

DATA AGGREGATION IN EXCEL

DATA AGGREGATION IN EXCEL

LEARNING OBJECTIVES

In today's lesson, we will:

1. Learn to apply Excel aggregation commands to our data, including:
 - a. **MIN, MAX, SUM, AVERAGE, COUNT**, and their conditional variants.
 - b. **COUNTIF, COUNTA, COUNTIFS**, and **COUNTBLANKS** for summarizing data.
2. Summarize data using **PivotTables** and manage common problems.
3. Explore data using **conditional formatting** for categorization and analysis.

OPENING

Aggregate functions summarize data in various ways. These functions are helpful on their own, but they also play an integral role in using **PivotTables**.

PivotTables are tremendously helpful when creating summaries of data!

DATA AGGREGATION IN EXCEL

GUIDED PRACTICE: HOW TO USE AGGREGATE FUNCTIONS

WHAT ARE AGGREGATE FUNCTIONS?

Aggregate functions allow you to ***summarize*** information using formulas in Excel.

Let's learn about Excel's aggregate functions and apply them to our scavenger hunt.

We'll be able to check our scavenger hunt answers with the new functions we're about to learn.

AGGREGATE FUNCTIONS: =MIN(...)

=MIN(. . .): Finds the **minimum value** of a range of numbers.

- **Syntax:** =MIN(number1, [number2], . . .)
- Solution to our scavenger hunt: =MIN(PART_3_Scav_hunt_DATA!B:B)

AGGREGATE FUNCTIONS: =MAX(...)

=MAX(. . .): Finds the **maximum value** of a range of numbers.

- **Syntax:** =MAX(number1, [number2], . . .)
- Solution to our scavenger hunt: =MAX(PART_3_Scav_hunt_DATA!0:0)

AGGREGATE FUNCTIONS: =SUM(...)

=SUM(. . .): Finds the **sum** of a range of numbers.

- **Syntax:** =SUM(number1, [number2], . . .)
- Solution to our scavenger hunt: =SUM(PART_3_Scav_hunt_DATA!E:E)

AGGREGATE FUNCTIONS: =AVERAGE(...)

=AVERAGE(. . .): Finds the **average** of a range of numbers.

- **Syntax:** =AVERAGE(number1, [number2], ...)
- Solution to our scavenger hunt: =AVERAGE(PART_3_Scav_hunt_DATA!AI:AI)

AGGREGATE FUNCTIONS: =COUNT(...)

=COUNT(. . .): Counts the **number of numeric values** in a range.

- **Syntax:** =COUNT(value1, [value2], ...)
- Solution to our scavenger hunt: =COUNT(PART_3_Scav_hunt_DATA!D:D)

AGGREGATE FUNCTIONS: =COUNTIF(...)

=COUNTIF(. . .): Counts the number of values in a range that meet the given criteria.

- Syntax: =COUNTIF(range, criteria)
- Solution to our scavenger hunt:
 - **Part A:** =COUNTIF(PART_3_Scav_hunt_DATA!B:B, "<25")
 - **Part B:** =COUNTIF(PART_3_Scav_hunt_DATA!AL:AL, "=High")

AGGREGATE FUNCTIONS: =COUNTIFS(...)

=COUNTIFS(. . .): Similar to **COUNTIF**, this counts the number of values in a range using specific criteria, except this method can take **many ranges with many criteria**.

- **Syntax:**

=COUNTIFS(criteria_range1, criteria1, [criteria_range2, criteria2]...)

- **Solution to our scavenger hunt:**

=COUNTIFS(PART_3_Scav_hunt_DATA!C:C, "<35", PART_3_Scav_hunt_DATA!D:D, ">=35")

AGGREGATE FUNCTIONS: =COUNTA(...)

=COUNTA(. . .): Counts the number of non-blank cells, not just the number of numeric cells.

- **Syntax:** =COUNTA(value1, [value2], ...)
- Solution to our scavenger hunt: =COUNTA(PART_3_Scav_hunt_DATA!A:A)-1

AGGREGATE FUNCTIONS: =COUNTBLANK (...)

