Egad! It’s Excel – Part 2

Data Boot Camp

Lesson 1.2

Quick Refresher

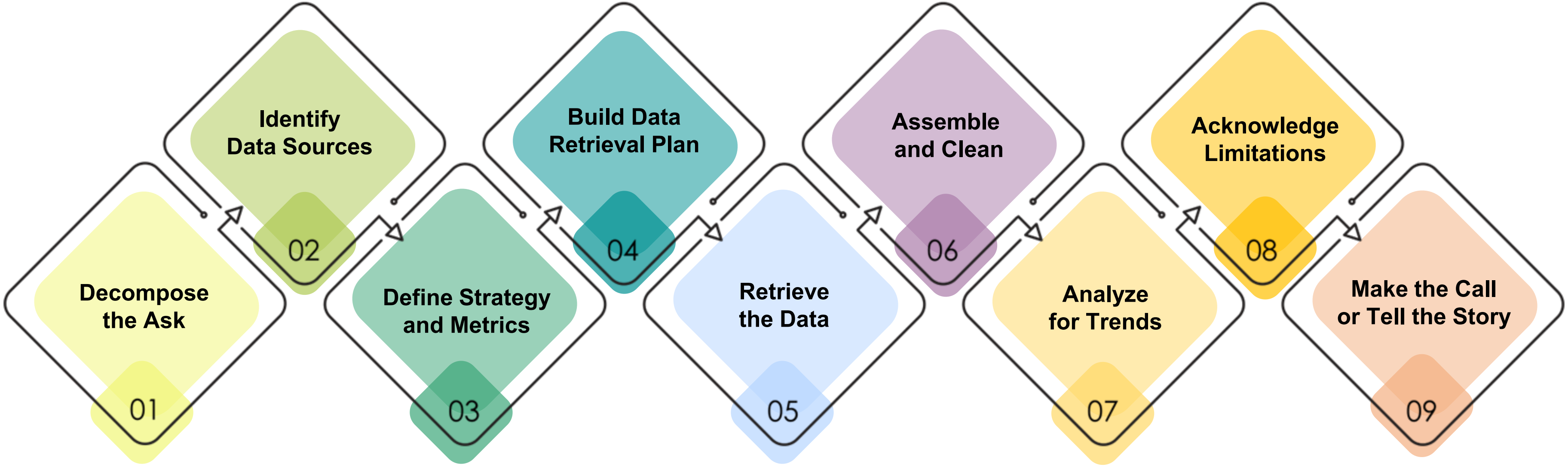
Data analytics is about what **two** things?

Fundamentally, data analytics is about **storytelling** and **truth-telling.**

What are the steps in the **Analytics Paradigm**?

Analytics Paradigm

Regardless of type or industry, this paradigm provides a repeatable pathway for effective data problem solving.

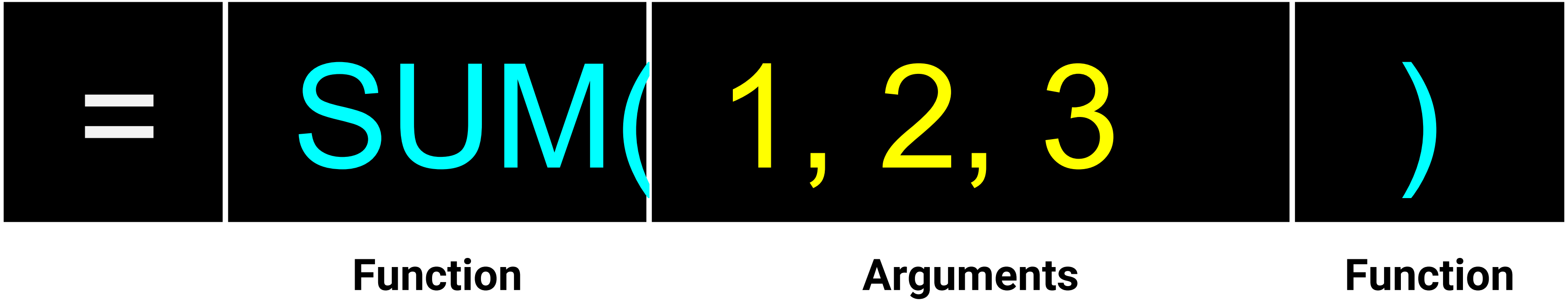


Let’s Start with the Basics

Formulas

Ooh…Coding! (Sort Of)

Excel has introduced you to a sort of proto-programming. When you write scripts, you will rely on functions (methods) that do something to or with arguments.



= SUM(1, 2, 3)

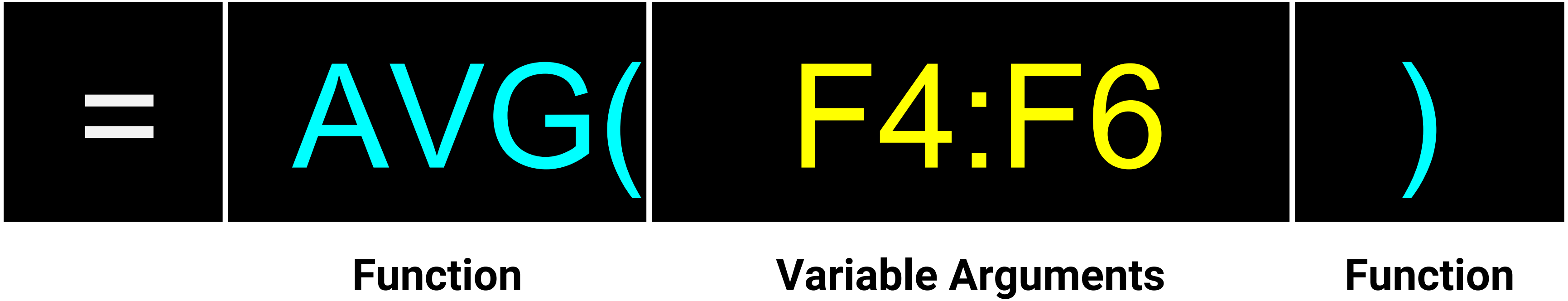
Function

Arguments

Function

Ooh…Coding! (Sort Of)

When we reference a range or a set of ranges, Excel is given a set of variable inputs. Excel will determine the actual values of these inputs prior to executing the function.



= AVG(F4:F6)

Function

Variable Arguments

Function

Ooh…Coding! (Sort Of)

What about this example?

Which is the function?

Which are the arguments?

```
= SUM( AVG(F4:F6), AVG(G4:G6) )
```

Ooh…Coding! (Sort Of)

Q

What about this example?

Which is the function?

Which are the arguments?

A

It depends.

```
= SUM( AVG(F4:F6), AVG(G4:G6) )
```

Ooh...Coding! (Sort Of)

What about this example?

Which is the function?

Which are the arguments?

The **AVG functions** take the provided ranges as their arguments.

```
= SUM(AVG(F4:F6), AVG(G4:G6) )
```

Ooh…Coding! (Sort Of)

Q

What about this example?

Which is the function?

Which are the arguments?

