**Time Series Analysis and Forecasting**

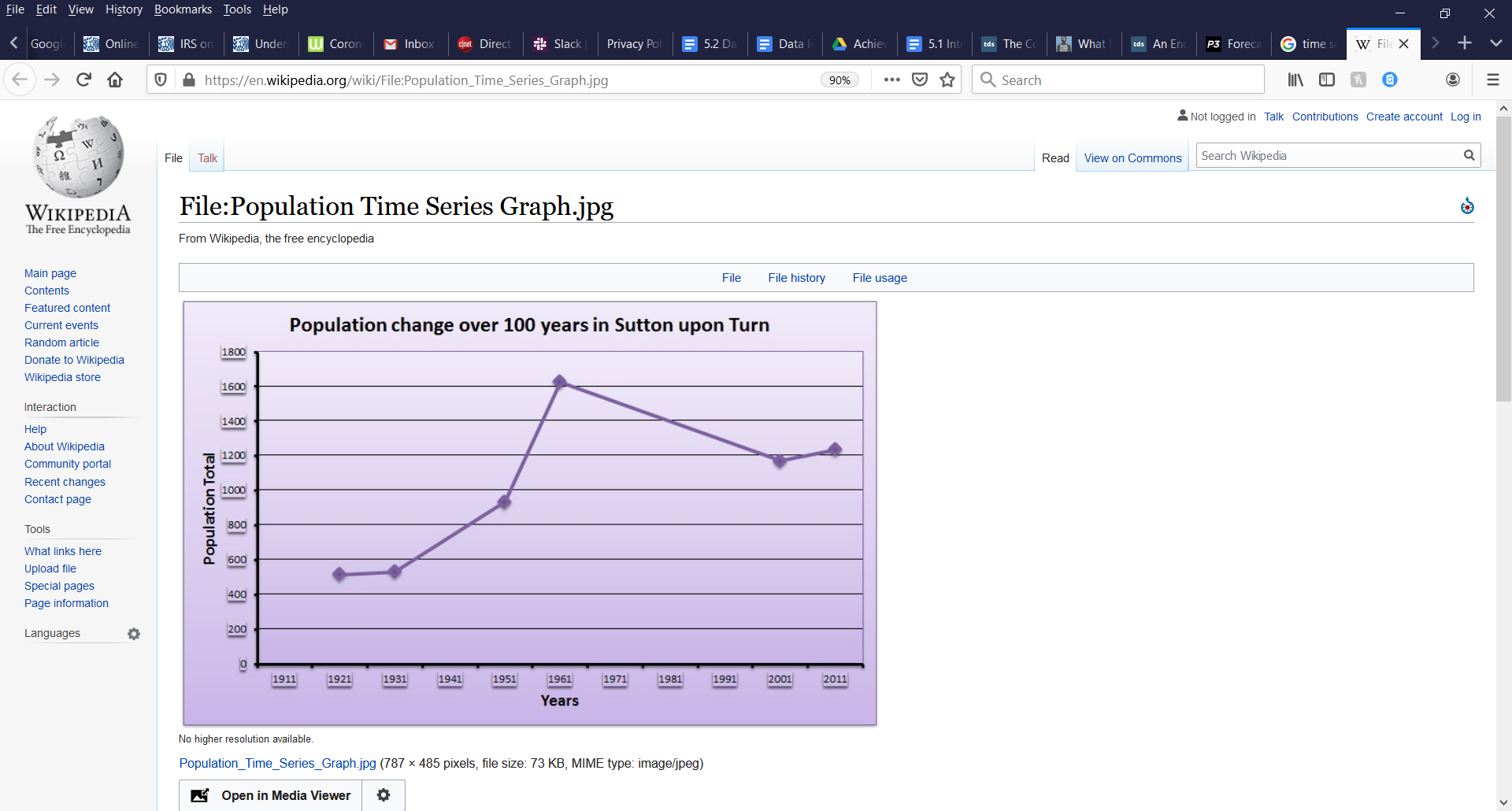
Learning Goals

* **Define forecasting and techniques of forecasting**
* **Discuss the Time Series forecasting technique and its application**
* **Explain the fundamental difference between forecasting and predictive analytics**
* **Describe time-series analysis**
* **Describe the role of forecasting in an analysis project**
* **Statistical approach to forecasting**

