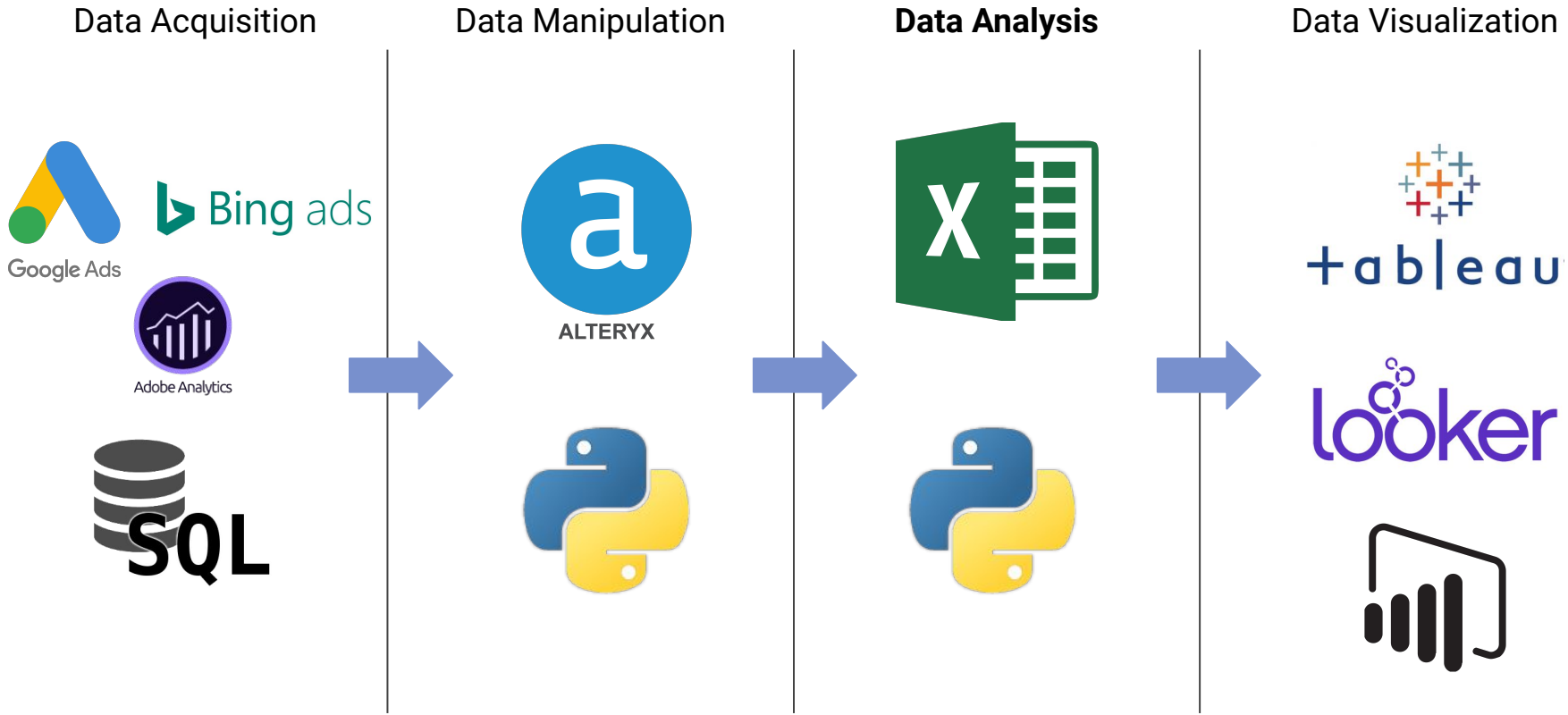
Excel Workshop for Beginners

How to 1-up your data analysis skills for industry

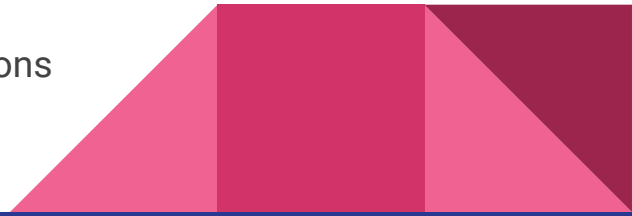
Tanner Arrizabalaga



The State of Business Analysis

What is Excel, and why Excel?

