Pandas:
Risk and Returns Over Time

FinTech

Lesson 3.3

© 2020 Trilogy Education Services, a 2U, Inc. brand. All Rights Reserved.

Class Objectives

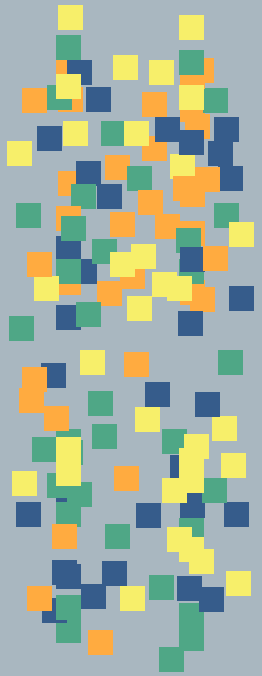
By the end of class, you will be able to:

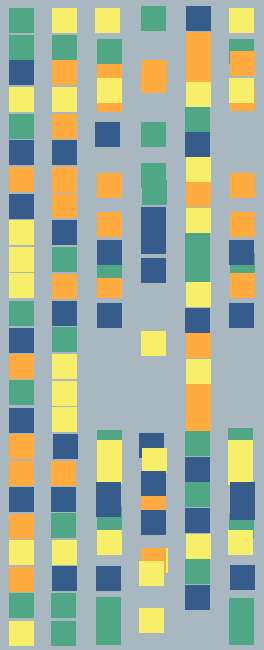
- Group data in a DataFrame to perform calculations on the grouped data.
- Manipulate datetime data in different formats: single variables, DataFrame columns, and series.
- Identify the calculations that can be done with datetime data & declare and use a DateTimeIndex.
- Calculate mean, median, and standard deviation using Pandas, and apply standard deviation to risk analysis use cases.
- Determine risk by identifying how stocks deviate from the mean.
- Describe Sharpe ratios and calculate them using Pandas DataFrames.

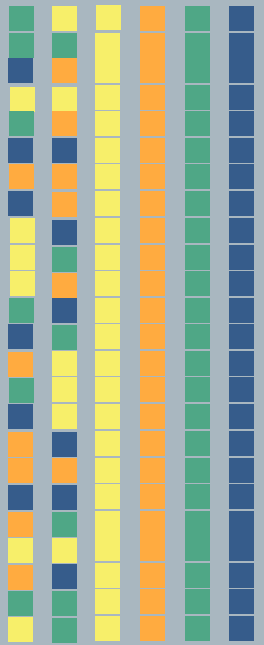
Sorting

Sorting

Data is not always organized in the best way for analysis. Sometimes, data needs to be cleaned and sorted.