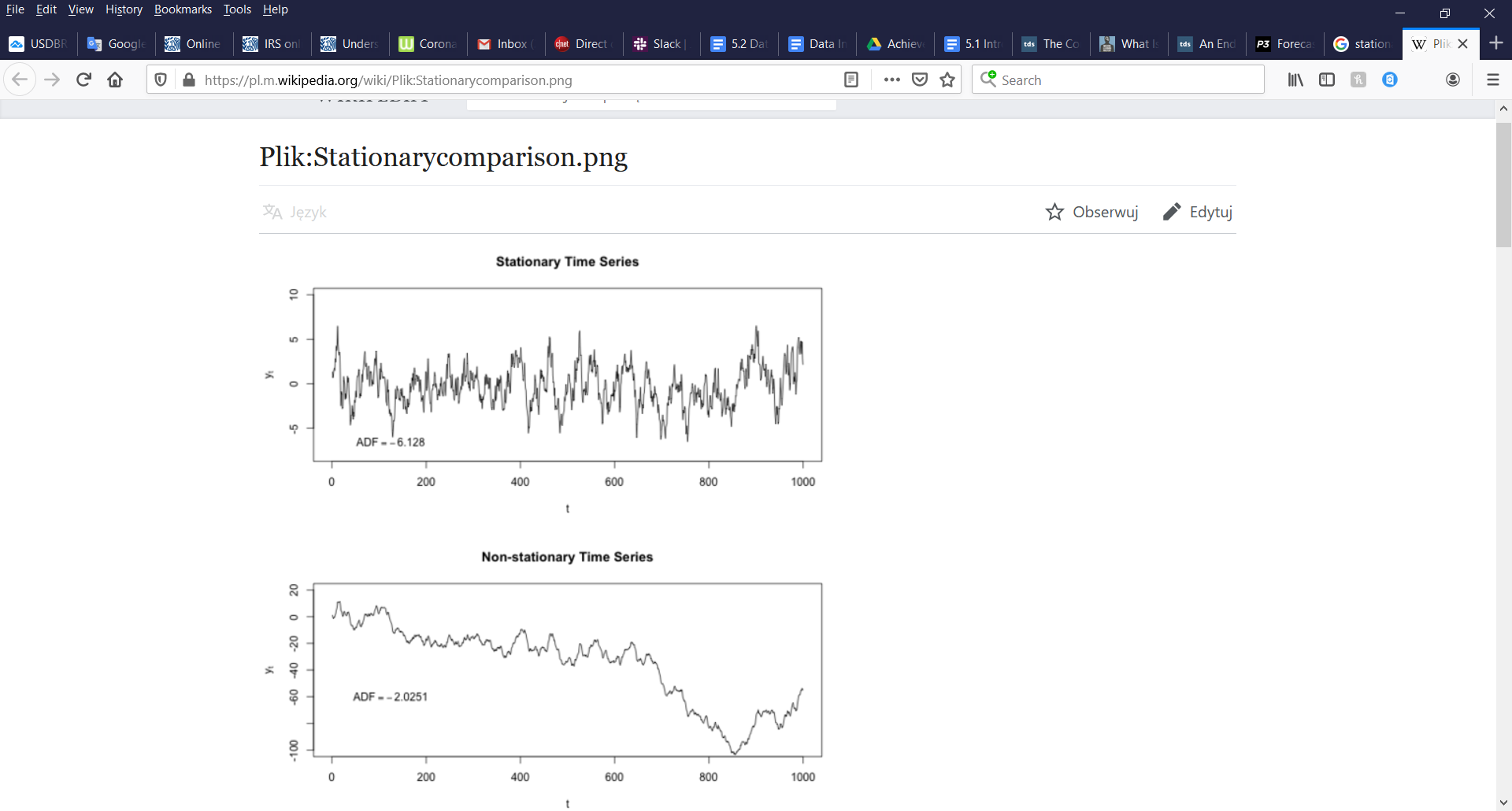
Introduction

It has been said in order to see the future, we first look to the past. Looking to the past provides a foundational point from which we can then measure the distance and direction to the present in order to anticipate a likely future. That is essentially what time series analysis and forecasting is all about. With predictive analysis, you learned some techniques and algorithms designed to anticipate a single action or outcome based on given variables. Time series analysis and forecasting may sound like similar concepts to predictive analysis, and they are! But the key difference is that time series and forecasting are used to predict broad level macro trends and patterns within larger populations over longer amounts of time. In this achievement we will define and discuss the concepts and techniques of time series analysis and forecasting. You will see examples of time series analysis and forecasting and their application in multiple industries. More importantly, you will develop an understanding of the technical skills needed for a data analyst to perform time series analysis and forecasting.

