ICDL
The Digital Skills Standard

ICDL Professional

DATA ANALYTICS - FOUNDATION

Syllabus 1.0

Learning Material

(MS Excel, Power BI)

Provided by:

Digital Fortress Global Services Limited

ICDL Data Analytics - Foundation

Organisations, as they increasingly embed technology in their activities, have available to them large sets of data associated with their key activities – for example, sales, customer service, production, and logistics. These activities often generate large streams of data even for small organisations. It is possible for organisations to exploit this data to monitor, to improve, and to innovate – but only if they apply sound data analytic techniques. This ICDL Data Analytics - Foundation module provides you with foundational knowledge and skills in data analytics, which will enable you to start exploiting your data to deliver business intelligence that can help your business reach its goals.

On completion of this module you will be able to:

- Understand the key concepts relating to the application of data analytics in business.
- Understand and apply key statistical analysis concepts.
- Import data into a spreadsheet and prepare it for analysis using data cleansing and filtering techniques.
- Summarise data sets using pivot tables and pivot charts.
- Understand and apply data visualization techniques and tools.
- Create and share reports and dashboards in a data visualization tool.

What are the benefits of this module?

This module gives you the essential knowledge and skills relating to key data analytics concepts, statistical analysis, data set preparation, data set summarisation and data visualization. Developed with input from subject matter experts and practising professionals, it develops a knowledge of key concepts and a practical understanding of how to use data analytics in business. Once you have developed the skills and knowledge set out in this book, you will be in a position to become certified in an international standard in this area - ICDL Data Analytics - Foundation.

For details of the specific areas of the ICDL Data Analytics - Foundation syllabus covered in each section of this book, refer to the ICDL Data Analytics - Foundation syllabus map at the end of the book.

How to use this book

This book covers the entirety of the ICDL Data Analytics – Foundation course. It introduces important concepts and sets out the specific steps associated with using different features of the application. You will also have the opportunity to practice some of these activities yourself using sample files provided in the Student folder. It is recommended that you do not save your changes to sample files, as you may want to practice an activity more than once.

Note: *The spreadsheet steps described in this book apply to Microsoft Excel 2016 as part of Microsoft Office 2016, purchased as a stand-alone product. Other versions of Microsoft Excel 2016, such as Microsoft Excel 2016 as part of a Microsoft Office 365 subscription, may include different functionality and steps. For example, the* ***Get & Transform Data*** *option in Microsoft Excel 2016 as part of an Office 365 subscription follows different steps for importing data.*

ICDL DATA ANALYTICS - FOUNDATION

LESSON 1 –
KEY CONCEPTS

In this section, you will learn how to:

- Identify the main types of data analytics
- Outline the business benefits of data analytics
- Identify the main phases of data analysis
- Recognise data protection considerations when analysing data

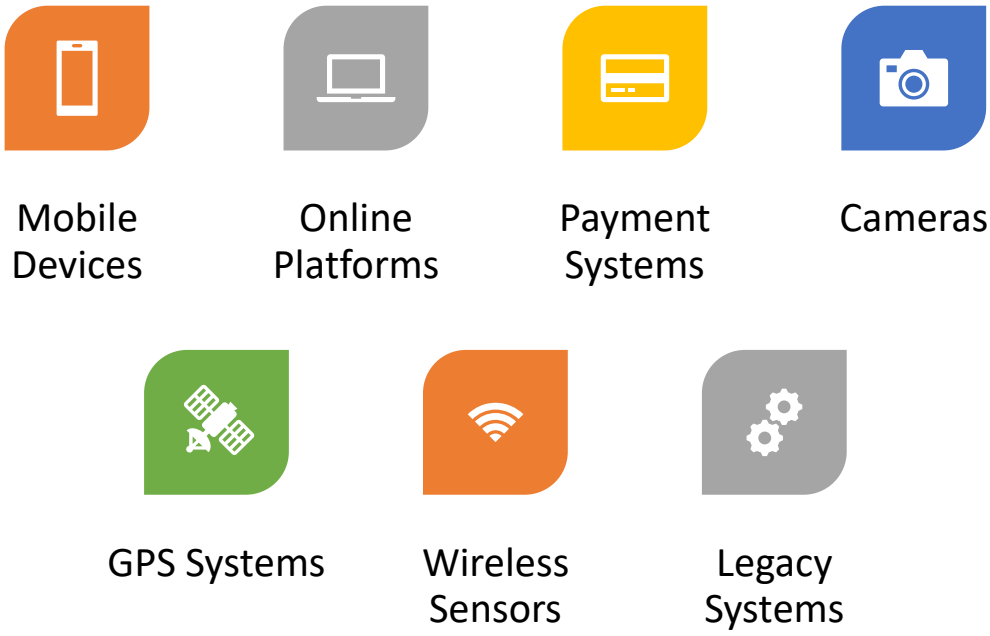
1.1 Types of Data Analytics

Concepts

Organisations can collect and process vast amounts of data such as sales figures, operating costs, logistics, customer satisfaction rates and online engagement metrics.

This data is collected from an increasing range of sources - for example, mobile devices, online platforms, payment systems, cameras, GPS systems, wireless sensors, and legacy systems. And this business data is stored in a wide variety of formats.

With vast amounts of data from different data streams and in different formats, organisations are using data analytics to find meaningful and useful insights that they can use to meet their organisational goals.



Data Sources

Data Analytics

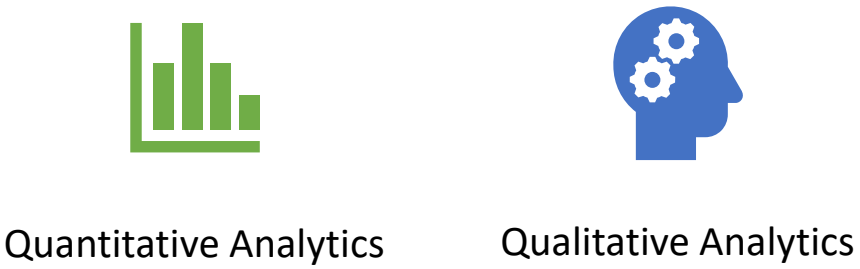
Data analytics refers to the techniques and processes used to collect, organise and examine data sets to create meaningful and useful information.

In recent times there has been a rapid rise in the use of data analytics across all types of organisations and sectors due to, among other things, technological advances in data analytics, increased computing power and better data storage capabilities. The use of data analytics can provide many benefits – for example, in business it can be used to gain competitive advantage, improve performance and increase profits.

There are different categories and types of data analytics, which although interrelated, have different purposes and provide different insights. In business deciding which type of data analytics is appropriate depends on a variety of factors, such as, the type of data being used, the stage of the workflow and your business requirements and objectives.

Quantitative and Qualitative Analytics

Two broad categories of data analytics are quantitative analytics and qualitative analytics.



Categories of Data Analytics

Quantitative analytics can be used to analyse information that can be measured and written down with numbers – for example, information like profits, costs, sales, number of repeat purchases, number of goods returned or number of complaints.

Quantitative analytics uses structured techniques to gather data, such as surveys, business reports and polls. And it uses mathematical and statistical methods to analyse the data gathered.

In quantitative analytics the data set is assessed objectively, providing broad and generalised insights. For example, you may calculate the percentage of customers who returned a product, the percentage of products that were faulty, or the number of sales in one store compared to another.

Qualitative analytics can be used to analyse information about attitudes, opinions and behaviours. This type of information is not easily measured or expressed as numbers – examples include, customer's opinions on a product or service, opinions on an organisation's reputation or attitudes towards a brand.

Qualitative analytics uses unstructured or semi-structured techniques to gather data such as, focus groups, discussions, interviews and observations.

In qualitative analytics the data set is assessed subjectively, providing narrow and specific insights. For example, you may need to analyse customer transcripts or interview notes to determine whether a customer likes your product or is satisfied with your service levels. You might have to pay close attention to the customer's language in order to draw conclusions, which will be subjective and depend on your interpretation of the customer's responses.

Qualitative analytics can also provide useful insights that you are unlikely to get from quantitative analytics. For example, you can use qualitative analytics to delve into the reasons for things, such as why some customers like your products and others don't.

Types of Data Analytics

Four types of data analytics, which can be broadly classified by their different purposes, are descriptive analytics, which is the simplest type, followed by diagnostic analytics, then predictive analytics and finally prescriptive analytics, which is the most complex type.

Types of Data Analytics

Descriptive Analytics

Descriptive analytics are used to find out **what happened in the past?**
Although considered the most basic form of data analytics, it still provides valuable insights into the past by summarising raw or historical data from multiple sources.

Descriptive analytics uses descriptive statistics, such as arithmetic operations, mean, median, and percentage.

Descriptive analytics can be used to create management reports providing insights into past performance. It allows you to see whether something happened as expected, for example if targets were met, but it doesn't provide reasons why.

Diagnostic Analytics

Diagnostic analytics are used to find out **why something happened in the past?** It takes a deeper look at the data to understand the root causes of events and to determine the factors that contributed to the outcome.

Diagnostic analytics use techniques such as drill-down, data discovery, and correlations. And it uses probabilities, likelihoods, and the distribution of outcomes for the analysis.

Diagnostic analytics can be used in social media marketing campaigns to assess levels of customer engagement through numbers of posts, mentions, followers, fans, page views, reviews and pins. This can give an insight into what worked in past campaigns and what did not.

Predictive Analytics

Predictive analytics are used to find out **what is likely to happen in the future?** It uses the findings of descriptive and diagnostic analytics to forecast the probability of a future outcome. The forecast is an estimate, the accuracy of which depends on the quality and constancy of the data.

Predictive analytics use techniques such as modelling to devise statistical or mathematical models of current and historical data.

Predictive analytics might be used by organisations to predict the impact of a proposed change, to predict customer purchasing trends, or to predict a customer's ability to repay a loan on time.

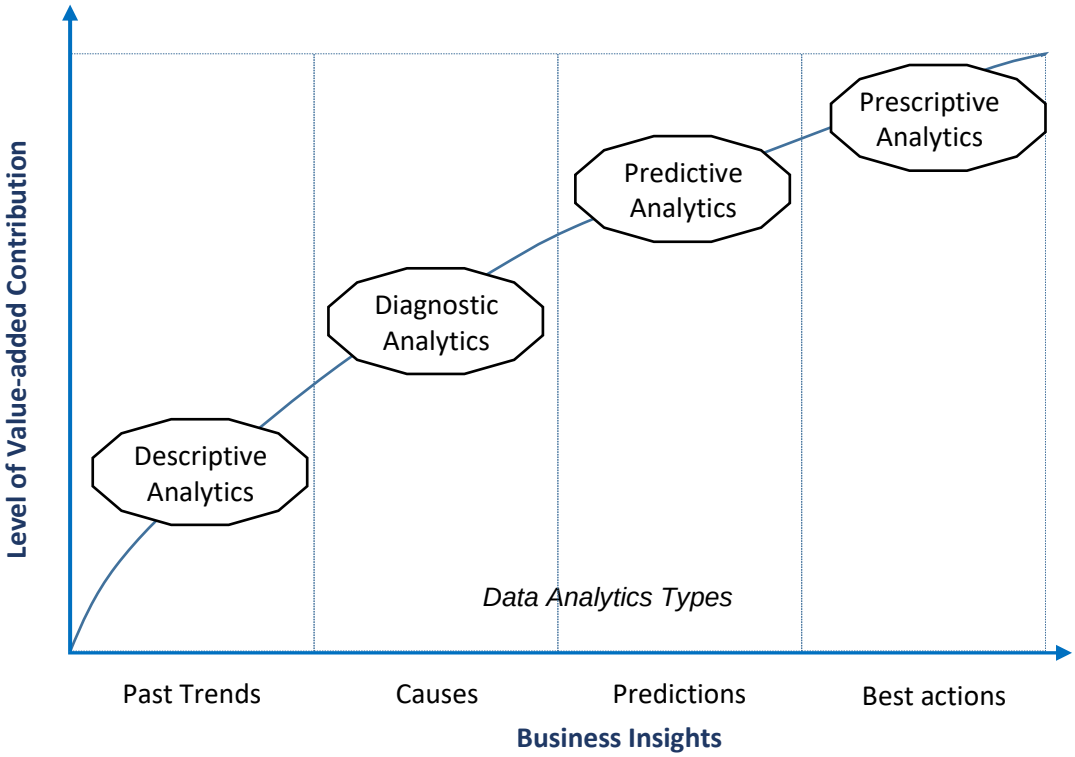
Prescriptive Analytics

Prescriptive analytics are used to identify **what is the best action to take now?** It is useful for avoiding problems that may arise in the future or for making the best use of trends. It is a relatively new and complex type of analytics.

Prescriptive analytics use the findings of predictive analytics combined with historical and transactional data, real-time data feeds from both internal and external sources, mathematical models and various business rules. Techniques include optimisation, simulation, and decision-analysis methods.

Larger organisations might use prescriptive analytics to optimise production, scheduling and inventory in the supply chain, to ensure delivery of the right products at the right time thereby optimising the customer experience.

The diagram below illustrates the value-added contribution and complexity of the various types of data analytics in relation to each other.



Case Example

The following is an online article that reports how a fashion retail company leverages their business using data analytics.

What are Zara's critical big data approaches?

Data is captured from POS terminals, e-commerce sales, customer surveys, PDA devices and RFID tags on the clothing. Store personnel are trained to capture customer preferences in their PDAs on every detail: buttons, zippers, color, cut and more. They enter and upload the feedback nightly and regional managers isolate their region's preferences and tastes.

The Wall Street Journal reports that "Zara has built its business on RFID tags." This gives Zara the ability to manage inventory cheaply at the same time it allows the company to observe the frequency of garments moving in and out of dressing rooms.

Zara gathers data from Instagram, surveys, and online social media, creating a central nervous system of their customers' fashion sensibility in each and every market.

Then, all the data above is compiled in the Arteixo, Spain data center 24 hours a day, analysts work, processing data to drive new design releases, shipping to stores twice a week. Zara generate weekly predictions for every single SKU they sell, offering their designs in tiny batches that run out—but rarely go on sale. Zara knows their customer inside out, and when they get it wrong, they can adapt faster than their competitors.

Source: Adam Nathan, Medium, Zara & Big Data: A 5-Minute Case Study, Oct 25, 2017

The various data analytics activities described in the article can be identified as belonging to one of the data analytics types - either descriptive, diagnostic, predictive or prescriptive.

Can you identify which activities belong to which type of analytics?

The table below outlines which data analytics activities belong to which type of data analytics:

Type of Data Analytics	Case Example Data Analytics Activities
Descriptive Analytics 	- Capturing data from POS (point of sale) terminals, e-commerce sales, customer surveys, PDA (personal digital assistant) devices, RFID (radio frequency identification) tags on clothing, store personnel's feedback, Instagram, surveys, and social media. - Compiling and analysing data.
Diagnostic Analytics 	- Processing data, based on descriptive analytics, to assess customers' feelings about new design releases.
Predictive Analytics 	- Generating prediction of Stock Keeping Unit (SKU) and shipping frequency to stores from Descriptive and Diagnostic Analytics.
Prescriptive Analytics 	- Identifying the popular units and prescribing the number to be offered so that they run out fast and price reductions on these items are minimised.

1.2 BUSINESS BENEFITS

Concepts

Knowing how to use and interpret data can help businesses reach their business goals and stay ahead of their competition. The insights gained from data analytics can provide businesses with the following benefits:

1. Identify Patterns And Trends

2. Improve Efficiency

3. Support Decision Making

4. Present Information Effectively

These benefits can generate positive outcomes for a business and its customers.

Identifying patterns/trends

The process of evaluating large sets of data containing varying data types can help businesses to uncover hidden patterns in the data.

Data analytics can be used to identify customer preferences and behaviour – for example, data gathered through loyalty card systems and online shopping platforms can provide information on purchasing patterns and trends in real time. This information can be used to provide personalised recommendations and targeted advertising to shoppers.

Data can also be analysed to identify market trends – for example, in an online retail environment, popular products and those becoming popular can be identified and this information used in supply chain management.

Organisations, such as social media platforms, can use data analytics as the basis of new business models where access to services are free in exchange for personal data. Data analytics are then used to generate revenue from the personal data, for example, by categorising customers according to their product preferences and using this information in targeted marketing campaigns.

Financial organisations may use data analytics to analyse past financial performance in order to find patterns and predict future financial performance. This information can be used in things like assessing risks when considering loan applications.

Improving efficiency

The insight gained by data analytics enables businesses to automate and optimize their business processes thus improving operational efficiency. For example, a chain of hotels can analyse historic data to predict when their hotels will be busiest. They can use this information to allocate staff and resources efficiently, thereby reducing costs.

In a manufacturing environment data analytics can be used to identify the output required at a given time period and the resources required to produce that quantity. With this analysis, production is more efficient and there is less waste.

Supporting decision making

Access to large volumes of analysed data allows businesses to make better and more informed business decisions and to set more effective business strategies.

For example, managers and decision-makers can make quicker and more effective choices using regularly updated dashboards and reports that display and/or summarise important performance metrics.

HR Service providers may use data analytics to help them decide whether to introduce a new HR policy. They may forecast the impact of the proposed HR policy on employee performance and only decide to introduce it if it has a positive effect on employee performance and well-being.

Healthcare providers may use data analytics to predict patient's medical needs. This information will help them when deciding how to allocate resources most effectively when planning for future demand.

Presenting information effectively

Data presented in easy-to-understand formats such as tables, charts, maps, and graphs help businesses to better interpret and understand their data.

Using visual data in reports and dashboards can make businesses more agile and help them find revealing insights faster and make decisions more quickly.

1.3 Data Analysis Process

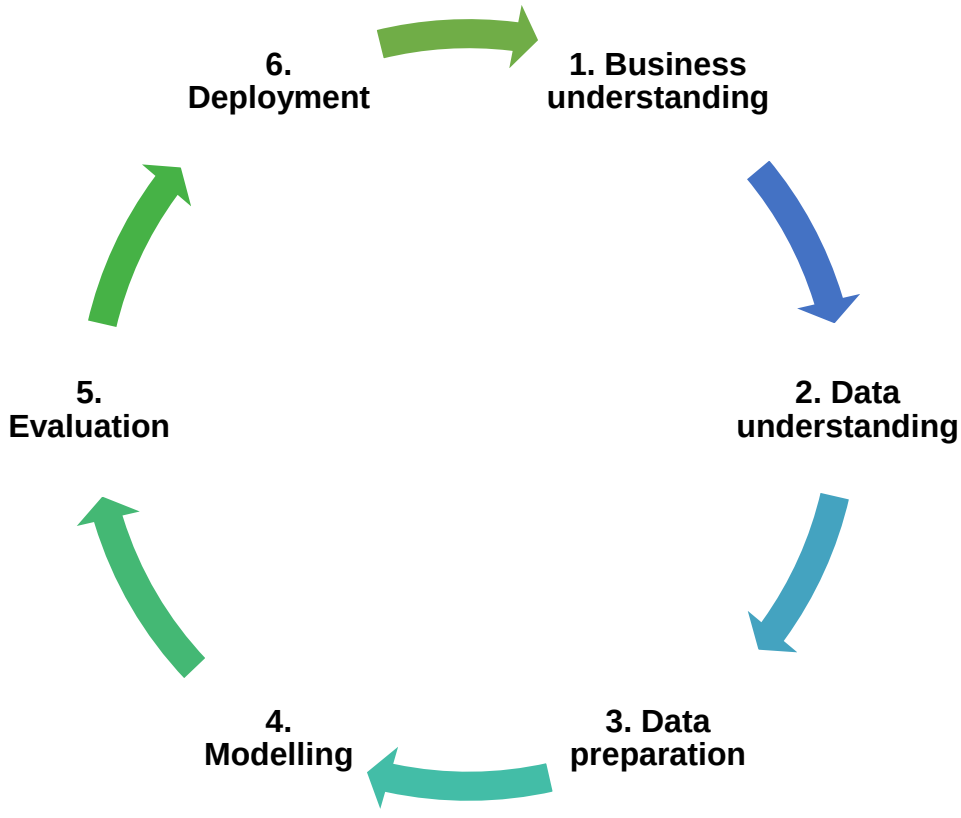
Concepts

Starting to use data analytics to gain business intelligence can be challenging. This can be the case when there is too much data as well as when there is too little data. Looking at data analysis as a process with different phases can help you get the most out of your organisations' data.

The data analysis process and phases will vary depending on the type of data analytics used and your business implementation. It can include tasks such as defining the question you are trying to answer; deciding what data you need to answer the question; determining if you have the data; collecting the data; getting the data ready for analysis (cleaning and transforming the data); analysing the data; interpreting the results; and using the results.

The Data Analysis Process *Type of Data Analytics* *Business Implementation*

For example, a type of data analytics known as data mining, which explores large data sets and generates predictions, typically follows a process with six phases.



Data Analysis Phases

These six phases are described in the table below.

Phase	Description
1. Business understanding 	In the initial phase the aim is to understand what the business goals are and what information is required to achieve those goals - for example, increase customer retention by 10% from last year. Then you create the related data analytics objective or objectives – for example predict how many customers will renew this year based on customer renewal rates over the last 3 years. And then you create a project plan to include defining timelines, how success will be measured and the tools and techniques.
2. Data understanding 	In the second phase the aim is to understand the data by considering the data requirements. You review and document the data you have acquired to establish if there are any data management or quality problems that need to be addressed. This involves collecting, describing, exploring and assessing the data.
3. Data preparation 	In the third phase the aim is to prepare the data for the modelling phase. You construct the final data set to use in your analysis from the initial raw data. This involves selecting, cleaning and transforming (constructing and integrating) the data.
4. Modelling 	In the fourth phase the aim is to find useful patterns in the data that can be useful to the business. To do this, you select and apply appropriate data modelling tools and techniques. This involves selecting modelling techniques, designing tests, building models and assessing models. This phase involves trial and error and usually takes multiple iterations.
5. Evaluation 	In the fifth phase the aim is to ensure that the results are useful for the business. To do this, you evaluate the results of the model in relation to the business success criteria.
6. Deployment 	In the sixth phase the aim is to use the results in the business to meet the business goals. To deploy the model, you integrate the results into the business. This phase includes planning the deployment, reporting the results and monitoring the results.

Note: *In this Foundation module, the focus is on the data understanding and data preparation phases and the associated tools in Microsoft Excel and Power BI.*

1.4 Data Protection Considerations

Concepts

Data protection is an area of law that is designed to protect the privacy rights of individuals in relation to the processing of their personal data. Personal data is typically defined as data that contains identifiers that can be used to identify an individual. Data protection laws govern how organisations use personal data and need to be complied with when analysing data that contains personal data.

Complying with Data Protection Regulations

Data protection regulations vary across the world. When analysing data that contains personal data, organisations need to ensure compliance with the appropriate data protection laws in their jurisdiction.

One example of a data protection law is the General Data Protection Regulation (GDPR). This regulation came into effect on 25th May 2018 in the European Economic Area (EEA). It applies to organisations established in any member state within the European Union. It also applies to organisations established outside the European Union, if they offer goods or services or monitor the behaviour of individuals within the European Union. Among other things, the GDPR requires that organisations processing personal data ensure that personal data is

- processed lawfully, fairly and in a transparent manner.
- only collected for specified, explicit and legitimate purposes.
- adequate, relevant and limited to what is necessary for processing.
- accurate and up-to-date.
- only kept for as long as it is needed for processing.
- stored securely.

Anonymising Personal Data

Complying with data protection laws may require a significant amount of effort from organisations. If using personal data is not necessary to achieve the objectives of the data analysis, it may make sense to avoid using personal data, for example, by not collecting it or by anonymising it. To make data truly anonymous, you need to ensure that it is impossible to link the data back to an individual. Some techniques for anonymising personal data include removing direct identifiers such as names and addresses or by reducing details such as using years of birth instead of dates of birth.

1.5 REVIEW EXERCISE

1. Which of the following data analytics types uses statistical calculations and aggregation of data?

   a) Descriptive b) Diagnostic

   c) Predictive d) Prescriptive

2. Which of the following data analytics types uses forecasting tools to predict future trends?

   a) Descriptive b) Diagnostic

   c) Predictive d) Prescriptive

3. Which of the following data analytics types is the most complex and very process-intensive?

   a) Descriptive b) Diagnostic

   c) Predictive d) Prescriptive

4. List 2 potential benefits for businesses embarking on data analytic activities.

   _____

   _____

5. Which of the following is NOT an example of quantitative analysis?

   a) Average price is $250k b) Longest time is 15 mins

   c) Soft and silky covers d) Number of registrations: 1074

6. Identify the last 2 phases of the data analytics process in data mining:

   _____

   _____

7. _____ personal data is the task of removing direct identifiers from personal data.

LESSON 2 –
STATISTICAL ANALYSIS

In this section, you will learn how to:

- Describe measures of central tendency of a data set
- Calculate central tendency values of a data set
- Describe some measures of variation of a data set
- Calculate the variation of a data set

2.1 SUMMARY STATISTICS INTRODUCTION

Concepts

To make sense of, and interpret large data sets of quantitative data, the data needs to be summarised or described in some way. Statistical analysis is a way of summarising or describing data to highlight typical values in the data, as well as how these vary.

Statistical analysis techniques for summarising or describing data are known as summary or descriptive statistics. The two main techniques are

- Measures of central tendency (averages).
- Measures of variation.

Measures of central tendency
These are summary statistical values that describe a data set by identifying the middle or central position within the data set. The number in the central position is seen as a typical representation of the numbers in the data set. The three most common measures of central tendency are:

- Mean
- Median
- Mode

Each of these measures calculates the location of the central point within a data set using a different method. Choosing the appropriate measure of central tendency depends on the type of data set and what you want to find out.

Measures of variation
These are statistical analysis techniques for interpreting and summarising large sets of quantitative data by examining the distribution of the data set. Rather than showing how data are similar, as is the case with measures of central tendency, measures of variation describe how the data varies or differs in a data set. The three most common measures of variation are:

- Range
- Variance
- Quartiles

Measures of variation are also commonly referred to as measures of dispersion, distribution, or spread.

Using a measure of central tendency and the relevant measure of variation for a data set data provides a better overview of the data than using one alone. In data analysis, this type of statistical analysis can be a first step in describing the data, before more complex analysis.

Case Example

Consider the following report to see if the measures of central tendency can be applied to solve the question presented:

Are our feet really getting bigger?

One high street shoe retailer has seen a marked increase in sales of larger sizes. But are our feet really getting bigger or are we coming round to the idea of "sensible shoes"?

Clodhopper, Big Foot, Yeti, flippers, clown's feet. You name it, Emma Supple has heard it from the patients who come knocking on the door of her foot care clinic.

