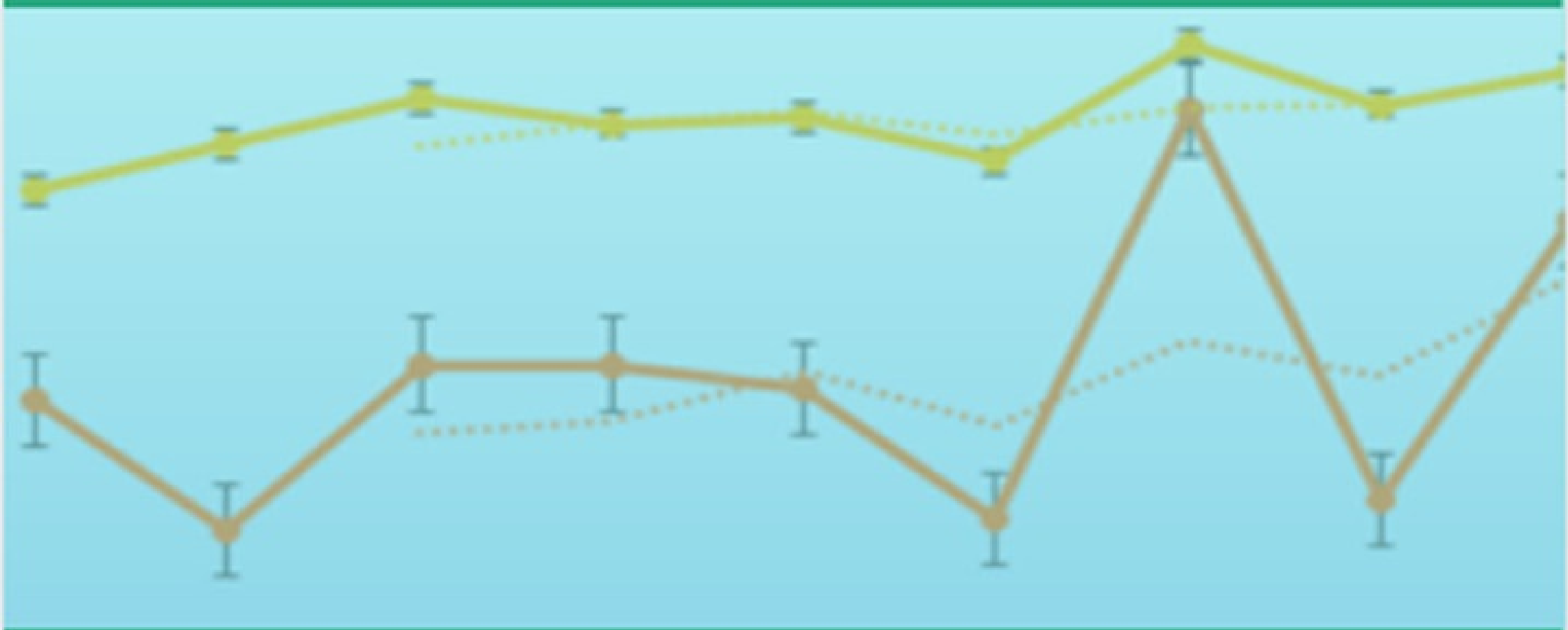
Data Analysis & Dashboarding with Microsoft® Excel®



Downloadable practce & solution workbooks
Step-by-step instructions and snapshots

JASMEEN KAUR

Data Analysis & Dashboarding with Microsoft® Excel®

Copyright © 2022 Jasmeen Kaur
All rights reserved.

Contents

Introduction

Lesson 1: Organizing Data

1.1 Clean Data

- 1. Remove non-printable characters and extra spaces
- 2. Parse data
  - Parse data with Convert Text to Columns Wizard
  - Parse data with functions
- 3. Convert dates stored in text format into dates
- 4. Convert numbers stored as text into numbers
  - With the error indicator
  - Changing the cell format
- 5. Change text to desired case
- 6. Spell Check
- 7. Remove duplicates
- List of functions used

1.2 Ranges

- Creating Named Ranges
- Using the Named Ranges
- Managing the Named Ranges

1.3 Tables

- Creating a Table
- Naming a table
- Managing a Table
- Exploring important properties of a table
  - 1. Calculated columns
  - 2. Structured Table References
  - 3. Sizing Handle
  - 4. Convert to Range
  - 5. Column Headers
  - 6. Tables expand or contract automatically
  - 7. Table totals
  - 8. Table Styles

1.4 Conditional formatting

- Highlight Cells Rules
- Top / Bottom Rules
- Data Bars
- Color Scales
- Icon Sets
- New Rule - Formatting based on User-defined functions
- Manage Rules

1.5 Sort and Filter Data

- Sorting
- Filtering

1.6 Data Validation

Lesson 2: Analyzing Data with Functions

2.1 Basic Functions

SUM
SUMIF
SUMIFS
AVERAGE
AVERAGEIF
AVERAGEIFS
COUNT
COUNTIF
COUNTIFS
MEDIAN
MAX
MIN
LARGE
SMALL
STDDEV.P
RANK.EQ
QUARTILE.INC
PERCENTILE.INC
List of basic functions
2.2 Logical, Date & Time, and Information Functions
Logical Functions
AND
OR
NOT
XOR
IF
IF and AND/OR
NESTED IF
IFERROR
Date and Time functions
DATE
DATEDIF
NETWORKDAYS
DATEVALUE
DAY, MONTH, YEAR
WEEKDAY, WEEKNUM
TODAY
TIME
TIMEVALUE
HOUR, MINUTE, SECOND
NOW
Information functions
2.3 Cell References
Relative cell references
Absolute cell references
Mixed cell references
2.4 Lookup Functions
VLOOKUP

Using VLOOKUP for exact matches
Using VLOOKUP for approximate matches
Using VLOOKUP to merge two tables
HLOOKUP
Using HLOOKUP for exact matches
Using HLOOKUP for approximate matches
INDEX
MATCH
INDEX and MATCH
TWO-WAY INDEX and MATCH
Array Formulas
2.5 Forecasting
Using exponential smoothing
Using linear regression
Lesson 3: Exploring & Visualizing Data
3.1 Create and Analyze a PivotTable
Creating a blank PivotTable
Adding fields to the blank PivotTable
Nesting PivotTable fields
Adding Filters to the PivotTable
Adding Slicers to the PivotTable
Making changes to a field in the Values Area
Adding multiple copies of a field in the Values Area
Changing PivotTable Design
3.2 PivotTable Tips and Tricks
3.3 Visualize Data with Charts
Inserting a chart
Chart types
1. Column Charts
2. Line Charts
3. Bar Charts
4. Area Charts
5. Pie Charts
6. Donut Charts
7. Scatter Charts
9. Treemap
10. Combo Charts
Formatting a chart
3.4 Advanced Charts
1. Picture Charts
2. Progress Doughnut Chart
3. Gantt Chart
3.5 Charting Best Practices
Lesson 4: Dashboarding
4.1 Identify Key Performance Indicators
KPIs
Creating KPIs
KPI reports

Building a KPI report
4.2 Create a Process Dashboard
4.3 Slicers, Timeline and PivotCharts
Create a named table
Insert a PivotTable
Insert a PivotChart
Insert Slicers and a Timeline
4.4 Steps to create a KPI dashboard
4.5 Create a Sales Dashboard
4.6 Dashboard Best Practices
Designing your dashboard
Distributing your dashboard
Printing your dashboard

Introduction

Data analysis is one of the most in-demand digital skills today. It helps companies rationally assess situations and make evidence-based business decisions. Microsoft Excel is the most popular data analysis tool used in business with estimates suggesting 750 million active Excel users. All business firms irrespective of their size and services use Excel to conduct diverse functions with 80-90% of the firms using spreadsheets for financial reporting and forecasting. According to a study conducted by Burning Glass Technologies and Capital One in 2017, more than 8 in 10 middle-skill jobs (82%) now require Excel skills. When interviewed, the salaries of those with analytical skills are 13-38% higher. Hence, a highly desirable combination of data analysis and Excel skills can help send the resume of a new job seeker to the top and can place middle-skilled workers in a highly favored position for future leadership.

This book aims to unlock the power of data using the data cleaning, analysis, visualization and dashboarding features available in Excel. The focus will be on walking the audience through the most powerful Excel features relevant to business data analysis.

The book will walk you through the complete lifecycle of data analysis using Excel, including cleaning data, analyzing and aggregating data, exploring aggregated data interactively, and visualizing and reporting key information and business insights to the management with professional and interactive dashboards. Practice workbooks will be available to download along with the book to help you follow the step-by-step instructions and snapshots to practice your newly acquired skills for data analysis and dashboarding in Excel. The snapshots are based on Microsoft® Excel® 2016 although they should work with Excel® 2013, Excel® 2019 or even Excel® 365.

The research[1] has shown that only studying individual Excel features in isolation results in amateur Excel users who understand features but not how they fit together. Instead of teaching Excel as a point-and-click tool, the book is designed in a way that previously covered features get recalled to ensure you solidify your understanding of how those features tie together as building blocks in the data analytics lifecycle. Towards the end, you will even develop a couple of cool interactive Excel dashboards from scratch!

Target Audience

The book is intended for:

- Anyone looking to pursue a career in data analysis and reporting
- Students and young adults who want to learn a skill set that is useful for any role in any industry
- Business professionals who know Excel but aren't confident about turning data into business insights

Although it is recommended that the audience has rudimentary functional knowledge of Excel, this book is also suitable for a beginner with basic computer skills. The book consists of four lessons providing you with solid theoretical understanding as well as the considerable practice of the data analysis tools and features in Excel.

Data Analytics Lifecycle

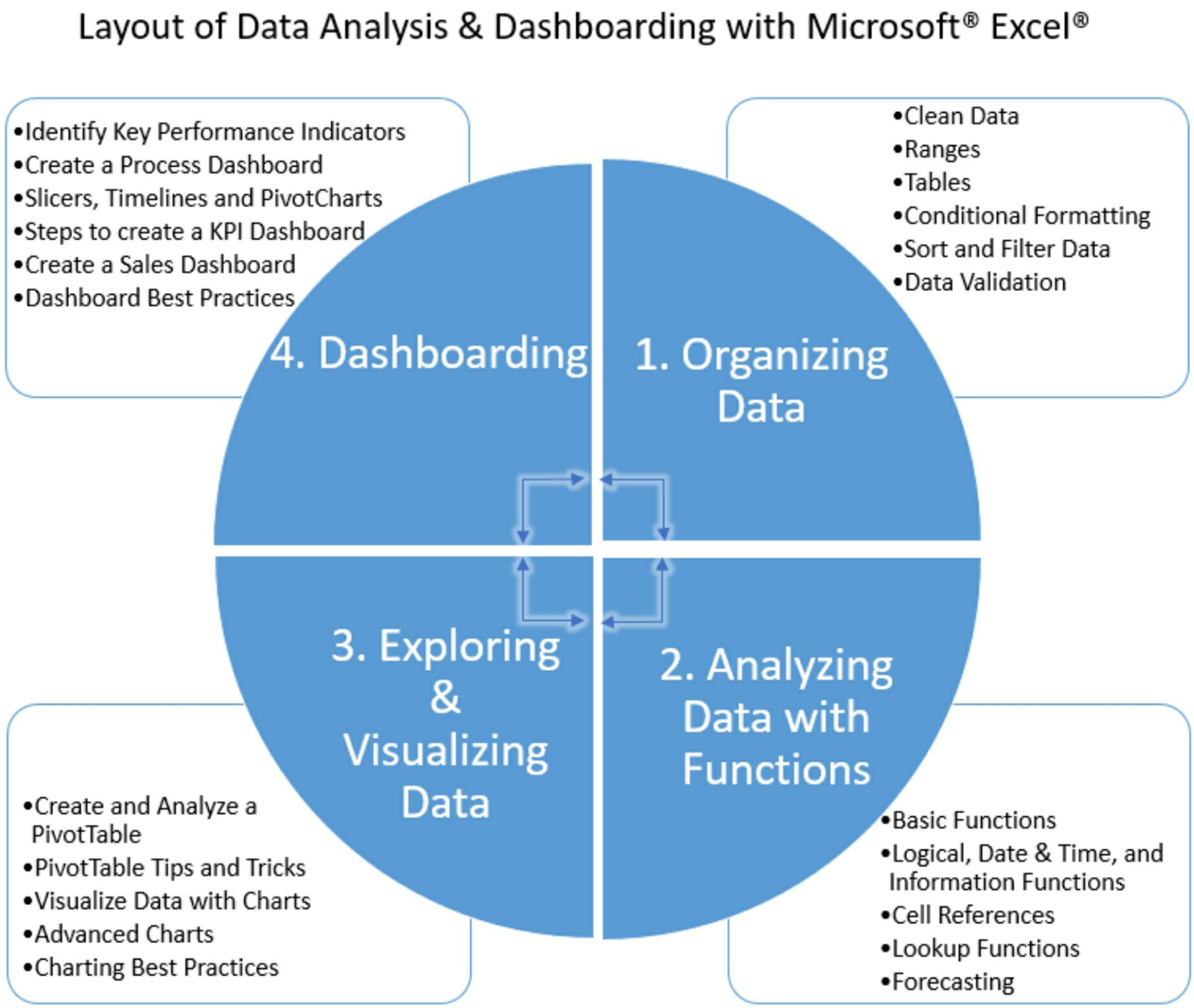
Microsoft Excel can be used to perform various data analysis functions, including data collection, structuring, aggregation, exploration, visualization and dashboarding. These capabilities of Excel can be employed to perform ad hoc as well as regular analysis and reporting. To get started with any analysis, **collecting data** relevant to the business problem or question and **organizing** it into a structure and format that makes analysis easy and efficient is paramount. While data collection refers to collecting and extracting data from several internal and external sources, including databases, spreadsheets, websites, etc., organizing data involves cleaning data to remove irrelevant information and arranging data in a structure that enables filtering and sorting of the data. Once the data is organized, there are a variety of Excel tools that may be utilized to analyze it. The choice of an ideal method or a combination of methods depends on the nature and complexity of the business requirements. One such method involves **analyzing data with functions**. Excel has some key simple and advanced functions that handle the majority of your Excel data analysis tasks. If the purpose is aggregating and **exploring data**, Excel PivotTables provide a fast, interactive and intuitive solution. They allow you to quickly explore and analyze raw data in Excel revealing powerful insights and trends. **Visualizing data** with Excel charts includes examining data in graphical format and obtaining additional insights regarding the trends within the data. **Reporting data** involves sharing the results of the data analysis in a format as required by the users to support their decision-making and further action via static reports or dynamic dashboards. Both utilize data visualization techniques, such as tables and charts which help in communicating the message clearly and efficiently to the users, however, dashboards are more interactive and drillable than static reports.

Data analysis is a cyclical and iterative process. Based on the trends you observe while visualizing data, you might need to go back to analyzing and aggregating it a bit differently. Based on the feedback by the users, you might

even need to collect more data. Any of the steps will need to be reiterated to meet new needs.

Layout of the book

The book contains four lessons with each of the lessons comprising of several chapters. Before we dive in, below is a brief description of each of the lessons:



Lesson 1

The lesson kicks off our journey through the data analytics lifecycle with steps to clean data. You will then learn about various Excel tools available to organize data in a format that makes data analysis easy. Topics to be covered include:

- Study a variety of functions to clean and format various fields within the data
- Create, use and manage named ranges to enhance analysis
- Create, use and manage and named tables to organize data
- Use conditional formatting to format and analyze data based on conditions
- Sort and filter data to retrieve the data you need
- Apply data validation for error-free data collection

Lesson 2

Excel has a large variety of functions. This lesson explores a set of functions most relevant to data analysis. Starting with basic functions – SUM, SUMIF, AVERAGE, AVERAGEIF, etc. critical to using Excel for data analysis, the course will move on to advanced functions – VLOOKUP, HLOOKUP, INDEX & MATCH used to obtain information from different parts of a workbook. You will also be introduced to cell references used to input functions and create excel formulae. A detailed study of logical, information, and data & time functions to help extract information and manipulate data to fulfill specific business requirements will follow. Toward the end of this lesson, we will also cover forecasting, a technique used to determine future trends based on historical data.

Lesson 3

This lesson starts with exploring PivotTables that are unarguably the quickest data summarization and aggregation tool. You will learn to create and modify PivotTables to answer a variety of business questions. Then, tips and tricks to optimize the use of PivotTables will be discussed.

Moving on, the lesson will cover how to create and modify charts in Excel. Charts are one of the most common ways to present data visually. Starting with common chart types - line, bar, pie charts, etc., we will then explore a few advanced charting techniques available in Excel. You will get to understand various charts and will build them on your own. We will wrap up with charting best practices to help design effective charts.

Lesson 4

To bring it all home, this lesson covers the presentation of the analysis through professional dashboards. The section introduces key performance indicators (KPIs) used to access business health and steps to design them. Then you will follow steps, including organizing data, performing calculations, visualizing data through charts and tables and adding interactivity with slicers, timelines and PivotCharts, to set up two dashboards from scratch. You will practice how tools and features covered throughout the program fit together to create an interactive dashboard. The lesson will wrap up with key dashboard best practices to design, distribute and print dashboards.

Happy learning!

Lesson 1: Organizing Data

Press the Ctrl key and click here to download practice & solution workbooks. Alternatively, copy and paste below link to your browser:

https://www.dropbox.com/sh/ri7tg8halmojk53/AACZFF0I9bbkqcEddQxkkv-fa?dl=0

1.1 Clean Data

You don't always have control over the format and type of data that you import from an external data source, such as a database, text file, or web page. Before the data can be analyzed, it often needs to be cleaned up. Excel has many features to help you get data in the desired format. The basic steps for cleaning data are discussed below. Depending on the format of the initial data set, the task may involve all or some of the steps. After importing the data from an external data source and before starting data cleaning, it is a good practice to keep a backup copy of the original data.

1. Remove non-printable characters and extra spaces

When you import data from another application, it can have non-printable characters and/or excess spaces. The excess spaces can be leading or trailing spaces and/or extra spaces between words. Follow the steps in the below *Practice* section to remove non-printable characters and excess spaces.

Practice

1. Open *Clean Data_Practice.xlsx* and go to *Clean Data* sheet. The sheet contains raw data with details of candidates who applied for a job competition: Candidate ID, Birthdate, Experience, Gender and Email. The columns seems to be delimited by either a comma (,) or a slash (/). The entire data, for now, is in a single column and needs to be formatted prior to performing any analysis.

	A
1	**Candidate ID,Birthdate,Experience/Gender/Email**
2	PUY-046,20636,18/Malee/bkearford0@qq.com
3	WDG-▯852,30511,19/Female/mmaclaren1@creativecommons.org
4	BMM-114,36838,20/Malee/agadaud2@omniture.com
5	NYU-778,20427, 10/Female/dmccurt3@exblog.jp
6	VAZ-841,25487,22/FEMALE/ctrueman4@1und1.de
7	OFO-006,26150,14/Female/svise5@tinypic.com
8	GHZ-▯563,21247,23/male/mcaves6@earthlink.net
9	KQB-103,19412,6/Male/bvolante7@addthis.com
10	EYG-806,27839,5/MALE/jaiskrigg8@networksolutions.com
11	AQR-9▯32 ,19632,19/Male/agarlandb@tuttocitta.it

2. Excel's *CLEAN* function can be used to delete non-printable characters and excess spaces can be removed using the *TRIM* function. These functions can also be used concurrently. Select cell B2. Enter the formula *=TRIM(CLEAN (A2))* and then press *Enter*.

SUM | =TRIM(CLEAN(A2))

	A (CLEAN(text))	B
1	**Candidate ID,Birthdate,Experience/Gender/Email**	**Candidate ID,Birthdate,Experience/Gender/Email**
2	PUY-046,20636,18/Malee/bkearford0@qq.com	=TRIM(CLEAN(A2))
3	WDG-▯852,30511,19/Female/mmaclaren1@creativecommons.org	
4	BMM-114,36838,20/Malee/agadaud2@omniture.com	
5	NYU-778,20427, 10/Female/dmccurt3@exblog.jp	
6	VAZ-841,25487,22/FEMALE/ctrueman4@1und1.de	
7	OFO-006,26150,14/Female/svise5@tinypic.com	
8	GHZ-▯563,21247,23/male/mcaves6@earthlink.net	
9	KQB-103,19412,6/Male/bvolante7@addthis.com	
10	EYG-806,27839,5/MALE/jaiskrigg8@networksolutions.com	
11	AQR-9▯32 ,19632,19/Male/agarlandb@tuttocitta.it	

3. Now select cell B2. Hover over the bottom right corner of the cell until green plus sign appears. Let's call it drag handle. Drag the handle down to fill rest of the rows with a copy of the formula. Extra spaces and unwanted non-printable characters, if any will be removed.

	A	B
1	Candidate ID,Birthdate,Experience/Gender/Email	Candidate ID,Birthdate,Experience/Gender/Email
2	PUY-046,20636,18/Malee/bkearford0@qq.com	PUY-046,20636,18/Malee/bkearford0@qq.com
3	WDG-852,30511,19/Female/mmaclaren1@creativecommons.org	WDG-852,30511,19/Female/mmaclaren1@creativecommons.org
4	BMM-114,36838,20/Malee/agadaud2@omniture.com	BMM-114,36838,20/Malee/agadaud2@omniture.com
5	NYU-778,20427, 10/Female/dmccurt3@exblog.jp	NYU-778,20427, 10/Female/dmccurt3@exblog.jp
6	VAZ-841,25487,22/FEMALE/ctrueman4@1und1.de	VAZ-841,25487,22/FEMALE/ctrueman4@1und1.de
7	OFO-006,26150,14/Female/svise5@tinypic.com	OFO-006,26150,14/Female/svise5@tinypic.com
8	GHZ-563,21247,23/male/mcaves6@earthlink.net	GHZ-563,21247,23/male/mcaves6@earthlink.net
9	KQB-103,19412,6/Male/bvolante7@addthis.com	KQB-103,19412,6/Male/bvolante7@addthis.com
10	EYG-806,27839,5/MALE/jaiskrigg8@networksolutions.com	EYG-806,27839,5/MALE/jaiskrigg8@networksolutions.com
11	AQR-932 ,19632,19/Male/agarlandb@tuttocitta.it	AQR-932 ,19632,19/Male/agarlandb@tuttocitta.it

2. Parse data

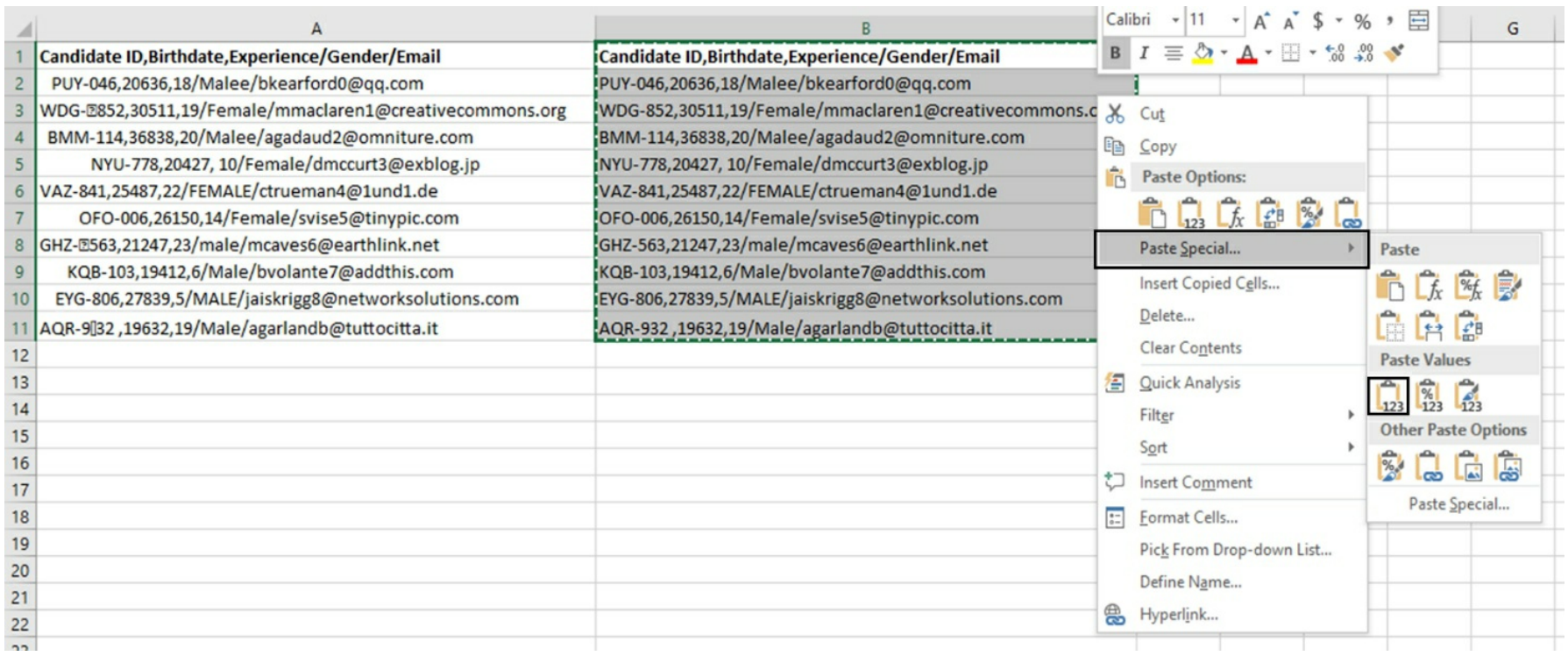
Parsing data means extracting it from one column into two or more columns. The data can be parsed by using the *Convert Text to Columns Wizard* or with help of parsing functions depending on the format of the data.

Parse data with Convert Text to Columns Wizard

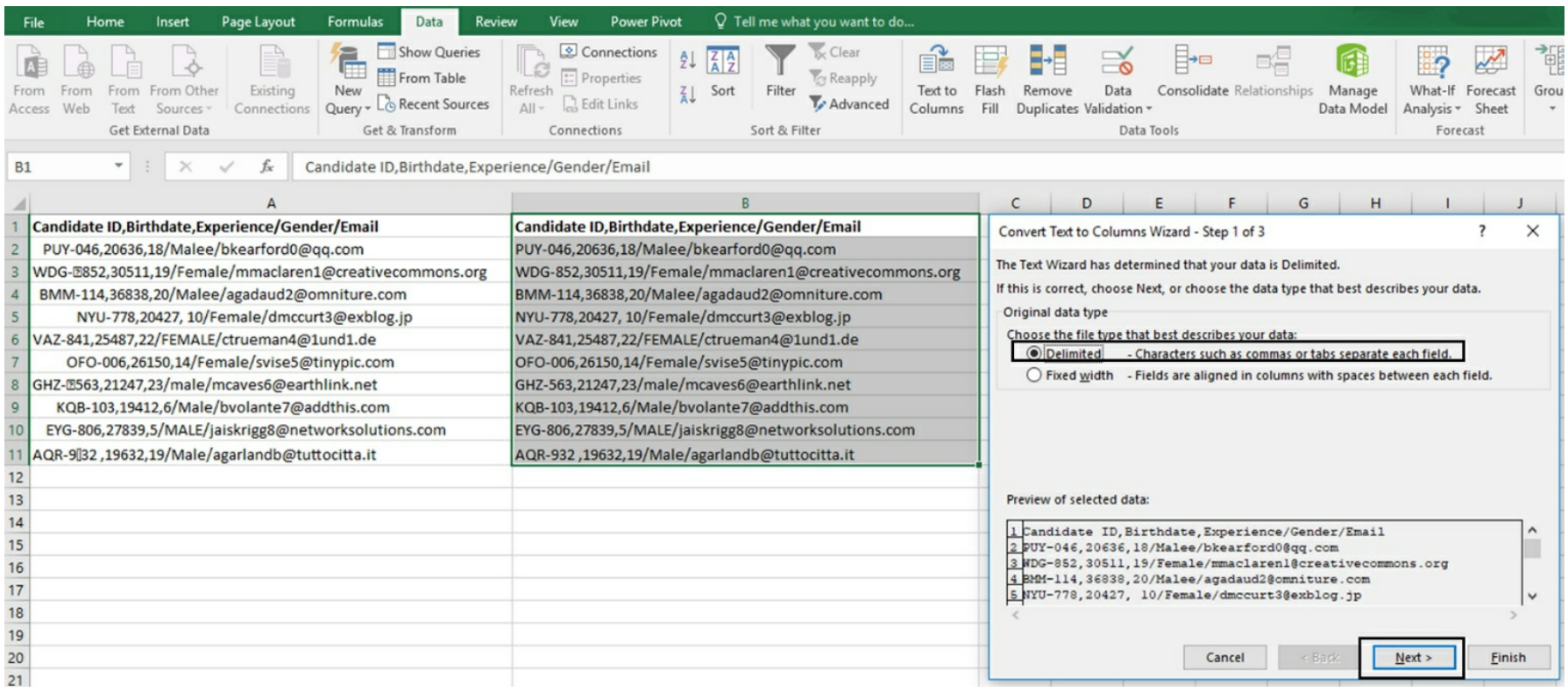
The data can be parsed using the *Convert Text to Columns Wizard* if imported data contains values delimited by spaces or characters. Your dataset is delimited by a comma (,) and a slash (/). Follow the below steps to parse the data.

Practice

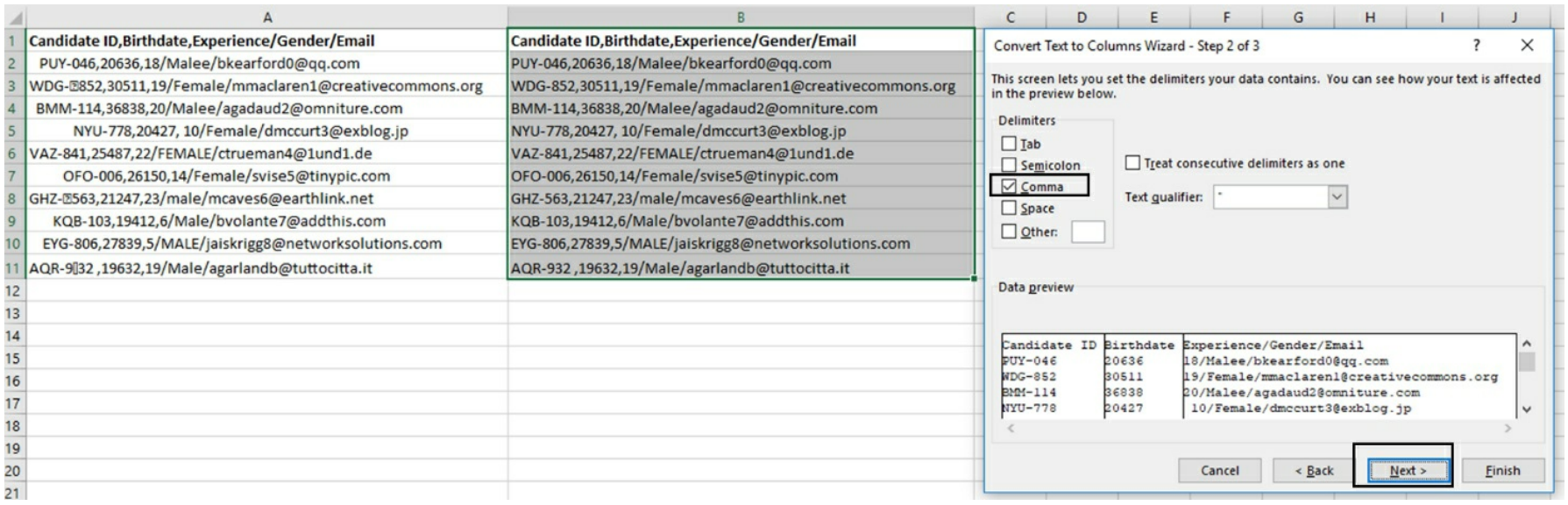
1. Continue working on the *Clean Data* sheet of the *Clean Data_Practice.xlsx*.
2. Cells of column B are embedded with functions. To make the column function free, select and copy range B1:B11, right-click and paste as values. Otherwise, functions get treated as input rather than data.



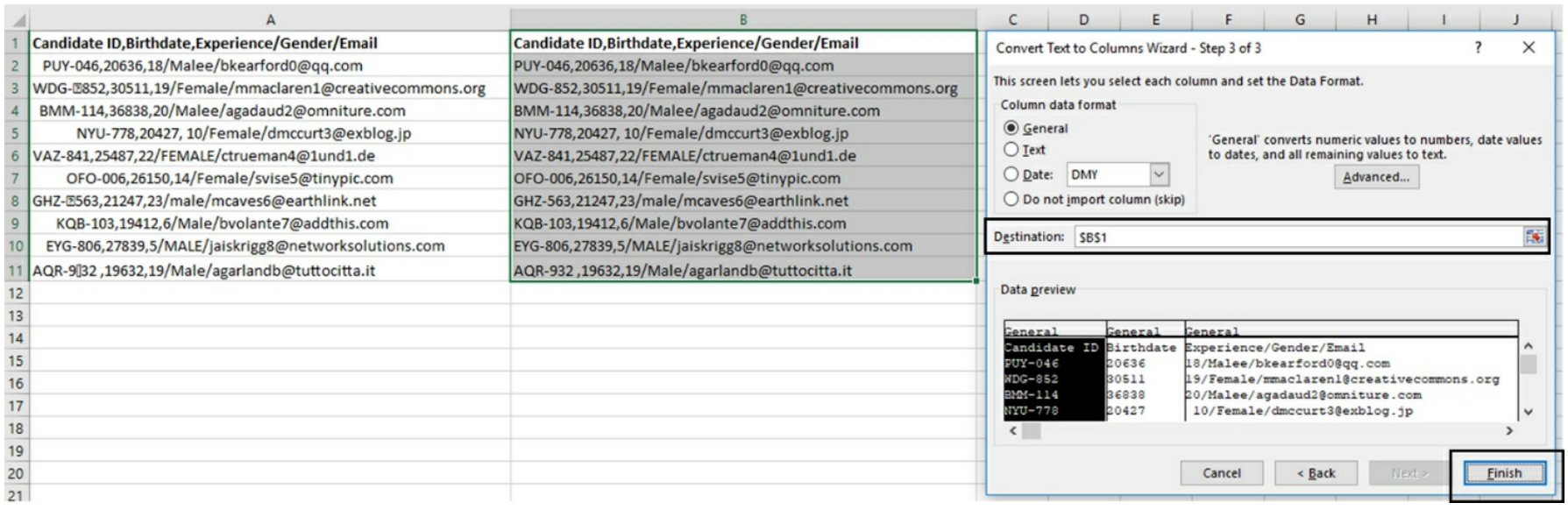
3. Now, keeping the range B1:B11 selected, go to the *Data* tab and under the *Data Tools* group, click on the *Text to Columns* icon. Step 1 of 3 of the *Convert Text to Columns Wizard* pops up.
4. Select *Delimited* and click *Next*.



5. For step 2 of the wizard, check the box next to *Comma* and uncheck the box next to *Tab*, if needed. A preview of the data becomes available on the window. If you're satisfied with the preview, click *Next*



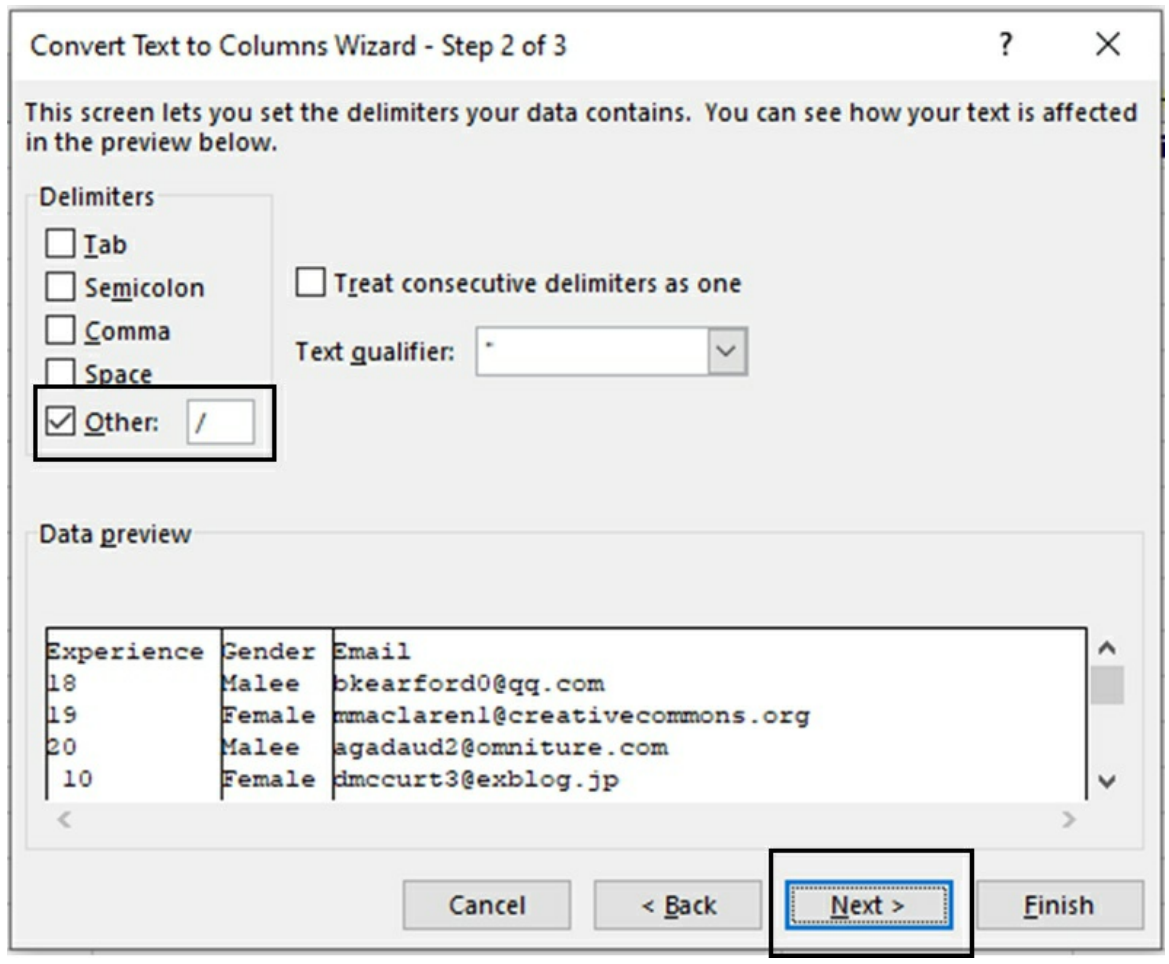
6. In step 3 of the wizard, select cell $B$1 in the *Destination* tab. In this step, you can also select and set the format for each of the columns in the data. A preview of the data becomes available on the window. If you're satisfied with the preview, click *Finish*.



7. Data previously in column B has now spread across columns B, C and D.

	A	B	C	D
1	Candidate ID,Birthdate,Experience/Gender/Email	Candidate ID	Birthdate	Experience/Gender/Email
2	PUY-046,20636,18/Malee/bkearford0@qq.com	PUY-046	20636	18/Malee/bkearford0@qq.com
3	WDG-852,30511,19/Female/mmaclaren1@creativecommons.org	WDG-852	30511	19/Female/mmaclaren1@creativecommons.org
4	BMM-114,36838,20/Malee/agadaud2@omniture.com	BMM-114	36838	20/Malee/agadaud2@omniture.com
5	NYU-778,20427, 10/Female/dmccurt3@exblog.jp	NYU-778	20427	10/Female/dmccurt3@exblog.jp
6	VAZ-841,25487,22/FEMALE/ctrueman4@1und1.de	VAZ-841	25487	22/FEMALE/ctrueman4@1und1.de
7	OFO-006,26150,14/Female/svise5@tinypic.com	OFO-006	26150	14/Female/svise5@tinypic.com
8	GHZ-563,21247,23/male/mcaves6@earthlink.net	GHZ-563	21247	23/male/mcaves6@earthlink.net
9	KQB-103,19412,6/Male/bvolante7@addthis.com	KQB-103	19412	6/Male/bvolante7@addthis.com
10	EYG-806,27839,5/MALE/jaiskrigg8@networksolutions.com	EYG-806	27839	5/MALE/jaiskrigg8@networksolutions.com
11	AQR-932 ,19632,19/Male/agarlandb@tuttocitta.it	AQR-932	19632	19/Male/agarlandb@tuttocitta.it

8. Now select range D1:D11, repeat steps 3 to 6 while checking the *Other* box and specifying "/" in step 2 of the Convert Text to Columns wizard and specifying $D$1 as destination in step 3.



9. Data previously in column D has now spread across columns D, E and F.

	B	C	D	E	F
1	Candidate ID	Birthdate	Experience	Gender	Email
2	PUY-046	20636	18	Malee	bkearford0@qq.com
3	WDG-852	30511	19	Female	mmaclaren1@creativecommons.org
4	BMM-114	36838	20	Malee	agadaud2@omniture.com
5	NYU-778	20427	10	Female	dmccurt3@exblog.jp
6	VAZ-841	25487	22	Female	ctrueman4@1und1.de
7	OFO-006	26150	14	Female	svise5@tinypic.com
8	GHZ-563	21247	23	Male	mcaves6@earthlink.net
9	KQB-103	19412	6	Male	bvolante7@addthis.com
10	EYG-806	27839	5	MALE	jaiskrigg8@networksolutions.com
11	AQR-932	19632	19	Male	agarlandb@tuttocitta.it

Parse data with functions

In absence of delimiters, LEFT, RIGHT or MID functions in combination with either FIND or SEARCH functions can be used to parse the data. At first, some of the formulas might look complex, but the logic is quite simple. The following example parses usernames and domains from email addresses, and progressively builds on the logic

behind these functions.

Practice

1. Continue working on the *Clean Data* sheet of the *Clean Data_Practice.xlsx*.
2. Name column G “Location of “@””. To extract the position of “@”, insert the following formula in G2, and then copy it down the column with the help of the fill handle:

=FIND(“@”, F2,1)

In this formula, FIND determines the position of “@” in the string stored in cell F2 starting from position 1 (from the beginning).

	F	G	H	I
1	**Email**	**Location of "@"**	**Email Username**	**Email domain**
2	bkearford0@qq.com	=FIND("@",F2,1)	=LEFT(F2,G2-1)	=RIGHT(F2,LEN(F2)-G2)
3	mmaclaren1@creativecommons.org			
4	agadaud2@omniture.com			
5	dmccurt3@exblog.jp			
6	ctrueman4@1und1.de			
7	svise5@tinypic.com			
8	mcaves6@earthlink.net			
9	bvolante7@addthis.com			
10	jaiskrigg8@networksolutions.com			
11	agarlandb@tuttocitta.it			

3. Name column H “Email Username”. To extract the username, insert the following formula in H2, and then copy it down the column:

=LEFT(F2,G2-1)

In this formula, the LEFT function extracts all the characters left to the position of “@” stored in cell G2. 1 is subtracted from the position to avoid extracting the @ itself.

4. Name column I “Email domain”. To extract the domain, insert the following formula in H2, and then copy it down the column:

=RIGHT(F2,LEN(F2)-G2)

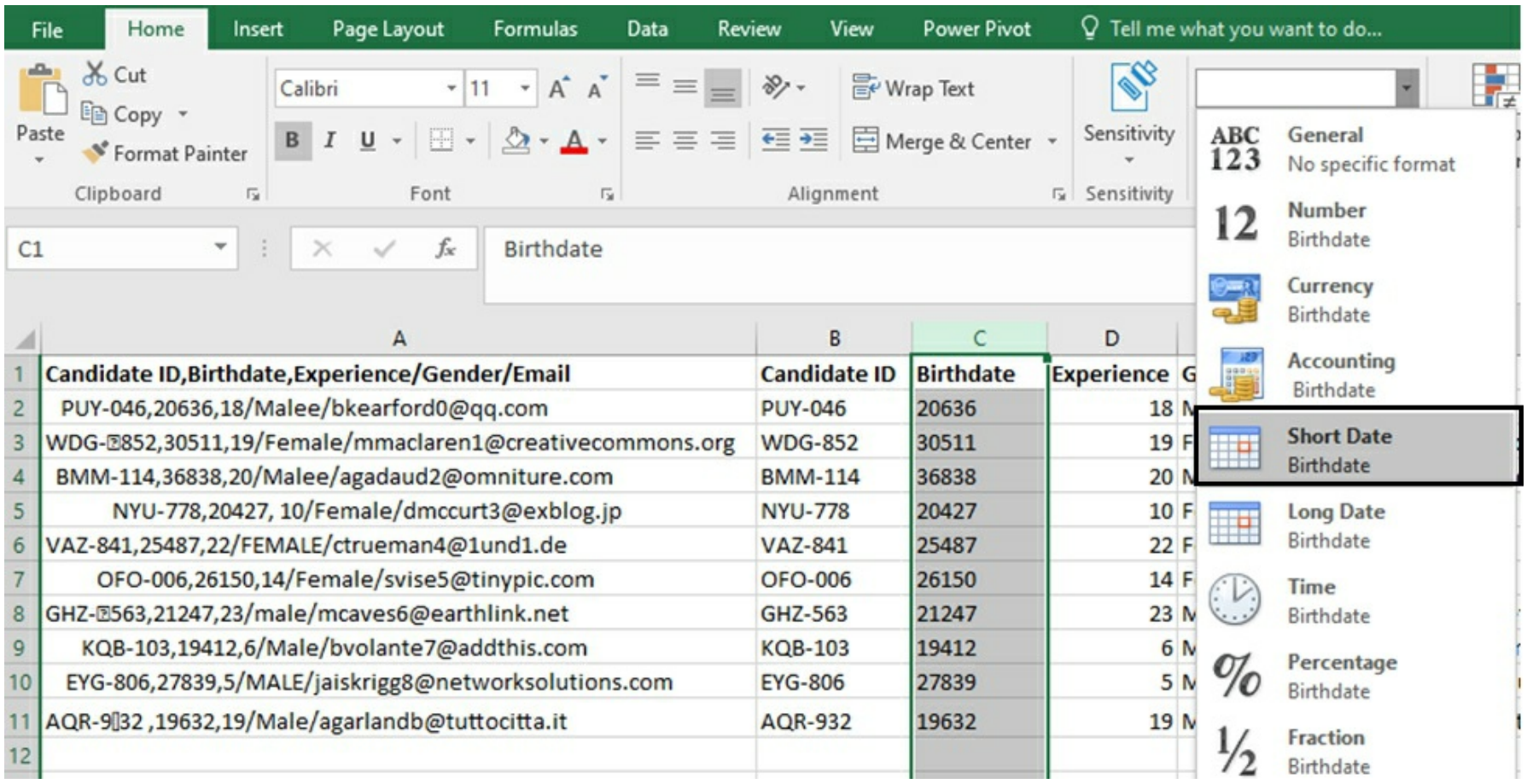
In this formula, the LEN function returns the total length of the email string stored in cell F2, from which the position of “@” (stored in cell G2) is subtracted. The difference is the number of characters to the right of @ that are extracted by the RIGHT function.

	F	G	H	I
1	**Email**	**Location of "@"**	**Email Username**	**Email domain**
2	bkearford0@qq.com	11	bkearford0	qq.com
3	mmaclaren1@creativecommons.org	11	mmaclaren1	creativecommons.org
4	agadaud2@omniture.com	9	agadaud2	omniture.com
5	dmccurt3@exblog.jp	9	dmccurt3	exblog.jp
6	ctrueman4@1und1.de	10	ctrueman4	1und1.de
7	svise5@tinypic.com	7	svise5	tinypic.com
8	mcaves6@earthlink.net	8	mcaves6	earthlink.net
9	bvolante7@addthis.com	10	bvolante7	addthis.com
10	jaiskrigg8@networksolutions.com	11	jaiskrigg8	networksolutions.com
11	agarlandb@tuttocitta.it	10	agarlandb	tuttocitta.it

3. Convert dates stored in text format into dates

Occasionally, data imported into Excel externally has dates stored in text format. Follow the below steps to convert a column containing dates in serial format into a date format.

1. Continue working on the *Clean Data* sheet of the *Clean Data_Practice.xlsx*.
2. Select column C named *Birthdate* by hovering over the column header until a down arrow appears and then clicking.
3. Go to the *Home* tab. Under the *Number* group, click on the drop-down and select *Short Date*.



4. The serial format will be converted into a date format

	A	B	C
1	**Candidate ID,Birthdate,Experience/Gender/Email**	**Candidate ID**	**Birthdate**
2	PUY-046,20636,18/Malee/bkearford0@qq.com	PUY-046	30/06/1956
3	WDG-852,30511,19/Female/mmaclaren1@creativecommons.org	WDG-852	14/07/1983
4	BMM-114,36838,20/Malee/agadaud2@omniture.com	BMM-114	08/11/2000
5	NYU-778,20427, 10/Female/dmccurt3@exblog.jp	NYU-778	04/12/1955
6	VAZ-841,25487,22/FEMALE/ctrueman4@1und1.de	VAZ-841	11/10/1969
7	OFO-006,26150,14/Female/svise5@tinypic.com	OFO-006	05/08/1971
8	GHZ-563,21247,23/male/mcaves6@earthlink.net	GHZ-563	03/03/1958
9	KQB-103,19412,6/Male/bvolante7@addthis.com	KQB-103	22/02/1953
10	EYG-806,27839,5/MALE/jaiskrigg8@networksolutions.com	EYG-806	20/03/1976
11	AQR-932 ,19632,19/Male/agarlandb@tuttocitta.it	AQR-932	30/09/1953

4. Convert numbers stored as text into numbers

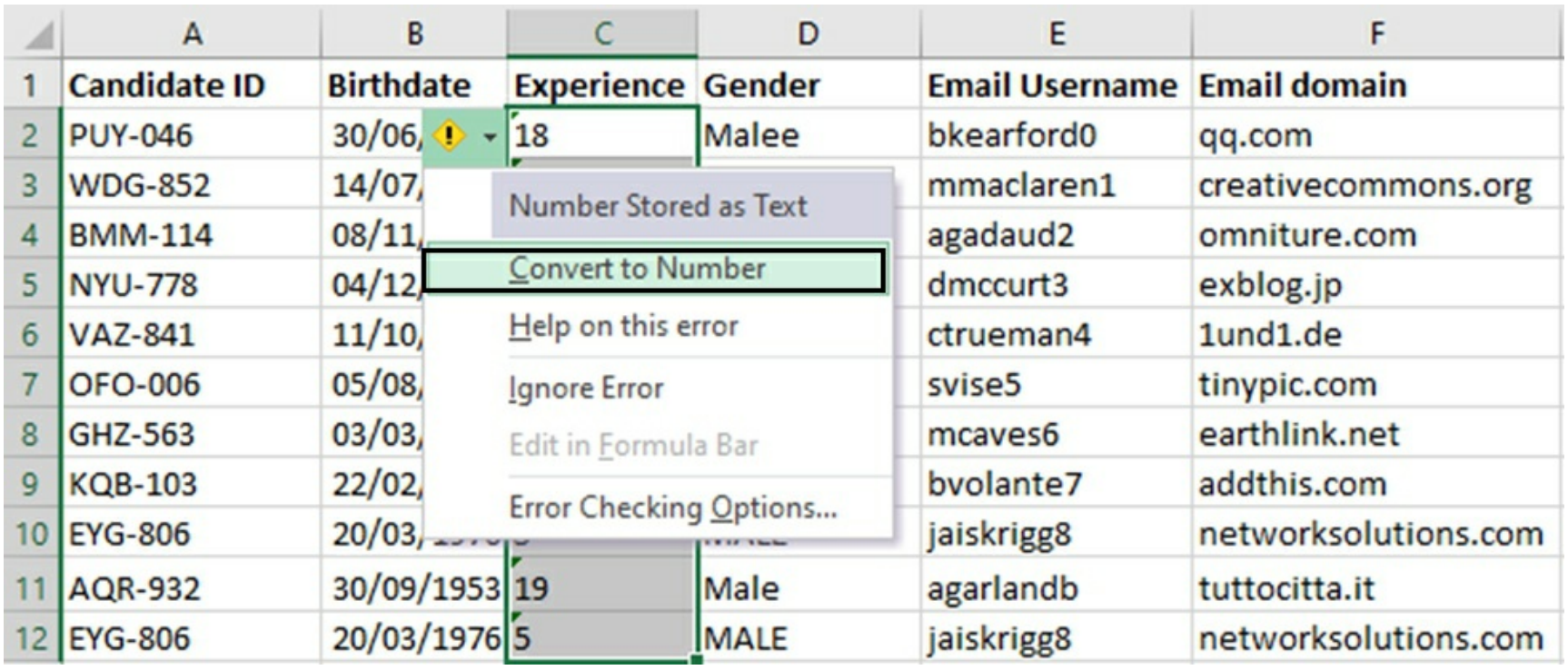
Sometimes imported data has numbers stored in text format. Numbers formatted as text are left-aligned instead of right-aligned in the cell, and are often marked with a green error indicator on the top left corner of the cell. We will discuss two methods to convert a column containing numbers in text format into a number format:

With the error indicator

If the cells display an error indicator (green triangle in the top left corner), converting text strings to numbers is a two-click process.

Practice

1. Open *Clean Data_Practice.xlsx* and go to *Convert Number* sheet. The sheet contains a bit cleaned up version of the data we've been dealing with.
2. The cells under the *Experience* column display an error indicator (green triangle in the top left corner) and are aligned to the left of the cell.
3. Select all the cells containing numbers formatted as text.
4. Click on the warning sign and select *Convert to Number*. This converts the numbers stored as text back to numbers.

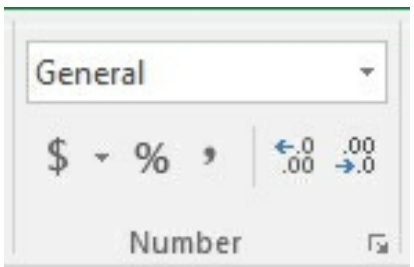


	A	B	C	D	E	F
1	Candidate ID	Birthdate	Experience	Gender	Email Username	Email domain
2	PUY-046	30/06/	18	Malee	bkearford0	qq.com
3	WDG-852	14/07/			mmaclaren1	creativecommons.org
4	BMM-114	08/11/			agadaud2	omniture.com
5	NYU-778	04/12/			dmccurt3	exblog.jp
6	VAZ-841	11/10/			ctrueman4	1und1.de
7	OFO-006	05/08/			svise5	tinypic.com
8	GHZ-563	03/03/			mcaves6	earthlink.net
9	KQB-103	22/02/			bvolante7	addthis.com
10	EYG-806	20/03/			jaiskrigg8	networksolutions.com
11	AQR-932	30/09/1953	19	Male	agarlandb	tuttocitta.it
12	EYG-806	20/03/1976	5	MALE	jaiskrigg8	networksolutions.com

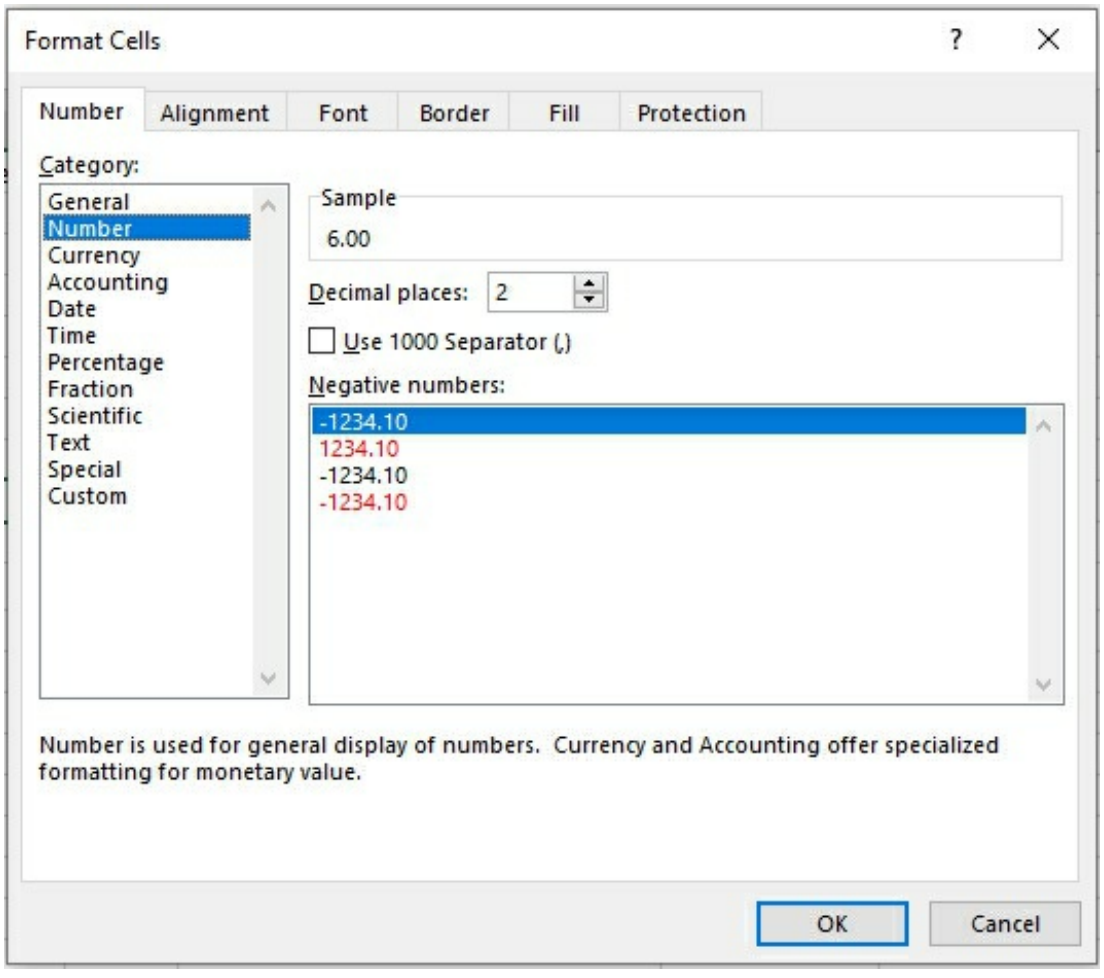
Changing the cell format

In some scenarios, you won't see a green error indicator appear in the upper-left corner of the cell with left-aligned numbers. In this case, you can change cell formatting to convert to number format with the below steps.

1. Select the cells that contain the numbers stored as text.
2. Go to the *Home* tab. In the *Number* group, click the Dialog Box Launcher next to *Number*.



3. A *Format Cells* dialog box will pop up. In the *Category* box, click the number format that you want to use and click OK. Selected cells will be converted to numbers.



5. Change text to desired case

To further clean up the data, UPPER, LOWER, or PROPER functions can be used to change the case of existing text to uppercase, lowercase, or proper case.

Practice

1. Continue working on the *Convert Number* sheet of the *Clean Data_Practice.xlsx*.

2. Select column E. Right click and select *Insert* to insert a new column to the right of column D.
3. Name the column E "Gender". To convert text in column D to proper case with first letter of every word capitalized, insert the following formula in cell E2, and then copy it down the column:

=PROPER(D2)

	A	B	C	D	E	F	G
1	**Candidate ID**	**Birthdate**	**Experience**	**Gender**	**Gender**	**Email Username**	**Email domain**
2	PUY-046	30/06/1956	18	Malee	=PROPER(D2)	bkearford0	qq.com
3	WDG-852	14/07/1983	19	Female		mmaclaren1	creativecommons.org
4	BMM-114	08/11/2000	20	Malee		agadaud2	omniture.com
5	NYU-778	04/12/1955	10	Female		dmccurt3	exblog.jp
6	VAZ-841	11/10/1969	22	Female		ctrueman4	1und1.de
7	OFO-006	05/08/1971	14	Female		svise5	tinypic.com
8	GHZ-563	03/03/1958	23	Male		mcaves6	earthlink.net
9	KQB-103	22/02/1953	6	Male		bvolante7	addthis.com
10	EYG-806	20/03/1976	5	MALE		jaiskrigg8	networksolutions.com
11	AQR-932	30/09/1953	19	Male		agarlandb	tuttocitta.it
12	EYG-806	20/03/1976	5	MALE		jaiskrigg8	networksolutions.com

4. To remove duplicate columns, select range E1 to E12, copy and paste as values at the same location to remove underlying formula and then delete column D.

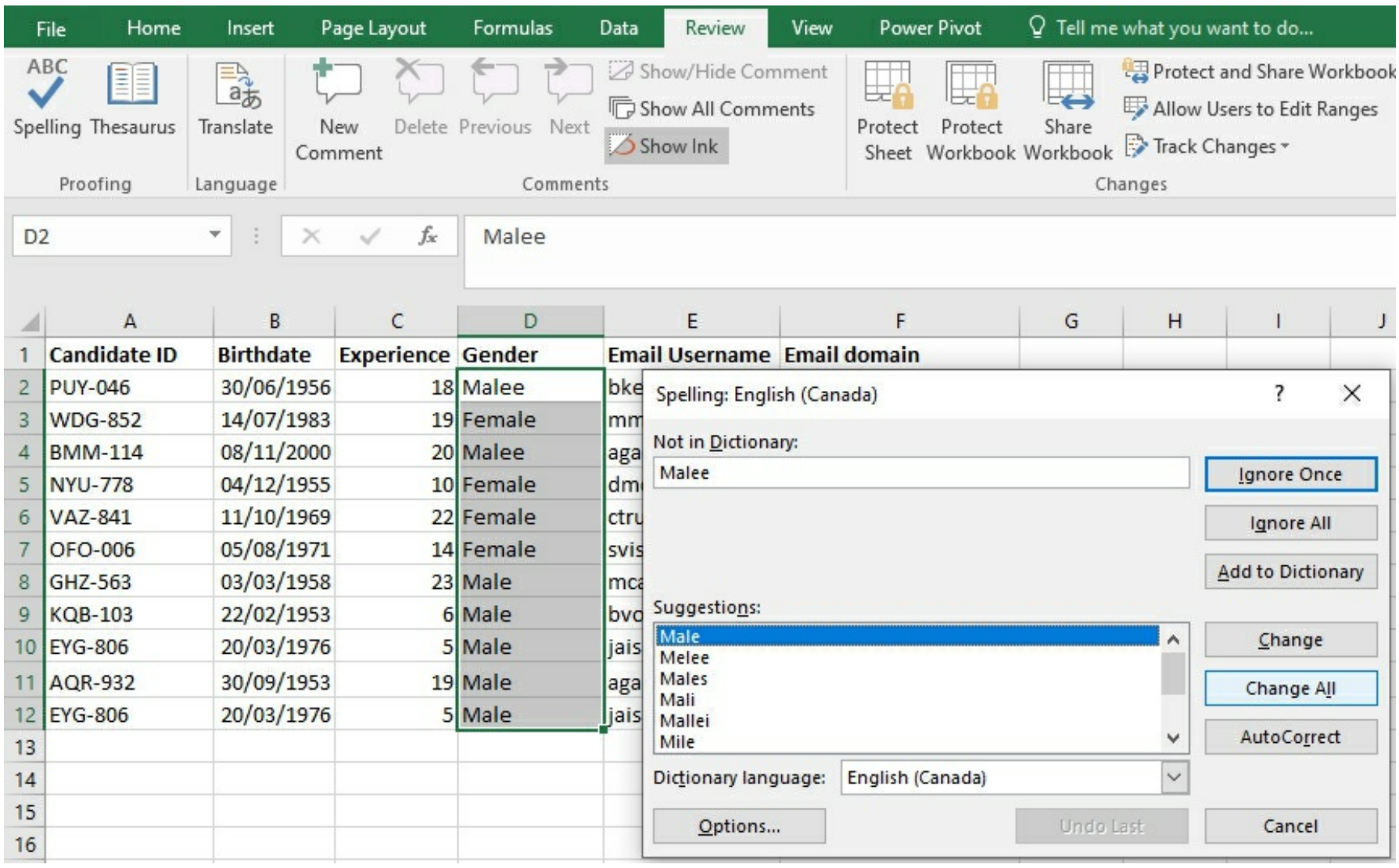
	A	B	C	D	E	F
1	**Candidate ID**	**Birthdate**	**Experience**	**Gender**	**Email Username**	**Email domain**
2	PUY-046	30/06/1956	18	Malee	bkearford0	qq.com
3	WDG-852	14/07/1983	19	Female	mmaclaren1	creativecommons.org
4	BMM-114	08/11/2000	20	Malee	agadaud2	omniture.com
5	NYU-778	04/12/1955	10	Female	dmccurt3	exblog.jp
6	VAZ-841	11/10/1969	22	Female	ctrueman4	1und1.de
7	OFO-006	05/08/1971	14	Female	svise5	tinypic.com
8	GHZ-563	03/03/1958	23	Male	mcaves6	earthlink.net
9	KQB-103	22/02/1953	6	Male	bvolante7	addthis.com
10	EYG-806	20/03/1976	5	Male	jaiskrigg8	networksolutions.com
11	AQR-932	30/09/1953	19	Male	agarlandb	tuttocitta.it
12	EYG-806	20/03/1976	5	Male	jaiskrigg8	networksolutions.com

6. Spell Check

The final data should be spelled and grammatically correct. Follow the below steps to spell check the *Gender* column.

Practice

1. Continue working on the *Convert Number* sheet of the *Clean Data_Practice.xlsx*.
2. Notice that under the *Gender* column, “Male” is misspelled as “Malee”. To correct this, select the column.
3. Go to the *Review* tab and under the *Proofing* group, click on *Spelling*. Alternatively, press the F7 key (as a shortcut). A spelling window will pop up.
4. Under the suggestions box, select the correct spelling and click on *Change All* to apply the change to all cells of the column.



5. A *Spell Check Complete* message box will pop up. Click OK.

All options available through the Spelling window are mentioned below:

- Ignore Once – To skip or ignore the currently selected word.
- Ignore All – To ignore all instances of the word.
- Add to Dictionary – To add spelling error term on your dictionary, so it doesn't prompt an error.
- Change – To change selected spelling errors with suggestions.
- Change All – To change all spelling errors with suggestions.
- AutoCorrect – To correct mistakes using Excel Autocorrect.

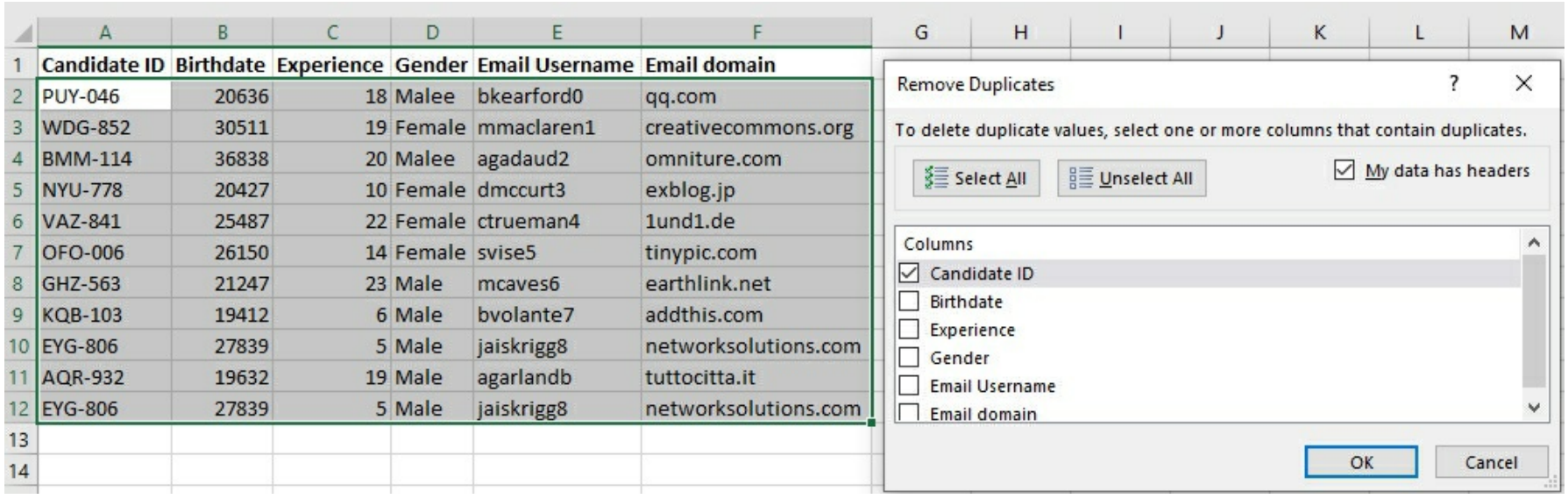
7. Remove duplicates

Lastly, you might want to remove duplicate rows from your data. The decision to remove or retain duplicate rows depends on the purpose of the analysis. Sometimes duplicate data is useful, while at other times it can yield incorrect results. When you remove duplicates, the duplicate data will be permanently deleted. Before you delete the duplicates, it's a good idea to backup the original data in case you need to refer back to it.

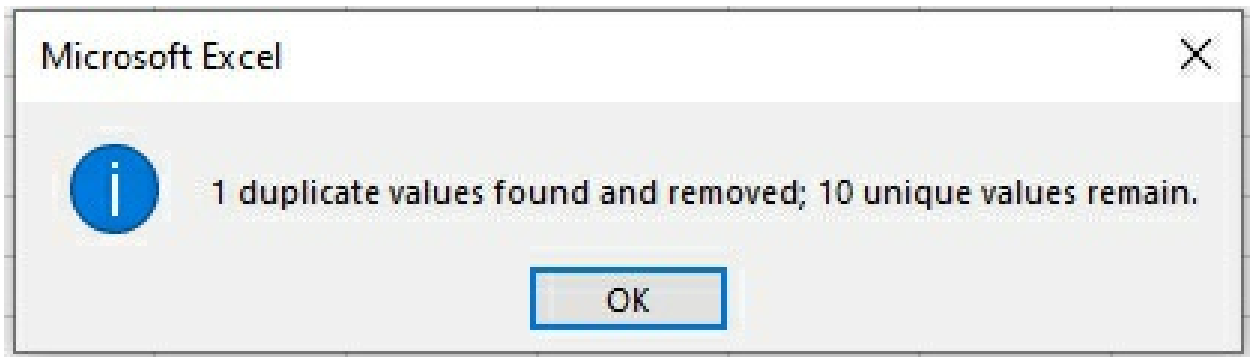
If you closely review the dataset on the *Convert Number* sheet, you will notice that rows 10 and 12 are duplicated. Follow the below steps to remove duplicates from the candidate data based on *Candidate ID*.

Practice

1. Continue working on the *Convert Number* sheet of the *Clean Data_Practice.xlsx*.
2. Select the entire range of data.
3. Go to the *Data* tab and under the *Data Tools* group, click *Remove Duplicates*.
4. A *Remove Duplicates* window will pop up. Under *Columns*, check (and uncheck, as needed) to select the *Candidate ID* column as you want to remove the duplicates based on this column and click OK.



5. Click OK on the message window displaying the number of rows removed and retained.



List of functions used

Function	Description
=CLEAN()	Removes all nonprintable characters from the text
=TRIM()	Removes all spaces from text except for single spaces between words.
=EXACT()	Compares two text strings and returns TRUE if they are exactly the same, FALSE otherwise.
=FIND()	Returns the starting position of one text string within another text string. FIND is case-sensitive.
=SEARCH()	Returns the number of the character at which a specific character or text string is first found, reading left to right. SEARCH is not case-sensitive.
=LEFT()	Returns the specified number of characters from the start of a text string
=RIGHT()	Returns the specified number of characters from the end of a text string
=MID()	Returns the specified number of characters from the middle of a text string starting at a specified position
=LEN()	Returns the number of characters in a text string
=LOWER()	Converts all letters in a text string to lowercase
=UPPER()	Converts all letters in a text string to uppercase
=PROPER()	Converts a string to a proper case; capitalizes the first letter in each word of a text value and all other letters to a lower case
=DOLLAR()	Converts a number to text, using the $ (dollar) currency format
=FIXED()	Rounds a number to the specified number of decimals and returns the result as text with or without commas
=TEXT()	Converts a value to text in a specific number format
=VALUE()	Converts a text string that represents a number to a number
=REPLACE()	Replaces part of text string with a specified, new text string
=SUBSTITUTE()	Substitutes new text for old text in a text string
=CONCATENATE()	Joins several text items into one text item
=CONCAT()	Joins several text strings into one text string
=REPT()	Repeats text a given number of times

1.2 Ranges

A named range is simply a single cell or a range of cells assigned with a name. The assigned name can be used in place of cell references in formulas and for defining the source data for graphs and data validation. Named ranges are useful in making the formulas easier to understand, type and debug, and help make the spreadsheets more dynamic and easy to update. For example, as below, *=Sales_Price+(Service_Tax*Sales_Price)* is easier to understand than *=C2+(C3*C2)*.

	A	B	C	D
1				
2		Sales Price	1500	
3		Service Tax	8%	
4				
5		Total Paid	1620	=C2+(C3*C2)
6		Total Paid	1620	=Sales_Price+(Service_Tax*Sales_Price)

Below syntax rules need to be followed for naming the ranges:

- Any combination of letters, numbers and symbols - underscores, backslashes, and periods. Other symbols including spaces are not allowed.
- Names are not case-sensitive.
- Names cannot mimic a valid cell reference, e.g., R2, D2.
- Names can begin with a character, underscore or backslash, but not a number.
- Name length can be up to 255 characters, however, it is advised to keep them short and easy to interpret, such as Interest_Rate, Sales_2019.

Creating Named Ranges

Named Ranges can be created in three ways:

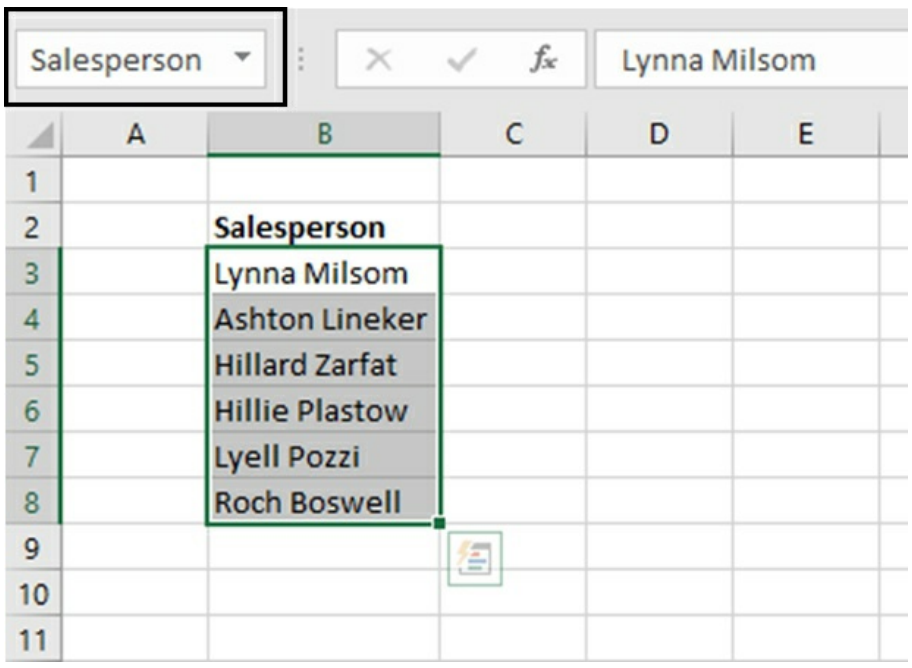
1. Using the Name box
2. Using the Define Name dialog box
3. Using the Create from Selection dialog box

1. Using the Name box

The fastest way to create a named range is by using the Name box situated to the left of the formula bar.

Practice

4. Open *Ranges_Practice.xlsx* and go to *Name Box* sheet.
5. Select cell range B3:B8. The range comprises names of sales personnel. Click on the Name box on the left of formula bar. Delete its contents. Type “Salesperson” and press enter to create the named range.

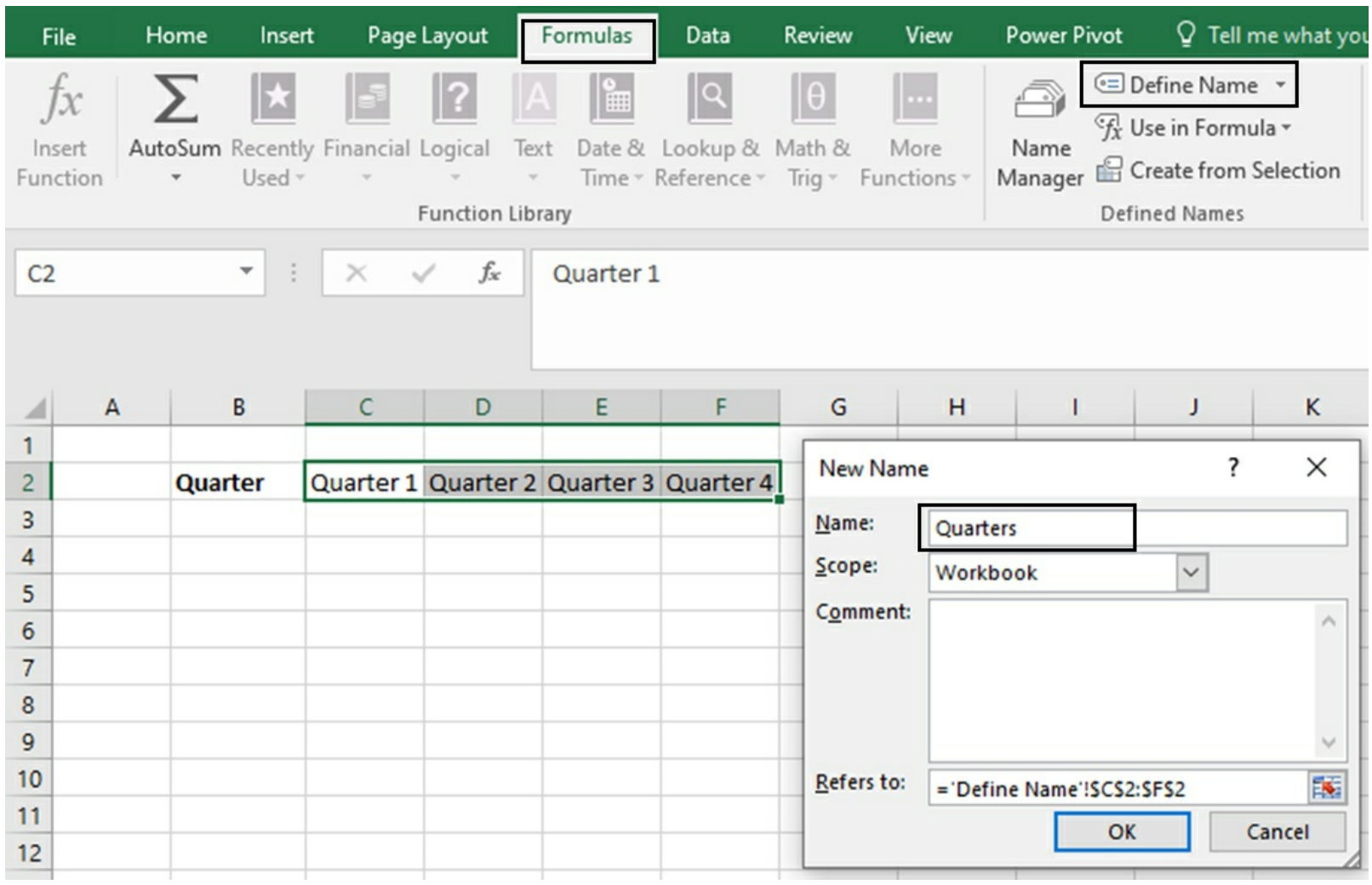


2. Using the *Define Name* dialog box

This is the second method of creating a named range by using a selection of cells in the worksheet.

Practice

1. Open *Ranges_Practice.xlsx* and go to *Define Name* sheet. Select cell range C3:F2. The range comprises quarters of a year.
2. Click on the *Formulas* tab on the top ribbon.
3. Click *Define Name* under the *Defined Names* group. A *New Name* dialog box will pop up.
4. Type the name "Quarters" in the *Name* box.
5. Confirm the range in the *Refers To* box is correct and click *OK*.

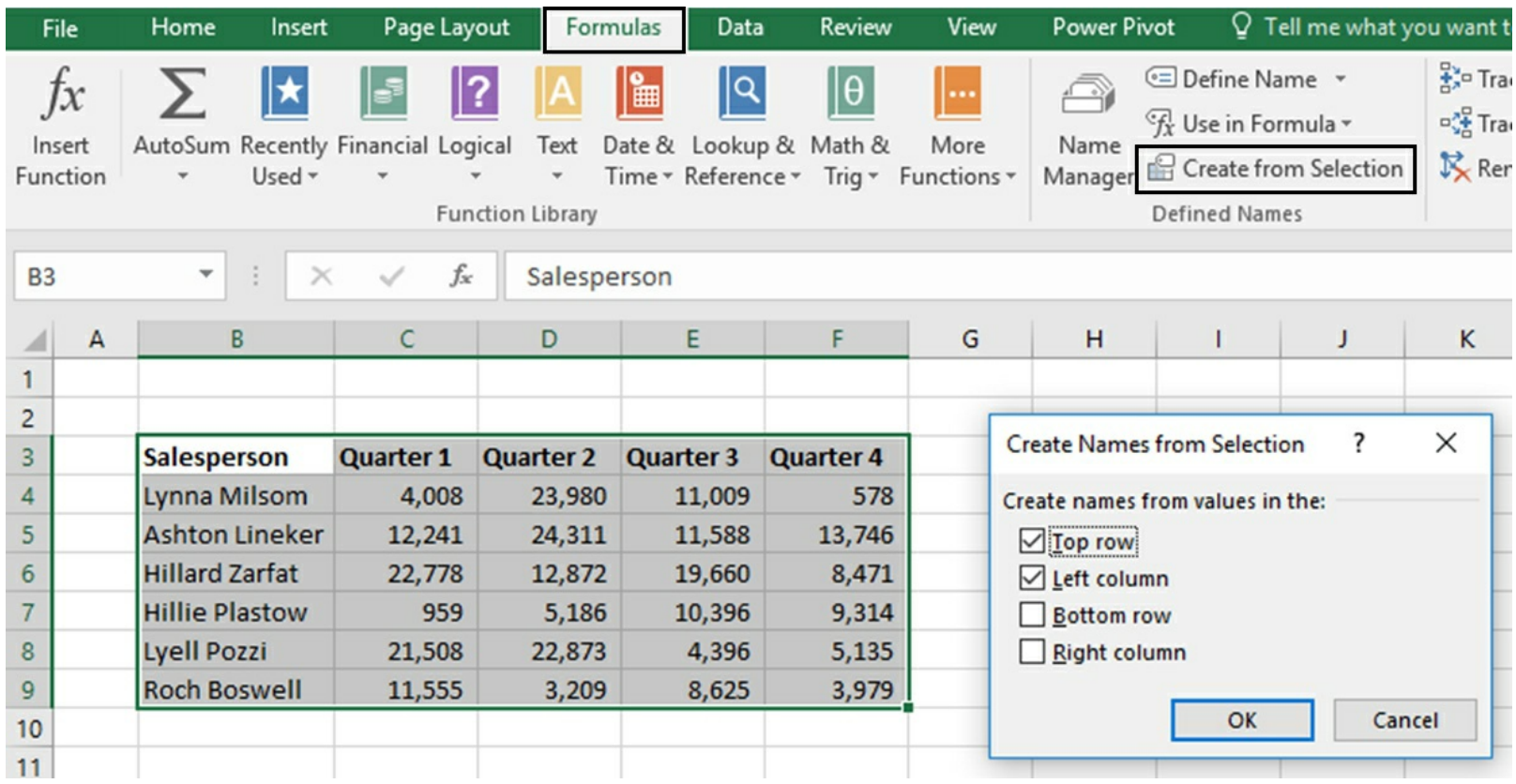


3. Using the Create from Selection dialog box

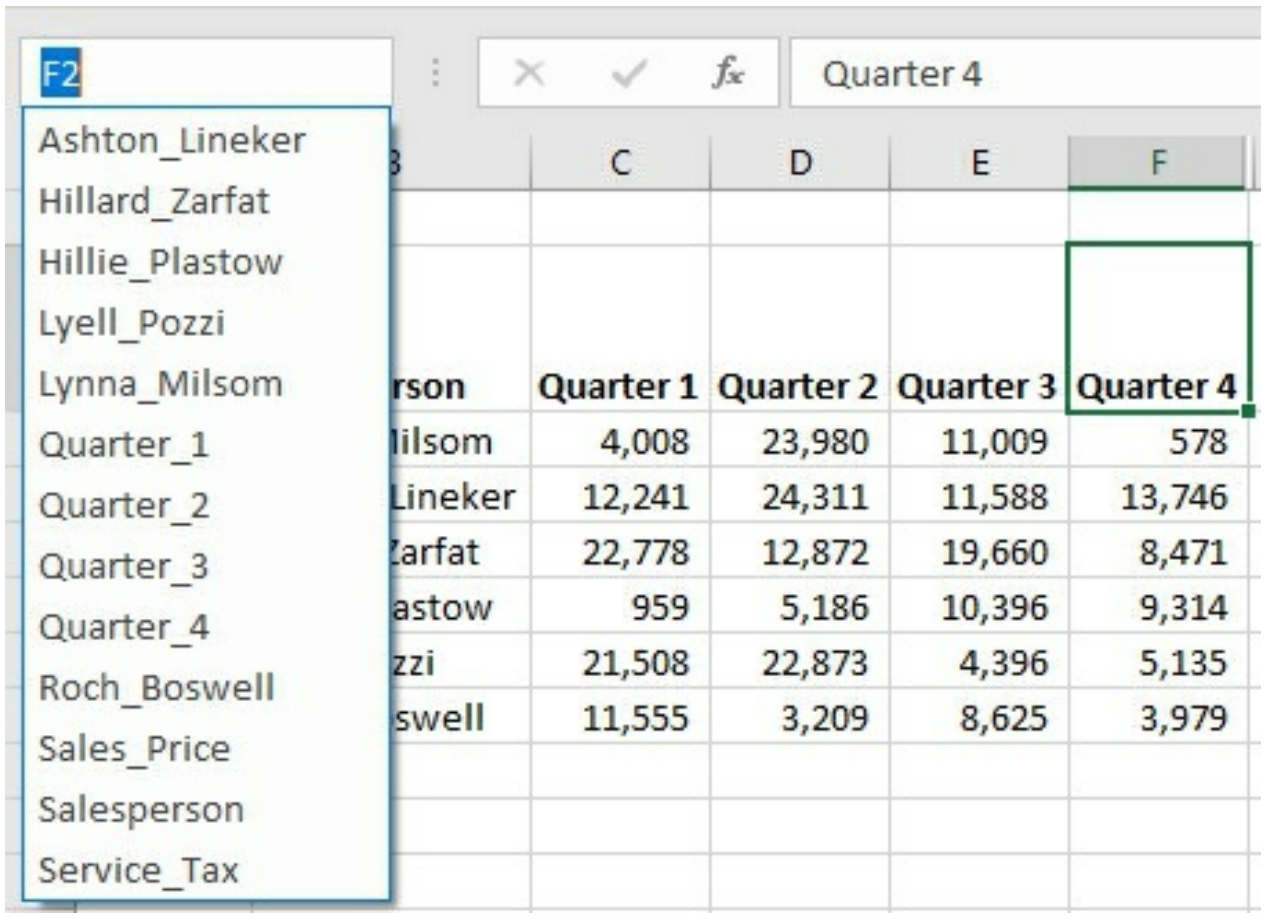
This is the third method of creating a named range by using a selection of cells in the worksheet.

Practice

1. Open *Ranges_Practice.xlsx* and go to *Create from Selection Box* sheet. Select cell range B2:F8. The range comprises of sales made by different salesperson in four quarters of a year. Names are listed as row headers and quarters as column headers.
2. Click on the Formulas tab. In the *Defined Names* group, click on *Create from Selection*. The *Create Names from Selection* dialog box pops up.
3. The selected range has a top row header and a left column header. So check *Top row* and *Left column* options. This selection depends on the location of row or column headers. If there is only a header row at the top of the table, then just *Top row* will be checked.



4. Click *OK*. The names are now created and can be viewed by clicking on the drop down icon to the right of the name box.

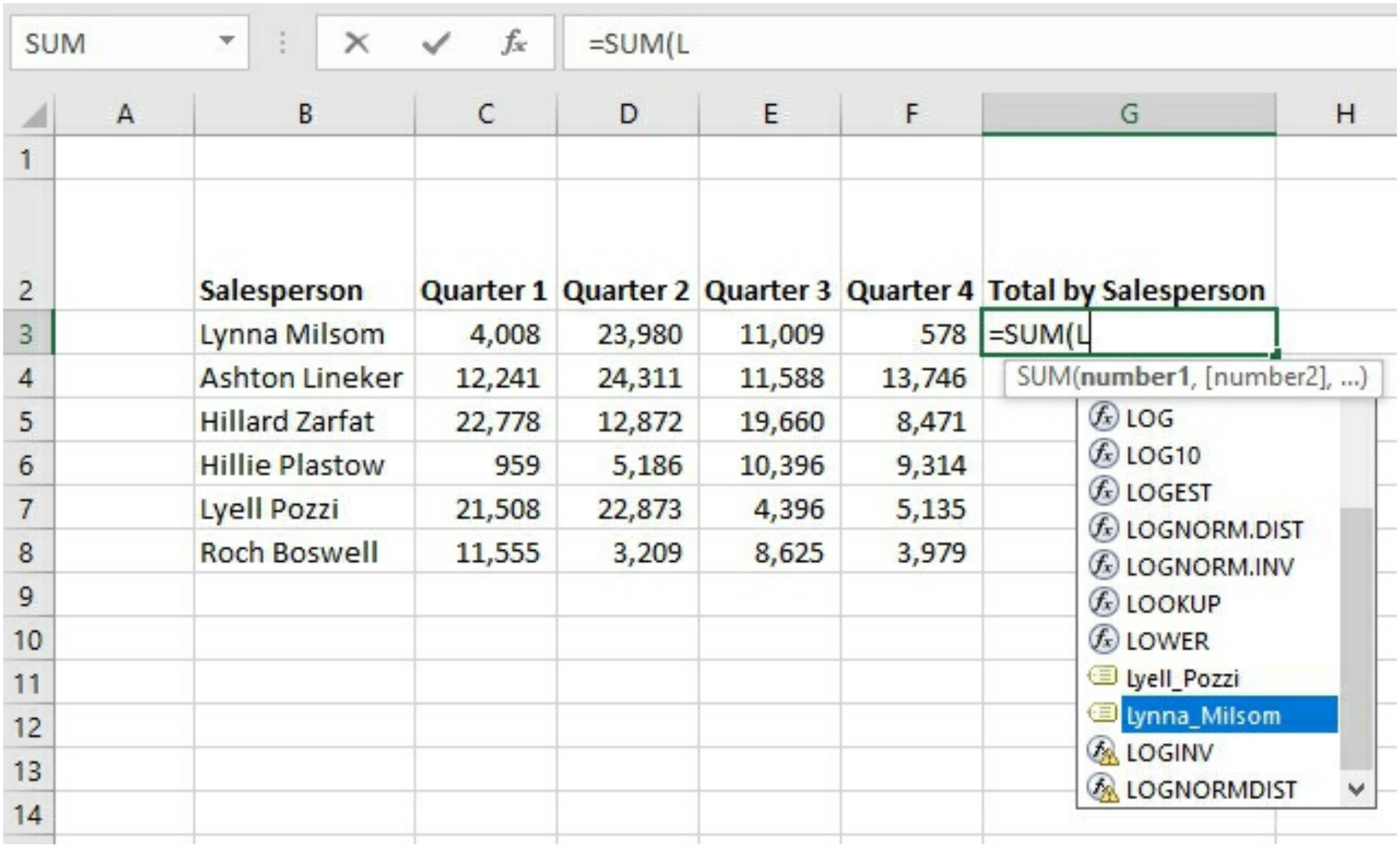


Using the Named Ranges

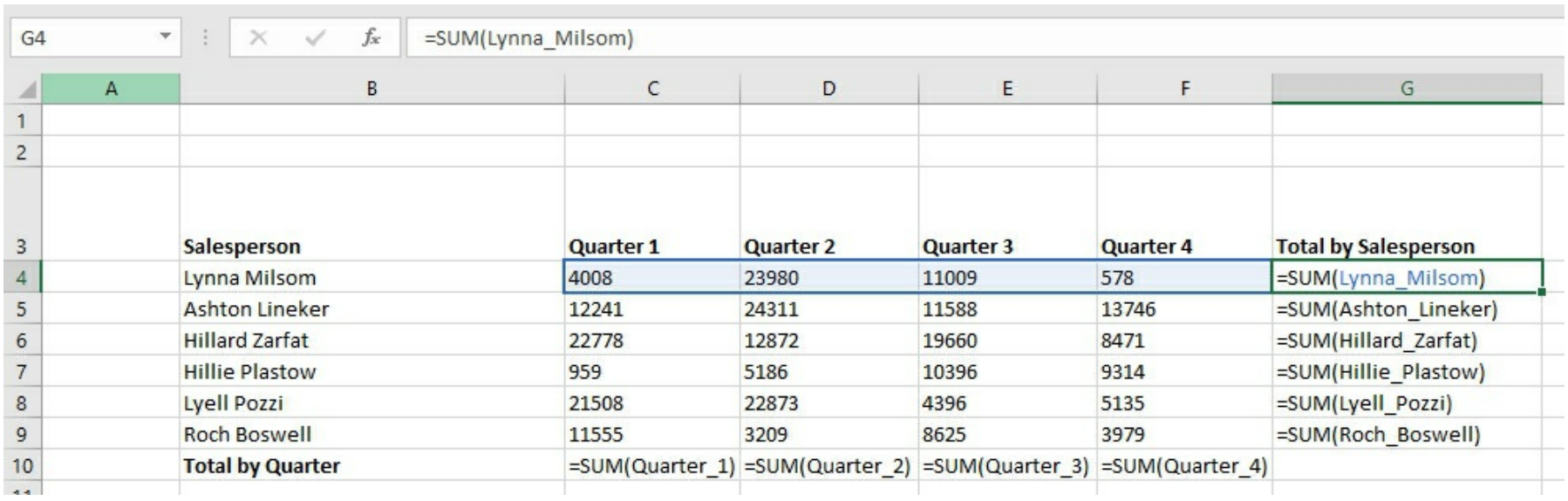
A Named Range will appear automatically in formulas on typing the first letter of the name. You can navigate to the name of your choice with help of down or up arrows and press the tab key to introduce the name into the formula.

Practice

1. Open *Ranges_Practice.xlsx* and go to *Using Named Ranges* sheet. The data on the sheet comprises of sales made by different salesperson of a company in four quarters of a year. Calculate total sales for each salesperson using the named ranges you just created. To do that, go to cell G3. Type the equal sign (=) and enter the syntax of SUM function but instead of numbers, press the letter "L" signifying the first letter of the salesperson's name (Lynna Milsom). Navigate to and select the name from the list that appears. Similarly, calculate sales for each salesperson with *=SUM(Salesperson_Name).*



2. Similarly, calculate sales for each quarter with =*SUM(Quarter_Number)*.



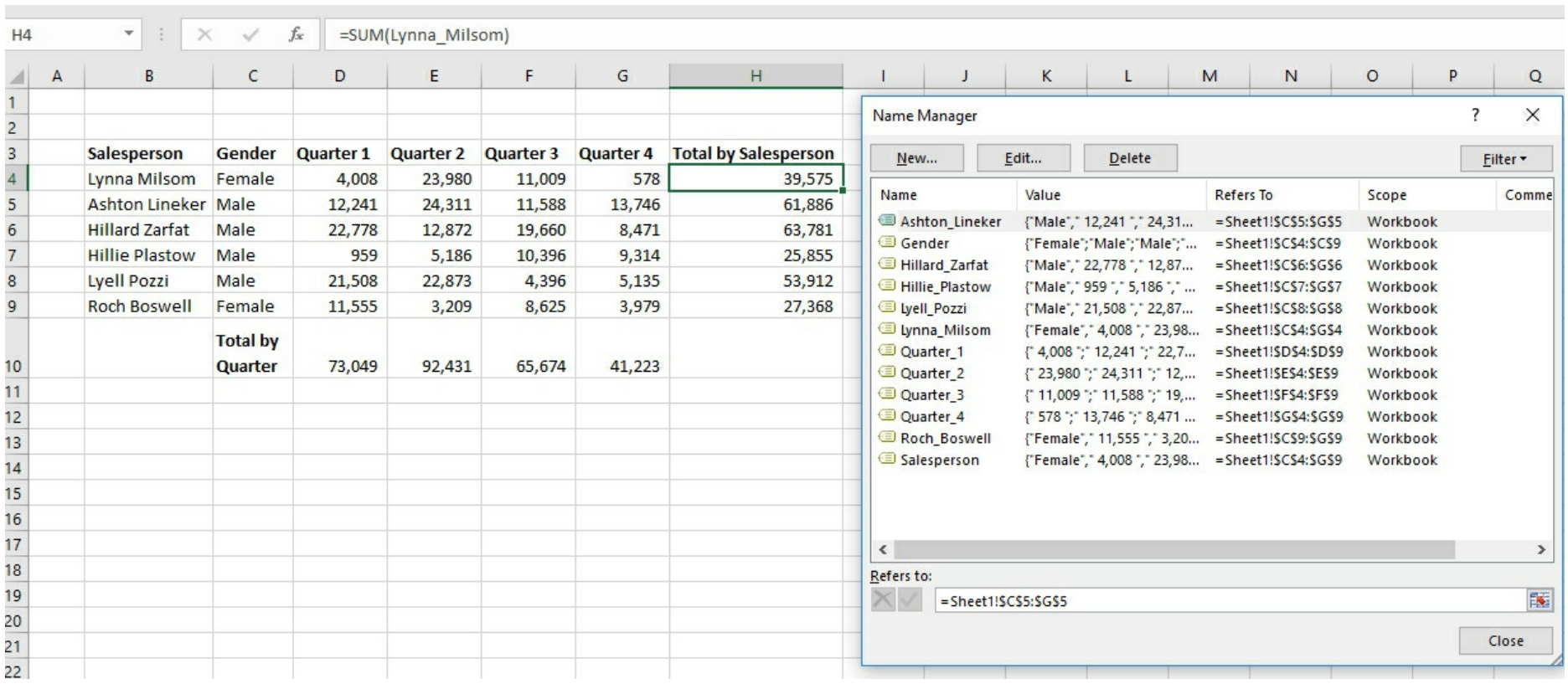
Managing the Named Ranges

The named ranges can be edited using the *Name Manager* box. To open the *Name Manager* box, click *Formulas*, then under the *Defined Names* group, click *Name Manager*. The box can also be accessed with the shortcut of Control + F3. It can also be used to view the named ranges.

The Name Manager provides icons to

- define a new name (*New* icon),
- edit an existing name (*Edit* icon),
- delete an existing name (*Delete* icon), and
- filter the existing names (*Filter* icon).

Selecting a named range and clicking on the *Edit* icon leads to an *Edit Name* pop-up that can be used to modify the range of a name that it refers to. The scope defaults to Workbook, which means the named range can be referenced within the workbook.



Start using named ranges in your Excel worksheets and you will quickly start appreciating their benefits!

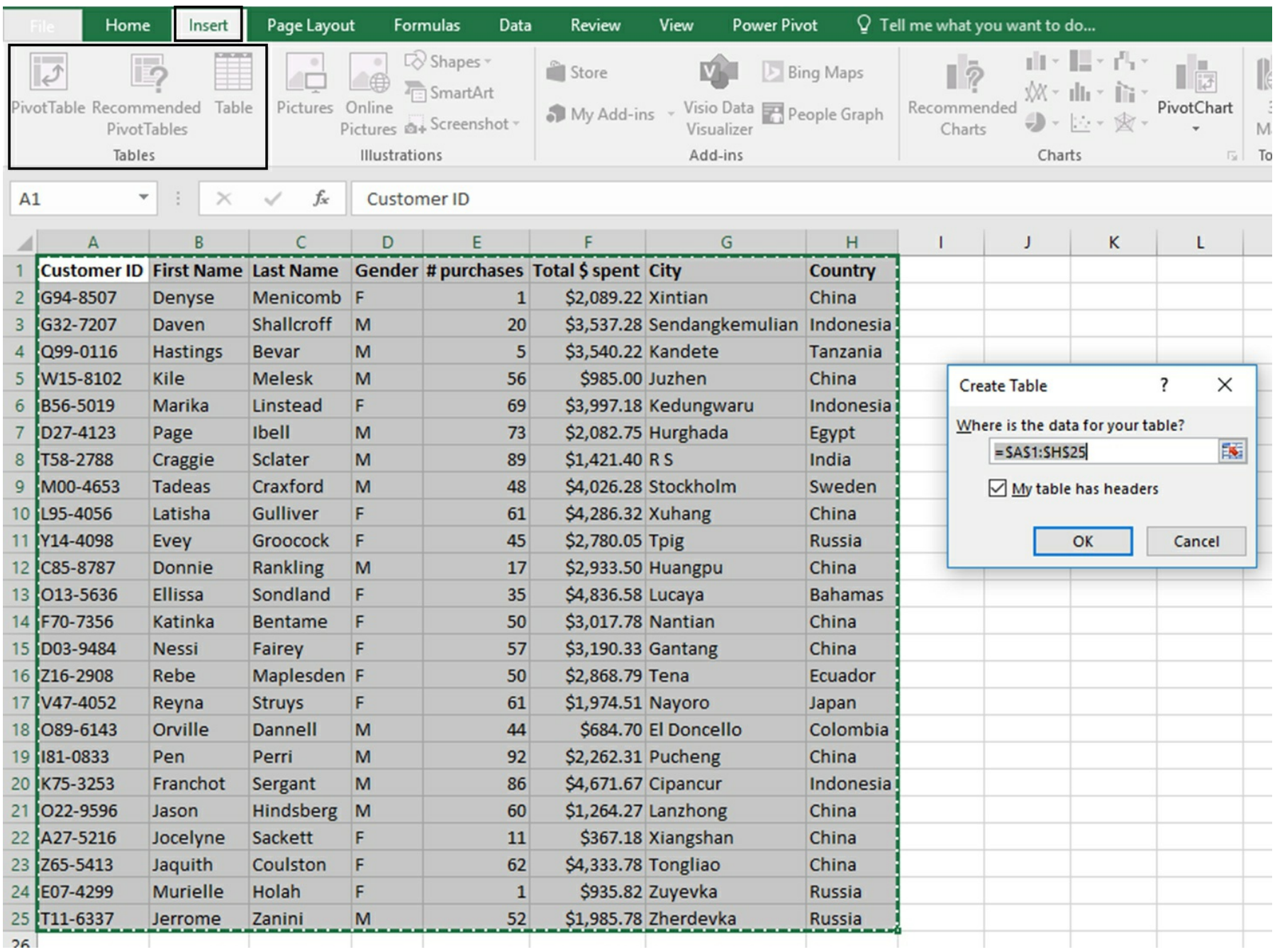
1.3 Tables

A table is a range of cells with structured data aimed to make analyzing and managing the data easier. A table contains a top row called a header row that describes the data contained in each column with the remaining rows containing information corresponding to unique records of the data. Tables are especially helpful when you need to append new data to the existing data on a regular basis and would prefer formulas that automatically update.

Creating a Table

Practice

1. Open *Tables_Practice.xlsx* and go to *Creating a Table* sheet. Select data on the sheet. The range has 24 records of customer data along with a header row on the top.
2. Click on the *Insert* tab. Under the *Tables* group, click *Table*. A *Create Table* dialog box will pop up. Alternatively, use the shortcut Ctrl+T after selecting the range.
3. Confirm the range selected in the *Where is the data for your table?* box is correct.



Customer ID	First Name	Last Name	Gender	# purchases	Total $ spent	City	Country
G94-8507	Denyse	Menicomb	F	1	$2,089.22	Xintian	China
G32-7207	Daven	Shallcroff	M	20	$3,537.28	Sendangkemulian	Indonesia
Q99-0116	Hastings	Bevar	M	5	$3,540.22	Kandete	Tanzania
W15-8102	Kile	Melesk	M	56	$985.00	Juzhen	China
B56-5019	Marika	Linstead	F	69	$3,997.18	Kedungwaru	Indonesia
D27-4123	Page	Ibell	M	73	$2,082.75	Hurghada	Egypt
T58-2788	Craggie	Sclater	M	89	$1,421.40	R S	India
M00-4653	Tadeas	Craxford	M	48	$4,026.28	Stockholm	Sweden
L95-4056	Latisha	Gulliver	F	61	$4,286.32	Xuhang	China
Y14-4098	Evey	Groocock	F	45	$2,780.05	Tpig	Russia
C85-8787	Donnie	Rankling	M	17	$2,933.50	Huangpu	China
O13-5636	Ellissa	Sondland	F	35	$4,836.58	Lucaya	Bahamas
F70-7356	Katinka	Bentame	F	50	$3,017.78	Nantian	China
D03-9484	Nessi	Fairey	F	57	$3,190.33	Gantang	China
Z16-2908	Rebe	Maplesden	F	50	$2,868.79	Tena	Ecuador
V47-4052	Reyna	Struys	F	61	$1,974.51	Nayoro	Japan
O89-6143	Orville	Dannell	M	44	$684.70	El Doncello	Colombia
I81-0833	Pen	Perri	M	92	$2,262.31	Pucheng	China
K75-3253	Franchot	Sergant	M	86	$4,671.67	Cipancur	Indonesia
O22-9596	Jason	Hindsberg	M	60	$1,264.27	Lanzhong	China
A27-5216	Jocelyne	Sackett	F	11	$367.18	Xiangshan	China
Z65-5413	Jaquith	Coulston	F	62	$4,333.78	Tongliao	China
E07-4299	Murielle	Holah	F	1	$935.82	Zuyevka	Russia
T11-6337	Jerrome	Zanini	M	52	$1,985.78	Zherdevka	Russia

4. Check the *My table has headers* box to specify that your table has a top row that contains name of each of the columns.
5. The selected range is converted into a table with the default style.

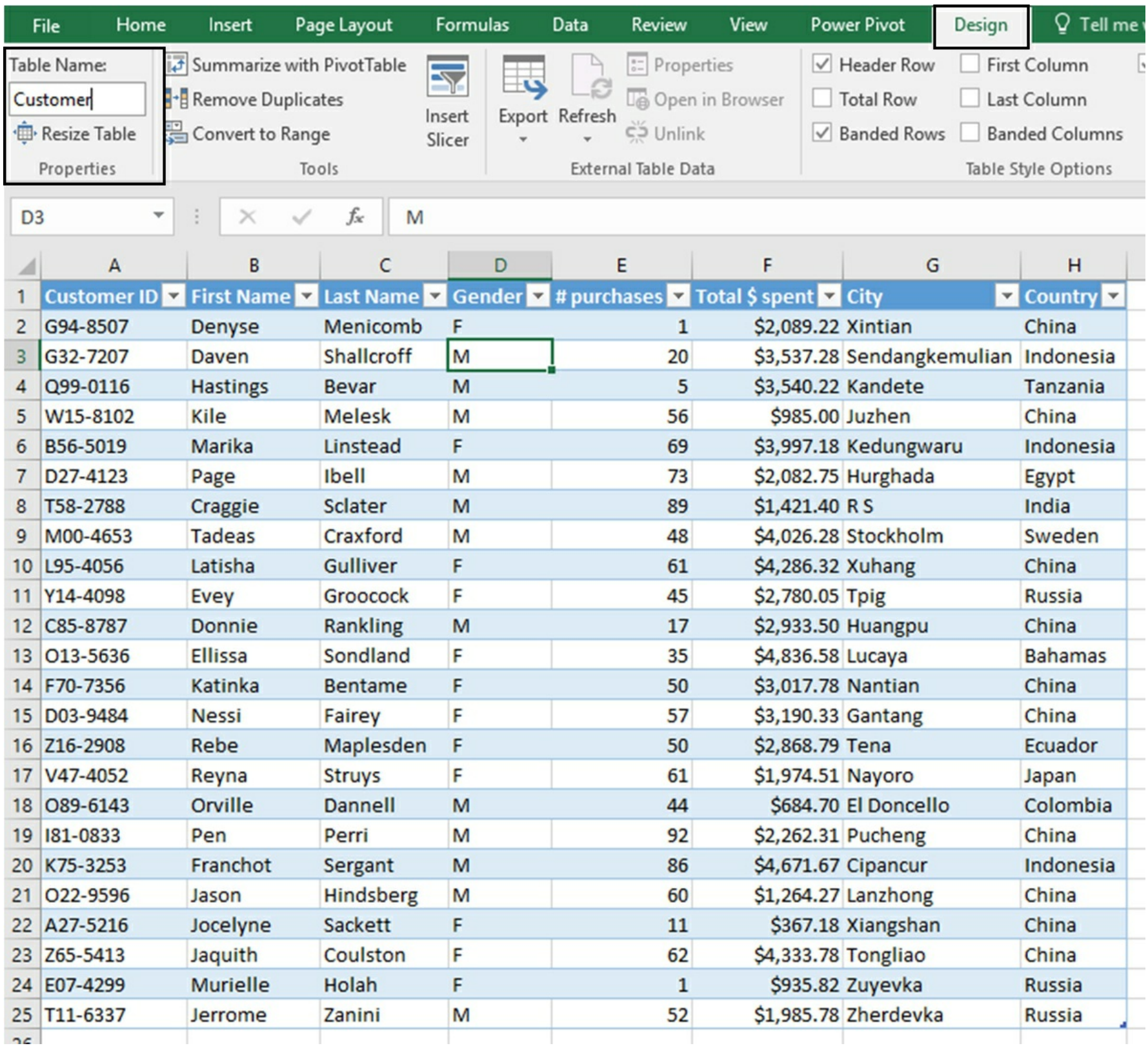
Customer ID	First Name	Last Name	Gender	# purchases	Total $ spent	City	Country
G94-8507	Denyse	Menicomb	F	1	$2,089.22	Xintian	China
G32-7207	Daven	Shallcroff	M	20	$3,537.28	Sendangkemulian	Indonesia
Q99-0116	Hastings	Bevar	M	5	$3,540.22	Kandete	Tanzania
W15-8102	Kile	Melesk	M	56	$985.00	Juzhen	China
B56-5019	Marika	Linstead	F	69	$3,997.18	Kedungwaru	Indonesia
D27-4123	Page	Ibell	M	73	$2,082.75	Hurghada	Egypt
T58-2788	Craggie	Sclater	M	89	$1,421.40	R S	India
M00-4653	Tadeas	Craxford	M	48	$4,026.28	Stockholm	Sweden
L95-4056	Latisha	Gulliver	F	61	$4,286.32	Xuhang	China
Y14-4098	Evey	Groocock	F	45	$2,780.05	Tpig	Russia
C85-8787	Donnie	Rankling	M	17	$2,933.50	Huangpu	China
O13-5636	Ellissa	Sondland	F	35	$4,836.58	Lucaya	Bahamas
F70-7356	Katinka	Bentame	F	50	$3,017.78	Nantian	China
D03-9484	Nessi	Fairey	F	57	$3,190.33	Gantang	China
Z16-2908	Rebe	Maplesden	F	50	$2,868.79	Tena	Ecuador
V47-4052	Reyna	Struys	F	61	$1,974.51	Nayoro	Japan
O89-6143	Orville	Dannell	M	44	$684.70	El Doncello	Colombia
I81-0833	Pen	Perri	M	92	$2,262.31	Pucheng	China
K75-3253	Franchot	Sergant	M	86	$4,671.67	Cipancur	Indonesia
O22-9596	Jason	Hindsberg	M	60	$1,264.27	Lanzhong	China
A27-5216	Jocelyne	Sackett	F	11	$367.18	Xiangshan	China
Z65-5413	Jaquith	Coulston	F	62	$4,333.78	Tongliao	China
E07-4299	Murielle	Holah	F	1	$935.82	Zuyevka	Russia
T11-6337	Jerrome	Zanini	M	52	$1,985.78	Zherdevka	Russia

Naming a table

Excel assigns a name in the Table# format (for example Table1, Table 2 and so on) to the table by default. The tables can be renamed to make them more meaningful and easy to reference. For naming a table, same syntax rules as named ranges apply.

