BUSINESS INTELLIGENCE

By Kushal P Wade

kushal.1889@gmail.com

BI Basics (Business Intelligence Basics):

- Business Intelligence (BI) refers to the use of technologies, applications, and practices to collect, integrate, analyze, and present business data to support decision-making processes within an organization.
- It is Data-Driven Decision Support System.
- The primary goal of BI is to help business users and decision-makers gain actionable insights from data and improve overall business performance.

kushal.1889@gmail.com

BI Basics
(Business Intelligence Basics):

kushal.1889@gmail.com

Information Gathering :

- Information gathering is the process of collecting relevant data from various sources to create a comprehensive understanding of a particular subject or problem.
- In the context of BI, information gathering involves identifying and extracting data from internal and external sources, including databases, spreadsheets, cloud services, websites, and more.
- The data collected must be accurate, relevant, and up-to-date to ensure meaningful analysis.

kushal.1889@gmail.com

Decision Making:

- BI plays a crucial role in enhancing decision-making processes within an organization.
- By providing accurate and timely insights, BI empowers decision-makers to make informed choices based on data-driven evidence rather than relying solely on intuition.
- BI tools often include interactive dashboards, ad-hoc querying capabilities, and data visualization to support decision-making at various levels of the organization.

kushal.1889@gmail.com

THE 2 SYSTEMS

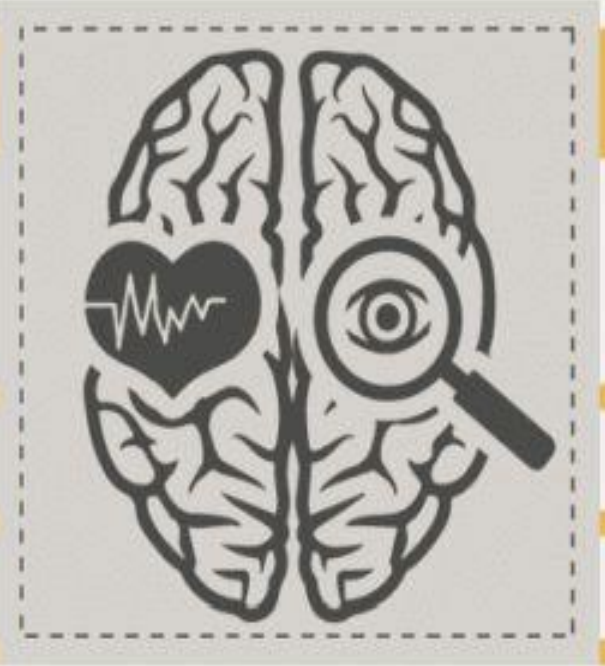
READINGRAPHICS
ACTIONABLE INSIGHTS IN ONE PAGE

System 1 (Fast Thinking)

Continuously scans our environment.

Fast but error-prone

Works automatically & effortlessly via shortcuts, **impulses and intuition.**



System 2 (Slow Thinking)

Used for specific problems, **only if necessary**

Takes effort to analyze, reason, solve complex problems, **exercise self-control**

Slow but reliable

1. Slow But Accurate

- You have a slow thinking system which you can think of as a spotlight of attention. You use this system when you really focus on and think about a problem step by step. This slow-thinking system is very accurate. However, it can only process so much information at any one time and requires a lot of energy.
- Many of us don't use our slow thinking system for most decisions. Instead, we use our fast system, otherwise known as our intuition.

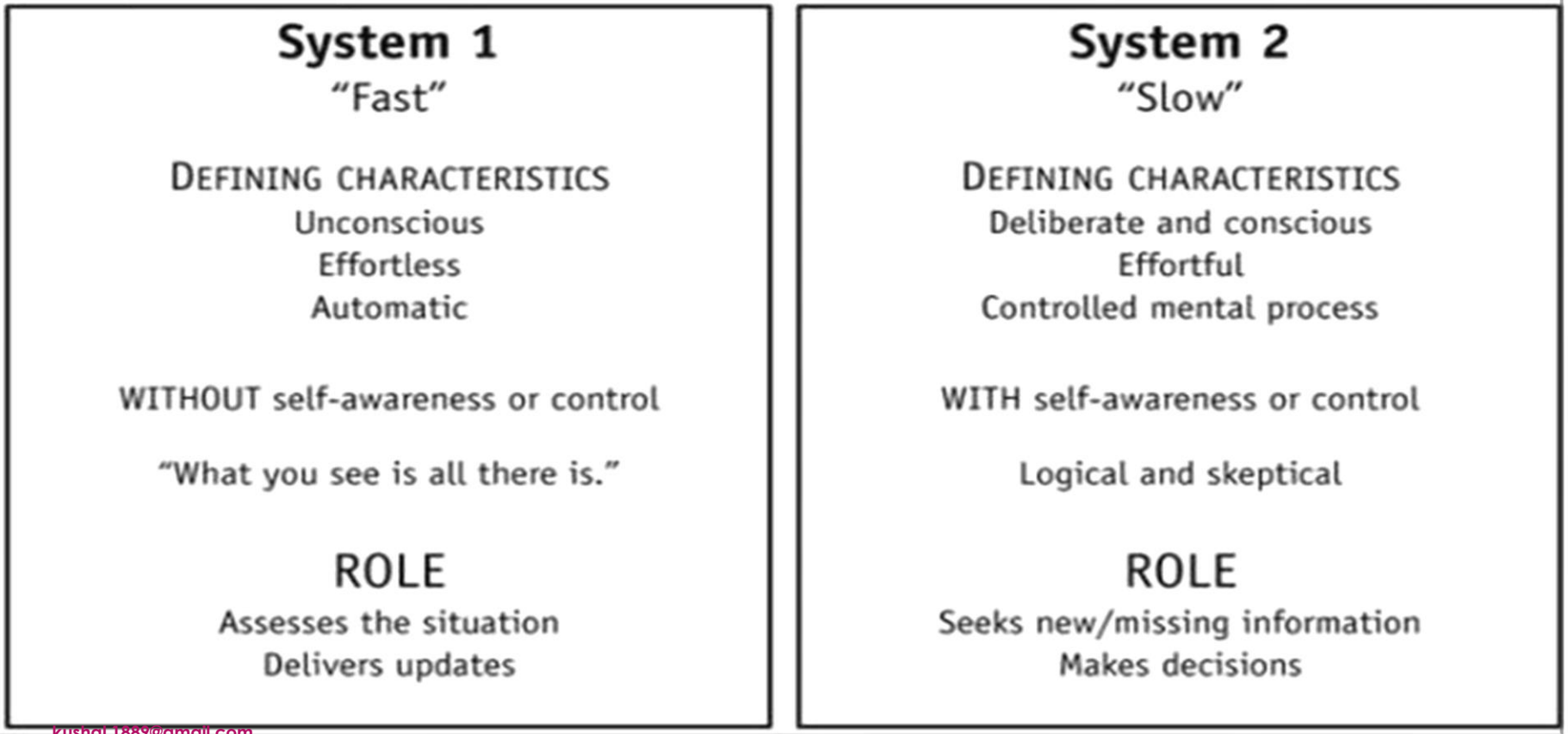
kushal.1889@gmail.com

2. Fast But Error-Prone

- Your fast-thinking system can take in massive amounts of data at once. It can also make snap decisions very quickly, with pretty high accuracy.
- However, your fast-thinking system has one big disadvantage. It is prone to logical fallacies and perceptual biases.
- If your slow thinking system is like a spotlight, your fast thinking system is like the area of a picture that isn't in focus. You don't have to burn much energy to use it, but you don't see things as clearly as you do with your slow-thinking system.

kushal.1889@gmail.com

Figure 1: A Comparison of System 1 and System 2 Thinking



Managing BI:

- Managing BI involves overseeing the entire BI infrastructure, which includes data storage, ETL (Extract, Transform, Load) processes, data warehouses, BI tools, and user access.

- Effective management ensures that data is accurate, secure, and accessible to the right users at the right time.

kushal.1889@gmail.com

User/Market Segmentation:

- BI user segmentation involves categorizing users based on their roles, responsibilities, and information needs within an organization.
- Different user segments may require access to specific data, reports, and dashboards tailored to their job functions.
- Understanding user segments helps BI teams deliver personalized and relevant insights to each group.

kushal.1889@gmail.com

User/Market Segmentation

- While market segmentation divides the entire target market into smaller subsets, customer segmentation takes your existing customer base and divides it into sections based on similar needs and behaviors.

- Keep in mind that customers will fit into more than one segment, depending on the segmentation variables you choose.

kushal.1889@gmail.com

Market Segmentation

Types of Market Segmentation:

- Demographic Segmentation
- Behavioral Segmentation
- Geographic Segmentation
- Psychographic Segmentation

kushal.1889@gmail.com

kushal.1889@gmail.com

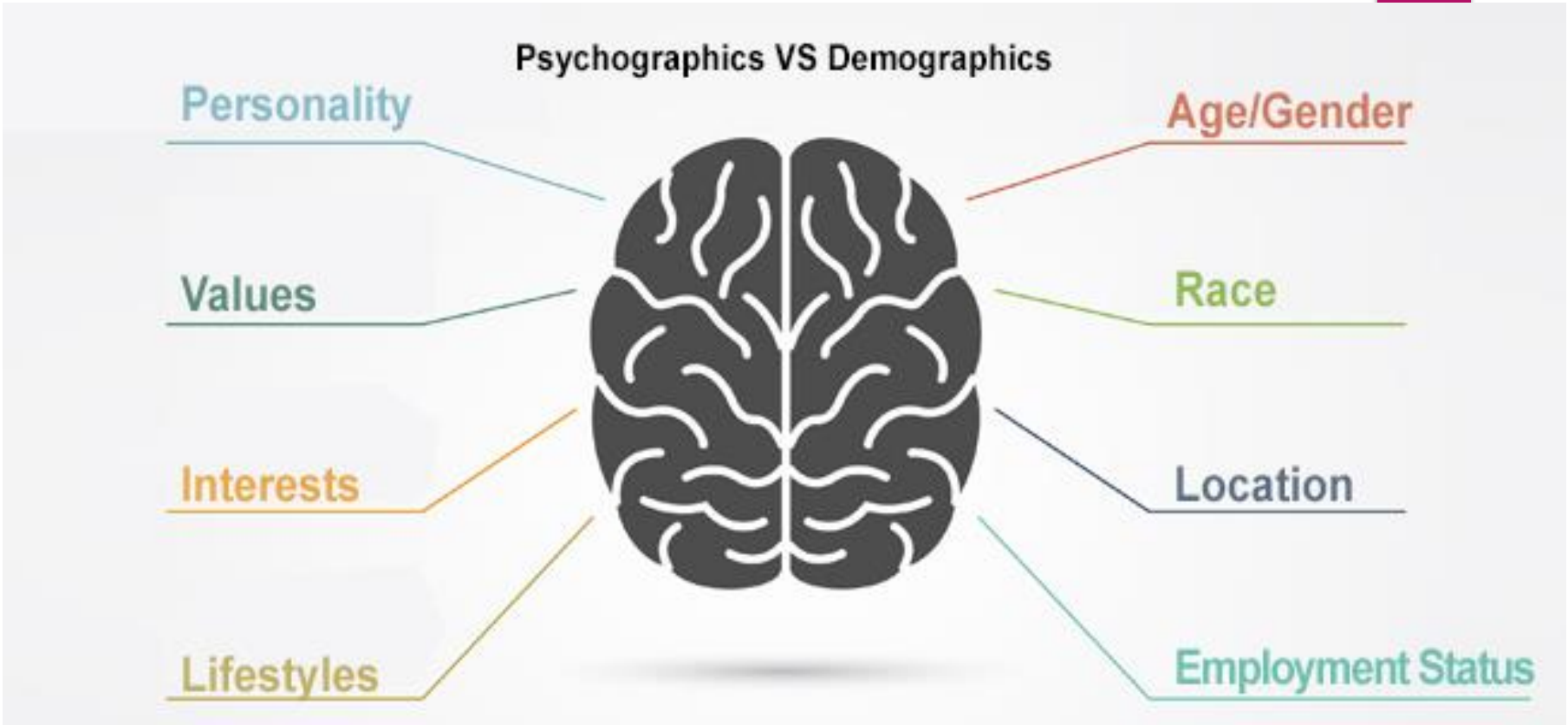
Behavioral Segmentation

- Online shopping habits
- Actions taken on a website
- Benefits sought
- Loyalty

kushal.1889@gmail.com

Geographic Segmentation, splitting up your market based on their location, is a basic but highly useful segmentation strategy

kushal.1889@gmail.com



kushal.1889@gmail.com

Gathering BI Requirements:

- Gathering BI requirements involves understanding the needs and expectations of stakeholders to determine the data and insights they require.