Their motivation is to find some relief from the pain of years spent squeezing into shoes that are too small.

The result of this self-imposed form of 21st Century foot binding is a host of podiatric injuries, ranging from corns, callouses and blisters to trapped nerves, toes which have been compressed to resemble claws and a condition called mallet toe.

Source: Jonathan Duffy, BBC News Magazine, 8 January 2010

In response to the question in the report “Are our feet really getting bigger?” you may want to check if feet have indeed grown bigger or is it that the sale of larger sizes of shoes has increased?

You know that shoe sizes are based on some measurements. Some things you may consider checking are:

- What is considered an **average** size of shoes?
- Is the **average** size a single value or a range of values?
- How far apart are the measurements of different sizes from what is considered an **average** size?

2.2 MEASURES OF CENTRAL TENDENCY

Concepts

Mean

The mean is the most popular measure of central tendency. This is commonly referred to as the average of a data set. The mean is calculated by summing up all the values in the data set and then dividing by the number of values in the data set. Since it incorporates all the values in the data set, any changes to the values will affect the mean.

Case Example

The following is the gross profit information for 10 fashion retail stores in a city:

Sample Data Set	
Store	**Gross Profit**
A	150 000
B	180 000
C	160 000
D	140 000
E	150 000
F	150 000
G	120 000
H	170 000
I	190 000
J	130 000
Total	1 540 000

As a potential investor in fashion retail, what representative profit value will you use for these 10 retail stores?

Calculating the mean gross profit for a group of stores:

Sample Data Set		Mean Calculation
Store	**Gross Profit**	
A	150 000	Total Gross Profit = 1 540 000 Number of Stores = 10 Mean = 1 540 000 / 10 = **154 000**
B	180 000	
C	160 000	
D	140 000	
E	150 000	
F	150 000	
G	120 000	
H	170 000	
I	190 000	
J	130 000	
Total	1 540 000	

The **mean gross profit** can also be visualised using a chart:

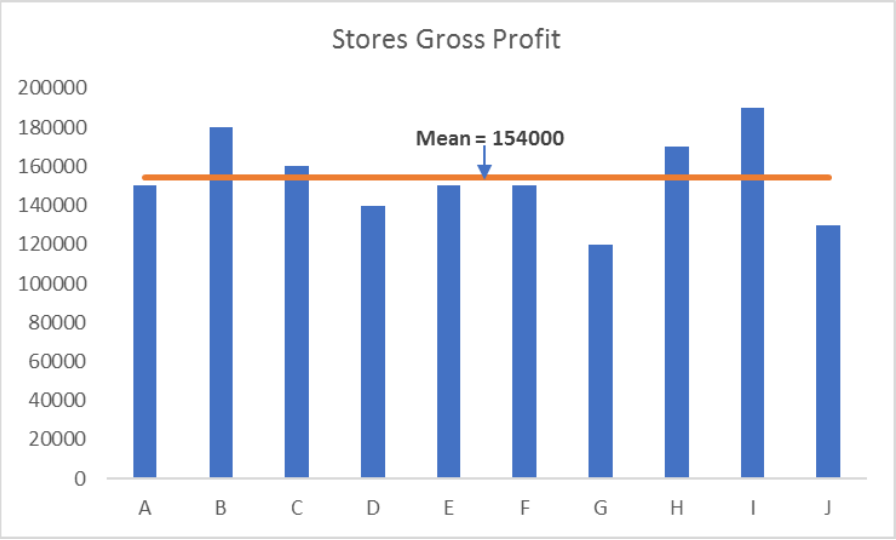


Chart Showing the Values in a Data set and the Mean

The mean may not always reflect the central position, especially when the data set has outliers. Outlier are values that are unusual compared to the rest of the data set by being particularly small or large. The mean value will be skewed by these values.

Case Example

In a survey on service quality across 10 restaurants in a city, the data collected on service response time is shown as follows:

Sample Data Set	
Customer	**Service Response Time (minutes)**
1	5
2	8
3	6
4	4
5	5
6	5
7	2
8	7
9	19
10	13
Total	74

Will a mean value of the service response time be representative of the service quality of the restaurants in that city?

Calculating the mean response time for service calls:

Sample Data Set		Mean Calculation
Customer	**Response Time (minutes)**	
1	5	Total Response Time = 74 Number of customers = 10 Mean = 74 / 10 = **7.4**
2	8	
3	6	
4	4	
5	5	
6	5	
7	2	
8	7	
9	19	
10	13	
Total	74	

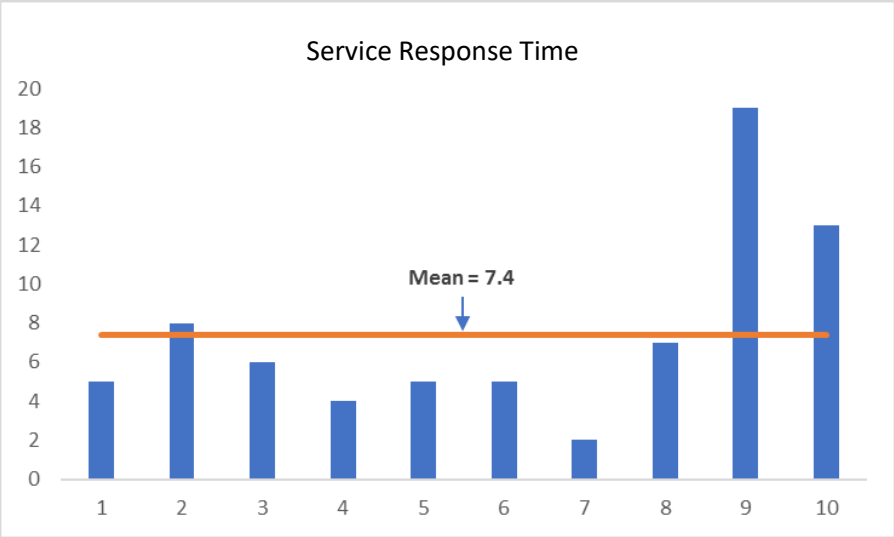


Chart Showing the Values in a Data set and the Skewed Mean

As in the example above, by inspecting the raw data, it suggests that the mean of 7.4 minutes, may not accurately reflect the typical response time, as most calls are in the 2 to 8 minutes range. The outliers (19 and 13 minutes in this case) tend to pull the mean away from the centre.

Concepts

Median

The median is the middle value in the data set distribution. It splits the data set in half. In an ordered set of data, the median is the middle value such that 50% of all data lies above or below it. Median is less affected by outliers and skewed data.

Note: *If calculating the median using a manual calculation it is necessary to have the data set arranged in ascending order before locating the median.*

Case Example

It is common for a business to experience sales fluctuations over different days in a month or a year. On some special occasions, such as Cyber Monday, some businesses will experience unusually high volume of sales; on other days, sales may be less than desirable. In this case, using the median sales value can better reflect a sales value on a typical day than using the mean sales.

The following sales information was collected for a small business over 10 days. What is the median sales value?

Sample Data Set	
Day	**Sales**
1	17 000
2	18 000
3	16 000
4	14 000
5	15 000
6	14 000
7	12 000
8	17 000
9	19 000
10	13 000

Calculating the median sales:

Sample Data Set		Sorted Data Set		Median Calculation
Day	**Sales**	**Day**	**Sales**	
1	17 000	7	12 000	
2	18 000	10	13 000	
3	16 000	4	14 000	Values in Data set = 10 (even)
4	14 000	6	14 000	
5	15 000	5	15 000	Median = Average of 5th & 6th values
6	14 000	3	16 000	
7	12 000	1	17 000	= (15 000 + 16 000) / 2
8	17 000	8	17 000	= **15 500**
9	19 000	2	18 000	
10	13 000	9	19 000	

Note*: If this data set had an odd number of values, the median would be identified by the middle value in the data set.*

Concepts

Mode

The mode is the value that occurs the most frequently in the data set. Mode is not affected by extreme values. It can be used for either numerical or qualitative data.

If there are no repeating values in the data set, then the data set does not have a mode. For some data sets, there can also be more than one mode. It is possible to have two modes (bimodal), three modes (trimodal) or more modes within larger data sets.

Mode is a very useful measure of central tendency in some industries; for example, a bookstore would use it to find out which books in its store are most in demand.

Case Example

The following results were collected in a customer satisfaction survey for a restaurant where 10 customers rated their satisfaction level from 1 to 5 with 5 being the highest rating. What is the mode level of satisfaction?

Sample Data Set – Customer Satisfaction Survey Results

5 4 5 3 5 4 5 5 2 5

Identifying the mode for a customer satisfaction survey:

Sample Data Set – Customer Satisfaction Survey Results					
Satisfaction level	5	4	3	2	1
Number of entries	6	2	1	1	0
Mode = **5**					

2.3 CALCULATING CENTRAL TENDENCY

Concepts

Spreadsheet applications provide functions for calculating the different central tendency values of data sets. In Microsoft Excel the functions are as follows:

- The **AVERAGE** function calculates the **mean** of a set of values in a data set. The AVERAGE function syntax is

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

- The **MEDIAN** function finds the **median** (middle number) of a set of values in a data set. The MEDIAN function syntax is

  MEDIAN(number1, [number2], ...)

- The **MODE** function finds the **mode** (the most frequently occurring number) in a numeric data set. The MODE function syntax is

  MODE(number1, [number2], ...)

The arguments can be a number or cell reference that refers to numeric values.

Steps

To calculate central tendency values using functions in Excel:

1. Open the **Orders1.xlsx** workbook.

	A	B	C	D
1	**OrderID**	**Order Date**	**Amount**	**Status**
2	ID1001	20/4/2018	$100	Delivered
3	ID1002	24/4/2018	$70	Cancelled
4	ID1003	26/4/2018	$130	Delivered
5	ID1004	2/5/2018	$250	Delivered
6	ID1005	2/5/2018	$90	Cancelled
7	ID1006	3/5/2018	$115	Delivered
8	ID1007	6/5/2018	$90	In transit
9	ID1008	10/5/2018	$120	Delivered
10	ID1009	18/5/2018	$210	Delivered
11	ID1010	18/5/2018	$130	In transit
12	ID1011	20/5/2018	$145	Delivered
13	ID1012	21/5/2018	$280	Delivered
14	ID1013	25/5/2018	$130	In transit
15	ID1014	3/6/2018	$80	Delivered
16	ID1015	4/5/2018	$110	Delivered

The data set shows order details such as order date, amount and status.

2. Click in the cell **G4** and type the formula **=AVERAGE(C2:C16)** to calculate the mean amount value.

3. Click in the cell **G6** and type the formula **=MEDIAN(C2:C16)** to calculate the median amount value.

   Note: *In Excel, you do not have to sort the data range for calculating median.*

4. Click in the cell **G8** and type the formula **=MODE(C2:C16)** to calculate the mode amount value.

5. The central tendency values of the data set are displayed in the respective cells. Save the file as **Orders1completed.xlsx** workbook.

	F	G	H
1	**Central Tendency Measures**		
2	**for the Amount Values**		
3			
4	Mean:	$137	
5			
6	Median:	$120	
7			
8	Mode:	$130	

2.4 MEASURES OF VARIATION

Concepts

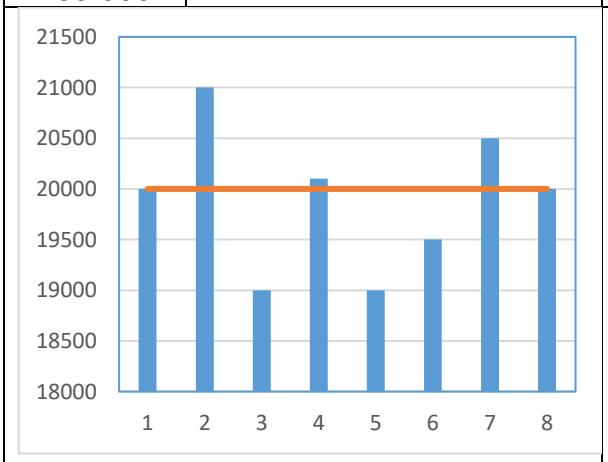
Measures of variation can be used in conjunction with measures of central tendency to give a more complete view of a data set. Each of these measures is used to determine the dispersion/scatter of values in a distribution.

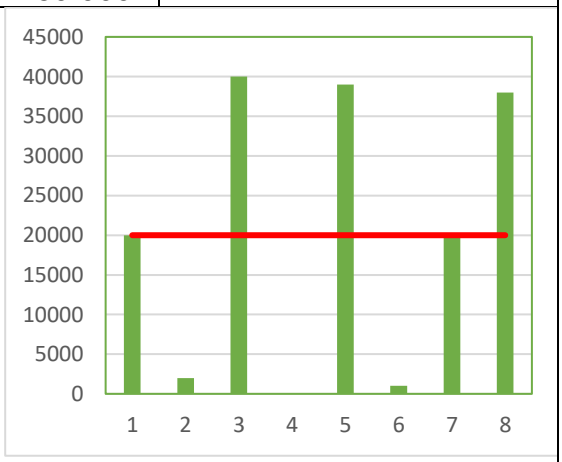
Case Example

Why would you want to measure the variation of a data set?

In the example below there are sample data sets for two retail stores. The mean, median, and mode of each store's daily sales all equal $20 000. However, there is significant difference between the two data sets. Store A's daily sales are much more consistent than those of Store B, which shows greater variation. This illustrates the need for measures of variation.

Sample Data Set 1 – Store A Sales		Sample Data Set 2 – Store B Sales	
20 000		20 000	
21 000		2 000	
19 000	Mean = 160 000 / 8	40 000	Mean = 160 000 / 8
20 100	= **20 000**	0	= **20 000**
19 900		39 000	
19 500	Median = **20 000**	1 000	Median = **20 000**
20 500		20 000	
20 000	Mode = **20 000**	38 000	Mode = **20 000**
Total = 160 000		**Total = 160 000**	





Concepts

Range

Range is one of the most basic measures of variation. It is the difference of the maximum and minimum values where **Maximum** is the largest value in the data set and **Minimum** is the smallest value in the data set. It is one way to describe how the data is dispersed or spread out in a data set

Range = Maximum-Minimum

Minimum, maximum, and range are very sensitive to outliers.

Case Example

The following is the sample data set for 10 retail stores. The calculation shows the minimum, maximum, and range gross profit

Sample Data set		Minimum, Maximum, and Range Calculation
Store	**Gross Profit**	
A	150 000	
B	180 000	
C	160 000	Minimum = **120 000**
D	140 000	
E	150 000	Maximum = **190 000**
F	150 000	
G	120 000	Range = 190 000 - 120 000
H	170 000	= **70 000**
I	190 000	
J	130 000	

Concepts

Variance

Variance measures how far a data set is spread out. Variance tells how representative the mean is of each of the values in the data set.

- The closer each individual value in the data set is to the mean, the smaller the variance will be.
- If each value is at the mean in a data set, the variance will equal zero, which indicates that there is no variation from the mean across the entire data set.
- The further each individual value in the data set is from the mean, the greater the variance will be, which indicates that the mean is not as typical of the individual values in the data set.

Technically variance is the average of the squared differences from the mean.

Case Example

The following is the new contracts volume handled by three sales representatives:

	Sample Data		
New Contracts Volume	**Sales Rep A**	**Sales Rep B**	**Sales Rep C**
	10	10	7
	8	10	6
	6	6	6
	4	1	6
	2	3	5
Total	**30**	**30**	**30**
Mean	= 30 / 5 = **6**	= 30 / 5 = **6**	= 30 / 5 = **6**

In the data set above, each sales representative's contracts volume has a mean equal to 6. However, the contracts volume differs with respect to how much the individual values vary from the mean.

Since the variance represents the average squared deviation from the mean, you need to determine which sales representative's contracts volume data set has the greatest variance and which one has the smallest variance.

- The contracts volume of Sales Rep B seems to be further from the mean value of 6, so it should have the largest variance.
- The mean value of 6 seems to be most typical of the contracts volume of Sales Rep C, so it should have the smallest variance.
- The contracts volume for Sales Rep A appears to be in between, so it should have a variance between Sales Rep B and Sales Rep C.

The following calculates the variance for the volume of contracts.

Sales Rep A Contracts Volume	Difference from Mean	Difference from Mean Squared	Sales Rep B Contracts Volume	Difference from Mean	Difference from Mean Squared	Sales Rep C Contracts Volume	Difference from Mean	Difference from Mean Squared
10	4	16	10	4	16	7	1	1
8	2	4	10	4	16	6	0	0
6	0	0	6	0	0	6	0	0
4	-2	4	1	-5	25	6	0	0
2	-4	16	3	-3	9	5	-1	1
	Total	40		Total	66		Total	2
Total of Difference from Mean Squared = 40 **Variance** = 40 / 5 = **8**			**Total of Difference from Mean Squared** = 66 **Variance** = 66 / 5 = **13.2**			**Total of Difference from Mean Squared** = 2 **Variance** = 2 / 5 = **0.4**		

The size of the variance does not have any special underlying standard interpretation. Variance, along with the mean, allows you to judge how well the mean serves as a measure of central tendency.

Typically, information for such statistics would be presented in a summary form as follows:

Summary Statistics	Contracts		
	Sales Rep A	Sales Rep B	Sales Rep C
Number of Contracts:	5	5	5
Mean of Contracts Volume:	6	6	6
Variance of Contracts Volume:	8	13.2	0.4

Using the summary, you can reach some conclusions about what each of these data sets would look like. For example, you can see that all three data sets are of the same size (5), and that all three have the same mean (6). The fact that the variance for Sales Rep C is only 0.4 indicates that most of the individual values in the data set should be very close to the value of 6. Similarly, you would assume that the individual values for Sales Rep B must be more diverse or spread out around the mean since the variance is much larger.

Concepts

Quartiles

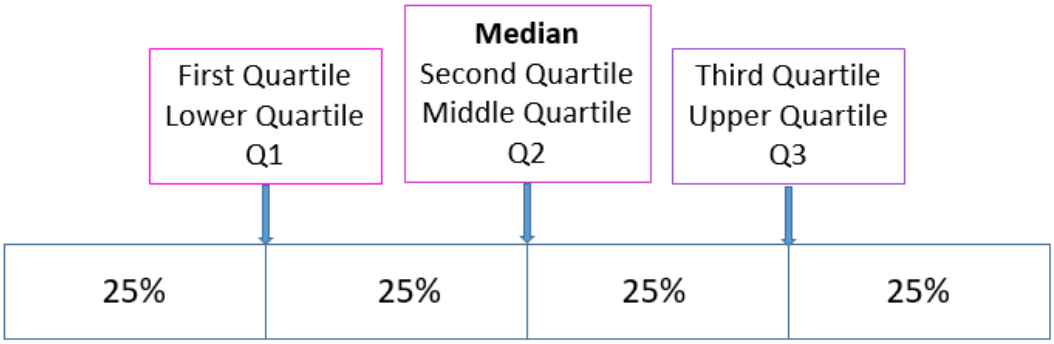
Quartiles in statistics are values that divide the data set into quarters according to where the values fall on the number line. They are often used in sales and surveys to divide population into groups.

The three quartiles that divide a data set into quarters are:

- The median of the lower half of the data set (25th percentile)
- The median of the data set (50th percentile)
- The median of the upper half of the data set (75th percentile)

The basic formula used to find a quartile by determining the value in the appropriate position in the ranked discrete data set are outlined below, where n is the number of values in the data set:

Quartile Position	Formula
First, Q_1	(n + 1) / 4
Second, Q_2 (also known as the median)	(n + 1) / 2
Third, Q_3	3 (n + 1) / 4



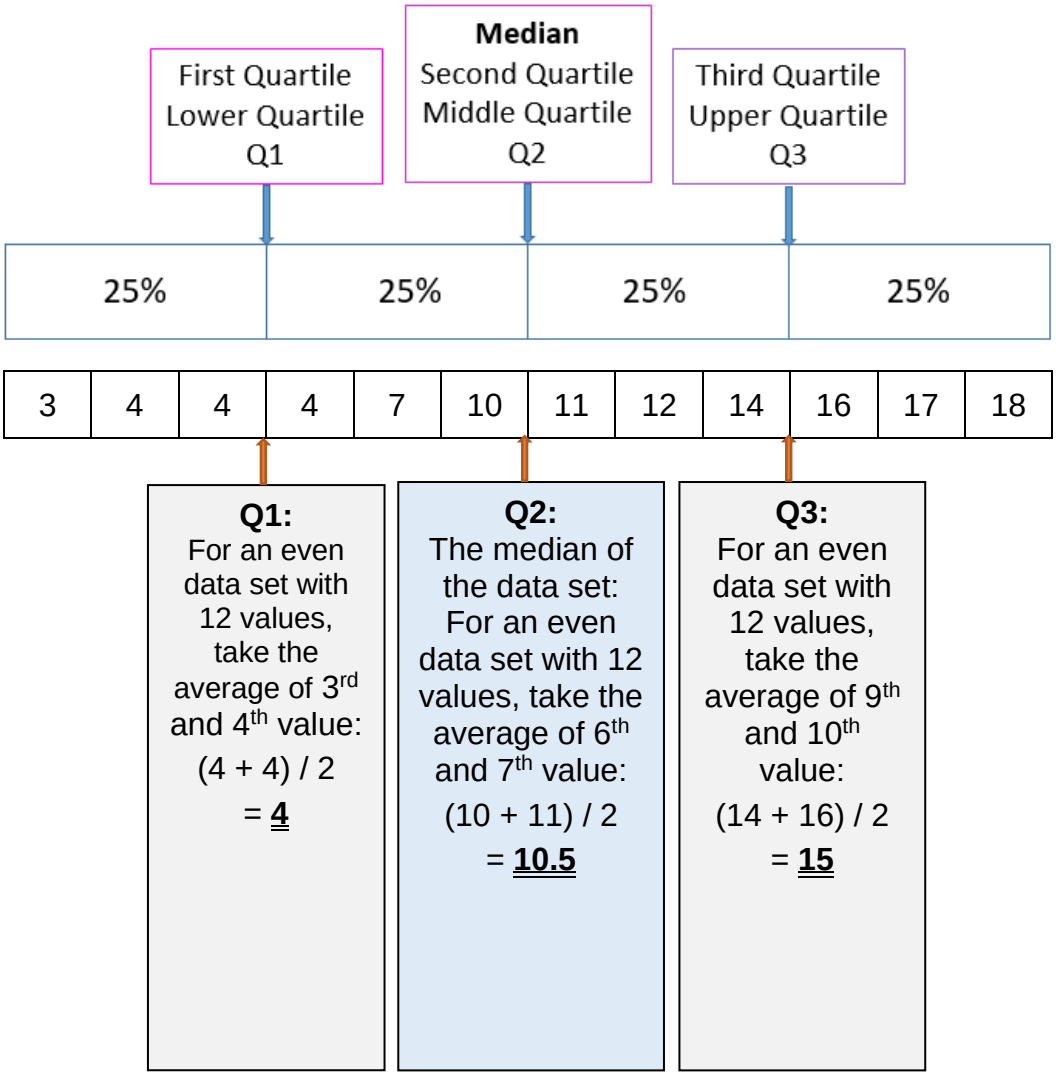
Quartile Positions

Case Example

The following calculates the quartiles for a new product's online order quantities for 12 days immediately after the first launch.

	Online Order Quantities											
Sample Data	16	14	18	4	17	10	11	12	4	3	4	18

The order quantities ranged from 3 to 18. How can we evaluate the variation within this set of data?



Quartile Values for an Even Set of Data

2.5 CALCULATING VARIATION

Concepts

Spreadsheet applications provide functions for calculating the different measures of variation (range, variance, and quartiles) of data sets. In Microsoft Excel the functions are as follows:

- The **MAX** and **MIN** functions are used to calculate the **range.** The syntax is

  MAX(number1, [number2],...)-MIN(number1,[number2],...)

  Arguments can be numbers, cell references or ranges containing numeric values.

- The **VAR.P** function finds the **variance** in an entire population. The syntax is

  VAR.P(number1, [number2], ...)

  Arguments can be numbers or names, arrays, or references that contain numbers.

- The **QUARTILE** function finds the **quartile** for a particular data set. The syntax is

  QUARTILE(array, quart)

  The array argument refers to the numeric data you want to analyse and is the range of cells containing the data set. The quart argument is the quartile value you want to return. There are 5 values for the quart argument as follows:

Quart	Return value
0	Minimum value
1	First quartile – 25th percentile
2	Median value – 50th percentile
3	Third quartile – 75th percentile
4	Maximum value

Steps

To calculate variation using functions in Excel:

1. Open the **Orders2.xlsx** workbook.
2. The data set shows order details such as order date, amount and status. Click in the cell **G4** and type the formula **=MAX(C2:C16)-MIN(C2:C16)** to calculate the range of the amount value.
3. Click in the cell **G7** and type the formula **=VAR.P(C2:C16)** to calculate the variance of the amount value.

Based on the variance value, complete the following statement to describe the amount values compared to the mean:

Most of the amount values in the data set are

_____

_____

***Answer**: Most of amount values in the data set are more diverse or spread out around the mean because the variance is much larger compared to the mean. This indicates that in this example the mean is not the best measure to determine the central tendency of the amount values.*

4. Click in the cell **G9** and type the formula **=QUARTILE(C2:C16,1)** to calculate the first quartile value.

5. Click in the cell **G10** and type the formula **=QUARTILE(C2:C16,2)** to calculate the second quartile value.

6. Click in the cell **G11** and type the formula **=QUARTILE(C2:C16,3)** to calculate the third quartile value.

   ***Note:** In Excel, you do not have to sort the data range for quartile calculations.*

   The range, variance and quartile values of the data set are displayed in the respective cells.

	F	G
1	**Variation Measures**	
2	**for the Amount Values**	
3		
4	Range:	$210
5		
6	Mean:	$137
7	Variance:	3586
8		
9	Q1:	95
10	Q2:	120
11	Q3:	137.5

Variation measures for order values

2.6 REVIEW EXERCISE

1. Open the **Cards.xlsx** workbook.

2. The data set shows credit card transactions of a retail store for a period of 3 months. Enter a formula in **I4** to calculate the mean amount of all the credit card transactions.

3. Enter a formula in **I6** to calculate the median amount of all the credit card transactions.

4. Enter a formula in **I8** to calculate the mode amount of all the credit card transactions.

5. Answer the following questions based on the calculation above:

   a. What is the most frequently charged amount?

   _____

   b. What is the average amount charged?

