Intro to Machine Learning and Time Series Analysis

FinTech

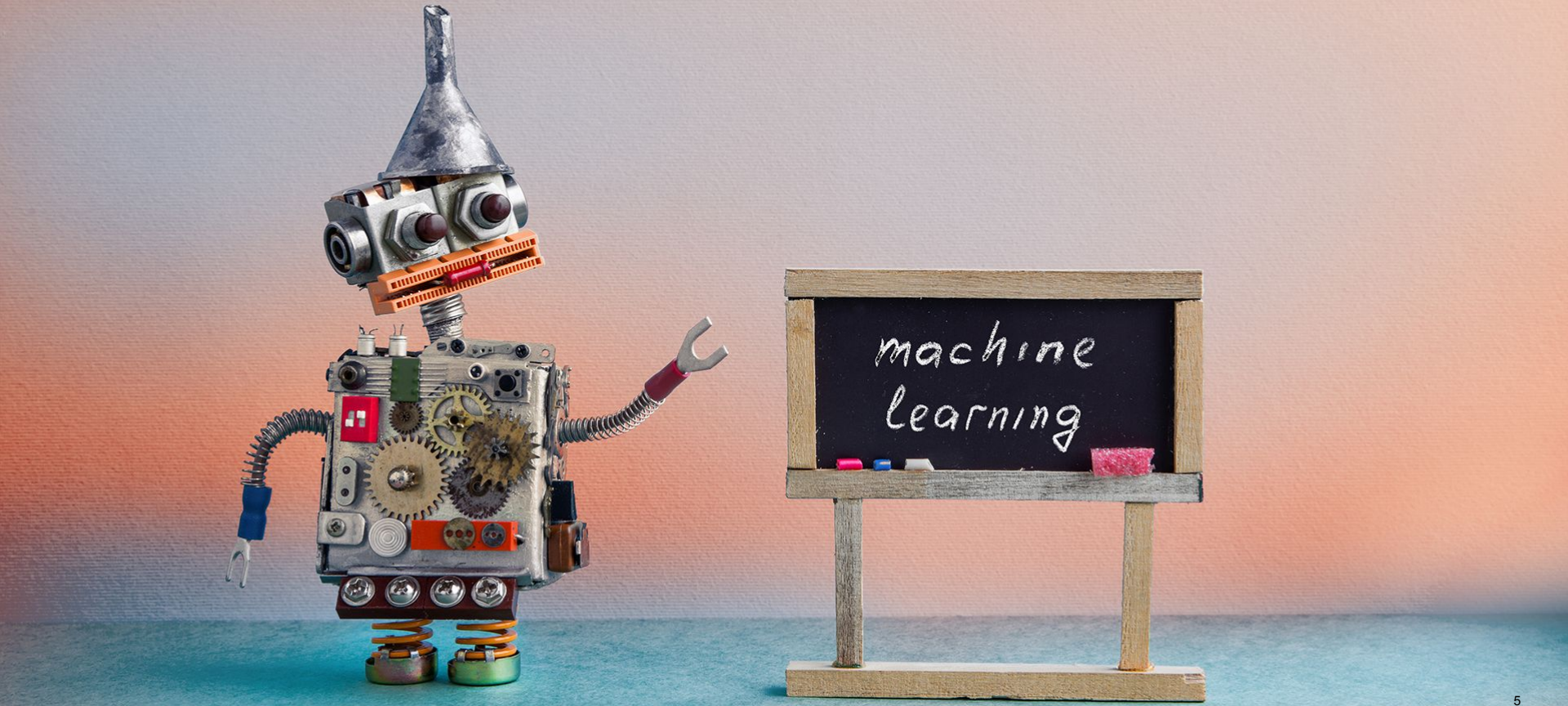
Lesson 10.1

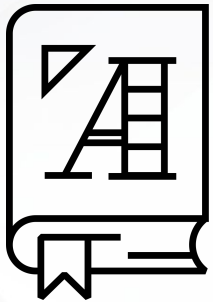
Class Objectives

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

Mysticism of Machine Learning

So It Begins...





Machine learning is an approach to programming that allows applications to learn from their inputs and make adjustments based on their outputs.

In a nutshell, **machine learning** develops statistical models that can make predictions or decisions on new data automatically.

Mysticism of Machine Learning

Machine learning can be used to predict:

- Loan eligibility, foreclosure rates, and fraud
- Disease diagnosis and prognosis
- Consumer segmentation and clustering
- Presidential election results
- Natural disaster and planetary climate impacts

Mysticism of Machine Learning

Machine Learning Models

01 Libraries

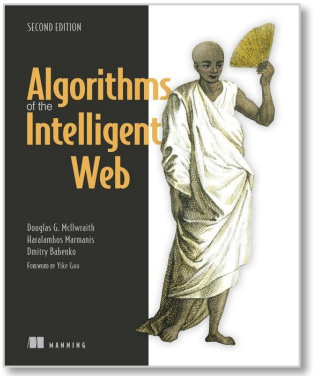
Models for machine learning are provided in libraries, just like other code we've used.

02 Pipelines

All machine learning pipelines use a **Model -> Fit-> Predict** paradigm. Once the model is fit, it can be used to make predictions.

Mysticism of Machine Learning

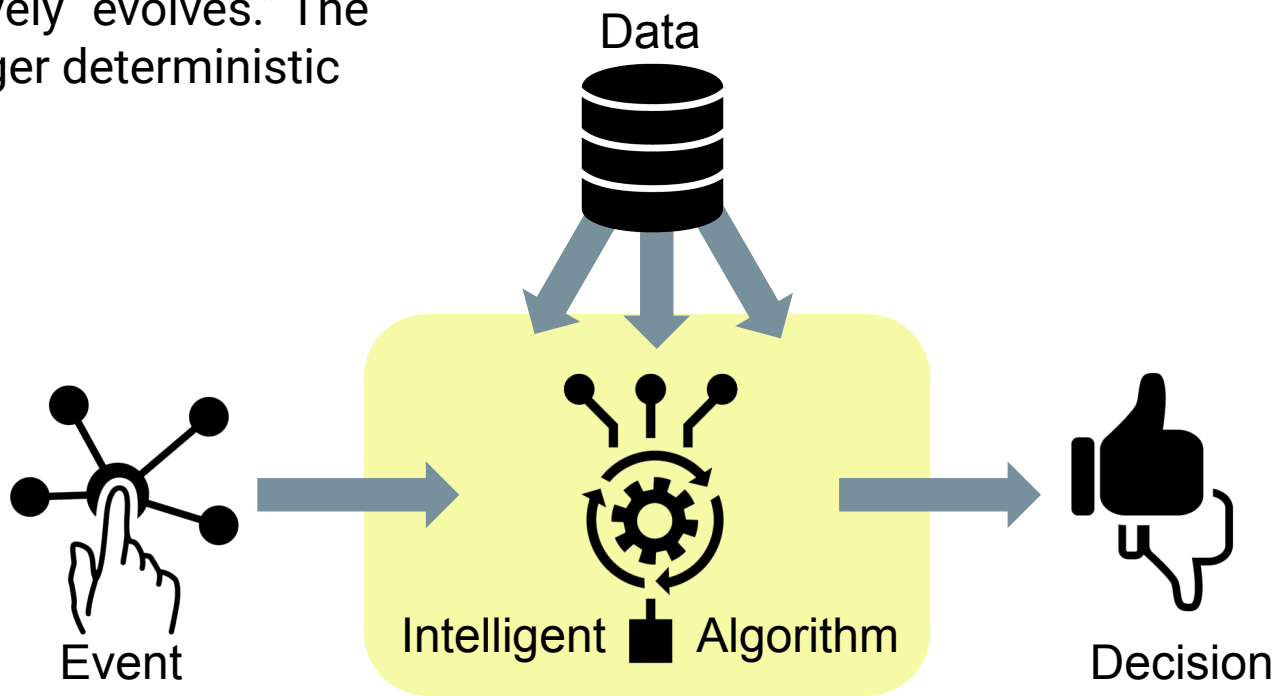
> **Intelligent algorithms** are ones that use data to modify its behavior. Intelligent algorithms differ in that they can change their behavior as they run, often resulting in a user experience that many would say is intelligent.
>
> *—Algorithms of the Intelligent Web, Second Edition*