- Excel is a **spreadsheet management program**
- Organize and record data in an easy-to-navigate way
- Do basic and complex mathematical functions (so you don’t have to)
- Analyze data and make forecasting predictions
- Widely used in fields of all kinds (finance, education, science, and even graphic design!)
- **De facto standard in the business world**
- Tableau vs. Excel in terms of visualization
    - Tableau is great for interactive, long-lasting dashboards
    - Excel is great for quick visualizations to understand distributions



Levels of Excel

Beginner (necessary for entry-level hires)

SUMIF/SUMIFS

COUNTIF / COUNTIFS

Data Filters, Data Sorting

Pivot Tables

Cell Formatting

Data validation

Excel shortcut keys

PIVOT Tables & PIVOT Reporting

Conditional Formatting

Advanced

Advanced Charting

Functions

Advanced Formulae

VLOOKUP

INDEX + MATCH

VBA & Macros

Data Tables, Simulations & solver

What are we learning today?

- Data Cleaning/Organization
  - Data Filters, Data Sorting
  - Cell Formatting
  - Data validation
- Aggregation Functions - COUNT(IFS), SUM(IFS), AVERAGE(IFS)
- Pivot Tables

Data Cleaning, Organization, and Understanding

Why should we clean data?

Data scientists spend 60% of their time on cleaning and organizing data

Tables -- Lingo + Format

	A	B	C
1	**First Name**	**Last Name**	**Budget**
2	Michael	Smith	$134,000
3	Samuel	Taylor	$70,000
4	Michael	Scott	$500

row index → the number/location that is associated with an entire row

column index → the letter/location that is associated with an entire column

cell → a single location in the sheet that may or may not contain a value, a combination of a row and cell index (Cell A2 contains “Michael)

Tables -- Lingo + Format (cont.)

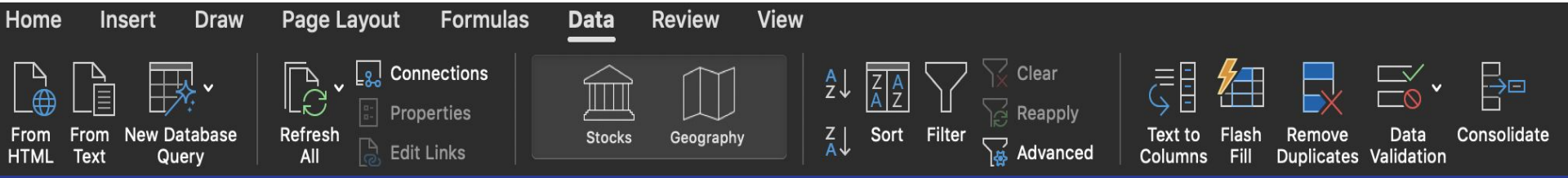
	A	B	C
1	**First Name**	**Last Name**	**Budget**
2	Michael	Smith	$134,000
3	Samuel	Taylor	$70,000
4	Michael	Scott	$500

column/feature → represents a particular variable **(First Name, Last Name, Budget)**

row/tuple → corresponds to a given record of the data set in question

How do we clean data? How do we organize data?

- Data tab in Excel is the place to go.
  - Sort
  - Filter
  - Text to columns
  - Flash Fill
  - Duplicate Handling



Demo

Functions

Functions -- Lingo + Format

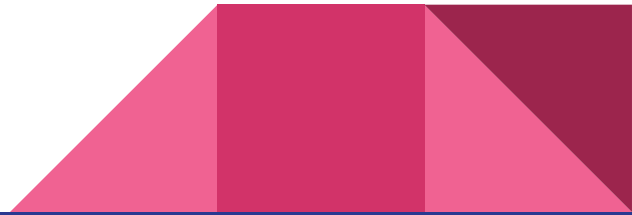
FUNCTION(value1, value2, range, ...)

(value1, value2, range, ...) → parameters we are passing in

FUNCTION → operation/operator we are applying

value1, value2, etc. → singular values in Excel, which are typically either a single value, or cell reference (E41, A5, etc.)

range → a list of values you want to apply a function to (A1:A5, or whole columns, like A:A)



COUNT - Syntax

COUNT() : Counts the number of cells in a range that contain numbers

The first item, cell reference, or range within which you want to count numbers.

=COUNT(value1, [value2], ...)

[argument] indicates **optional** argument

EX: =COUNT(C1,C2,C3) → 3

=COUNT(C1:C3) → 3

COUNTIFS - Syntax

COUNTIFS() : Counts the number of cells specified by a given set of conditions or criteria.

The range of cells you want evaluated for the **criteria**.