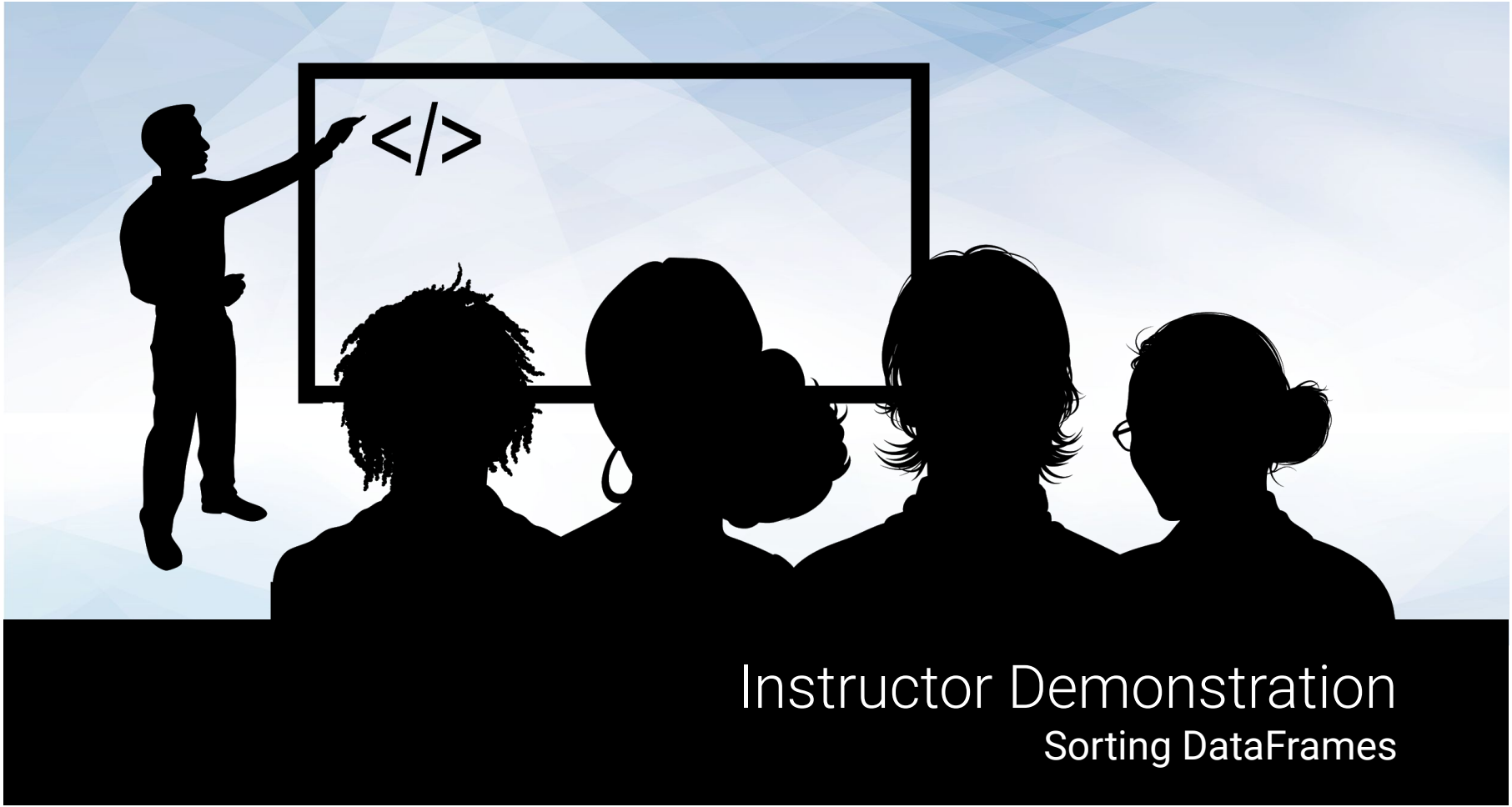
Sorting

The `sort_values` function in Pandas can be used to sort a DataFrame. Sorting data helps improve visual representation of data.

Data can be sorted in ascending or descending order.

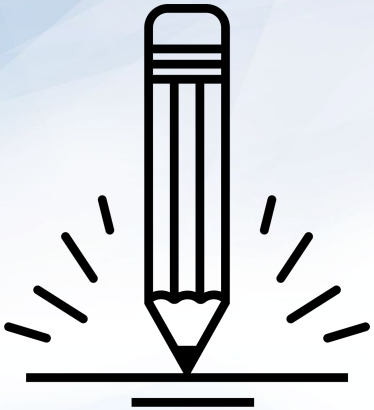
```
sort_values(ascending=True)
```

Consider dates: would you rather see dates sorted or randomly listed?



Instructor Demonstration

Sorting DataFrames



Activity: Out of Sorts

In this activity, you will analyze data for a single ticker to calculate daily returns for the year 2019. The data will then be sorted in descending order to identify the top five performing days for returns.

(Instructions sent via Slack.)

Suggested Time:
10 Minutes

Time’s Up! Let’s Review.

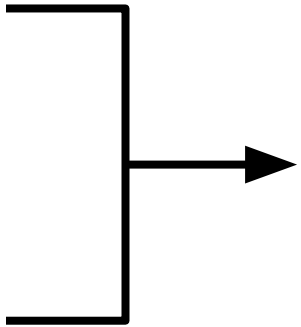
Grouping

Grouping

A key component of data analysis is grouping data. **Grouping** allows for similar data to be aggregated or manipulated as groups.

Example aggregations that can be done on groups are adding, summing, determining min and max, etc.

Category	Sales
a	1
a	2
b	10
b	9



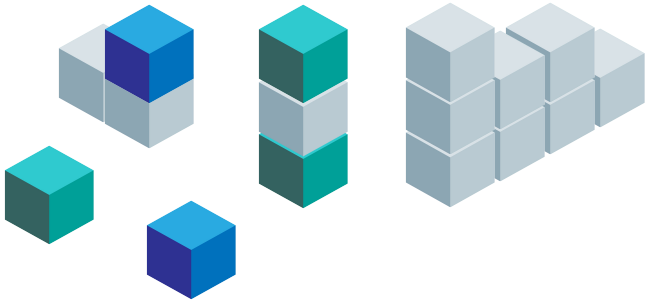
Category	Sales
a	3
b	19

Grouping

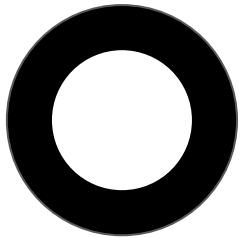
Behind the scenes, the Pandas `groupby` function does the following:

Splits the data into groups based on certain criteria.

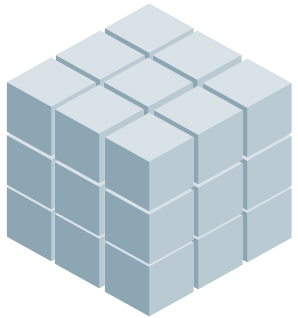
Applies a function to each group independently.