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

Mysticism of Machine Learning

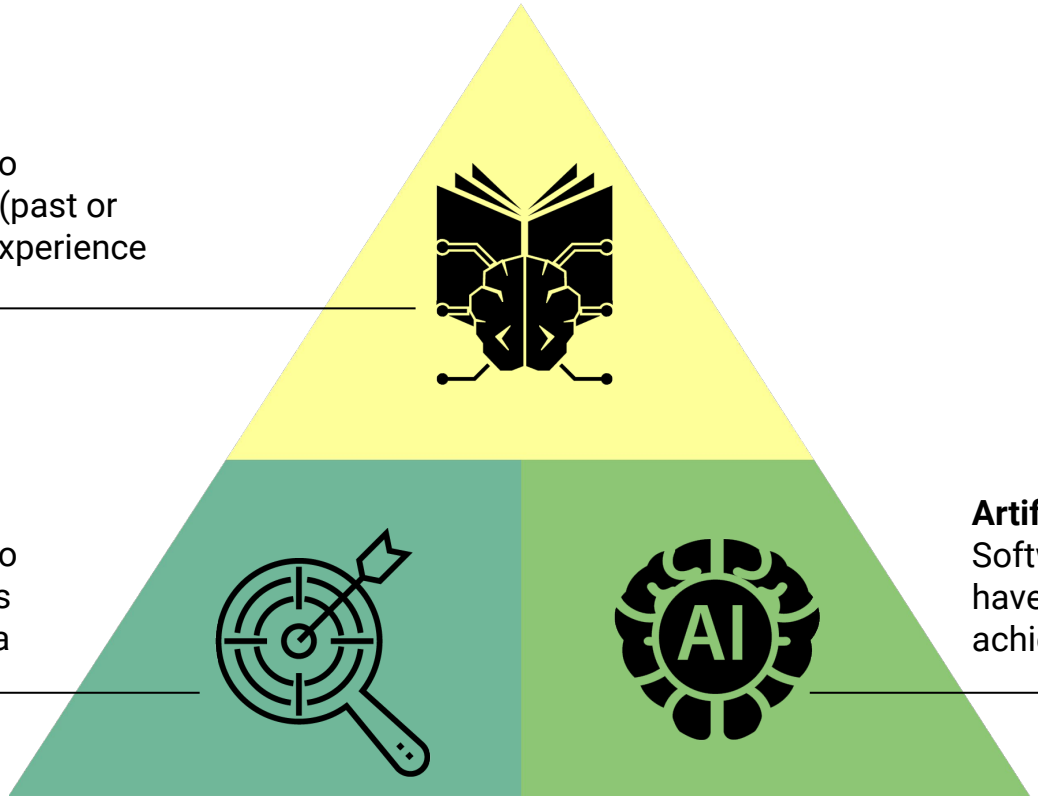
An **intelligent algorithm** responds to data so that the algorithm gets better, and effectively "evolves." The decision is no longer deterministic given the event.



Mysticism of Machine Learning

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



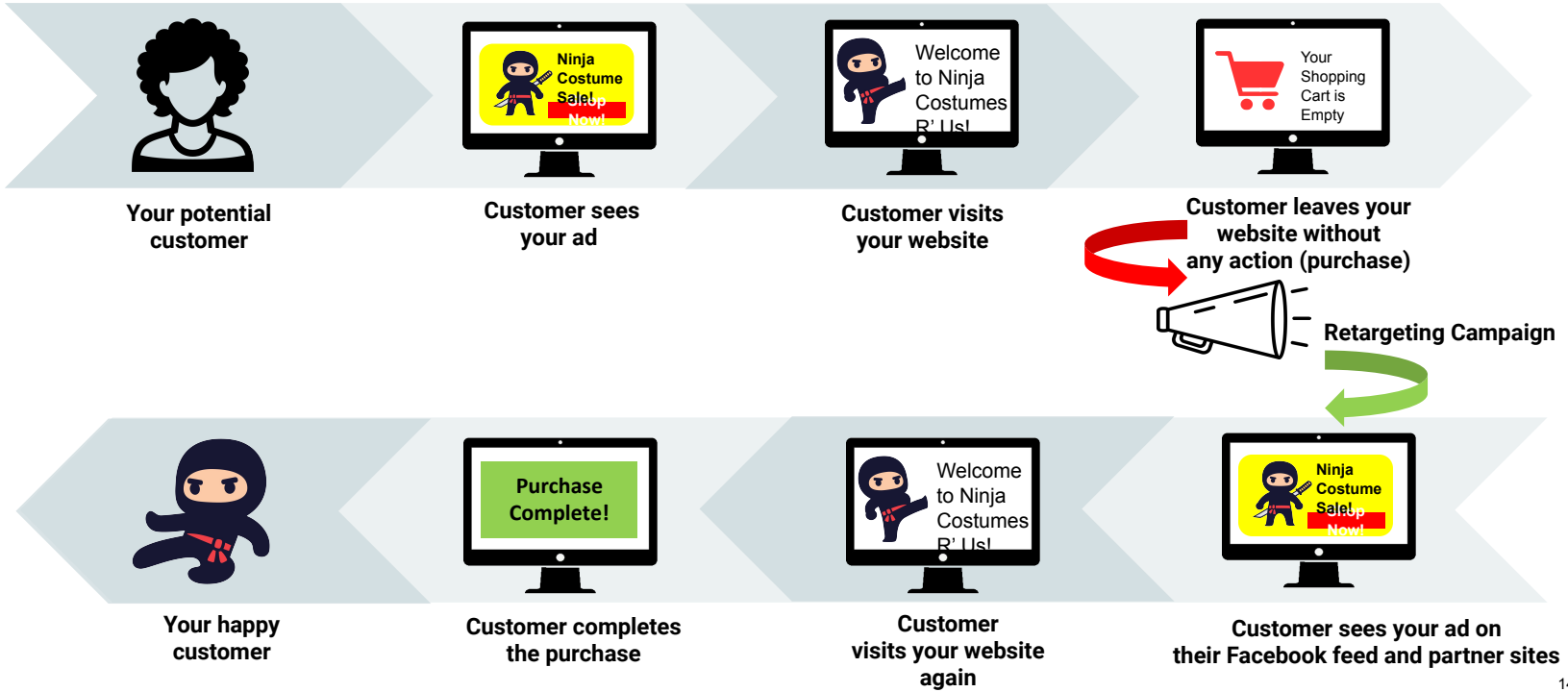
Artificial Intelligence
Software (and machines) that have a series of options to achieve a particular goal

