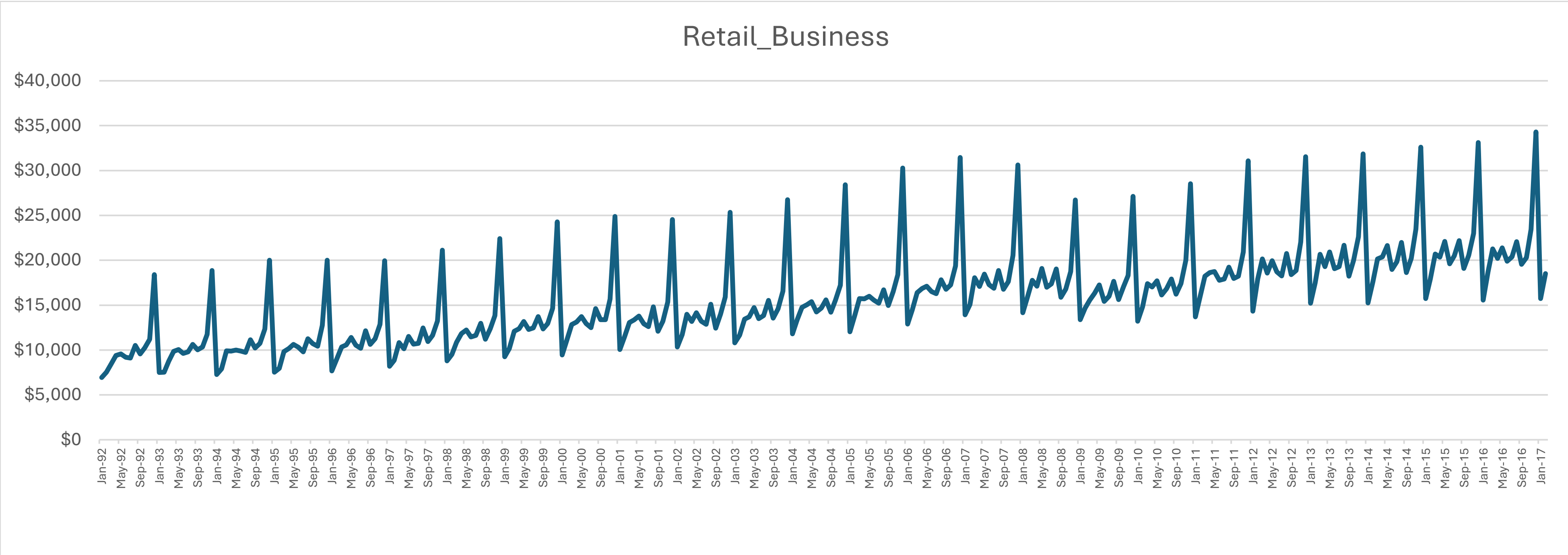
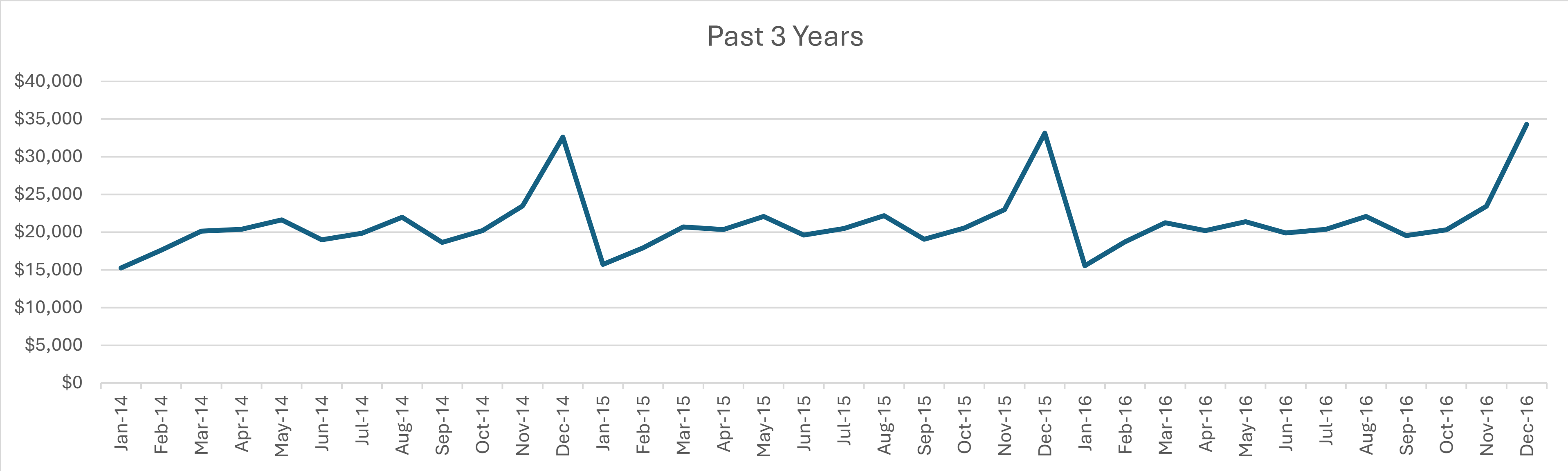
1. **Create a time series using the instructions provided in the Exercise.**
2. **Observe the pattern of the line in your time series and answer the following questions:**
    - **What characteristics does the pattern display (e.g., seasonality, stationarity)? Write a short paragraph to explain your answer.**
    - **What advice might you give your client based on this time series. Why?**