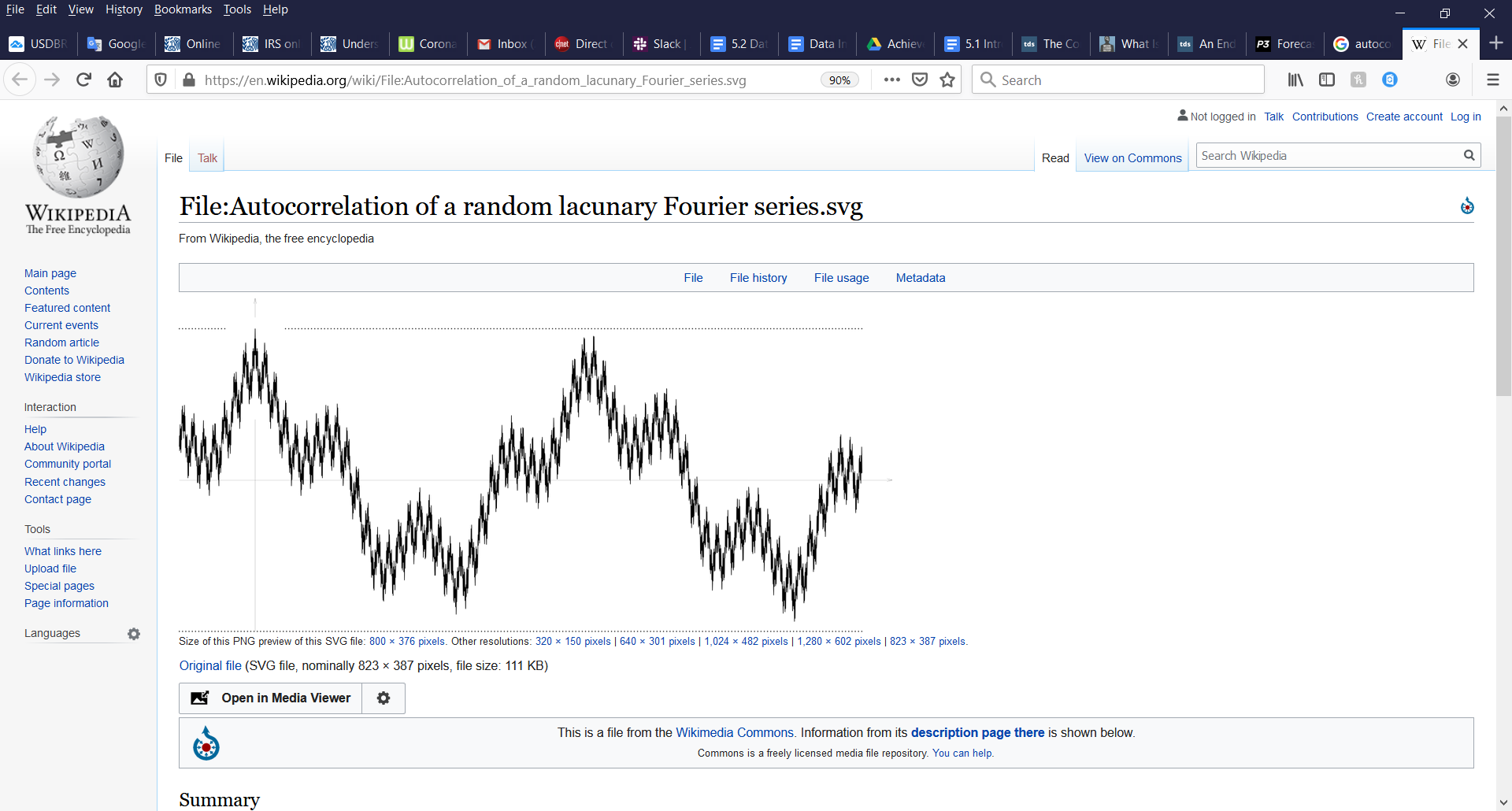
Time Series

A time series is represented by a set of data points arranged in order by time of occurrence. When the data points are arranged by time and built into a graphic representation, the user may then visualize the data to recognize the direction, or trend in the data being measured. Whereas other statistical methods measure the relationship between two chosen variables, the time series always has time as the independent variable. In the image to the right, the dependent variable is “Population Total” and the independent variable is time, in this case measured in years. The goal of this time series is to reveal the trend in total population over time so that a forecast can be made.