- This process requires close collaboration between business users and BI analysts to identify key performance indicators (KPIs), data dimensions, and visualization preferences to build a BI solution that addresses business needs effectively.

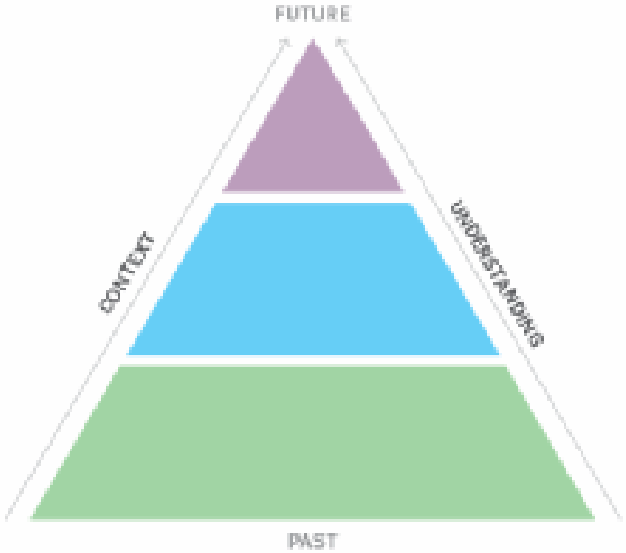
kushal.1889@gmail.com

Content and Knowledge Management:

- Content and knowledge management in BI refer to organizing and maintaining the repository of data, reports, dashboards, and analytical models.
- This ensures that data is consistent, up-to-date, and easily accessible to users.
- Knowledge management involves documenting insights, best practices, and data definitions to share knowledge across the organization.

kushal.1889@gmail.com

Knowledge triangle



- **KNOWLEDGE:** Awareness and understanding of information, gained through experience.
- **INFORMATION:** Data that has been organized and contextualized, often stored in documents and other files, such as invoices.
- **DATA:** Raw facts and figures, such as names and numbers in spreadsheets.

kushal.1889@gmail.com

A strategic approach to BI:

- A strategic approach to BI involves aligning BI initiatives with an organization's overall business objectives.
- It requires defining clear goals, identifying key performance indicators, and prioritizing projects that will have the most significant impact on the organization's success.
- A strategic approach helps ensure that BI efforts contribute to the organization's growth and competitive advantage.

kushal.1889@gmail.com

The significance of visual analytics:

- Visual analytics is a critical component of BI that uses graphical representations, charts, and interactive dashboards to present data visually.
- The significance of visual analytics lies in its ability to make complex data more understandable and accessible to a broader audience.
- Visualizations help identify patterns, trends, and outliers quickly; enabling better and faster decision-making.

kushal.1889@gmail.com

Information Visualization:

- Information visualization is the process of transforming raw data into graphical representations such as charts, graphs, heatmaps, and other visual elements.

- The goal is to make data more intuitive and easier to comprehend, allowing users to gain insights at a glance.

kushal.1889@gmail.com

Data representation :

- Data representation refers to the various formats used to present data.
- It can be numerical (e.g., tables), visual (e.g., charts), or textual (e.g., summaries).
- The choice of data representation depends on the nature of the data and the insights sought by users.

kushal.1889@gmail.com

Data Collection & Binding :

- Data collection involves gathering data from multiple sources and integrating it into a centralized repository, often a data warehouse, to create a unified view.

- Data binding refers to the process of connecting the data to the BI tools and visualizations, enabling users to interact with the data and extract meaningful insights.

kushal.1889@gmail.com

Structured Data :

- Structured data is organized and formatted in a consistent manner, making it easily searchable and analysable.

- It fits neatly into predefined data models and is usually stored in relational databases or spreadsheets.

kushal.1889@gmail.com

Unstructured Data :

- Unstructured data lacks a predefined data model and doesn't fit neatly into traditional databases.
- It includes things like text documents, emails, social media posts, images, and videos.
- Analyzing unstructured data requires specialized tools like natural language processing and machine learning algorithms to extract valuable insights.

kushal.1889@gmail.com

Pivot Table:

- PivotTable is an extremely powerful tool that you can use to slice and dice data.
- Knowledge of PivotTable features helps us to start with exploring, analyzing, and reporting data based on the requirements.

kushal.1889@gmail.com

Pivot Table:

- A PivotTable is an interactive way to quickly summarize large amounts of data. You can use a PivotTable to analyze numerical data in detail and answer unanticipated questions about your data.

- A PivotTable is specially designed for Querying large amounts of data in many user-friendly ways.

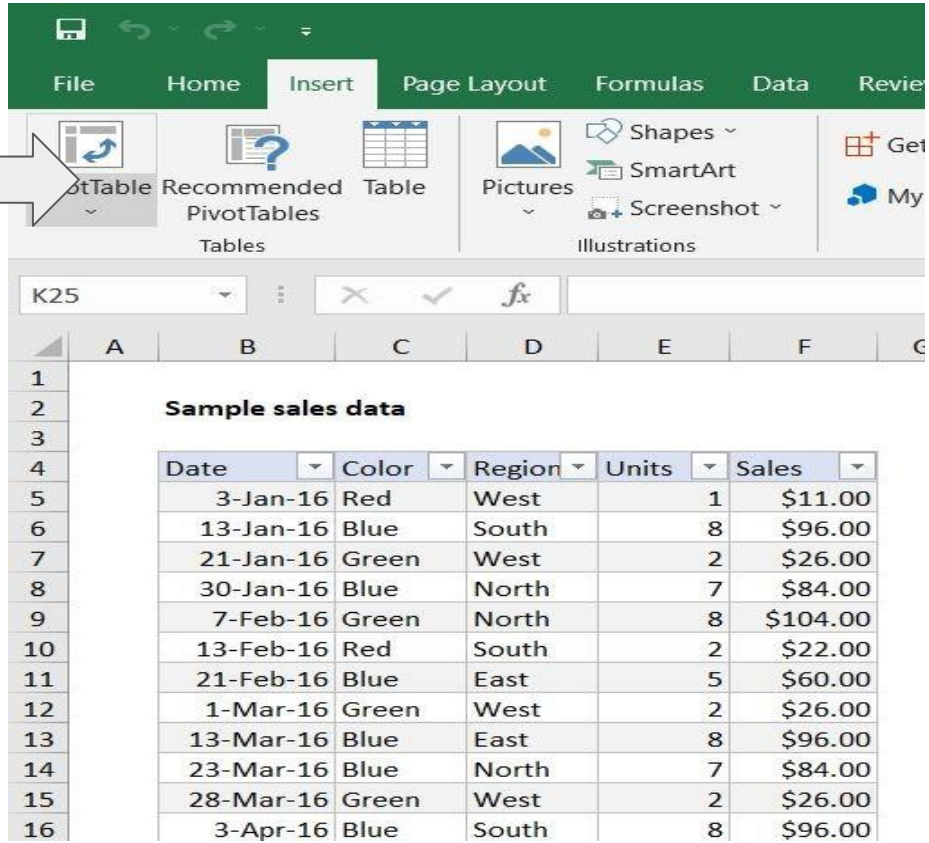
kushal.1889@gmail.com

Features of Pivot Table:

- Creating a PivotTable is extremely simple and fast
- Enabling churning of data instantly by simple dragging of fields, sorting and filtering and different calculations on the data.
- Arriving at the suitable representation for your data as you gain insights into it.
- Ability to create reports on the fly.
- Producing multiple reports from the same PivotTable in a matter of seconds.
- Providing interactive reports to synchronize with the audience.

Pivot Table

1. To start off, select *any cell in the data* and click Pivot Table on the Insert tab of the ribbon:

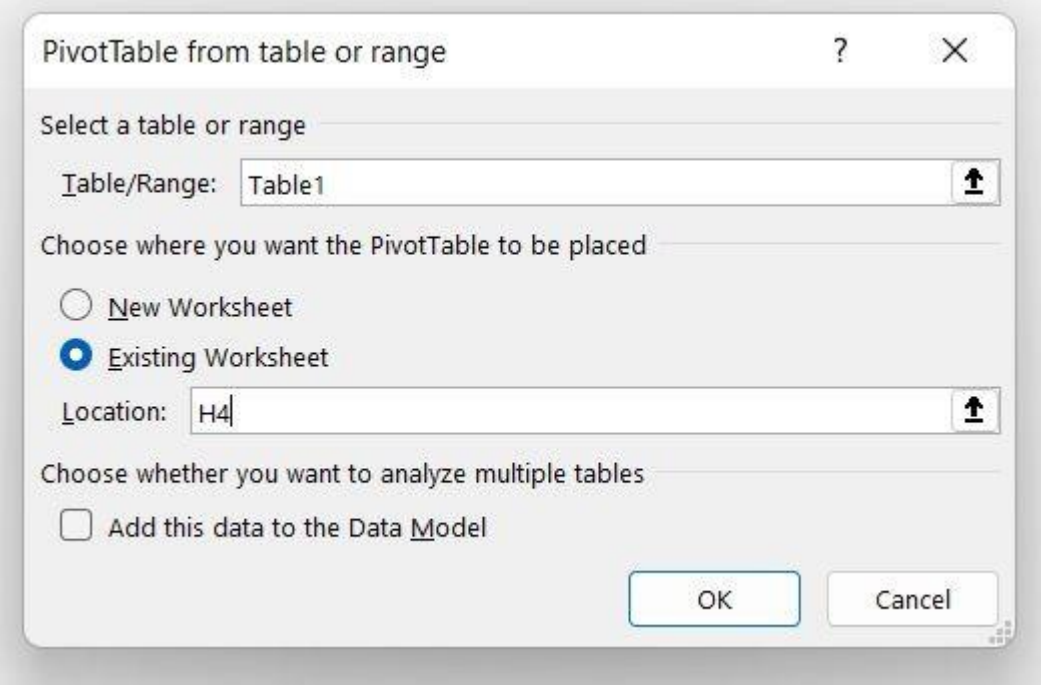


Sample sales data

Date	Color	Region	Units	Sales
3-Jan-16	Red	West	1	$11.00
13-Jan-16	Blue	South	8	$96.00
21-Jan-16	Green	West	2	$26.00
30-Jan-16	Blue	North	7	$84.00
7-Feb-16	Green	North	8	$104.00
13-Feb-16	Red	South	2	$22.00
21-Feb-16	Blue	East	5	$60.00
1-Mar-16	Green	West	2	$26.00
13-Mar-16	Blue	East	8	$96.00
23-Mar-16	Blue	North	7	$84.00
28-Mar-16	Green	West	2	$26.00
3-Apr-16	Blue	South	8	$96.00