There are very regular spikes in the data, so this is showing seasonality.

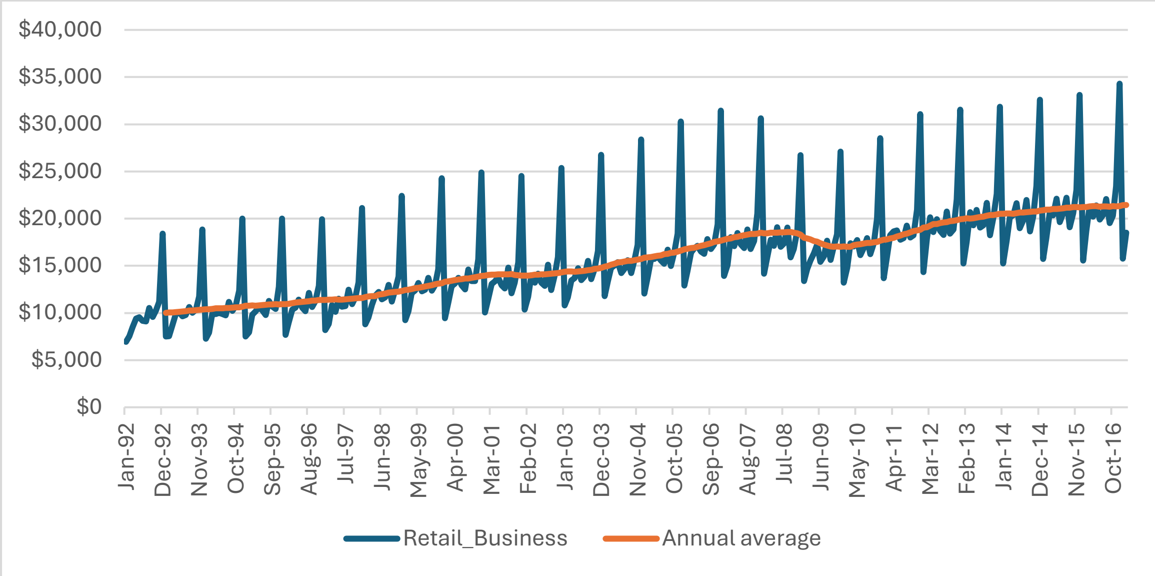
At this scale it is hard to see what the precise pattern is, so I have looked just at the past 3 years.



