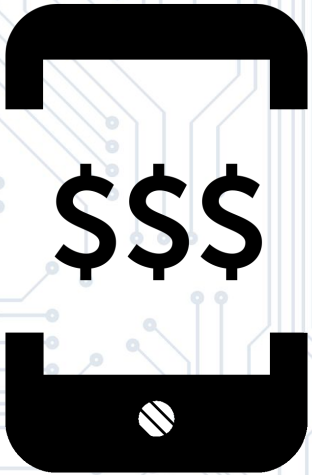


Intro to Machine Learning and Time Series Analysis

FinTech

Lesson 10.1

Class objectives

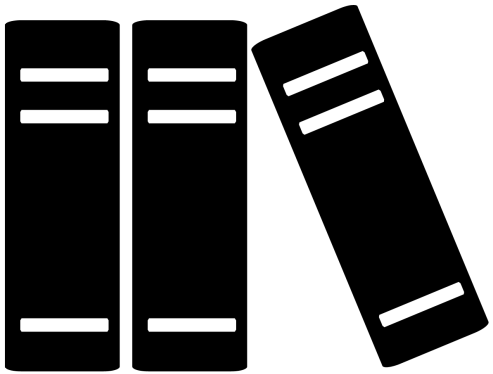
By the end of today's class you will understand:

- Intro to Machine Learning
- Time Series Fundamentals
- Time Series Decomposition
- Hodrick-Prescott Filter
- Autocorrelation and Partial Autocorrelation Analysis

Mysticism of Machine Learning

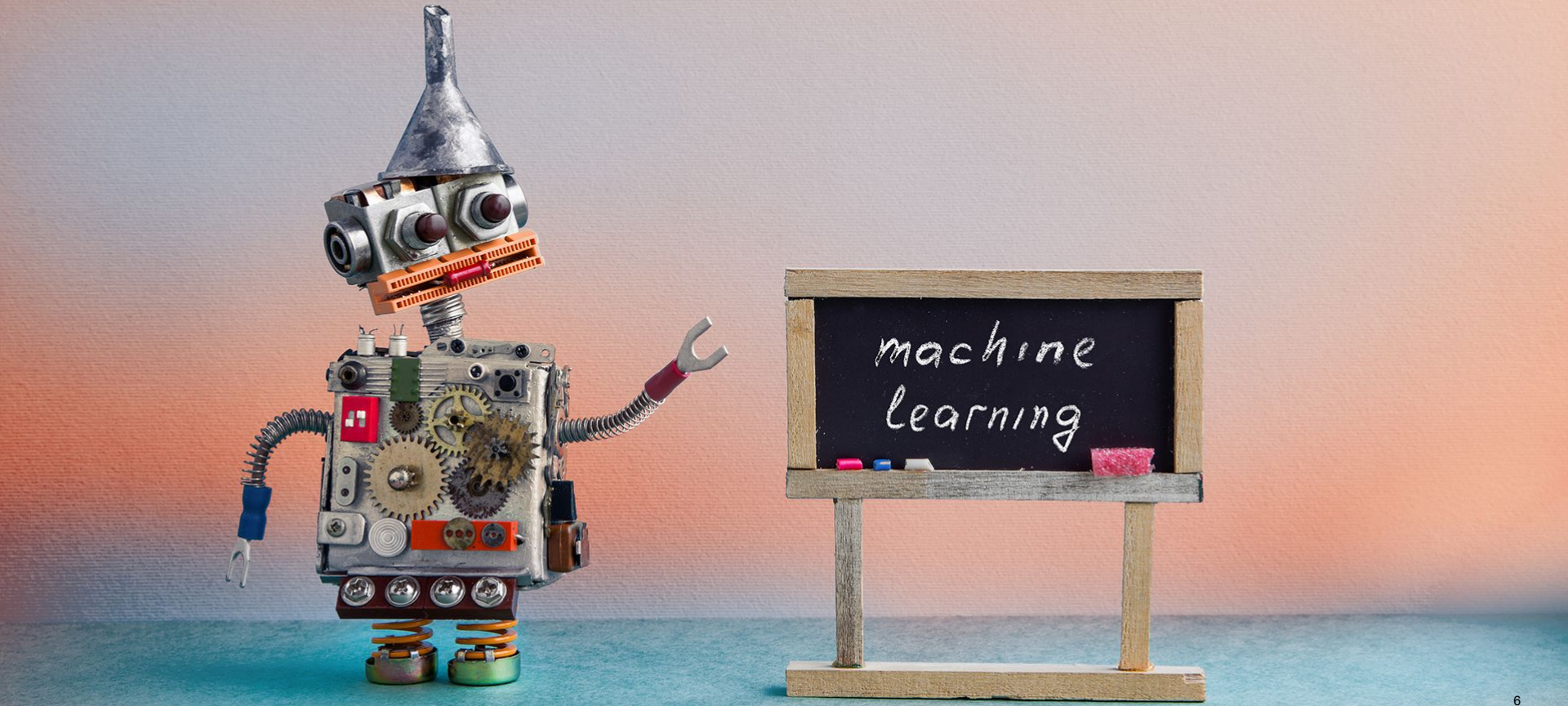
Machine Learning in a Nutshell

Libraries



Statistics

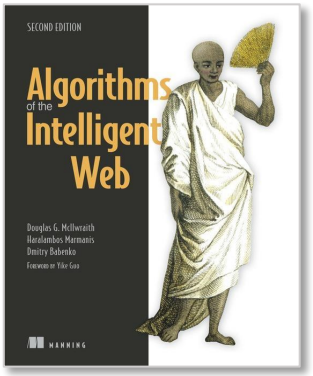
So It Begins...