=COUNTIFS(criteria range, criteria, [criteria range, criteria…])

Condition or criteria that defines which cells will be used to count.

EX: =COUNTIFS(D2:D9, 6.65)

SUM - Syntax

SUM() : Adds all the numbers in a range of cells.

First number/range to sum

=SUM(number1, number2, [...])

Second number/range to sum

EX: =SUM(D1, D2:D9, 10)

EX: =SUM(15, 20, A:A)

SUMIFS - Syntax

SUMIFS() : Adds the cells specified by a given set of conditions or criteria.

Actual cells to sum

Condition or criteria that defines which cells will be used to find the sum.

=SUMIFS(sum_range, criteria_range, criteria, [...])

Range of cells for **criteria** to be applied to

EX: =SUMIFS(D2:D9, A2: A9, “>5.5”, B4, 10)

AVERAGE - Syntax

AVERAGE() : Returns the average (arithmetic mean) of its arguments

First number/range to use in average

=AVERAGE(number1, number2, [...])

Second number/range to use in average

EX: =AVERAGE(D1, D2:D9, 10)

AVERAGEIFS - Syntax

AVERAGEIFS() : Finds average(arithmetic mean) for the cells specified by a given set of conditions or criteria.

Actual cells to be used to find the average.

Range of cells for **criteria** to be applied to

=AVERAGEIFS(average_range, criteria_range, criteria, [...])

Condition or criteria that defines which cells will be used to find the average.

EX: =AVERAGEIFS(D2:D9, A2: A9, “>5.5”, B4, 10)

Functions to Know

Text Functions

“Tanner Arrizabalaga”

Function	Description
UPPER	Converts text to uppercase
LOWER	Converts text to lowercase
PROPER	Capitalizes the first letter in each word
TRIM	Removes extra spaces from text
LEN	Returns the length of text
FIND	Returns the location of text in a string -- 4
LEFT	Extracts text from the left of a string -- “Tanner”
RIGHT	Extracts text from the right of a string -- “Arrizabalaga”
MID	Extracts text from inside a string -- “ner Arr”
CONCATENATE	Joins text together

Date Functions

Function	Description
TODAY	Returns the current date
YEAR	Returns the year from a date
MONTH	Returns the month from a date
DAY	Returns the day as a number (1-31) from a date
NOW	Returns the current date and time

Information Functions

Function	CHECK IF
ISERROR	...has an error
ISNA	...returns #N/A
ISNUMBER	...is a number
ISEVEN	...is even
ISODD	...is odd

Math Functions

Function	Description
SUM	Adds numbers together
AVERAGE	Returns average of a group of numbers
MIN	Returns smallest value
MAX	Returns largest value
ROUND	Rounds a number to a given number of digits
SUMPRODUCT	Multiplies, then sums arrays

Logical Functions

Function	Description
IF	Tests for a specific condition
AND	Tests multiple conditions with AND
OR	Tests multiple conditions with OR
COUNT	Counts numbers
COUNTA	Counts numbers of non-blank cells
COUNTIF	Counts cells that match criteria
SUMIF	Sums numbers in a range that match criteria
AVERAGEIF	Returns the average of numbers that match criteria