Practice

1. Continue working on the *Creating a Table* sheet of the *Tables_Practice.xlsx*. To look at the name of the table you just created, click on any cell on the table to activate the *Table Tools* tab on the top ribbon.
2. Under the *Table Tools* tab, click on *Design*.
3. In the *Table Name* box under the *Properties* group, rename the table to “Customer” and hit enter.



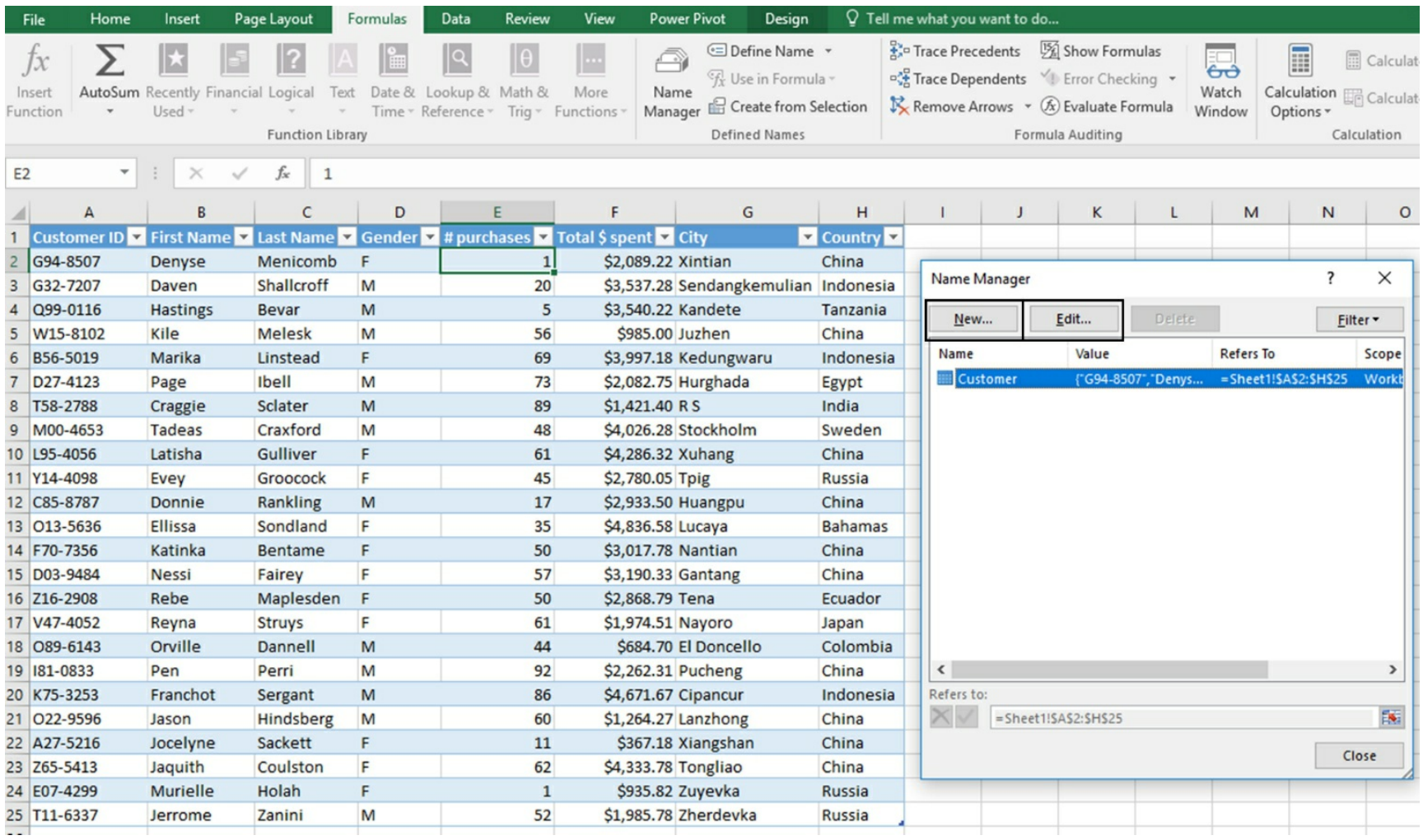
	A	B	C	D	E	F	G	H
1	Customer ID	First Name	Last Name	Gender	# purchases	Total $ spent	City	Country
2	G94-8507	Denyse	Menicomb	F	1	$2,089.22	Xintian	China
3	G32-7207	Daven	Shallcroff	M	20	$3,537.28	Sendangkemulian	Indonesia
4	Q99-0116	Hastings	Bevar	M	5	$3,540.22	Kandete	Tanzania
5	W15-8102	Kile	Melesk	M	56	$985.00	Juzhen	China
6	B56-5019	Marika	Linstead	F	69	$3,997.18	Kedungwaru	Indonesia
7	D27-4123	Page	Ibell	M	73	$2,082.75	Hurghada	Egypt
8	T58-2788	Craggie	Sclater	M	89	$1,421.40	R S	India
9	M00-4653	Tadeas	Craxford	M	48	$4,026.28	Stockholm	Sweden
10	L95-4056	Latisha	Gulliver	F	61	$4,286.32	Xuhang	China
11	Y14-4098	Evey	Groocock	F	45	$2,780.05	Tpig	Russia
12	C85-8787	Donnie	Rankling	M	17	$2,933.50	Huangpu	China
13	O13-5636	Ellissa	Sondland	F	35	$4,836.58	Lucaya	Bahamas
14	F70-7356	Katinka	Bentame	F	50	$3,017.78	Nantian	China
15	D03-9484	Nessi	Fairey	F	57	$3,190.33	Gantang	China
16	Z16-2908	Rebe	Maplesden	F	50	$2,868.79	Tena	Ecuador
17	V47-4052	Reyna	Struys	F	61	$1,974.51	Nayoro	Japan
18	O89-6143	Orville	Dannell	M	44	$684.70	El Doncello	Colombia
19	I81-0833	Pen	Perri	M	92	$2,262.31	Pucheng	China
20	K75-3253	Franchot	Sergant	M	86	$4,671.67	Cipancur	Indonesia
21	O22-9596	Jason	Hindsberg	M	60	$1,264.27	Lanzhong	China
22	A27-5216	Jocelyne	Sackett	F	11	$367.18	Xiangshan	China
23	Z65-5413	Jaquith	Coulston	F	62	$4,333.78	Tongliao	China
24	E07-4299	Murielle	Holah	F	1	$935.82	Zuyevka	Russia
25	T11-6337	Jerrome	Zanini	M	52	$1,985.78	Zherdevka	Russia

Managing a Table

Table names can be managed with the Name Manager similar to named ranges.

Practice

1. Continue working on the *Creating a Table* sheet of the *Tables_Practice.xlsx*. Click anywhere on the *Customer* table.
2. Go to the *Formulas* tab and under *Defined Name* group, click *Name Manager*.
3. The *Name Manager* dialog box pops up with all the tables in the workbook. You can use the Name Manager to create a new table or edit an existing table.



	Customer ID	First Name	Last Name	Gender	# purchases	Total $ spent	City	Country
2	G94-8507	Denyse	Menicomb	F	1	$2,089.22	Xintian	China
3	G32-7207	Daven	Shallcroff	M	20	$3,537.28	Sendangkemulian	Indonesia
4	Q99-0116	Hastings	Bevar	M	5	$3,540.22	Kandete	Tanzania
5	W15-8102	Kile	Melesk	M	56	$985.00	Juzhen	China
6	B56-5019	Marika	Linstead	F	69	$3,997.18	Kedungwaru	Indonesia
7	D27-4123	Page	Ibell	M	73	$2,082.75	Hurghada	Egypt
8	T58-2788	Craggie	Sclater	M	89	$1,421.40	R S	India
9	M00-4653	Tadeas	Craxford	M	48	$4,026.28	Stockholm	Sweden
10	L95-4056	Latisha	Gulliver	F	61	$4,286.32	Xuhang	China
11	Y14-4098	Evey	Groocock	F	45	$2,780.05	Tpig	Russia
12	C85-8787	Donnie	Rankling	M	17	$2,933.50	Huangpu	China
13	O13-5636	Ellissa	Sondland	F	35	$4,836.58	Lucaya	Bahamas
14	F70-7356	Katinka	Bentame	F	50	$3,017.78	Nantian	China
15	D03-9484	Nessi	Fairey	F	57	$3,190.33	Gantang	China
16	Z16-2908	Rebe	Maplesden	F	50	$2,868.79	Tena	Ecuador
17	V47-4052	Reyna	Struys	F	61	$1,974.51	Nayoro	Japan
18	O89-6143	Orville	Dannell	M	44	$684.70	El Doncello	Colombia
19	I81-0833	Pen	Perri	M	92	$2,262.31	Pucheng	China
20	K75-3253	Franchot	Sergant	M	86	$4,671.67	Cipancur	Indonesia
21	O22-9596	Jason	Hindsberg	M	60	$1,264.27	Lanzhong	China
22	A27-5216	Jocelyne	Sackett	F	11	$367.18	Xiangshan	China
23	Z65-5413	Jaquith	Coulston	F	62	$4,333.78	Tongliao	China
24	E07-4299	Murielle	Holah	F	1	$935.82	Zuyevka	Russia
25	T11-6337	Jerrome	Zanini	M	52	$1,985.78	Zherdevka	Russia

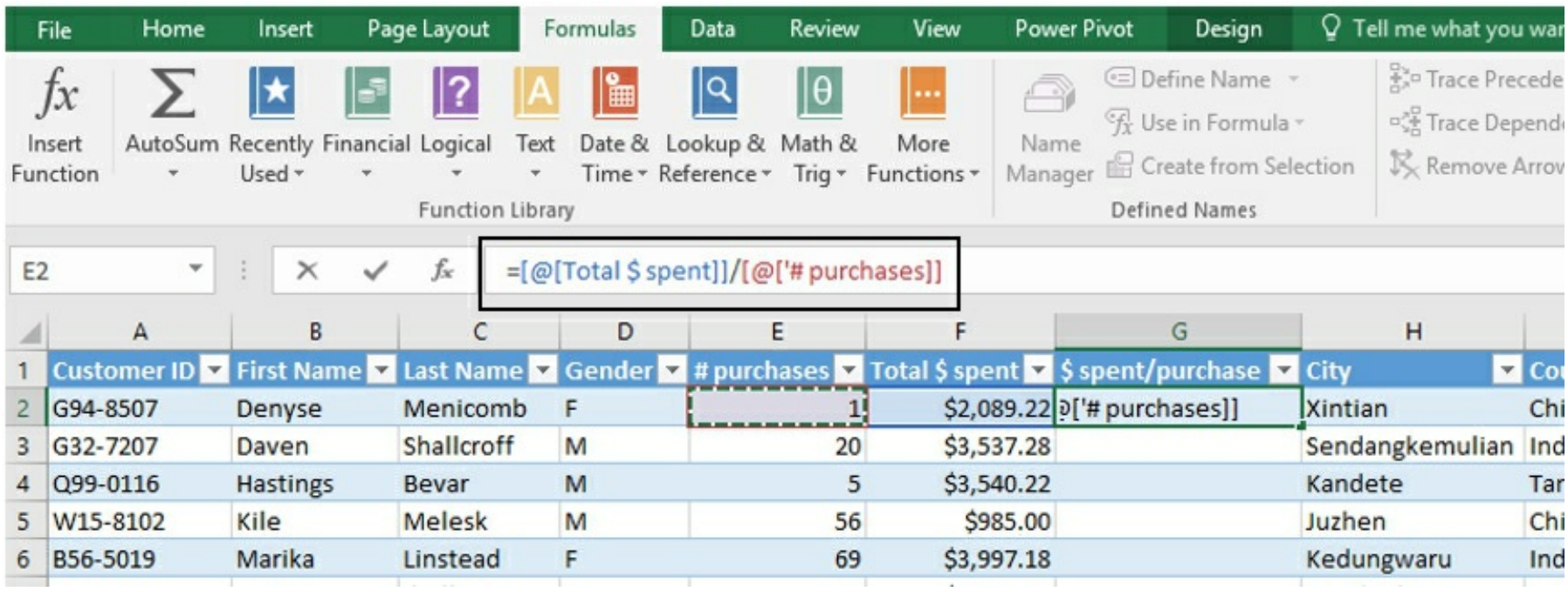
Exploring important properties of a table

1. Calculated columns

In an Excel table, a formula is entered in one cell of the column, and then that formula gets automatically populated to the rest of the column by itself. This saves time, especially if you have a lot of rows. When you change a formula, the change will also auto-populate to the rest of the calculated column.

Practice

1. Continue working on the *Creating a Table* sheet of the *Tables_Practice.xlsx*. To calculate $spent/purchase in the *Customer* table, insert a new column to the right of the *Total $ spent* column. Name the column *$spent/purchase*.
2. Click on cell G2 in the column. Type the equal sign (=). Then click on cell F2, type the division sign (/) and click on cell E2. In the formula bar, you will notice **structured table reference**s which we will discuss in a bit. Press Enter.



	Customer ID	First Name	Last Name	Gender	# purchases	Total $ spent	$ spent/purchase	City
2	G94-8507	Denyse	Menicomb	F	1	$2,089.22	['# purchases]]	Xintian
3	G32-7207	Daven	Shallcroff	M	20	$3,537.28		Sendangkemulian
4	Q99-0116	Hastings	Bevar	M	5	$3,540.22		Kandete
5	W15-8102	Kile	Melesk	M	56	$985.00		Juzhen
6	B56-5019	Marika	Linstead	F	69	$3,997.18		Kedungwaru

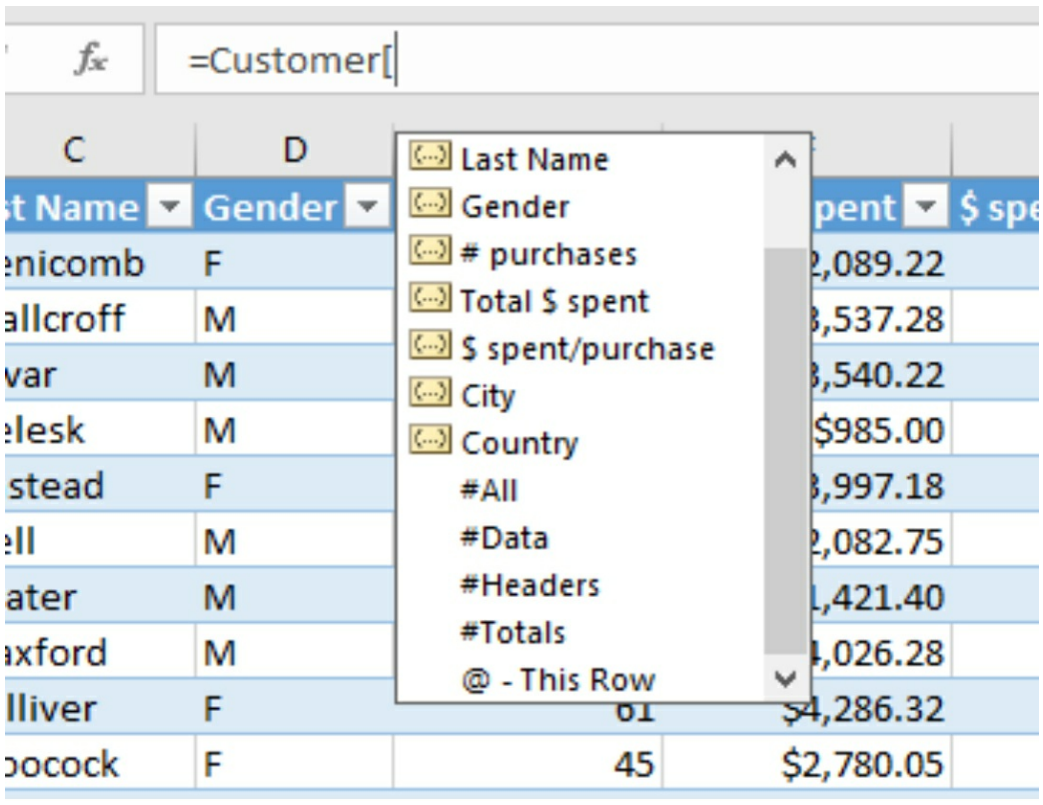
3. The formula will propagate automatically to the other cells in the column.

G3 =[@[Total $ spent]]/[@['# purchases]]

	A	B	C	D	E	F	G	H	I
1	Customer ID	First Name	Last Name	Gender	# purchases	Total $ spent	$ spent/purchase	City	Country
2	G94-8507	Denyse	Menicomb	F	1	$2,089.22	$2,089.22	Xintian	China
3	G32-7207	Daven	Shallcroff	M	20	$3,537.28	$176.86	Sendangkemulian	Indonesia
4	Q99-0116	Hastings	Bevar	M	5	$3,540.22	$708.04	Kandete	Tanzania
5	W15-8102	Kile	Melesk	M	56	$985.00	$17.59	Juzhen	China
6	B56-5019	Marika	Linstead	F	69	$3,997.18	$57.93	Kedungwaru	Indonesia
7	D27-4123	Page	Ibell	M	73	$2,082.75	$28.53	Hurghada	Egypt
8	T58-2788	Craggie	Sclater	M	89	$1,421.40	$15.97	R S	India
9	M00-4653	Tadeas	Craxford	M	48	$4,026.28	$83.88	Stockholm	Sweden

2. Structured Table References

In the above example, while entering the formula to calculate $spent/purchase, instead of explicit cell references as =F2/E2, the formula appeared as =[@[Total $ spent]]/[@[# purchases]]. These combinations of table and column names are called structured table references. When an Excel table is created and assigned a name, Excel also assigns a name to each of the column headers in the table. A structured table reference begins with the table's name followed by reference specifiers enclosed in square brackets []. To get a list of the available reference specifiers, enter the table's name in a formula and then enter the open square bracket. Excel will display a list of valid specifiers in the auto-complete list in addition to all column headers.

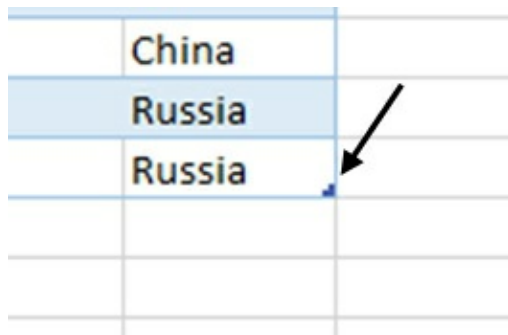


Any of the following reference specifiers can be used in structured references:

Reference specifier:	Refers to:	Example
#All	The entire column, including column headers, data, and totals (if any).	=Customer[[#All],[Total $ spent]]; all cells in the Total $ spent column, including the column header.
#Data	The data rows.	= Customer[[#Data],[Total $ spent]]; all the cells in the Total $ spent column, excluding column header.
#Headers	The header row.	=Customer[#Headers]; headers of the Customer table.
#Totals	The total row. If none exists, it returns null.	
@ or #This Row or @[Column Name]	The cells in the same row as the formula.	=[@[Total $ spent]]; cell corresponding to Total $ spent column in the same row as formula

3. Sizing Handle

You can resize a table to add or remove rows/columns by dragging the blue-colored sizing handle on the bottom right of the table, as needed. The table will be resized automatically.

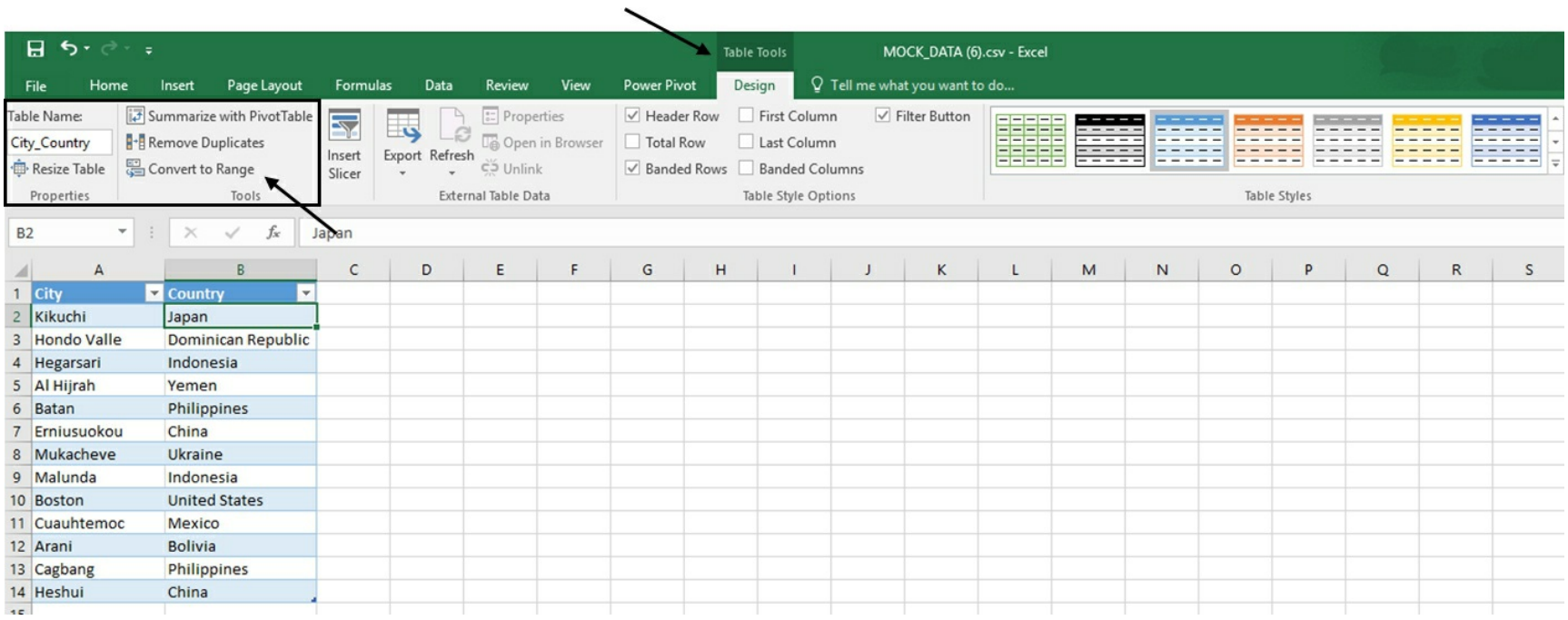


4. Convert to Range

Sometimes the table format is not the best for what you need to do. A table can be easily converted to normal range without losing the table style.

Practice

1. Open *Tables_Practice.xlsx* and go to *Convert to Range* sheet. There is a table named *City_Country* on the sheet. The table has 13 records. Click anywhere on the table to view *Table Tools Design* tab on the top ribbon.
2. On the *Table Tools Design* tab, under *Tools* group, click *Convert to Range*.
3. Confirm the *Do you want to convert the table to a normal range?* pop up with a *Yes*.
4. The table will now be converted to a normal cell range (not a named range).



5. Column Headers

While scrolling down a large table with many rows, as the column headers scroll off the top of the table, worksheet columns names are automatically replaced with column headers. Hence the column headers stay visible without needing to freeze the header row.

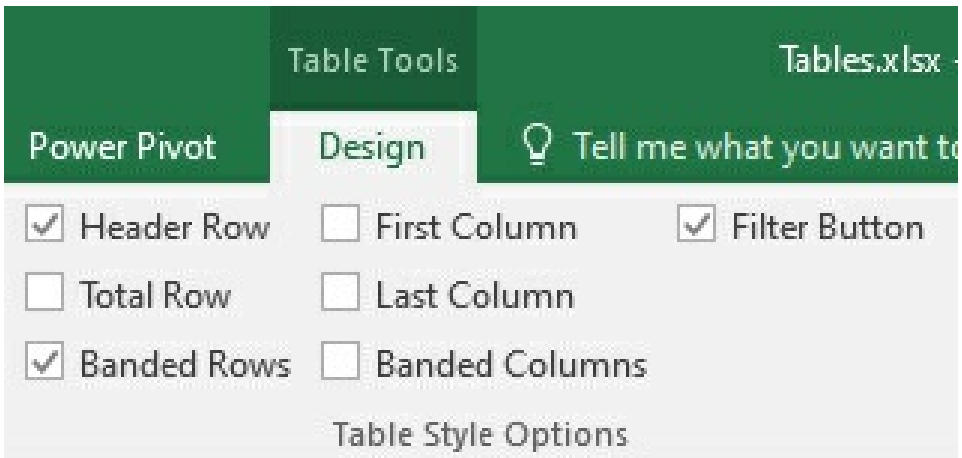
	Customer ID	First Name	Last Name	Gender	# purchases	Total $ spent	$ spent/purchase	City	Country
22	A27-5216	Jocelyne	Sackett	F	11	$367.18	$33.38	Xiangshan	China
23	Z65-5413	Jaquith	Coulston	F	62	$4,333.78	$69.90	D	China
24	E07-4299	Murielle	Holah	F	1	$935.82	$935.82	Zuyevka	Russia
25	T11-6337	Jerrome	Zanini	M	52	$1,985.78	$38.19	Zherdevka	Russia

6. Tables expand or contract automatically

When new rows or columns are appended to an Excel Table, the table automatically expands to include them. Similarly, a table automatically contracts when rows or columns are deleted.

7. Table totals

You can add a total row that allows you to display summary calculations in the last row of your table. To enable the total row, go to the *Table Tools Design* tab and check the *Total Row* box under *Table Style Options* group.



8. Table Styles

Changing the style of a table is easy and there are several styles to choose from. Just go to the *Table Tools Design* tab and select a new style from *Table Styles* group. More styles are available by clicking the down arrows.

1.4 Conditional formatting

Conditional formatting is an excellent tool to visualize data of interest in a spreadsheet. With conditional formatting, you can specify the formatting of one or more cells based on conditions or rules. The conditions can be based on the contents of the selected cell and other cell(s) within the workbook. The cells that meet the defined conditions would be formatted as specified.

Conditional formatting options are available through the *Home* tab on the top ribbon. We will discuss them one by one.

Highlight Cells Rules

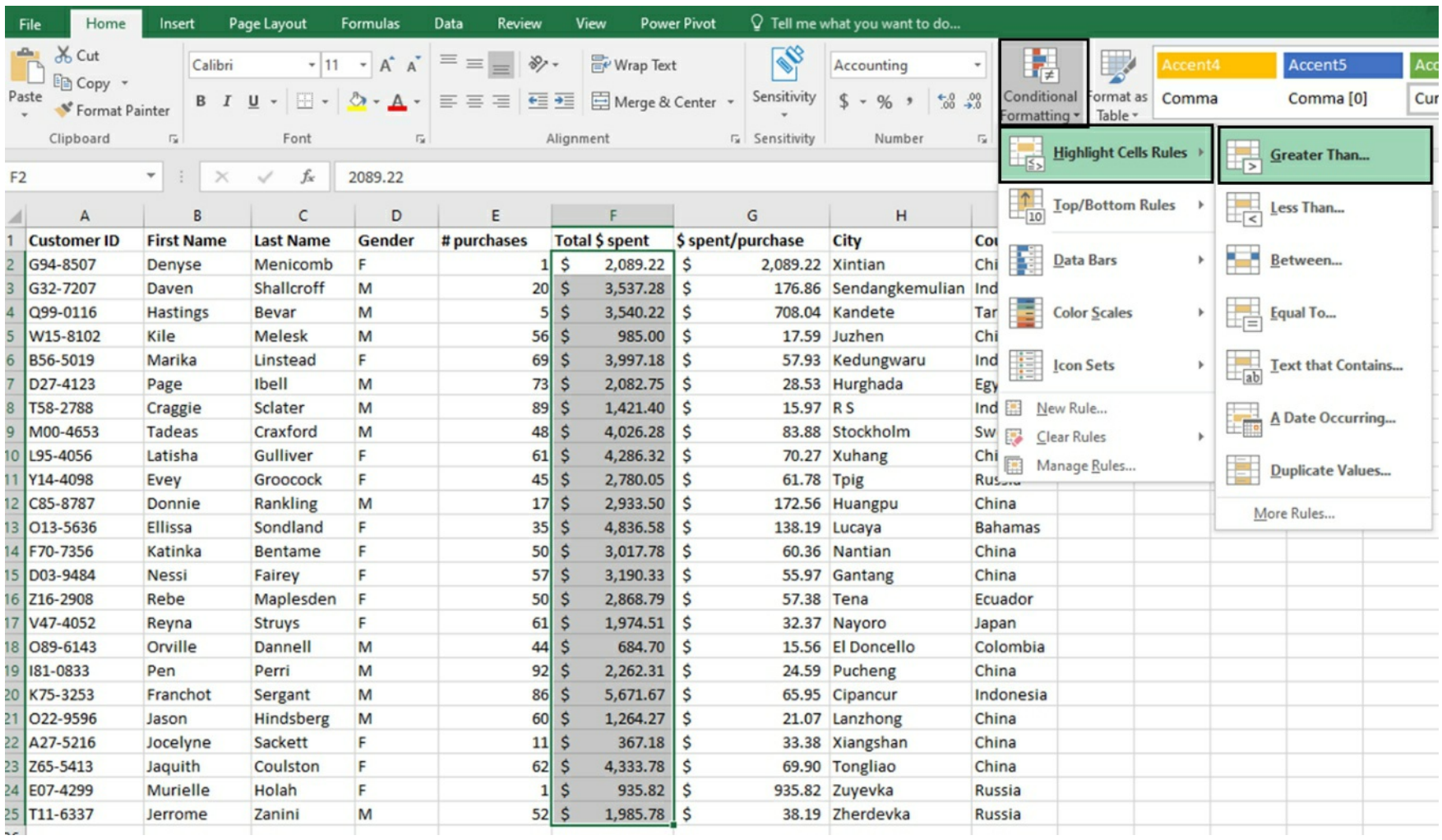
Highlight Cells Rules can be used to format cells that meet any of the following criteria −

1. Cells contain numbers within given numerical values (Greater Than, Less Than, Between, Equal To).
2. Cells contain text that has a given text string.
3. Cells contain dates relative to the current date (Yesterday, Today, Tomorrow, in the last 7 days, Last week, This week, Next week, Last month, This Month, Next Month).
4. Cells contain duplicate values.

Practice

Open *Conditional Formatting_Practice*.xlsx. The workbook contains a sheet named *Conditional Formatting* with 24 records of customer data. Follow the below steps to format cells in the Total $ spent column with the $ value (1) greater than 5000, and (2) less than 500.

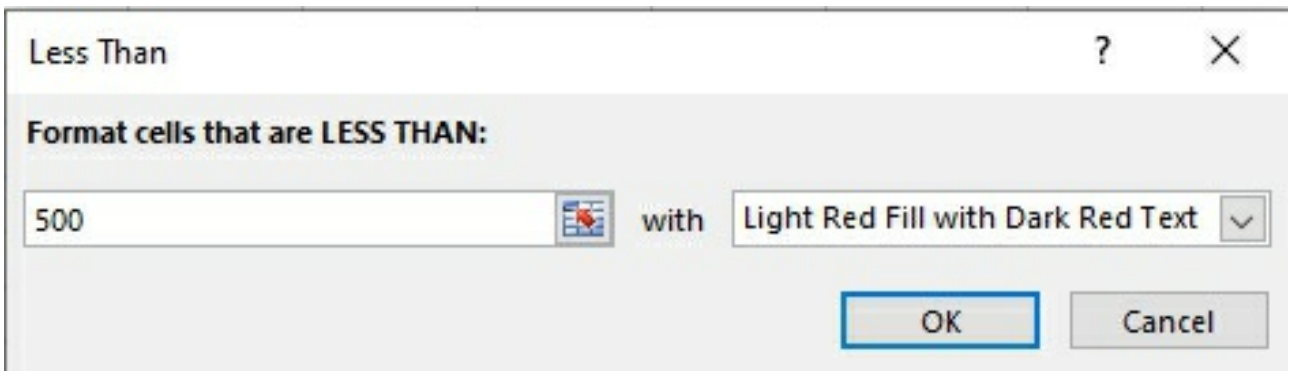
1. Select data values under *Total $ spent* column (F2:F25) on the *Conditional Formatting* sheet.
2. Go to *Home* tab. Under the *Styles* group, select *Conditional Formatting* and click on *Highlight Cell Rules* from the dropdown.



	A	B	C	D	E	F	G	H
1	Customer ID	First Name	Last Name	Gender	# purchases	Total $ spent	$ spent/purchase	City
2	G94-8507	Denyse	Menicomb	F	1	$ 2,089.22	$ 2,089.22	Xintian
3	G32-7207	Daven	Shallcroff	M	20	$ 3,537.28	$ 176.86	Sendangkemulian
4	Q99-0116	Hastings	Bevar	M	5	$ 3,540.22	$ 708.04	Kandete
5	W15-8102	Kile	Melesk	M	56	$ 985.00	$ 17.59	Juzhen
6	B56-5019	Marika	Linstead	F	69	$ 3,997.18	$ 57.93	Kedungwaru
7	D27-4123	Page	Ibell	M	73	$ 2,082.75	$ 28.53	Hurghada
8	T58-2788	Craggie	Sclater	M	89	$ 1,421.40	$ 15.97	R S
9	M00-4653	Tadeas	Craxford	M	48	$ 4,026.28	$ 83.88	Stockholm
10	L95-4056	Latisha	Gulliver	F	61	$ 4,286.32	$ 70.27	Xuhang
11	Y14-4098	Evey	Groocock	F	45	$ 2,780.05	$ 61.78	Tpig
12	C85-8787	Donnie	Rankling	M	17	$ 2,933.50	$ 172.56	Huangpu
13	O13-5636	Ellissa	Sondland	F	35	$ 4,836.58	$ 138.19	Lucaya
14	F70-7356	Katinka	Bentame	F	50	$ 3,017.78	$ 60.36	Nantian
15	D03-9484	Nessi	Fairey	F	57	$ 3,190.33	$ 55.97	Gantang
16	Z16-2908	Rebe	Maplesden	F	50	$ 2,868.79	$ 57.38	Tena
17	V47-4052	Reyna	Struys	F	61	$ 1,974.51	$ 32.37	Nayoro
18	O89-6143	Orville	Dannell	M	44	$ 684.70	$ 15.56	El Doncello
19	I81-0833	Pen	Perri	M	92	$ 2,262.31	$ 24.59	Pucheng
20	K75-3253	Franchot	Sergant	M	86	$ 5,671.67	$ 65.95	Cipancur
21	O22-9596	Jason	Hindsberg	M	60	$ 1,264.27	$ 21.07	Lanzhong
22	A27-5216	Jocelyne	Sackett	F	11	$ 367.18	$ 33.38	Xiangshan
23	Z65-5413	Jaquith	Coulston	F	62	$ 4,333.78	$ 69.90	Tongliao
24	E07-4299	Murielle	Holah	F	1	$ 935.82	$ 935.82	Zuyevka
25	T11-6337	Jerrome	Zanini	M	52	$ 1,985.78	$ 38.19	Zherdevka

3. Select *Greater Than*. A *Greater Than* box will pop up. Enter *5000* as a value and select *Green Fill with Dark Green Text* from the dropdown.

4. Repeat with clicking on *Less Than* this time by entering 500 and selecting *Light Red Fill with Dark Red Text* from the dropdown.



5. The data will be highlighted based on the given conditions and the corresponding formatting.

Total $ spent
$ 2,089.22
$ 3,537.28
$ 3,540.22
$ 985.00
$ 3,997.18
$ 2,082.75
$ 1,421.40
$ 4,026.28
$ 4,286.32
$ 2,780.05
$ 2,933.50
$ 4,836.58
$ 3,017.78
$ 3,190.33
$ 2,868.79
$ 1,974.51
$ 684.70
$ 2,262.31
$ 5,671.67
$ 1,264.27
$ 367.18
$ 4,333.78
$ 935.82
$ 1,985.78

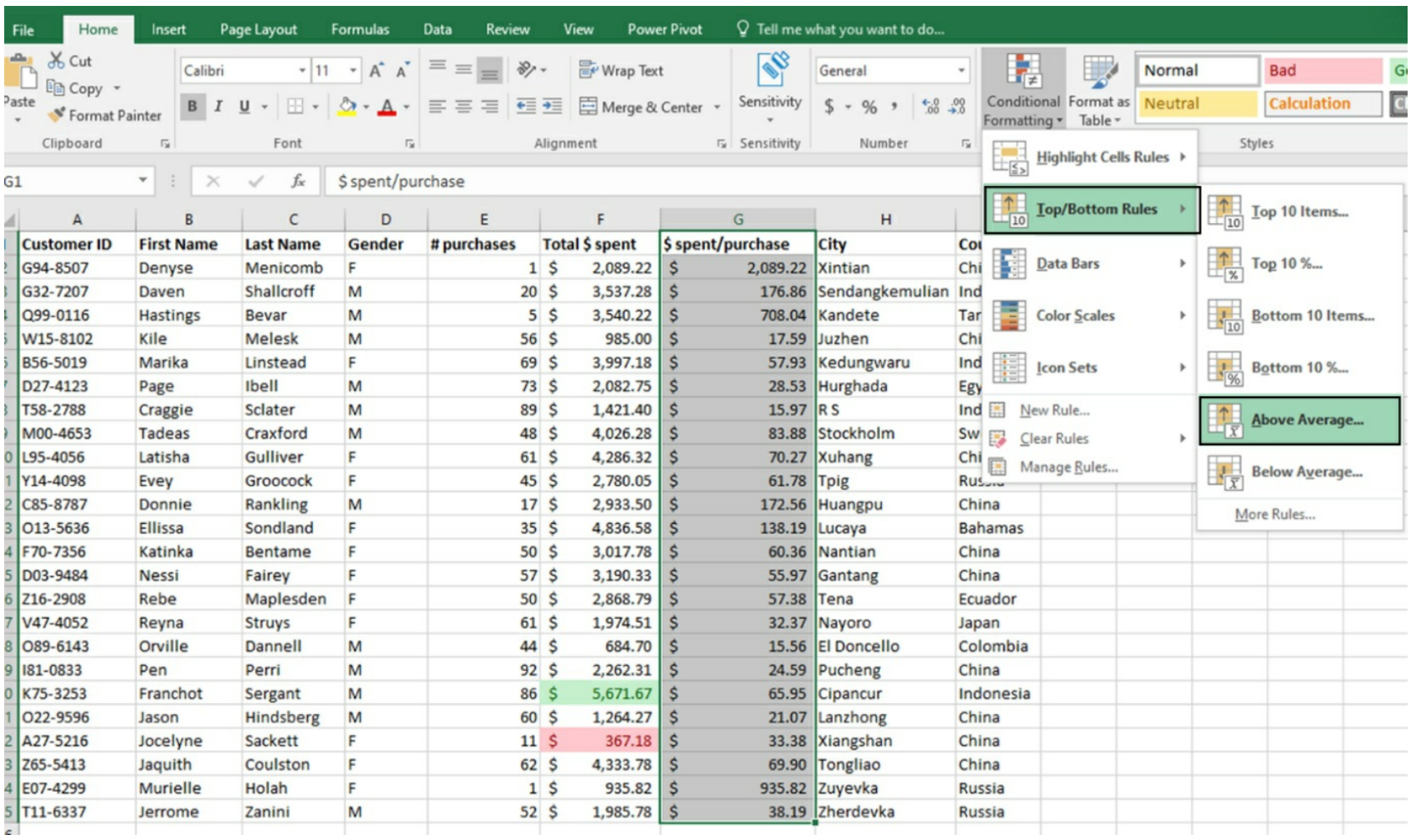
Top / Bottom Rules

Top/Bottom Rules can be used to format cells that contain numbers representing Top 10 items, Top 10%, Bottom 10 items, Bottom 10%, Above Average or Below Average.

Practice

Continue working on the *Conditional Formatting* sheet and follow the below steps to format cells in the *$ spent/purchase* column with the $ value (1) above average, and (2) below average.

1. Select data values under the *$ spent/purchase* column (G2:G25) on the sheet.
2. Go to *Home* tab. Under the *Styles* group, select *Conditional Formatting* and click on *Top/Bottom Rules* from the dropdown.



3. Select *Above Average*. An *Above Average* box will pop up. Select *Green Fill with Dark Green Text* from the dropdown.

4. Repeat with clicking on *Below Average* this time by entering *Red Text* from the dropdown.

5. The data will be highlighted based on the given conditions and the corresponding formatting.

$ spent/purchase	
$	2,089.22
$	176.86
$	708.04
$	17.59
$	57.93
$	28.53
$	15.97
$	83.88
$	70.27
$	61.78
$	172.56
$	138.19
$	60.36
$	55.97
$	57.38
$	32.37
$	15.56
$	24.59
$	65.95
$	21.07
$	33.38
$	69.90
$	935.82
$	38.19

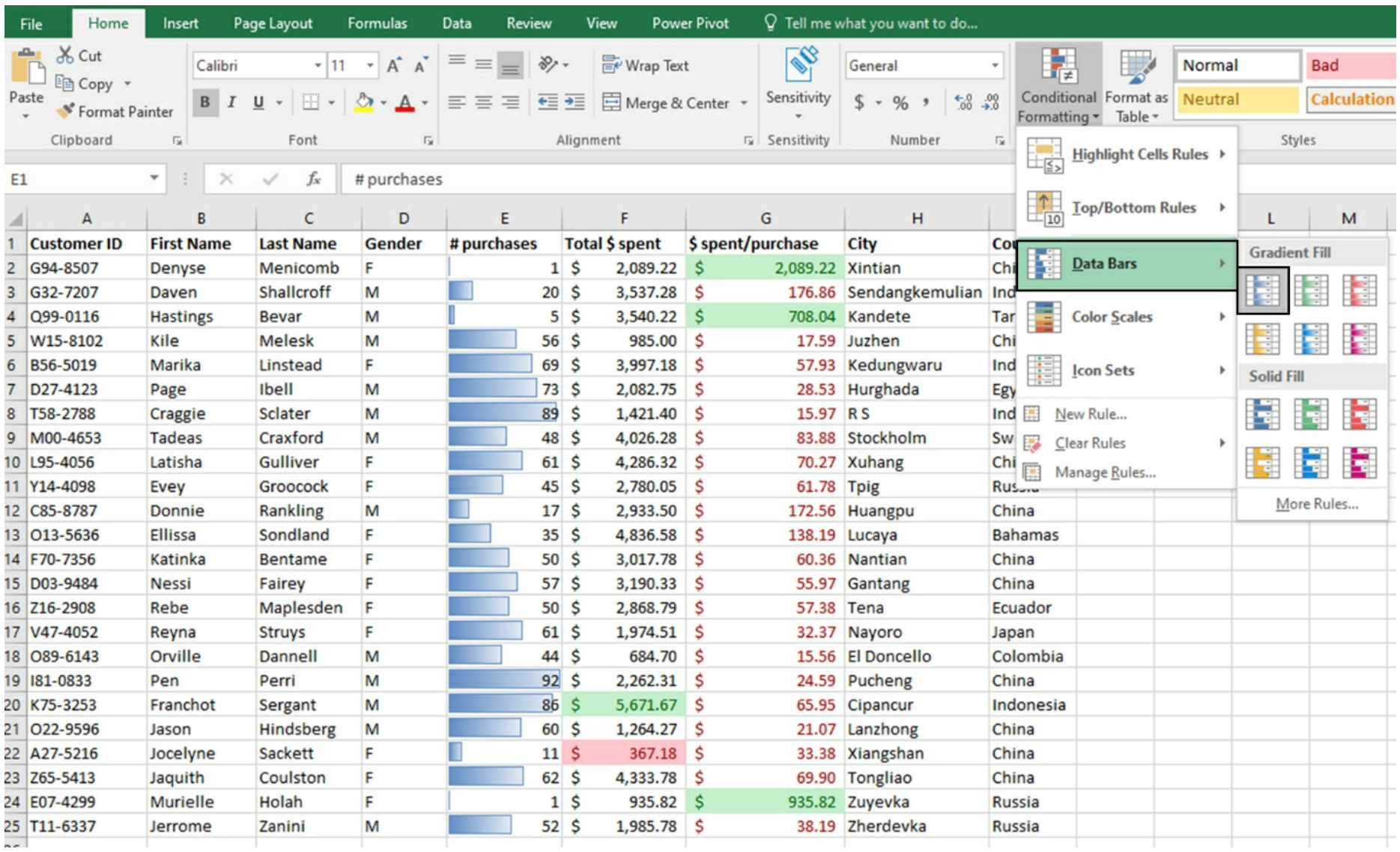
Data Bars

Data bars help to visualize the value of a cell comparative to other cells. The length of the bar represents the magnitude of the cell value. A longer bar represents a higher value, and a shorter bar represents a lower value.

Practice

Continue working on the *Conditional Formatting* sheet and follow the below steps to visualize data bars in the *# purchases* column.

1. Select data values under the *# purchases* column (E2:E25) on the *Conditional Formatting* sheet.
2. Go to *Home* tab. Under the *Styles* group, select *Conditional Formatting* and click on *Data Bars* from the dropdown.
3. Under the *Gradient Fill* group, select the first color style (Blue Data Bar). The data bars will appear.



	A	B	C	D	E	F	G	H
1	Customer ID	First Name	Last Name	Gender	# purchases	Total $ spent	$ spent/purchase	City
2	G94-8507	Denyse	Menicomb	F	1	$ 2,089.22	$ 2,089.22	Xintian
3	G32-7207	Daven	Shallcroff	M	20	$ 3,537.28	$ 176.86	Sendangkemulian
4	Q99-0116	Hastings	Bevar	M	5	$ 3,540.22	$ 708.04	Kandete
5	W15-8102	Kile	Melesk	M	56	$ 985.00	$ 17.59	Juzhen
6	B56-5019	Marika	Linstead	F	69	$ 3,997.18	$ 57.93	Kedungwaru
7	D27-4123	Page	Ibell	M	73	$ 2,082.75	$ 28.53	Hurghada
8	T58-2788	Craggie	Sclater	M	89	$ 1,421.40	$ 15.97	R S
9	M00-4653	Tadeas	Craxford	M	48	$ 4,026.28	$ 83.88	Stockholm
10	L95-4056	Latisha	Gulliver	F	61	$ 4,286.32	$ 70.27	Xuhang
11	Y14-4098	Evey	Groocock	F	45	$ 2,780.05	$ 61.78	Tpig
12	C85-8787	Donnie	Rankling	M	17	$ 2,933.50	$ 172.56	Huangpu
13	O13-5636	Ellissa	Sondland	F	35	$ 4,836.58	$ 138.19	Lucaya
14	F70-7356	Katinka	Bentame	F	50	$ 3,017.78	$ 60.36	Nantian
15	D03-9484	Nessi	Fairey	F	57	$ 3,190.33	$ 55.97	Gantang
16	Z16-2908	Rebe	Maplesden	F	50	$ 2,868.79	$ 57.38	Tena
17	V47-4052	Reyna	Struys	F	61	$ 1,974.51	$ 32.37	Nayoro
18	O89-6143	Orville	Dannell	M	44	$ 684.70	$ 15.56	El Doncello
19	I81-0833	Pen	Perri	M	92	$ 2,262.31	$ 24.59	Pucheng
20	K75-3253	Franchot	Sergant	M	86	$ 5,671.67	$ 65.95	Cipancur
21	O22-9596	Jason	Hindsberg	M	60	$ 1,264.27	$ 21.07	Lanzhong
22	A27-5216	Jocelyne	Sackett	F	11	$ 367.18	$ 33.38	Xiangshan
23	Z65-5413	Jaquith	Coulston	F	62	$ 4,333.78	$ 69.90	Tongliao
24	E07-4299	Murielle	Holah	F	1	$ 935.82	$ 935.82	Zuyevka
25	T11-6337	Jerrome	Zanini	M	52	$ 1,985.78	$ 38.19	Zherdevka

4. To remove the data bars, select range E2:E25. Under the *Styles* group on the *Home* tab, select *Conditional Formatting*, click on *Clear Rules* and select *Clear rules from Selected Cells* from the dropdown.

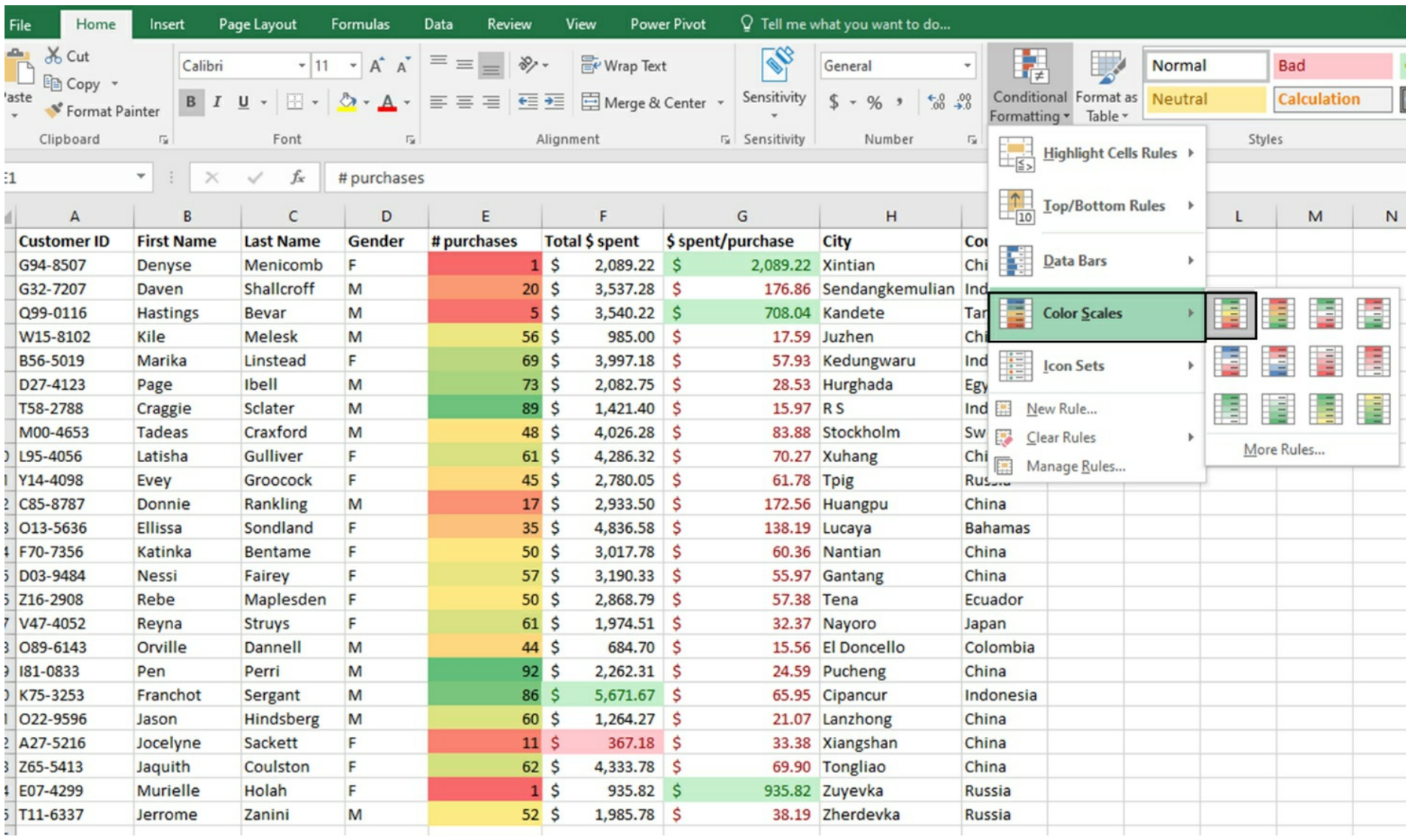
Color Scales

Color scales are another tool to help to visualize the value of a cell comparative to other cells. A two-color scale helps to compare a range of cells by using a shade of two colors. The shade of the color represents higher or lower values. Similarly, a three-color scale helps you compare a range of cells by using a shade of three colors. The shade of the color represents higher, middle, or lower values. For example, in a green, yellow, and red color scale, higher value cells have a green color, middle value cells have a yellow color, and lower value cells have a red color.

Practice

Continue working on the *Conditional Formatting* sheet and follow the below steps to visualize color scales in the *# purchases* column.

1. Select data values under the *# purchases* column (E2:E25) on the sheet.
2. Go to *Home* tab. Under the *Styles* group, select *Conditional Formatting* and click on *Color Scales* from the dropdown.
3. Select the first color scale style (Green - Yellow - Red Color Scale).



Customer ID	First Name	Last Name	Gender	# purchases	Total $ spent	$ spent/purchase	City	Country
G94-8507	Denyse	Menicomb	F	1	$ 2,089.22	$ 2,089.22	Xintian	Chi
G32-7207	Daven	Shallcroff	M	20	$ 3,537.28	$ 176.86	Sendangkemulian	Ind
Q99-0116	Hastings	Bevar	M	5	$ 3,540.22	$ 708.04	Kandete	Tar
W15-8102	Kile	Melesk	M	56	$ 985.00	$ 17.59	Juzhen	Chi
B56-5019	Marika	Linstead	F	69	$ 3,997.18	$ 57.93	Kedungwaru	Ind
D27-4123	Page	Ibell	M	73	$ 2,082.75	$ 28.53	Hurghada	Egy
T58-2788	Craggie	Sclater	M	89	$ 1,421.40	$ 15.97	R S	Ind
M00-4653	Tadeas	Craxford	M	48	$ 4,026.28	$ 83.88	Stockholm	Sw
L95-4056	Latisha	Gulliver	F	61	$ 4,286.32	$ 70.27	Xuhang	Chi
Y14-4098	Evey	Groocock	F	45	$ 2,780.05	$ 61.78	Tpig	Rus
C85-8787	Donnie	Rankling	M	17	$ 2,933.50	$ 172.56	Huangpu	China
O13-5636	Ellissa	Sondland	F	35	$ 4,836.58	$ 138.19	Lucaya	Bahamas
F70-7356	Katinka	Bentame	F	50	$ 3,017.78	$ 60.36	Nantian	China
D03-9484	Nessi	Fairey	F	57	$ 3,190.33	$ 55.97	Gantang	China
Z16-2908	Rebe	Maplesden	F	50	$ 2,868.79	$ 57.38	Tena	Ecuador
V47-4052	Reyna	Struys	F	61	$ 1,974.51	$ 32.37	Nayoro	Japan
O89-6143	Orville	Dannell	M	44	$ 684.70	$ 15.56	El Doncello	Colombia
I81-0833	Pen	Perri	M	92	$ 2,262.31	$ 24.59	Pucheng	China
K75-3253	Franchot	Sergant	M	86	$ 5,671.67	$ 65.95	Cipancur	Indonesia
O22-9596	Jason	Hindsberg	M	60	$ 1,264.27	$ 21.07	Lanzhong	China
A27-5216	Jocelyne	Sackett	F	11	$ 367.18	$ 33.38	Xiangshan	China
Z65-5413	Jaquith	Coulston	F	62	$ 4,333.78	$ 69.90	Tongliao	China
E07-4299	Murielle	Holah	F	1	$ 935.82	$ 935.82	Zuyevka	Russia
T11-6337	Jerrome	Zanini	M	52	$ 1,985.78	$ 38.19	Zherdevka	Russia

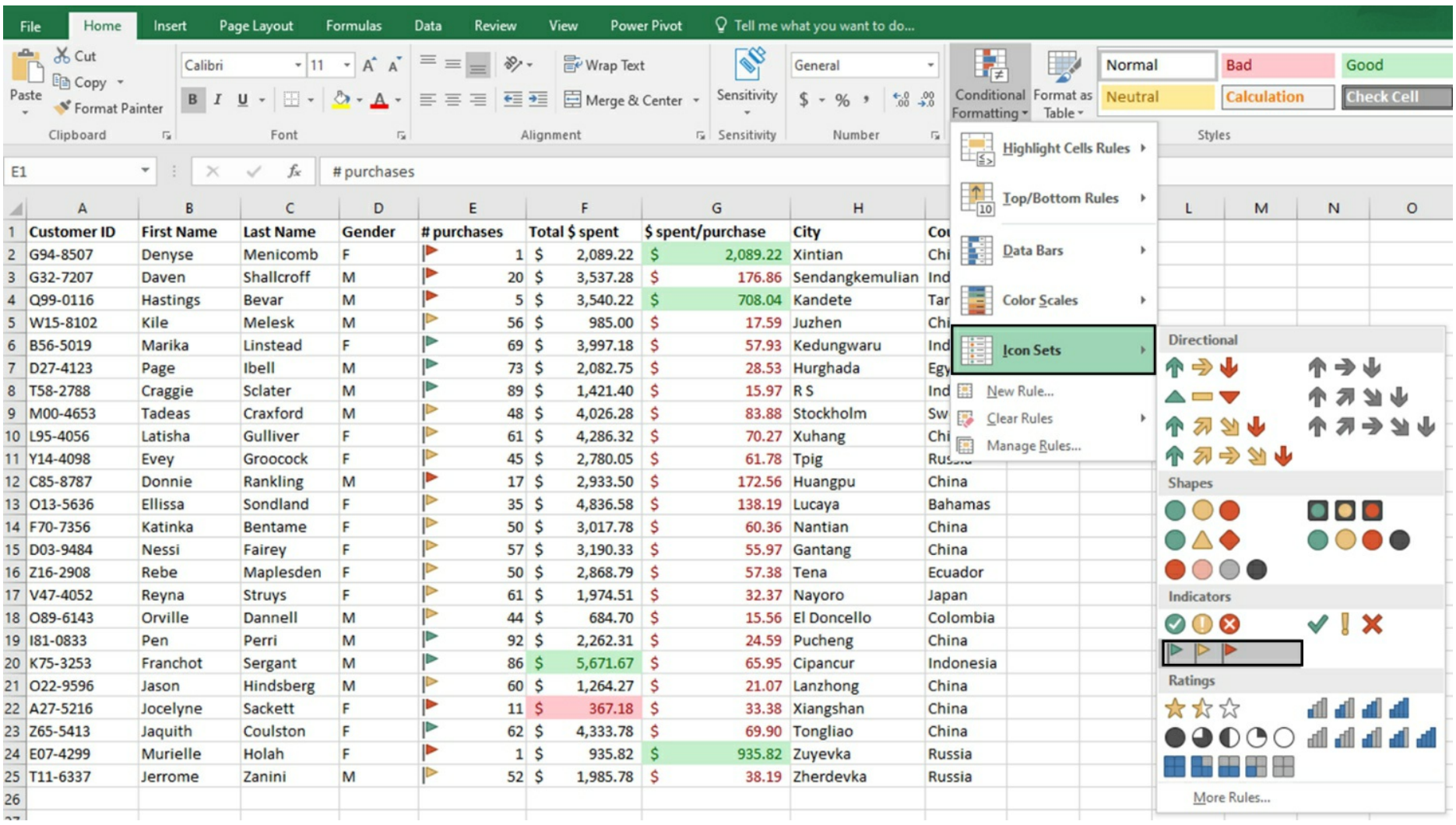
Icon Sets

Icon sets can be used to group data into three to five categories based on the values. Each icon represents a range of values. For example, in the 3 Arrows icon set, the green up arrow represents higher values, the yellow sideways arrow represents middle values, and the red down arrow represents lower values.

Practice

Continue working on the *Conditional Formatting* sheet and follow the below steps to insert icon sets in the *# purchases* column.

1. Select data values under the # purchases column (E2:E25) on the sheet.
2. Go to *Home* tab. Under the *Styles* group, select *Conditional Formatting* and click on *Icon Sets* from the dropdown.
3. Under the *Directional* group, select the first color scale style (3 Arrows (Colored)).



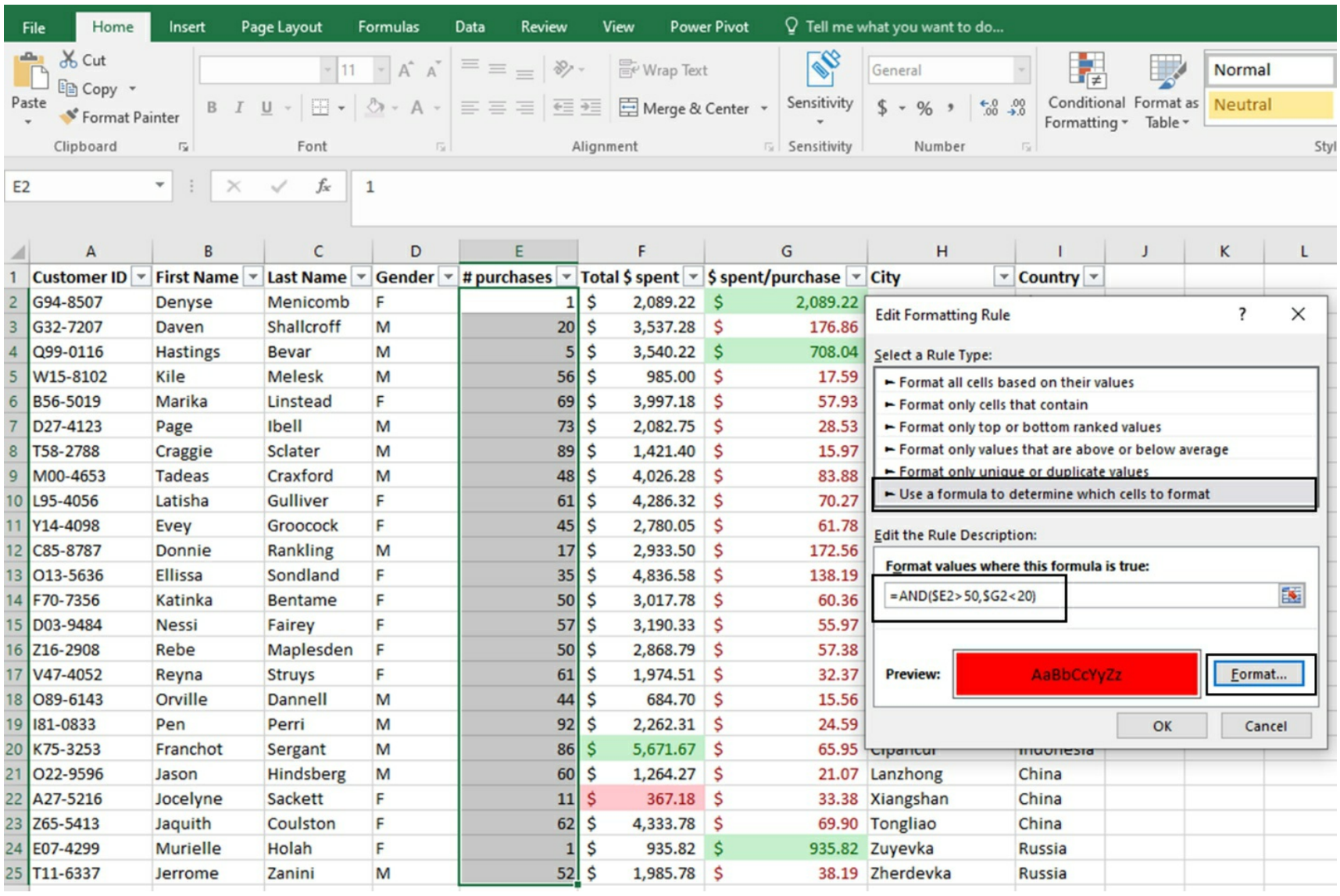
New Rule - Formatting based on User-defined functions

The New Rule option helps to set a user-defined formula as a condition to format a cell.

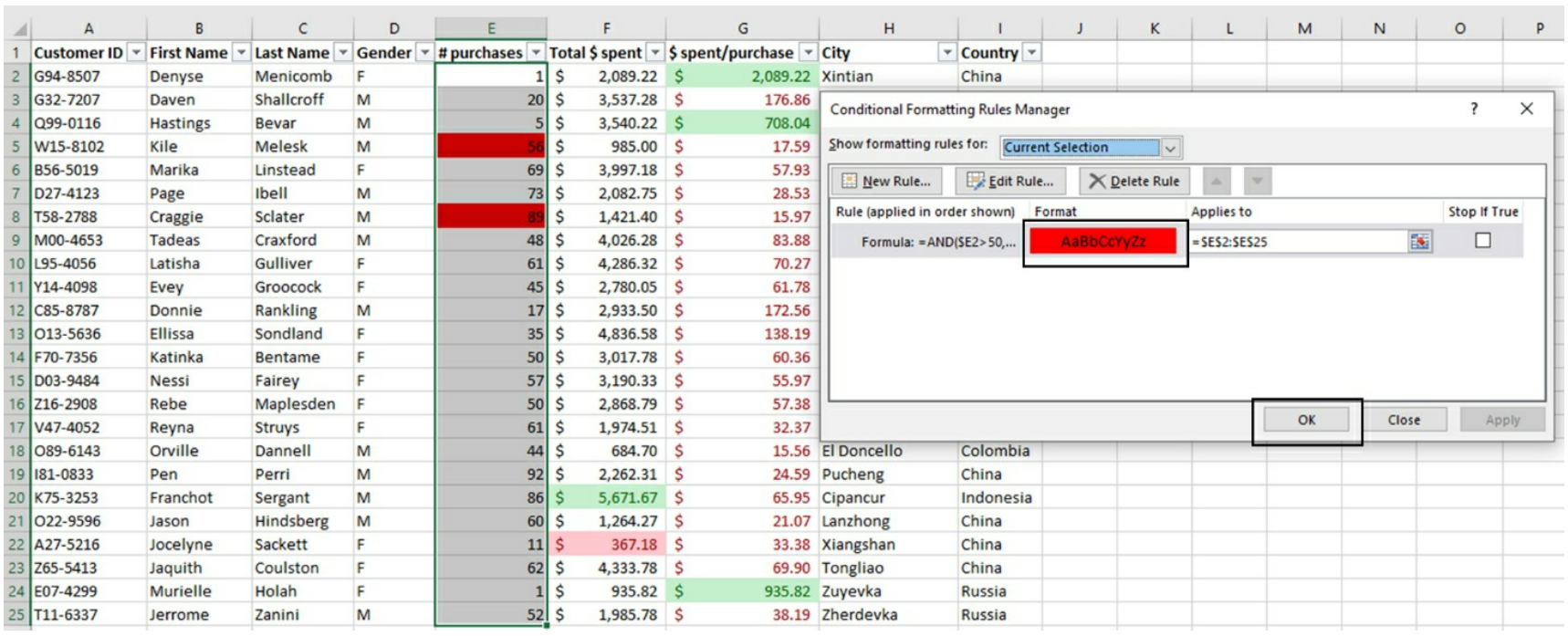
Practice

Continue working on the *Conditional Formatting* sheet and follow the below steps to change the background of cells in the *# purchases* column where >50 purchases were made while spending <20 $ per purchase.

1. Select data values under the *# purchases* column (E2:E25) on the sheet.
2. Go to *Home* tab. Under the *Styles* group, select *Conditional Formatting* and click on *New Rule* from the dropdown. The *New Formatting Rule* dialog box pops up.
3. For *Select a Rule Type*, select *Use a formula to determine which cells to format*. *Edit the Rule Description* box appears.
4. In the format values where this formula is true box, enter the formula *=AND($E2>50,$G2<20)*. The dollar sign ($) locks the reference to columns E and G, and the AND ensures both conditions are TRUE. In rows where the AND function returns TRUE, the conditional formatting will be applied.

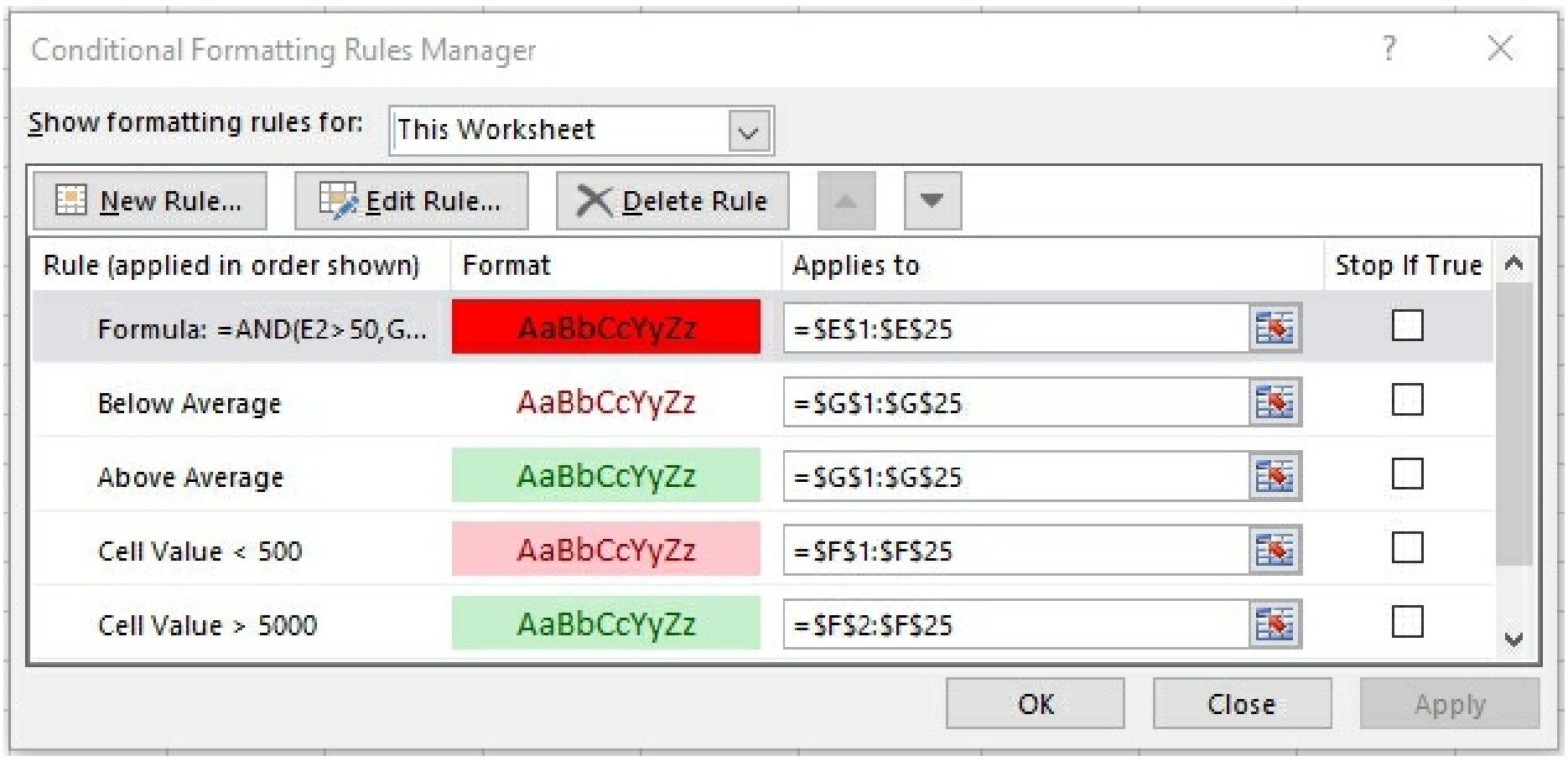


5. Click *Format*. A *Conditional Formatting Rules Manager* box will appear. Change the fill (background color) of the cell that meets the requirements of the formula to red. Click *OK* and then *Apply*.



Manage Rules

The conditioning rules can be managed through the *Conditional Formatting Rules Manager* pop-up available under *Conditional Formatting* icon under *Home*. You can see formatting rules for the current selection, for the entire current worksheet, for the other worksheets in the workbook. You can add new rules or edit or delete existing rules through the pop-up.


Conditional Formatting Rules Manager
Show formatting rules for:
This Worksheet
New Rule...
Edit Rule...
Delete Rule
Rule (applied in order shown)
Format
Applies to
Stop If True
Formula: =AND(E2>50,G...
AaBbCcYyZz
=$E$1:$E$25
Below Average
AaBbCcYyZz
=$G$1:$G$25
Above Average
AaBbCcYyZz
=$G$1:$G$25
Cell Value < 500
AaBbCcYyZz
=$F$1:$F$25
Cell Value > 5000
AaBbCcYyZz
=$F$2:$F$25
OK
Close
Apply

1.5 Sort and Filter Data

Sorting

Sorting data means arranging data in an order of interest, for example, in alphabetical order, highest to lowest, or order rows by colors or icons or custom lists. You can also sort the data by more than one column or row. Sorting helps to find and organize the data in the order you want. Hence, it aids in visualizing and understanding the data better to draw effective conclusions.

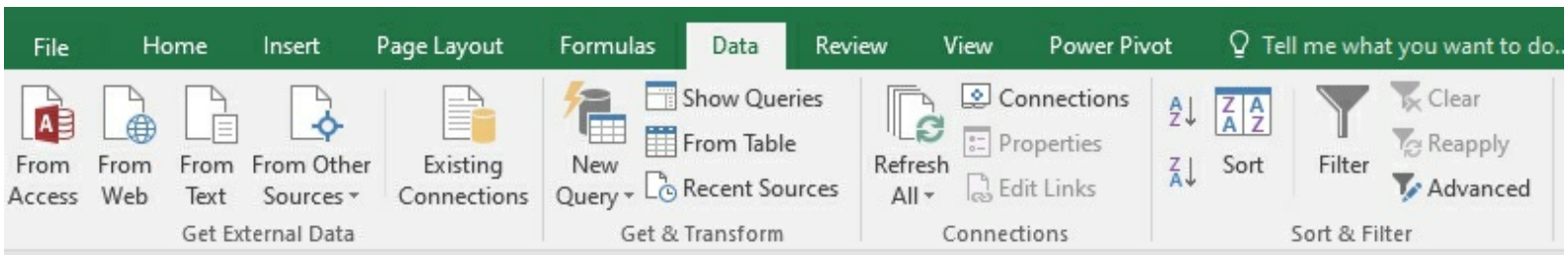
Practice

1. Open *Sort & Filter Data_Practice.xlsx* and go to *Sorting* sheet. You will continue to work with the customer data.

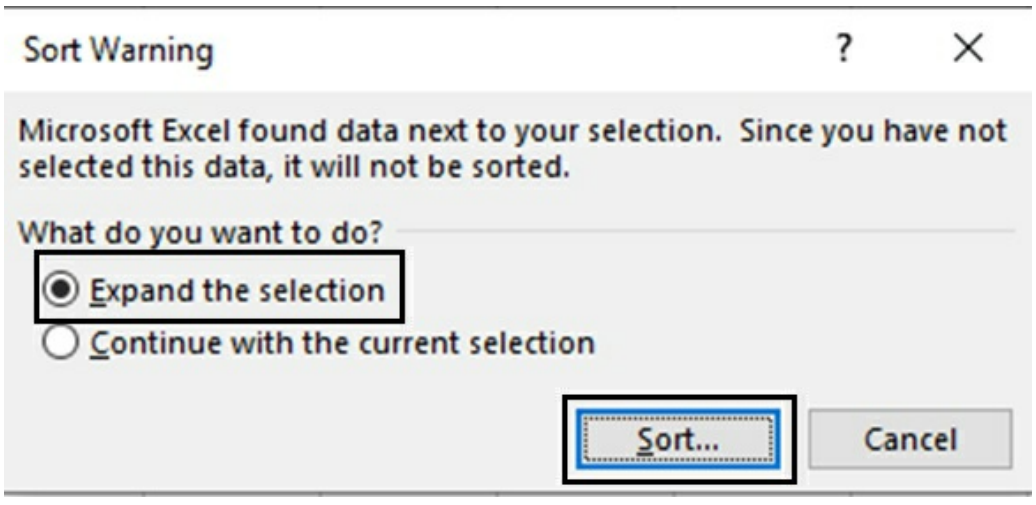
	A	B	C	D	E	F	G	H	I
1	Customer ID	First Name	Last Name	Gender	# purchases	Total $ spent	$ spent/purchase	City	Country
2	G94-8507	Denyse	Menicomb	F	1	$ 2,089.22	$ 2,089.22	Xintian	China
3	G32-7207	Daven	Shallcroff	M	20	$ 3,537.28	$ 176.86	Sendangkemulian	Indonesia
4	Q99-0116	Hastings	Bevar	M	5	$ 3,540.22	$ 708.04	Kandete	Tanzania
5	W15-8102	Kile	Melesk	M	56	$ 985.00	$ 17.59	Juzhen	China
6	B56-5019	Marika	Linstead	F	69	$ 3,997.18	$ 57.93	Kedungwaru	Indonesia
7	D27-4123	Page	Ibell	M	73	$ 2,082.75	$ 28.53	Hurghada	Egypt
8	T58-2788	Craggie	Sclater	M	89	$ 1,421.40	$ 15.97	R S	India
9	M00-4653	Tadeas	Craxford	M	48	$ 4,026.28	$ 83.88	Stockholm	Sweden
10	L95-4056	Latisha	Gulliver	F	61	$ 4,286.32	$ 70.27	Xuhang	China
11	Y14-4098	Evey	Groocock	F	45	$ 2,780.05	$ 61.78	Tpig	Russia
12	C85-8787	Donnie	Rankling	M	17	$ 2,933.50	$ 172.56	Huangpu	China
13	O13-5636	Ellissa	Sondland	F	35	$ 4,836.58	$ 138.19	Lucaya	Bahamas
14	F70-7356	Katinka	Bentame	F	50	$ 3,017.78	$ 60.36	Nantian	China
15	D03-9484	Nessi	Fairey	F	57	$ 3,190.33	$ 55.97	Gantang	China
16	Z16-2908	Rebe	Maplesden	F	50	$ 2,868.79	$ 57.38	Tena	Ecuador
17	V47-4052	Reyna	Struys	F	61	$ 1,974.51	$ 32.37	Nayoro	Japan
18	O89-6143	Orville	Dannell	M	44	$ 684.70	$ 15.56	El Doncello	Colombia
19	I81-0833	Pen	Perri	M	92	$ 2,262.31	$ 24.59	Pucheng	China
20	K75-3253	Franchot	Sergant	M	86	$ 5,671.67	$ 65.95	Cipancur	Indonesia
21	O22-9596	Jason	Hindsberg	M	60	$ 1,264.27	$ 21.07	Lanzhong	China
22	A27-5216	Jocelyne	Sackett	F	11	$ 367.18	$ 33.38	Xiangshan	China
23	Z65-5413	Jaquith	Coulston	F	62	$ 4,333.78	$ 69.90	Tongliao	China
24	E07-4299	Murielle	Holah	F	1	$ 935.82	$ 935.82	Zuyevka	Russia

2. Select any cell within the *First Name* column (column B). Go to the *Data* tab. In the *Sort & Filter* group,

   - To quickly sort the names in ascending order, select Sort A to Z icon (with downward arrow).
   - To quickly sort the names in descending order, select Sort Z to A icon (with downward arrow).



If a Sort Warning pops up, select *Expand the selection* and then *Sort* to make sure that Excel includes all columns when sorting the data, not just the column selected.

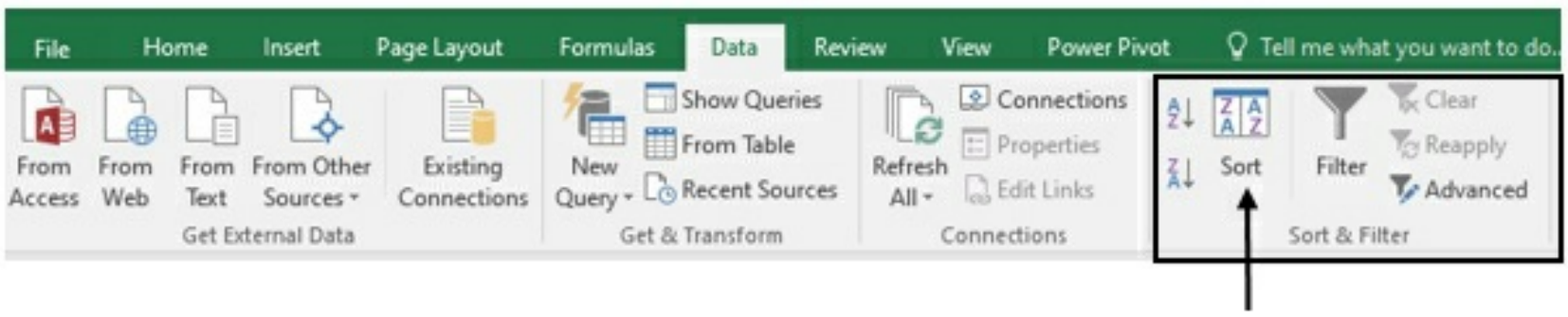


3. Select any cell within the *# purchases* column (column E). Go to the *Data* tab. In the *Sort & Filter*

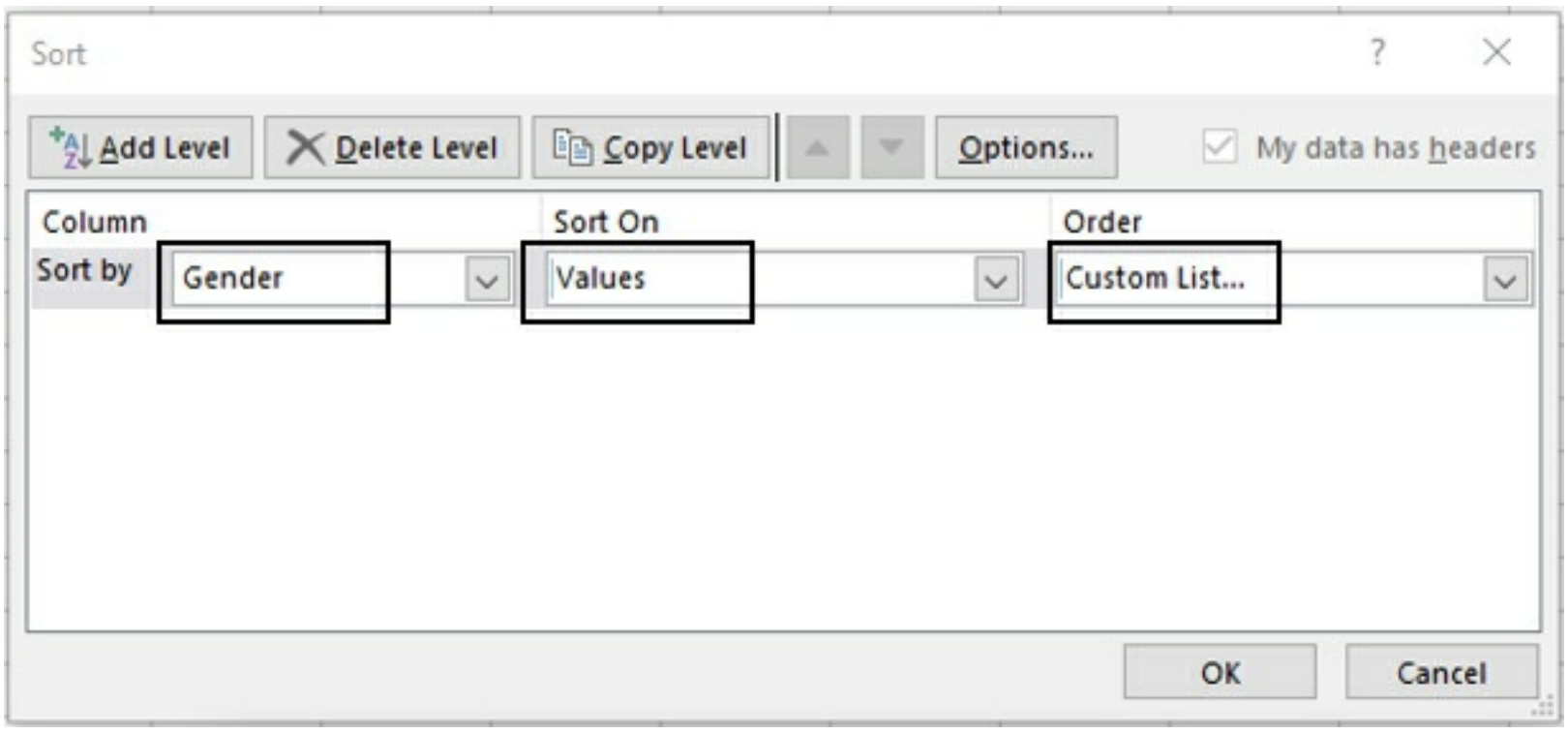
group,

- To quickly sort the numbers from smallest to largest, select Sort A to Z icon (with downward arrow).
- To quickly sort the numbers from largest to smallest, select Sort Z to A icon (with downward arrow).

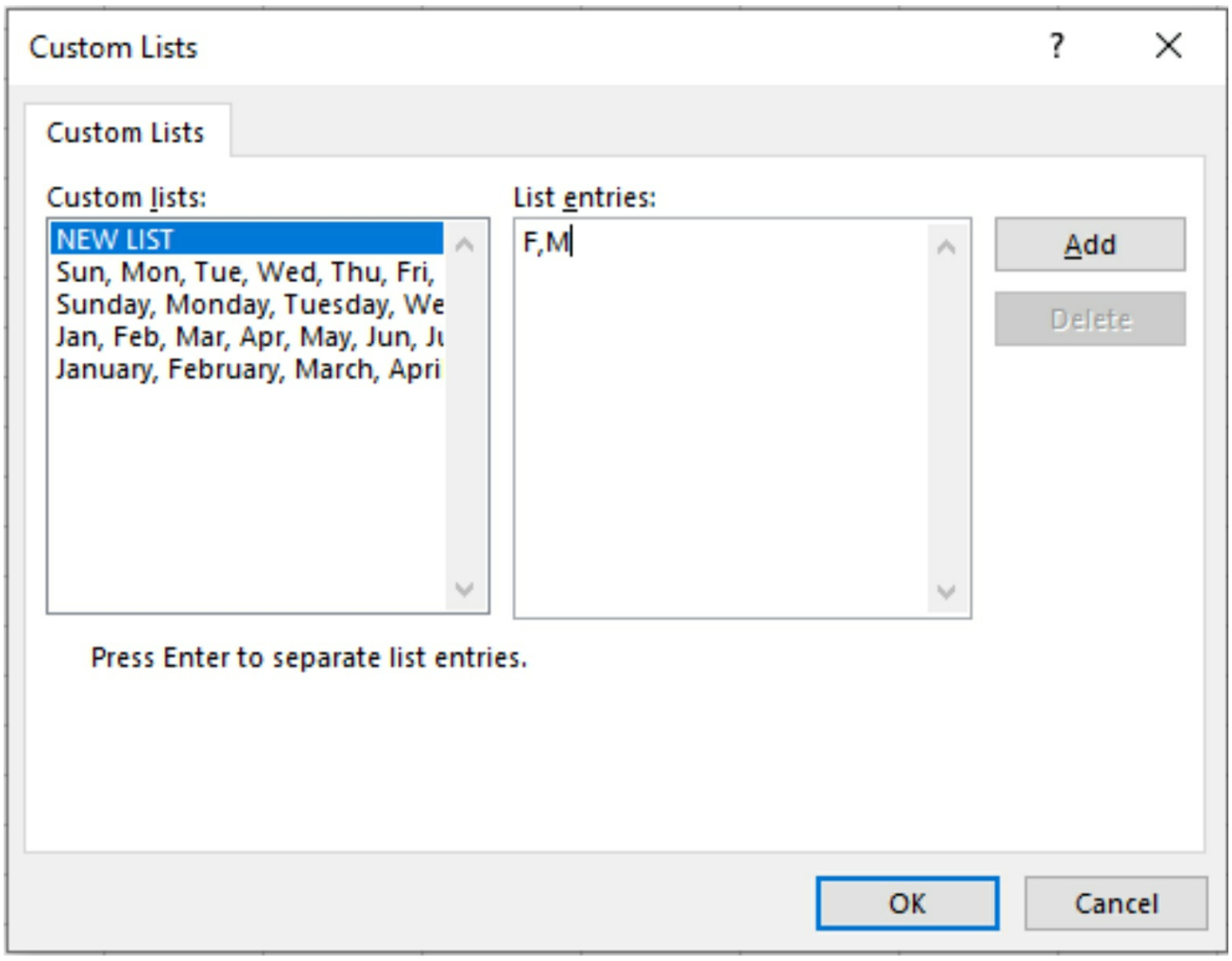
4. Now, let's learn how to sort by a combination of a custom list and font color. A custom list helps to sort in a user-defined order. Let's create a custom list to sot the *Gender* column (column D). Select a cell in the column.
5. Go to the *Data* tab and under *Sort & Filter* group, select *Sort*.



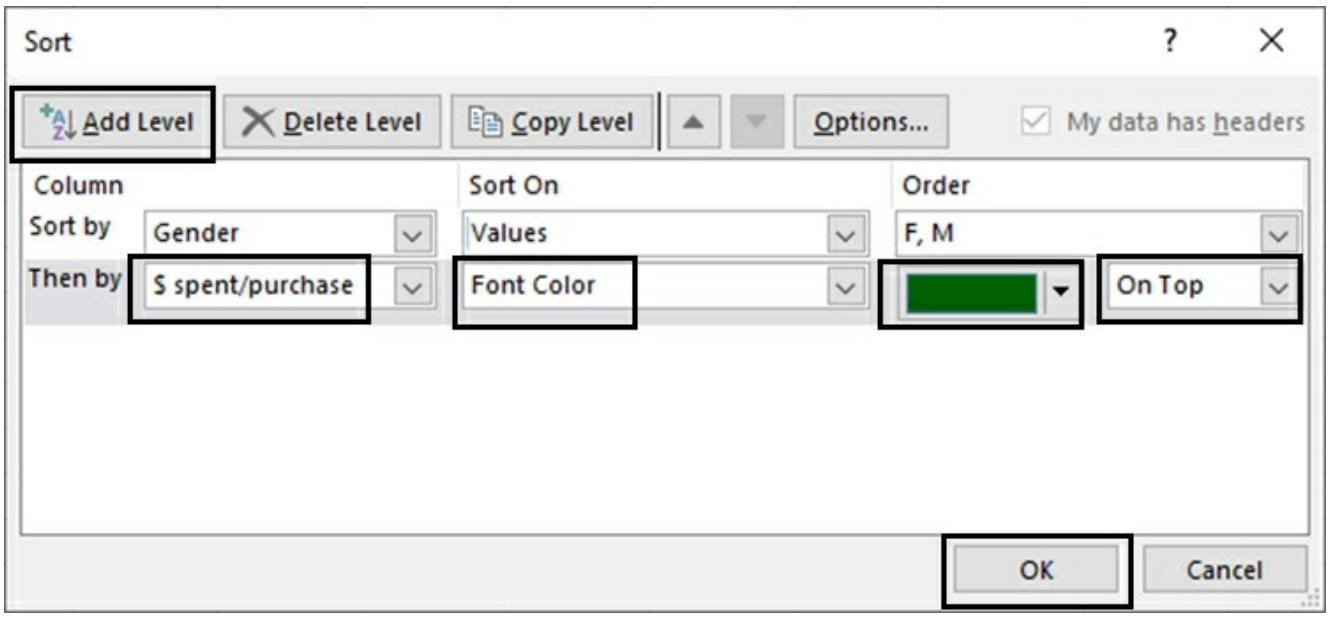
6. The *Sort* dialog box will pop up. For *Sort by*, select *Gender* from the dropdown, for *Sort On* choose *Values* and for *Order* choose *Custom List.*



7. The *Custom Lists* dialog box pops up. Select *NEW LIST* from the *Custom Lists:* box. In *List entries:* box, type *F, M* to sort the females prior to males. Click *Add* to save the new list. The new list will be added to the *Custom lists:* box. Select the new list and click *OK*.



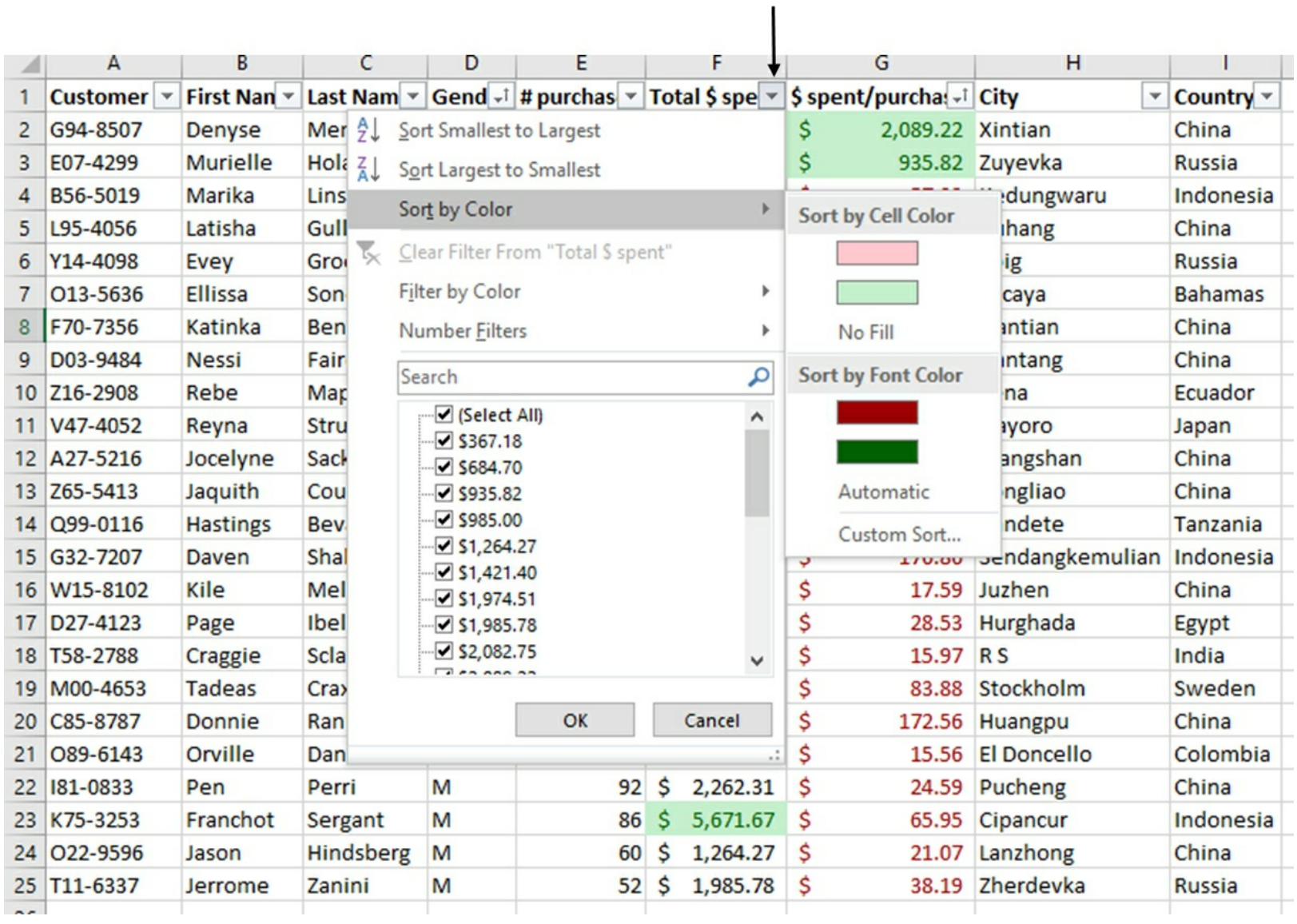
8. The *Custom Lists* dialog box will close. The *Sort* dialog box remains open. Now, let's sort the *$ spent/purchase* column (column G) by font color. On the *Sort* dialog box, click *Add Level*.
9. For *Then by*, select *$ spent/purchase* from the dropdown, for *Sort On*, select *Font Color* and for *Order*, select Green color and *On Top* from the dropdowns.



10. Click OK on the Sort dialog box. The worksheet will be sorted first by custom order supplied (F followed by M) and then by font color (green font first).

	A	B	C	D	E	F	G	H	I
1	Customer	First Nan	Last Nam	Gend	# purchas	Total $ spe	$ spent/purcha	City	Country
2	G94-8507	Denyse	Menicomb	F	1	$ 2,089.22	$ 2,089.22	Xintian	China
3	E07-4299	Murielle	Holah	F	1	$ 935.82	$ 935.82	Zuyevka	Russia
4	B56-5019	Marika	Linstead	F	69	$ 3,997.18	$ 57.93	Kedungwaru	Indonesia
5	L95-4056	Latisha	Gulliver	F	61	$ 4,286.32	$ 70.27	Xuhang	China
6	Y14-4098	Evey	Groocock	F	45	$ 2,780.05	$ 61.78	Tpig	Russia
7	O13-5636	Ellissa	Sondland	F	35	$ 4,836.58	$ 138.19	Lucaya	Bahamas
8	F70-7356	Katinka	Bentame	F	50	$ 3,017.78	$ 60.36	Nantian	China
9	D03-9484	Nessi	Fairey	F	57	$ 3,190.33	$ 55.97	Gantang	China
10	Z16-2908	Rebe	Maplesden	F	50	$ 2,868.79	$ 57.38	Tena	Ecuador
11	V47-4052	Reyna	Struys	F	61	$ 1,974.51	$ 32.37	Nayoro	Japan
12	A27-5216	Jocelyne	Sackett	F	11	$ 367.18	$ 33.38	Xiangshan	China
13	Z65-5413	Jaquith	Coulston	F	62	$ 4,333.78	$ 69.90	Tongliao	China
14	Q99-0116	Hastings	Bevar	M	5	$ 3,540.22	$ 708.04	Kandete	Tanzania
15	G32-7207	Daven	Shallcroff	M	20	$ 3,537.28	$ 176.86	Sendangkemulian	Indonesia
16	W15-8102	Kile	Melesk	M	56	$ 985.00	$ 17.59	Juzhen	China
17	D27-4123	Page	Ibell	M	73	$ 2,082.75	$ 28.53	Hurghada	Egypt
18	T58-2788	Craggie	Sclater	M	89	$ 1,421.40	$ 15.97	R S	India
19	M00-4653	Tadeas	Craxford	M	48	$ 4,026.28	$ 83.88	Stockholm	Sweden
20	C85-8787	Donnie	Rankling	M	17	$ 2,933.50	$ 172.56	Huangpu	China
21	O89-6143	Orville	Dannell	M	44	$ 684.70	$ 15.56	El Doncello	Colombia
22	I81-0833	Pen	Perri	M	92	$ 2,262.31	$ 24.59	Pucheng	China
23	K75-3253	Franchot	Sergant	M	86	$ 5,671.67	$ 65.95	Cipancur	Indonesia
24	O22-9596	Jason	Hindsberg	M	60	$ 1,264.27	$ 21.07	Lanzhong	China
25	T11-6337	Jerrome	Zanini	M	52	$ 1,985.78	$ 38.19	Zherdevka	Russia

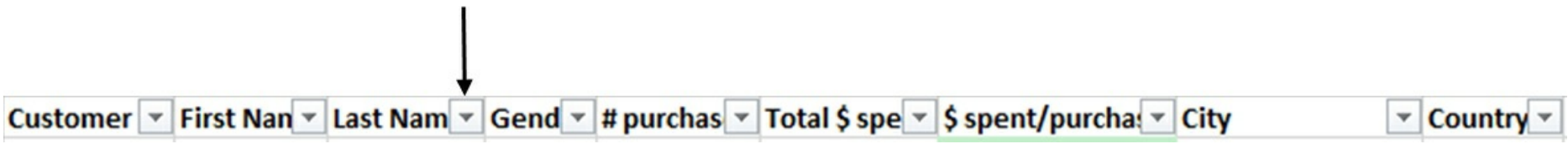
11. Lastly, as a tip, there is an alternative method to access sort. Click on the dropdown box embedded in the column header of the *$ spent/purchase* column (column G). If the dropdown box is not visible, go to the *Data* tab and under the *Sort & Filter* group, click on *Filter* to make it visible across all headers. The dropdown is called AutoFilter and offers options for sorting as well as filtering and will be discussed in greater detail in the next section.



Filtering

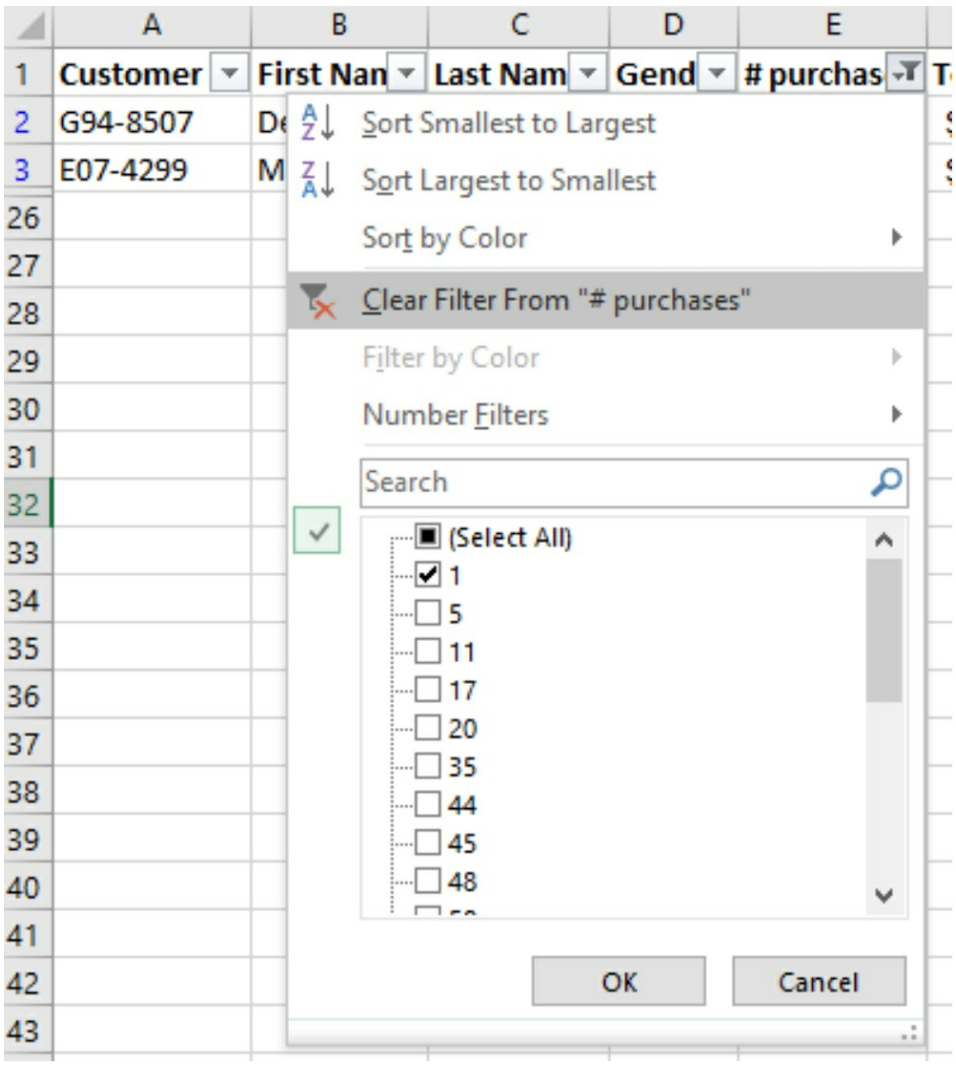
If your worksheet contains hundreds of rows, it can be difficult to find information on specific rows quickly. Filters can be used to display only the data that meets your criteria, allowing you to view only what you need. Filters are cumulative, which means you can apply multiple filters to help narrow down your results.

Excel provides AutoFilter as an easy way to turn the values in an Excel column into filters based on the column's contents. AutoFilter is represented by the arrow next to the column header.



If the AutoFilter is not visible on your sheet, you can turn it on with simple steps. Click any cell within your range, go to the *Data* tab, and under *Sort & Filter* group, click *Filter*.

Once you filter data in a range of cells or tables, you can also clear the filter to redisplay all of the data. To clear a filter, click on the AutoFilter next to the column header of the filtered column and choose *Clear Filter From "COLUMN NAME"*.

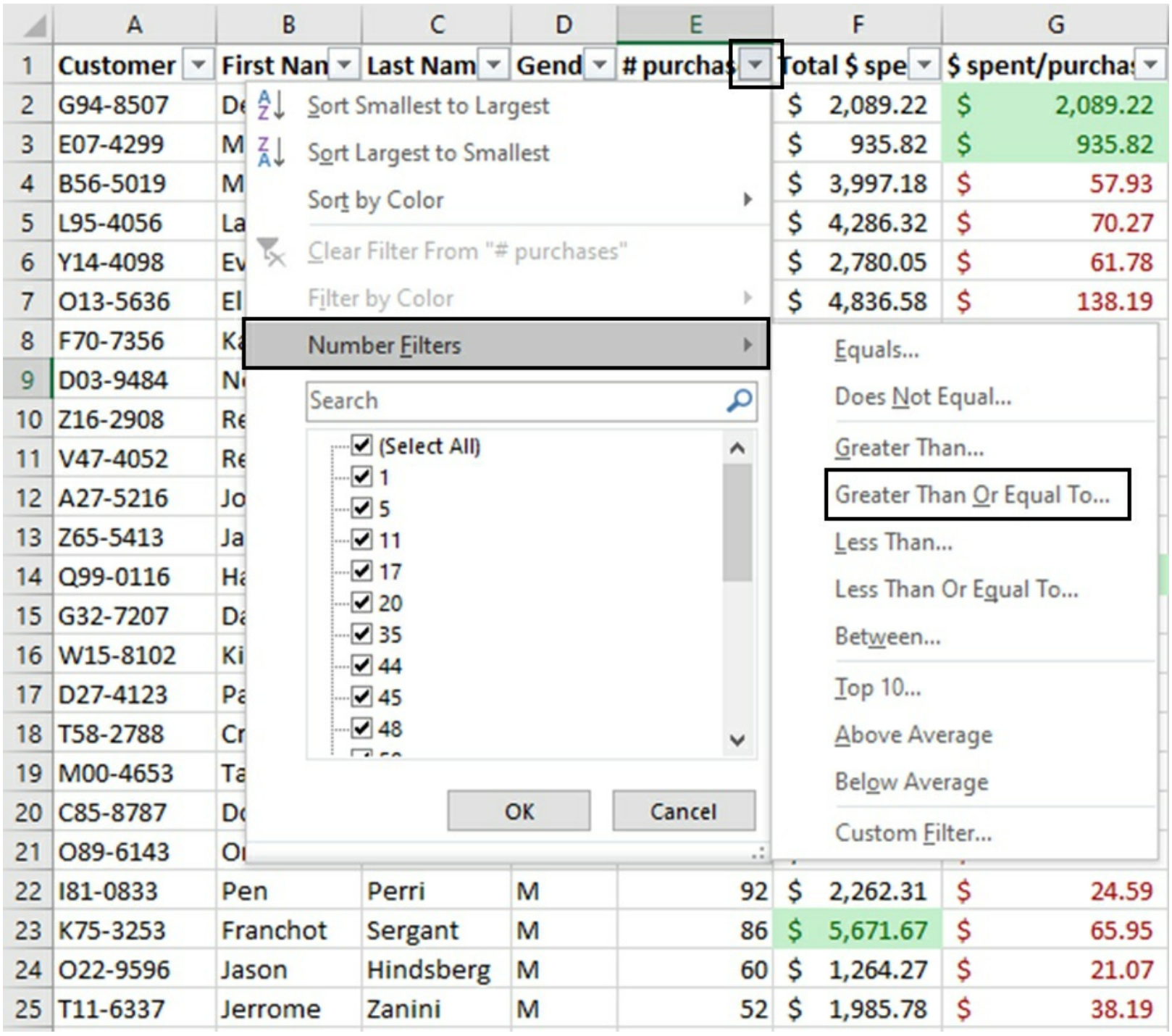


Practice

1. Open *Sort & Filter Data_Practice.xlsx* and go to *Filtering* sheet. You will continue to work with the customer data.