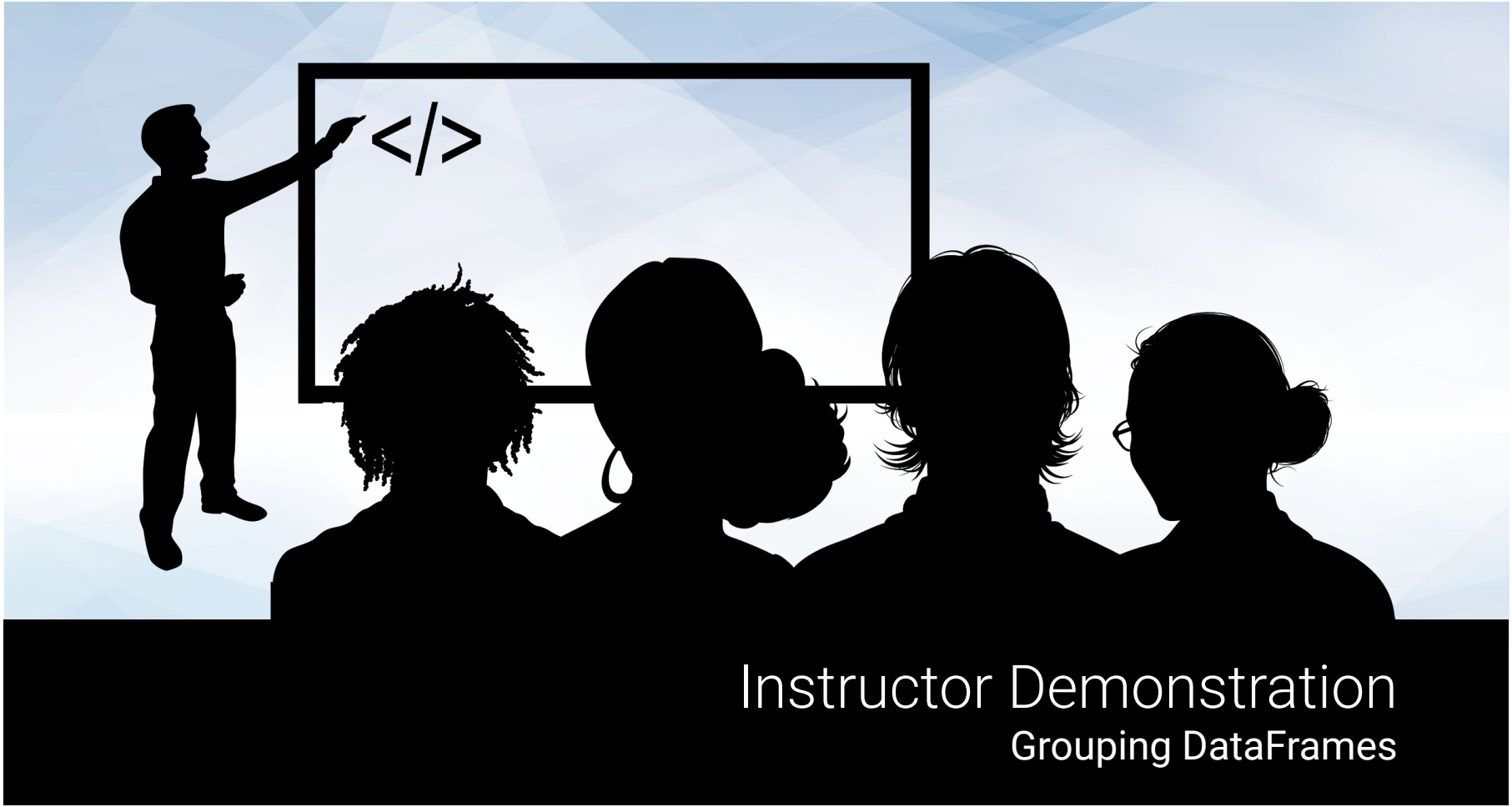
Splitting Data



Applying a Function

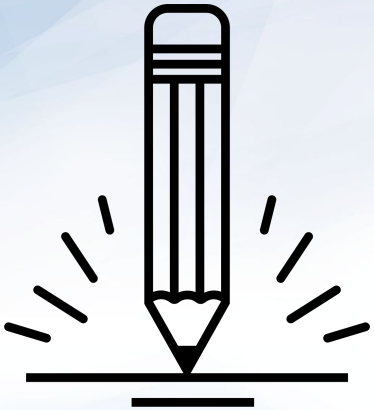


Combining Results



Instructor Demonstration

Grouping DataFrames



Activity: Group Dynamics

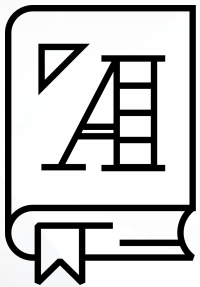
In this activity, you will load in cryptocurrency data, group data by each crypto, perform aggregations to analyze price trends, and then plot the results.

(Instructions sent via Slack.)

Suggested Time:
15 Minutes

Time’s Up! Let’s Review.