   _____

6. Enter a formula in **I10** to calculate the range of all the credit card transactions.

7. Enter the appropriate quartile formulas in the range **I12:I15**.

   The results are displayed as follows:

	H	I
3	**Statistical Summaries**	
4	Mean:	93
5		
6	Median:	88
7		
8	Mode:	65
9		
10	Range:	199
11		
12	Q1:	47.75
13	Q2:	88
14	Q3:	140
15	Q4:	200

8. A sales manager wants to calculate the range of performance of his sales associates. What must he do?
   a. Identify the central sales amount.
   b. Find the sales amount that occurs the most.
   c. Identify how far from average each sales amount is.
   d. Subtract the lowest sales amount from the highest sales amount.

LESSON 3 –
IMPORTING DATA SETS

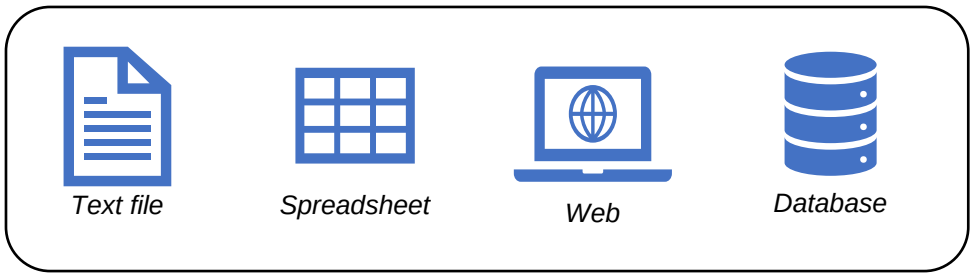
In this section, you will learn how to:

- Import data into a spreadsheet application from a text file
- Import data into a spreadsheet application from a spreadsheet
- Import data into a spreadsheet application from a website table
- Import data into a spreadsheet application from a database table

3.1 Importing Data Sets Introduction

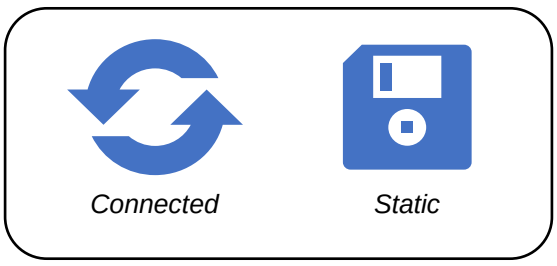
Concepts

The data sets that you want to analyse may be stored in various formats in different programs and sources. They may contain a few rows of data or thousands of rows of data. Common sources of data may include text files, spreadsheets, the web and databases.



You can bring data stored in different formats in various programs into a spreadsheet application such as Microsoft Excel, where you can perform further analysis activities, generate reports and create dashboards. In some cases, you will need to export data from the program or system where it resides before you can import it into Excel. In other cases, you can import data from external sources directly into Excel from within Excel.

When you are importing data into Excel you can make a permanent connection to the external data source so that the imported data can be updated (refreshed). Imported data, when refreshed, is updated with any changes made to the external data source since the last refresh. You can also import data without making a connection to the external data source, in which case the imported data remains static regardless of any changes made to the external data source.



3.2 IMPORTING DATA FROM TEXT FILES

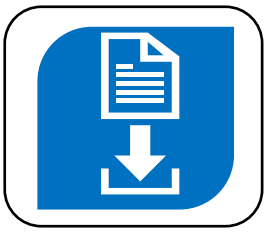
Concepts

In many cases when data is exported from other programs or systems, it is saved in a text file format. Two commonly used text file formats are:

- A comma-separated values (CSV) text file:
  - This contains one page of text with little or no formatting. Each column of data is separated by delimiters, which are typically commas, but they can be another character such as a semi colon, a tab or space. The file extension is .csv.

- A delimited text file:
  - This is similar to a CSV file, except that the columns of data are typically separated by the TAB character rather than another character delimiter. Columns are all the same width in this file type, so you can manually insert columns where you want. The file extension is .txt.

You can import data from a text file into Microsoft Excel by importing the data as an external data range. This makes a permanent connection to the source file that allows the data to be refreshed. Or you can import data from a text file into Microsoft Excel by opening the file in Excel. This doesn't make a permanent connection to the source file, which means the data is static and can't be updated.



Steps

To import data from a **text file** into a **worksheet** in **Excel**:

1. Open the **Customers.xlsx** workbook.

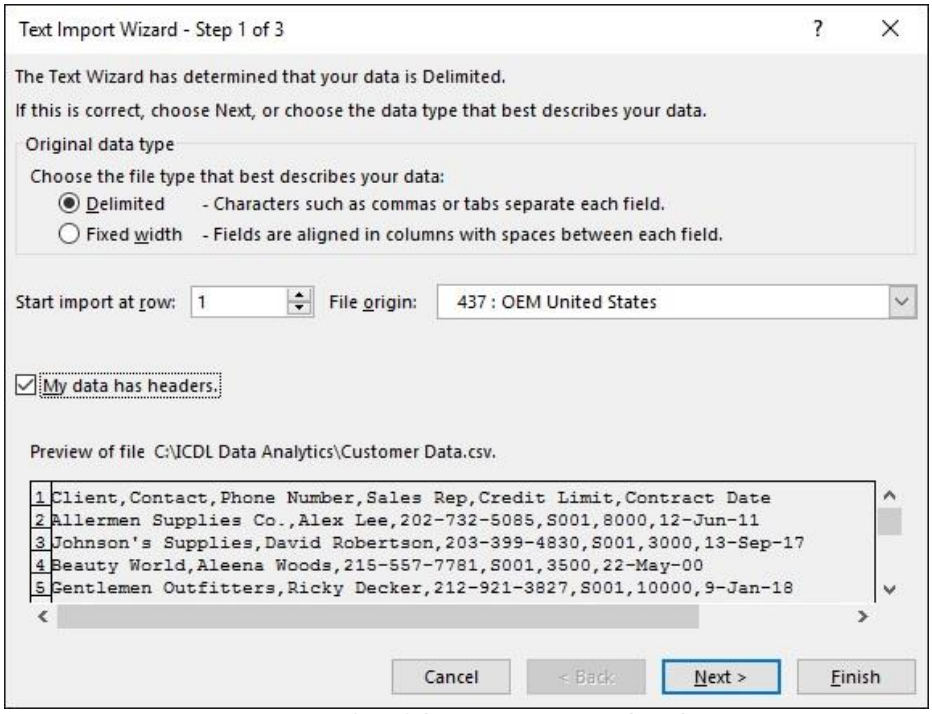
2. On the **Data** tab, in the **Get External Data** group, click **From Text**.

3. Locate and select the file **Customer Data.csv** file.

4. Click **Import**.

5. In the first dialog box of the **Text Import Wizard**, select **Delimited** under **Original data type** and check the option **My data has headers**.

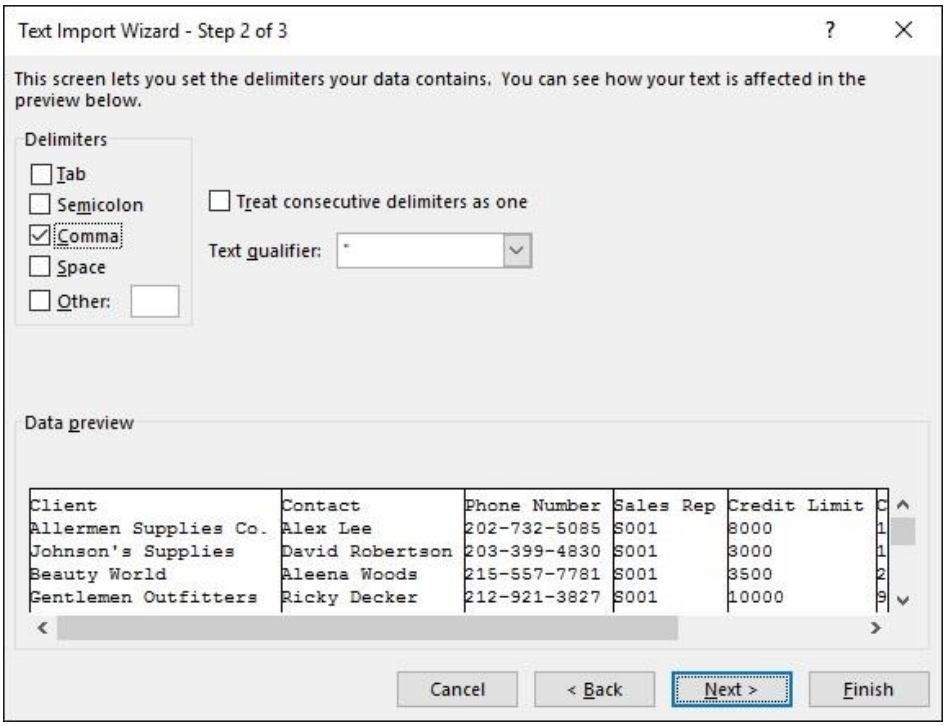
   ***Note**: If importing a .txt file, select the **Fixed width** radio button and follow the steps in the **Text Import Wizard**.*



Step 1 of 3 of the Text Import Wizard

6. Click **Next**.

7. In the second dialog box of the **Text Import Wizard**, check the option **Comma** under **Delimiters**.

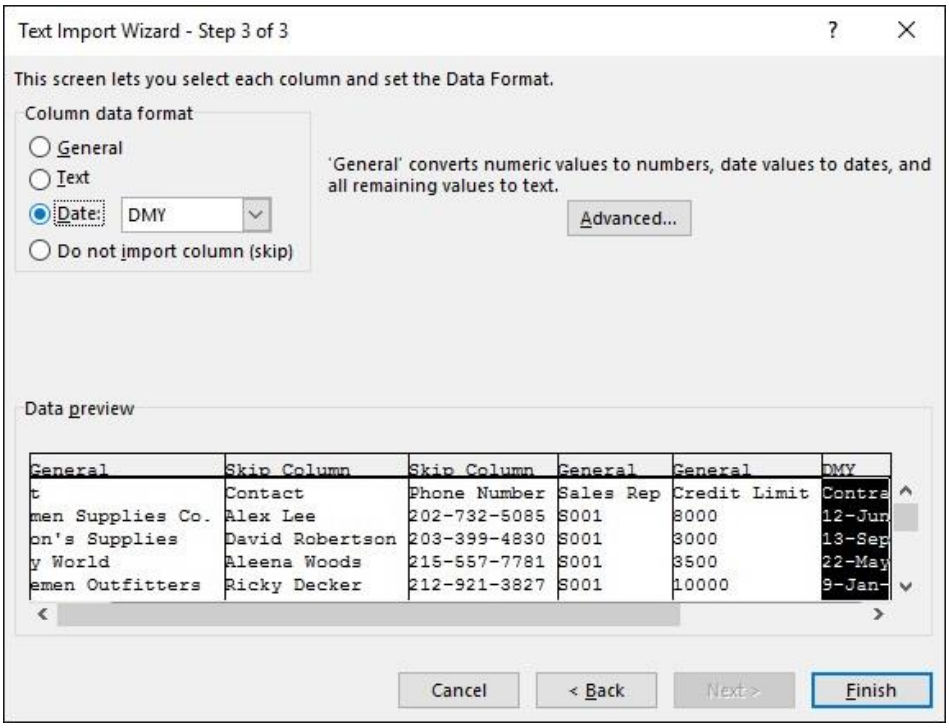


Step 2 of 3 of the Text Import Wizard

8. Click **Next**.

9. In the third dialog box of the **Text Import Wizard**, select the required data format for the columns and specify the columns to skip.

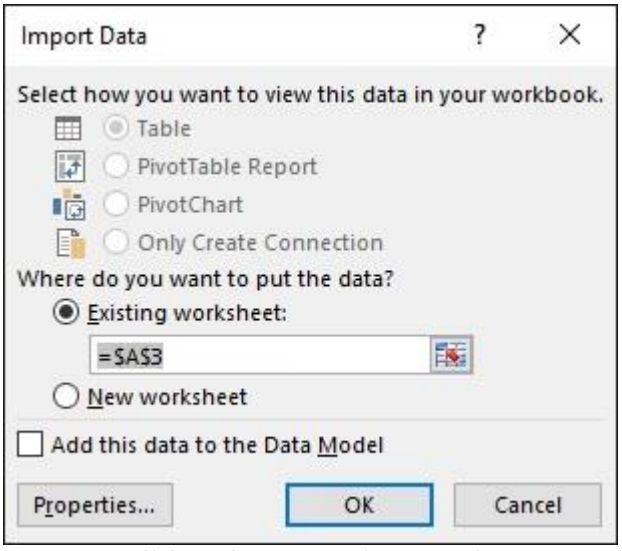
    a. Select the **Contact** column and select the option **Do not import column (skip)**.
    b. Select the **Phone Number** column and select the option **Do not import column (skip)**.
    c. Select the **Contract Date** column and select the option **Date**.
    d. Ensure the date format is **DMY**.



Step 3 of 3 of the Text Import Wizard

10. Click **Finish** when you have made your selections.

11. In the **Import Data** dialog box, ensure the option **Existing worksheet** is selected and click on cell **A3**.



Specifying Placement of Imported Data

12. Click **OK**.

The data is imported into the worksheet.

	A	B	C	D
1	**Customer Details**			
2				
3	Client	Sales Rep	Credit Limit	Contract Date
4	Allermen Supplies Co.	S001	8000	12-Jun-11
5	Johnson's Supplies	S001	3000	13-Sep-17
6	Beauty World	S001	3500	22-May-00
7	Gentlemen Outfitters	S001	10000	9-Jan-18
8	Wheelers	S001	5000	7-Apr-11
9	Joe's Sweets	S002	1500	22-Mar-00
10	Marty's Mart	S002	7000	15-Dec-09
11	Showcase Cosmetics	S002	1900	1-Dec-09
12	Colours World	S002	3000	9-Nov-09
13	Sports & Leisure Supply House	S002	4300	15-Oct-00

Imported Comma Delimited Data

***Note**:* To refresh the data, right-click on the data set and choose **Refresh**.

*If there is active content in the workbook, the **Message Bar** displays security alerts. To proceed, click **Enable Content**.*

***Note:** These steps apply to Microsoft Excel 2016 as part of Microsoft Office 2016 purchased as a stand-alone product. Other versions of Microsoft Excel 2016, such as Microsoft Excel 2016 as part of an Office 365 subscription, may include different functionality and steps. For example, the **Get & Transform Data** option in Microsoft Excel 2016 in Microsoft Office 365 follows different steps for importing data.*

For versions of Excel 2016 that include the **Get & Transform Data** option, the steps to import a text file are:

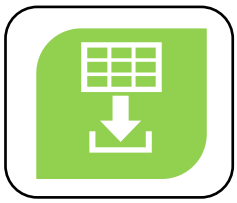
1. In the workbook, click **From Text/CSV** in the **Get & Transform Data** group on the **Data** tab.
2. Browse to the location of the file and select the file **Customer Data.csv**.
3. Click **Import**.
4. Click **Load**.

The Get & Transform functionality automatically detects column delimiters including identifying column names and data types.

3.3 IMPORTING DATA FROM SPREADSHEETS

Concepts

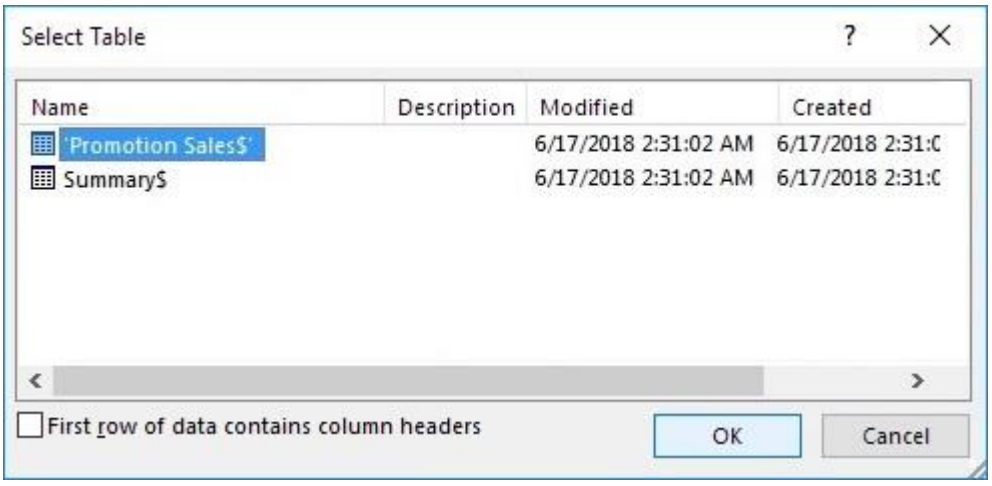
When you are managing and analysing data you may want to import data from one workbook into a different workbook in Microsoft Excel – for example, if you have existing sales or finance reports. You can do this and create a dynamic connection between the two workbooks. This allows you to refresh the connection between the workbooks to ensure the imported data is up-to-date.



Steps

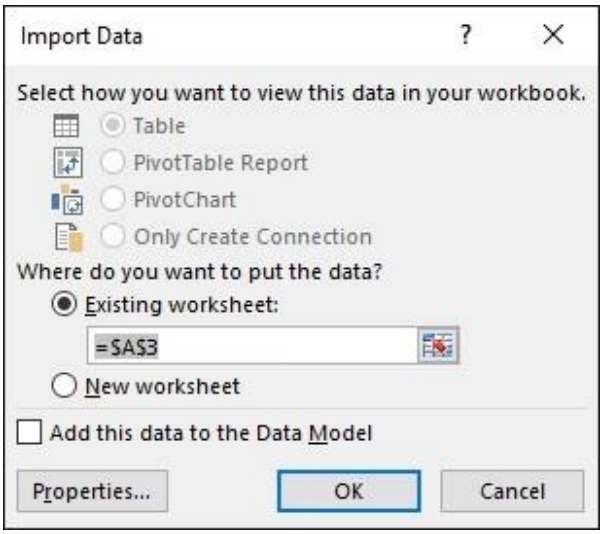
To import data from another **spreadsheet** into a **worksheet** in **Excel**:

1. Open the **Trips.xlsx** workbook.
2. On the **Data** tab, in the **Get External Data** group, click **Existing Connections**.
3. In the **Existing Connections** dialog box, click **Browse for More**.
4. Locate and select the **Ticket Sales.xlsx** workbook.
5. Click **Open**.
6. In the **Select Table** dialog box, select **Promotion Sales**.



Selecting Worksheet to Import

7. Click **OK**.
8. In the **Import Data** dialog box, ensure the option **Existing worksheet** is selected and click on cell **A3**.



Specifying Placement of Imported Data

9. Click **OK**.

The data set is imported in table format.

	A	B	C	D	E	F	G	H
1	Trip Details - Year 2017							
2								
3	Date	First	Last	Office	Destination	Amount	Tickets	Commission
4	3/1/2017	Antonio	Rommero	Spring Lake Park	Washington, D.C.	399	1	Yes
5	16/2/2017	LeAnne	Chang	Maplewood	Los Angeles	295	2	No
6	19/2/2017	Julie	Fredricks	Minneapolis	New York	349	2	Yes
7	27/2/2017	Janelle	Olness	Fargo	Washington, D.C.	399	2	Yes
8	27/3/2017	Anne	Olson	Brooklyn Center	New York	349	1	Yes
9	27/3/2017	Anna	Torkleson	Brainerd	New York	349	2	Yes
10	30/3/2017	Maria	Rodriguez	Hibbing	Washington, D.C.	399	1	Yes
11	20/4/2017	Darlene	Clay	St. Paul	St. Louis	299	5	Yes
12	30/4/2017	Jason	Trent	Hibbing	Dallas	187	2	Yes
13	5/5/2017	Brian	Kipp	Duluth	Norfolk	469	1	Yes

Imported Spreadsheet Data

***Note**: To refresh the data, right-click on the data set and choose **Refresh**.*

***Note:** These steps apply to Microsoft Excel 2016 as part of Microsoft Office 2016 purchased as a stand-alone product. Other versions of Microsoft Excel 2016, such as Microsoft Excel 2016 as part of an Office 365 subscription, may include different functionality and steps. For example, the **Get & Transform Data** option in Microsoft Excel 2016 in Microsoft Office 365 follows different steps for importing data.*

For versions of Excel 2016 that include the **Get & Transform Data** option, the steps to import a spreadsheet are:

1. In the workbook, click **Get Data** in the **Get & Transform Data** group on the **Data** tab.
2. Click **Get File** and then click **Get Workbook**.
3. Browse to the location of the file and select the file, **Ticket Sales.xlsx**.
4. Click **Import**.
5. Click **Promotion Sales**.
6. Click **Load**.

3.4 Importing Data from Website Tables

Concepts

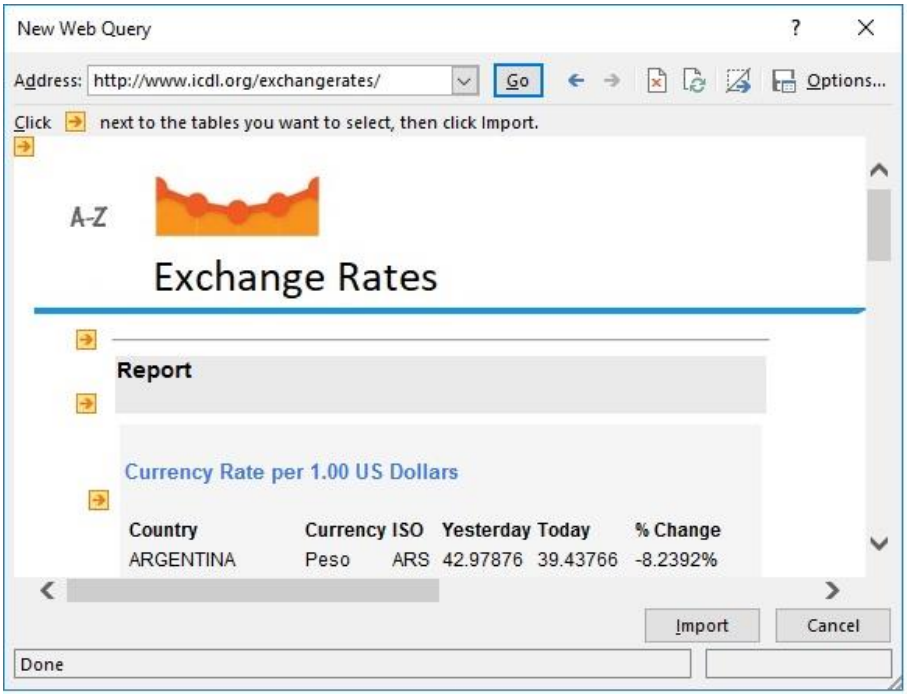
You may want to import information from web pages into Microsoft Excel for analysis. For example, you may want to import stock information directly from a web page or company sales figures from an internal company web page. Depending on your needs, you can import data that you can refresh, or you can import data and make it static.

To import a web table into Excel, you can use the **Web Query** feature in Excel. This feature finds all the tables on a web page and allows you to select the tables you want to import. The tables from the web page can be dynamically updated in the spreadsheet.

Steps

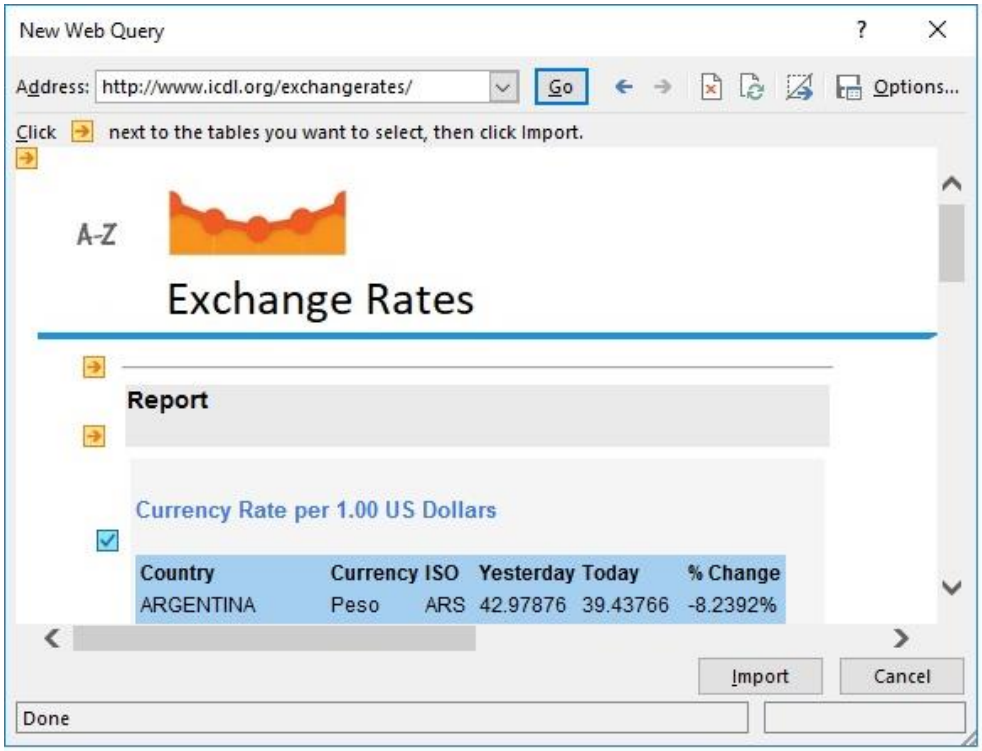
To import data from a **web page** into a **worksheet** in **Excel**:

1. Open the **Rates.xlsx** workbook.
2. On the **Data** tab, in the **Get External Data** group, click **From Web**.
3. In the **New Web Query** dialog box, click in the **Address** text box and type the website address **http://www.icdl.org/exchangerates** and click **Go**.



New Web Query Dialog Box

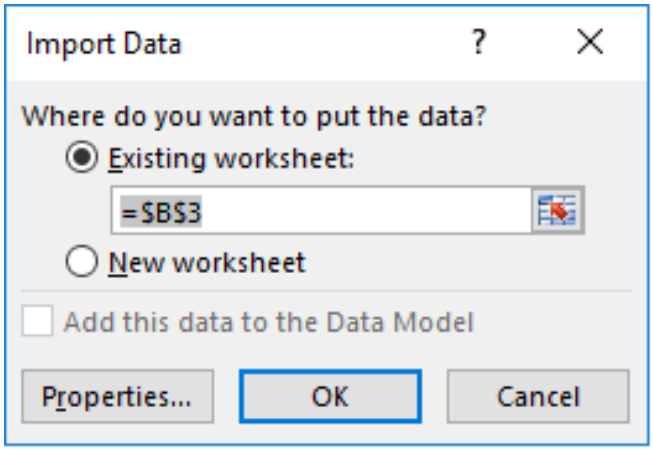
4. Click the table selection arrow next to the table to import.



