GENERAL ASSEMBLY

AUTOCORRELATION AND TIME SERIES DATA

Joseph Nelson, Data Science Immersive

AGENDA

- Time Series Data Quick Review
- Trend and Seasonality
- Autocorrelation
- Code Along

TIME SERIES DATA

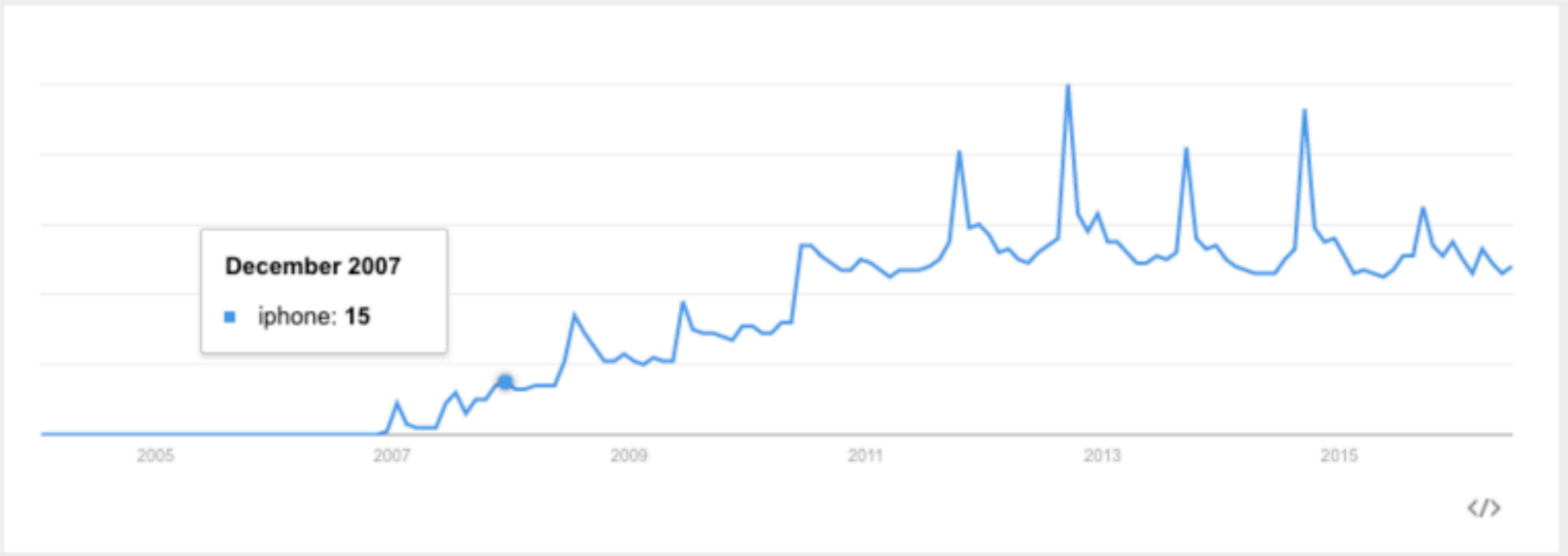
- A univariate time series is a sequence of measurements of the same variable collected over time. Most often, the measurements are made at regular time intervals.
- What are some real world scenarios where Time Series Data Analysis is useful?

TREND AND SEASONANLITY

- What constitutes a trend in data?
- Is Linearity required for trend?

TREND AND SEASONANLITY

- What constitutes a trend in data?
- Is Linearity required for trend?
- Trend may "change direction" when it goes from an increasing trend to a decreasing trend. Trend can only be measured in the scope of the data collected, though there may be trends that are un-measureable if the data is not complete.



TREND AND SEASONANLITY

- **When there are patterns that repeat over known, fixed periods** of time within the data set it is considered to be seasonality, seasonal variation, periodic variation, or periodic fluctuations (all different terms, but they mean the same thing).
- A seasonal pattern exists when a series is influenced by factors relating to the cyclic nature of time - i.e. time of month, quarter, year, etc.
- Seasonality is always of a fixed and known period, otherwise it is not truly seasonality, and must be either attributed to another factor or counted as a set of anomalous events in the data.

AUTOCORRELATION

- While in previous weeks, our analyses has been concerned with the correlation between two or more variables (height and weight, education and salary, etc.), in time series data, autocorrelation is a measure of how correlated a variable is with itself. Specifically, autocorrelation measures how closely related earlier variables are with variables occurring later in time.

$$r_k = \frac{\sum_{t=k+1}^{n} (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^{n} (y_t - \bar{y})^2}$$

AUTOCORRELATION

- To compute autocorrelation, we fix a lag k which is the delta between the given point and the prior point used to compute the correlation.
- With a k value of 1, we'd compute how correlated a value is with the prior one. With a k value of 10, we'd compute how correlated a variable is with one 10 time points earlier.
- We will be using d gstore, Rossmann. This data contains the well as whether a sale or holiday affected th

$$r_k = \frac{\sum_{t=k+1}^{n} (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^{n} (y_t - \bar{y})^2}$$

CODE ALONG

- We will be using data made available by a German drugstore, Rossmann. This data contains the daily sales made at the drugstore as well as whether a sale or holiday affected the sales data.