Multi-Indexing



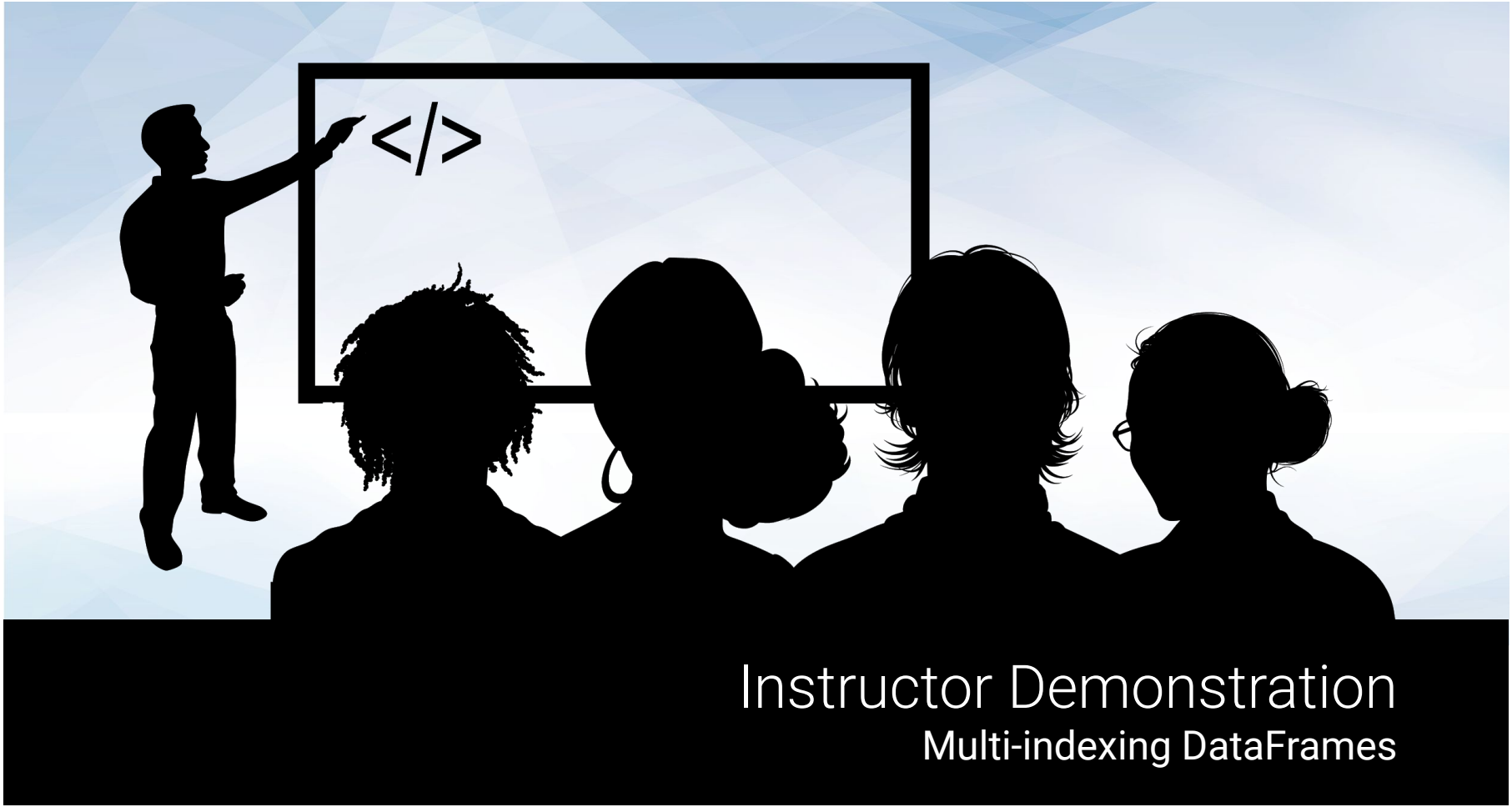
Multi-indexing is the process of creating more than one index for a DataFrame.

Multi-Indexing

Sometimes, more than one index is needed to perform a data lookup. For example, it is common to use multi-indexing when working with dates. This allows data to be accessed by day, month, and/or year.

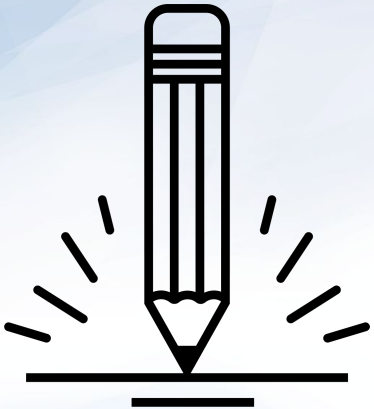
	one			two		
	a	b	c	a	b	c
0	-1.401530	0.626666	-0.165782	-0.361173	-1.139887	-0.027251
1	1.201998	-0.665546	-0.554207	0.644199	0.572868	-1.552404
2	-1.201190	-1.428929	1.226697	0.162548	1.481702	0.721526
3	-1.622470	0.541475	-0.482980	-1.970389	1.974586	0.165966

Multi-indexing improves data storage, lookup, manipulation, and assignment.



Instructor Demonstration

Multi-indexing DataFrames



Activity: Indexing Fever

In this activity, you will use hierarchical indexes to gain access to historical stock data.

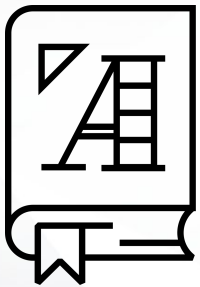
(Instructions sent via Slack.)

Suggested Time:
15 Minutes

Time’s Up! Let’s Review.

Break

Concatenation

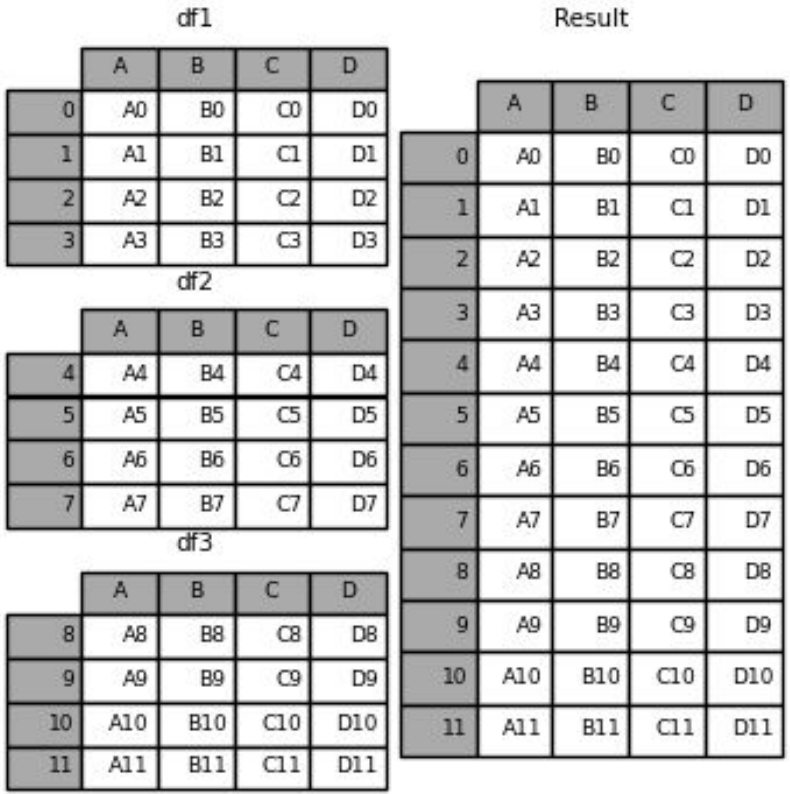


Concatenation is the process of appending data from one object with another.