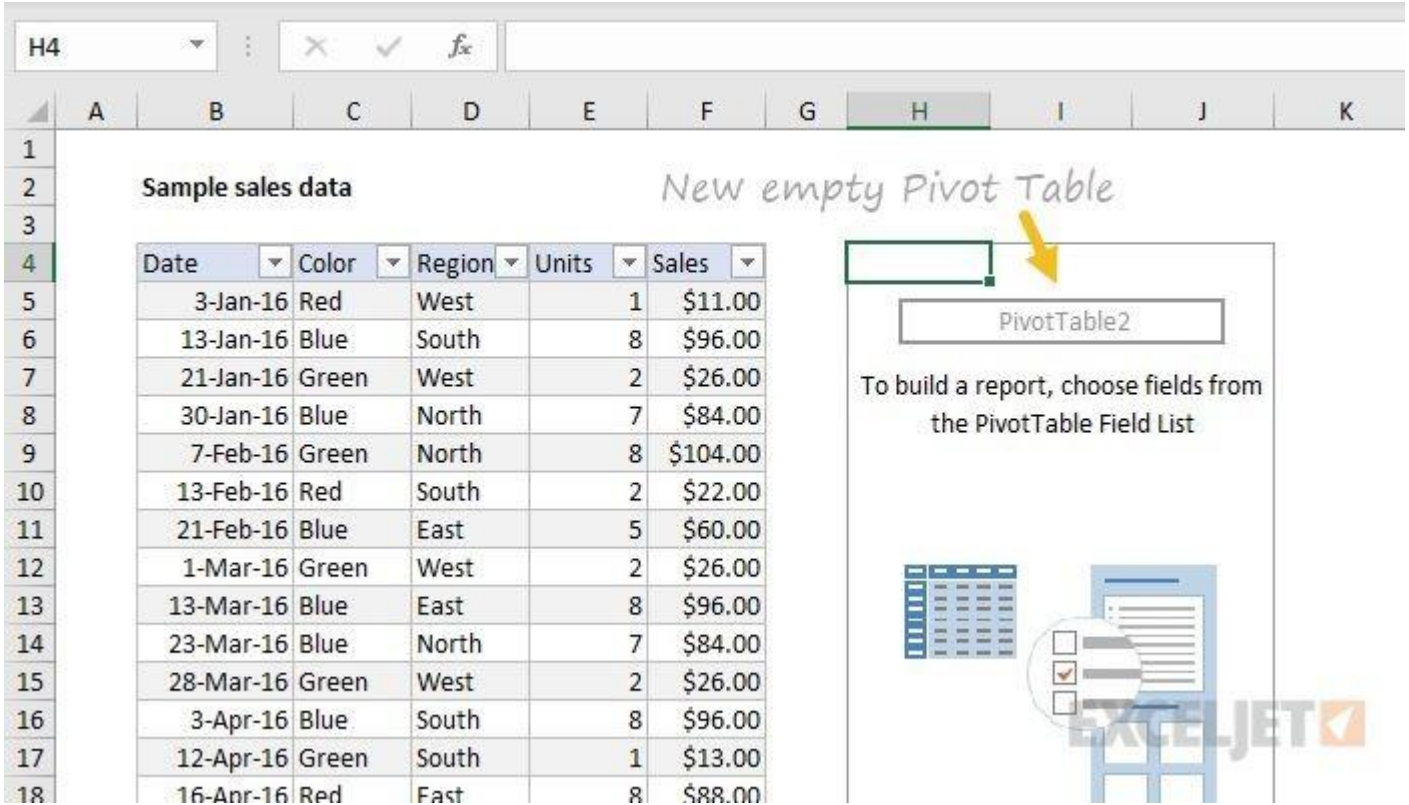
The sample data contains 452 records with 5 fields of information: Date, Color, Units, Sales, and Region. This data is perfect for a pivot table.

2. Override the default location and enter H4 to place the pivot table on the current worksheet:



Excel will display the Create Pivot Table window. Notice the data range is already filled in. The default location for a new pivot table is New Worksheet.

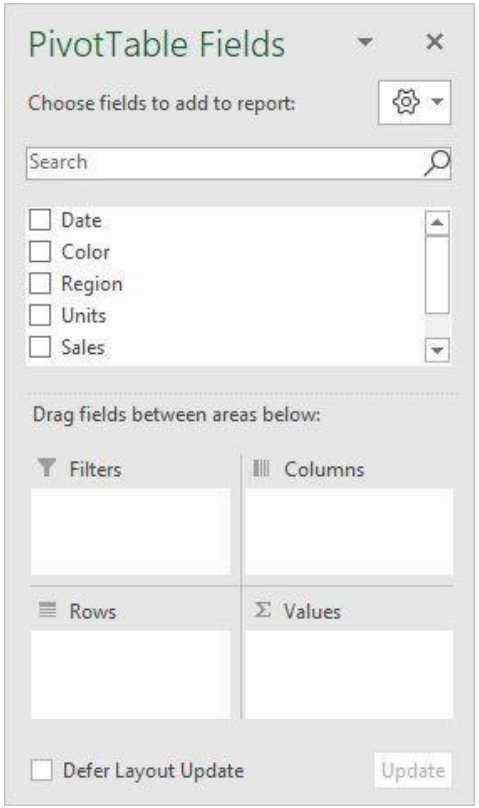
3. Click OK, and Excel builds an empty pivot table starting in cell H4.



Date	Color	Region	Units	Sales
3-Jan-16	Red	West	1	$11.00
13-Jan-16	Blue	South	8	$96.00
21-Jan-16	Green	West	2	$26.00
30-Jan-16	Blue	North	7	$84.00
7-Feb-16	Green	North	8	$104.00
13-Feb-16	Red	South	2	$22.00
21-Feb-16	Blue	East	5	$60.00
1-Mar-16	Green	West	2	$26.00
13-Mar-16	Blue	East	8	$96.00
23-Mar-16	Blue	North	7	$84.00
28-Mar-16	Green	West	2	$26.00
3-Apr-16	Blue	South	8	$96.00
12-Apr-16	Green	South	1	$13.00
16-Apr-16	Red	East	8	$88.00

Excel also displays the PivotTable Fields pane, which is empty at this point. Note all five fields are listed, but unused:

To build a pivot table, drag fields into one the Columns, Rows, or Values area. The Filters area is used to apply global filters to a pivot table.

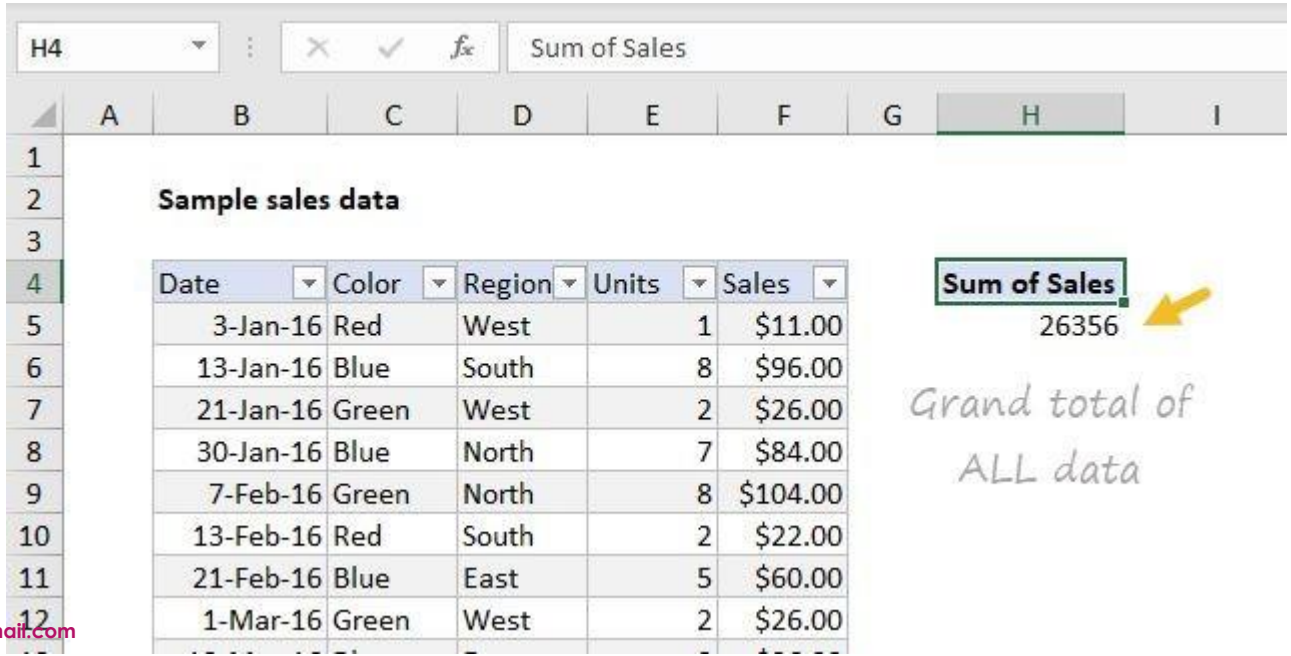


kushal.1889@gmail.com

Add fields

1. Drag the Sales field to the Values area.

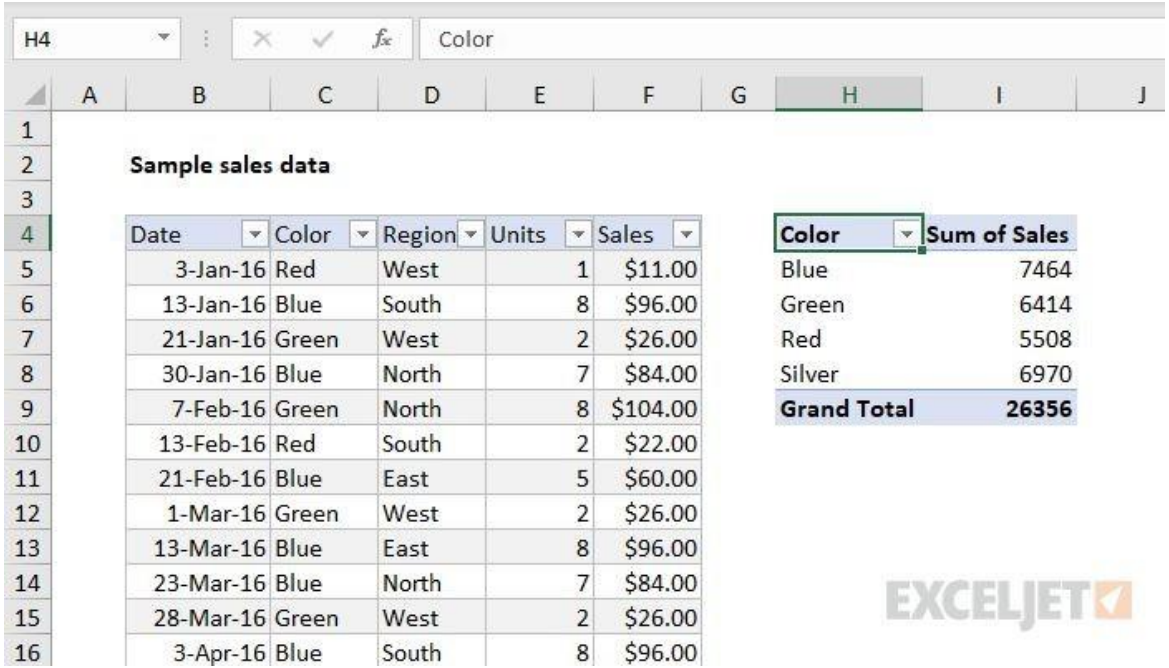
Excel calculates a grand total, 26356. This is the sum of all sales values in the entire data set:



2. Drag the Color field to the Rows area.

Excel breaks out sales by Color. You can see Blue is the top seller, while Silver comes in last:

Notice the Grand Total remains 26356. This makes sense, because we are still reporting on the full set of data.