A

This is a **nested function**. We'll be doing plenty of complex nests in this class.

```
= SUM(AVG(F4:F6), AVG(G4:G6) )
```

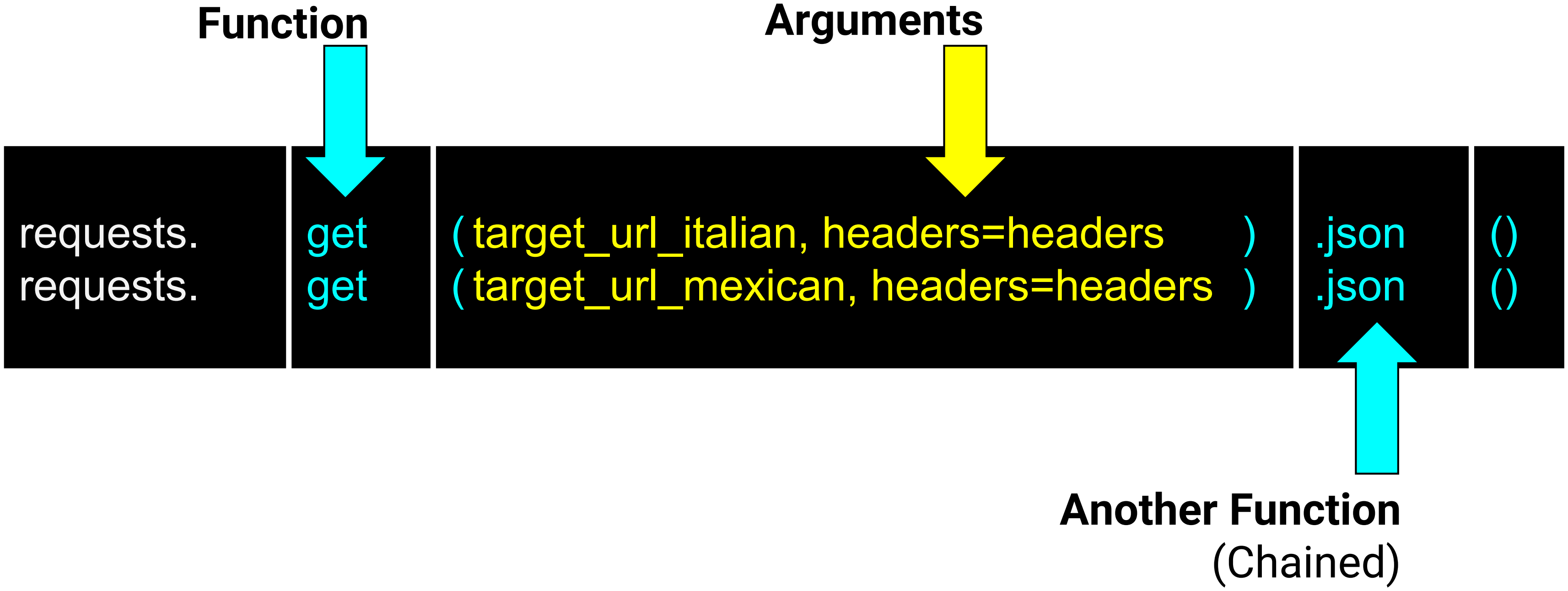
You Can Code Too!

Here’s a Python snippet from the last class.

```
requests.get(target_url_italian,        headers=headers).json()
requests.get(target_url_mexican,        headers=headers).json()
```

You Can Code Too!

Syntax and capabilities may differ across technologies and platforms, but fundamental concepts remain the same.

