Data Analytics

Referencing and Lookups

GENERAL ASSEMBLY

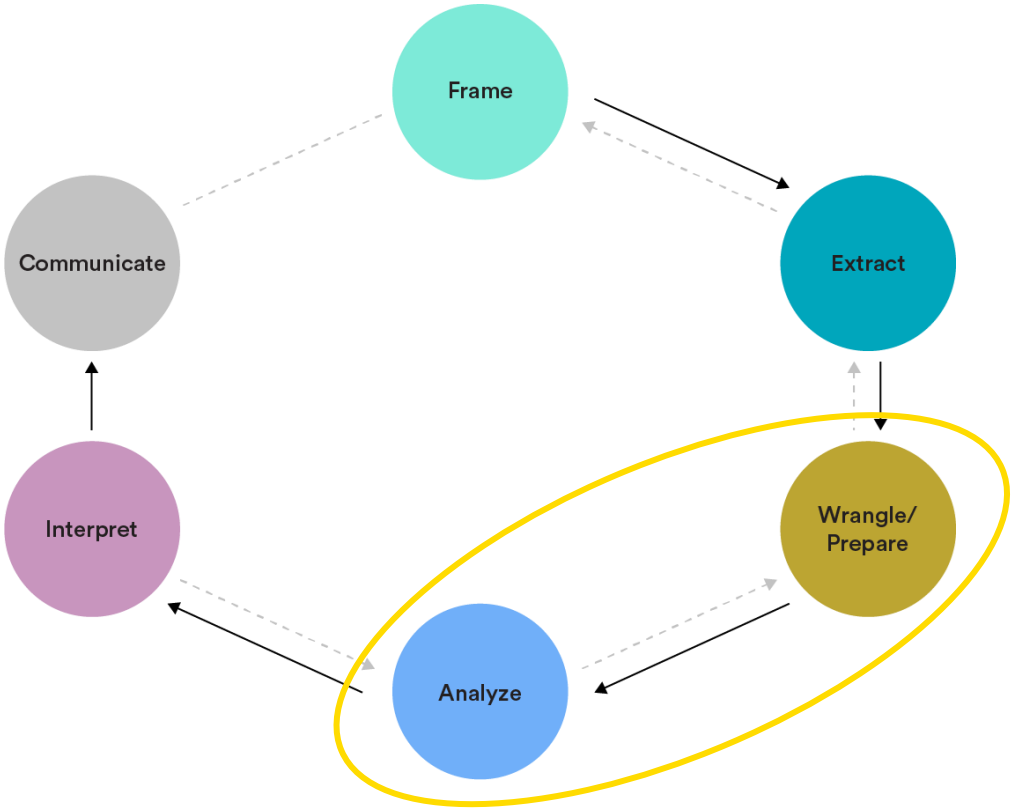
Our Learning Goals

- Build relationships between cells in Excel.
- Manipulate data sets using VLOOKUP.
- Look up values in other tables using INDEX and MATCH.

Where We Are in the DA Workflow

Wrangle/Prepare: Clean and prepare relevant data.

Analyze: Structure, comprehend, and visualize data.



Discussion:

Looking Up Information

Referencing is all about looking something up elsewhere and comparing it with what's in front of you.

While we're still investigating the reason behind the increase in returns, let's open up all of your Superstore files and look into **those attributes you think will have the biggest impact on returns.**

Discussion:

Referencing Information

Based on the attributes we identified, we now know that we need to:

Reference customers with frequent returns in a region and match them up with attributes of their sale.

What's an efficient way to go about this?

