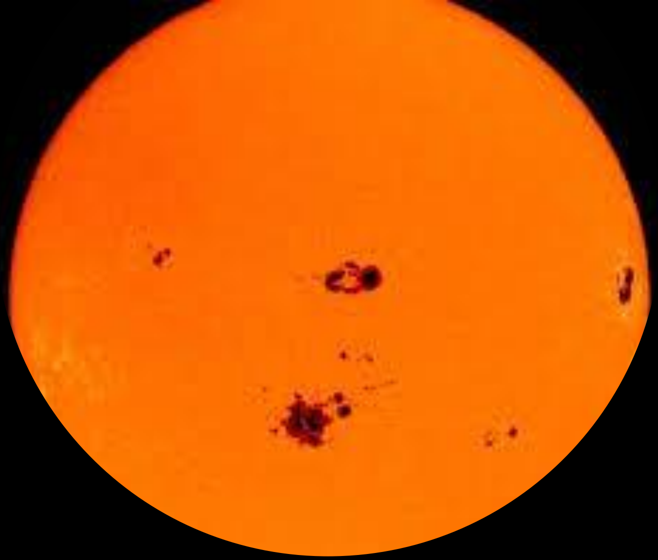
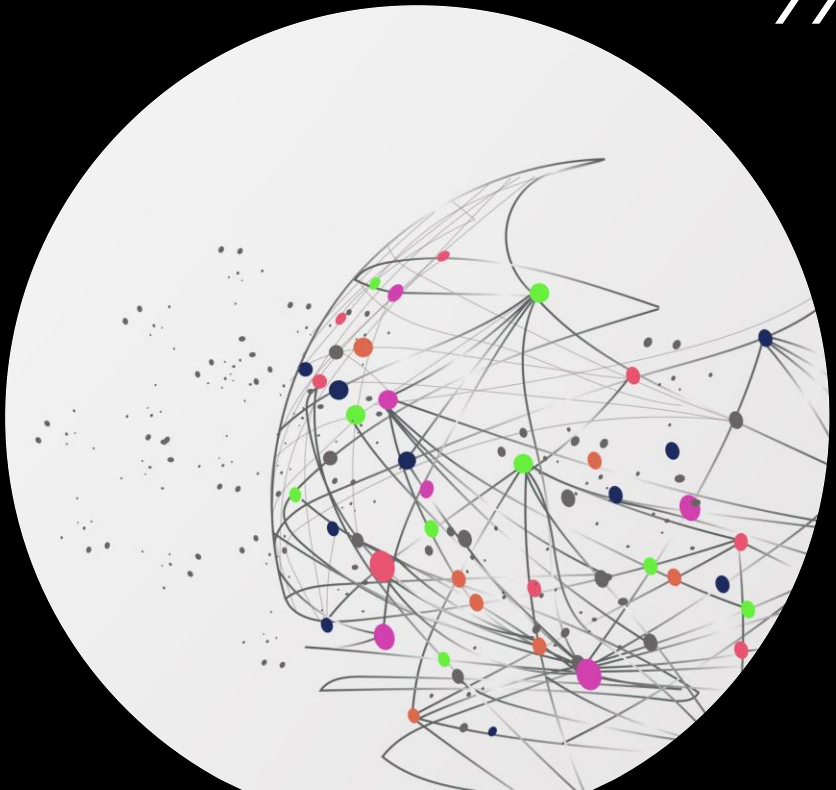
Sunspots Data