=COUNTBLANK(. . .): Counts the number of blank cells in the range. This is the complement to COUNTA.

- **Syntax:** =COUNTBLANK(range)
- Solution to our scavenger hunt:
=COUNTBLANK(PART_3_Scav_hunt_DATA! A1 : AM1446)

DATA AGGREGATION IN EXCEL

INDEPENDENT PRACTICE: USING AGGREGATE FUNCTIONS

BREAK OUT ACTIVITY: USING AGGREGATE FUNCTIONS



EXERCISE

DIRECTIONS

1. Open
“**PART_3_Independent_Activity_DATA.**”
2. Based on your experience, choose to complete either the BASE or STRETCH tab (20 min).

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

DELIVERABLE

Complete BASE or STRETCH tab in “*Day 1 - Data Analytics with Excel*”

DATA AGGREGATION IN EXCEL

INTRODUCTION: PIVOT TABLES

INTRODUCTION TO PIVOTTABLES

PivotTables allow you to quickly create **dynamic** aggregations, slices, and filters.

Unlike static charts, this means a PivotTable allows you to use your data as an active source — without having to write your own formulas.

[PivotTables are an extensive topic](#) — entire books have been written about them — but in this lesson learn enough to use them efficiently.

DATA AGGREGATION IN EXCEL

DEMO: PIVOT TABLE COMPONENTS

DIAMONDS DATASET: DESCRIPTION

We'll demonstrate how to use PivotTables on the Diamonds dataset. This dataset has 10 variables:

1. **Price:** Price in U.S. dollars (\$326–\$18,823).
2. **Carat:** Weight of the diamond (0.2–5.01).
3. **Cut:** Quality of the cut (fair, good, very good, premium, ideal).
4. **Color:** Diamond color, from J (worst) to D (best).
5. **Clarity:** A measurement of how clear the diamond is (I1 (worst), SI1, SI2, VS1, VS2, VVS1, VVS2, IF (best)).
6. **X:** Length in mm (0–10.74).
7. **Y:** Width in mm (0–58.9).
8. **Z:** Depth in mm (0–31.8).
9. **Depth:** Total depth percentage = $z / \text{mean}(x, y) = 2 * z / (x + y)$ (43–79).
10. **Table:** Width of top of diamond relative to widest point (43–95).

CREATE A PIVOTTABLE

1. Open the spreadsheet labeled “**Diamond_Inventory.xlsx.**”
2. Create a PivotTable by clicking “**PivotTable**” on the Insert ribbon under the Data heading.
3. Make sure you have a cell selected in the diamond data to pre-populate the range.
4. Verify that the table/range auto-selected is the diamonds data.
 - a. If not, you can change it before moving on.
5. Decide where you want to put the PivotTable.
 - a. Let’s create a new worksheet for this example (which is the default option).
6. Click OK.

COMPONENTS OF A PIVOTTABLE

PivotTables have **four** primary components: **filters**, **rows**, **columns**, and **values**.

1. **Filters:** *Which data* should we include in our PivotTable?
 - a. We'll start by filtering out any data we don't want to include.
2. **Rows:** What *unique data values* do we want to have as rows in our table?
 - a. These values must exist in the data. If they don't, you'll need to create them first.

COMPONENTS OF A PIVOTTABLE

PivotTables have **four** primary components: **filters**, **rows**, **columns**, and **values**.

3. **Columns:** What *unique data values* do we want to have as columns in our table?
 - a. These values must already exist.
4. **Values:** These are the values that will be in the cells of our table.
 - b. To create them, we'll have to tell Excel how we want them aggregated using the aggregation functions we've already learned.

CREATING A PIVOTTABLE

Let's demonstrate by finding the *average price per color* in our Diamonds data:

1. Drag "color" to Rows.
2. Drag "price" to Values.
3. To fix the average price, click "i" next to "Sum of price" in Values.
4. Change "Summarize by" to "Average."
5. This will properly take the average, but let's make sure our formatting is correct.
 - a. These averages are dollar amounts, so go to "Format > Cells > Number" and select Currency.