	A	B	C	D	E	F	G	H	I
1	Customer ID	First Name	Last Name	Gender	# purchases	Total $ spent	$ spent/purchase	City	Country
2	G94-8507	Denyse	Menicomb	F	1	$ 2,089.22	$ 2,089.22	Xintian	China
3	G32-7207	Daven	Shallcroff	M	20	$ 3,537.28	$ 176.86	Sendangkemulian	Indonesia
4	Q99-0116	Hastings	Bevar	M	5	$ 3,540.22	$ 708.04	Kandete	Tanzania
5	W15-8102	Kile	Melesk	M	56	$ 985.00	$ 17.59	Juzhen	China
6	B56-5019	Marika	Linstead	F	69	$ 3,997.18	$ 57.93	Kedungwaru	Indonesia
7	D27-4123	Page	Ibell	M	73	$ 2,082.75	$ 28.53	Hurghada	Egypt
8	T58-2788	Craggie	Sclater	M	89	$ 1,421.40	$ 15.97	R S	India
9	M00-4653	Tadeas	Craxford	M	48	$ 4,026.28	$ 83.88	Stockholm	Sweden
10	L95-4056	Latisha	Gulliver	F	61	$ 4,286.32	$ 70.27	Xuhang	China
11	Y14-4098	Evey	Groocock	F	45	$ 2,780.05	$ 61.78	Tpig	Russia
12	C85-8787	Donnie	Rankling	M	17	$ 2,933.50	$ 172.56	Huangpu	China
13	O13-5636	Ellissa	Sondland	F	35	$ 4,836.58	$ 138.19	Lucaya	Bahamas
14	F70-7356	Katinka	Bentame	F	50	$ 3,017.78	$ 60.36	Nantian	China
15	D03-9484	Nessi	Fairey	F	57	$ 3,190.33	$ 55.97	Gantang	China
16	Z16-2908	Rebe	Maplesden	F	50	$ 2,868.79	$ 57.38	Tena	Ecuador
17	V47-4052	Reyna	Struys	F	61	$ 1,974.51	$ 32.37	Nayoro	Japan
18	O89-6143	Orville	Dannell	M	44	$ 684.70	$ 15.56	El Doncello	Colombia
19	I81-0833	Pen	Perri	M	92	$ 2,262.31	$ 24.59	Pucheng	China
20	K75-3253	Franchot	Sergant	M	86	$ 5,671.67	$ 65.95	Cipancur	Indonesia
21	O22-9596	Jason	Hindsberg	M	60	$ 1,264.27	$ 21.07	Lanzhong	China
22	A27-5216	Jocelyne	Sackett	F	11	$ 367.18	$ 33.38	Xiangshan	China
23	Z65-5413	Jaquith	Coulston	F	62	$ 4,333.78	$ 69.90	Tongliao	China
24	E07-4299	Murielle	Holah	F	1	$ 935.82	$ 935.82	Zuyevka	Russia

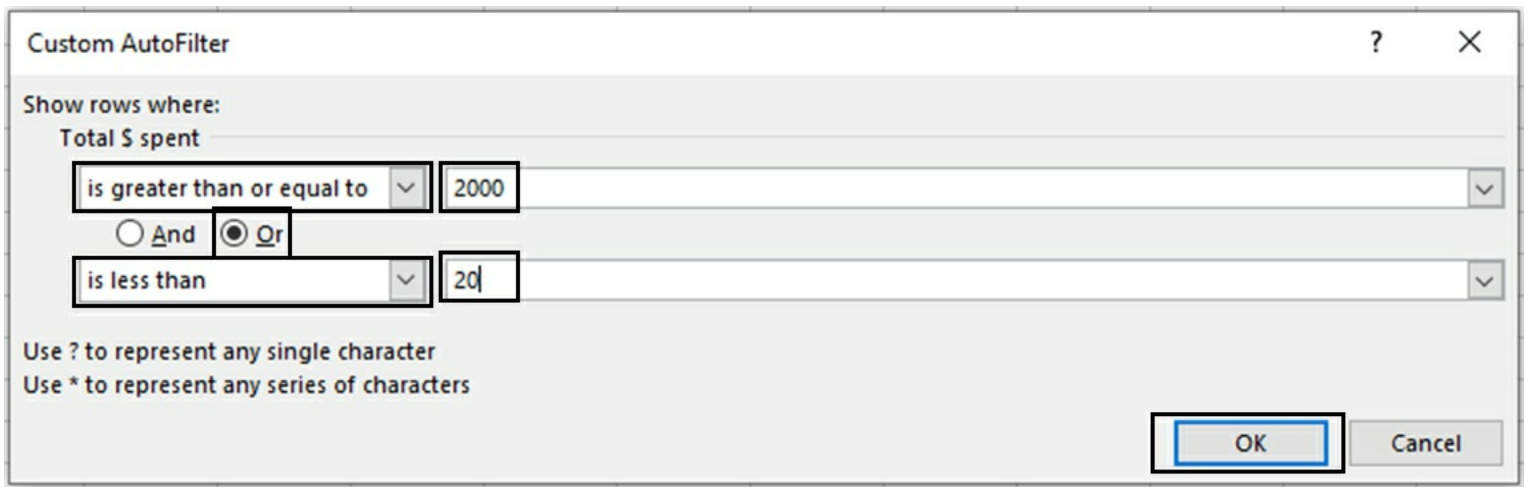
2. Follow previously described steps on turn on the AutoFilter if it isn't on already.
3. Click the AutoFilter on *#Total $ spent* column and select *Number Filters*. Various operations available to filter the column appear. Click *Greater Than Or Equal To.*



4. A *Custom AutoFilter* box pops up. Follow the below steps to filter rows where *Total $ Spent* values is either greater than or equal to 2000, or less than 20:
    - Ensure *is greater than or equal to* is selected on the first dropdown and enter the value

2000 next to it.

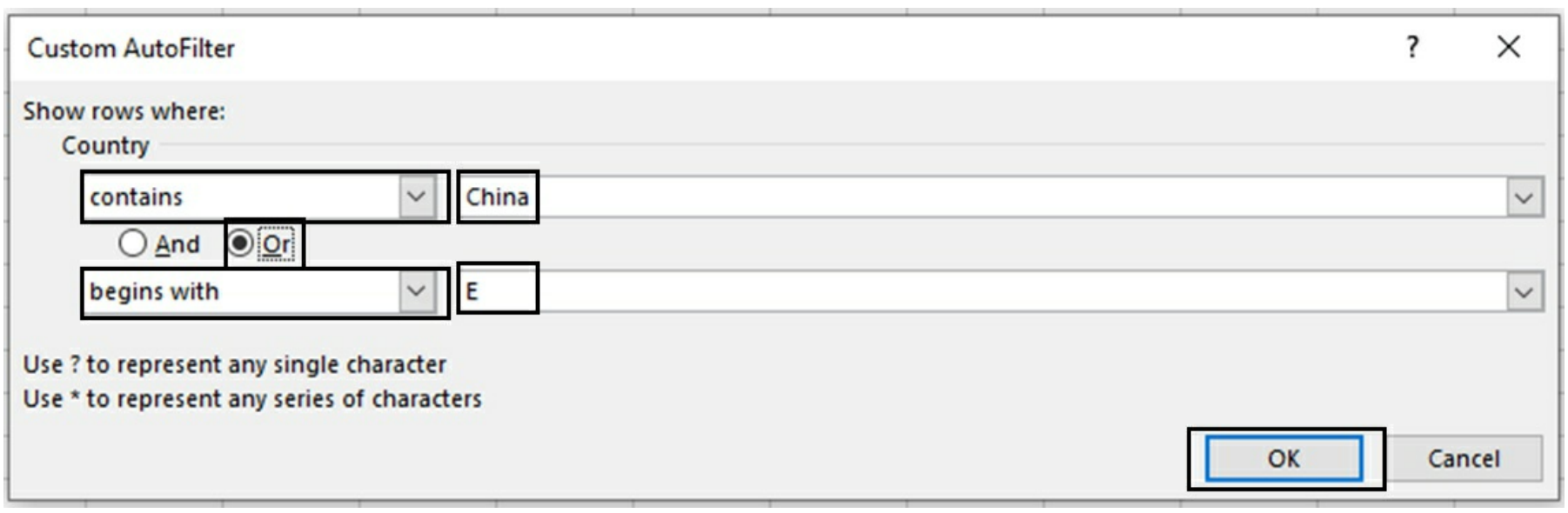
- Select *Or*.
- Select *is less than* on the second dropdown and enter the value *20* next to it. Click *OK*.



5. Verify appropriate rows have been filtered.

	A	B	C	D	E	F	G	H	I
1	Customer	First Nan	Last Nam	Gend	# purchas	Total $ spe	$ spent/purcha	City	Country
2	G94-8507	Denyse	Menicomb	F	1	$ 2,089.22	$ 2,089.22	Xintian	China
4	B56-5019	Marika	Linstead	F	69	$ 3,997.18	$ 57.93	Kedungwaru	Indonesia
5	L95-4056	Latisha	Gulliver	F	61	$ 4,286.32	$ 70.27	Xuhang	China
6	Y14-4098	Evey	Groocock	F	45	$ 2,780.05	$ 61.78	Tpig	Russia
7	O13-5636	Ellissa	Sondland	F	35	$ 4,836.58	$ 138.19	Lucaya	Bahamas
8	F70-7356	Katinka	Bentame	F	50	$ 3,017.78	$ 60.36	Nantian	China
9	D03-9484	Nessi	Fairey	F	57	$ 3,190.33	$ 55.97	Gantang	China
10	Z16-2908	Rebe	Maplesden	F	50	$ 2,868.79	$ 57.38	Tena	Ecuador
13	Z65-5413	Jaquith	Coulston	F	62	$ 4,333.78	$ 69.90	Tongliao	China
14	Q99-0116	Hastings	Bevar	M	5	$ 3,540.22	$ 708.04	Kandete	Tanzania
15	G32-7207	Daven	Shallcroff	M	20	$ 3,537.28	$ 176.86	Sendangkemulian	Indonesia
17	D27-4123	Page	Ibell	M	73	$ 2,082.75	$ 28.53	Hurghada	Egypt
19	M00-4653	Tadeas	Craxford	M	48	$ 4,026.28	$ 83.88	Stockholm	Sweden
20	C85-8787	Donnie	Rankling	M	17	$ 2,933.50	$ 172.56	Huangpu	China
22	I81-0833	Pen	Perri	M	92	$ 2,262.31	$ 24.59	Pucheng	China
23	K75-3253	Franchot	Sergant	M	86	$ 5,671.67	$ 65.95	Cipancur	Indonesia

6. You have already filtered the data to show the records with *Total $ Spent values* >= 2000, or < 20. Now let's narrow it down further and filter these records that belong to customers residing in either China or a country whose name starts with the letter 'E'.
7. Click the AutoFilter on the *Country* column and select *Text Filters*. Various operations available to filter the column appear. Click *Contains*.
8. A *Custom AutoFilter* box pops up. Follow the below steps to filter rows where the Country is either China or starts with the letter 'E':
    - Ensure *contains* is selected on the first dropdown and type *China* next to it.
    - Select *Or*.
    - Select *begins with* on second dropdown and type *E* next to it. Click *OK*.



9. Verify appropriate rows have been filtered.

	A	B	C	D	E	F	G	H	I
1	Customer	First Nan	Last Nam	Gend	# purchas	Total $ spe	$ spent/purcha	City	Country
2	G94-8507	Denyse	Menicomb	F	1	$ 2,089.22	$ 2,089.22	Xintian	China
5	L95-4056	Latisha	Gulliver	F	61	$ 4,286.32	$ 70.27	Xuhang	China
8	F70-7356	Katinka	Bentame	F	50	$ 3,017.78	$ 60.36	Nantian	China
9	D03-9484	Nessi	Fairey	F	57	$ 3,190.33	$ 55.97	Gantang	China
10	Z16-2908	Rebe	Maplesden	F	50	$ 2,868.79	$ 57.38	Tena	Ecuador
13	Z65-5413	Jaquith	Coulston	F	62	$ 4,333.78	$ 69.90	Tongliao	China
17	D27-4123	Page	Ibell	M	73	$ 2,082.75	$ 28.53	Hurghada	Egypt
20	C85-8787	Donnie	Rankling	M	17	$ 2,933.50	$ 172.56	Huangpu	China
22	I81-0833	Pen	Perri	M	92	$ 2,262.31	$ 24.59	Pucheng	China

1.6 Data Validation

Sometimes the data to be analyzed is manually entered by one or more users. To ensure all data entries are accurate and consistent, you may want to control information entered into specific cells. When setting up a workbook for data entry, data validation can be used to restrict the type of data or the values that users enter into a cell. For example, you can:

- Show an input message on values allowed to be entered
- Create a drop-down list of predefined items for a cell
- Restrict type and range of entries, such as a date range or whole numbers only
- Create custom rules for what can be entered
- Show a warning message when incorrect data has been entered

Practice

Suppose a landscaping company wants to keep track of customer inquiries as well as the employees they get allocated to. This will help them to keep a track of the type of enquiries coming in as well as the workload of their staff.

You will create an enquiry inventory form for the company with below columns. The data validation constraints associated with each of the columns are also mentioned, where applicable. Entered data is validated against these constraints and accepted only when it meets the validation criteria. Otherwise, the user will get an error message.

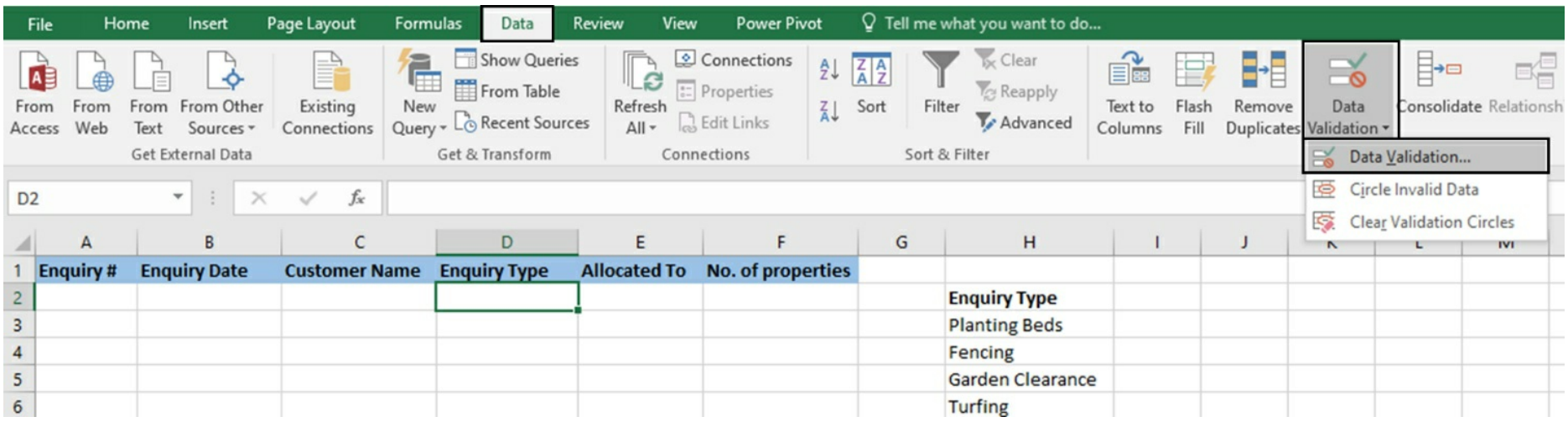
1. Enquiry #
2. Enquiry Date – enter a date greater than the data entry start date
3. Customer Name
4. Enquiry Type – select from a given list
5. Allocated To – select from a given list
6. No. of properties – enter a whole number

Follow the below steps to create the enquiry inventory form embedded with data validation:

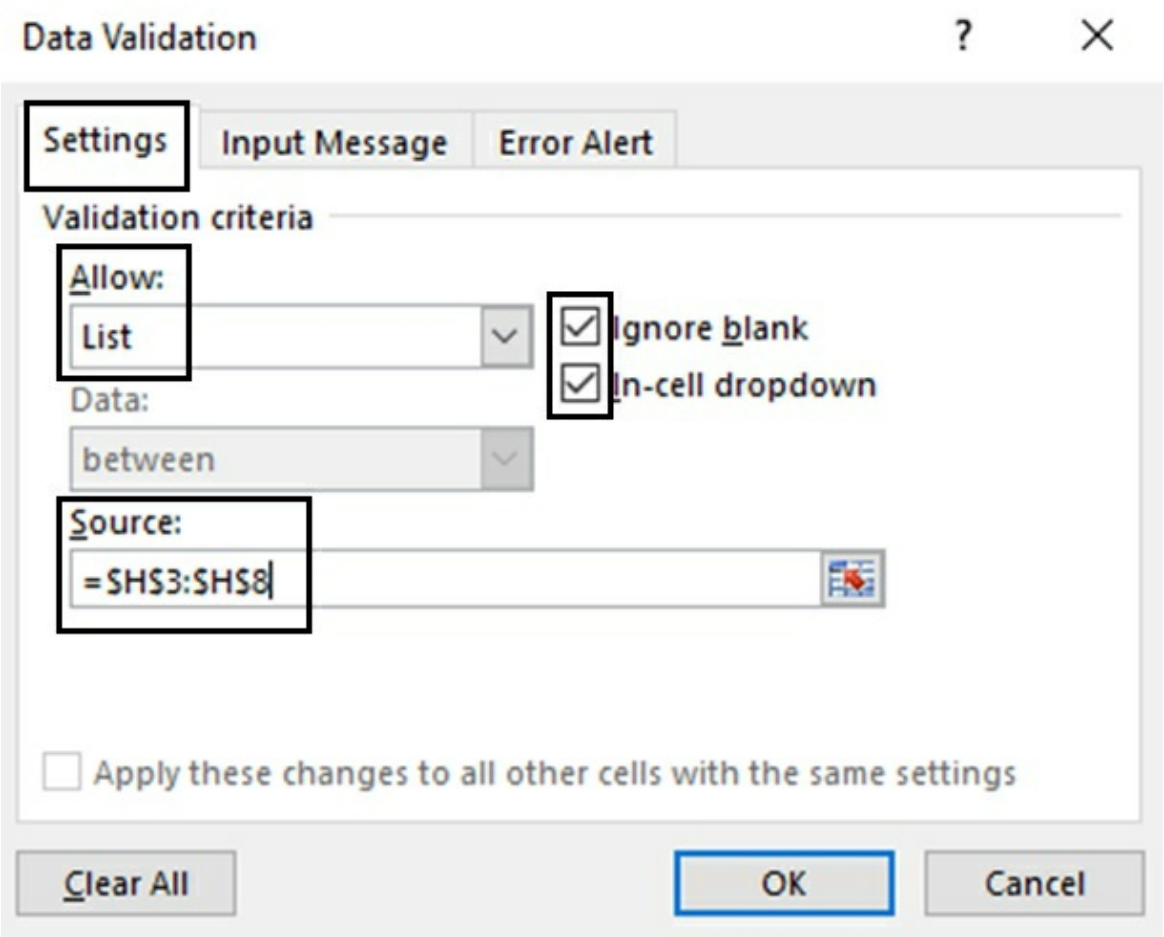
1. Open *Data Validation_Practice.xlsx* and go to *Data Validation* sheet. You will notice the sheet already has the above column headers in columns A through F. In addition, it also has eligible Enquiry Type and Employee lists as well start date for data entry.

	A	B	C	D	E	F	G	H
1	Enquiry #	Enquiry Date	Customer Name	Enquiry Type	Allocated To	No. of properties		
2								**Enquiry Type**
3								Planting Beds
4								Fencing
5								Garden Clearance
6								Turfing
7								Walkway
8								Driveway
9								
10								
11								
12								**Employee List**
13								Freemon Sansbury
14								Tymothy Kleisle
15								Pammy Servante
16								Codee Joris
17								Rosabel Watmore
18								Larissa Huzzay
19								Gilberta Hintzer
20								Leelah Chaffer
21								Ruthann Just
22								Annora Stower
23								
24								**Start Date**
25								07/12/2021

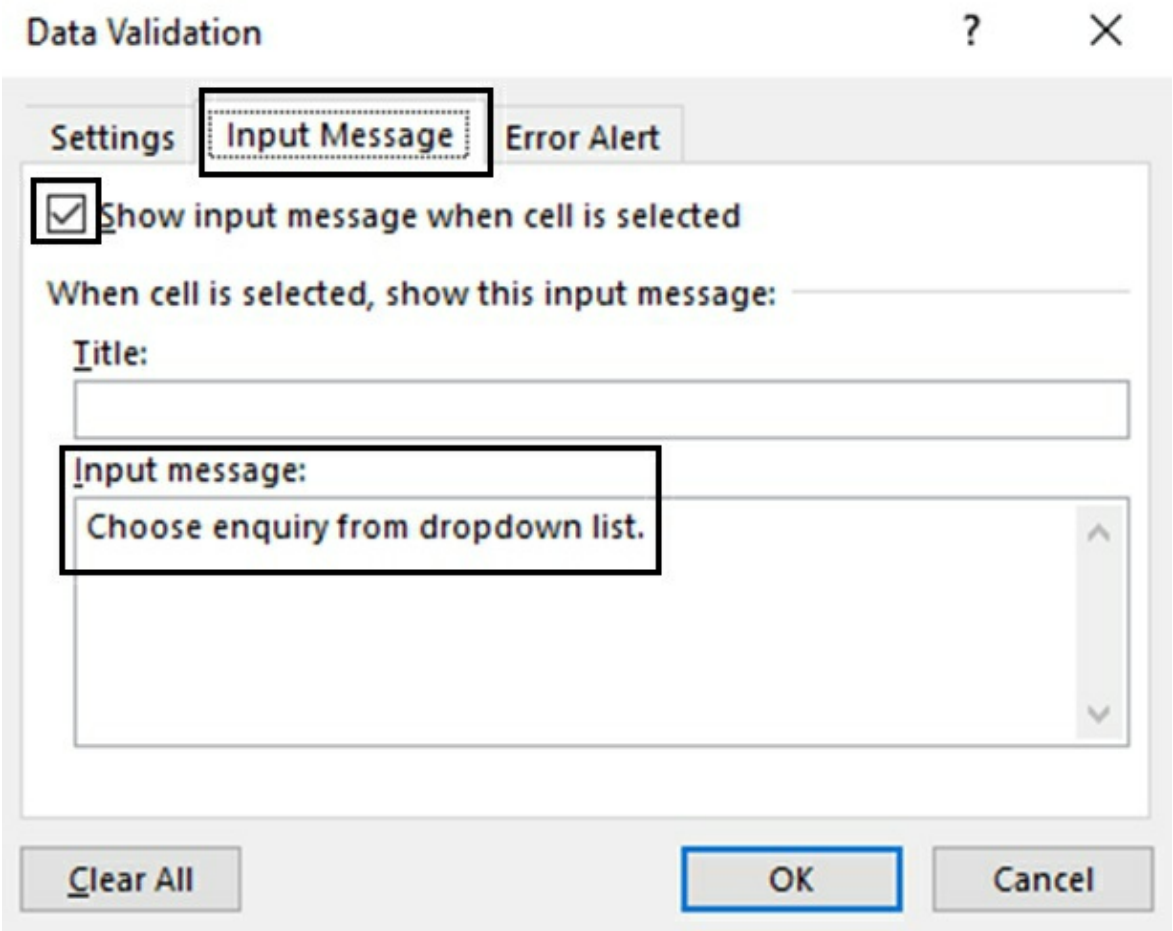
2. Let's start with column D – *Enquiry Type*. Click the first cell under the column (D2). Then go to the *Data* tab on the top ribbon.
3. Under the *Data Tools* group, click on *Data Validation* and select *Data Validation* from the drop-down list. A *Data Validation* dialog box will pop up.



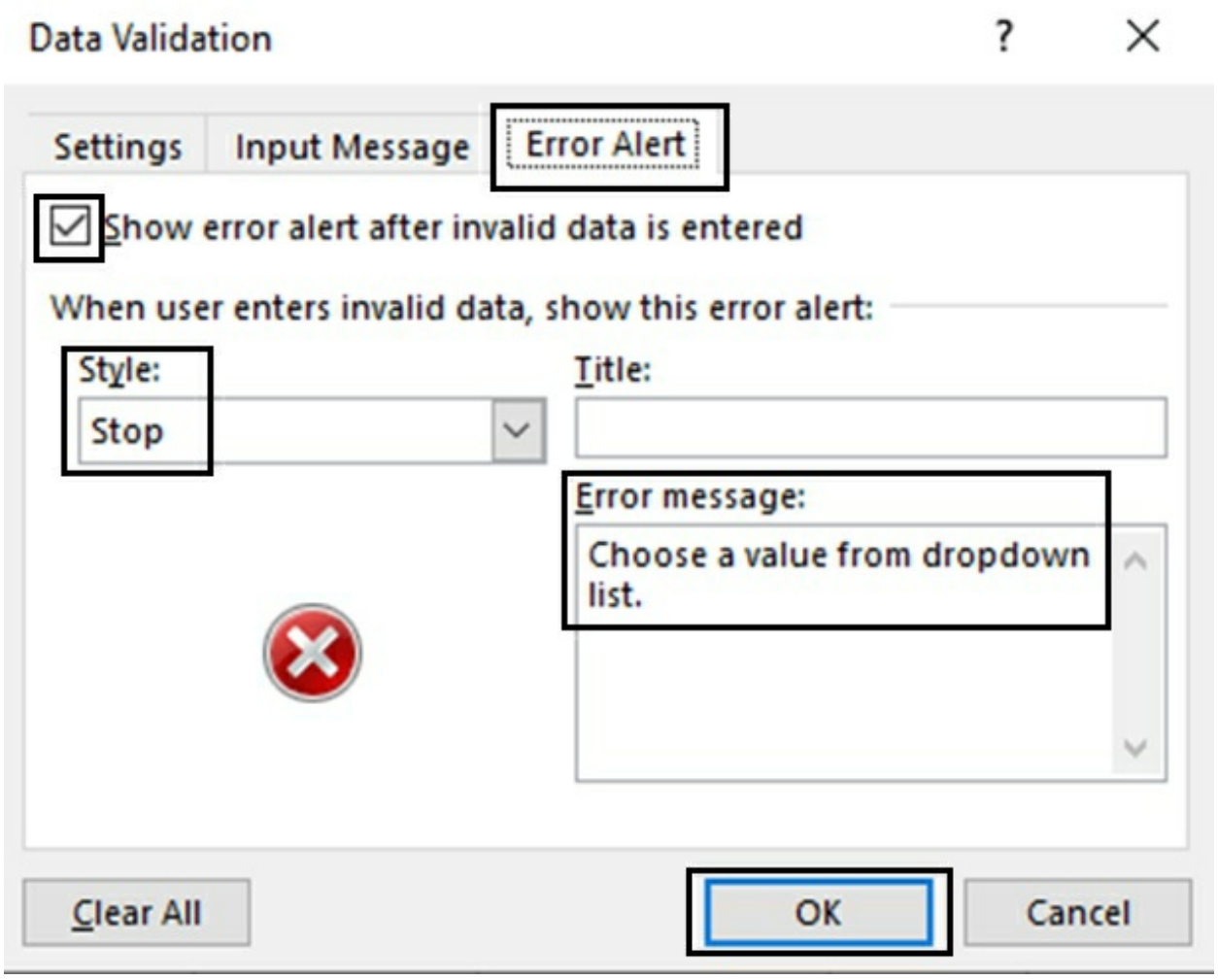
4. The *Settings* tab on the dialog box sets the validation criteria. For *Allow:*, select *List* from the dropdown. For *Source*, select the range *$H$3:$H$8* that contains eligible enquiry types. Check the boxes *Ignore blank* and *In-cell dropdown*.



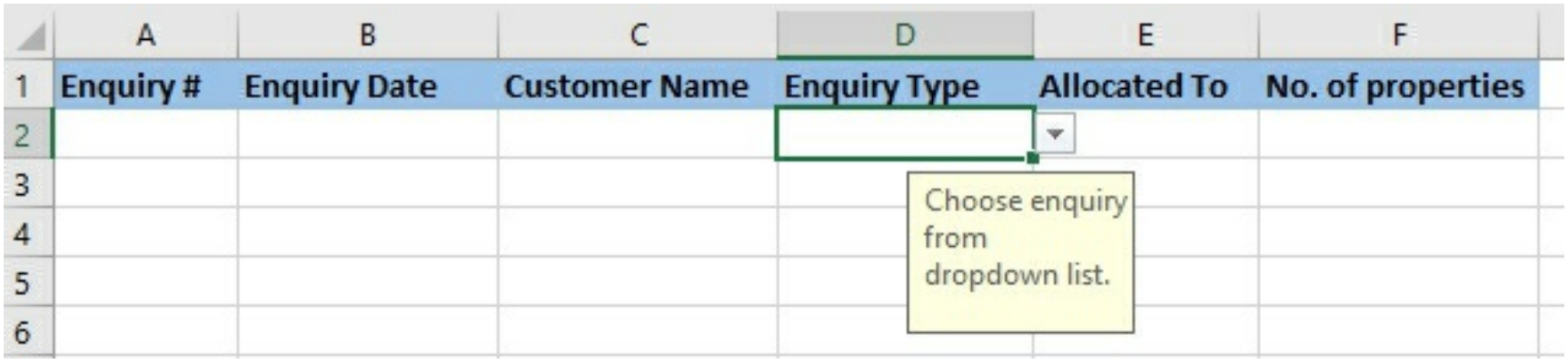
5. Now, go to the *Input Message* tab on the dialog box that sets the message displayed when the cells are selected. Check *Show input message when cell is selected* box. In the *Input message:* box, enter *Choose enquiry from the dropdown*.



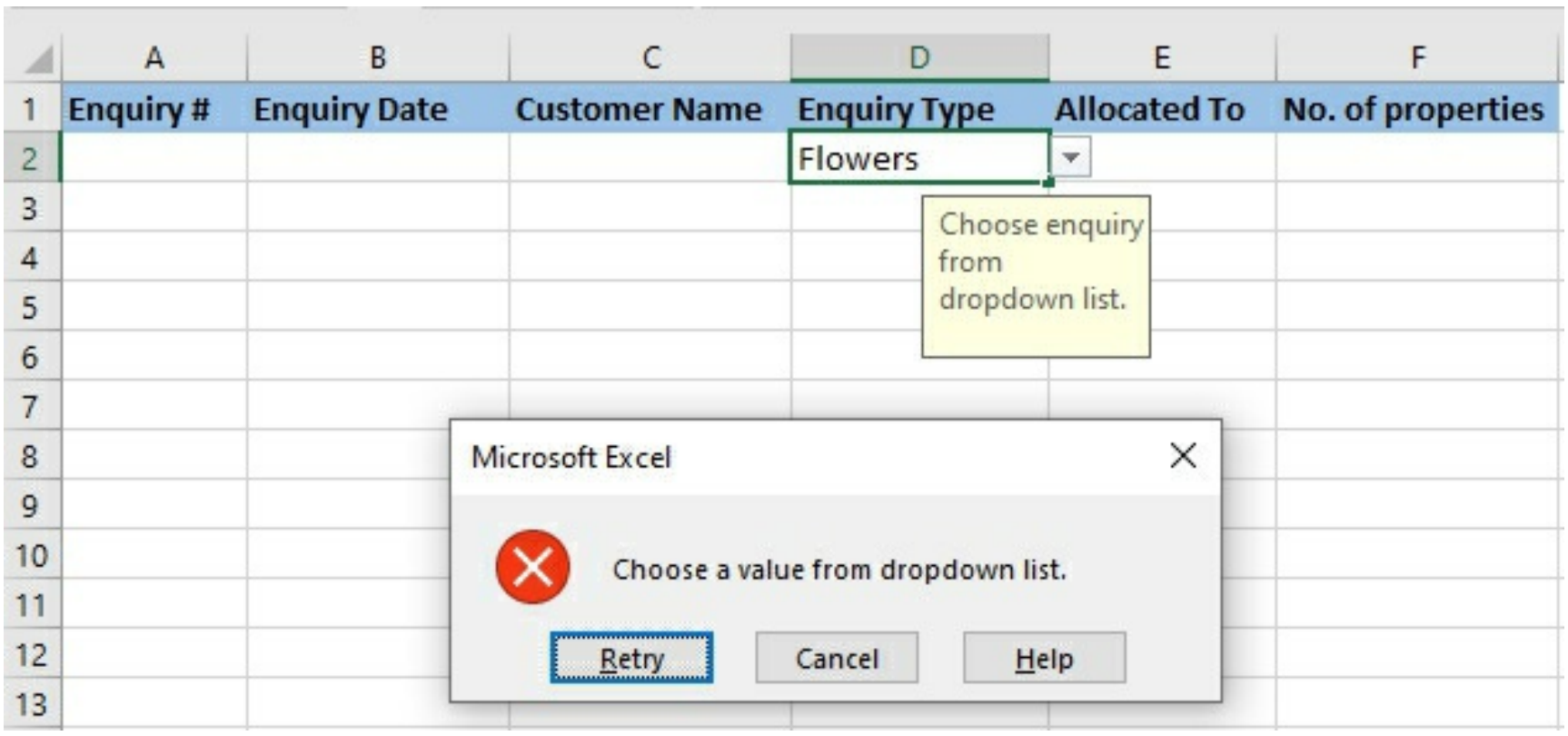
6. Now, go to the *Error Alert* tab on the dialog box that sets the error message to be displayed when wrong data is entered into the cell. Check the *Show error alert after invalid data is entered* box. For *Style*, select *Stop* from the dropdown. In the *Error message:* box, type *Choose a value from dropdown list* and click *OK*.



7. To verify that data validation has been set for the cell D2, click on the cell. The Input message set in step 5 as well as a dropdown button appears.



8. Instead of choosing a value from the dropdown, double click on the cell and type *Flowers* and hit *Enter*. The error message set in step 6 will show up. Click on *Retry* and choose a value from the dropdown. Data validation criteria for the cell have been successfully assigned.



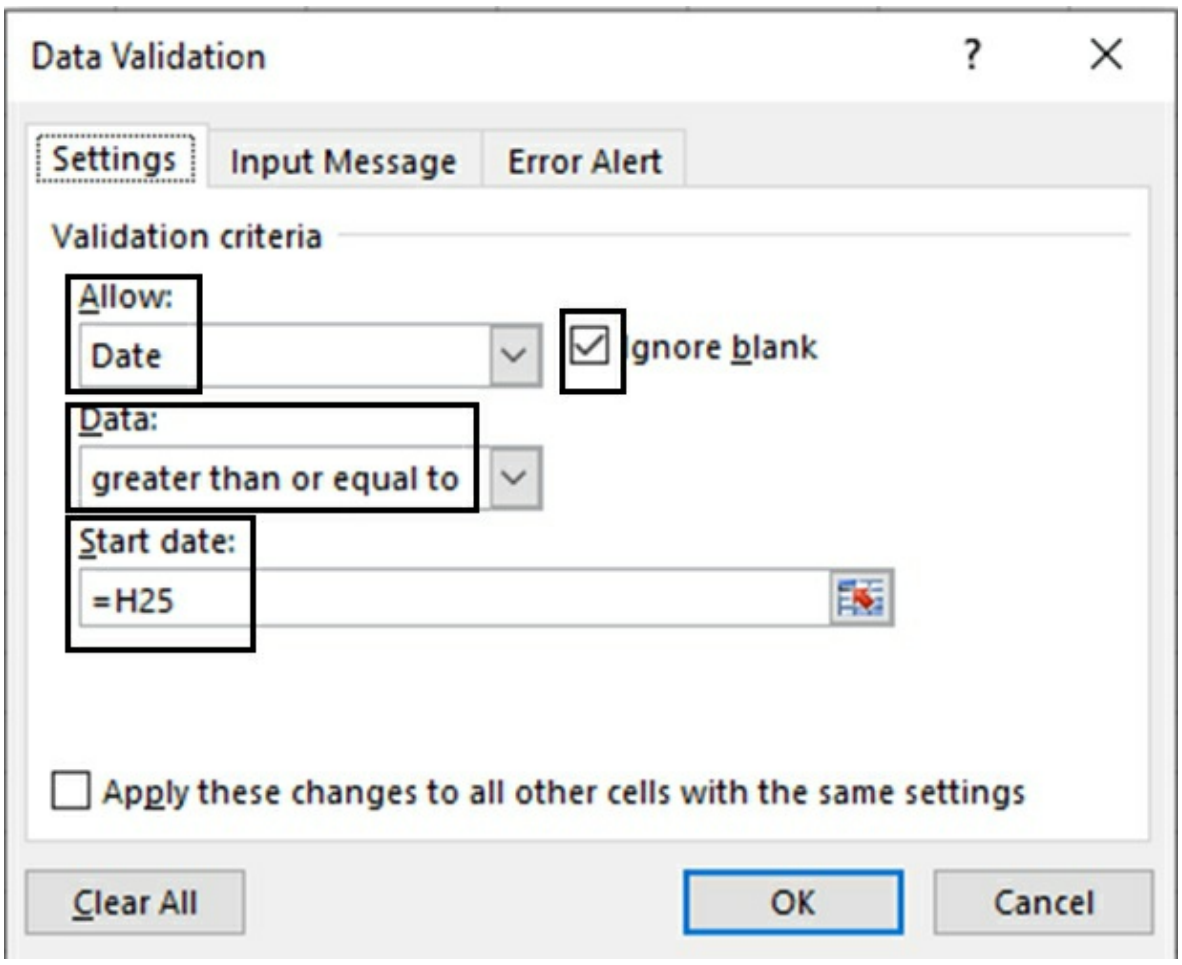
9. To assign the criteria for the column, click on the cell D2 and hover over the bottom right of the cell until the drag handle (+ symbol) appears. Pull it down to as many rows as needed. Hence, data validation criteria can be dragged down just like a formula.

10. Now to set validation criteria for column E – *Allocated To* based on the list stored in the range $H$13:$H$22, click the first cell under the column (E2), and follow steps 3-9 with the below changes:

- On *Settings* tab (step 4), for source, select range *$H$13:$H$22*.
- On *Input Message* tab (step 5), type *Choose employee from dropdown list*.

11. *Enquiry Date* in Column B should be on or after the date value stored in the cell $H$25. We will assume that this is the start date of data entry and any date prior to this start date can't be entered. To set this criterion, follow the previously mentioned steps to open the *Data Validation* dialog box.

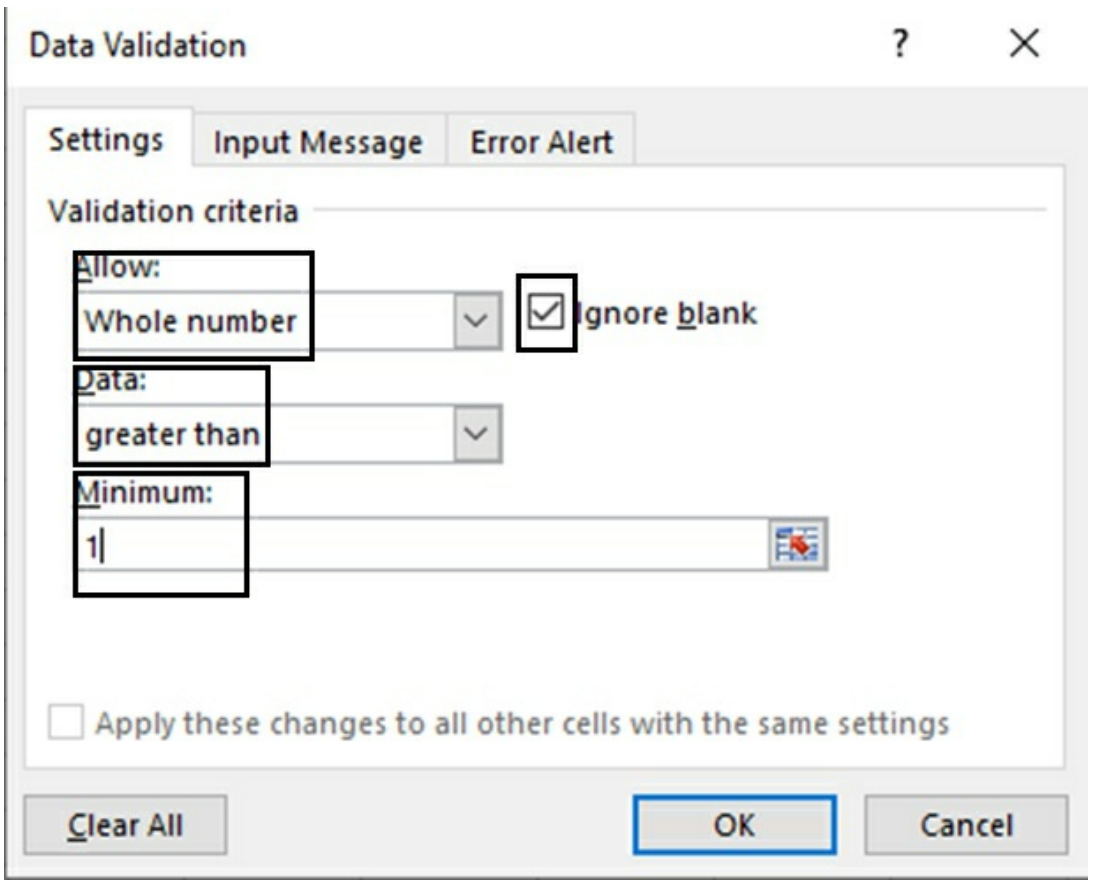
12. On the *Settings* tab, for *Allow:*, select *Date* from the dropdown. For *Data*, select *greater than or equal to* and for *Start date:*, select the cell *$H$25* that contains the start date. Check the *Ignore blank* box. Follow previously mentioned steps to set appropriate input message and error alert, as well as to drag the validation criteria down to as many rows as needed.



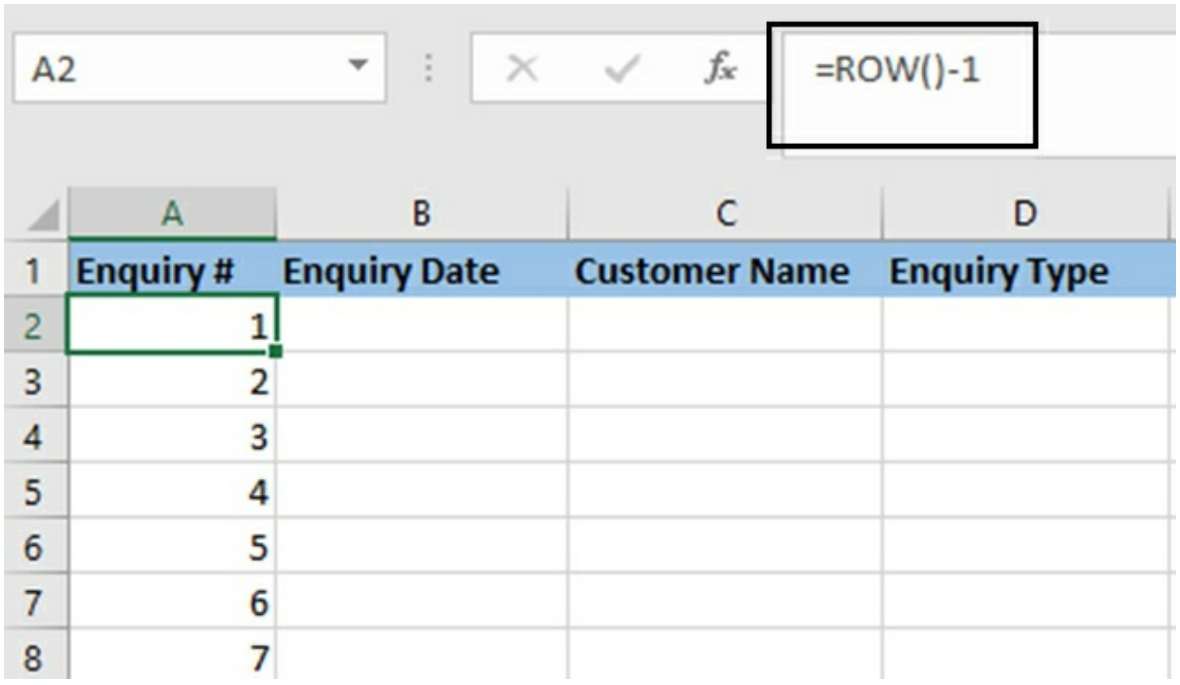
13. No. of properties in column F should be a whole number. To set this criterion, follow the steps to open the *Data Validation* dialog box.

14. On the *Settings* tab, for *Allow:*, select *Whole number* from the dropdown. For

Data, select *greater than* and for *Minimum*, enter *1*. Check the *Ignore blank* box. Follow previously mentioned steps to set appropriate input message and error alert, as well as to drag the validation criteria down to as many rows as needed.



15. For column A Inquiry #, enter formula *=ROW()-1* in cell A2, and drag it down to as many rows as needed. The ROW() returns the number of the row in which the formula appears and 1 is subtracted to account for the header row which appears in row 1.



16. As a clean-up, hide column H that contains dropdown values and format and shade cells and columns as needed.

Lesson 2: Analyzing Data with Functions

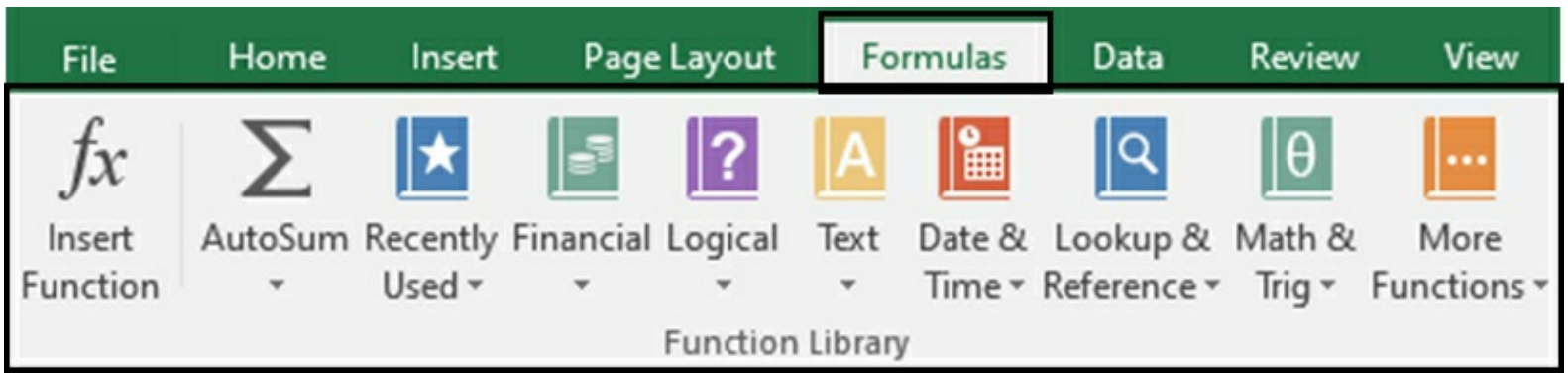
Press the Ctrl key and click here to download practice & solution workbooks. Alternatively, copy and paste below link to your browser:

https://www.dropbox.com/sh/bgyi9plpbc34b7z/AACMDeza4vcgsU4900CgzuNza?dl=0

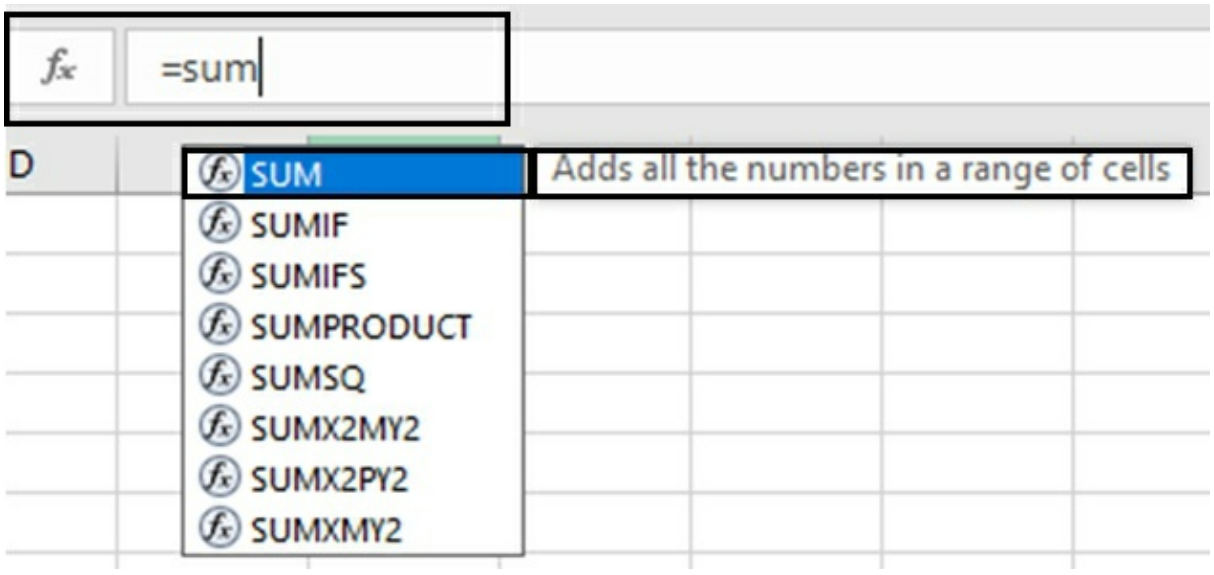
2.1 Basic Functions

Microsoft Excel functions have made data analysis accessible to everyone. Many Excel functions can be used to aggregate and summarize the data to draw key insights from the data that can assist management and executive decision-making. Mastering these Excel functions is extremely important to start using Excel as a data summarization and analysis tool.

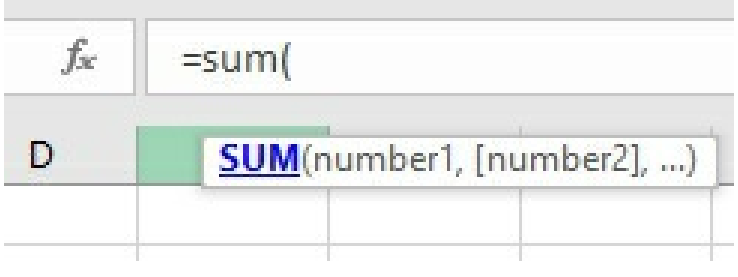
You can find all available Excel functions in the **Function Library** on the *Formulas* tab.



You aren't expected to memorize the name and syntax of hundreds of functions currently available through Excel. As soon as you type the = sign in the formula bar and first letter in the formula, Excel provides a helpful list of functions that start with that letter, as well as a pop-up indicating what the currently highlighted function from that list does.



As soon as you choose the function of interest or type the function's full name, Excel also prompts the function's syntax and arguments. Further help on the function is available by clicking on the name of function in the prompt.



Below we will discuss simple yet helpful Excel functions that are a necessary skill for everyone who wishes to start deriving insights from an Excel spreadsheet.

For the *Practice* sections, the dataset in the *Basic Functions_Practice*.xlsx will be used. The dataset on the *Basic Functions* sheet contains a number of Audi A5 and A6, and BMW 4 Series Gran Coupe and 5 series cars sold every quarter in 2016 and 2017.

SUM

The SUM function performs the basic arithmetic operation of addition. It adds all the numbers in a given range of cells.

Syntax: SUM(number1, [number2], …)

In the syntax of all Excel functions, an argument enclosed in [square brackets] is optional, other arguments are required. It implies that your formula should include at least 1 number, reference to a cell or a range of cells.

Practice: To calculate the total number of Audi and BMW cars sold over all four quarters of all years, type formula =SUM(E2:E33) in cell I2 and press enter.

SUMIF

The SUMIF function performs addition based on one criterion. It adds values in a range that meets user-specified criteria.

Syntax: SUMIF(range, criteria, [sum_range])

Practice: To calculate the total number of Audi and BMW cars sold in the first quarter of all years, type formula =SUMIF(D2:D33,"Q1",E2:E33) in cell I3 and press enter.

In the *range* argument, the range of cells to be tested against the criteria (D2:D33) is entered, in the *criteria* argument - the criteria ("Q1"), and in the *[sum_range]* argument - the cells to be summed (E2:E33). Note that *[sum_range]* argument is optional. If omitted, Excel will add the cells specified in the *range* argument.

SUMIFS

The SUMIFS function can be used to perform addition based on one or more criteria. It adds the values in a range that meet one or multiple user-specified criteria.

Syntax: SUMIFS(sum_range, criteria_range1, criteria1, [criteria_range2, criteria2], ...)

Practice: To calculate total number of Audi and BMW cars sold in first quarter of 2016, enter formula =SUMIFS(E2:E33,D2:D33,"Q1",C2:C33,"2016") in cell I4.

In the *sum_range* argument, the cells to be summed (E2:E33) are entered, in the *criteria_range1* argument, the range of cells to be tested against the criteria 1 (D2:D33), in the *criteria1* argument - the first criteria ("Q1"), in the *criteria_range2* argument, the range of cells to be tested against the criteria 2 (C2:C33) and in the *criteria2* argument - the second criteria ("2016"). Additional ranges and their respective criteria can be added thereafter.

AVERAGE

The AVERAGE function calculates an average or arithmetic mean. It returns the average of all the numbers in a given range of cells. Average is the sum of all numbers divided by their count. It is the most commonly used summary statistic to summarize a set of observations.

Syntax: AVERAGE(number1, [number2], …)

Practice: To calculate the average number of Audi and BMW cars sold quarterly over all years, type formula =AVERAGE(E2:E33) in cell I6 and press enter.

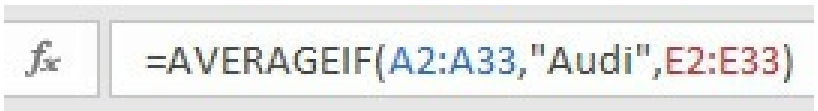
AVERAGEIF

The AVERAGEIF function calculates the average based on one criterion. It returns the average of all the values in a range that meets a user-specified criteria.

Syntax: AVERAGEIF(range, criteria, [average_range])

Practice: To calculate average number of Audi cars sold quarterly, enter formula =AVERAGEIF(A2:A33,"Audi",E2:E33) in cell I7.

In the *range* argument, the range of cells to be tested against the criteria (A2:A33) is entered, in the *criteria* argument - the criteria ("Audi"), and in the optional [average_range] argument - the cells to be averaged (E2:E33).



AVERAGEIFS

The AVERAGEIFS function calculates the average based on one or more criteria. It returns the average of all the values in a range that meet one or multiple user-specified criteria.

Syntax: AVERAGEIFS(average_range, criteria_range1, criteria1, [criteria_range2, criteria2], ...)

Practice: To calculate average number of Audi cars sold quarterly in 2016, enter formula =AVERAGEIFS(E2:E33,A2:A33,"Audi",C2:C33,"2016") in cell I8.

In the *average_range1* argument, the cells to be averaged (E2:E33) are entered, in the *criteria_range1* argument, the range of cells to be tested against the criteria 1 (A2:A33), in the *criteria1* argument - the first criteria ("Audi"), in the *criteria_range2* argument, the range of cells to be tested against the criteria 1 (C2:C33) and in the *criteria2* argument - the second criteria ("2016"). Additional ranges and their respective criteria can be added thereafter.

=AVERAGEIFS(E2:E33,A2:A33,"Audi",C2:C33,"2016")

COUNT

The COUNT function counts the number of cells in a range that contains numbers. It provides a count of all the numbers in a given range of cells.

Syntax: COUNT(value1, [value2], ...)

Practice: To count all cells with data on all cars sold, type formula =COUNT(E2:E33) in cell I10 and press enter.

=COUNT(E2:E33)

COUNTIF

The COUNTIF function counts the number of cells based on one criterion. It returns the count of cells in a range with the values that meet a user-specified criteria.

Syntax: COUNTIF(range, criteria)

Practice: To count all cells with data on Audi cars sold, enter formula =COUNTIF(A2:A33,"Audi") in cell I11.

In the *range* argument, the group of cells you want to count (A2:A33) is entered and in the *criteria* argument - the criteria ("Audi").

=COUNTIF(A2:A33,"Audi")

COUNTIFS

The COUNTIFS function counts the number of cells based on one or more criteria. It returns the count of cells in a range with the values that meet one or multiple user-specified criteria.

Syntax: COUNTIFS(criteria_range1, criteria1, [criteria_range2, criteria2]…)

Practice: To count all cells with data on Audi A5 cars sold, enter formula =COUNTIFS(A2:A33, "Audi",B2:B33, "A5") in cell I12.

In the *criteria_range1* argument, the range of cells to be tested against the criteria 1 (A2:A33), in the *criteria1* argument - the first criteria ("Audi"), in the *criteria_range2* argument, the range of cells to be tested against the criteria 1 (B2:B33) and in the *criteria2* argument - the second criteria ("A5"). Additional ranges and their respective criteria can be added thereafter.

=COUNTIFS(A2:A33, "Audi",B2:B33, "A5")

MEDIAN

The MEDIAN function calculates the median or the number in the middle of the set of given numbers. If you were to sort the data from smallest to largest, the median represents the data point at the middle of the distribution. 50% of the data points are larger than the median and the rest 50% are smaller than the median. If the data contains outliers median is a better measure of summarization compared to average, as outliers, unlike average, do not impact median. Outliers are extreme data points that distort the value of the average. For example, the average of 16, 12, 13, 6, 12 and -50 is 1.5, whereas the median is 12. Median is hence not impacted by the outlier (-50).

Syntax: MEDIAN(value1, [value2], ...)

Practice: To calculate median quarterly car sales of any car make and model, enter formula =MEDIAN(E2:E33) in cell I14.

=MEDIAN(E2:E33)

MAX

The MAX function returns the largest value in a set of values in a given range of cells.

Syntax: MAX(number1, [number2], ...)

Practice: To calculate maximum quarterly car sales of any car make and model, enter formula =MAX(E2:E33) in cell I16.

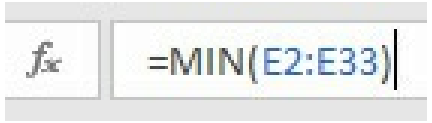
=MAX(E2:E33)

MIN

The MIN function returns the smallest value in a set of values in a given range of cells.

Syntax: MIN(number1, [number2], ...)

Practice: To calculate minimum quarterly car sales of any car make and model, enter formula =MIN(E2:E33) in cell I17.



LARGE

The LARGE function returns the k-th largest value in a set of values in a given range of cells.

Syntax: LARGE(array, k)

Practice: To calculate the second largest quarterly car sales of any car make and model, enter formula =LARGE(E2:E33,2) in cell I19. In the *array* argument, the range of data for which you want to determine the k-th largest value (E2:E33) is entered and in the *k* argument, the position (2 – second from the largest) in the range of data to return.

SMALL

The SMALL function returns the k-th smallest value in a set of values in a given range of cells.

Syntax: SMALL(array, k)

Practice: To calculate second smallest quarterly car sales of any car make and model, enter formula =SMALL(E2:E33,2) in cell I20. In the *array* argument, the range of data for which you want to determine the k-th largest value (E2:E33) is entered and in the *k* argument, the position (2 – second from the smallest) in the range of data to return.

STDDEV.P

The STDDEV.P function returns standard deviation of an entire population. Standard deviation is the measure of the dispersion of data from its average or mean. It measures the absolute variability of a distribution. A high standard deviation means values in the data set tend to be dispersed over a wide range, while a lower standard deviation indicates values close to the mean. While reporting average and standard deviation together, “average ± Standard Deviation” and “mean (Standard Deviation)” notations are commonly used.

Syntax: STDEV.P(number1,[number2],...)

Practice: To calculate the standard deviation of quarterly car sales of any car make and model, enter formula =STDEV.P(E2:E33) in cell I22.

RANK.EQ

The RANK.EQ function returns the rank of a number in a list of numbers.

Syntax: RANK.EQ(number,ref,[order])

Practice: In 2016 Q1, 9,740 Audi A6 cars (cell E10) were sold. To determine the rank of 9,740 in quarterly car sales if the data is ordered ascendingly, enter formula =RANK.EQ(9740,E2:E33,1) in cell I24.

In the *number* argument, the number whose rank needs to be found (9,740) is entered, in the *ref* argument, the range of cells to find the rank from and in the optional *[order]* argument, a number specifying whether to rank descendingly (0) or ascendingly (1).

QUARTILE.INC

The QUARTILE.INC function returns the quartile of a range of cells. Quartiles divide a range of ordered data into three according to size. Generally, the data is arranged from smallest to largest. 25% of the values in ordered data are lower than the first quartile, also known as the 25th percentile. The second quartile, also known as the median or

the 50th percentile, represents the middle of the data; 50% of the data is below this value and 50% above. 75% of the values in ordered data are lower than the third quartile, also known as the 75th percentile. The QUARTILE.INC function, can provide the three quartiles and the minimum and maximum values in the dataset and can be used to create a "five-point summary" of the dataset. This can help you see at a glance where the middle of the data is and how spread out it is.

Syntax: QUARTILE.INC(array,quart)

Practice: To calculate the first quartile (25th percentile) of quarterly car sales, enter the formula QUARTILE.INC(E2:E33,1) in cell I26.

In the *array* argument, the range of numeric values for which the quartile value needed to be determined (E2:E33) is entered and in the *quart* argument, the quart value from 0 to 4 (0 - Minimum value, 1- First quartile, 2 - Median value, 3 - Third quartile, 4 - Maximum value).

PERCENTILE.INC

The PERCENTILE.INC function returns the k-th percentile of a given set of data. The percentile tells you how a value compares to the other values in a given data set. In simple words, when a value is at the k-th percentile, this means that it is higher than k% of the other values in the group. Hence, PERCENTILE.INC function can be used to summarize data beyond quartiles, minimum and maximum.

Syntax: PERCENTILE.INC(array,k)

Practice: To calculate first quartile (25th percentile) of quarterly car sales, enter formula =PERCENTILE.INC(E2:E33,0.25) or =PERCENTILE.INC(E2:E33,25%) in cell I28. In the *array* argument, the range of numeric values for which the percentile value needed to be determined (E2:E33) is entered and in the *k* argument, the percentile value between 0 and 1 (0.25) or 0 and 100% (25%).

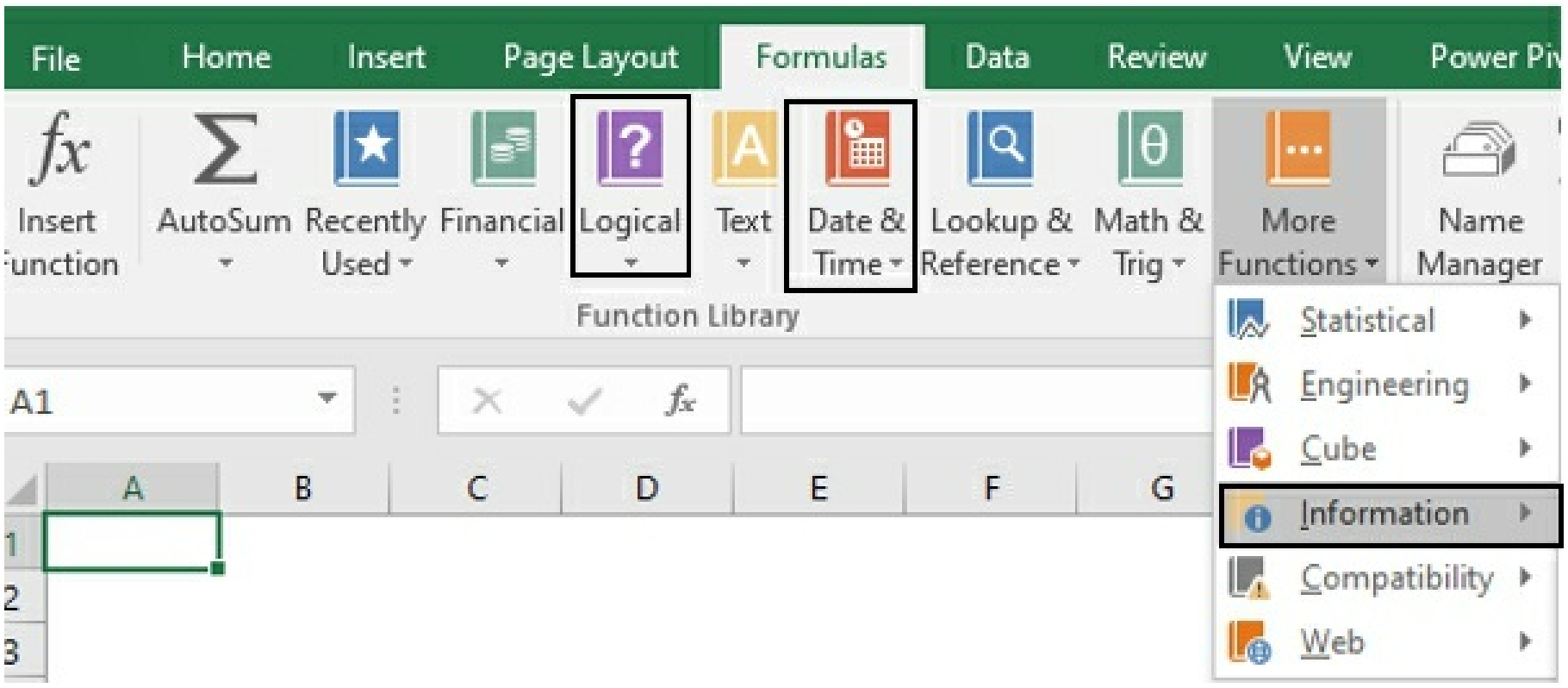
List of basic functions

Function	Description
=SUM()	Adds all the numbers in a range of cells
=SUMIF()	Adds the cells specified by a given condition or criteria
=SUMIFS()	Adds the cells specified by a given set of conditions or criteria
=AVERAGE()	Returns the average (arithmetic mean) of its arguments.
=AVERAGEA()	Returns the average of its arguments, including numbers, text, and logical values
=AVERAGEIF()	Returns the average (arithmetic mean) of all the cells in a range that meet a given criteria
=AVERAGEIFS()	Returns the average (arithmetic mean) of all cells that meet multiple criteria
=COUNT()	Counts the number of cells in a range that contains numbers
=COUNTA()	Counts the number of cells in a range that are not empty
=COUNTBLANK()	Counts the number of blank (empty) cells within a range
=COUNTIF()	Counts the number of cells within a range that meet the given criteria
=COUNTIFS()	Counts the number of cells within a range that meet multiple criteria
=MEDIAN()	Returns the median or the number in the middle of a range of data
=MODE.SNGL()	Returns the most repetitive or frequently occurring value in an array or a range of data
=MAX()	Returns the largest value a range of data
=MIN()	Returns the smallest value a range of data

=LARGE()	Returns the k-th largest value a range of data
=SMALL()	Returns the k-th smallest value a range of data
=STDDEV.S()	Estimates standard deviation based on a sample
=STDDEV.P()	Estimates standard deviation based on the entire population given as arguments
=RANK.EQ()	Returns the rank of a number a range of data. If more than one value has the same rank, the top rank is returned. This is the same as the RANK function.
=QUARTILE.INC()	Returns the quartile of the data set, based on percentile values from 0..1, exclusive. The QUARTILE.INC function is inclusive and has "greater than or equal to" behavior. For example, 25% of the data is greater than or equal to the value of first quartile (25th percentile)
=QUARTILE.EXC()	Returns the quartile of the data set, based on percentile values from 0..1, inclusive. QUARTILE.EXC is exclusive, and has "greater than" behavior. For example, 25% of the data is greater than the value of first quartile (25th percentile)
=PERCENTILE.INC()	Returns the k-th percentile of values in a range

2.2 Logical, Date & Time, and Information Functions

On the *Formulas* tab, Excel built-in functions are organised into groups (text, logical, financial, etc.). In this chapter, we will discuss, commonly used functions included in the *Logical*, *Data & Time* and *Information* groups.



Logical Functions

Excel's logical functions (AND, OR, NOT and XOR) are some of the most popular and useful to draw conclusions from the data. They compare values in different cells against one or more conditions and perform actions dependent upon the outcome of the comparison. These functions return either TRUE or FALSE when their arguments are evaluated.

We will study the four logical functions - AND, OR, NOT and XOR, with two "conditional" functions IF and IFERROR, as in practice, they are very likely to be used in combination. We will work with the employee dataset on *Logical* sheet of *Logical, Information, and Date & Time functions_Practice.xlsx*. The dataset provides information on calls received, answered and resolved by select employees of a call centre, as well as ratings submitted by the customers served after those calls. Let's get started!

AND

The AND function returns TRUE if all of the arguments evaluate to TRUE. Hence, it tests several conditions and returns TRUE only if all of them are met.

Syntax: AND(logical1, [logical2], …)

Practice: To determine if an employee resolved more than 300 calls and also received a rating of greater than 2, enter formula =AND(D3>300,E3>2) in cell F3 and drag down the formula.

In the *logical1* argument, first logical condition – Calls Resolved greater than 300 (D3>300) is entered and in the *logical2* argument, second logical condition Customer Rating greater than 3 (E3>2).

OR

The OR function returns TRUE if any argument evaluates to TRUE. Hence, it tests several conditions and returns TRUE if at least one of them is met.

Syntax: OR(logical1, [logical2], …)

Practice: To determine if an employee either resolved more than 350 calls or received a rating of greater than 3, enter formula =OR(D3>350,E3>3) in cell G3 and drag down the formula.

In the *logical1* argument, first logical condition – Calls Resolved greater than 350 (D3>350) is entered and in the *logical2* argument, second logical condition Customer Rating greater than 3 (E3>3).

NOT

The NOT function returns the reversed logical value of its argument, i.e. if the argument is FALSE, then TRUE is returned and vice versa. This function is not used as extensively as AND and OR functions.

Syntax: NOT(logical1, [logical2], …)

XOR

The XOR function is a logical Exclusive OR function. Like NOT function, it is also not commonly used. The concept of 'Exclusive Or' may be a bit difficult to grasp at first. An XOR function contains two logical statements, returns (1) TRUE if either argument evaluates to TRUE, or (2) FALSE if both arguments are TRUE or neither is TRUE. When more logical statements are added, the XOR function in Excel results in (1) TRUE if an odd number of the arguments evaluate to TRUE, or (2) FALSE if is the total number of TRUE statements is even, or if all statements are FALSE.

Syntax: XOR(logical1, [logical2], …)

IF

The IF function evaluates a certain condition and returns the value or calculation you specify if the condition is TRUE, and another value or calculation if the condition is FALSE.

Syntax: IF(logical_test, [value_if_true], [value_if_false])

where, *logical_test*: The logical condition to check.

value_if_true: The value or calculation to implement if the condition is met, or is true.

value_if_false: The value or calculation to implement if the condition is not met, or is false.

Practice: Suppose you want to send an appreciation email to employees who received a rating of greater than 3 and assess the performance of others. To determine which employees to appreciate or assess, enter formula =IF(E3>3,"Appreciation","Assessment") in cell H3 and drag down the formula.

In the *logical_test* argument, logical condition – Customer Rating greater than 3 (E3>3) is entered, in the *value_if_true*, value to be performed if the condition is met ("Appreciation") and in the *value_if_false*, value to be performed if the condition is not met ("Assessment").

```
fx  =IF(E3>3,"Appreciation","Assessment")
```

IF and AND/OR

Excel AND and OR functions by themselves have limited usage however in combination with other Excel functions, especially IF, these functions provide a significant advantage in analysing data.

Practice: Suppose you want to amend the criteria to appreciate the employees to include the employees who resolved more than 300 calls and received a rating of greater than or equal to 3. To determine which employees deserve an appreciation, enter formula =IF(AND(D3>300,E3>=3),"Appreciation","Assessment") in cell I3.

In the *logical_test* argument, logical condition containing the AND function (AND(D3>300,E3>=3)) is entered, in the *value_if_true*, value to be performed if the condition is met ("Appreciation") and in the *value_if_false*, value to be performed if the condition is not met ("Assessment").

```
fx  =IF(AND(D3>300,E3>=3),"Appreciation","Assessment")
```

Similarly, you can use a combination of IF & OR functions. In case you have to evaluate your data based on several sets of multiple conditions, you can also employ both AND & OR functions in combination with IF at the same time.

NESTED IF

If you need to create more elaborate logical tests for your data, you can include additional IF statements in the *value_if_true* and *value_if_false* arguments of your Excel IF formulas. These multiple IF functions are called nested IF functions and they may prove particularly useful if you want your formula to return 3 or more different results.

Practice: Suppose you take action based on the customer ratings of the employees, and use the below criteria to do so.

Training: Customer Rating of 2 or less (<=2)

Discussion: Customer Rating between 2 and 3 (>2 and <=3)

Appreciate: Customer Rating more than 3 (>3)

To determine what action to take, enter formula =IF(E3<=2,"Training",IF(AND(E3>2, E3<=3),"Discussion","Appreciate")) in cells J3 and drag down the formula. Note that if the employee rating doesn't pass the criteria of 2 or less as well as between 2 and 3, it automatically means the rating is >3.

```
=IF(E3<=2,"Training",IF(AND(E3>2, E3<=3),"Discussion","Appreciate"))
```

Just one nested IF function is sufficient in this case. However, you can nest more IF functions based on your analysis.

IFERROR

The IFERROR function checks if a formula results in an error. If FALSE, it returns the calculation of the formula. If TRUE, it returns another specified value or calculation.

Syntax: =IFERROR(VALUE,value_if_error)

Practice: Suppose you want to calculate the Call Resolution Rate defined as total calls resolved divided by the total calls answered expressed as a percentage. To calculate it, enter the formula (D3/C3)*100 in cell K3 and drag it down. The calculation for the last employee shows an error. Click on cell K8 to investigate the error. The calls answered for the employee are 0 which seems like a data entry error that you would want to flag. To display the cells with formulas resulting in an error as “CHECK”, revise the formula in cell K3 to =IFERROR((D3/C3)*100,"CHECK") and drag down. The error is now displaying as “CHECK”.

```
=IFERROR((D3/C3)*100,"CHECK")
```

Date and Time functions

The data obtained from different sources might contain date values. In this section, you will understand how to prepare data that contains date and time values for analysis and calculations.

For Practice sections, we will employ the *Date & Time* sheet of *Logical and Date & Time Functions_Practice.xlsx*. Let’s get started!

DATE

If your data has Year, Month and Day in separate columns, you will need to combine them in a date format to perform any calculations. The DATE function can be used for this. It returns a date from the year, month, and day.

Syntax: DATE(year, month, day)

Practice: To calculate the date from specified year month and day, enter formula =DATE(B3,C3,D3) in cell E3 on the *Date & Time* sheet.

```
=DATE(B3,C3,D3)
```

This calculates a serial number of the date. Excel stores date as sequential serial numbers so that they can be used in calculations. To convert the serial number to a date, go to the *Home* tab and click on the dropdown under the *Number* group. Now, change the format of the cell from “General” to “Short Date”.

DATEDIF

The DATEDIF function returns the number of days, months or years between two dates.

Syntax: =DATEDIF(start_date, end_date, unit)

Practice: To calculate the number of days between the specified start date and end date, enter formula =DATEDIF(B7,C7,"d") in cell D7.Unit can be specified as “y” to calculate the number of years between two dates and as “m” to calculate the number of months.

```
=DATEDIF(B7,C7,"d")
```

NETWORKDAYS

The NETWORKDAYS function returns the number of working days between two dates. Working days include Mon-Fri and exclude the weekends.

Syntax: =NETWORKDAYS(start_date,end_date,[holidays])

Practice: To calculate the number of working days between the specified start date and end date, enter formula =NETWORKDAYS(B7,C7) in cell E7.

```
=NETWORKDAYS(B7,C7)
```

DATEVALUE

Issues with date formatting are very common when data is imported into Excel from outside sources. The imported date entries appear to be in a correct date format, but they don't behave like dates as Excel stores them as text. Such date entries are left-aligned in a cell by default and usually have a leading apostrophe, visible in the formula bar on clicking the cell.

The DATEVALUE function is used to convert such date entries stored as text into the date's corresponding serial number.

Syntax: DATEVALUE(date_text)

Practice: Click on cell B11. The date in the cell though appears in a date format is being stored as text. This is confirmed by the left-alignment of the date in the cell and the leading apostrophe (') visible in the formula bar.

To convert it into a serial format, enter formula =DATEVALUE(B11) in cell C11 and change the format of the cell to short date as described previously.

DAY, MONTH, YEAR

Extracting day, month and year from the date is sometimes required to perform calculations, draw graphs, create PivotTables, etc. The DAY function is used to return the day as a number (1-31) from a date, the MONTH function returns the month as a number (1-12) from a date and the YEAR function returns the year of a date.

Syntax: =DAY(serial_number); MONTH(serial_number); YEAR(serial_number)

Practice: Enter the below formulas in cells C15, D15 and E15 respectively to determine the day, month and year of the given date specified in Excel serial number format in cell B15.

=DAY(B15)

=MONTH(B15)

=YEAR(B15)

WEEKDAY, WEEKNUM

The WEEKDAY function returns the day of the week as a number (1-7), and the WEEKNUM function returns the week number in a year (1-52).

Syntax: =WEEKDAY(serial_number,[return_type]; =WEEKNUM(serial_number,[return_type])

Practice: Enter below formulas in cells F15 and G15 respectively to determine weekday and week number of the date specified in Excel serial number format in cell B15.

=WEEKDAY(B15)

=WEEKNUM(B15)

Please note that *return_type* argument is optional and specifies first day of the week. 1 is default for Sunday, 2 for Monday, and so on.

TODAY

If you need to perform calculations based on today's date, you can calculate it using TODAY function. It returns the current date formatted as a date. TODAY function can be used in combination with DATEDIF and NETWORKDAYS functions to calculate the difference in days of a specific day from today's date.

Syntax: TODAY()

Practice: Enter formula =TODAY() in cell B18 to return today's date.

TIME

If hours, minutes and seconds are displayed on separate columns in your data, the TIME function can be used to combine them as Excel decimal numbers and then format them as time.

Syntax: TIME(hour,minute,second)

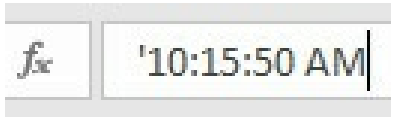
Practice: To calculate the date from specified hours, minutes and seconds, enter formula =TIME(B22,C22,D22) in cell E22. This calculates Excel decimal number of the time. Excel stores time as decimal numbers to make the calculations based on them easier. To convert the decimal number to time, go to the *Home* tab and click on the dropdown under the *Number* group. Now, change the format of the cell from "General" to "Time".

TIMEVALUE

The TIMEVALUE Function converts a time stored as text into the time's corresponding decimal number.

Syntax: =TIMEVALUE(time_text)

Practice: To convert left-aligned time value stored as text in cell B26, enter formula =TIMEVALUE(B26) in cell C26 and change the format of the cell to time as described previously.



HOUR, MINUTE, SECOND

The HOUR function returns the hour as a number (0-23), the MINUTE function returns the minute as a number (0-59) and the SECOND function returns the second as a number (0-59).

Syntax: =HOUR(serial_number; =MINUTE(serial_number; =SECOND(serial_number)

Practice: Enter below formulas in cells C15, D15 and E15 respectively to determine day, month and year of the specified in Excel serial number format in cell B30.

=HOUR(B30)

=MINUTE (B30)

= SECOND (B30)

NOW

The NOW function returns the current date and time.

Syntax: NOW()

Practice: Enter formula =NOW() in cells B33 to return the current time and date.

Information functions

Information functions are used to get information about the worksheets and their contents, for example, whether a cell has an error, the type of value in a cell, the current row number, etc. Here is a list of the frequently used information formulas in Excel.

	Function	Description	Syntax
	Error Information		
1.	ISERROR	Tests if cell value is an error. Returns TRUE or FALSE.	=ISERROR(value)
2.	ISNA	Tests if cell value is #N/A. Returns TRUE or FALSE.	=ISNA(value)
	Cell Content Information		
3.	ISNUMBER	Tests if cell contains a number. Returns TRUE or FALSE.	=ISNUMBER(value)
4.	ISODD	Tests if cell contains an odd number. Returns TRUE or FALSE.	=ISODD(number)
5.	ISEVEN	Tests if cell contains an even number. Returns TRUE or FALSE.	=ISEVEN(number)
6.	ISTEXT	Tests if cell contains text. Returns TRUE or FALSE.	=ISTEXT(value)
7.	ISBLANK	Tests if cell is blank. Returns TRUE or FALSE.	=ISBLANK(value)
8.	ISFORMULA	Tests if cell contains a formula. Returns TRUE or FALSE.	=ISFORMULA(reference)
	Row and Column Information		
9.	ROW	Returns the row number of a cell reference.	=ROW(reference)
10.	COLUMN	Returns the column number of a cell reference.	=COLUMN(reference)

2.3 Cell References

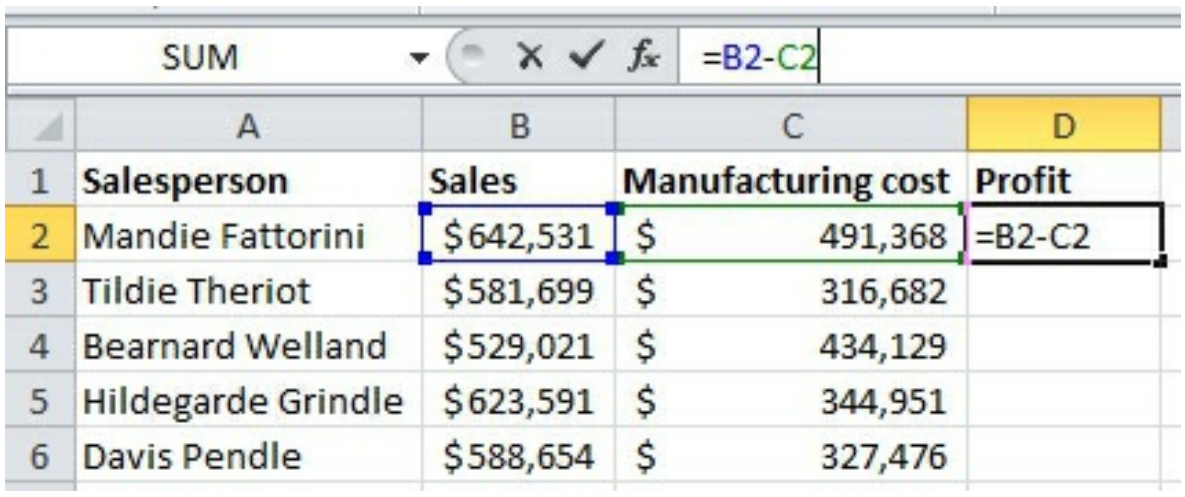
Before we move forward with this lesson, it will be worthwhile to study cell references used in Excel formulas. Excel mainly uses three types of cell references to create formulas – relative, absolute and mixed. They behave differently when copied. We will discuss them one by one.

Relative cell references

Relative cell references are the default cell references in formulas. When copied, they automatically change based on the relative position of rows and columns. Let's understand what this means with the help of a practice exercise.

Practice

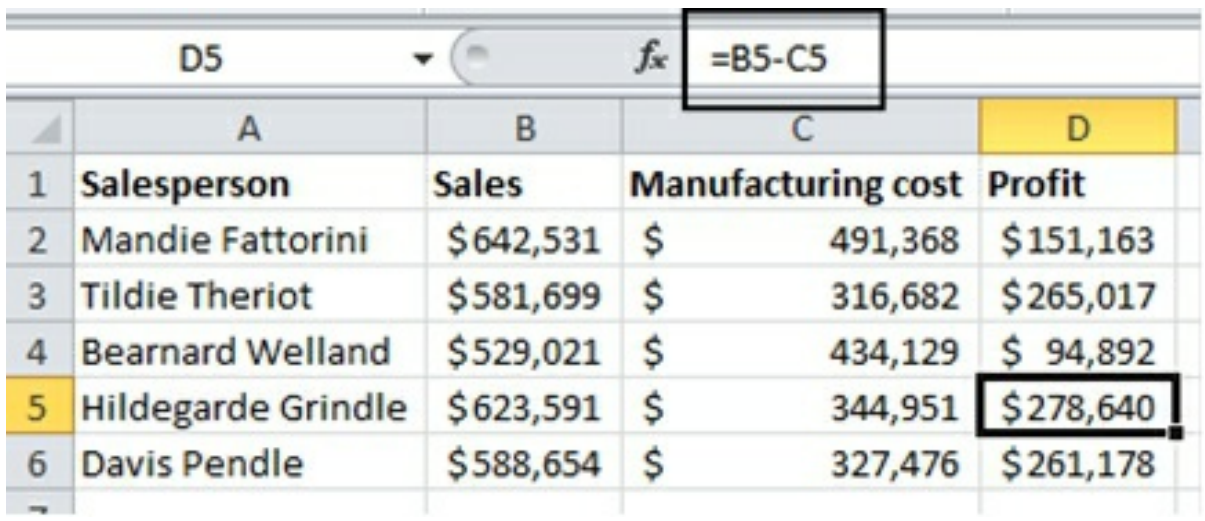
1. Open *Cell References_Practice.xlsx* and go to the *Relative* sheet. The tab contains data on 5 sales personnel of a company. The data consists of sales made by these personnel and the manufacturing costs associated with those sales, as well as an empty *Profit* column. Suppose we want to calculate profit generated by the employees.
2. *In cell D2 under the Profit header, enter formula =B2-C2 and hit the enter key. The formula subtracts expenditure costs for the first employee from the sales to calculate the profit generated by the employee.*



SUM | =B2-C2

	A	B	C	D
1	Salesperson	Sales	Manufacturing cost	Profit
2	Mandie Fattorini	$642,531	$ 491,368	=B2-C2
3	Tildie Theriot	$581,699	$ 316,682	
4	Bearnard Welland	$529,021	$ 434,129	
5	Hildegarde Grindle	$623,591	$ 344,951	
6	Davis Pendle	$588,654	$ 327,476	

3. Instead of creating a new formula for each employee, we can copy this formula to the other rows. To do this, select cell D2 and hover the cursor over to its bottom right until a + sign (drag handle) appears. Double click on the drag handle to copy the formula down.
4. The formula will be copied to the rest of the rows containing data with cell references relative to that row. To check, double-click on any other cell in the *Profit* column. The relative cell references should be different for the cell depending on cell's row.



D5 | =B5-C5

	A	B	C	D
1	Salesperson	Sales	Manufacturing cost	Profit
2	Mandie Fattorini	$642,531	$ 491,368	$151,163
3	Tildie Theriot	$581,699	$ 316,682	$265,017
4	Bearnard Welland	$529,021	$ 434,129	$ 94,892
5	Hildegarde Grindle	$623,591	$ 344,951	$278,640
6	Davis Pendle	$588,654	$ 327,476	$261,178

Absolute cell references

Absolute cell references are used when a cell reference in a formula is intended to remain constant. Unlike relative references, absolute references do not change when copied. An absolute reference is designated in a formula by inserting a dollar sign ($) before the column and row. As an alternative, you can enter the formula and press the F4 key on your keyboard to make an absolute reference. Let's practice!

Practice

1. Continue working on *Cell References_Practice.xlsx* and go to the *Absolute* sheet. As you'll notice, we will work with the same dataset as the previous exercise. The dataset now has the profit column filled and an additional, empty Bonus column. It also has a cell indicating bonus percentage which is set at

10% of the profit generated by each of the employees. The aim is to calculate bonus for each of the employees.

2. In cell E2 under the *Bonus* header, enter formula =$G$3*D2 and hit enter. The formula calculates the bonus portion corresponding to 10% of the profit generated by each of the employees. Don't forget that instead of typing $ sings, the cell G3 can be converted into an absolute reference by pressing the F4 key after typing.

SUM | =$G$3*D2

	A	B	C	D	E	F	G
1	Salesperson	Sales	Manufacturing cost	Profit	Bonus		
2	Mandie Fattorini	$642,531	$ 491,368	$151,163	=$G$3*D2		Bonus
3	Tildie Theriot	$581,699	$ 316,682	$265,017			10%
4	Bearnard Welland	$529,021	$ 434,129	$ 94,892			
5	Hildegarde Grindle	$623,591	$ 344,951	$278,640			
6	Davis Pendle	$588,654	$ 327,476	$261,178			

3. Now, select cell E2, hover the cursor over to its bottom right until the drag handle appears and double click on it to copy the formula down.
4. The formula will be copied to the rest of the rows containing data with absolute referencing to the cell containing the constant bonus percentage. To check, double-click on any other cell in the *Bonus* column. Note that the absolute reference to G3 which holds the bonus percentage does not change, while the reference to profit changes with each new row.

E5 | =$G$3*D5

	A	B	C	D	E	F	G
1	Salesperson	Sales	Manufacturing cost	Profit	Bonus		
2	Mandie Fattorini	$642,531	$ 491,368	$151,163	$ 15,116		Bonus
3	Tildie Theriot	$581,699	$ 316,682	$265,017	$ 26,502		10%
4	Bearnard Welland	$529,021	$ 434,129	$ 94,892	$ 9,489		
5	Hildegarde Grindle	$623,591	$ 344,951	$278,640	$ 27,864		
6	Davis Pendle	$588,654	$ 327,476	$261,178	$ 26,118		

Mixed cell references

A mixed reference has both absolute and relative components. See below for examples:

$A1 - Allows the row reference to change, but locks the column reference.

A$1 - Allows the column reference to change, but locks the row reference.

Absolute references are used to keep a row or a column constant. The $ sign precedes either the column or the row reference making it absolute. Similarly, relative references are used to keep a row or a column absolute. To make a cell reference mixed, enter the cell reference and keep pressing the F4 key on your keyboard to switch from relative to absolute to a desired mixed cell reference.

Practice

1. Continue working on *Cell References_Practice.xlsx* and go to the *Mixed* sheet. We are continuing with the same dataset. Now we want to calculate bonuses set at 5%, 10% and 12% of the profit generated by each of the employees to decide optimal bonus to hand over that will agree with our company budget. We will calculate the bonus at different percentages and see what percentage makes sense.
2. In cell E3 under the *Bonus 1* header, enter formula =$D3*E$2 and hit enter. Don't forget that the cell D3 can be converted into a mixed reference with column locked ($D3) reference by entering the cell address and pressing the F4 key twice. Similarly, cell E3 can be converted into a mixed reference with row locked (D$3) by entering the cell address and pressing the F4 key three times.

SUM =$D3*E$2

	A	B	C	D	E	F	G
1					Bonus 1	Bonus 2	Bonus 3
2	Salesperson	Sales	Manufacturing cost	Profit	5%	10%	12%
3	Mandie Fattorini	$642,531	$ 491,368	$151,163	=$D3*E$2		
4	Tildie Theriot	$581,699	$ 316,682	$265,017			
5	Bearnard Welland	$529,021	$ 434,129	$ 94,892			
6	Hildegarde Grindle	$623,591	$ 344,951	$278,640			
7	Davis Pendle	$588,654	$ 327,476	$261,178			

3. This formula is carefully constructed with two mixed references so that it can be copied across the range E3:G7 without manual changes. Select cell E3, hover the cursor over to its bottom right until the drag handle appears and drag the handle to the right up to cell G3 to copy the formula.

E3 =$D3*E$2

	A	B	C	D	E	F	G
1					Bonus 1	Bonus 2	Bonus 3
2	Salesperson	Sales	Manufacturing cost	Profit	5%	10%	12%
3	Mandie Fattorini	$642,531	$ 491,368	$151,163	$ 7,558	$ 15,116	$ 18,140
4	Tildie Theriot	$581,699	$ 316,682	$265,017			
5	Bearnard Welland	$529,021	$ 434,129	$ 94,892			
6	Hildegarde Grindle	$623,591	$ 344,951	$278,640			
7	Davis Pendle	$588,654	$ 327,476	$261,178			

4. Explore the newly filled cells. For cells in row 3, the reference to $D3 has the column locked to make sure the formula continues to refer to profit from column D as it is copied to the right, but the row reference to different bonus percentages changes dynamically.

SUM =$D3*F$2

	A	B	C	D	E	F	G
1					Bonus 1	Bonus 2	Bonus 3
2	Salesperson	Sales	Manufacturing cost	Profit	5%	10%	12%
3	Mandie Fattorini	$642,531	$ 491,368	$151,163	$ 7,558	=$D3*F$2	$ 18,140
4	Tildie Theriot	$581,699	$ 316,682	$265,017			
5	Bearnard Welland	$529,021	$ 434,129	$ 94,892			
6	Hildegarde Grindle	$623,591	$ 344,951	$278,640			
7	Davis Pendle	$588,654	$ 327,476	$261,178			

5. Now select range E3:G3, hover the cursor over to its bottom right until the drag handle appears and drag the cursor down to fill the remaining rows.

E3 =$D3*E$2

	A	B	C	D	E	F	G
1					Bonus 1	Bonus 2	Bonus 3
2	Salesperson	Sales	Manufacturing cost	Profit	5%	10%	12%
3	Mandie Fattorini	$642,531	$ 491,368	$151,163	$ 7,558	$15,116.30	$ 18,140
4	Tildie Theriot	$581,699	$ 316,682	$265,017	$ 13,251	$26,501.70	$ 31,802
5	Bearnard Welland	$529,021	$ 434,129	$ 94,892	$ 4,745	$ 9,489.20	$ 11,387
6	Hildegarde Grindle	$623,591	$ 344,951	$278,640	$ 13,932	$27,864.00	$ 33,437
7	Davis Pendle	$588,654	$ 327,476	$261,178	$ 13,059	$26,117.80	$ 31,341

6. Explore the newly filled cells. For cells in column E, the reference to E$2 has the row locked so that as the formula is copied down, it will continue to refer to the bonus percentage values in row 2 but the column reference to profits changes dynamically.

SUM =$D6*E$2

	A	B	C	D	E	F	G
1					Bonus 1	Bonus 2	Bonus 3
2	Salesperson	Sales	Manufacturing cost	Profit	5%	10%	12%
3	Mandie Fattorini	$642,531	$ 491,368	$151,163	$ 7,558	$15,116.30	$ 18,140
4	Tildie Theriot	$581,699	$ 316,682	$265,017	$ 13,251	$26,501.70	$ 31,802
5	Bearnard Welland	$529,021	$ 434,129	$ 94,892	$ 4,745	$ 9,489.20	$ 11,387
6	Hildegarde Grindle	$623,591	$ 344,951	$278,640	=$D6*E$2	$27,864.00	$ 33,437
7	Davis Pendle	$588,654	$ 327,476	$261,178	$ 13,059	$26,117.80	$ 31,341

2.4 Lookup Functions

In this chapter, we will study four advanced Excel functions – VLOOKUP, HLOOKUP, INDEX and MATCH. These functions are a great tool in data analysis. VLOOKUP and HLOOKUP are used to find values in a table or range of data, INDEX is used to obtain a value or the reference to a value from within a table or range, and MATCH is used to obtain the relative position of a specified item in a range of cells. These functions can also be combined to get the desired results based on the aim of the analysis. Towards the end, we will also discuss array formulas that can perform multiple calculations on one or several items simultaneously.

Let's get started!

VLOOKUP

VLOOKUP stands for vertical lookup. The VLOOKUP function searches for a specified value in the leftmost column of a given range or table and returns a value of a specified number of columns to the right from the found value.

Syntax: =VLOOKUP(lookup_value,table_array,col_index_num,[range_lookup])

Arguments:

- *lookup_value* – The value you want to look up.
- *table_array* – The range that contains both *lookup_value* and the value you want the vlookup to return. The column containing the *lookup_value* must be the left-most column in the range.
- *col_index_num* – The column number of the *table_array* from which you want to return a value.
- *range_lookup* – TRUE or FALSE. FALSE searches for an exact match. TRUE searches for the nearest match that is less than or equal to the *lookup_value*. If TRUE, the left-most column of the *table_array* must be sorted ascendingly for the vlookup to function properly.

Using VLOOKUP for exact matches

The argument *range_lookup* is set to FALSE if an exact match with the *lookup_value* needs to be returned.

Practice:

1. Open *Lookup Functions_Practice.xlsx* and go to the *VLOOKUP* sheet. We are working with employee dataset that provides information on employee ID, name, title, department and salary of select employees of a company. Suppose you want to automate the display of the salary (in cell H6) of any employee ID you enter into cell H5.
2. Click on cell H6 and enter the below formula into the cell.

=VLOOKUP(H5,$A$1:$E$29,5,FALSE)

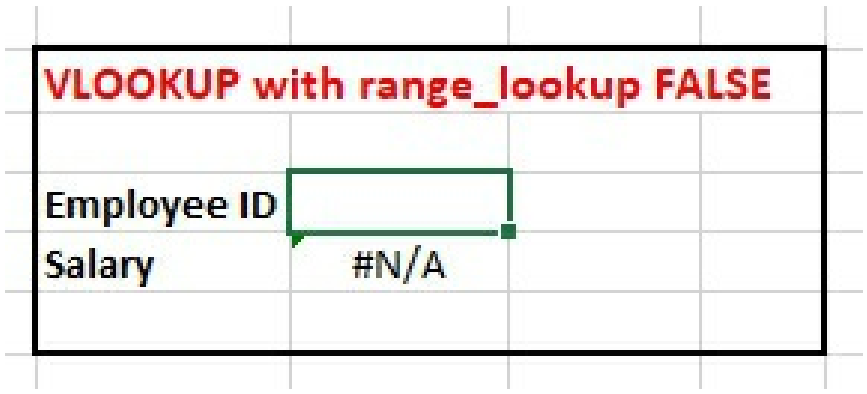
A brief description of each of the arguments is below.

- In the *lookup_value* argument, a cell reference for the value that needs to be searched is entered. In this instance, it is H5, the cell designated to enter employee ID for which salary needs to be searched.
- In the *table_array* argument, the lookup table needed to be searched is entered. In this case, it is $A$1:$E$29. The first column containing the *lookup_value* (Employee ID) is in column 1. When working in a spreadsheet, by default, cell references are relative and adjust as the formula is dragged down through the rows. However, when working with the VLOOKUP functions, we want the lookup table to stay consistent. Hence, we use absolute references by placing dollar signs ($) before the column as well as row reference. This can be alternatively done by placing the cursor in or next to the reference and pressing F4 key.
- In the *col_index_num* argument, column number in the *table_array* or vlookup table that contains the return value is entered. In this case, the column reference would be 5 as the employee's salary column is column 5 in the lookup table.
- The *range_lookup* argument is used to indicate whether you want the exact or approximate match of the return value. In this case, we want the exact match (FALSE).

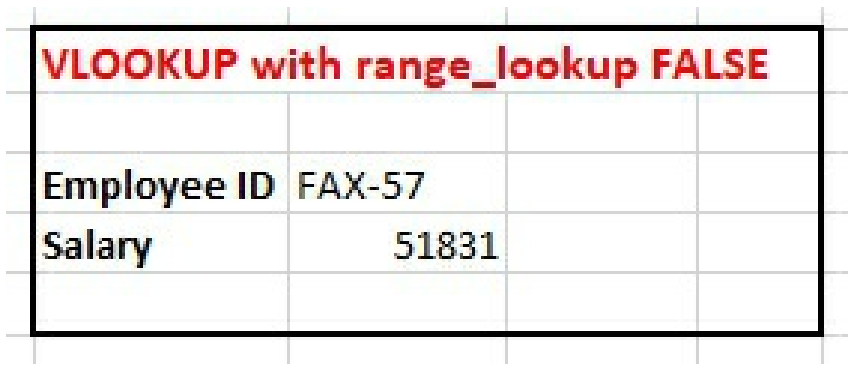
SUM | =VLOOKUP(H5,$A$1:$E$29,5,FALSE)

	A	B	C	D	E	F	G	H	I	J
1	**Employee ID**	**Name**	**Title**	**Department**	**Salary**					
2	QJA-07	Evie Base	Accounting Assistant I	Accounting	$109,605					
3	HMO-83	Jorrie MacTrustram	Tax Accountant	Accounting	$ 95,332		**VLOOKUP with range_lookup FALSE**			
4	IQZ-97	Clio Pitson	Community Outreach Specialist	Marketing	$ 70,726					
5	XTT-54	Lina Jacob	Information Systems Manager	IT	$ 94,406		**Employee ID**			
6	CIG-82	Lauren Pollak	Programmer Analyst IV	IT	$100,424		**Salary**	5,FALSE)		
7	AZW-23	Elizabeth Lapides	Speech Pathologist	Operations	$ 87,053					
8	FAX-57	Hadlee Hounihan	Design Engineer	IT	$ 51,831					
9	GXM-51	Junie Klimushev	Accounting Assistant I	Accounting	$104,663					
10	QPH-99	Cindy Breznovic	Speech Pathologist	Operations	$101,465					
11	TSN-47	Niccolo Scrymgeour	Web Designer II	IT	$ 61,136					
12	EIU-27	Waly Scogin	Environmental Specialist	Operations	$ 47,888					
13	APC-45	Jayme Sambrok	Sales Associate	Sales	$ 83,806					
14	GZH-10	Skelly Dignall	Administrative Assistant III	Operations	$ 69,125					
15	SDN-09	Carolus Adame	Analyst Programmer	IT	$105,932					
16	YOZ-39	Bendix Ventum	Cost Accountant	Accounting	$111,417					
17	JII-95	Valentina Sutheran	Physical Therapy Assistant	Operations	$102,695					
18	VSD-61	Gratiana Zapatero	Design Engineer	Engineering	$113,671					
19	LBJ-50	Paulie O'Carmody	Senior Editor	Operations	$ 85,182					
20	WCC-51	Trumaine Andres	Senior Cost Accountant	Accounting	$ 43,942					
21	VNY-84	Brenden Peskett	Human Resources Assistant I	Human Resources	$107,285					
22	HBE-39	Carling Durrans	Research Nurse	Operations	$ 81,899					
23	MNL-93	Benn Le Lievre	Payment Adjustment Coordinator	Accounting	$ 84,469					
24	APS-61	Gerri Pavolillo	Senior Sales Associate	Sales	$106,024					
25	MTX-57	Shantee Durnin	Senior Financial Analyst	Accounting	$104,051					
26	MOK-62	Dee Ivchenko	Biostatistician IV	Operations	$ 51,581					
27	LPN-62	Jess Bradman	Administrative Assistant III	Human Resources	$ 55,810					
28	UGP-19	Carina Poge	Financial Analyst	Accounting	$ 72,707					
29	ZTH-55	Margery Boom	Structural Engineer	Engineering	$ 59,847					

3. Press enter. You will get an error message as cell H5 is currently empty.



4. Now, enter any employee ID in H5. The corresponding salary will be returned in cell H6.



In brief, the VLOOKUP formula searched the column 1 of the vlookup table (Employee ID) until the given employee ID is found and then returns the corresponding value in column 5 (Salary) of the same row.

Using VLOOKUP for approximate matches

The argument *range_lookup* is set to TRUE if an approximate match with the *lookup_value* needs to be returned.

Practice

1. Continue working on *VLOOKUP* sheet of the *Lookup Functions_Practice.xlsx*. Suppose you want to automate display of bonus (in cell H22) of any approximate salary value you enter into cell H21 based on a lookup table that relates salary to bonus.
2. Click on cell H22 and enter below formula into the cell.

=VLOOKUP(H21,$H$16:$I$19,2,TRUE)

A brief description of each of the arguments is below.

- The *lookup_value* argument is H21, the cell designated to enter salary for which bonus needs to be searched.
- In the *table_array* argument, the lookup table needed to be searched is entered. In this case, it is

$H$16:$I$19. The first column contains the *lookup_value* (Employee Salary). Note that it is sorted in ascending order so that you can get an approximate match. In case of an approximate match, a value corresponding to the largest value that is less than the *lookup_value* is returned. Any salary that is >= 0 and < 50,000 will return a bonus amount of 3%. And salary >= 50,000 and < 100,000 will return 5%. And so on.

- In the *col_index_num* argument, column number in the *table_array* or lookup table that contains the return value is entered. In this case, it is 2 as a bonus is stored in column 2 in the lookup table.
- The *range_lookup* argument is used to indicate whether you want the exact or approximate match of the return value. In this case, we want an approximate match and the argument is set as TRUE.

SUM =VLOOKUP(H21,$H$16:$I$19,2,TRUE)

	A	B	C	D	E	F	G	H	I	J
1	Employee ID	Name	Title	Department	Salary					
2	QIA-07	Evie Base	Accounting Assistant I	Accounting	$109,605					
3	HMO-83	Jorrie MacTrustram	Tax Accountant	Accounting	$ 95,332		VLOOKUP with range_lookup FALSE			
4	IQZ-97	Clio Pitson	Community Outreach Specialist	Marketing	$ 70,726					
5	XTT-54	Lina Jacob	Information Systems Manager	IT	$ 94,406		Employee ID	FAX-57		
6	CIG-82	Lauren Pollak	Programmer Analyst IV	IT	$100,424		Salary	51831		
7	AZW-23	Elizabeth Lapides	Speech Pathologist	Operations	$ 87,053					
8	FAX-57	Hadlee Hounihan	Design Engineer	IT	$ 51,831					
9	GXM-51	Junie Klimushev	Accounting Assistant I	Accounting	$104,663					
10	QPH-99	Cindy Breznovic	Speech Pathologist	Operations	$101,465					
11	TSN-47	Niccolo Scrymgeour	Web Designer II	IT	$ 61,136					
12	EIU-27	Waly Scogin	Environmental Specialist	Operations	$ 47,888					
13	APC-45	Jayme Sambrok	Sales Associate	Sales	$ 83,806		VLOOKUP with range_lookup TRUE			
14	GZH-10	Skelly Dignall	Administrative Assistant III	Operations	$ 69,125					
15	SDN-09	Carolus Adame	Analyst Programmer	IT	$105,932			Salary	Bonus	
16	YOZ-39	Bendix Ventum	Cost Accountant	Accounting	$111,417			0	3%	
17	JII-95	Valentina Sutheran	Physical Therapy Assistant	Operations	$102,695			50,000	5%	
18	VSD-61	Gratiana Zapatero	Design Engineer	Engineering	$113,671			100,000	6%	
19	LBJ-50	Paulie O'Carmody	Senior Editor	Operations	$ 85,182			150,000	8%	
20	WCC-51	Trumaine Andres	Senior Cost Accountant	Accounting	$ 43,942					
21	VNY-84	Brenden Peskett	Human Resources Assistant I	Human Resources	$107,285		Salary	120,000		
22	HBE-39	Carling Durrans	Research Nurse	Operations	$ 81,899		Bonus	,2,TRUE)		
23	MNL-93	Benn Le Lievre	Payment Adjustment Coordinator	Accounting	$ 84,469					
24	APS-61	Gerri Pavolillo	Senior Sales Associate	Sales	$106,024					
25	MTX-57	Shantee Durnin	Senior Financial Analyst	Accounting	$104,051					
26	MOK-62	Dee Ivchenko	Biostatistician IV	Operations	$ 51,581					
27	LPN-62	Jess Bradman	Administrative Assistant III	Human Resources	$ 55,810					
28	UGP-19	Carina Poge	Financial Analyst	Accounting	$ 72,707					
29	ZTH-55	Margery Boom	Structural Engineer	Engineering	$ 59,847					

3. Press enter. You will get an error message as cell H21 is currently empty.
4. Now, enter any salary in H21. The corresponding bonus will be returned in cell H22.

VLOOKUP with range_lookup TRUE

	Salary	Bonus
	0	3%
	50,000	5%
	100,000	6%
	150,000	8%
Salary	120,000	
Bonus	6%	

In brief, the VLOOKUP formula searches the column 1 of the vlookup table (Salary) and returns the value corresponding to the largest value that is less than the *lookup_value*. For example, as the salary of 120,000 is >100,000 and < 150,000, 6% corresponding to 100,000 is returned.

Using VLOOKUP to merge two tables

VLOOKUP can also be used to merge two tables. It is in fact an important application of the VLOOKUP function. To merge tables, you can use the VLOOKUP function to lookup and retrieve data from one table to the other. To use VLOOKUP this way, both tables must share a common column.

In the example below, let's refer to the main table as *Salary Table*, the table on top right as *Department Table* and

the one on bottom right as *Bonus Table*. VLOOKUP with exact match is used to pull Department ID into column F of the *Salary Table* from the *Department Table* based on the *Department* column which is common between the two tables.

Also, VLOOKUP with approximate match is used to calculate bonus in column G of the *Salary Table* from the *Bonus Table* based on the *Salary* column which is common between the two tables.

	A	B	C	D	E	F	G	H	I	J
1	Employee ID	Name	Title	Department	Salary	Department ID	Bonus (%)			
2	QIA-07	Evie Base	Accounting Assistant I	Accounting	109605	=VLOOKUP(D2,$I$6:$J$12,2,FALSE	=VLOOKUP(E2,$I$18:$J$21,2,TRUE)			
3	HMO-83	Jorrie MacTrustram	Tax Accountant	Accounting	95332					
4	IQZ-97	Clio Pitson	Community Outreach Specialist	Marketing	70726					
5	XTT-54	Lina Jacob	Information Systems Manager	IT	94406				Department	Department ID
6	CIG-82	Lauren Pollak	Programmer Analyst IV	IT	100424				Accounting	AX
7	AZW-23	Elizabeth Lapides	Speech Pathologist	Operations	87053				Engineering	EX
8	FAX-57	Hadlee Hounihan	Design Engineer	IT	51831				Human Resources	HX
9	GXM-51	Junie Klimushev	Accounting Assistant I	Accounting	104663				IT	IX
10	QPH-99	Cindy Breznovic	Speech Pathologist	Operations	101465				Marketing	MX
11	TSN-47	Niccolo Scrymgeour	Web Designer II	IT	61136				Operations	OX
12	EIU-27	Waly Scogin	Environmental Specialist	Operations	47888				Sales	SX
13	APC-45	Jayme Sambrok	Sales Associate	Sales	83806					
14	GZH-10	Skelly Dignall	Administrative Assistant III	Operations	69125					
15	SDN-09	Carolus Adame	Analyst Programmer	IT	105932					
16	YOZ-39	Bendix Ventum	Cost Accountant	Accounting	111417					
17	JII-95	Valentina Sutheran	Physical Therapy Assistant	Operations	102695				Salary	Bonus
18	VSD-61	Gratiana Zapatero	Design Engineer	Engineering	113671				0	0.03
19	LBJ-50	Paulie O'Carmody	Senior Editor	Operations	85182				50000	0.05
20	WCC-51	Trumaine Andres	Senior Cost Accountant	Accounting	43942				100000	0.06
21	VNY-84	Brenden Peskett	Human Resources Assistant I	Human Resources	107285				150000	0.08
22	HBE-39	Carling Durrans	Research Nurse	Operations	81899					
23	MNL-93	Benn Le Lievre	Payment Adjustment Coordinator	Accounting	84469					
24	APS-61	Gerri Pavolillo	Senior Sales Associate	Sales	106024					
25	MTX-57	Shantee Durnin	Senior Financial Analyst	Accounting	104051					
26	MOK-62	Dee Ivchenko	Biostatistician IV	Operations	51581					
27	LPN-62	Jess Bradman	Administrative Assistant III	Human Resources	55810					
28	UGP-19	Carina Poge	Financial Analyst	Accounting	72707					
29	ZTH-55	Margery Boom	Structural Engineer	Engineering	59847					

See *VLOOKUP – Merge Tables* sheet of the *Lookup Functions_Practice.xlsx* to study the example more thoroughly.