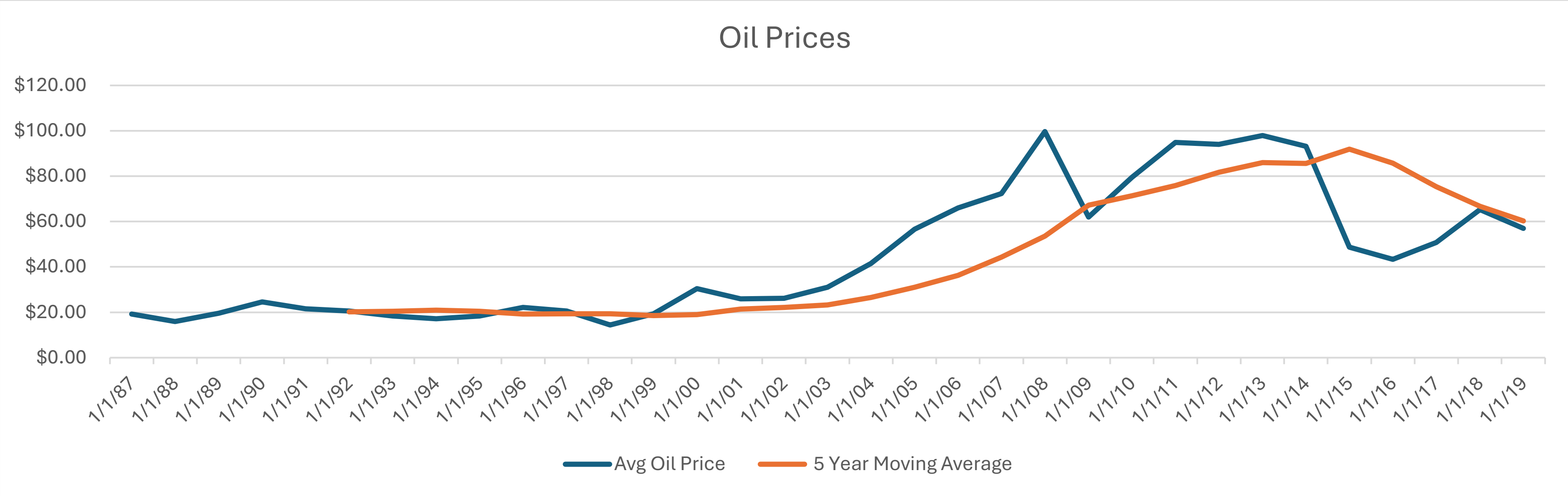
Now it is clear that the main spike occurs in December, but with growth from September each year

There are also smaller spikes in May and August.

The data does not show stationarity, as there is a clear upward trend over time. Plotting the annual average shows this clearly



3. **Create a simple moving average using the instructions in the Exercise.**
4. **Observe the pattern/trend of the oil price line in relation to the five-year moving average line and answer the following questions:**
    - **Is there a certain characteristic to the pattern and trend? Make sure to provide a short explanation for your answer.**
    - **Explain how the moving average affects oil price volatility and how it makes forecasting easier.**



There is no seasonality to this data, there is no repeating pattern every x number of years.

The overall trend is an increase in price, but more recently – from about 2009 – the rate has first slowed, and then – after 2015 - fallen.