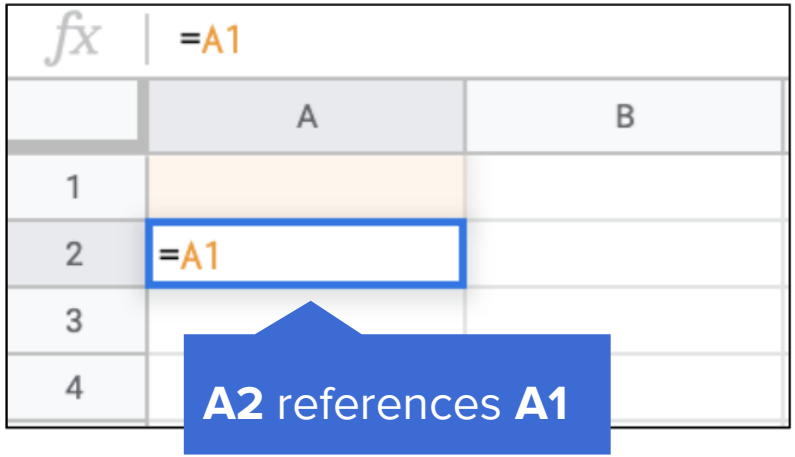
Referencing and Lookups

Getting Started With VLOOKUP

Referencing and VLOOKUP

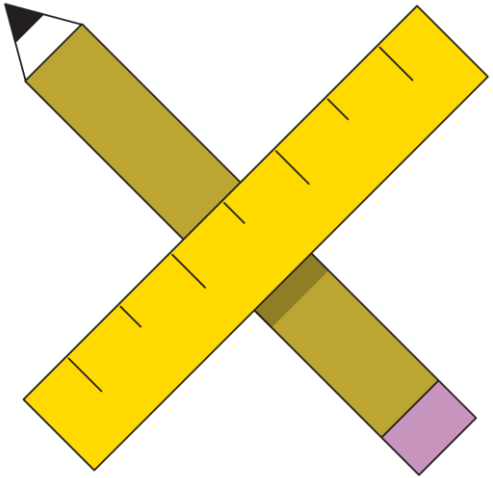
Referencing, in its basic form, means pulling the value of one cell into another cell.

With VLOOKUP, we will sometimes need to **reference across files** and **lock references** to make the formula function properly.



Referencing With Advanced Excel Tools

VLOOKUP, **HLOOKUP**, **XLOOKUP**, and **INDEX/MATCH** are often considered advanced tools that **increase efficiency** while **reducing data integrity issues**.



What Is VLOOKUP?

V stands for “vertical.”

VLOOKUP is:

- A function that works with **data formatted in columns**.
- A function that finds or “looks up” the value in one column of data and **returns the corresponding value** from another column (and usually another table).

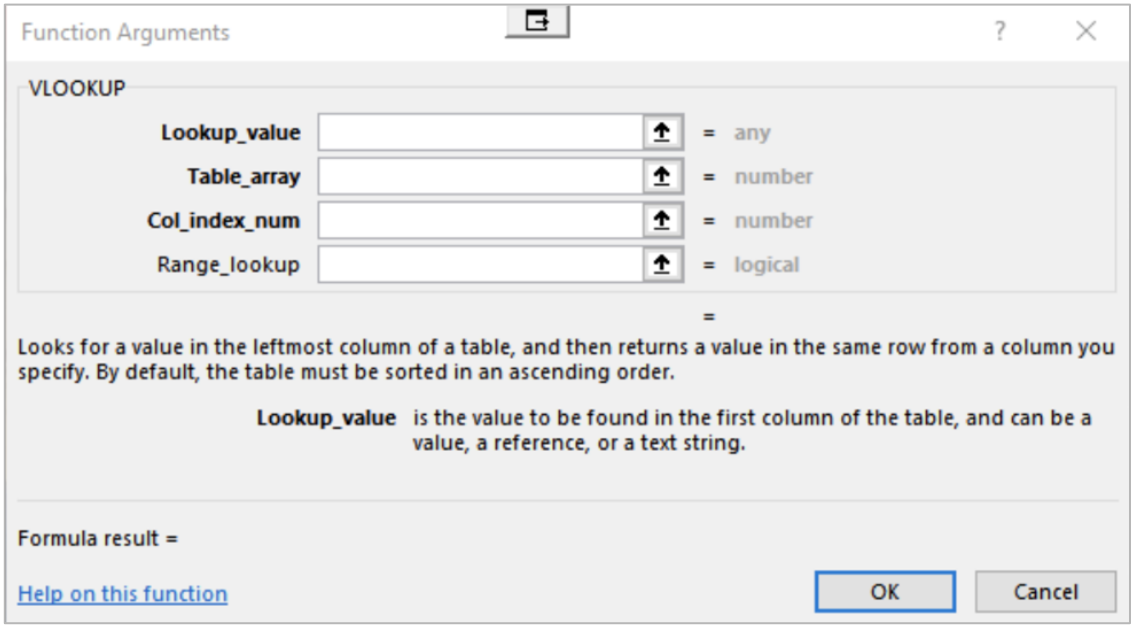
Building a VLOOKUP Statement

When building a VLOOKUP statement:

- In the cell, type “=VL.”
- Double click on the function name that’s presented for syntax.

Building a VLOOKUP Statement

Alternatively, click ***fx*** in the menu ribbon and select **Function** > VLOOKUP.



VLOOKUP Syntax

Lookup_value is the value that will be used to match data. It's usually an identifier and it must exist in both worksheets.

```
=VLOOKUP(lookup_value, table_array, col_index_num, [range_lookup])
```

Table_array is the table from which you want to retrieve data.

Col_index_num is the number of the column from the left side of the table_array from which you want to retrieve data.

VLOOKUP Syntax | Range_lookup

```
=VLOOKUP(lookup_value, table_array, col_index_num, [range_lookup])
```

range_lookup defines whether or not the lookup_value is an approximate match or an exact match of the value you are comparing it to in the left-most column of the table_array.

TRUE: Approximate match is needed.*

FALSE: An exact match is required.

In the work of a data analyst, it is rare to have all the data you need right in your data set.

Guided Walk-Through:

Pulling Data From Another Worksheet

Let’s combine data from another worksheet into the Superstore data set and explore the concept of a “lookup table” as well as how to use one to categorize our quantitative data.

1. Open your workbook: **Lesson 07_Superstore Workbook**.
2. In order to combine these two lists, we’ll create a third column in “**Returns**” with “**Customer Names”** — we can use **VLOOKUP** to “look up” one value in another table and return another column in that row.

Guided Walk-Through:

Pulling Data From Another Worksheet (Cont.)

Let's try simplifying what's required by naming a range to lookup.

1. Insert a column to the right of the "**order_id**" column in "**Returns**". The new column should be column B, with the header **"customer_name"**

2. In cell B2 of "**Returns**", enter: `=VLOOKUP(A2,orders!A:AA,27,FALSE)`.

   The customer name should populate cell **B2**.

3. Double click the bottom right corner of the cell — or click/drag — to replicate this function down the entire column.

Guided Walk-Through:
Using a VLOOKUP

Now, we'll examine if certain **states** have more returns than others.

1. In our "Returns" worksheet, add another column after Column **B**.
2. Name this "**state**" by typing this in Cell **C1**.
3. In Cell **C2**, enter:
   `=VLOOKUP(A2,orders!A:W,23,FALSE)`
4. Expand this formula to all rows by double clicking the bottom-right corner of the cell.

Now we have the states in our "Returns" data set!

Solo Exercise:

Moving Data With VLOOKUP

10 minutes

Now let’s say we want to bring in more information from the “Orders” data set. We don’t need everything, just a few columns.

On your own, use **VLOOKUP** to bring the “**category,**” “**sales,**” and “**profit**” columns from the “**Orders**” workbook to the “**Returns**” workbook.

Remember to start from the “Order_ID” column!

Solo Exercise:

Moving Data With VLOOKUP — Solutions

Category:

```
=VLOOKUP(A2,orders!A:J,10,FALSE)
```

Sales:

```
=VLOOKUP(A2,orders!A:N,14,FALSE)
```

Profit:

```
=VLOOKUP(A2,orders!A:O,15,FALSE)
```

Referencing and Lookups

Creating Categorical Variables

What Are Categorical Variables?

A categorical value, aka, a nominal variable, typically has **two or more non-ordinal categories** (values).

Hair color is a categorical variable that has a number of categories (e.g., blonde, brown, red...), but there is no inherent order to these categories.

Creating Categorical Values

Often, it's helpful to create categorical values from **numeric values**.