	A	B	C	D	E	F	G	H	I	J
1	Employee ID	Name	Title	Department	Salary	Department ID	Bonus (%)			
2	QIA-07	Evie Base	Accounting Assistant I	Accounting	$ 109,605	AX	6%			
3	HMO-83	Jorrie MacTrustram	Tax Accountant	Accounting	$ 95,332	AX	5%			
4	IQZ-97	Clio Pitson	Community Outreach Specialist	Marketing	$ 70,726	MX	5%			
5	XTT-54	Lina Jacob	Information Systems Manager	IT	$ 94,406	IX	5%		Department	Department ID
6	CIG-82	Lauren Pollak	Programmer Analyst IV	IT	$ 100,424	IX	6%		Accounting	AX
7	AZW-23	Elizabeth Lapides	Speech Pathologist	Operations	$ 87,053	OX	5%		Engineering	EX
8	FAX-57	Hadlee Hounihan	Design Engineer	IT	$ 51,831	IX	5%		Human Resources	HX
9	GXM-51	Junie Klimushev	Accounting Assistant I	Accounting	$ 104,663	AX	6%		IT	IX
10	QPH-99	Cindy Breznovic	Speech Pathologist	Operations	$ 101,465	OX	6%		Marketing	MX
11	TSN-47	Niccolo Scrymgeour	Web Designer II	IT	$ 61,136	IX	5%		Operations	OX
12	EIU-27	Waly Scogin	Environmental Specialist	Operations	$ 47,888	OX	3%		Sales	SX
13	APC-45	Jayme Sambrok	Sales Associate	Sales	$ 83,806	SX	5%			
14	GZH-10	Skelly Dignall	Administrative Assistant III	Operations	$ 69,125	OX	5%			
15	SDN-09	Carolus Adame	Analyst Programmer	IT	$ 105,932	IX	6%			
16	YOZ-39	Bendix Ventum	Cost Accountant	Accounting	$ 111,417	AX	6%			
17	JII-95	Valentina Sutheran	Physical Therapy Assistant	Operations	$ 102,695	OX	6%		Salary	Bonus
18	VSD-61	Gratiana Zapatero	Design Engineer	Engineering	$ 113,671	EX	6%		0	3%
19	LBJ-50	Paulie O'Carmody	Senior Editor	Operations	$ 85,182	OX	5%		50,000	5%
20	WCC-51	Trumaine Andres	Senior Cost Accountant	Accounting	$ 43,942	AX	3%		100,000	6%
21	VNY-84	Brenden Peskett	Human Resources Assistant I	Human Resources	$ 107,285	HX	6%		150,000	8%
22	HBE-39	Carling Durrans	Research Nurse	Operations	$ 81,899	OX	5%			
23	MNL-93	Benn Le Lievre	Payment Adjustment Coordinator	Accounting	$ 84,469	AX	5%			
24	APS-61	Gerri Pavolillo	Senior Sales Associate	Sales	$ 106,024	SX	6%			
25	MTX-57	Shantee Durnin	Senior Financial Analyst	Accounting	$ 104,051	AX	6%			
26	MOK-62	Dee Ivchenko	Biostatistician IV	Operations	$ 51,581	OX	5%			
27	LPN-62	Jess Bradman	Administrative Assistant III	Human Resources	$ 55,810	HX	5%			
28	UGP-19	Carina Poge	Financial Analyst	Accounting	$ 72,707	AX	5%			
29	ZTH-55	Margery Boom	Structural Engineer	Engineering	$ 59,847	EX	5%			

HLOOKUP

HLOOKUP stands for horizontal lookup. The HLOOKUP function searches for a specified value in the top row of a table, and returns a value a specified number of rows down from the found value. It is the same as a VLOOKUP, except it looks up values horizontally instead of vertically. Out of the two, VLOOKUP is more commonly used.

Syntax: =HLOOKUP(lookup_value,table_array,row_index_num,range_lookup)

Arguments:

- *lookup_value* – The value you want to search for.
- *table_array* – The table from which to return data or the hlookup table
- row_index_num – The row number from which to retuen data.

- range_lookup – A Boolean to indicate if exact (FALSE) or approximate match (TRUE) is required.

Using HLOOKUP for exact matches

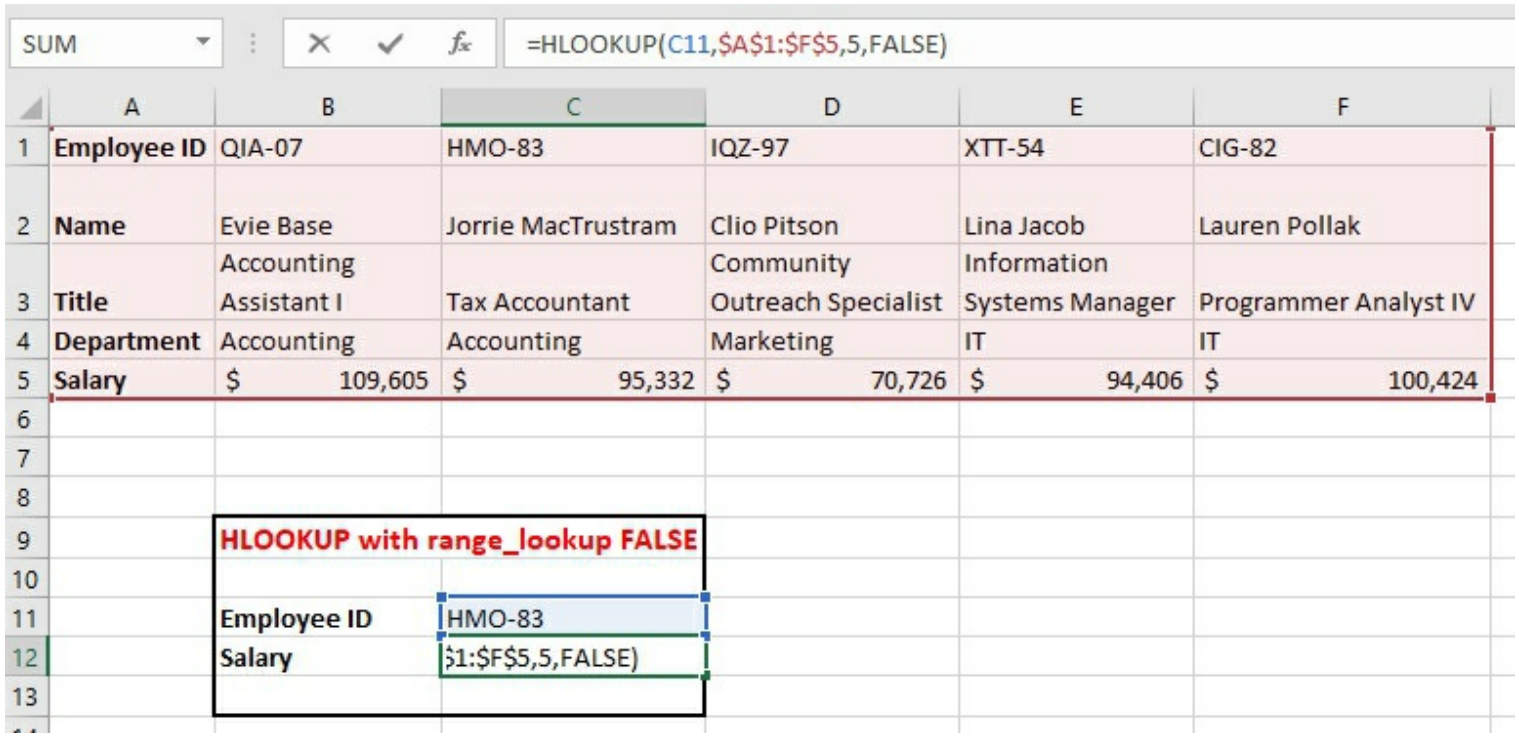
The argument *range_lookup* is set to FALSE if an exact match with the *lookup_value* needs to be returned.

Practice:

1. Continue working on *Lookup Functions_Practice.xlsx* and go to the *HLOOKUP* sheet. Suppose you want to automate display of the salary (in cell C12) of any employee ID you enter into cell C11.
2. Click on cell C12 and enter below formula into the cell.

=HLOOKUP(C11,$A$1:$F$5,5,FALSE)

3. Enter any Employee ID in C11. The corresponding salary will be returned in cell C12.



In brief, the HLOOKUP formula searches the row 1 (Employee ID) of the hlookup table ($A$1:$F$5) until the given employee ID is found and then returns the corresponding value in row 5 (Salary) of the same column.

Using HLOOKUP for approximate matches

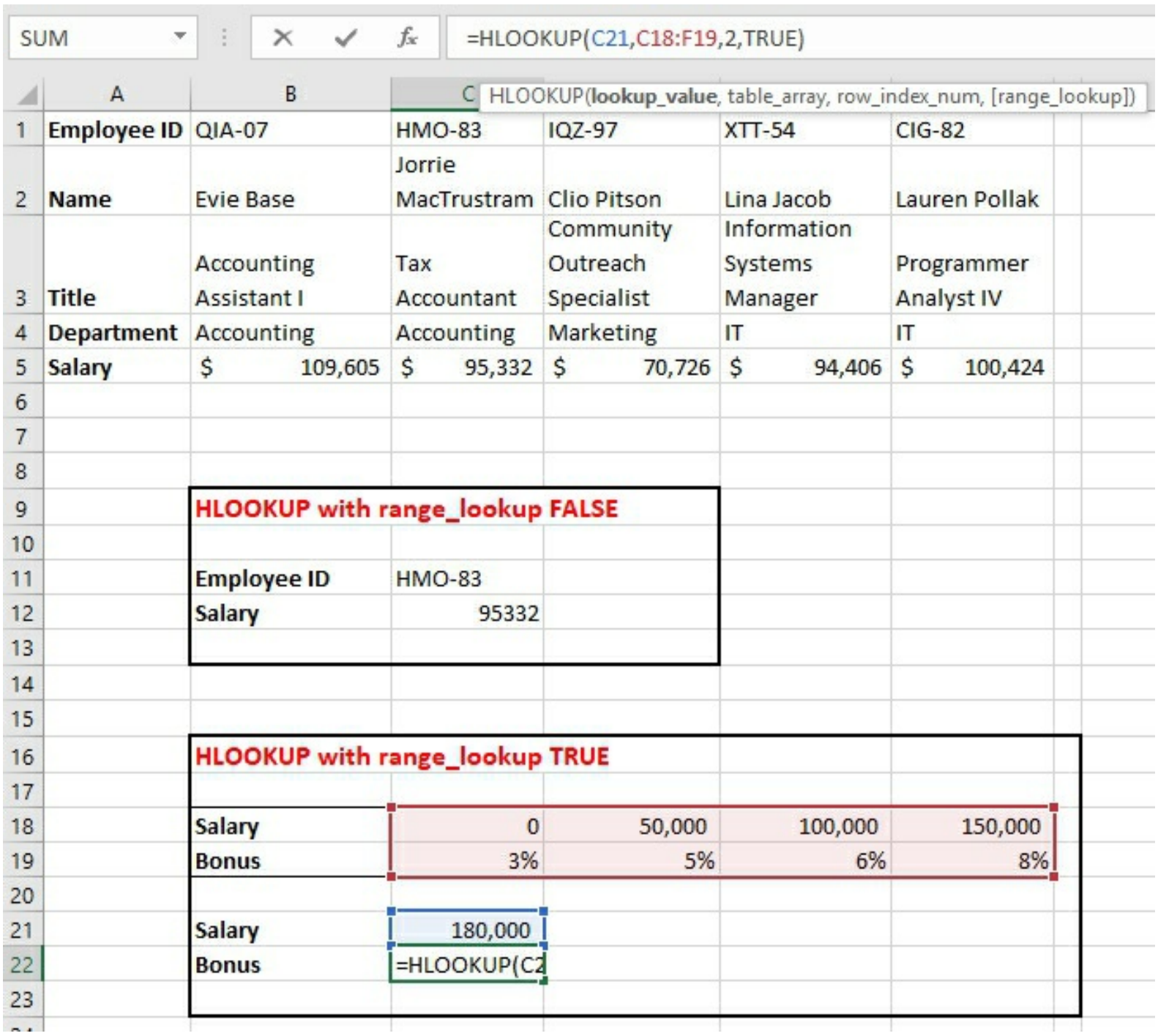
The argument *range_lookup* is set to TRUE if an approximate match with the *lookup_value* needs to be returned.

Practice

1. Continue working on *HLOOKUP* sheet of the *Lookup Functions_Practice.xlsx*. Suppose you want to automate display of the bonus (in cell C22) of any employee ID you enter into cell C21.
2. Click on cell C22 and enter below formula into the cell.

=HLOOKUP(C21,$C$18:$F$19,5,TRUE)

3. Enter any salary in C21. The corresponding bonus will be returned in cell C22.



In brief, the HLOOKUP formula searches the row 1 (Salary) of the vlookup table ($C$18:$F$19) and returns the value corresponding to the largest value that is less than the *lookup_value* is returned. As the salary of 180,000 is >150,000, 8% corresponding to 150,000 is returned.

INDEX

The INDEX function returns a cell value from a list or table based on its column and row numbers. There are two ways to use the INDEX function: *array* form is used to return the value of a specified cell or array of cells, and *reference* form is used to return a reference to specified cells. We will be focusing on the array form.

Syntax: =INDEX (array, row_num, [col_num])

Arguments:

- *array* – A range of cells.
- *row_num* – The index of the row with return value in the array.
- *column_num* – The index of the column with return value in the array.

Practice

1. Continue working on *Lookup Functions_Prcatice.xlsx* and go to the *INDEX* sheet. We are now working with a dataset containing calls received, answered, resolved as well as customer ratings received by six call centre employees. Suppose you want to display calls answered by Davis Pendle.
2. Click on cell I4 and enter below formula into the cell.

=INDEX(B2:F7,5,3)

3. Check if the function has returned the correct value.

In brief, the INDEX function returns the value at the intersection of the fifth row and third column in the range B2:F7 (first argument). When the range is two-dimensional, both the row and column number are specified. When the range is one-dimensional, only row number is needed.

SUM | =INDEX(B2:F7,5,3)

	A	B	C	D	E	F	G	H	I
1		**Employee Name**	**Calls Received**	**Calls Answered**	**Calls Resolved**	**Customer Rating**			
2		Mandie Fattorini	585	424	389	4		**INDEX**	
3		Tildie Theriot	549	479	312	3			
4		Bearnard Welland	547	453	224	2		Davis Pendle - Calls Answered	F7,5,3)
5		Hildegarde Grindle	501	444	344	3			
6		Davis Pendle	530	467	388	2			
7		Dona Ebdon	568	0	500	2			

MATCH

In practice, we don't know the location (row and column index) of a value of interest in a range, hence limiting the utility of INDEX function. However, the MATCH function is used to find the position of a value in a range and can then be used in combination with the INDEX function. Let's focus on the MATCH function for now. The MATCH function searches for an item in a range of cells and returns a number representing its position in the range.

Syntax: =MATCH(lookup_value,lookup_array,[match_type])

Arguments:

- *lookup_value* – The value you want to search for.
- *lookup_array* – An array of data either one column wide or one row high that you want to search in.
- *match_type* – 0, -1 or 1 specifies return value if an exact match is not found. 0 returns an exact match, 1 returns the exact match or next smallest and -1 returns the exact match or next largest.

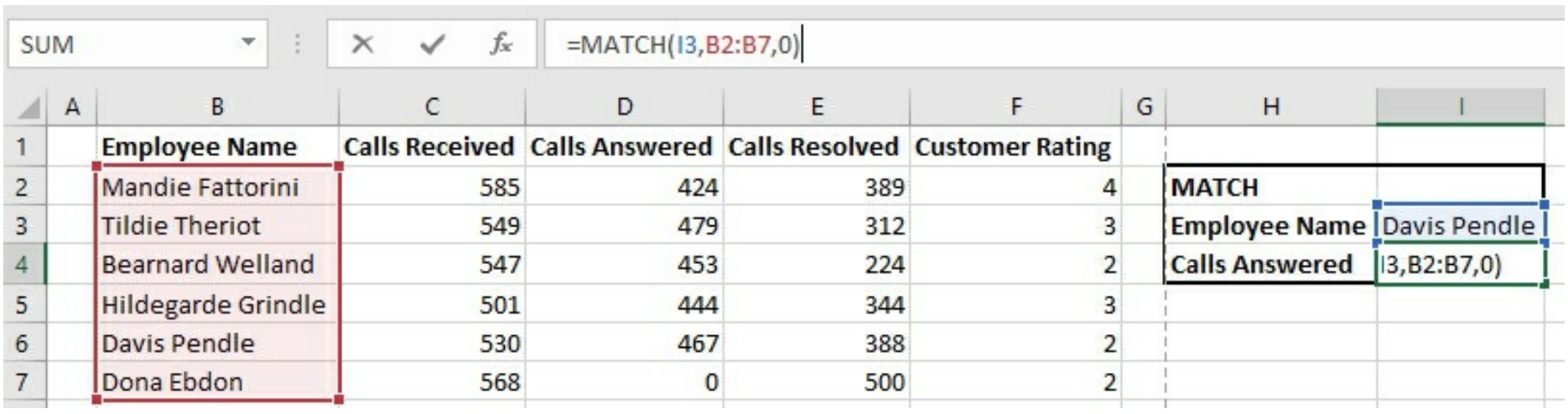
Practice

1. Continue working on *Lookup Functions_Practice.xlsx* and go to the *MATCH* sheet. Suppose you want to display calls answered by Davis Pendle.
2. Click on cell I4 and enter below formula into the cell.

   =MATCH(I3,B2:B7,0)

3. Check if the function has returned the correct position number in the array.

In brief, the MATCH function returns the position number of a specified value (I3, Davis Pendle) in the given array (B2:B7) upon finding the exact match.



INDEX and MATCH

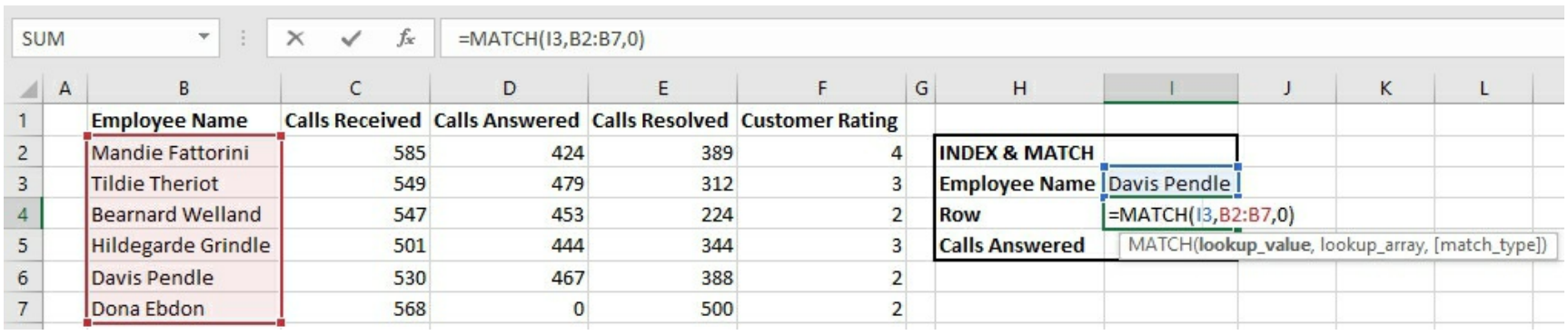
Now that we've covered the basics of both INDEX and MATCH, let's combine the two functions in a single formula.

Practice

1. Continue working on *Lookup Functions_Practice.xlsx* and go to the *INDEX & MATCH* sheet. Suppose you want the formula to return the number of calls answered by the employee whose name is entered in cell I3.
2. We know that we can give INDEX formula a row and column number to retrieve a value at the intersection of a given row and column. To be able to use the INDEX, first, let's retrieve row number with the MATCH function. Click on cell I4 and enter below formula into the cell.

= MATCH(I3,B2:B7,0)

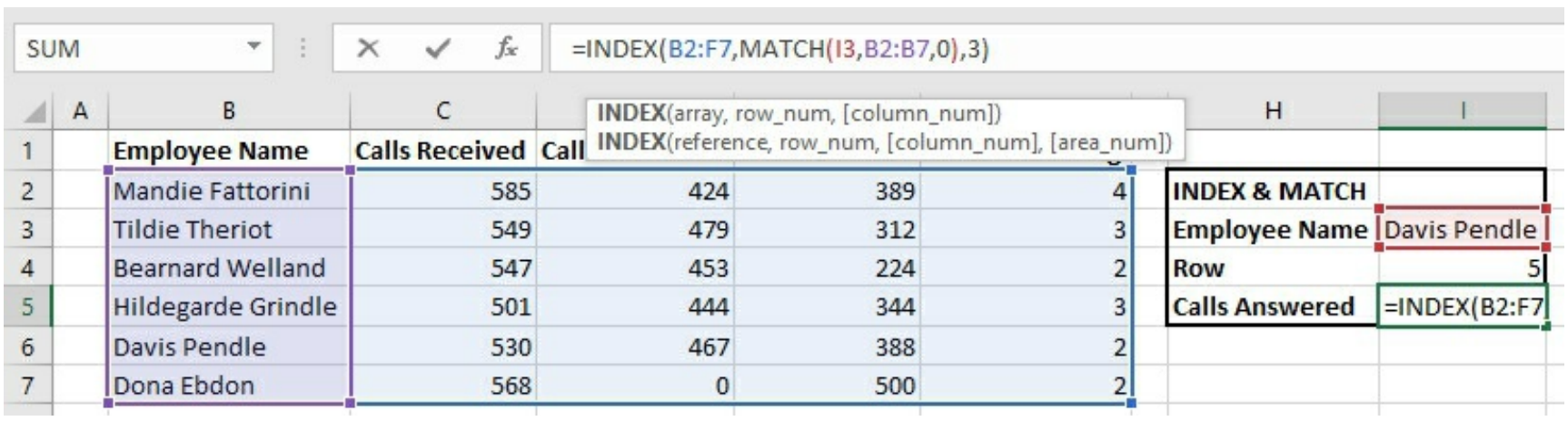
The formula returns 5 as position number of the name entered in cell I3 (Davis Pendle).



3. Now, click on cell I5 and enter the below formula into the cell.

=INDEX(B2:F7,MATCH(I3,B2:B7,0),3)

We have simply replaced the *row_num* 5 in the INDEX function we previously used with the MATCH function from step 3. We also manually supplied 3 as *column_num* of Calls Answered column. The function returns the value at the intersection of row 5 and column 3 of the specified range (B2:F7).



TWO-WAY INDEX and MATCH

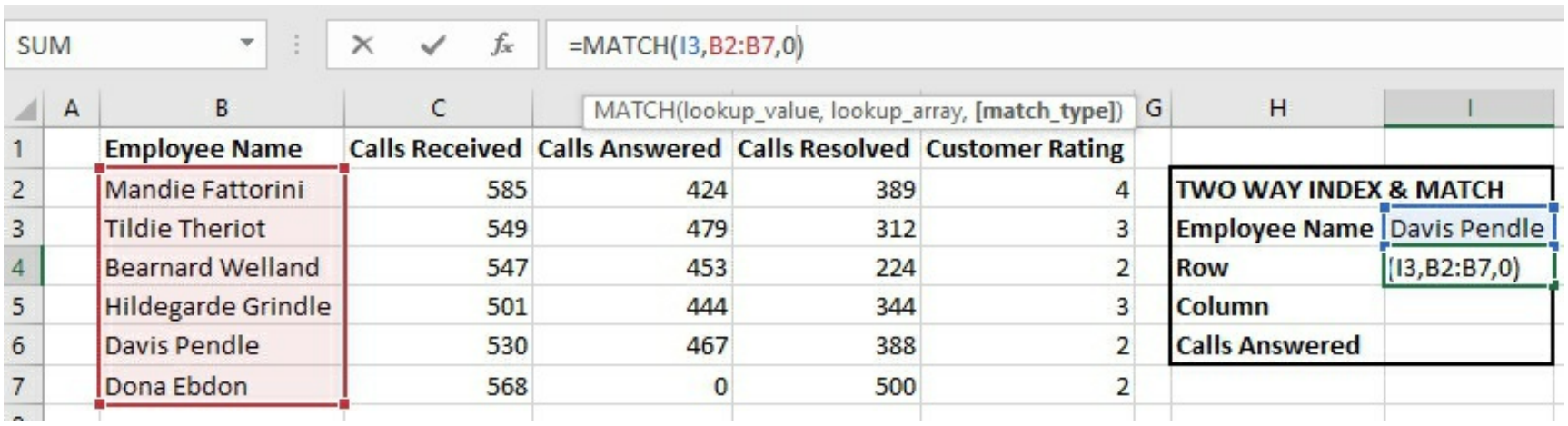
Above we used the MATCH function to find the row number dynamically, but provided the column number manually. Now we will make the formula fully dynamic by using another match function to dynamically determine the column number. Hence, we will use two MATCH functions – one to get the row position, and the other to get the column position.

Practice

1. Continue working on *Lookup Functions_Practice.xlsx* and go to the *TWO WAY INDEX & MATCH* sheet. Suppose we want the formula to return the number of calls answered by the employee whose name is entered in cell I3.
2. First, let's retrieve row number with the MATCH function. Click on cell I4 and enter below formula into the cell.

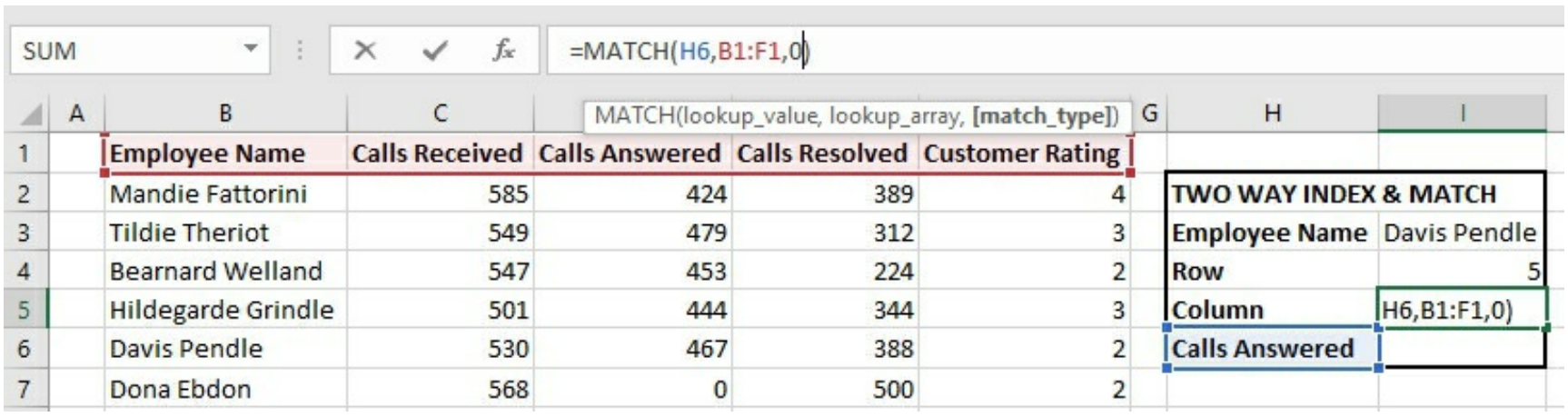
= MATCH(I3,B2:B7,0)

The formula returns 5 as position number of the name entered in cell I2 (Davis Pendle).



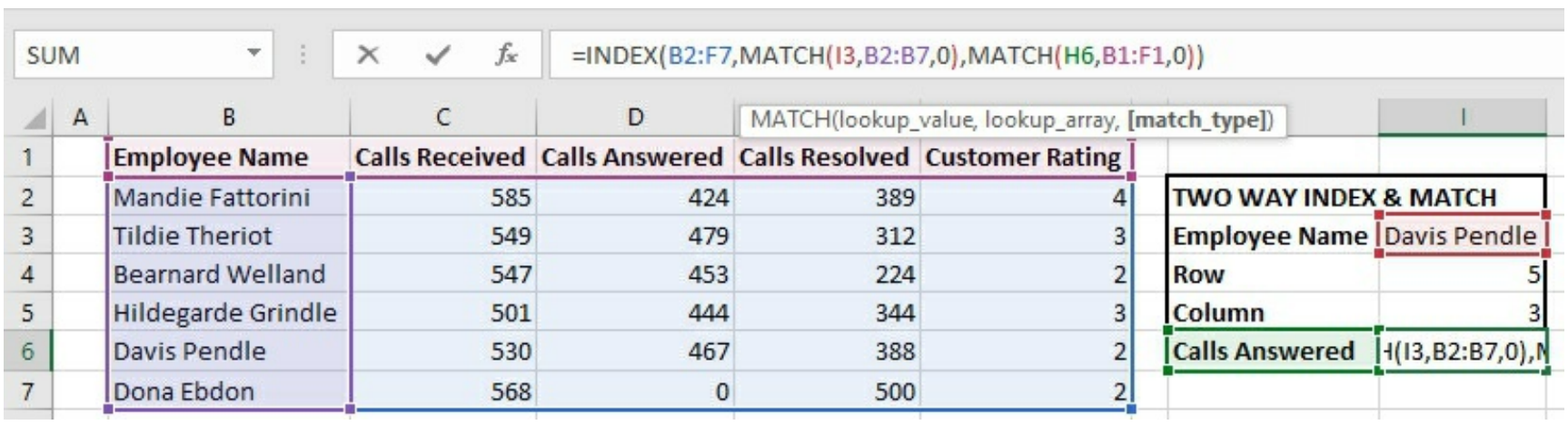
3. Now, let's retrieve column number with the MATCH function. Click on cell I5 and enter below formula into the cell. The formula returns 3 as the position number of the column titled Calls Answered.

=MATCH(H6,B1:F1,0)



4. Now, click on cell I6 and enter below formula into the cell.

=INDEX(B2:F7,MATCH(I3,B2:B7,0),MATCH(H6,B1:F1,0))



We have replaced both *row_num* and *column_num* arguments with two MATCH functions. The MATCH functions return 5 as row number and 3 as column number, and the INDEX formula returns the value at the intersection of row 5 and column 3 of the specified range (B2:F7).

These functions especially VLOOKUP and INDEX-MATCH are essential to learn data analysis in Excel. While we have covered VLOOKUP extensively, consider diving deeper into INDEX-MATCH functions beyond what is covered in this lesson as it has many useful capabilities such as left lookup which VLOOK is not capable of.

Array Formulas

In addition to the lookup functions, we will also briefly cover array formulas in this chapter. An array is a collection of items in a single row or column or in multiple rows and columns. The items can be text or numbers. An array formula is capable of processing several items in an array at the same time instead of just one. In other words, an array formula evaluates all individual values in an array and performs multiple calculations on one or several items based on the formula.

Practice

1. Continue working on *Lookup Functions_Practice.xlsx* and go to the *Array Formulas* sheet. Suppose you want to calculate Average Calls Abandoned. To be able to do this with regular formulae, you will need to insert an additional column to calculate *Calls Abandoned* by subtracting *Calls Answered* from *Calls Received*, and then average values in the column. Alternatively, you can use an array formula!
2. In cell H3, type the below formula.

=AVERAGE(C2:C7-D2:D7)

3. Press the keyboard shortcut CTRL + SHIFT + ENTER to complete the array formula. Hence, array formulas are entered by pressing CTRL and SHIFT keys in addition to the ENTER key. Once you do this, the formula is surrounded with {curly braces}, which indicate an array formula.

H3 | {=AVERAGE(C2:C7-D2:D7)}

	A	B	C	D	E	F	G	H
1		Employee Name	Calls Received	Calls Answered	Calls Resolved	Customer Rating		
2		Mandie Fattorini	585	424	389	4		Average Calls Abandoned
3		Tildie Theriot	549	479	312	3		91
4		Bearnard Welland	547	453	224	2		
5		Hildegarde Grindle	501	444	344	3		
6		Davis Pendle	530	467	388	2		
7		Dona Ebdon	568	468	500	2		

The formula subtracts the values in each row of the specified arrays (cells C2 through D7), and averages the subtracted values to provide an output.

2.5 Forecasting

Forecasting is a technique to establish future trends based on historical data. It can be used to make predictions about the future, for example anticipate future expenses or sales. However, forecasting bases the prediction for the future on the pattern of what happened in the recent past. Hence, an impact of any external factors, for example impact on sales of a product of release of a similar product at half the price by a competitor, is not accounted for and the future is not predicted definitively. Hence, it is important to note that forecasting provides insights into "what could happen" instead of "what will happen" for decision making.

For forecasting, the input data is arranged in two columns or series:

1. Time series – A series of date or time entries taken at successive equally spaced points in time, for example hourly, daily, monthly, yearly etc.
2. Data series – Numeric data values corresponding to time series that are to be forecasted.

Microsoft Excel offers two different methods for forecasting based on the data:

1. Exponential smoothing
2. Linear Regression

Both these methods rely on Excel functions to forecast a value based on existing values and trends.

Using exponential smoothing

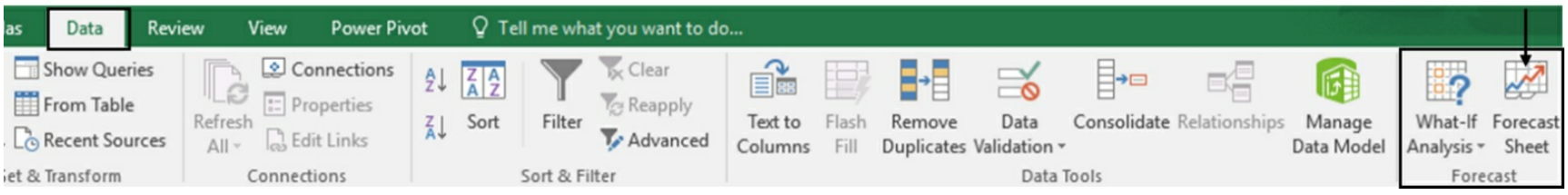
Exponential smoothing is used to smooth out irregularities (peaks and valleys) to easily recognize trends and forecast future values based on those trends. Hence, it smoothens out the data by eliminating minor deviations and is best suited for non-linear datasets that contain seasonal patterns. A seasonal pattern occurs when a time series is affected by seasonal factors such as the time of the year or the day of the week, for example elevated temperatures during summer months in the northern hemisphere.

Practice

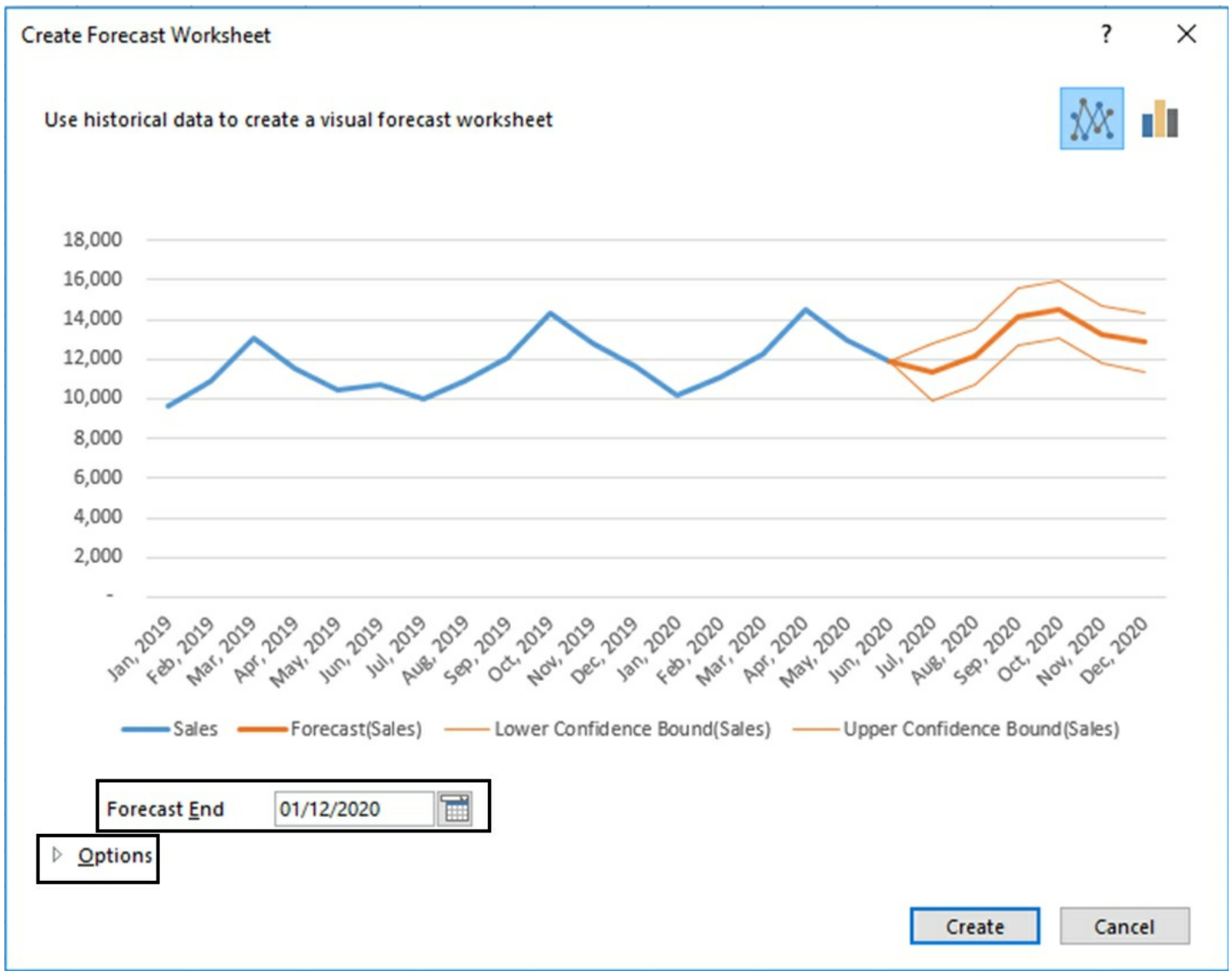
1. Open *Forecasting_Practice*.xlsx and go to the *Exponential Smoothing* sheet. The sheet contains monthly sales between Jan, 2019 to Jun, 2020. *Month* is the time series and *Sales* is the data series. We will forecast sales up to Dec, 2020.
2. Select both columns, including headers (range A1:B19).

	A	B
1	Month	Sales
2	Jan, 2019	9,643
3	Feb, 2019	10,862
4	Mar, 2019	13,077
5	Apr, 2019	11,567
6	May, 2019	10,426
7	Jun, 2019	10,698
8	Jul, 2019	9,998
9	Aug, 2019	10,867
10	Sep, 2019	12,090
11	Oct, 2019	14,301
12	Nov, 2019	12,791
13	Dec, 2019	11,654
14	Jan, 2020	10,198
15	Feb, 2020	11,067
16	Mar, 2020	12,290
17	Apr, 2020	14,501
18	May, 2020	12,991
19	Jun, 2020	11,854

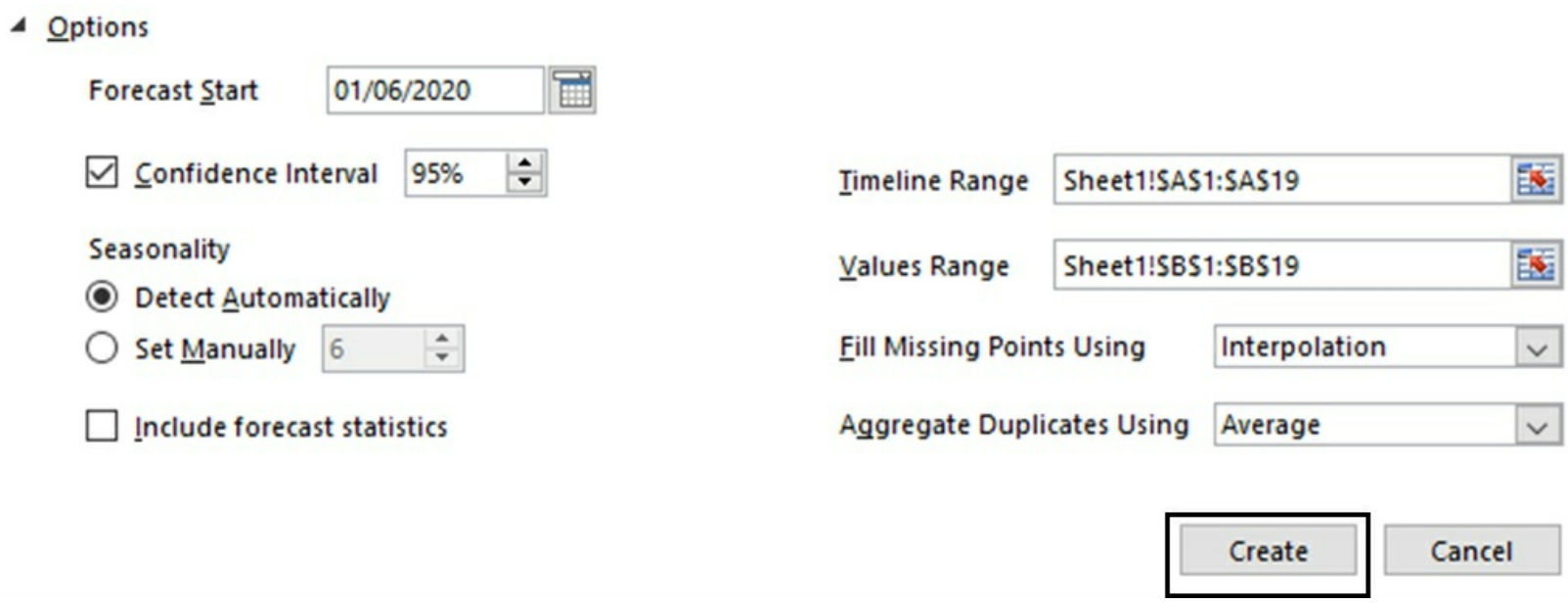
3. Go to the *Data* tab and under the *Forecast* group, select *Forecast Sheet*.



4. The *Create Forecast Worksheet* window opens up with a preview of the forecast as part of a line graph. You can toggle between the default line graph or column chart depending on your preference. For *Forecast End*, enter or select 01/12/2020.



5. Now click on *Options* to explore what's available. You can change any of the default options of your forecast.



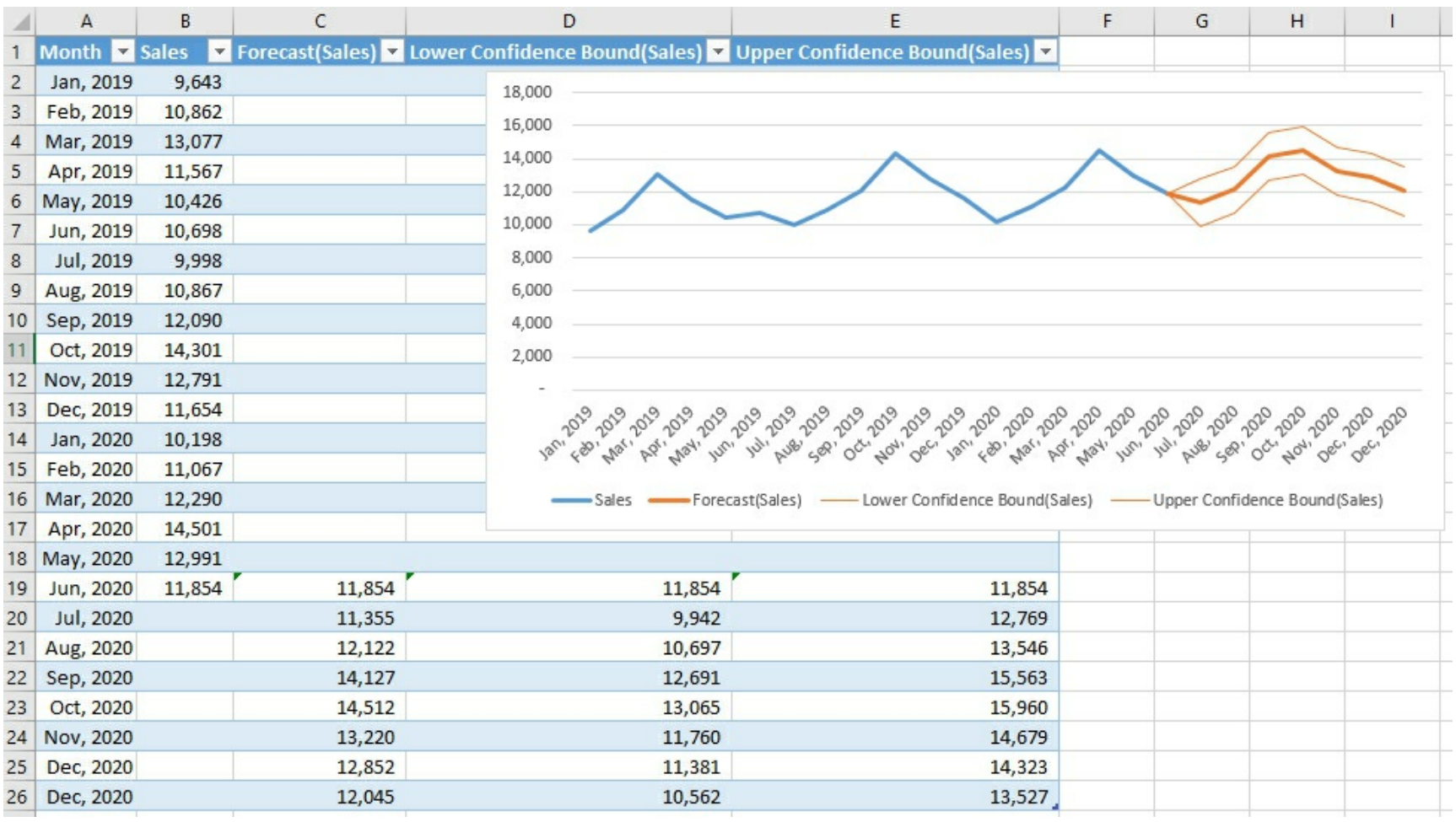
Following settings are available to be configured:

- *Forecast Start* – Used to change the start date for forecasting.
- *Confidence Interval* – Confidence interval sets a certain level of confidence a reasonable range of values in which the predicted values should fall. The default level is *95%*, meaning 95% of future values are expected to fall within the range. Hence, confidence interval helps to determine forecast accuracy. On the line graph, it is represented by two thin lines around the forecast line, and on the column chart by the error bar values.
- *Seasonality* –Excel identifies the seasonal cycle automatically but this option can be used to set it manually. Seasonality is the length of the seasonal pattern in which fluctuations occur at regular intervals. For example, if sales make a significant jump in the month of December each year, the

seasonality is set to 12.

- *Include Forecast Statistics* – Used to include additional statistical information on the forecast calculated by the FORECASE.ETS.STAT function. Check this box if you want Excel to generate additional statistics such as smoothing constants and error metrics.
- *Timeline Range* – Used to edit the range used for time series.
- *Values Range* – Used to edit the range used for data series.
- *Fill Missing Points Using* – Used to set method for handling missing points. The default is *Interpolation*, where the missing points are filled based on the weighted average of neighboring points. Alternatively, you can select *Zeros* to treat the missing points as zero values.
- *Duplicate Aggregates Using* – Used to set method for handling duplicate values with the same timestamp are calculated. The default is *Average* with *Median*, *Max*, *Min*, etc. available as alternate options.

6. Finally, click on *Create* to create the forecast.
7. Excel creates a table with original and predicted values as well as lower and upper confidence intervals of the predicted values on a new sheet along with a chart that visually presents this data.



	A	B	C	D	E
1	Month	Sales	Forecast(Sales)	Lower Confidence Bound(Sales)	Upper Confidence Bound(Sales)
2	Jan, 2019	9,643			
3	Feb, 2019	10,862			
4	Mar, 2019	13,077			
5	Apr, 2019	11,567			
6	May, 2019	10,426			
7	Jun, 2019	10,698			
8	Jul, 2019	9,998			
9	Aug, 2019	10,867			
10	Sep, 2019	12,090			
11	Oct, 2019	14,301			
12	Nov, 2019	12,791			
13	Dec, 2019	11,654			
14	Jan, 2020	10,198			
15	Feb, 2020	11,067			
16	Mar, 2020	12,290			
17	Apr, 2020	14,501			
18	May, 2020	12,991			
19	Jun, 2020	11,854	11,854	11,854	11,854
20	Jul, 2020		11,355	9,942	12,769
21	Aug, 2020		12,122	10,697	13,546
22	Sep, 2020		14,127	12,691	15,563
23	Oct, 2020		14,512	13,065	15,960
24	Nov, 2020		13,220	11,760	14,679
25	Dec, 2020		12,852	11,381	14,323
26	Dec, 2020		12,045	10,562	13,527

8. Notice that sales value for Jun, 2020 (last month in data series) is duplicated across all columns. This provides a starting point for prediction and confidence interval lines on the graph to generate an uninterrupted, continuous line.
9. Now, click on cell C20 which contains predicted sales for July, 2020. Now go to the formula bar and click anywhere on the FORECAST.ETS function to display syntax of the formula.

10. Compare formula displayed in cell C20 with its syntax. The *target_date* argument represents the time point for which to forecast a value (stored in A20). The *values* argument represents the data series (B2:B19). The *timeline* argument represents the time series (A2:A19). 1 in the *seasonality argument* tells Excel to detect seasonality automatically. 1 in the *data_completion* argument tells Excel to complete missing points as the average of the neighboring points. The *aggregation* argument is omitted, which means that duplicate values with the same time stamp are to be averaged.

Using linear regression

Linear regression can be used to forecast linear data without seasonality or other fluctuations. This method can also

be used for small and simple data sets that do not have enough historical data to detect seasonality. Microsoft Excel does not provide a built-in feature to perform linear forecasting automatically. It is instead performed using the FORECAST.LINEAR function.

The FORECAST.LINEAR function predicts a future value based on existing values. The future values are predicted based on existing values, known as x-values (time series) and y-values (data series) using linear regression.

FORECAST.LINEAR(x, known_y's, known_x's)

The *x* argument represents the data point for which you want to predict a value. The *known_y's* argument represents the data series range and the *known_x's* arguments the time series range.

Practice

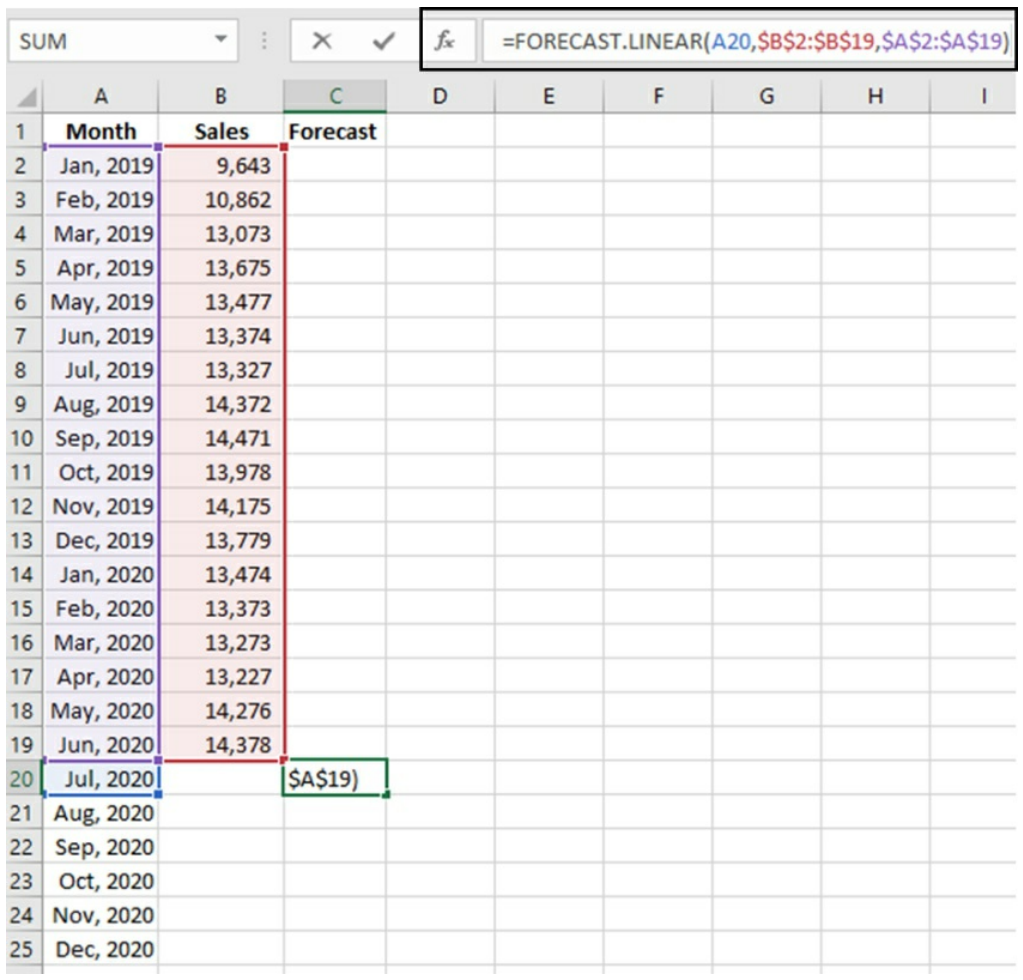
1. Continue working on *Forecasting_Prcatice.xlsx* and go to the *Linear Regression* sheet. There are three columns – *Month* (time series), *Sales* (data series) and *Forecast* (for forecasted values).
2. Click on cell C20 under the Forecast column and enter the below formula:

=FORECAST.LINEAR(A20,$B$2:$B$19,$A$2:$A$19)

where, A20 is the target month for the prediction, $B$2:$B$19 is the input data series range and $A$2:$A$19 is the time series range.

Notice that both ranges for date and time series are locked with absolute cell references to prevent them from changing when the formula is copied down.

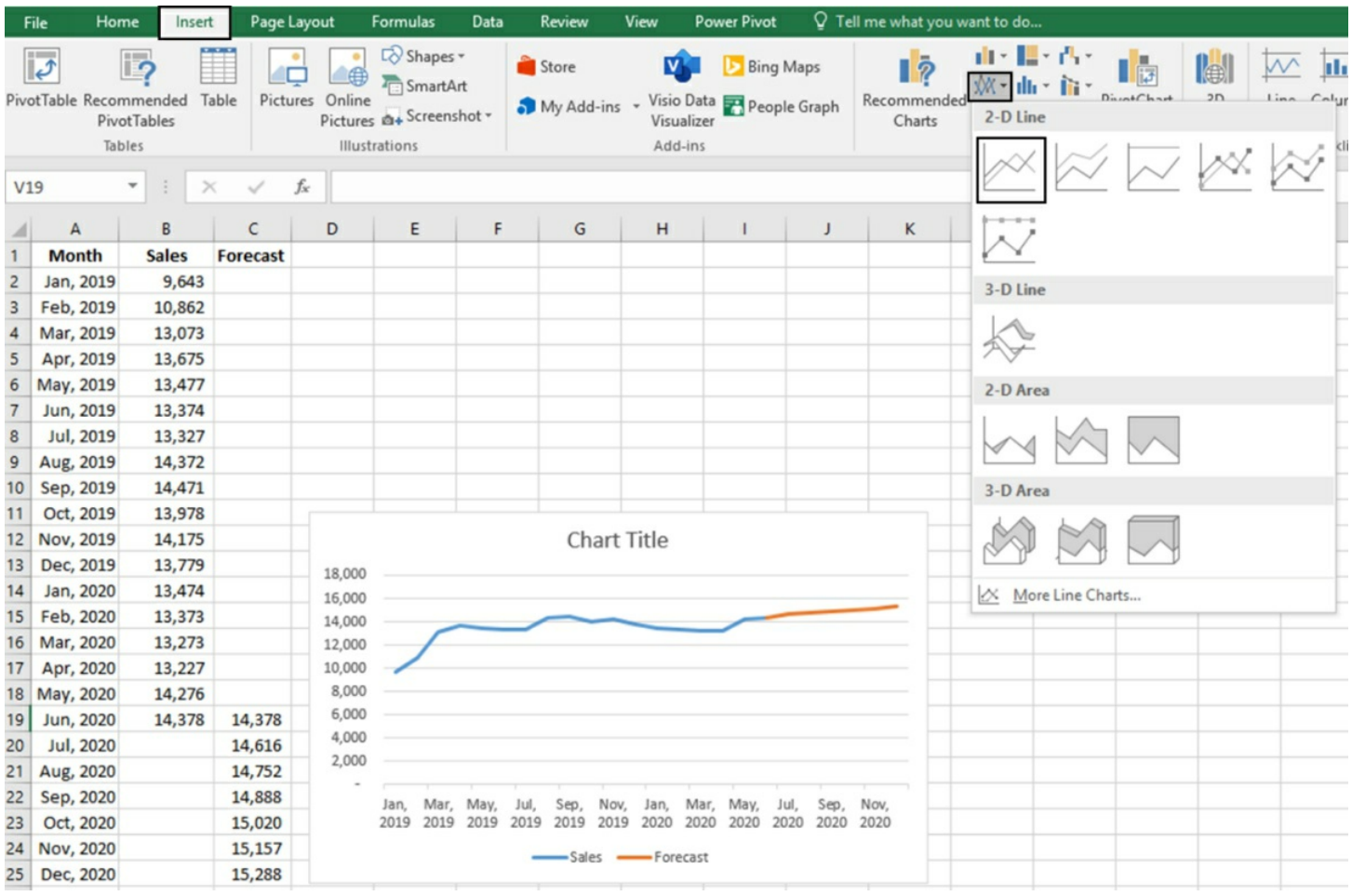
3. Copy the formula down by selecting cell C20, hovering to the bottom right of the cell and double-clicking on the drag handle (+ sign) once it appears. Now, we have our forecasted values.



4. To draw the graph, first copy the sales value (stored in cell B19) for Jun, 2020 (last month in data series) and paste it in the *Forecast* column (cell C19). This will help generate an uninterrupted line.

	A	B	C
1	Month	Sales	Forecast
2	Jan, 2019	9,643	
3	Feb, 2019	10,862	
4	Mar, 2019	13,073	
5	Apr, 2019	13,675	
6	May, 2019	13,477	
7	Jun, 2019	13,374	
8	Jul, 2019	13,327	
9	Aug, 2019	14,372	
10	Sep, 2019	14,471	
11	Oct, 2019	13,978	
12	Nov, 2019	14,175	
13	Dec, 2019	13,779	
14	Jan, 2020	13,474	
15	Feb, 2020	13,373	
16	Mar, 2020	13,273	
17	Apr, 2020	13,227	
18	May, 2020	14,276	
19	Jun, 2020	14,378	14,378
20	Jul, 2020		14,616
21	Aug, 2020		14,752
22	Sep, 2020		14,888
23	Oct, 2020		15,020
24	Nov, 2020		15,157
25	Dec, 2020		15,288

5. Now, select all three columns containing the data. Go to the *Insert* tab and in the *Charts* group, click on *Insert Line or Area Chart* icon and choose the first chart type (Line) in the first group (2-D Line). A line graph that visually represents the original and predicted values will be created.



Lesson 3: Exploring & Visualizing Data

Press the Ctrl key and click here to download practice & solution workbooks. Alternatively, copy and paste below link to your browser:

https://www.dropbox.com/sh/cj5gxxwt39l9j5c/AACndc4mevbQgvyPYeHkED1Fa?dl=0

3.1 Create and Analyze a PivotTable

For data analysis, PivotTables are one of the most useful features in Excel. They can be used to summarize a large amount of data with very little effort and time.

Summary statistics provided by a PivotTable can be used for reporting and data analysis. As PivotTables can be created interactively, you can view the same data from different perspectives within no time.

In this chapter, we'll discuss key features of PivotTable and practice them side-by-side using a set of data. Let’s get started!

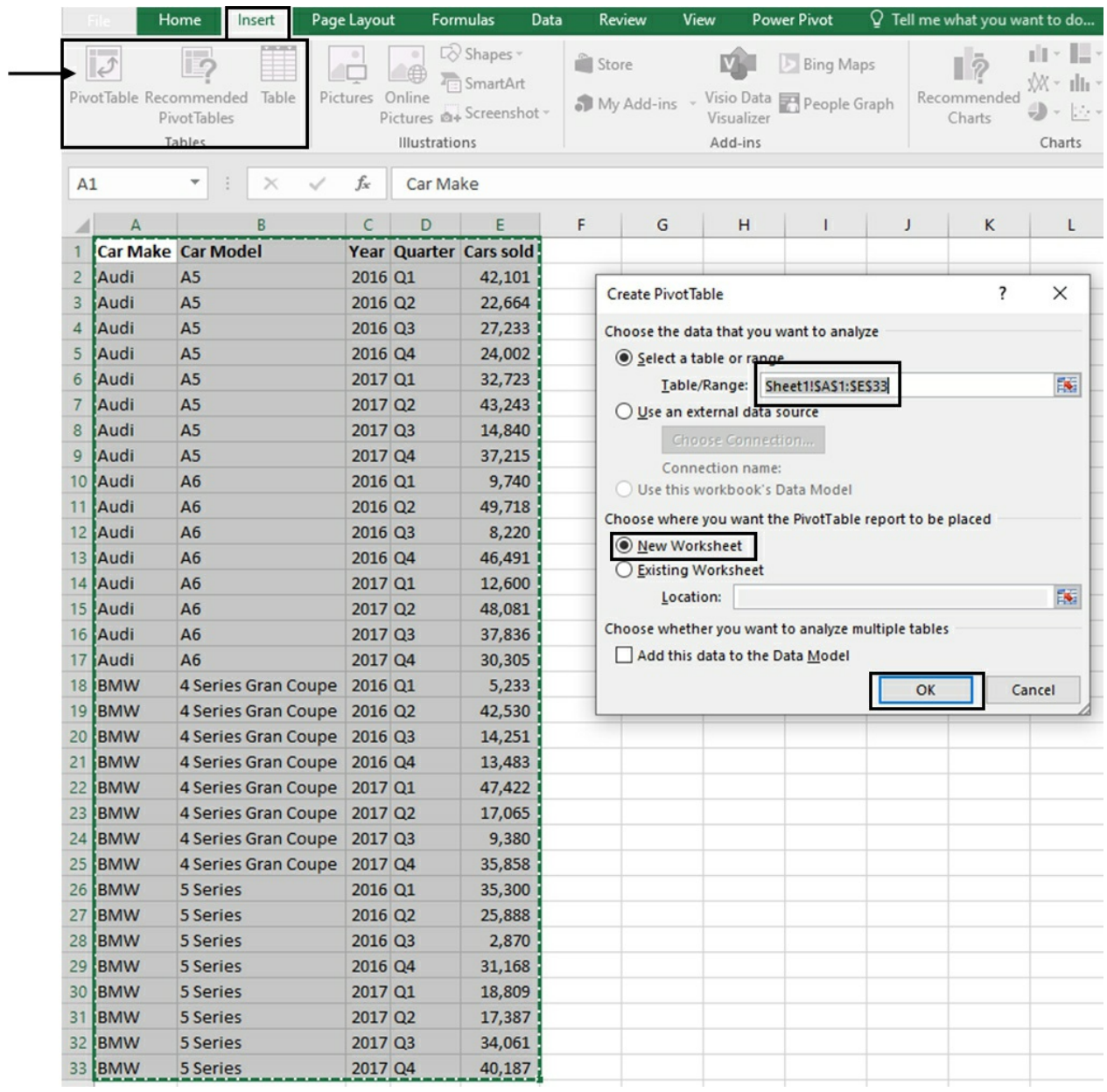
Creating a blank PivotTable

Before creating the PivotTable, ensure that your data is organized in rows and columns, with each row containing information on one record. Each column should have a short and unique header and contain only one format of data, such as dates, currency, or text. Creating a PivotTable takes just a minute!

Practice

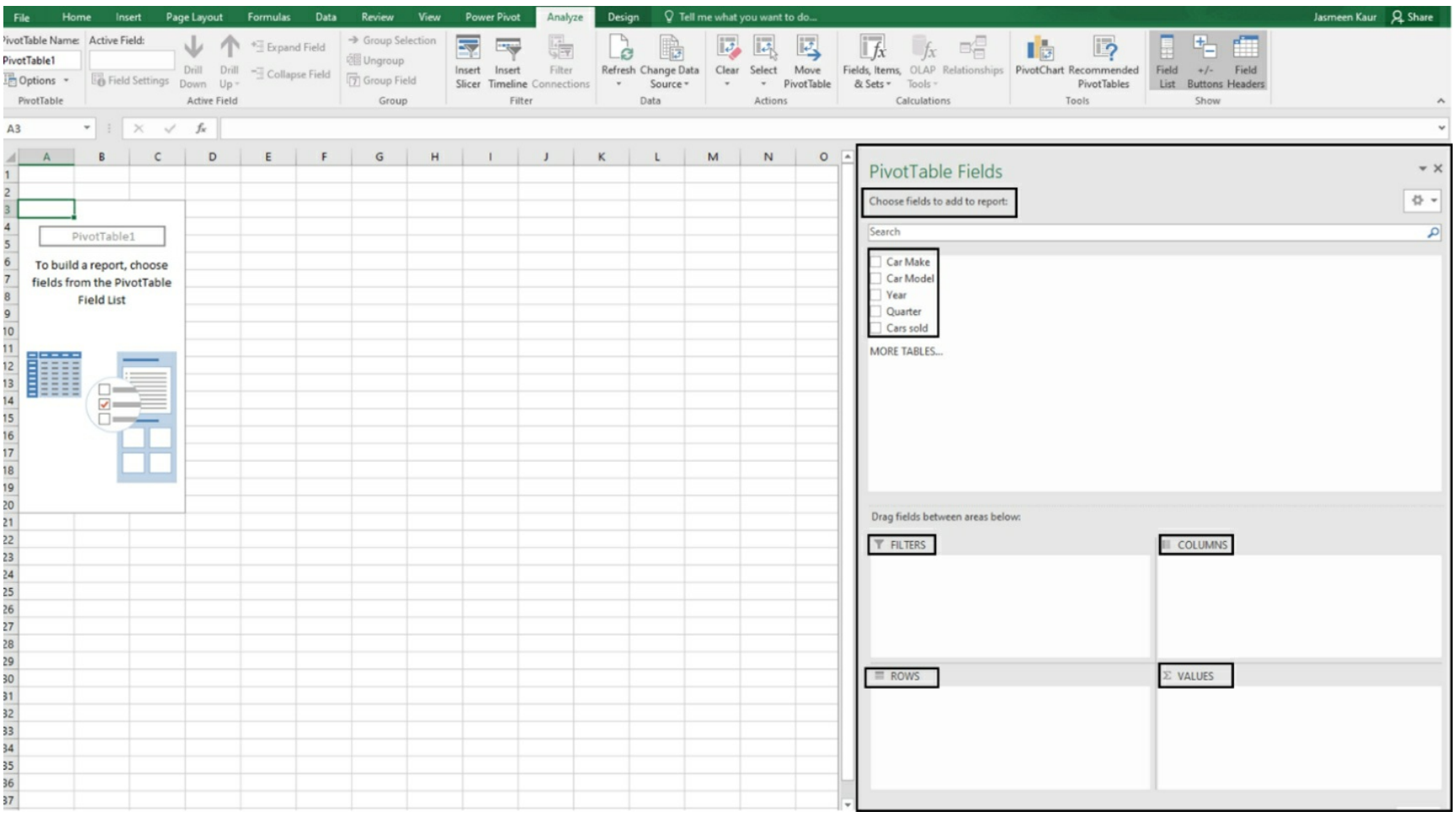
For this lesson, you will use dataset placed in *DATA* sheet of the *PivotTables_Practice.xlsx*. The dataset includes select models of Audi and BMW cars sold on a quarterly basis in 2016 and 2017. Let's say we want to calculate total cars sold broken down by Car Make (Audi, BMW) and Year. It could be time consuming and difficult to do; each Car Make and Year appears on multiple rows, and we would need to total all of their different car makes individually. Fortunately, a PivotTable can instantly summarize the data in a way that will make it much easier to read.

1. Open *PivotTables_Practice.xlsx* and go to the *DATA* sheet. To create a PivotTable, select entire dataset (A1:E33) on the sheet.
2. Go to the *Insert* tab and under the *Tables* group, select *PivotTable*. The *Create PivotTable* dialog box will pop up.
3. In the dialog box,
    a. for *Table/Range* under *Choose the data that you want to analyze*, confirm the range assigned is same as selected in step 1, and
    b. for *Choose where you want the PivotTable report to be placed*, select *New Worksheet*.



	Car Make	Car Model	Year	Quarter	Cars sold
2	Audi	A5	2016	Q1	42,101
3	Audi	A5	2016	Q2	22,664
4	Audi	A5	2016	Q3	27,233
5	Audi	A5	2016	Q4	24,002
6	Audi	A5	2017	Q1	32,723
7	Audi	A5	2017	Q2	43,243
8	Audi	A5	2017	Q3	14,840
9	Audi	A5	2017	Q4	37,215
10	Audi	A6	2016	Q1	9,740
11	Audi	A6	2016	Q2	49,718
12	Audi	A6	2016	Q3	8,220
13	Audi	A6	2016	Q4	46,491
14	Audi	A6	2017	Q1	12,600
15	Audi	A6	2017	Q2	48,081
16	Audi	A6	2017	Q3	37,836
17	Audi	A6	2017	Q4	30,305
18	BMW	4 Series Gran Coupe	2016	Q1	5,233
19	BMW	4 Series Gran Coupe	2016	Q2	42,530
20	BMW	4 Series Gran Coupe	2016	Q3	14,251
21	BMW	4 Series Gran Coupe	2016	Q4	13,483
22	BMW	4 Series Gran Coupe	2017	Q1	47,422
23	BMW	4 Series Gran Coupe	2017	Q2	17,065
24	BMW	4 Series Gran Coupe	2017	Q3	9,380
25	BMW	4 Series Gran Coupe	2017	Q4	35,858
26	BMW	5 Series	2016	Q1	35,300
27	BMW	5 Series	2016	Q2	25,888
28	BMW	5 Series	2016	Q3	2,870
29	BMW	5 Series	2016	Q4	31,168
30	BMW	5 Series	2017	Q1	18,809
31	BMW	5 Series	2017	Q2	17,387
32	BMW	5 Series	2017	Q3	34,061
33	BMW	5 Series	2017	Q4	40,187

4. Click *OK*. A blank *PivotTable* and a *PivotTable Fields* pane will appear at the right of the window.



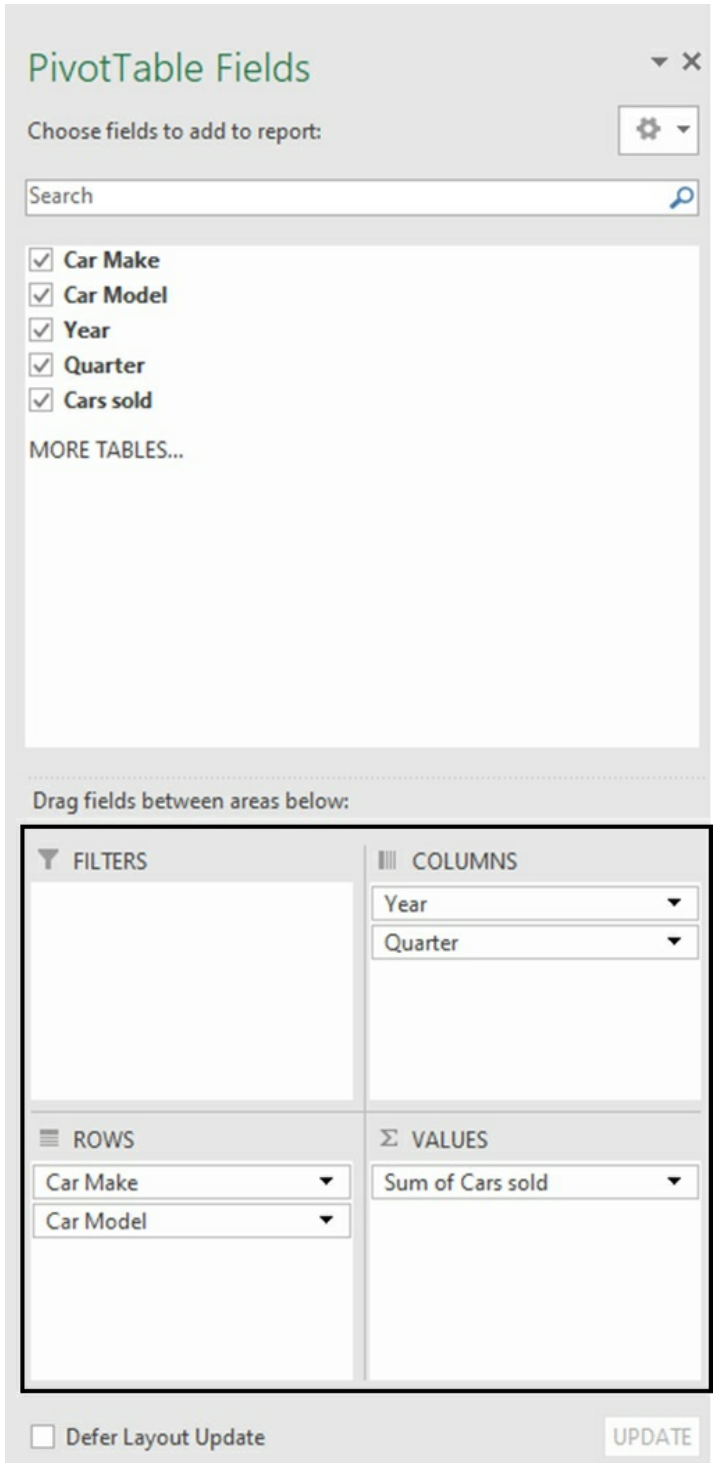
Adding fields to the blank PivotTable

The PivotTable Fields pane comprises of fields and areas. Fields represent the columns in your data and areas decide the layout of the PivotTable report. There are 4 areas available – ROWS, COLUMNS, Σ VALUES and FILTERS.

- In ROWS area, fields we want to appear as rows are placed.
- In COLUMNS area, fields we want to appear as columns are placed.
- Σ VALUES area is used to assign the fields you want to summarize the data by.
- In FILTERS area, place fields by which we want to filter the summarized data.

Practice

5. Drag and drop *Year* and then *Quarter* under *Columns*, *Car Make* and then *Car Model* under *Rows* area and *Cars Sold* under *Σ Values*. While doing so, keep an eye on how the PivotTable report changes as each of the fields is added. Notice that *Sum of Cars Sold* is calculated automatically under the *Σ Values* area.



6. Study the pivot report. Observe that:
    a. Years and quarters are in columns in the order they are placed on fields pane. Each year is nested with respective quarters.
    b. Car makes and car models are in rows in the order they are placed on fields pane. Each car make is nested with respective car models.
    c. Report is summarized by sum of cars sold

Sum of Cars sold	Column Labels										
	2016				2016 Total	2017				2017 Total	Grand Total
Row Labels	Q1	Q2	Q3	Q4		Q1	Q2	Q3	Q4		
Audi	**51841**	**72382**	**35453**	**70493**	**230169**	**45323**	**91324**	**52676**	**67520**	**256843**	**487012**
A5	42101	22664	27233	24002	116000	32723	43243	14840	37215	128021	244021
A6	9740	49718	8220	46491	114169	12600	48081	37836	30305	128822	242991
BMW	**40533**	**68418**	**17121**	**44651**	**170723**	**66231**	**34452**	**43441**	**76045**	**220169**	**390892**
4 Series Gran Coupe	5233	42530	14251	13483	75497	47422	17065	9380	35858	109725	185222
5 Series	35300	25888	2870	31168	95226	18809	17387	34061	40187	110444	205670
Grand Total	**92374**	**140800**	**52574**	**115144**	**400892**	**111554**	**125776**	**96117**	**143565**	**477012**	**877904**

Nesting PivotTable fields

If you have more than one field in any of the areas, then the fields are nested under each other in the order you place the fields in that area. Change the order by dragging the fields and observe how nesting changes.

Practice

7. Continuing on the PivotTable you just created, expand or collapse categories of the nested field (*Car Model*) by clicking on + and - icons placed next to each category of the *Car Make*.
8. Move fields between *Column* and *Row* areas and observe changes to the layout of the report.

Adding Filters to the PivotTable

You can assign a filter to one of the fields so that you can dynamically change the PivotTable based on the values of

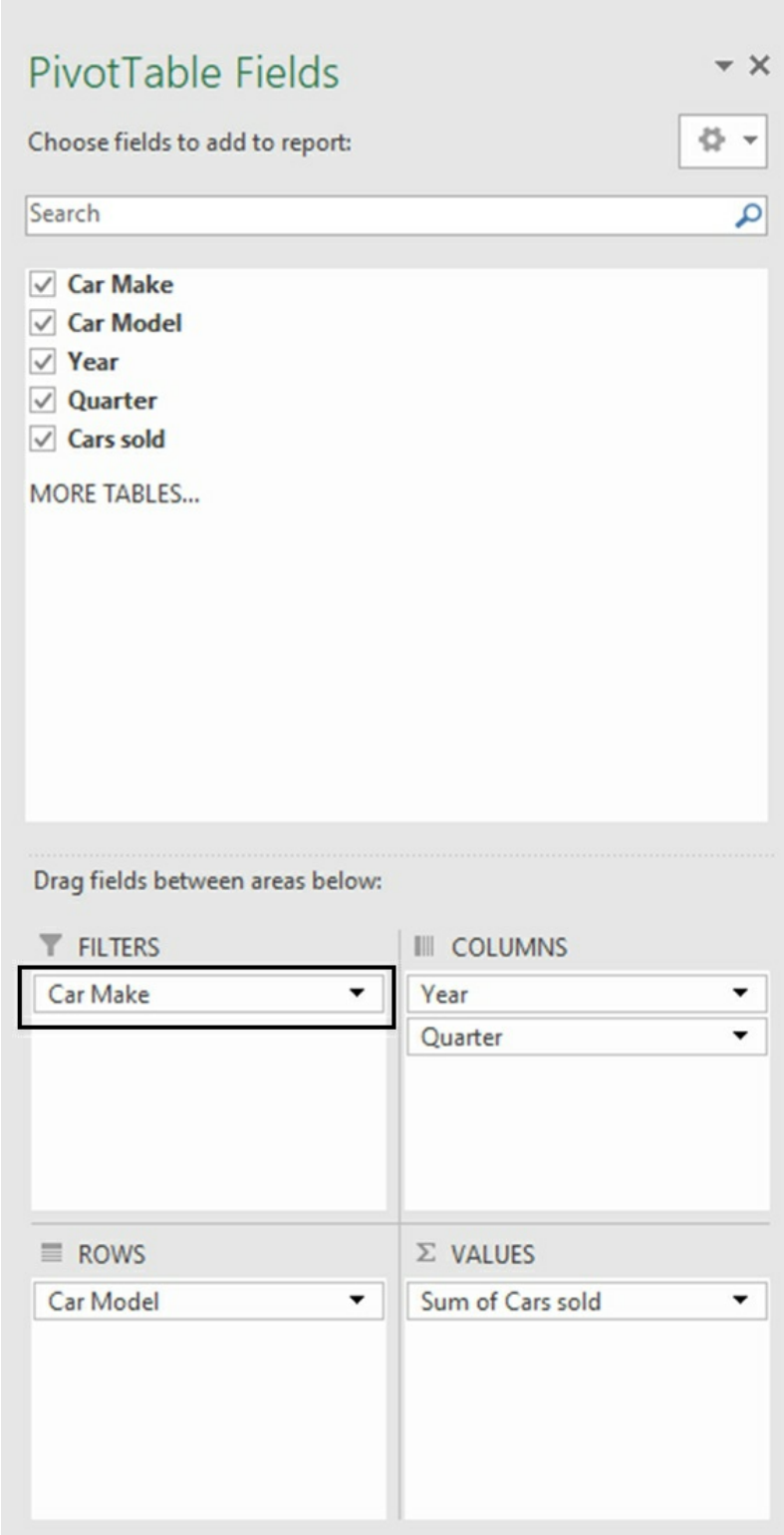
that field.

Practice

9. Go back to the PivotTable and revert it back to the below format, if needed.

Sum of Cars sold	Column Labels										
	2016				2016 Total	2017				2017 Total	Grand Total
Row Labels	Q1	Q2	Q3	Q4		Q1	Q2	Q3	Q4		
Audi	**51841**	**72382**	**35453**	**70493**	**230169**	**45323**	**91324**	**52676**	**67520**	**256843**	**487012**
A5	42101	22664	27233	24002	116000	32723	43243	14840	37215	128021	244021
A6	9740	49718	8220	46491	114169	12600	48081	37836	30305	128822	242991
BMW	**40533**	**68418**	**17121**	**44651**	**170723**	**66231**	**34452**	**43441**	**76045**	**220169**	**390892**
4 Series Gran Coupe	5233	42530	14251	13483	75497	47422	17065	9380	35858	109725	185222
5 Series	35300	25888	2870	31168	95226	18809	17387	34061	40187	110444	205670
Grand Total	**92374**	**140800**	**52574**	**115144**	**400892**	**111554**	**125776**	**96117**	**143565**	**477012**	**877904**

10. Drag *Car Make* from *Rows* to *Filters* in the PivotTable areas.



11. The filter with the label as *Car Make* appears above the PivotTable with default as '(All)'.

Car Make	(All)										
Sum of Cars sold	Column Labels										
	2016				2016 Total	2017				2017 Total	Grand Total
Row Labels	Q1	Q2	Q3	Q4		Q1	Q2	Q3	Q4		
4 Series Gran Coupe	5233	42530	14251	13483	75497	47422	17065	9380	35858	109725	185222
5 Series	35300	25888	2870	31168	95226	18809	17387	34061	40187	110444	205670
A5	42101	22664	27233	24002	116000	32723	43243	14840	37215	128021	244021
A6	9740	49718	8220	46491	114169	12600	48081	37836	30305	128822	242991
Grand Total	**92374**	**140800**	**52574**	**115144**	**400892**	**111554**	**125776**	**96117**	**143565**	**477012**	**877904**

12. Click the arrow in the box to the right of the default. From the drop-down list select *Audi* and click *OK*.

13. Only the data pertaining to *Audi* car make will be summarized.

Car Make	Audi										
Sum of Cars sold	Column Labels										
	2016				2016 Total	2017				2017 Total	Grand Total
Row Labels	Q1	Q2	Q3	Q4		Q1	Q2	Q3	Q4		
A5	42101	22664	27233	24002	116000	32723	43243	14840	37215	128021	244021
A6	9740	49718	8220	46491	114169	12600	48081	37836	30305	128822	242991
Grand Total	**51841**	**72382**	**35453**	**70493**	**230169**	**45323**	**91324**	**52676**	**67520**	**256843**	**487012**

Adding Slicers to the PivotTable

Slicers provide buttons that you can click to filter PivotTables they are linked to. They clearly indicate current filters and hence can be used to get a better view of the items data was filtered for.

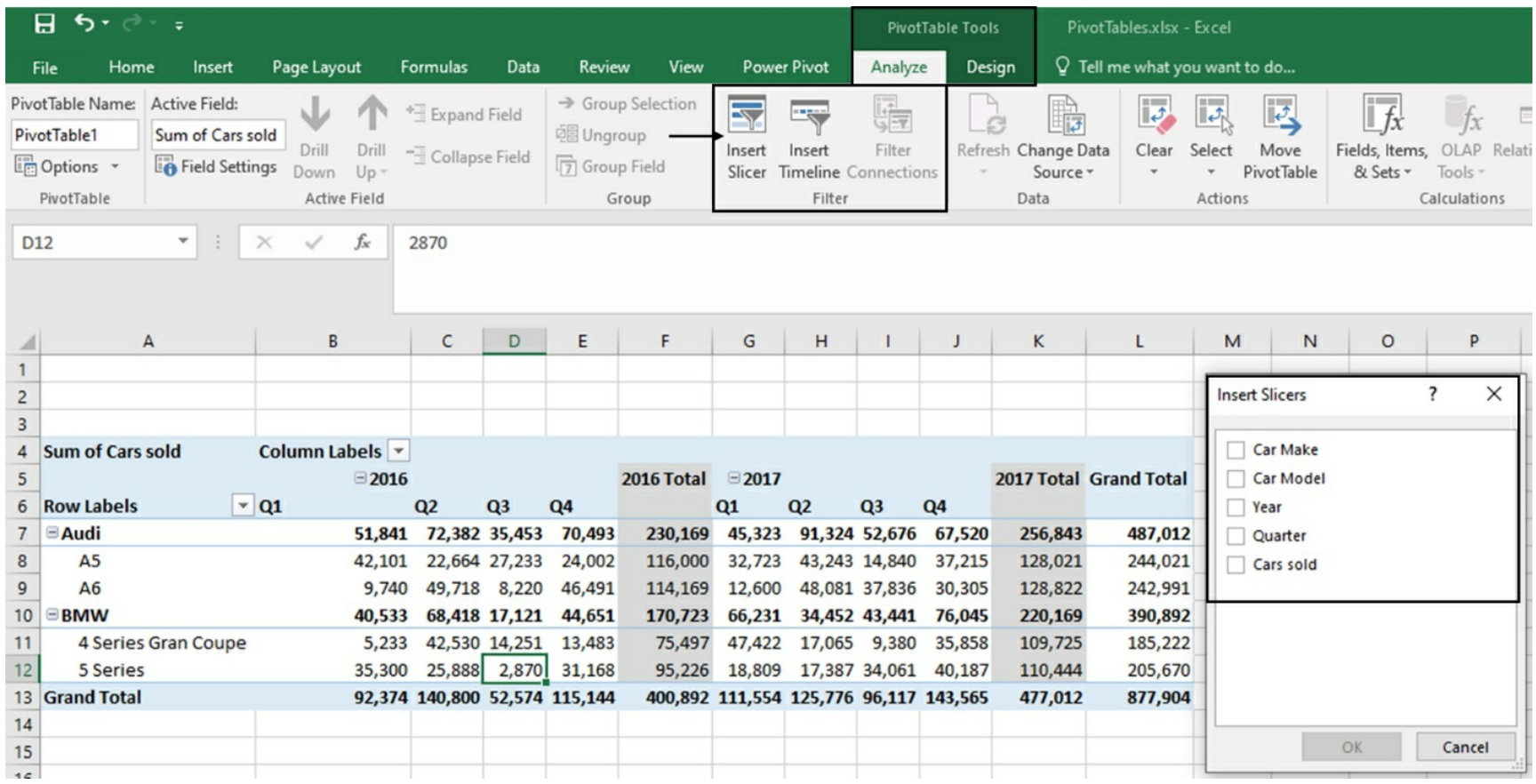
Practice

14. Click anywhere on the PivotTable. This will display *PivotTable Tools* options on the top ribbon.

15. Under *PivotTable Tools*, go to *Analyze* tab and under the *Filter* group, click on *Insert Slicer* icon.

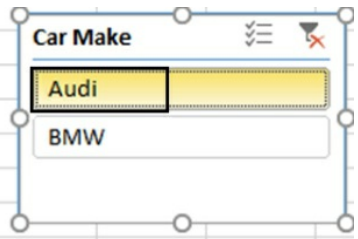
16. The *Insert Slicers* box pops up and contains all the fields from the data.

17. Check the field *Car Make* and click *OK*. A slicer with both values for Car Make (Audi and BMW) selected by default. *Slicer Tools* is displayed on the ribbon to provide options to modify the slicer.



18. Select *Audi* for *Car Make*. The selected item in the slicer gets highlighted. PivotTable report will be summarized for Audi.

Average of Cars sold	Column Labels										
	2016				2016 Total	2017				2017 Total	Grand Total
Row Labels	Q1	Q2	Q3	Q4		Q1	Q2	Q3	Q4		
Audi	25,921	36,191	17,727	35,247	28,771	22,662	45,662	26,338	33,760	32,105	30,438
A5	42,101	22,664	27,233	24,002	29,000	32,723	43,243	14,840	37,215	32,005	30,503
A6	9,740	49,718	8,220	46,491	28,542	12,600	48,081	37,836	30,305	32,206	30,374
Grand Total	25,921	36,191	17,727	35,247	28,771	22,662	45,662	26,338	33,760	32,105	30,438



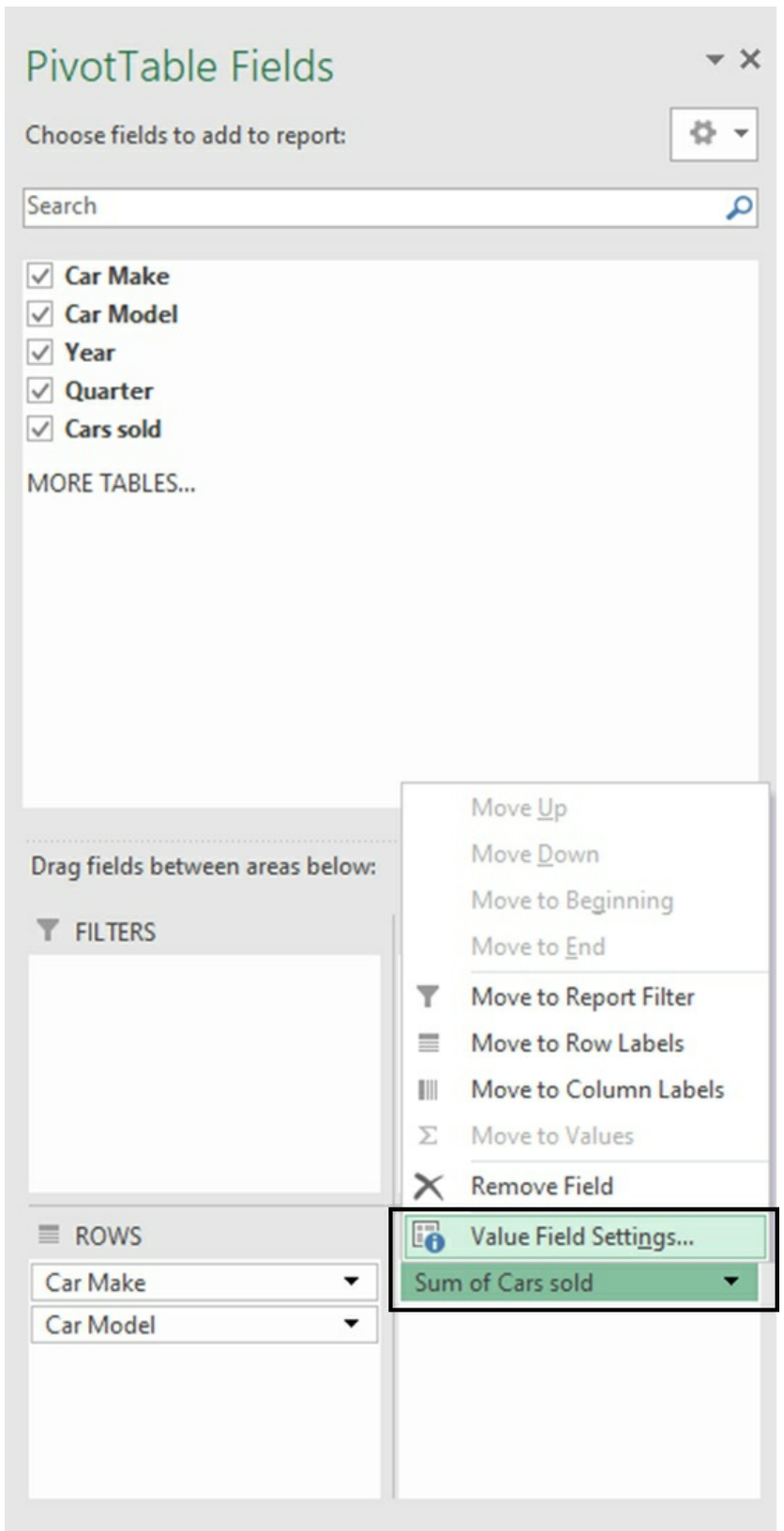
19. Now press the Ctrl key and select both *Audi* and *BMW* (while keeping the Ctrl key pressed). PivotTable report will be summarized for both car makes.

Making changes to a field in the Values Area

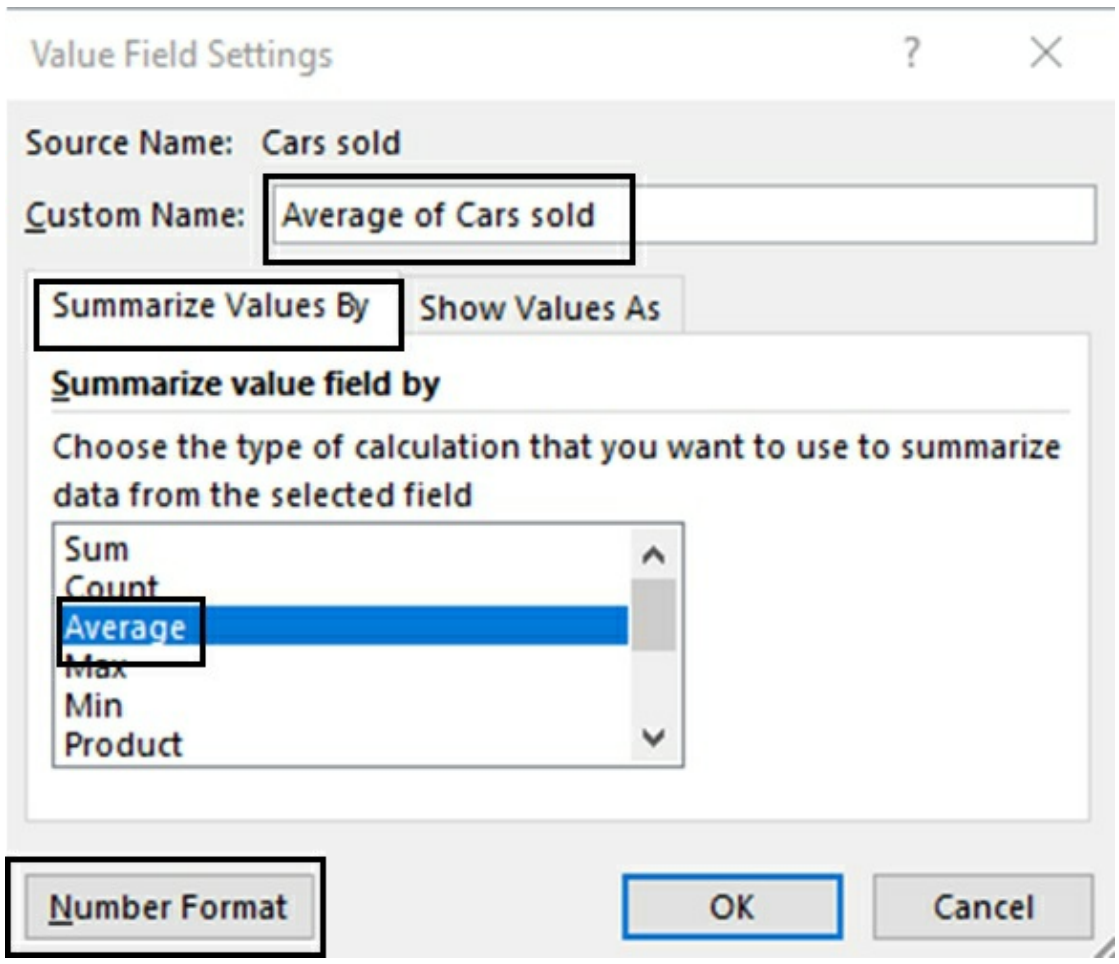
In our practice, we have summarized the report using the default, the *Sum of Cars Sold*. However, you can also use other calculations as needed.

Practice

20. Go back to the PivotTable and on the *PivotTable Fields* pane, click on *Sum of Cars Sold* field. From the drop-down that appears, select *Value Field Settings*.

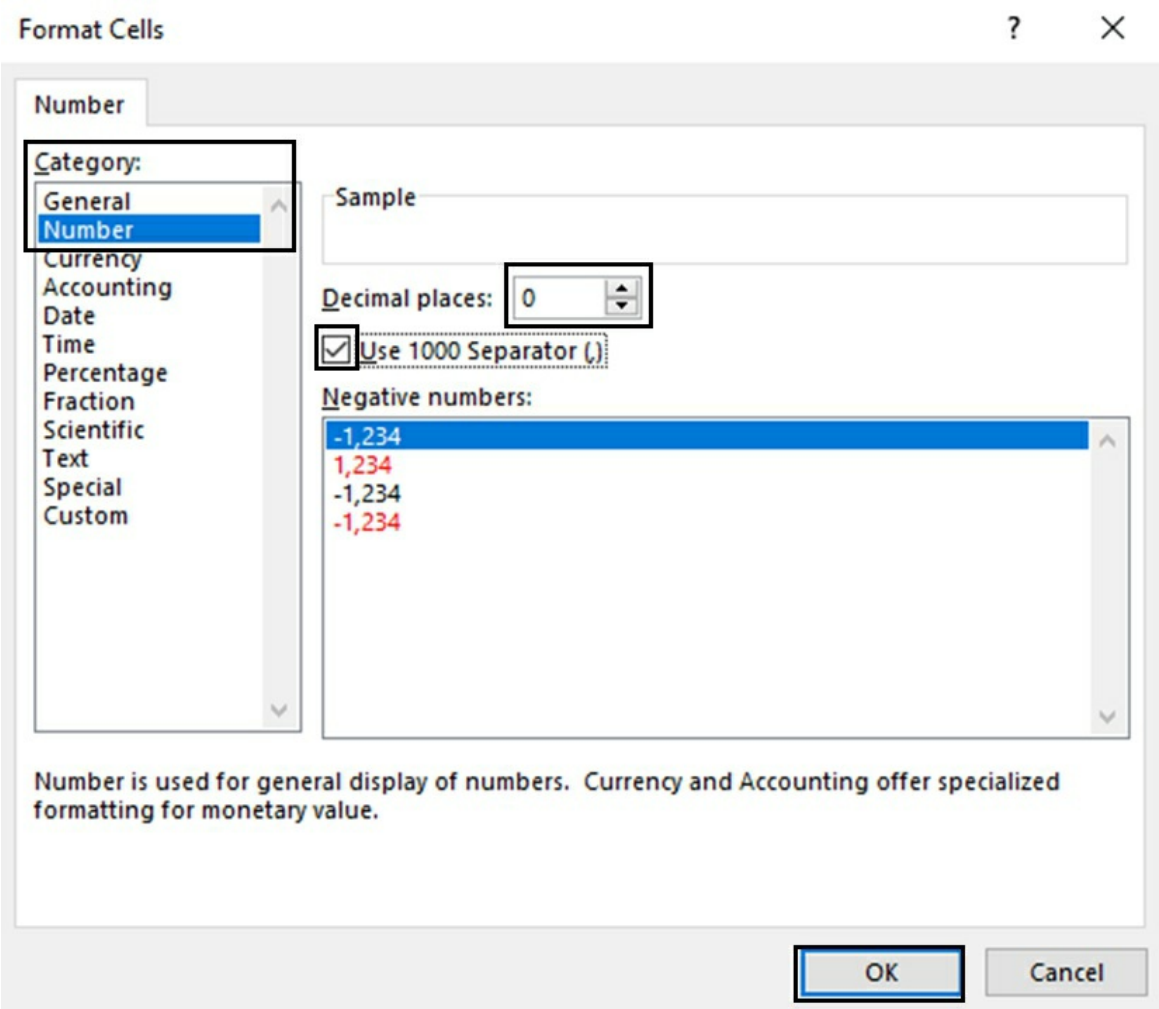


21. A *Value Field Settings* box appears. On *Summarize Values By* tab, for *Summarize value filed by*, study types of calculations available that can be applied to the selected filed and select *Average*. Notice that the *Custom Name* of the field automatically changes to *Average of Cars Sold*.



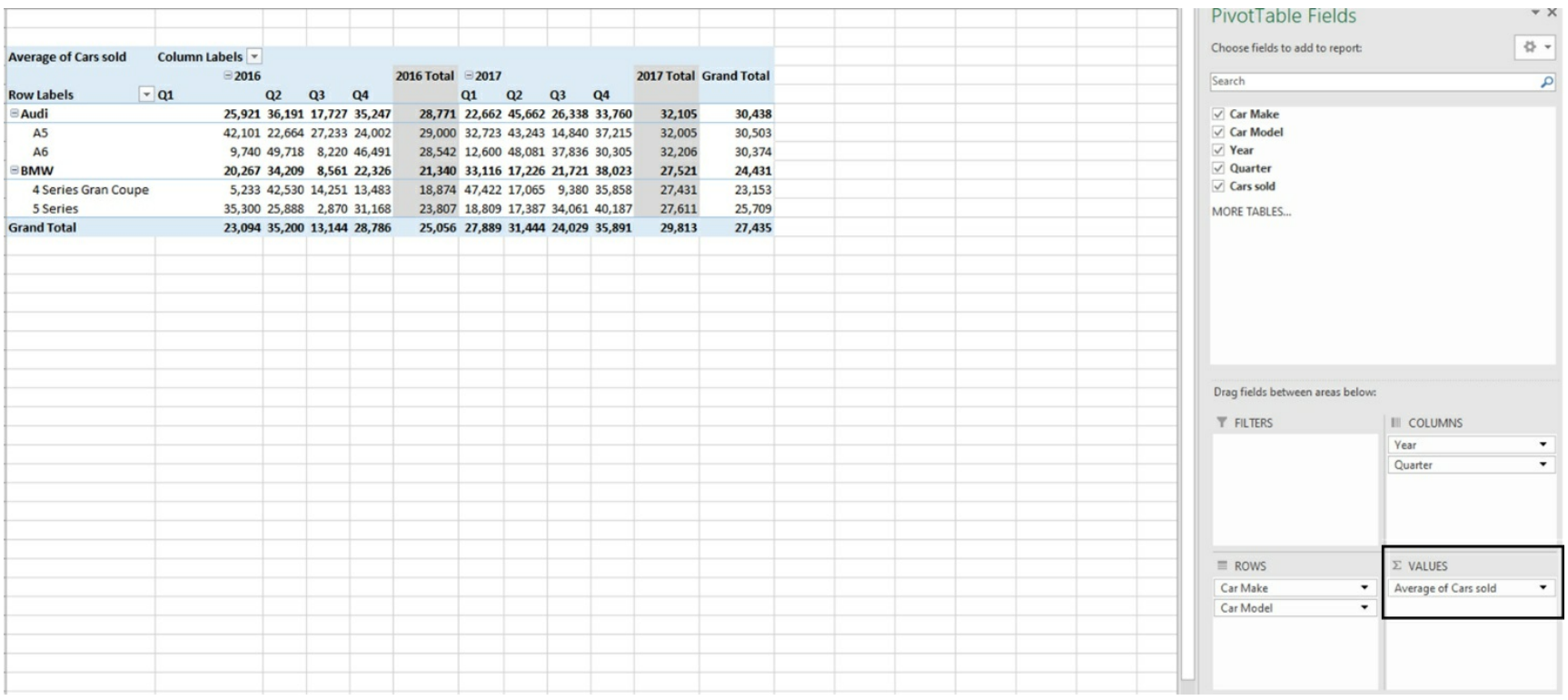
22. Now, click on the *Number Format* icon on bottom left of the box to assign a format to the calculated values.

23. A *Format Cells* box appears. For *Category*, select *Number*. Change *Decimal places* to *0* and check *Use 1000 Separator (,)* box. Click *OK* to close the box.



24. Now, click *OK* again to close the *Value Field Settings* box.

25. The PivotTable now summarizes the report by average of cars sold.



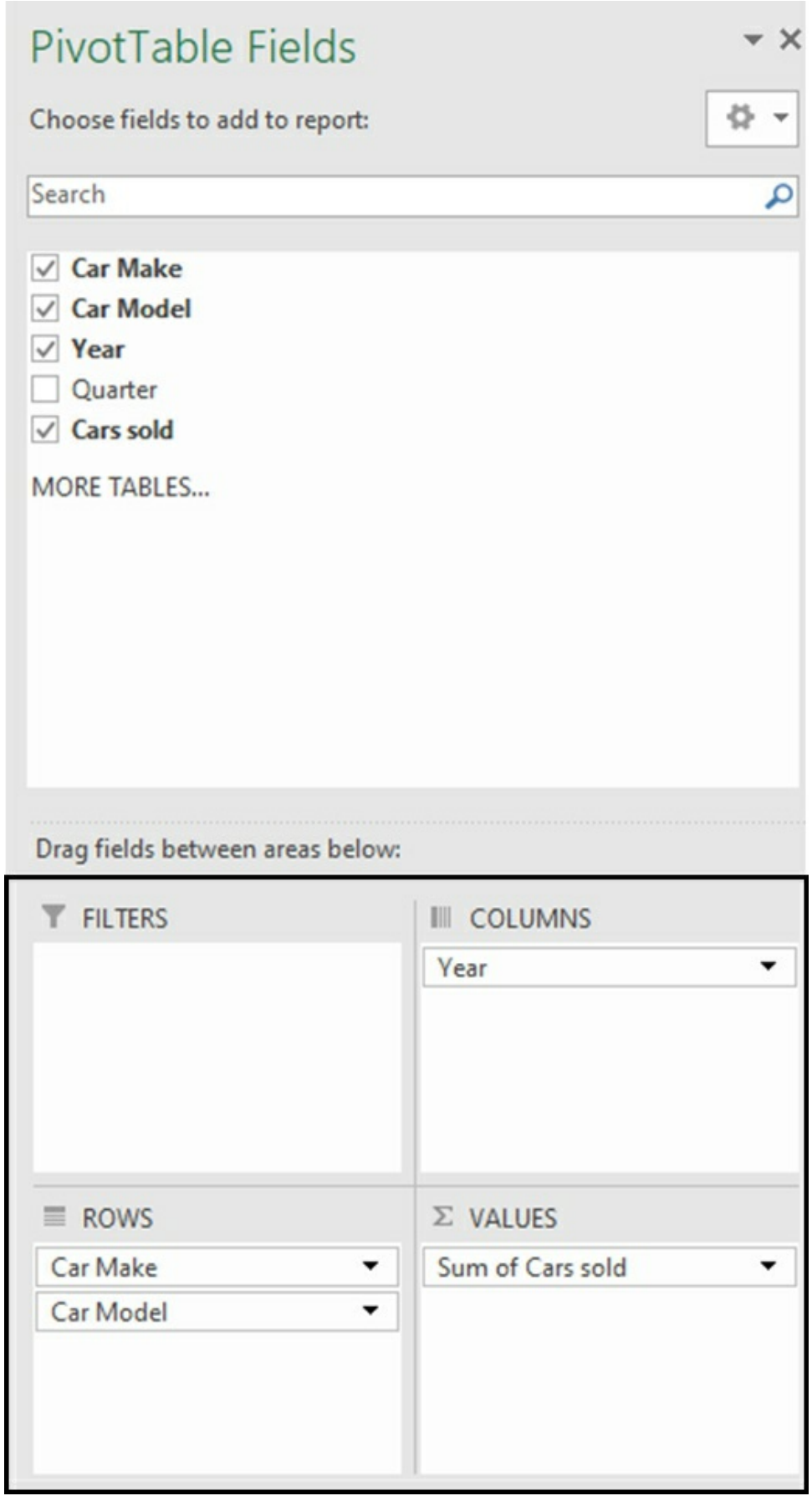
Average of Cars sold	Column Labels										
	2016				2016 Total	2017				2017 Total	Grand Total
Row Labels	Q1	Q2	Q3	Q4		Q1	Q2	Q3	Q4		
Audi	25,921	36,191	17,727	35,247	28,771	22,662	45,662	26,338	33,760	32,105	30,438
A5	42,101	22,664	27,233	24,002	29,000	32,723	43,243	14,840	37,215	32,005	30,503
A6	9,740	49,718	8,220	46,491	28,542	12,600	48,081	37,836	30,305	32,206	30,374
BMW	20,267	34,209	8,561	22,326	21,340	33,116	17,226	21,721	38,023	27,521	24,431
4 Series Gran Coupe	5,233	42,530	14,251	13,483	18,874	47,422	17,065	9,380	35,858	27,431	23,153
5 Series	35,300	25,888	2,870	31,168	23,807	18,809	17,387	34,061	40,187	27,611	25,709
Grand Total	23,094	35,200	13,144	28,786	25,056	27,889	31,444	24,029	35,891	29,813	27,435

Adding multiple copies of a field in the Values Area

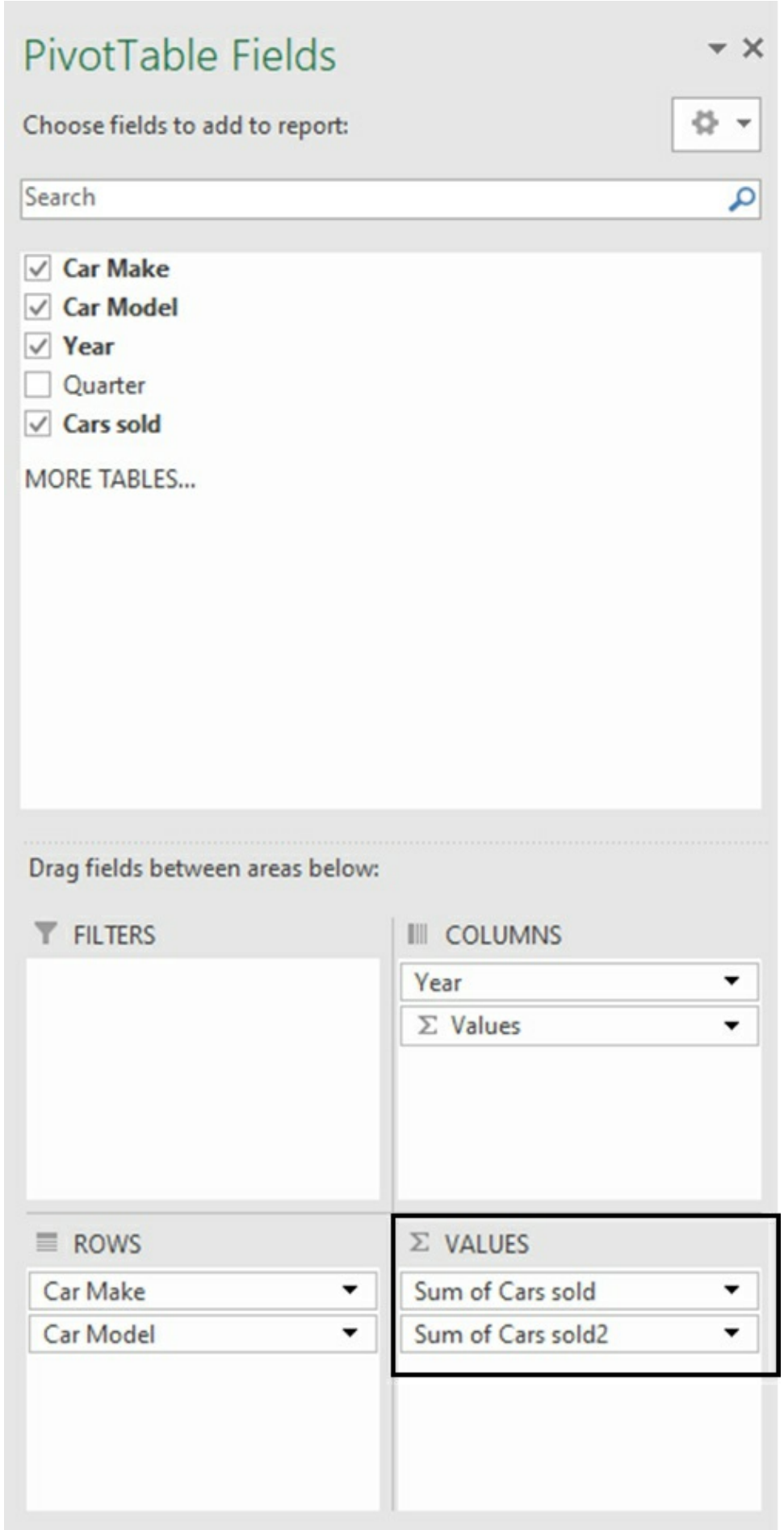
Sometimes, you'll want to summarize the same field with more than one type of calculations. For example, suppose in addition to summarizing the report by sum of cars sold, you also want to calculate percentages of the total cars sold. This is done by adding the field twice and altering the calculation.