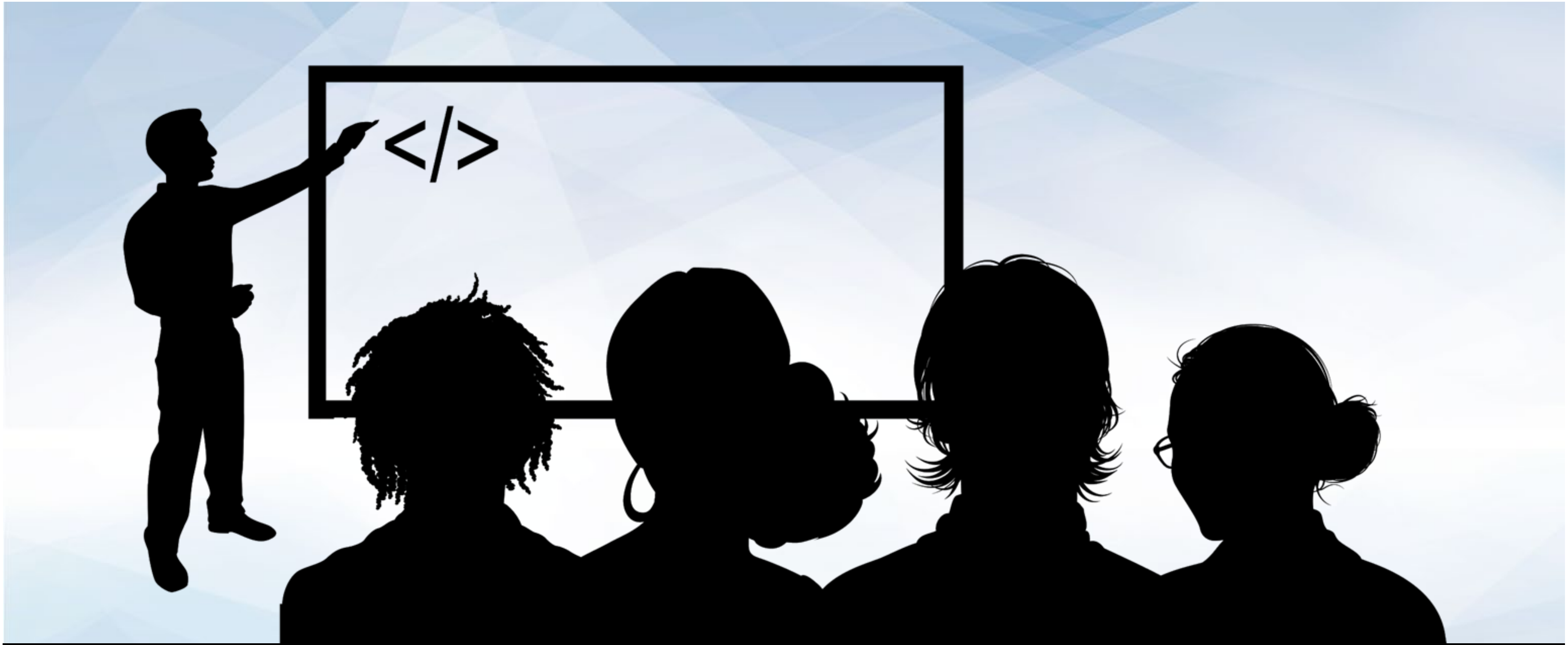


Function

Arguments

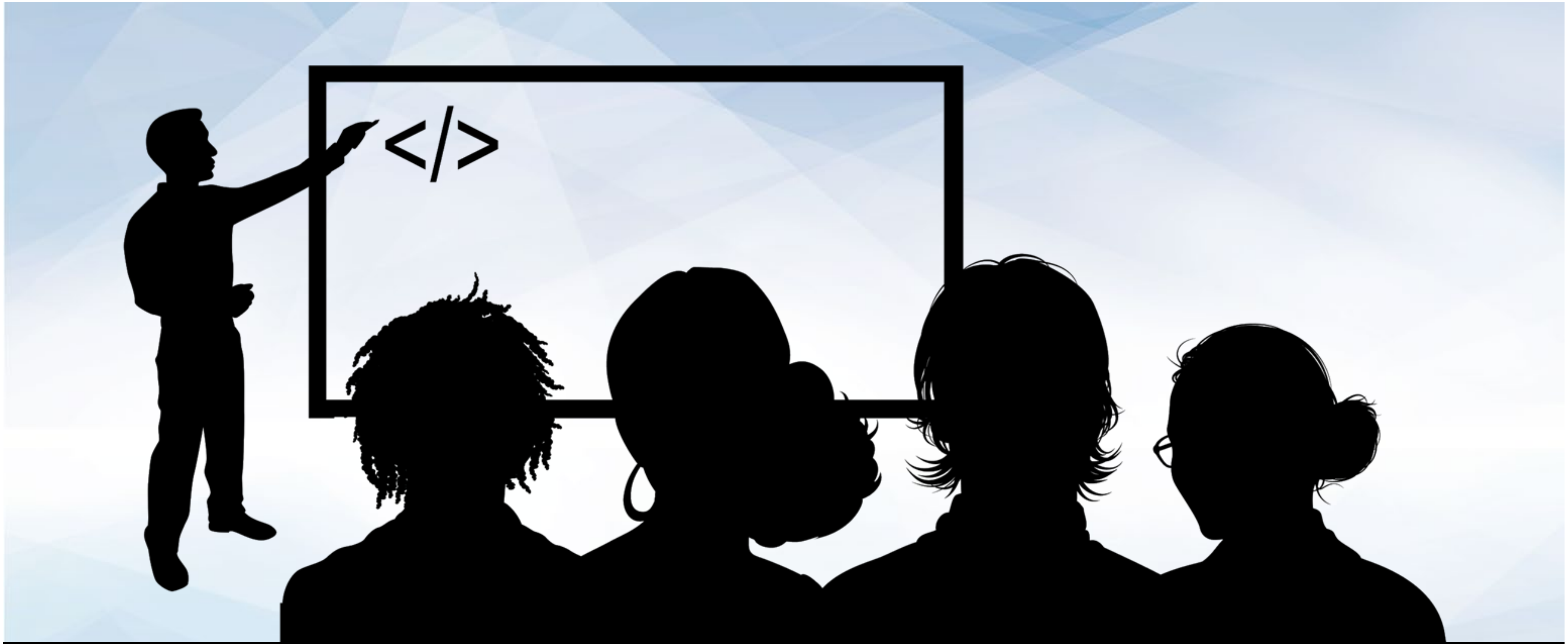
```
requests. get ( target_url_italian, headers=headers  ) .json ()
requests. get ( target_url_mexican, headers=headers ) .json ()
```

Another Function
(Chained)



Instructor Demonstration

Excel Playground

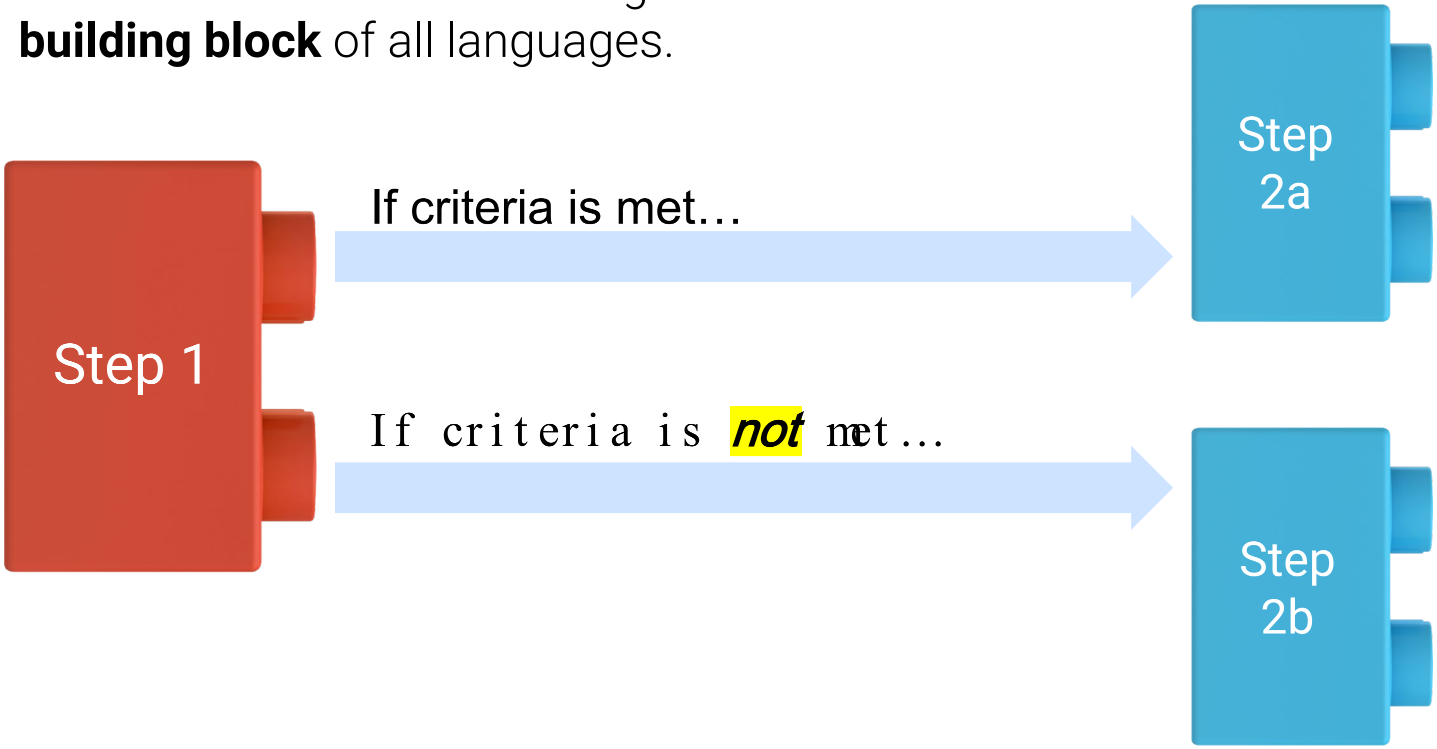


Instructor Demonstration

Named Ranges

Conditionals: If This, Then That

Conditionals present a way to control the flow of logic based on certain criteria being met. This is a **core building block** of all languages.