Mysticism of Machine Learning

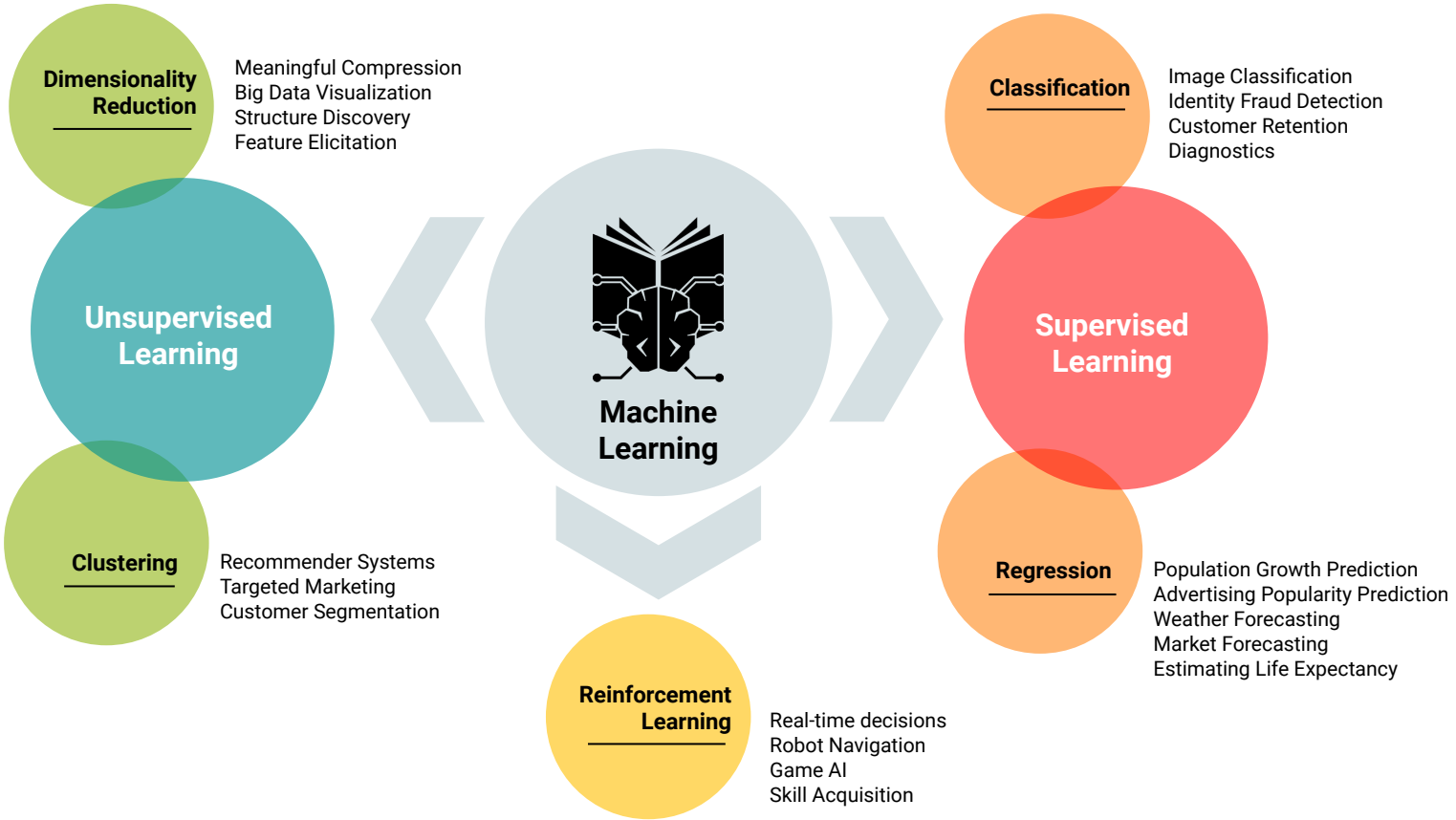
Artificial Intelligence

Mysticism of Machine Learning

Predictive Analytics



Mysticism of Machine Learning



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

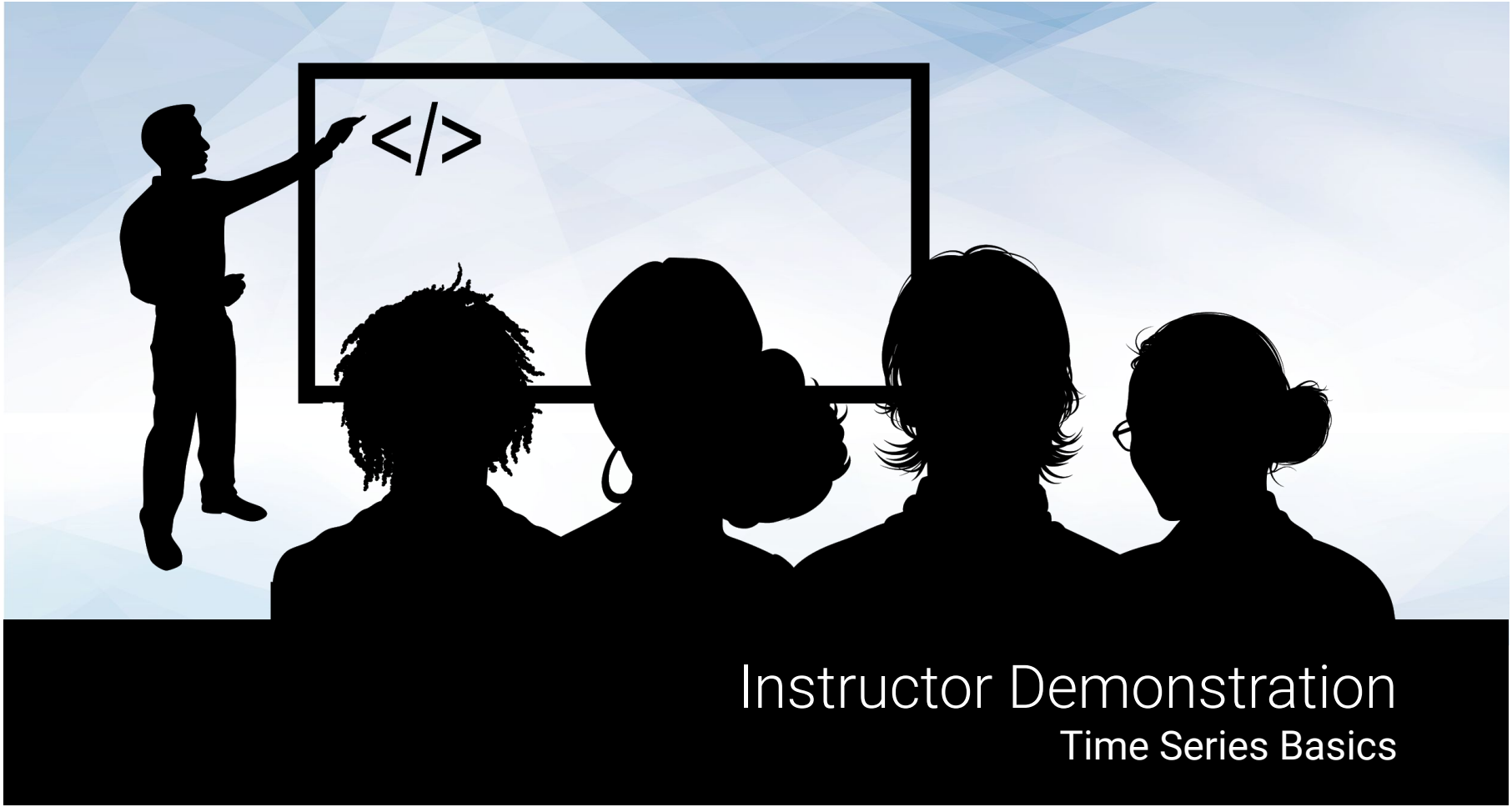
Pandas DateTimeIndex can be used to help with this.

```
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()
```