Concatenation at a Glance

DataFrames can be concatenated so that the records from two or more DataFrames are combined.

Concatenation can be done by column, so that columns from one DataFrame are placed adjacent to columns from another DataFrame.



df1

	A	B	C	D
0	A0	B0	C0	D0
1	A1	B1	C1	D1
2	A2	B2	C2	D2
3	A3	B3	C3	D3

df2

	A	B	C	D
4	A4	B4	C4	D4
5	A5	B5	C5	D5
6	A6	B6	C6	D6
7	A7	B7	C7	D7

df3

	A	B	C	D
8	A8	B8	C8	D8
9	A9	B9	C9	D9
10	A10	B10	C10	D10
11	A11	B11	C11	D11

Result

	A	B	C	D
0	A0	B0	C0	D0
1	A1	B1	C1	D1
2	A2	B2	C2	D2
3	A3	B3	C3	D3
4	A4	B4	C4	D4
5	A5	B5	C5	D5
6	A6	B6	C6	D6
7	A7	B7	C7	D7
8	A8	B8	C8	D8
9	A9	B9	C9	D9
10	A10	B10	C10	D10
11	A11	B11	C11	D11

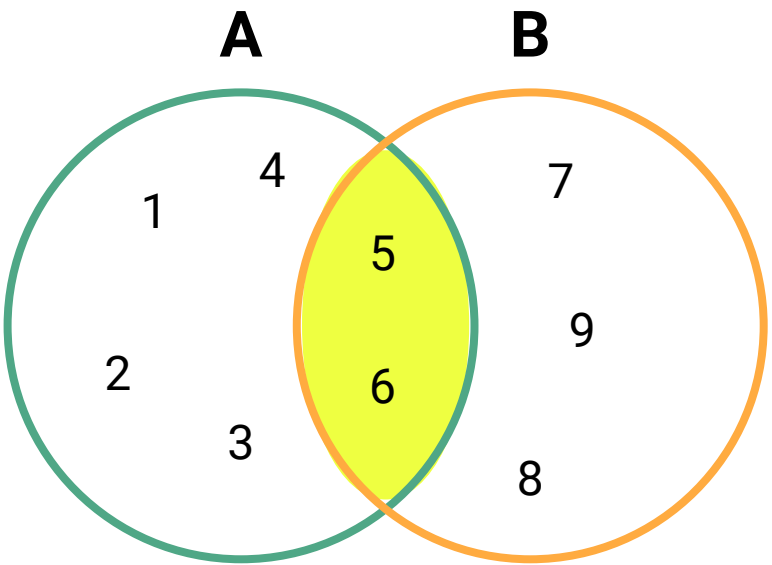
Concatenation

Pandas has a `concat` function that can be used to combine more than one DataFrame.

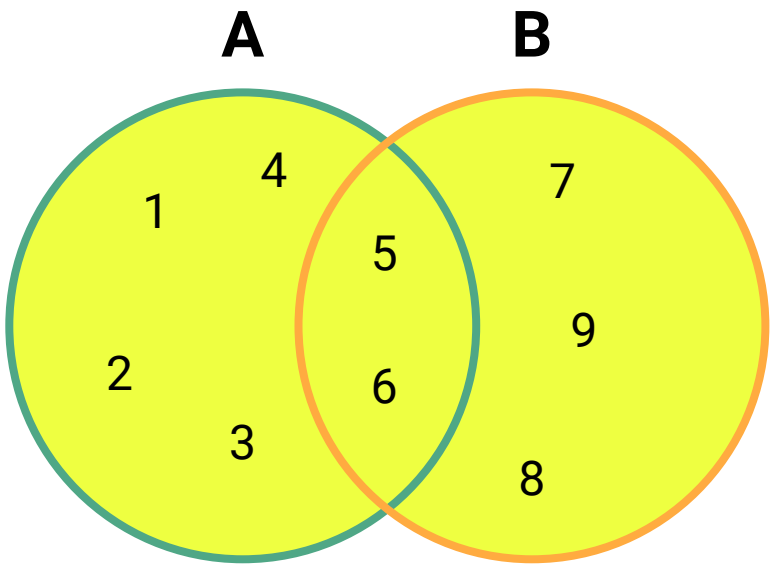
Parameters:

- A list of DataFrames to be joined
- The `axis` to join on (by column or row)
- The `join` operation (inner vs. outer)