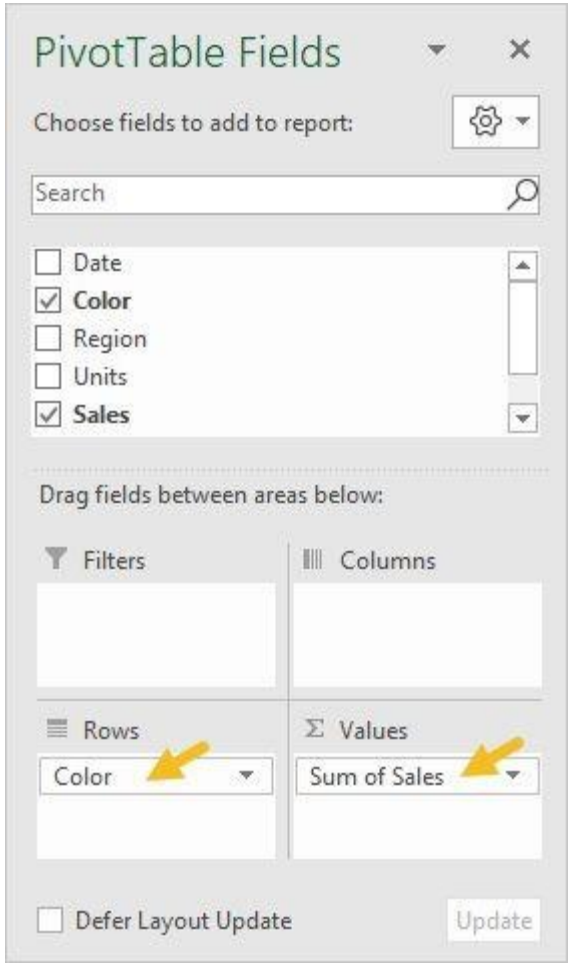


Date	Color	Region	Units	Sales
3-Jan-16	Red	West	1	$11.00
13-Jan-16	Blue	South	8	$96.00
21-Jan-16	Green	West	2	$26.00
30-Jan-16	Blue	North	7	$84.00
7-Feb-16	Green	North	8	$104.00
13-Feb-16	Red	South	2	$22.00
21-Feb-16	Blue	East	5	$60.00
1-Mar-16	Green	West	2	$26.00
13-Mar-16	Blue	East	8	$96.00
23-Mar-16	Blue	North	7	$84.00
28-Mar-16	Green	West	2	$26.00
3-Apr-16	Blue	South	8	$96.00

Color	Sum of Sales
Blue	7464
Green	6414
Red	5508
Silver	6970
Grand Total	**26356**

kushal.1889@gmail.com

You can see Color is a Row field, and Sales is a

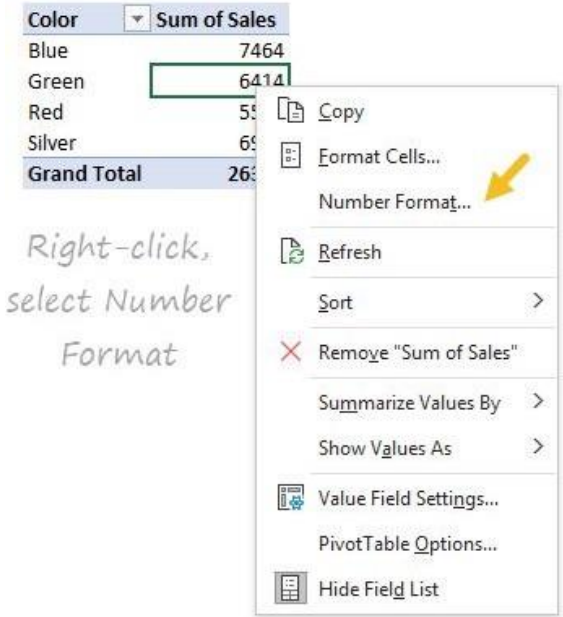


kushal.1889@gmail.com

Number formatting

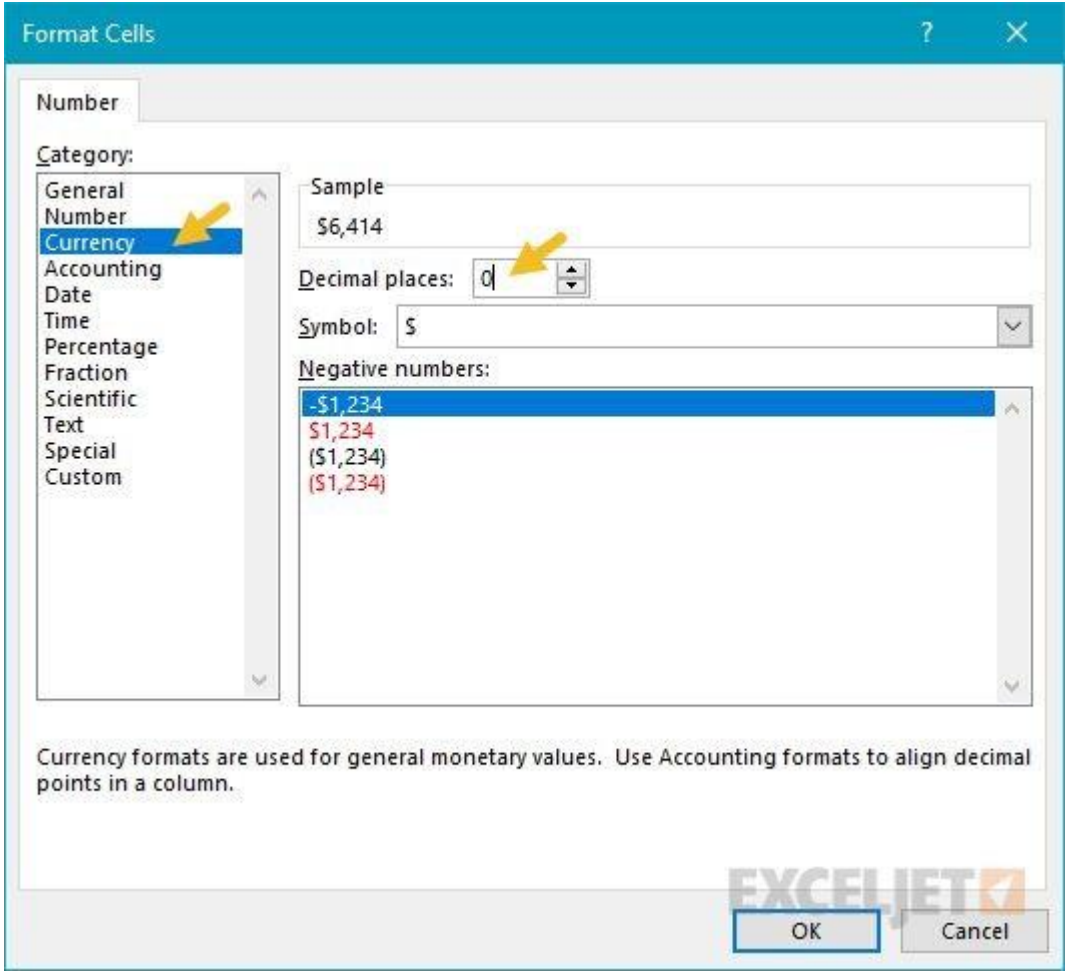
Pivot Tables can apply and maintain number formatting automatically to numeric fields. This is a big time-saver when data changes frequently.

1. Right-click any Sales number and choose Number Format:



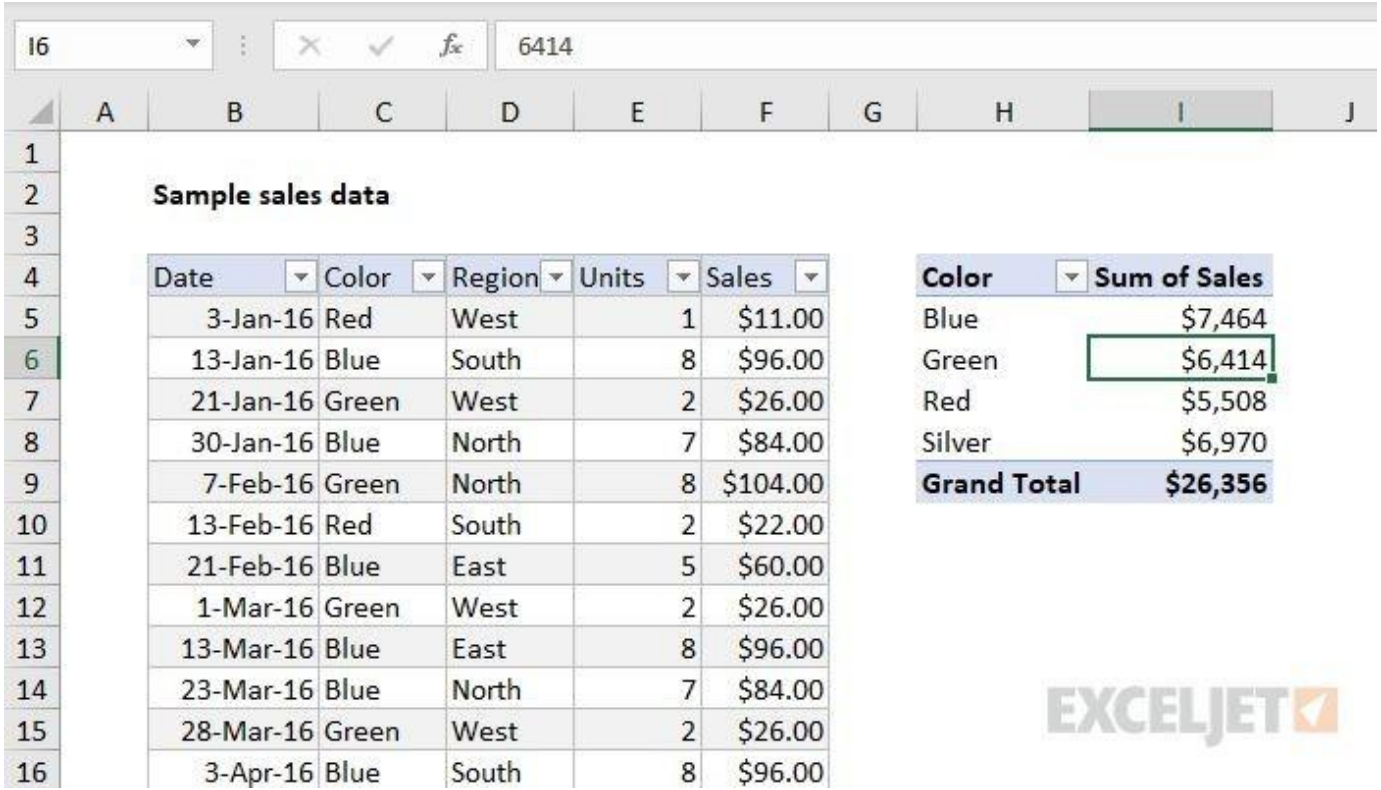
kushal.1889@gmail.com

2. Apply Currency formatting with zero decimal places, the click OK:



kushal.1889@gmail.com

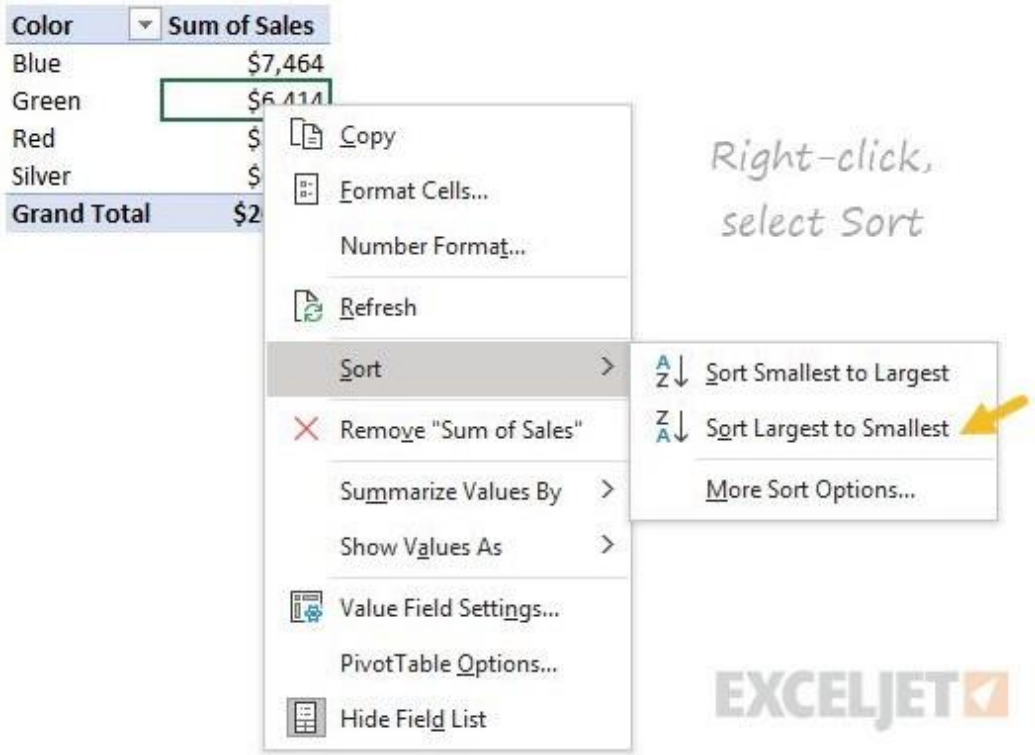
- In the resulting pivot table, all sales values have Currency format applied
- Currency format will continue to be applied to Sales values, even when the pivot table is reconfigured, or new data is added.:



kushal.1889@gmail.com

Sorting by value

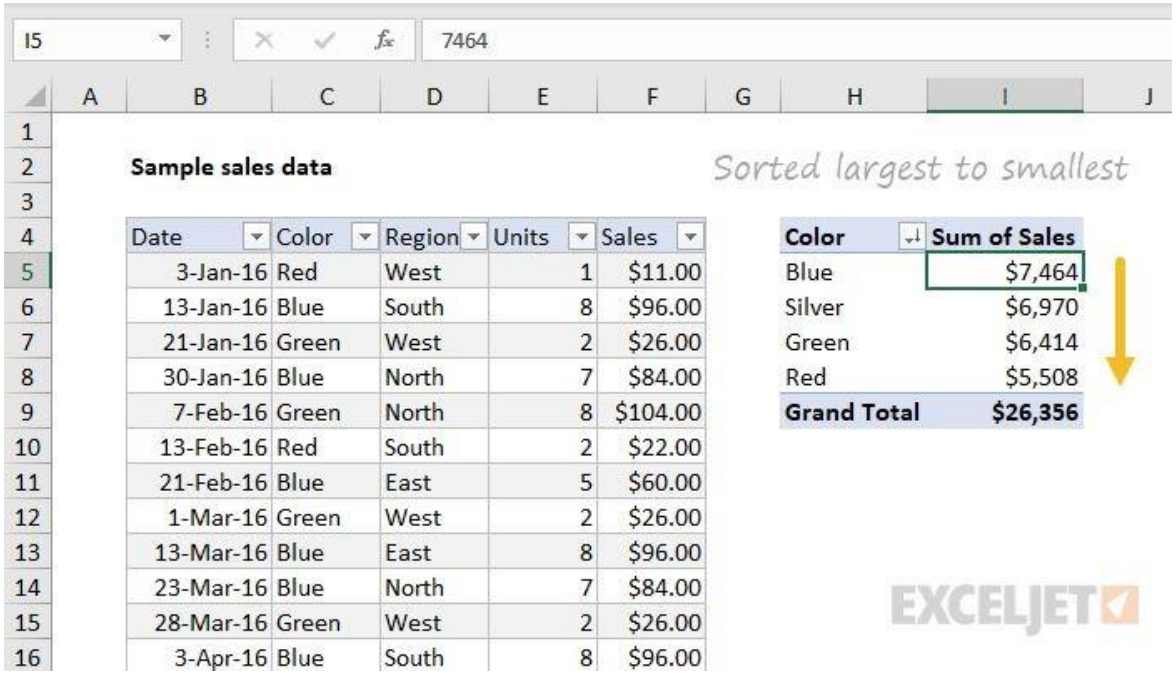
1. Right-click any Sales value and choose Sort > Largest to Smallest.



kushal.1889@gmail.com

Excel now lists top-selling colors first. This sort order will be maintained when data changes, or when the pivot table is reconfigured

.



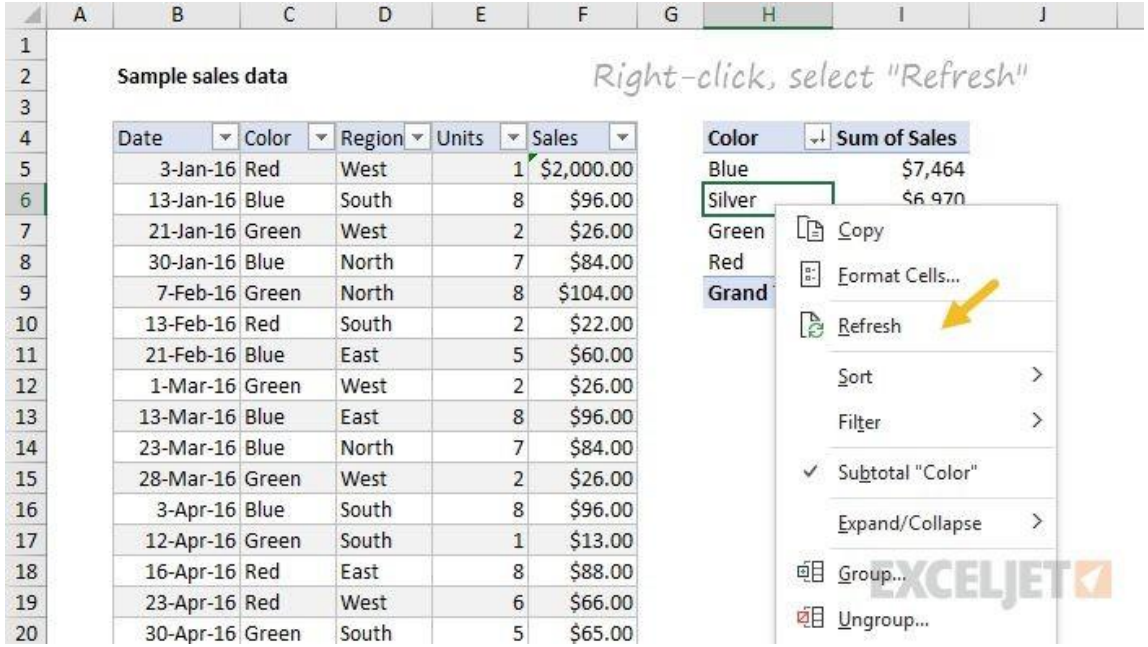
kushal.1889@gmail.com

Refresh data

Pivot table data needs to be "refreshed" in order to bring in updates. To reinforce how this works, we'll make a big change to the source data and watch it flow into the pivot table.

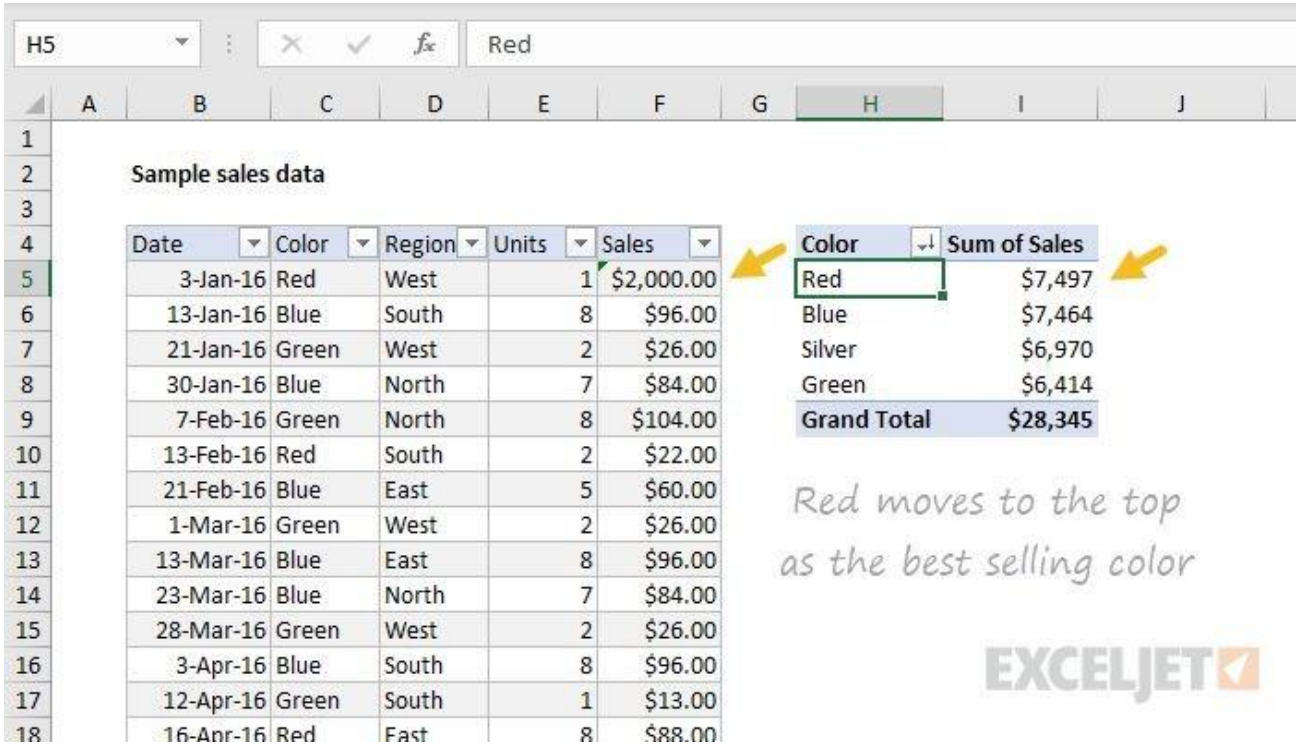
1. Select cell F5 and change $11.00 to $2000.

2. Right-click anywhere in the pivot table and select "Refresh".



kushal.1889@gmail.com

Notice "Red" is now the top selling color, and automatically moves to the top:



kushal.1889@gmail.com

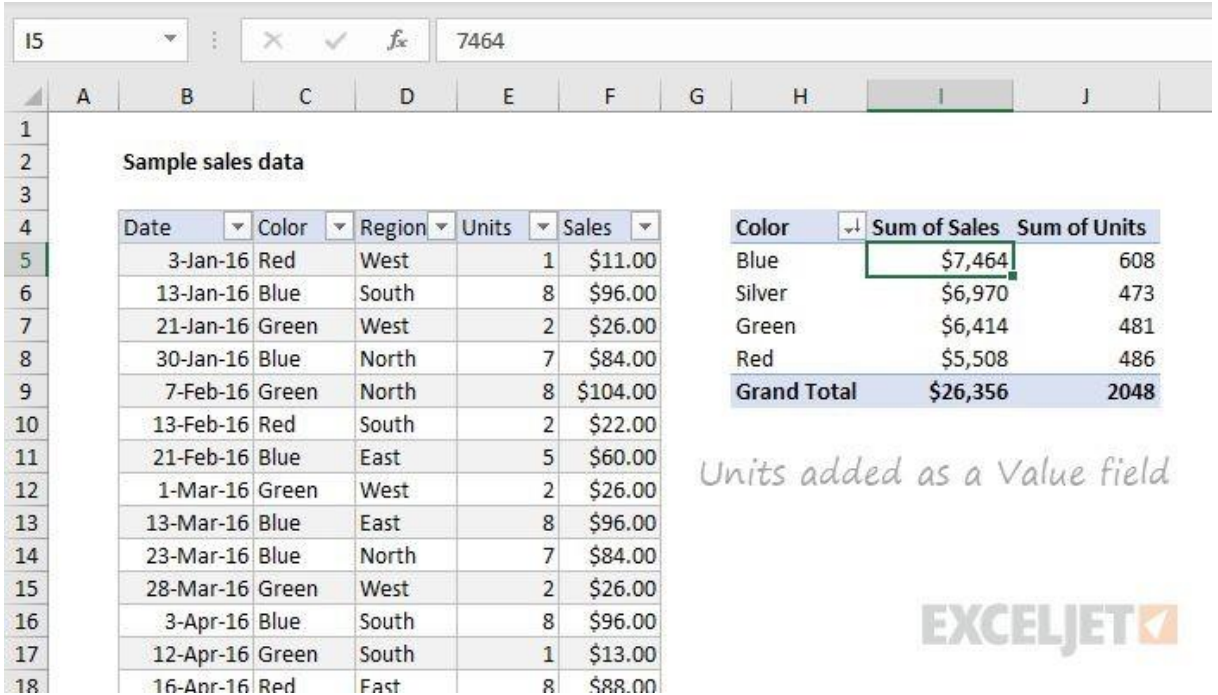
3. Change F5 back to $11.00 and refresh the pivot again.