Basic Definitions

Intelligent Algorithms (Definition)

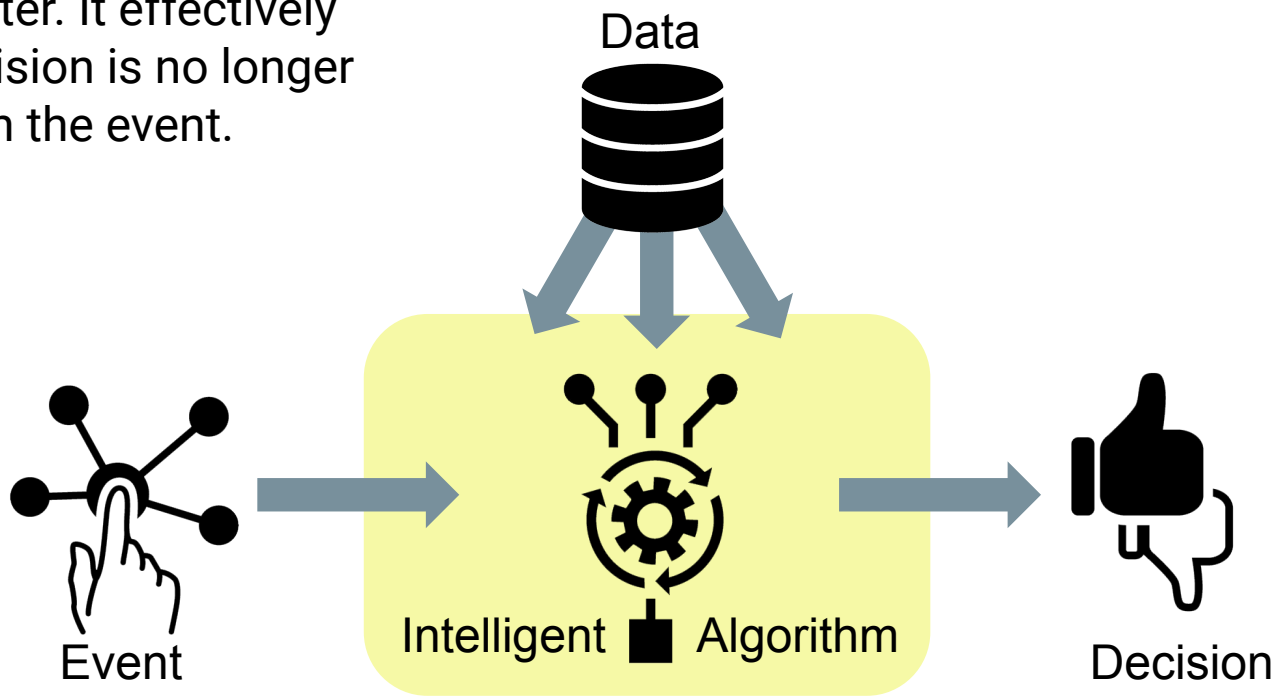
> Intelligent algorithms are ones that use data to modify its behavior. Intelligent algorithms differ in that they can change their behavior as they run, often resulting in a user experience that many would say is intelligent.
>
> —*Algorithms of the Intelligent Web, Second Edition*



Algorithms of the Intelligent Web, Second Edition
by *Douglas G. McIlwraith Haralambos Marmanis Dmitry Babenko*
Publisher: Manning Publications
Release Date: August 2016

Intelligent Algorithms (Diagram)