By Jonah Zembower





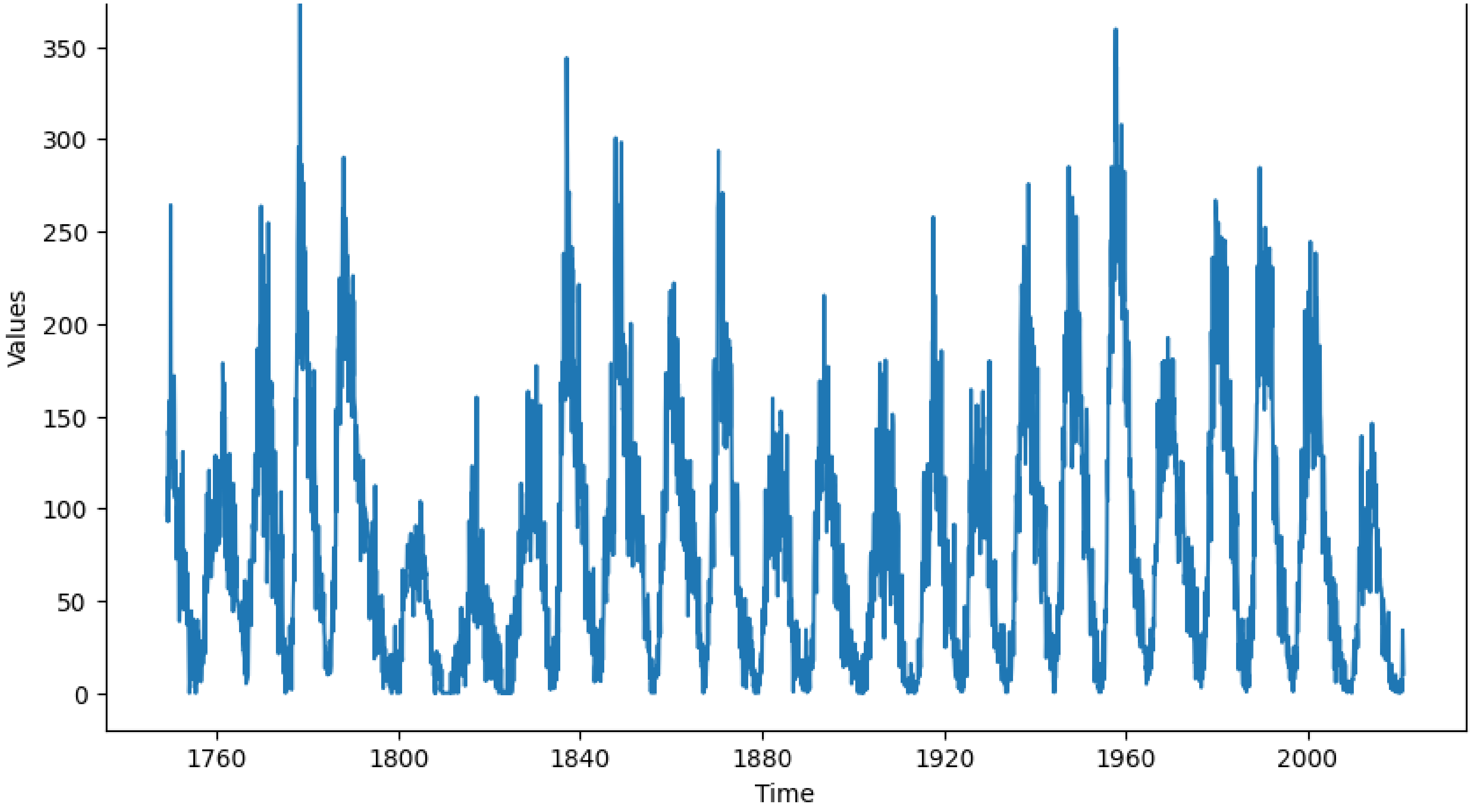
Sunspots Dataset

- This dataset contains historical records of sunspots. Typical sunspot datasets include measurements such as the date of observation, sunspot numbers, and sometimes details about solar cycles. Such data is crucial for studying solar activity and its cycles over time.

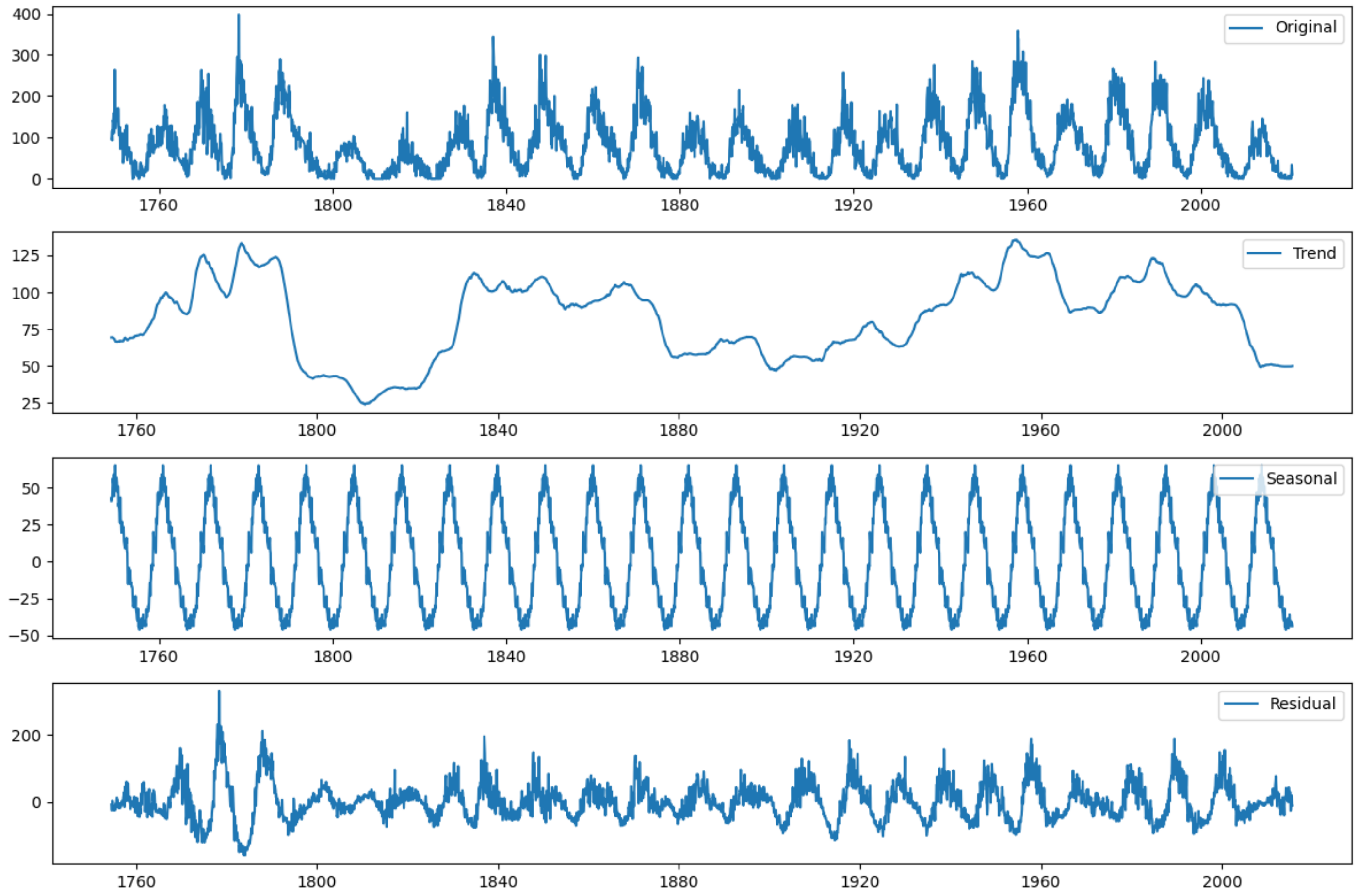
Creating a Data frame to Split into Monthly Values

