AI Boot Camp

Introduction to Time Series

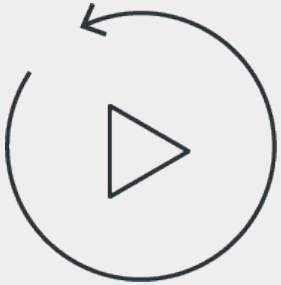
Module 8 Day 1

Class Objectives

By the end of class, you will be able to:

1. Recognize the importance of time data.
2. Manipulate time series data using pandas.
3. Use exploratory data analysis techniques on time series data.
4. Identify time series patterns in stock market data by using advanced slicing techniques and time pattern identification methods.
5. Use visualizations to identify relationships within time series data.

Welcome



Let’s **recap**

What do you know about **machine learning**?

Recap: Machine Learning

1. A machine learning algorithm automatically adapts to improve the accuracy and precision of outcomes and predictions, eliminating the need for manual configuration of inputs or adjustments to the algorithm.

2. Machine learning algorithms can learn autonomously, so developers don’t need to code for every conceivable scenario.

We use machine learning when we develop a **statistical model** of **existing data** that can automatically **make predictions or decisions** about **new data**.

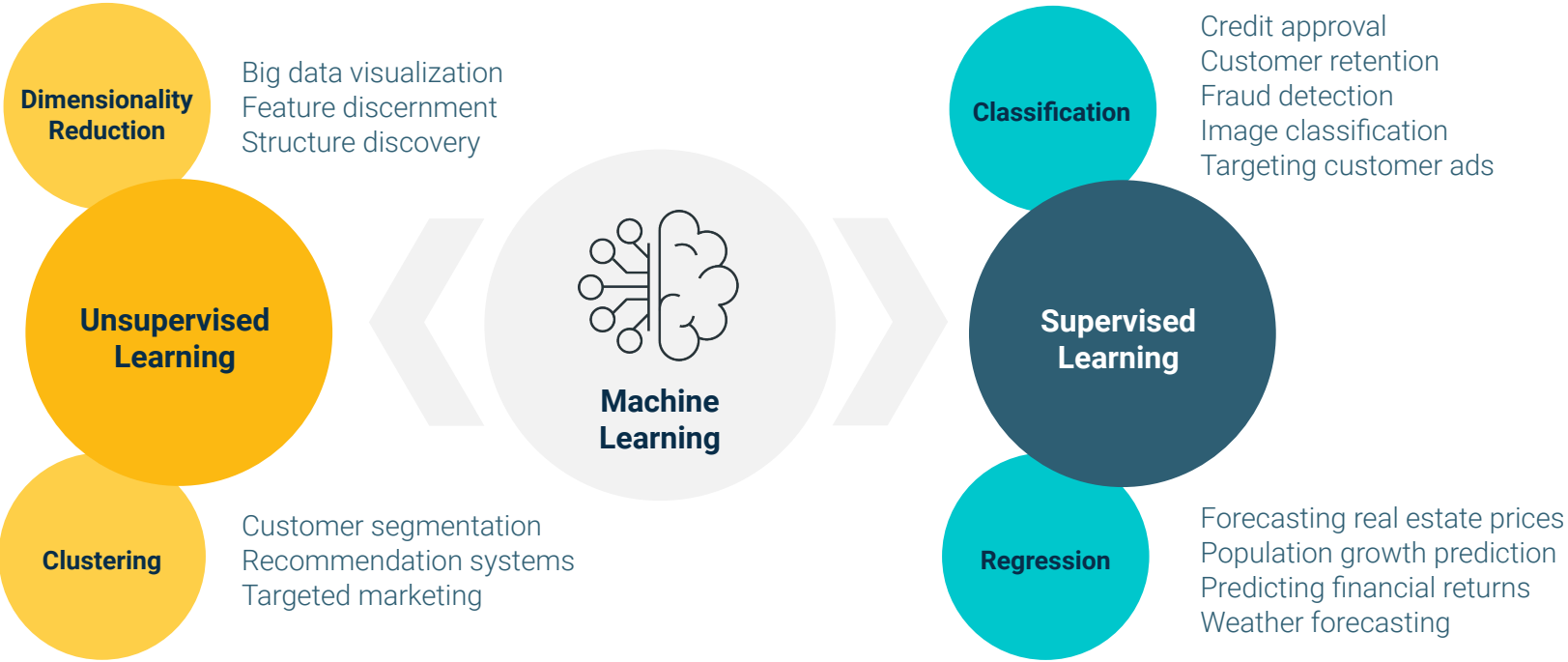
Recap: Machine Learning

1. All machine learning pipelines follow a **Model-Fit-Predict** paradigm where we use a dataset or data model to fit, or train, the algorithm.

