Data Analytics

# Aggregating Data With PivotTables

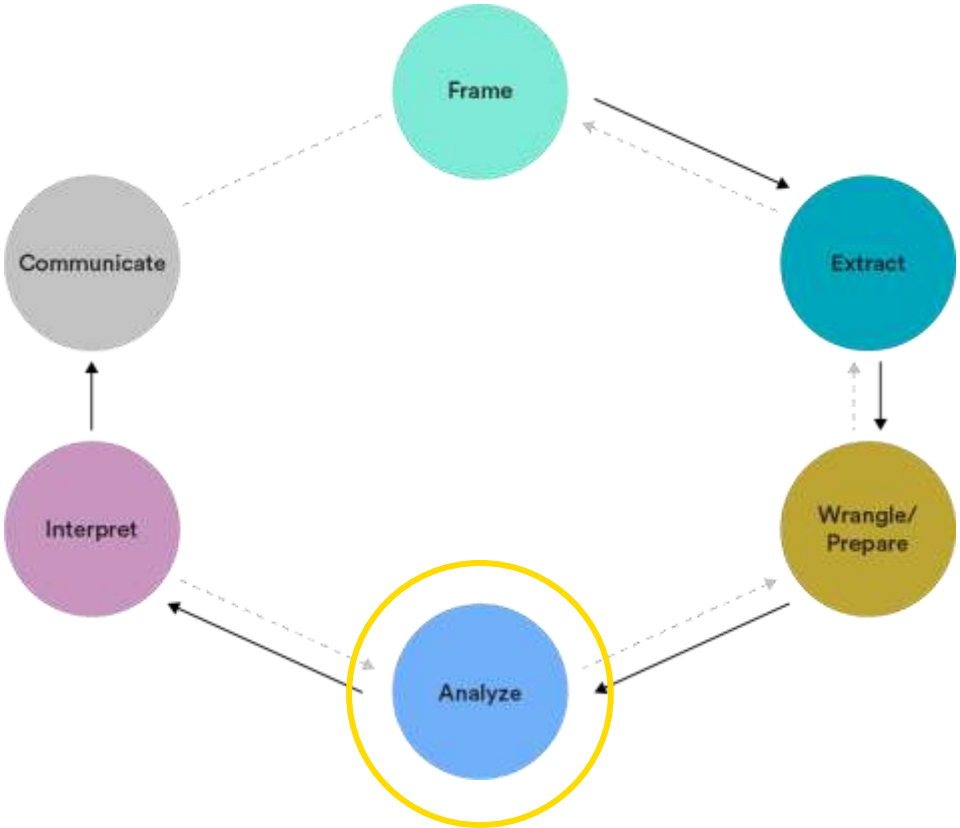
GENERAL ASSEMBLY

# Our Learning Goals

- Apply Excel aggregation functions to data sets.
- Use PivotTables to summarize data.
- Create modifiable data structures using regular tables to increase efficiency.
- Use named ranges to easily reference data subsets.

# Where We Are in the DA Workflow

**Analyze:** Structure, comprehend, and visualize data.



**Discussion:**

# Preparing Your Data for Analysis

As we've continued to investigate Superstore's increase in returns, we've looked into **referencing customers with frequent returns in a region and matching them up with the salespeople who process these returns.**

As a next step, it makes sense to look deeper into the average sales of certain products across regions and segments.

It's great to have so much data at hand, but what happens after cleaning? How do you prepare cleaned data for analysis?

Aggregating Data With PivotTables

# Aggregate Functions

**Discussion:**

# When Counting Only Gets You So Far

The regional sales director needs to know the number of sales that have been reviewed yet remain unapproved. Looking at the spreadsheet on the right, the ask seems simple enough — you can even count it by hand.

But what if there were over 800 rows of data? What could you do to speed up this process while ensuring your calculation is accurate?

| Reviewed by Sales | Reviewed by Management | Approved | Shipped |
|---|---|---|---|
| x | x | x | x |
| x | x | x | x |
| x | x | x | |
| x | x | x | x |
| x | x | x | x |
| x | x | x | x |
| x | x | x | x |
| x | x | | x |
| x | x | x | |
| x | x | x | x |
| x | | | |
| x | x | x | x |
| x | x | x | x |
| x | x | | x |

# Meet Aggregate Functions

- **Aggregate functions summarize data** using formulas in Excel.
- They play an integral role when using **PivotTables**.

| Commonly Used Aggregate Functions | |
|---|---|
| ● Minimum value<br>● Maximum value<br>● SUM<br>● SUMIF<br>● Average | ● COUNT<br>● COUNTIF<br>● COUNTIFS<br>● COUNTA<br>● COUNTBLANK |

# Aggregate Functions | Average and COUNT

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

Finds the **average** of a range of numbers.

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

Counts the **number of numeric values** in a range.

# Aggregate Functions | Minimum and Maximum Values

```
=MIN(number1, [number2], ...)
```

Finds the **minimum value** of a range of numbers.

```
=MAX(number1, [number2], ...)
```

Finds the **maximum value** of a range of numbers.