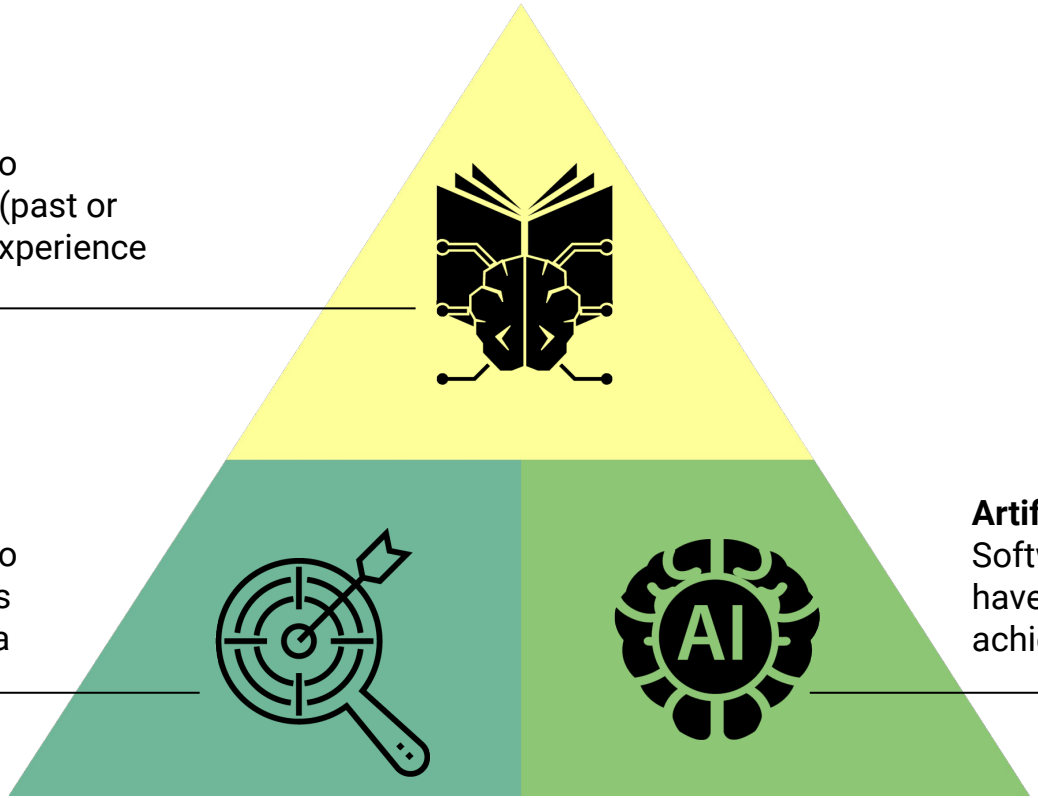
Intelligent algorithms are ones that respond to data such that the algorithm gets better. It effectively “evolves.” The decision is no longer deterministic given the event.



Intelligent Algorithms (Triad)

Machine Learning
Capability of software to generalize phenomena (past or future) based on past experience

Predictive Analytics
Capability of software to predict future outcomes based on historical data

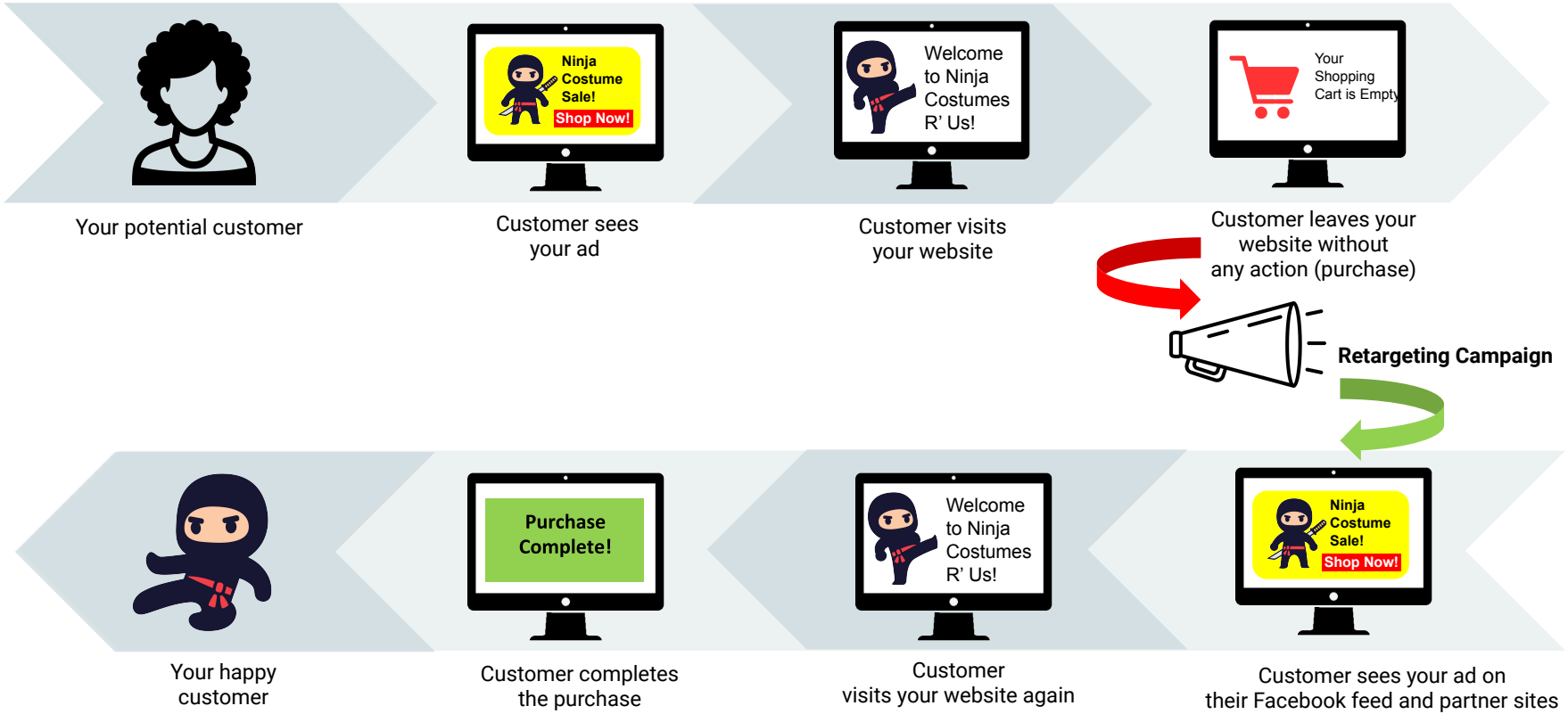


Artificial Intelligence
Software (and machines) that have a series of options to achieve a particular goal

