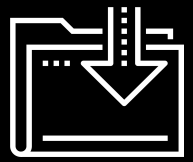
Data Fundamentals in Excel

Data Boot Camp

Lesson 1.2



A Few Admin Things

Class Repository and Zoom Video Feed

Class Git Repository

Classroom content and Challenge assignments

Class Videos

Automatically uploaded, on-demand videos

Quick Refresher

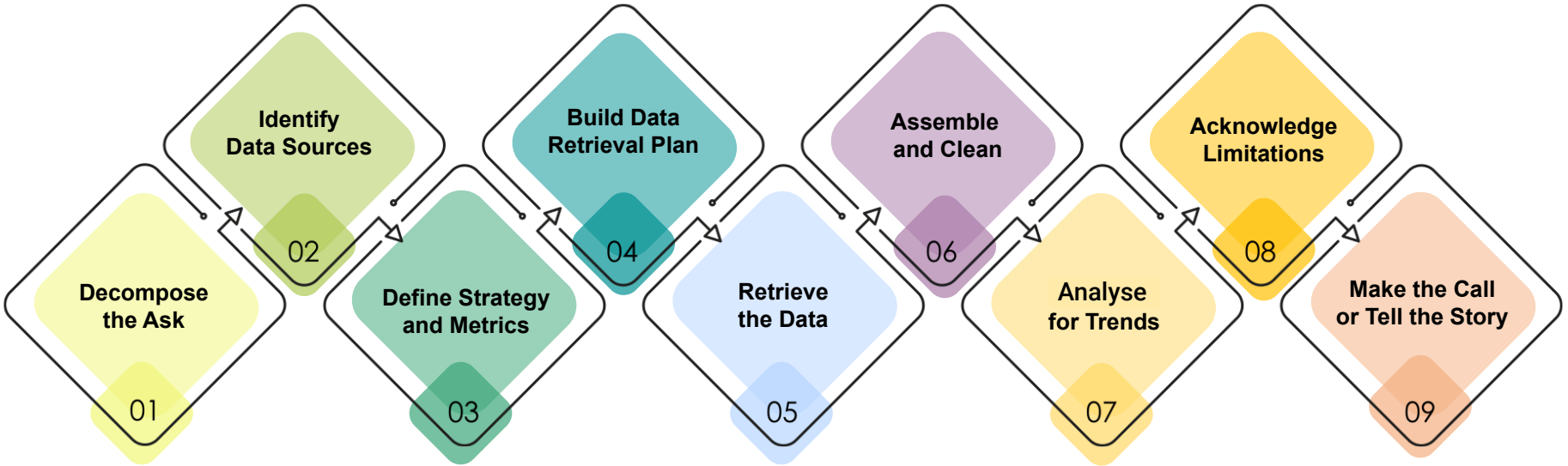
Data analytics is about what two things?

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

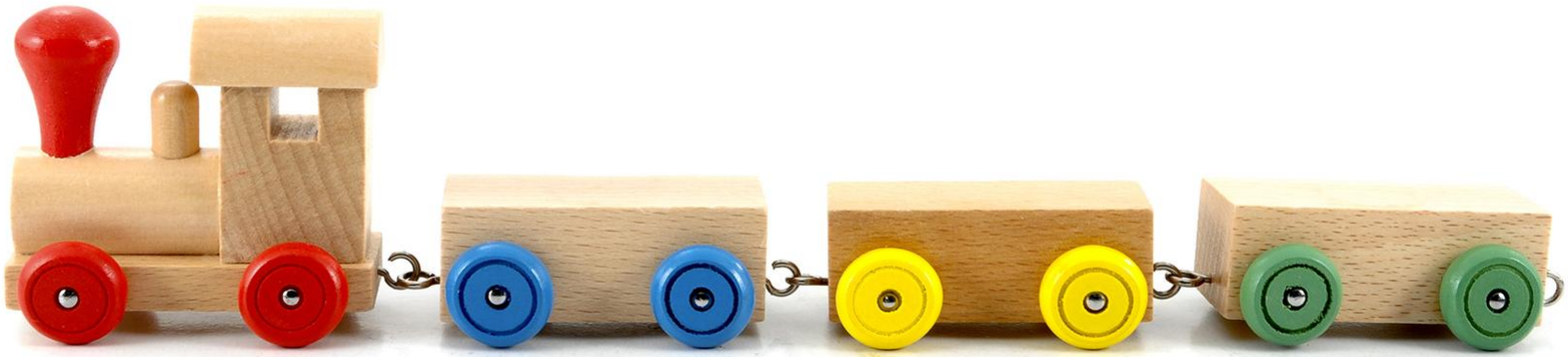
What are the steps in the analytics paradigm?