For example, a test that is scored 0–100 could be **classified** as A, B, C, D, or F, depending on the score.

Discussion:

Categorical Values in the Superstore Data Set

Knowing that it's helpful to create categorical values from **numeric values**, let's take a look at the "Orders" data.

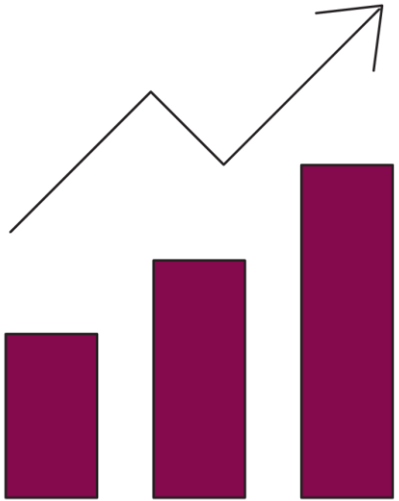
To help us solve the returns problem, which columns should we use to create categorical values that will help us analyze our data?

Computers Out:

Categorical Values in the Profit Margins

Let's practice assigning categorical values to the profit margins.

1. In the "Returns" workbook, in Cell **J1**, type "profit_margin."
2. In Cell **J2**, enter the formula "**=I2/H2**" and expand.
3. Classify our profit margins as either *low*, *medium*, or *high*.



Guided Walk-Through:
Margin Lookup

Refer to the steps below:

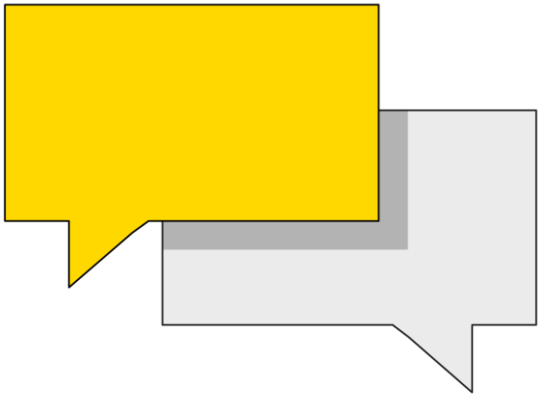
1. In your workbook, create a new worksheet called “**margin_lookup**”
2. In Cells **A1**, **A2**, and **A3**, enter values -5, 0, and 0.3.
3. In Cells **B1**, **B2**, and **B3**, enter values low, medium, and high.
4. On the original "Returns" worksheet, type “**Margin Category**” in Cell **K1**.
5. In **K2**, complete the lookup:
   `=VLOOKUP(J2, 'margin_lookup'!$A$1:$B$3, 2, TRUE)`.
6. Expand to all rows.

Discussion:

VLOOKUP and Absolute Cell References

Remember absolute cell references?

- What happens if we try that last formula without references?
- Can anyone explain what's happening to the formula?



Share your answers with the class!

Referencing and Lookups

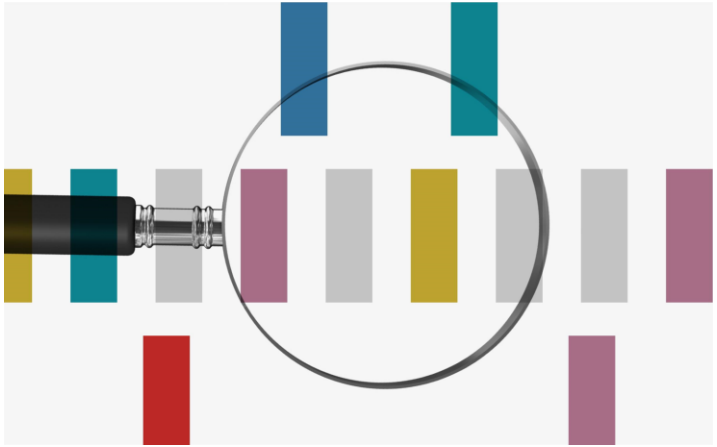
Other LOOKUPs

What Is HLOOKUP?

H stands for “horizontal.”