2. Once trained, the model and algorithm can be used to make actual predictions.

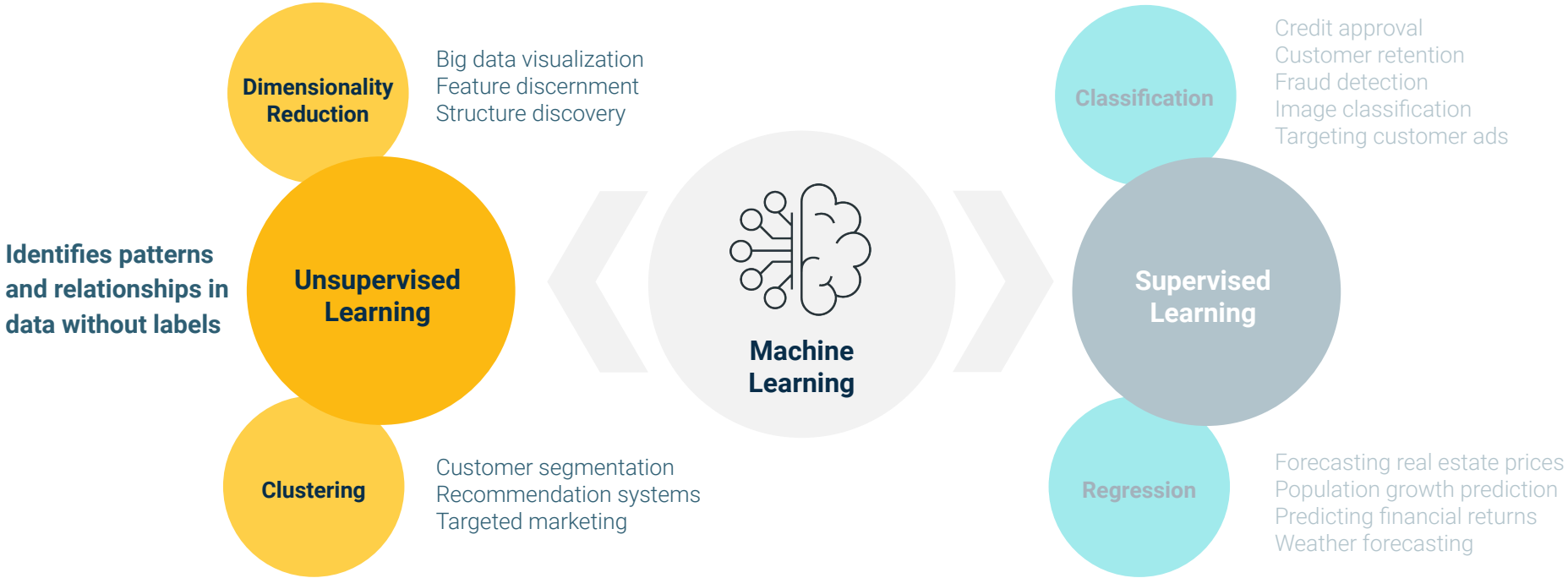
Recap: Machine Learning

We've learned that machine learning has **two main approaches**:



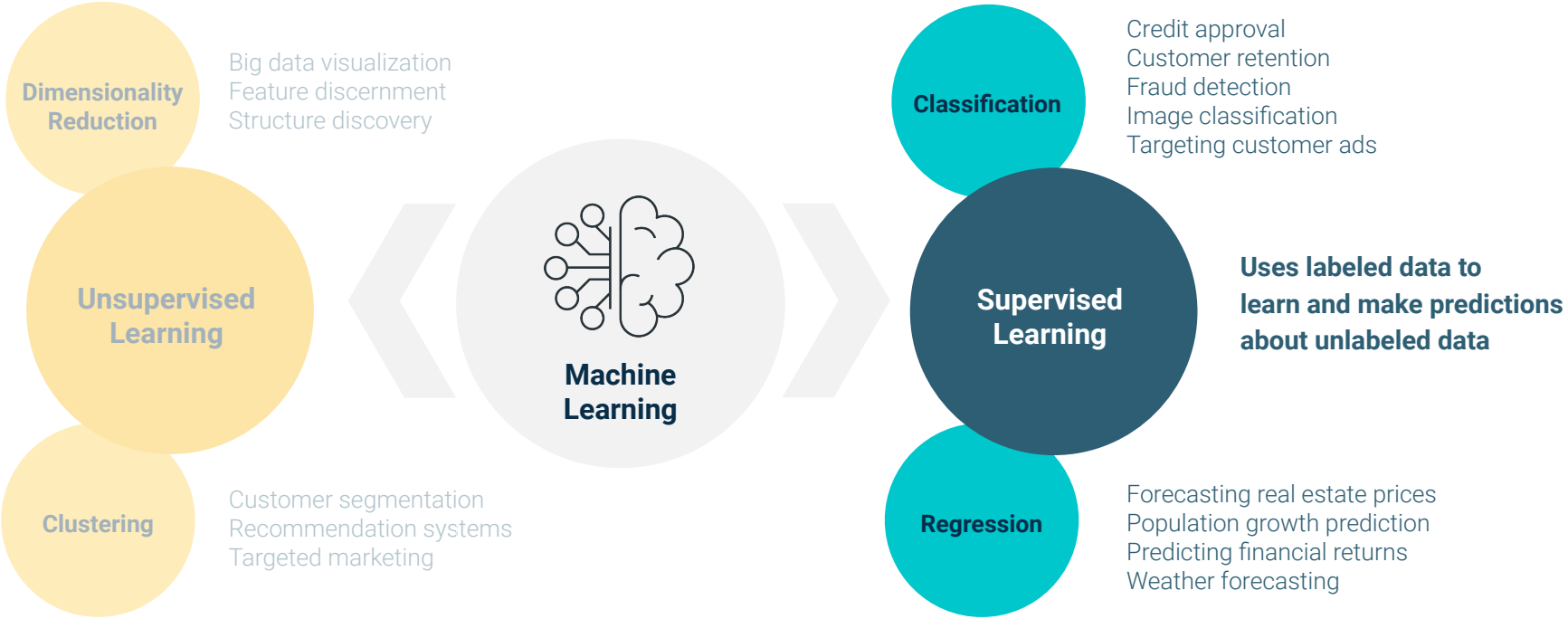
Recap: Machine Learning

Unsupervised learning is when an intelligent algorithm learns using data without labels.



Recap: Machine Learning

Supervised learning uses labeled data to learn.

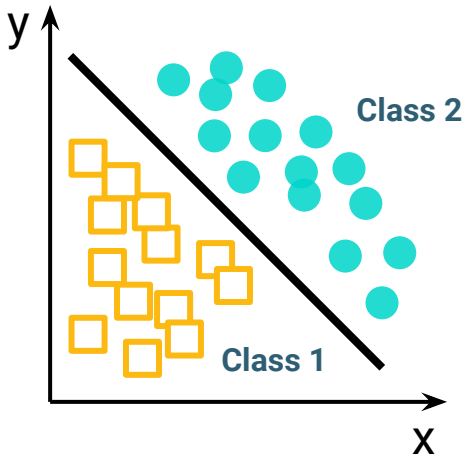


Recap: Machine Learning

Classification and **regression** are types of supervised learning. Both are used to make predictions.

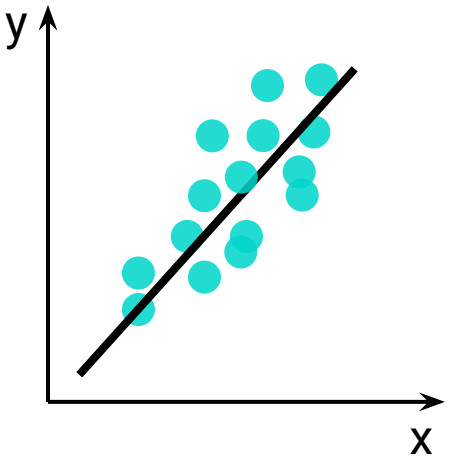
Classification

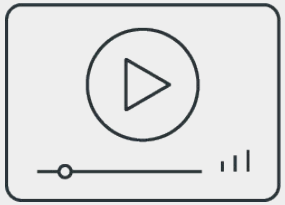
(Classifying outcomes as classes/groups)



Regression

(Predicting outcomes using continuous numbers)





Time for a quick **video**


How AI's, Like
ChatGPT,
Learn

How Machines Learn

The following machine learning examples were mentioned in the video:

01

An algorithm decides what price you are willing to pay at a particular moment.