Demo

Match & Index

MATCH - Syntax

MATCH(): Find the position of a value in a column/a row

The value that you want to match in the lookup_array

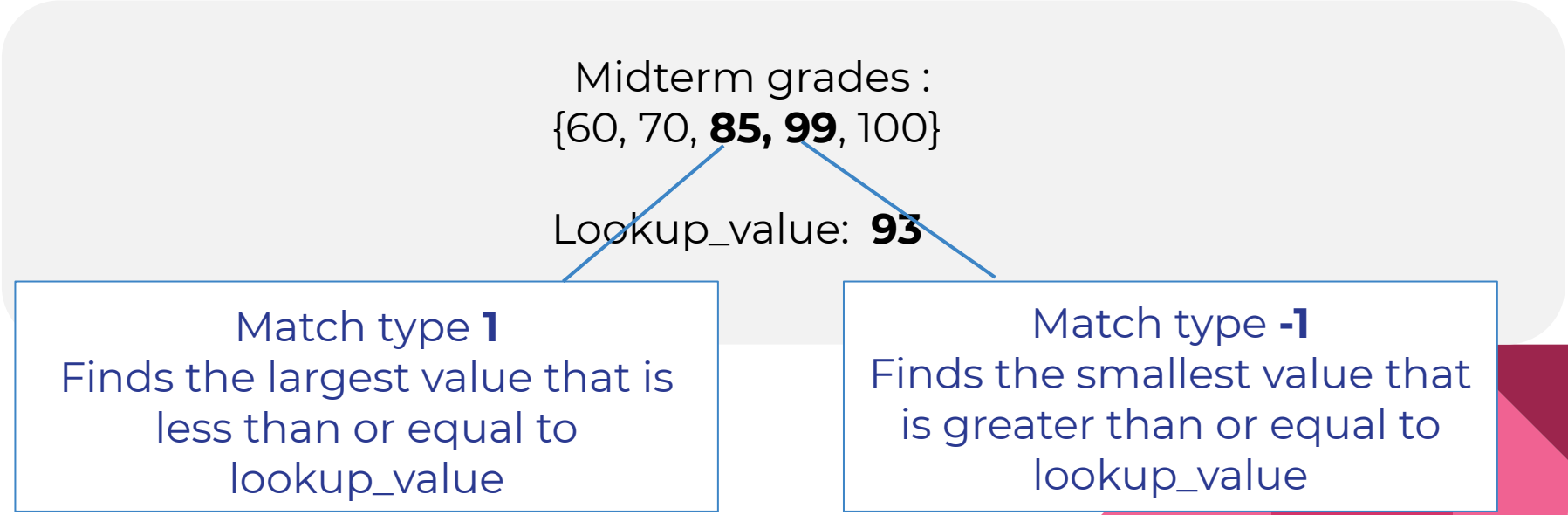
=MATCH(Lookup_value, Lookup_array, [match_type])

The range of cells being researched

Optional argument for specifying for the type of the match {-1,0,1} (for numbers only)

MATCH - Syntax

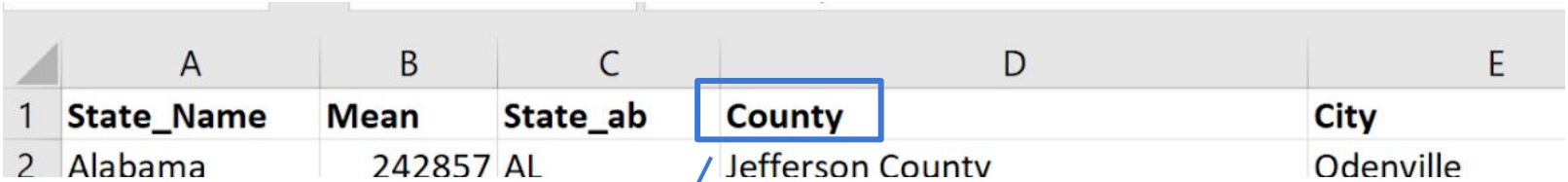
MATCH(): Find the position of a value in a column/a row



MATCH - Syntax

MATCH(): Find the position of a value in a column/a row

Q: What is the *column index* of County?



	A	B	C	D	E
1	**State_Name**	**Mean**	**State_ab**	**County**	**City**
2	Alabama	242857	AL	Jefferson County	Odenville

=MATCH("County", A1:E1, 0)

=4

MATCH - Syntax

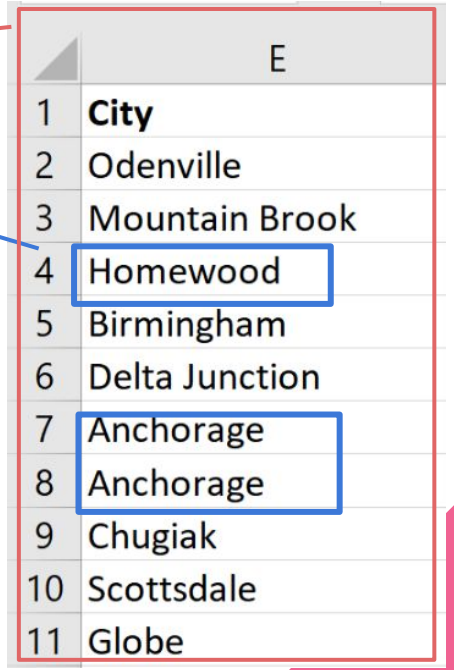
MATCH(): Find the position of a value in a column/a row

=MATCH("Homewood", E**1**:E11, 0)

=4

=MATCH("Anchorage", E1:E11, 0)

=7 (**first value** equal to look_up value)



	E
1	**City**
2	Odenville
3	Mountain Brook
4	Homewood
5	Birmingham
6	Delta Junction
7	Anchorage
8	Anchorage
9	Chugiak
10	Scottsdale
11	Globe

INDEX- Syntax

INDEX(): Use row and column indices to return the value from a reference

A range of cells or an array constant

=INDEX (array, row_num, column_num)

Selects the row in array from which to return a value

Selects the column in array from which to return a value

INDEX- Syntax

INDEX(): Use row and column indices to return the value from a reference

Q: what is the area of land in Delta Junction (city)?