Practice

26. Revert to the original PivotTable that summarizes the report by the sum of cars sold. To reduce the width of PivotTable, remove *Quarter* from *Columns* area by either clicking on it selecting *Remove Field* or simply selecting it and dragging out of the area.



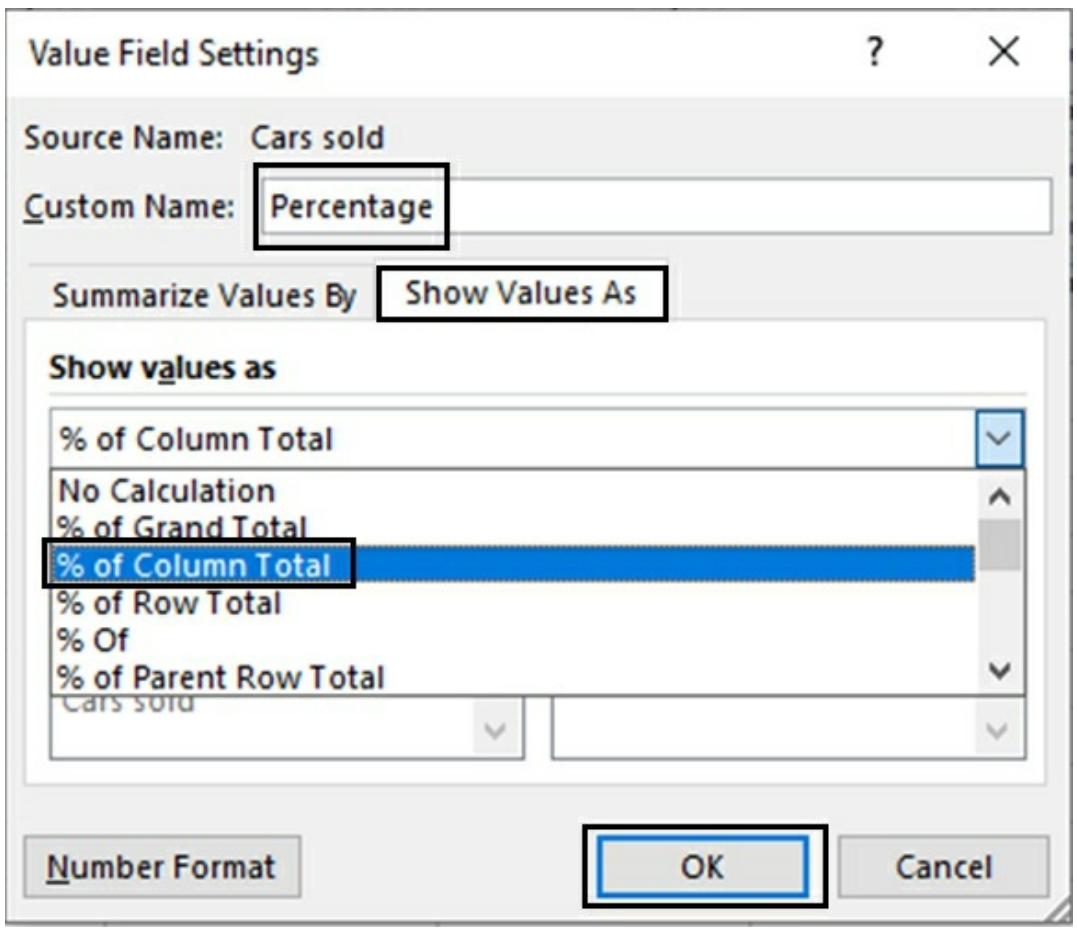
27. Drag the *Cars Sold* field (again) to the *Σ Values* area.



28. Click on the newly added field. From the drop-down that appears, select *Value Field Settings*.

29. On the *Value Field Settings* box, go to *Show Values As* tab. Change *Custom Name* to *Percentage*.

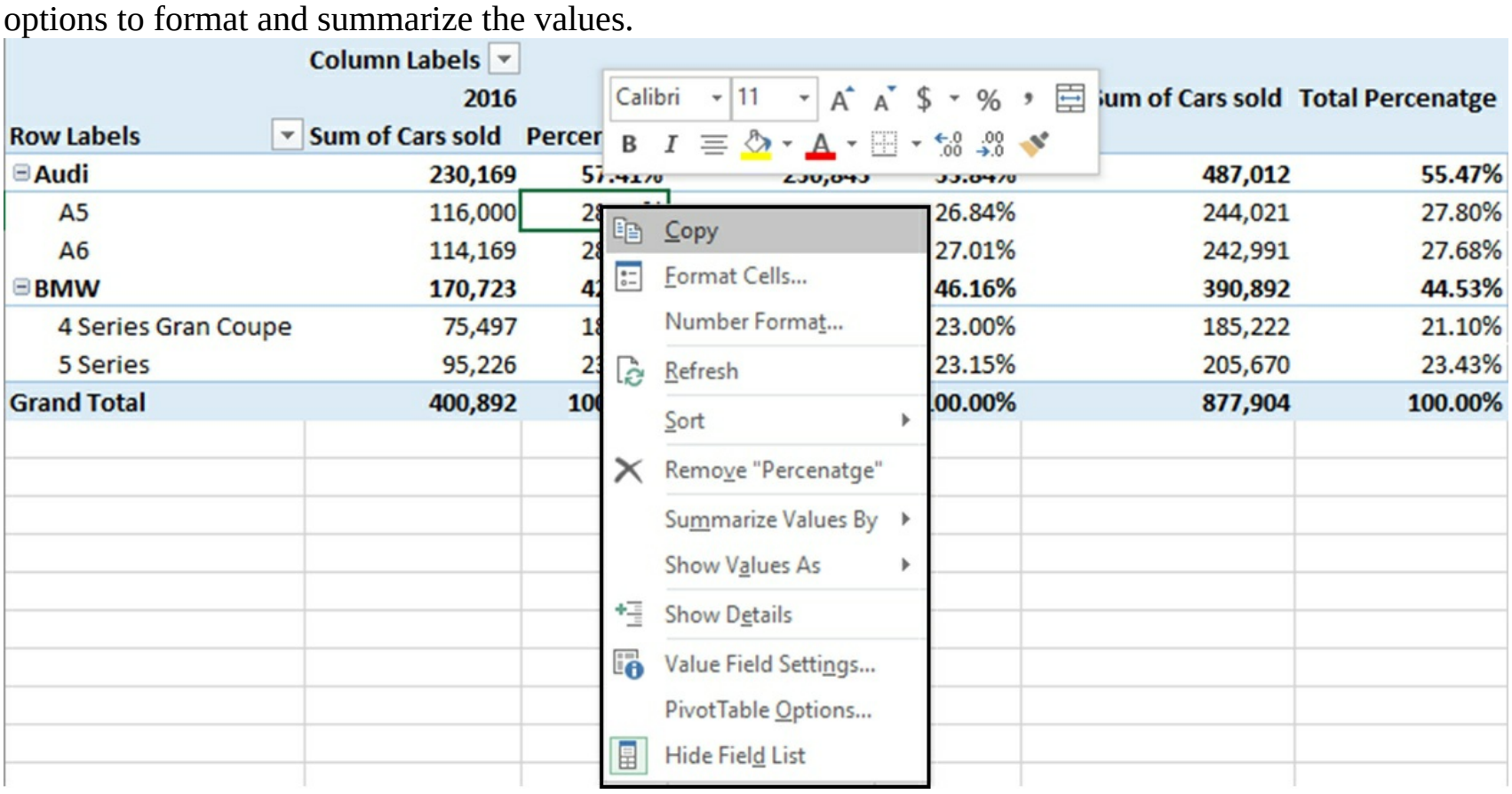
30. Under *Show Values As*, click on the down arrow to display a dropdown. From the dropdown, study types of calculations available that can be applied to the selected filed and select *% of Column Total*. Click *OK*.



31. The report is now summarized by sum as well as the percentage of cars sold.

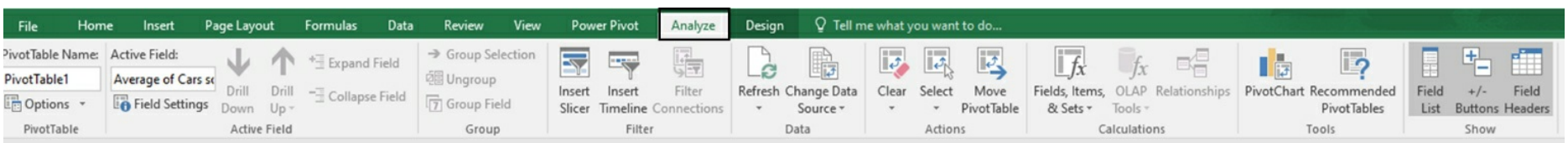
Row Labels	Column Labels 2016 Sum of Cars sold	2016 Percenatge	2017 Sum of Cars sold	2017 Percenatge	Total Sum of Cars sold	Total Percenatge
Audi	**230,169**	**57.41%**	**256,843**	**53.84%**	**487,012**	**55.47%**
A5	116,000	28.94%	128,021	26.84%	244,021	27.80%
A6	114,169	28.48%	128,822	27.01%	242,991	27.68%
BMW	**170,723**	**42.59%**	**220,169**	**46.16%**	**390,892**	**44.53%**
4 Series Gran Coupe	75,497	18.83%	109,725	23.00%	185,222	21.10%
5 Series	95,226	23.75%	110,444	23.15%	205,670	23.43%
Grand Total	**400,892**	**100.00%**	**477,012**	**100.00%**	**877,904**	**100.00%**

Please note that *Value Field Settings* are also available on right-clicking on a value field along with other options to format and summarize the values.

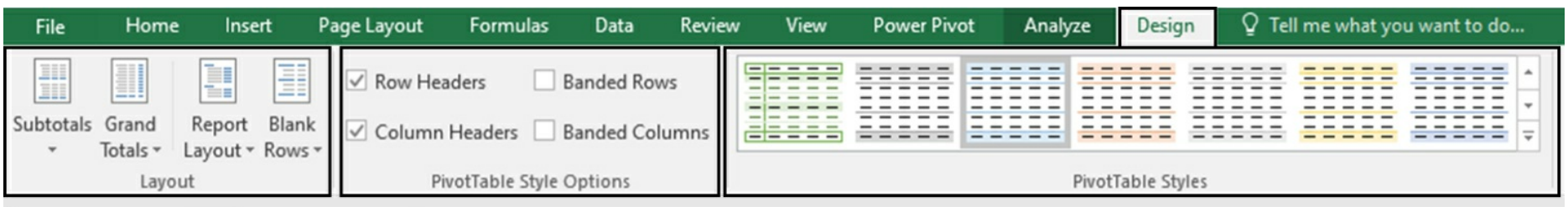


Changing PivotTable Design

You might have already noticed that clicking on the PivotTable displays PivotTable Tools on the top ribbon. There are two tabs under the PivotTable Tools – Analyze and Design. Commands on the **Analyze** tab help to explore the data in the PivotTable.



Design tab commands provide various layout, style options and styles.



1. **Layout:** You can select PivotTable layout based on your preferences for the following −

   a. **Subtotals:** By default, subtotals are placed at the top of a group. You have options available to change this placement of the subtotals or not display them at all.

   b. **Grand Totals:** By default, grand totals are displayed for both rows and columns. You have options available to display them for rows or columns only or not display at all.

   c. **Report Layout:** By default, a PivotTable is displayed in Compact Form layout. The **Compact Form** displays items from different row area fields in one column and uses indentation to distinguish between the items from different fields. Two other layouts are Tabular and Outline. **Tabular Form** displays one column per field and provides space for field headers. **Outline Form** is similar to tabular form but displays subtotals at the top of every group.

   d. **Blank Rows:** To make your PivotTable report more readable, you can insert a blank row after each item. You can remove the blank rows anytime later.

2. **Styles and style options**: You have several PivotTable styles to choose from based on your preferences. Additional style options are available to further enhance the design of a PivotTable.

Practice

32. Click anywhere on the PivotTable. This will display *PivotTable Tools* options on the top ribbon.

33. Go to the *Design* tab and explore try all layout and style options we just discussed and structure your PivotTable in a format you like.

3.2 PivotTable Tips and Tricks

PivotTables are a great reporting and data analysis feature available through Excel. As you learnt in the previous chapter, you can create a basic PivotTable in no time and begin interactively exploring your data to draw key insights and conclusions. Below are some tips and tricks to help you use this extremely useful feature efficiently.

1. Organize your data effectively

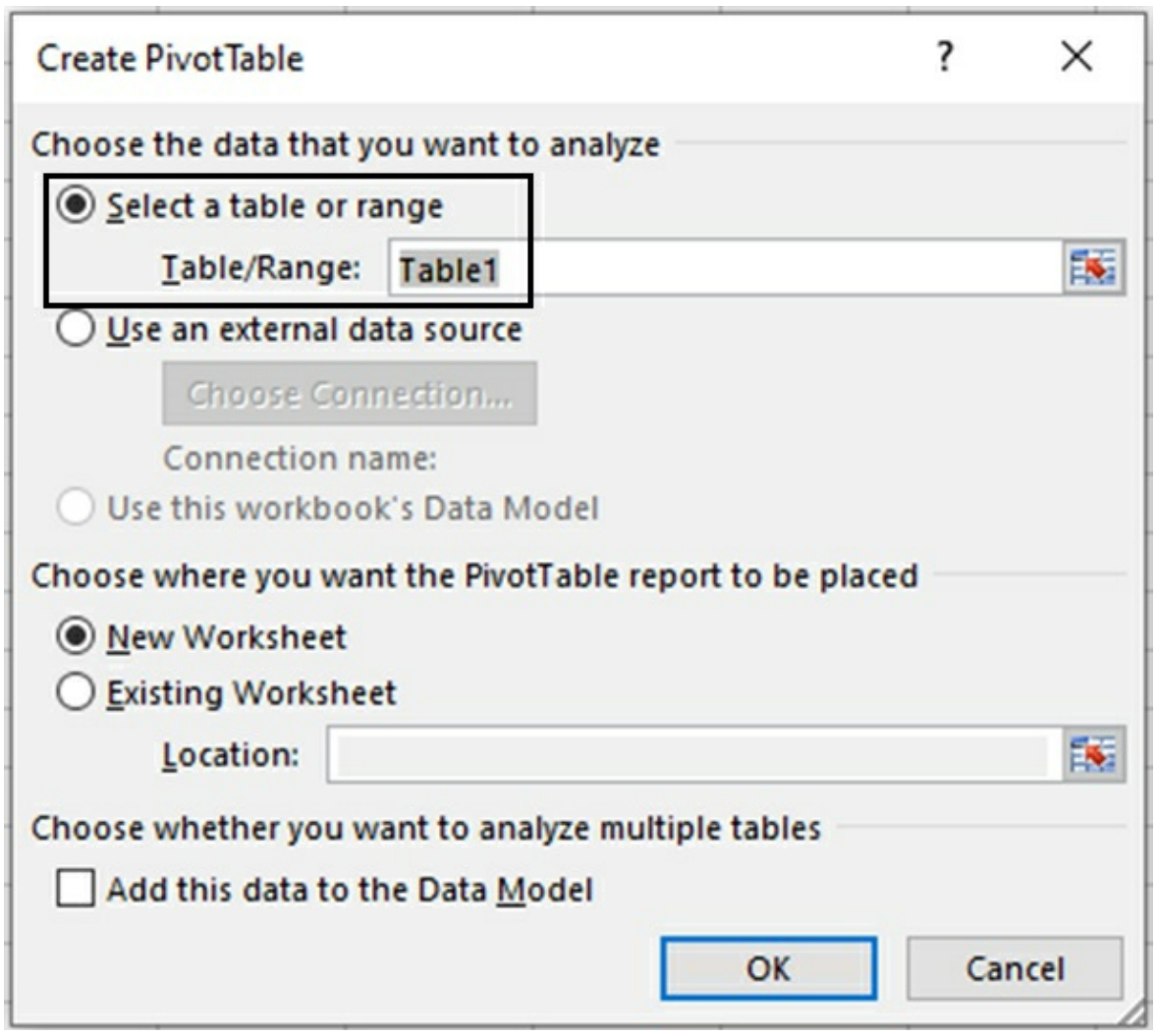
Before creating the PivotTable, ensure that your data is organized in rows and columns. Each column should have headers (on first row only) to indicate the data it contains. Each row should contain information on one record. The columns should contain only one format of data, such as dates, currency, or text. The source data shouldn't have any totals, subtotals or blank rows.

2. Plan your PivotTable report

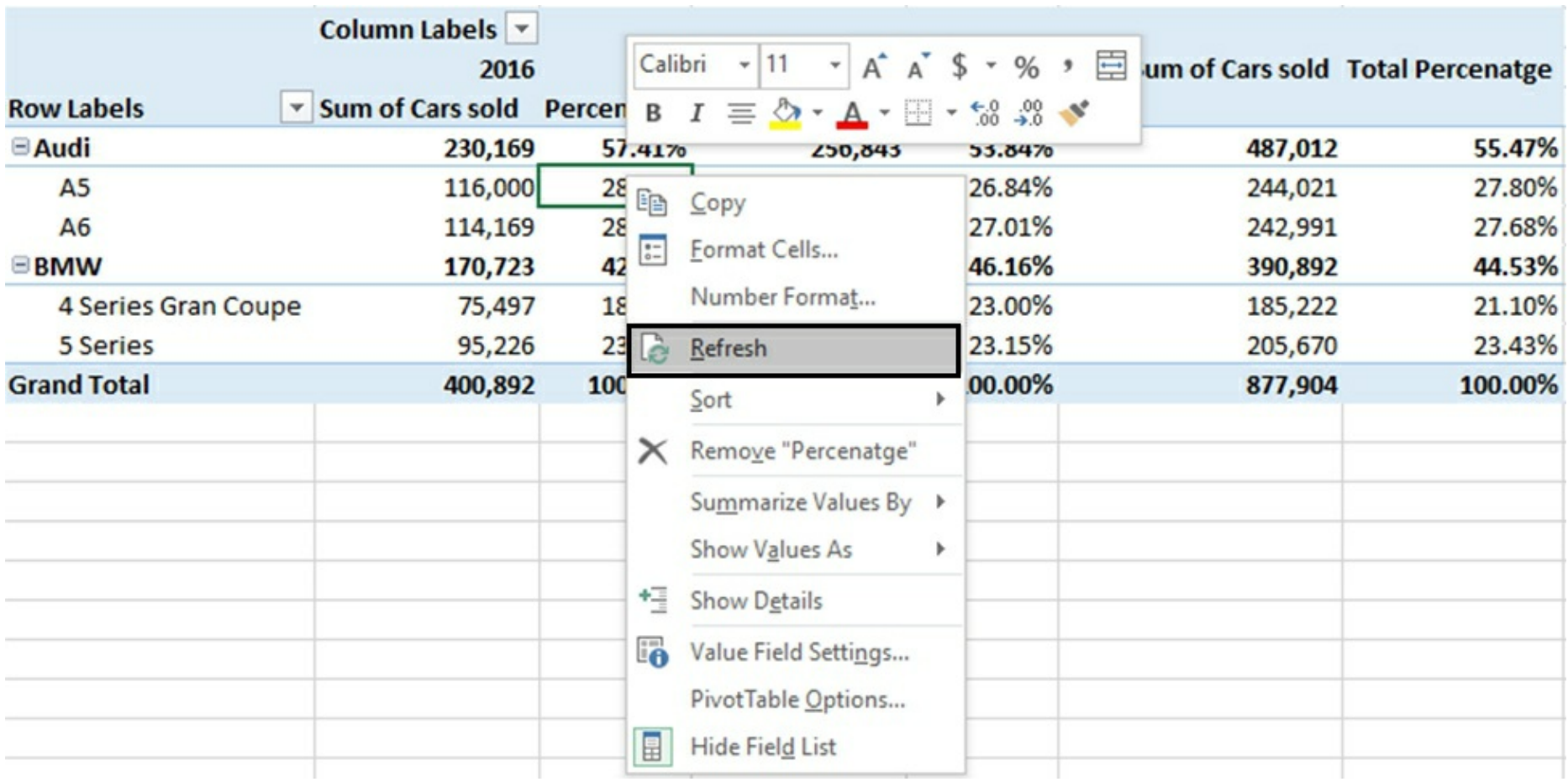
Just like any other reporting and data analysis endeavour, think about the questions you need to answer before you start building your PivotTable. Based on your purpose, sketch out an initial design. This will help guide you through the large number of choices you have available through the PivotTables. Keep things focused on the questions you need to answer. While dragging and removing fields on the field pane and watching PivotTable dynamically change will also be helpful, having an initial design will save you a lot of unproductive rabbit holes.

3. Make the PivotTable dynamic

If you envision appending more rows or columns to the source data after you have built your initial PivotTable, it will be a better idea to convert your source data into a named table. We have discussed steps to create and name a table in Chapter 2.3. If you want to insert a PivotTable based on a named table, click anywhere on the table, go to *Insert* tab and under the *Tables group*, click on *PivotTable*. The name of the table will automatically display as source data on the *Create PivotTable* box.

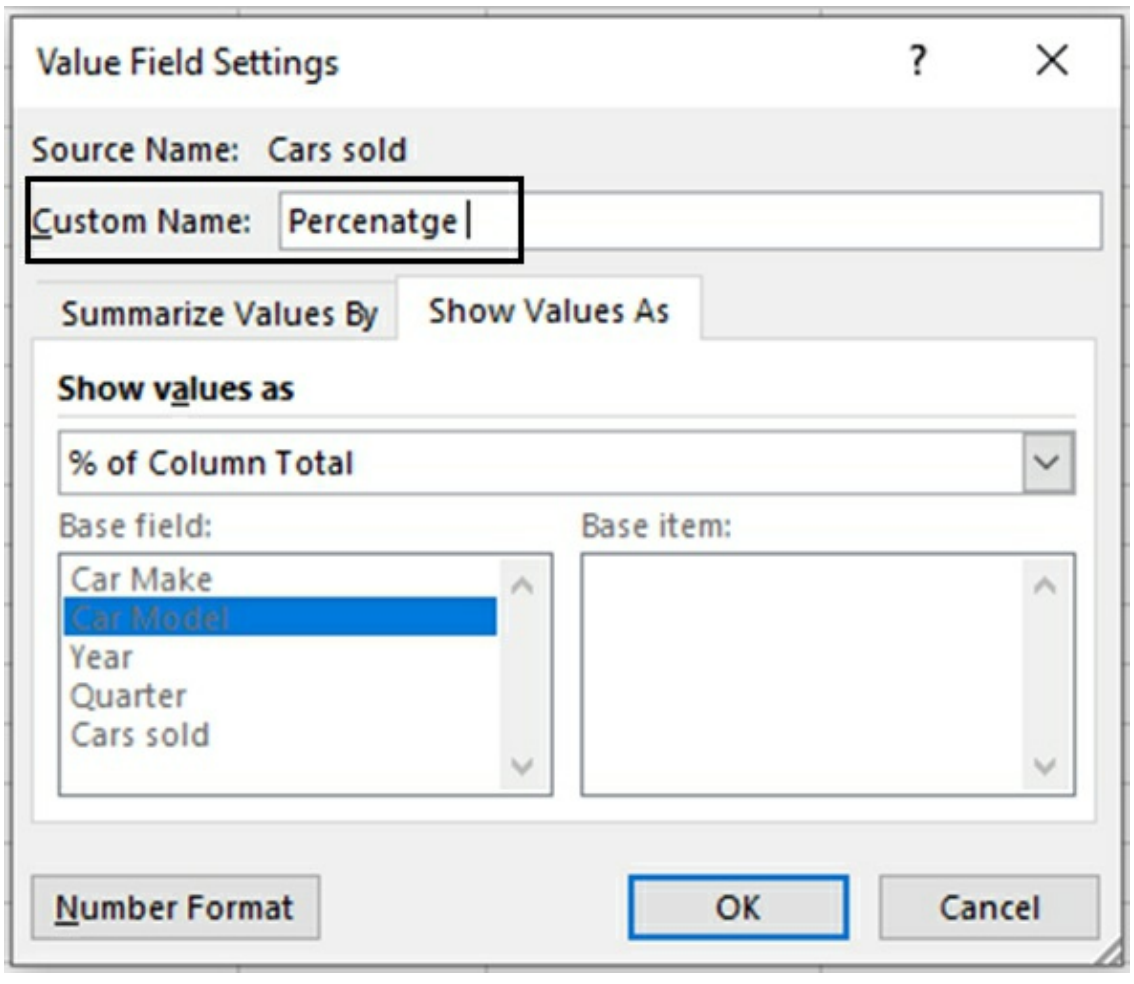


Each time you append more data to your table, the table range will automatically adjust to include appended data. To include newly appended data into your PivotTable, all you have to do is right click anywhere on the PivotTable and select *Refresh*.

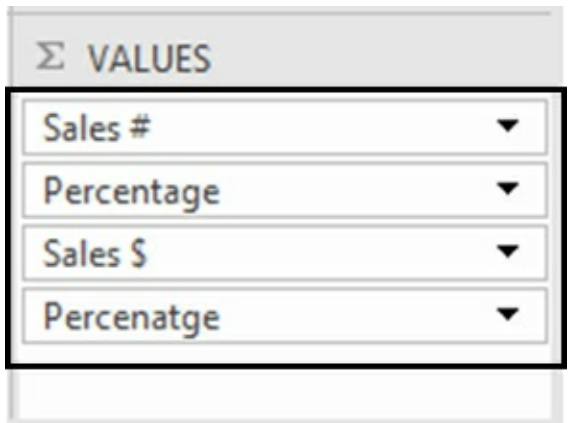


4. Add spaces to field names if want to repeat them multiple times

When you try to rename fields, you might run into a problem if you try to use exactly the same field name that already appears in the data. For example, suppose you have *sales #* and *sales $* as two separate fields in your PivotTable and you want to list their respective percentages next to them in additional columns. As we have studied in the previous chapter, you bring in *sales #* into the *Σ Values* area right below the *sales #*, change it to calculate appropriate percentage, change the name of the filed to Percentage. This will lead to a *Percentage* column placed between *sales #* and *sales $* columns. However, if you want to insert such a column for sales $ and want to name it same Excel will show an error that the field already exists. To avoid, this, just add a space to the end of the field name (*Percentage* in this case).



Excel will no longer show any error and you won't see the difference. If you want to use a filed name a third time, simply add another space.



5. Copy your PivotTable as needed

What if you have built your first PivotTable, but you want to see a different view of the same data and compare them side by side. Instead of creating a new PivotTable from scratch, you can copy and paste the previous PivotTable on the same sheet or a different sheet. The latter can also be done by just copying the sheet containing the existing PivotTable. You can then rearrange fields of the new PivotTable to create the new view.

6. Show or hide subtotals and grand totals

PivotTables show totals for both rows and columns by default. There may be times, however, when the inclusion of subtotals and grand totals either doesn't make sense or just hinders a clear view of your report. If this is the case, you can easily disable one or both of them through options available through *Subtotals* and *Grand Totals* icons under the *Layout* group on *Design* tab of *PivotTable Tools*. We have already covered this in the previous chapter.

7. Get rid of default headings

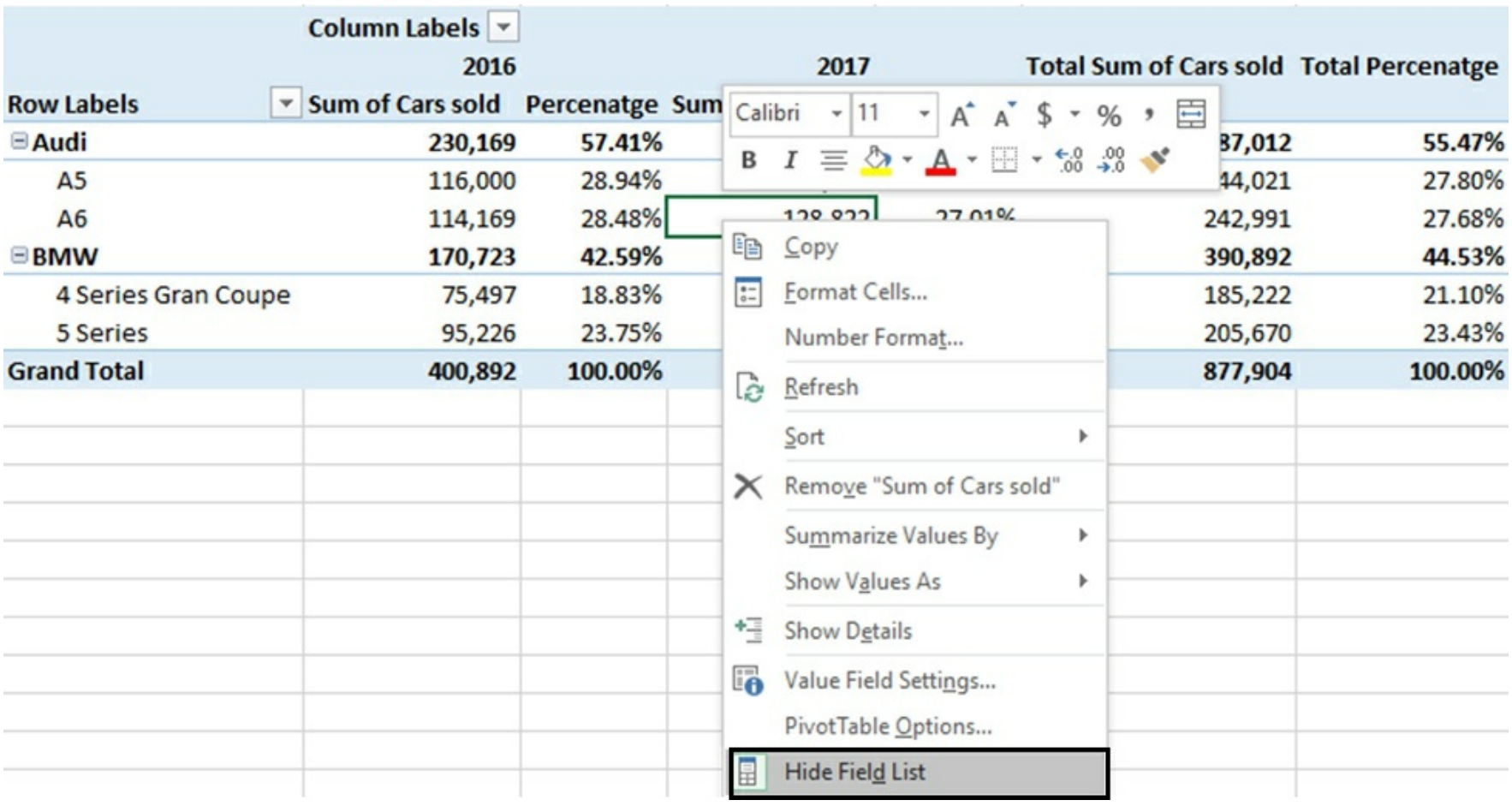
The default layout for the new PivotTable is the compact form. This layout displays "Row Labels" and "Column Labels" as default headings in the PivotTable. To get rid of these headings, one way is to change the PivotTable layout from compact to outline or tabular forms. This will cause the PivotTable to display the actual field names as headings in the PivotTable.

8. View data behind any calculated value

You can easily view or extract the data that makes up a calculation, for example, a sum, on a PivotTable by double clicking on it. Excel will add a new sheet to your workbook that will contain the data used to calculate the value.

9. Hide the PivotTable Fields pane

If you're planning to distribute the PivotTable report, you can choose to hide the *PivotTable Fields* pane on the right. Simply right click anywhere on the PivotTable and select *Hide Field List*. This option is also available through *Analyze* tab of *PivotTable Tools* under the *Show* group.

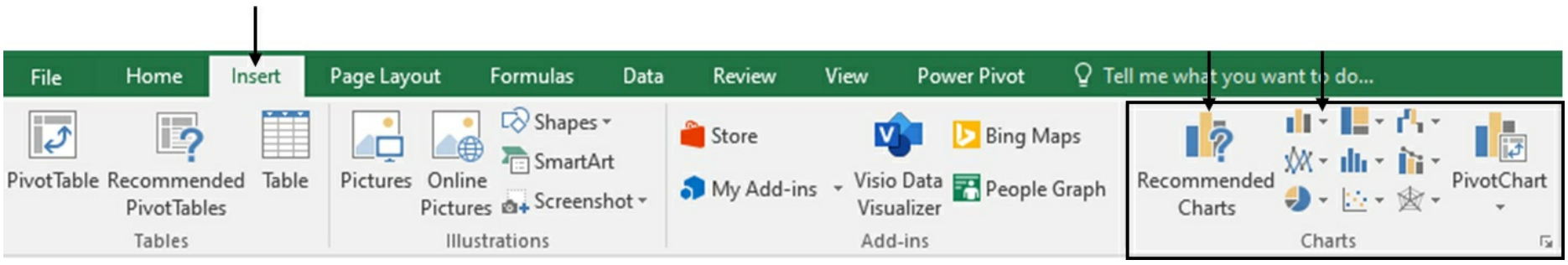


3.3 Visualize Data with Charts

In Microsoft® Excel, a chart is a tool to visualize data graphically. Graphical visualization allows the audience to perceive trends and patterns in the data more easily. Excel charts can be employed for data analysis, presentation, and reporting. In this chapter, we will learn how to insert and modify charts for various purposes. We will also discuss the features of each chart type that make them better suited for specific tasks.

Inserting a chart

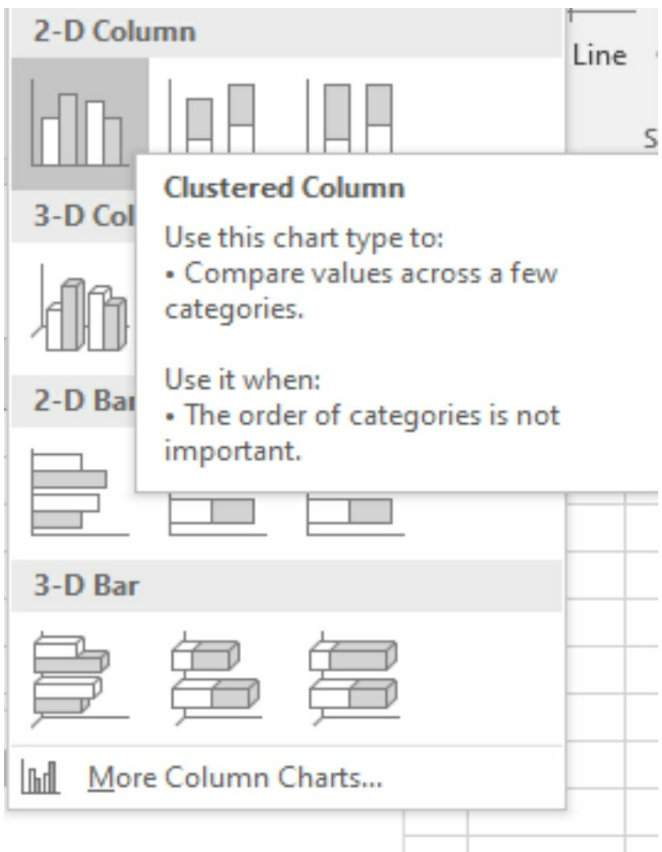
Excel charts are available through the *Insert* tab under the *Charts* group. The *Charts* group is used to choose the desired type of chart to visualize the selected data. The sub-types of each type of chart are available by clicking on the arrow beside the chart icon. Excel also includes an option to browse *Recommended Charts*, which shows you the chart types that best fit your data. By default, the chart will appear directly on the spreadsheet where your data is.



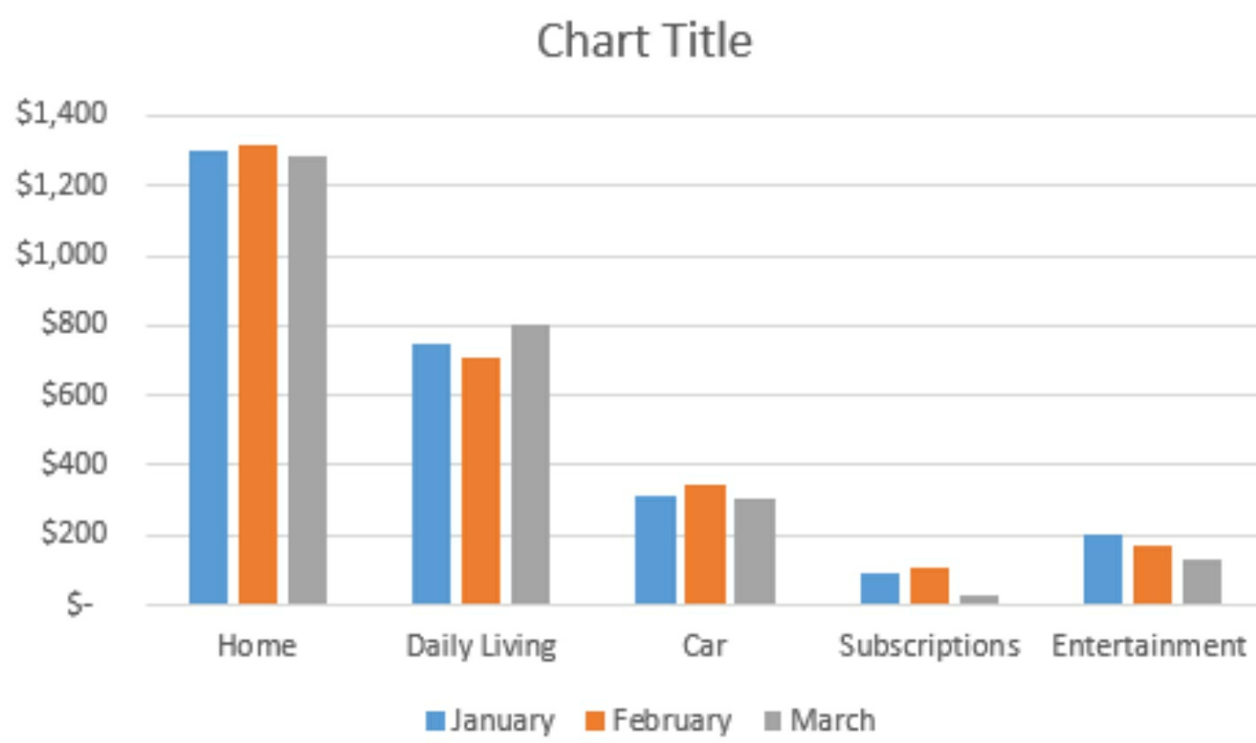
Practice

Let's learn how to insert a clustered column chart. The details about the chart will be discussed later in the chapter. For now, just focus on the steps.

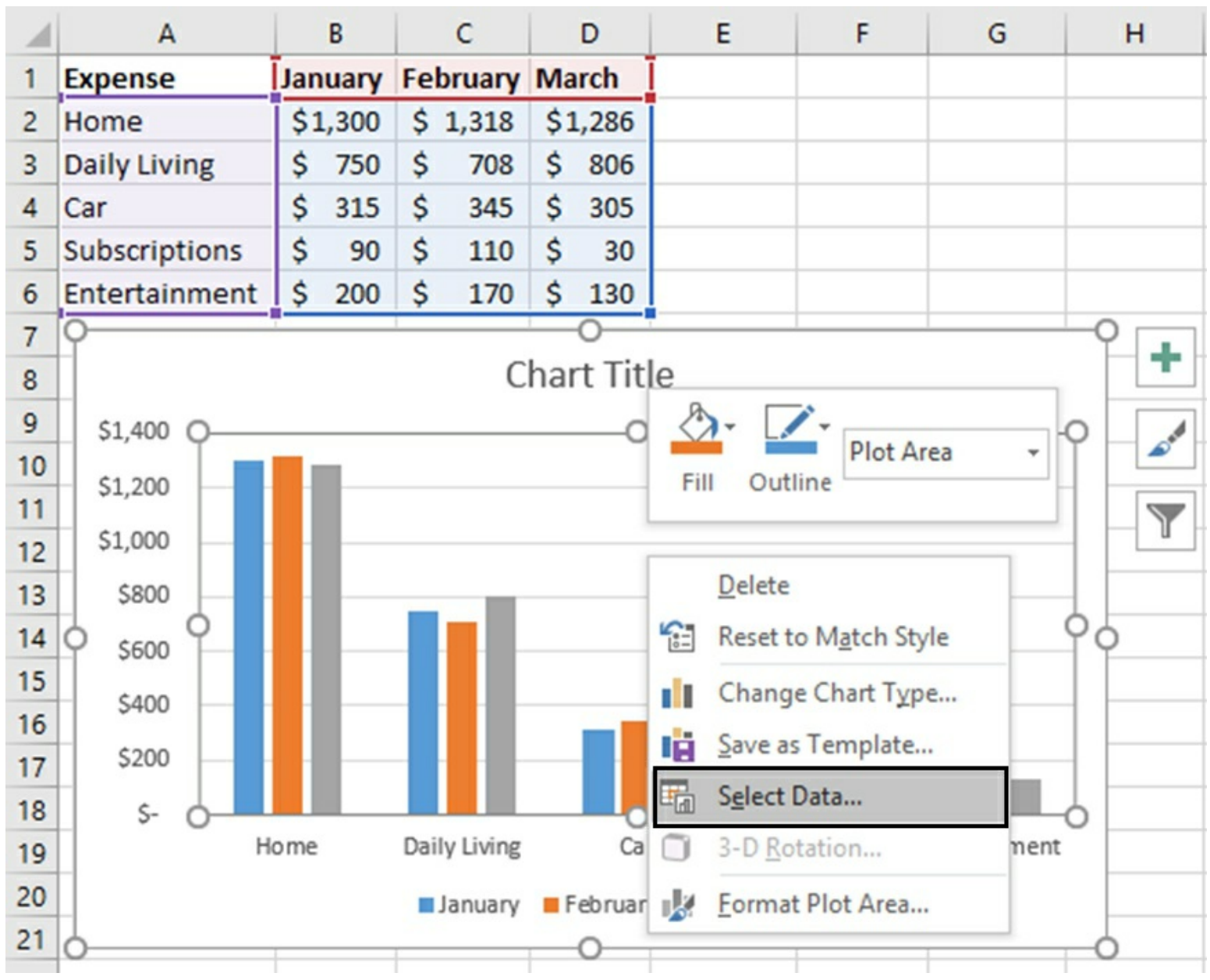
1. Open *Charts_Practice.xlsx* and go to the *Column* sheet. The sheet contains data on personal expenses of an individual for the months of January, February and March broken down by categories.
2. Select the data including the column titles and row labels (A1:D6).
3. Go to the *Insert* tab and under the *Charts* group, click on the *Column Chart* icon. The 2-D column, 3-D Column chart options will be displayed. Hover through the options slowly to view chart names and previews on the sheet.



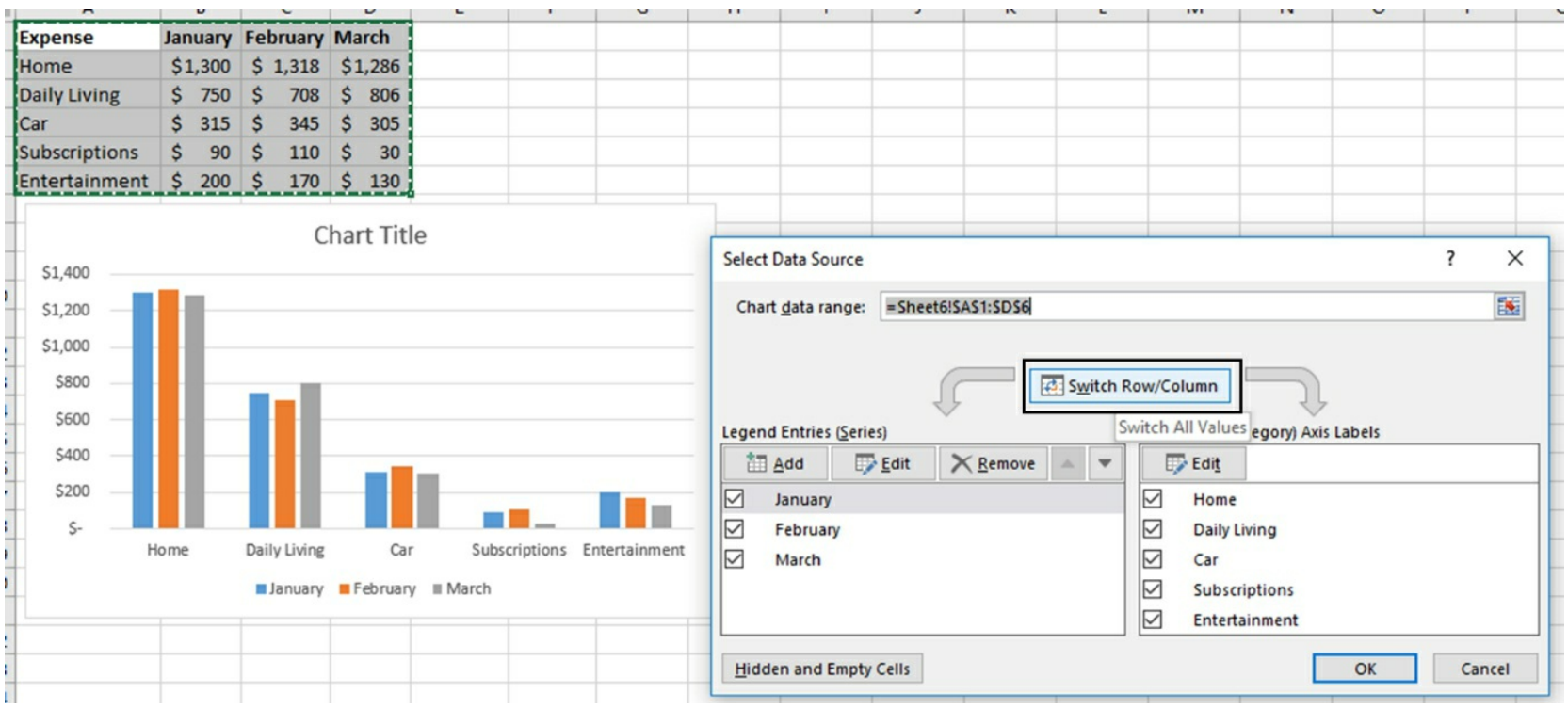
4. Choose *Clustered Column*. The chart will be inserted into the worksheet.



5. In the chart, the expenses are grouped by expense category with columns representing each month. Suppose your aim is to compare different expense categories in the same month. It is a bit difficult to do with this graph. However, we could switch the series and category so the chart will group the data by month with columns for each expense category. To do this, select and right click the chart and choose *Select Data*. The *Select Data Source* window will pop up.



6. Click on *Switch Row/Column*. Now, click OK.



7. The rows and columns will be switched. The chart contains the same data—it's just organized differently to meet your needs. Click on the chart title twice to place it in editing mode and rename it to *2D Clustered Column*.

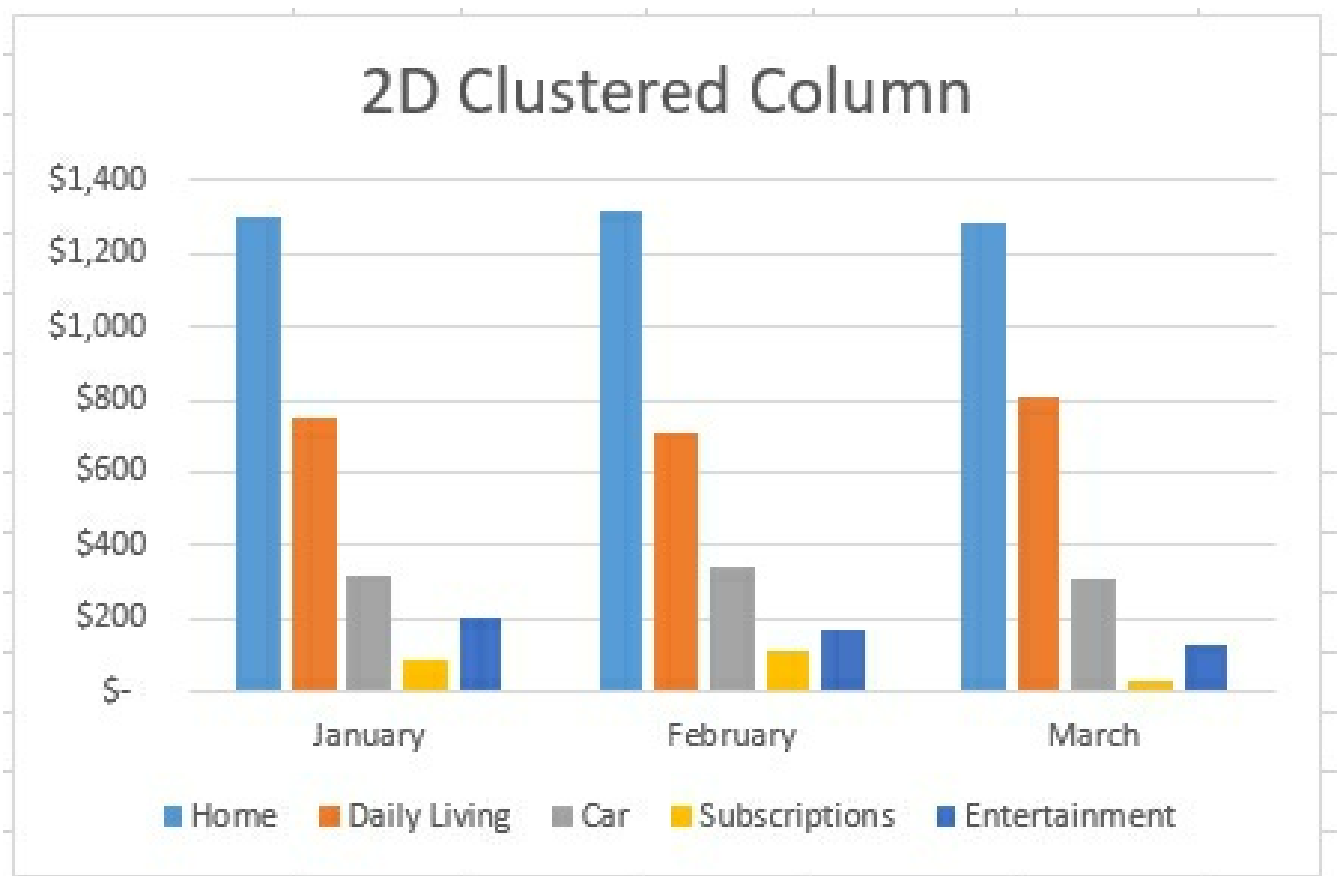


Chart types

We will now discuss below the most popular Excel chart types one by one:

1. Column Charts
2. Line Charts
3. Bar Charts
4. Area Charts
5. Pie Charts
6. Donut Charts
7. Scatter Charts
8. Bubble Charts
9. Tree map Charts
10. Combo Charts

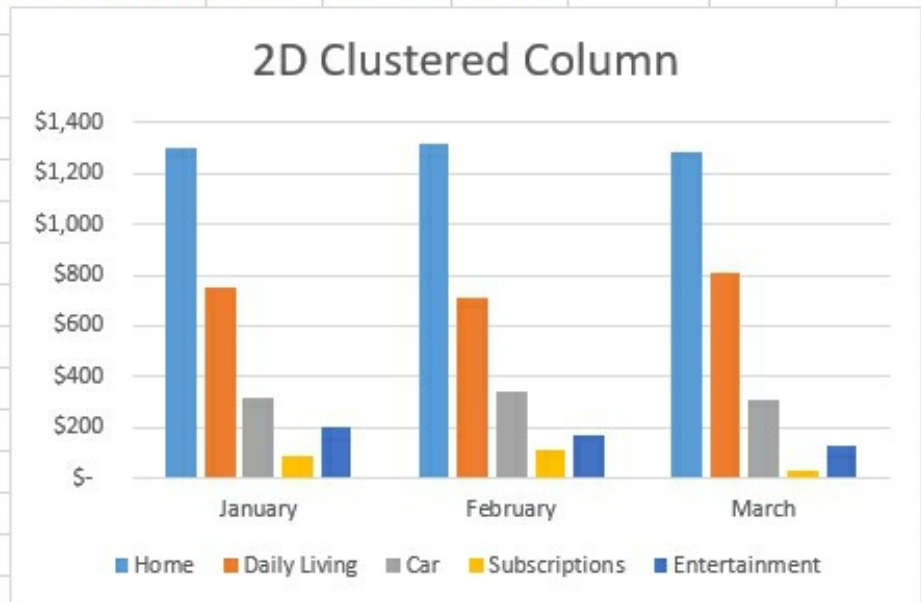
1. Column Charts

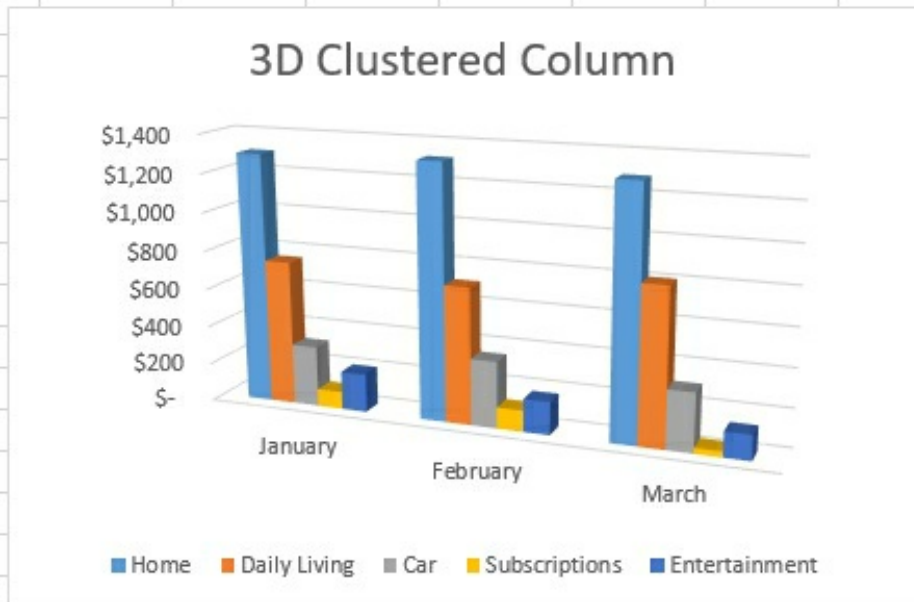
Column charts are one of the most commonly used chart types. They are used to make a comparison across several categories or data series. These categories or data series are shown as vertical columns on the x-axis and the height

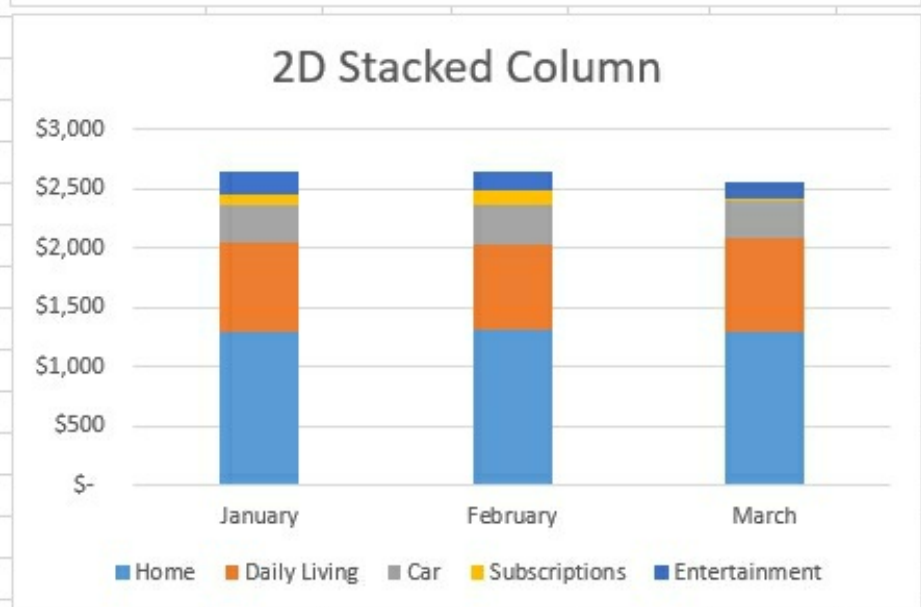
of the column represents their magnitude.

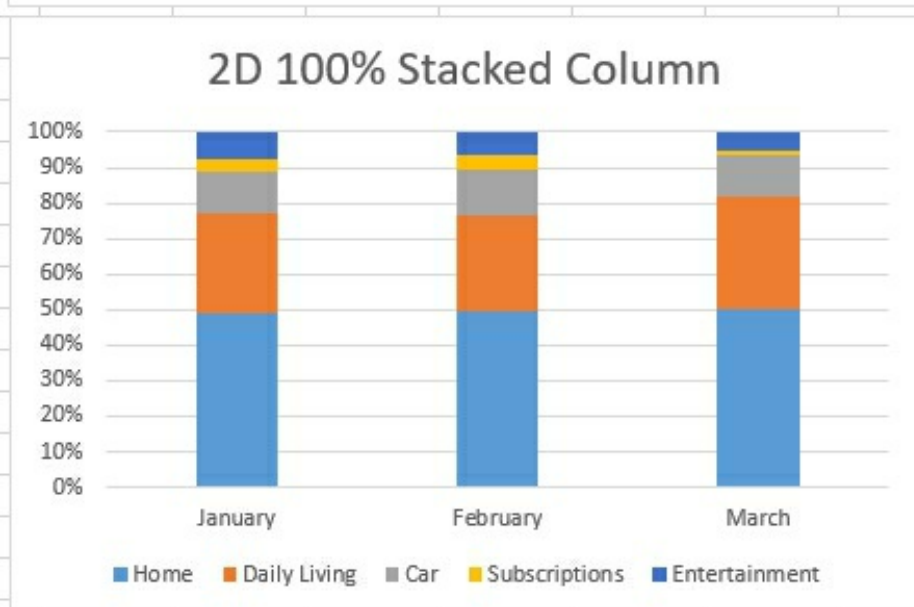
With column charts, it's easy to add data labels at the top of each bar. The data labels should be avoided if there are too many categories.

Expense	January	February	March
Home	$1,300	$ 1,318	$1,286
Daily Living	$ 750	$ 708	$ 806
Car	$ 315	$ 345	$ 305
Subscriptions	$ 90	$ 110	$ 30
Entertainment	$ 200	$ 170	$ 130









Clustered Column Chart

A clustered column chart is used when comparison across multiple series is of more value than the total. In the clustered column chart, data series are clustered in multiple vertical columns grouped by category. Each data series shares the same axes. Hence, they enable comparison of multiple series within a category as well as comparison of same series across several categories.

As the number of series increases, they become hard to interpret quickly.

A 3-D clustered column chart shows the clustered columns in 3-D format.

Stacked Column Chart

A stacked column chart is used when category totals are of more value than the comparison across multiple series. In a stacked column chart, data series are stacked one on top of each other in vertical columns. They show the contribution of each component toward the total of category.

Stacked column charts can display multiple series and categories in a compact manner, and are useful to show change over time via easy comparison of the total column lengths. They also help to show the change over time of a component's contribution to the overall group. Comparison across categories become difficult as the number of series within a category increases. They overall become visually complex to interpret as the number of series and categories increases.

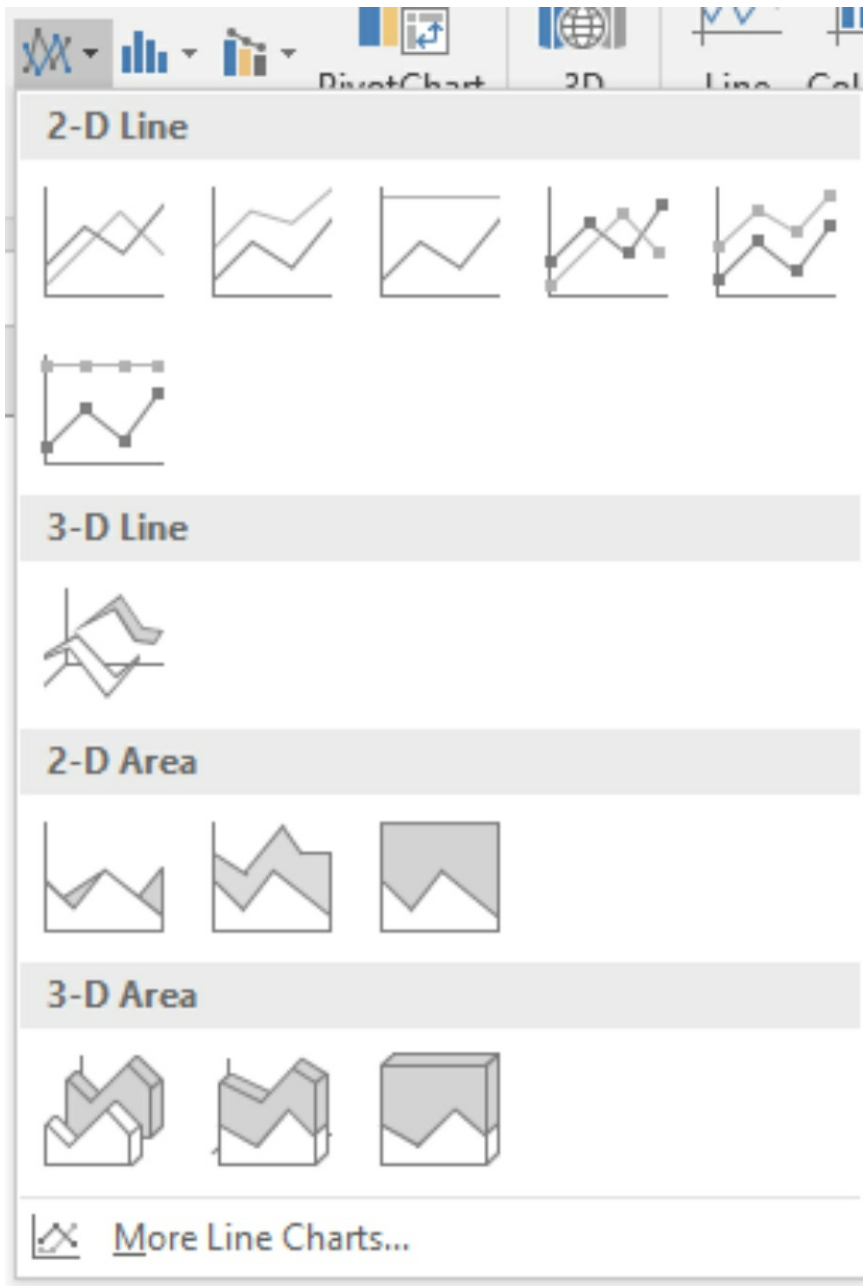
A 3-D stacked column chart shows the stacked columns in 3-D format.

100% Stacked Column Chart

A 100% stacked column chart is used when the percent contribution of each of the series within a category is of value. A 100% stacked column chart displays the relative percentage of multiple data series in stacked columns representing multiple categories, where the total of stacked columns always equals 100%. They show the relative proportion of each component within a category.

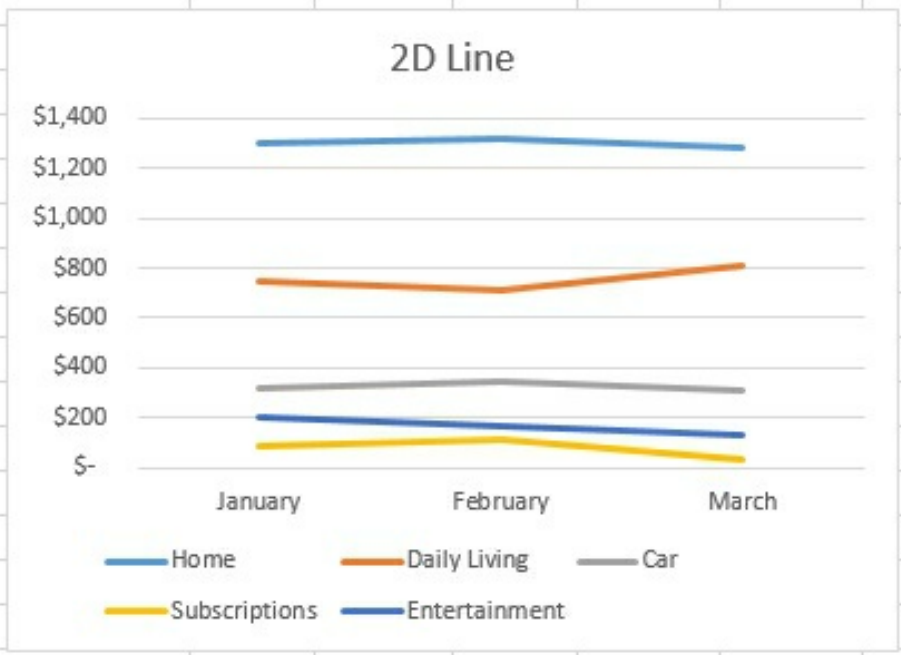
100% stacked column charts can display and compare relative proportions (contributions) of multiple series within categories in a compact manner. They also help to show the change over time of a component's proportion in the overall group. However, it's generally difficult to compare the relative size of the components that make up each column.

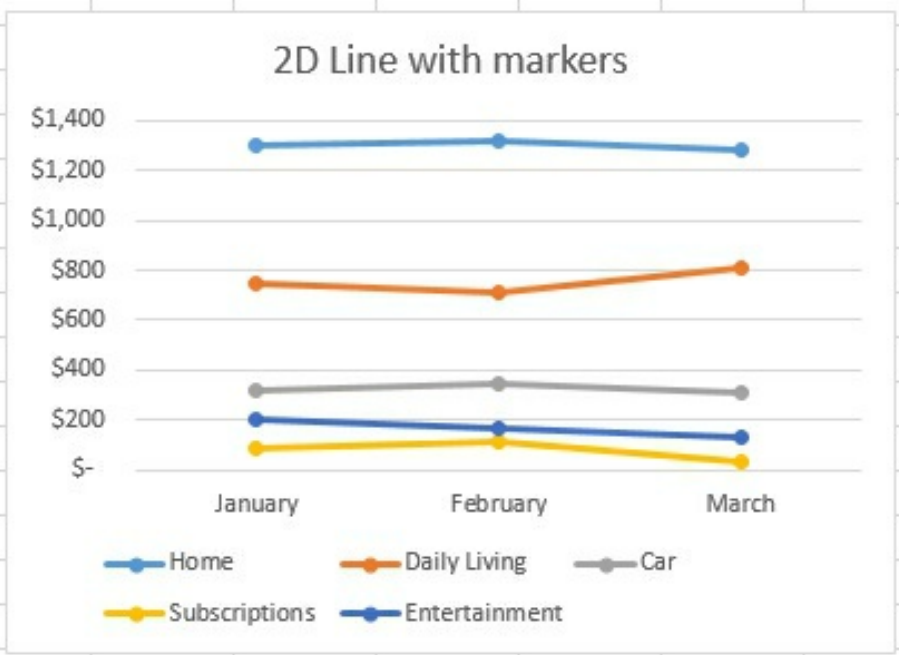
2. Line Charts



A line chart is used to show change or trends over time. A line chart plots multiple series of data sharing the same axes as separate lines. The biggest advantage of line charts over column charts is that they can handle more series without becoming too cluttered.

Expense	January	February	March
Home	$1,300	$ 1,318	$1,286
Daily Living	$ 750	$ 708	$ 806
Car	$ 315	$ 345	$ 305
Subscriptions	$ 90	$ 110	$ 30
Entertainment	$ 200	$ 170	$ 130





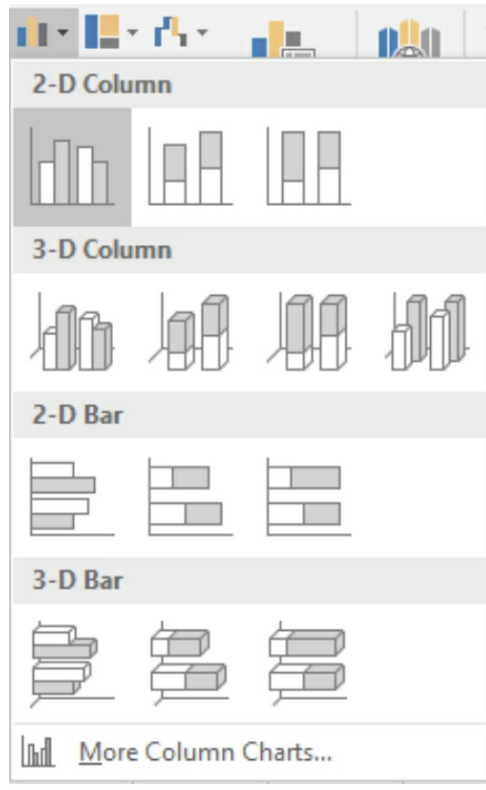
Line Chart with markers

The line chart with markers is a line chart with individual data values identified with a marker. The markers draw

attention to the data points along the line. If there are many categories, use a line chart without markers to avoid cluttering.

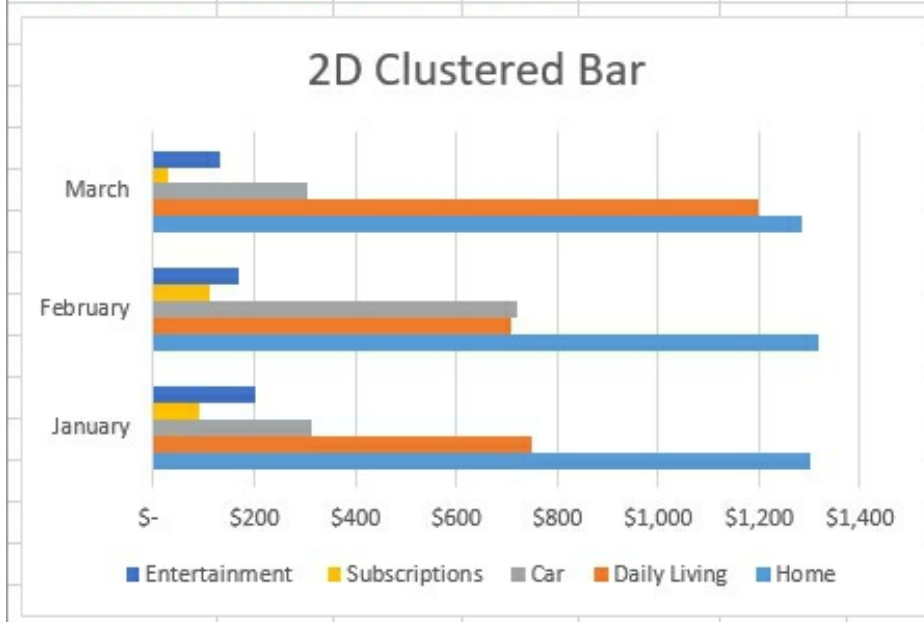
We will skip the stacked line and 100% stacked line graphs as they are not commonly used.

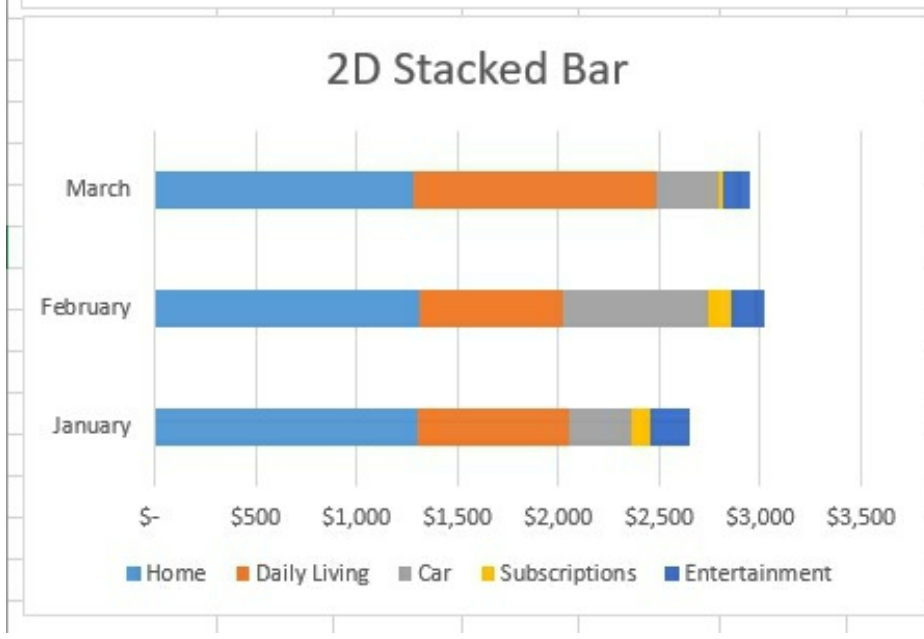
3. Bar Charts

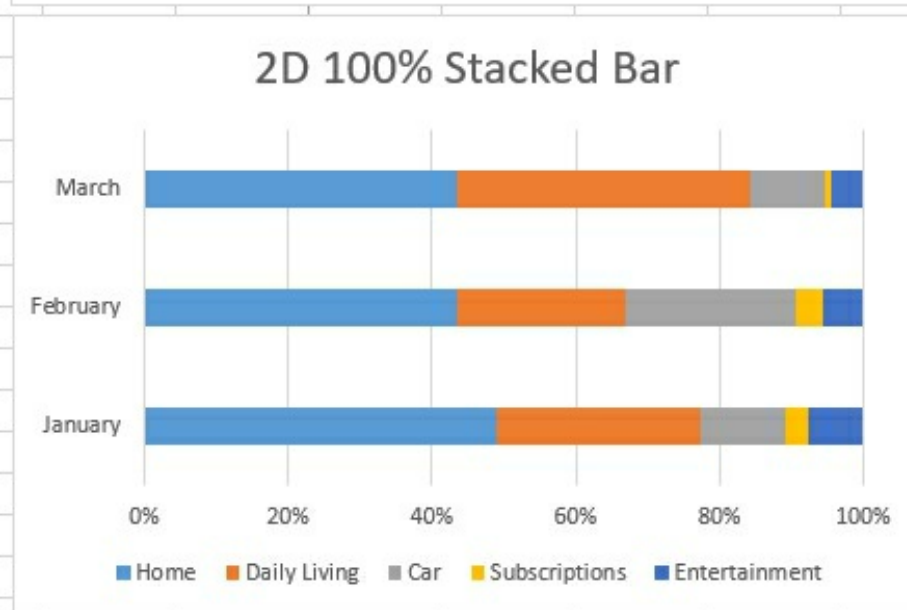


Bar charts plot data using horizontal bars, so they are easy to read and compare. Bar charts are effectively the same as column charts. Both process data the same way; the only difference is that column charts display data as vertical columns, whereas bar charts display data as horizontal bars.

Expense	January	February	March
Home	$1,300	$ 1,318	$1,286
Daily Living	$ 750	$ 708	$1,200
Car	$ 315	$ 719	$ 305
Subscriptions	$ 90	$ 110	$ 30
Entertainment	$ 200	$ 170	$ 130







Clustered Bar Chart

In the clustered bar chart, data series are clustered in multiple horizontal columns grouped by category. Hence, they

enable the comparison of multiple series within a category as well as comparison of the same series across several categories.

A 3-D clustered bar chart shows the clustered bars in 3-D format.

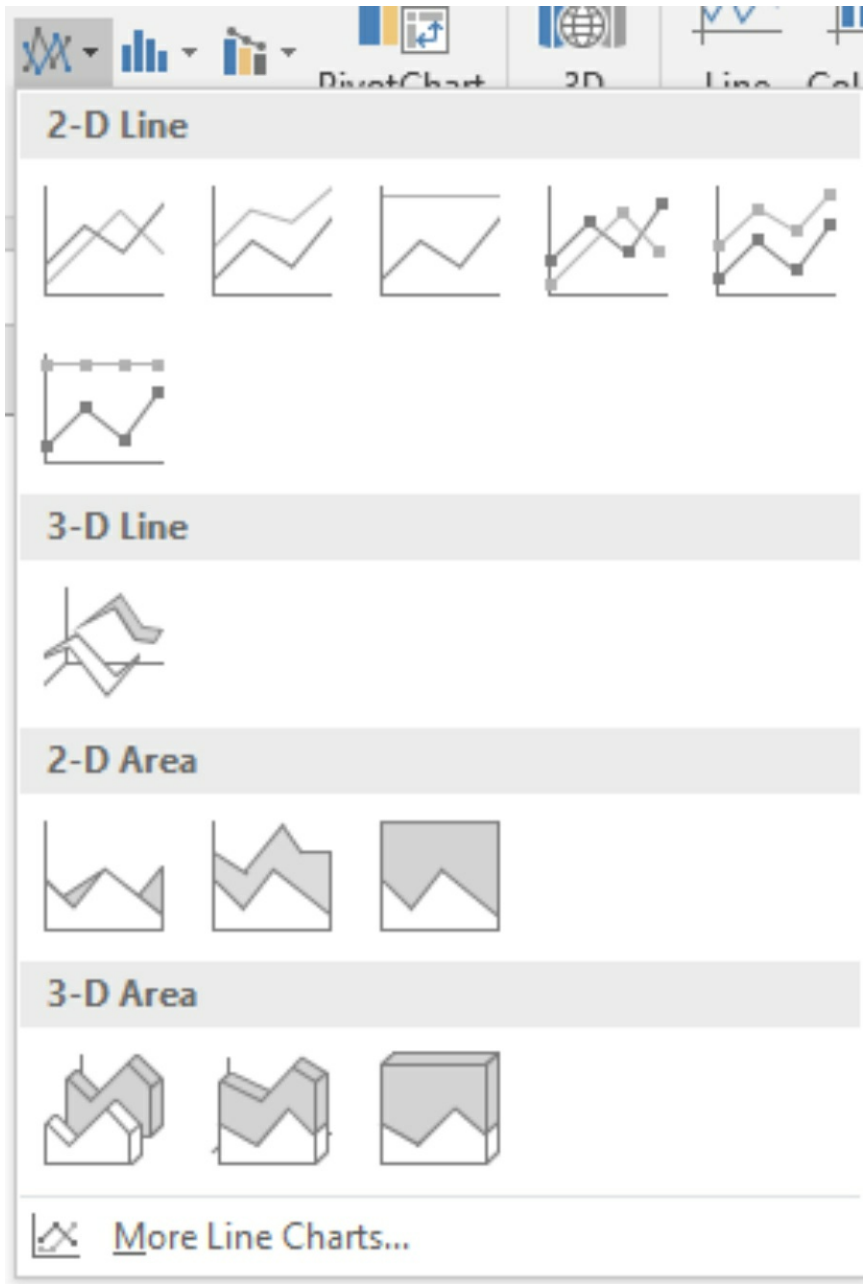
Stacked Bar Chart

In a stacked bar chart, data series are stacked on top of each other in horizontal bars. A 3-D stacked bar chart shows the stacked bars in 3-D format.

100% Stacked Bar Chart

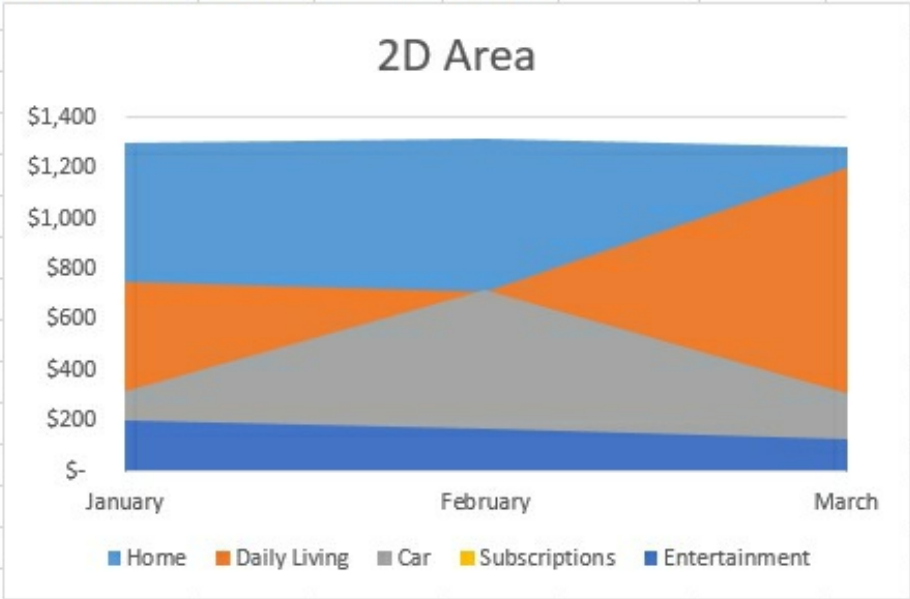
A 100% stacked bar chart is used to display the percent contribution of each of the series within a category.

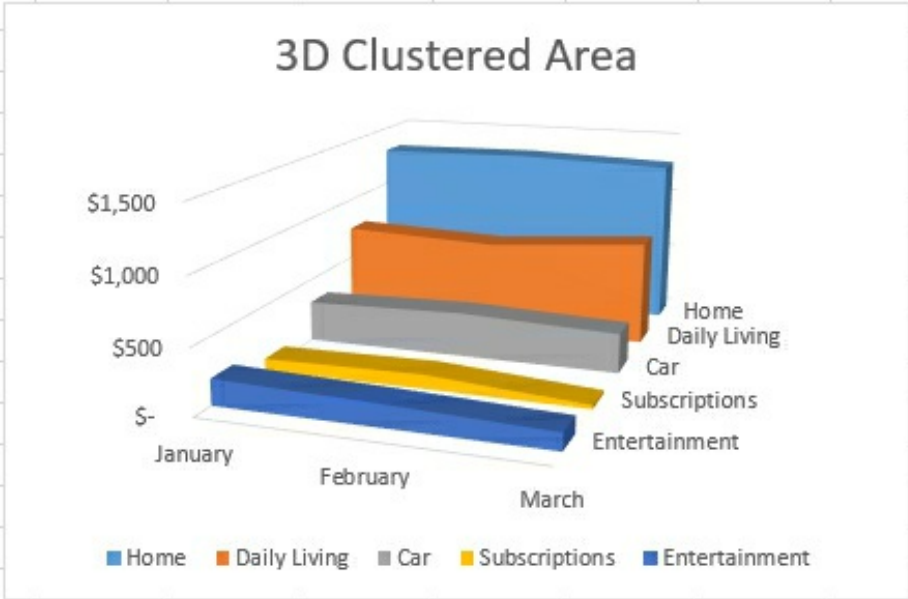
4. Area Charts

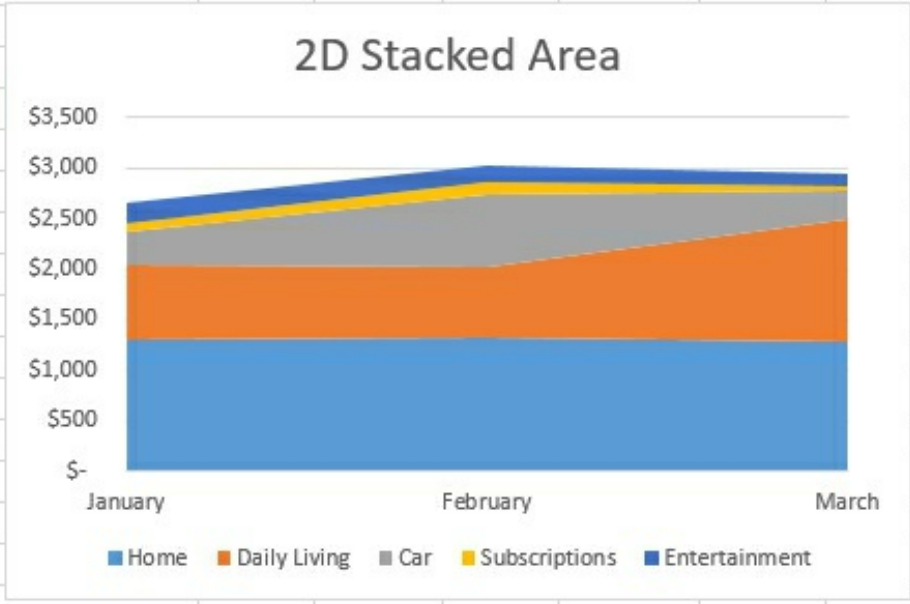


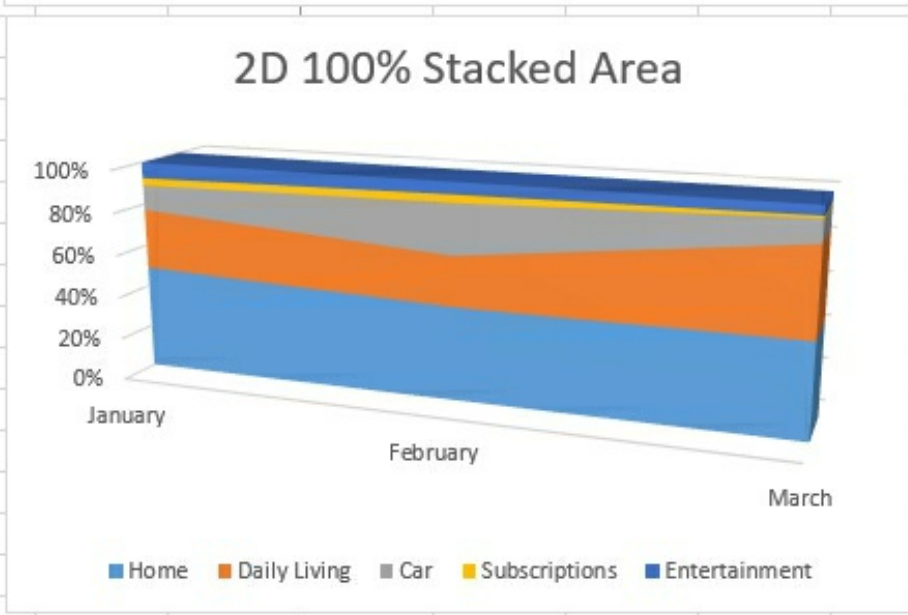
Area chart is essentially a line chart with the area below a series filled with a color. An area chart plots change over time. If there are two or more series that intersect each other, then data from one series can be hidden behind data from another series. As a rule, consider using a line chart instead of a non-stacked area chart if the series intersect. A 3-D area chart shows the clustered bars in 3-D area.

Expense	January	February	March
Home	$1,300	$ 1,318	$1,286
Daily Living	$ 750	$ 708	$1,200
Car	$ 315	$ 719	$ 305
Subscriptions	$ 90	$ 110	$ 30
Entertainment	$ 200	$ 170	$ 130









Stacked Area Chart

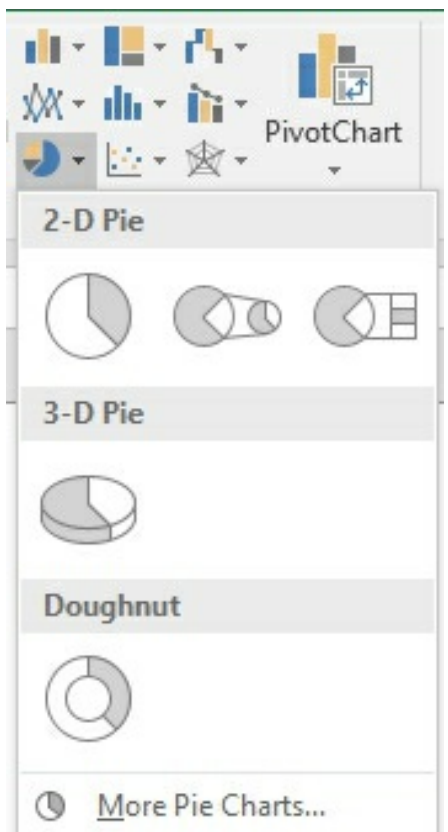
A stacked area chart shows data series plotted with filled areas stacked on top of each other. The chart draws attention to the change in various data series over time across categories.

100% Stacked Area Chart

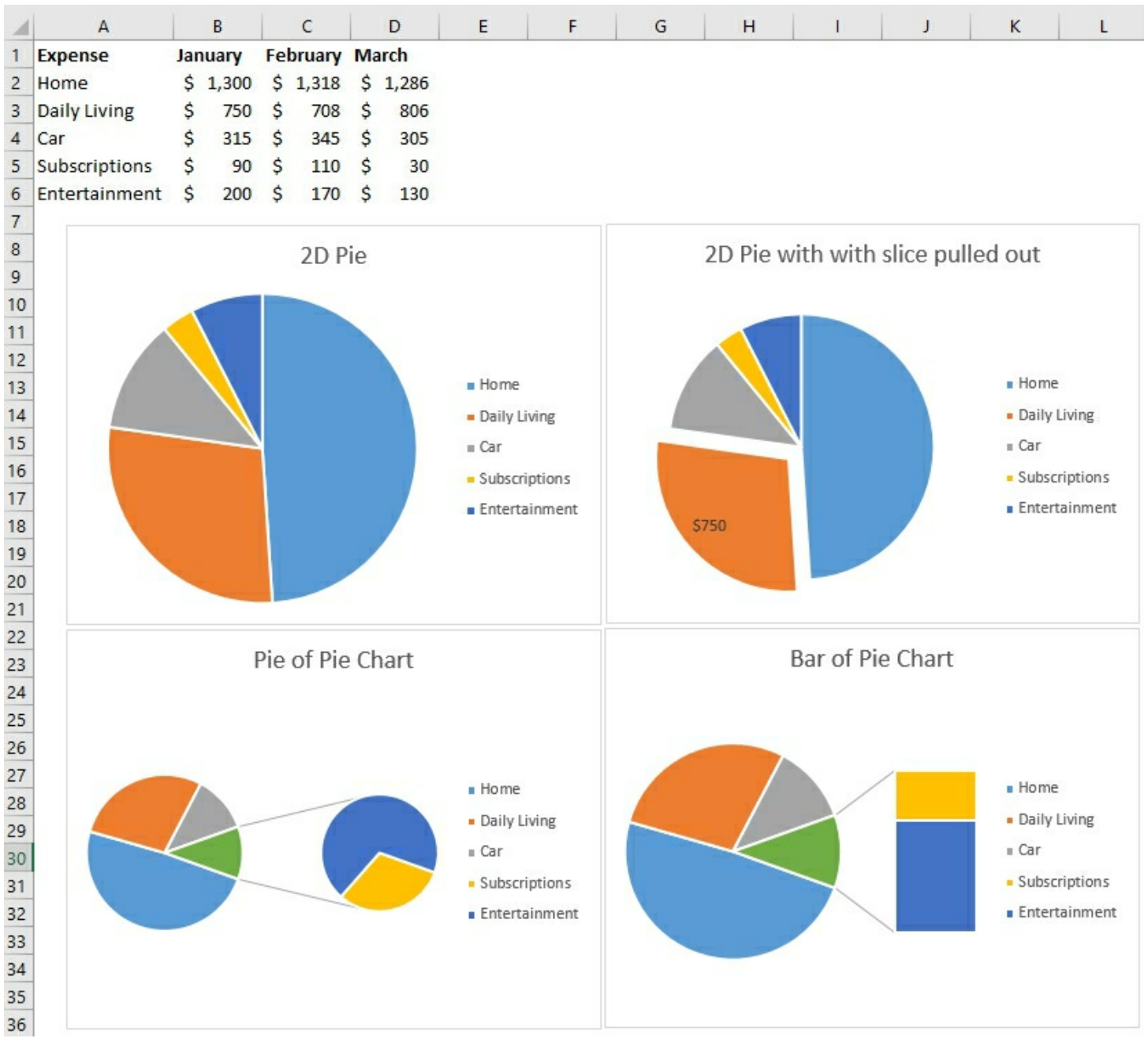
A 100% stacked area chart shows data plotted as areas and stacked so that the cumulative area amounts to 100%. They show the percentage each component contributes towards a group.

The charts show the progression and composition of data series within categories. However, they are a bit non-intuitive and not commonly used.

5. Pie Charts



A pie chart looks like a pie! Each piece of the pie represents the proportion of individual values in a data series relative to the sum of the values. Hence the data points are shown as automatically calculated percentages of the series. They shouldn't be used when categories do not total to 100%.



Expense	January	February	March
Home	$ 1,300	$ 1,318	$ 1,286
Daily Living	$ 750	$ 708	$ 806
Car	$ 315	$ 345	$ 305
Subscriptions	$ 90	$ 110	$ 30
Entertainment	$ 200	$ 170	$ 130

Pie of pie and bar of pie charts

Pie charts work best to display data with a small number of categories. The pie of pie and bar of pie chart provide a way to add additional categories to a pie chart without generating a pie chart too complex to read. Pie of pie chart shows pie charts with smaller values pulled out into a secondary pie, whereas bar of pie chart shows smaller values pulled out into a secondary stacked bar chart. Both make multiple small values easier to distinguish.

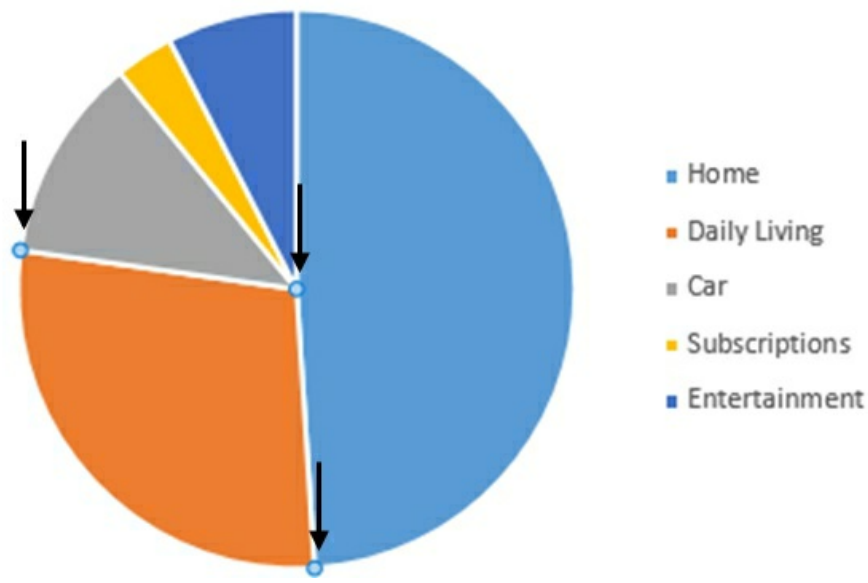
What if you're plotting a pie chart to include in a presentation and want a slice of the pie chart to stand out so that you can focus the discussion on it? You can easily pull the slice manually to emphasize it. Let's practice!

Practice

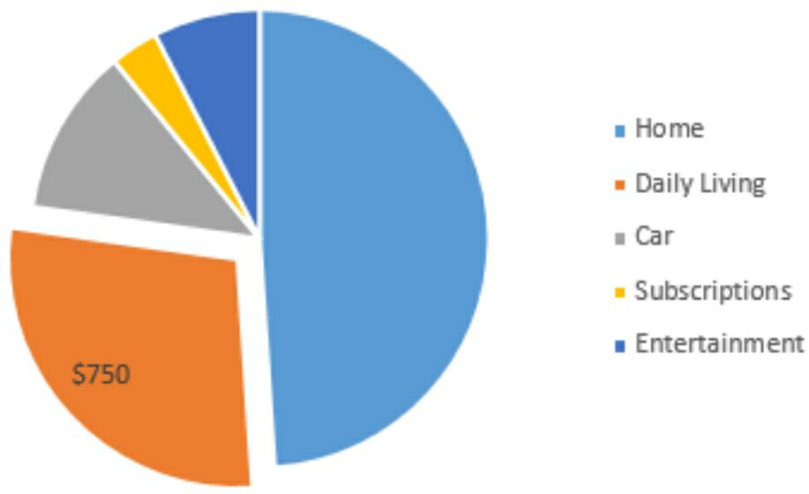
1. Continue working on *Charts_Practice.xlsx* and go to the *Pie* sheet. The sheet contains data on

personal expenses for January broken down by categories and a pie chart based on the data.

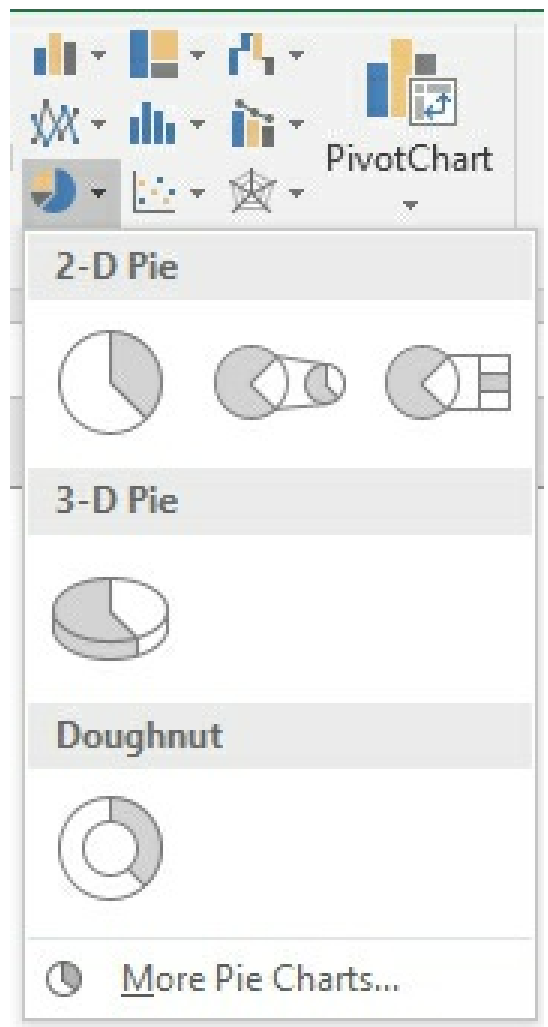
2. Suppose you want to emphasize on the proportion of the budget spent on subscriptions. Click on the pie chart to select it. Then click on the pie representing subscriptions. Three small circles will appear around the pie.



3. Now pull it out of the chart slightly. Right click on the pie and select *Add Data Labels*. From the *Add Data Labels* options that become visible, again select *Add Data Labels*. You have now emphasized the pie of interest.

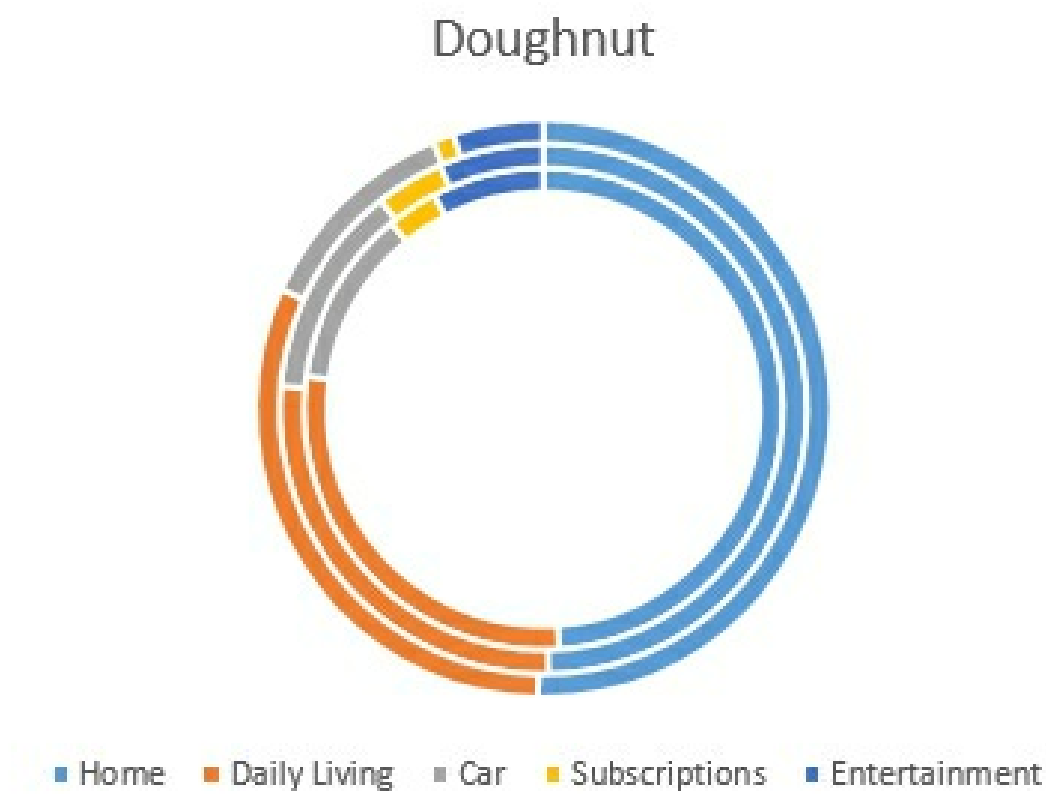


6. Donut Charts

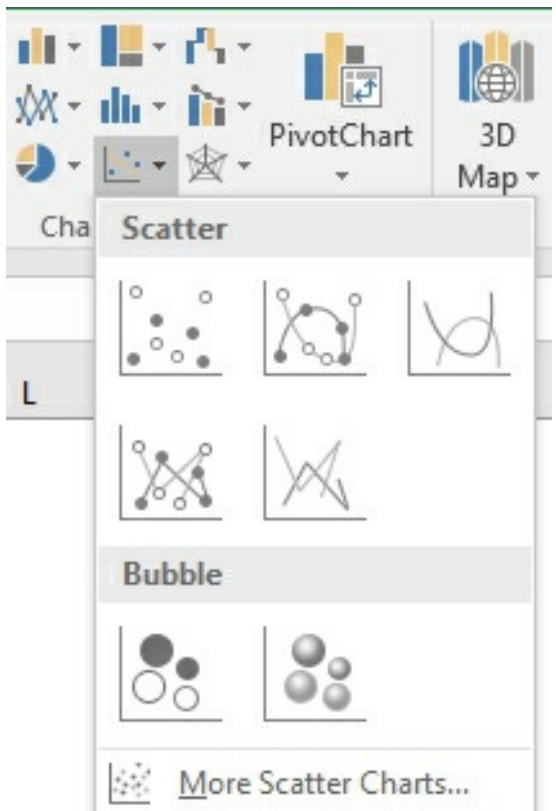


The pie charts have a limitation of displaying only one series of data at a time. A donut chart does not have that limitation. Multiple series can be displayed in the donut chart as concentric rings. However, they become more

complex to read as the total number of series increases.



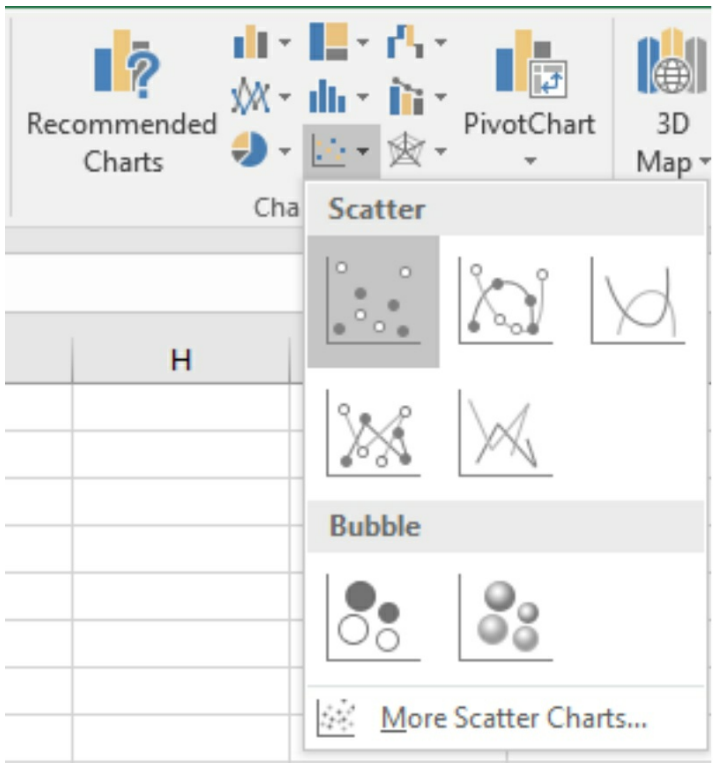
7. Scatter Charts



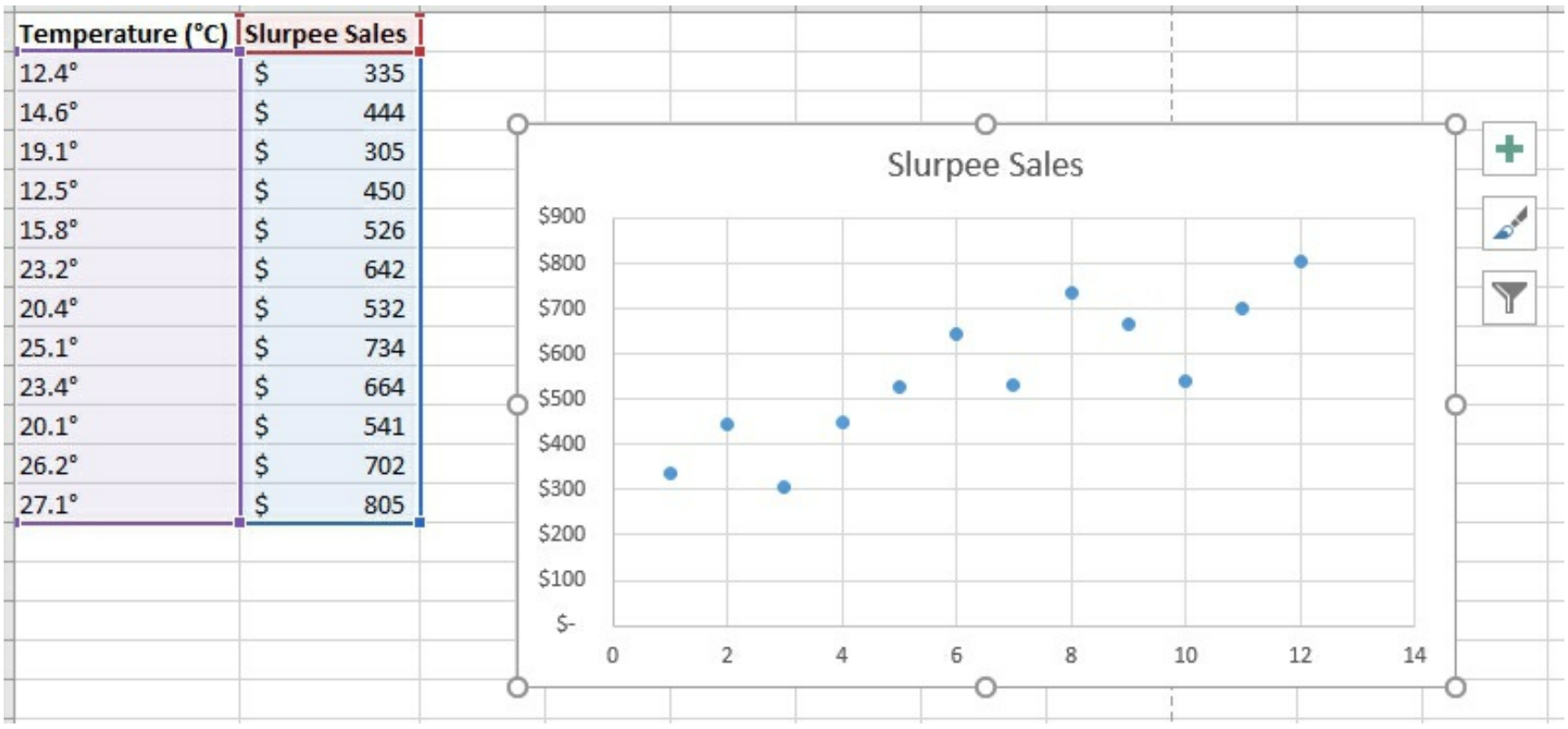
A scatter plot is used to show the relationship between two variables both containing a set of numeric values. A scatter plot plots one variable on the vertical axis and the other on the horizontal axis with each data point then plotted as a discrete point on the chart. Scatter charts are typically used for showing and comparing numeric values. Let's practice!