	January	February	March	April	May	June	July	August	September	October	November	December
0	96.7	104.3	116.7	92.8	141.7	139.2	158.0	110.5	126.5	125.8	264.3	142.0
1	122.2	126.5	148.7	147.2	150.0	166.7	142.3	171.7	152.0	109.5	105.5	125.7
2	116.7	72.5	75.5	94.0	101.2	84.5	110.5	99.7	39.2	38.7	47.5	73.3
3	58.3	83.3	118.3	98.8	99.5	66.0	130.7	48.8	45.2	77.7	62.7	66.7
4	73.3	53.3	76.2	63.3	60.0	52.8	36.7	65.0	46.7	41.7	33.3	11.2
...	...	...	...	...	...	...	...	...	...	...	...	...
267	57.0	56.4	54.1	37.9	51.5	20.5	32.4	50.2	44.6	33.4	21.4	18.5
268	26.1	26.4	17.7	32.3	18.9	19.2	17.8	32.6	43.7	13.2	5.7	8.2
269	6.8	10.7	2.5	8.9	13.1	15.6	1.6	8.7	3.3	4.9	4.9	3.1
270	7.7	0.8	9.4	9.1	9.9	1.2	0.9	0.5	1.1	0.4	0.5	1.5
271	6.2	0.2	1.5	5.2	0.2	5.8	6.1	7.5	0.6	14.4	34.0	21.8