The main aspects that are relevant to interpreting the trend are stationarity, seasonality, and autocorrelation. Stationarity occurs in a time series when the statistical attributes do not vary over time. In other words, the means and variance remain the same. Stationarity is easily recognized in a visual as the lack of change in variance and mean over time is evident.



An example of autocorrelation is shown below. An autocorrelation plot occurs when data point values are always equal at a given time interval, which reflects in the plot by showing a repeating pattern.



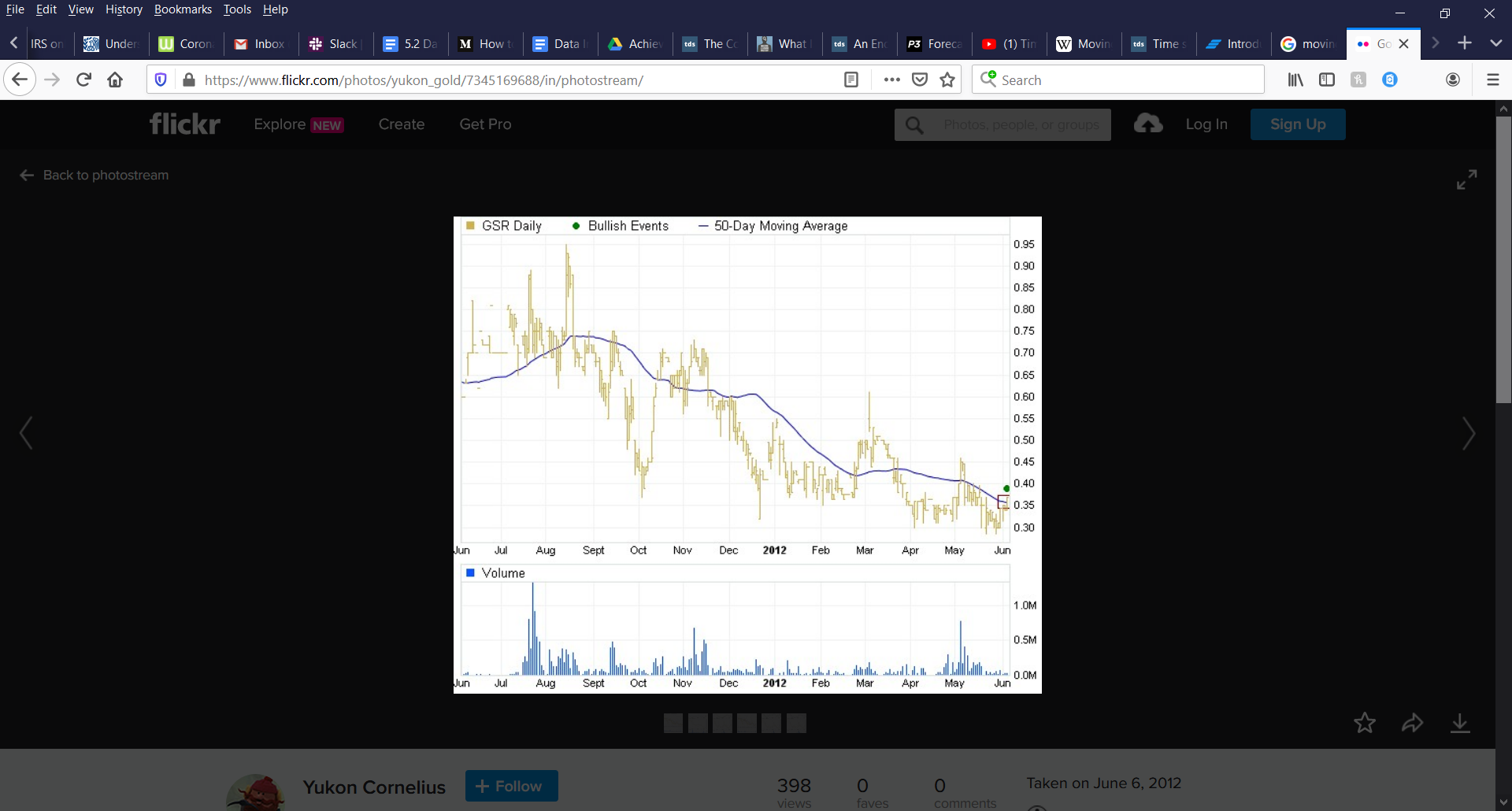
Regression analysis is a common forecasting technique; however, it cannot be applied to autocorrelated data because the regression assumes the data isn’t autocorrelated.