Practice:

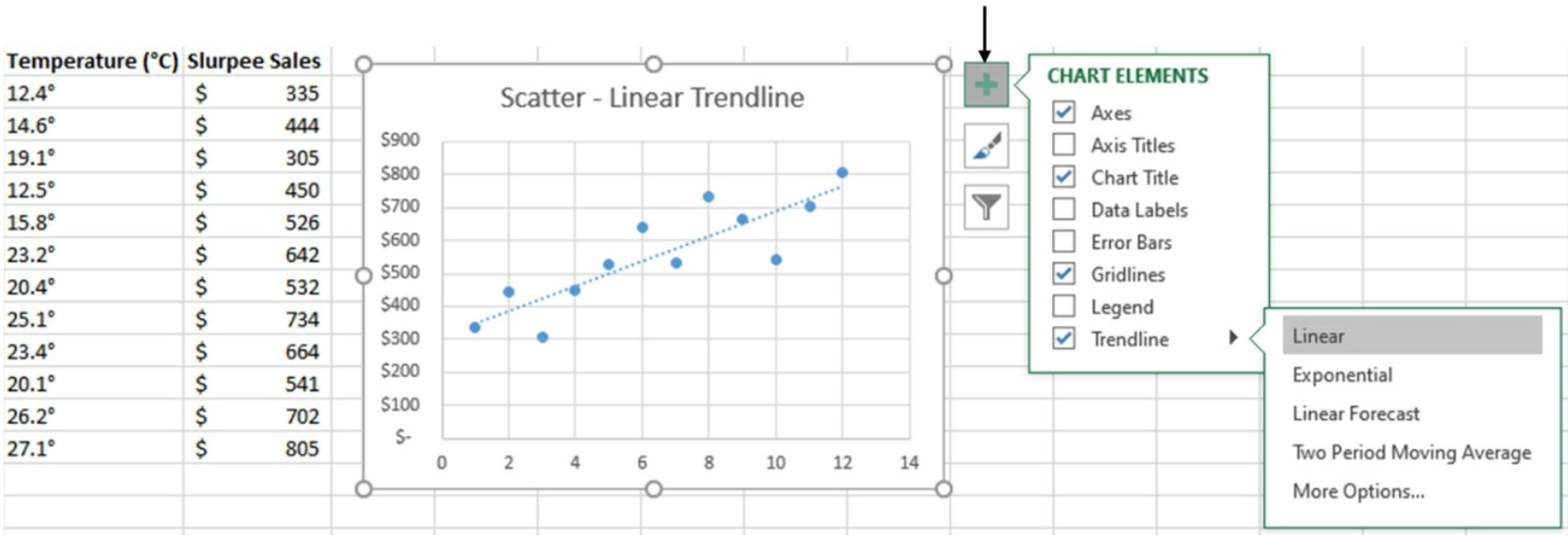
1. Continue working on *Charts_Practice.xlsx* and go to the *Scatter* sheet. The sheet contains data on daily temperature and respective Slurpee sales during summer days.
2. Select the data including the column titles and row labels (A1:B13).
3. Go to the *Insert* tab and under the *Charts* group, select *Insert Scatter (X,Y) or Bubble Chart* icon. From the displayed options, under the *Scatter* group, select *Scatter Chart*.



4. A Scatter Chart will be inserted. Rename the title to Slurpee Sales.

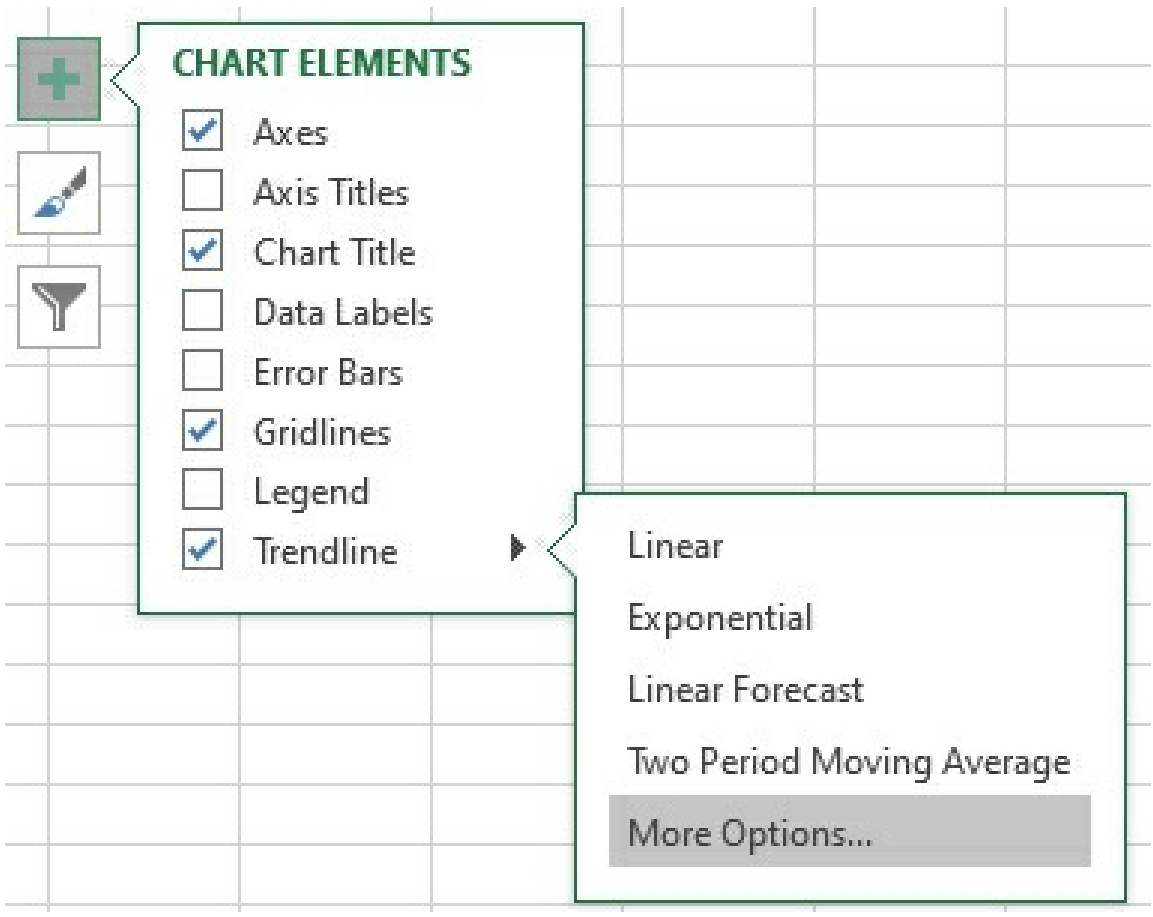


5. There seems a linear trend between the temperature and the Slurpee sales. To make the trend more visual, a **linear trendline** can be added to the chart. To add the trendline, select the chart and click on the + icon to the top right of the chart. Select Trendline and choose *Linear* from one of the suggested types. Rename the title of the chart to "Scatter – Linear Trendline".

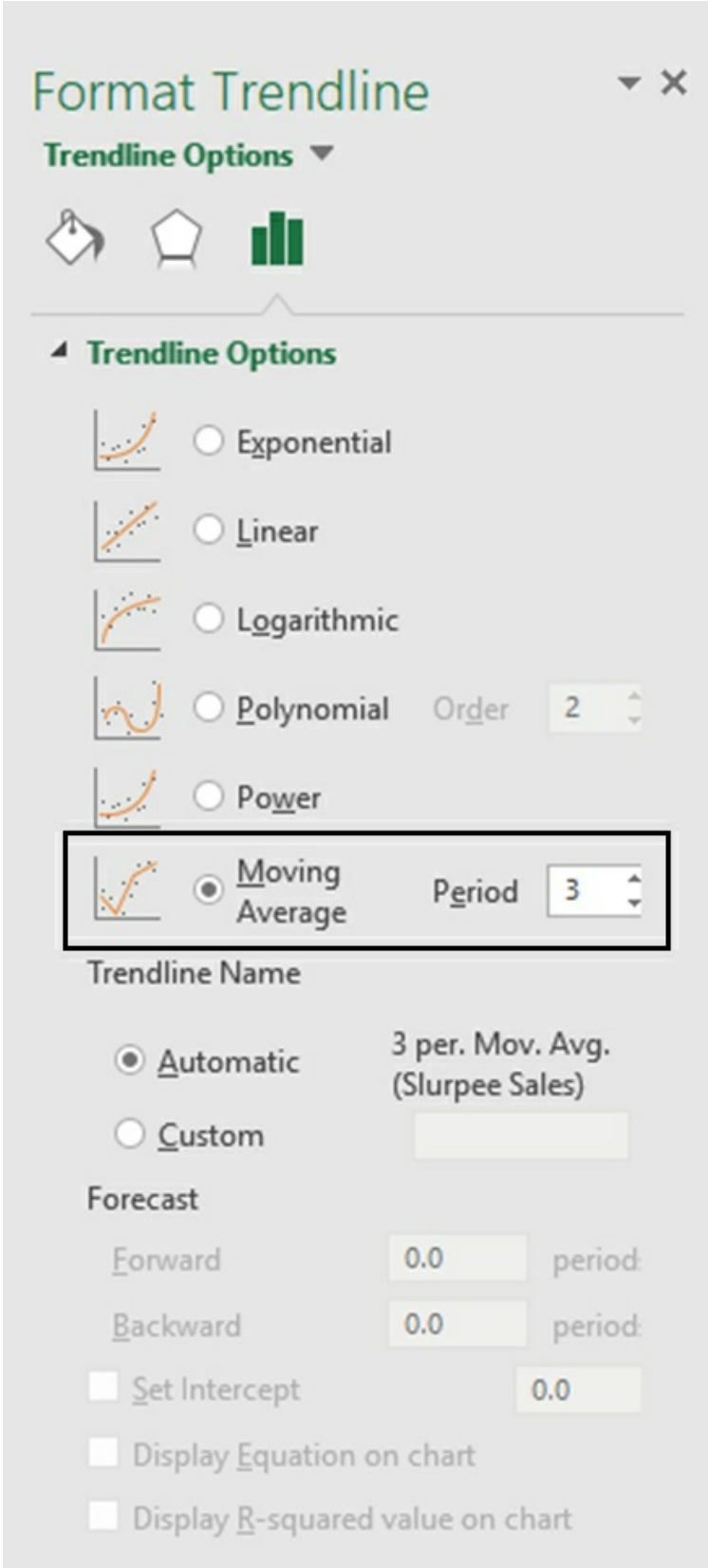


6. Now copy and paste the chart right next to the copied chart. Rename the title of the copied chart to "Scatter – Moving Average". Now we will learn how to insert a moving average trendline in the chart. A **moving average trendline** smoothes out fluctuations in data to show a pattern or trend more clearly. A moving average averages a specific number of data points, set by the Period option, and uses the average value as a point in the line. If Period is set to 3, for example, then the average of the first three data points is used as the first point in the moving average trendline. The average of the

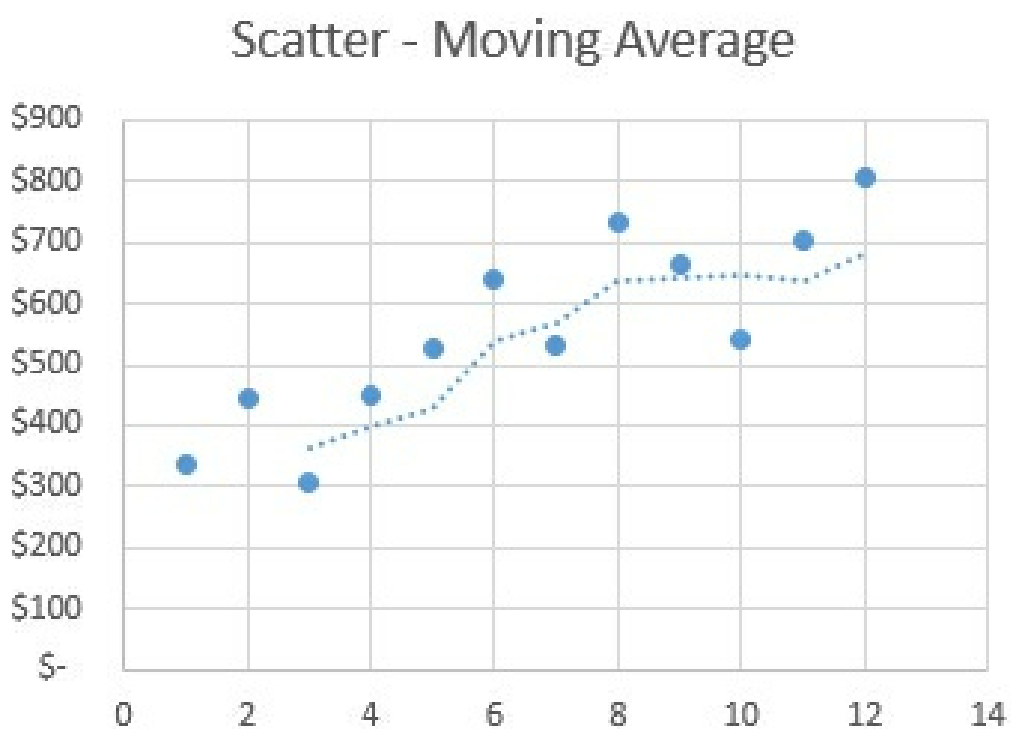
second, third and fourth data points is used as the second point in the trendline, and so on. Hence, the name moving average. To change the trendline to a moving average trendline**,** click on the + icon to the top right of the chart. Choose *More Options* from one of the suggested types.



7. A *Format Trendline* window will be displayed on the extreme right of the screen. Under *Trendline Options*, select *Moving Average* and set the *Period* to *3*.



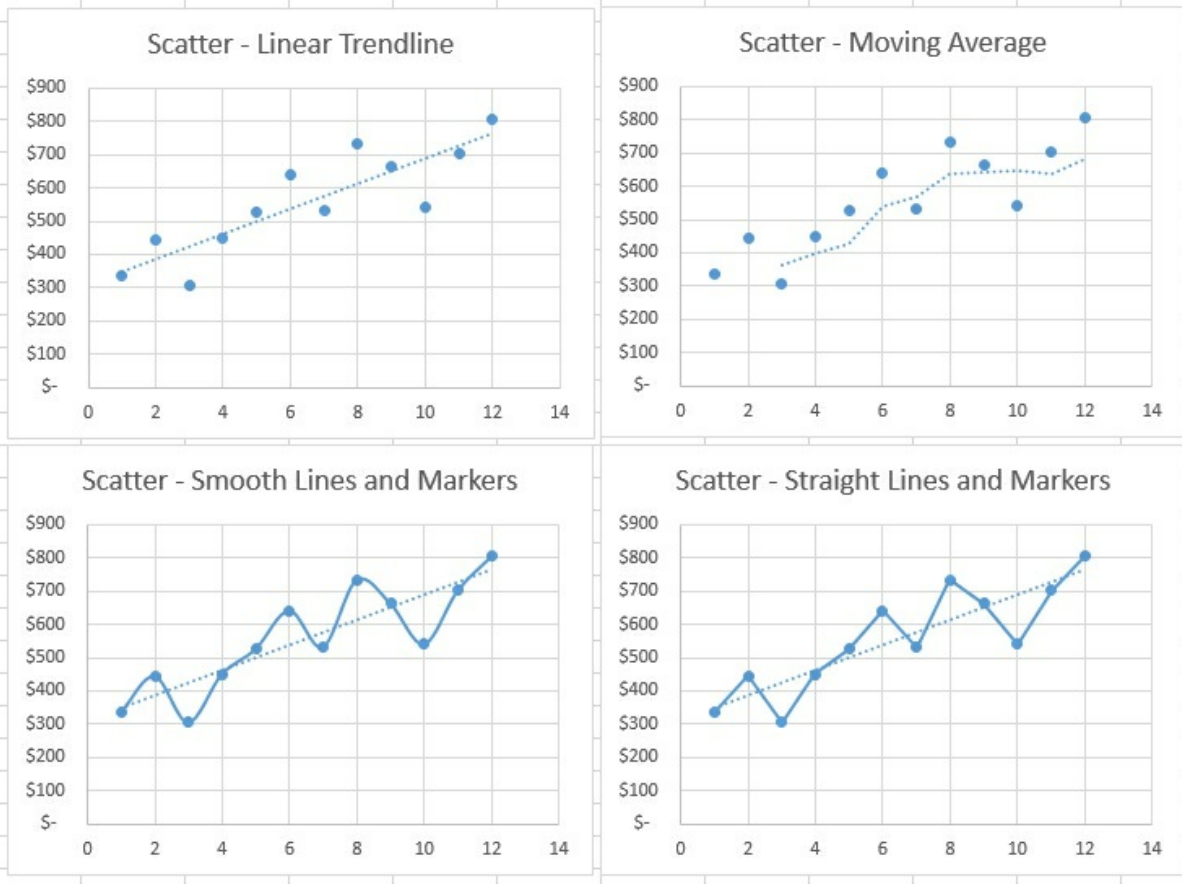
8. The linear trendline will change to a moving average trendline. Close the window.



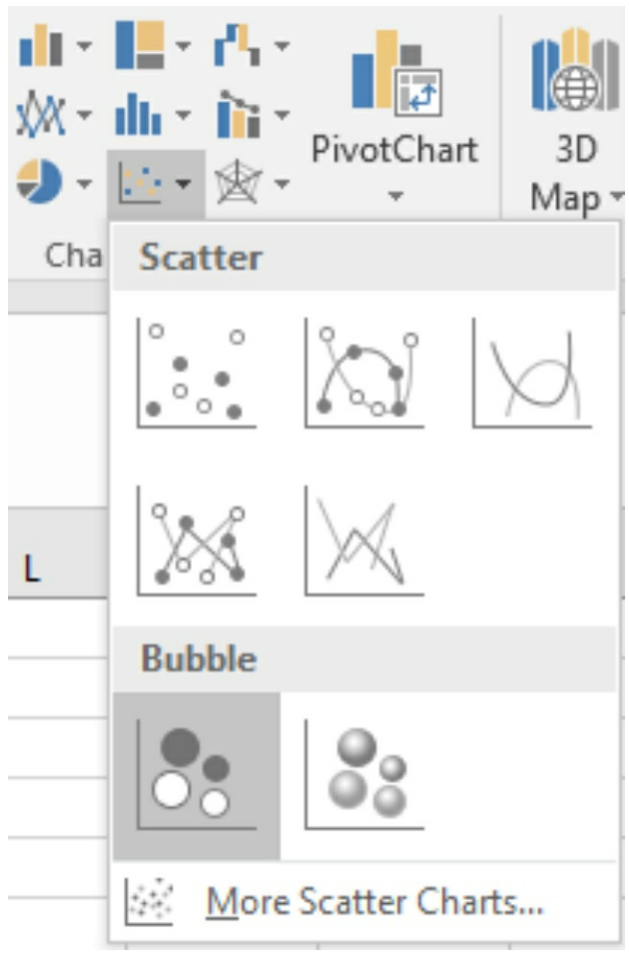
Scatter charts with smooth/straight lines and markers

Scatter charts with smooth lines and markers display a smooth curve that connects the data points, whereas data points on scatter charts with straight lines and markers are connected by straight lines.

Temperature (°C)	Slurpee Sales
12.4°	$ 335
14.6°	$ 444
19.1°	$ 305
12.5°	$ 450
15.8°	$ 526
23.2°	$ 642
20.4°	$ 532
25.1°	$ 734
23.4°	$ 664
20.1°	$ 541
26.2°	$ 702
27.1°	$ 805

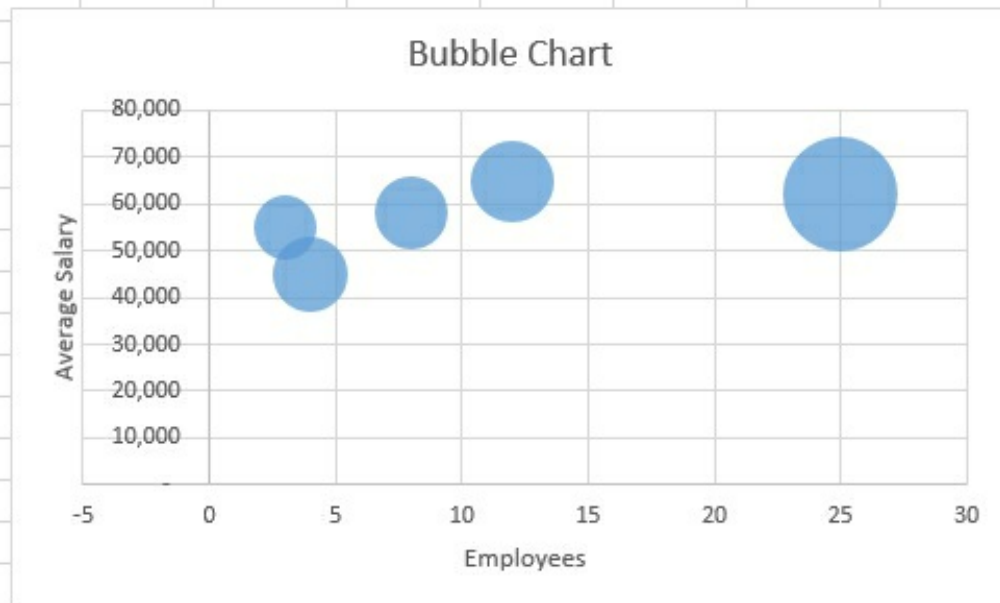


8. Bubble Charts

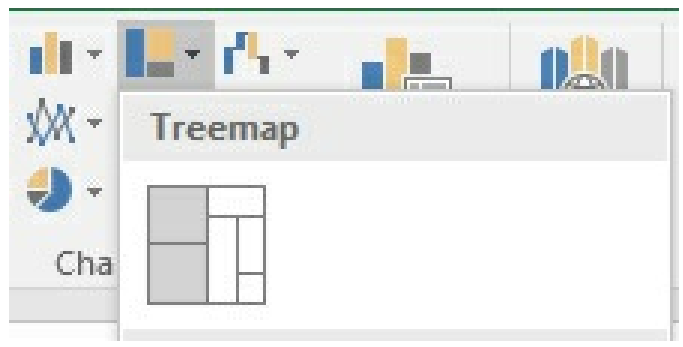


A bubble chart is a variation of a scatter chart in which the data points are replaced with bubbles, and an additional dimension of the data is represented in the size of the bubbles. The bubble chart can be used instead of a scatter chart if data contains three data series each containing a set of numeric values. The sizes of the bubbles are determined by the numeric values in the third data series.

Department	Employees	Average Salary	Average Bonus
Sales	12	65,000	2,500
Marketing	8	58,000	2,000
HR	3	55,000	1,500
Administration	4	45,000	2,200
Operations	25	62,000	5,000

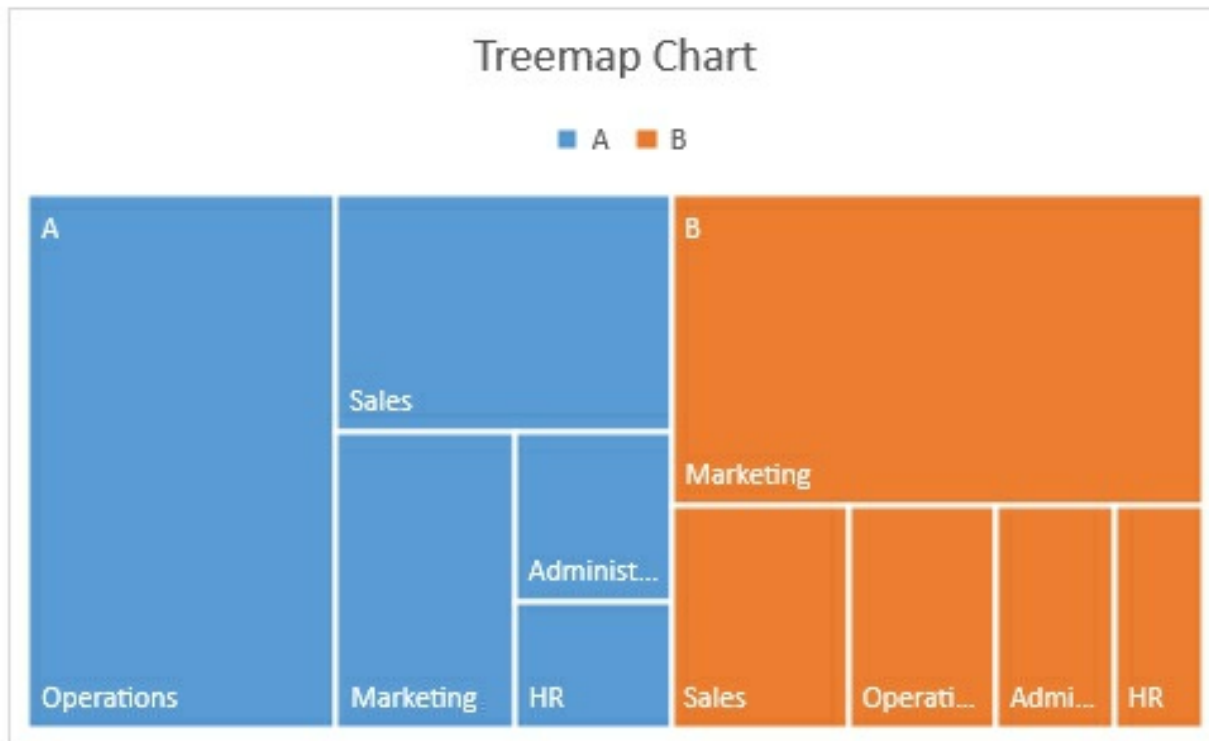


9. Treemap

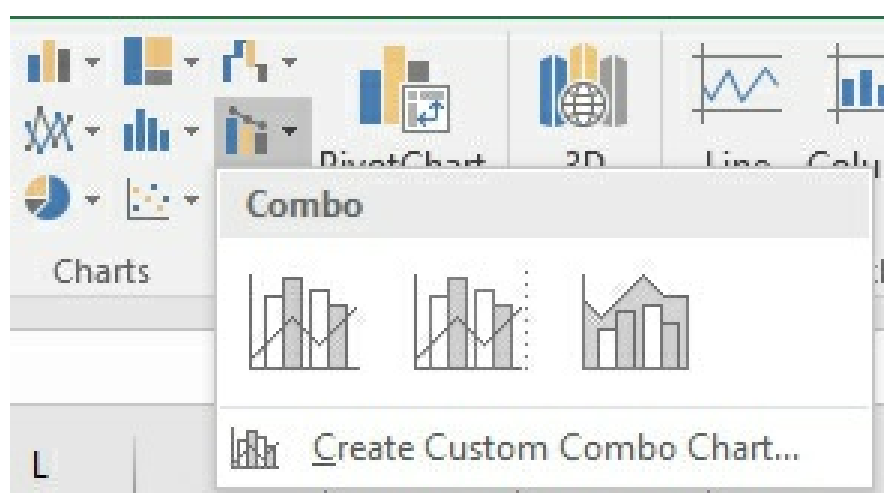


A treemap chart displays hierarchical data. The data is displayed in form of rectangles colored based on the first level of hierarchy and sized proportionally to values of the second level of hierarchy. The color of the rectangles is determined by the first series in the data and size is determined by the second series. Treemap charts are used to compare proportions within and across the hierarchy.

Company	Department	Employees
A	Sales	12
A	Marketing	8
A	HR	3
A	Administration	4
A	Operations	25
B	Sales	6
B	Marketing	25
B	HR	3
B	Administration	4
B	Operations	5

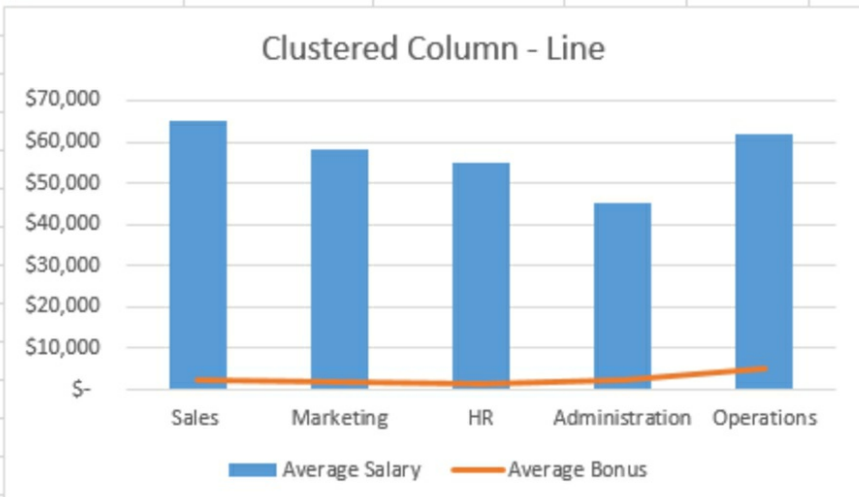


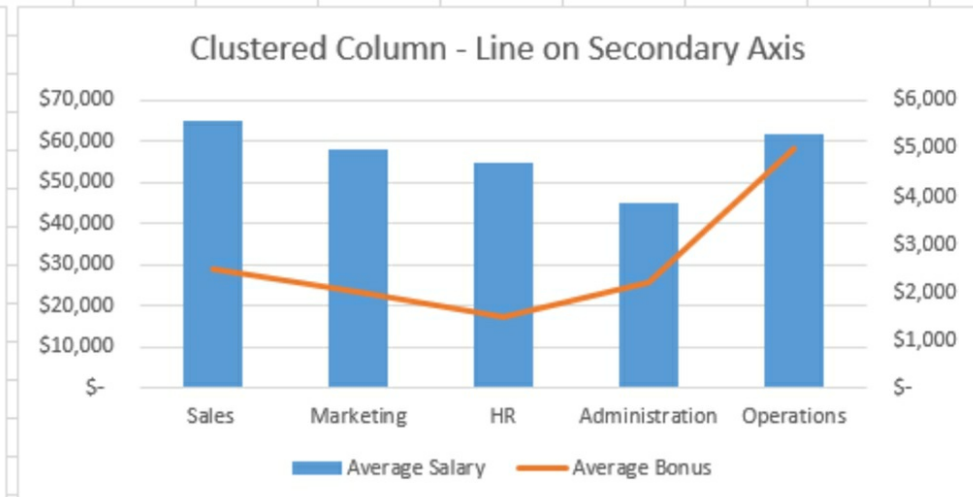
10. Combo Charts



A combo chart is a hybrid of two chart types, most commonly a column chart and a line chart. *Clustered Column – Line* chart combines a clustered column and a line chart showing one series as columns and the other as lines in the same chart. *A Clustered Column – Line on Secondary Axis* chart can be used if the scales of the series are widely different. In this chart, a secondary y-axis is used to scale values of the second data series generally representing the line on the chart.

Department	Average Salary	Average Bonus
Sales	$ 65,000	$ 2,500
Marketing	$ 58,000	$ 2,000
HR	$ 55,000	$ 1,500
Administration	$ 45,000	$ 2,200
Operations	$ 62,000	$ 5,000

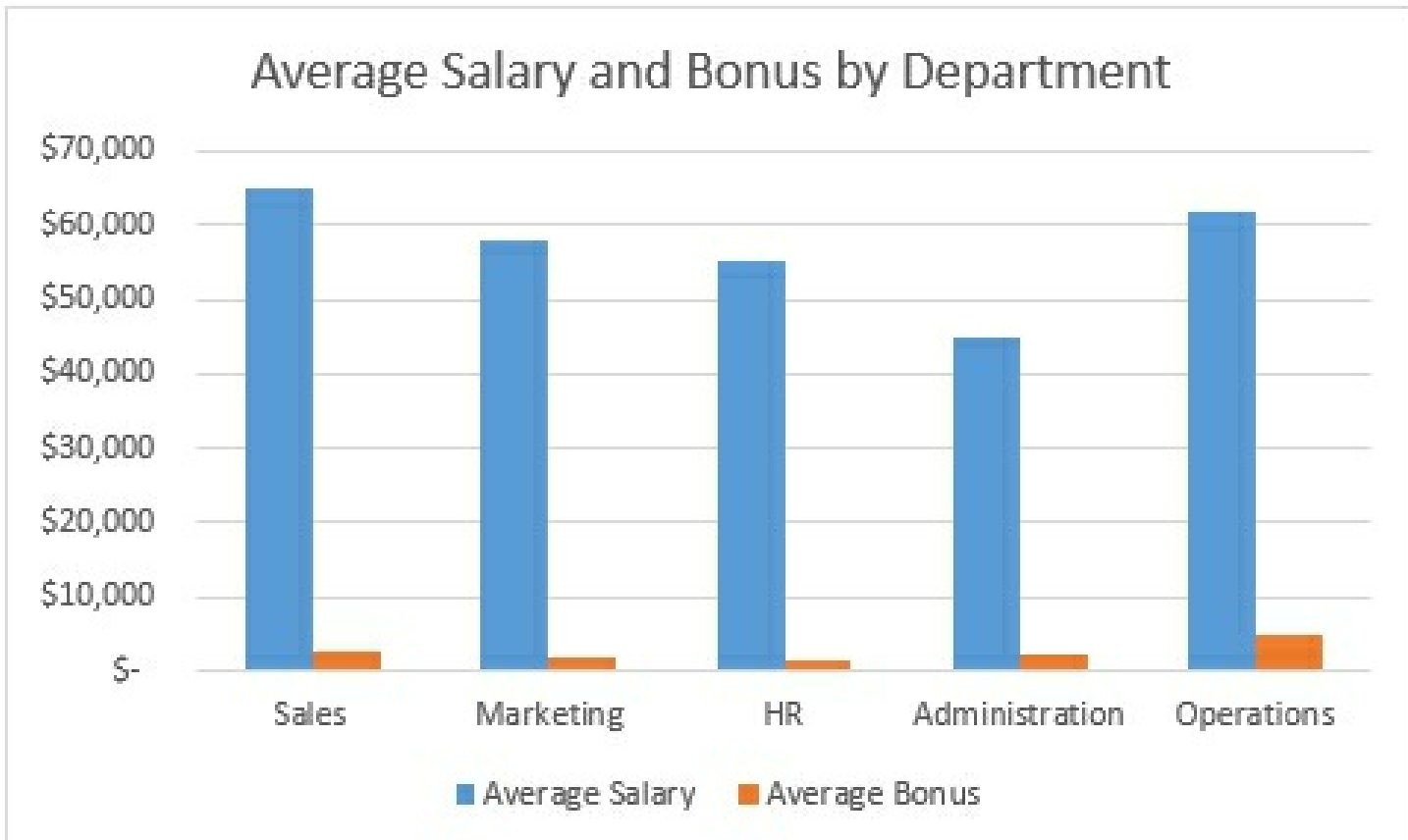




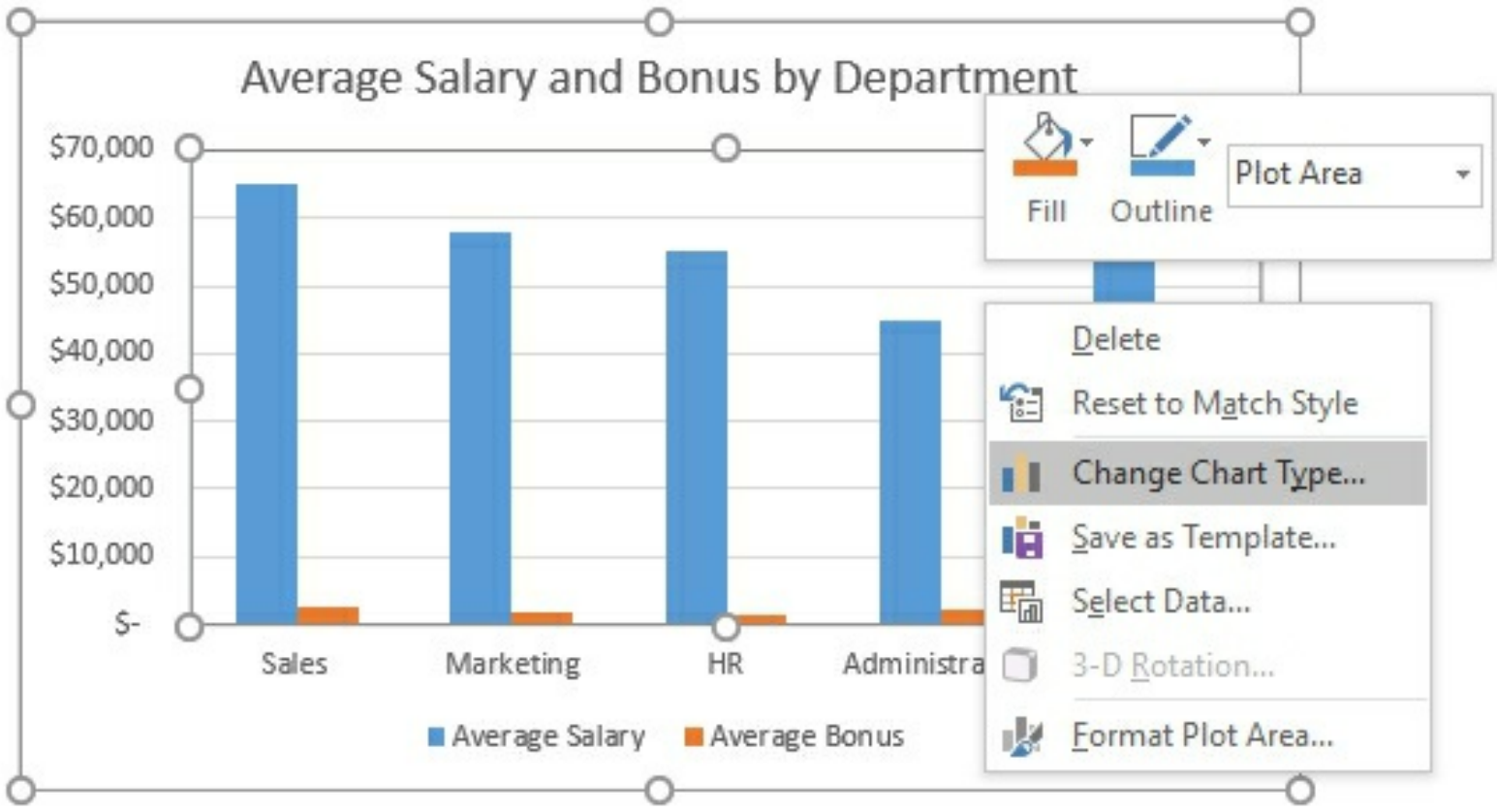
You can also change an existing chart to a combo chart (or any other chart type). Let's see how it's done.

Practice

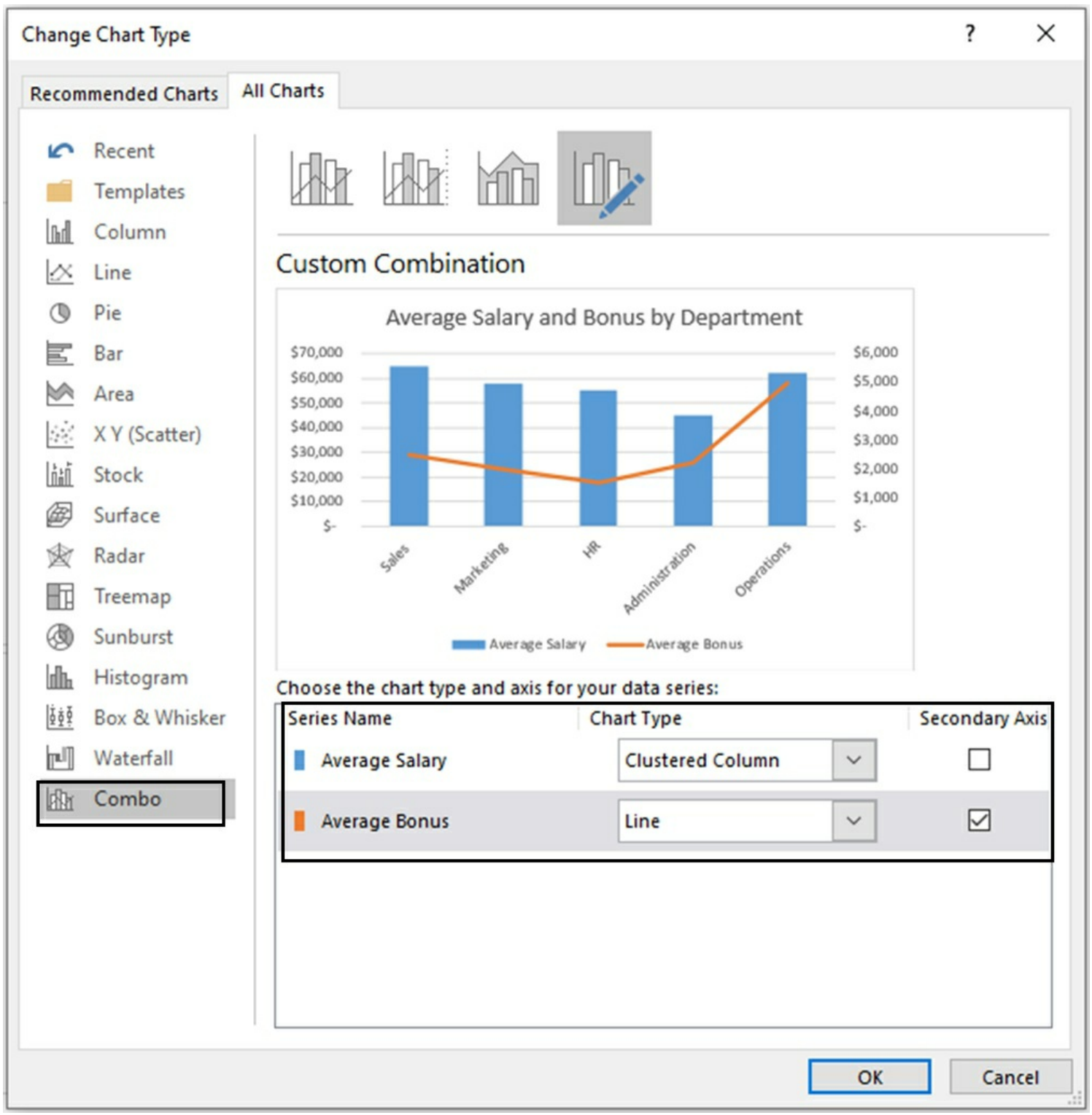
1. Continue working on *Charts_Practice.xlsx* and go to the *Combo* sheet. The sheet contains data on the average salary and bonus of different departments of a company and a 2-D column chart based on the data.



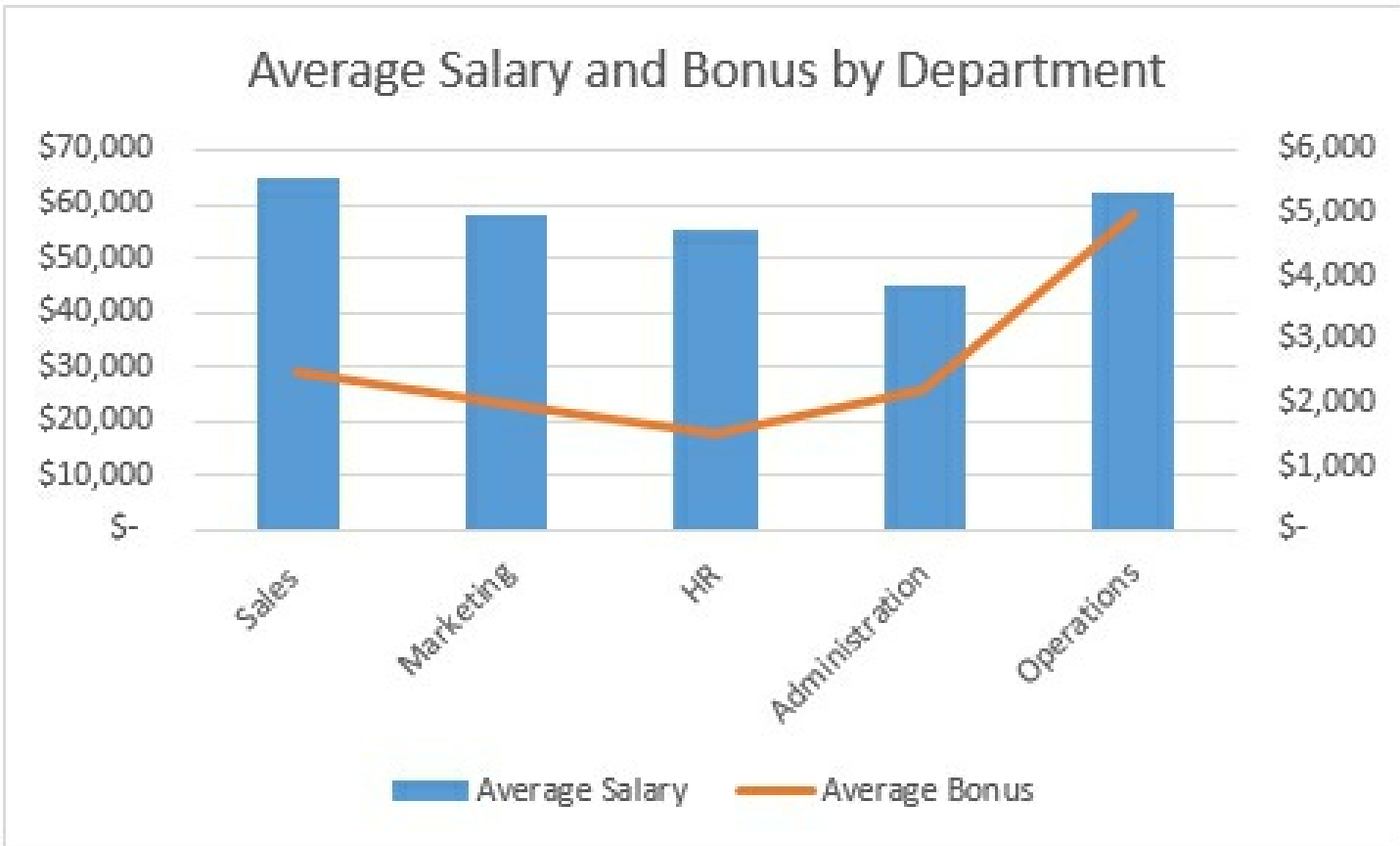
2. You can notice that because there is a wide difference between the scale of salary and bonus, as such, bonuses are hard to compare. This can be sorted by using a combo chart. To change the chart to a combo chart, select the chart, right click and select *Change Chart Type*.



3. The *Change Chart Type* dialog box will appear. Under *All Charts* tab, select *Combo* from the left pane. Ensure *Clustered Column* is listed as *Chart Type* for *Average Salary* and *Line* for *Average Bonus*. To scale the Average Bonus line on secondary axis, check the corresponding box under *Secondary Axis* and click OK.

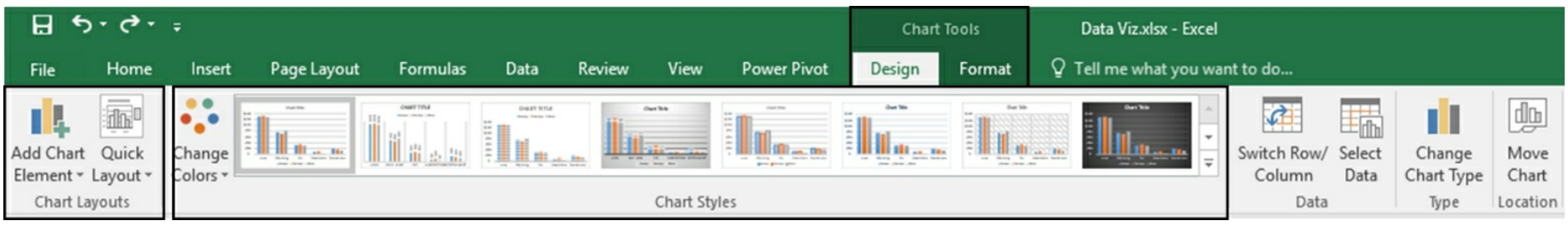


4. The chart will update to a combo chart.

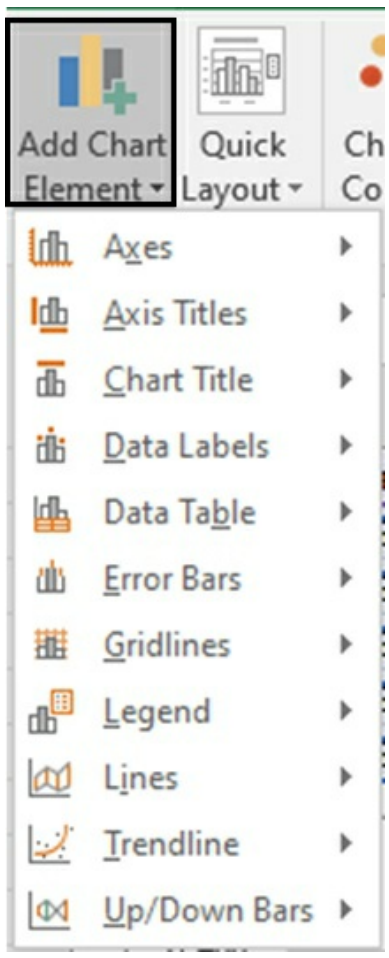


Formatting a chart

After inserting a chart, there are several options available to format it. When the chart is selected, you will see additional tabs - *Design* and *Format*, grouped as *Chart Tools*. A chart's layout and style can be edited easily using these tabs.



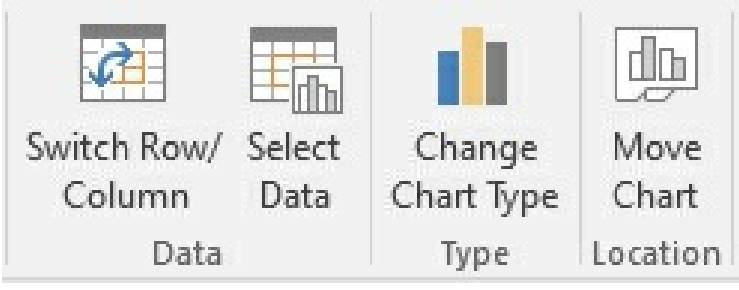
- You can add **chart elements**—such as chart titles, legends, and data labels—to add more context to the chart. To add a chart element, click the *Add Chart Element* command on the *Design* tab, then choose the desired element from the drop-down menu. To rename a chart element, like a chart title, simply double-click the placeholder for the element and type the new name.



- You can use one of Excel's predefined layouts to add chart elements as well. Click the *Quick Layout* command on the *Design* tab and choose the desired layout from the drop-down menu.

- Excel also offers different chart styles which allow to quickly modify the look and feel of a chart. To change the chart style, select the desired style from the *Chart styles* group on the *Design* tab.

- You can also access *Switch Row/Colum*n, *Select Data*, *Change Chart Type* options directly from the *Design* tab in addition to the alternate routes you followed previously in the chapter. In addition, there is a *Move Chart* option to move the chart to another worksheet in the workbook.



- You can also use the chart formatting shortcut buttons available by clicking on the chart to quickly add chart elements, change the chart style and color, and filter the chart data.

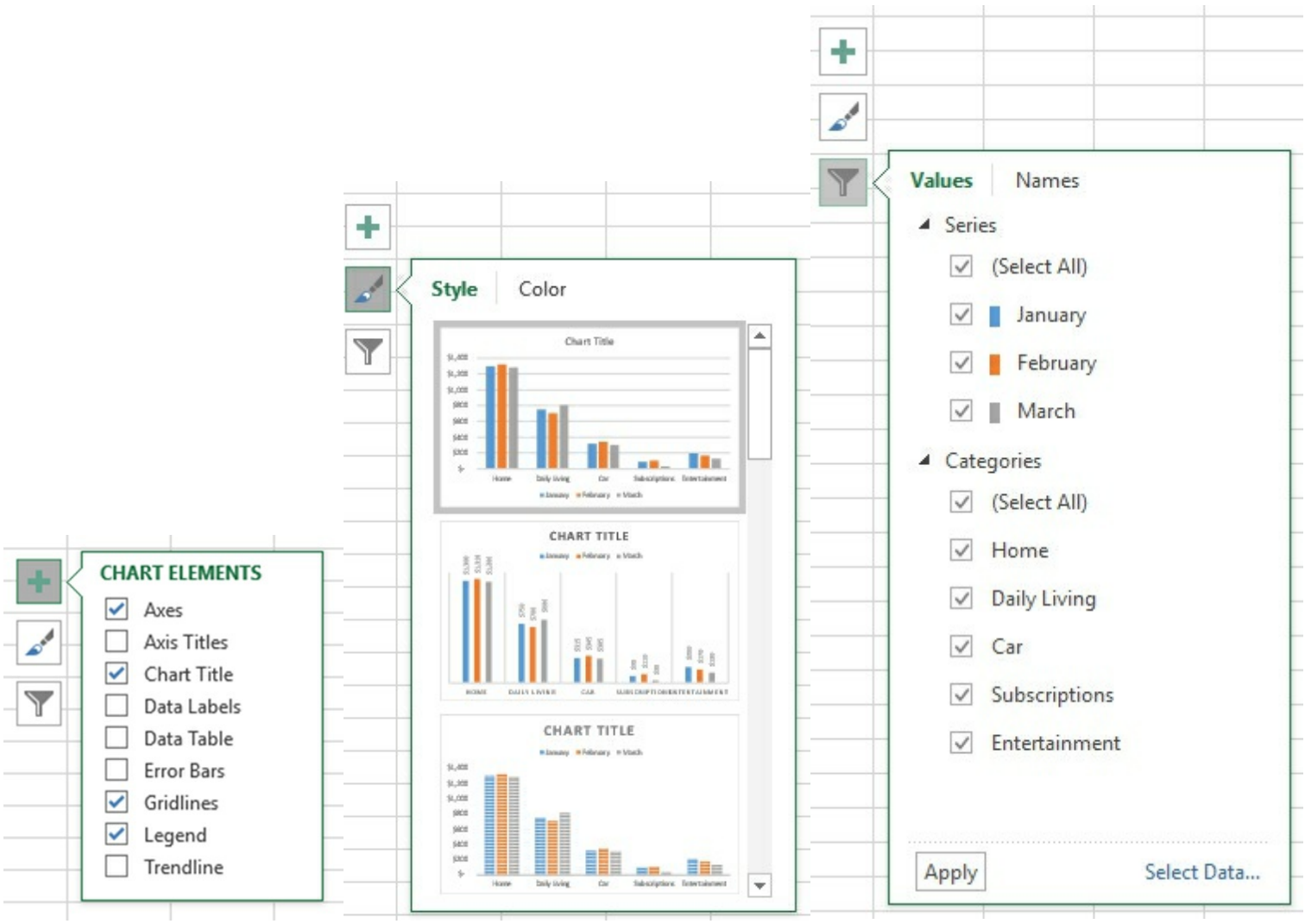


CHART ELEMENTS
Axes
Axis Titles
Chart Title
Data Labels
Data Table
Error Bars
Gridlines
Legend
Trendline
Style
Color
Chart Title
CHART TITLE
CHART TITLE
Values
Names
Series
(Select All)
January
February
March
Categories
(Select All)
Home
Daily Living
Car
Subscriptions
Entertainment
Apply
Select Data...

3.4 Advanced Charts

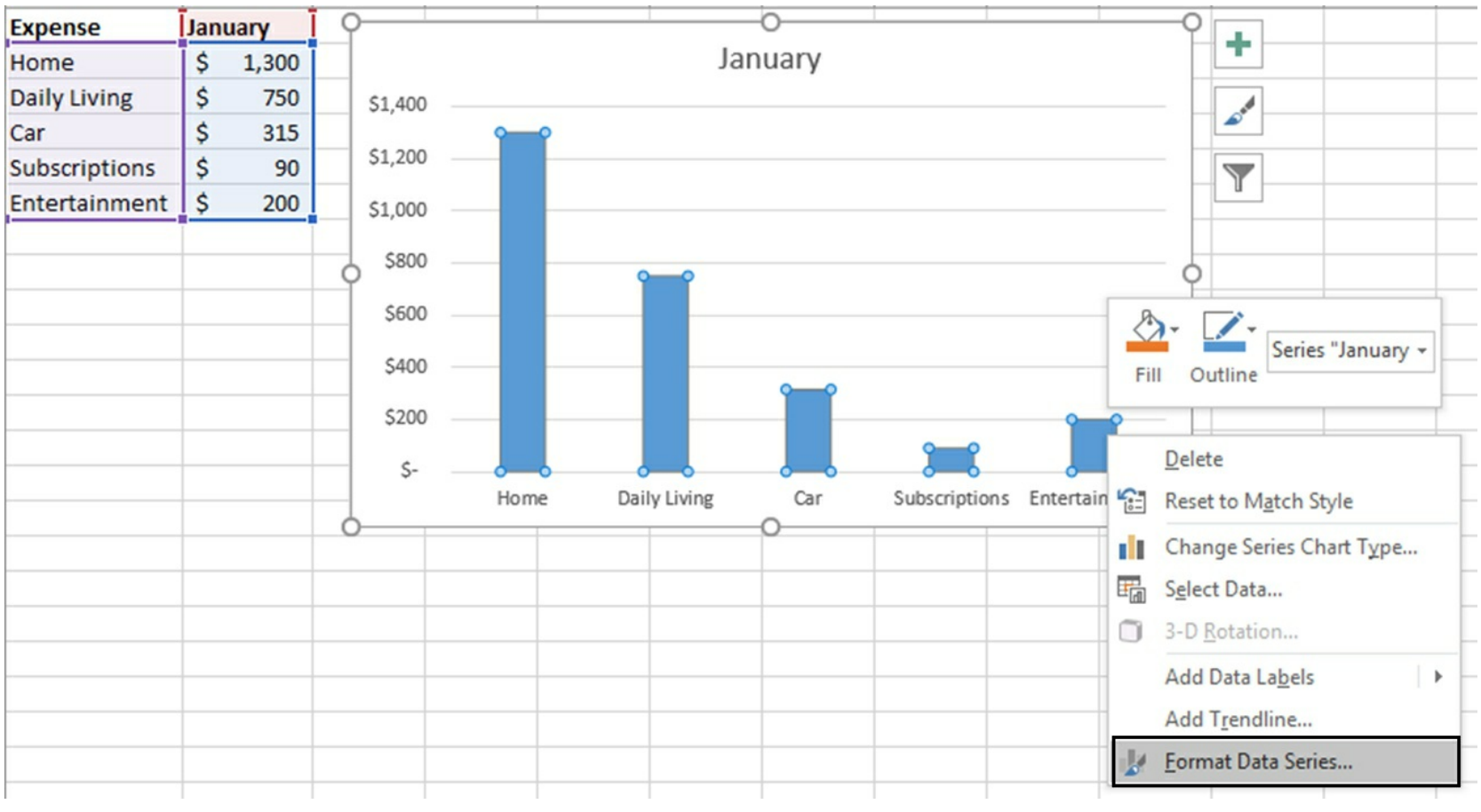
In the previous chapter, we discussed the most commonly used Excel charts and their usage in detail. Apart from the default charts available in Excel, there are many advanced charts that you can easily create and use in your day-to-day work. In this chapter, we will discuss three examples of advanced charts that can be useful when creating reports and dashboards.

1. Picture Charts

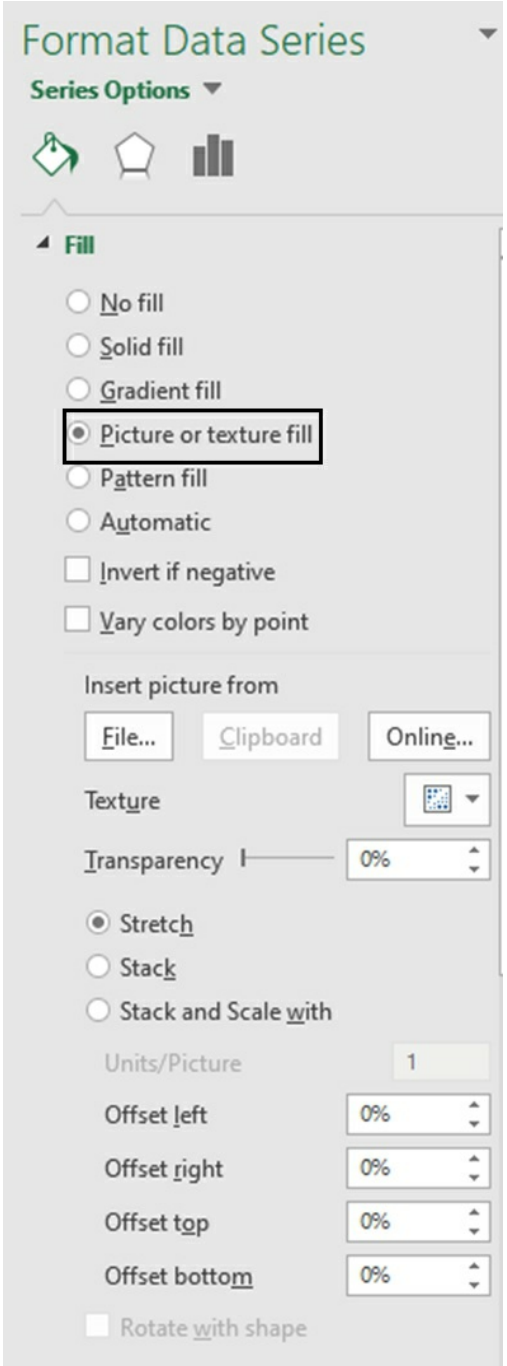
You can enhance a chart by inserting a picture instead of a data marker (column, bar or bubble) in the chart area. This will help deliver your message strongly. Let's practice!

Practice

1. Continue working on *Advanced Charts_Practice.xlsx* and go to the *Picture* sheet. The sheet contains data on personal expenses for January broken down by categories and a column chart based on the data.
2. Right click on a column on the column chart and select *Format Data Series*.



3. A *Format Data Series* pane will pop up on the right of the screen. Under the *Fill and Line* tab, expand options under *Fill*, if needed and select *Picture or texture fill*. Under *Insert picture from*, click on *File* and provide the .jpg file you downloaded as part of lesson downloads (*3.4 Picture.jpg*).



4. The picture will appear in place of columns in the chart.

2. Progress Doughnut Chart

The progress doughnut chart displays the percentage of completion providing detail to your audience on how close a team might be to completing a task or achieving a target. It is hence used for reporting the progress of a target or a project. Let's get into it!

Practice

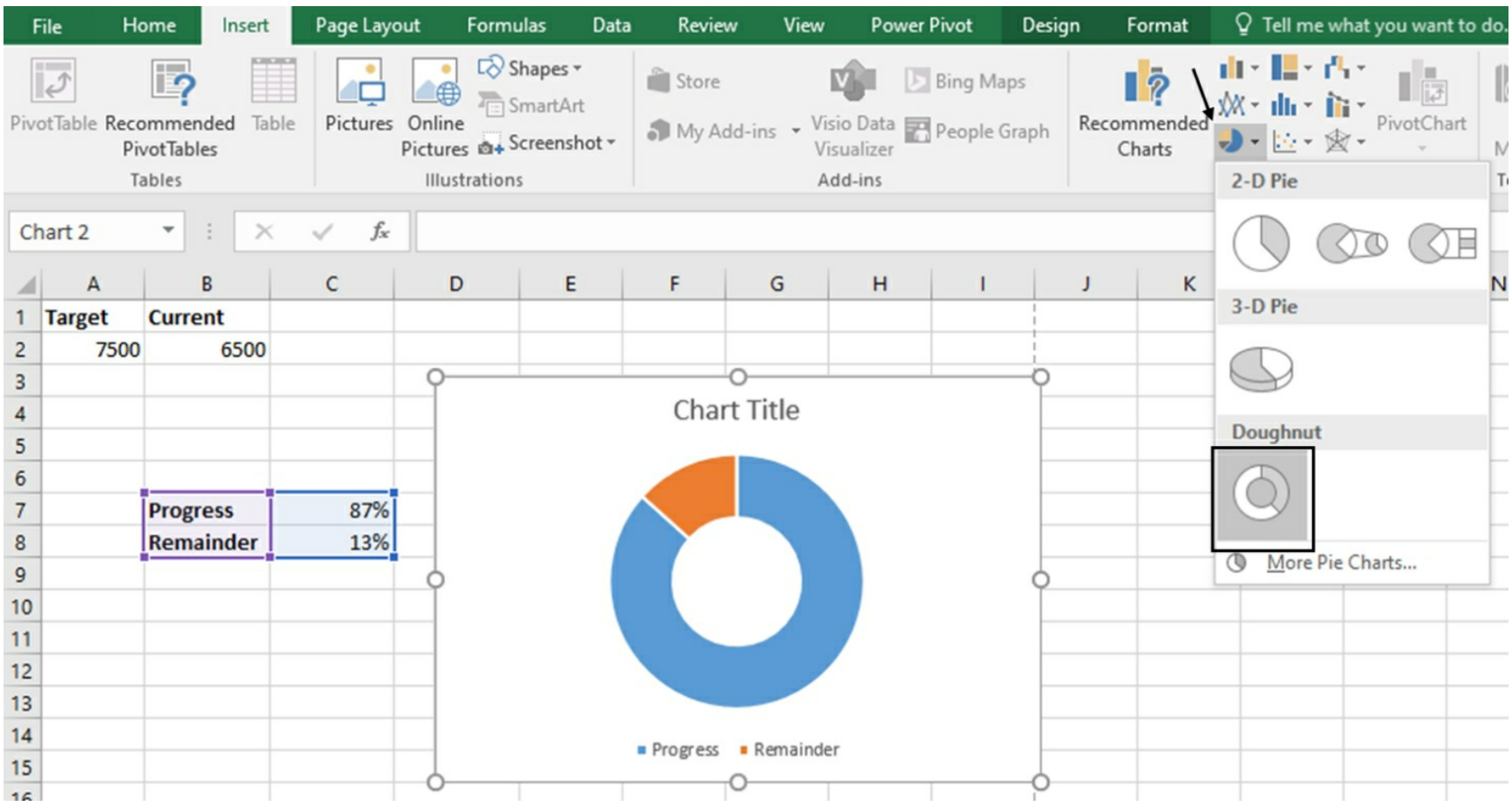
1. *Continue working on Advanced Charts_Practice.xlsx and go to the Progress sheet.*
2. We will first calculate progress in percentage. To do that, enter the below formula in cell C7 and format the cell as *Percentage*.

=B2/A2

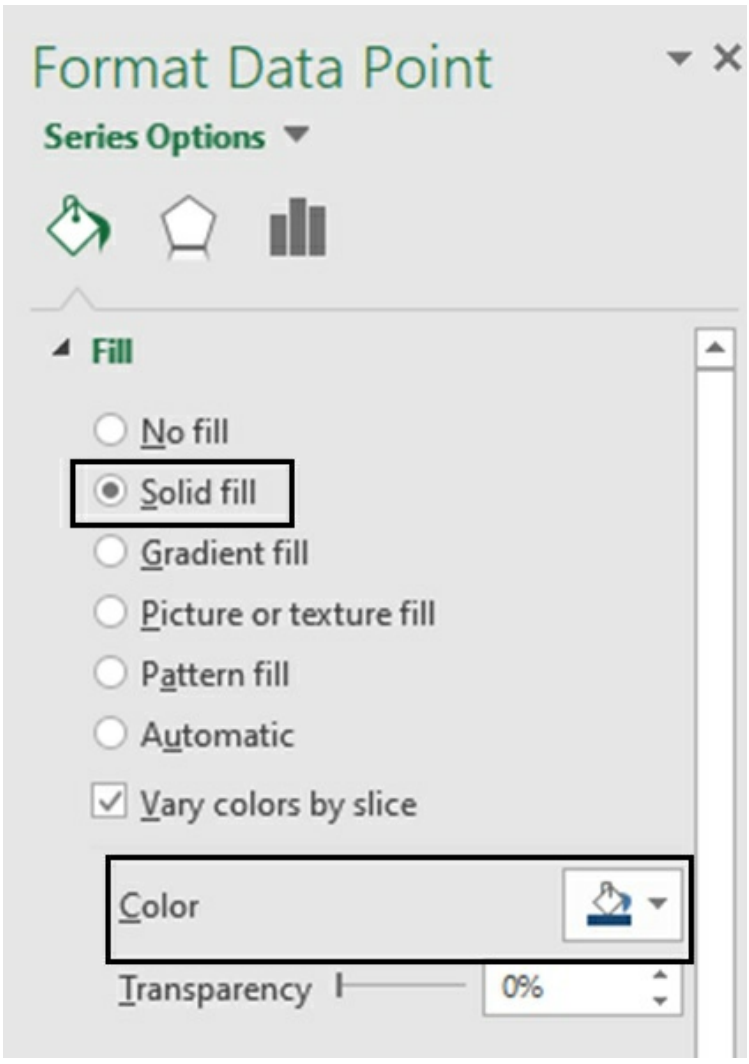
3. To calculate remainder work in percentage, enter the below formula in cell C8 and format the cell as *Percentage.*

=1-C7

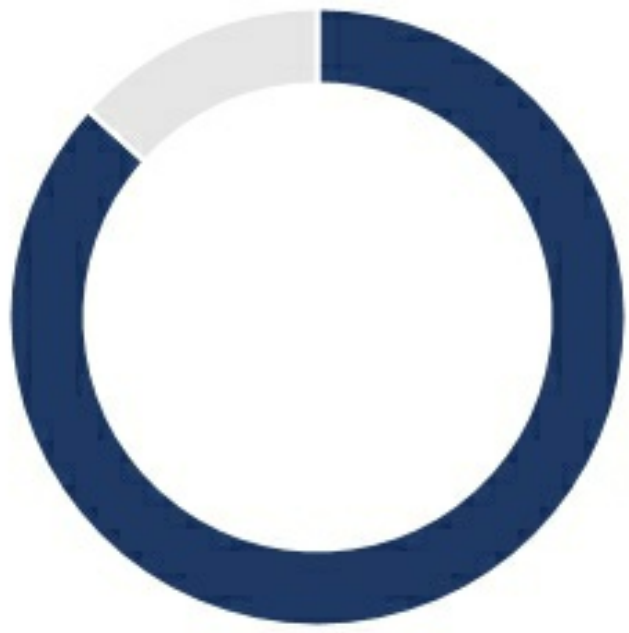
4. Now to insert the doughnut chart, select the data range B7:C8, go to the *Insert* tab and select *Doughnut Chart* from the Pie Chart drop-down menu.



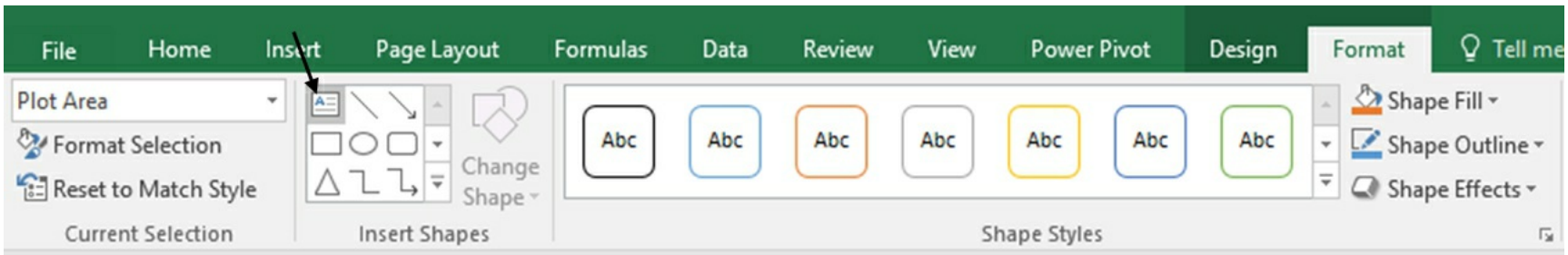
5. The doughnut chart will be inserted on the sheet.
6. Now select the legend and press the delete key to remove the legend.
7. To change the colors of the progress and remainder bars, left-click on the progress bar twice to select it. A *Format Data Point* pane will pop up on the left. Under the *Fill and Line* tab, expand options under *Fill*, if needed, select *Solid fill* and change the *Color* to a bold color, for example, dark blue.



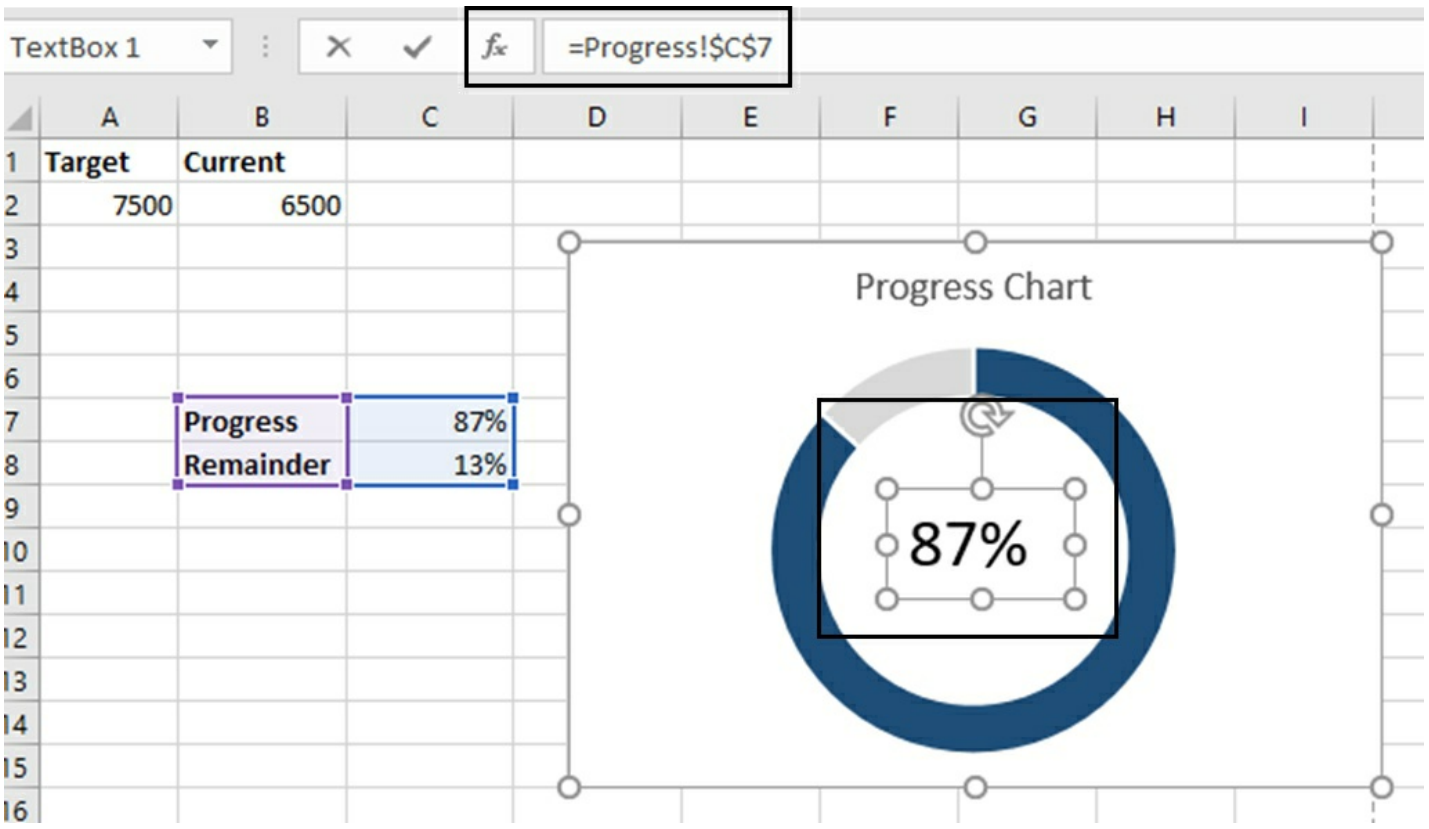
8. Repeat the above step for the remainder bar and select a light color, for example, gray.
9. Rename the chart title to *Progress Chart*. The doughnut chart should now look more like a progress chart.



10. The last step is to add a label with the percentage progress value. This will be done by adding a *Text Box* shape to the chart that displays the percentage in the middle of the circle. To do that, select the chart to display Chart Tools on the top ribbon. Go to the *Format* tab and select the text box shape from the *Insert Shapes* group.



11. Draw the text box inside the chart. This will add the text box to the chart. Now, select the text box and in the formula bar, type the equal sign (=), select cell C7 and press enter. Change the formatting of the text box to meet your needs.



3. Gantt Chart

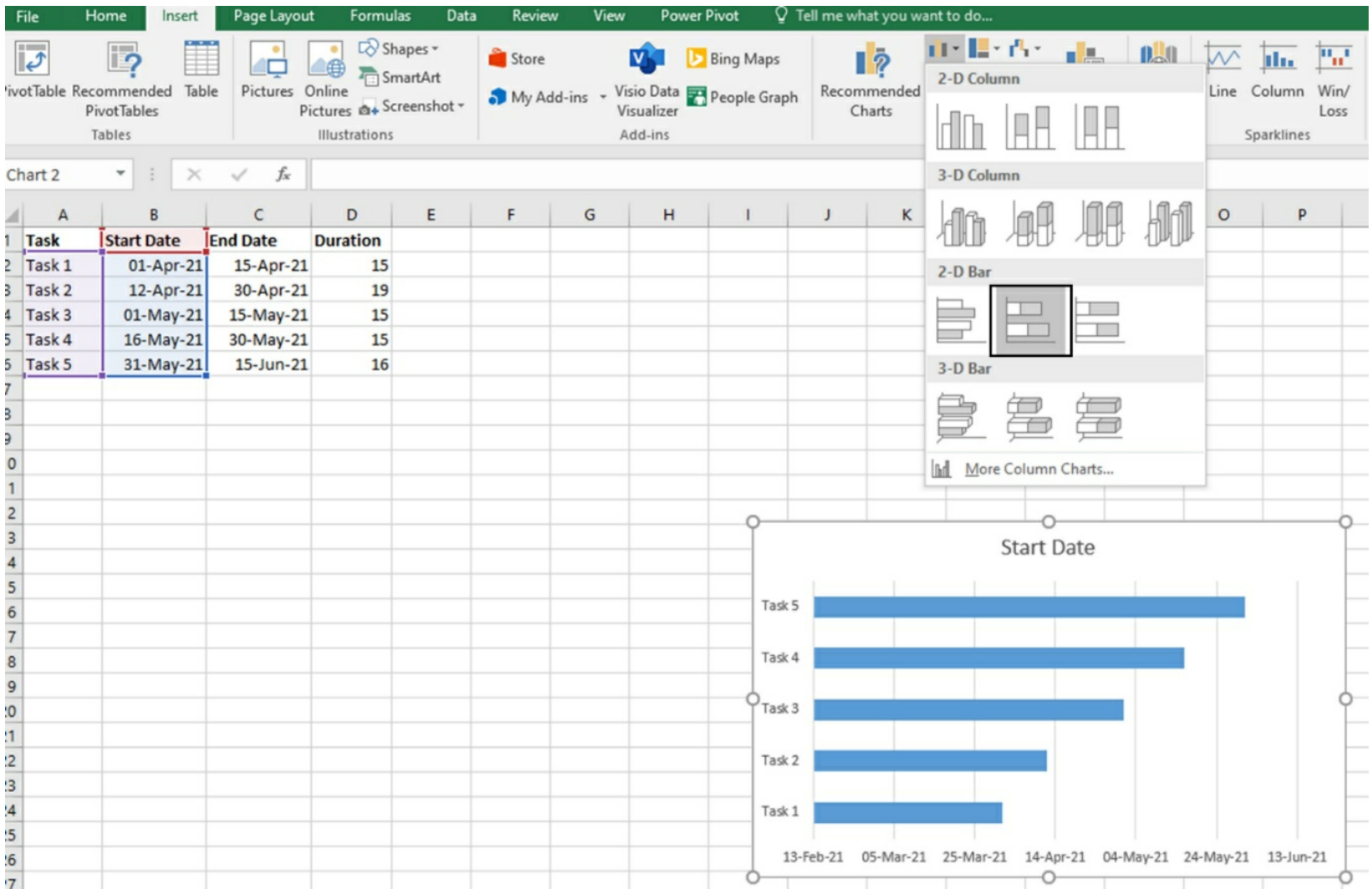
The Gantt chart represents tasks within a project in the form of cascading horizontal bar charts. As it breaks down individual tasks according to their start and finish dates as well as duration, it helps to track the tasks against their schedule. Excel does not have a default Gantt chart template. However, we can quickly create a Gantt chart in Excel by using the bar graph and some formatting. Let's see how it's done!

Practice

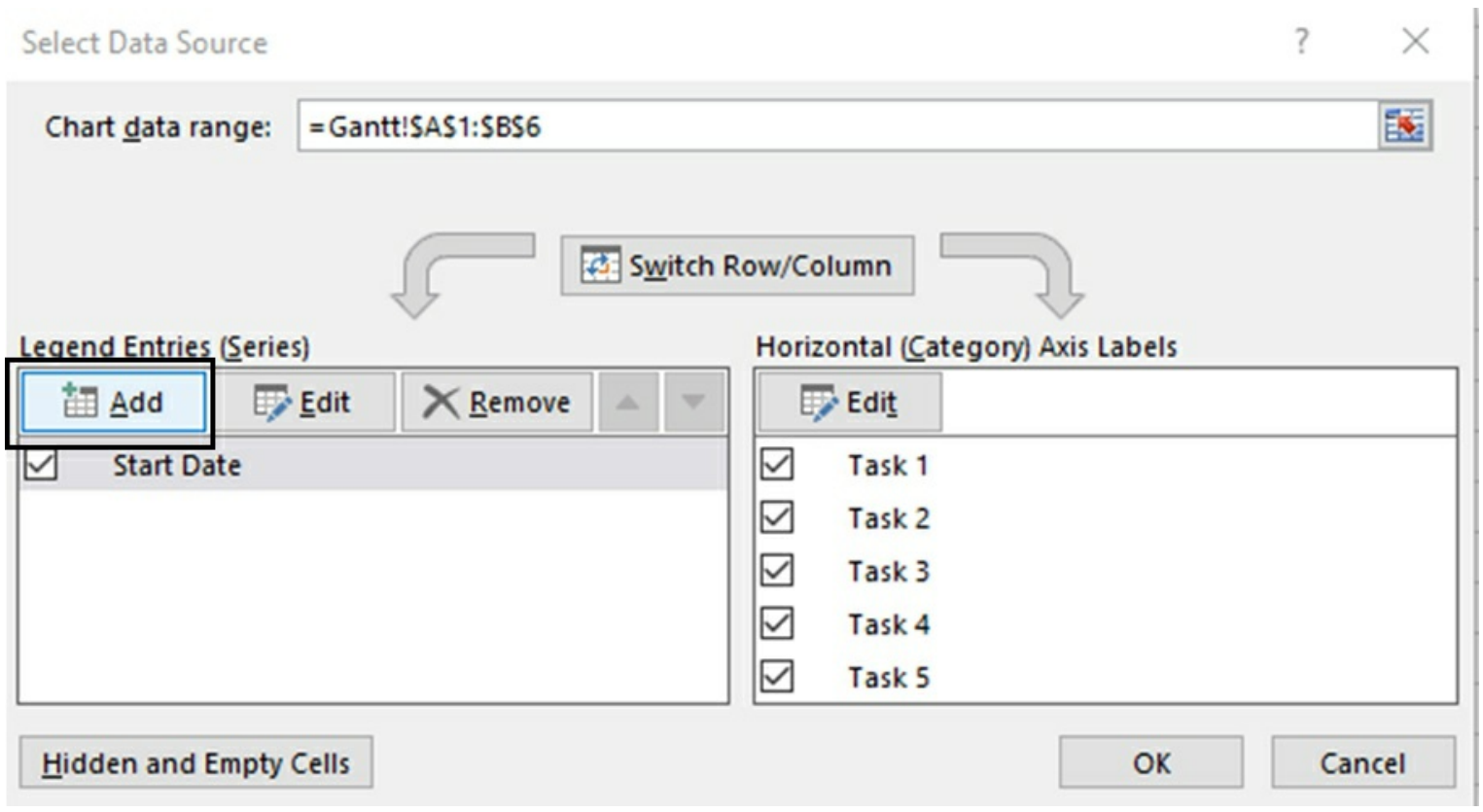
1. Continue working on *Advanced Charts_Practice.xlsx* and go to the *Gantt* sheet. The sheet contains data on start date and end date of individual tasks in a project, as well as their duration calculated

based on those dates.

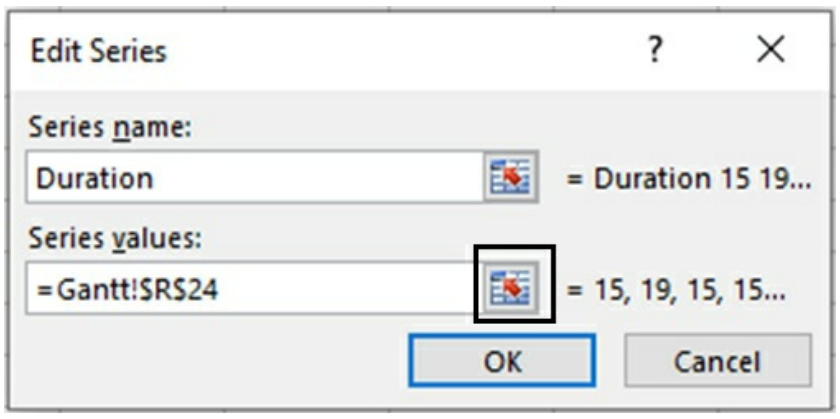
2. Select data in columns A and B (A1:B5). Go to the *Insert* tab and under the *Charts* group, click on *Insert Column or Bar Chart* icon and select *Stacked Bar Chart* under the *2-D Bar* group. A bar graph will be inserted into the spreadsheet.



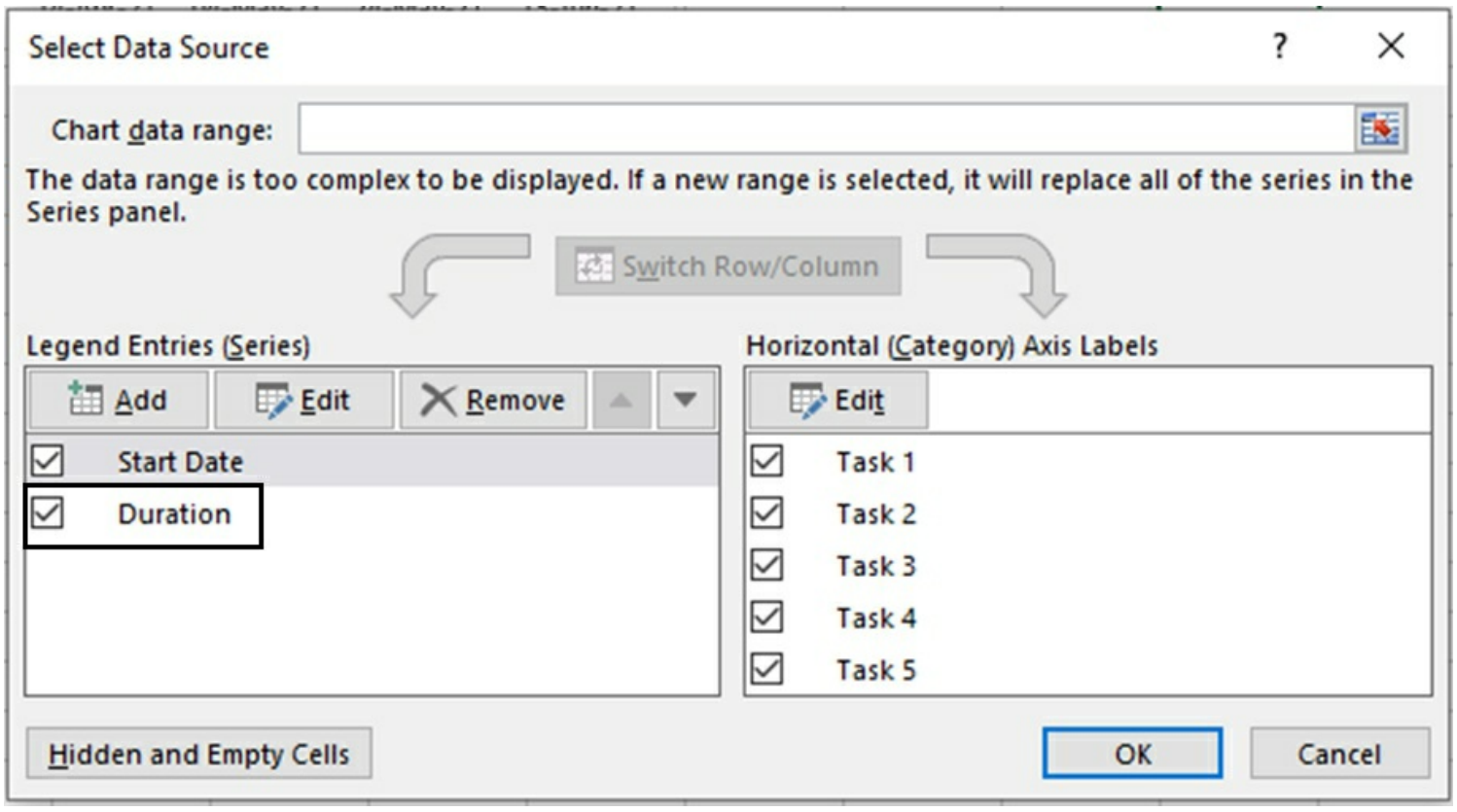
3. Now you need to add the duration data series into your chart. To do that, right-click anywhere on the chart and choose *Select Data*. The *Select Data Source* window will open. Click the *Add* button.



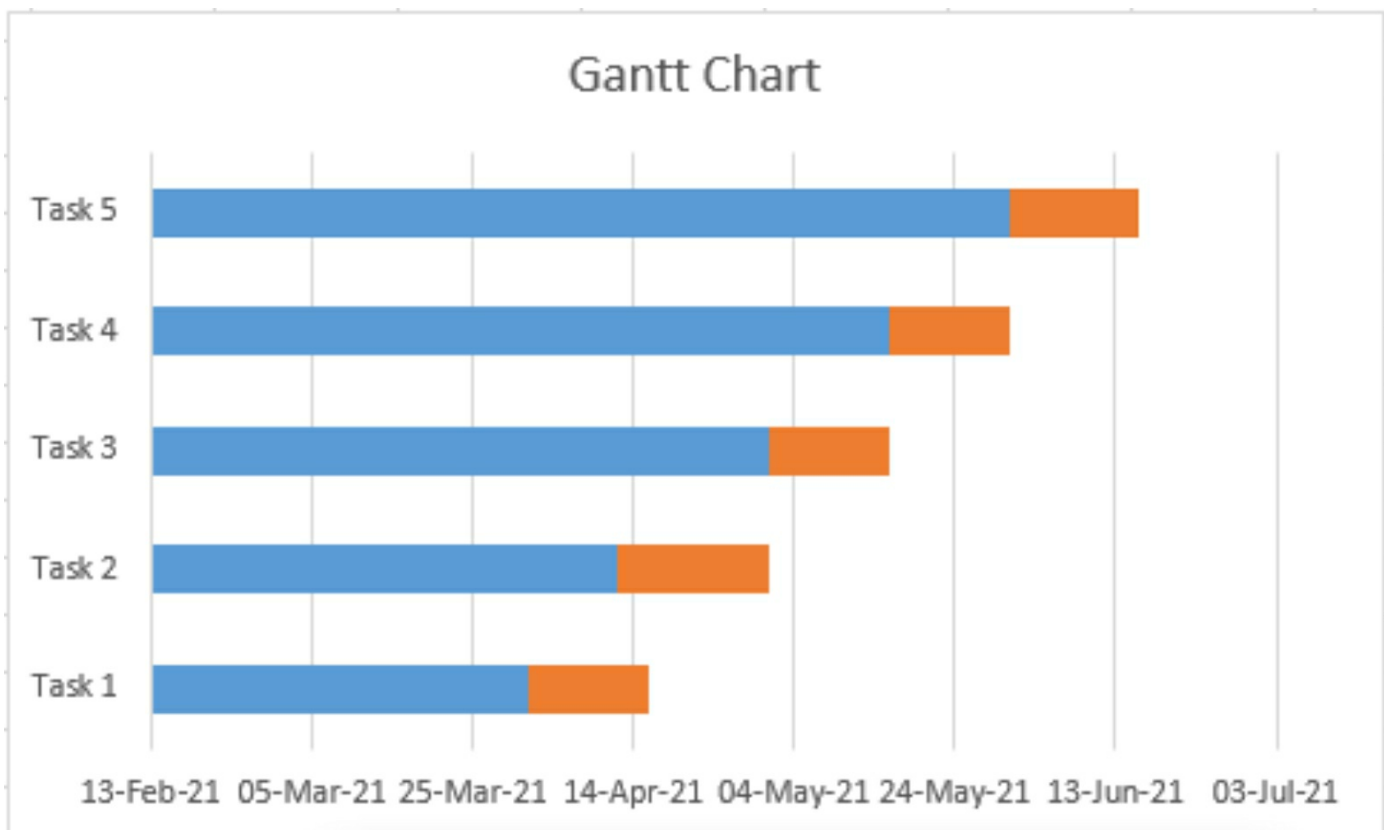
4. The *Edit Series* window will open. For *Series name*, type "Duration". For *Series values*, click the range selection icon (containing a red arrow) next to the Series values field. A second *Edit Series* window will open. Select cell range containing project Duration data (D2:D6). Make sure you have not mistakenly included the header or any empty cell. Close the second window. Click *OK* on the main *Edit Series* window.



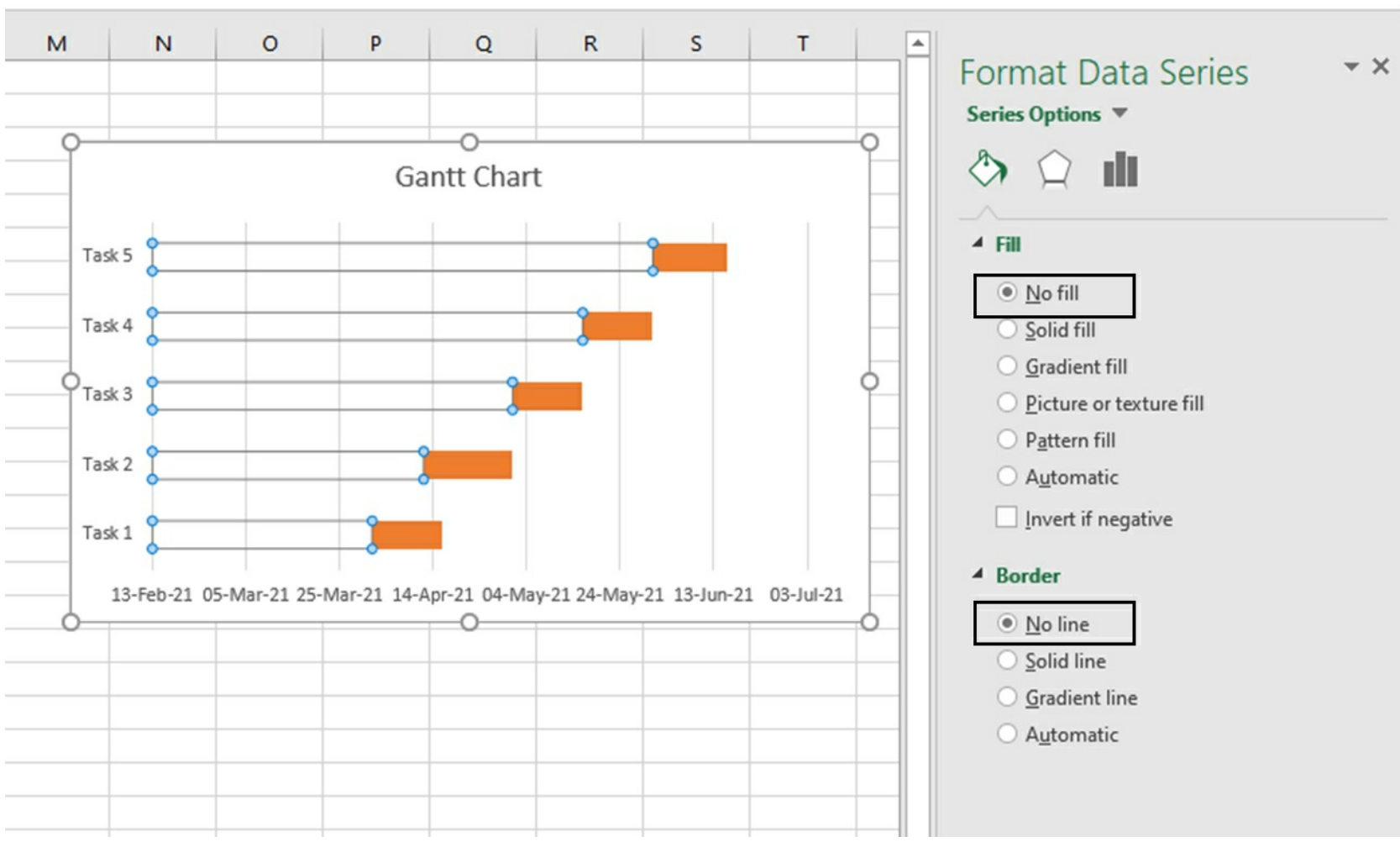
5. Now you will be back to the *Select Data Source* window. Ensure the Duration data series has been added and click *OK*.



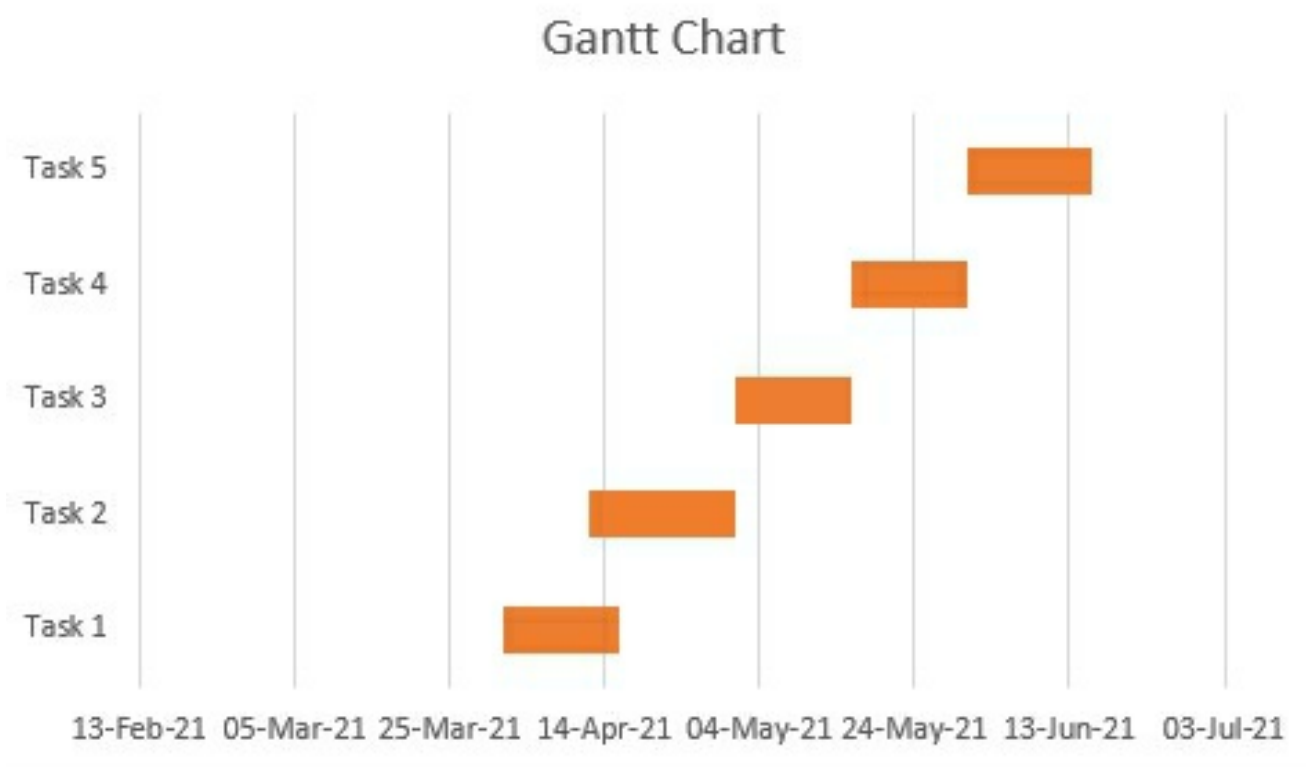
6. This will result in a stacked bar chart similar to the below.



7. Now, right click on any blue bar in the chart to select them all, right click and choose *Format Data Series*.
8. The *Format Data Series* window will pop up. For *Fill*, select *No Fill* and for *Border*, select *No Line*.



9. This will result in a Gantt chart that can be formatted further to meet your needs.



3.5 Charting Best Practices

Charts help the audience visualize complex data, reveal patterns, provide context, and describe relationships within data. In order to make a chart effective, the basics of good chart design can't be overlooked. We will discuss key best practices on how to design charts that present patterns clearly. Let's get started.

1. Define a clear purpose: Your chart should answer a specific question posed for the target audience. Whatever the intent, for example, track performance, monitor customer behavior, and measure effectiveness of processes, take time at the beginning to clearly define the purpose of the chart and the insights you want the audience to take away from the chart.

2. Know your audience: Tailor your chart to the needs and expertise of the target audience. It should allow the audience to view and process data easily and quickly. Take into account the familiarity of the audience to the charts being used. Use simple charts if the audience is not accustomed to viewing charts on a regular basis.

3. Select correct chart type: There are so many different types of charts. Choosing the right chart for visualizing your data depends primarily on the type of data you need to visualize. Select the chart that visualizes your data most effectively and efficiently conveys the essential information in your data. Below are a few pointers on how different types of charts can be used to represent different type of information:

- Use line graphs to measure variations or patterns over time. They are excellent for displaying both large and small changes.
- Bar and column charts are employed to assess varying values across different categories. They can be used to track changes over time as well, but are best used only when those changes are significant.
- Scatter plots are used to display the relationship between two variables.
- Use pie and donut charts if the proportion of various categories in the data is of interest.
- Use heatmaps to display and compare hierarchical data graphically.

4. Readability: Ensuring consistency and clarity in chart's design aspects, such as axis or gridlines, can also enhance the overall readability of the chart. Below are a few tips on how to improve the readability of the charts:

- Provide context, where needed. Use color, arrows, text, and other visual cues to help viewers interpret information on the graph at first glance.
- For multiple categories with long names, use a bar graph and arrange data in order of interest (usually, descending).
- To understand a line chart with a legend, readers are forced to scan back and forth between the legend and the chart itself. Where applicable, label lines individually instead of including a legend.
- Label your chart effectively. Include axes labels. Include value labels on bars, columns, and pie slices to enhance interpretability.
- Keep the chart simple. Avoid unnecessary complexity. For example, avoid borders and background colors or use light colors when they are needed. Avoid or space out gridlines.
- Don't overload the chart with too much data. Play with the level of detail in your chart to find the balance that's right for your audience.
- The numerical axis (usually the y-axis), more often than not, starts at zero. There however are exceptions, for example, if you need to chart changes in stock prices of a company during a day. With a zero baseline, it will be almost impossible to see the ups and downs of the data. Hence, the zero baseline will hinder chart's usage. But please remember that as soon as you prune your axis you're zooming into data, so do so only when absolutely necessary.
- The chart should pass the squint test. Looking at your graph, squint your eyes until all of the text and numbers are blurred. Can you still comprehend all elements of the chart? If not, identify what is getting lost and go back and emphasize it through size and color adjustments.

5. Make efficient use of colors: Colors can be a useful tool in the charts as you can effectively communicate important information about your data through the use of different color combinations. Below are key considerations to take into account to use color in your charts:

- Use white or very light backgrounds.
- Use medium gray axis lines and borderlines, and light gray gridlines if they are necessary.
- Use lighter fill colors and no borders for areas and bars.
- Use bolder colors for lines and markers.

- Use darker colors for data labels but medium grays for axis labels and titles.
- Use colors for emphasis. While it is possible to assign different colors to each bar in a chart, do so for select bars that look different to draw attention to them. This will also simplify less important information.
- Avoid gradients, shadows, glows.
- Don’t use too many colors.

6. Avoid 3-D charts: 3D space makes it too difficult to compare the values, volumes and areas to one another as it distorts the visual relationship and leads to mistakes in data interpretation. Hence, 3-D charts reduce comprehension and should be avoided.

Lesson 4: Dashboarding

Press the Ctrl key and click here to download practice & solution workbooks. Alternatively, copy and paste below link to your browser:

https://www.dropbox.com/sh/j03v20n2lr373l3/AACiDgGjn3XyakQbfRF5pnk6a?dl=0

4.1 Identify Key Performance Indicators

In today's age, all businesses irrespective of their size collect data to monitor performance, streamline processes, and understand their customers on a deeper level. All this data needs to be analyzed and reported in form of business insights to the management so that it can be employed as key evidence to make informed decisions. The key question here is what should be reported from the data that provides value to the management and how. One of the most effective means of doing this is by identifying key performance indicators (KPIs) specific to your business and reporting them in form of interactive dashboards.

In this lesson, we'll study KPIs and their types and discuss how they are created and reported as dashboards. We will also practice creating a couple of cool dashboards! The purpose of this lesson is to bring all the concepts studied in previous lessons home. So, we will be constantly referring to previous chapters and the concepts covered in them especially in the *Practice* sections. Ensure to leave some additional time for the *Practice* sessions and take this as an opportunity to review some of the previously covered content.

KPIs

A KPI is a data point measured or calculated to assess the performance of an organization or a project. KPIs deliver critical performance-based information that enables the organization and its stakeholders to understand how various aspects of the business are performing. This helps with making necessary changes or decisions in a timely manner and seeing their impact in the future by monitoring the KPIs. Some examples of KPIs grouped based on their function are mentioned below:

1. **Input KPIs:** Input KPIs are used to measure resources needed for a business process or project. Some examples include staff time, number of staff, budget.
2. **Process KPIs:** Process KPIs are used to determine the efficiency of activities involved to achieve the outcome. KPIs focused around customer service by phone fall under this category. These include calls resolved, calls abandoned, resolution rate etc.
3. **Output KPIs:** Output KPIs measure the outcome of a process or business activity. Examples include revenues, profits and new customers acquired.
4. **Financial KPIs:** Financial KPIs measure economic stability, growth, and business viability. Common examples include gross profit, net profit and budget variance.
5. **Project KPIs:** These KPIs measure the progress of a project. Examples include % time spent, % completion and % budget spent.

Creating KPIs

Every business depending on its focus and size will have its own set of KPIs to determine the performance of the organization. To understand which KPIs will prove to be the most valuable to your business, sit down with your management team and key stakeholders and discuss which top three to five business data points would they want to monitor on a regular basis. These KPIs should instantly tell them what they need to know about the progress of the organization. A good KPI is well-defined, specific, measurable and relevant to your business. Don't try to create too many measures – keep it simple to avoid confusion and contradiction. Create thorough definitions of KPIs for review and approval by the management team. Discuss how frequently they should be reported to the management.

KPIs also need to be managed. This involves changing and evolving KPIs as business conditions and priorities change. Regularly review the KPIs and if they no longer serve what your organization is trying to achieve, collaborate with the management team to revise them.

KPI reports

A KPI report is a management tool that facilitates the organization and presentation of the business KPIs. These reports help companies to identify strengths, weaknesses and trends and make appropriate changes to achieve organizational goals.

There are three types of KPI reports: operational, strategic and analytical.

1. **Operational KPI Reports**

An operational KPI report is used by operational managers to monitor real-time or transactional data against key KPIs. They provide information on the day-to-day activities of an organization to help make decisions or take action. For example, in a sales department, an operational KPI report will cover opportunities created and

opportunity-to-win ratio of each of the sales employees.

2. **Strategic KPI Reports**

A strategic KPI report is used by executives to monitor the overall performance of company KPIs. Strategic reports are updated less frequently than the operational reports and summarize KPIs over a period of time (month, quarter, year).

3. **Analytical KPI Reports**

An analytical KPI report is created for a specific area of the organization to analyze a key trend in one or more KPIs more thoroughly. This allows users to investigate trends, predict outcomes, discover insights and establish targets based on insights into historical data.

Building a KPI report

KPI reports can be static or interactive. A static report is a text or table-based tool. It seldom contains visual or interactive parts. An interactive report, on the other hand, allows users to investigate the data by drilling down the individual KPIs dynamically. As KPIs reported in the form of an interactive dashboard offer additional visibility into the data they report, we will focus on them for the rest of the lesson.

A KPI dashboard is a visual representation of KPIs that allows users to quickly view and analyze data in one place. It presents data in a user-friendly, digestible format that makes it easy to extract important information and act on it. In the past, Excel reporting involved generating multiple static reports for different users depending on their needs. However, now Excel offers features like dropdowns, PivotTables, PivotCharts, slicers and timelines to automatically expand and contract data to display only the information that users want to see. Excel is an excellent choice for creating and distributing dashboards due to its ease of use and universal availability.

In the next chapter, we will practice creating our first dashboard that incorporates interactivity with help of dropdowns. Then, we will discuss PivotCharts, slicers and timelines in greater detail. This will provide a look and feel of dashboard features before we dive deeper.

4.2 Create a Process Dashboard

As discussed earlier, a process dashboard is a management tool that showcases essential information on how different activities are performing across a business. It provides high-maturity, KPI-intensive data necessary for process analysis and decision making.