# Aggregate Functions | SUM and SUMIF

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

Finds the **sum** of a range of numbers.

```
=SUMIF(range, criteria, [sum_range])
```

Finds the **sum of values in a range** that meet the given criteria.

# Aggregate Functions | COUNTIF and COUNTIFS

```
=COUNTIF(range, criteria)
```

Counts the number of **values in a range** that meet the given criteria.

```
=COUNTIFS(criteria_range1, criteria1,
[criteria_range2, criteria2]…)
```

Counts the number of values in a range using specific criteria and can take in **many ranges with many criteria**.

# Aggregate Functions | COUNTA and COUNTBLANK

```
=COUNTA(value1,[value2], ...)
```

Counts the number of **non-blank cells**, not just the number of numeric cells.

```
=COUNTBLANK(range)
```

Counts the number of **blank cells** in the range.

**Group Exercise:**

# Using Aggregate Functions | Your Turn

**20 minutes**

In a small group, practice together the tasks below. You may use your existing analysis worksheet or create a new one. Ready, set, go!

- **MIN** to find the lowest sale.
- **MAX** to find the highest order quantity.
- **AVERAGE** to find the average profit margin.
- **SUM** to find the total sum of profits.
- **SUMIF** to find the total sales in the “Technology” category.
- **COUNT** to count the number of orders.
- **COUNTIF** to count the number of orders from the “Home Office” customer category.
- **COUNTIFS** to count the number of furniture orders in the “Consumer” customer category.
- **COUNTA** to count the number of returns (use the “Reason_returned” column).
- **COUNTBLANK** to count the number of orders that were NOT returned.

**Group Exercise:**

# Using Aggregate Functions | Solutions

| Formula | Value |
|---|---|
| =MIN(orders!O:O) | $0.44 |
| =MAX(orders!R:R) | 14 |
| =AVERAGE(orders!Q:Q) | 1% |
| =SUM(orders!P:P) | $ 23,248.58 |
| =SUMIF(orders!K:K,"Technology",orders!O:O) | $ 2,264,181.92 |
| =COUNT(orders!C:C) | 24059 |
| =COUNTIF(orders!AC:AC,"Home Office") | 4791 |
| =COUNTIFS(orders!AC:AC,"Consumer",orders!K:K,"Furniture") | 2327 |
| =COUNTA(orders!D:D)-1 | 1209 |
| =COUNTBLANK(orders!D2:D24060) | 22850 |

Aggregating Data With PivotTables

# PivotTables

**Discussion:**

# Connecting Hypotheses to Analysis

Let's take a few minutes to revisit our hypothesis-driven questions.

1. List out hypotheses:
    a. Do different types of customer segments make more returns?
    b. Do different categories get returned more often?
    c. Does shipping speed impact whether an item gets returned?

2. How would we answer these?

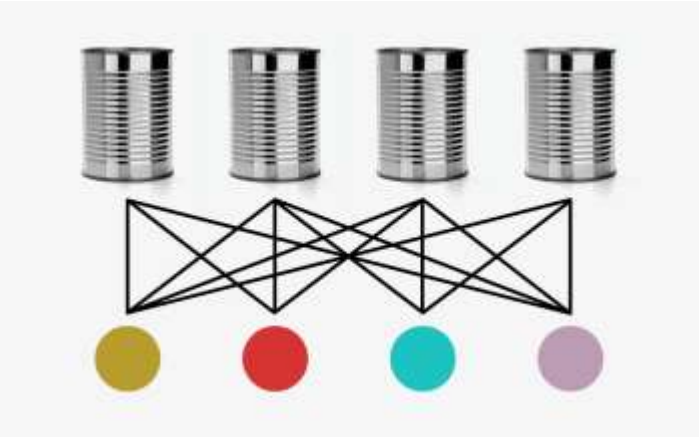
**Use "class_exercises" tab to keep track of the above!**

# Meet PivotTables

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

Unlike static charts, they enable you to use your data as an **active source** — without having to write your own formulas.



# Four Components of a PivotTable

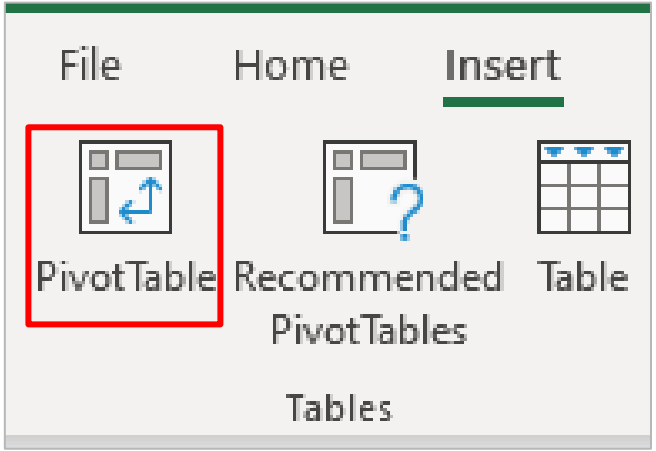
| Components | Questions to Ask of Data |
|---|---|
| **Filter** | Which data should we include in our PivotTable? |
| **Rows** | What unique data values do we want to have as rows in our table? |
| **Columns** | What unique data values do we want to have as columns in our table? |
| **Values** | What values will be in the cells of our table? |