HLOOKUP is closely related to VLOOKUP, but instead of working with data sorted into columns, **HLOOKUP** work with **data sorted in rows**.

Because this is usually a difficult way to arrange data, it is *not* often used.



HLOOKUP Syntax

The **value** to look for in the first column of a table.

The **table** from which to retrieve a value.

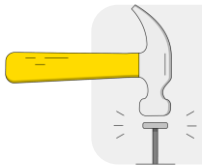
```
=HLOOKUP(lookup_value, table_array, row_index_num, [range_lookup])
```

The **row** in a table from which to retrieve a value.

Range_lookup: optional.
TRUE = Approximate match (default).
FALSE = Exact match.

Limitations of VLOOKUP/ HLOOKUP

- These LOOKUPs are **unidirectional** and must work with indicies fixed to the left side (VLOOKUP) or top (HLOOKUP) of the work area.
- Because the VLOOKUP references a col_index, it's **unable to dynamically update** whenever you insert a column or columns in the table_array.



Let's see what happens when we "break" the VLOOKUP by inserting a column.

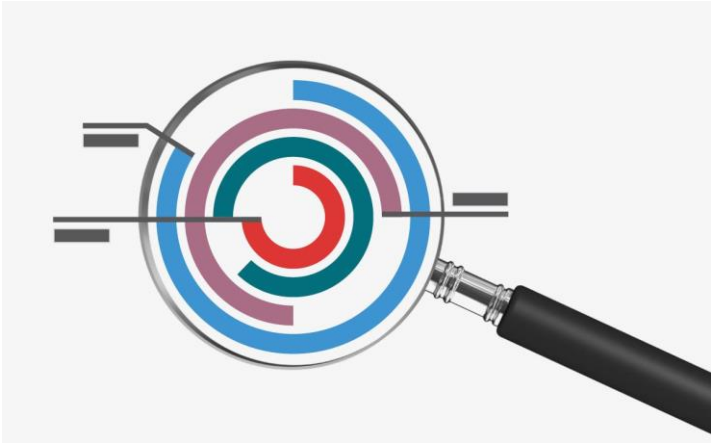
Referencing and Lookups

XLOOKUP

What Is XLOOKUP?

XLOOKUP is closely related to the other LOOKUPs, but it…

- Only requires the **lookup_value**, **lookup_array**, and **return_array** arguments.
- *Does not* require the return_array to be to the right of the lookup_array (it can be on either side).
- The match_mode argument is *optional* and defaults to exact.



XLOOKUP Syntax

The **value** to look for in the first column of a table.

Array that contains the answer you want to return.

=**XLOOKUP**(lookup_value, lookup_array, return_array)

Array in which the lookup_value can be found.

Guided Walk-Through:

Basic XLOOKUP

XLOOKUP was introduced to a select set of customers in August 2019 and officially launched to all Office 365* customers in January 2023.

A basic XLOOKUP accepts three arguments:

=XLOOKUP(lookup_value, lookup_array, return_array)

*You will only be able to use XLOOKUP with Office 365.

Guided Walk-Through:
“If Not Found” XLOOKUP

“If not found” XLOOKUP allows us to **set a return value** if a match isn’t found, negating the use of “IFERROR” or “IFNA.”

“If not found” XLOOKUP accepts four arguments:

=XLOOKUP(lookup_value, lookup_array, return_array, **value_if_not_found**) .

Guided Walk-Through:

XLOOKUP Search Mode

XLOOKUP Search Mode allows us to specify the direction of the lookup.

=**XLOOKUP**(lookup_value, lookup_array, return_array, value_if_not_found, **match_mode**, **search_mode**).

- **Match_mode** specifies exact or inexact match.
- **Search_mode** specifies the direction of the lookup: 1 is first item to last and -1 is last item to first.

Guided Walk-Through:
Nested XLOOKUP

A **nested XLOOKUP** can perform vertical AND horizontal lookups, just like an INDEX/MATCH.

=XLOOKUP(lookup_value, lookup_range, XLOOKUP(lookup_value, lookup_range, return_range**))**.