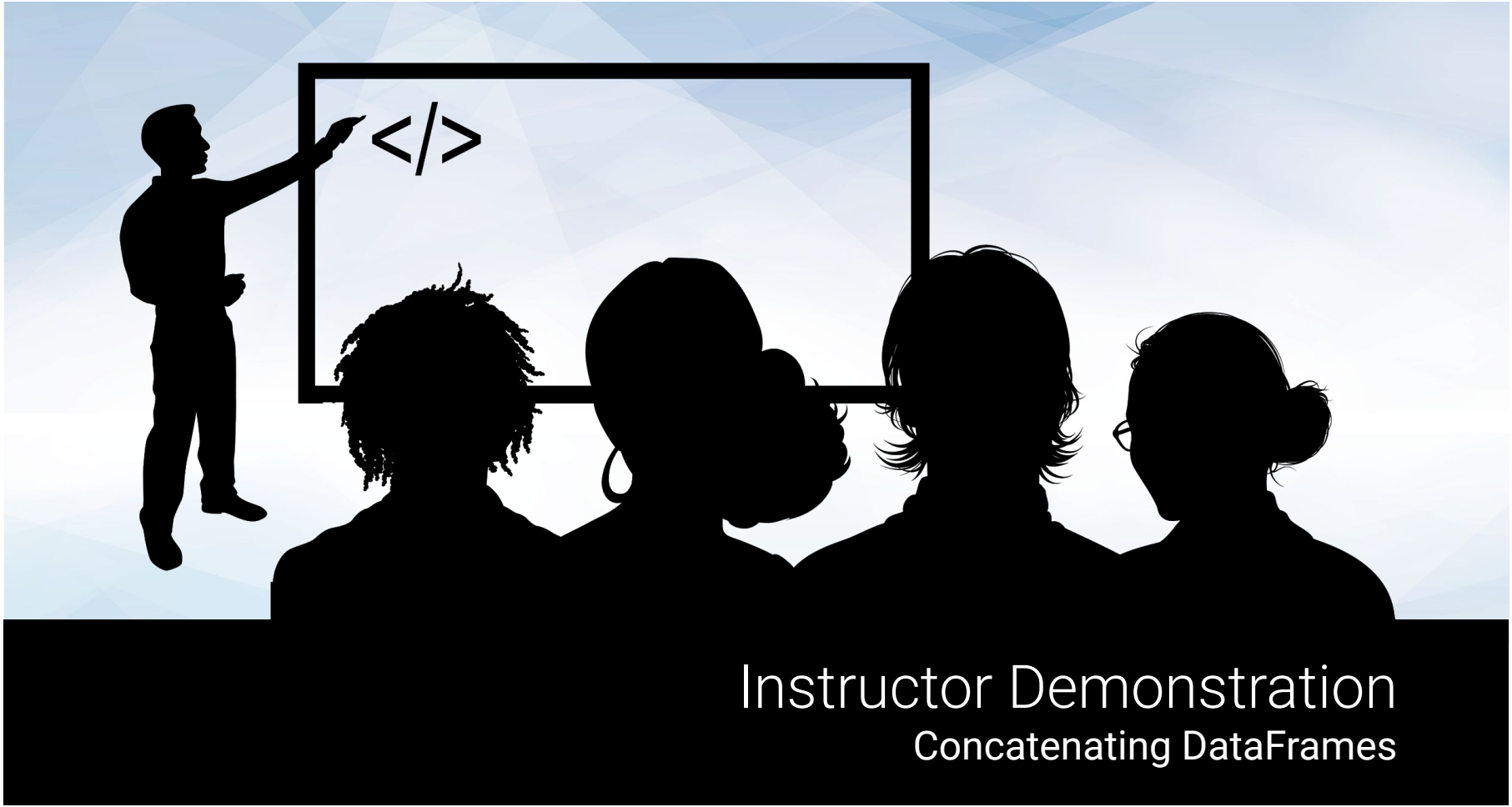
Join Operation: Inner Vs. Outer



Inner

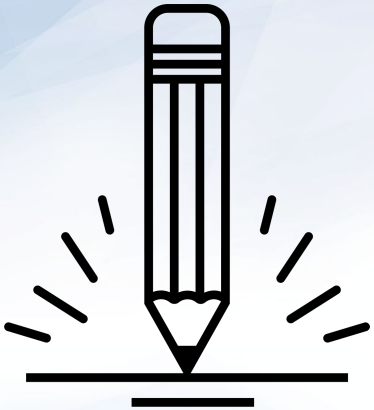


Outer



Instructor Demonstration

Concatenating DataFrames



Activity:
Mastering Concatenation

In this activity, you will combine multiple DataFrames using the `concat` function.

(Instructions sent via Slack.)

Suggested Time:
15 Minutes

Time’s Up! Let’s Review.

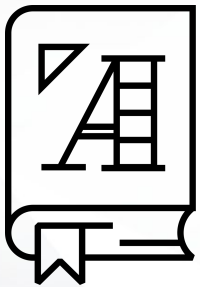
Standard Deviation and Risk

Risk Analysis

An essential aspect of analyzing portfolio and stock data is determining **risk**.

- With daily returns calculated, and data from multiple portfolios combined into one dataset, you can complete a holistic analysis of stock data.
- One component of risk is calculating the **mean** performance, or price, of a stock. The second is calculating the **standard deviation**.