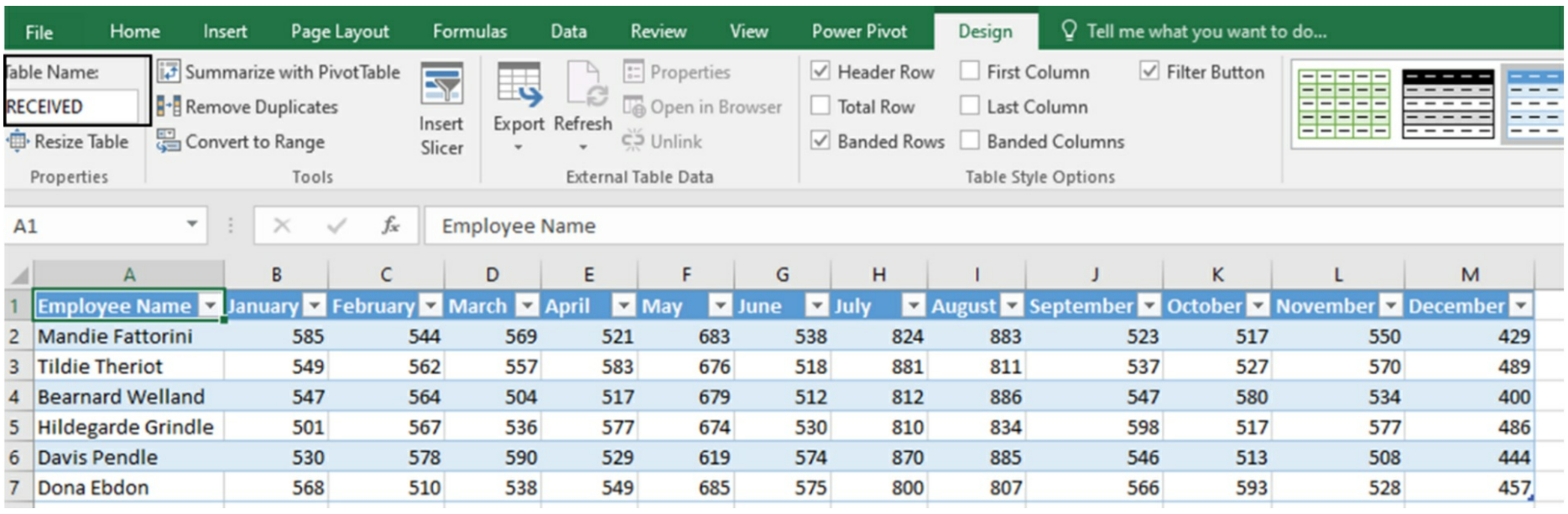
For example, call centers intend to track relevant process KPIs to identify insights that will lead to improved customer support with the help of professional reports and interactive dashboards. Suppose the management team of a small call center with six employees wants you to create an interactive KPI dashboard with the 2019 data. You sat down with the management team and determined that the below mentioned KPIs are of most value to them.

1. Abandon Rate – overall percentage of calls abandoned or not answered.
2. Resolution Rate – overall percentage of calls resolved
3. Customer Rating – overall customer rating

You are provided with four data sets with calls received, calls answered calls resolved and customer rating for each month of 2019 and each employee. You would follow a series of steps to create a dashboard.

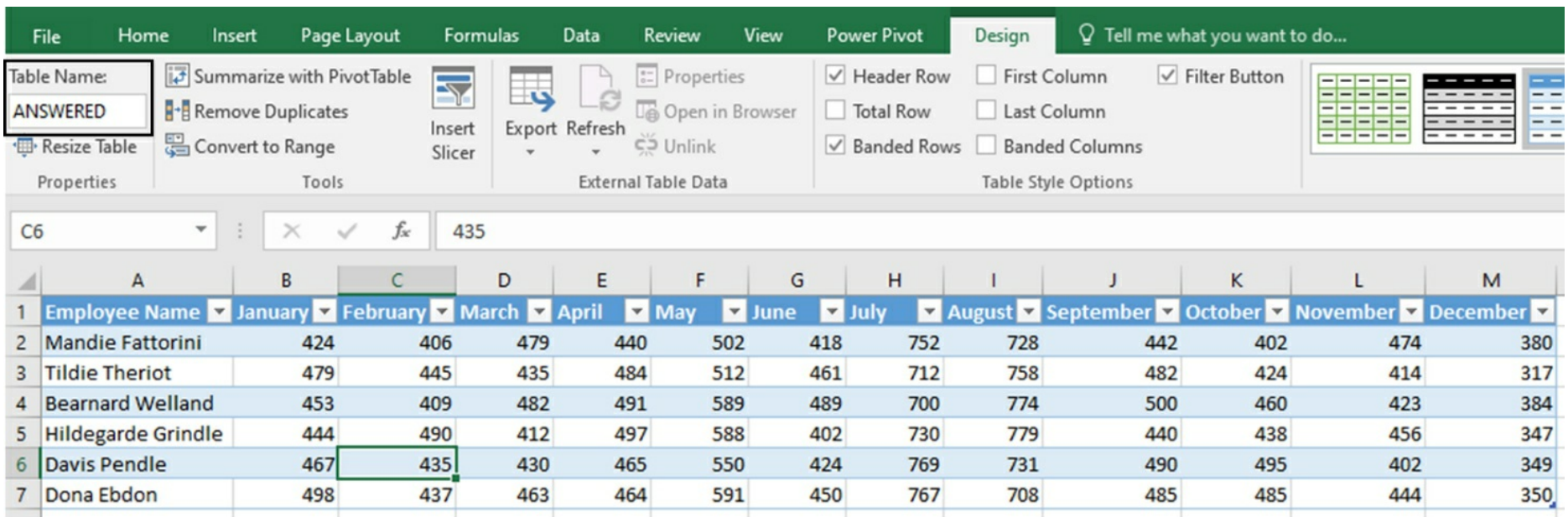
Practice

1. Open *Process Dashboard_Practice*.xlsx and go to the *Calls Received* sheet. Convert the data on the tab to a named table called RECEIVED. We had thoroughly discussed the steps to do so in Chapter 1.3. Refer to the chapter, if needed.



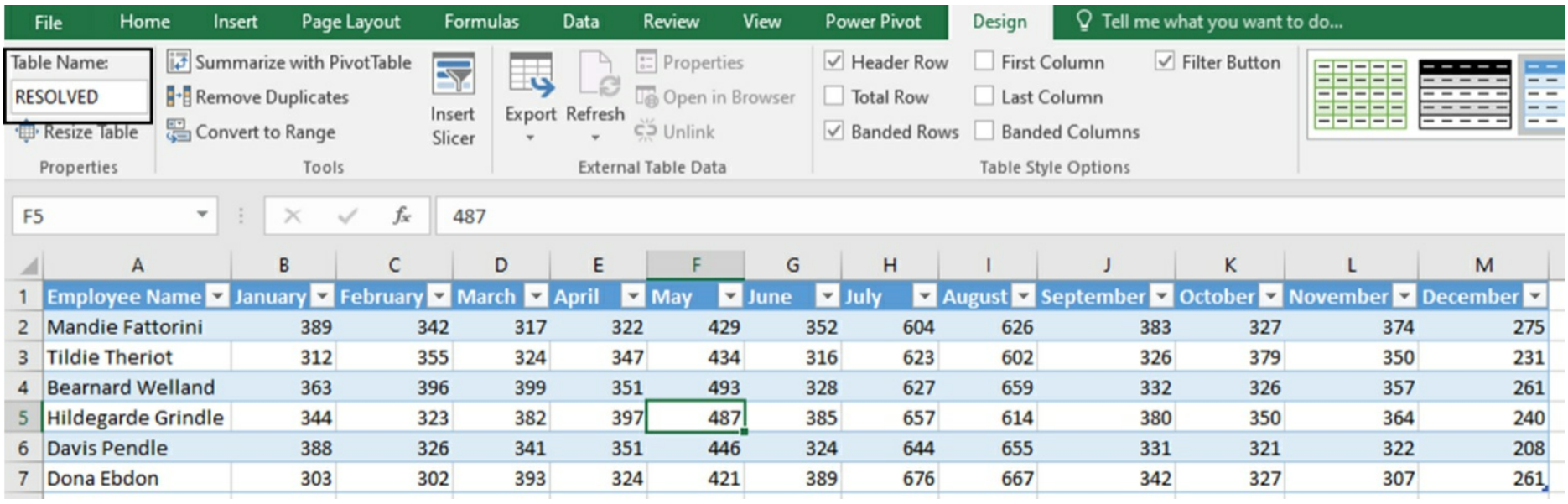
Employee Name	January	February	March	April	May	June	July	August	September	October	November	December
Mandie Fattorini	585	544	569	521	683	538	824	883	523	517	550	429
Tildie Theriot	549	562	557	583	676	518	881	811	537	527	570	489
Bearnard Welland	547	564	504	517	679	512	812	886	547	580	534	400
Hildegarde Grindle	501	567	536	577	674	530	810	834	598	517	577	486
Davis Pendle	530	578	590	529	619	574	870	885	546	513	508	444
Dona Ebdon	568	510	538	549	685	575	800	807	566	593	528	457

2. Now, go to the *Calls Answered* sheet. Covert the data on the tab to a named table called ANSWERED.



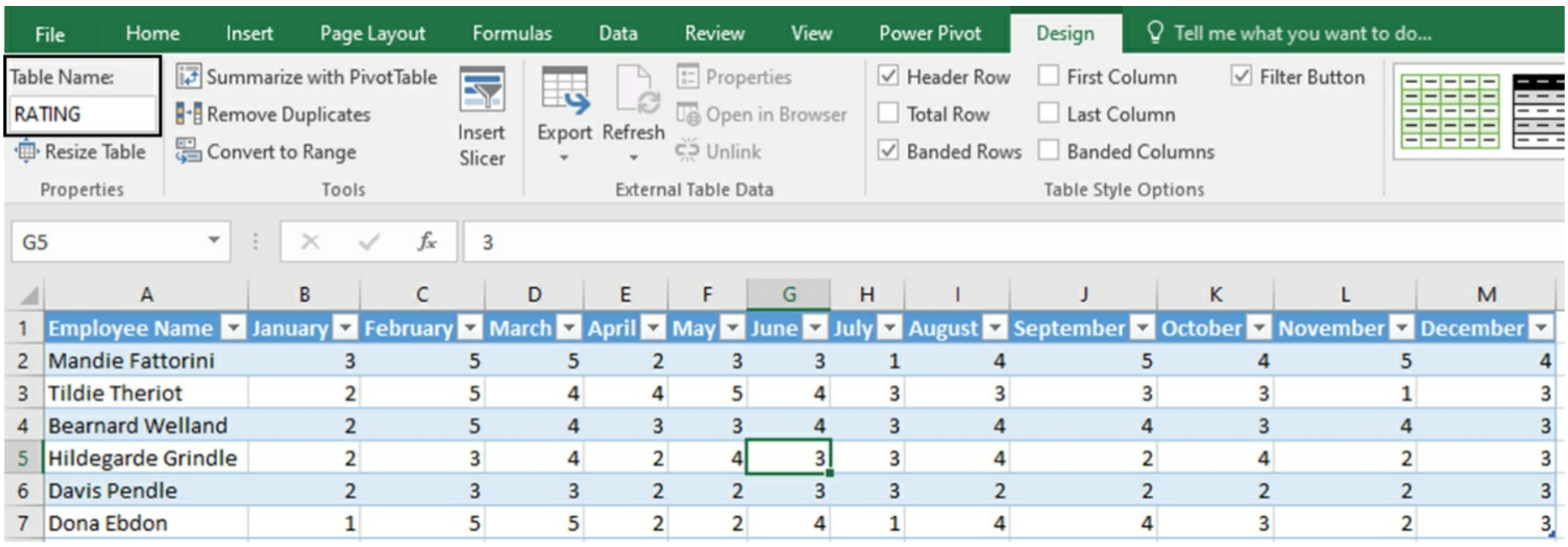
Employee Name	January	February	March	April	May	June	July	August	September	October	November	December
Mandie Fattorini	424	406	479	440	502	418	752	728	442	402	474	380
Tildie Theriot	479	445	435	484	512	461	712	758	482	424	414	317
Bearnard Welland	453	409	482	491	589	489	700	774	500	460	423	384
Hildegarde Grindle	444	490	412	497	588	402	730	779	440	438	456	347
Davis Pendle	467	435	430	465	550	424	769	731	490	495	402	349
Dona Ebdon	498	437	463	464	591	450	767	708	485	485	444	350

3. Go to the *Calls Resolved* sheet. Covert the data on the tab to a named table called RESOLVED.



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Employee Name	January	February	March	April	May	June	July	August	September	October	November	December
2	Mandie Fattorini	389	342	317	322	429	352	604	626	383	327	374	275
3	Tildie Theriot	312	355	324	347	434	316	623	602	326	379	350	231
4	Bearnard Welland	363	396	399	351	493	328	627	659	332	326	357	261
5	Hildegarde Grindle	344	323	382	397	487	385	657	614	380	350	364	240
6	Davis Pendle	388	326	341	351	446	324	644	655	331	321	322	208
7	Dona Ebdon	303	302	393	324	421	389	676	667	342	327	307	261

4. Go to the *Customer Rating* sheet. Covert the data on the tab to a named table called RATING.

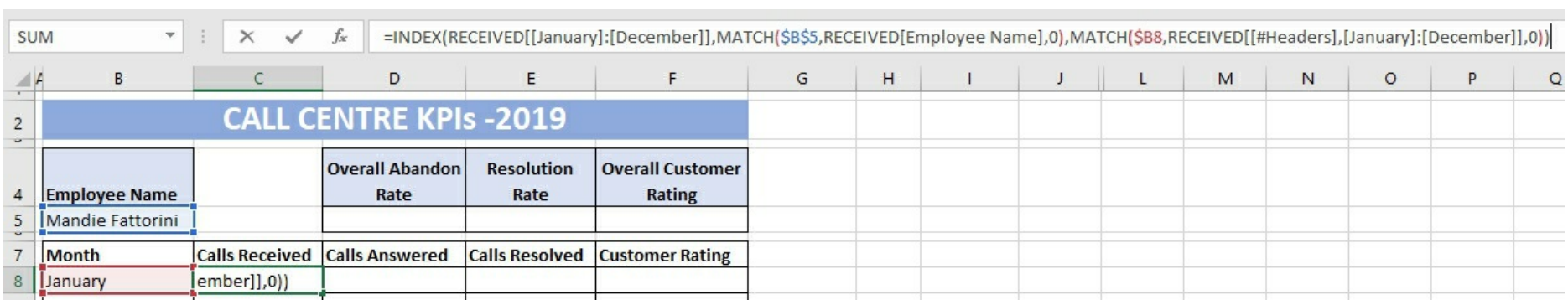


	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Employee Name	January	February	March	April	May	June	July	August	September	October	November	December
2	Mandie Fattorini	3	5	5	2	3	3	1	4	5	4	5	4
3	Tildie Theriot	2	5	4	4	5	4	3	3	3	3	1	3
4	Bearnard Welland	2	5	4	3	3	4	3	4	4	3	4	3
5	Hildegarde Grindle	2	3	4	2	4	3	3	4	2	4	2	3
6	Davis Pendle	2	3	3	2	2	3	3	2	2	2	2	3
7	Dona Ebdon	1	5	5	2	2	4	1	4	4	3	2	3

5. Now go to the *KPIs* sheet that is essentially a dashboard template. You will notice that names of the employees are placed in range K4:K9 in column K. With the data validation methodology we studied in Chapter 1.6, embed cell B5 with a drop-down list that takes employee names stored in the range K4:K9. Refer to the chapter as needed.

6. To bring in calls received by the selected employee, enter and copy down below formula in cell C8 of the KPI's sheet.

=INDEX(RECEIVED[[January]:[December]],MATCH($B$5,RECEIVED[Employee Name],0),MATCH($B8,RECEIVED[[#Headers],[January]:[December]],0))



I know the formula looks scary but believe it or not, we have studied it before. It is a two-way INDEX and MATCH formula that we discussed in Chapter 2.4. It is a bit longer because we are using it with a named table and its structured references. We had covered the structured table references in Chapter 1.3.

Let's break the formula down a little to help understand it easily. As we know, the INDEX formula returns a value at the intersection of a row number and a column number in an array.

Syntax: =INDEX (array, row_num, [col_num])

=INDEX(RECEIVED[[January]:[December]],MATCH($B$5,RECEIVED[Employee Name],0),MATCH($B8,RECEIVED[[#Headers],[January]:[December]],0))

Hence,

RECEIVED[[January]:[December]] refers to the input array which is the range B2:M7 on *Calls Received* sheet, the data under columns January through December of RECEIVED table.

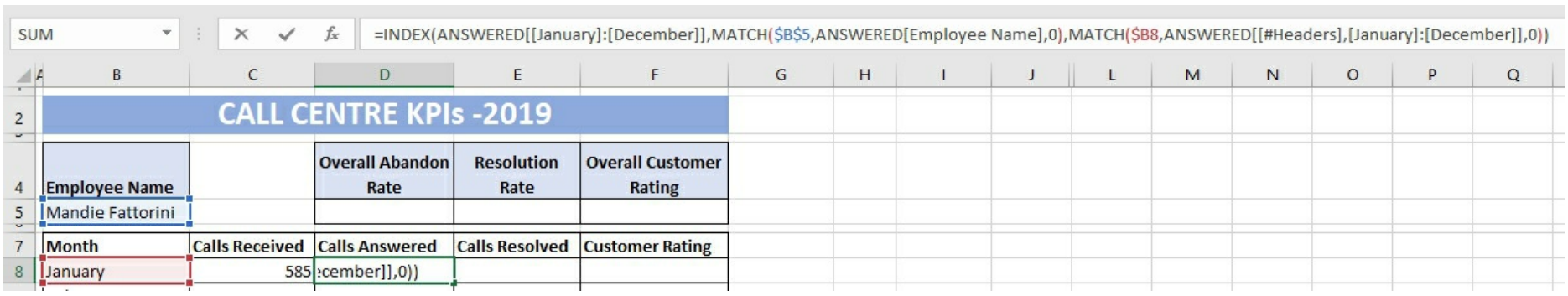
MATCH($B$5,RECEIVED[Employee Name],0) takes employee name stored in cells B5 of *KPIs* sheet as input and returns the row number matching the name of the employee in the *Employee Name* column of RECEIVED table. Note that reference to cell B5 is locked with absolute references discussed in Chapter 2.3.

MATCH($B8,RECEIVED[[#Headers],[January]:[December]],0) takes month from column A of *KPIs* sheet as input and returns the column number matching the headers of RECEIVED table. Note that $B8 is a mixed reference that locks the column reference but allows the row reference to change, as discussed in Chapter 2.3. This will help to drag the formula down efficiently.

The INDEX function then returns the value at the intersection of row and column numbers produced by the MATCH functions.

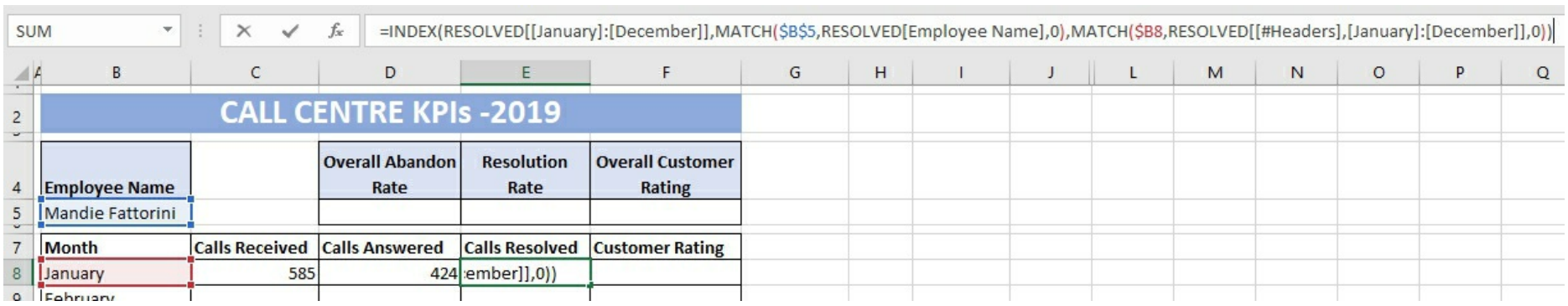
7. To bring in calls answered by the selected employee, enter and copy down below formula in cell D8 of the *KPIs* sheet.

=INDEX(ANSWERED[[January]:[December]],MATCH($B$5,ANSWERED[Employee Name],0),MATCH($B8,ANSWERED[[#Headers],[January]:[December]],0))



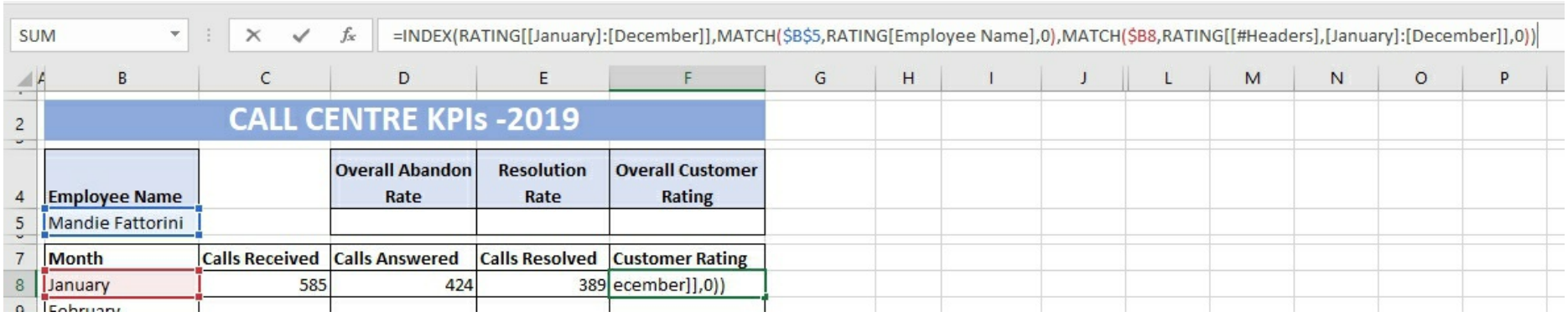
8. To bring in calls resolved by the selected employee, enter and copy down below formula in cell E8 of the KPI's sheet.

=INDEX(RESOLVED[[January]:[December]],MATCH($B$5,RESOLVED[Employee Name],0),MATCH($B8,RESOLVED[[#Headers],[January]:[December]],0))



9. To bring in Customer Rating for the selected employee, enter and copy down below formula in cell F8.

=INDEX(RESOLVED[[January]:[December]],MATCH($B$5,RESOLVED[Employee Name],0),MATCH($B8,RESOLVED[[#Headers],[January]:[December]],0))

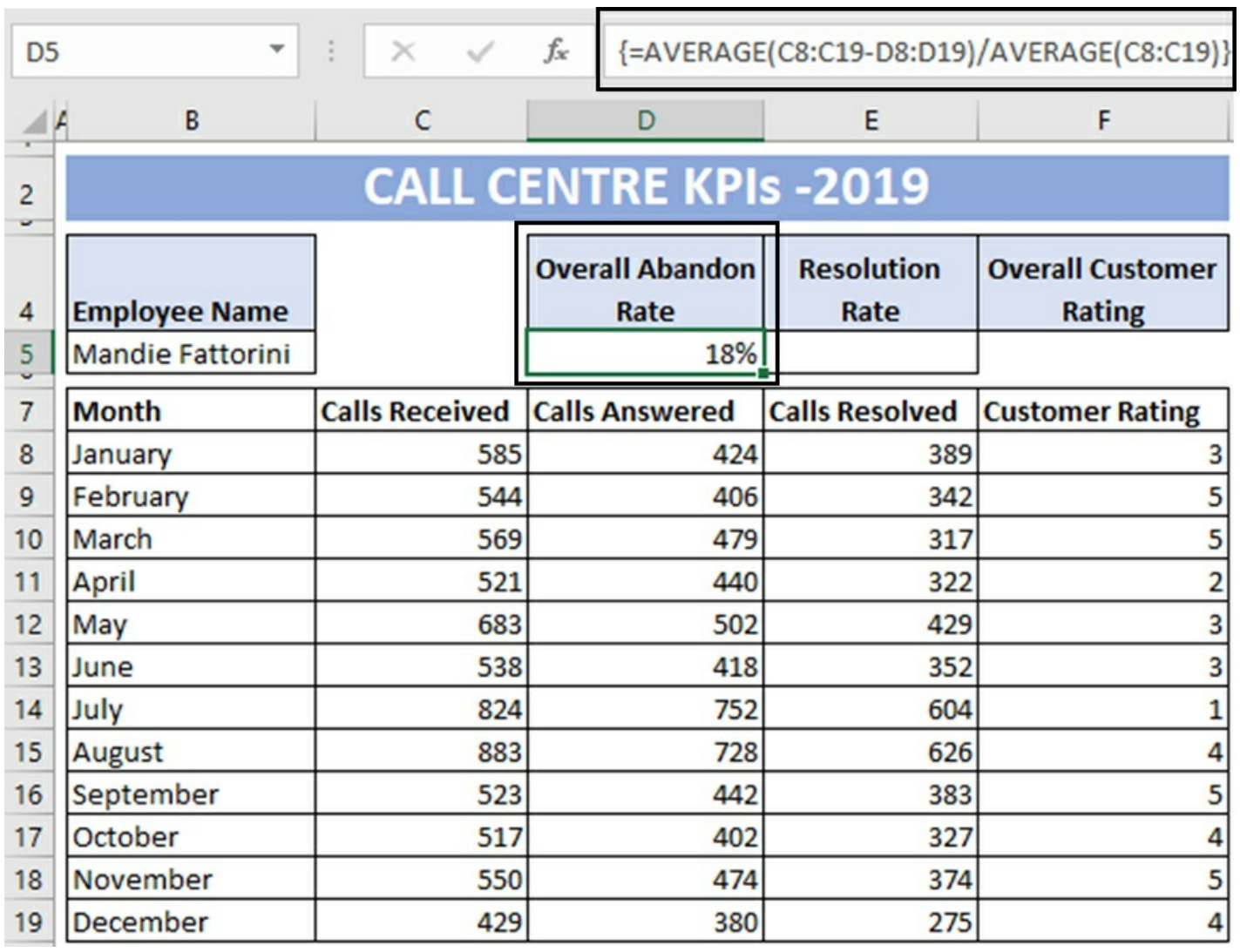


10. Select another employee from the *Employee Name* dropdown. The Template should respond dynamically to the change and calls received, calls answered, calls resolved, and customer rating data should update for the selected employee.

11. Now, we have to calculate **Overall Abandon Rate** for the selected employee. It is defined as overall percentage of calls abandoned or not answered throughout the year. To calculate it, we need to first calculate the difference between calls received and calls abandoned over the months, average it, and then divide it by the average of calls received over the months. Instead of performing these calculations one by one, we will employ an array formula to compute the overall abandon rate in one go. Type below array formula in cell D5 and press Ctrl+Shift+Enter to enter it. It will introduce curly brackets around the formula. We have covered array formulas in Chapter 2.4. Note that cell D5 is formatted as a percentage.

=AVERAGE(C8:C19-D8:D19)/AVERAGE(C8:C19)

The formula calculates the average abandoned calls by averaging the difference between *Calls Received* and *Calls Answered* over the months (AVERAGE(C23:C34-D23:D34)) and then dividing it by the average of *Calls Received* (AVERAGE(C23:C34)).



12. Now, to calculate **Resolution Rate** for selected employee, type the below formula in cell E5 and press enter. Note that cell E5 is formatted as a percentage.

=AVERAGE(E8:E19)/AVERAGE(D8:D19)

Resolution Rate is defined as overall percentage of calls resolved during the year. The formula divides the average of *Calls Resolved* column by *Calls Answered* column.

SUM | =AVERAGE(E8:E19)/AVERAGE(D8:D19)

	B	C	D	E	F
2	CALL CENTRE KPIs -2019				
4	Employee Name		Overall Abandon Rate	Resolution Rate	Overall Customer Rating
5	Mandie Fattorini		18%	GE(D8:D19)	
7	Month	Calls Received	Calls Answered	Calls Resolved	Customer Rating
8	January	585	424	389	3
9	February	544	406	342	5
10	March	569	479	317	5
11	April	521	440	322	2
12	May	683	502	429	3
13	June	538	418	352	3
14	July	824	752	604	1
15	August	883	728	626	4
16	September	523	442	383	5
17	October	517	402	327	4
18	November	550	474	374	5
19	December	429	380	275	4

13. To calculate Overall Customer Rating for selected employee, type below formula in cell F5 and press enter. Note that cell F5 is formatted as a percentage.

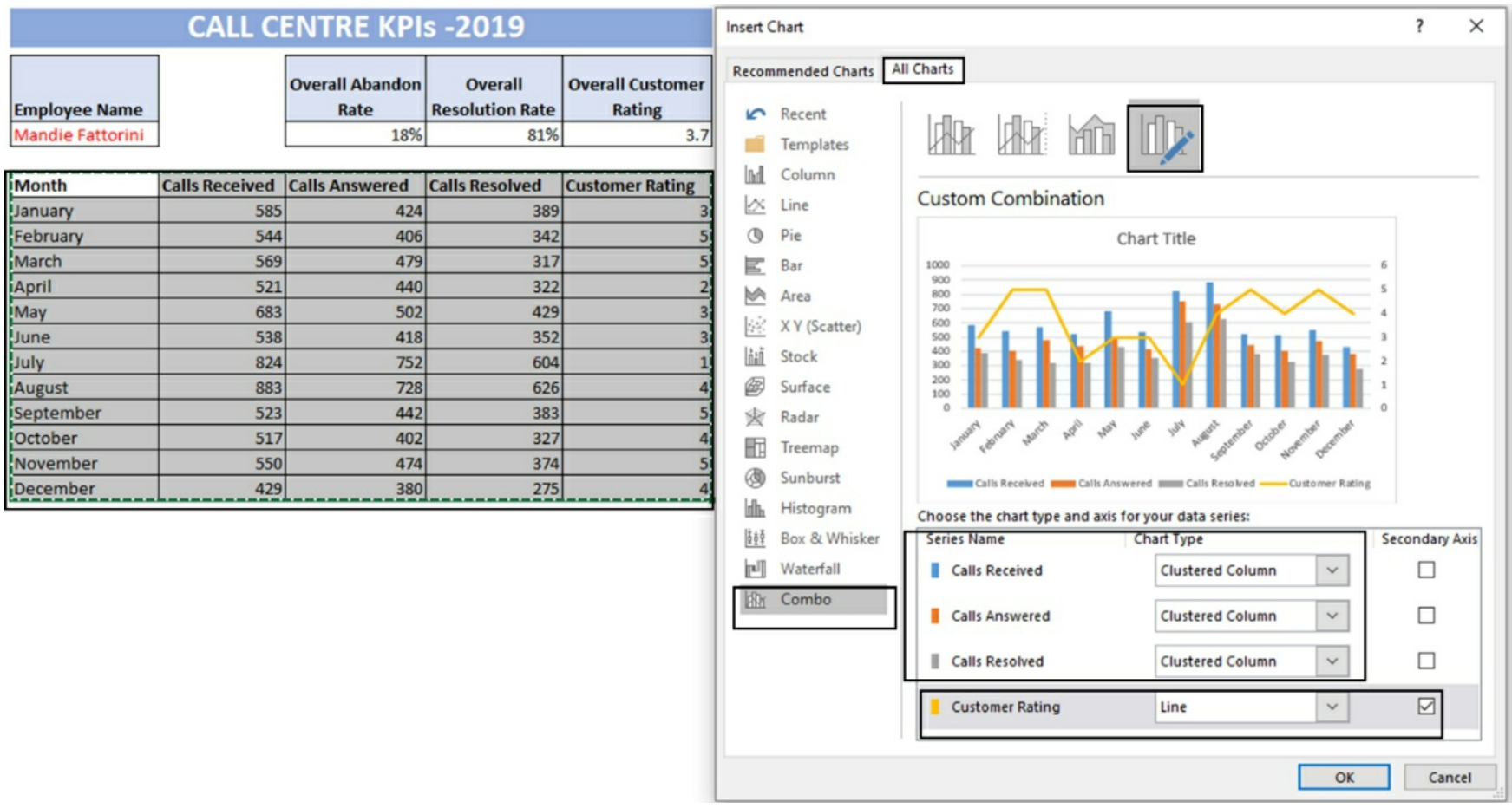
=AVERAGE(F8:F19)

The formula averages the *Customer Rating* column.

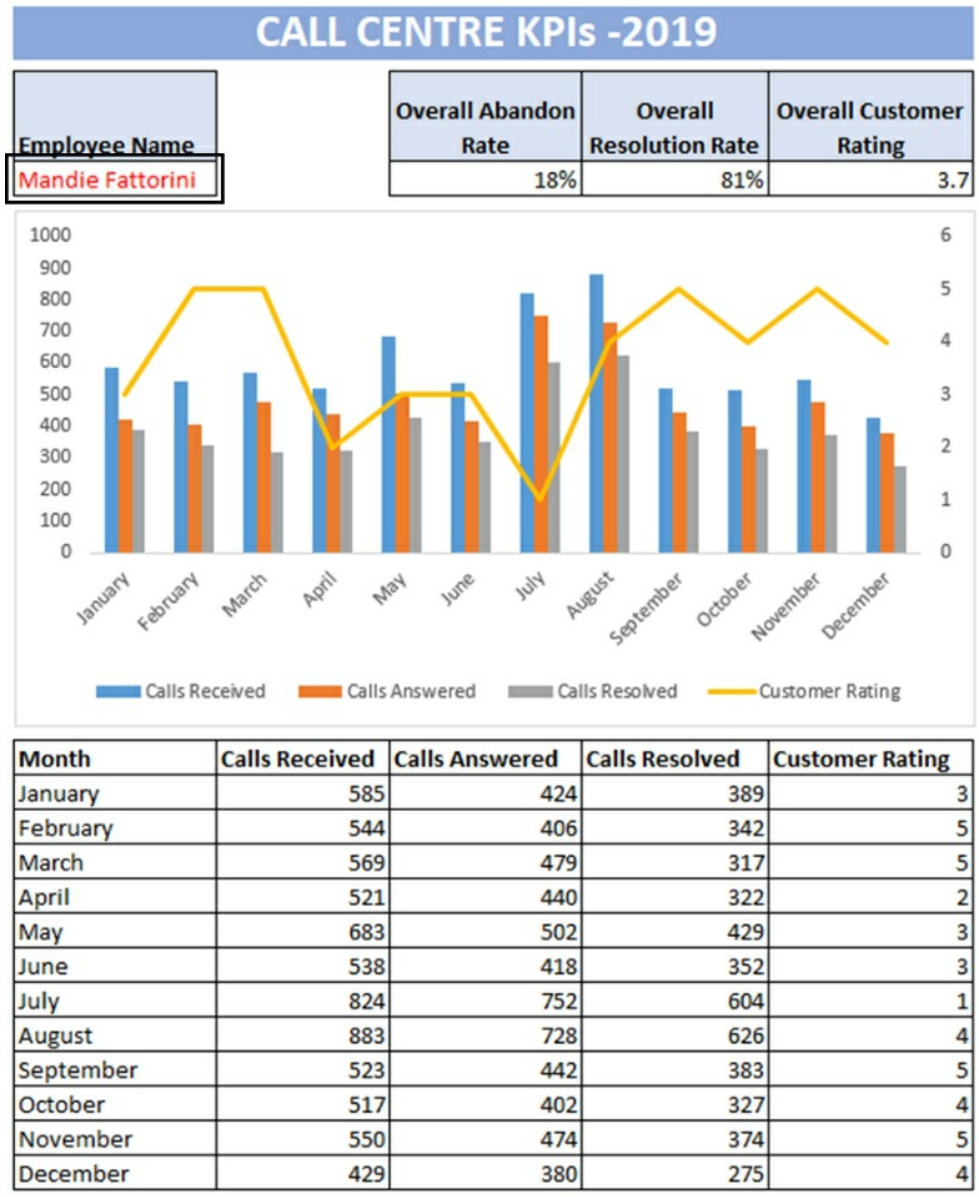
SUM | =AVERAGE(F8:F19)

	B	C	D	E	F
2	CALL CENTRE KPIs -2019				
4	Employee Name		Overall Abandon Rate	Resolution Rate	Overall Customer Rating
5	Mandie Fattorini		18%	81%	=AVERAGE(F8:F19)
7	Month	Calls Received	Calls Answered	Calls Resolved	Customer Rating
8	January	585	424	389	3
9	February	544	406	342	5
10	March	569	479	317	5
11	April	521	440	322	2
12	May	683	502	429	3
13	June	538	418	352	3
14	July	824	752	604	1
15	August	883	728	626	4
16	September	523	442	383	5
17	October	517	402	327	4
18	November	550	474	374	5
19	December	429	380	275	4

14. What's a dashboard without a graph! Select range B7:F19 and insert a combo chart with *Calls Received*, *Calls Answered* and *Calls Received* as clustered columns and *Customer Rating* as a line on the secondary axis. Refer to Chapter 3.3 for detailed instructions, as needed.



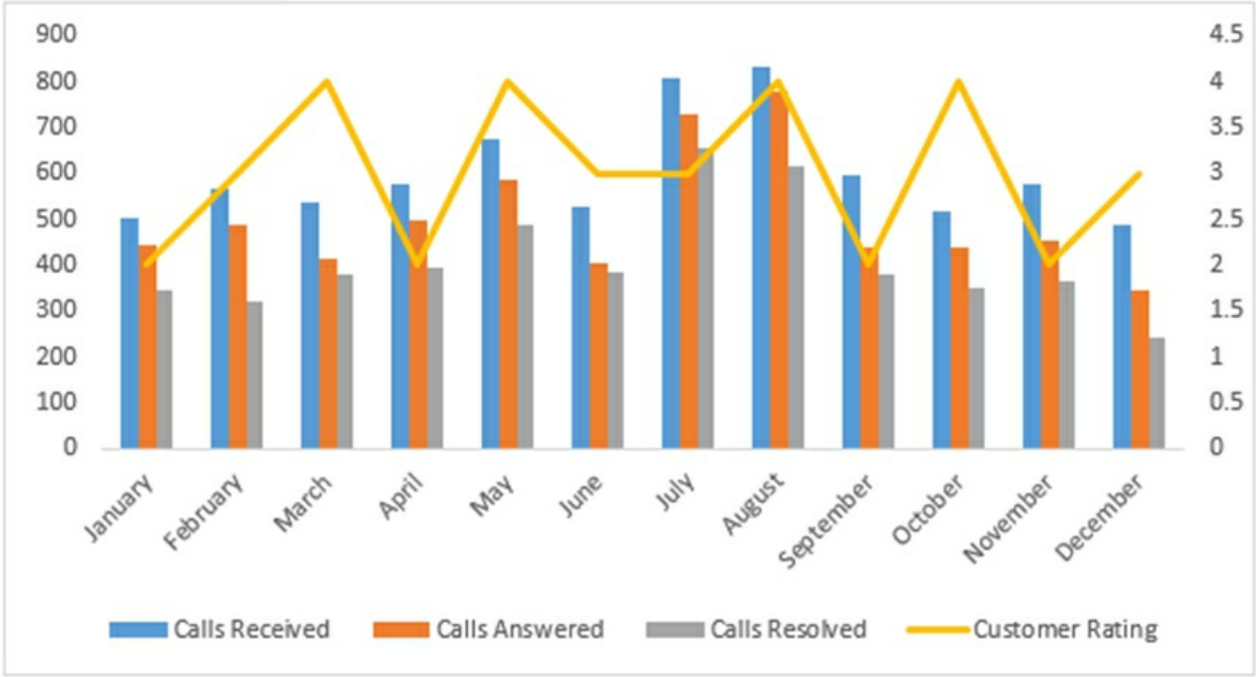
15. Delete the title of the chart. If you prefer, embed the chart between the overall KPI list and the data table. To do that, cut and paste the table with data (range B7:F19) a few rows further down and move the chart towards the middle.



16. Select another employee to ensure the formulae and the chart update. Viola! You are ready to impress the management team with your simple, yet impressive dashboard!

CALL CENTRE KPIs -2019

Employee Name	Overall Abandon Rate	Overall Resolution Rate	Overall Customer Rating
Hildegarde Grindle	16%	82%	3.0



Month	Calls Received	Calls Answered	Calls Resolved	Customer Rating
January	501	444	344	2
February	567	490	323	3
March	536	412	382	4
April	577	497	397	2
May	674	588	487	4
June	530	402	385	3
July	810	730	657	3
August	834	779	614	4
September	598	440	380	2
October	517	438	350	4
November	577	456	364	2
December	486	347	240	3

4.3 Slicers, Timeline and PivotCharts

In this chapter, we'll discuss how to use PivotTables and PivotCharts to create a simple, but dynamic dashboard. Then we'll add slicers and a timeline to give users the ability to quickly filter the data they want to look at. Slicers and timelines enable PivotTable and PivotCharts to automatically expand and contract displaying only the information that users want to see.

As we discussed in Chapter 3.1, a **PivotTable** can be used to summarize, analyze, explore, and present raw, row-level data. **PivotCharts** complement PivotTables by providing a visualization of the data in a PivotTable. **Slicers** are visual filters that filter the PivotTables and the PivotCharts. They contain a set of buttons representing categories in the data. Clicking on a button filters the data for that category. The slicers also highlight the filter being used. **Timelines** are like slicers, but they are specifically used with date fields. They allow you to easily filter on ranges of dates by days, months, quarters or years. The dates appear in a horizontal line going from oldest to newest as you go from left to right on the timeline.

You will create these dashboard elements one by one and assemble them in form of an interactive dashboard.

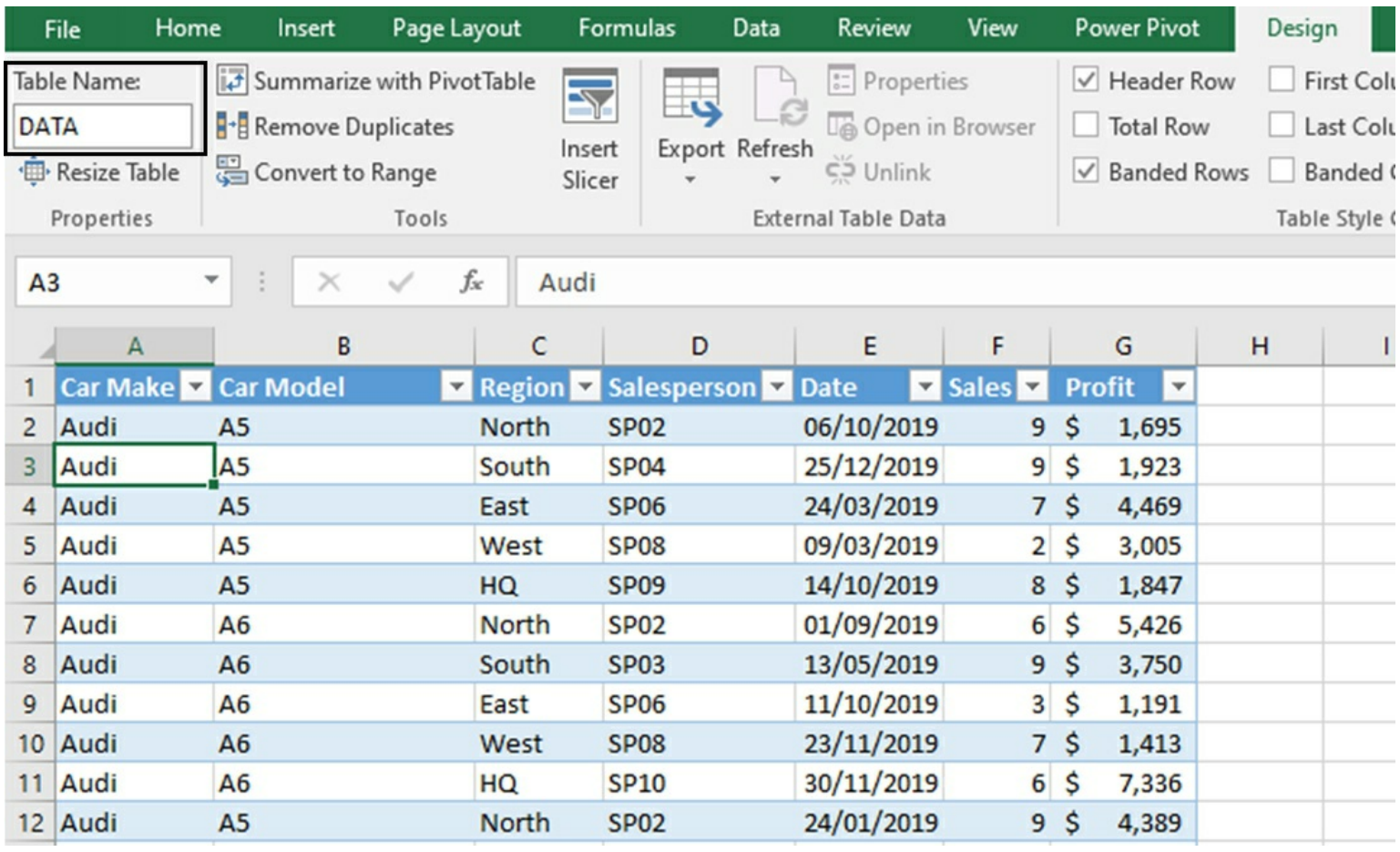
Practice

1. Open *Slicers, Timeline and PivotCharts_Practice.xlsx* and go to the *Raw Data* sheet. The sheet contains data containing individual records of number of sales made and profit generated by several salespersons operating in different regions and selling different models of Audi and BMW. Does this sound familiar? The dataset is an expanded version of the data we used to create PivotTables in Chapter 3.1. A screenshot of the data is below.

	A	B	C	D	E	F	G
1	Car Make	Car Model	Region	Salesperson	Date	Sales	Profit
2	Audi	A5	North	SP02	06/10/2019	9	$ 1,695
3	Audi	A5	South	SP04	25/12/2019	9	$ 1,923
4	Audi	A5	East	SP06	24/03/2019	7	$ 4,469
5	Audi	A5	West	SP08	09/03/2019	2	$ 3,005
6	Audi	A5	HQ	SP09	14/10/2019	8	$ 1,847
7	Audi	A6	North	SP02	01/09/2019	6	$ 5,426
8	Audi	A6	South	SP03	13/05/2019	9	$ 3,750
9	Audi	A6	East	SP06	11/10/2019	3	$ 1,191
10	Audi	A6	West	SP08	23/11/2019	7	$ 1,413
11	Audi	A6	HQ	SP10	30/11/2019	6	$ 7,336
12	Audi	A5	North	SP02	24/01/2019	9	$ 4,389
13	Audi	A5	South	SP04	13/09/2019	9	$ 6,509
14	Audi	A5	East	SP06	27/12/2019	3	$ 3,568
15	Audi	A5	West	SP07	28/04/2019	1	$ 2,463
16	Audi	A5	HQ	SP10	13/05/2019	8	$ 5,725
17	Audi	A6	North	SP02	27/05/2019	8	$ 5,323
18	Audi	A6	South	SP03	07/08/2019	1	$ 4,007
19	Audi	A6	East	SP06	28/10/2019	9	$ 9,380
20	Audi	A6	West	SP07	03/01/2019	3	$ 4,475
21	Audi	A6	HQ	SP09	28/11/2019	1	$ 5,948
1992	BMW	4 Series Gran Coupe	North	SP01	28/08/2019	6	$ 9,768
1993	BMW	4 Series Gran Coupe	South	SP03	26/06/2019	1	$ 6,071
1994	BMW	4 Series Gran Coupe	East	SP05	20/09/2019	5	$ 3,861
1995	BMW	4 Series Gran Coupe	West	SP07	14/05/2019	3	$ 2,466
1996	BMW	4 Series Gran Coupe	HQ	SP10	04/01/2019	10	$ 2,262
1997	BMW	5 Series	North	SP02	29/08/2019	2	$ 6,258
1998	BMW	5 Series	South	SP04	08/10/2019	10	$ 5,258
1999	BMW	5 Series	East	SP06	27/10/2019	5	$ 8,831
2000	BMW	5 Series	West	SP08	02/08/2019	10	$ 3,066
2001	BMW	5 Series	HQ	SP09	01/12/2019	9	$ 6,233

Create a named table

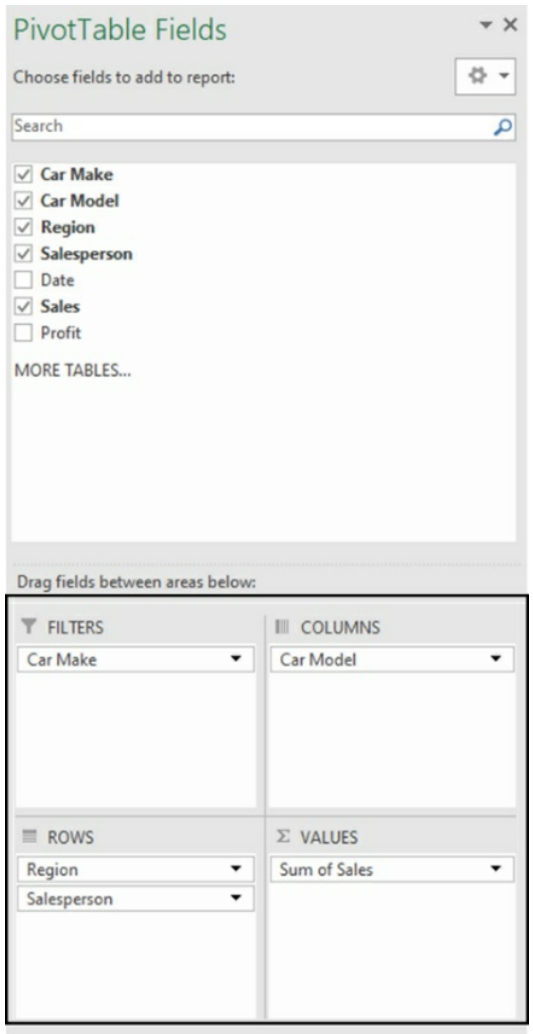
2. Now covert the data into a named table. We had thoroughly discussed the steps to do so in Chapter 1.3. Refer to the chapter if needed.
3. Name the table DATA.



	A	B	C	D	E	F	G
1	Car Make	Car Model	Region	Salesperson	Date	Sales	Profit
2	Audi	A5	North	SP02	06/10/2019	9	$ 1,695
3	Audi	A5	South	SP04	25/12/2019	9	$ 1,923
4	Audi	A5	East	SP06	24/03/2019	7	$ 4,469
5	Audi	A5	West	SP08	09/03/2019	2	$ 3,005
6	Audi	A5	HQ	SP09	14/10/2019	8	$ 1,847
7	Audi	A6	North	SP02	01/09/2019	6	$ 5,426
8	Audi	A6	South	SP03	13/05/2019	9	$ 3,750
9	Audi	A6	East	SP06	11/10/2019	3	$ 1,191
10	Audi	A6	West	SP08	23/11/2019	7	$ 1,413
11	Audi	A6	HQ	SP10	30/11/2019	6	$ 7,336
12	Audi	A5	North	SP02	24/01/2019	9	$ 4,389

Insert a PivotTable

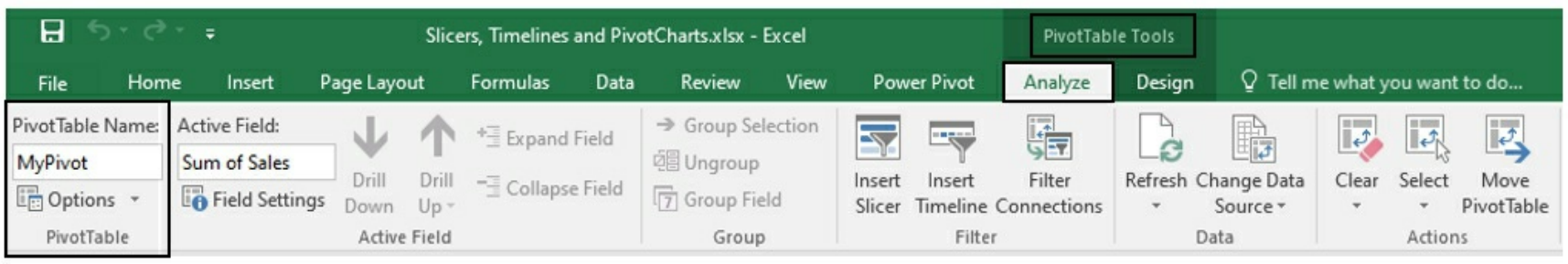
4. Now you will create a PivotTable that summarizes sales by the salesperson by *Region*, *Salesperson* and *Car Model*. We also want to filter the PivotTable summary by *Car Make*. To a blank PivotTable, you will start with selecting any cell within your DATA table, going to the *Insert* tab and then selecting *PivotTable*. Refer to Chapter 3.1 for detailed instructions, as needed.
5. Add the PivotTable fields as shown in the below screenshot to the blank PivotTable and format as desired.



6. If you want, spend some time making any necessary adjustments to style, report layout and general formatting. You have now created a summary of sales by *Region* and *Salesperson* nested as rows and *Car Model* as columns with *Car Make* as a filter.

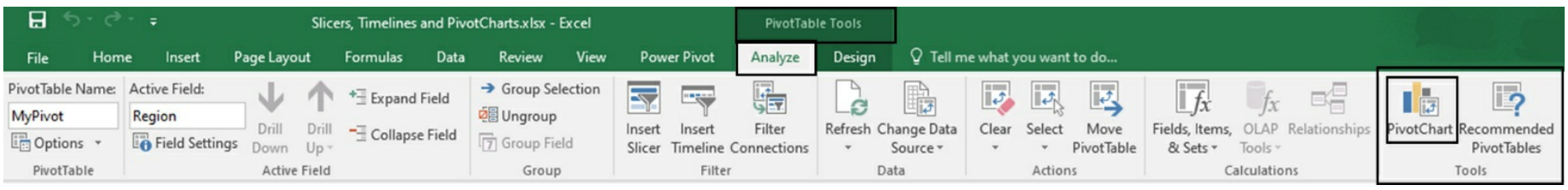
	A	B	C	D	E	F
1	Car Make	(All)				
2						
3	Sum of Sales	Column Labels				
4	Row Labels	4 Series Gran Coupe	5 Series	A5	A6	Grand Total
5	East	490	535	577	542	2144
6	SP05	266	199	291	283	1039
7	SP06	224	336	286	259	1105
8	HQ	591	531	561	530	2213
9	SP09	283	263	275	280	1101
10	SP10	308	268	286	250	1112
11	North	554	540	552	594	2240
12	SP01	315	263	260	275	1113
13	SP02	239	277	292	319	1127
14	South	526	551	558	511	2146
15	SP03	264	269	258	293	1084
16	SP04	262	282	300	218	1062
17	West	538	511	491	569	2109
18	SP07	301	231	256	265	1053
19	SP08	237	280	235	304	1056
20	Grand Total	2699	2668	2739	2746	10852

7. Excel will allocate the name *PivotTable1* to the PivotTable by default. To rename it to *MyPivot*, click anywhere on the PivotTable to enable *PivotTable Tools* tab. On the tab, go to *Analyze* and under the *PivotTable* group, enter the new name *MyPivot* in the *PivotTable Name* box. This is good practice for when you have multiple PivotTables and PivotCharts that you will eventually want to connect using slicers and a timeline.

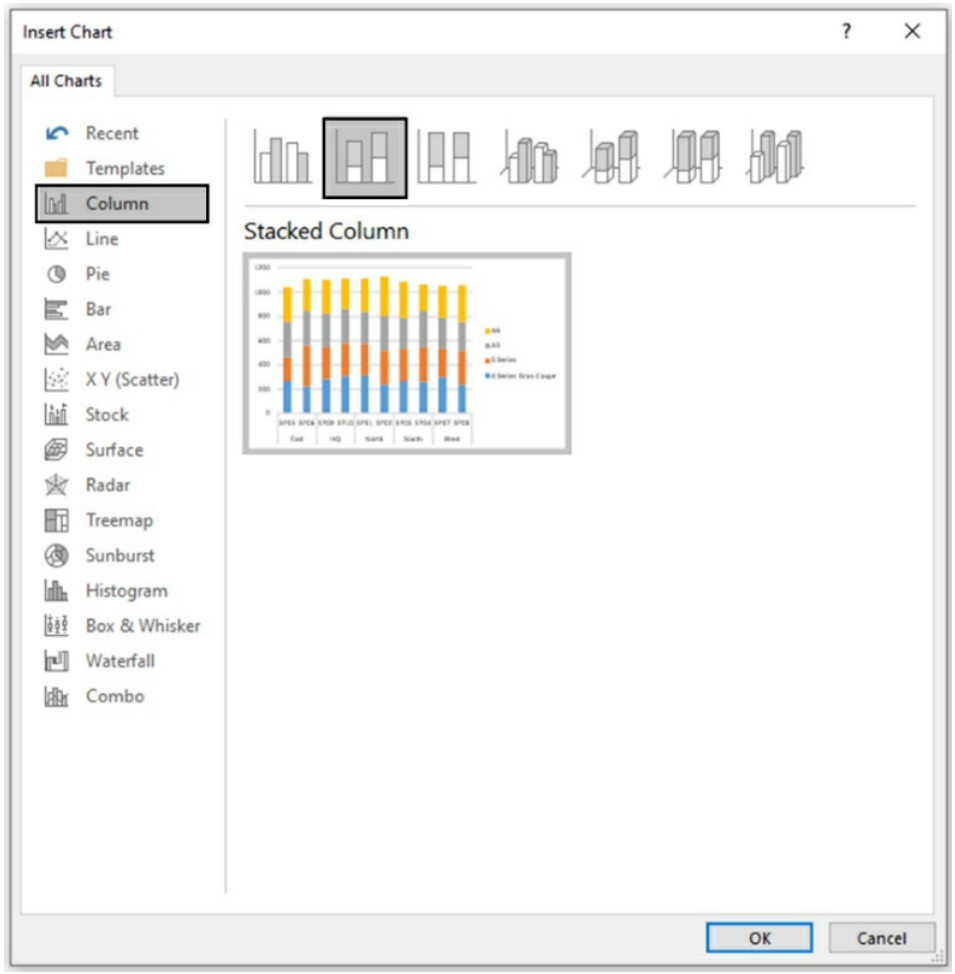


Insert a PivotChart

8. Click anywhere in your PivotTable to enable *PivotTable Tools* tab. On the tab, go to *Analyze* and under the *Tools* group, select *PivotChart*.

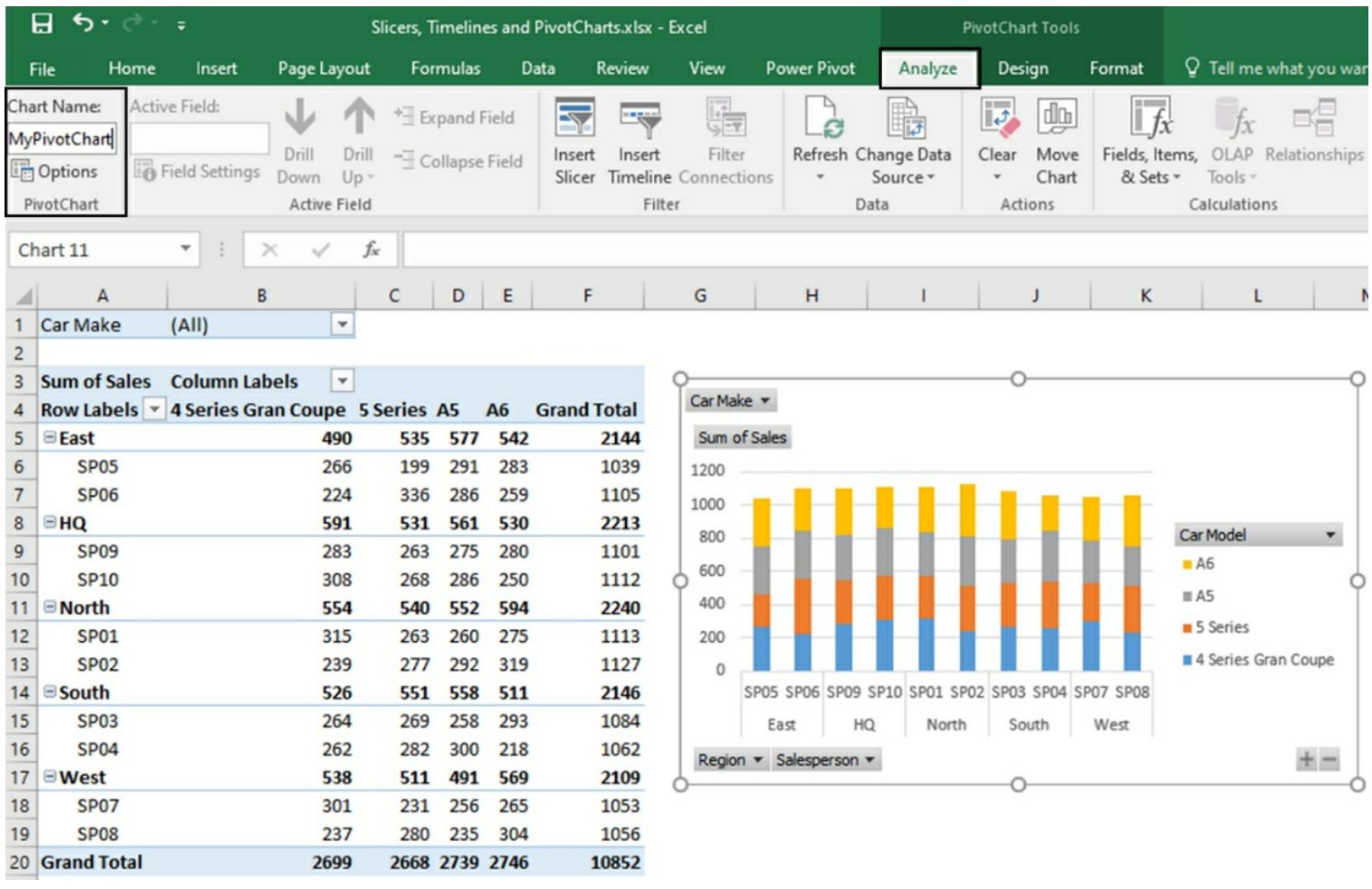


9. Go to *Column* group (if needed), choose a *Stacked Column* chart and click *OK*.



10. A stacked column chart with sales as a stacked column will appear.

11. Rename your PivotChart. To rename it to *MyPivotChart*, click anywhere on the PivotChart to enable *PivotTable Tools* tab. On the tab, go to *Analyze* and under the *PivotChart* group, enter the new name *MyPivotChart* in the *Chart Name* box.



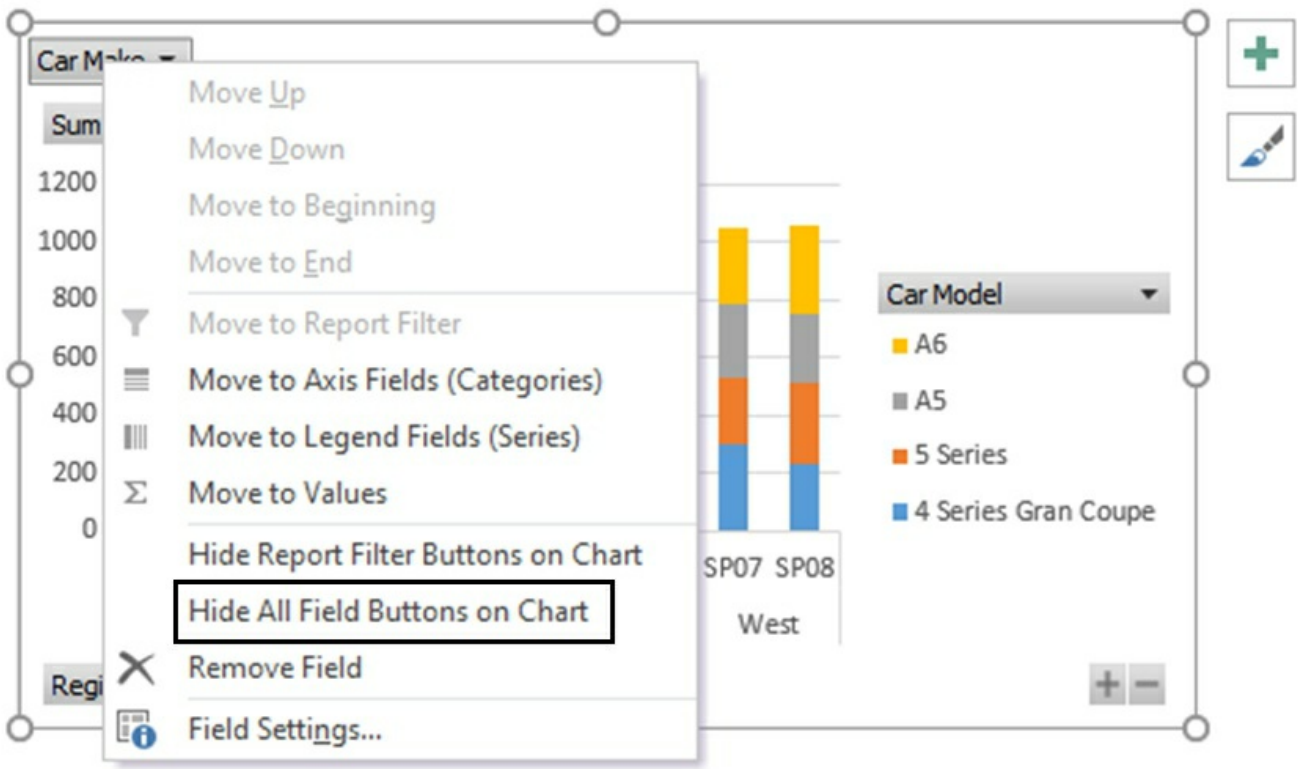
12. Select the chart and size and format as desired from the *Design* and *Format* tabs under *PivotChart Tools*.

13. Now, click on the downward arrow next to *Car Make* header placed at the top left of the PivotChart, select *BMW* and click *OK*. Both PivotTable and PivotChart will only show sales of BMW Cars. Now remove the filter from the PivotTable (not PivotChart). Notice that changes to the PivotTable are immediately reflected in the PivotChart and vice versa.

	A	B	C	D
1	Car Make	BMW		
2				
3	Sum of Sales	Column Labels		
4	Row Labels	4 Series Gran Coupe	5 Series	Grand Total
5	East	490	535	1025
6	SP05	266	199	465
7	SP06	224	336	560
8	HQ	591	531	1122
9	SP09	283	263	546
10	SP10	308	268	576
11	North	554	540	1094
12	SP01	315	263	578
13	SP02	239	277	516
14	South	526	551	1077
15	SP03	264	269	533
16	SP04	262	282	544
17	West	538	511	1049
18	SP07	301	231	532
19	SP08	237	280	517
20	Grand Total	2699	2668	5367

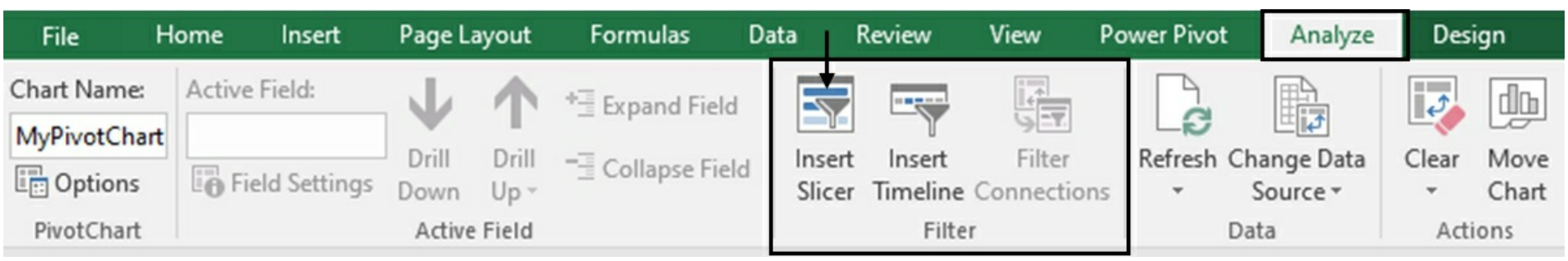
14. Filters are also available for *Region*, *Salesperson* and *Car Model*. Play around with them.

15. One drawback of standard PivotChart filters is that they don't show what fields are being used to filter the PivotChart. As we will be using slicers to filter PivotChart, we will remove them from the chart. To do so, right click on any of the filters and select *Hide All Field Buttons on Chart*.

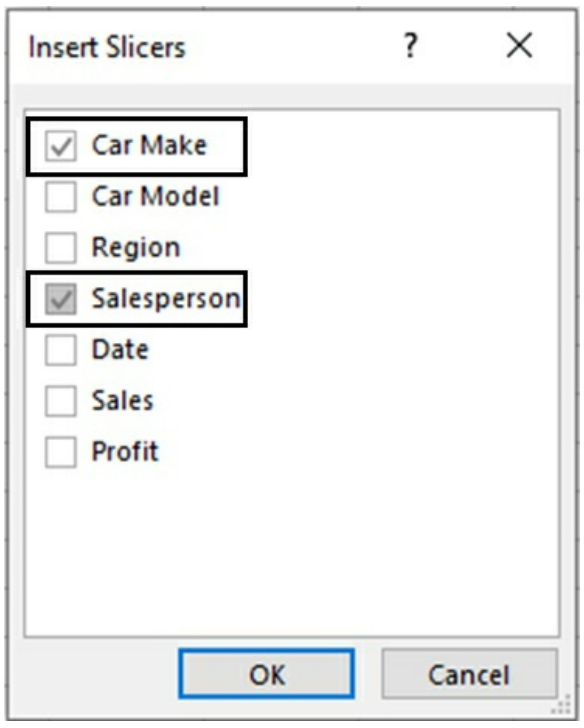


Insert Slicers and a Timeline

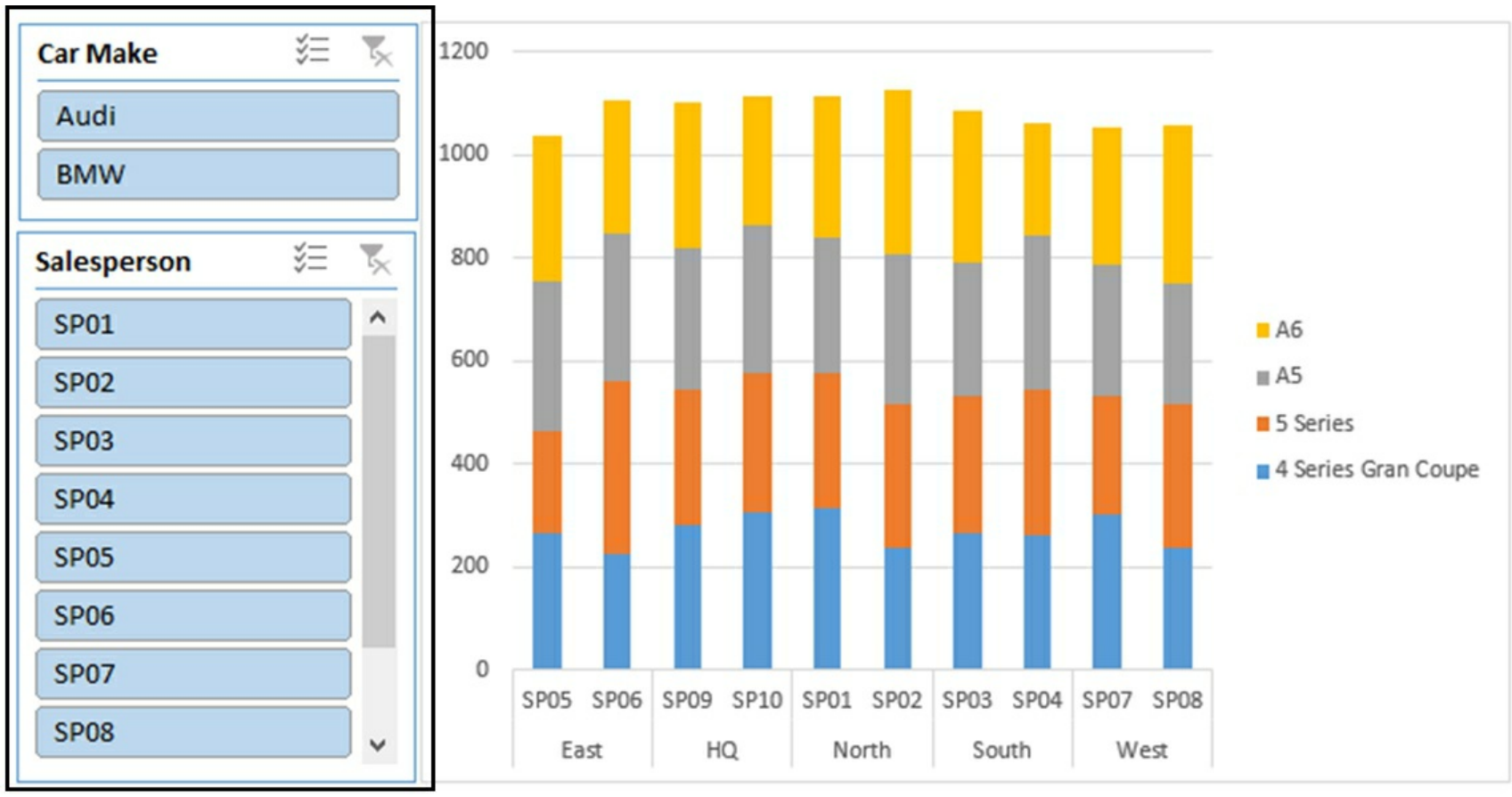
16. Now, you will insert slicers for *Car Make* and *Salesperson* and a timeline to allow you to filter your PivotTable and PivotChart. Slicers have buttons for each of the fields that stay visible and are highlighted only for the current selection. Hence, you always know what fields are being used to filter the PivotTable and PivotChart. To insert slicers, click anywhere on the PivotChart to enable *PivotChart Tools* tab. On the tab, go to *Analyze* and under the *Filter* group, click on *Insert Slicer*.



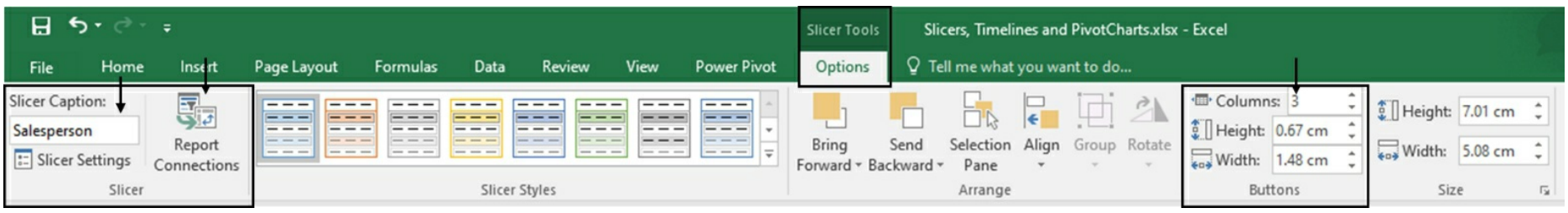
17. An *Insert Slicers* window will pop up. Select *Car Make* and *Salesperson*, and click *OK*.



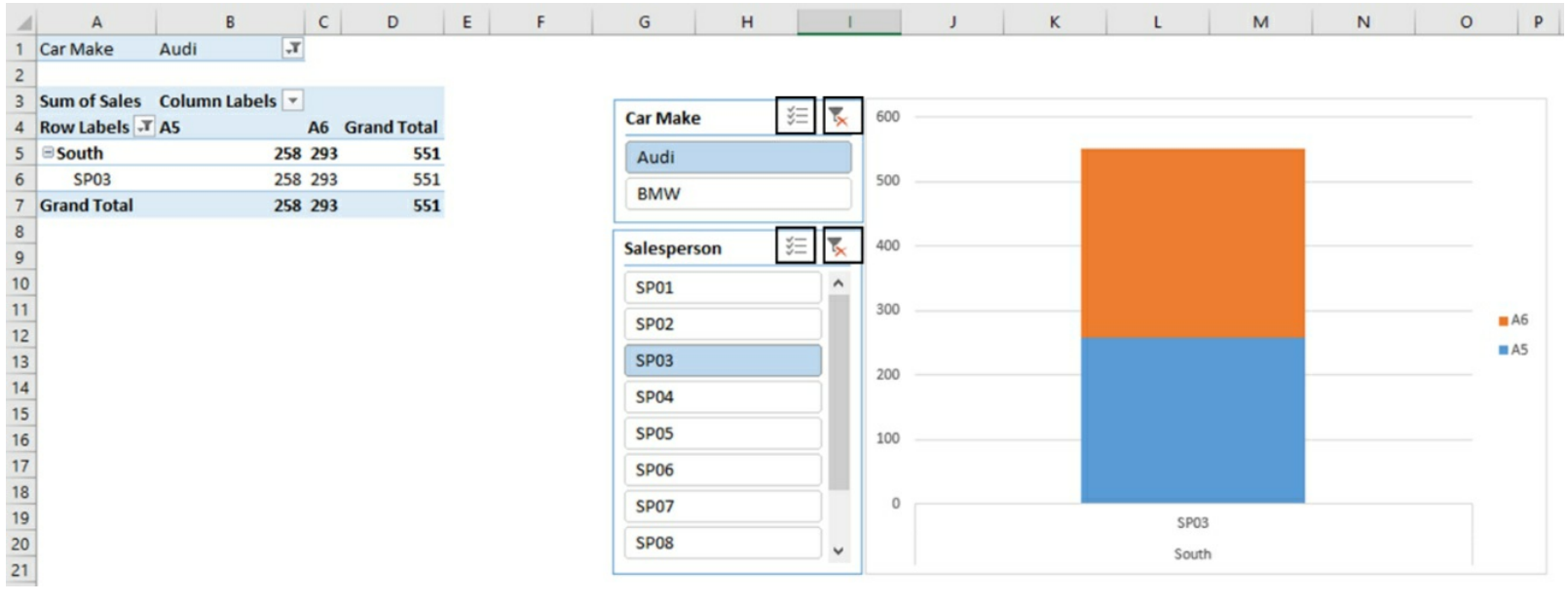
18. The slicers will be added to the middle of the screen stacked on top of each other. Arrange them on the left of the PivotChart. Move the PivotChart and resize, as necessary.



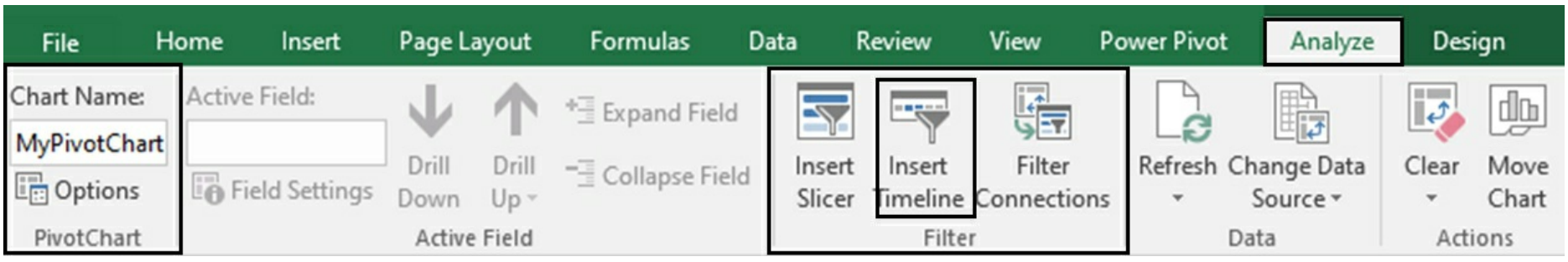
19. Click on *Salesperson* slicer to enable *Slicer Tools* on the top ribbon with *Options* tab. On the tab, under *Buttons* group, increase columns to *3*. Now fields in the slicer will display in 3 columns. Under *Slicer* group, notice slicer name in the *Slicer Caption* box. Explore *Report Connections* from the *Slicer* group and format style of the slicer as desired from *Slicer Styles* group.



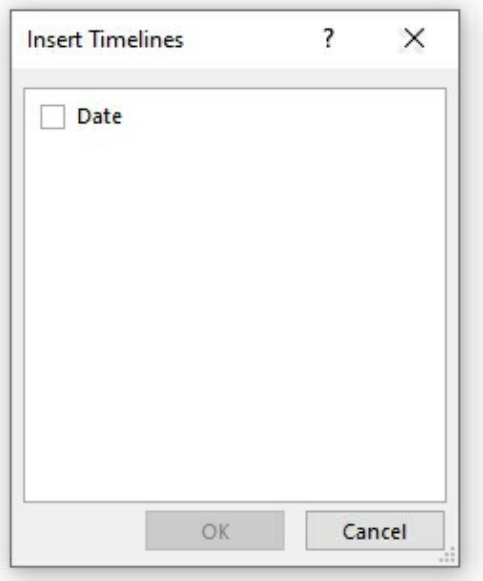
20. Select and unfilter slicer buttons to see how the PivotChart and the PivotTable behave.



21. Now to add a timeline, click anywhere on the PivotChart to enable *PivotChart Tools* tab. On the tab, go to *Analyze* and under the *Filter* group, click on *Insert Timeline*. An *Insert Timelines* window will pop up.

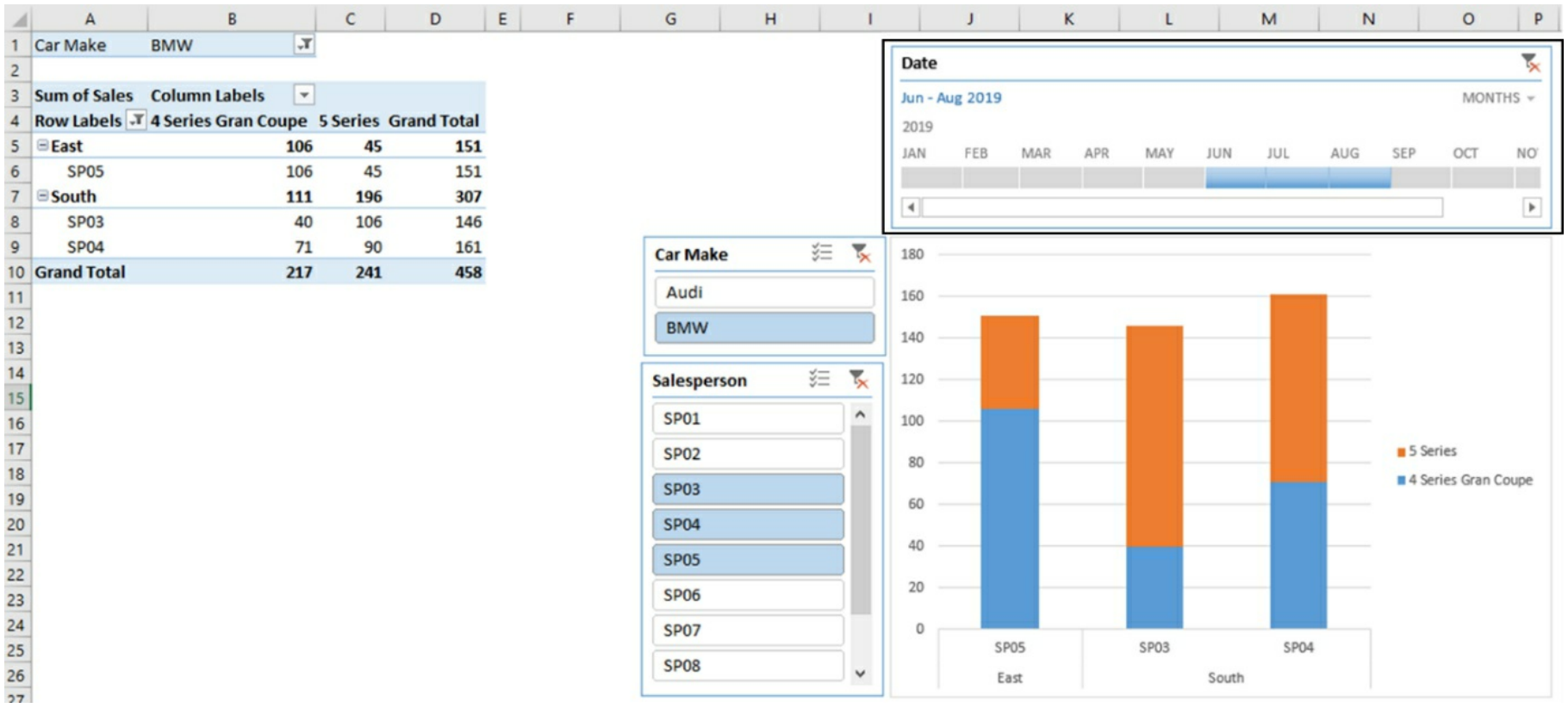


22. Only *Date* will appear as an option as it is the only column in the data containing date values. Select *Date* and click *OK*. The timeline will be added toward the middle of the screen.

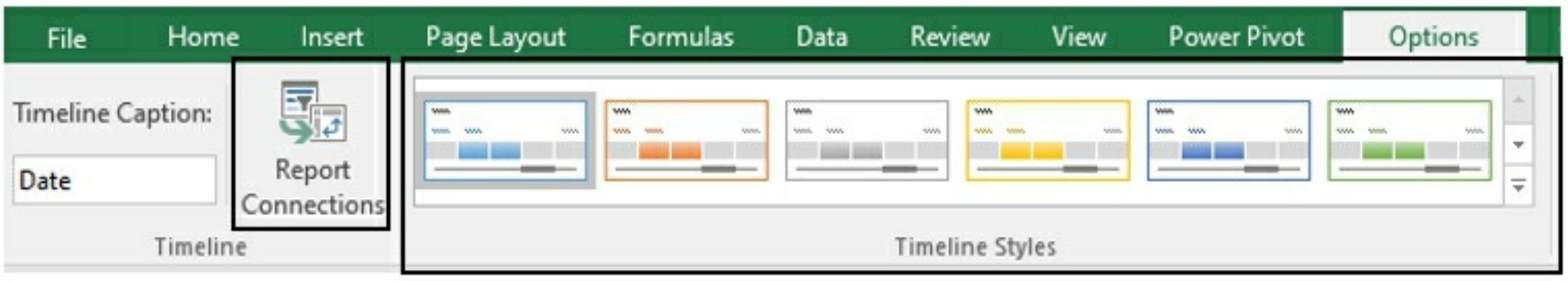


23. On the timeline, if needed change *All Periods* filter to *MONTHS*. Other options are YEARS, QUARTERS and DAYS.

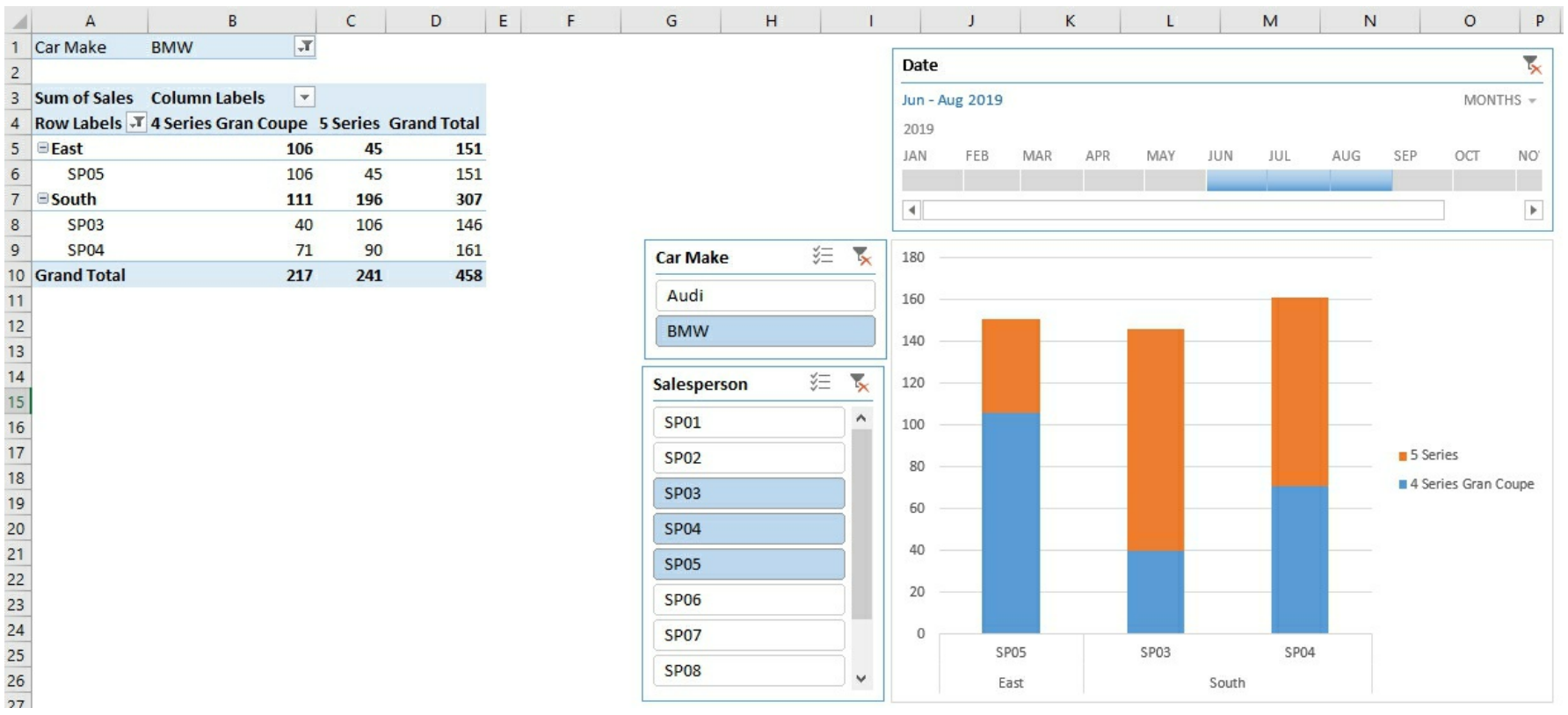
24. Arrange the timeline on top of the PivotChart. Move the PivotChart and resize, as necessary.



25. Click on the timeline to enable *Timeline Tools* on the top ribbon with *Options* tab. Explore *Report Connections* and *Timeline Styles*.



26. Choose specific months on the timeline to see how the PivotChart and the PivotTable behave. Now for the selected month, choose specific *Car Make* and *Salesperson* from the slicers to see how PivotTable and PivotChart change. When you select different buttons on your slicers and a specific period on your timeline, your PivotTable and PivotChart will change dynamically. Congratulations on creating another interactive dashboard!



4.4 Steps to create a KPI dashboard

Until now, you might have intuitively picked up some of the steps to create a KPI dashboard from the *Practice* tasks. I would still like to summarize a general process to create a KPI dashboard with a predetermined set of KPIs. I wanted to keep it until after we have discussed some of the concepts involved in the process. We will go through the process step by step:

1. **Identify and extract data**

Once you have consulted with your executive management and decided upon a set of KPIs that will be useful to them, the next stage is to identify and extract data on those KPIs. For this, start with exploring the data your organization currently houses. Extract the needed data into Excel. Based on the KPI, the data collection method might also include surveys, questionnaires, and interviews. If the data is currently not being collected, you will need to create and implement new data collection points in relevant processes of your business.

2. **Set up your Excel dashboard workbook**

Before creating dashboards in Excel, you need to import all KPI data into a single Excel workbook. You can copy and paste the data. If your existing spreadsheet is a bit messy, clean it up and ensure that things are organized into their appropriate rows and columns. Remember to structure data in form of named table(s). Organize your workbook to include:

- one sheet for your dashboard,
- one or more sheets for the raw data, and
- one or more sheets for calculations

3. **Plan your KPI dashboard**

After identifying and collecting KPI data, you'll be ready to create a KPI report in Excel and customize it to your requirements. Before building the dashboard, take some time to look at your data and figure out what you want to highlight. Do you need to display all the information? What kind of story are you trying to communicate? Make a rough sketch of the dashboard on paper and then work your way backward to think how to structure and analyze data in a manner that it easily feeds into the dashboard.

You would also want to tailor your dashboard to the audience it is meant for. Your target audience should be able to not only view, understand and process the data being displayed easily, but also interact with it to filter the data they want to see.

Once you have a preliminary vision for your dashboard in mind, referring to your raw sketch, create a raw dashboard template on your dashboard sheet. The key elements of the template will include PivotTables, PivotCharts, slicers, timelines etc.

4. **Introduce dashboard elements**

Based on your initial sketch, start introducing dashboard elements to the dashboard sheet. These include:

- Excel formulas like SUMIF, COUNT, VLOOKUP, INDEX, MATCH etc.
- PivotTables
- PivotCharts
- Non-charting objects, like slicers, timelines, key values etc.

The elements you want to add to your dashboard will help you group similar data together to produce a cohesive layout.

5. **Introduce interactivity**

You have a couple of different options to add interactivity to your dashboard:

- **Data validation list:** As we practiced in Chapter 1.6, you can use a data validation or dropdown list to create interactive charts. With this drop-down list, users can select the criteria they want to filter on, and the dashboard will automatically update to reflect data for that criteria.
- **Slicers or Timelines:** Slicers and timelines allow you to quickly filter your PivotTables and PivotCharts, so you can see just the information that's meaningful to you. We discuss these in Chapter 3.3.

6. **Finalize the dashboard**

Once your dashboard is functionally complete, arrange it the way that makes sense to you by making final adjustments. For instance, you might want to add a report title or a background. Test each of your slicers and timelines to make sure that your PivotTables and PivotCharts behave appropriately.

7. **Distribute the dashboard**

Before distributing your dashboard to a wider audience, you might want to ensure that your users have the same Excel version as you. If they use an older version, ensure that it is compatible with all formulas and dashboard elements you used in the dashboard. Hide sheets with raw data and protect your workbook to disable any editing. Provide necessary instruction to the users on how to navigate the dashboard. It is a good practice to also include an instruction sheet providing the purpose and data behind each of the dashboard elements in detail.

4.5 Create a Sales Dashboard

In this chapter, we will create an interactive sales dashboard using the same dataset we used previously in Chapter 4.3. A sales dashboard is a type of operational dashboard. The dataset contains individual records of number of sales and profit generated by ten salespersons operating in five regions selling two models each of Audi (A5 and A6) and BMW (4 Series Gran Coupe and 5 Series). The final product will look like the below screenshot. You will notice that in addition to four slicers and one timeline, there are four dashboard elements:

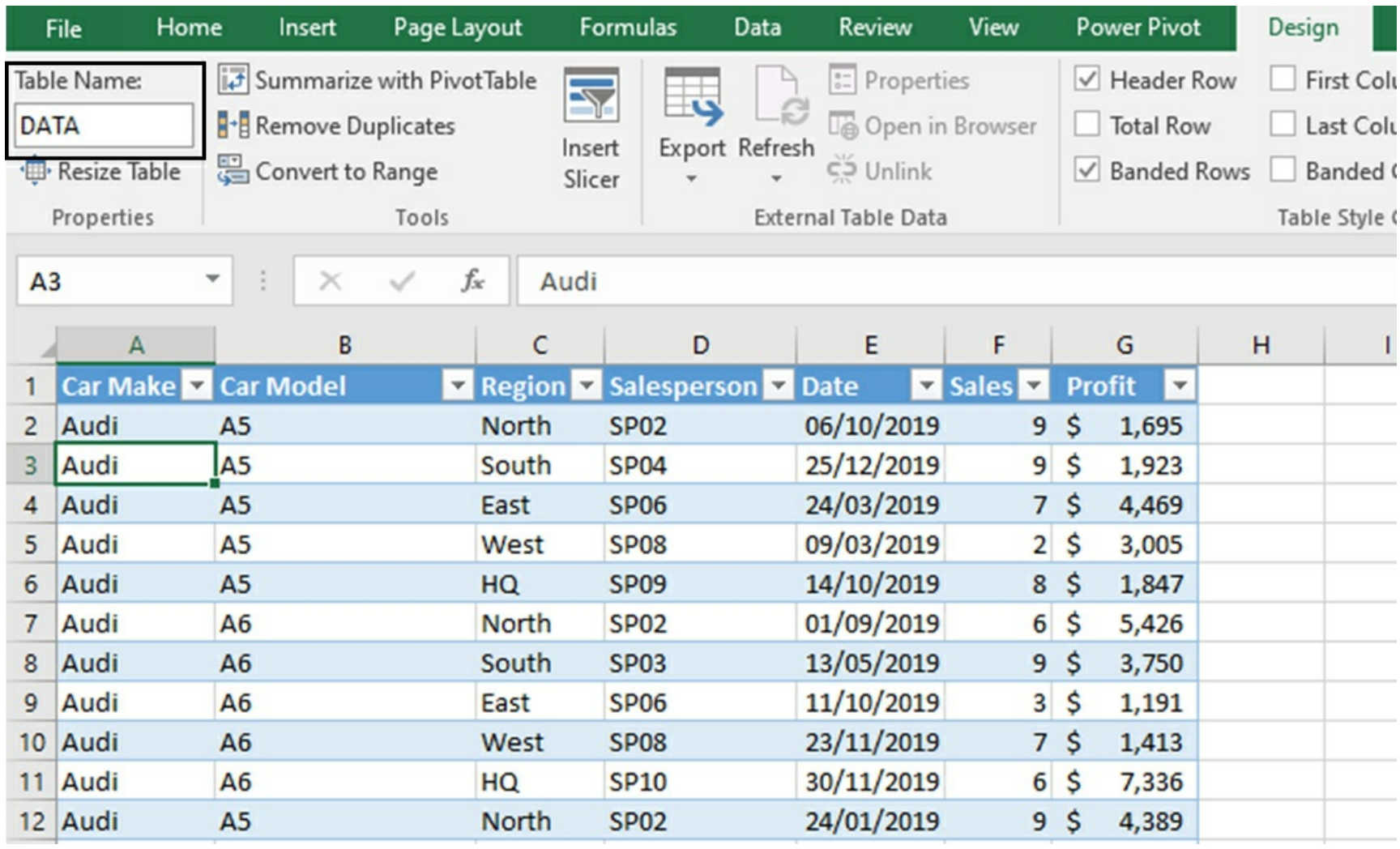
1. Two key values (2019 Sales & 2019 Profit)
2. One graph that breaks sales and profit by month
3. A sales pie chart
4. A profit pie chart

Keep the layout of the final product in the back of your mind as you follow steps to create it. In real life, you will have a rough sketch of the dashboard to refer to instead of a final product.

Practice

Create a named table and insert required sheets

1. Open *Sales Dashboard_Practice*.xlsx and go to the *DATA* sheet. Covert the data placed on the sheet into a named table called *DATA*.

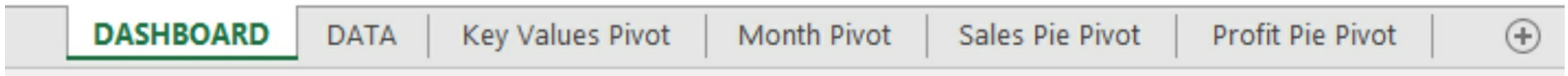


	A	B	C	D	E	F	G
1	Car Make	Car Model	Region	Salesperson	Date	Sales	Profit
2	Audi	A5	North	SP02	06/10/2019	9	$ 1,695
3	Audi	A5	South	SP04	25/12/2019	9	$ 1,923
4	Audi	A5	East	SP06	24/03/2019	7	$ 4,469
5	Audi	A5	West	SP08	09/03/2019	2	$ 3,005
6	Audi	A5	HQ	SP09	14/10/2019	8	$ 1,847
7	Audi	A6	North	SP02	01/09/2019	6	$ 5,426
8	Audi	A6	South	SP03	13/05/2019	9	$ 3,750
9	Audi	A6	East	SP06	11/10/2019	3	$ 1,191
10	Audi	A6	West	SP08	23/11/2019	7	$ 1,413
11	Audi	A6	HQ	SP10	30/11/2019	6	$ 7,336
12	Audi	A5	North	SP02	24/01/2019	9	$ 4,389

2. We will be deriving the dashboard elements (except slicers and timelines) as below:

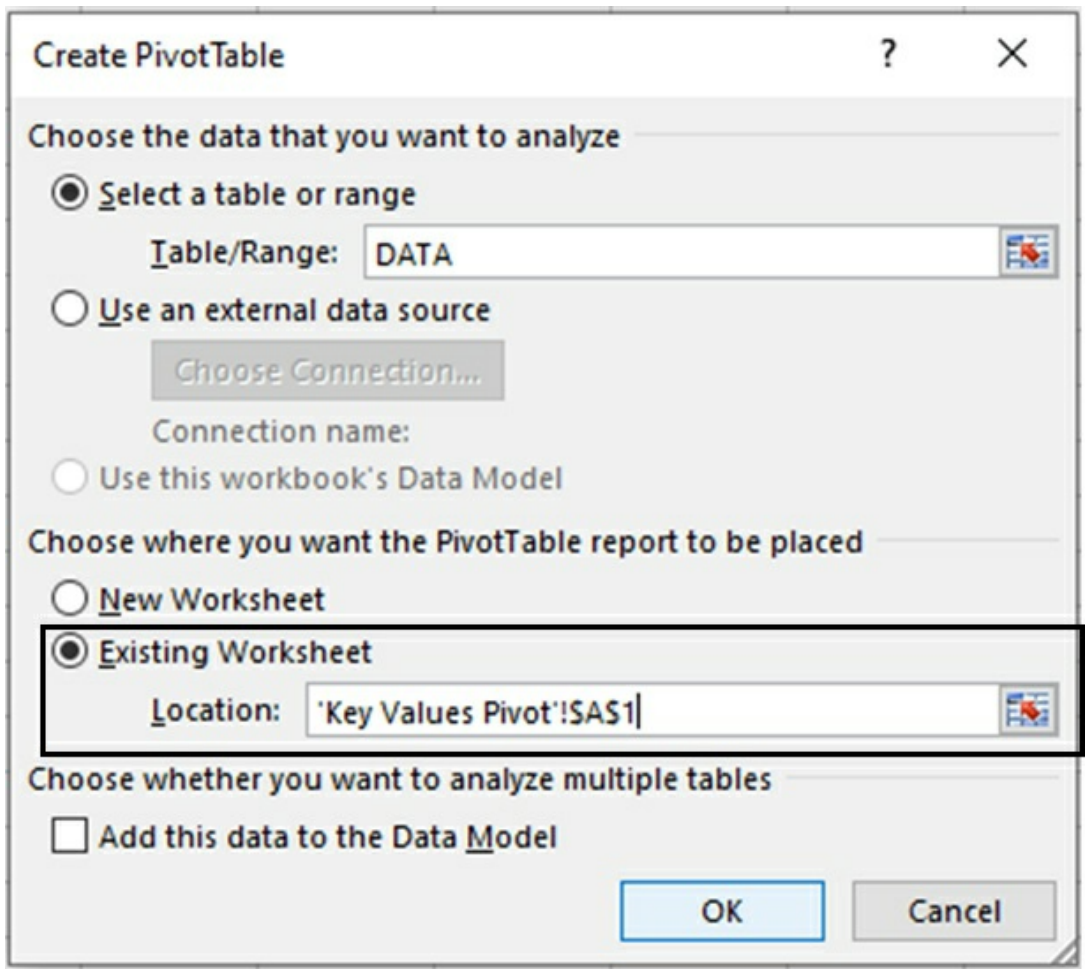
    1. **Key Values:** Both key values will be derived from a PivotTable named *KeyValuePivot* to be placed on a sheet named *Key Values Pivot* with help of a GETPIVOTDATA function.
    2. **Sales and Profit by Month Line Graph:** The sales and profit by month graph is actually a PivotChart named *MonthLineGraph* linked to a PivotTable to be placed on a sheet named *Month Pivot*.
    3. **Sales Pie Chart:** The sales pie chart is a PivotChart named *SalesPieChart* linked to a PivotTable to be placed on a sheet named *Sales Pie Pivot*.
    4. **Profit Pie Chart:** The sales pie chart is a PivotChart *ProfitPieChart* linked to a PivotTable to be placed on a sheet named *Profit Pie Pivot*.

The dashboard will be placed on a sheet named DASHBOARD. Hence, the next step is to insert the required sheets – *DASHBOARD*, *Key Values Pivot*, *Month Pivot*, *Sales Pie Pivot* and *Profit Pie Pivot*.

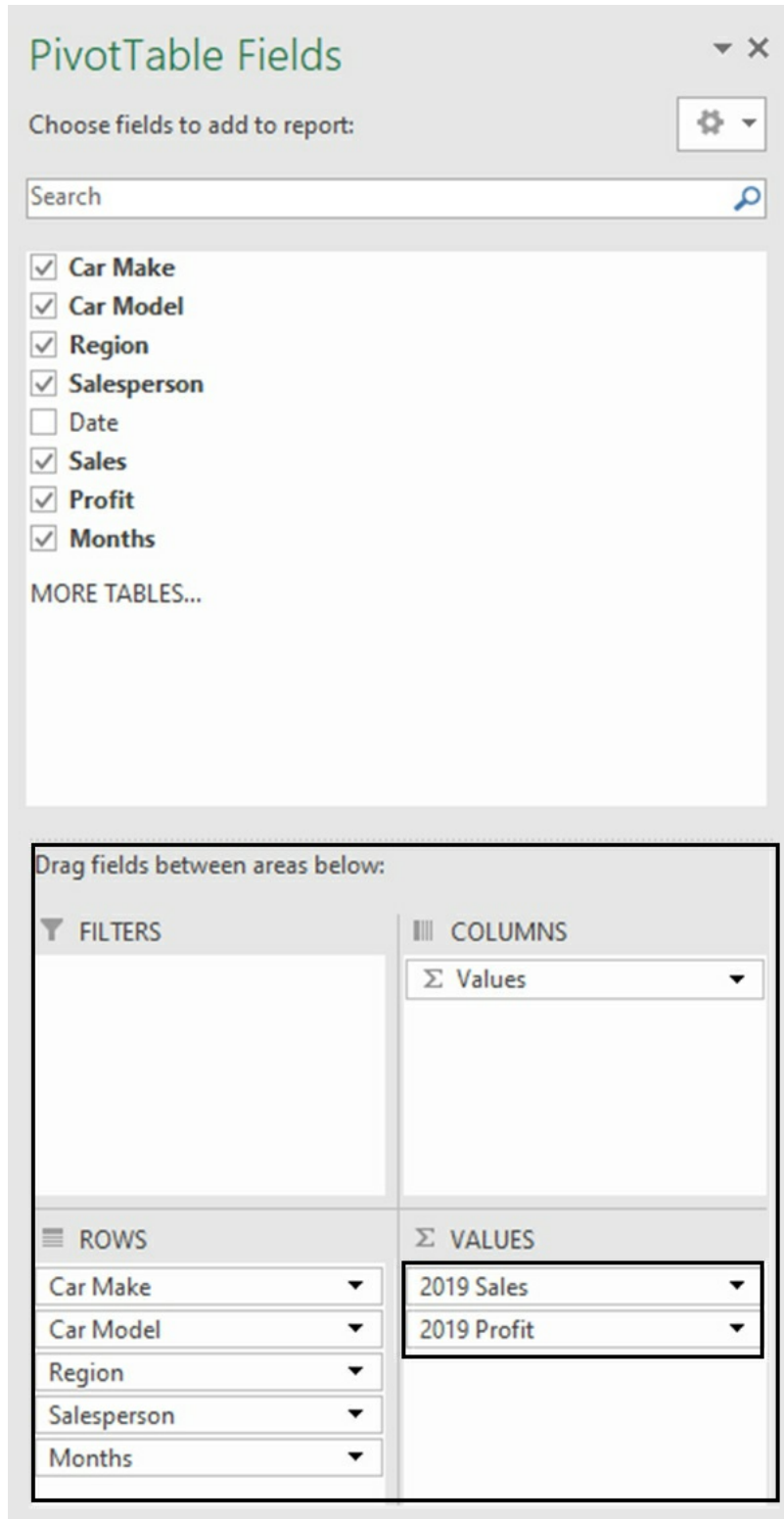


Insert Key Values PivotTable

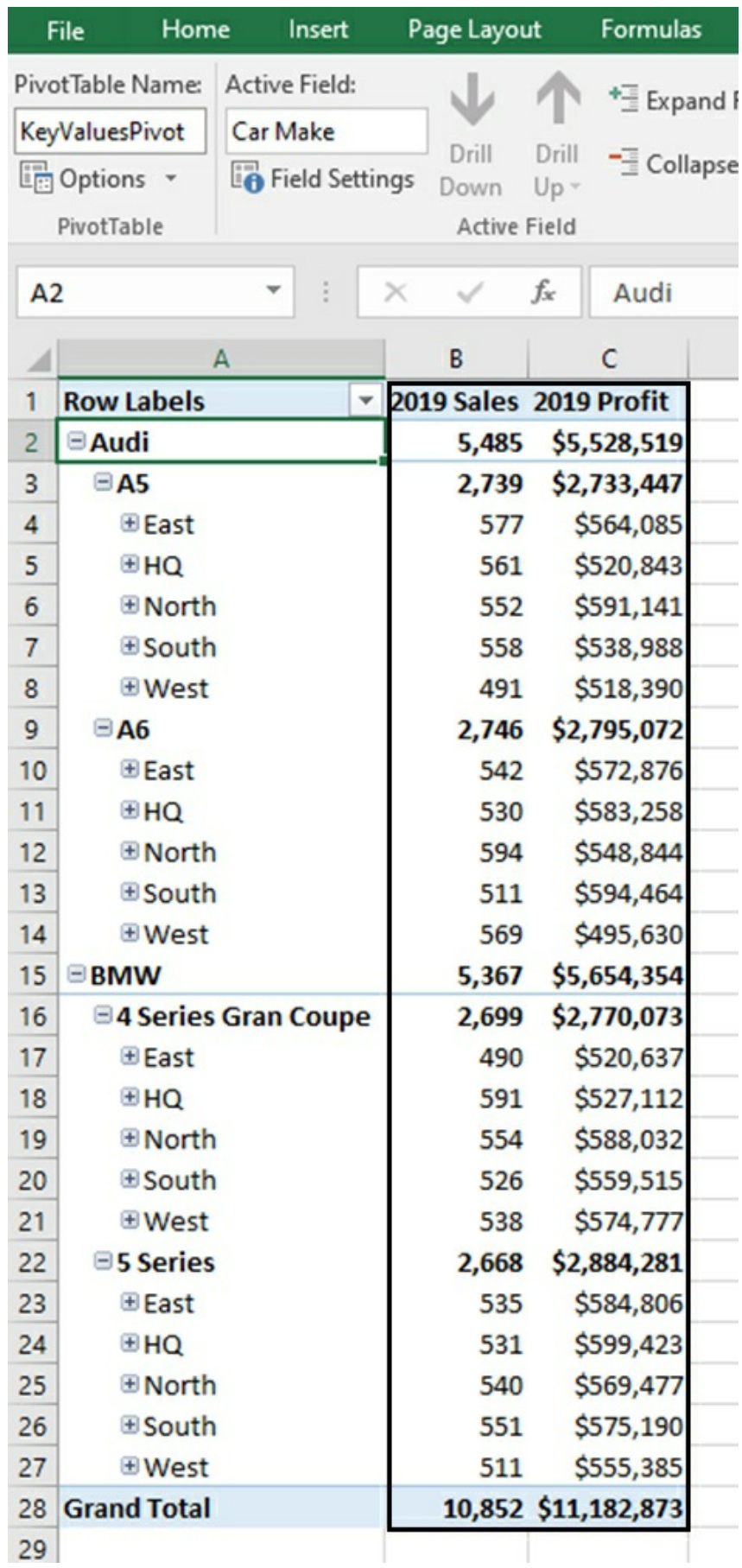
3. We will start will inserting key values into the *DASHBOARD* sheet. They are to be derived from a PivotTable named *KeyValuePivot* with help of a GETPIVOTDATA function. Let's start with creating the PivotTable. Note the final dashboard has slicers for *Car Make*, *Car Model*, *Region* and *Salesperson*. Hence, these four columns should be included in the PivotTable.
4. To insert a blank PivotTable, select any cell within your *DATA* table, go to the *Insert* tab and then select *PivotTable*. On the *Create PivotTable* window, choose *Existing Worksheet* under *Choose where you want the PivotTable report to be placed*. Then to designate *Key Values Pivot* sheet for *Location*, select the *Key Value Pivot* sheet while keeping the *Create PivotTable* window open, select cell A1 on the sheet and click *OK*.