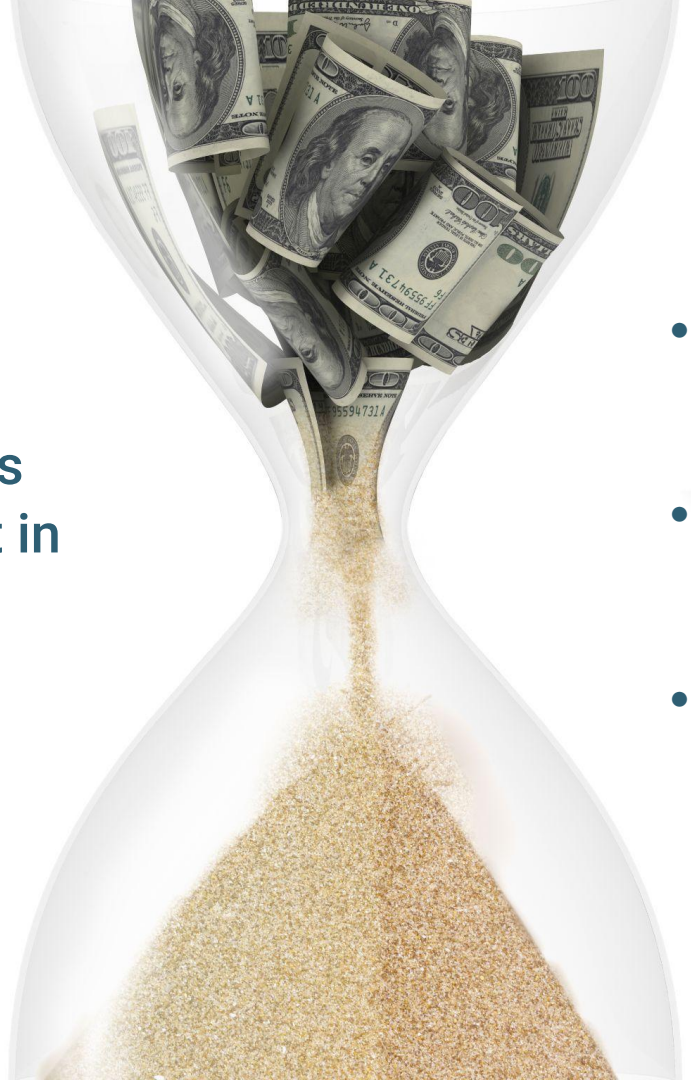
02

An algorithm predicts which financial transactions are fraudulent.

03

Algorithms continuously trade against other algorithms in the stock market.

The concept of time is an important element in decision-making.



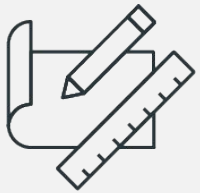
- Your willingness to pay for something might depend on the time of day (late-night shopping) or year (holiday season).
- Financial transactions that occur very late at night or very early in the morning have a greater chance of being fraudulent.
- Profitable strategies might appear and disappear depending on market conditions that change constantly over time.

Recap: Machine Learning

For a machine learning model to make predictions about all these outcomes, such as when a particular stock trade would be profitable, it requires plenty of historical data on both good *and* bad outcomes.

To make good decisions, a model needs many examples to learn from.

Many practical applications for machine learning models involve **supervised learning** of **time series data**.



Introduction to
Time Series Data

Introduction to Time Series Data

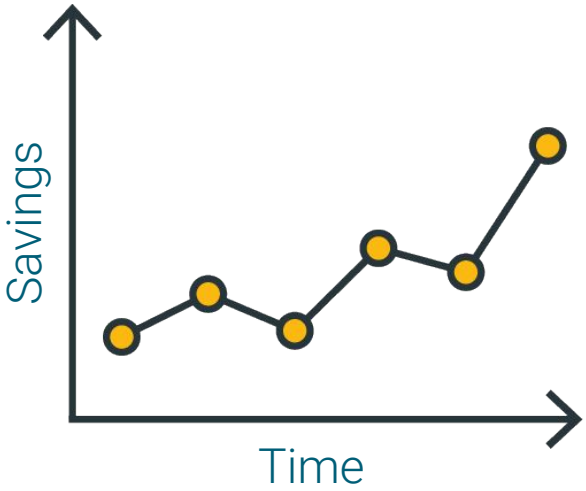
Our lives are based on the results of decisions that we make at specific times.

For example: Time might drive us to buy warm clothes in the winter or search for the perfect vacation spot in the summer.

Introduction to Time Series Data

Time also impacts how the world works and behaves.

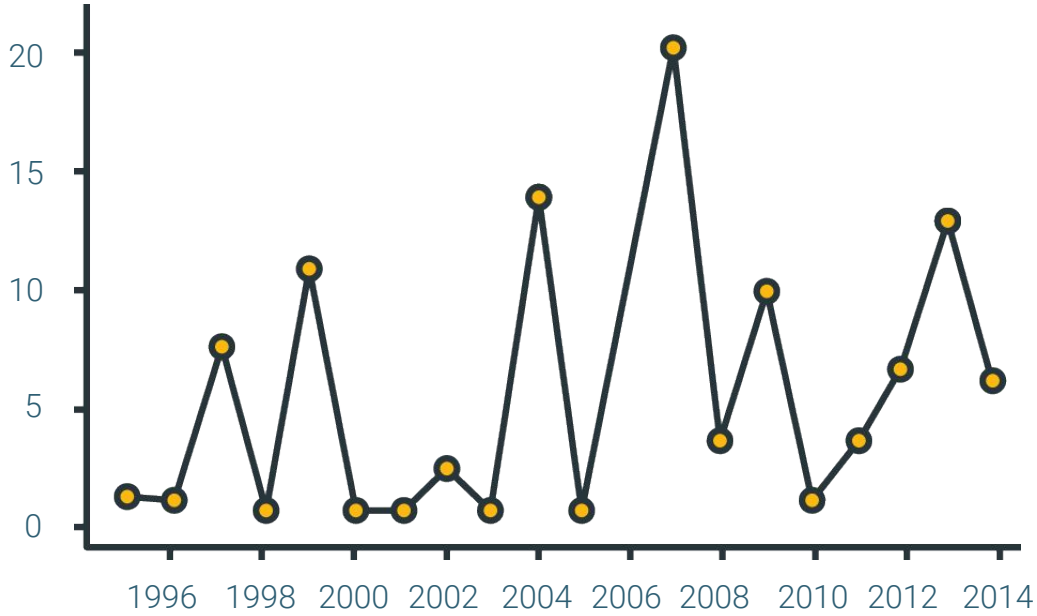
For example: When you make an investment or deposit money into your savings account, the value of the money can increase over time.