Step 1

If criteria is met…

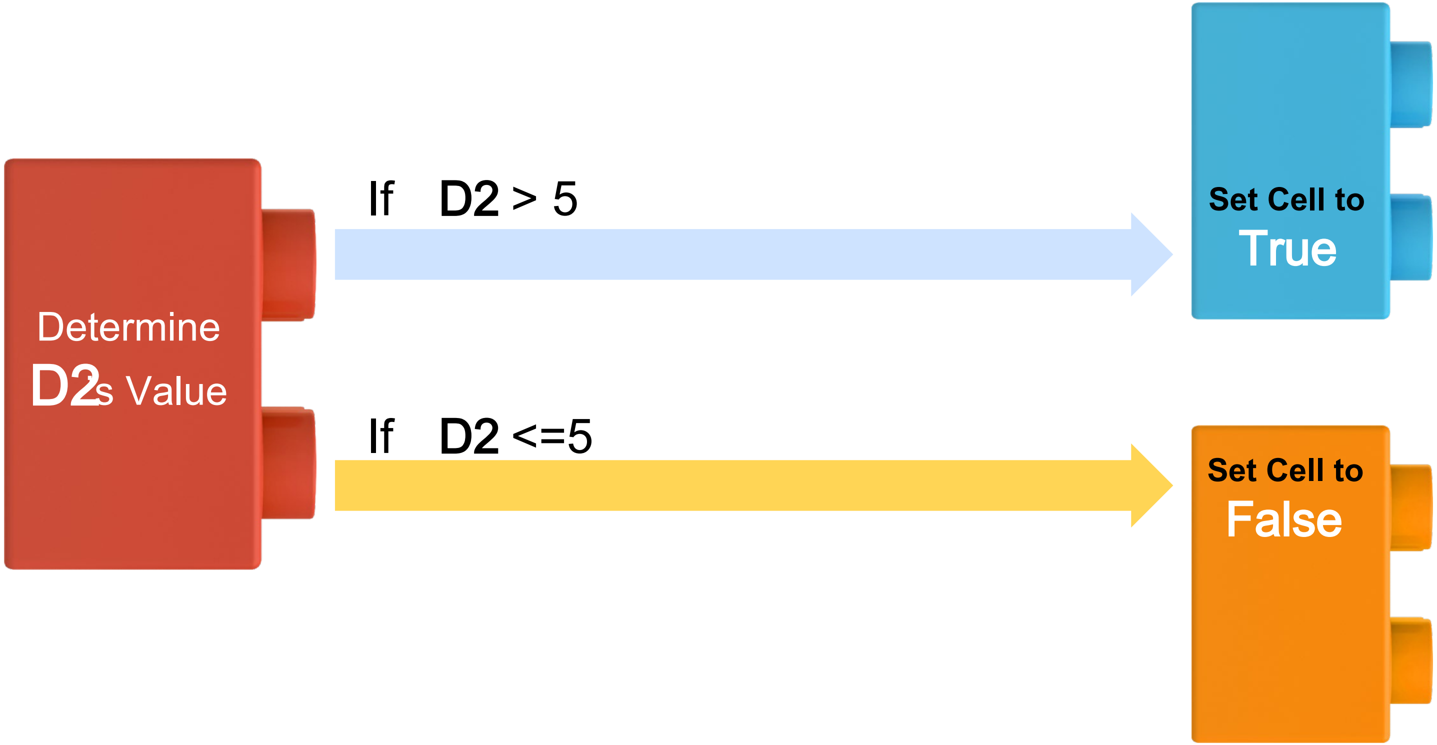
Step 2a

If criteria is *not* met…

Step 2b

Conditionals: If This, Then That

=IF(D2>5, TRUE, FALSE)



Determine D2s Value

If D2 > 5

Set Cell to True

If D2 <=5

Set Cell to False

But what if we want to **combine** conditions?

AND, NOT, OR

Ooh...Coding! (Sort Of)

Q

But what if we want to **combine** conditions?

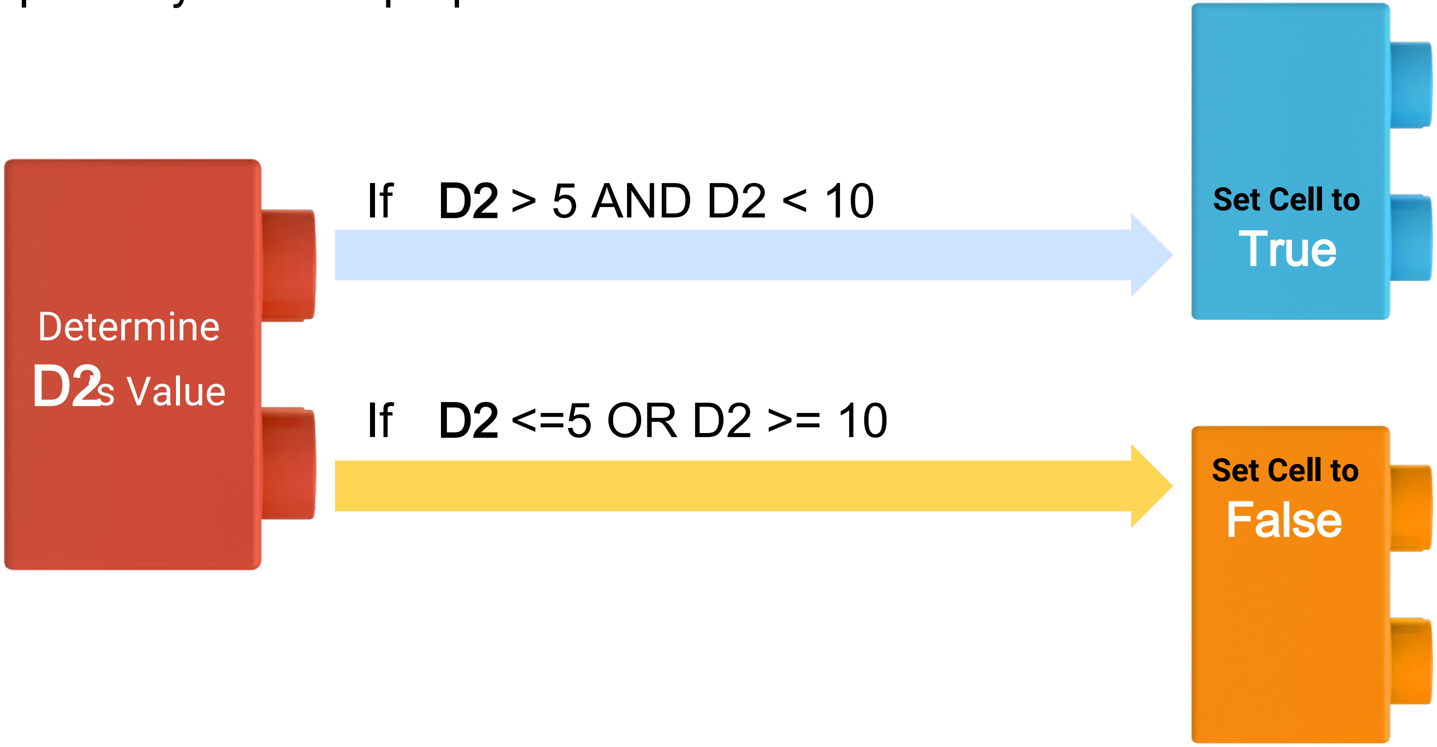
A

AND, NOT, OR

```
=IF( AND(D2>5, D2<10),TRUE,FALSE)
```

Conditionals: If This, Then That

Nesting conditionals can quickly become a very convoluted (but necessary) part of your data prep.