Analytics Paradigm

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



Let’s Start with the Basics



Instructor Demonstration

Excel Playground

Excited to get started?

Formulas

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`.

Coding (Sort Of)

When we reference a range or a set of ranges, Excel gets a set of `variable` inputs. Excel determines the values of these inputs before running the function.

Coding (Sort Of)

What about this example?
Which is the function?
Which are the arguments?

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

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) )
```

Coding (Sort Of)

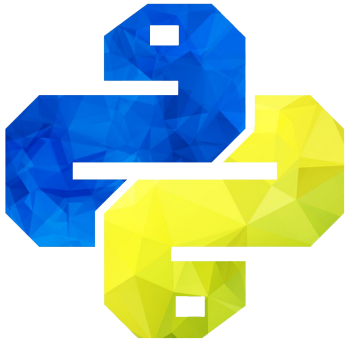
What about this example?
Which is the function?
Which are the arguments?

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

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

You Can Also Code

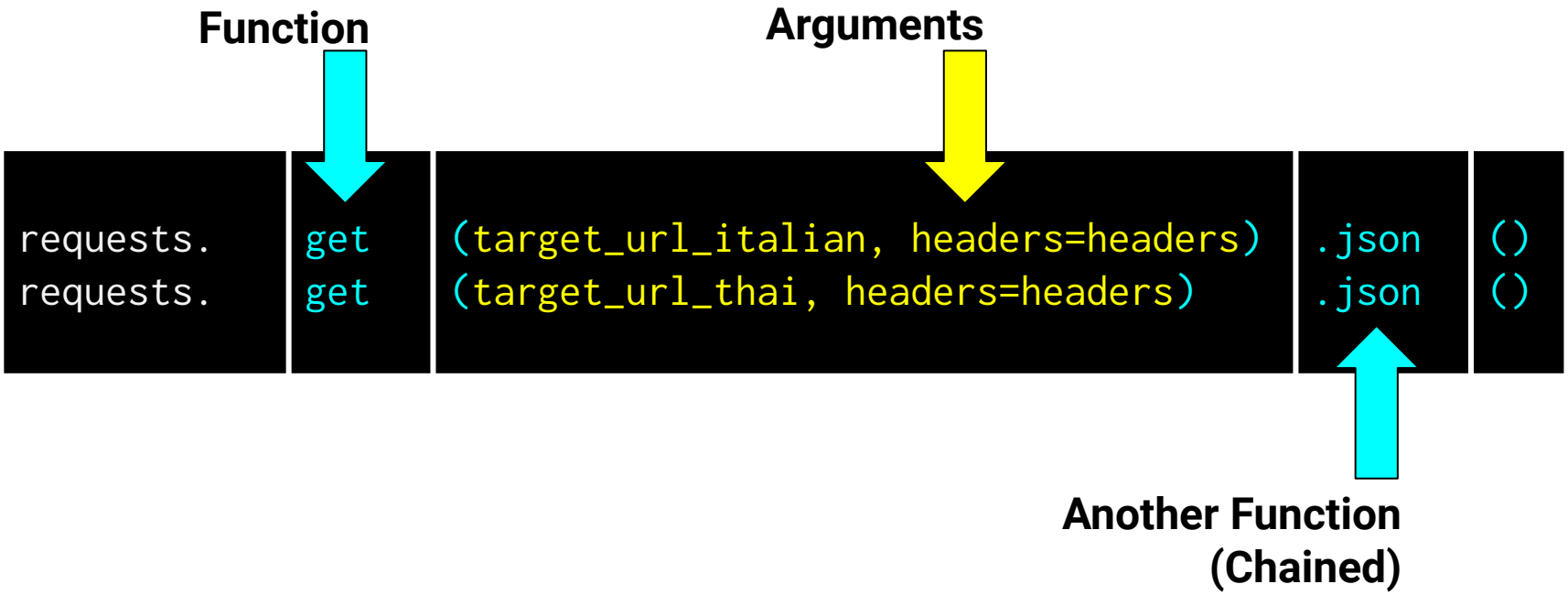
Here's a Python snippet from the last class.



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

You Can Also Code

The syntax and capabilities might differ across technologies and platforms, but the fundamental concepts remain the same.



Instructor Demonstration

Named Ranges

There are multiple ways to select data in a formula

Most of us learned to select a range of cells to input into a function.

There are multiple ways to select data in a formula

But, we can name a range of values to make interpreting formulas easier.