Looking at the Time Series of the Data

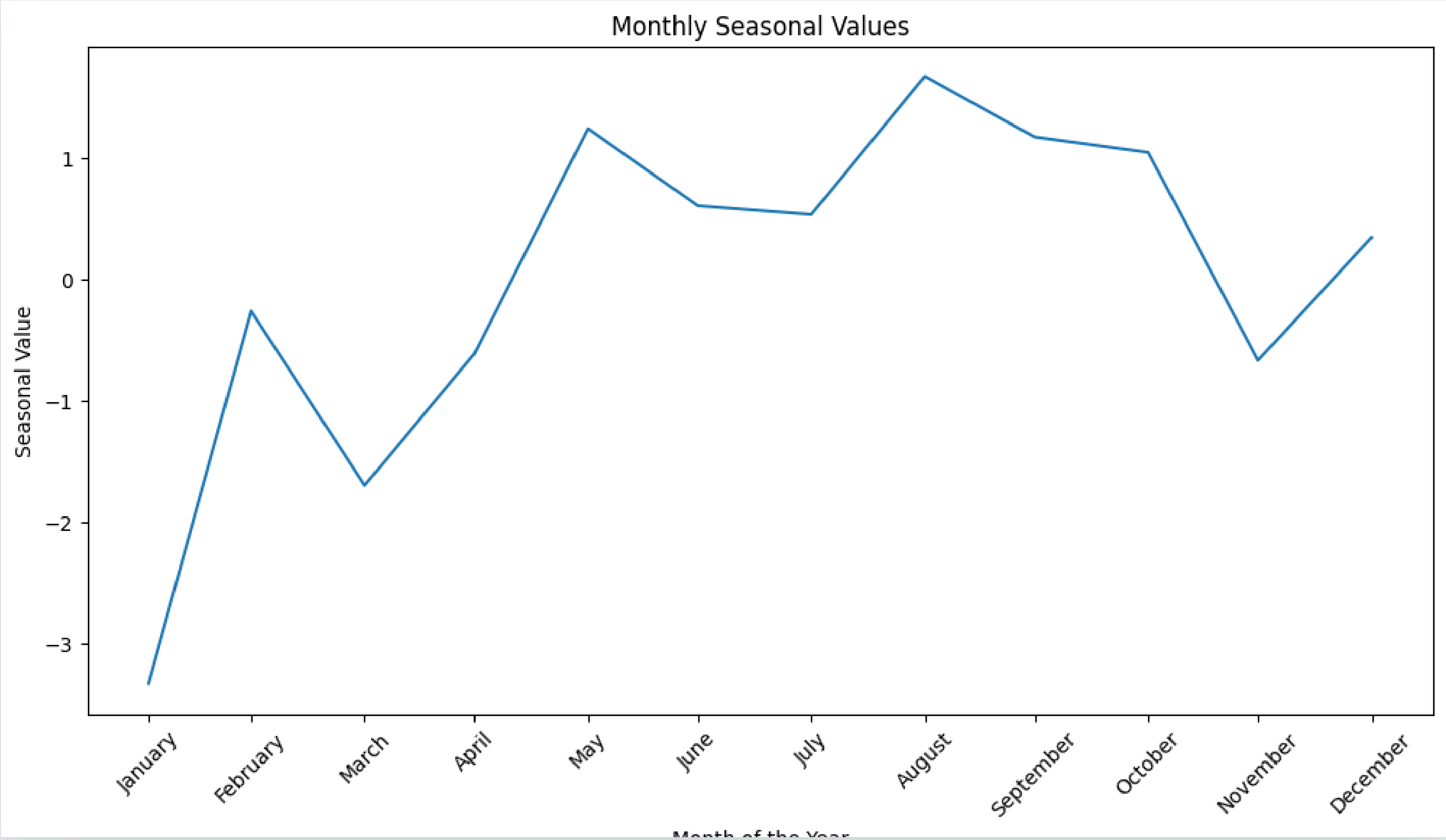


Seasonal Decomposition



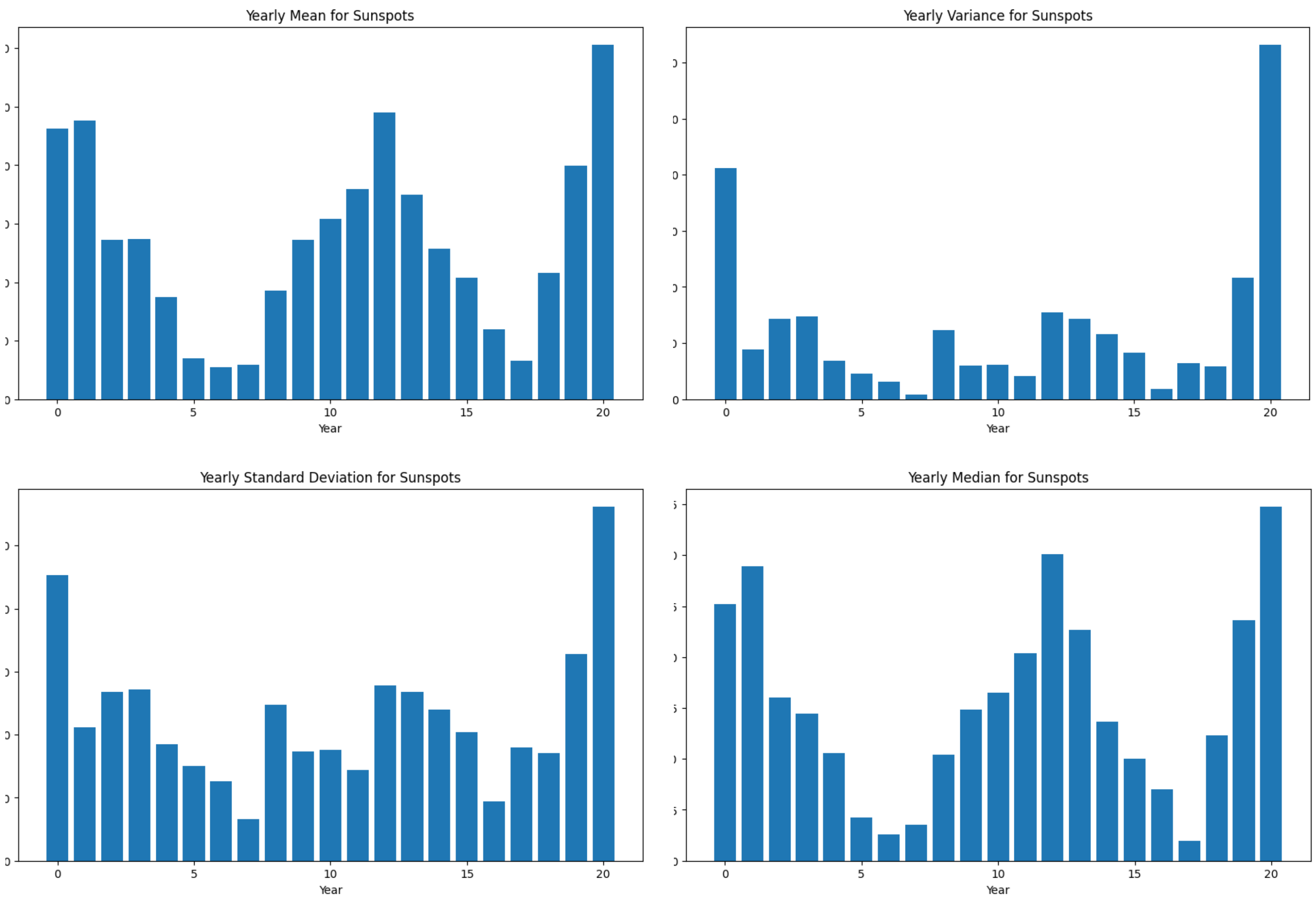
- The seasonal component aligns with the 11 year solar cycle.
- Residuals appear mostly random without any discernible pattern.

Looking at Monthly Values for a Year



- General increase throughout the year in the trend.

Looking at the Yearly Mean, Variance, Standard Deviation, and Median



Building the ARIMA Model

Components of Box-Jenkins Models:

p: Autoregressive (AR) order for non-seasonal component.

d: Differencing order for non-seasonal component.

q: Moving average (MA) order for non-seasonal component.

P: Autoregressive (AR) order for seasonal component.