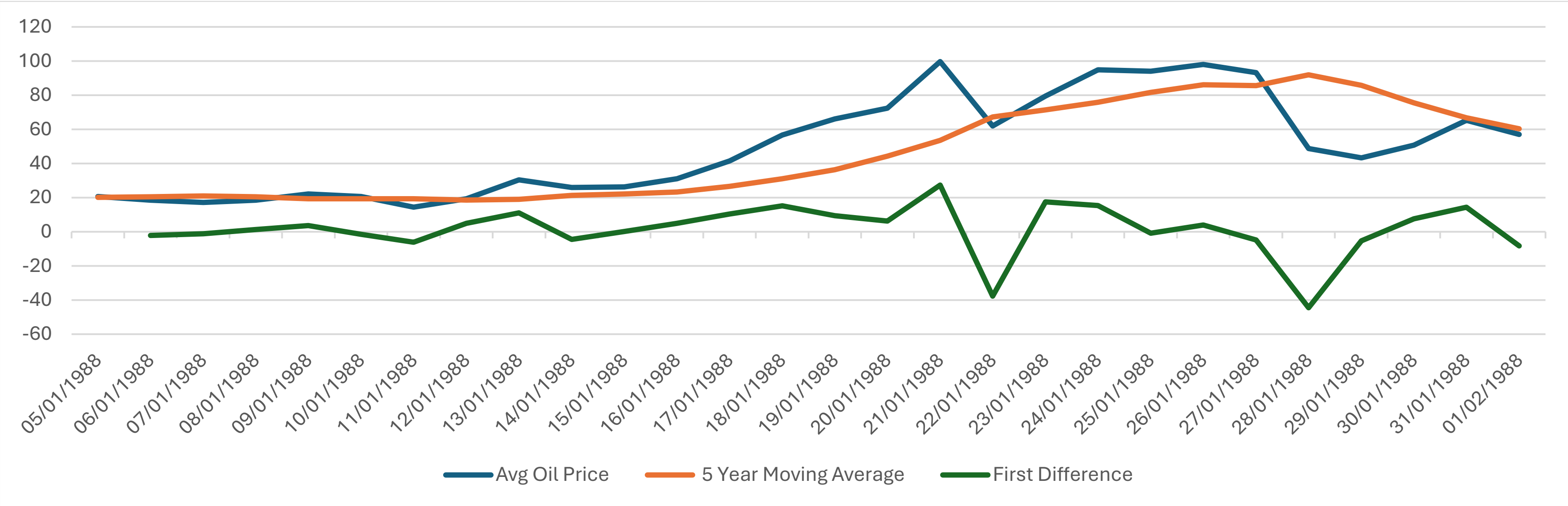
Even looking at annual averages there is quite a lot of fluctuation in the data, without the 5 year average, it would be harder to see the trends.

5. **This Exercise mainly looked at non-stationary time series. Briefly explain why you might convert a non-stationary time series into a stationary time series before applying a forecasting model. (If you need help answering this question, check out the Resources above.)**

“it’s easier to make predictions if the statistical properties of the data remain constant” – therefore we can make more accurate predictions with a stationary time series. Some statistical techniques rely on stationary data.

One way to do this is with Differencing – at it’s simplest subtracting each observation from the one before it.

Looking at this we see it removes the upward trend – this is now stationary data, with a value that is around 0 (we may need to do it again to flatten more of the fluctuations)



W can see this in the averages:

Avg Oil Price	5 Year Moving Average	First Difference
49.55	45.85	1.35

6. **There are lots of other forecasting models, such as the Autoregressive Integrated Moving Average (ARIMA) model, which you'll have an opportunity to explore using Python in Achievement 6.**
    - **Do some research on the ARIMA model and one other model not covered in this Exercise; Facebook Prophet is one example that's become popular in recent years.**

ARIMA

https://www.ibm.com/think/topics/arima-model

ARIMA stands for Autoregressive Integrated Moving Average and it's a technique for time series analysis and for forecasting possible future values of a time series.

Autoregressive modeling and Moving Average modeling are two different approaches to forecasting time series data. ARIMA integrates these two approaches, hence the name.

https://www.datacamp.com/tutorial/arima?dc_referrer=https%3A%2F%2Fwww.google.com%2F

Autoregressive (AR) part

The Autoregressive (AR) component builds a trend from past values in the AR framework for predictive models. For clarification, the 'autoregression framework' works like a regression model where you use the lags of the time series' own past values as the regressors.

Integrated (I) part

The Integrated (I) part involves the differencing of the time series component keeping in mind that our time series should be stationary, which really means that the mean and variance should remain constant over a period of time. Basically, we subtract one observation from another so that trends and seasonality are eliminated. By performing differencing we get stationarity. This step is necessary because it helps the model fit the data and not the noise.

Moving average (MA) part

The moving average (MA) component focuses on the relationship between an observation and a residual error. Looking at how the present observation is related to those of the past errors, we can then infer some helpful information about any possible trend in our data.

We can consider the residuals among one of these errors, and the moving average model concept estimates or considers their impact on our latest observation. This is particularly useful for tracking and trapping short-term changes in the data or random shocks. In the (MA) part of a time series, we can gain valuable information about its behavior which in turn allows us to forecast and predict with greater accuracy.