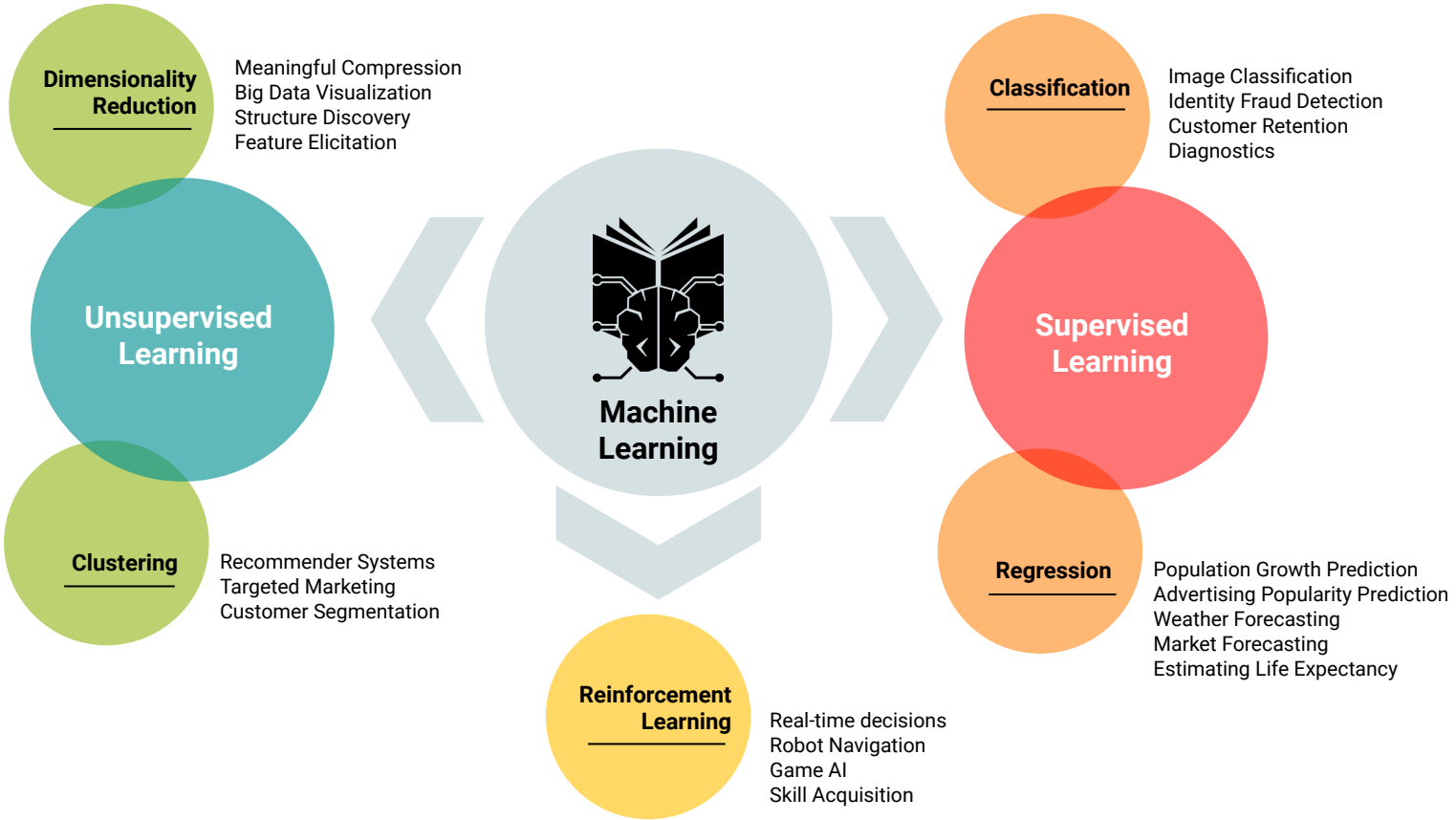
Artificial Intelligence (Example)