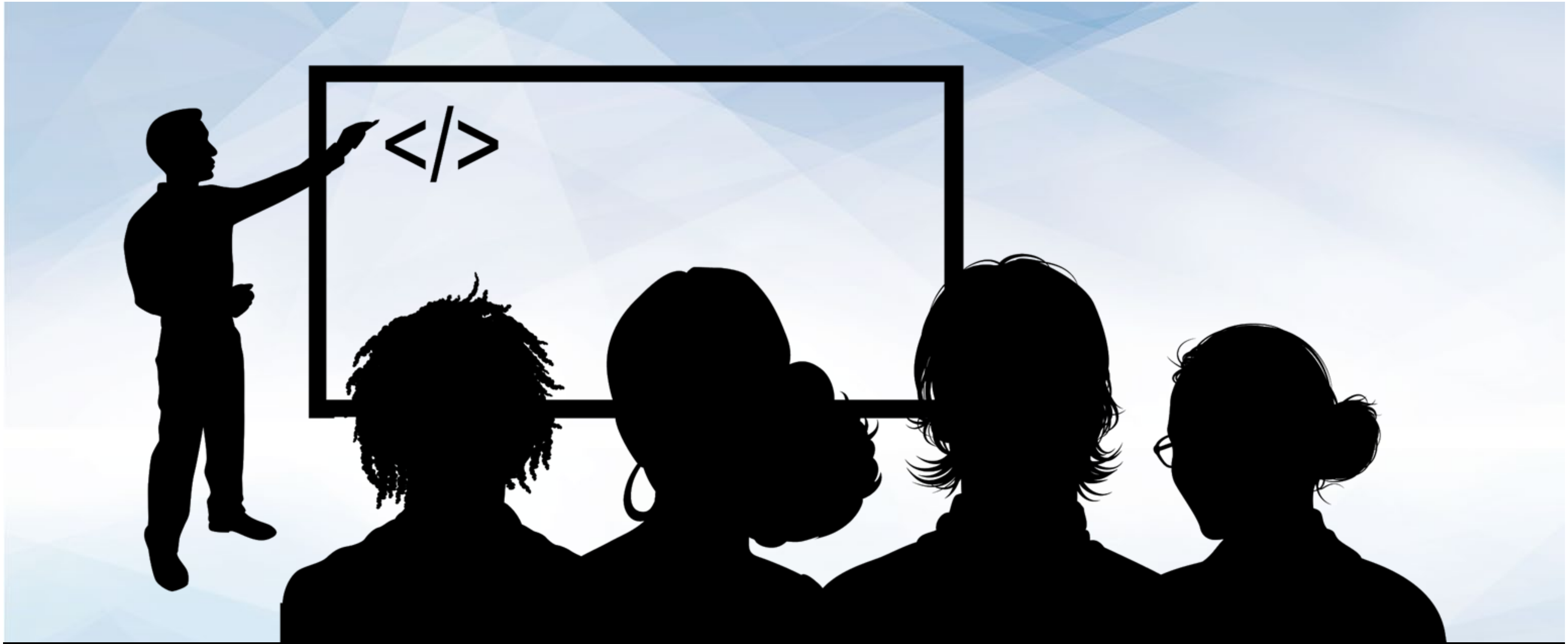
Determine D2s Value

If **D2** > 5 AND D2 < 10

Set Cell to True

If **D2** <=5 OR D2 >= 10

Set Cell to False



Instructor Demonstration

Color Counter

Everybody Do: Gradebook

Create a formula that calculates the final grade for a student based on their previous exams and papers.

Suggested Time:
7 minutes

Activity: Gradebook

When making this calculation:

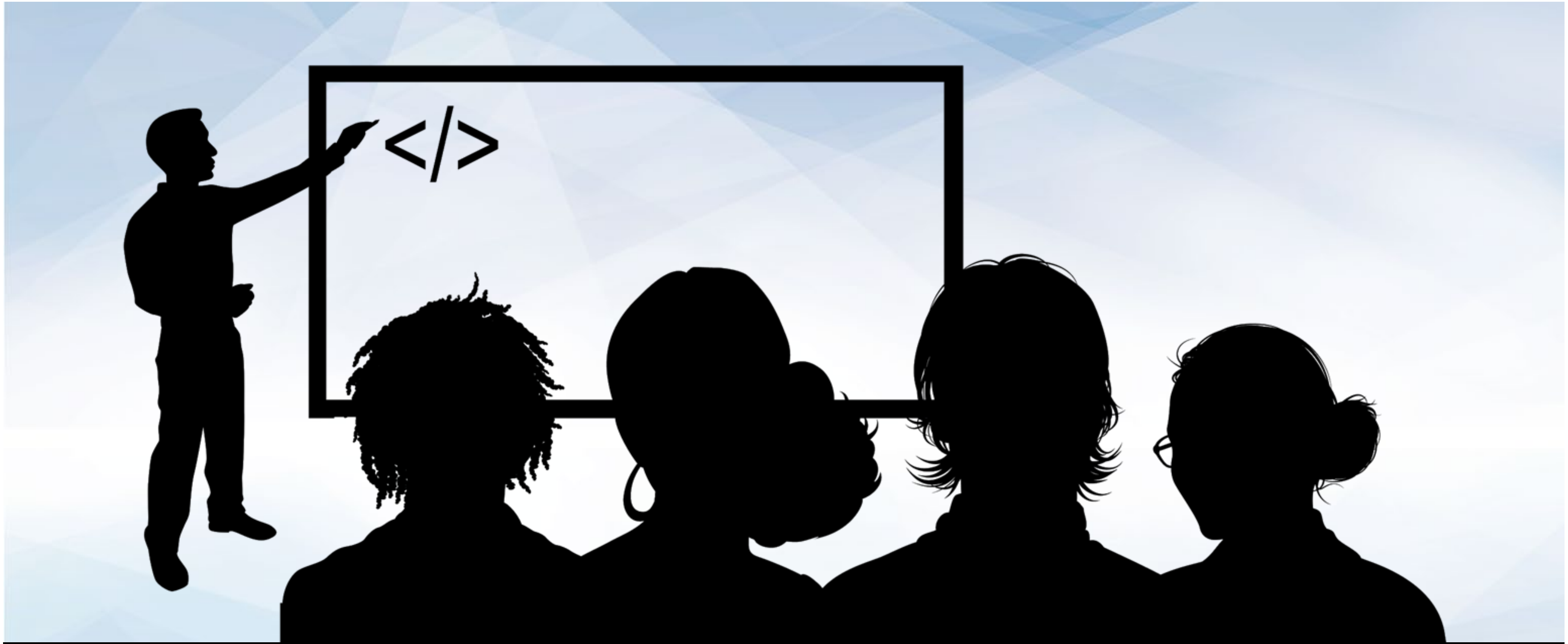
- Consider every paper and exam to be equal in weight; each should comprise one-fourth of the overall grade.
- Round the result to the nearest integer.
- Using conditionals, create a formula that returns PAS if a student's final grade is greater than or equal to 60. If a student's final grade is below 60, the formula should return FAIL.

Bonus:

Create a nested IF() formula that returns a letter grade based on a student's final grade.

- Greater than or equal to 90 = A
- Greater than or equal to 80 and less than 90 = B
- Greater than or equal to 70 and less than 80 = C
- Greater than or equal to 60 and less than 70 = D
- Anything less than 60 = F

Suggested Time: 7 minutes



Instructor Demonstration

Multiple Worksheets

Everybody Do:
Apples and Oranges

Apples and oranges are two completely different fruits whose data sets are being stored on two completely different worksheets. In this activity, you will take two sheets of data and compare their values against each other to determine which is the ultimate fruit!

Suggested Time:
7 minutes

Activity: Apples and Oranges

Instructions:

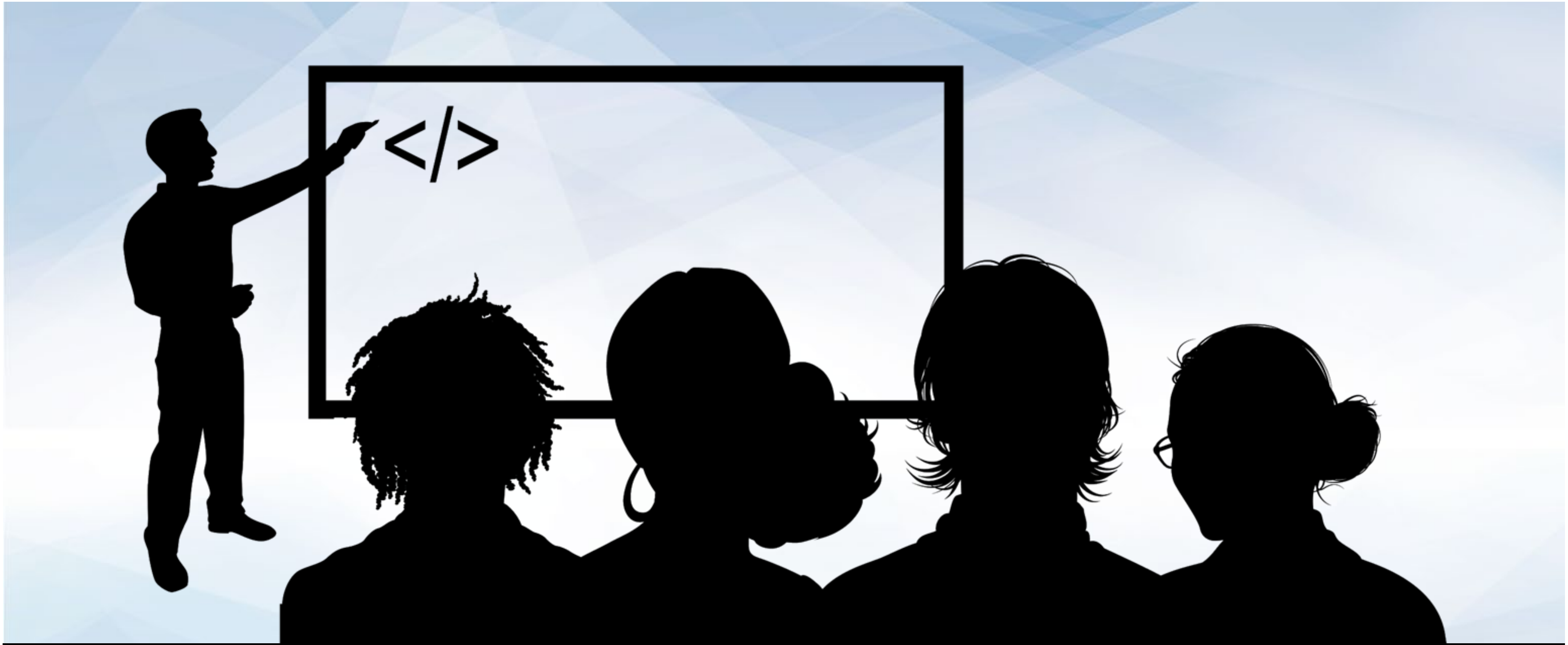
- In each Apples and Oranges worksheet, calculate the average values of the data contained in each column.
- Create a new worksheet that will be used to compare characteristics of apples and oranges against each other.
- Create 5 columns with the headers *Durability, Ease of Use, Taste, Juiciness,* and *Aerodynamics*.
- Create 3 rows with the headers *Apples, Oranges,* and *Winner*.
- Pull in the averages data from the Apples and Oranges worksheets, placing them in their respective cells in the new worksheet.
- Using a conditional formula, determine whether Apples or Oranges hold the greater average value per column and print the winner in the *Winner* row.

Bonus:

- Write a formula to determine the overall winner by counting which has the most wins.
- Make another worksheet that calculates the averages. These averages should be populated in the new worksheet without referencing the averages calculated in the original worksheets.

Suggested Time: 7 minutes

Time’s Up! Let’s Review.



Instructor Demonstration

Formatting

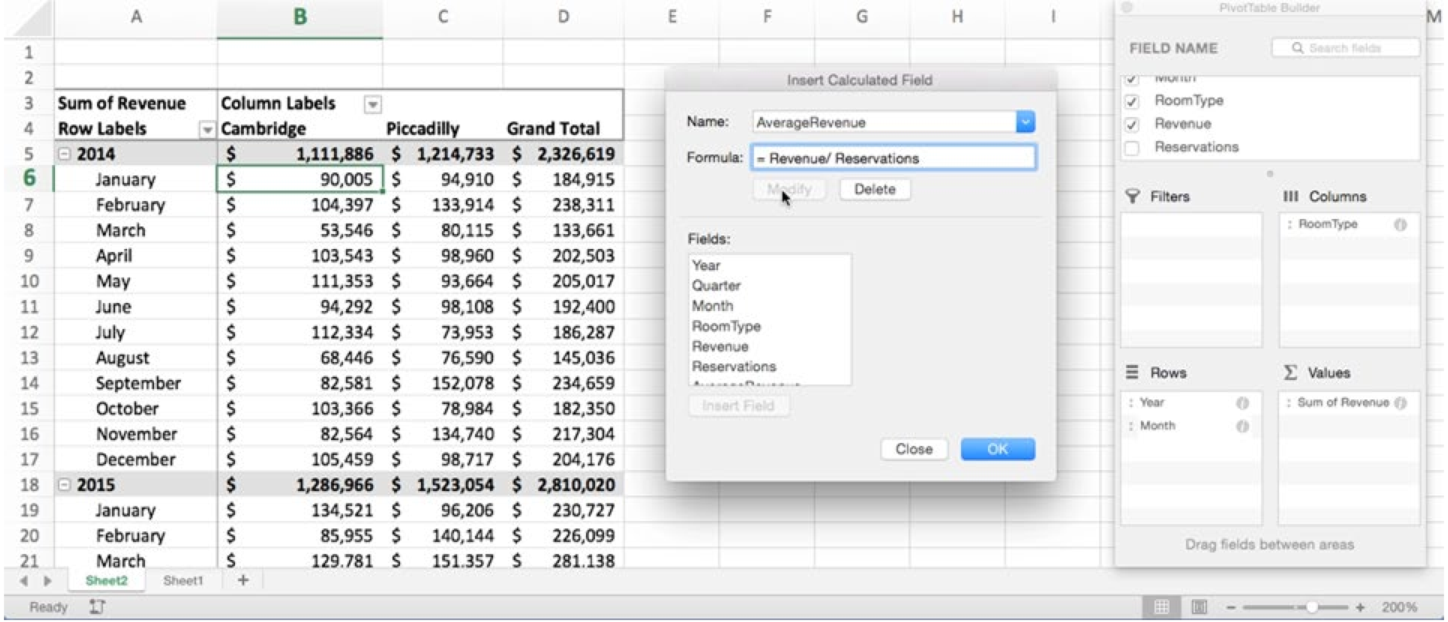
Take a Break!

Time’s Up! Let’s Review.

Pivot Tables

Get Pivot With It

Pivot tables are one of the most important data visualization concepts to master in this class. (Don't worry. They are a cinch to deal with.)



Sum of Revenue	Column Labels		
Row Labels	Cambridge	Piccadilly	Grand Total
⊟ 2014	**$ 1,111,886**	**$ 1,214,733**	**$ 2,326,619**
January	$ 90,005	$ 94,910	$ 184,915
February	$ 104,397	$ 133,914	$ 238,311
March	$ 53,546	$ 80,115	$ 133,661
April	$ 103,543	$ 98,960	$ 202,503
May	$ 111,353	$ 93,664	$ 205,017
June	$ 94,292	$ 98,108	$ 192,400
July	$ 112,334	$ 73,953	$ 186,287
August	$ 68,446	$ 76,590	$ 145,036
September	$ 82,581	$ 152,078	$ 234,659
October	$ 103,366	$ 78,984	$ 182,350
November	$ 82,564	$ 134,740	$ 217,304
December	$ 105,459	$ 98,717	$ 204,176
⊟ 2015	**$ 1,286,966**	**$ 1,523,054**	**$ 2,810,020**
January	$ 134,521	$ 96,206	$ 230,727
February	$ 85,955	$ 140,144	$ 226,099
March	$ 129,781	$ 151,357	$ 281,138