**Computers Out:**

# Creating a PivotTable

In your Lesson 4 workbook, select any cell in the “Clean Data” tab. And then…

- Create a PivotTable by clicking “**PivotTable**” on the “Insert” ribbon.
- Verify that the table/range auto-selected is the “Orders” data. If not, you can change it before moving on.



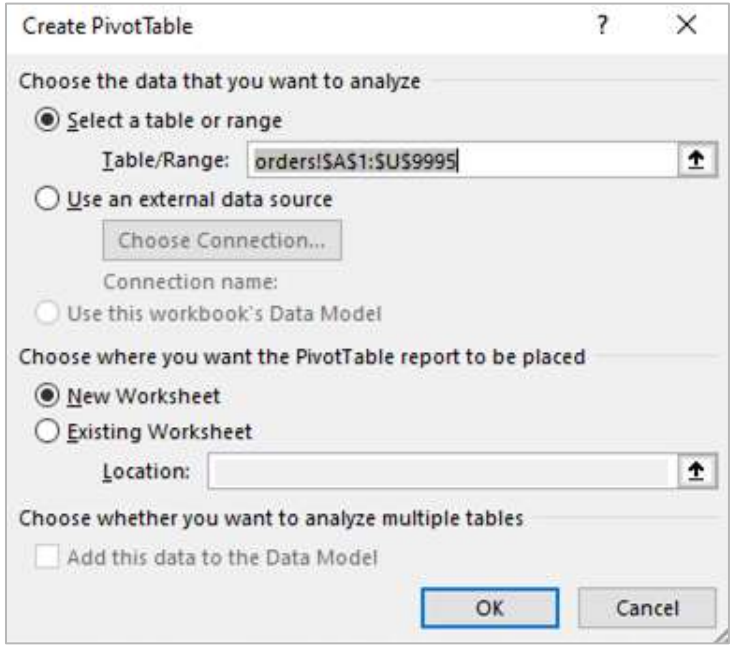

 

**Computers Out:**

# Creating a PivotTable | Setting It All Up

Decide where you want to put the PivotTable, and...

- Create a new worksheet for this example (this is the default option).
- Click “OK.”



**Guided Walk-Through:**

# Finding the Average Sales | PivotTable Field

Let's find the *average sale per customer segment* in our "Orders" data:

1. Drag *"segment"* to the **Rows** field.
2. Drag *"sales"* to **Values** field.

Your PivotTable Fields list should now look like this.

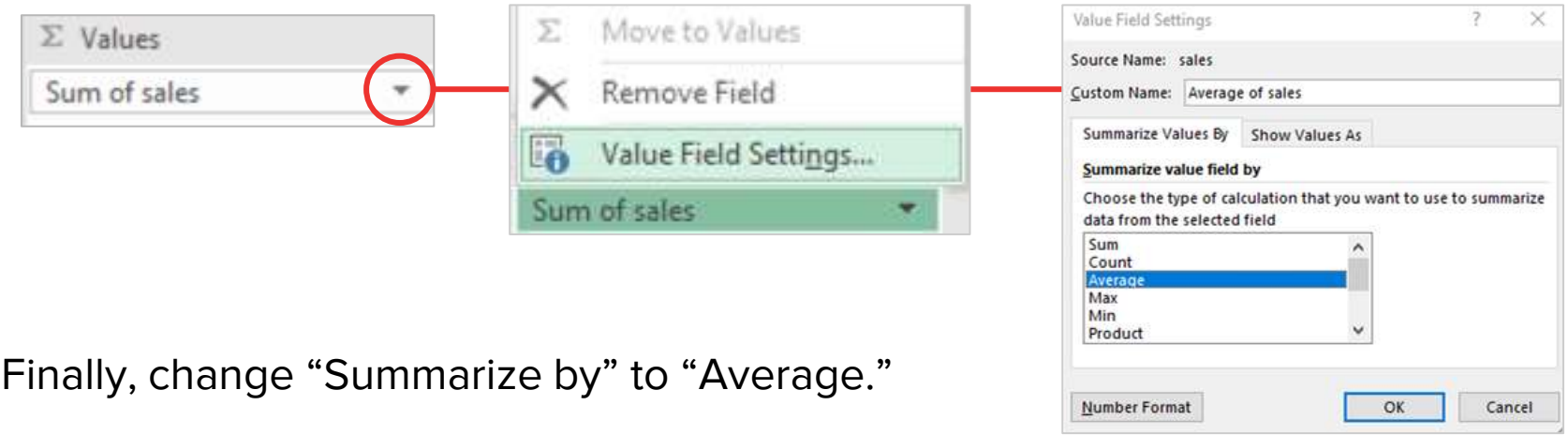

 