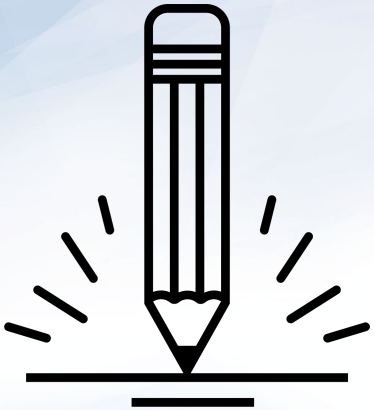
Instructor Demonstration

Time Series Basics

Resampling Options

Helpful resource can be found here

Alias	Description
B	Business day
D	Calendar day
W	Weekly
M	Month end
Q	Quarter end
A	Year end
BA	Business year end
AS	Year start
H	Hourly frequency
T, min	Minutely frequency
S	Secondly frequency
L, ms	Millisecond frequency
U, us	Microsecond frequency
N, ns	Nanosecond frequency



Activity: Time Series Basics

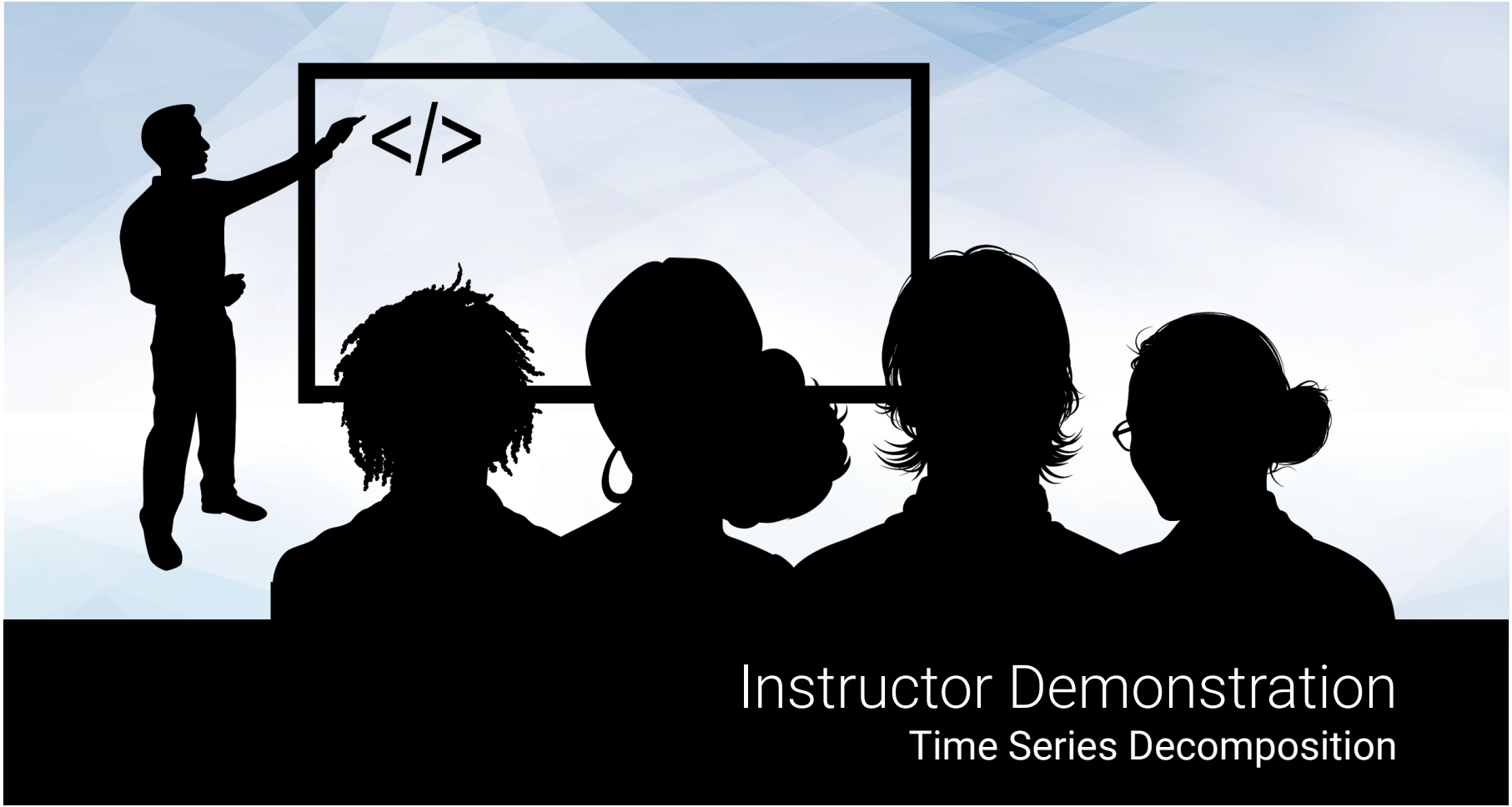
In this activity, you will practice the basics of time series manipulation in Pandas.

(Instructions sent via Slack.)

Suggested Time:
15 minutes

Time’s Up! Let’s Review.