```
=AVG(A1:A10)
```

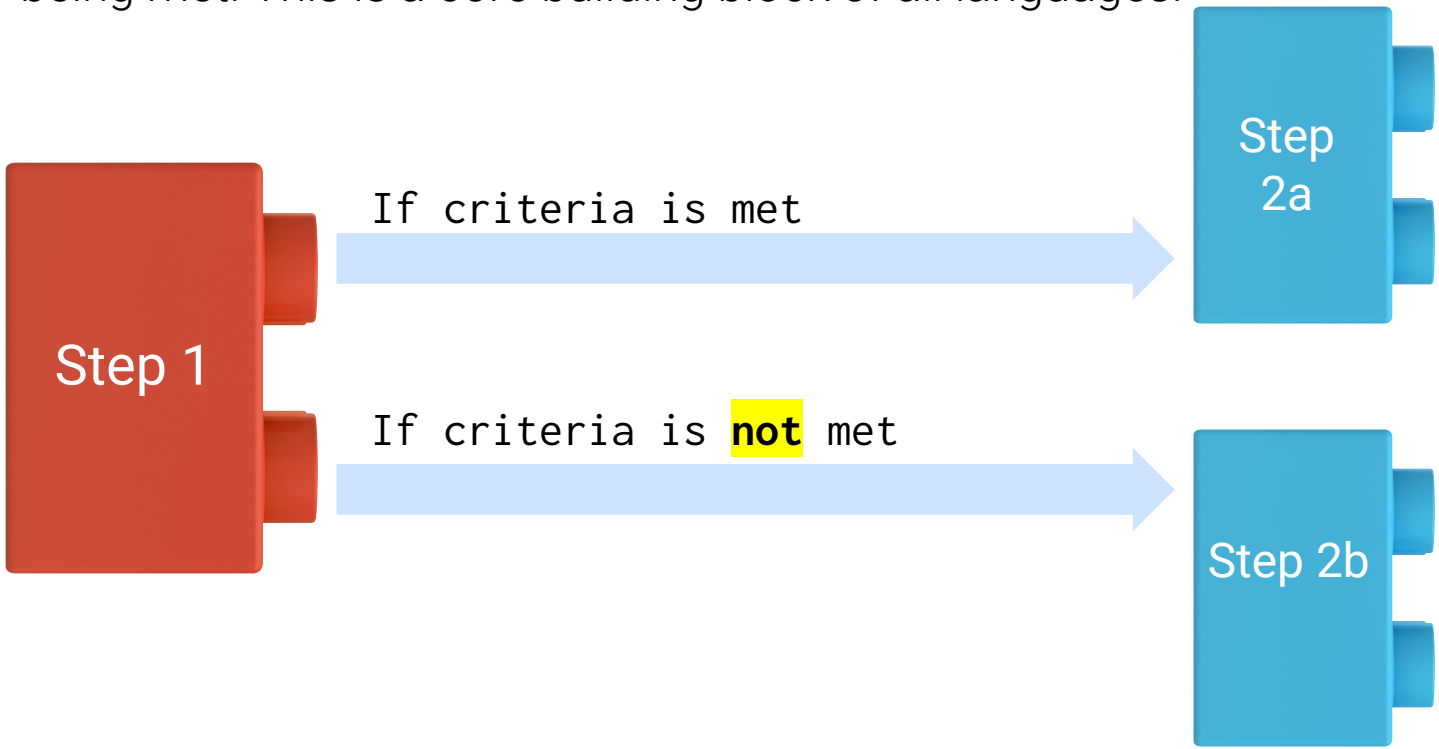
```
=AVG(prices)
```

Instructor Demonstration

Colour Counter