5. Add the PivotTable fields as shown in the below screenshot to the blank PivotTable and format as desired. For *Σ Values*, note to rename *Sum of Sales* and *Sum of Profit* fields as *2019 Sales* and *2019 Profit* respectively to reciprocate with the title of our key values. If the PivotTable fields pane is not visible, activate it by clicking on the *Field List* from PivotTable tools. Refer to Chapter 3.1, as needed. To display Months under *Σ Rows*, start with dragging the *Date* field under *Σ Rows*. Excel will automatically group the *Date* into *Quarters*, *Months* and *Years*. Drag all other fields except *Months* out of the PivotTable fields pane. If the groups don't show up automatically, drag the *Date* field under *Σ Rows*, right click any of the dates in the PivotTable and click *Group*. From the *Grouping* window that pops up, select the grouping of your choice (*Months* in this case) and click *OK*.



6. Format *2019 Sales* as Number without decimals with comma separator enabled, and *2019 Profit* as Currency without decimals. Also, rename the PivotTable to *KeyValuesPivot*. Refer to Chapter 3.1, as needed.



	A	B	C
1	Row Labels	2019 Sales	2019 Profit
2	Audi	5,485	$5,528,519
3	A5	2,739	$2,733,447
4	East	577	$564,085
5	HQ	561	$520,843
6	North	552	$591,141
7	South	558	$538,988
8	West	491	$518,390
9	A6	2,746	$2,795,072
10	East	542	$572,876
11	HQ	530	$583,258
12	North	594	$548,844
13	South	511	$594,464
14	West	569	$495,630
15	BMW	5,367	$5,654,354
16	4 Series Gran Coupe	2,699	$2,770,073
17	East	490	$520,637
18	HQ	591	$527,112
19	North	554	$588,032
20	South	526	$559,515
21	West	538	$574,777
22	5 Series	2,668	$2,884,281
23	East	535	$584,806
24	HQ	531	$599,423
25	North	540	$569,477
26	South	551	$575,190
27	West	511	$555,385
28	Grand Total	10,852	$11,182,873
29			

7. Now type 2019 Sales and 2019 Profit in cells E2 and F2 of the *Key Values Pivot* sheet. They will act as headers for our key values. Make the headers bold. In order to calculate these key values from *KeyValuesPivot*, we will

    - type formula =GETPIVOTDATA("2019 Sales",$A$1) in cell E3, and
    - type formula =GETPIVOTDATA("2019 Profit",$A$1) in cell F3.

The GETPIVOTDATA functions return the grand totals of the PivotTable. They takes name of the PivotTable field that contains the data that you want to retrieve (2019 Sales and 2019 Profit) and a reference to any cell in the PivotTable ($A$1 cell of *KeyValuesPivot*) and return the grand totals.

E3 =GETPIVOTDATA("2019 Sales",$A$1)

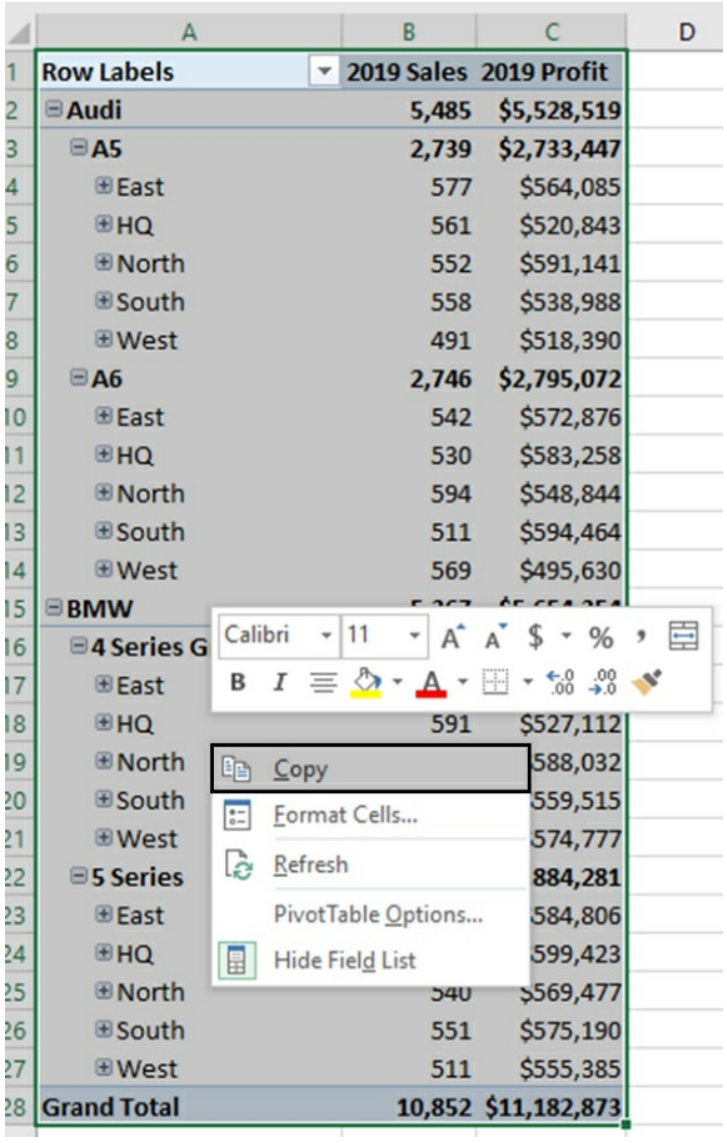
	A	B	C	D	E	F
1	Row Labels	2019 Sales	2019 Profit			
2	Audi	5,485	$5,528,519		2019 Sales	2019 Profit
3	A5	2,739	$2,733,447		s",$A$1)	
4	East	577	$564,085			
5	HQ	561	$520,843			
6	North	552	$591,141			
7	South	558	$538,988			
8	West	491	$518,390			
9	A6	2,746	$2,795,072			
10	East	542	$572,876			
11	HQ	530	$583,258			
12	North	594	$548,844			
13	South	511	$594,464			
14	West	569	$495,630			
15	BMW	5,367	$5,654,354			
16	4 Series Gran Coupe	2,699	$2,770,073			
17	East	490	$520,637			
18	HQ	591	$527,112			
19	North	554	$588,032			
20	South	526	$559,515			
21	West	538	$574,777			
22	5 Series	2,668	$2,884,281			
23	East	535	$584,806			
24	HQ	531	$599,423			
25	North	540	$569,477			
26	South	551	$575,190			
27	West	511	$555,385			
28	Grand Total	10,852	$11,182,873			

8. Confirm that the PivotTable grand totals match the output of the GETPIVOTDATA functions. Now, click on the arrow to the right of Row Labels on the top left of PivotTable and filter for *Audi*. Observe that grand totals of the PivotTable and also the output of the GETPIVOTDATA functions will change dynamically. Remove the filter to return to the unfiltered PivotTable.

	A	B	C	D	E	F
1	Row Labels	2019 Sales	2019 Profit			
2	Audi	5,485	$5,528,519		2019 Sales	2019 Profit
3	A5	2,739	$2,733,447		10,852	$ 11,182,873
4	East	577	$564,085			
5	HQ	561	$520,843			
6	North	552	$591,141			
7	South	558	$538,988			
8	West	491	$518,390			
9	A6	2,746	$2,795,072			
10	East	542	$572,876			
11	HQ	530	$583,258			
12	North	594	$548,844			
13	South	511	$594,464			
14	West	569	$495,630			
15	BMW	5,367	$5,654,354			
16	4 Series Gran Coupe	2,699	$2,770,073			
17	East	490	$520,637			
18	HQ	591	$527,112			
19	North	554	$588,032			
20	South	526	$559,515			
21	West	538	$574,777			
22	5 Series	2,668	$2,884,281			
23	East	535	$584,806			
24	HQ	531	$599,423			
25	North	540	$569,477			
26	South	551	$575,190			
27	West	511	$555,385			
28	Grand Total	10,852	$11,182,873			

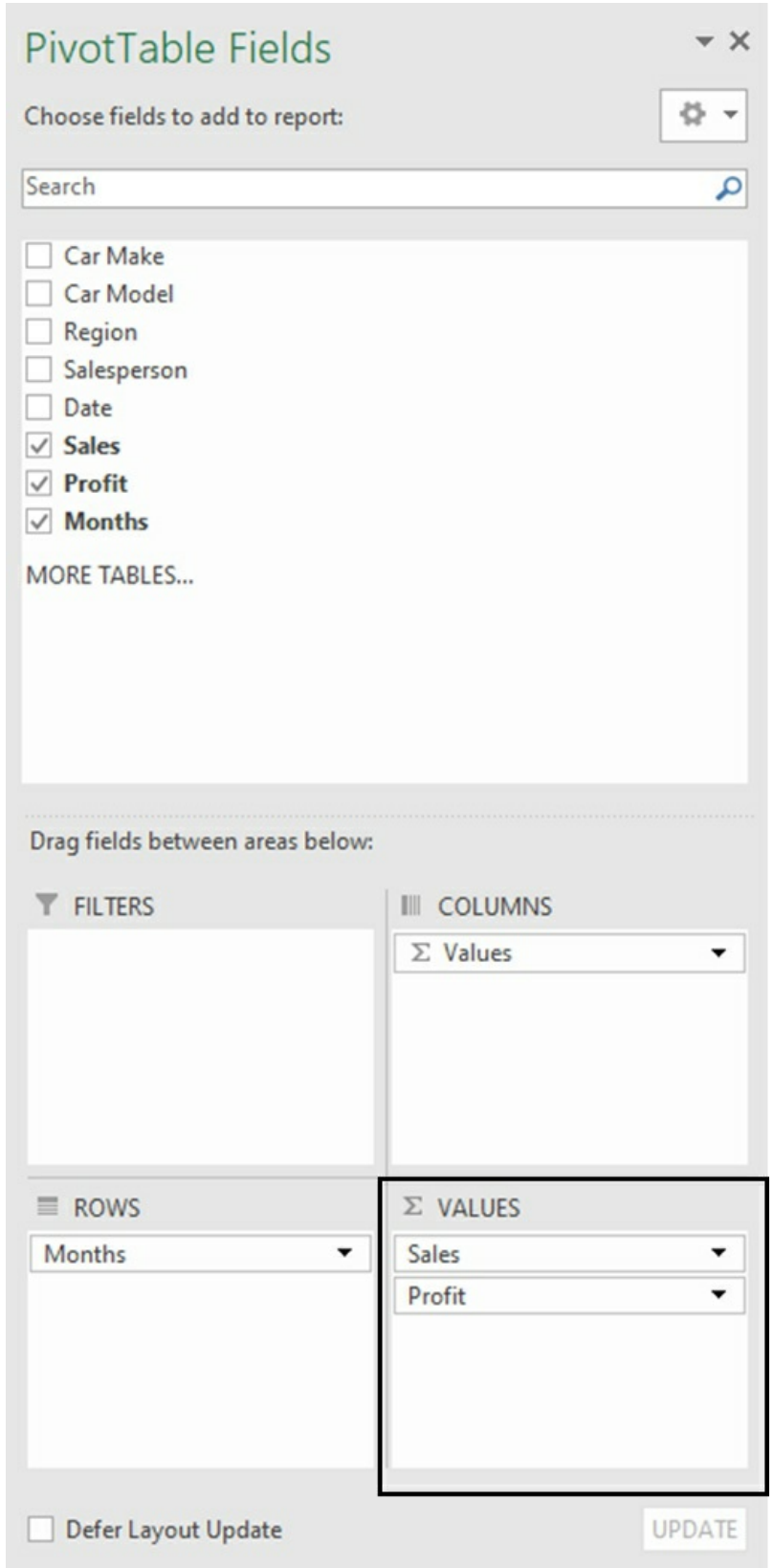
Insert Sales and Profit by Month PivotTable and Line PivotChart

9. Instead of creating another PivotTable, we will copy and paste the previous PivotTable and modify its fields. Copy *KeyValuesPivot* and paste it on cell $A$1 of the *Month Pivot* sheet.

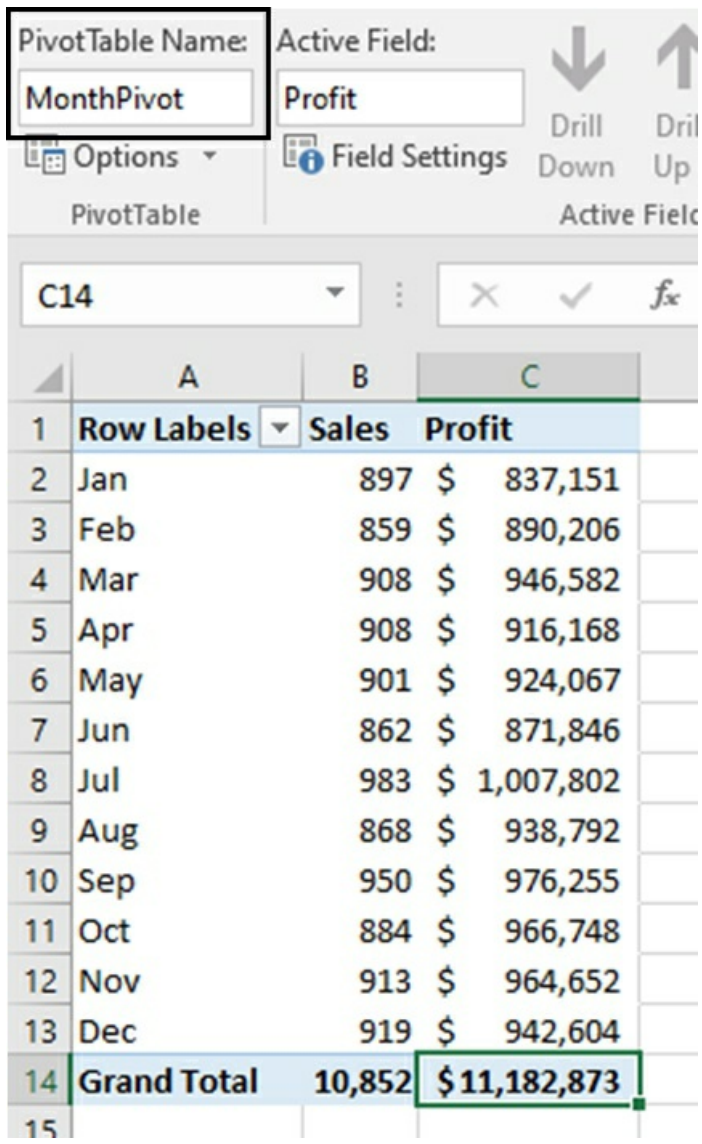


Row Labels	2019 Sales	2019 Profit
Audi	5,485	$5,528,519
A5	2,739	$2,733,447
East	577	$564,085
HQ	561	$520,843
North	552	$591,141
South	558	$538,988
West	491	$518,390
A6	2,746	$2,795,072
East	542	$572,876
HQ	530	$583,258
North	594	$548,844
South	511	$594,464
West	569	$495,630
BMW		
4 Series G		
East		
HQ	591	$527,112
North		588,032
South		559,515
West		574,777
5 Series		884,281
East		584,806
HQ		599,423
North	540	$569,477
South	551	$575,190
West	511	$555,385
Grand Total	10,852	$11,182,873

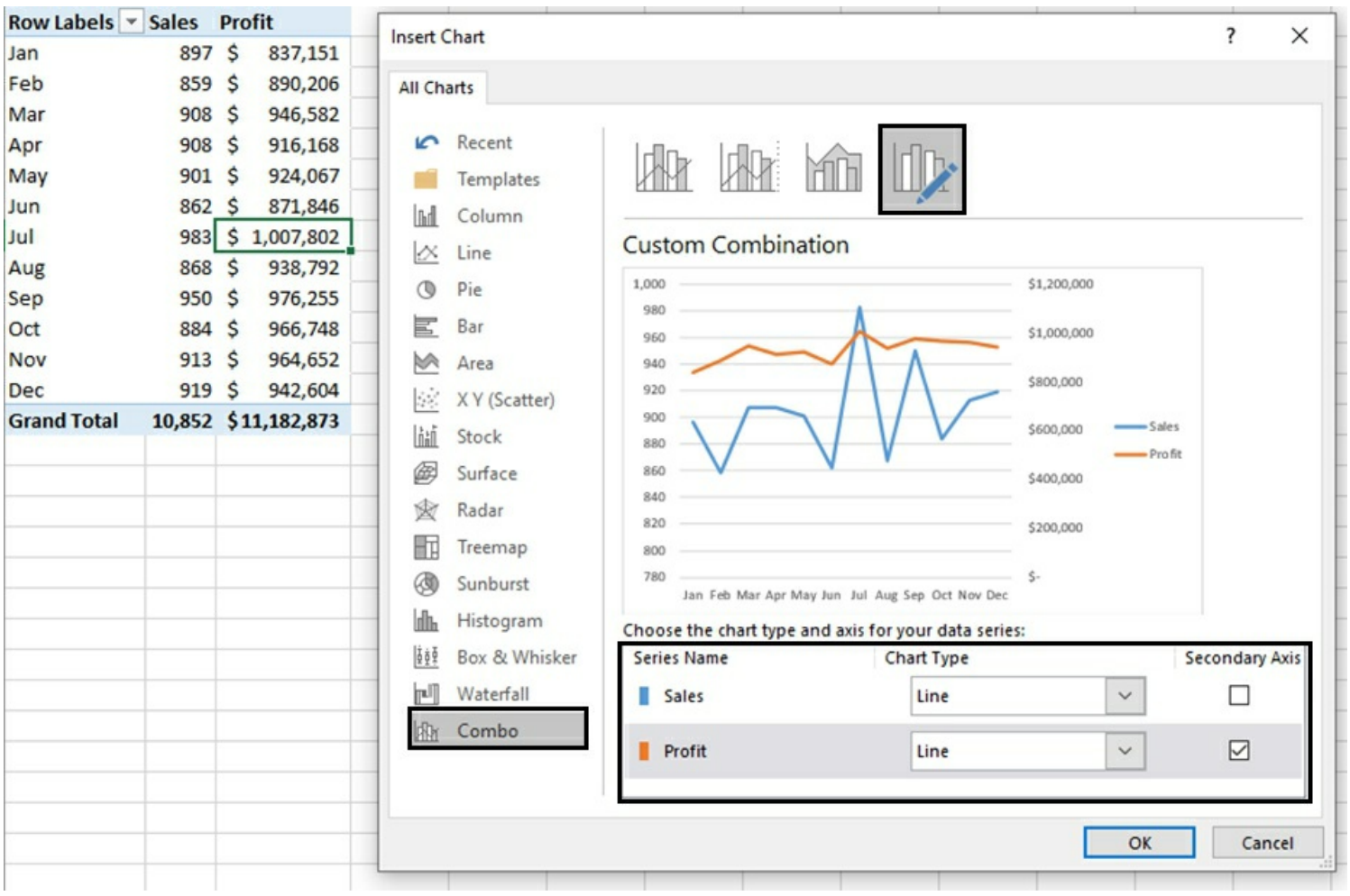
10. Modify the PivotTable fields as shown in the below screenshot. For *Σ Values*, we want to rename *Sum of Sales* and *Sum of Profit* fields as *Sales* and *Profit* respectively as they will be used as legend in the line PivotChart. The problem is that as PivotTable already has columns called Sales and Profit, it won't let us use these names. A neat trick is to add a space at the end of the existing names to make them "different" and Excel will no longer show an error!



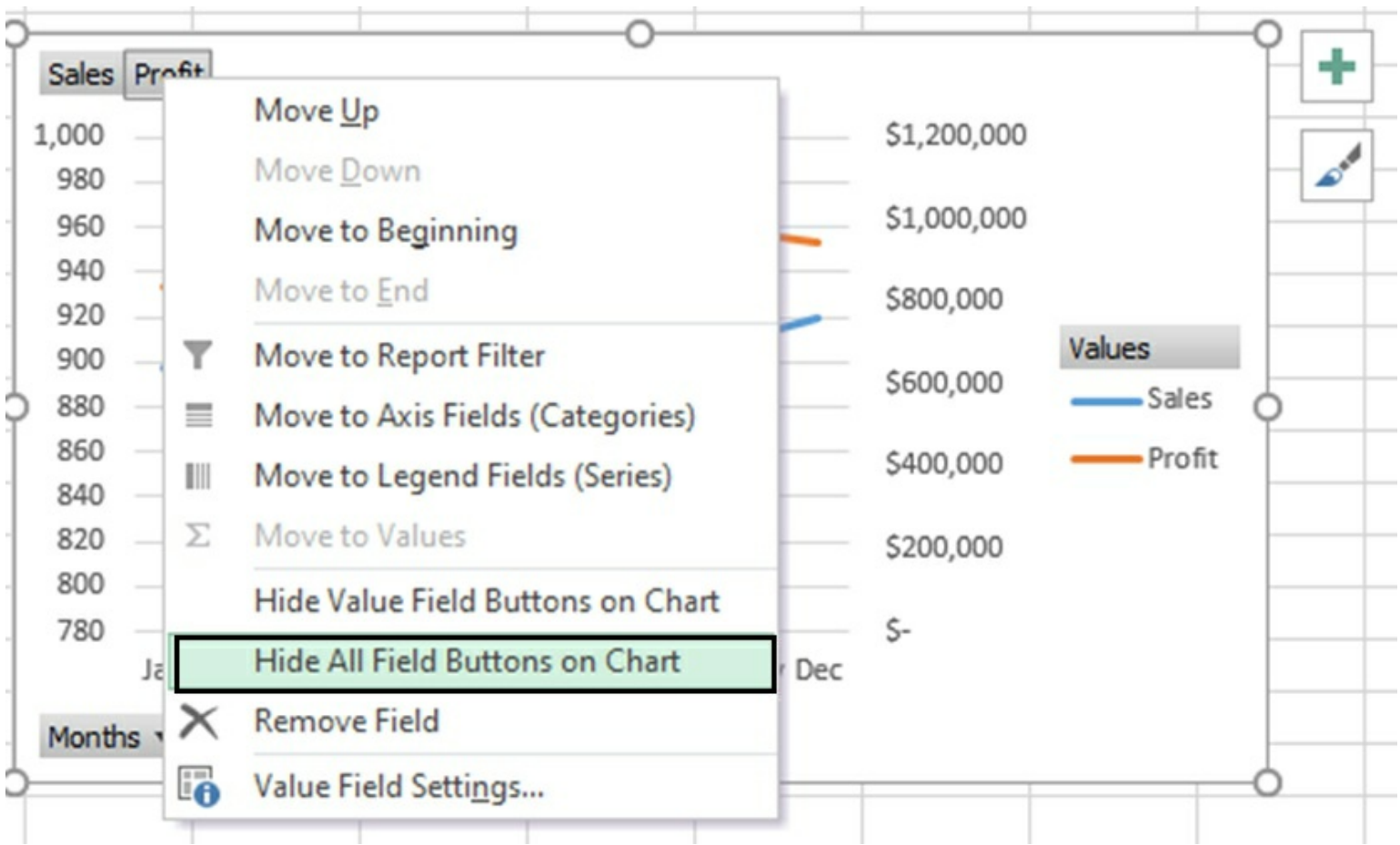
11. Rename the PivotTable to *MonthPivot*.



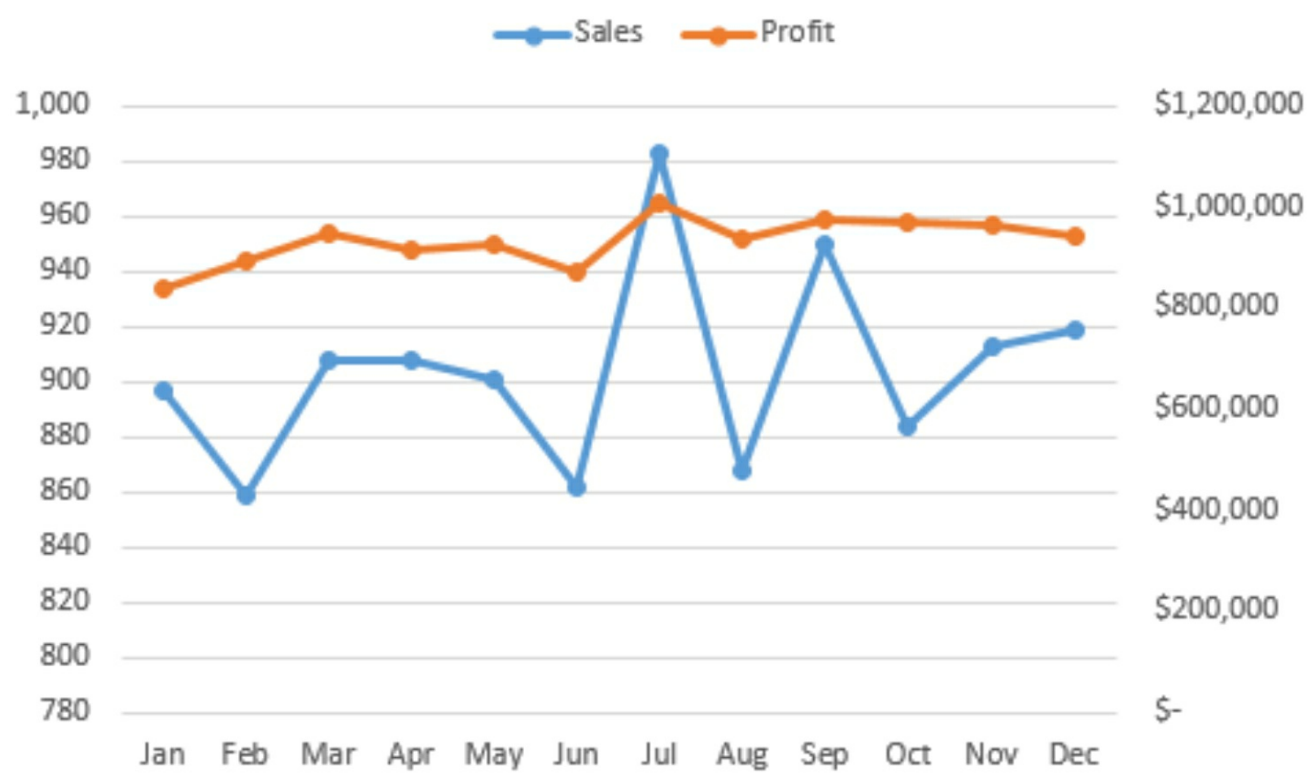
12. To insert the PivotChart, click anywhere in your PivotTable to enable *PivotTable Tools* tab. On the tab, go to *Analyze* and under the *Tools* group, select *PivotChart*. Go to *Combo* group and insert a custom combo chart with *Sales* as a line on the primary axis and *Profit* as a line on the secondary axis.



13. Right click on any of the filters on the PivotChart and select *Hide All Field Buttons* on the chart.

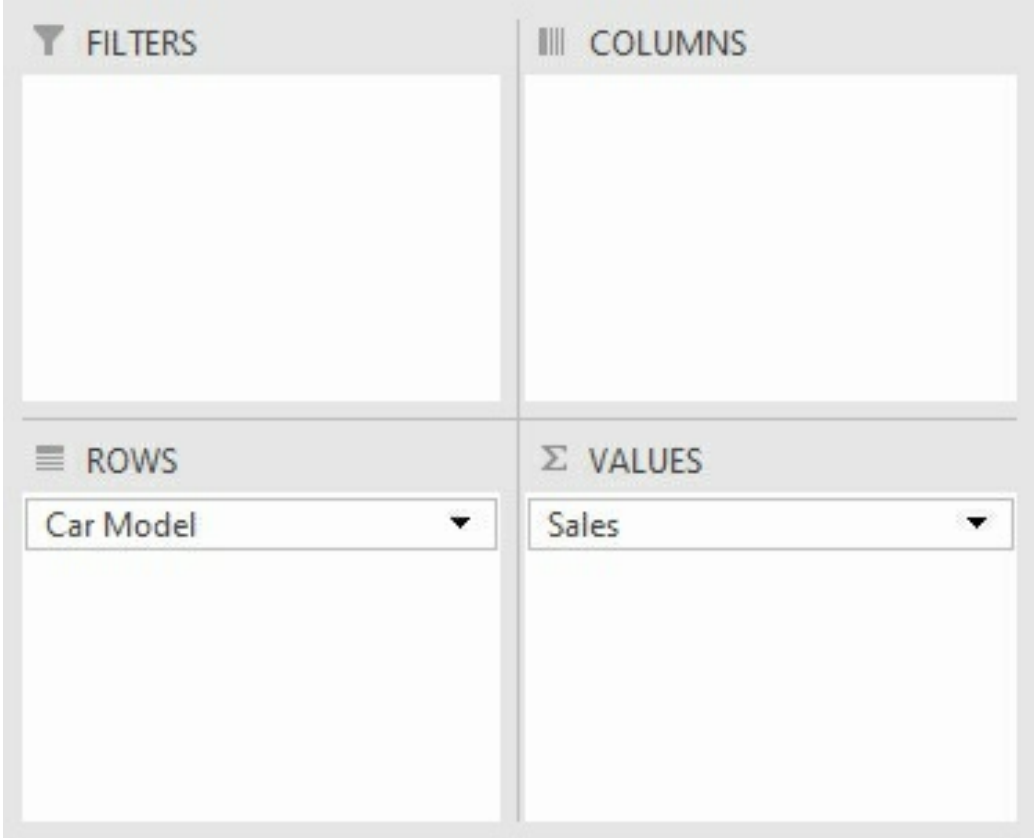


14. To finish, rename the chart as *MonthLineGraph*.

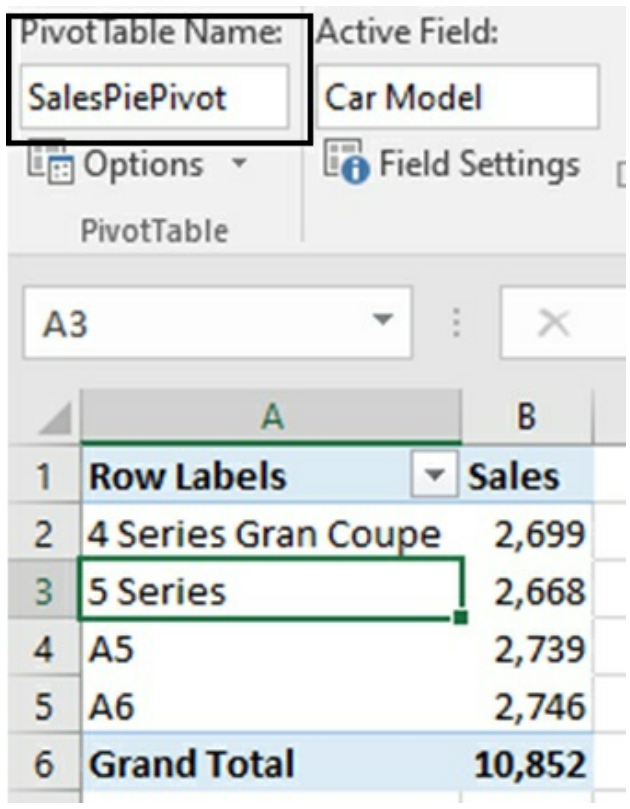


Insert Sales PivotTable and Pie PivotChart

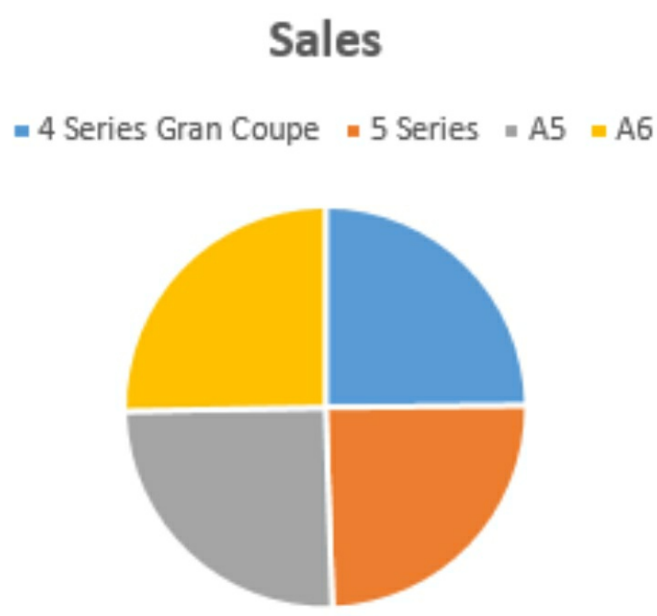
15. Copy *MonthPivot* and paste it on cell $A$1 of the *Sales Pie Pivot* sheet. Modify the PivotTable fields as shown in the below screenshot. For *Σ Values*, rename *Sum of Sales* fields as *Sales*.



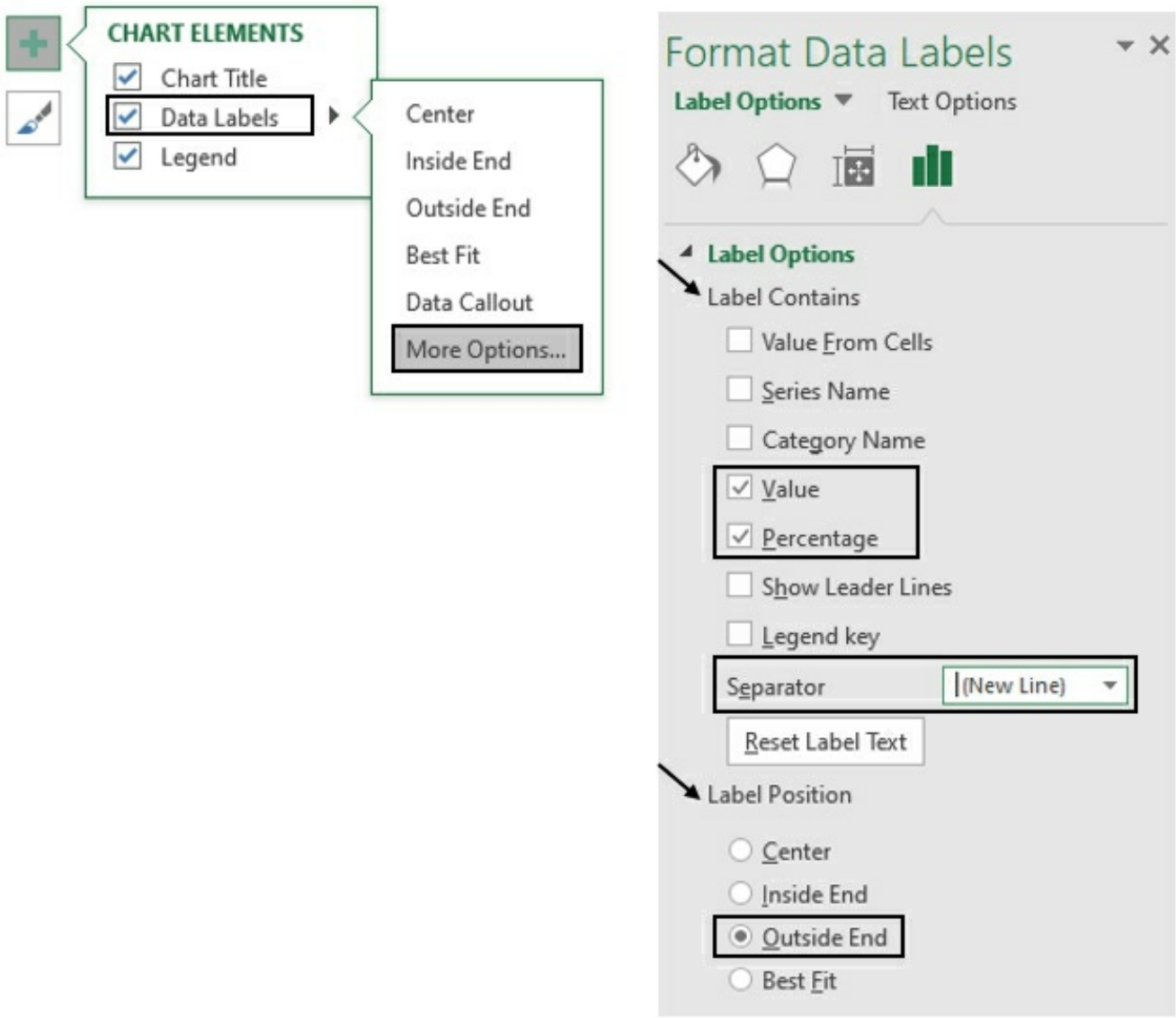
16. Rename the PivotTable to *SalesPiePivot*.



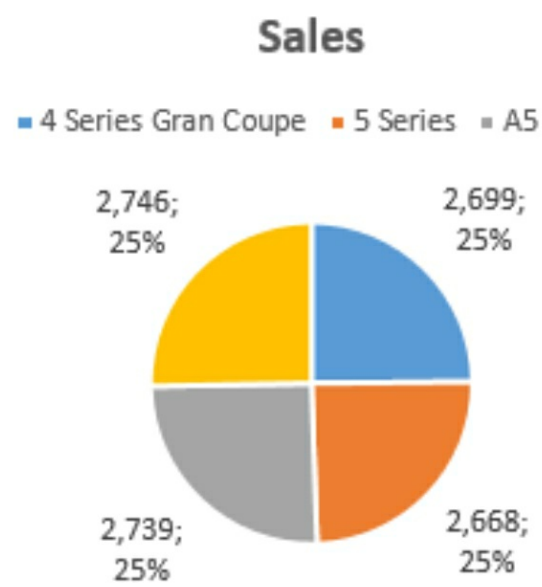
17. To insert the PivotChart, click anywhere in your PivotTable to enable *PivotTable Tools* tab. On the tab, go to *Analyze* and under the *Tools* group, select *PivotChart*. Go to *Pie* group and insert the first pie chart from the options.



18. We now want to include sales values as well as percentages on the pie chart. To do so, select the chart, click on the plus sign (Chart Elements) next to the chart in the upper right corner, check *Data Labels*, click on the arrow next to it and select *More Options* from the drop-down menu. A *Format Data Labels* window will appear on the right. For *Label Options*, check *Value* and *Percentage*, for *Separator*, select *New Line* to move the second label to a new line, and for *Label Position*, select *Outside End*. Note that *Add Chart Element* command is also available through the *Design* tab.

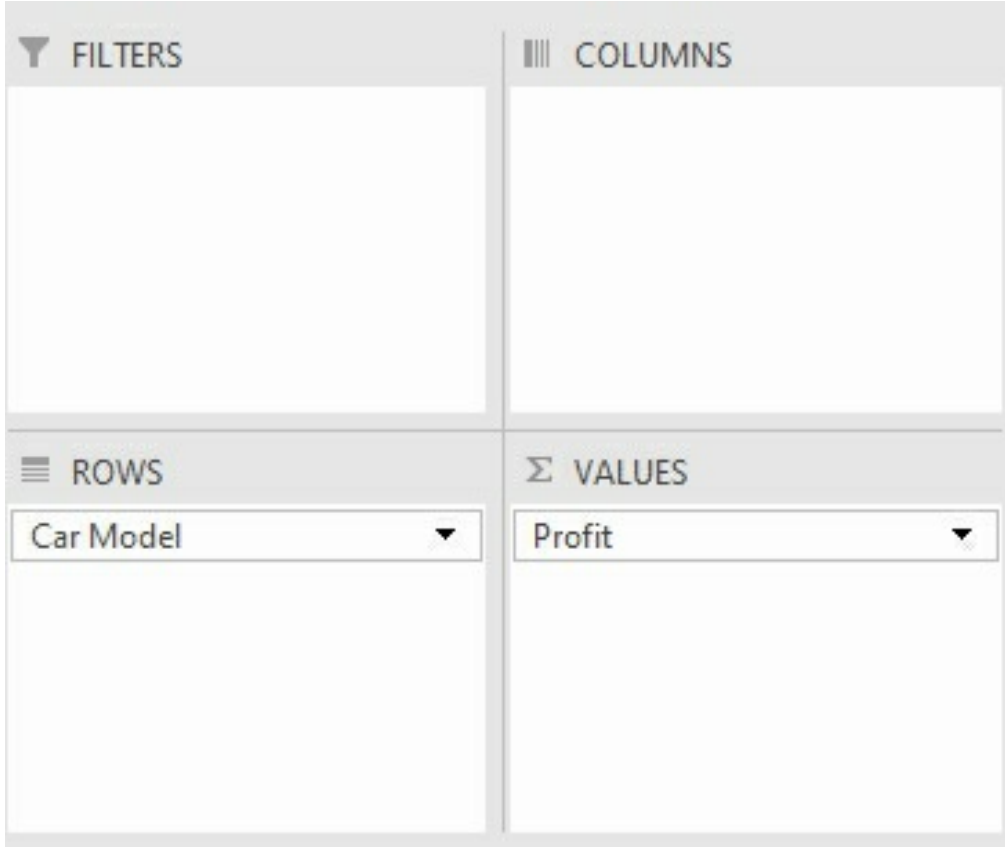


19. Rename chart title to *Sales* and place chart legend on top of the chart. Rename the chart to *SalesPieChart*.

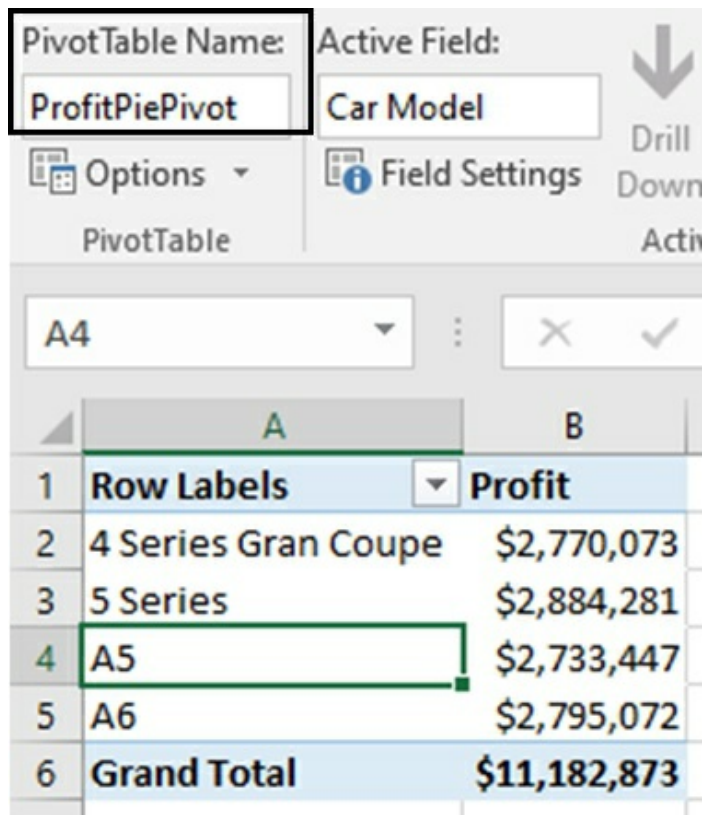


Insert Profit PivotTable and Pie PivotChart

20. Copy *SalesPiePivot* and paste it on cell $A$1 of the *Profit Pie Pivot* sheet. Modify the PivotTable fields as shown in the below screenshot. For *Σ Values*, we want to rename *Sum of Profit fields* as Profit. To do so, add two spaces at the end of the existing name Profit.

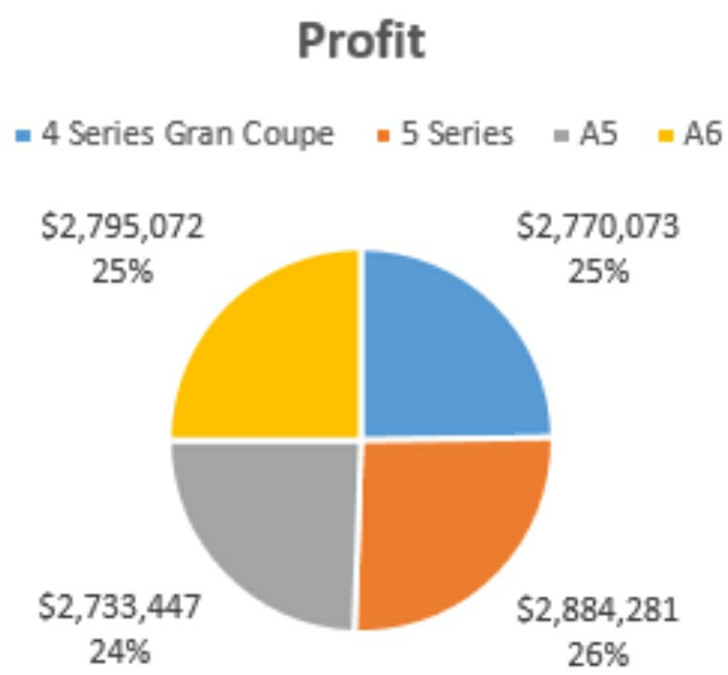


21. Rename the PivotTable to *ProfitPiePivot*.



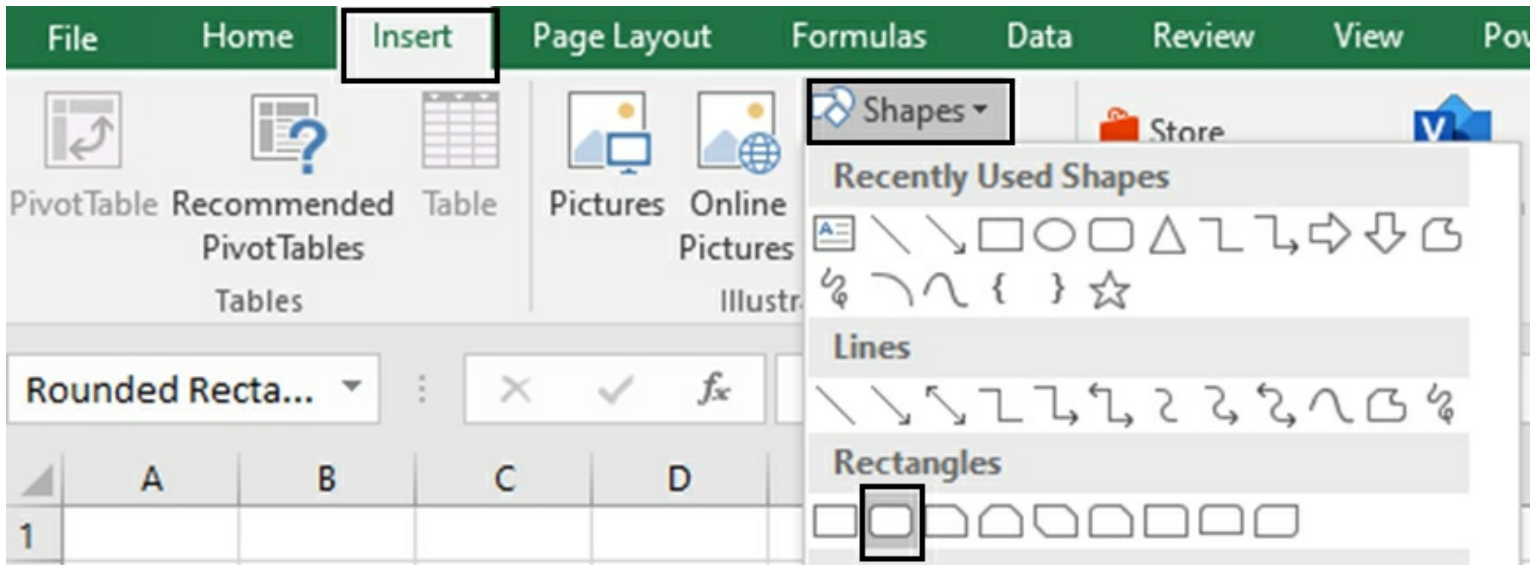
22. Insert the pie PivotChart, introduce outside end labels for profit values as well as percentages on the PivotChart.

23. Rename chart title to *Profit* and place chart legend on top of the chart. Remember to rename the chart to *ProfitPieChart*.

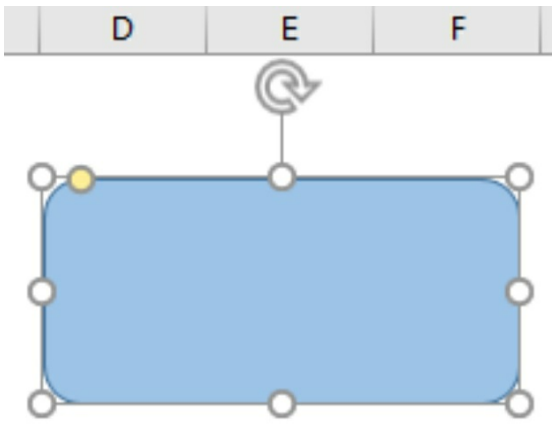


Assemble the Dashboard

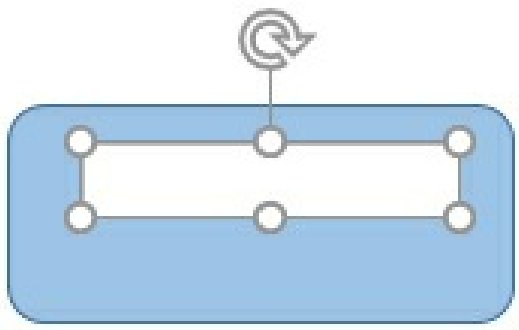
24. Now that we have all dashboard elements ready, we will start assembling the actual dashboard on the *DASHBOARD* sheet. Let's start with the key values. Go to the *DASHBOARD* sheet and click on the *Insert* tab. On the tab, under *Illustrations* group, click on *Shapes* and select a *Rounded Rectangle* from the dropdown.



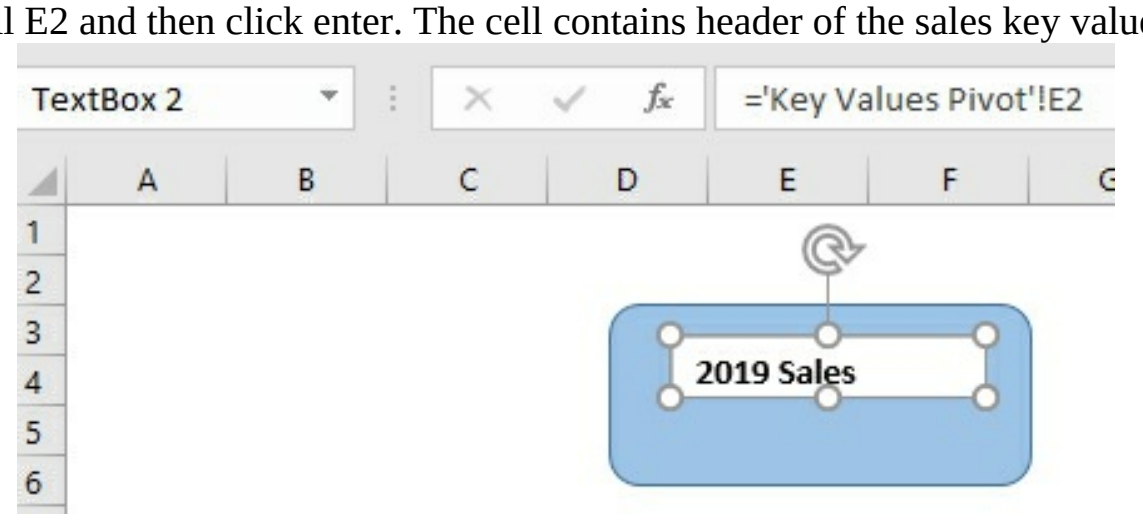
25. Draw the rounded rectangle toward the top left of the sheet.



26. Go to *Insert* tab. Under the *Text* group, click on *Text Box* and draw a text box on the upper part of the rounded rectangle.



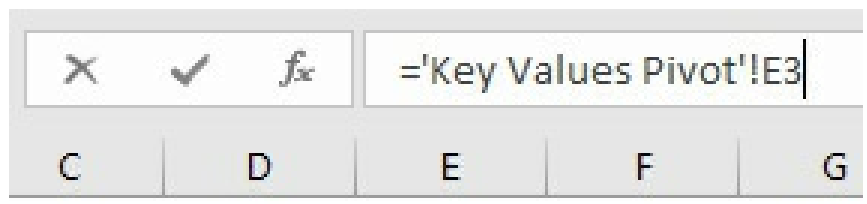
27. Select the text box, go to the formula bar, type '=', go the *Key Value Pivot* sheet =, select cell E2 and then click enter. The cell contains header of the sales key value.



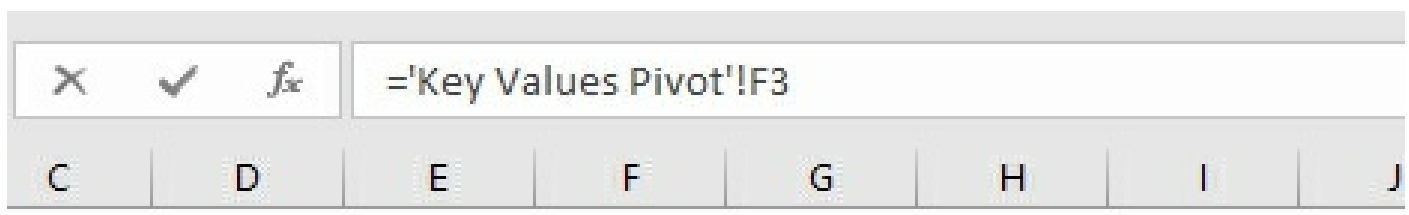
28. Select the text box, if needed to display the *Drawing Tools* tab. Align the text toward the centre of the box. Increase the size of the text to 14. On the *Drawing Tools* tab under *Format*, change the *Shape Fill* to *No Fill* and *Shape Outline* to *No Outline*.

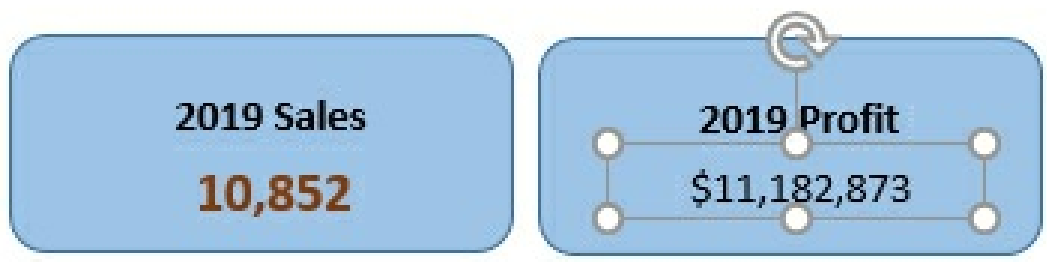
29. Now, copy the text box, and paste and align it on the lower part of the rounded rectangle. Copy and pasting will ensure that the size and alignment of the boxes stay consistent.

30. Now, select the lower text box. Go to the formula bar and change cell reference from ='Key Values Pivot'!E2 to ='Key Values Pivot'!E3. The E3 cell contains the grand total of sales as extracted from the *KeyValuesPivot*. Format the content of the text box as desired, for example, make it bold or assign a different color to it.

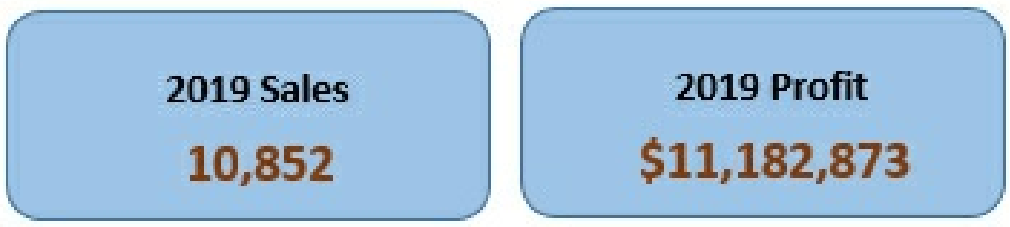


31. Now, our sales key value is ready. Instead of creating profit key value from scratch, we will just copy the sales key value, paste it, and change cell references of the two text boxes. To ensure the sales key value is copied properly, we will first group all of its elements together. To do that, keep the Ctrl key pressed and select rectangular shape and then both text boxes one by one, right click, hover over *Group* to see the options and then, click on *Group*. This will group all three elements together. Now copy the group, paste it next to the sales key value and change cell references of upper and lower text boxes to cells ='Key Values Pivot'!F2 and ='Key Values Pivot'!F3 that contain profit key value header and grand total respectively.

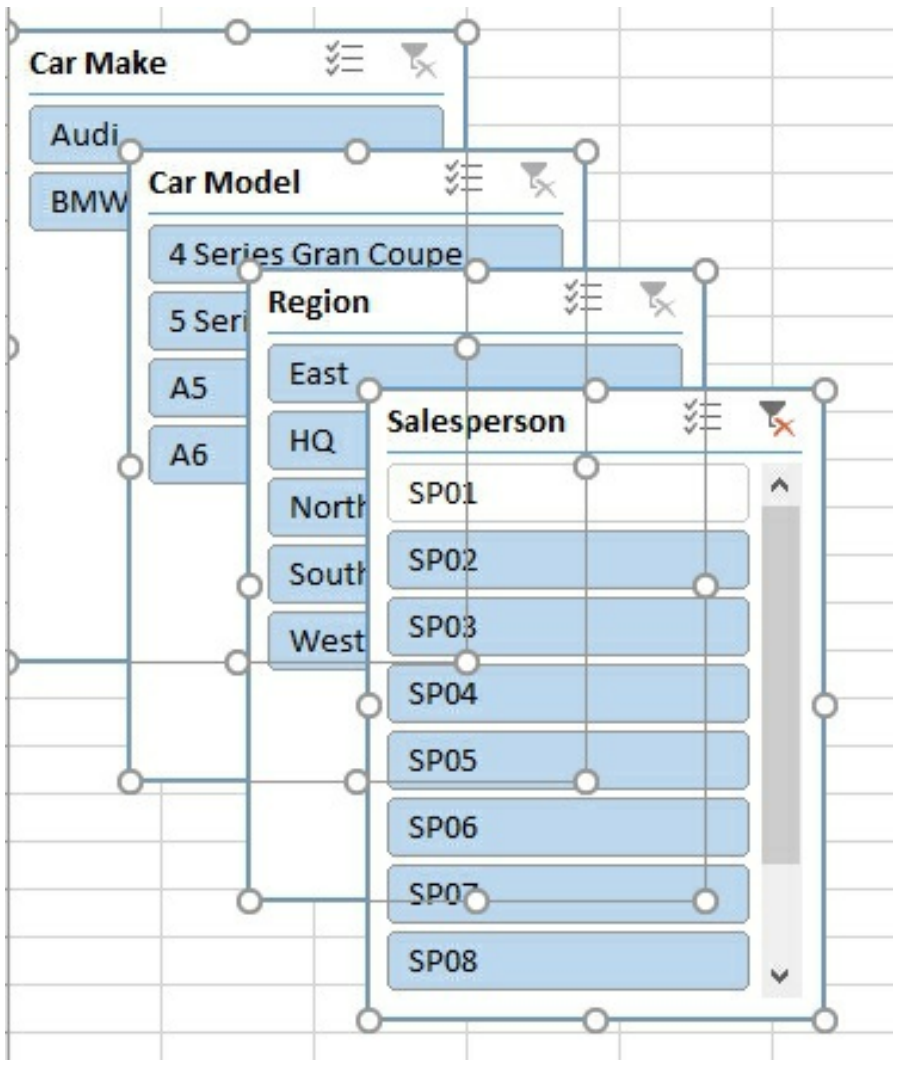




32. Format text box with profit value same as the box with sales value. Now have our key values ready!

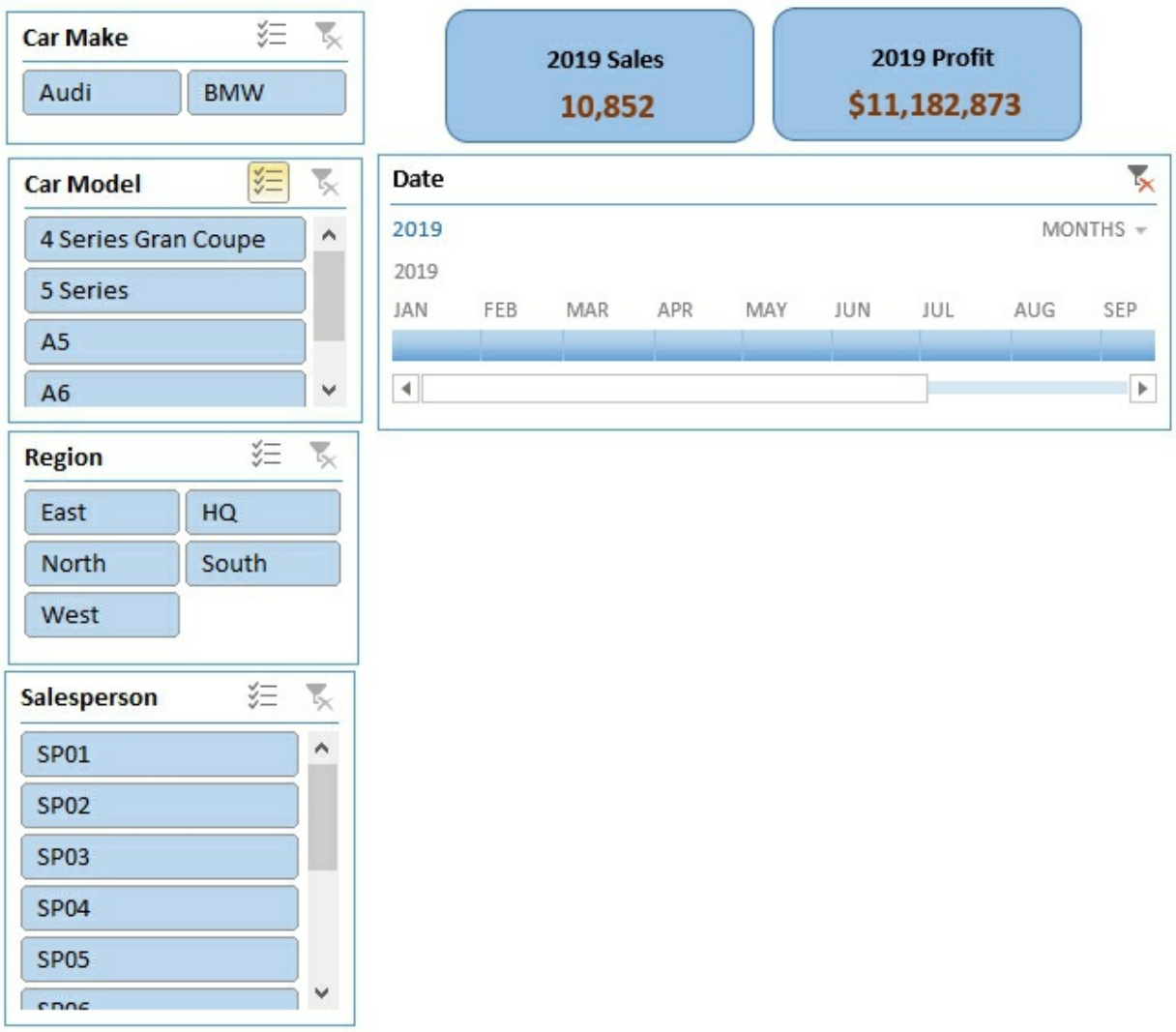


33. Now, let's move on to the slicers! Click on any of the PivotTables in the workbook to active the *PivotTable Tools* tab. On the tab, click on *Analyze*, and under the *Filter* group, select *Insert Slicer*. On the *Insert Slicers* window, check *Car Make*, *Car Model*, *Region* and *Salesperson*, and click *OK*. This will insert four slicers on the sheet same as the PivotTable. Keeping the Ctrl key pressed, select them one by one, cut them from the current sheet and paste on the *DASHBOARD* sheet.



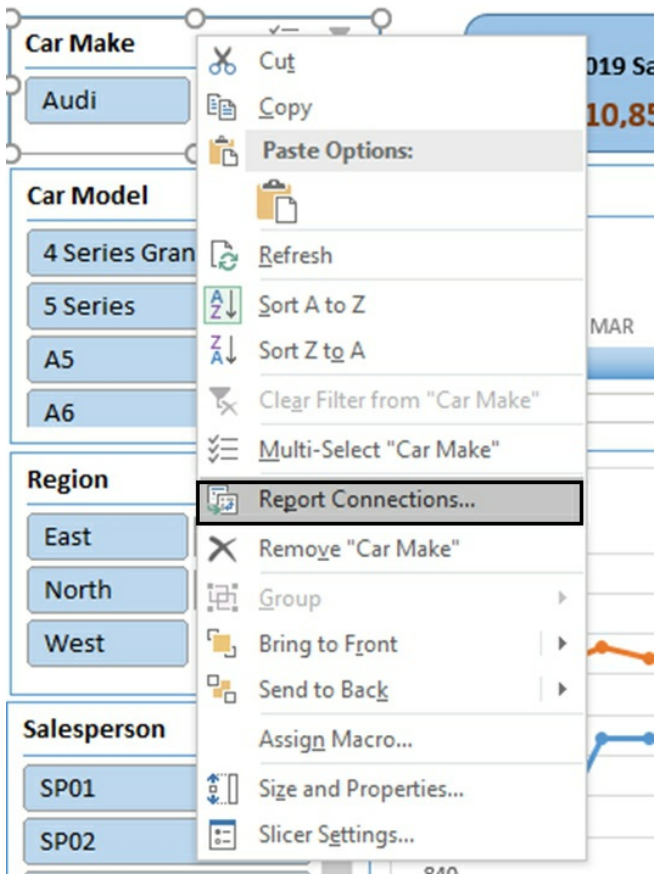
34. Align the slicers to the left of the sheet below one another and place key values next to them. Change the number of columns in the slicers, as desired.

35. Similarly, insert a timeline using *Date* and place it right below the key values. Ensure *Months* are selected as period on the timeline. For steps, refer to Chapter 5.3, if needed.

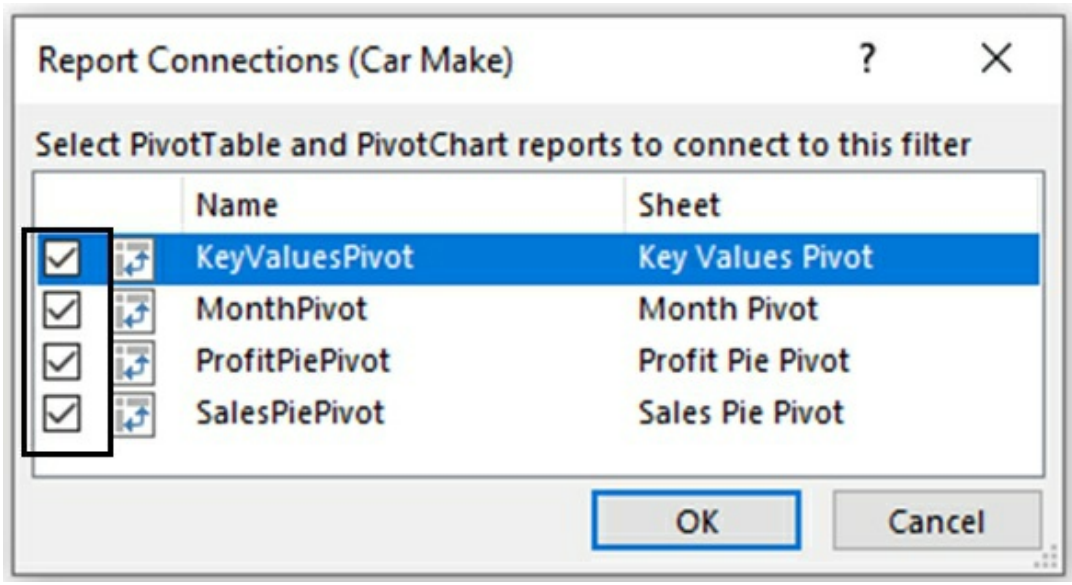


36. Now, cut *MonthLineGraph* PivotChart from the *Month Pivot* sheet and paste it right below the timeline on the *DASHBAORD* sheet. Similarly, cut *SalesPieChart* PivotChart from the *Sales Pie Pivot* sheet and *ProfitPieChart* PivotChart from the *Profit Pie Pivot* sheet one by one, and paste them to the left of the dashboard.

37. Slicers will only be connected to the PivotTable you used to create them, so you need to connect them to the rest of the PivotTables feeding into your dashboard. Select each slicer, right click and click on *Report Connections*. Alternatively, you can go to Slicer Tools > Options > Report Connections. A *Report Connections* window will pop up.



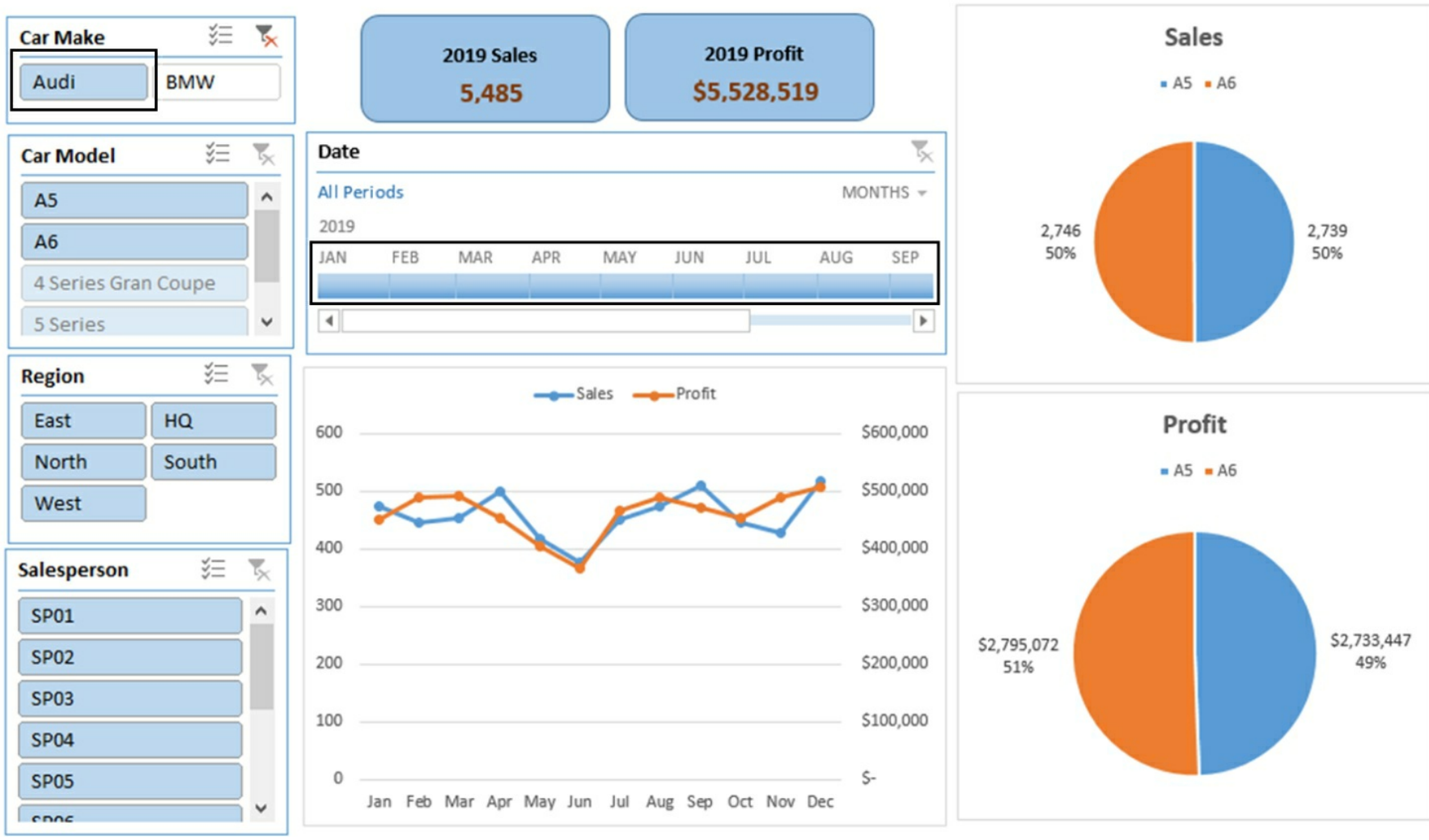
38. On the *Report Connection* window, check *KeyValuePivot*, *MonthPivot*, *ProfitPiePivot* and *SalesPiePivot* and click *OK*.



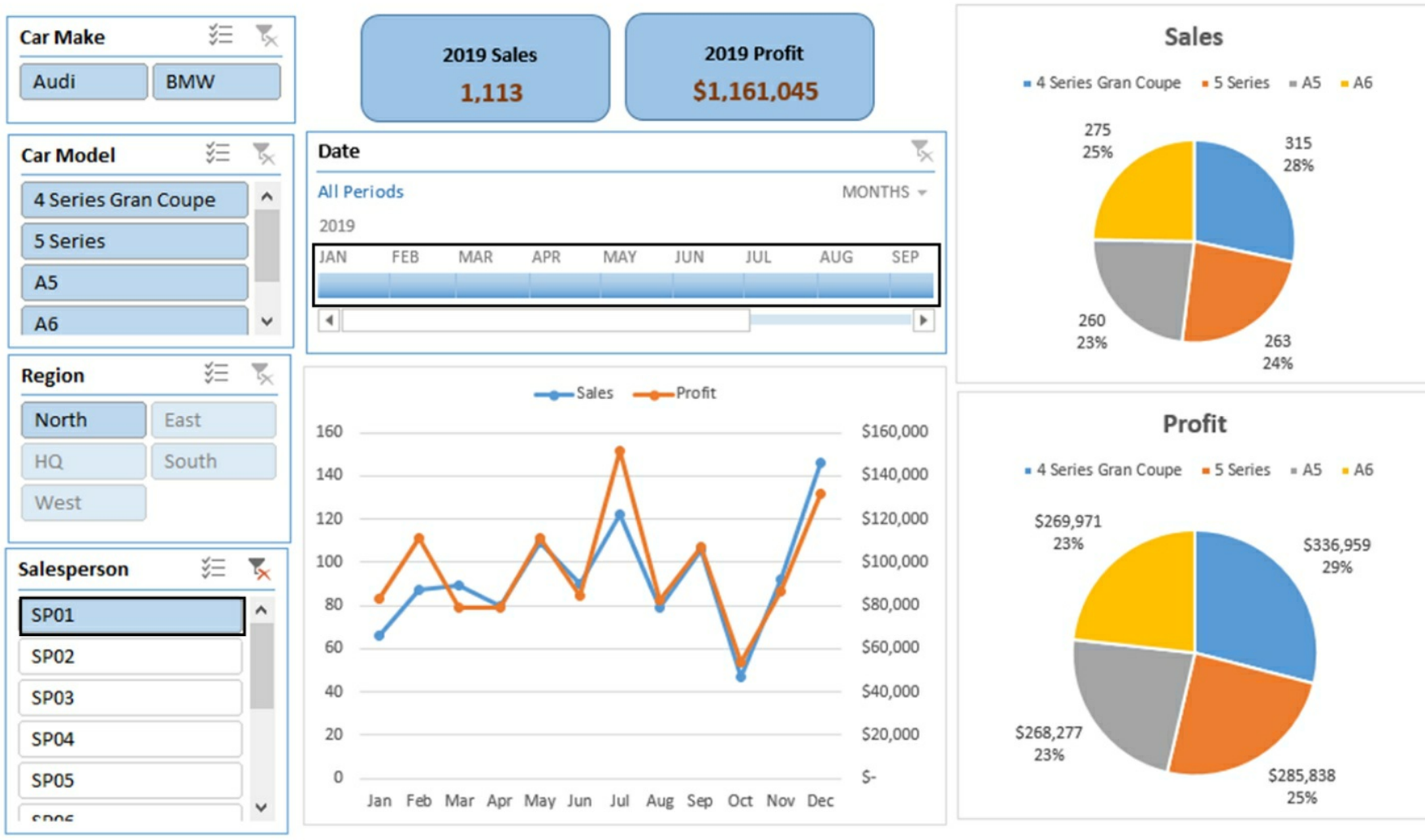
39. Similarly, connect the timeline to *KeyValuePivot*, *MonthPivot*, *ProfitPiePivot* and *SalesPiePivot* and click *OK*.

40. Now to make your dashboard look clean, go to the *DASHBOARD* sheet, click on the *View* tab, and uncheck gridlines to hide them.

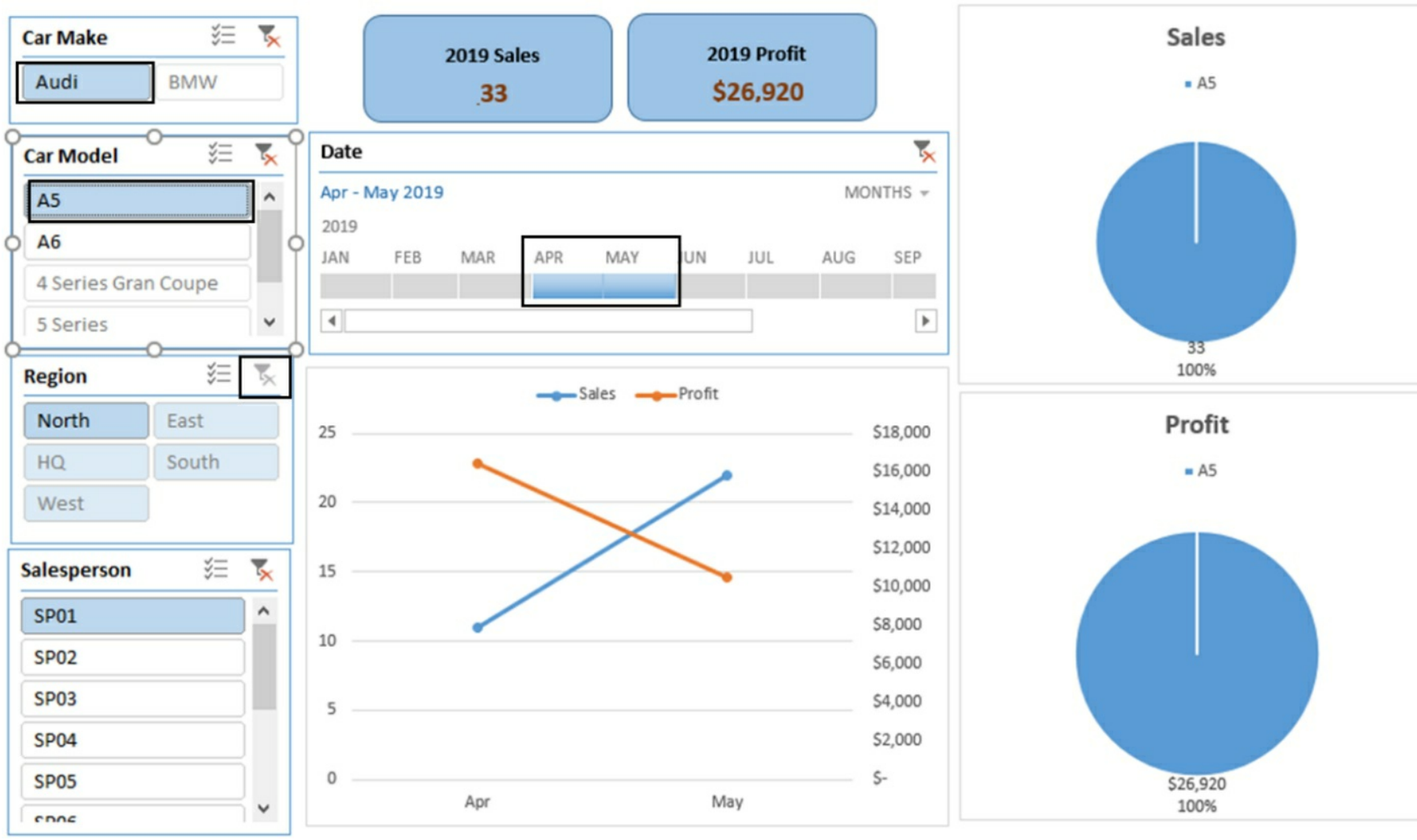
41. Before we start celebrating, how about we play around with our final product a little bit! Can you show me the sales and profit trend for just Audi cars in 2019?



42. How about monthly trend of sales and profit for salesperson SP01 in 2019?



43. Let's check the monthly trend of sales and profit for SP01 specifically for Audi A5s during April and May of 2019.



44. If you were able to explore answers to each of these questions, your dashboard is fully functional.

4.6 Dashboard Best Practices

An Excel dashboard is a useful decision-making tool that uses dynamic graphs, charts and tables to summarize and report KPIs. Dashboard best practices are critical while designing, distributing and printing your dashboards. In this chapter, we'll go over these best practices to ensure you create user-friendly and easily accessible Excel dashboards.

Designing your dashboard

Let's start with discussing key best practices on how to design dashboards that meet the needs of your users.

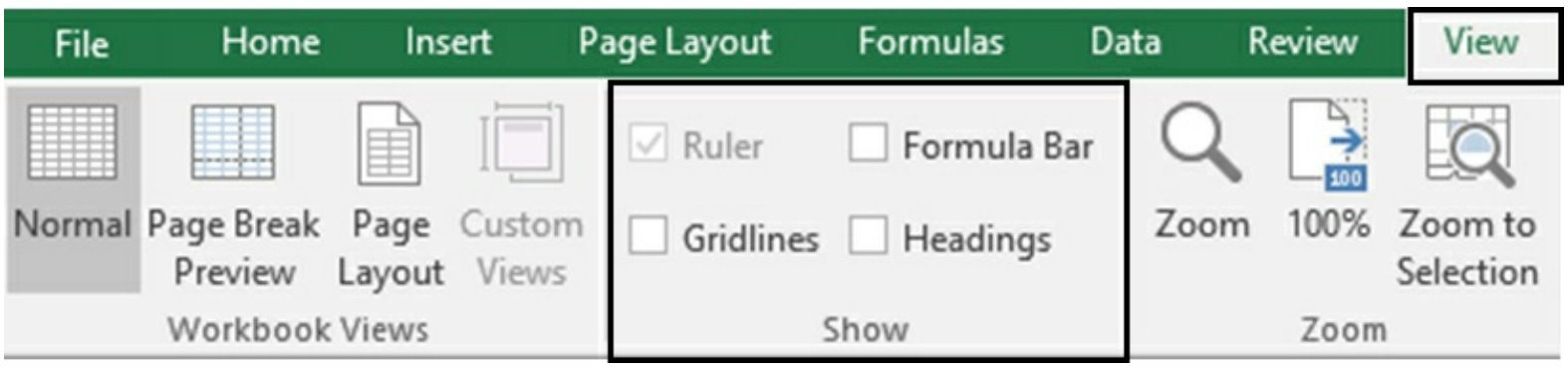
1. **Know your audience:** Tailor your dashboard to the needs and expertise of the target audience. Ensure that the audience is able to view, filter and interpret data being displayed on the dashboard easily and quickly. The dashboard should inform them and empower them to make decisions. Choose the complexity of your design based on the skill set of your audience.
2. **Highlight key data at top:** Add key data points on the top to highlight and focus them on the dashboard. Key values are a great way to highlight such data points.
3. **Keep the layout simple:** Design your dashboard in a way that it provides a streamlined summary of the key information, but don't clutter it. Remember the objective of the dashboard is to answer a question and support decision-making. Remove everything that deviates from this objective. Stay consistent with colors, labelling and formatting.
4. **Choose the right type of elements:** Choose the right type of dashboard elements from your available options – PivotTables, PivotCharts, slicers, timelines, key values etc. It's important to understand what type of information you want to convey and choose a dashboard element that is suited to the task. Use appropriate charts. If your audience is familiar with the combination charts, you can use them to save space.
5. **Design efficiently:** Add a dashboard title for context. If needed, add footnotes that help understand the dashboard and the data it contains a bit better. PivotTables can't overlap one another. To avoid errors, ensure to allow enough space between them so that they can expand and contract when filters are added, removed, or changed.

Distributing your dashboard

Once your dashboard design is complete, make some final adjustments prior to distributing it to ensure your dashboard looks seamless and is protected from any changes made by the users. You're welcome to open the Sales Dashboard we created in the previous chapter and practice these adjustments to better understand them.

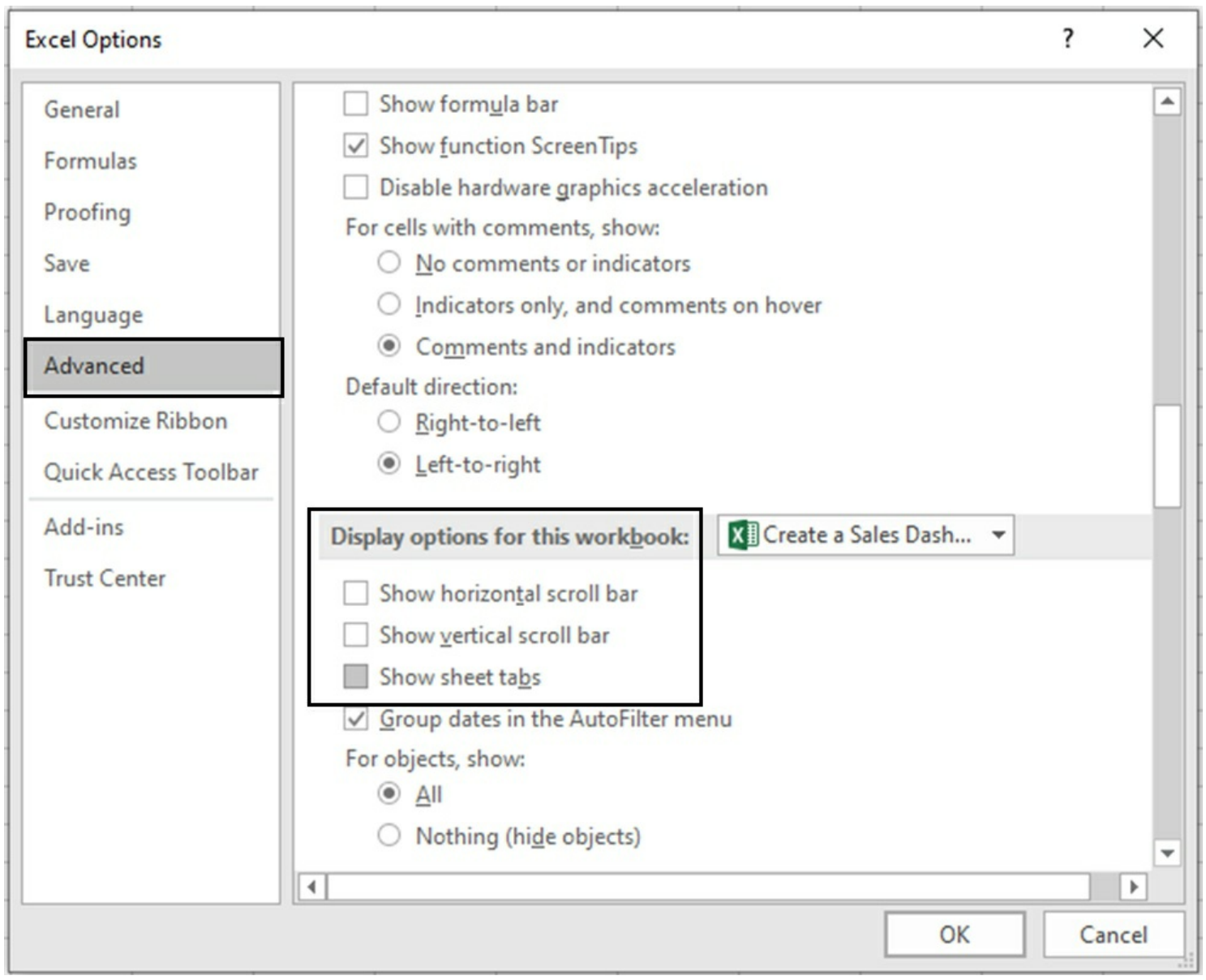
1. **Hide unnecessary features**

Get rid of row and column labels, gridlines and formula bar. On the *View* tab, under *Show* group, unselect *Headings*, *Gridlines* and *Formula*.



Hide all data sheets other than the sheet on which the dashboard is placed. Right click on the sheet name on the tab and click hide. To hide them together, click on the first sheet, hold the shift key and click on the last one to select all of them together, then right-click and select *Hide*. To unhide the sheet, right-click on any sheet name visible on the tab, select *Unhide* and then select the sheet you want to unhide.

Alternatively, you can hide the sheet tabs altogether. Go to *File* tab and click on *Options* to open the *Excel Options* window. On the window, click on *Advanced*, scroll down to *Display Options for this workbook* and unselect *Show sheet tabs*. If you choose this option, it will be a good idea to unselect *Show horizontal scroll bar* and *Show vertical scroll bar* to ensure the users don't scroll away. You can unhide them to update the dashboard or as needed.



2. Test slicers and timelines

Test your slicers and timelines to make sure that your PivotTables and PivotCharts work properly. Ensure that PivotTables don't overlap each other, otherwise, Excel will display an error message. These issues will need to be fixed before you distribute your dashboard.

3. Version proofing

Ideally, the users of your dashboard should have the same version of MS Excel that the dashboard is built in. If not, ensure that all formulas you have used to build the dashboard are compatible with the version users are using. Rework incompatible formulas, if needed.

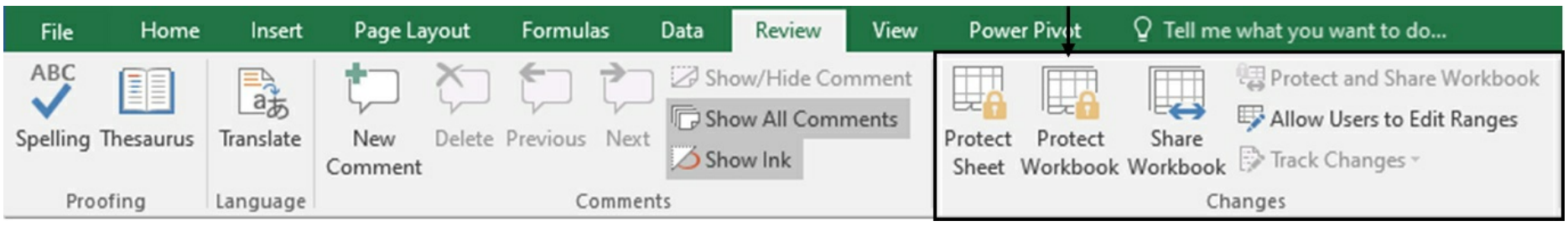
4. Provide context

Provide a user guide on how to use the dashboard as well as definitions of data points included in the dashboard.

You could include a separate worksheet with how-to-steps and definitions. It will help people navigate and understand your dashboard.

5. **Protect your dashboard**

Protect your workbook by going to the *Review* tab and selecting *Protect Workbook* under the *Changes* group. Insert an optional password in the window that pops up. Please note that while this will ensure users can navigate and filter the dashboard as they need without accidentally messing up the dashboard, it is not a bulletproof method to prevent users from accessing any confidential data included on one of the hidden sheets. There are plenty of methods available on the internet to crack workbook protection.

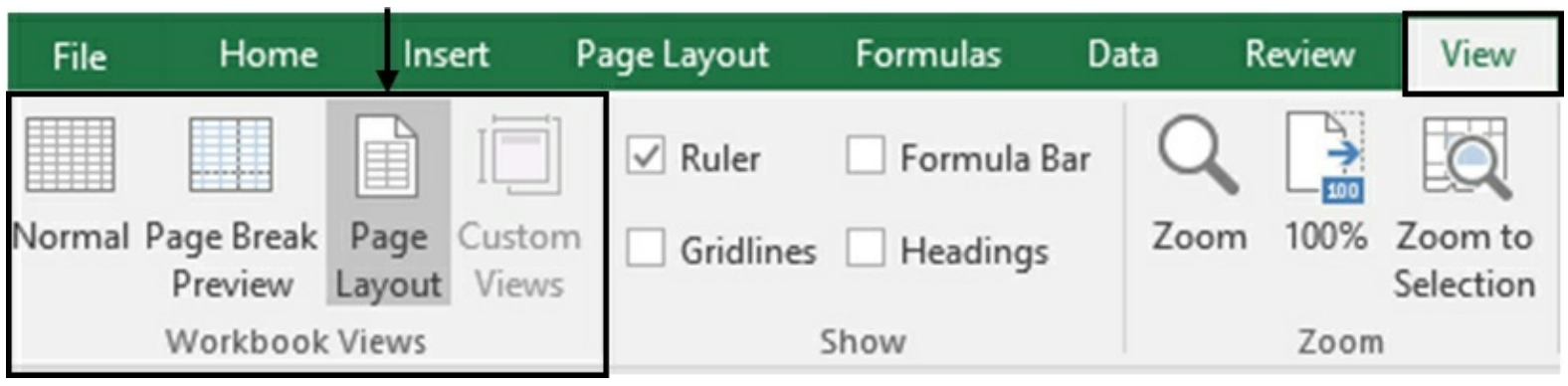


Printing your dashboard

Your users might want to print your dashboard! Yes, it can happen. Though interactivity is the heart and soul of a dashboard, a user might want to print a specific dashboard view to take to a meeting or to mark up. However, spreadsheets aren't easily printable. Make the below adjustments to make sure that someone can easily print the dashboard. In addition to the dashboards, these adjustments can be employed to print any spreadsheet. Again, feel free to apply these adjustments to the Sales Dashboard for practice.

1. **Switch to Page Layout**

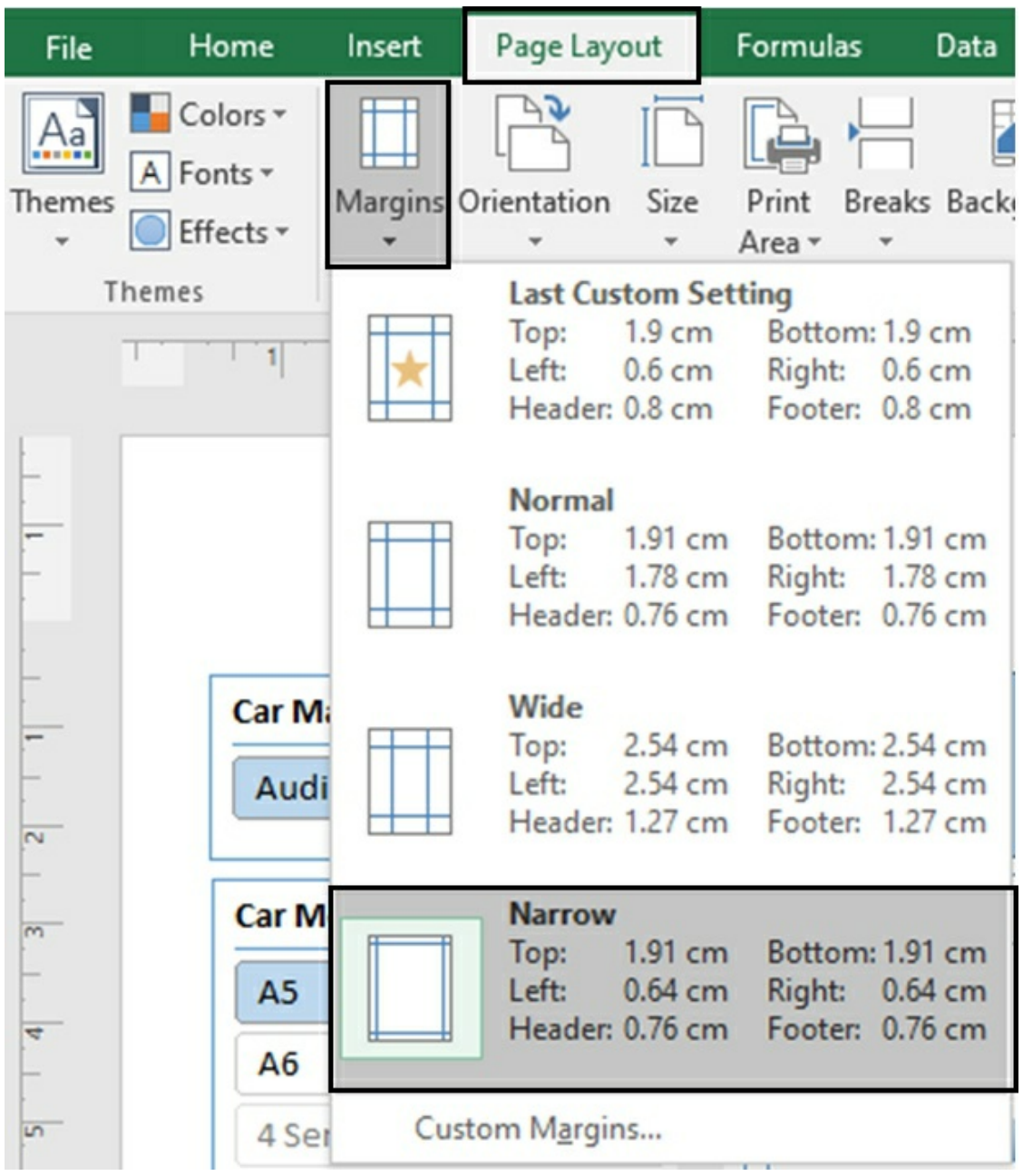
When setting up your dashboard for printing, it is very useful to switch from *Normal Layout* to *Page Layout* as it helps to view how content on the dashboard splits over multiple sheets. To switch to the Page Layout, go to the *View* tab and under *Workbook Views* group, select *Page Layout*.



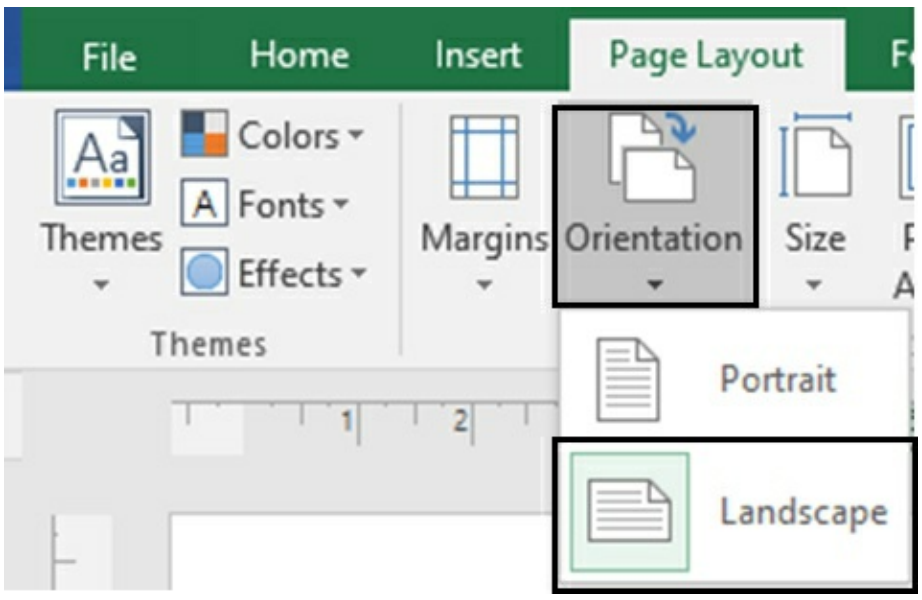
2. **Adjust Page Orientation and Margins**

As you review your dashboard in Page Layout, you might want to adjust the orientation and the margins to make sure your dashboard prints properly. In most cases, you would want it to print on a single page as opposed to other Excel reports that might expand to several pages.

To change the margins, click on the *Page Layout* tab and under the *Page Setup* group, click *Orientation*. You have an option to choose from *Normal*, *Wide* or *Narrow* settings, *Last Custom Setting* and *Custom Margins*. The *Custom Margins* options lets you set your own margins. For Sales Dashboard, set *Margins* to *Narrow*.

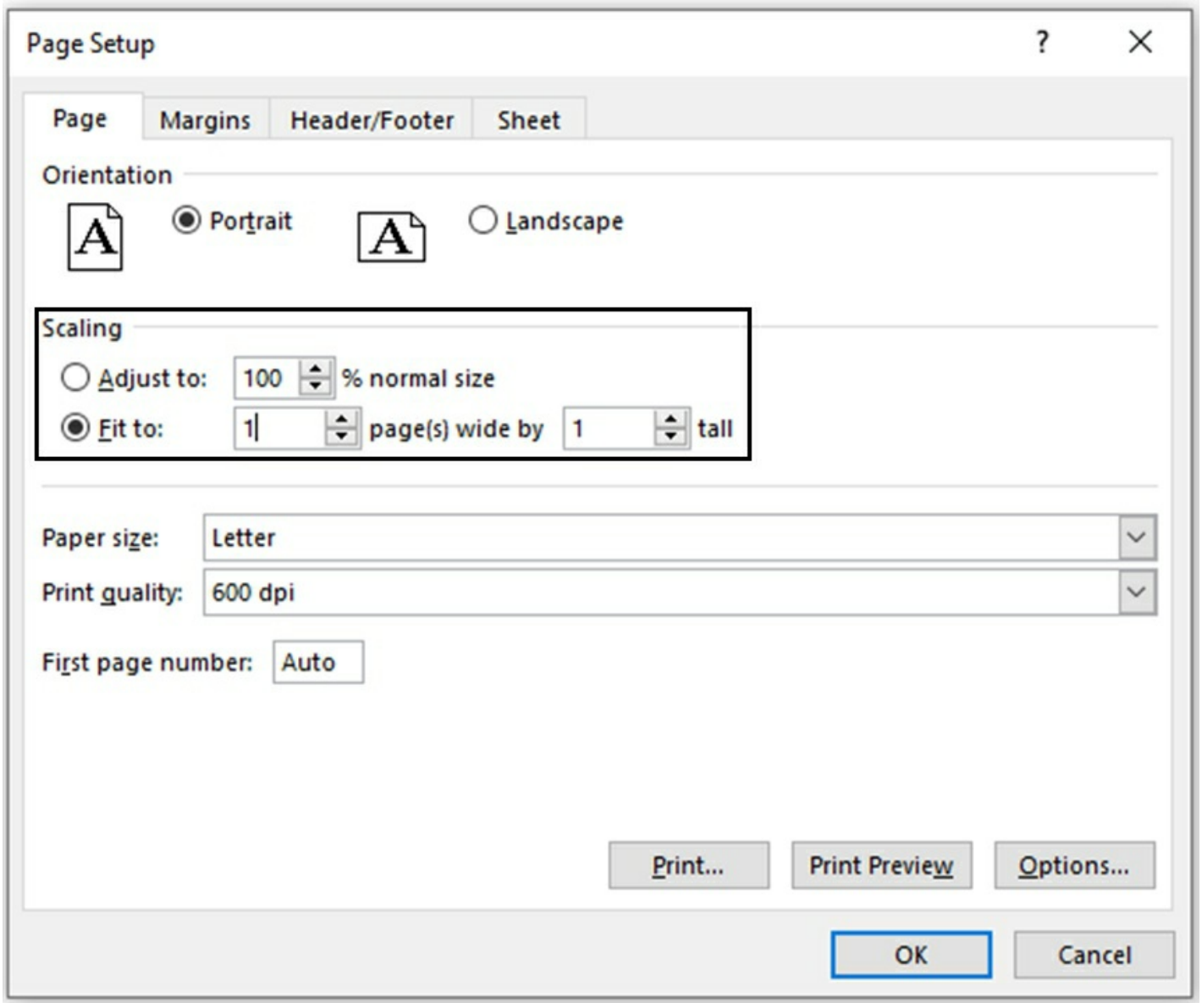


To change the *Page Orientation*, click on the *Page Layout* tab and under the *Page Setup* group, click *Orientation*. You have an option to choose from *Portrait* and *Landscape* orientations. For Sales Dashboard, set *Orientation* to *Landscape*.



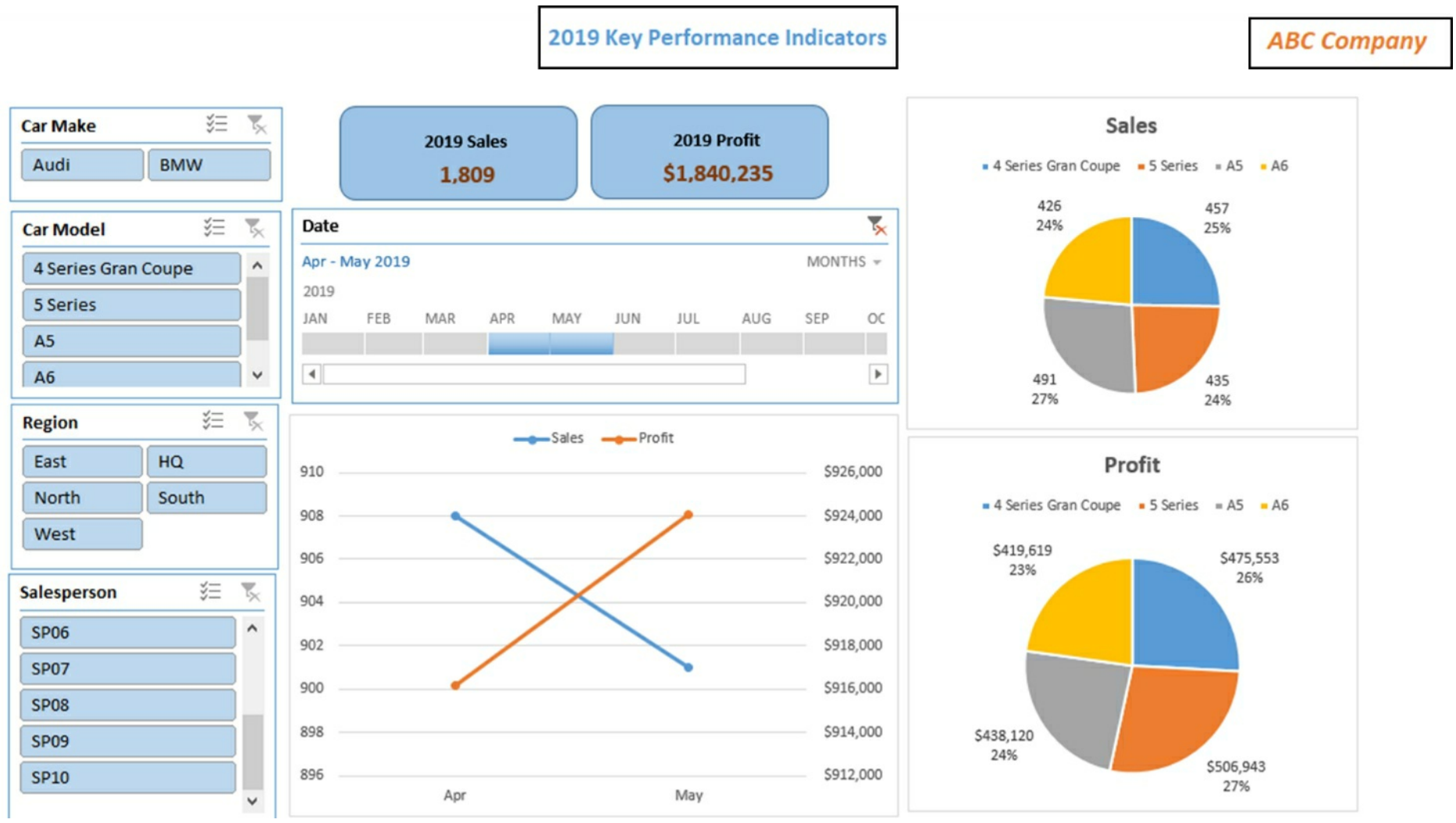
3. Use Scale to Fit

If your dashboard is just over one page, scale it to fit on one sheet. Go to the *Page Layout* tab and under the *Scale to Fit* group, click on the expansion arrow in the bottom right corner. A *Page Setup* window will pop up. In the window, under *Scaling*, select *Fit to*. By default, 1 is specified for both page width and length. If not, change both to 1 and click *OK*.



4. Utilize headers and footers to add context

You can include the title of the dashboard, company name etc. as header and author's name, date of dashboard update or succinct footnotes etc. as footers. The easiest way to do so is to view your dashboard in Page Layout and click the left, center, or right header or footer text box at the top or the bottom of the page to add or edit corresponding text.



5. Adjust the print area

By default, Excel prints the entire dashboard placed in your spreadsheet. If you want to only print a certain part of your dashboard, you'll need to adjust the print area. First, select the content you want to print. Next, go to the *Page Layout* tab and under the *Page Setup* group, click on *Print Area* and then *Set Print Area*.

Only the selected section of the dashboard will now print. If you want to print the entire dashboard, you will have to first clear the previous print area using the *Print Area* option.

Congratulations on finishing this book! Give yourself pat on the back and call it a day! Remember to practice, practice and practice to keep improving your skills in Data Analysis and Reporting with Excel. Please also remember to recommend the book to other keen learners!

---

[1] Garrett, N. (2015). Textbooks for Responsible Data Analysis in Excel. Journal of Education for Business