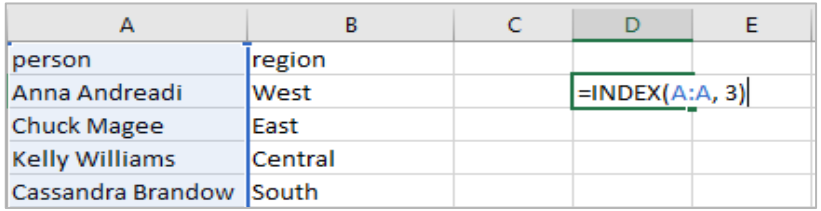
Don’t forget that a nested XLOOKUP will need **two** end parentheses!

Referencing and Lookups

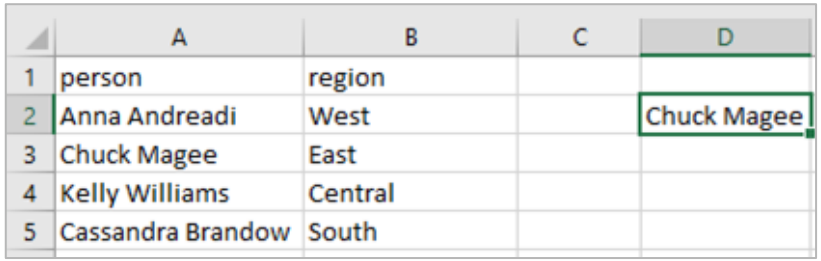
INDEX and MATCH

INDEX

INDEX is a function that returns **the value at the intersection of a row and column** in a given range.



	A	B	C	D	E
1	person	region			
2	Anna Andreadi	West		=INDEX(A:A, 3)	
3	Chuck Magee	East			
4	Kelly Williams	Central			
5	Cassandra Brandow	South			



	A	B	C	D
1	person	region		
2	Anna Andreadi	West		Chuck Magee
3	Chuck Magee	East		
4	Kelly Williams	Central		
5	Cassandra Brandow	South		

Syntax:

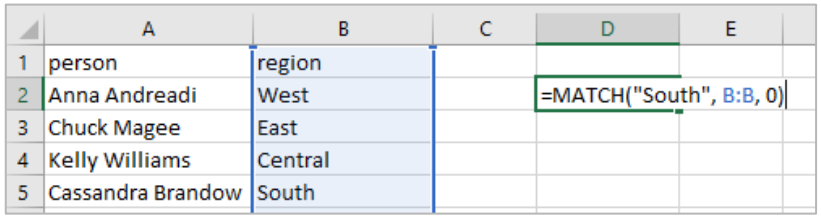
```
=INDEX(Array, Row_num, Column_num)
```

Example: I want to know the name of the person in the **third** row of the "**person**" column of my "Sales" table.

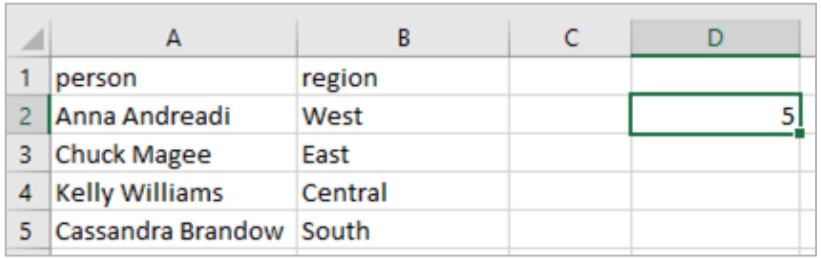
=INDEX(A:A, 3)

MATCH

Returns **the position of an item in an array** that matches a value.