Guided Walk-Through:

# Finding the Average Sales | Value Field Settings

3. To change the “Sale” value from “sum” to “average,”click “i” next to “Sum of sales” in Values.

On a PC, click the small triangle, then click “Value Field Settings…”



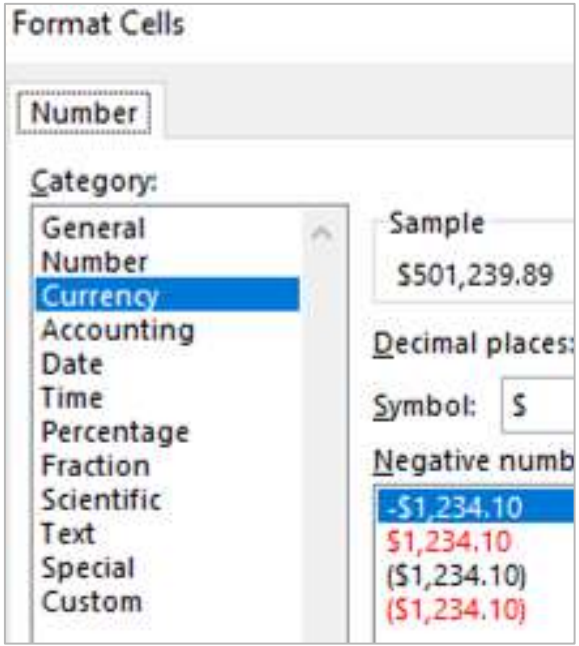

Finally, change “Summarize by” to “Average.”

## Guided Walk-Through:
# Finding the Average Sales | Formatting

This will properly take in the average sales, but first let's make sure our formatting is correct.

These averages are dollar amounts, so

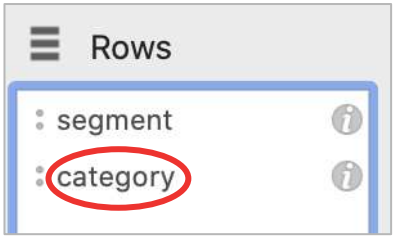
- Go to "Number Format > Number."
- Select "Currency" (or click the "$" button).



## Guided Walk-Through:
# Finding the Average Sales *and* Categories

Next, let's find the average sale per segment AND per category. Drag "category" into the Rows field below region.




| Row Labels | Average of sales |
|---|---|
| **Furniture** | **$405.10** |
| Consumer | $413.18 |
| Corporate | $393.34 |
| Home Office | $402.04 |
| **Office Supplies** | **$124.11** |
| Consumer | $120.02 |
| Corporate | $133.44 |
| Home Office | $120.69 |
| **Technology** | **$484.01** |
| Consumer | $458.78 |
| Corporate | $509.39 |
| Home Office | $512.51 |
| **Grand Total** | **$248.11** |

For every segment, we now see the **average for each segment and each category**. But, what if we want category and *then* segment?

Let's change the order and look at the results.

Guided Walk-Through:

# Finding the Average Sales *and* Categories

It'd be much easier to parse and compare this data if the segments were presented down the rows and the categories were presented across the columns.

Let's move "category" to Columns.

| Segment | Category | | | |
|---|---|---|---|---|
| Average of sales | Column Labels | | | |
| Row Labels | Furniture | Office Supplies | Technology | Grand Total |
| Consumer | $413.18 | $120.02 | $458.78 | $242.59 |
| Corporate | $393.34 | $133.44 | $509.39 | $255.23 |
| Home Office | $402.04 | $120.69 | $512.51 | $251.67 |
| **Grand Total** | **$405.10** | **$124.11** | **$484.01** | **$248.11** |

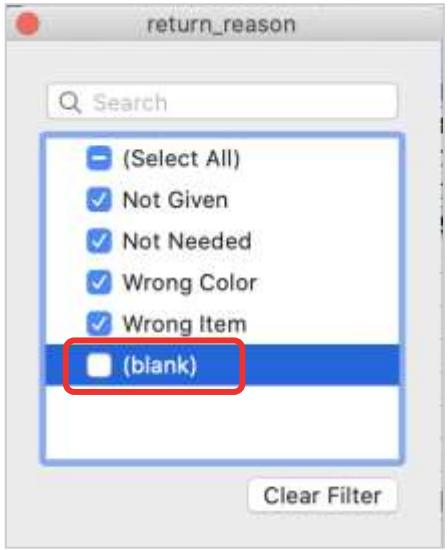
 

## Guided Walk-Through:
# **Average Sales** by Segment, Category, *and* Returned

Ready to get even more info from our data? Let's find the *average sales per segment and category for only returns.*

The structure of our PivotTable is going to remain the same, but we need to filter some of the data out of the calculations.

- Let's move "return_reason" to the Filter field.