Selecting a Web Table

5. Click **Import**.

6. In the **Import Data** dialog box, ensure the option **Existing worksheet** is selected and click on cell **B3**.



Specifying Placement of Imported Web Table

7. Click **OK**.

The data is imported into the worksheet in Excel.

	A	B	C	D	E	F	G
1	**Rates Web Table**						
2							
3		Country	Currency	ISO	Yesterday	Today	% Change
4		ARGENTINA	Peso	ARS	42.97876	39.43766	-8.24%
5		AUSTRALIA	Dollar	AUD	1.391824	1.396352	0.33%
6		AUSTRIA	Euro	EUR	0.901678	0.879489	-2.46%
7		BELGIUM	Euro	EUR	0.901678	0.879489	-2.46%
8		BRAZIL	Real	BRL	3.882046	3.76415	-3.04%
9		CANADA	Dollar	CAD	1.327511	1.328991	0.11%
10		CHILE	Peso	CLP	763.73	702.8475	-7.97%
11		CHINA	Yuan	CNY	6.966121	6.791871	-2.50%
12		CZECH REP.	Koruna	CZK	22.4297	22.51965	0.40%
13		DENMARK	Krone	DKK	6.68725	6.522686	-2.46%
14		EUROPEAN UNION	Euro	EUR	0.901678	0.879489	-2.46%

Imported Web Table

Excel will attempt to import the web data as it appears in the web table. However, the less structured the web data, the more formatting will be required in Excel to prepare the data for analysis.

***Note**: To refresh the data, right-click on the data set and click **Refresh**.*

***Note:** These steps apply to Microsoft Excel 2016 as part of Microsoft Office 2016 purchased as a stand-alone product. Other versions of Microsoft Excel 2016, such as Microsoft Excel 2016 as part of an Office 365 subscription, may include different functionality and steps. For example, the **Get & Transform Data** option in Microsoft Excel 2016 in Microsoft Office 365 follows different steps for importing data.*

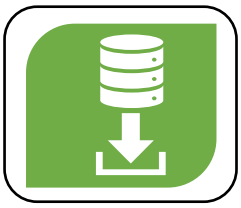
For versions of Excel 2016 that include the **Get & Transform Data** option, the steps to import a web table are:

1. In the workbook, click **From Web** in the **Get & Transform Data** group on the **Data** tab.
2. Enter the URL http://www.icdl.org/exchangerates and click **OK**.
3. Select the table to import.
4. Click **Load**.

3.5 Importing Data from Database Tables

Concepts

A database management system (DBMS) provides a structured method of storing data by preventing redundancy and maintaining integrity. In general, they are used to store large amounts of data in tables and have extensive querying abilities that break data down into manageable chunks. However, the interface of a DBMS is limited, which means it can be time consuming to manipulate data into meaningful reports.

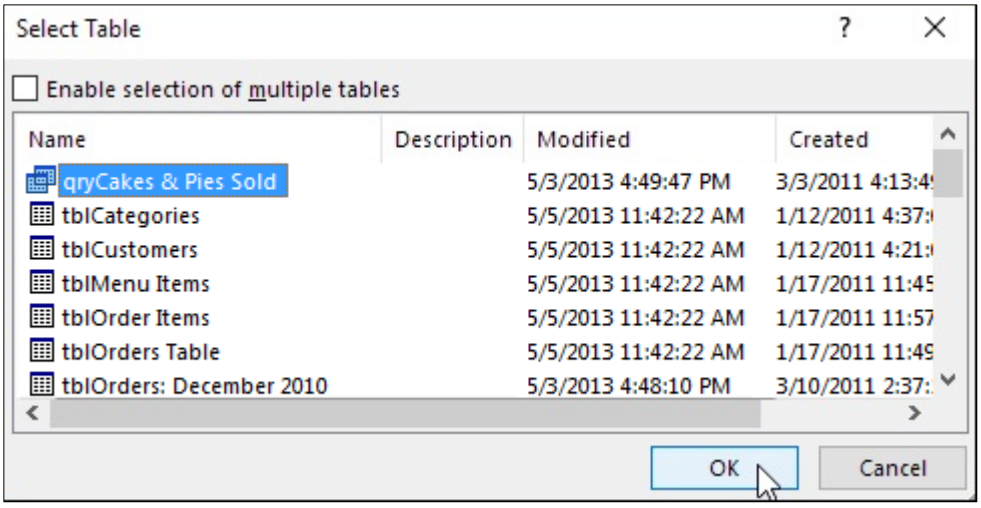


You can import database tables from a DBMS such as Microsoft Access into Microsoft Excel. In Excel you can more easily filter, chart, and analyse a data set to produce meaningful reports. Many organisations use Microsoft Excel and Access almost interchangeably, feeding data between the two applications.

Steps

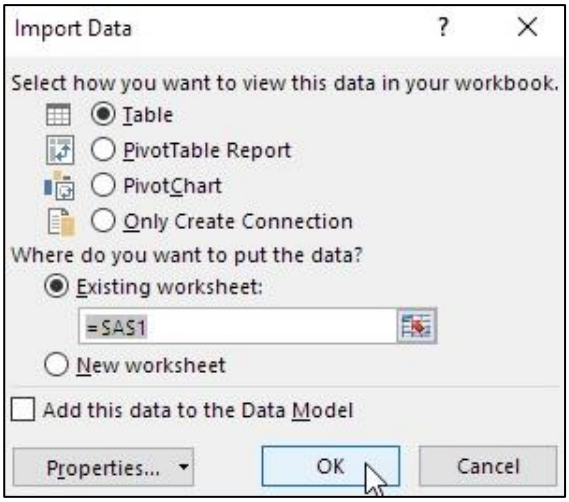
To import data from a **database table** into a **worksheet** in **Excel**:

1. Open the **Customer Details.xlsx** workbook.
2. On the **Data** tab, in the **Get External Data** group, click **From Access**.
3. Locate and select the **BakeryDatabase** database.
4. Click **Open**.
5. Select the table **tblCustomers** and Click **OK**.



Select Table Dialog Box

6. In the **Import Data** dialog box, ensure the option **Table** is selected.



Specifying Placement of Imported Database Table

7. Click **OK**.

   The table is imported into the worksheet.

	A	B	C	D
1	ID	First Name	Last Name	Street Address
2	1	Tracey	Beckham	7 East Walker Dr.
3	2	Lucinda	George	789 Brewer St.
4	3	Jerrod	Smith	211 St. George Ave
5	4	Brett	Newkirk	47 Hillsborough St.
6	5	Chloe	Jones	23 Solo Ln.
7	6	Quinton	Boyd	4 Cypress Cr.
8	7	Alex	Hinton	1011 Hodge Ln.
9	8	Nisha	Hall	123 Huntington St.
10	9	Hillary	Clayton	2516 Newman

Imported Database Table

***Note**: To refresh the data, right-click on the data set and choose **Refresh**.*

***Note:** These steps apply to Microsoft Excel 2016 as part of Microsoft Office 2016 purchased as a stand-alone product. Other versions of Microsoft Excel 2016, such as Microsoft Excel 2016 as part of an Office 365 subscription, may include different functionality and steps. For example, the **Get & Transform Data** option in Microsoft Excel 2016 in Microsoft Office 365 follows different steps for importing data.*

For versions of Excel 2016 that include the **Get & Transform Data** option, the steps to import a database table are:

1. Click **Get Data** in the **Get & Transform Data** group on the **Data** tab and select **From Database**, **From Microsoft Access Database**.
2. Browse to the location of the file and select the file.
3. Click **Import**.
4. Select the table to import.
5. Click **Load**.

3.6 REVIEW EXERCISE

1. Create a new workbook.
2. Import the **Real Estate.csv** comma delimited text file.
3. Specify that the text file contain headers.
4. Specify the Delimiter as **Comma**.
5. Skip the columns **Sale Date**, **Latitude** and **Longitude**.
6. Accept all the other default settings and place the data from cell **A1** onwards.
7. Save the workbook as **Real Estate Imported**.

LESSON 4 –
SHAPING DATA SETS

In this section, you will learn how to:

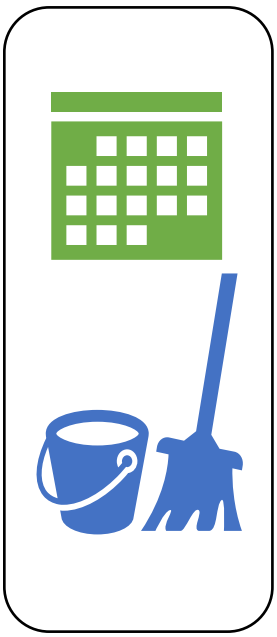
- Remove duplicate data
- Validate that given values belong to a reference data set using the VLOOKUP function
- Validate that given values belong to a specific range using one or more IF functions
- Extract values from a string using text functions

4.1 SHAPING DATA SETS INTRODUCTION

Concepts

To get meaningful information from data analysis you need to ensure that you are working with a good quality data set. For example, data imported into Microsoft Excel from external sources, such as text files, spreadsheets, web pages or databases may contain errors or inconsistencies or may not be arranged as needed.

You will often need to **shape and prepare** the data set before you can analyse it, which happens in the **data preparation** phase of the **data analytics process**. Before you begin shaping and preparing the data set it is good practice to save a backup copy of the original data set. This includes **cleaning** the data to fix any issues that may affect the quality of the analysis. The cleaning tasks will depend on the type of data and its quality but may involve:



- Fixing spelling mistakes.
- Identifying and removing duplicate data.
- Finding and replacing text
- Fixing incorrect formatting.
- Changing the case of text.
- Removing extra spaces from text.
- Fixing number or date and time formats.
- Merging and splitting columns.
- Transforming and rearranging columns and rows.
- Reconciling data by comparing or matching values.

Spreadsheet applications contain built-in features for some of the cleaning tasks and it is good practice to complete these types of tasks first. For example, in Excel you can

- clean up spelling mistakes using the **Spell Checker** tool.
- fix incorrect formatting or replace text using the **Find and Replace** tool.
- remove duplicate rows using the **Remove Duplicates** tool.

For other tasks, such as extracting data, you may need to manipulate imported data columns, which may involve changing imported values into new values using formulas.

4.2 Removing Duplicate Data

Concepts

Data duplication can occur in data sets when records are entered more than once or consolidated from multiple sources. Duplicates may occur for one or more fields. If you want to limit the records to unique data for some fields or unique records for the entire data set, you can use Excel's built-in **Remove Duplicates** tool. This tool finds and removes exact duplicates.

Steps

To remove duplicates using Excel's built-in **Remove Duplicates** tool:

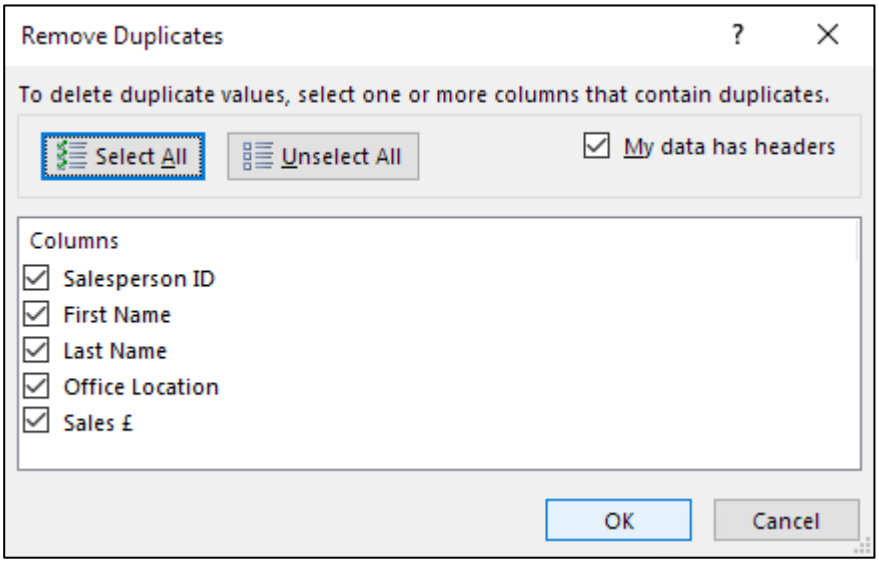
1. Open the **Duplicate Records.xlsx** workbook.

   It is a good idea to copy the original data set to another worksheet or workbook before removing duplicate values, as the tool will permanently delete duplicates.

2. Copy the range **A4:E11** and paste the copied range starting from cell **G4** onwards.

3. Select any cell within the data set that starts from cell G4.

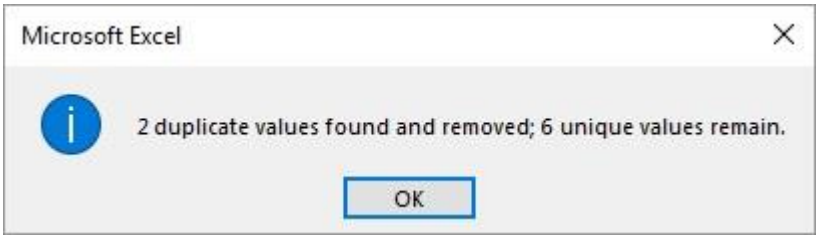
4. On the **Data** tab, in the **Data Tools** group, click **Remove Duplicates**.



Remove Duplicates Dialog Box

5. Click **OK**.

   A message will appear to indicate how many duplicate values were removed, or how many unique values remain.



Remove Duplicates Message

6. Click **OK**.

Salesperson ID	First Name	Last Name	Office Location	Sales £
A147	Roberta	Miller	London	250,500
A148	Marie	Jones	Cardiff	175,900
A149	Ann	Connolly	Glasgow	164,800
A150	Michael	Green	Birmingham	210,400
A151	Paula	West	Bristol	175,600
A152	James	Milner	London	295,300

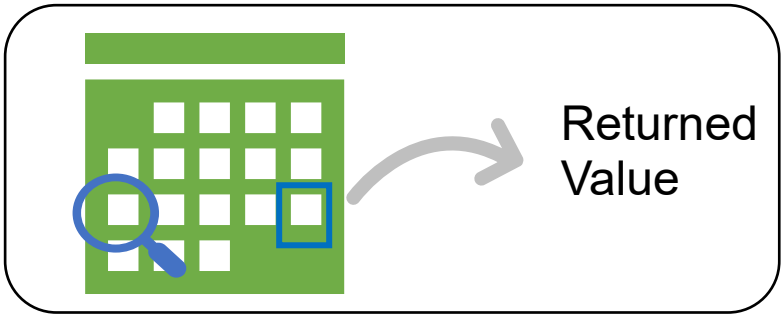
Unique Records

When duplicates are removed, the first occurrence of the value in the data set is kept, but other identical values are deleted.

4.3 VALIDATING DATA USING VLOOKUP

Concepts

In Excel you can use the **VLOOKUP** function to validate data, which can save a huge amount of time when reconciling data. The **VLOOKUP** function looks up a value in the first column of a specified range of cells and returns a value from a specified column in the same row.



You can use the **VLOOKUP** function to look up given values to quickly find specific values in a data set. You can use this to automatically verify that you are using the correct data. For example, you can use the **VLOOKUP** function to verify if a value, such as a country code matches a list of values in a reference data set, such as a list of approved country codes.

And you can perform calculations or display results using the values returned. For example, you can use the **VLOOKUP** function to find a commission rate in a table based on a specified value in the data set. You can calculate the amount of commission to pay based on the specified value multiplied by the rate returned by the **VLOOKUP** function.

The syntax for the **VLOOKUP** function is as follows:

VLOOKUP(lookup_value, table_array, col_index_num, [range_lookup])

- **lookup_value** is the value you want to look for in a specified range of cells.
- **table_array** is the range of cells containing the value you want to return.
- **col_index_num** is the column number in the range containing the value you want to return.
- **range_lookup** specifies whether you want to return an exact match (**FALSE**) or an approximate match (**TRUE** or omitted). **VLOOKUP** in exact match mode is useful when using a unique key as a lookup value, for example, an order ID. **VLOOKUP** in approximate match mode is useful when looking for the best match for a given lookup value – for example finding the right commission rate based on a sales figure.

The VLOOKUP function searches vertically down the first column of the specified table array to find the specified lookup value and then looks across the corresponding row to find the specified column and returns the value from the cell where the row and column meet.

Steps

To use the VLOOKUP function to validate data:

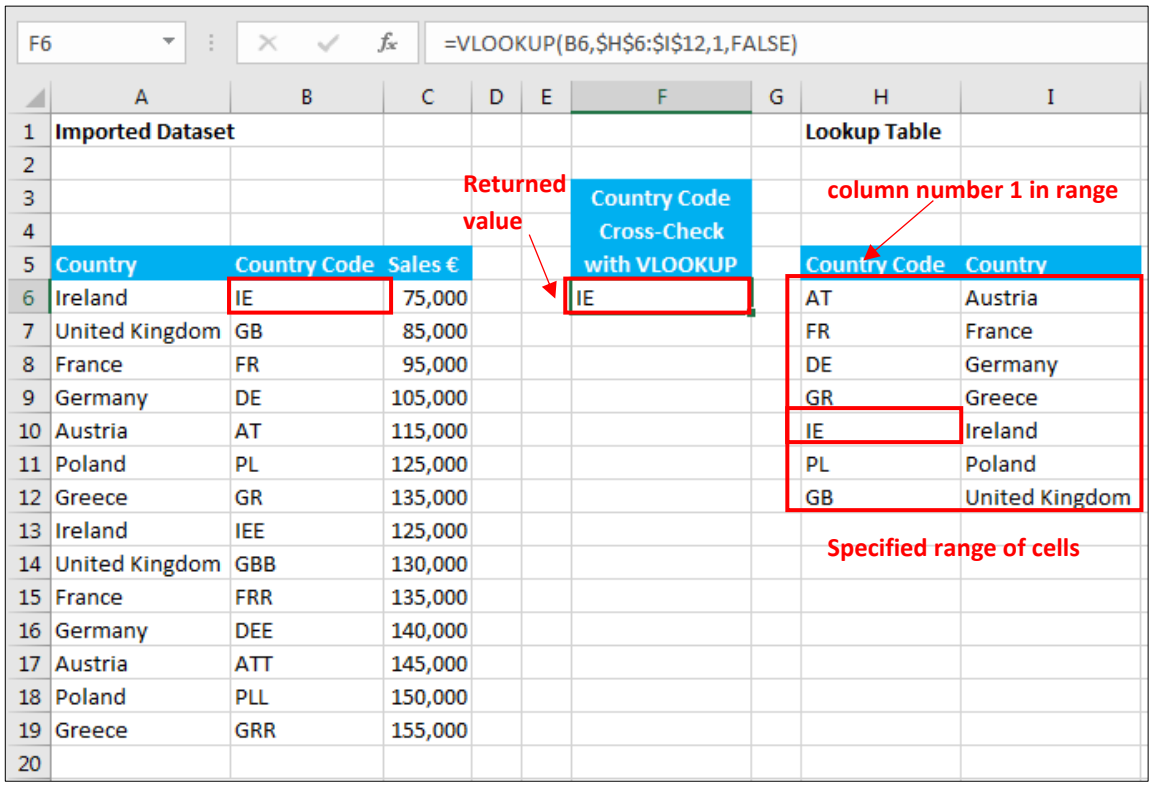
1. Open the **Value Validation.xlsx** workbook.
   The **VLOOKUP** worksheet contains imported regional sales data in the range **A5:C19**. Some of the entries in the **Country Code** column in this range are inconsistent. Using the **VLOOKUP** function, you want to validate whether a given country code in the **Country Code** column exists in the reference data set range (table array) **H6:I12**. You want to specify that only an exact match is returned.

	A	B	C	D	E	F	G	H	I
1	**Imported Dataset**							**Lookup Table**	
2									
3	Data set to validate					Country Code Cross-Check with VLOOKUP			
4								Reference data set	
5	Country	Country Code	Sales €					Country Code	Country
6	Ireland	IE	75,000					AT	Austria
7	United Kingdom	GB	85,000					FR	France
8	France	FR	95,000					DE	Germany
9	Germany	DE	105,000					GR	Greece
10	Austria	AT	115,000					IE	Ireland
11	Poland	PL	125,000					PL	Poland
12	Greece	GR	135,000					GB	United Kingdom
13	Ireland	IEE	125,000						
14	United Kingdom	GBB	130,000						
15	France	FRR	135,000						
16	Germany	DEE	140,000						
17	Austria	ATT	145,000						
18	Poland	PLL	150,000						
19	Greece	GRR	155,000						

Imported Data with Values to Validate Against a Reference Data set

2. Select the cell **F6**.

3. Type **=VLOOKUP(B6, $H$6:$I$12, 1, FALSE)** and press **Enter**.



	A	B	C	D	E	F	G	H	I
1	**Imported Dataset**							**Lookup Table**	
2									
3									
4									
5	**Country**	**Country Code**	**Sales €**					**Country Code**	**Country**
6	Ireland	IE	75,000			IE		AT	Austria
7	United Kingdom	GB	85,000					FR	France
8	France	FR	95,000					DE	Germany
9	Germany	DE	105,000					GR	Greece
10	Austria	AT	115,000					IE	Ireland
11	Poland	PL	125,000					PL	Poland
12	Greece	GR	135,000					GB	United Kingdom
13	Ireland	IEE	125,000						
14	United Kingdom	GBB	130,000						
15	France	FRR	135,000						
16	Germany	DEE	140,000						
17	Austria	ATT	145,000						
18	Poland	PLL	150,000						
19	Greece	GRR	155,000						
20									

The **VLOOKUP** function looks up the value specified in the cell **B6** (**IE**) in the **first column** of the specified range of cells (**H6:I12**). It finds **IE** in cell **H10**. It returns the value from **row 10** in column 1, specified by the column number index. For the range **H6:I12** this is column **H**.

***Note**: The $ sign before the row and column coordinates makes the cell references absolute so they won't change.*

4. Copy the formula to the cell range **F7:F19**.

   If the imported country code exists in the first column of the reference data set, the function will return the country code otherwise it will return **#N/A** to indicate that the value is not found.

	A	B	C	D	E	F	G	H	I
1	**Imported Dataset**							**Lookup Table**	
2									
3						Country Code Cross-Check with VLOOKUP			
4		Values not found							
5	Country	Country Code	Sales €					Country Code	Country
6	Ireland	IE	75,000			IE		AT	Austria
7	United Kingdom	GB	85,000			GB		FR	France
8	France	FR	95,000			FR		DE	Germany
9	Germany	DE	105,000			DE		GR	Greece
10	Austria	AT	115,000			AT		IE	Ireland
11	Poland	PL	125,000			PL		PL	Poland
12	Greece	GR	135,000			GR		GB	United Kingdom
13	Ireland	IEE	125,000			#N/A			
14	United Kingdom	GBB	130,000			#N/A			
15	France	FRR	135,000			#N/A			
16	Germany	DEE	140,000			#N/A			
17	Austria	ATT	145,000			#N/A			
18	Poland	PLL	150,000			#N/A			
19	Greece	GRR	155,000			#N/A			

Cross-checking with VLOOKUP

To generate the correct country code for the countries listed in the range A13:A19:

1. Select the cell **B13**.
2. Type **=VLOOKUP(A13, $H$6:$I$12, 1)** and press **Enter**.
3. Copy the formula to the cell range **B14:B19**.
4. Check that the generated data is the same as that shown in the range **A24** onwards.

	A	B	C
21	**Corrected Dataset**		
22			
23	Country	Country Code	Sales €
24	Ireland	IE	75,000
25	United Kingdom	GB	85,000
26	France	FR	95,000
27	Germany	DE	105,000
28	Austria	AT	115,000
29	Poland	PL	125,000
30	Greece	GR	135,000
31	Ireland	IE	125,000
32	United Kingdom	GB	130,000
33	France	FR	135,000
34	Germany	DE	140,000
35	Austria	AT	145,000
36	Poland	PL	150,000
37	Greece	GR	155,000

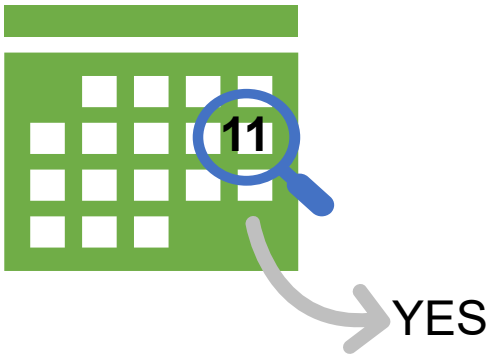
Corrected Data set

4.4 VALIDATING DATA USING IF FUNCTIONS

Concepts

In Excel you can use the **IF** function to validate the data set – for example to check whether values are in a particular range. The **IF** function is one of the most popular and useful functions in Excel. You use an **IF** statement to evaluate a condition and to return one value if the condition is met, and another value if the condition is not met.

For example:
Check **IF** the value is **greater than 10**,
IF it is **greater than 10** return **YES**,
IF it is **not greater than 10** then return **NO**



The syntax for the **IF** function is as follows:

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

- logical_**test** is a value or logical expression that can be evaluated as **TRUE** or **FALSE**.
- **value_if_true** is the value to return when logical_test evaluates to **TRUE**. This argument is optional.
- **value_if_false** is the value to return when logical_test evaluates to **FALSE**. This argument is optional.

When analysing data, you often need to evaluate several conditions at the same time, which means you need to create more complex logical tests.

To create more complex logical tests, you can use **AND** or **OR** functions in the logical test of an **IF** function. If the logical test contains the **AND** function, **TRUE** is returned if all the conditions are met; otherwise **FALSE** is returned. If the logical test contains the **OR** function, **TRUE** is returned if any of the conditions is met; otherwise **FALSE** is returned. You can combine **AND/OR** functions and use multiple logical functions to meet the requirements of your data analysis.

You can also use multiple **IF** functions together to create complex logical tests. Multiple **IF** functions used together in one formula are known as nested **IF** functions. They are useful for returning 3 or more different results. Additional **IF** statements are included in the **value_if_true** and **value_if_false** arguments of the **IF** function.

Steps

To use one or more **IF** functions to validate data:

1. Open the **Value Validation.xlsx** workbook.

   The **IF** worksheet contains regional sales data in the range **A5:B10**. These values need to be validated to check whether they are within the maximum range specified in cell **F6**, which is 800,000.

	A	B	C	D	E	F
1	**Validate that given values belong to a specified range using one or more if functions**					
2						
3				Checked with IF function		
4						
5	Location	Sales		Sales		Maximum Range
6	France	795,000				800,000
7	Germany	823,000				
8	Poland	500,600				
9	Greece	685,200				
10	Austria	475,600				

Sales Data to Validate Against a Maximum Range

2. Select the cell **D6**.

3. Type **=IF(B6>$F$6,"Error",B6)** and press **Enter**.

   If the sales amount exceeds the maximum range, the function will return the message **Error**, otherwise it will return the sales amount itself.