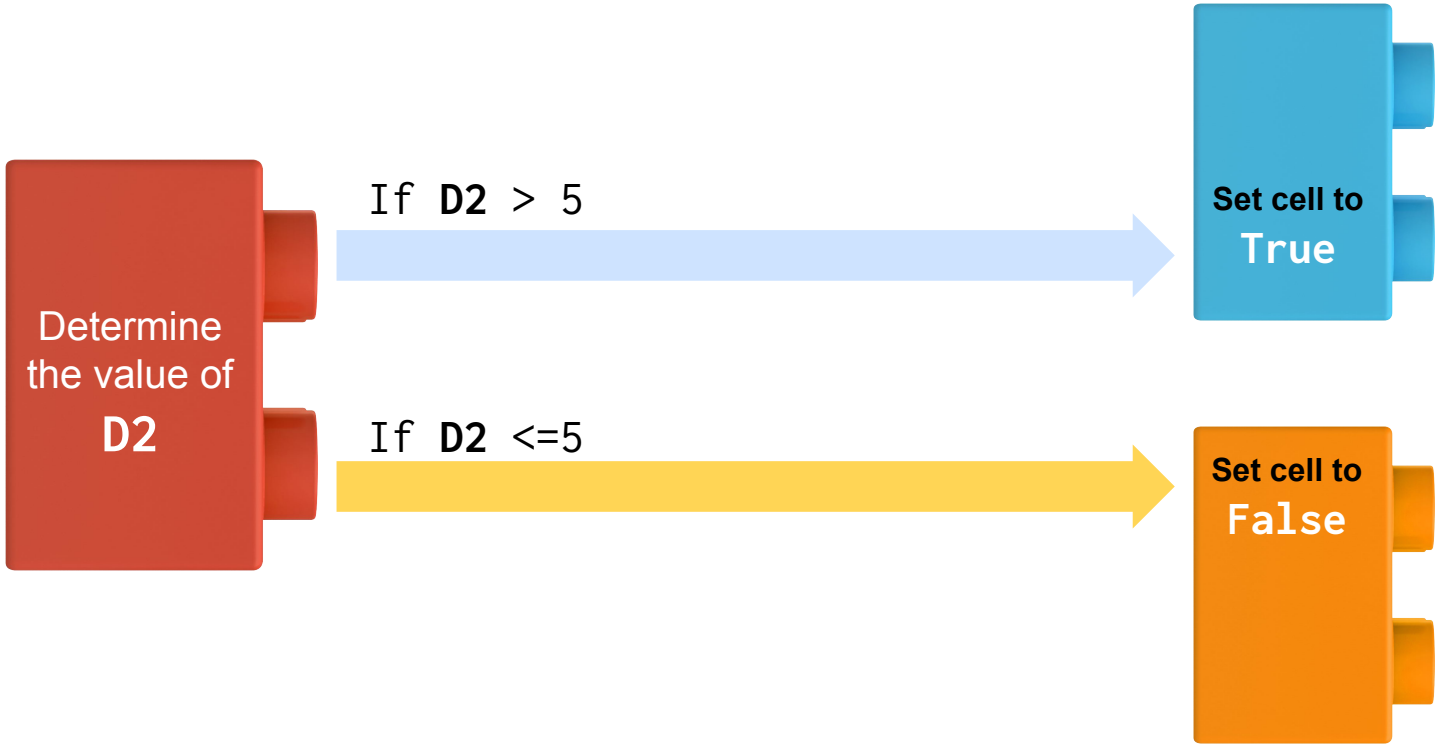
Conditionals: If This, Then That

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.



Conditionals: If This, Then That