The Importance of Time in Finance

Time is essential in financial decision-making, so we need to learn how to manipulate, analyze, and understand data that is measured over time.

This type of data is called **time series data**.

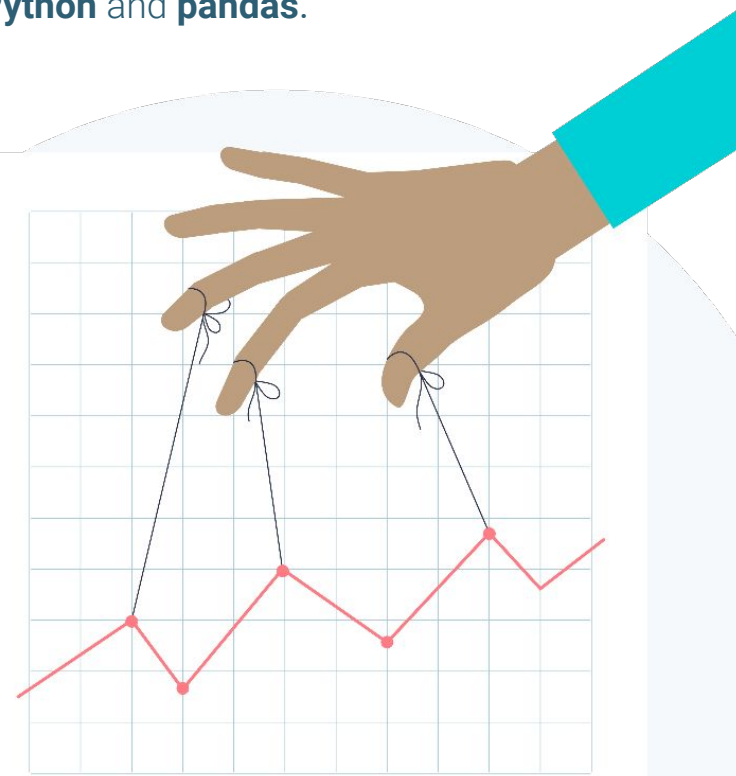


Introduction to Time Series Data

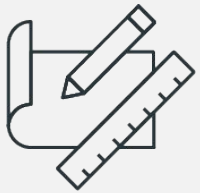
In this class, you will learn how to manipulate time series data with **Python** and **pandas**.

You will learn to use these tools' advanced capabilities to analyze and work with data in multiple formats from different sources.

Armed with this knowledge, you can construct more intricate time series models, including those capable of making predictions across varying time scales, such as days, weeks, years, and even different time zones.



Questions?



The Importance of **Time**

Aspects of Time Data

Time plays a role in a multitude of analysis tasks, including:

Discovering

Discovering an EKG (heart monitoring) reading at a particular point in time.

Aggregating

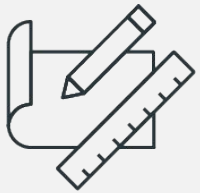
Aggregating the daily revenue for a firm that has sales in different countries or regions.

Forecasting

Forecasting tomorrow's weather, sales, or even traffic conditions.

Aspects of Time Data

To model and predict future events in time ...	We need the ability to handle everything that relates to time data.
When dealing with data from across the world ...	We need the ability to convert time zones.
To recognize patterns in time data ...	We need the ability to resample our data, converting hourly data to daily data or daily data to weekly data.



Using Pandas to **Work with Time Data**

Python and Pandas

In today's class, you'll learn about the intricacies of manipulating time series data with **Python** and **pandas**.

The concepts of dates and times in a programming language can be complicated.