Insert Calculated Field

Name: AverageRevenue

Formula: = Revenue/ Reservations

Fields:
Year
Quarter
Month
RoomType
Revenue
Reservations

PivotTable Builder

FIELD NAME

☑ Month
☑ RoomType
☑ Revenue
□ Reservations

Filters

Columns: RoomType

Rows: Year, Month

Values: Sum of Revenue

Drag fields between areas

Get Pivot With It

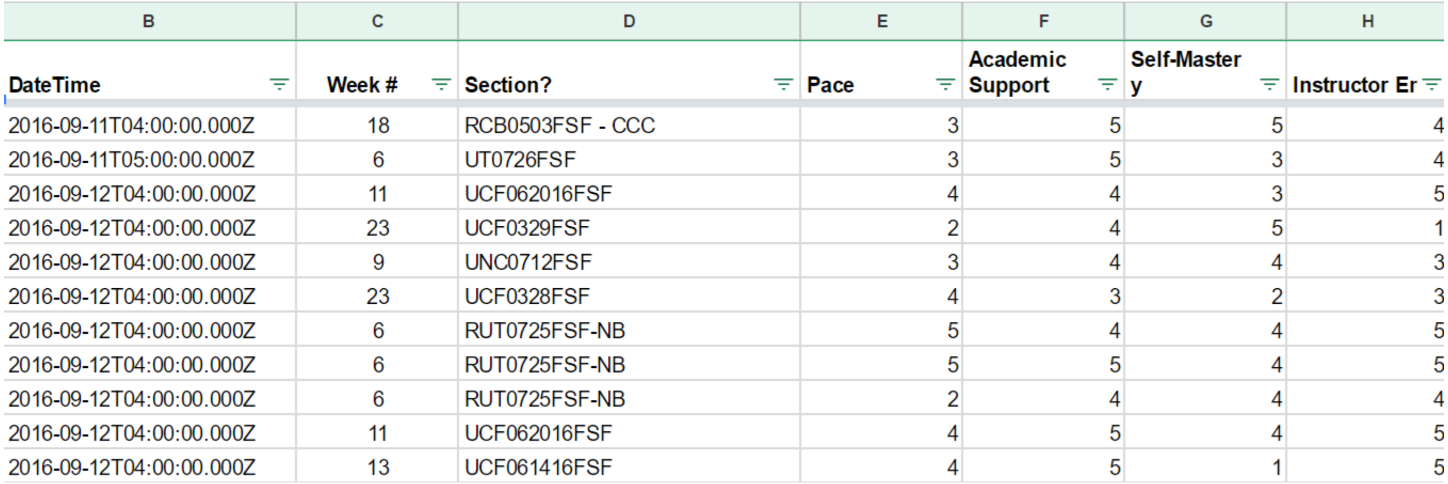
In essence, a pivot table is a **summative** analytic tool that allows us to perform aggregate functions that allow any combination of fields. (The term *pivot table* comes from the fact that we are pivoting along a data axis).

Seller	Qty. Sold	Date
Joseph	$42.50	1/1/17
Jacob	$65.00	1/3/17
Jacob	$5.25	1/6/17
Joseph	$125.00	1/6/17
Jacob	$3.50	1/7/17
Matt	$32.00	1/9/17

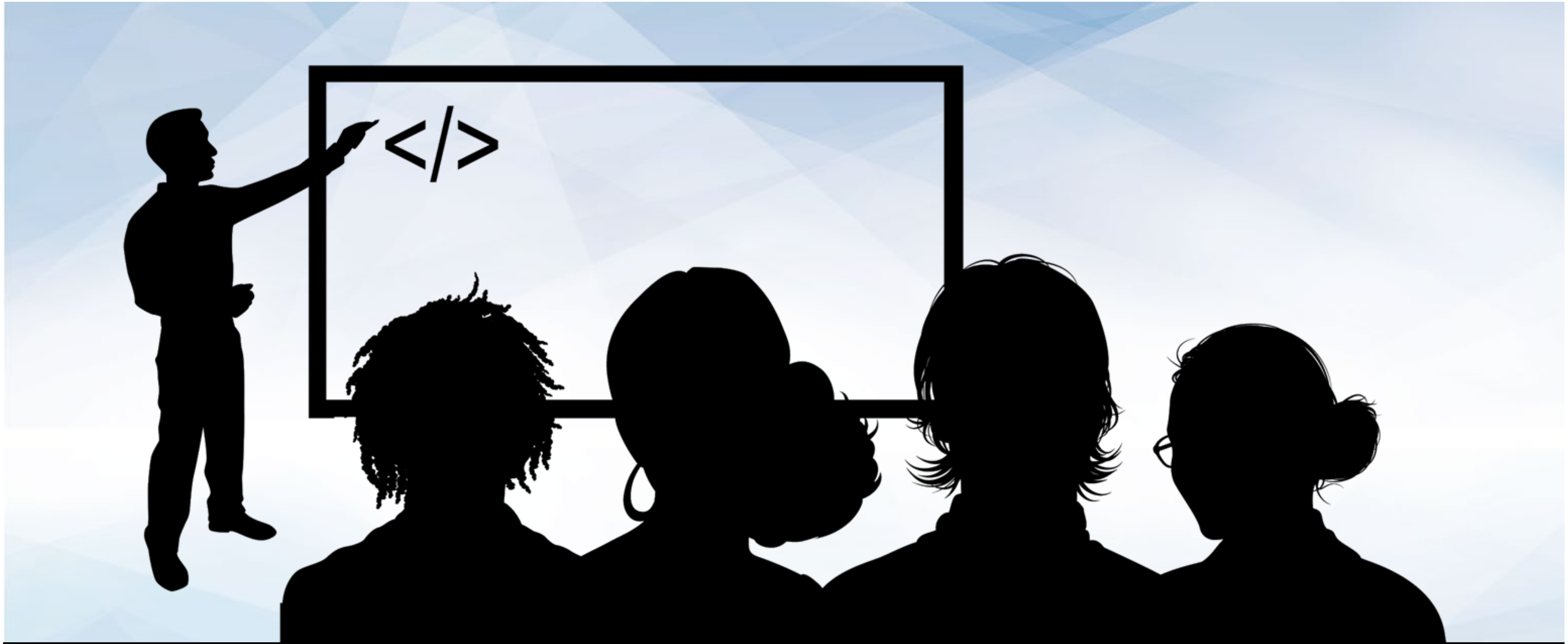
Seller	Total Sold
Joseph	$167.50
Jacob	$73.75
Matt	$32.00

Word to the Wise: Keep It Flat!

Modern Business Intelligence (BI) tools like Tableau, Sisense, and Salesforce work best if data is stored in flat CSVs—meaning column headers represent fields (vertically) on the spreadsheet. This is largely because all of these technologies heavily utilize pivot tables as a tool for their visualizations. **Don’t try to confuse this simplicity. “Spreadsheet magic” is a nightmare to analyze.**



B	C	D	E	F	G	H
DateTime	Week #	Section?	Pace	Academic Support	Self-Mastery	Instructor Er
2016-09-11T04:00:00.000Z	18	RCB0503FSF - CCC	3	5	5	4
2016-09-11T05:00:00.000Z	6	UT0726FSF	3	5	3	4
2016-09-12T04:00:00.000Z	11	UCF062016FSF	4	4	3	5
2016-09-12T04:00:00.000Z	23	UCF0329FSF	2	4	5	1
2016-09-12T04:00:00.000Z	9	UNC0712FSF	3	4	4	3
2016-09-12T04:00:00.000Z	23	UCF0328FSF	4	3	2	3
2016-09-12T04:00:00.000Z	6	RUT0725FSF-NB	5	4	4	5
2016-09-12T04:00:00.000Z	6	RUT0725FSF-NB	5	5	4	5
2016-09-12T04:00:00.000Z	6	RUT0725FSF-NB	2	4	4	4
2016-09-12T04:00:00.000Z	11	UCF062016FSF	4	5	4	5
2016-09-12T04:00:00.000Z	13	UCF061416FSF	4	5	1	5



Instructor Demonstration

Pivot Tables

Everybody DO:
Top Songs Pivot Table

In this activity, you will use a 5000 row spreadsheet containing data for the top 5000 songs from 1901 onward. Using pivot tables, you will uncover which artists have the most songs in the top 5000, the song titles, and the year each song was released.