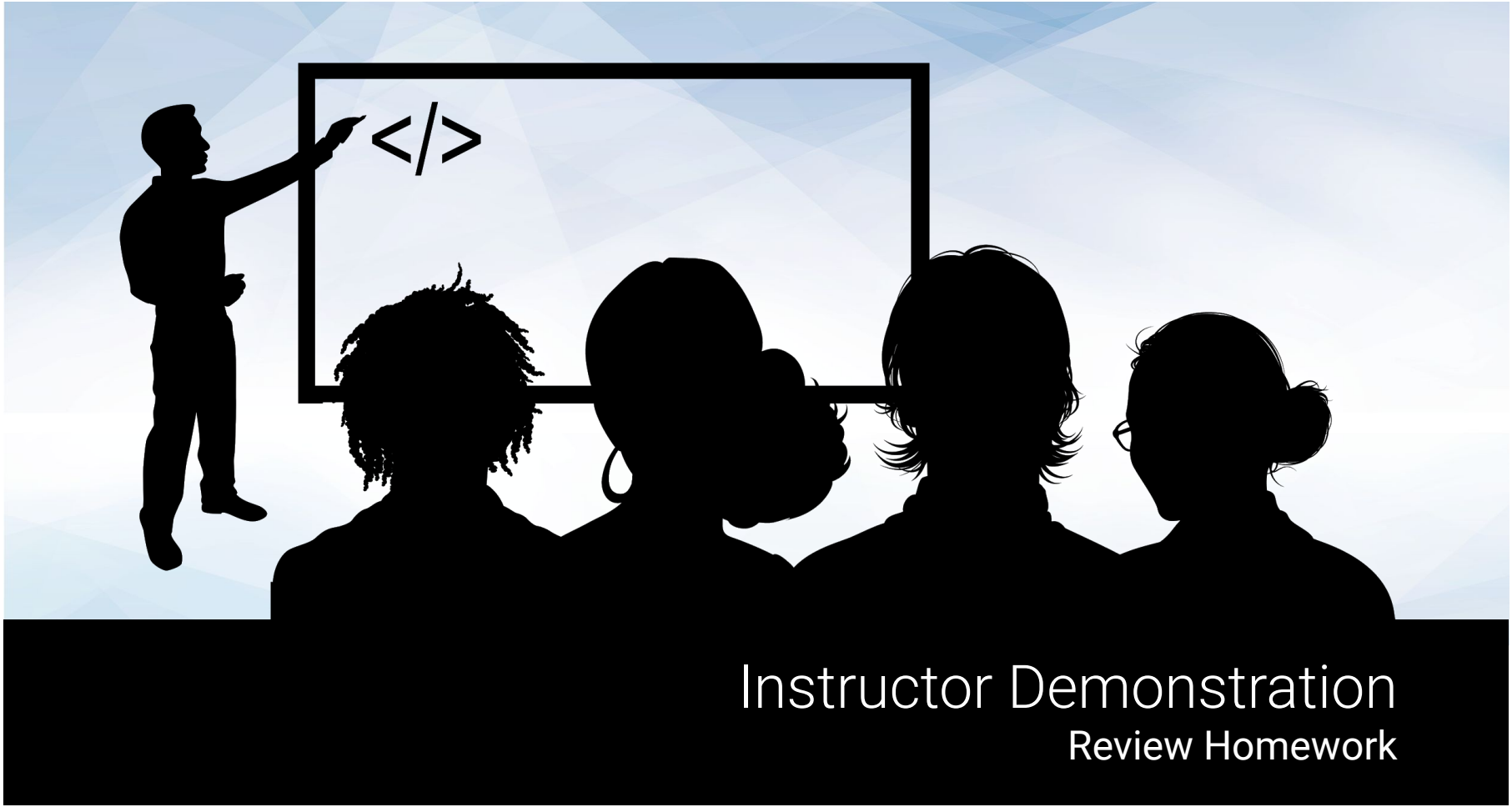
Predictive Analytics (Example)

How retargeting ads work:



Machine Learning (Categories)





Instructor Demonstration

Review Homework

Time Series Basics

Time Series Basics

Working with time series data requires a return to the basics

Data needs to be sliced and diced at various time frequencies in order to analyze data points as a time series

E.g day, week, month, year

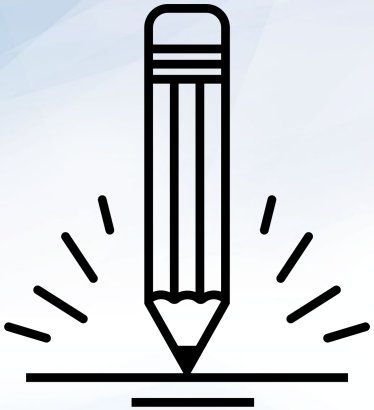
Pandas DateTimeIndex index can be used to help with this

```
df.loc[2019]
```

Time Series Basics

The Pandas resample function can also be used to slice and dice data, once a DateTimeIndex has been created

```
weekly = df['Close'].resample('W').mean()
```