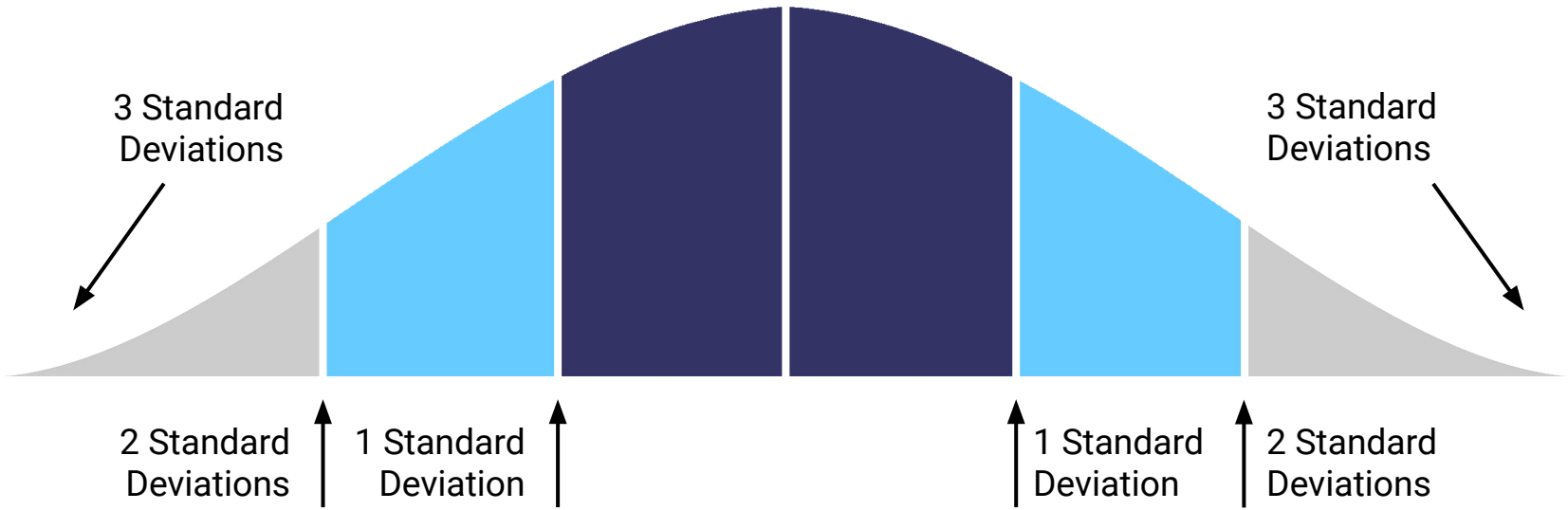
The **mean** can be used to determine the average value of a portfolio or stock over time.



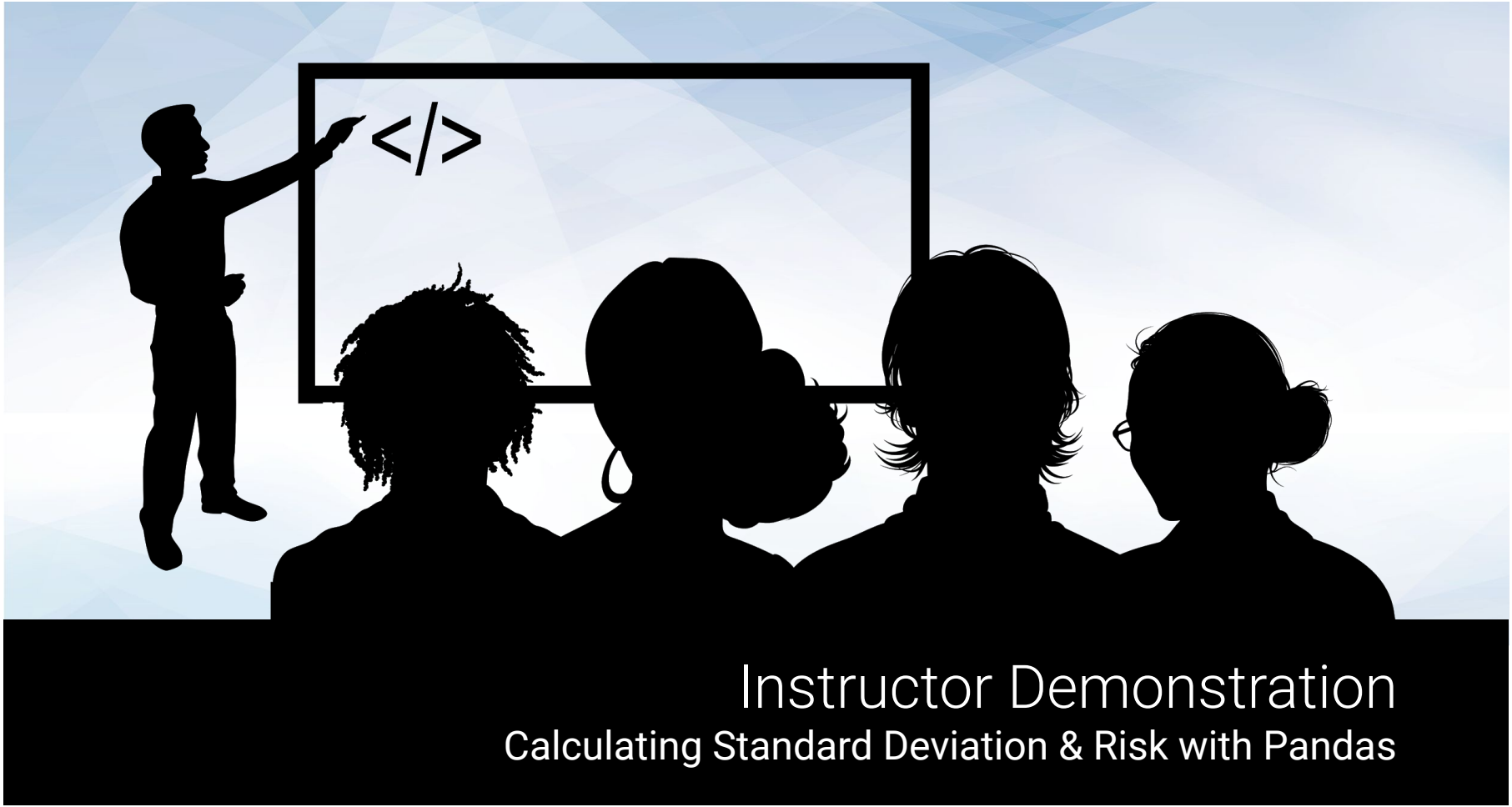
Standard deviation identifies precisely how far away a value is from the average.