	A	B	C	D	E
1	person	region			
2	Anna Andreadi	West		=MATCH("South", B:B, 0)	
3	Chuck Magee	East			
4	Kelly Williams	Central			
5	Cassandra Brandow	South			



	A	B	C	D
1	person	region		
2	Anna Andreadi	West		5
3	Chuck Magee	East		
4	Kelly Williams	Central		
5	Cassandra Brandow	South		

Syntax:

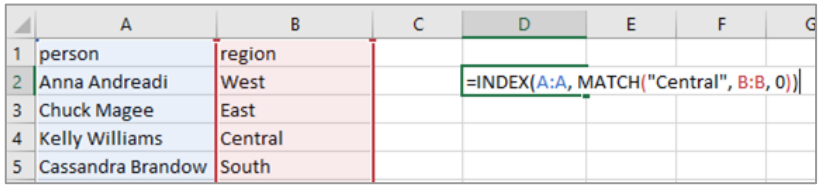
```
=MATCH(Lookup_value, Lookup_array, Match_type)
```

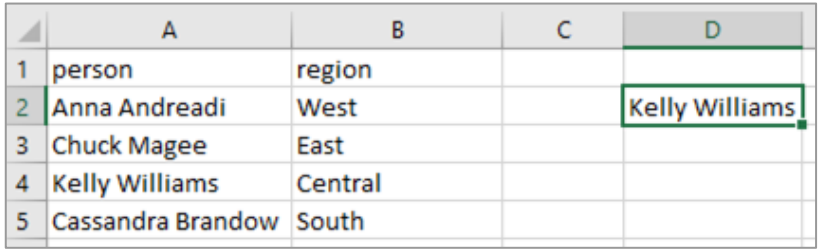
Example: I want to find the first occurrence of the words “**South**” in the **"regions"** column.

=MATCH("South", B:B, 0)

INDEX/MATCH LOOKUP

Specifies lookup column + **returns the value column independently**.





Syntax:

```
=INDEX(Return_value_range,
MATCH(Lookup_value,
Lookup_value_range, Match_type))
```

Example: I want to return the name of the person whose region is "Central."

=INDEX(A:A, MATCH("Central", B:B, 0))

Guided Walk-Through:
INDEX/MATCH Demo, Part 1

Open the “**index_demo”** sheet, and let’s prepare the worksheet with these steps.

1. In Cell **E2**, type: `=INDEX(A:A, 4)`
2. In Cell **E3**, type: `=MATCH(“Central”, B:B, 0)`

For now, we’ll always use a “0” as the third argument. The “0”parameter requires an **exact** match. In the next section, we’ll see how to use an **inexact** match.

Guided Walk-Through:
INDEX/MATCH Demo, Part 2

We will now combine **INDEX** and **MATCH** into a nested formula.

3. In Cell **I2**, type: `=INDEX(A:A,MATCH(H2,B:B,0))`.
4. Copy it down to **I3**.

The inner **MATCH** looks up the “region” of interest in the B:B column, returning the matching row number. The row number from the **MATCH** above is then used in the **INDEX** to look up and return a value in Column A:A.

Guided Walk-Through:

Revising VLOOKUPS With INDEX/MATCH

Let's recreate our VLOOKUPs from earlier using INDEX/MATCH. In a new column, use **INDEX**/**MATCH** to look up the customer names. Open up the "Returns" sheet and go to the "Customer Names" column.

1. In Row 2 of the "Customer Name" column, type:
   `=INDEX(orders!AA:AA,MATCH(A2,orders!A:A,0))`
2. Copy this down to all rows.

Partner Exercise:

VLOOKUP and INDEX/MATCH Practice

10 minutes

Complete the **class_exercises** tab.

1. Redo the VLOOKUPs we did earlier with INDEX/MATCH for the category, sales, and profit columns.
2. Redo the profit margin and margin category columns.

You may work with a partner, checking in with each other after answering each question.

Referencing and Lookups

Wrapping Up

Recap

Today, we...

- Built relationships between cells in Excel.
- Manipulated data sets using VLOOKUP.
- Looked up values in other tables using INDEX and MATCH.

Looking Ahead

Homework:

- VLOOKUP and INDEX/MATCH

Up Next: Aggregating Data With PivotTables

Additional Resources

- Excel Keyboard Shortcuts
- Relative, Absolute, and Mixed Cell References
- How to Use INDEX MATCH
- Using INDEX MATCH
- F4 No Longer Changes Absolute Cell References
- Explaining XLOOKUP
- VLOOKUP vs. INDEX MATCH


GA