```
=IF(D2>5,TRUE,FALSE)
```



What if we want to combine conditions?

AND, NOT, OR

Coding (Sort Of)

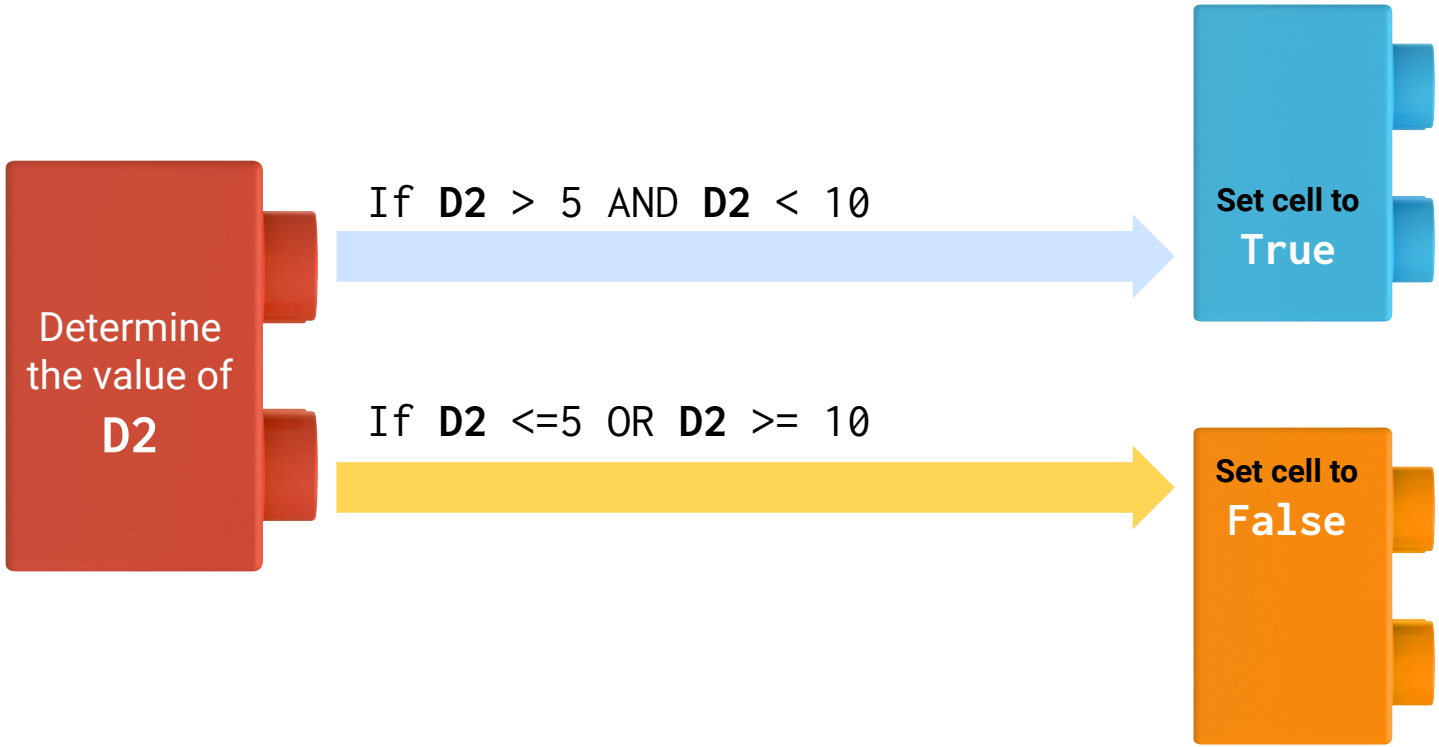
What if we want
to **combine** conditions?

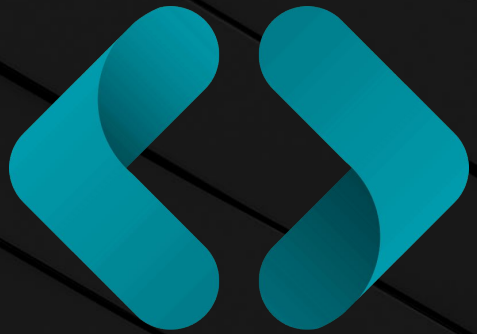
AND, NOT, OR

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

Conditionals: If This, Then That

Nesting conditionals is powerful but can quickly become convoluted.





Time to <code>

Activity: Grade Book

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

Suggested Time:

20 minutes

Activity: Grade Book

To do:	• Create a formula that calculates the final grade for a student based on their previous exams and papers.
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 “PASS” if a student’s final grade is greater than or equal to 60. If a student’s final grade is less than 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 • Less than 60 = F

Time’s Up! Let’s Review.

Instructor Demonstration

Measures of Central Tendency

What are measures of central tendency?

Values that describe the centre of a dataset.

Central Tendency

The three most common measures of central tendency:

Mean

The arithmetic average.

To calculate: Divide the sum of all the values by the number of values.

Median

The middle value of a dataset.