Suggested Time:
10 minutes

Challenge: Top Songs Pivot Table

Instructions:

- Select all of the data in your worksheet and create a new pivot table.
- Make a pivot table that can be filtered by year and contains two rows: *Artist* and *Name*.
- All of an artist’s songs should be listed below their name.
- Update your pivot table to contain values for:
  - How many songs an artist has in the top 5000
  - The sum of the final_score of their songs.
- Sort your pivot table by descending sum of the final_score.

Suggested Time: 10 minutes

Lookups

Look It Up with Lookups

Assume this table is gigantic. How would we **retrieve** the population of a specific planet for use in another formula?

Planet	**Population**
Zeelo	5020
Merinoa	380
Cardboard Box	2
…	…
Asteroid 9	95

Look It Up with Lookups

Q

Assume this table is gigantic. How would we **retrieve** the population of a specific planet for use in another formula?

A

=vlookup(<value>, <full table>,
<column to retrieve>,<match parameter>)

Planet	Population
Zeelo	5020
Merinoa	380
Cardboard Box	2
…	…
Asteroid 9	95

Look It Up with Lookups

What will this yield?
=vlookup(“Asteroid 9”, Planets, 3, FALSE)

Planet	Population	Species
Zeelo	5020	Zoltans
Merinoa	380	Murphies
Cardboard Box	2	Hambones
…	…	
Asteroid 9	95	Asterisks

Look It Up with Lookups

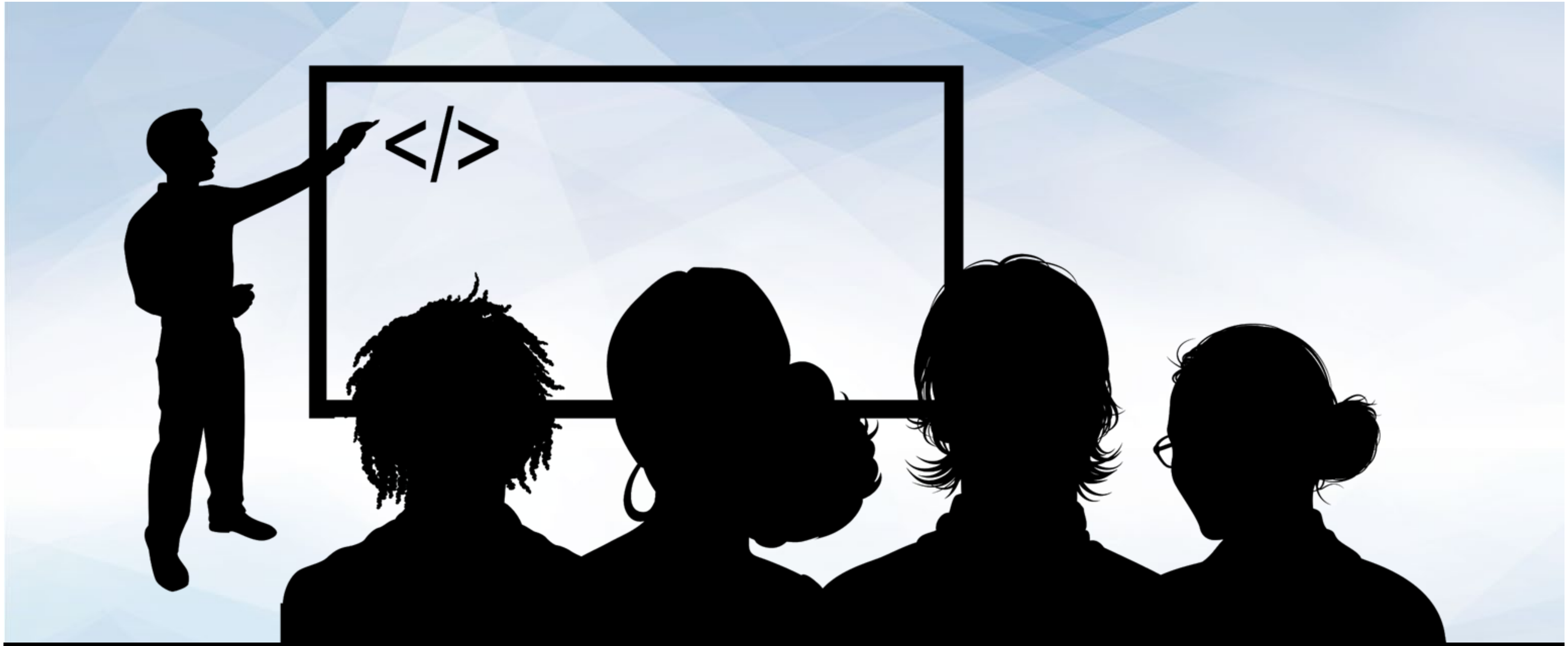
Q

What will this yield?
=vlookup(“Asteroid 9”, Planets, 3, FALSE)

Planet	Population	Species
Zeelo	5020	Zoltans
Merinoa	380	Murphies
Cardboard Box	2	Hambones
…	…	
Asteroid 9	95	Asterisks

A

Asterisks



Instructor Demonstration

Lookups

Partner Activity: Product Pivot

A small company selling electronics and electronic media has asked our class to create a table that visualizes the cost of their recent orders. Using lookups, create a pivot table that serves this purpose.

Suggested Time:
10 minutes

Activity: Product Pivot

Determine the Product Price of each row in the Orders sheet by using a VLOOKUP() that references each row's Product ID. (The Product Price of a row does not include shipping.)

Determine the Shipping Price of each row in the Orders sheet by using a VLOOKUP() that references each row's Product ID.

Select all data on the Orders sheet and create a new pivot table that calculates the sum of both Product Price and Shipping Price for each Order Number and Product ID.

Suggested Time: 10 minutes

Time’s Up! Let’s Review.

Extra Challenge: McDonald's Formatting

Despite attempts to convince people otherwise, the McDonald's menu is not well known for having the healthiest food options of all time. Looking at a massive table filled with nutrition facts is a pretty daunting task, though, so it's up to the class to make this spreadsheet just a little bit easier to understand.

Suggested Time:
10 minutes

Extra Challenge: McDonald’s Formatting

Instructions:

- Apply conditional formatting to all values in columns C and D that formats cells based on their values, and modifies the color of cells so that lighter shades occur for low values, while darker shades occur for high values.
- Apply conditional formatting to columns F, H, K, M, O, Q, T, U, V, and W that turns a cell green when its value is less than 50, and red when it is greater than or equal to 50.
- Add a thin-line border to all values in the table.
- Add a thick horizontal border that underscores the top row of the table.
- Add a thick vertical border down the right side of column A, and apply text wrapping to all values in that column.

Suggested Time: 10 minutes

Questions?