Notice that a dropdown for "return_reason" has been added above the PivotTable.

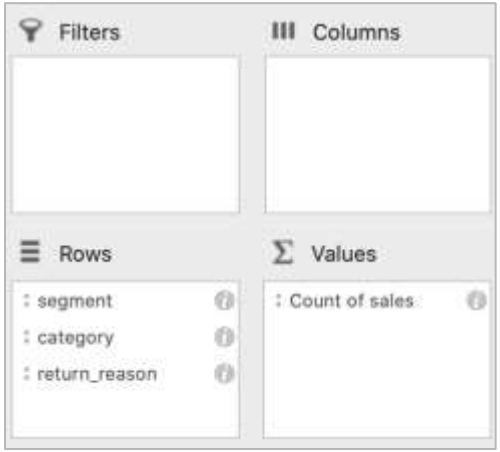
- De-select "Blank" and notice that the numbers now change to only display the orders that have been returned.

Guided Walk-Through:

# **Number of Records** by Segment, Category, *and* Returned

Let's identify the *number of orders by segment, category, and returned:*

1. Move "segment," "category," and "return_reason" to Rows.
2. Change "Average of sales" to "Count of sales."
3. Reformat this into a number.




**Note:** When doing a count, you can use any of the variables for Values, as long as every member of that column has values of some kind.

Aggregating Data With PivotTables

# Counting in PivotTables

## Discussion:
# True or False?

Take a look at this PivotTable from our “Orders” data set.

True or false: **The consumer segment of the business sold 12,207 dollars worth of products.**

Be ready to defend your answer.

| Row Labels | Count of sales |
|---|---|
| Consumer | 12207.0 |
| Corporate | 7061.0 |
| Home Office | 4791.0 |
| **Grand Total** | **24059.0** |

**Discussion:**

# The Answer Is...

The consumer segment of the business sold 12,207 dollars worth of products — **false**.

This is a tricky concept in Excel:

- **12,207** is the *number of orders* (or sales) from the consumer segment.
- You can think of it as the number of times customers went to the cash register — this number has nothing to do with the dollar amount at the register or the number of products brought to the register.

You can also interpret it as: **There were 12,207 orders placed by consumers.**

## Discussion:
# What Are We Counting?

This PivotTable is counting the **number** of **cells** in the sales column that contain a value.

Now that we know this, determine whether the following statements are true or false.

| Row Labels | Count of sales |
|---|---|
| Consumer | 12207.0 |
| Corporate | 7061.0 |
| Home Office | 4791.0 |
| **Grand Total** | **24059.0** |

1. **There is a total of 24,059 sales across segments**.
2. **“Home Office” generated the lowest revenue among the three segments.**

**Discussion:**

# The Answers Are…

1. There is a total of 24,059 sales across segments — **true!**
    a. This can also be interpreted as: **There is a total of 24,059 rows that have value in the "Sales" column for all there segments.**
2. "Home Office" generated the lowest revenue among the three segments — **false!**
    a. Because the number of rows does not represent dollars, the count of sales for "Home Office" does not reflect the amount of revenue (dollars) generated.

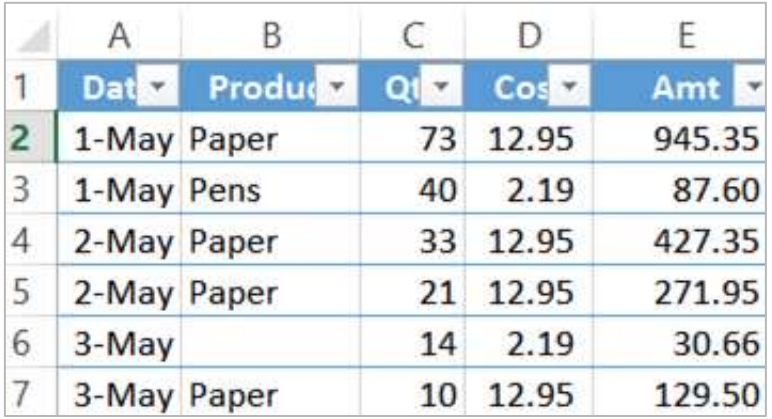
Aggregating Data With PivotTables

# Regular Tables

# Why Regular Tables?

Regular tables can be used with PivotTables to create more efficient workbooks. Here is why:

- Regular tables tell Excel the location of your data set.
- Regular tables are **modifiable**, meaning you can add/remove rows and columns.
- **Formulas are easier to read** in regular tables and **stay consistent** across all your data.
- Visualizations will **automatically update**.



| | A | B | C | D | E |
|---|---|---|---|---|---|
| 1 | Dat | Produ | Qt | Cos | Amt |
| 2 | 1-May | Paper | 73 | 12.95 | 945.35 |
| 3 | 1-May | Pens | 40 | 2.19 | 87.60 |
| 4 | 2-May | Paper | 33 | 12.95 | 427.35 |
| 5 | 2-May | Paper | 21 | 12.95 | 271.95 |
| 6 | 3-May | | 14 | 2.19 | 30.66 |
| 7 | 3-May | Paper | 10 | 12.95 | 129.50 |