Seasonality refers to fluctuations that occur at periodic intervals. For example, seasonality is commonly observed in retail sales during the Christmas holiday season sales peak, then decline in the following months. Seasonality is also easily observed in the fluctuations in temperatures from season to season, hence the term seasonality.

Forecasting

Forecasting may be a synonym to predicting in the English language, but when it comes to data analytics these are two very distinct concepts. As stated in the introduction, predictive analysis attempts to predict a single individual or entity will do on a very short timeline. Forecasting with time series seeks to use identify broader trends over longer periods of time to anticipate the actions of larger populations. Think of predicting as micro and forecasting as macro in terms of events and trends they seek to discover. Time series forecasting seeks to make a model that fits historical data and subsequently uses the model to forecast future values. For example, NETFLIX may seek to forecast their revenue for the next quarter and would likely create a time series showing their last several quarterly revenue incomes, then use statistics and time to model a forecast. But in order to predict what new films or series you are likely to watch next, they would have to build a predictive model. Forecasting has tremendous utility in the business and science industries by helping decision makers plan and formulate broader strategies.

Moving Averages

The moving average is a forecasting technique used for modeling a single variable. The moving average seeks to “smooth” out the fluctuations between time intervals by averaging a subset of a numerical series in a given time frame. Moving averages are commonly used by financial analysts as a technical analysis tool to identify price action trends in securities or other financial assets. In the illustration to the right, Goldstrike Resources historical stock price is represented by the yellow line. The blue line represents the moving average of the stock price. In this case, it’s a 50 day moving average. It is computed by finding the average closing price of the previous 50 days of closing prices. For longer term price forecasts, you would use a moving average with a greater number of days. Modeling data into a moving average can better reveal the overall trend as it smooths out the day to day volatility in stock price action. The moving average works to give the financial analyst insight into whether the security is overpriced or cheap by where the price is at any given time in relation to the moving average. Or a moving average could be just one variable within a larger model that the financial analyst uses to decide to buy or sell an asset.