4. Copy the formula to the cell range **D7:D10**.

   In this example one location, Germany, has exceeded the maximum range.

	A	B	C	D	E	F
5	Location	Sales		Sales		Maximum Range
6	France	795,000		795,000		800,000
7	Germany	823,000		Error		
8	Poland	500,600		500,600		
9	Greece	685,200		685,200		
10	Austria	475,600		475,600		

Validate for Maximum Range

The **IF** worksheet also contains regional sales data in the range **A15:B20**. These values need to be validated to check whether they are within the minimum and maximum range specified in cells **F16** and **G16**.

	A	B	C	D	E	F	G
15	Location	Sales		Sales		Minimum Range	Maximum Range
16	France	795,000				250,000	800,000
17	Germany	823,000					
18	Poland	500,600					
19	Greece	685,200					
20	Austria	475,600					

Sales Data to Validate Against Minimum and Maximum Range

5. Select the cell **D16**.

6. One way to check whether the values are in the minimum and maximum range specified in cells **F16** and **G16** is to use the AND logical test.

   To do this, type **=IF(AND(B16>=$F$16,B16<=$G$16),B16,"Error")** and press **Enter**.

   If the sales amount is within the minimum and maximum range, the function will return the sales amount, otherwise the message **Error** will be returned.

7. Copy the formula to the cell range **D17:D20**.

   In this example the sales value for Germany is outside the minimum and maximum range.

	A	B	C	D	E	F	G
15	Location	Sales		Sales		Minimum Range	Maximum Range
16	France	795,000		795,000		250,000	800,000
17	Germany	823,000		Error			
18	Poland	500,600		500,600			
19	Greece	685,200		685,200			
20	Austria	475,600		475,600			

Validate for Minimum and Maximum Range

***Note**: Another way to check whether the values are in the minimum and maximum range specified in cells F16 and G16 is to use a nested IF function. For example, in cell D16, enter either of the following:*

- ***=IF(B16>$G$16,"Error",IF(B16<$F$16,"Error", IF(B16<$G$16,B16)))***
- ***=IF(B16>$G$16,"Error",IF(B16<$F$16,"Error", IF(B16>$F$16,B16)))***

4.5 EXTRACTING VALUES USING TEXT FUNCTIONS

Concepts

One of the main tasks in cleaning data is to extract parts of data and parse it to separate columns. A common task is to divide the text in a single cell into multiple columns – for example, you may want to split the first name and last name contained in one cell into two separate columns or split an address contained in one cell into several columns.

The table below outlines the common functions in Excel used for extracting text.

Function	Syntax and Description
LEFT	**LEFT(text, [num_chars])** Extracts a specified number of characters from the left side of a string. For example:=LEFT("Ireland",3) returns "Ire".
RIGHT	**RIGHT(text, [num_chars])** Extracts a specified number of characters from the right side of a string. For example: =RIGHT("Ireland",3) returns "and"
MID	**MID(text, start_num, num_chars)** Extracts a specified number of characters from a specified starting position in a string. For example: =MID("Ireland",2,3) returns "rel".
LEN	**LEN(text)** Returns the length of a string as a number of characters. For example: =LEN("Ireland") returns 7.
FIND	**FIND (find_text, within_text, [start_num])** Returns the position, as a number, of one text string inside another string. If the text is not found, FIND returns a #VALUE error. For example: =FIND("land","Ireland",1) returns 4.

The functions can be combined (nested) in Excel to perform particular tasks. The table below outlines some common combinations of these functions for extracting text.

Functions	Syntax and Description
LEFT and LEN	**LEFT(text,LEN(text)-n)** Removes the last n characters from a text string. This is useful when there are values with variable lengths and you want to remove a specified number of characters from the right. For example: =LEFT("IrelandIE",LEN("IrelandIE")-2) returns Ireland.
LEFT and FIND	**LEFT(text,FIND(character,text)-1)** Extracts characters from the left side of a string up until a specified character. For example: =LEFT("Ireland_IE",FIND("_","Ireland_IE")-1) returns Ireland.

Using these functions to clean data involves using the formula to extract the data into a new column, filling down the new column, and converting the formulas in the new column to values.

Steps

To use **LEFT**, **RIGHT** and **LEN** functions to extract values:

1. Open the **Text Functions.xlsx** workbook.

   The cell range **A4:A10** contains the imported country and its relevant country codes. The data is however combined. You need to split the country and the code into separate columns.

	A
1	**Imported Data**
2	
3	Country
4	IrelandIE
5	United KingdomGB
6	FranceFR
7	GermanyDE
8	AustriaAT
9	PolandPL
10	GreeceGR

Imported Country and Code Data

***Note**: To extract the country into the cell **C4**, you need to extract all the entries in cell **A4** except for the last 2 characters. You can do this using the **LEFT** and **LEN** functions. Use the **LEN** function to find the length of the entry in **A4** and then subtract 2.*

2. Select the cell **C4**.

3. Type **=LEFT(A4, LEN(A4)-2)** and press **Enter**.

4. Copy the formula to the cell range **C5:C10**.

5. To extract the country code into the cell **D4**, you need to extract the last 2 characters in cell **A4** using the **RIGHT** function. Select the cell **D4**.

6. Type **=RIGHT(A4, 2)** and press **Enter**.

7. Copy the formula to the cell range **D5:D10**.

	A	B	C	D
1	**Imported Data**			
2				
3	Country		Country Extracted	Country Code Extracted
4	IrelandIE		Ireland	IE
5	United KingdomGB		United Kingdom	GB
6	FranceFR		France	FR
7	GermanyDE		Germany	DE
8	AustriaAT		Austria	AT
9	PolandPL		Poland	PL
10	GreeceGR		Greece	GR

Imported Data Parsed into Separate Columns

8. Copy the range **C4:D10** and paste as values starting from cell **F4** onwards.

	F	G
1	**Cleansed Data**	
2		
3	Country	Country Code
4	Ireland	IE
5	United Kingdom	GB
6	France	FR
7	Germany	DE
8	Austria	AT
9	Poland	PL
10	Greece	GR

Cleansed Data

4.6 REVIEW EXERCISE

1. Open the **Shipping.xlsx** workbook.

   The data set shows order and shipping details for a period of seven days.

   The **Shipping Details** in column **B** consists of shipping days and status.

2. Remove duplicate records in the range **A3:D22**.

3. Enter a formula in **H4** to extract the first number from **Shipping Details** in **B4**.

4. Enter a formula in **I4** to extract the shipping status from **Shipping Details** in **B4**.

   Hint: Use the **LEN** function to find the length of the entry in **B4** and then subtract 4.

5. Enter a formula in **J4** to evaluate whether the **Order Value** in **C4** is 1000 onwards. Display the **Order Value** if the condition is met, otherwise display 0.

   The **Order ID** in column **D** consists of **customer ID**, **country code**, **order month** and **year**.

6. Enter a formula in **K4** to extract the first 3 numbers from **Order ID** in **D4**.

7. Enter a formula in **L4** to extract the 2 characters after the hyphen from **Order ID** in **D4**.

8. Using the table range **P6:Q10**, calculate the commission amount payable based on the **Order Value** (**Order Value * Rate**).

   Hint: Use the **VLOOKUP** function in approximate match mode to find the rate based on the **Order Value**.

9. Copy the formulas in row 4 (in columns H to M) to the rest of the rows in the data set.

G	H	I	J	K	L	M
Date	Shipping Days	Status	Taxable Order Value	Customer ID	Country Code	Commission Amount
1-Jun-2018	3	Shipped	$3,500	316	SG	$105.00
2-Jun-2018	1	Shipped	$4,500	328	UK	$180.00
2-Jun-2018	3	Returned	$3,200	333	AU	$96.00
3-Jun-2018	7	Shipped	$1,250	379	JP	$37.50
3-Jun-2018	2	Lost	$0	340	MY	$0.00
3-Jun-2018	4	Returned	$5,260	395	FR	$210.40
3-Jun-2018	3	Shipped	$0	338	IT	$0.00
4-Jun-2018	1	Shipped	$1,100	363	UK	$33.00
4-Jun-2018	4	Shipped	$2,470	322	GR	$74.10
5-Jun-2018	3	Shipped	$3,000	360	UK	$90.00
5-Jun-2018	1	Shipped	$0	311	PL	$0.00
6-Jun-2018	6	Shipped	$5,340	328	FR	$213.60
7-Jun-2018	4	Shipped	$2,550	347	IN	$76.50
7-Jun-2018	3	Lost	$7,580	376	UK	$454.80
7-Jun-2018	5	Shipped	$0	326	BR	$0.00
7-Jun-2018	4	Shipped	$6,900	349	US	$345.00
7-Jun-2018	4	Delayed	$4,560	301	CN	$182.40

LESSON 5 –
FILTERING DATA SETS

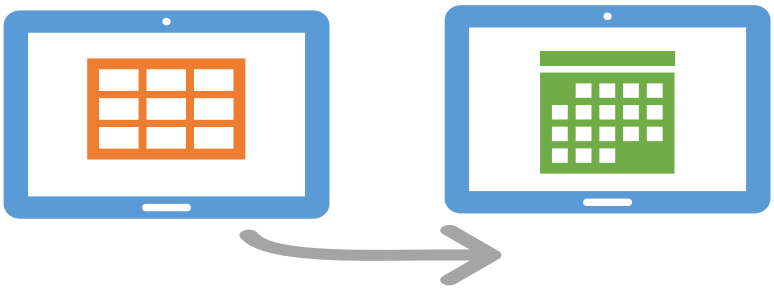
In this section, you will learn how to:

- Format data set as table
- Insert and use table slicers

5.1 Formatting Data Sets as Tables

Concepts

Formatting a data set as a table in Excel helps to add functionality to a worksheet and makes managing and analysing data easier. A table is also an excellent source for a pivot table.



The visual design and layout of a table facilitates the summary and presentation of key numerical information. You can use predefined or custom table styles to quickly format table data in Excel and there are options to modify the table display.

Before formatting a data set as table, you need to organise the data set. Some guidelines include:

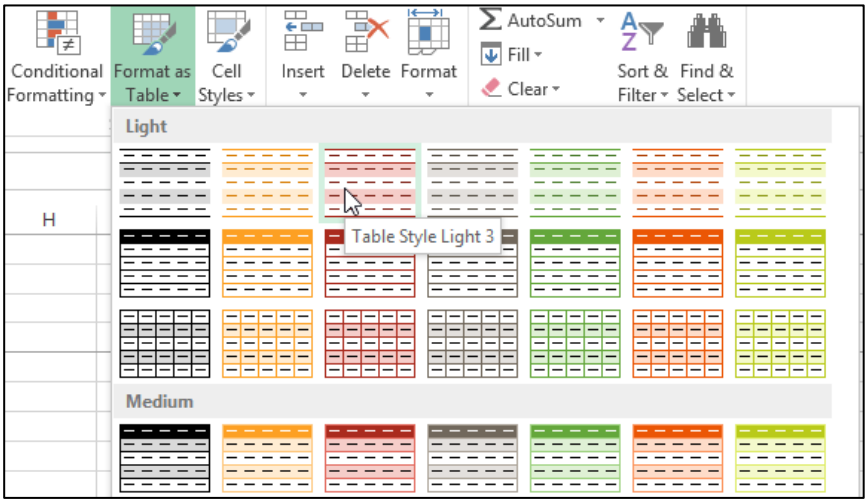
- Organise the data in rows and columns, with each row containing information about one record, such as a sales entry, or inventory transaction.
- Ensure each column contains a descriptive and unique heading in the first row of the data set.
- Ensure each column contains one type of data, such as dates, currency, or text.
- Ensure there are no blank rows and no absolutely blank columns.

Steps

To convert a data set into a table in Excel:

1. Open the **Sales.xlsx** workbook.
2. Click on any cell in the data set.
3. On the **Home** tab, in the **Styles** group, click **Format as Table**.

4. Select a table style from the drop-down menu.



Selecting a Table Style

5. Click **OK i**n the **Format As Table** dialog box.

***Note**: The data range for the table is automatically highlighted, and the range address is listed.*

Format as Table Dialog Box

The cell range is formatted in the selected table style.

	A	B	C	D	E	F	G	H
1	Brewers Enterprise							
2								
3								
4	SalesRep	Date	Mont	Sale	Product	Regior	Yea	Profi
5	Elaine Woods	8/12/2017	Dec	1215	Green Tea	Northeast	2017	450
6	Thomas Lee	5/2/2018	Feb	630	Latte	Southwest	2018	315
7	James Carter	5/12/2017	Dec	1620	Mocha	Northeast	2017	900
8	Frank Edwards	4/2/2018	Feb	1440	Flat White	Central	2018	600
9	Susan Edwards	5/11/2017	Nov	752	Long Black	Northeast	2017	396
10	Jayne Michaels	3/3/2018	Mar	1089	Hot Chocolate	Southwest	2018	605

Data set Formatted as Table

6. On the **Design** tab, check or uncheck the desired options in the **Table Style Options** group.

   These options can affect the table style in various ways, depending on the type of content in the table.

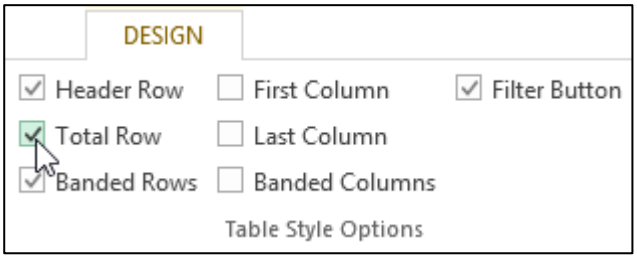


Table Style Options Group

- **Header Row** - Apply or remove formatting from the first row in the table.
- **Total Row** - Add **SUBTOTAL** functions like **SUM**, **AVERAGE**, **COUNT**, **MIN** or **MAX** to the table from a drop-down selection.
- **Banded Rows** - Display odd and even rows with alternating shading for ease of reading.
- **First Column** - Apply or remove formatting from the first column in the table.
- **Last Column** - Apply or remove formatting from the last column in the table.
- **Banded Columns** - Display odd and even columns with alternating shading for ease of reading.
- **Filter Button** - Toggle **AutoFilter** on and off. With filtering enabled in the header row, table data can quickly be sorted or filtered.

5.2 USING TABLE SLICERS

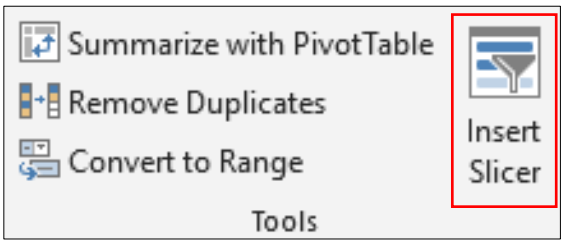
Concepts

Slicers are visual filters that provide buttons that you can quickly click to filter the data in a table. Slicers show at a glance the current filter that is applied. This makes it easy to understand the information that is shown in a filtered table.

Steps

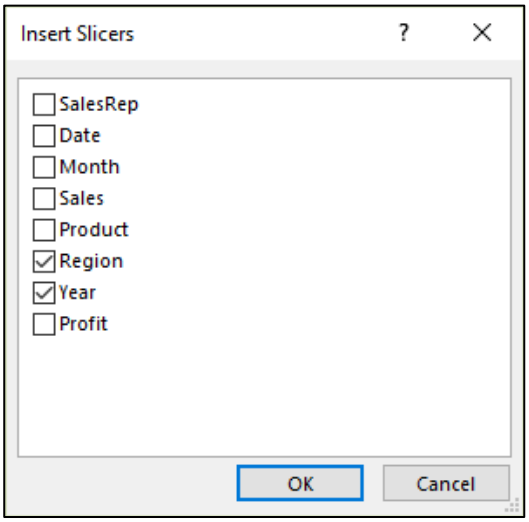
To insert and use a slicer on a table:

1. Open the **Slicers.xlsx** workbook.
2. Click on any cell in the table.
3. On the **Table Tools**, **Design** tab, in the **Tools** group, click **Insert Slicer**.



Insert Slicer icon in Tools group

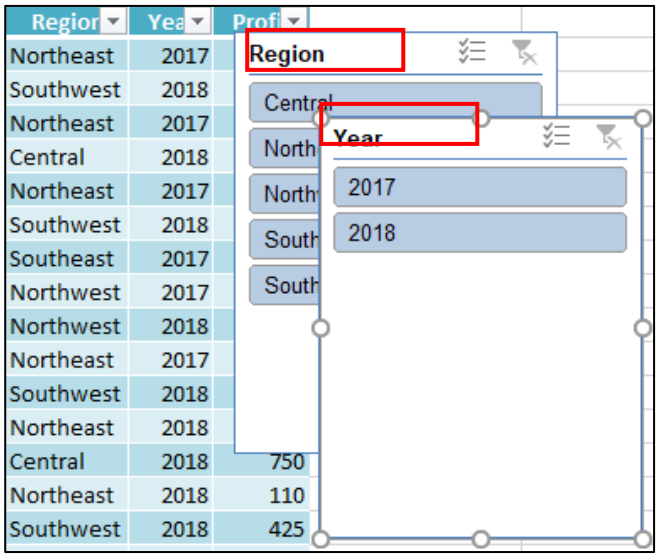
4. In the **Insert Slicers** dialog box, check the fields (columns) for which you want to create a slicer. In this case select the **Region** and **Year** fields (columns).



Selecting the Columns to Create Slicers

5. Click **OK**.

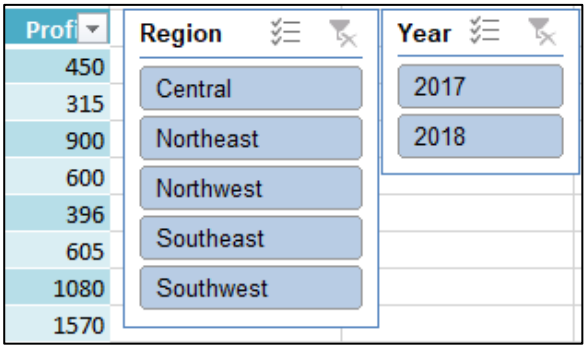
A slicer is displayed for every field (column) selected.



Inserted Slicers

6. Click and drag to position the slicer accordingly.

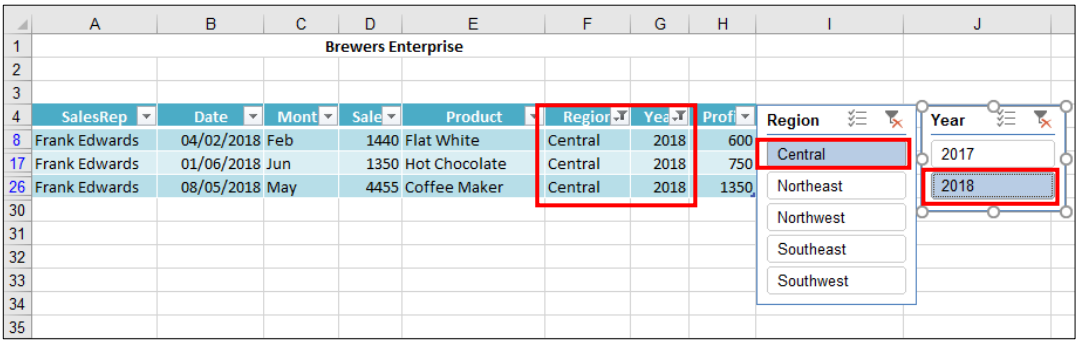
7. Resize the slicers as required using the sizing handles.



Slicers Moved and Resized

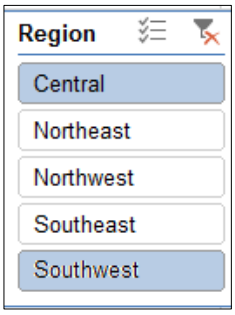
8. In each slicer, click the items on which you want to filter.

   The selected buttons are shown in the slicers and the table columns are filtered according to the selections.



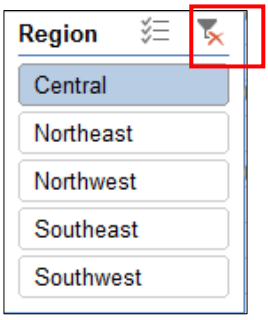
Using the Slicers

9. To select more than one item, press **CTRL**, and then click the items on which you want to filter.



Selecting multiple items

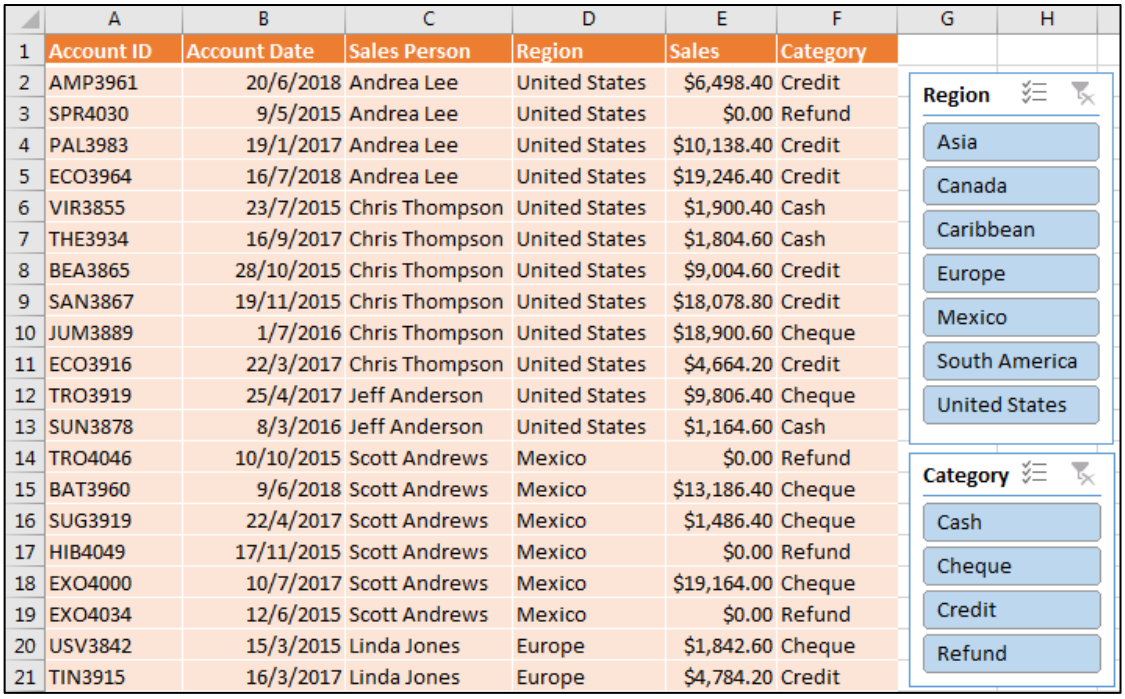
10. To clear the filter, use the **Clear Filter** button.



Clear Filter Button

5.3 REVIEW EXERCISE

1. Open the **Regional Sales.xlsx** workbook.
2. Format the data set range **A1:F58** as table using one of the medium styles.
3. Remove the banded row formatting.
4. Remove the **AutoFilter** options.
5. Insert slicers using the field item **Region** and **Category**.
6. Move and resize the slicers so that they appear at the right of the table.



	A	B	C	D	E	F
1	Account ID	Account Date	Sales Person	Region	Sales	Category
2	AMP3961	20/6/2018	Andrea Lee	United States	$6,498.40	Credit
3	SPR4030	9/5/2015	Andrea Lee	United States	$0.00	Refund
4	PAL3983	19/1/2017	Andrea Lee	United States	$10,138.40	Credit
5	ECO3964	16/7/2018	Andrea Lee	United States	$19,246.40	Credit
6	VIR3855	23/7/2015	Chris Thompson	United States	$1,900.40	Cash
7	THE3934	16/9/2017	Chris Thompson	United States	$1,804.60	Cash
8	BEA3865	28/10/2015	Chris Thompson	United States	$9,004.60	Credit
9	SAN3867	19/11/2015	Chris Thompson	United States	$18,078.80	Credit
10	JUM3889	1/7/2016	Chris Thompson	United States	$18,900.60	Cheque
11	ECO3916	22/3/2017	Chris Thompson	United States	$4,664.20	Credit
12	TRO3919	25/4/2017	Jeff Anderson	United States	$9,806.40	Cheque
13	SUN3878	8/3/2016	Jeff Anderson	United States	$1,164.60	Cash
14	TRO4046	10/10/2015	Scott Andrews	Mexico	$0.00	Refund
15	BAT3960	9/6/2018	Scott Andrews	Mexico	$13,186.40	Cheque
16	SUG3919	22/4/2017	Scott Andrews	Mexico	$1,486.40	Cheque
17	HIB4049	17/11/2015	Scott Andrews	Mexico	$0.00	Refund
18	EXO4000	10/7/2017	Scott Andrews	Mexico	$19,164.00	Cheque
19	EXO4034	12/6/2015	Scott Andrews	Mexico	$0.00	Refund
20	USV3842	15/3/2015	Linda Jones	Europe	$1,842.60	Cheque
21	TIN3915	16/3/2017	Linda Jones	Europe	$4,784.20	Credit

7. Use the slicers to filter for **credit** transactions for the regions **Europe** and **United States**.

	A	B	C	D	E	F
1	Account ID	Account Date	Sales Person	Region	Sales	Category
2	AMP3961	20/6/2018	Andrea Lee	United States	$6,498.40	Credit
4	PAL3983	19/1/2017	Andrea Lee	United States	$10,138.40	Credit
5	ECO3964	16/7/2018	Andrea Lee	United States	$19,246.40	Credit
8	BEA3865	28/10/2015	Chris Thompson	United States	$9,004.60	Credit
9	SAN3867	19/11/2015	Chris Thompson	United States	$18,078.80	Credit
11	ECO3916	22/3/2017	Chris Thompson	United States	$4,664.20	Credit
21	TIN3915	16/3/2017	Linda Jones	Europe	$4,784.20	Credit
22	THE3873	17/1/2016	Linda Jones	Europe	$9,662.20	Credit
23	BBD3906	12/12/2016	Linda Jones	Europe	$4,788.40	Credit
32	BVI3909	15/1/2017	Suzy Liam	United States	$19,004.60	Credit
56	MYS3999	28/6/2017	Tom Baker	United States	$26,184.20	Credit
57	ECO3996	30/5/2017	Tom Baker	United States	$9,785.40	Credit

LESSON 6 –
PIVOT TABLE DATA AGGREGATION

In this section, you will learn how to:

- Change the method of aggregation for a value
- Display multiple aggregation values
- Use built-in calculations

6.1 SUMMARISING DATA INTRODUCTION

Concepts

When you have prepared your data, you can start analysing it. In Excel, you can perform simple analysis using statistical methods like sum, count, average, maximum and minimum to summarise the data in a selected range. You can also perform simple analysis using options like sorting, filtering and table slicers.