**Image source**: Contextures.com

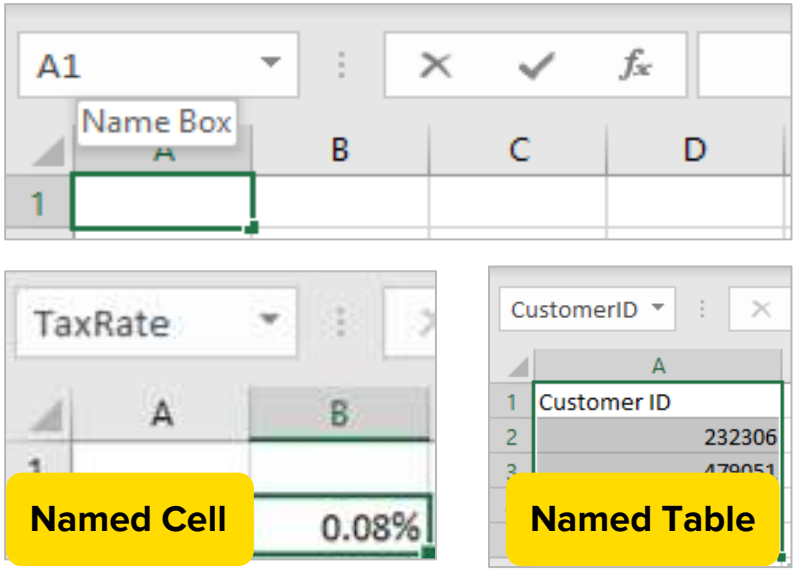
# Using Named Ranges With Tables

**Naming ranges** is a simple way to identify, define, or refer to a single cell or group of cells, to be identified in a formula as **a table array**.

**Step 1:** Select the cell(s) you want to name.

**Step 2:** Go to Name Box and type in a name, then press "enter."

**Note** that the Name Box does not allow spaces.

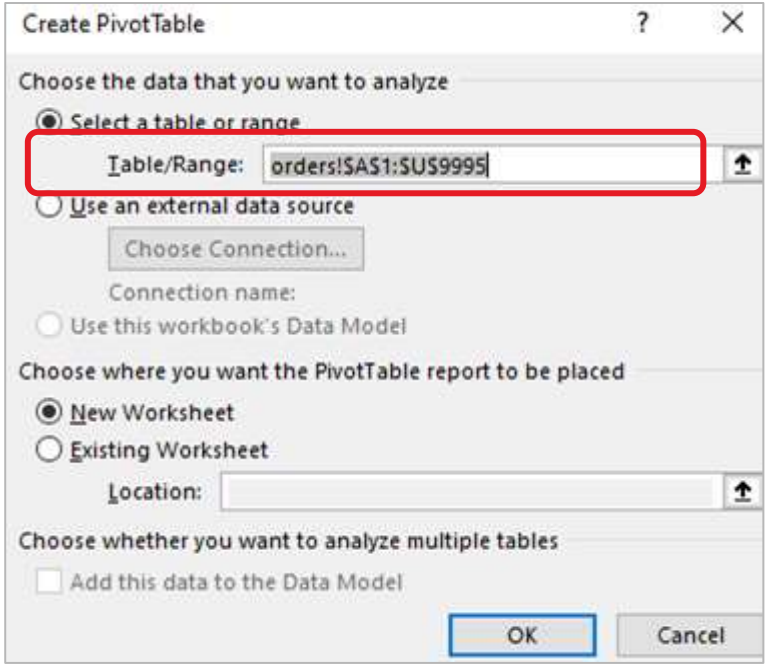


# Creating a Dynamic Named Range

Remember this from when we were first setting up a PivotTable?

This is what the PivotTable dialog box looks like without a table.



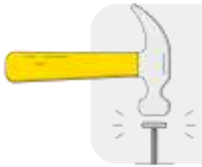
**Discussion:**

# Regular Tables | Let's Just Say...

Let's say we added 10 rows of data, all in the "Consumer" customer segment. To do this, you'd paste these rows below the data set, and they'd turn blue to signify they are now part of the table.

Now, what happens when you click "Refresh"?

**Hint:** As long as everything is connected, it will update with the "Refresh" button.



Copying and pasting values breaks connections. Use **cell references** instead if possible.

Aggregating Data With PivotTables

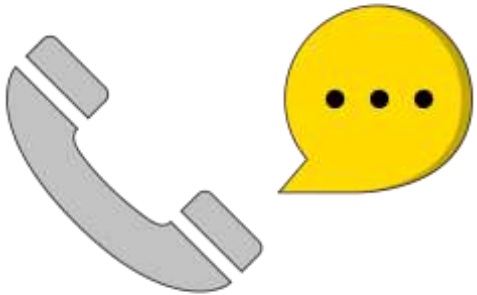
# PivotTable Slicers

# A Special Request

The head of sales is on the phone and wants to know *the sum of sales by segment*. As you create a PivotTable to answer her question, she asks you another one:

**Can you also break down the *sum of sales by product category*? I need to know how each category is doing by consumer segment.**

What can we build to enable her to find the answers she needs?