Note: changing F5 to $2000 is not realistic, but it's a good way to force a change you can easily see in the pivot table. Try changing an existing color to something new, like "Gold" or "Black". When you refresh, you'll see the new color appear. You can use undo to go back to original data and pivot.

Second value field

You can add more than one field as a Value field.

1. Drag Units to the Value area to see Sales and Units together:



Date	Color	Region	Units	Sales
3-Jan-16	Red	West	1	$11.00
13-Jan-16	Blue	South	8	$96.00
21-Jan-16	Green	West	2	$26.00
30-Jan-16	Blue	North	7	$84.00
7-Feb-16	Green	North	8	$104.00
13-Feb-16	Red	South	2	$22.00
21-Feb-16	Blue	East	5	$60.00
1-Mar-16	Green	West	2	$26.00
13-Mar-16	Blue	East	8	$96.00
23-Mar-16	Blue	North	7	$84.00
28-Mar-16	Green	West	2	$26.00
3-Apr-16	Blue	South	8	$96.00
12-Apr-16	Green	South	1	$13.00
16-Apr-16	Red	East	8	$88.00

Color	Sum of Sales	Sum of Units
Blue	$7,464	608
Silver	$6,970	473
Green	$6,414	481
Red	$5,508	486
Grand Total	**$26,356**	**2048**

kushal.1889@gmail.com

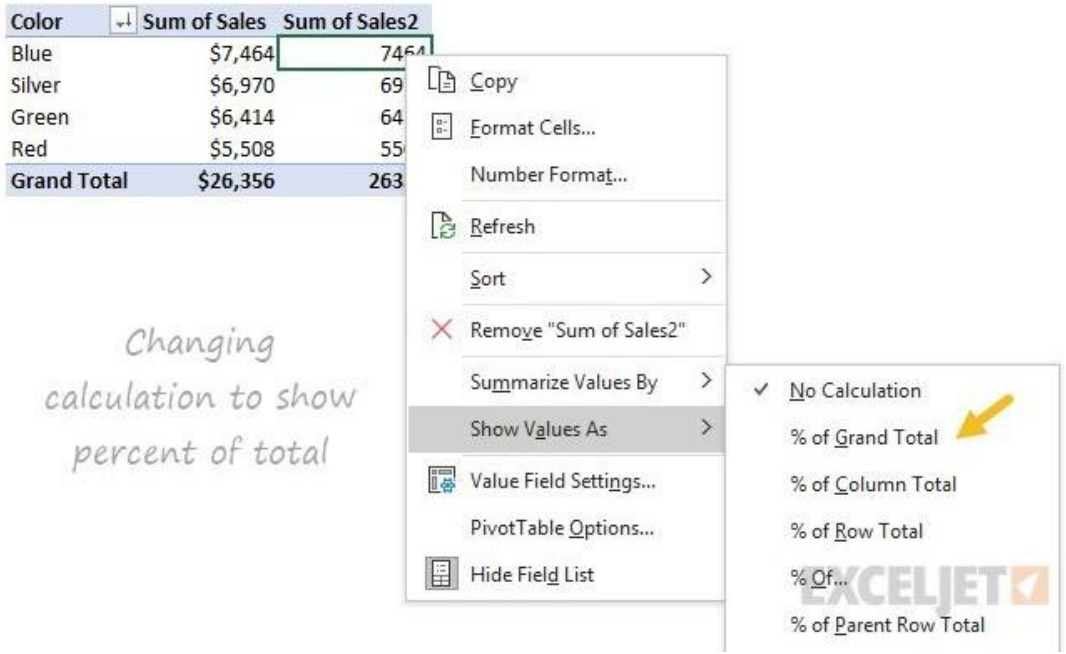
Percent of total

There are different ways to display values. One option is to show values as a percent of total. If you want to display the same field in different ways, add the field twice.

1. Remove the Units from the Values area

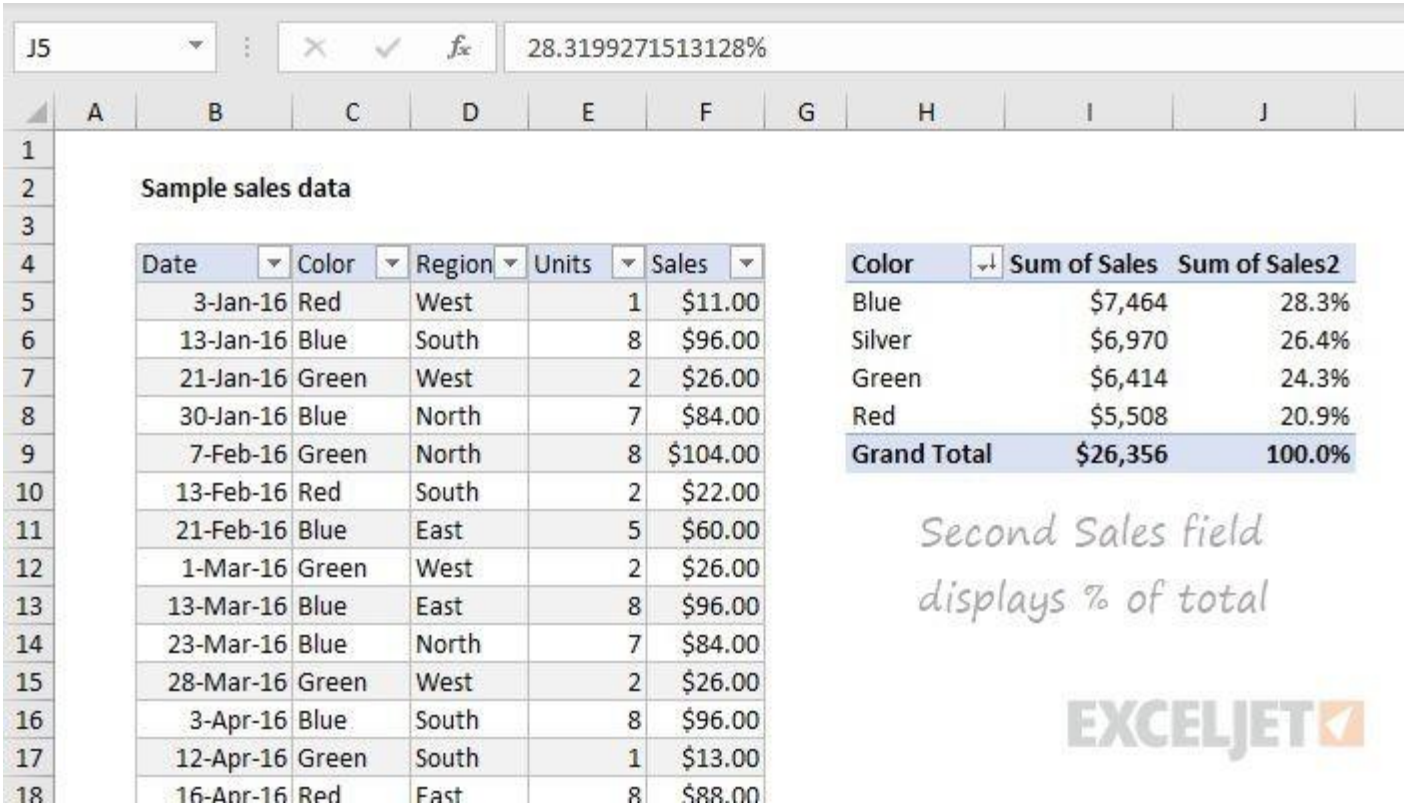
2. Add the Sales field (again) to the Values area.

3. Right-click the second instance and choose "% of grand total":



kushal.1889@gmail.com

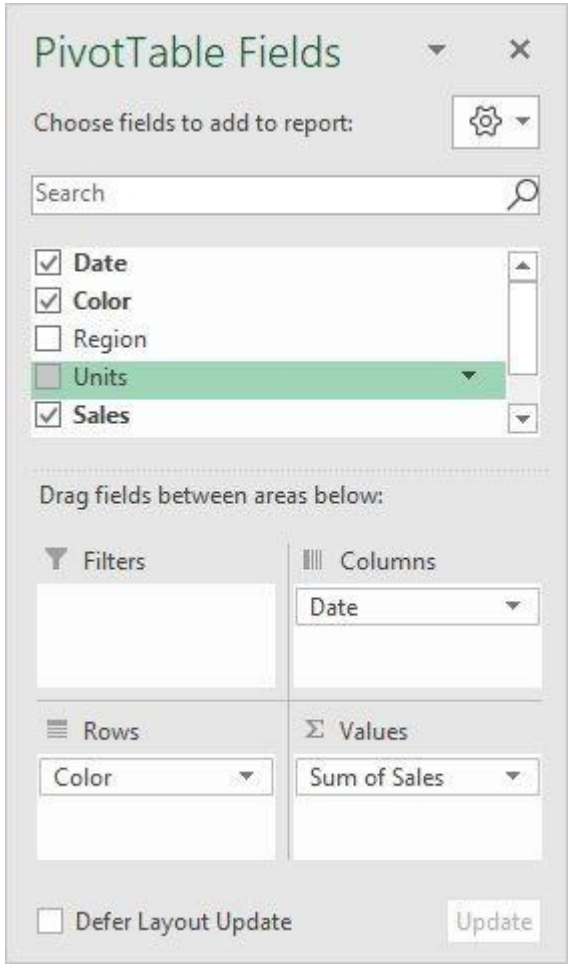
The result is a breakdown by color along with a percent of total:



Date	Color	Region	Units	Sales
3-Jan-16	Red	West	1	$11.00
13-Jan-16	Blue	South	8	$96.00
21-Jan-16	Green	West	2	$26.00
30-Jan-16	Blue	North	7	$84.00
7-Feb-16	Green	North	8	$104.00
13-Feb-16	Red	South	2	$22.00
21-Feb-16	Blue	East	5	$60.00
1-Mar-16	Green	West	2	$26.00
13-Mar-16	Blue	East	8	$96.00
23-Mar-16	Blue	North	7	$84.00
28-Mar-16	Green	West	2	$26.00
3-Apr-16	Blue	South	8	$96.00
12-Apr-16	Green	South	1	$13.00
16-Apr-16	Red	East	8	$88.00

Color	Sum of Sales	Sum of Sales2
Blue	$7,464	28.3%
Silver	$6,970	26.4%
Green	$6,414	24.3%
Red	$5,508	20.9%
Grand Total	**$26,356**	**100.0%**

Note: the number format for percentage has also been adjusted to show 1 decimal.