CREATING A PIVOTTABLE

Next, let's find the *average price per color and cut*:

1. If we want to calculate the average price *per color and per cut*, drag “cut” into Rows.
2. For every color, we now see the price per each color and each cut.
3. What if we wanted cut and then color? Change the order and see the results.
4. It'd be much easier to parse and compare this data if the colors were down the rows and the cuts were across the columns.
5. Move “cut” to Columns.

CREATING A PIVOTTABLE

Next, let's find the *average price per color and cut for the top three best clarities*:

1. For this, the structure of our PivotTable is going to remain the same, but we need to *filter* some of the data out of the calculations.
2. Move “clarity” to Filter.
3. Notice that a dropdown for clarity has been added above the PivotTable. Use the dropdown to filter out only VVS1, VVS2, and IF.

CREATING A PIVOTTABLE

Last, let's identify the *number of diamonds by cut, color, and clarity*:

1. Move “cut,” “color,” and “clarity” to Rows.
2. Change “Average of price” to “Count of price.”
3. Reformat this into a number.
 - a. When doing a count, you can often use any of the variables for Values, as long as they have values of some kind.
 - b. Think of this as counting the number of values for a given column. Because the table is a rectangle, this also gives you the number of observations.

PIVOTTABLES SUMMARY

- PivotTables work well with this data set **because the data is not aggregated.**
- Each row defines only one observation (one diamond) and contains many **categorical variables.**
- Categorical variables are especially important for PivotTables because a PivotTable will always turn a variable into a categorical variable when it's placed as a Row or Column.

DATA AGGREGATION IN EXCEL

GUIDED PRACTICE: PIVOT TABLES WITH ACS DATA

PIVOTTABLES

Unlike the Diamonds dataset, our ACS data has been aggregated by census tract. Therefore, we'll need to keep the following consideration in mind to ensure our PivotTables are accurate and streamlined:

- Any values that we want in the rows or columns should be **discrete, categorical** variables.

In the case of our ACS data, this means *we'll need to create categorical data*, like we did when calculating density in the last lesson.

GUIDED PRACTICE: PIVOT TABLES



EXERCISE

DIRECTIONS

1. Open “**PART_4_acs_DATA_clean**”
2. Follow the three prompts.

DELIVERABLE

Answers to the prompts provided.

PIVOTTABLES WITH ACS DATA: BREAK OUT NO. 1

Question 1: Do rates of unemployment differ significantly by density?

Note: Remember, we created a “density_group” definition in our last lesson.

- Rows: “density_group”
- Values: unemployment rate

PIVOTTABLES WITH ACS DATA: BREAK OUT NO. 2

Question 2: What are the minimum, average, and maximum sizes (in sq. km) for census groups by density?

- Rows: “density_group”
- Values:
 - Min of “area_sqkm”
 - Average of “area_sqkm”
 - Max of “area_sqkm”

PIVOTTABLES WITH ACS DATA: BREAK OUT NO. 3

Question 3: What percentage of each county lives within each type of density group?

- Rows: “county”
- Columns: “density_group”
- Values: total population
 - Set “Show Data As” to “% of Row.”

DATA AGGREGATION IN EXCEL

COMMON PROBLEMS & SOLUTIONS FOR PIVOTTABLES

COMMON PIVOT TABLE SOLUTIONS



10 pivot table problems and easy fixes

DATA AGGREGATION IN EXCEL

INDEPENDENT PRACTICE: USING PIVOTTABLES

BREAK OUT ACTIVITY: USING AGGREGATE FUNCTIONS



EXERCISE

DIRECTIONS

1. Open “**PART_4_Independent_Activity_DAT**”
2. Based on your experience, choose to complete either the BASE or STRETCH tab (20 min).