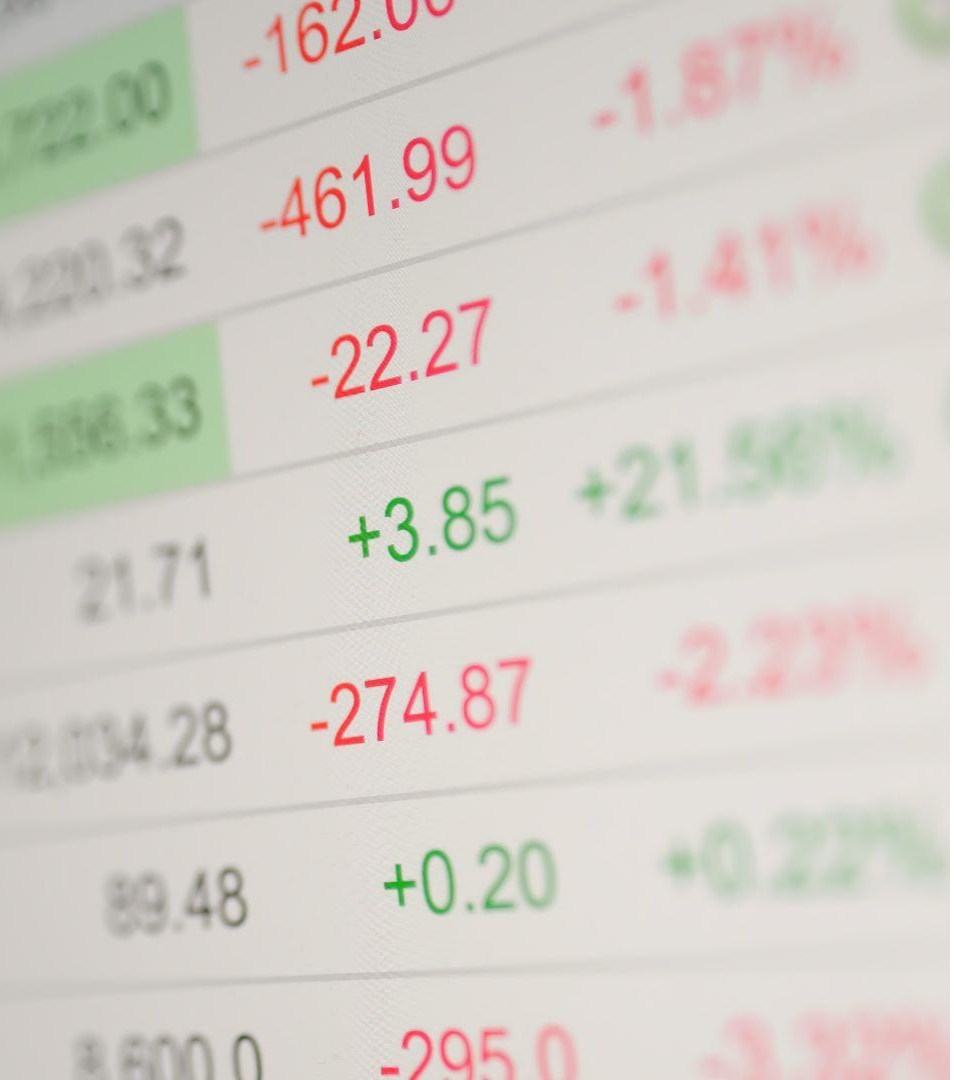
This is especially true when we consider the globe with its various time zones and levels of granularity.

These levels range from daily data to data measured at the microsecond.

Also, there are many ways to format and store dates and times on computers.

As you learned in earlier units, Python and pandas supply functions that help us work with dates and times in a DataFrame.

These functions use* datetime *objects, which will also make it easier to work with time series models.



A common example of time series data involves the stock market.

We can measure the price of each stock at specific intervals throughout the trading day, such as every minute, every hour, or every day (which will give us the closing price).

In this demo, we'll analyze the **S&P 500**, which is the index of the top 500 public stocks in the United States.

Instructor **Demonstration**

Using Pandas to Work with Time Data

What is Coordinated Universal Time?

UTC

Coordinated Universal Time (UTC—the initialism places “Coordinated” last) is a time standard that anyone in the world can use to specify an exact moment in time regardless of the location.

UTC doesn’t adjust for daylight saving time, which is what differentiates it from Greenwich Mean Time (GMT).

Coordinated Universal Time (UTC)

Whenever a timestamp includes a plus sign (+) or a minus sign (−), the number after the sign indicates the number of hours that we need to add or subtract from UTC to get the correct time zone.

Common time zones for stock data include:

`-05:00`

New York
standard time

`-04:00`

New York
daylight saving time

The UTC time standard also matches London's time zone at `+00:00`.

Coordinated Universal Time (UTC)

The **Time Zone Database** contains the time zone codes used by Python and other programming languages.

- This database is an international collaboration project that the Internet Assigned Numbers Authority (IANA) supports.
- The database is updated to reflect changes that governments make to time zone boundaries, UTC offsets, and daylight saving rules.
- You are encouraged to learn more about the Time Zone Database and how it's managed.

Using Pandas for `datetime` Objects

With pandas, we can use **`datetime`** objects to perform mathematical and other programming operations on dates and times.

Calling the pandas **`to_datetime`** function and passing it a parameter of “today” returns an object called Timestamp, which contains the following parts:

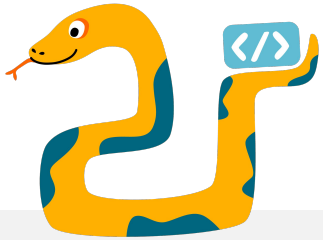
- The **`date`** and **`time`** information of the user’s current date in the format of year-month-day
- The time in the format of hours-minutes-seconds-milliseconds

Using Pandas for `datetime` Objects

`Timestamp` is the pandas equivalent of the Python `datetime` object.