Model identification

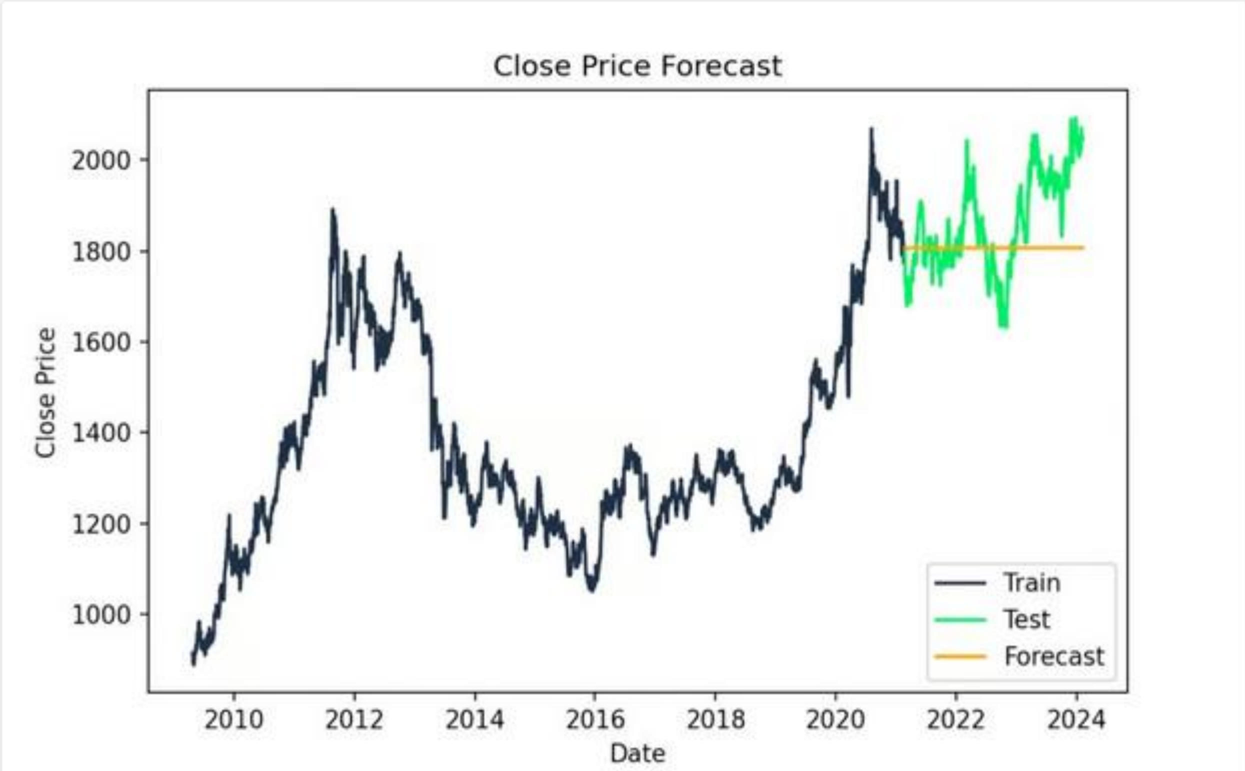
When we build an ARIMA model, we have to consider the *p*, *d*, and *q* terms that go into our ARIMA model.

- The first parameter, *p*, is the number of lagged observations. By considering *p*, we effectively determine how far back in time we go when trying to predict the current observation. We do this by looking at the autocorrelations of our time series, which are the correlations in our series at previous time lags.
- The second parameter, *d,* refers to the order of differencing, which we talked about. Again, differencing simply means finding the differences between consecutive timesteps. It is a way to make our data stationary, which means removing the trends and any changes in variance over time. *d* indicates differencing at which order you get a process stationary.
- The third parameter *q* refers to the order of the moving average (MA) part of the model. It represents the number of lagged forecast errors included in the model. Unlike a simple moving average, which smooths data, the moving average in ARIMA captures the relationship between an observation and the residual errors from a moving average model applied to lagged observations.