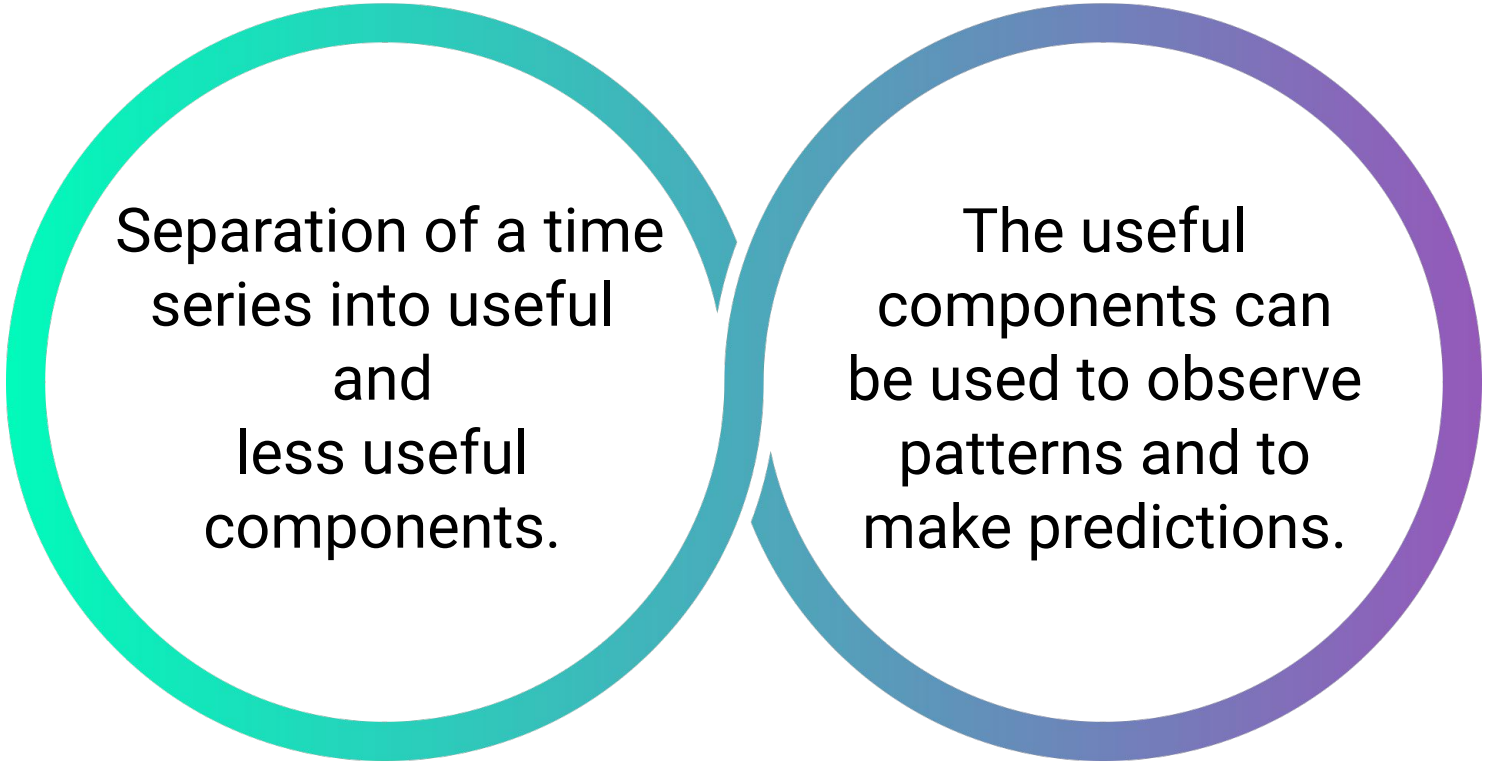
Time Series Decomposition



Instructor Demonstration

Time Series Decomposition

Time Series Decomposition



Time Series Decomposition

These are the components of time series decomposition:

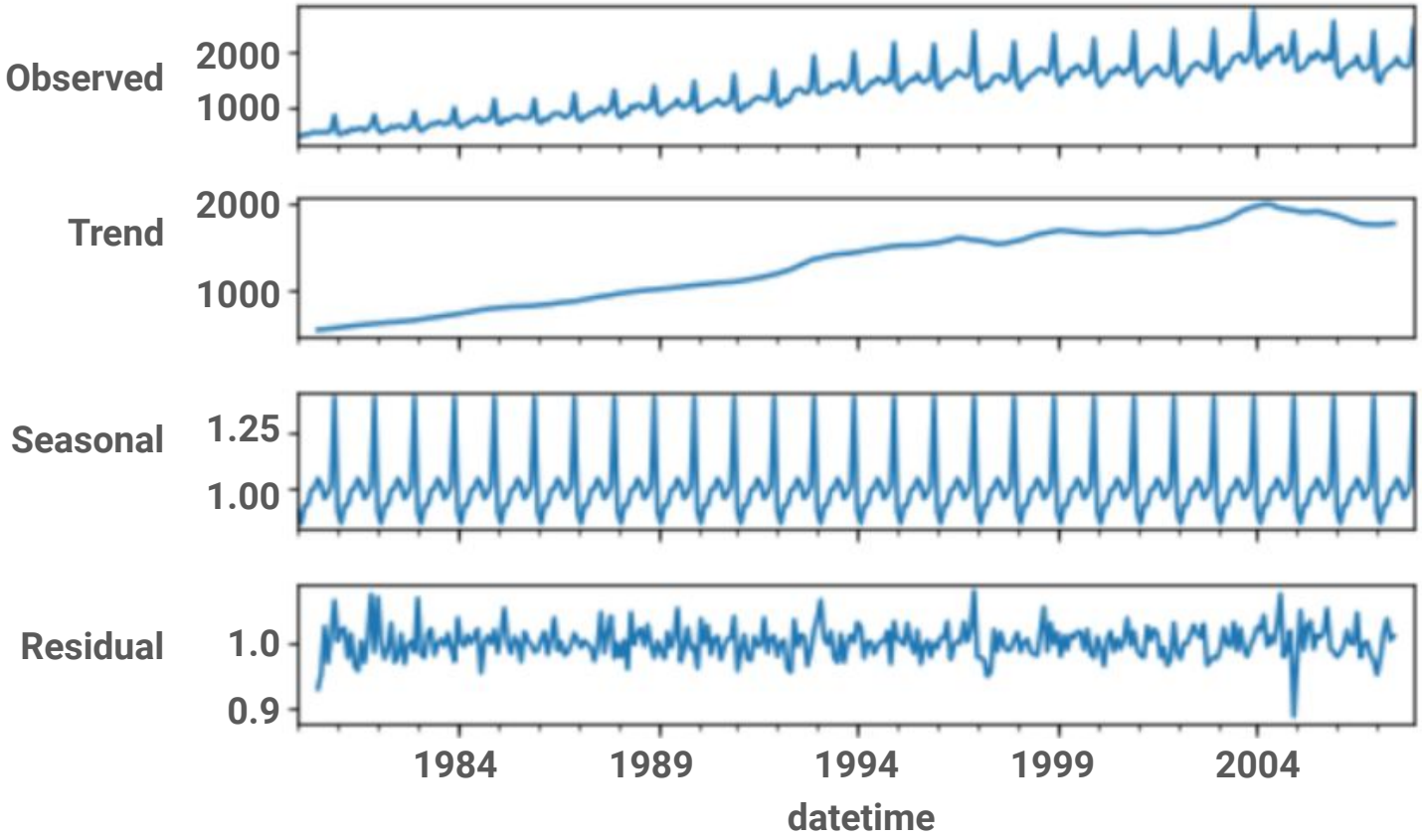
01 **Level:** What is the average value of the series?