Sometimes you may need more powerful analysis to answer specific questions and show comparisons, patterns, and trends. For example, you may need to find out

- Which store sold the most units this month?
- What percentage of the yearly sales took place in the first quarter?
- How did this year's sales compare with last year's sales?

In Excel, you can use a pivot table to calculate, summarise, and analyse large data sets with thousands of rows quickly and easily. A pivot table is useful tool for summarising data in two or more ways - for example, to find the number of units sold both by store and by month. There are many ways you can work with pivot tables to analyse data to find the information you are interested in.

6.2 CHANGING AGGREGATION METHODS

Concepts

A pivot table presents a summary of the underlying source data by applying summary functions to the data. By default, the data in a pivot table are summed but you can change the way the data is summarised to suit the purpose of your analysis. You can determine whether the data is summed, counted, averaged and so forth by changing the summary function, also known as the aggregation method.

For example, if you are analysing the sales data across multiple stores and the total sales amount by store is shown, you can change this to show the average sales amount by store instead.

The summary functions (aggregation methods), available in Excel are outlined in the table.

Aggregate Method	Description
Sum	The sum of the values. This is the default function for adding numerical values.
Count	The number of values. This is the default function for counting entries in the number fields.
Average	The average (mean) of values.
Max	The largest value.
Min	The lowest value.

Steps

To change the aggregation method used in a pivot table:

1. Open the **Cards Pivot.xlsx** workbook and go to the **Pivot1** worksheet.

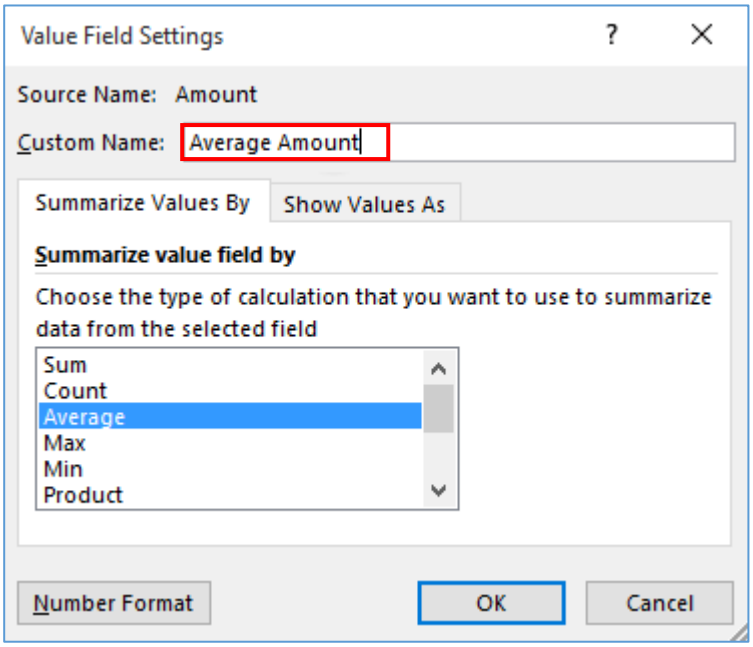
2. The PivotTable shows the **total credit card transaction value** for each **department**. Click on any value in the **Sum of Amount** item in the PivotTable.

3. On the **Analyze** tab, in the **Active Field** group, click **Field Settings**.

4. Select the function **Average** in the **Summarize value field by** list box.

5. Click in the **Custom Name** text box and type **Average Amount** to change the name.

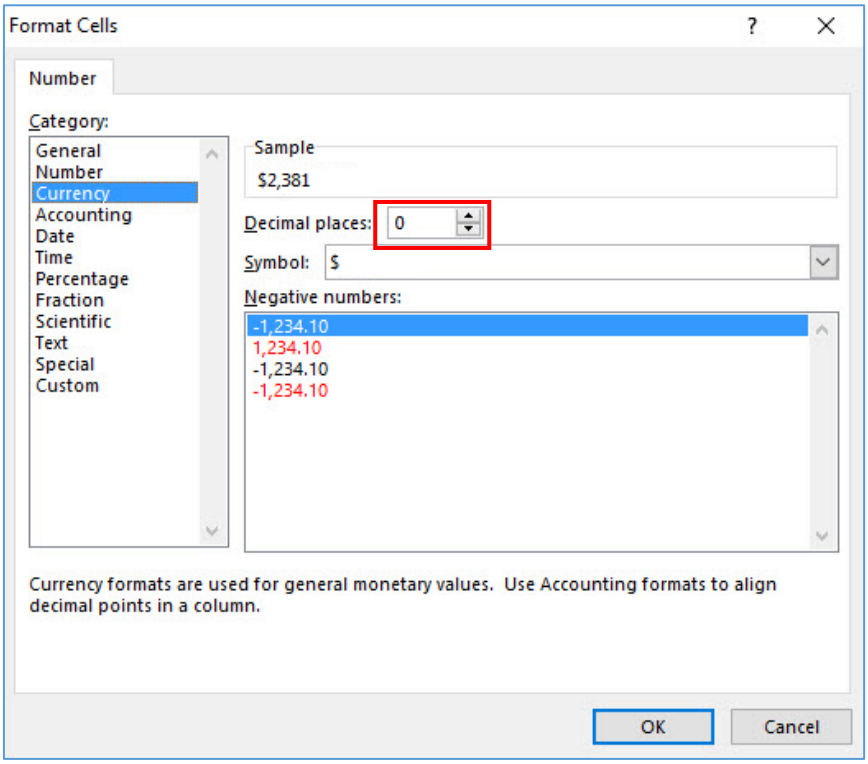
   Note: *Every field in a PivotTable has a name. Fields in the row, column, and filter areas inherit their names from the heading in the source data. Fields in the value section are given default names such as* ***Sum of Amount****. Select the function before renaming the field because the name will change to a default name when a function is selected, e.g. when selecting* ***Average****, the name will reflect* ***Average of Amount****.*



Changing the Aggregation Method

6. You can also change the format of the value. Click **Number Format** to set the format of the value.

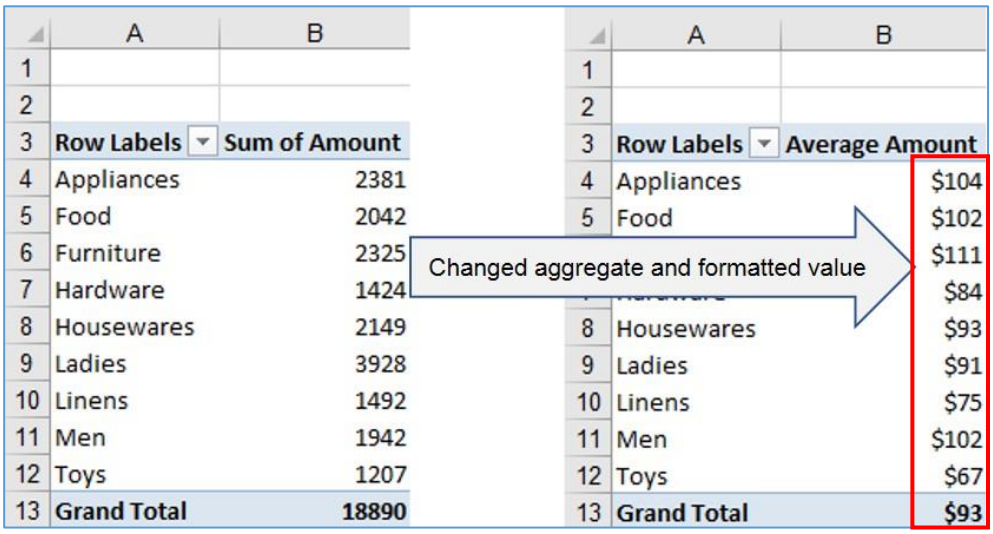
7. Select **Currency** and set the **Decimal places** to **0**.



Formatting the Value Field

8. Click **OK** to close the **Format Cells** dialog box.

9. Click **OK** to close the **Values Field Settings** dialog box.



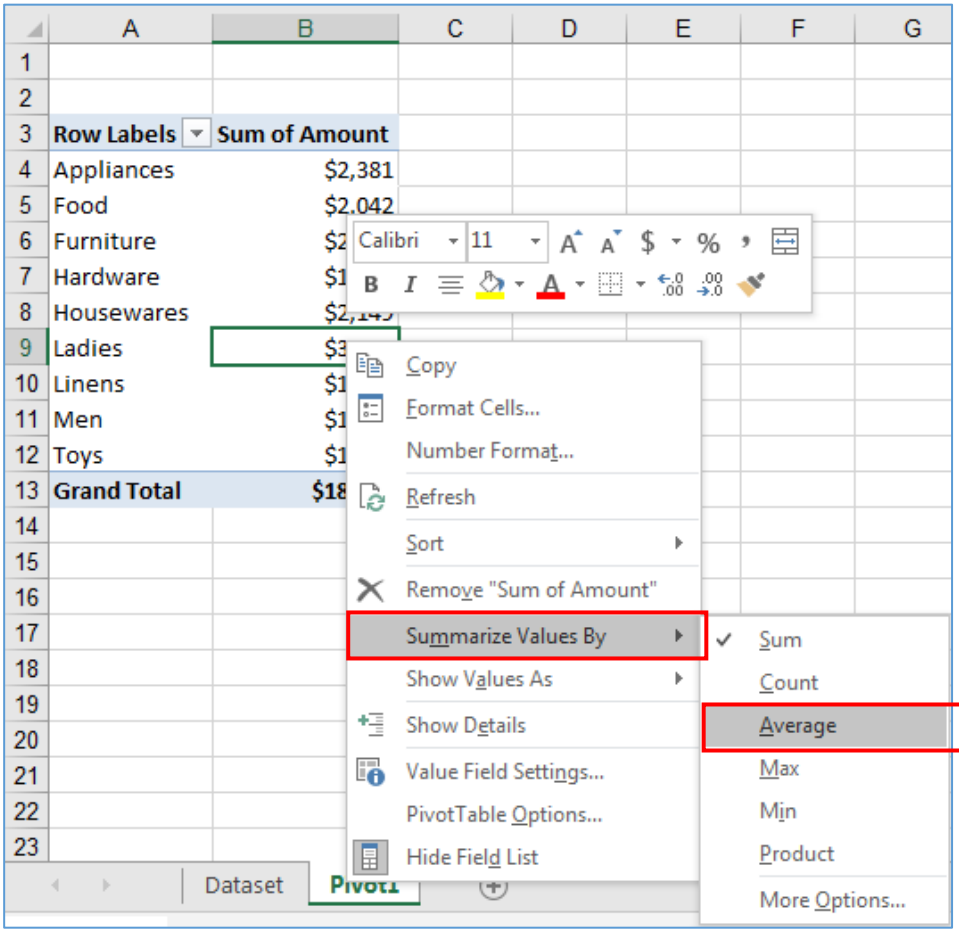
	A	B
1		
2		
3	**Row Labels**	**Sum of Amount**
4	Appliances	2381
5	Food	2042
6	Furniture	2325
7	Hardware	1424
8	Housewares	2149
9	Ladies	3928
10	Linens	1492
11	Men	1942
12	Toys	1207
13	**Grand Total**	**18890**

	A	B
1		
2		
3	**Row Labels**	**Average Amount**
4	Appliances	$104
5	Food	$102
6	Furniture	$111
7	Hardware	$84
8	Housewares	$93
9	Ladies	$91
10	Linens	$75
11	Men	$102
12	Toys	$67
13	**Grand Total**	**$93**

Original and Formatted Aggregate Value

The aggregate data is summarized using the average aggregation method. The pivot table shows the average amount of credit card transactions by department formatted as dollars with no decimal places.

***Note**: Right-click a value field, choose **Summarize Values By** and select a function to switch quickly between aggregation methods.*



	A	B
3	**Row Labels**	**Sum of Amount**
4	Appliances	$2,381
5	Food	$2,042
6	Furniture	$2
7	Hardware	$1
8	Housewares	$2,149
9	Ladies	$3
10	Linens	$1
11	Men	$1
12	Toys	$1
13	**Grand Total**	**$18**

Using the Right-Click Method to Change Aggregation Method

6.3 DISPLAYING MULTIPLE AGGREGATION VALUES

Concepts

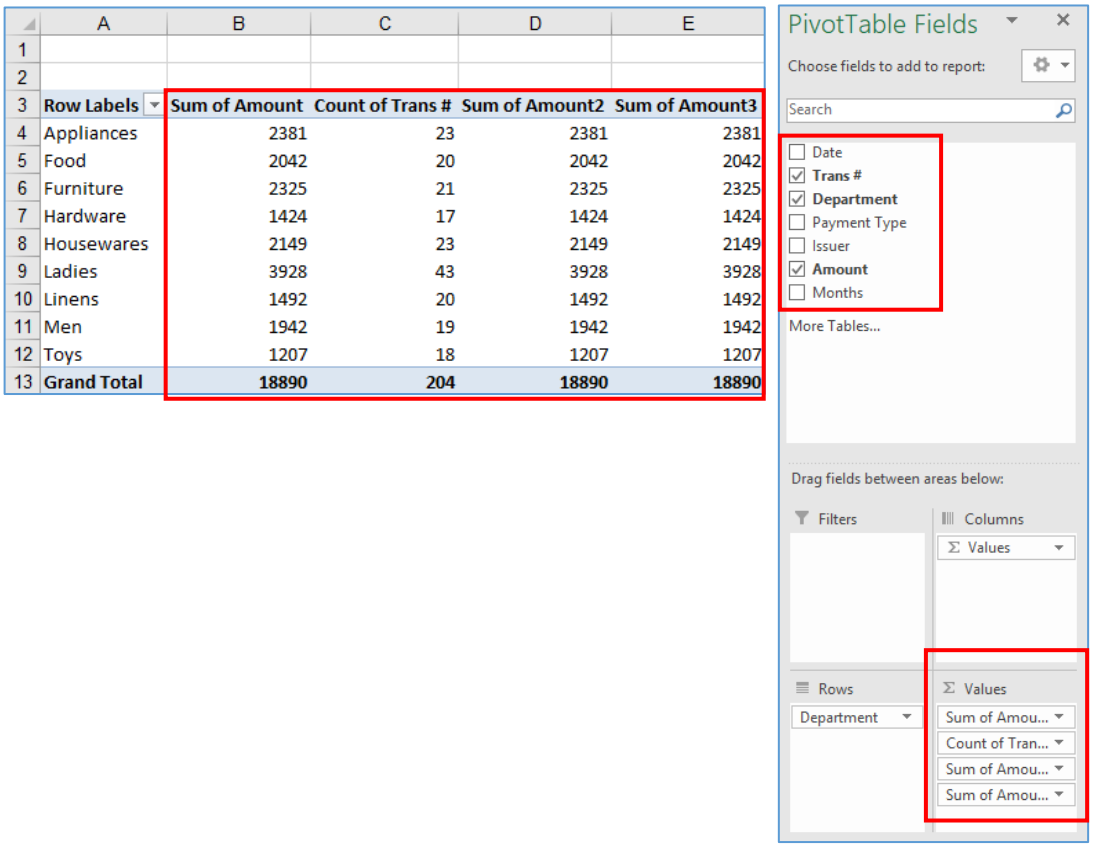
For an overview of different summary values on a data set, you can display multiple summary calculations (aggregation values) in a pivot table, such as **Sum**, **Count**, **Average**, **Max, and Min**.

For example, you may want to analyse the sales performance of multiple stores and compare several aspects of their performance at one time. You may want to show the total amount of sales by store, the total number of sales by store, the average sale amount by store, the highest sale amount by store, and the lowest sale amount by store. You can display all these aggregation values using the **Sum**, **Count**, **Average**, **Max,** and **Min** calculations at the same time in a pivot table.

Steps

To show multiple aggregation values in a pivot table:

1. Open the **Cards Pivot.xlsx** workbook and go to the **Pivot2** worksheet.
2. Click on any cell in the PivotTable.
3. To display **Sum**, **Count**, **Max** and **Min** aggregation values, drag and drop the following fields into the **Σ Values** area in the **PivotTable Fields** panel:
    a. **Amount**: This will automatically display as **Sum** aggregate.
    b. **Trans #**: This will automatically display as **Count** aggregate.
    c. **Amount**: This will automatically display as **Sum of Amount 2** aggregate.
    d. **Amount**: This will automatically display as **Sum of Amount 3** aggregate.



Row Labels	Sum of Amount	Count of Trans #	Sum of Amount2	Sum of Amount3
Appliances	2381	23	2381	2381
Food	2042	20	2042	2042
Furniture	2325	21	2325	2325
Hardware	1424	17	1424	1424
Housewares	2149	23	2149	2149
Ladies	3928	43	3928	3928
Linens	1492	20	1492	1492
Men	1942	19	1942	1942
Toys	1207	18	1207	1207
Grand Total	**18890**	**204**	**18890**	**18890**

4. Right click the field **Sum of Amount2**.

5. Select **Summarize Values By** and select **Max**. This changes the aggregate values for the field **Sum of Amount2** from **total sum** to **maximum**.

6. Right click the field **Sum of Amount3**.

7. Select **Summarize Values By** and select **Min**. This changes the aggregate values for the field **Sum of Amount3** from **sum** to **minimum**.

8. To change the field names, right click each of the four fields. (**Sum of Amount**, **Count of Trans#**, **Max of Amount2**, **Min of Amount3**).

9. Select **Value Field Settings**.

10. Change the **names** in the **Custom Name** text box as follows:

Field	Function	Custom Name
Sum of Amount	Sum	Total
Count of Trans #	Count	Transactions
Max of Amount2	Max	Highest
Min of Amount3	Min	Lowest

	A	B	C	D	E
1					
2					
3	**Row Labels**	**Total**	**Transactions**	**Highest**	**Lowest**
4	Appliances	2381	23	189	2
5	Food	2042	20	200	20
6	Furniture	2325	21	199	6
7	Hardware	1424	17	169	1
8	Housewares	2149	23	194	11
9	Ladies	3928	43	196	6
10	Linens	1492	20	155	10
11	Men	1942	19	192	9
12	Toys	1207	18	189	13
13	**Grand Total**	**18890**	**204**	**200**	**1**

Labelled Multiple Aggregation Values

The four aggregation values are displayed in the pivot table with customised field names.

The key data summary of the transactions for the different departments are displayed in the pivot table to facilitate analysis.

6.4 USING BUILT-IN CALCULATIONS

Concepts

The PivotTable command contains built-in calculations, also known as custom calculations, that you can use to quickly add calculations to a pivot table. You can present values in different ways such as percentage calculations, difference from specific values, running totals, and ranked. This functionality is provided in the **Show Values As** option.

For example, when analysing the sales performance of multiple stores, you may see the sales of each store by year, but you want to show the difference between two years' sales. Maybe you want to compare yearly sales totals to see if there are any trends, maybe indicating problems that need to be addressed or successes that need to be capitalised. You can do this using a built-in (custom) calculation – in this example **Difference From**.

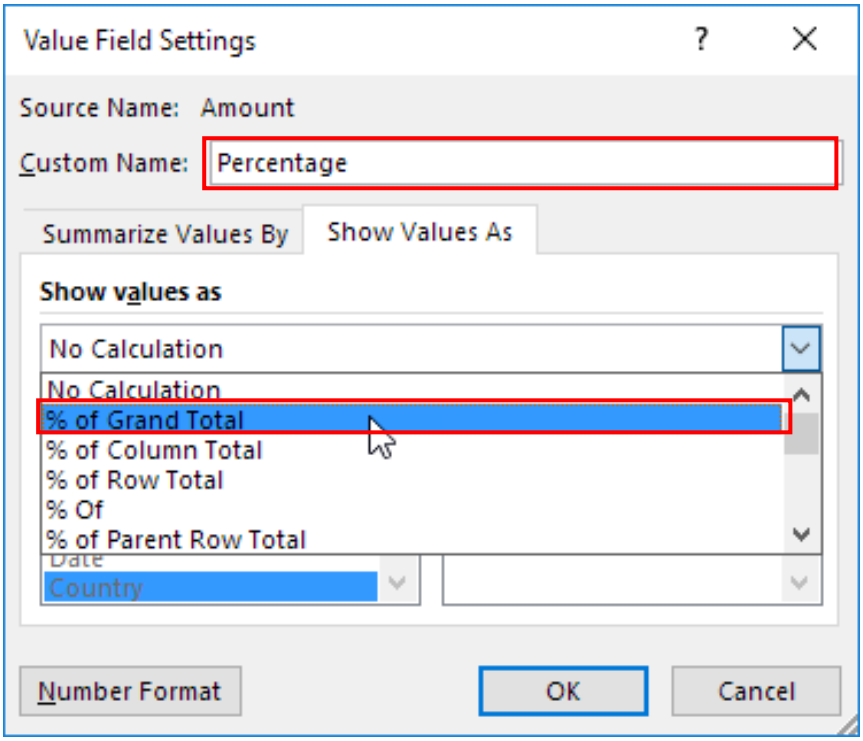
The table below outlines the commonly used built-in (custom) calculations in the **Show Value As** option.

Show Value As	Description
No Calculation	Displays the original value.
% of Grand Total	Displays values as a percentage of the grand total of all the values or data points.
% of Column Total	Displays all the values in each column or series as a percentage of the total for the column or series.
% of Row Total	Displays the value in each row or category as a percentage of the total for the row or category.
Difference From	Displays values as the difference between two values. For example, the difference between this year's sales value and last year's sales value.
% Difference From	Displays values as a percentage of the difference between two values. For example, the difference between this year's sales value and last year's sales value as a percentage. The percentage change is a useful figure for examining trend data.
Running Total in	Displays cumulative or running totals of pivot table cell values. For example, cumulative sales over a period of time such as the year-to-date.
Rank Smallest to Largest	Displays the rank of selected values in a specific field from smallest to largest, listing the smallest item in the field as 1, and each larger value with a higher rank value.
Rank Largest to Smallest	Displays the rank of selected values in a specific field, listing the largest item in the field as 1, and each smaller value with a higher rank value.

Steps

To show values as a percentage of grand total using the **% of Grand Total custom** calculation:

1. Open the **Cards Pivot.xlsx** workbook and go to the **Pivot3** worksheet.
2. Click on any value in the **Sum of Amount** item in the PivotTable.
3. On the **Analyze** tab, in the **Active Field** group, click **Field Settings**.
4. Select the **Show Values As** tab in the **Value Field Settings** dialog box.
5. Click the **Show values as** drop-down list and select **% of Grand Total**.
6. Type **Percentage** in the **Custom Name** box to modify the name.



Showing Value as Percentage

7. Click **Number Format** to format the number.
8. Select the **Percentage** category and set the **Decimal places** to **1**.

9. Click **OK** to close the **Format Cells** dialog box.

10. Click **OK** to close the **Value Field Settings** dialog box.

The Pivot table shows the amount for each department as a percentage of the Grand Total.

	A	B	C
1			
2			
3	**Row Labels**	**Total Amount**	**Percentage**
4	Appliances	2381	12.6%
5	Food	2042	10.8%
6	Furniture	2325	12.3%
7	Hardware	1424	7.5%
8	Housewares	2149	11.4%
9	Ladies	3928	20.8%
10	Linens	1492	7.9%
11	Men	1942	10.3%
12	Toys	1207	6.4%
13	**Grand Total**	**18890**	**100.0%**

Amount Values Shown as Percentage of Grand Total

To rank the **number of transactions** for each department from highest to lowest using the **Rank Largest to Smallest** custom calculation (with the highest value being ranked as 1):

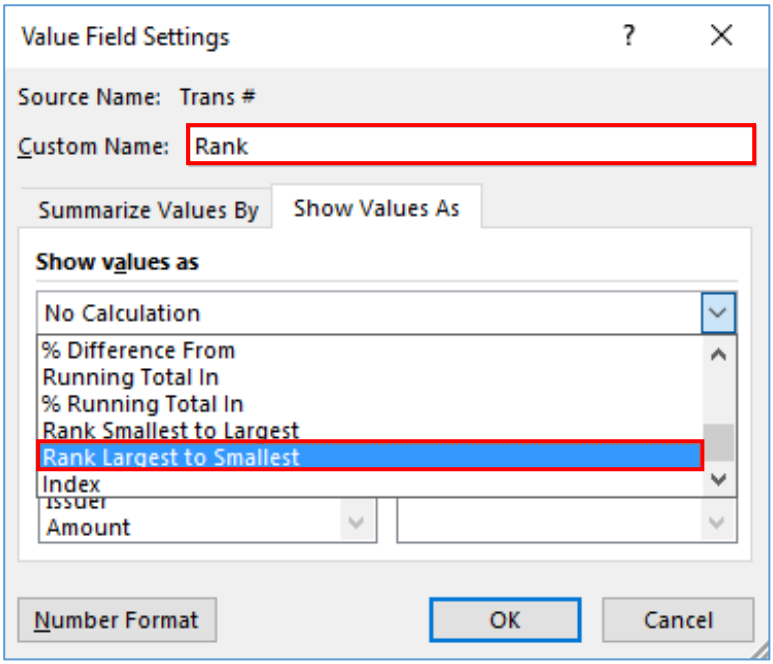
1. Click on any value in the **Count of Trans #** item in the PivotTable in the **Pivot3** worksheet in the **Cards Pivot** workbook.

2. On the **Analyze** tab, in the **Active Field** group, click **Field Settings**.

3. Select the **Show Values As** tab in the **Value Field Settings** dialog box.

4. Click the **Show values as** drop-down list and select **Rank Largest to Smallest**.

5. Type **Rank** in the **Custom Name** box to modify the name.



Ranking Values

***Note**: The default base field is **Department** as the ranking is done by comparing the number of transactions across the departments.*

6. Click **OK**.

The number of transactions are ranked from highest to lowest (with the highest value being ranked as 1).