	E	F	G	H
1	**City**	**Place**	**Type**	**ALand**
2	Odenville	Argo	Town	27893577
3	Mountain Brook	Autaugaville	Track	5012892
4	Homewood	Autaugaville	Track	2006630
5	Birmingham	Autaugaville	Track	12429198
6	Delta Junction	Whitestone	CDP	18298887
7	Anchorage	Akutan city	Track	85707800

=INDEX(E1:H7, MATCH("Delta Junction", E1:E7, 0), MATCH("ALand", E1:H1,0))

=INDEX(E1:H7, 6, 4) =18298887

Demo

VLookup

VLOOKUP - Syntax

VLOOKUP() : Looks for a value in the **leftmost column** of a table, and then returns a value **in the same row** from a column you specify.

Value to be found in the first column of the table

Column number in table_array from which the matching value should be returned

=VLOOKUP(lookup_value, table_array, col_index_num, range_lookup)

Table from which data is retrieved

Find closest match = TRUE; Find an exact match = FALSE.

VLOOKUP - Notes

- **VLOOKUP** retrieves data based on two columns:

  - One column contains the **look-up value**

  - Another column contains the **return value**

- The **table_array** starts with the column containing the **look-up value**

- **VLOOKUP** only looks to the right!

  - **Return value** can appear in any columns to the right

  - For values to the left: **MATCH & INDEX**

VLOOKUP - Example

Question: Find the city name of the place with id 502876.

	A	B	C	D	E
1	State_Name	id	State_ab	County	City
2	Alabama	101920	AL	Jefferson County	Odenville
3	Alabama	1022616	AL	Autauga County	Mountain Brook
4	Alabama	1022606	AL	Autauga County	Homewood
5	Alabama	102526	AL	Autauga County	Birmingham
6	Alaska	201700	AK	Southeast Fairbanks Census Area	Delta Junction
7	Alaska	20257	AK	Aleutians East Borough	Anchorage
8	Alaska	202377	AK	Aleutians East Borough	Anchorage
9	Alaska	202177	AK	Aleutians East Borough	Chugiak
10	Arizona	4023283	AZ	Apache County	Scottsdale
11	Arizona	401629	AZ	Gila County	Globe
12	Arizona	4023173	AZ	Apache County	Phoenix
13	Arizona	4022623	AZ	Apache County	Paradise Valley
14	Arkansas	502876	AR	Arkansas County	Rogers
15	Arkansas	5022286	AR	Arkansas County	Little Rock
16	Arkansas	5011037	AR	Fulton County	Brockwell
17	Arkansas	502526	AR	Arkansas County	Little Rock
18	California	60224419	CA	Alameda County	San Diego
19	California	60220539	CA	Alameda County	Huntington Beach
20	California	60213769	CA	Alameda County	Los Angeles
21	California	60227449	CA	Alameda County	Menlo Park
22	Colorado	8021630	CO	Adams County	Cherry Hills Village
23	Colorado	8022030	CO	Adams County	Boulder
24	Colorado	8022150	CO	Adams County	Superior
25	Colorado	802120	CO	Adams County	Highlands Ranch

=VLOOKUP(502876, B1:E25, 4, FALSE)

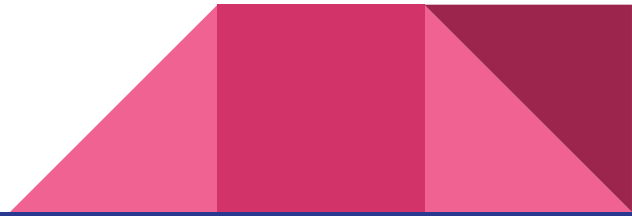
Alternate:

=VLOOKUP(502876, B1:E25, MATCH("City", B1: E1, 0), FALSE)

Pivot Tables

What are Pivot Tables?

- All of the formulas, combined in a nice representational format
- Instead of explaining, let’s get right into demonstration.



Thank you!

Questions?