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

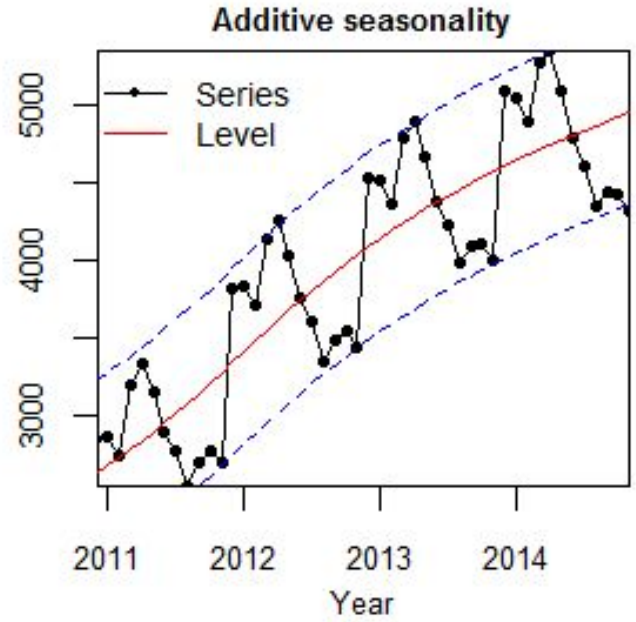
03 **Periodicity:** Do patterns occur in cycles?

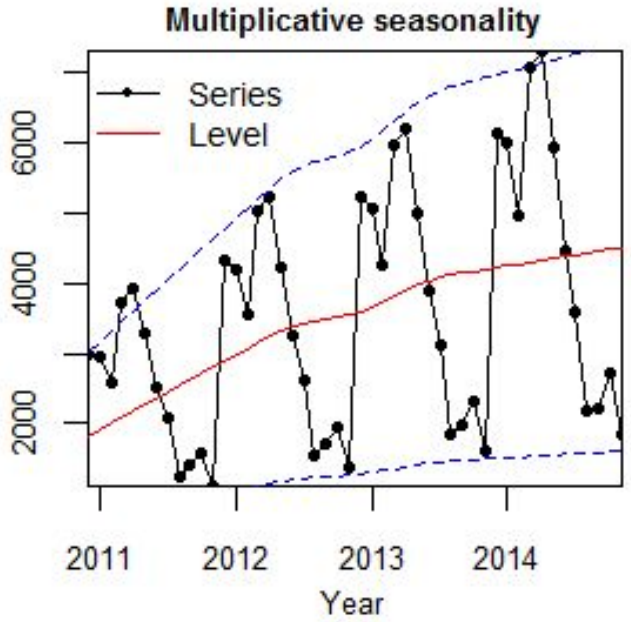
04 **Residual:** How much noise exists in the data?