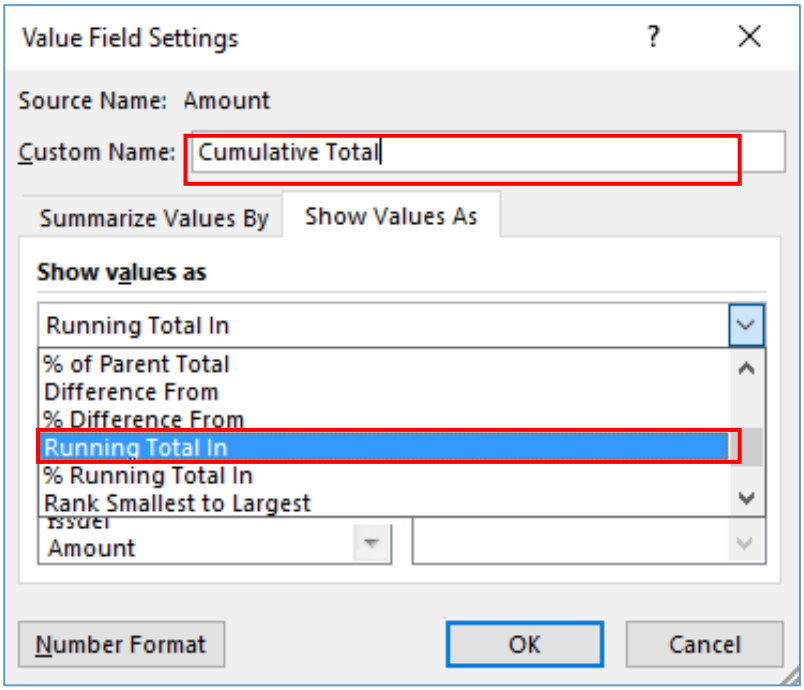
	A	B	C	D
1				
2				
3	**Row Labels**	**Total Amount**	**Percentage**	**Rank**
4	Appliances	2381	12.6%	2
5	Food	2042	10.8%	4
6	Furniture	2325	12.3%	3
7	Hardware	1424	7.5%	7
8	Housewares	2149	11.4%	2
9	Ladies	3928	20.8%	1
10	Linens	1492	7.9%	4
11	Men	1942	10.3%	5
12	Toys	1207	6.4%	6
13	**Grand Total**	**18890**	**100.0%**	

Number of Transactions Ranked from Highest to Lowest

To add another value item to show the **cumulative total over a period of time** – in this example a **cumulative daily total**:

1. Go to the **Pivot4** worksheet.
2. Click on any cell in the PivotTable.
3. In the **PivotTable Fields** panel, drag and drop the **Amount** field into the **Σ Values** area.

   ***Note**: If necessary, set the aggregation method for the field to **Sum**.*
4. Click on any value in the **Sum of Amount** item in the PivotTable.
5. On the **Analyze** tab, in the **Active Field** group, click **Field Settings**.
6. Select the **Show Values As** tab in the **Value Field Settings** dialog box.
7. Click the **Show values as** drop-down list and select **Running Total In**.
8. Ensure **Date** is set as the **Base field**.
9. Type **Cumulative Total** in the **Custom Name** box to modify the name.



Running Total

10. Click **OK**.

The cumulative total up to each date is shown in the same row as the date.

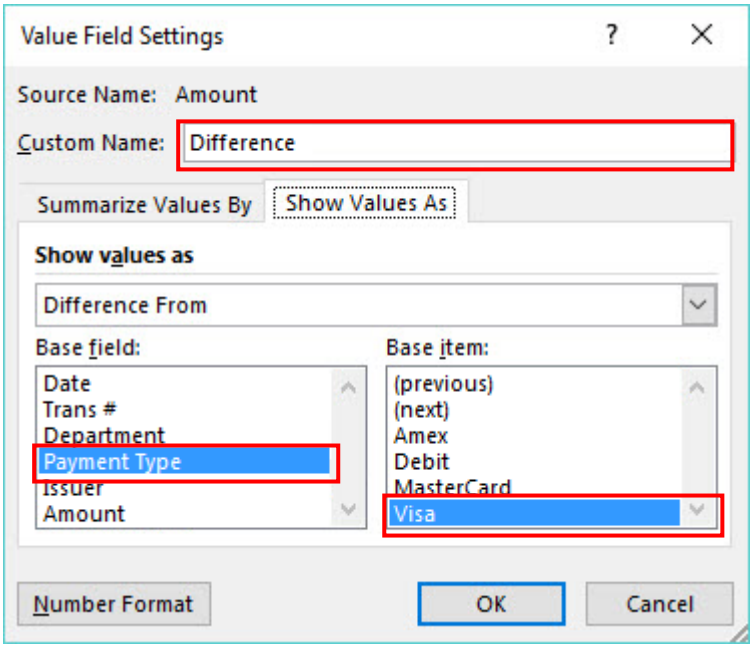
	A	B	C
1			
2			
3	**Row Labels**	**Daily Total**	**Cumulative Total**
4	2-Apr-18	357	357
5	3-Apr-18	203	560
6	4-Apr-18	365	925
7	5-Apr-18	154	1079
8	6-Apr-18	105	1184
9	7-Apr-18	154	1338
10	8-Apr-18	104	1442

Cumulative Daily Total

To add another value item to show the **difference between one value** and **another value** in a pivot table - in this example, to show the difference between **the total amount of Visa card transactions** and **the other card types** (Amex, Debit and MasterCard):

1. Go to the **Pivot5** worksheet.
2. Click on any cell in the PivotTable.
3. In the **PivotTable Fields** panel, drag and drop the **Amount** field into the **Σ Values** area.

   ***Note**: If necessary, set the aggregation method for the field to **Sum**.*
4. Click on any value in the **Sum of Amount** item in the PivotTable.
5. On the **Analyze** tab, in the **Active Field** group, click **Field Settings**.
6. Select the **Show Values As** tab in the **Value Field Settings** dialog box.
7. Click the **Show values as** drop-down list and select **Difference From**.
8. Ensure **Payment Type** is selected in the **Base field** list box.
9. Select **Visa** in the **Base item** list box to define Visa as the value to be subtracted from the other values.
10. Type **Difference** in the **Custom Name** box.



Difference From

11. Click **OK**.

The difference between the total amount for each card and the Visa card is shown.

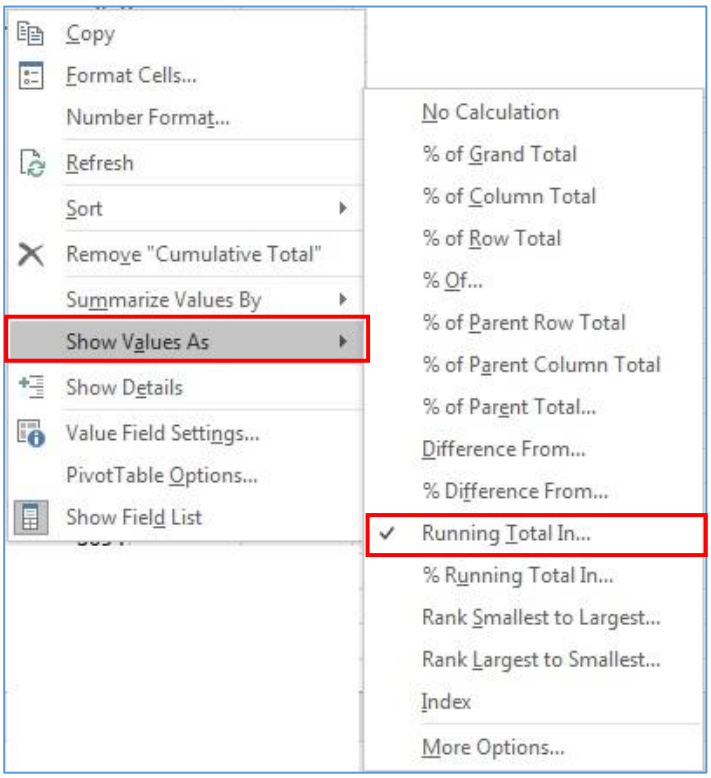
The resulting pivot table shows **Visa** as the preferred type of payment in this context and the gap in transactional values by other payment methods compared to it.

	A	B	C
1			
2			
3	**Row Labels**	**Total Amount**	**Difference**
4	Amex	1399	-12603
5	Debit	906	-13096
6	MasterCard	2583	-11419
7	Visa	14002	
8	**Grand Total**	**18890**	

Comparing Differences between Visa and Other Card Payments

***Note**: The **Difference From** calculation subtracts one value in the pivot table from another and shows the result (that is the difference from the value of the Base item in the Base field). To see the percentage difference, use **% Difference From**. The percentage change figure is useful for analysing trend data.*

*To switch quickly between the various built-in calculations, right-click a value field, choose **Show Values As**, and select a calculation from the context menu (for example, Running Total In). The required result will be shown in the pivot table.*



Using the Right-Click Method to Apply a Built-in Calculation

6.5 REVIEW EXERCISE

1. Open the **Aggregate.xlsx** workbook and go to the **Trips** worksheet.

2. Create a PivotTable in the build area provided from cell **K4** onwards using the appropriate fields and aggregation methods to summarise the **average amount of tickets sold for each office** and the **number of tickets sold for each office**.

3. Label the average amount as **Average Cost** and format the values as currency without decimal.

4. Label the number of tickets sold as **Total Tickets**.

	K	L	M
4	**Row Labels**	**Average Cost**	**Total Tickets**
5	Blaine	$219	6
6	Bloomington	$258	10
7	Brainerd	$280	7
8	Brooklyn Center	$246	11
9	Duluth	$312	17
10	Edina	$370	14
11	Fargo	$278	11
12	Hibbing	$411	7
13	Mankato	$372	8
14	Maplewood	$310	7
15	Minneapolis	$326	33
16	Rochester	$216	13
17	Spring Lake Park	$333	9
18	St. Cloud	$446	9
19	St. Louis Park	$228	3
20	St. Paul	$310	15
21	St. Peter	$282	8
22	Two Harbors	$255	11
23	**Grand Total**	**$307**	**199**

5. Go to the **Shipping** worksheet.

6. Create a PivotTable in the build area provided from cell **G4** onwards using the appropriate fields and built-in calculations to show the **daily total of the Order Value** and the **running total of the Order Value**.

7. Label the daily total as **Daily Sum**.

8. Label the running total as **Running Sum**.

	G	H	I
4	Row Labels	Daily Sum	Running Sum
5	1-Jun-2018	3500	3500
6	2-Jun-2018	7700	11200
7	3-Jun-2018	7760	18960
8	4-Jun-2018	3570	22530
9	5-Jun-2018	3780	26310
10	6-Jun-2018	5340	31650
11	7-Jun-2018	22180	53830
12	Grand Total	53830	

9. Create a PivotTable in cell **G15** onwards using the data set in the range **A3:D20** to show the **percentage of the order value for each shipping detail**.

10. Label the column as **Orders Share**.

	G	H
15	Row Labels	Orders Share
16	1 - Shipped	11.85%
17	2 - Lost	0.84%
18	3 - Lost	14.08%
19	3 - Returned	5.94%
20	3 - Shipped	13.56%
21	4 - Delayed	8.47%
22	4 - Returned	9.77%
23	4 - Shipped	22.14%
24	5 - Shipped	1.10%
25	6 - Shipped	9.92%
26	7 - Shipped	2.32%
27	Grand Total	100.00%

11. Go to the **Sales** worksheet.

12. Create a pivot table in the build area provided from cell **H4** onwards using the appropriate fields and built-in calculations to **rank 2018 Sales** for each **Market from largest to smallest**.

13. Label the column as **2018 Rank**.

	H	I
4	Row Labels	2018 Rank
5	Asia	4
6	Europe	1
7	North America	2
8	South America	3
9	Grand Total	

LESSON 7 –
PIVOT TABLE FREQUENCY ANALYSIS

In this section, you will learn how to:

- Automatically group date, time and numeric data items
- Manually create custom groups
- Rename groups
- Ungroup data items

7.1 Grouping Date, Time and Numeric Data

Concepts

When analysing large sets of data in pivot tables, you may want to look at subsets of the data to find patterns and trends. You can do this by organizing the data into groups.

If your pivot table shows how often something occurs for a particular field in a data set, this is a frequency distribution table. You can put data into groups to create a grouped frequency distribution table. This provides useful information on how data in the data set is distributed across groups. The group with the highest frequency for the data set is known as the modal group.

If the data in your pivot table has date or time fields, you can group the data by specified time periods. The groups available are seconds, minutes, hours, day, month, quarter, or year. In the example below, the first pivot table lists daily transaction totals and numbers. The transaction totals and numbers in the second pivot table have been grouped by week. In the second pivot table you can see at a glance that the highest total occurred in week 30/4/2018-6/5/2018 and the highest number (or frequency) of transactions occurred 11/6/2018-17/6/2018.

	A	B	C
1			
2			
3	Row Labels	Total	Transactions
4	2-Apr-18	357	4
5	3-Apr-18	203	2
6	4-Apr-18	365	4
7	5-Apr-18	154	2
8	6-Apr-18	105	1
9	7-Apr-18		
10	8-Apr-18		
11	9-Apr-18	231	3
12	10-Apr-18	51	1
13	11-Apr-18	2	1
14	12-Apr-18	235	2
15	13-Apr-18	246	3
16	14-Apr-18	395	4
17	15-Apr-18	234	2

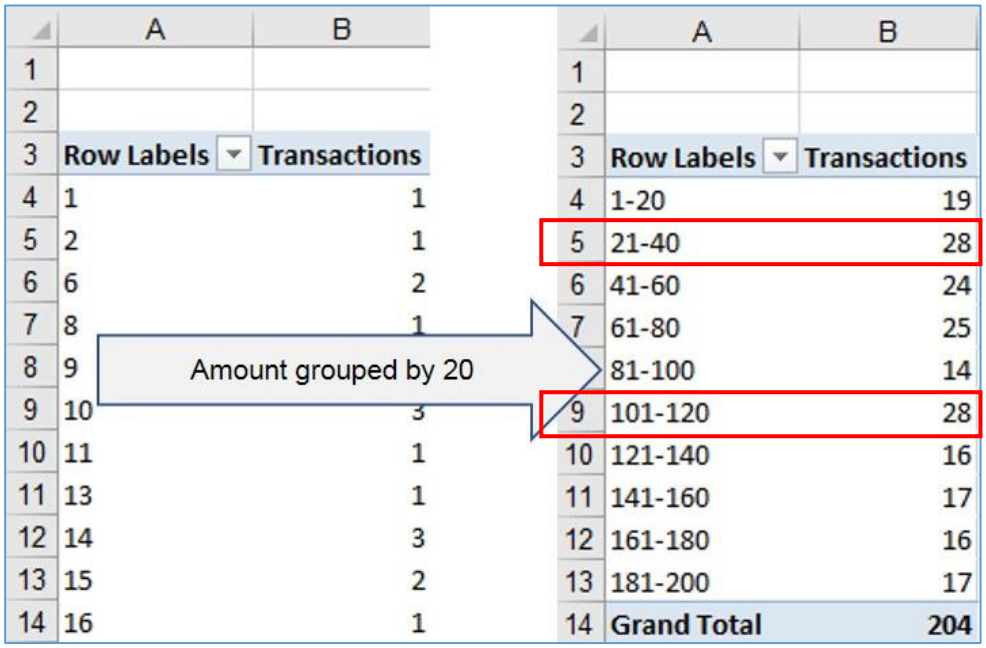
Dates grouped by 7 days

	A	B	C
1			
2			
3	Row Labels	Total	Transactions
4	2/4/2018 - 8/4/2018	1442	17
5	9/4/2018 - 15/4/2018	1394	16
6	16/4/2018 - 22/4/2018	884	10
7	23/4/2018 - 29/4/2018	1566	17
8	30/4/2018 - 6/5/2018	1953	18
9	7/5/2018 - 13/5/2018	1403	16
10	14/5/2018 - 20/5/2018	1286	14
11	21/5/2018 - 27/5/2018	1360	14
12	28/5/2018 - 3/6/2018	1767	18
13	4/6/2018 - 10/6/2018	1619	16
14	11/6/2018 - 17/6/2018	1767	19
15	18/6/2018 - 24/6/2018	1538	15
16	25/6/2018 - 30/6/2018	911	14
17	Grand Total	18890	204

Data Grouped by Week

***Note**: You can also group date and time data into hierarchical groups such as by week and day or by year and quarter.*

If the data in your pivot table contains numeric values, you can group the data by a specified interval. In the example below, the first pivot table shows individual transaction amounts and the number of transactions, or frequency for each amount. In the second pivot table the transaction amounts have been grouped into intervals of 20. In the second pivot table you can see at a glance that the most frequently occurring transaction amounts occurred in the 21-40 and the 101-120 ranges.



Row Labels	Transactions
1	1
2	1
6	2
8	1
9	3
10	3
11	1
13	1
14	3
15	2
16	1

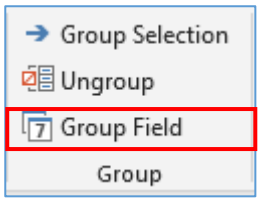
Row Labels	Transactions
1-20	19
21-40	28
41-60	24
61-80	25
81-100	14
101-120	28
121-140	16
141-160	17
161-180	16
181-200	17
Grand Total	**204**

Transaction Amount Grouped by 20

Steps

To group data in a pivot table by date:

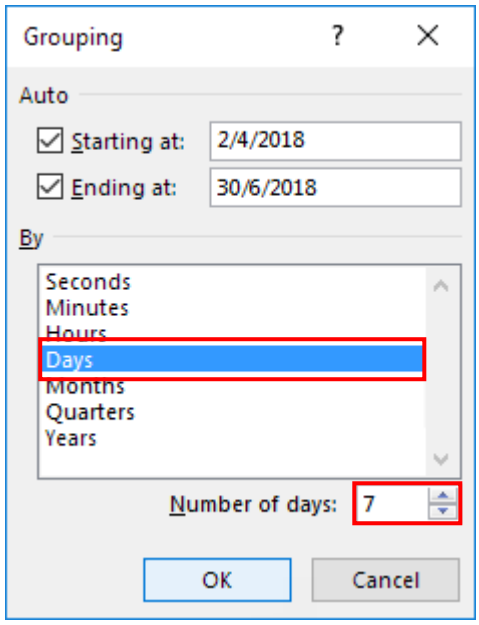
1. Open the **Cards Group.xlsx** workbook and go to the **Group1** worksheet.
2. Click on any date value below **Row Labels** in the PivotTable.
3. On the **Analyze** tab, in the **Group** group, click **Group Field**.



Grouping Dates

4. In the **Grouping** dialog box, in the **By** list box, select the period of time to group by – in this case, to view a weekly report deselect the **Months** option, select **Days** and type **7** in the **Number of Days** box.

   ***Note**: There is no weekly option, so to group the data by week you need to select **Days** and set the number of days as **7**. The week range is determined by the date in the **Starting at** box, so adjust this if necessary. By default, the **Starting at** and **Ending at** dates are the first and last dates in the pivot table.*



Grouping by a Week

5. Click **OK**.

The daily report is presented as a weekly report in the Pivot Table.

	A	B	C
1			
2			
3	**Row Labels**	**Total**	**Transactions**
4	2-Apr-18	357	4
5	3-Apr-18	203	2
6	4-Apr-18	365	4
7	5-Apr-18	154	2
8	6-Apr-18	105	1
9	7-Apr-18		
10	8-Apr-18		
11	9-Apr-18	231	3
12	10-Apr-18	51	1
13	11-Apr-18	2	1
14	12-Apr-18	235	2
15	13-Apr-18	246	3
16	14-Apr-18	395	4
17	15-Apr-18	234	2

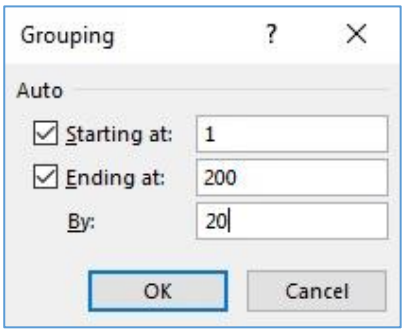
Dates grouped by 7 days

	A	B	C
1			
2			
3	**Row Labels**	**Total**	**Transactions**
4	2/4/2018 - 8/4/2018	1442	17
5	9/4/2018 - 15/4/2018	1394	16
6	16/4/2018 - 22/4/2018	884	10
7	23/4/2018 - 29/4/2018	1566	17
8	30/4/2018 - 6/5/2018	1953	18
9	7/5/2018 - 13/5/2018	1403	16
10	14/5/2018 - 20/5/2018	1286	14
11	21/5/2018 - 27/5/2018	1360	14
12	28/5/2018 - 3/6/2018	1767	18
13	4/6/2018 - 10/6/2018	1619	16
14	11/6/2018 - 17/6/2018	1767	19
15	18/6/2018 - 24/6/2018	1538	15
16	25/6/2018 - 30/6/2018	911	14
17	**Grand Total**	**18890**	**204**

Dates Grouped by 7 Days

To group numeric values by a specified interval:

1. Open the **Cards Group.xlsx** workbook and go to the **Group2** worksheet.
2. Click on any numerical value in the **Rows** area in the pivot table.
3. On the **Analyze** tab, in the **Group** group, click **Group Field**.
4. In the **Grouping** dialog box, in the **By** text box, type the range or interval to group by – in this case, **20** - and click **OK**.



Grouping Numbers

***Note**: By default, the starting and ending values are the lowest and highest values in the pivot table, so adjust this if necessary.*

The transactions are grouped by an interval of 20 in the Pivot Table.

	A	B
1		
2		
3	**Row Labels**	**Transactions**
4	1	1
5	2	1
6	6	2
7	8	1
8	9	
9	10	3
10	11	1
11	13	1
12	14	3
13	15	2
14	16	1

Amount grouped by 20

	A	B
1		
2		
3	**Row Labels**	**Transactions**
4	1-20	19
5	21-40	28
6	41-60	24
7	61-80	25
8	81-100	14
9	101-120	28
10	121-140	16
11	141-160	17
12	161-180	16
13	181-200	17
14	**Grand Total**	**204**

Amount Values Grouped by 20

7.2 CREATING CUSTOM GROUPS

Concepts

There are many ways that pivot tables automatically group data, but you may also want to manually create custom groups.

You will need to manually create custom groups when the data you want to group are not date, time or numeric items, or the group you want to create does not appear in the underlying data. For example, you may have a pivot table that shows a breakdown of credit card transactions by various banks. You want to further group the transactions into groups of banks, but those groups don't appear in the underlying data. In this case. you will need to manually create your own custom groups of banks.

When you group data items, the group is given a name by default. You can rename groups to define names that are more relevant for your purposes.

Steps

To manually group data items in custom groups:

1. Open the **Cards Group.xlsx** workbook and go to the **Group3** worksheet.

2. Select the items to add to the first group – in this case, the issuers listed are to be grouped by country, so select the following issuers to be part of a **USA** group:
    - **American Express**
    - **Bank of America**
    - **CapitalOne**
    - **Chase Manhattan**
    - **CitiBank**
    - **Seattle FirstBank**
    - **Wells Fargo**

    ***Note**: To select multiple non-contiguous items, press **CTRL** and click on the items.*

	A	B
1		
2		
3	**Row Labels**	**Total Amount**
4	American Express	$1,399
5	Bank of America	$911
6	Bank of Montreal	$2,342
7	CapitalOne	$1,202
8	Chase Manhattan	$1,148
9	CitiBank	$2,207
10	Lloyds Banking Group	$1,222
11	Santandar	$997
12	Santandar of Scotland	$941
13	ScotiaBank	$1,824
14	Seattle FirstBank	$1,301
15	Toronto Dominion Bank	$1,043
16	VanCity Credit Union	$1,206
17	Wells Fargo	$1,147
18	**Grand Total**	**$18,890**

Manually Select Items to Be Grouped

3. On the **Analyze** tab, in the **Group** group, click **Group Selection**.

	A	B
3	**Row Labels**	**Total Amount**
4	**Group1**	**$9,315**
5	American Express	$1,399
6	Bank of America	$911
7	CapitalOne	$1,202
8	Chase Manhattan	$1,148
9	CitiBank	$2,207
10	Seattle FirstBank	$1,301
11	Wells Fargo	$1,147

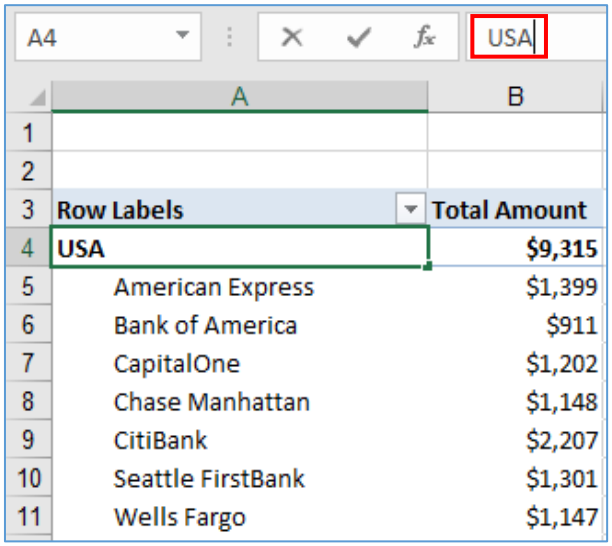
Selected Issuers Grouped under One Category

4. Repeat the selecting and grouping for the following items:
   - **Bank of Montreal**
   - **ScotiaBank**
   - **Toronto Dominion Bank**
   - **VanCity Credit Union**

5. Repeat the selecting and grouping for the following items:
   - **Lloyds Banking Group**
   - **Santandar**
   - **Santandar of Scotland**

To rename groups:

1. For each manual grouping a new name is automatically assigned (for example **Group1, Group2** and **Group3)**. To change the group name to a meaningful label, click on the group name and type a new name (for example, **USA**).



Renaming Grouped Field

2. To hide the details in each group, click the **Collapse** button next to the group name.

Row Labels	Total Amount
USA	**$9,315**
Group2	**$6,415**
Bank of Montreal	$2,342
ScotiaBank	$1,824
Toronto Dominion Bank	$1,043
VanCity Credit Union	$1,206
Group3	**$3,160**
Lloyds Banking Group	$1,222
Santandar	$997
Santandar of Scotland	$941
Grand Total	**$18,890**

Renaming Grouped Field

3. Rename **Group2** as **Canada**, **Group3** as **Others** and collapse both groups.

7.3 UNGROUPING DATA

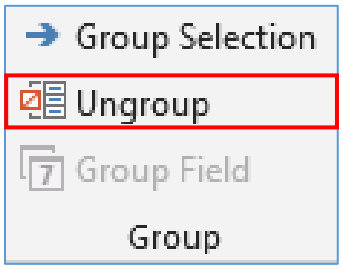
Concepts

Data that are grouped by applying selected criteria or manually can be easily ungrouped.

Steps

To ungroup data in a grouped pivot table:

1. Go to the grouped pivot table in the **Group3** worksheet in **Cards Group.xlsx** workbook.
2. Select a group label – in this case, the country label **USA**.
3. On the **Analyze** tab, in the **Group** group, click **Ungroup** to show the original data.



Ungrouping data

Note**: Alternatively, you can right-click a group label and select **Ungroup.