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

DELIVERABLE

Complete BASE or STRETCH tab in
“L4_independent_activty_p2.xlsx”

DATA AGGREGATION IN EXCEL

CONDITIONAL FORMATTING ANALYSIS

CONDITIONAL FORMATTING ANALYSIS

Finally, we'll briefly cover some pointers for conditional formatting:

1. Use conditional formatting to draw attention to specific parts of the data *whenever they meet a predefined condition*.
2. Conditional formatting is a great way to *communicate a range* or spectrum, or to *visually demonstrate proportions within data*.
3. Conditional formatting can also assist in showing values that have fallen outside the expected range, helping to *identify outliers*.

CONDITIONAL FORMATTING ANALYSIS

1. Access controls for conditional formatting on the top menu.
2. Create highlighting “rules” to manage the desired criteria.

The screenshot displays the Microsoft Excel interface with the 'Conditional Formatting' menu open. The menu options include: Highlight Cells Rules, Top/Bottom Rules, Data Bars, Color Scales, Icon Sets, New Rule..., Clear Rules, and Manage Rules... The background shows a table titled '05 Diamond_Inventory - Excel' with columns A through H. The table contains 14 rows of diamond inventory data.

	A	B	C	D	E	F	G	H
4	0.23	Good	E	VS1	56.9	65	327	4.05
5	0.29	Premium	I	VS2	62.4	58	334	4.2
6	0.31	Good	J	SI2	63.3	58	335	4.34
7	0.24	Very Good	J	VVS2	62.8	57	336	3.94
8	0.24	Very Good	I	VVS1	62.3	57	336	3.95
9	0.26	Very Good	H	SI1	61.9	55	337	4.07
10	0.22	Fair	E	VS2	65.1	61	337	3.87
11	0.23	Very Good	H	VS1	59.4	61	338	4
12	0.3	Good	J	SI1	64	55	339	4.25
13	0.23	Ideal	J	VS1	62.8	56	340	3.93
14	0.22	Premium	F	SI1	60.4	61	342	3.88

BREAK OUT ACTIVITY: CONDITIONAL FORMATTING



EXERCISE

DIRECTIONS

1. Open the “Diamond Inventory” worksheet.
2. Add conditional formatting for columns indicating:
 - Table size, based on numerical rating using data bars.
 - Cut designation, by ranking the categories with individual colors.

You may work with a partner, checking in with each other after answering each question (20 min).

DELIVERABLE

The completed “Diamond Inventory” worksheet with conditional formatting implemented.

CONDITIONAL FORMATTING ANALYSIS: DIGGING DEEPER

How to build a search box
with conditional formatting

Search:

DATA AGGREGATION IN EXCEL

CONCLUSION

REVIEW: DATA AGGREGATION IN EXCEL

What did we accomplish in today's lesson?

1. We learned how to use and apply various **aggregate functions** in order to summarize our data, including: **MIN**, **MAX**, **SUM**, **AVERAGE**, **COUNT**, **COUNTIF**, **COUNTA**, **COUNTIFS**, and **COUNTBLANKS**.
2. We defined how to use **PivotTables** and explained some of their many benefits:
 - a. Saving time, creating aggregations, exploring slices of data
 - i. *Note: PivotTable logic is an important topic that comes up again in later units, so practice and review materials accordingly!*
3. We applied **conditional formatting** with color and bars to our data set to quickly highlight specific characteristics.

DATA AGGREGATION IN EXCEL

RESOURCES

- The Excel documentation on the Microsoft website is helpful when looking up the arguments and behavior of Excel functions: <https://support.office.com/en-us/excel>
- Common PivotTable issues and solutions: https://youtu.be/P_o_zuvGYyU
- ExcelJet on conditional formatting using a control window:
<https://youtu.be/aAOrhNHGQdg>

DATA AGGREGATION IN EXCEL

CITATIONS

- Diamonds data set from ggplot2's R package:
<http://docs.ggplot2.org/0.9.3.1/diamonds.html>