Activity: Time Series Basics

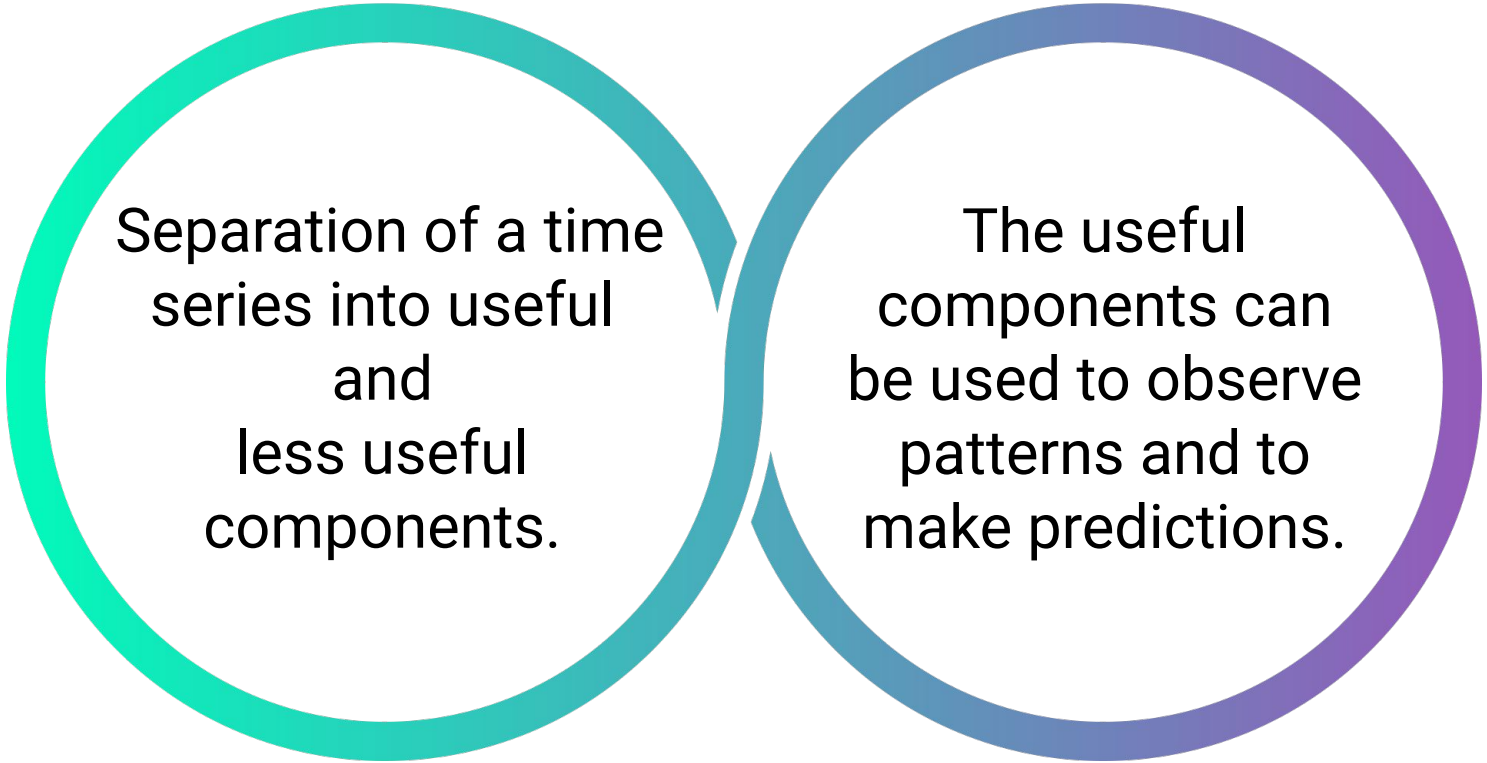
Instructions sent via Slack.

Suggested Time:
15 minutes

Time’s Up! Let’s Review.

Time Series Decomposition

Time Series Decomposition



Time Series Components

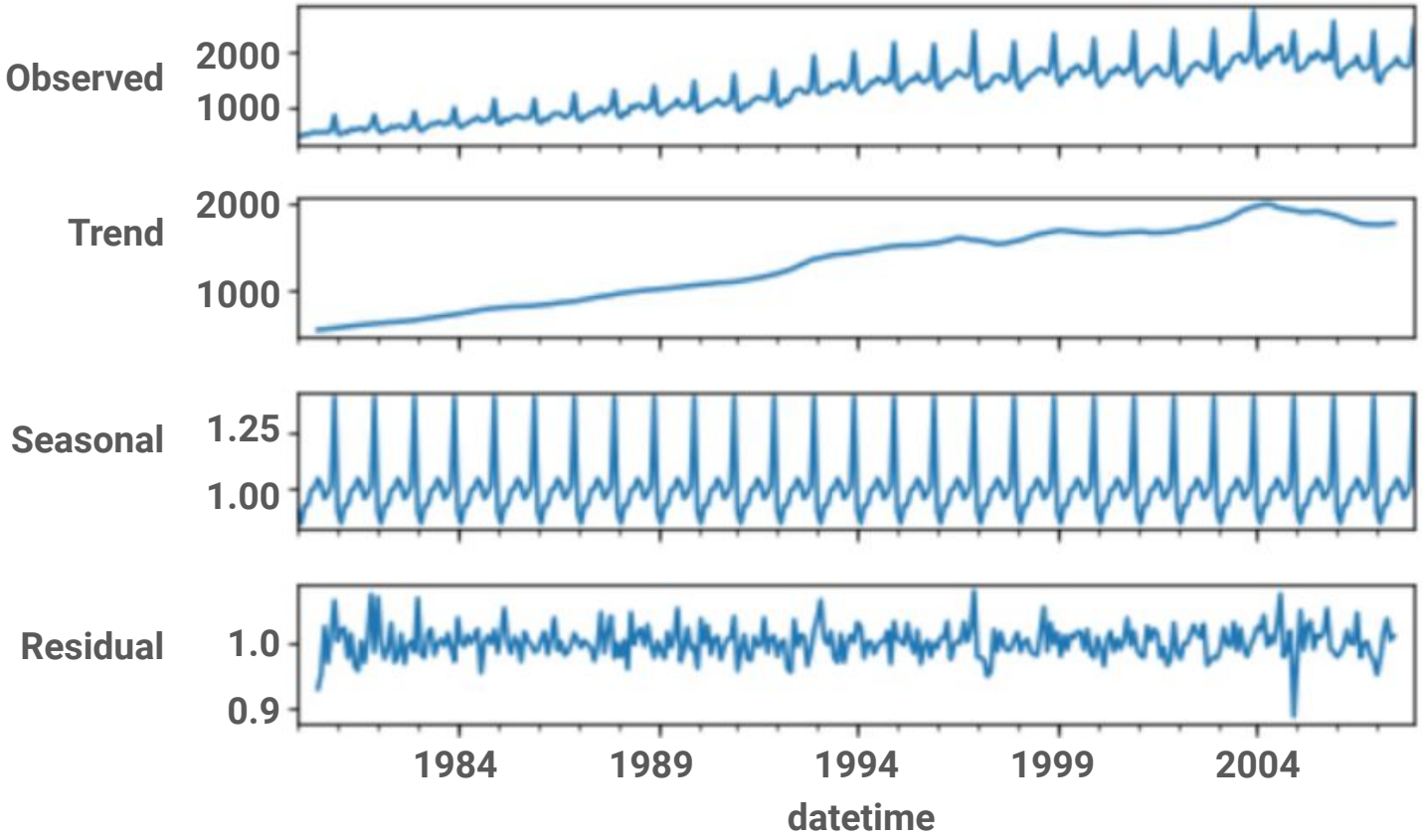
01 **Level:** What is the average value of the series?