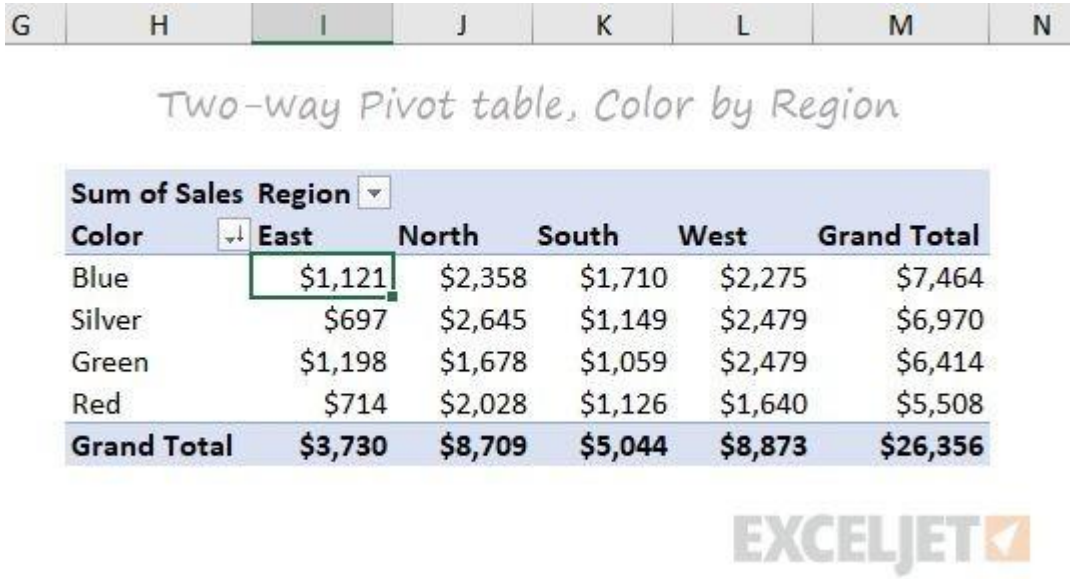
Two-way Pivot

Pivot tables can plot data in various two-dimensional arrangements.

1. Drag the Date field out of the columns area

2. Drag Region into the Columns area.

Excel builds a two-way pivot table that breaks down sales by color and region:



kushal.1889@gmail.com

3. Swap Region and Color (i.e. drag Region to the Rows area and Color to the Columns area).

Excel builds another two-dimensional pivot table:

Two-way Pivot table, Region by Color

Sum of Sales	Color				
Region	Blue	Silver	Green	Red	Grand Total
East	$1,121	$697	$1,198	$714	$3,730
North	$2,358	$2,645	$1,678	$2,028	$8,709
South	$1,710	$1,149	$1,059	$1,126	$5,044
West	$2,275	$2,479	$2,479	$1,640	$8,873
Grand Total	**$7,464**	**$6,970**	**$6,414**	**$5,508**	**$26,356**

Again notice total sales ($26,356) is the same in *all pivot tables above*. Each table presents a different view of the *same data*, so they all sum to the *same total*.

The above example shows how quickly you can build different pivot tables from the same data. You can create many other kinds of pivot tables, using all kinds of data.

Key Pivot Table benefits

- **Simplicity:** Basic pivot tables are very simple to set up and customize. There is no need to learn complicated formulas

- **Speed:** You can create a good-looking, useful report with a pivot table in minutes. Even if you are very good with formulas, pivot tables are faster to set up and require much less effort.

- **Flexibility:** Unlike formulas, pivot tables don't lock you into a particular view of your data. You can quickly rearrange the pivot table to suit your needs. You can even clone a pivot table and build a separate view

kushal.1889@gmail.com

Key Pivot Table benefits

- **Accuracy:** As long as a pivot table is set up correctly, you can rest assured results are accurate. In fact, a pivot table will often highlight problems in the data faster than any other tool.
- **Formatting:** A Pivot table can apply automatically apply consistent number and style formatting, even as data changes.
- **Updates:** Pivot tables are designed for ongoing updates. If you base a pivot table on an Excel Table, the table resizes as needed with new data. All you need to do is click Refresh, and your pivot table will show you the latest.

kushal.1889@gmail.com

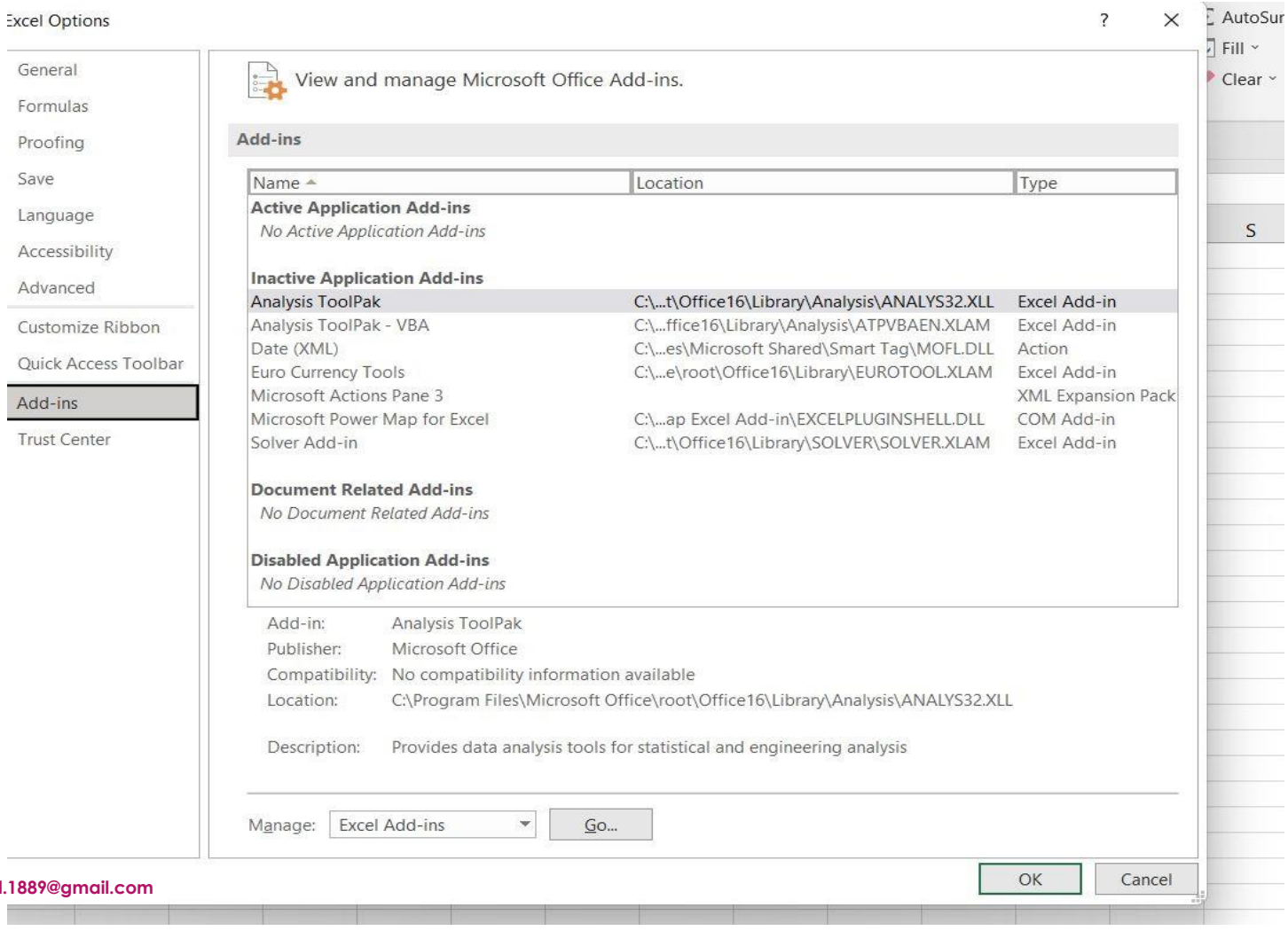
Key Pivot Table benefits

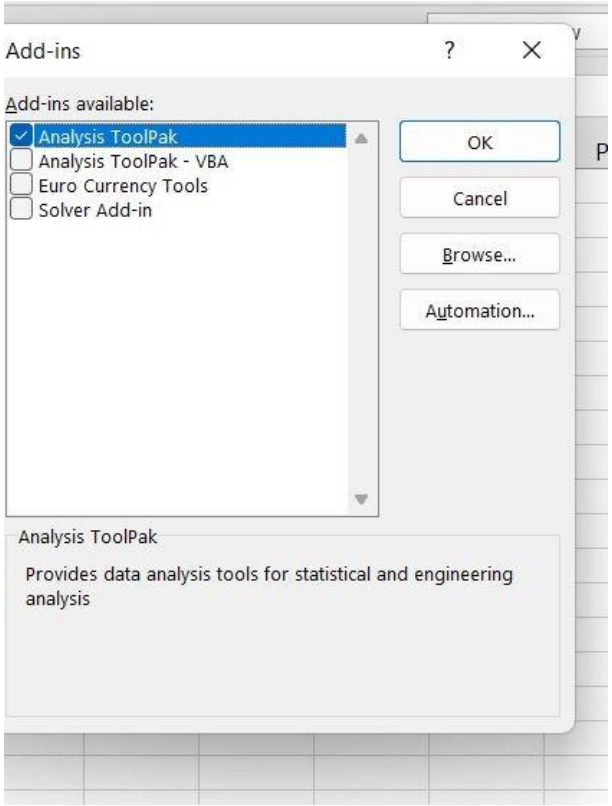
- **Filtering:** Pivot tables contain several tools for filtering data. Need to look at North America and Asia, but exclude Europe? A pivot table makes it simple.
- **Charts:** Once you have a pivot table, you can easily create a pivot chart.

kushal.1889@gmail.com

Analysis ToolPak in Excel

kushal.1889@gmail.com

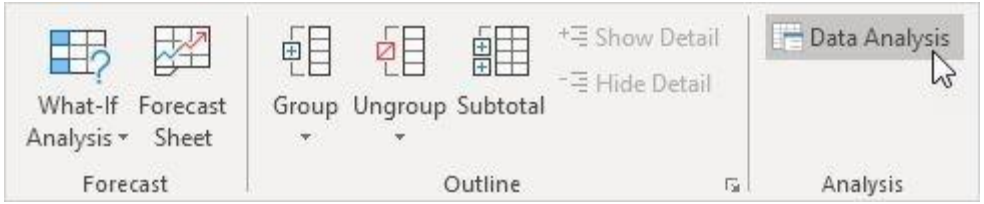




kushal.1889@gmail.com

To use the Analysis Toolpak add-in in Excel to quickly generate correlation coefficients between multiple variables, execute the following steps.

1. On the Data tab, in the Analysis group, click Data Analysis.

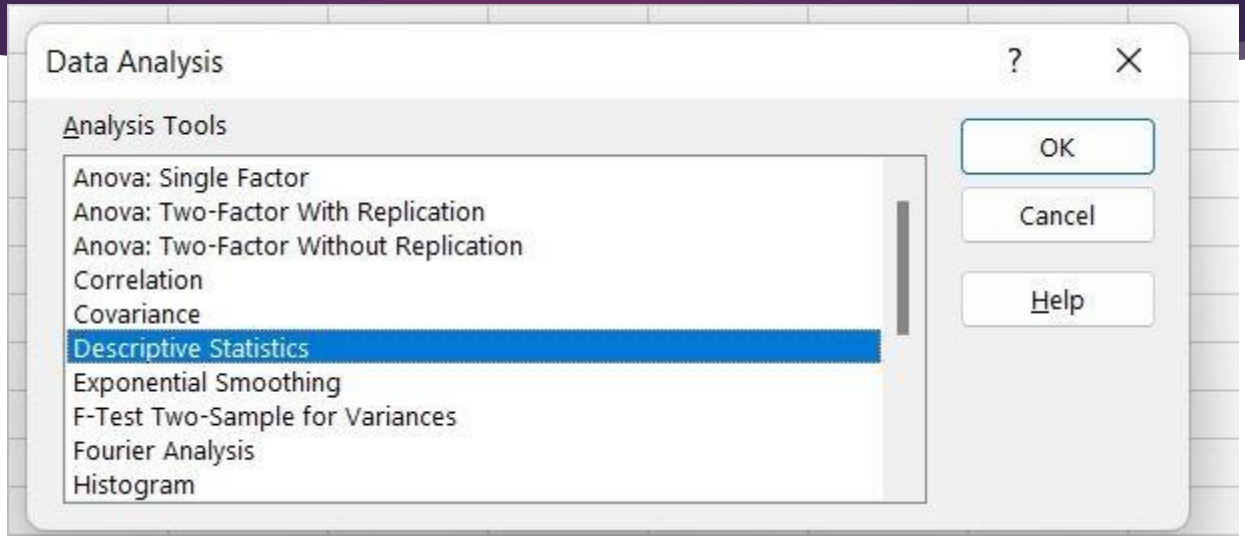


Note: can't find the Data Analysis button? Click here to load the Analysis ToolPak add-in.