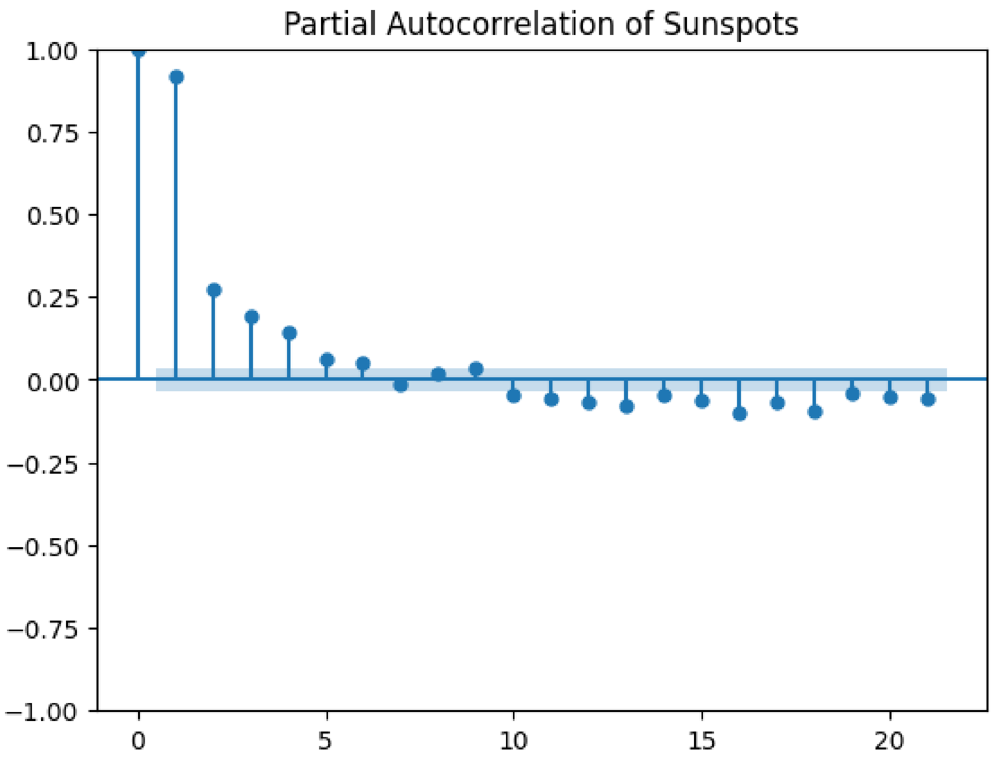
D: Differencing order for seasonal component.

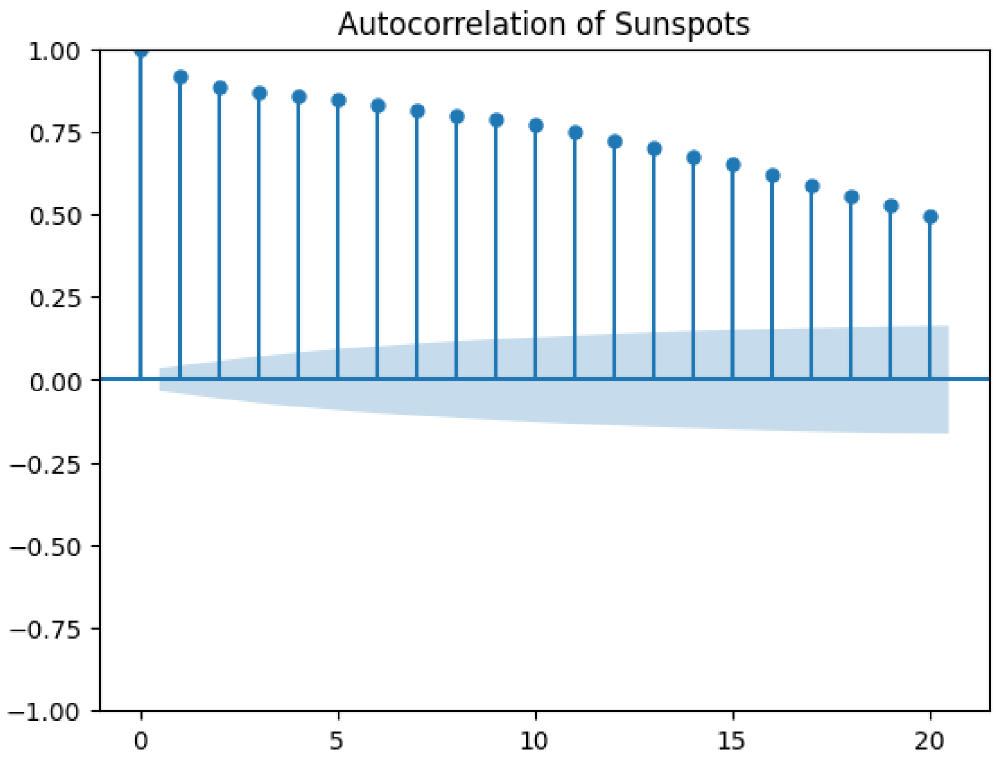
Q: Moving average (MA) order for seasonal component.

s: Seasonal period (e.g., 12 for monthly data, 4 for quarterly data).