Finding the ARIMA terms

We use tools like ACF (Autocorrelation Function) and PACF (Partial Autocorrelation Function) to determine the values of *p*, *d*, and *q*. The number of lags where ACF cuts off is *q*, and where PACF cuts off is *p*. We also have to choose the appropriate value for *d* by creating a situation where, after differencing, the data resembles white noise. For our data, we choose 1 for both *p* and *q* because we see a significant spike in the first lag for each.

Ultimate output is something like this:



ARIMA forecast actual vs. predicted values. Image by Author

Facebook Prophet

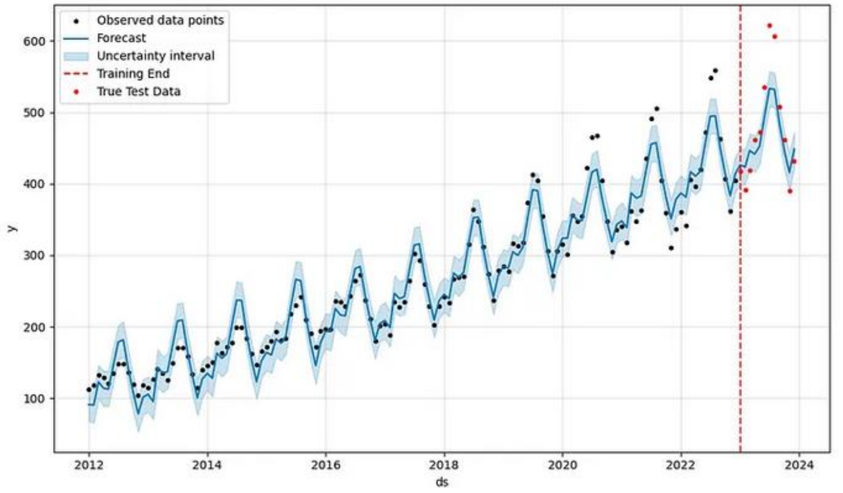
https://github.com/facebook/prophet

Prophet is a procedure for forecasting time series data based on an additive model where non-linear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects. It works best with time series that have strong seasonal effects and several seasons of historical data. Prophet is robust to missing data and shifts in the trend, and typically handles outliers well.

https://medium.com/data-science/getting-started-predicting-time-series-data-with-facebook-prophet-c74ad3040525

Prophet is particularly well-suited for business forecasting applications, and it has gained popularity due to its ease of use and effectiveness in handling a wide range of time series data. As with every tool, keep in mind that while Prophet is powerful, the choice of forecasting method depends on the specific characteristics of the data and the goals of the analysis. In general, it is not granted that Prophet performs better than other models. However, Prophet comes with some useful features e.g., a reflection of seasonality change pre- and post-COVID or treating lockdowns as one-off holidays.

Prophet can be imported into Python for use, and produce outputs a little like this:



Plotted results for the time series analysis incl. true test data and the prediction. Image by author

As we can see it handles seasonality and trends well.

- **Imagine you have to explain these models to a colleague who's unfamiliar with them. Write two short paragraphs (1 for each model) without going into the technical details. Include links to the resources you found during research.**

If you have some data that tells you what's happened over time in the past – say how many cars you have sold each month - you can use statistical modelling techniques to predict what is most likely to happen in the future.

We can use something like AIMRA. TO make that work your data really needs to be what we call "Stationary" – the average values don't really change over time. But if they do, there's a way to still make it work. Then it looks at all the previous individual values, and what the trend of those values has been over time. Combining what it can understand from both of those produces a prediction for the future – and that can be quite a complex, it'll predict what will happen each month, and that prediction should just as varied as the actual results.

Facebook have produced a product called Prophet which does a similar thing. It's different in a couple of ways. First, it's designed to handle data that isn't stationary and data that changes in predictable ways, like you always sell more cars when a new reg plate comes out. Second, it's a little bit more of a black box in terms of exactly how its doing what tit does, but there's loads of ways you can tweak it so that you're confident in the prediction.