Standard Deviation

The `std` Pandas function is used to calculate standard deviation for a DataFrame. Standard deviation can be used to determine the risk associated with an investment. Standard deviation is also used to calculate how much returns have been distributed from the average.



The greater the standard deviation, the greater the risk (and the potential) for a larger payout.



Instructor Demonstration

Calculating Standard Deviation & Risk with Pandas

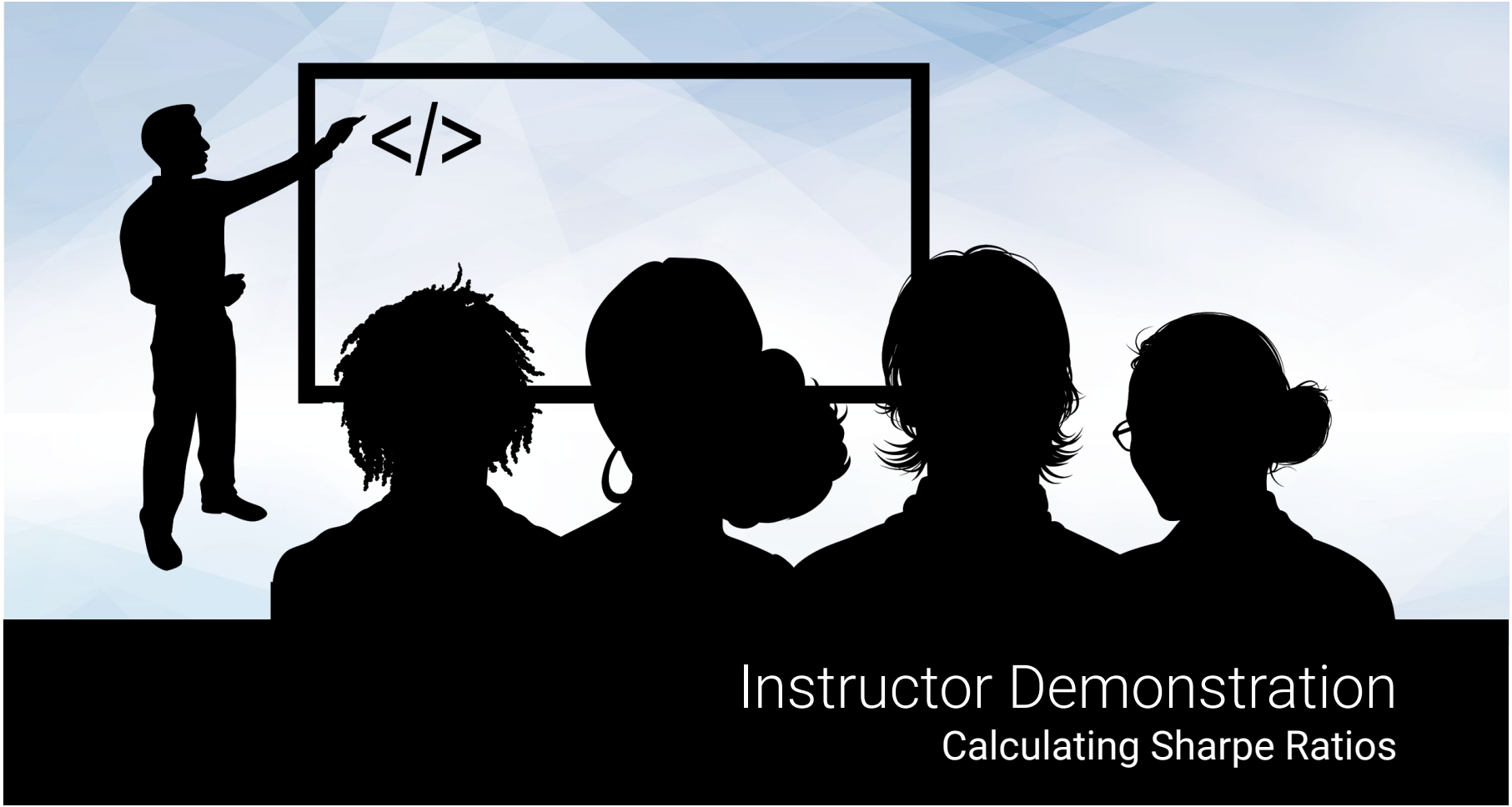
Sharpe Ratios

Sharpe Ratios

While standard deviation calculates how dispersed a set of values are, **Sharpe ratios** identify how much excess reward is associated with an investment after risk has been accounted for.

Sharpe ratios are calculated by dividing annualized average returns by annualized standard deviation.

$$\text{Sharpe Ratio} = \frac{R_p - R_f}{O_p}$$



Instructor Demonstration

Calculating Sharpe Ratios

Activity: Risky Business

In this activity, you will prep data and use standard deviation and Sharpe ratios to analyze cryptocurrency portfolio performance.

(Instructions sent via Slack.)

Suggested Time:
15 Minutes

Time’s Up! Let’s Review.

Congratulations!

You just levelled up and acquired the following skills:

- Data consolidation
- Data segmentation
- Data profiling and quality
- Investment and portfolio risk analysis

Questions?

The End