Looking at AR and MA Components

- AR order selection [1,2,3,4,16]
- MA order selection [1,2,3,4,5]
- d order selection [0,1]





ARIMA Selection

Used automated process to select based on the AIC and BIC

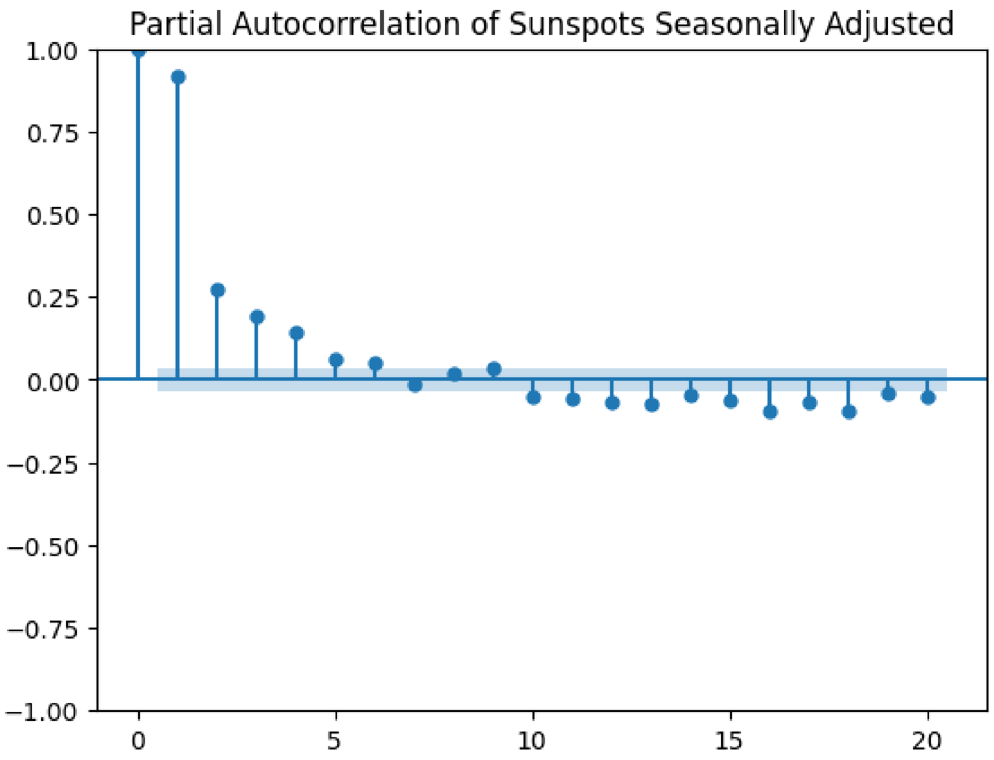
Found:

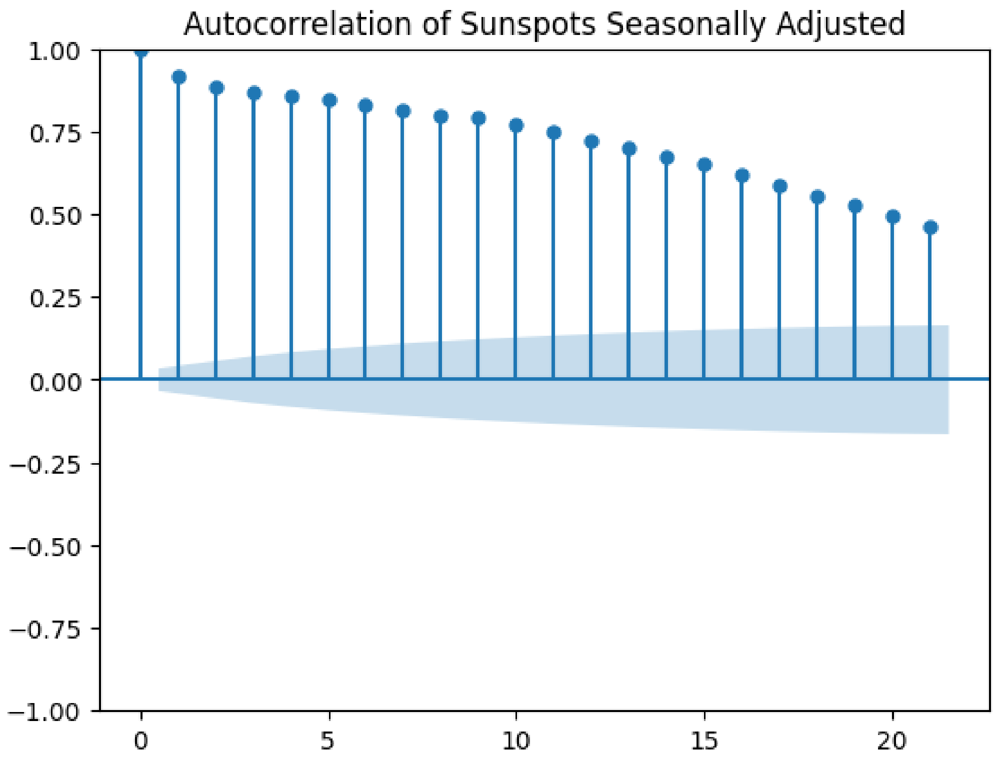
ARIMA (16,0,1)

- AIC: 30145.431603
- BIC: 30270.160889

Looking at P,D,Q for removing Seasonality

- P (AR) order selection [1,2,3,4,16]
- Q (MA) order selection [1,2,3,4,5]
- D order selection [0,1]