02 **Trend:** Is there an overall direction of movement?

03 **Periodicity:** Do patterns occur in cycles?

04 **Residual:** How much noise exists in the data?

Time Series Decomposition in Action



Time Series Decomposition in Action

- The `observed data` panel is decomposed into the next three elements.
- An upward trend is observed in the data.
- A seasonality is also observed.
- The residual components are the leftovers when trend and seasonality are removed.

AdobeStock_56885471

Exponentially-Weighted Moving Average

Exponentially Weighted Moving Average (EWMA)

EWMA is an approach used to “denoise” or “smooth” out time series data so that trends and predictions can be made

01

EWMA involves calculating the average of the last n prices

02

Weights are added to the averages based on the recency of the data

- Recent data is weighted more heavily
- Weighting decreases exponentially for previous prices/time periods

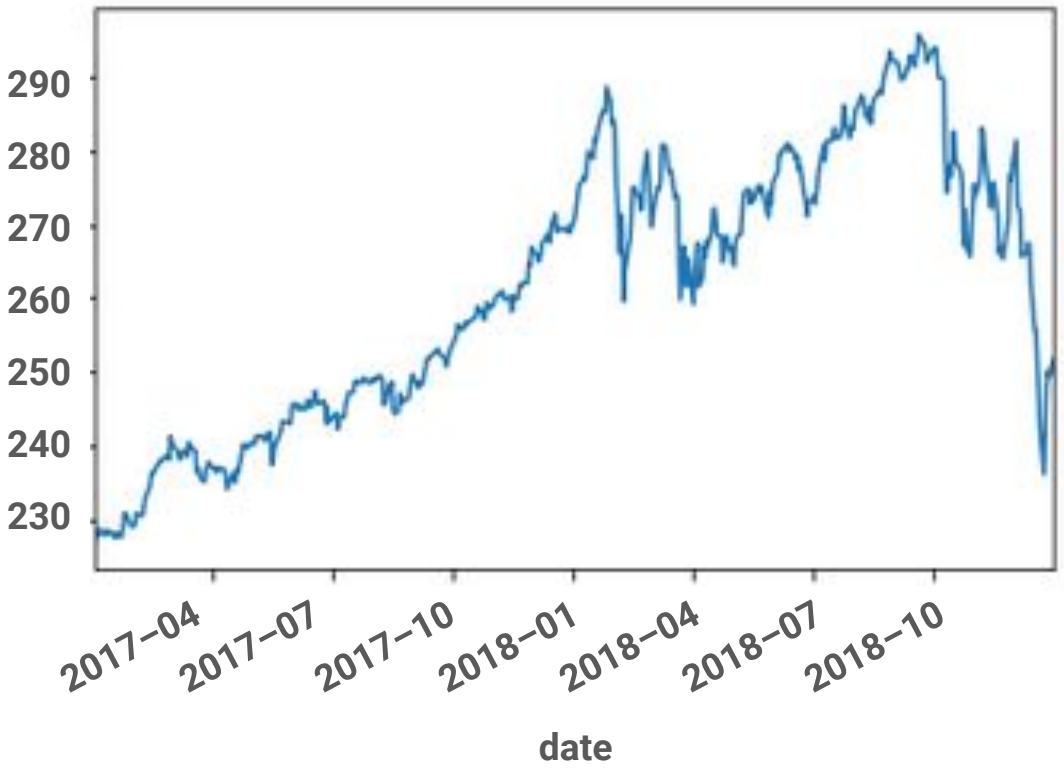
03

Requires past average values to be stored in memory

Exponentially Weighted Moving Average (EWMA)

EWMA is used to highlight trends and illustrate the price trajectory for an investment

In which direction is the price moving?



Hodrick-Prescott Filter

Hodrick-Prescott Filter: A mathematical function that separates a time series into trend and non-trend components.