To calculate: Sort the dataset, and then find the centre.

Mode

The most frequent value of a dataset.

To calculate: Count the frequency of each value in a dataset, and then determine the most frequent value.

Instructor Demonstration

Measuring the Measures


Break

Instructor Demonstration

Formatting

Formatting in Excel Falls into Two Categories

Data Formatting

- Changes the way a value gets represented in a cell
- Helps with interpretation, or adding context to the range of values

Examples

- Date and time
- Currency
- Percentage
- Scientific notation

Style Formatting

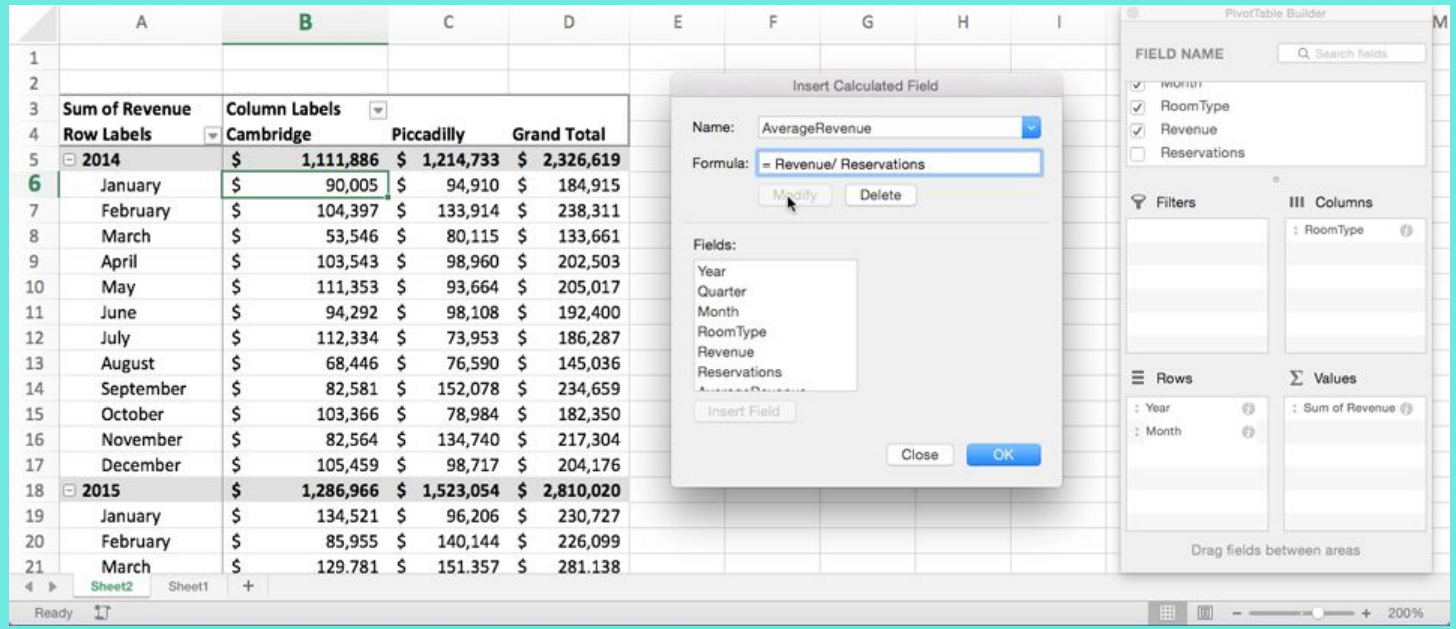
- Changes the way the cell and text get viewed
- Can include font colour, cell highlighting, borders, and more
- Can be performed manually or by using formulas and logic (conditional formatting)

Instructor Demonstration

Pivot Tables

Introduction to Pivot Tables

The pivot table is one of the most important data visualisation concepts to become proficient with in this class. (Don't worry. We'll explain them.)



Introduction to Pivot Tables

In essence, a pivot table is a **summative** analytic tool that we can use 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).

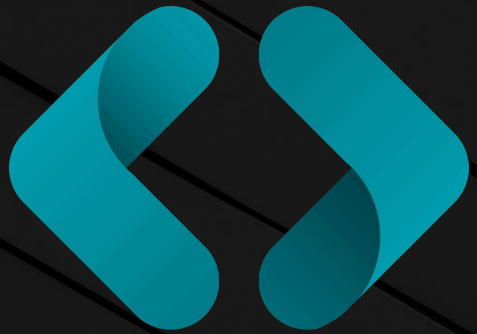
Seller	Qty Sold	Date
Joseph	$42.50	1/1/17
Jacob	$65.00	1/3/17
Jacob	$5.25	1/6/17
Joseph	$125.00	1/6/17
Jacob	$3.50	1/7/17
Matt	$32.00	1/9/17

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

Keep It Flat

Modern business intelligence (BI) tools, like those from Tableau, Sisense, and Salesforce, work best if data is stored in flat CSVs—meaning that column headers represent fields (vertically) on the spreadsheet. This is largely because all these technologies heavily use pivot tables as a tool for their visualisations. **Don't try to confuse this simplicity. "Spreadsheet magic" is a nightmare to analyse.**

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



Time to <code>

Activity: 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 that each song was released.

Suggested Time:

20 minutes

Top-Songs Pivot Table Instructions

- Select all of the data in your worksheet, and then create a new pivot table.
- Make a pivot table that can be filtered by year and contains two rows: 'artist' and 'name'.
- Note that all of an artist's songs should be listed underneath 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`.

Time's Up! Let's Review.

Instructor Demonstration

Lookups

Look It Up with Lookups

Assume that 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

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