SARIMA Selection

Used automated processes to select based on AIC and BIC

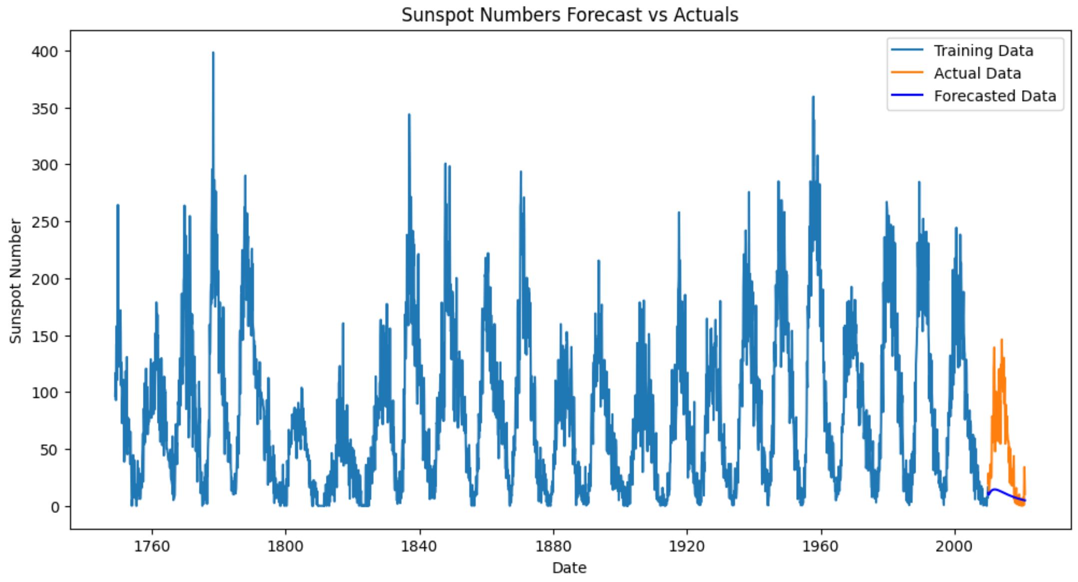
Found:

SARIMA (3,0,2) (0,0,1,12)

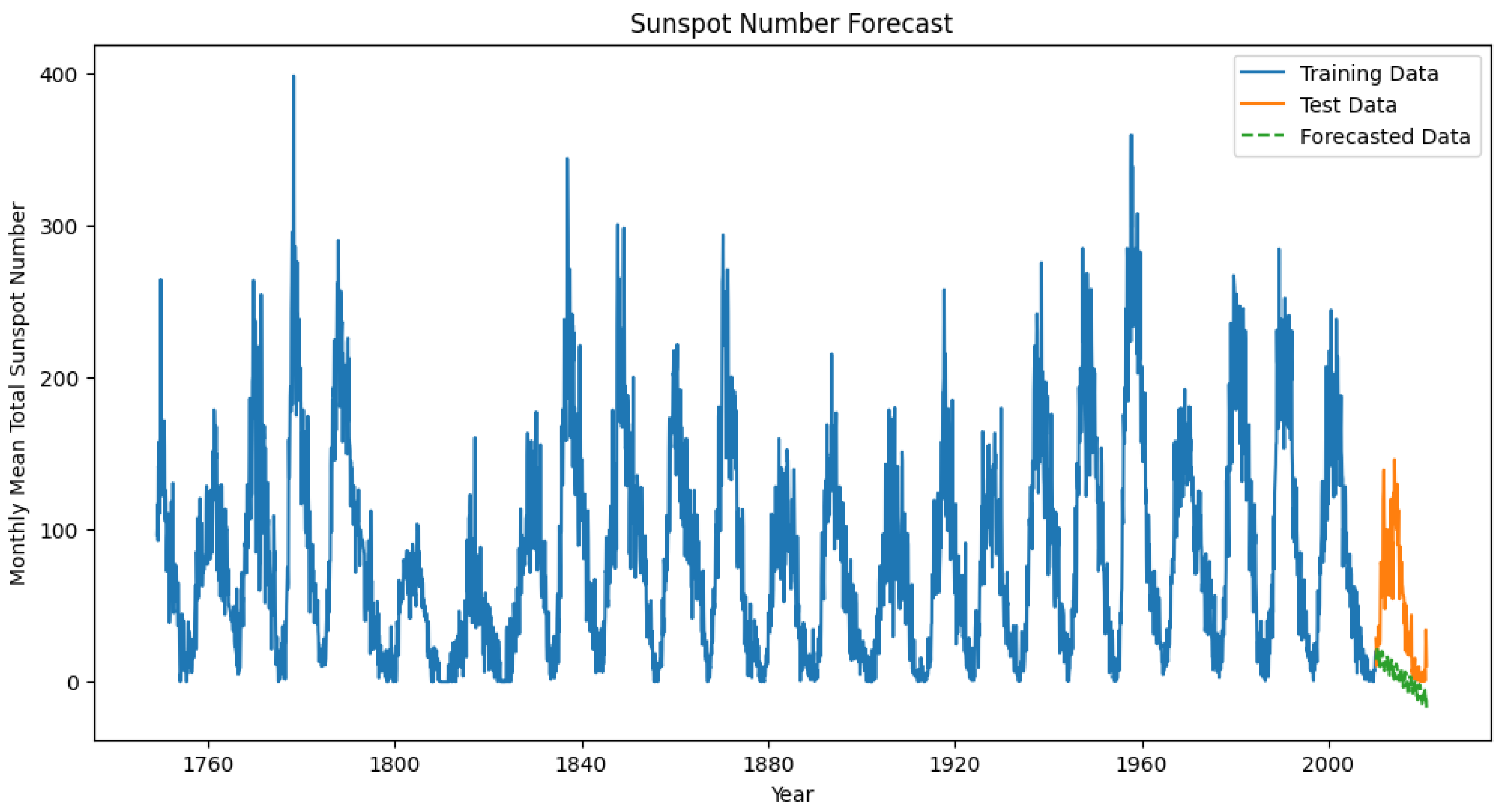
- AIC: 30181.10394832894
- BIC: 30229.83206866262