Time Series Decomposition



Multiplicative vs Additive Time Series Data





Time Series Decomposition

- 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

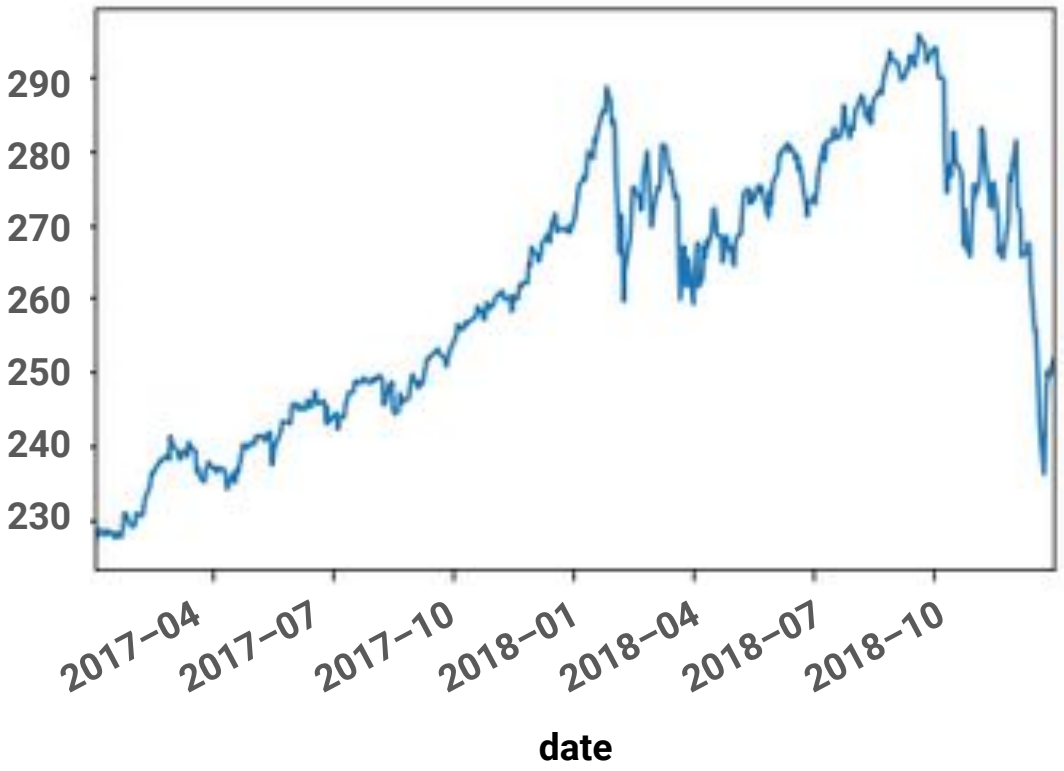
03

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

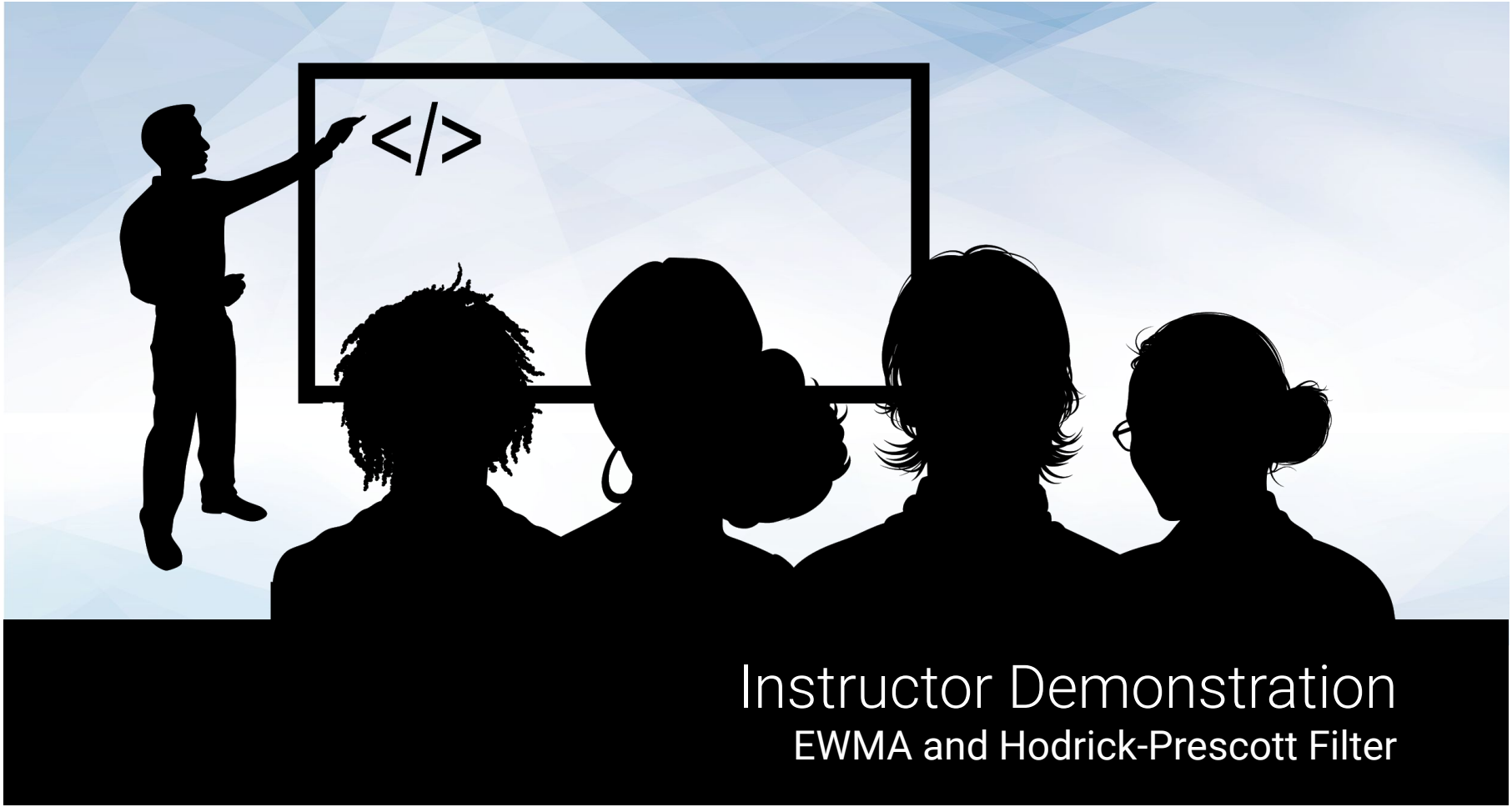
Filters out short-term fluctuations.

Hodrick-Prescott Filter

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

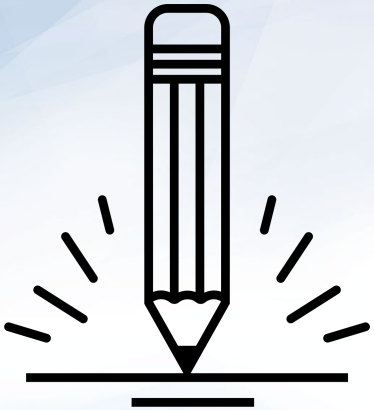
time series value - trend = cyclic element

Difference in trend over time = volatility



Instructor Demonstration

EWMA and Hodrick-Prescott Filter



Activity: CA Macroeconomics

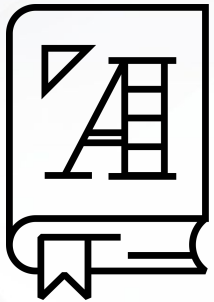
In this activity, you will use the Hodrick-Prescott filter to identify macroeconomic trends in Canada for the period 2004 to 2010.

(Instructions sent via Slack.)

Suggested Time:
15 minutes

Time’s Up! Let’s Review.