```
=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)
```

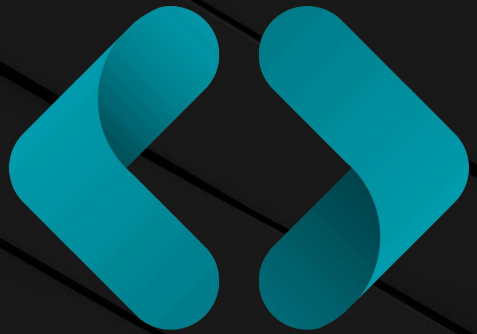
Planet	Population	Species
Zeelo	5020	Zoltans
Merinoa	380	Murphies
Cardboard Box	2	Hambones
…	…	
Asteroid 9	95	Asterisks

Look It Up with Lookups

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



Time to <code>

Partner Activity: Product Pivot

An independent artist who sells their product designs in an online store wants to visualise the cost of their recent orders. Using lookups, create a pivot table for the artist.

Suggested Time:

15 minutes

Partner 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 "Shipping Priority".

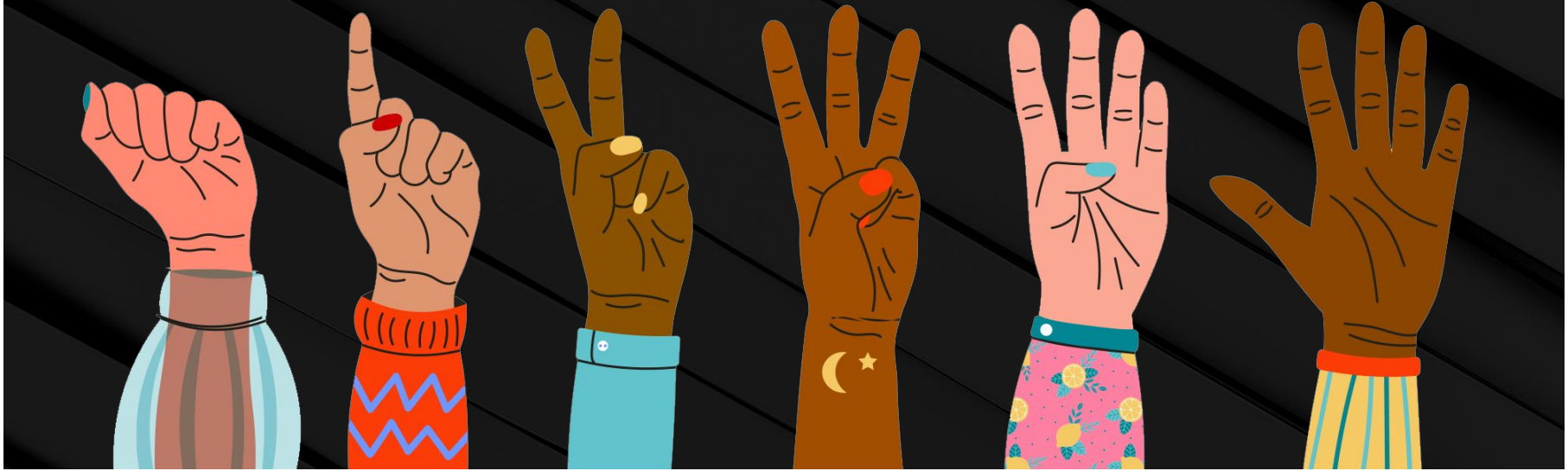
Select all of the data on the "Orders" sheet, and create a new pivot table that calculates the sum of both "Product Price" and "Shipping Price" for each "Order Number" and "Product ID".

Time’s Up! Let’s Review.

Questions?

FIST TO FIVE:

Who feels comfortable with pivot tables in Excel?



FIST TO FIVE:

Who feels comfortable with the measures of central tendency?