`Timestamp`

=



`datetime`

This pandas function is:

- Used for entries that make up the pandas `DatetimeIndex` and other time-oriented data structures.
- Convenient when we want to use an API to pull data that ranges from a particular time in the past to today.

Instructor **Demonstration**

Using the Time-Related Functions & Converting UTC Data to a Specific Time Zone

Questions?

Activity:

Inspecting Time Zones in Stock Data

In this activity, you will load historical stock data about Tesla Inc. (TSLA) to practice your `datetime` data transformation skills.

Suggested time:
20 minutes

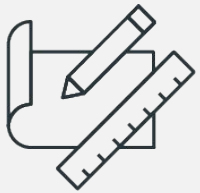
Time’s up!

Let’s review

Questions?

Instructor **Demonstration**

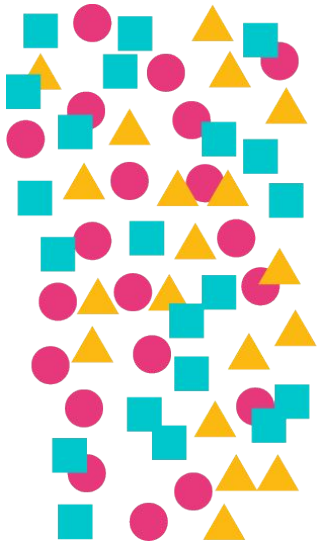
Analyzing Market Data Across Time



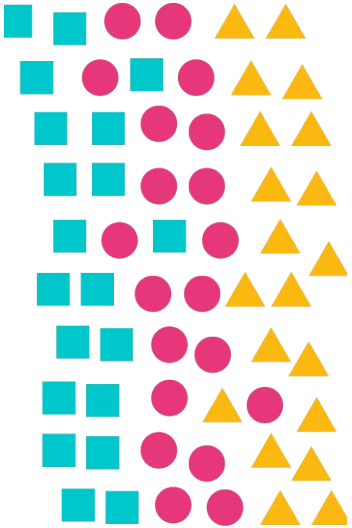
Time Series **Patterns**

Patterns

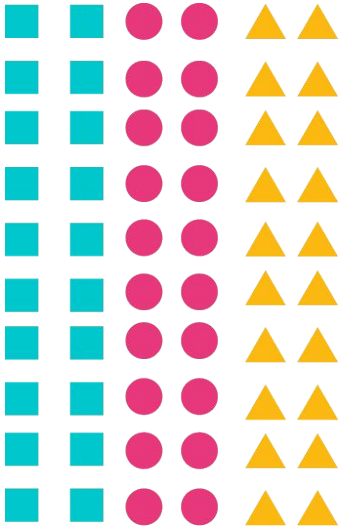
The more we understand the patterns in our data, the better we become at training and building models that involve this data.



Big data



Analytics



Decisions

Patterns of behavior often result from everyday human activities and behaviors.

Examples include going to lunch or feeling excited about the market opening or closing.

These identifiable patterns in our data are known as **common intraday patterns** and **time-of-day patterns**.

Stock Market Patterns

Some examples of time series patterns in the stock market include:

1. A surge in trading volume at the opening of the market at 9:30am Eastern Time.
2. A spike in trading volume when the market closes its positions at lunchtime.
3. A daily high or low price that's tested around lunchtime.
4. Increases in trading volume and the potential for price movement between 2pm and 3pm Eastern Time.
5. A final push before the market closes, just before 4pm Eastern Time.

Let’s try to find some patterns in our **S&P 500 data** by visualizing them with plots!

Instructor **Demonstration**

Checking the Closing Prices

Slicing Time Data

Slicing time data shows the advantage of occasionally zooming in on our data to identify trends.

If we examined the entire dataset, we would have difficulty observing a lunchtime trend.

By zooming in to a specific time, smaller trends like this become more apparent.

When we identify trends, we can incorporate them—big or small—into our time series models for better accuracy.

Instructor **Demonstration**

Further Time Slicing with Pandas

Break

15 mins

Activity:

Visualizing Stock Data

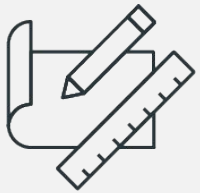
In this activity, you will convert a date column to `datetime` and perform slicing functions on various dates.

Suggested time:
20 minutes

Time’s up!

Let’s review

Questions?



Exploring **Time Series Data**

Exploring Time Series Data

In this section, we'll learn how to use pandas to explore time series data for seasonal patterns.

One of the goals of time series analysis is to support better decision-making by **understanding how the series events behave**.

