Documentation for My Denver Crime Forecasting

I recently completed the Udemy course: Python for Time Series Analysis and Forecasting with an intense feeling of excitement. On my path to becoming a data scientist, this moment feels like a major milestone. I’ve never felt so eager to apply what I’ve learned.

This is my first real project and I would welcome efficiency and general improvement suggestions. There is probably a much more elegant way to digitise the data than the method I used, which was cumbersome, but functional.

This project does two things:

* Examines which crime categories vary seasonally
* Makes forecasts of future crime rates for selected crime categories

Here’s the breakdown

1. A list of references I used during construction
2. Library Imports
3. DataFrame creation from the CSV data file
4. Set the index to datetime format
5. DataFrame cleanup – remove unwanted columns
6. Digitisation, Step 1 – create a template for translating the categorical entries to numeric
7. Digitisation, Step 2 – make a time series of the crime category to use as a basis for digitising the individual crime categories
8. Digitisation, Step 2 – make time series for individual crime categories and transform them into digital
9. Add the digitised crime categories to the DataFrame
10. Import Seasonality data CSV file – I requested the temperature data for a Denver zip code from <https://www.climate.gov/maps-data/dataset/past-weather-zip-code-data-table> and made a CSV
11. Suspecting burglary is seasonal, I overlaid the two data sets
12. Resample the crime category data to a weekly basis to match the temperature data
13. Examine all the crime categories for seasonal variation
14. Preparing for application of forecasting: test for trends and auto-regression/auto-correlation
15. Use StatsModel Seasonal Decomposition to break out the seasonal element
16. Test Larceny against seasonal temperatures (with an 80 point boost to get a good overlap)
17. Test whether Larceny is truly seasonal:  
    Remove the seasonal element and test for randomness  
    (it should be random if the seasonal influence has been correctly identified and removed)
18. Model the data with ARIMA – use the ACF/PACF plots to select model order
19. Test against other orders to see if we have the best model
20. Use the modelled data to make forecasts
21. Try modelling with STL seasonal decomposition, forecast and test residuals
22. Try Holts-Winters exponential smoothing, forecast and test residuals