Hodrick-Prescott Filter

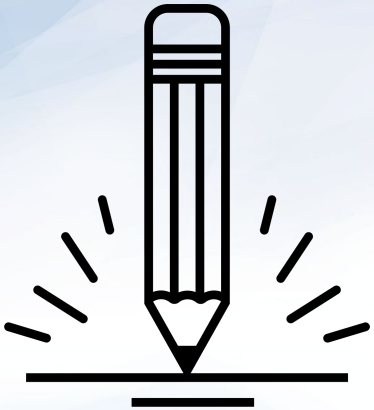
Filters out short-term fluctuations

Hodrick-Prescott Filter

$$\min_{\tau} \left(\sum_{t=1}^{T} (y_t - \tau_t)^2 + \lambda \sum_{t=2}^{T-1} [(\tau_{t+1} - \tau_t) - (\tau_t - \tau_{t-1})]^2 \right)$$

time series value - trend = cyclic element

Difference in trend over time = volatility



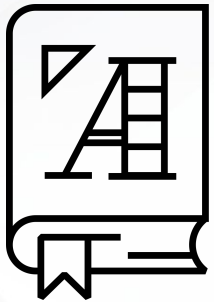
Activity: You've got a FRED

Instructions sent via Slack.

Suggested Time:
15 minutes

Time’s Up! Let’s Review.

Autocorrelation

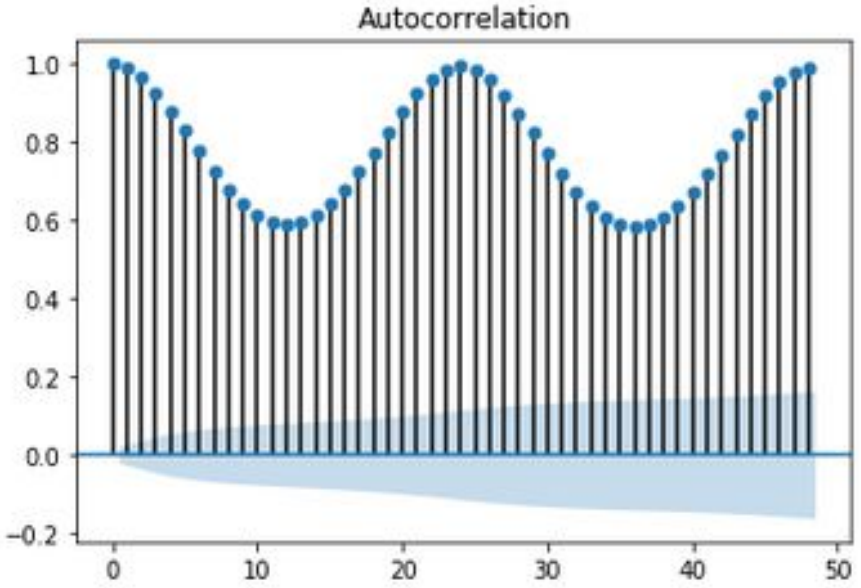


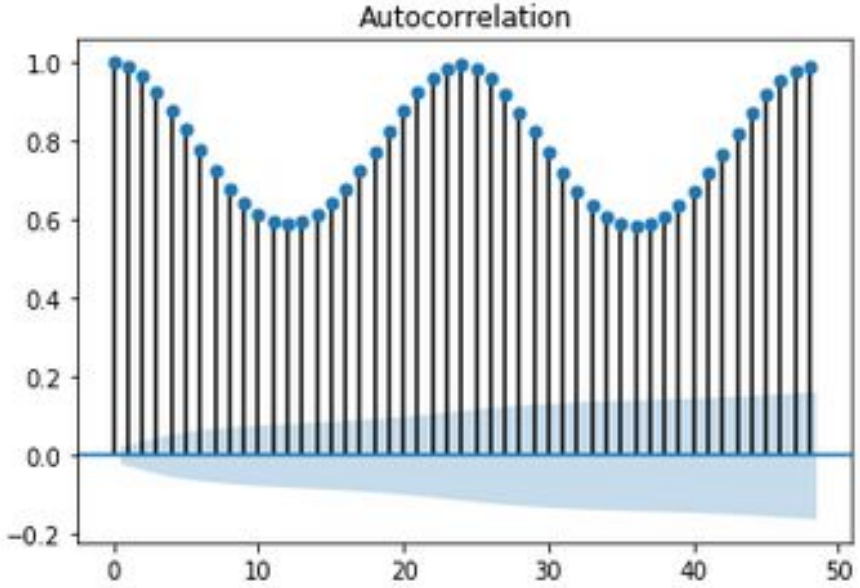
Autocorrelation is a measure of how closely current values correlate with past values

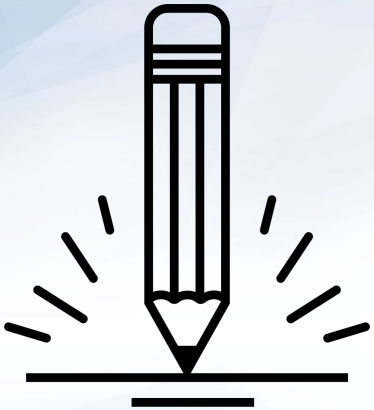
For example, **autocorrelation** is used to determine to what extent today's prices correlate with yesterday's prices

Autocorrelation

```
df.Temperature.autocorr(lag=1)
sm.graphics.tsaplots.plot_acf(df.Temperature,
```







Activity: Euro ETFs

Instructions sent via Slack.

Suggested Time:
15 minutes

Time’s Up! Let’s Review.

Questions?