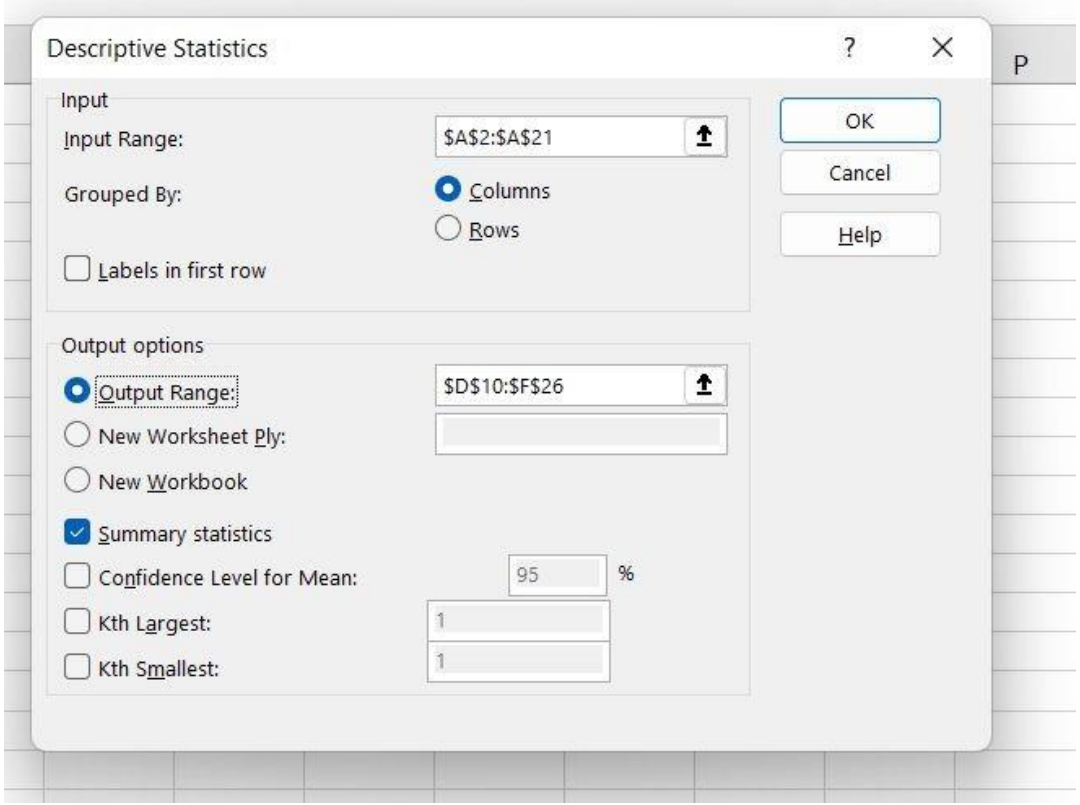
2. Select Correlation and click OK.

kushal.1889@gmail.com

Descriptive Statistics in Excel



kushal.1889@gmail.com



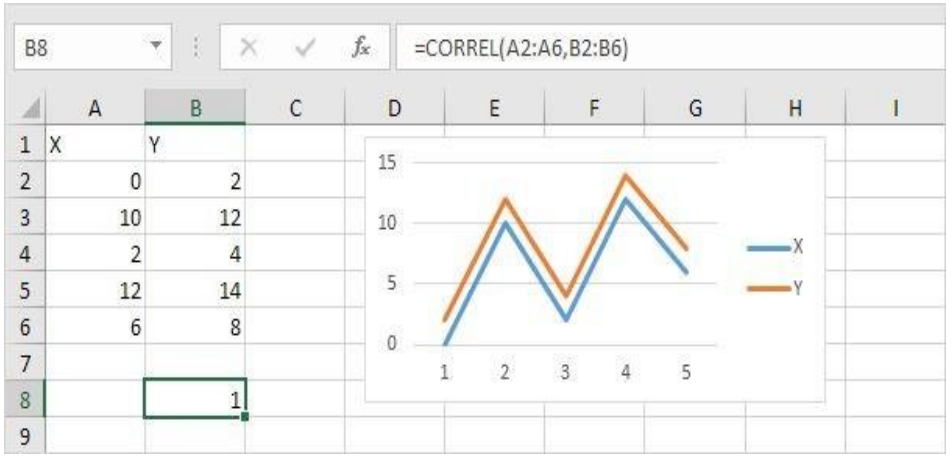
	Column1
Mean	47.25
Standard E	7.879311685
Median	55
Mode	87
Standard D	35.23735309
Sample Va	1241.671053
Kurtosis	-1.715300569
Skewness	-0.003594223
Range	96
Minimum	4
Maximum	100
Sum	945
Count	20
	0

kushal.1889@gmail.com

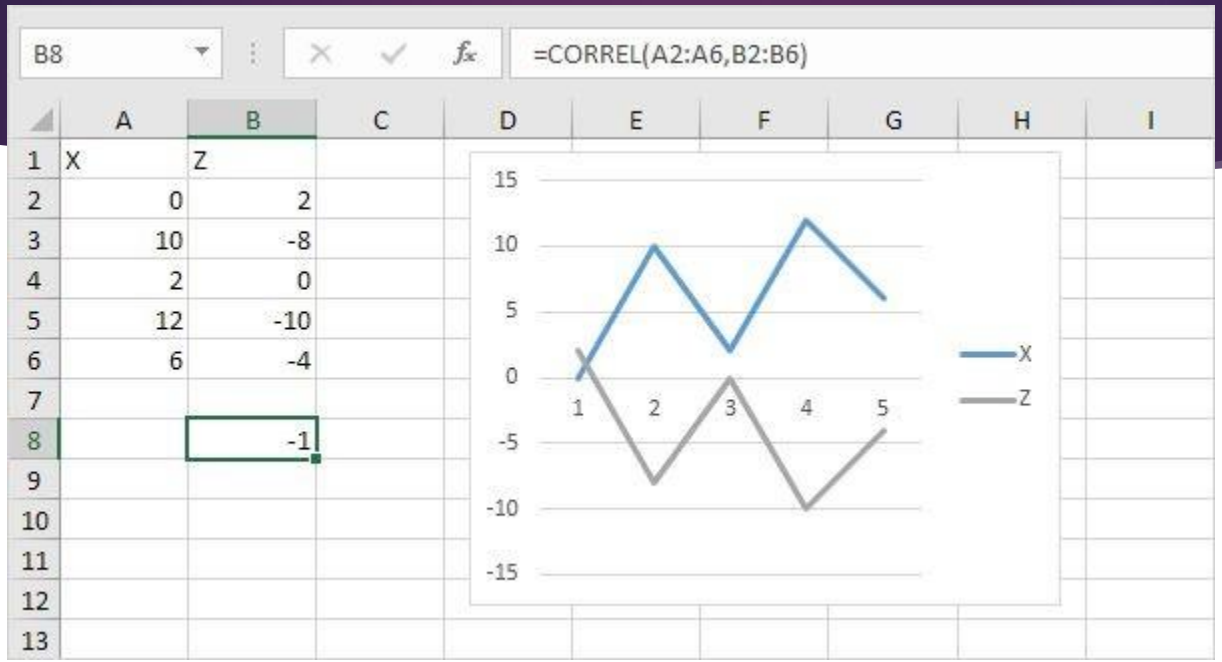
Correlation in Excel

The correlation coefficient (a value between -1 and +1) tells you how strongly two variables are related to each other. We can use the CORREL function or the Analysis Toolpak add-in in Excel to find the correlation coefficient between two variables.

- A correlation coefficient of +1 indicates a perfect positive correlation. As variable X increases, variable Y increases. As variable X decreases, variable Y decreases.

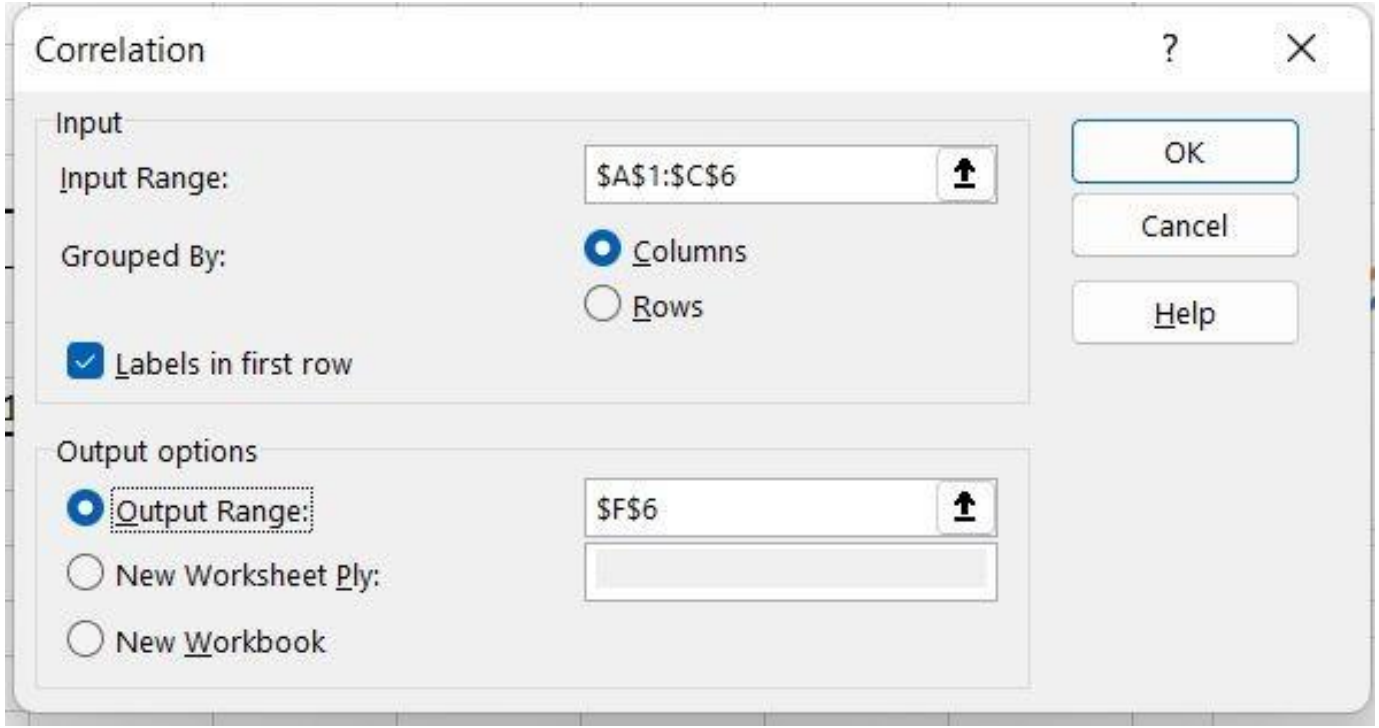


kushal.1889@gmail.com



- A correlation coefficient near 0 indicates no correlation.

A correlation coefficient of -1 indicates a perfect negative correlation. As variable X increases, variable Z decreases. As variable X decreases, variable Z increases.



kushal.1889@gmail.com

Conclusion: variables A and C are positively correlated (0.91). Variables A and B are not correlated (0.19). Variables B and C are also not correlated (0.11) . You can verify these conclusions by looking at the graph.

	A	B	C
1	A	B	C
2	0	2	2
3	14	6	11
4	1	8	3
5	10	5	13
6	5	6	4

	A	B	C
A	1		
B	0.191516	1	
C	0.909268	0.108893	1

kushal.1889@gmail.com