistically significant. This indicates that the demand for visits to Yosemite and the surrounding parks are complementary, and that more visitors to Yosemite would lead to more visitors to the surrounding parks.

How can we use these results in the business world? For example, managers of nearby parks may be able to attract more visitors by conducting joint campaigns with Yosemite Park.

Finally, we discuss the limitations of our analysis and possible future works.

Next steps:

- We would like to work on VAR model
- We would like to work on hierarchical model also
- We hope to get this done next week