Forecast vs. Actuals using SARIMA

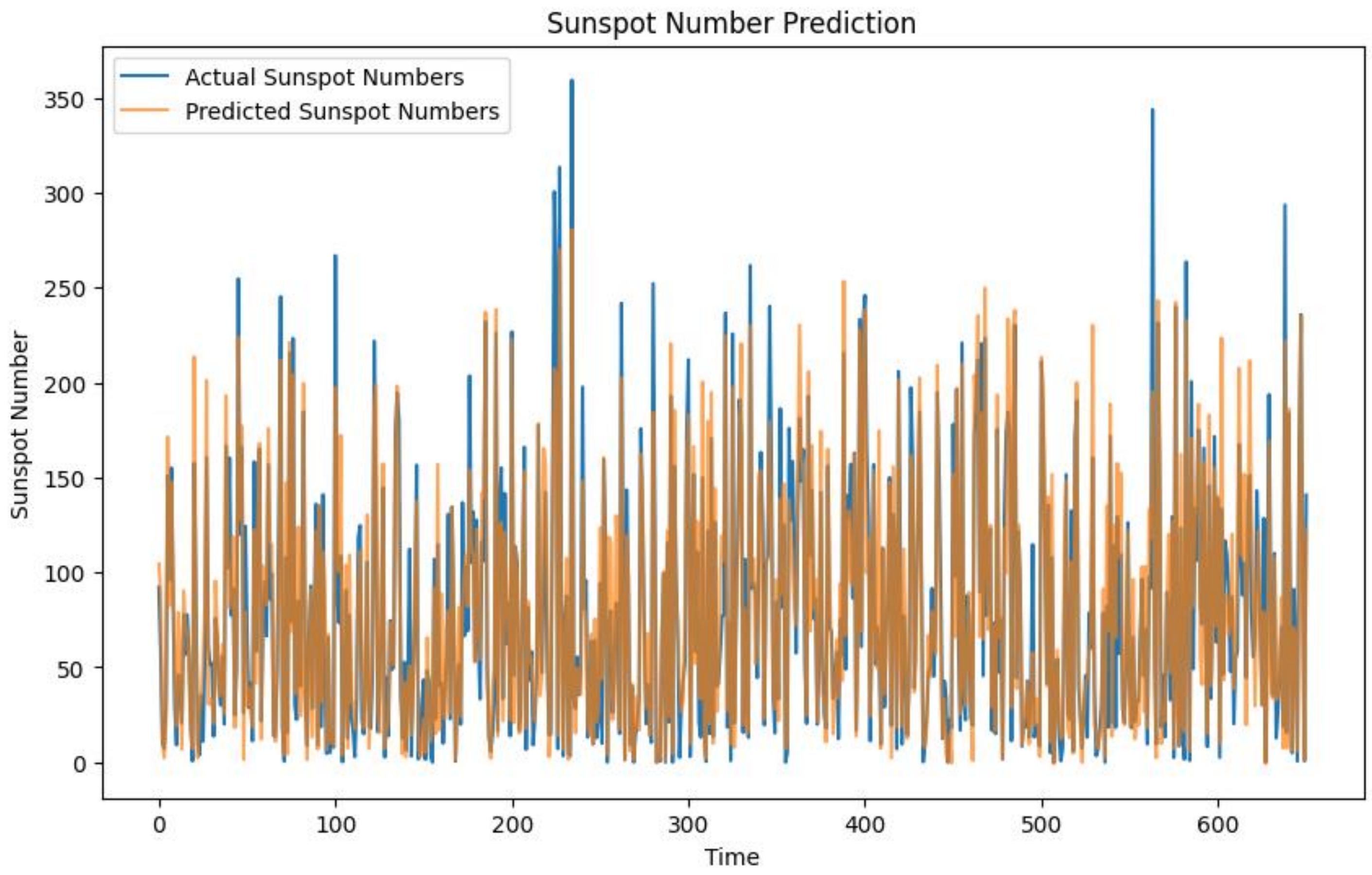
- We can see the forecast may have not accurately predicted the complex seasonality and trends in the data.



Using Holt's Winters Method



Using Neural Networks



Cross Validation of Neural Network

- Average MSE: 0.003776115895043971
- Average RMSE: 0.06127926364414369

Conclusion

- Improvement Opportunities:
  - Increasing Model Complexity: Adding more LSTM layers or neurons, or adjusting the model architecture (e.g., more dense layers, dropout for regularization).
  - Feature Engineering: Including additional features that might help predict sunspot numbers, such as prior cycle lengths or other solar activity indicators.
  - Hyperparameter Tuning: Experimenting with different learning rates, number of epochs, or batch sizes.
  - Data Augmentation: Increasing the dataset size or using data augmentation techniques to provide more training examples.