Autocorrelation

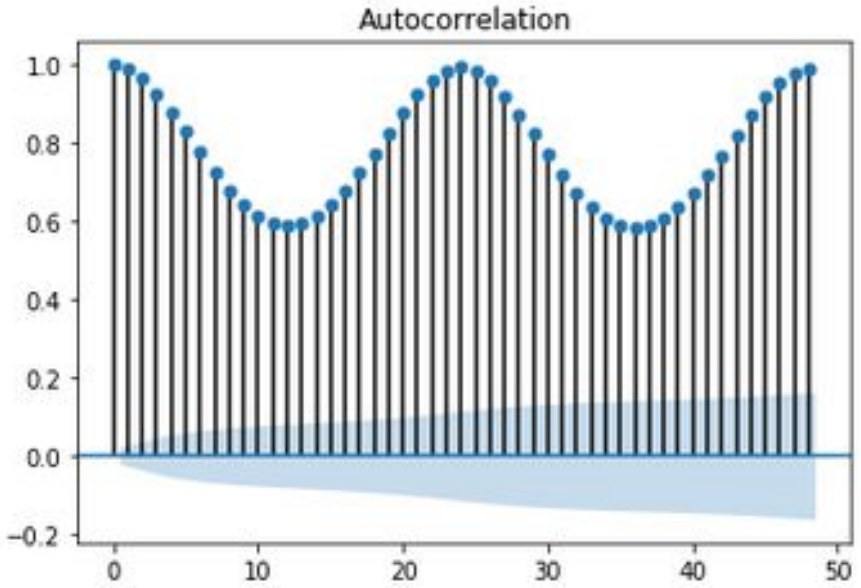


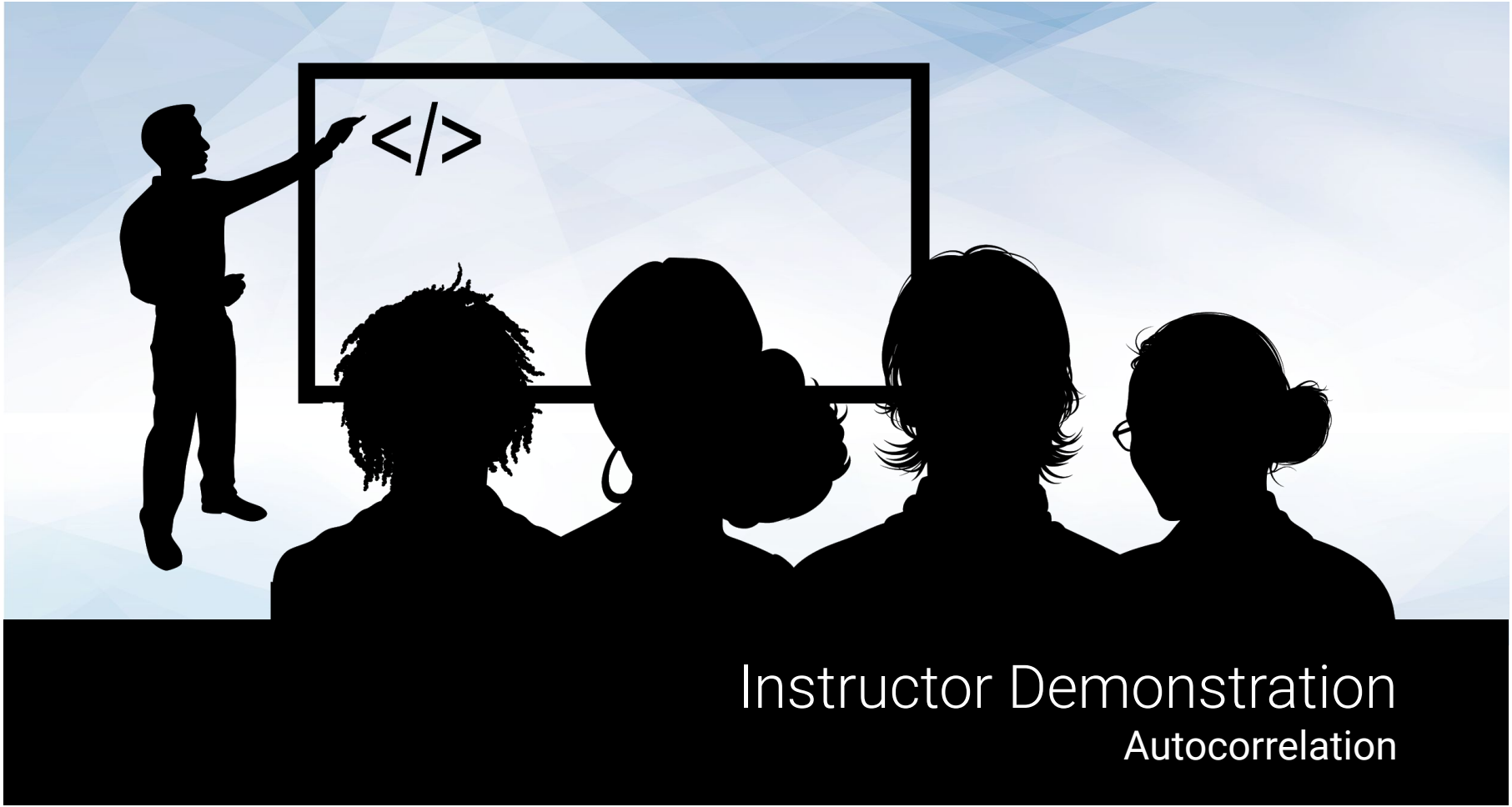
Autocorrelation is a measure of how closely current values correlate with past values.

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

Autocorrelation

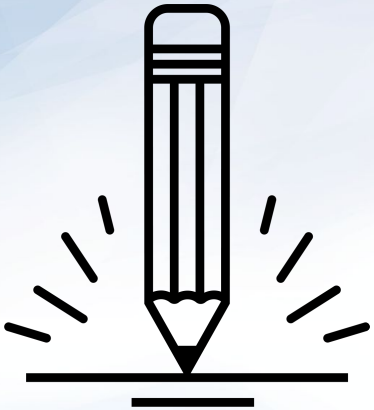
```
df.Temperature.autocorr(lag=1)
sm.graphics.tsaplots.plot_acf(df.Temperature,
```





Instructor Demonstration

Autocorrelation



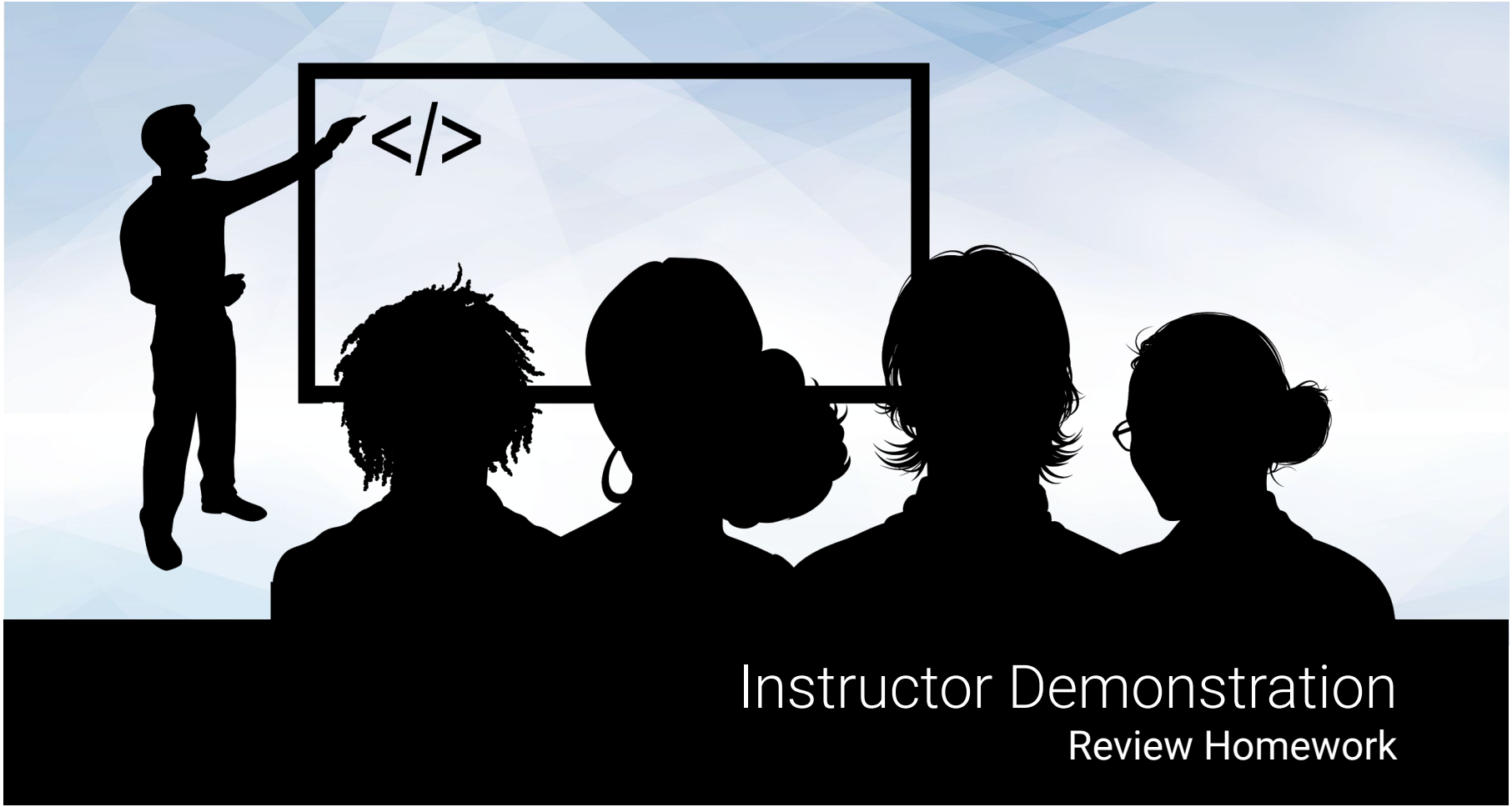
Activity: Euro ETFs

In this activity, you will examine a time series of bid-ask spreads of an ETF for autocorrelation.

(Instructions sent via Slack.)

Suggested Time:
15 minutes

Time’s Up! Let’s Review.



Instructor Demonstration

Review Homework

Questions?