Exploring Time Series Data

For example, using home sales data, we can identify the best time to sell a property based on increasing demand or higher sale prices.

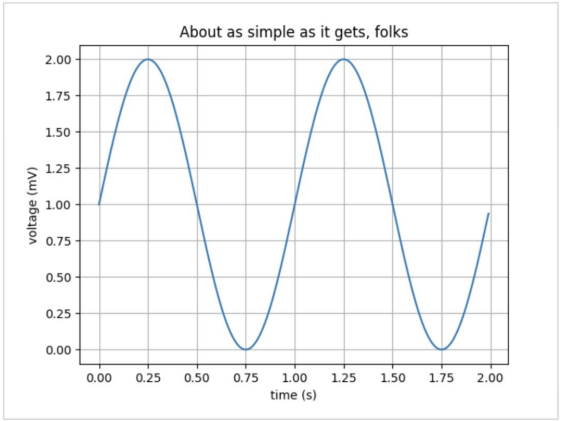
To make such decisions, we need to identify patterns in a visualization depicting financial data over time.

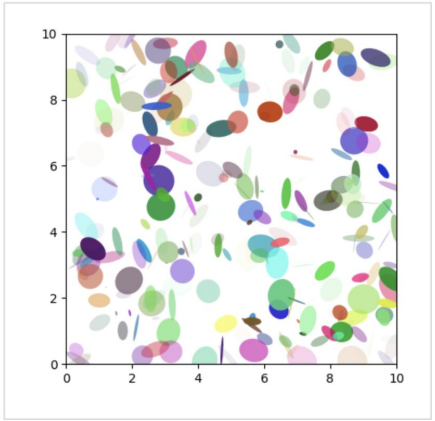
Exploring Time Series Data

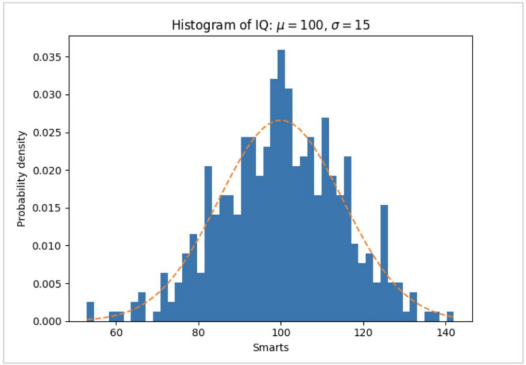
In this demonstration, you will:

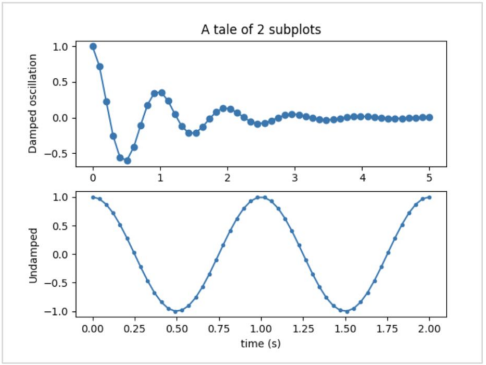
Create and interpret time series visualizations such as line plots.

Analyze time series data to recognize relationships.









Instructor **Demonstration**

Loading and Preprocessing Time Series Data

Instructor **Demonstration**

Analyzing Time Data from a Quarterly Perspective

Analyzing Time Data from a Quarterly Perspective

Now, we'll analyze the homes-sold series from a quarterly perspective.

We'll use the DatetimeIndex attributes and the groupby function to compute the total home sales per quarter, as the following code shows:

```
# Compute the total home sales per quarter
quarterly_sales =
df_home_sales["homes_sold"].groupby(by=[df_home_sales.index.quarter]).sum()

# Display total home sales per quarter
quarterly_sales
```

Questions?

Activity:

Analyzing Time Series Patterns in the S&P 500 Index

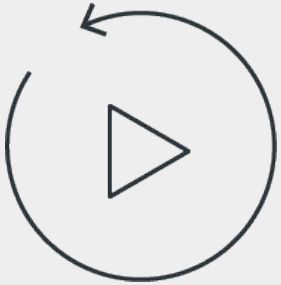
In this activity, you will use your newly developed skills to visualize and analyze time series patterns in the S&P 500 volume data.

Suggested time:
20 minutes

Time’s up!

Let’s review

Questions?



Let’s **recap**

Recap

After today's lesson you are able to:

1. Recognize the importance of time data.
2. Manipulate time series data using pandas.
3. Use exploratory data analysis techniques on time series data.
4. Identify time series patterns in stock market data by using advanced slicing techniques and time pattern identification methods.
5. Use visualizations to identify relationships within time series data.

Next

In the next lesson, you'll learn about using **correlations** to identify whether two time series with seasonal patterns have a relationship—and whether that relationship is predictable. You will also learn how to forecast time series and interpret the forecasting results.

Questions?

The End