# Meet PivotTable Slicers

Slicers provide buttons that can be clicked to **filter tables**.

In addition to quick filtering, slicers also **indicate the current filtering state**, making it easy to understand what exactly is being displayed.

## Guided Walk-Through:
# Let's Try it!

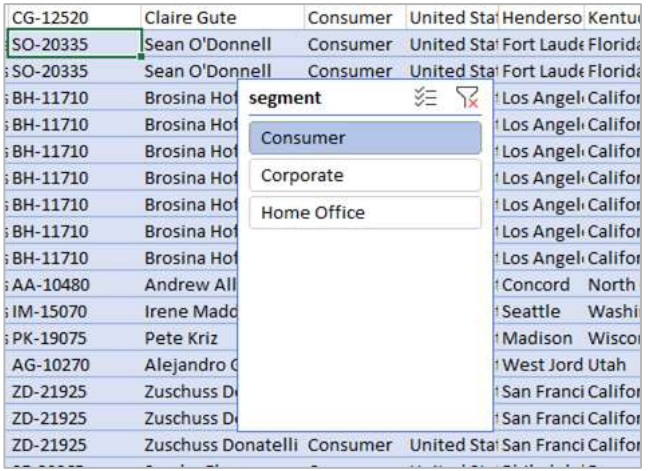
Go to the "Orders" table and select any cell. Then…

1. Insert a slicer by clicking the "Slicer" icon in the "Insert" tab of the ribbon.
2. Select "Segment."

   A box should appear over the table with all of the subsets of the "Segment" column.

1. Select "Consumer."

   The table is now only displaying data from the "Consumer" segment.



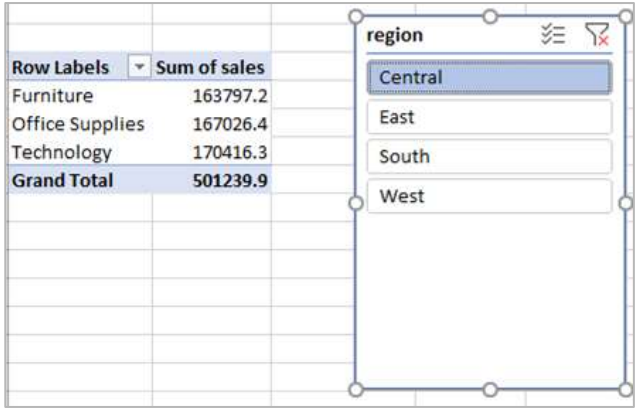
## Computers Out:
# How Can We Do This Quickly and Correctly?

The head of sales is on the phone and wants to know *the sum of sales by customer segment*. As you create a PivotTable to answer her question, she throws in another one:

**Can you also break down the *sum of sales by product category*? I need to know how each category is doing by region.**

You'll need your PivotTable to do this.

Let's find out how!



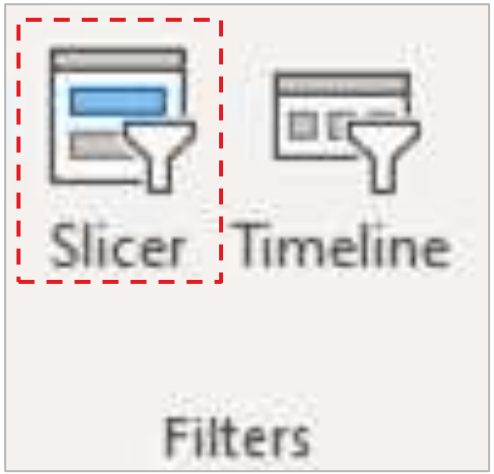

| Row Labels | Sum of sales |
|---|---|
| Furniture | 163797.2 |
| Office Supplies | 167026.4 |
| Technology | 170416.3 |
| **Grand Total** | **501239.9** |

Guided Walk-Through:

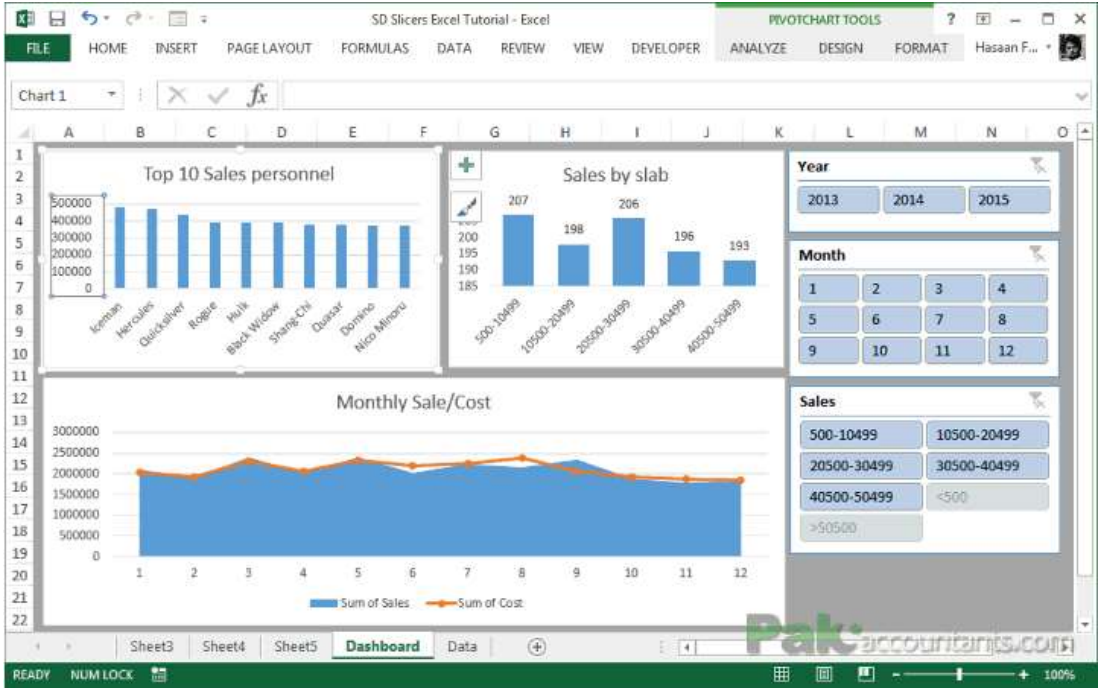
# Using PivotTable Slicers

To filter your PivotTable by *region*, insert a PivotTable slicer by clicking on the slicer icon.



## Real Cases:
# Slicers on Display




Take a look at this dashboard.

What benefit(s) of using the slicer can you identify here?

**Image source**: Pak Accountants Slicer How-To

# What Can Slicers Do

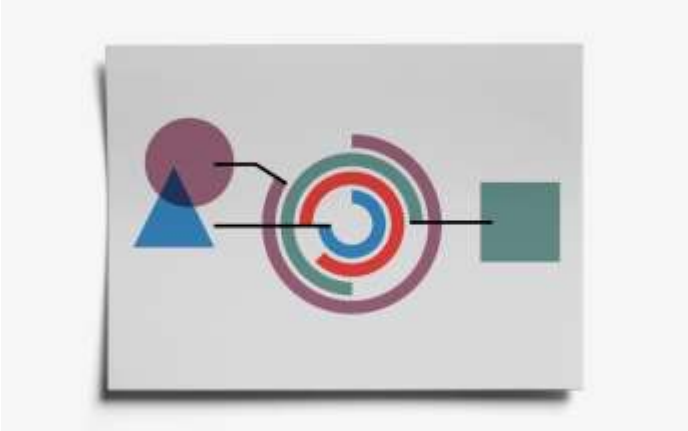
...that filters can't?

Slicers **can be applied to multiple tables.**

Multiple slicers can be applied to **the same data set.**

Slicers **can be moved around in your workbook.**

Slicers are more useful in **creating intuitive and interactive dashboards**.



To ensure an accurate, streamlined PivotTable, keep in mind that...

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

**Solo Exercise:**

# PivotTables Practice | Prompts

**8 minutes**

In your existing analysis sheet or in a new worksheet, answer these questions:

1. Do rates of sales differ significantly by profit margin category?
   (**Hint:** Remember, we created a margin category definition in the last lesson.)

2. What are the minimum, average, and maximum profits for orders by margin category? (**Extra practice**: Convert all to currency.)

3. What percentage of all orders occurred in each customer category, broken down by profit margin category?

4. What is the highest performing subcategory by sales in the “Consumer” customer category? (**Hint**: Use a slicer!)

**Solo Exercise:**

# PivotTables Practice | Methods and Solutions

How did you reach your answers? Check out the following methods:

| | |
|---|---|
| **1.**<br>Rows: margin_category<br>Values: Count of Sales | **2.**<br>Rows: profit_margin_category<br>Values: Min of Profit, Max of Profit, and Average of Profit |
| **3.**<br>Rows: profit_margin_category<br>Columns: segment<br>Values: count of order ID (set to show values as % of grand total) | **4.**<br>Rows: subcategory<br>Values: sum of sales<br>Slicer: Consumer segment |

Aggregating Data With PivotTables

# Wrapping Up

# Recap

**Today, we...**

- Applied **aggregate functions** in order to summarize our data.
- Used **PivotTables** to summarize data.
- Followed an efficient workflow in Excel with tables.

# Looking Ahead

**Next Class:** Communicating With Excel

# Additional Resources

- Ten PivotTable Problems and Easy Fixes
- Blog Post: Tables and Linking Data Structures in Excel
- Blog Post: Grouping in PivotTables