Weighted Moving Averages

Weighted moving averages work very similarly to the moving averages. The difference is that in a weighted moving average, the most recent data is weighted more heavily than older data in the time series, whereas in a simple moving average, the data points at each time interval are all weighted equally. On a 30 day weighted moving average, you may choose to weigh most recent 15 days by 1.5 and the older data by 0.5. How much more the data analyst weighs the most recent data is dependent on the subject matter of what you are modeling, but the sum of the weights should be 1.

Exponential Smoothing

Exponential smoothing forecasting technique can support data with a lack of obvious trend or pattern. Exponential smoothing is similar to a weighted moving average in that it forecasts based on the sum of weighted historical data points, but it uses a formula to exponentially decrease the weights of past data points with time. As with the weighted moving average, the sum of previous weights should equal 1. Due to the lack of seasonality or pattern, the exponential smoothing technique is preferred, but is better for short term forecasts.

Application of Time Series Analysis and Forecasting

Time Series Analysis and Forecasting are used extensively in economics, science, supply chain, medicine and many others. The time dimension of this analytical technique not only guides planning, but allows one to make correlations between events, time and their impact on a variable. The two examples provided below show how time series are used in different industries.

Inventory Optimization

Keeping inventory at an optimal level has a large impact on a business’s bottom line. Businesses that sell massive amounts of goods incur the expenses of keeping that inventory stored somewhere. They also miss out on revenue when they run out of a good before resupplying. Inventory optimization is a business concept that uses time series analysis and forecasting to avoid these kinds of inventory issues. In the example time series chart below, a local pharmacy can see the sales data of two products for which it keeps an inventory. Using time series, the pharmacy can easily identify the seasonality impacts on the product sales of cough drops and sun screen. The pharmacy now has a better idea of how much of each product they should be keeping in stock depending on the time of year.

Environment

Air pollution is a global health issue. The World Health Organization estimates that 4.6 million people die per year from health conditions attributable to air pollution. As the world population grows and more areas of the world industrialize, the amount of air pollution would increase if not for legislation that limits the amount of harmful pollutants put into the air by automobiles and industry. Mexico City, due to being surrounded by mountains and being one of the largest cities in the world has long suffered from the effects of air pollution. Scientists and legislators seek to know if the laws they work to implement are having the desired effects and time series analysis does exactly that.

With time series analysis, scientists can help to validate the efficiency of air pollution laws by tracking a variable such as deaths per 100,000 per year. They can also use this data to predict future air quality levels and numbers of people expected to develop health conditions related to air quality assuming the legislation remains in place.

Conclusion

The time series is one of the most common but overlooked areas of data analysis. Modeling a variable’s relationship with time helps us to identify trends, plan and evaluate the impact of policies and business decisions. In this achievement, we have discussed the concepts of time series, analyzing and modeling time series data, and using common statistics to make forecasts. You have learned how stationarity, autocorrelation, and seasonality can affect your analysis and forecast. Most importantly, you have seen where and how different industries use and analyze time series to make decisions and measure the impact of those decisions. The simplicity of the time series/forecasting combined with the positive impact on planning, deciding, and evaluating ensure the skills and techniques of time series analysis will be an important part of your data analytics skills for much time to come.

Resources

<https://powerpivotpro.com/2017/08/difference-forecasting-predictive-analytics/>

<https://towardsdatascience.com/the-complete-guide-to-time-series-analysis-and-forecasting-70d476bfe775>

<https://www.influxdata.com/blog/autocorrelation-in-time-series-data/>

<https://medium.com/data-dump/supply-chain-advanced-analytics-hacking-customer-demand-forecasting-inventory-optimization-201f9224cf2f>

<https://ourworldindata.org/air-pollution>

<https://www.sciencedaily.com/terms/air_pollution.htm>

Images

<https://en.wikipedia.org/wiki/File:Population_Time_Series_Graph.jpg>

<https://pl.m.wikipedia.org/wiki/Plik:Stationarycomparison.png>

<https://en.wikipedia.org/wiki/File:Autocorrelation_of_a_random_lacunary_Fourier_series.svg>

<https://www.flickr.com/photos/yukon_gold/7345169688/in/photostream/>