7.4 REVIEW EXERCISE

1. Open the **Sales Group.xlsx** workbook.
2. Go to the **Group1** worksheet.
3. Ungroup the current automatic date values in the **Row** area.
4. Group the dates in the **Row** area by **Year** and **Quarter**.

	A	B
3	Row Labels	Total Sales
4	⊟2017	$24,805
5	Qtr1	$1,620
6	Qtr3	$19,598
7	Qtr4	$3,587
8	⊟2018	$30,933
9	Qtr1	$16,322
10	Qtr2	$14,611
11	Grand Total	$55,738

***Note**: If the **Total Sales** is not visible for the **Year**, set the **Subtotals** to **Automatic** in the **Years field settings**.*

5. Go to the **Group2** worksheet.
6. Group the sales amount in the **Row** area by every thousand, starting from **0** to **8000**.
7. Select the following labels in the Column area, group the selection and rename the group accordingly.

Labels to Select and Group	Resulting Group	New Group Name
Northeast and **Northwest**	Group 1	North
Southeast and **Southwest**	Group 2	South

8. Remove the original **Region** field from the **Columns** area in the **Field List** pane.

	A	B	C	D	E
3	No. of Sales	Column Labels			
4	Row Labels	Central	North	South	Grand Total
5	0-999		2	5	7
6	1000-1999	3	6	2	11
7	3000-3999		1	2	3
8	4000-4999	1			1
9	5000-5999		1		1
10	7000-8000		2		2
11	Grand Total	4	12	9	25

LESSON 8 –
FILTERING PIVOT TABLES

In this section, you will learn how to:

- Use report filters
- Use slicers
- Use timelines

8.1 USING REPORT FILTERS

Concepts

When you have summarized data in a pivot table in Excel, you may want to create reports that analyse certain parts of the data. You can create these reports using one or more **report filters**. A **report filter** can be used to display selected items in a pivot table. Items selected in the filter are displayed in the pivot table and items that are not selected are hidden. Report filters are displayed above the pivot table for easy access.

For example, you might want to show the sales amounts for a specific year instead of all years or you might want to show the sales amounts for one or two regions instead of all regions. The pivot table below has **report filters** for **Account Date** and **Region** which are showing information for the year **2017** and the region **United States**.

	H	I
1	Account Date	2017
2	Region	United States
3		
4	**Row Labels**	**Total Sales**
5	Andrea Lee	$10,138.40
6	Chris Thompson	$6,468.80
7	Jeff Anderson	$9,806.40
8	Suzy Liam	$19,004.60
9	Tom Baker	$40,760.00
10	**Grand Total**	**$86,178.20**

Account Date and Region Report Filters

You may need to create a pivot table report for each unique item in a field. For example, you may need to create a report for each region or each year. You can do this by creating a **report filter**, filtering by each item in the **report filter** list, and copying the output for each item into a new worksheet. A quicker way to do this is to create a **report filter** and use the **Show Report Filter Pages** feature. This feature automates the process of applying a filter for each item and copying the results into separate worksheets. It also names the new worksheet after the item. This saves lots of time, especially if you are creating a large number of reports.

Steps

To add a report filter to a pivot table in Excel:

1. Open the **Sales Report.xlsx** workbook.

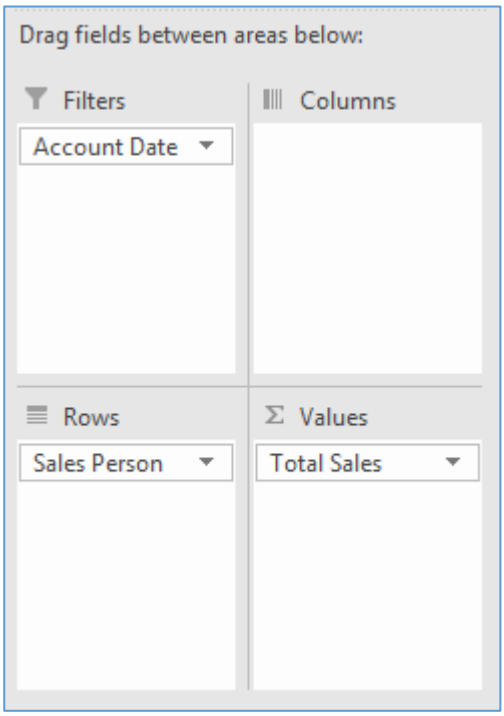
2. Click on any cell in the pivot table to open the **PivotTable Fields** panel.

Total Sales	Column Labels				
Row Labels	2015	2016	2017	2018	Grand Total
Andrea Lee	$0.00		$10,138.40	$25,744.80	$35,883.20
Cathy Smith	$4,440.20	$11,847.40	$17,202.00		$33,489.60
Chris Thompson	$28,983.80	$18,900.60	$6,468.80		$54,353.20
Duncan Bogg	$1,920.60		$20,561.00	$25,237.40	$47,719.00
Jason Patron	$0.00		$1,964.80	$8,022.60	$9,987.40
Jeff Anderson		$1,164.60	$9,806.40		$10,971.00
Kiran Dev	$0.00		$4,184.20	$27,992.60	$32,176.80
Linda Jones	$20,314.40	$14,450.60	$11,371.00		$46,136.00
Scott Andrews	$0.00		$20,650.40	$13,186.40	$33,836.80
Suzy Liam		$1,166.60	$19,004.60	$4,659.20	$24,830.40
Tom Baker	$0.00	$18,805.80	$40,760.00	$3,846.80	$63,412.60
Grand Total	**$55,659.00**	**$66,335.60**	**$162,111.60**	**$108,689.80**	**$392,796.00**

Pivot Table in Sales Report

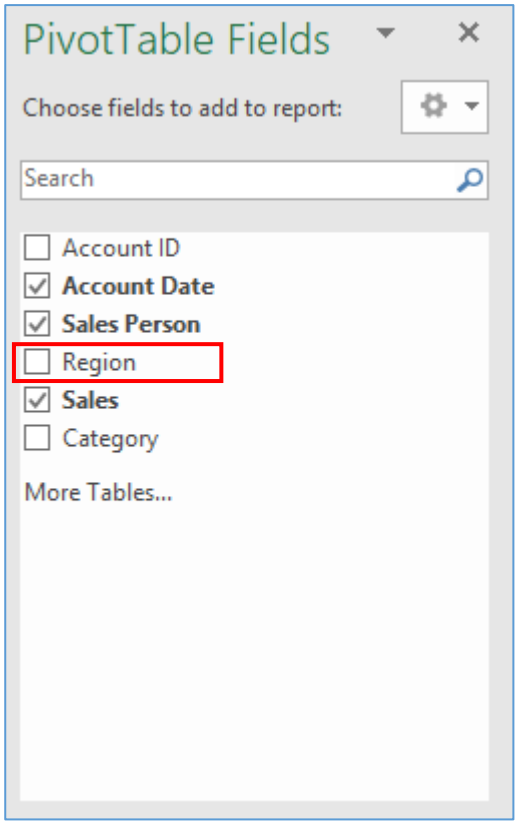
3. To add the account date as a filter, move the **Account Date** field from the **Columns** area to the **Filters** area in the **PivotTable Fields** panel.

   ***Note**: The Account Date has been grouped by year.*



Moving Account Date from Columns to the Filters area

4. To add the **Region** as a filter, drag and drop the **Region** field to the **Filters** area.



Moving the Region field to the Filters area.

The applied report filters (**Account Date** and **Region**) appear above the pivot table.

	H	I
1	Account Date	(All)
2	Region	(All)
3		
4	**Row Labels**	**Total Sales**
5	Andrea Lee	$35,883.20
6	Cathy Smith	$33,489.60
7	Chris Thompson	$54,353.20
8	Duncan Bogg	$47,719.00
9	Jason Patron	$9,987.40
10	Jeff Anderson	$10,971.00
11	Kiran Dev	$32,176.80
12	Linda Jones	$46,136.00
13	Scott Andrews	$33,836.80
14	Suzy Liam	$24,830.40
15	Tom Baker	$63,412.60
16	**Grand Total**	**$392,796.00**

Adding Fields to the Filters Area

5. To use the report filters to view the summary for the **United States** in the year **2017**, select the year **2017** from the **Account Date** drop-down list and click **OK**.

6. And select the region **United States** from the **Region** drop-down list and click **OK**.

	H	I
1	Account Date	2017
2	Region	United States
3		
4	**Row Labels**	**Total Sales**
5	Andrea Lee	$10,138.40
6	Chris Thompson	$6,468.80
7	Jeff Anderson	$9,806.40
8	Suzy Liam	$19,004.60
9	Tom Baker	$40,760.00
10	**Grand Total**	**$86,178.20**

PivotTable with Report Filters Applied

7. To clear the filters from the pivot table, select **All** from the **Account Date** drop-down list and click **OK**.

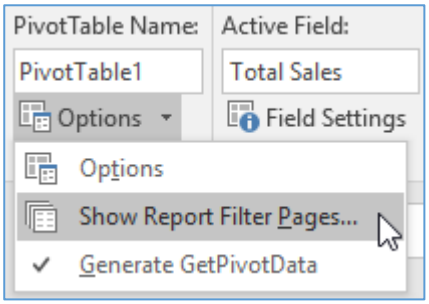
8. And select **All** from the **Region** drop-down list and click **OK**.

   ***Note**: To remove the report filters from the pivot table, remove the **fields** from the **Filters** area.*

To use the **Show Report Filter Pages** feature:

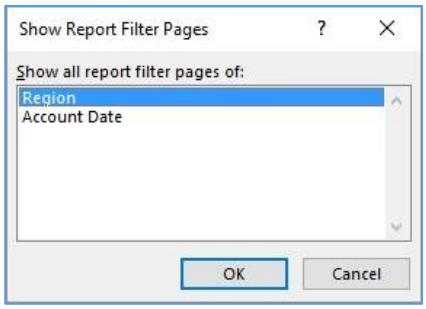
1. Click on any cell in the pivot table (with report filters applied).

2. On the **Analyze** tab, in the **PivotTable** group, click the **Options** down-pointing arrow, and click **Show Report Filter Pages**.



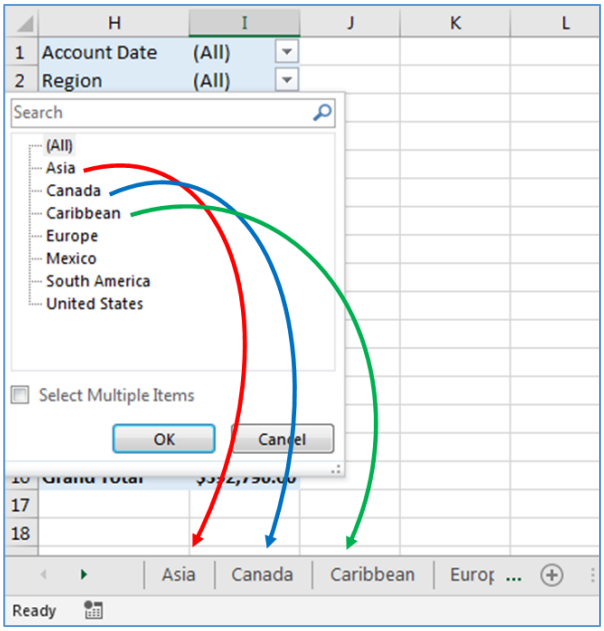
Show Report Filter Pages Command

3. Select the required filter field, in this case **Region**, and click **OK**.



Show Report Filter Pages Dialog Box

A worksheet is created for each item in the **Region** report filter list with a pivot table filtered for that item. The worksheet is also renamed to match the item name.



New Worksheets Created for Each Item

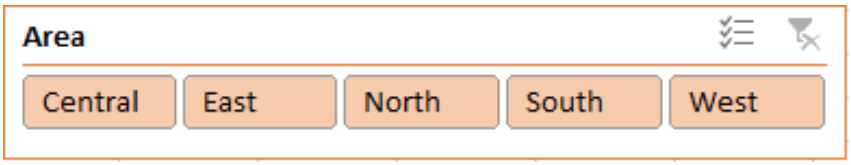
***Note:** When changes are made to the original pivot table after the application of the filters, those changes will **NOT** automatically be updated in the pivot tables in the new report filter pages.*

8.2 USING PIVOT TABLE SLICERS

Concepts

As well as using report filters, you can also use slicers to filter data visually in a pivot table. The slicer feature provides buttons that you can click to quickly and easily filter the pivot table data, making reports interactive.

For example, you may have a pivot table that shows the total amount spent by different types of customers on different product categories and you create a slicer that allows you to filter this information by different areas. The slicer below has buttons that allow you to filter by the areas Central, East, North, South and West.



A pivot table slicer

Slicers also indicate the current filtering state, which makes it easy to understand what is shown in a filtered pivot table. When you select an item in a slicer, that item is included in the filter and the data for that item is displayed in the report.

You can also customise slicers by changing slicer options. For example, you can change the name of the slicer for formulas, the header name displayed in the slicer, the order in which the items are displayed in the slicer, the colour of the slicer, the number of columns, the button size and the slicer size.

You can also connect a slicer to multiple pivot tables to create useful cross-filtered reports. A slicer can only be connected to multiple pivot tables when the pivot tables have the same source data range.

Steps

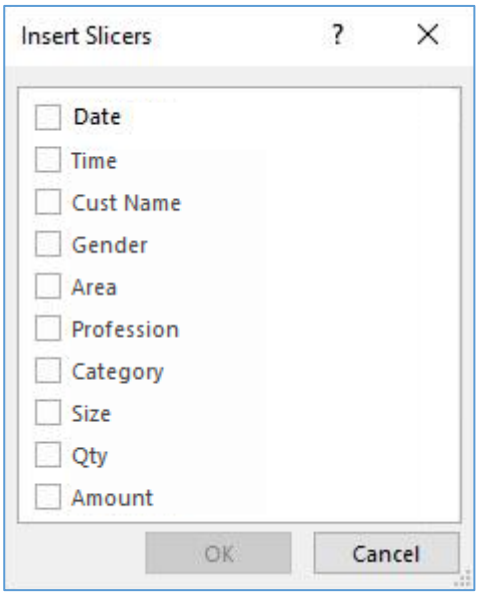
To apply slicers to a Pivot Table:

1. Open the **Pivot Slicers.xlsx** workbook and go to the **Slicer1** worksheet.

2. The pivot table shows total amount spent by different types of customers on different product categories. Click on any value in the pivot table.

3. On the **Analyze** tab, in the **Filter** group, click **Insert Slicer**.

The **Insert Slicers** dialog box appears.

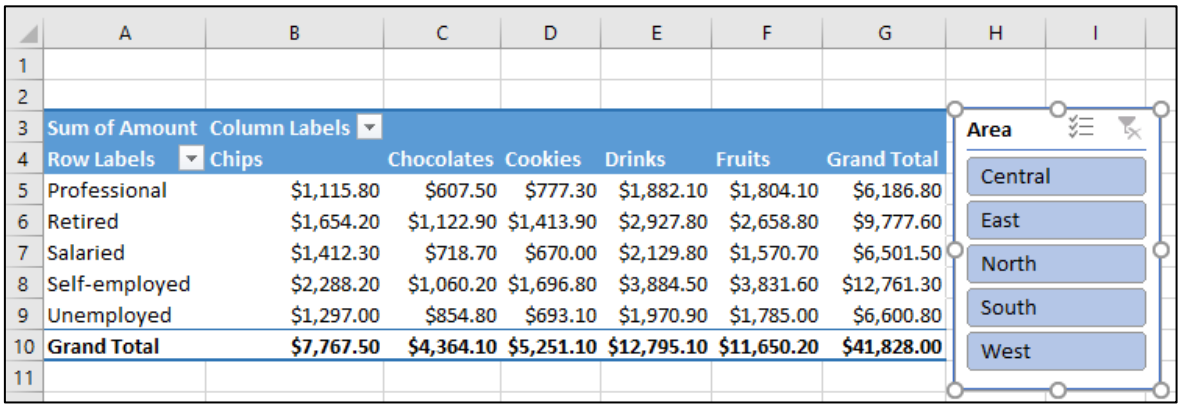


Insert Slicers Dialog Box

4. Select the relevant field, in this case, select **Area** and click **OK**.

   The slicer appears on the worksheet alongside the PivotTable (in a layered display if you have more than one slicer).

5. Click and drag the slicer to position it, for example, to the right of the pivot table. Resize the slicer as necessary using the sizing handles.



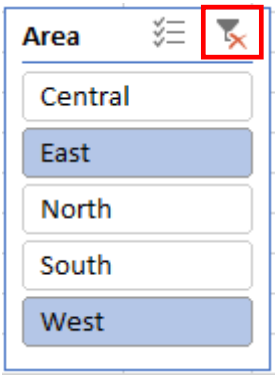
Sum of Amount	Column Labels					
Row Labels	Chips	Chocolates	Cookies	Drinks	Fruits	Grand Total
Professional	$1,115.80	$607.50	$777.30	$1,882.10	$1,804.10	$6,186.80
Retired	$1,654.20	$1,122.90	$1,413.90	$2,927.80	$2,658.80	$9,777.60
Salaried	$1,412.30	$718.70	$670.00	$2,129.80	$1,570.70	$6,501.50
Self-employed	$2,288.20	$1,060.20	$1,696.80	$3,884.50	$3,831.60	$12,761.30
Unemployed	$1,297.00	$854.80	$693.10	$1,970.90	$1,785.00	$6,600.80
Grand Total	**$7,767.50**	**$4,364.10**	**$5,251.10**	**$12,795.10**	**$11,650.20**	**$41,828.00**

Pivot table with a Slicer

6. Use the slicer to filter the pivot table by doing one of the following:

   a. To select a single item, click the item button.
   b. To select multiple contiguous items, click the first item button, press **SHIFT**, and click the last item button.
   c. To select multiple non-contiguous items, press **CTRL** and click on the item buttons.

7. Click the **Clear Filter** icon at the top right of the slicer to clear the filter.

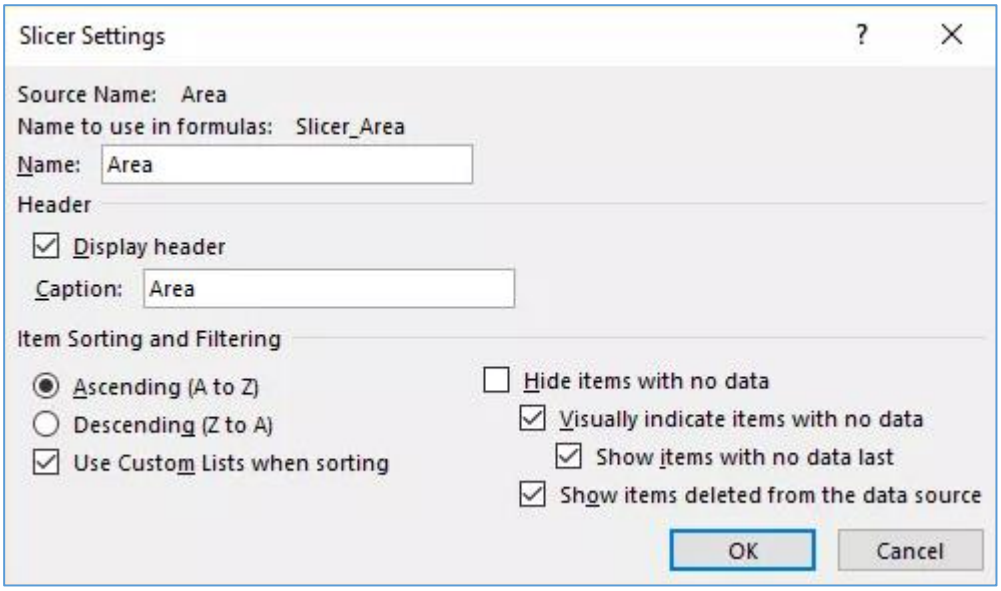


Clear Filter in Slicer

To change the slicer settings:

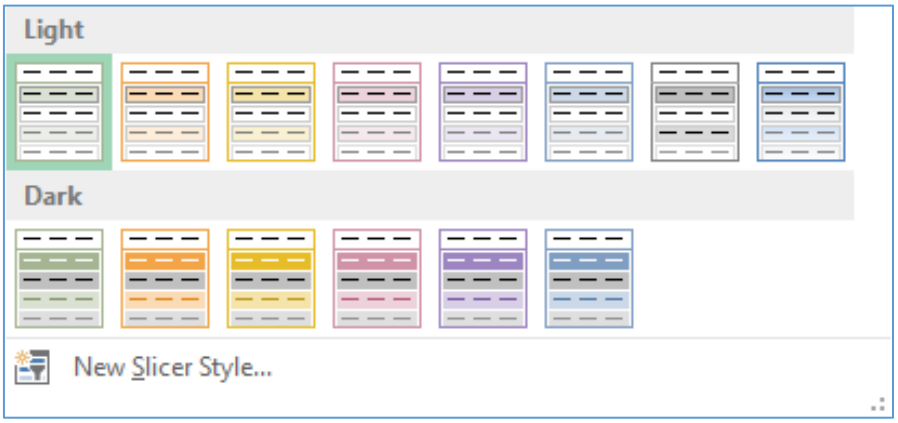
1. On the **Options** tab, in the **Slicer** group, click **Slicer Settings**.

2. Change the Slicer's **Name** and display options for **Header** and **Item Sorting and Filtering** if required.



Setting Slicer Options

3. On the **Options** tab, in the **Slicer Styles** group, select a style for the slicer – in this case **Light Orange, Slicer Style Light 2**.



Applying Slicer Style

4. On the **Options** tab, in the **Buttons** group, select the number of columns – in this case, type **5** in the **Columns** box, and change the button height and width as necessary.

5. The slicer dimension may need to be adjusted to ensure all the buttons are visible. On the **Options** tab, in the **Size** group, change the overall size of the slicer height and width as necessary.

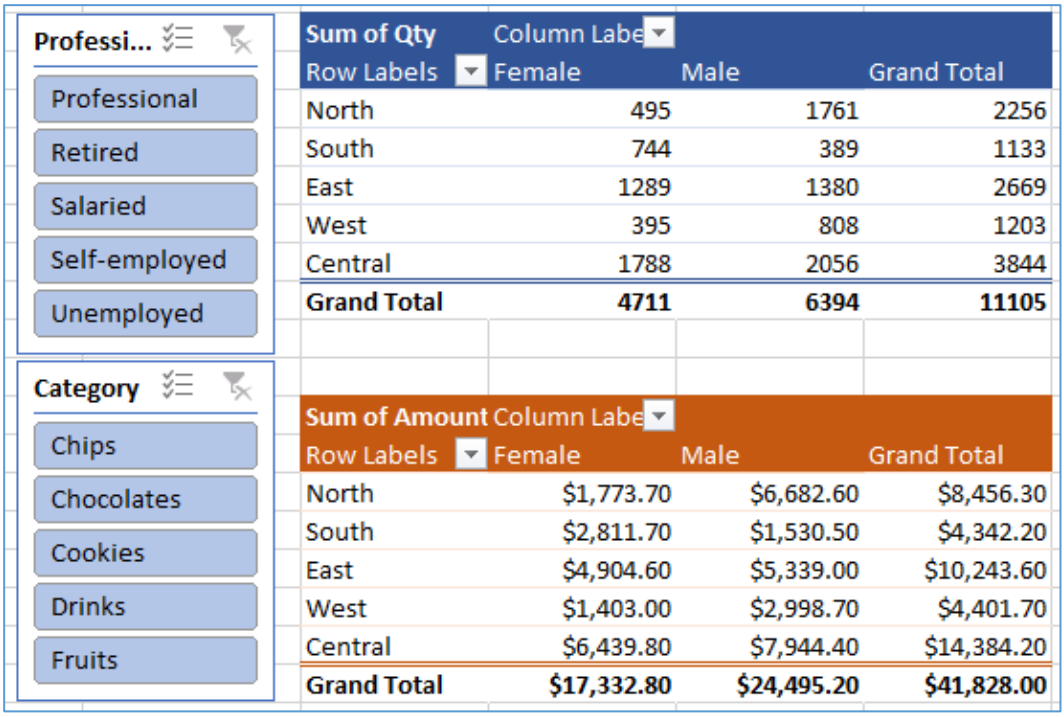
   The slicer style is changed to light orange and the number of columns is changed from 1 to 5.

Formatted Slicer

***Note:** To delete a slicer, click on the slicer to select it, and press **Delete**. Or right-click the slicer and select **Remove <Name of Slicer>** from the context menu.*

To connect slicers across multiple pivot tables:

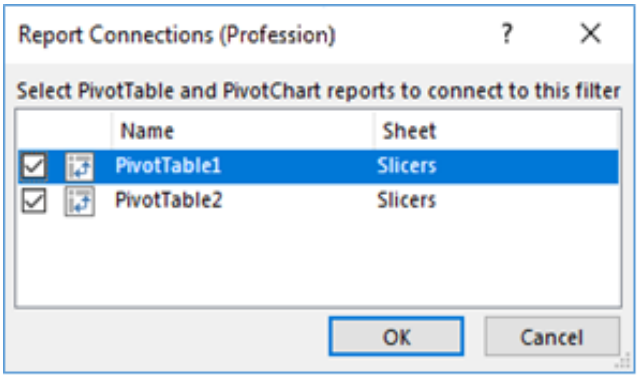
1. Open the **Share Slicers.xlsx** workbook and go to the **Slicers** worksheet.
2. Select the **Profession** slicer next to the top pivot table.



Sum of Qty	Column Labe		
Row Labels	Female	Male	Grand Total
North	495	1761	2256
South	744	389	1133
East	1289	1380	2669
West	395	808	1203
Central	1788	2056	3844
Grand Total	**4711**	**6394**	**11105**

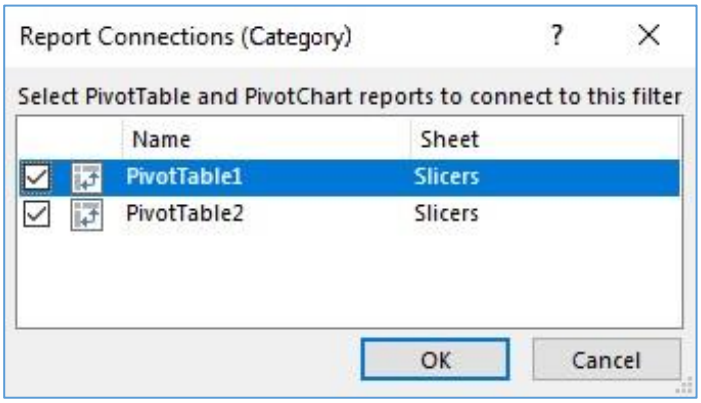
Sum of Amount	Column Labe		
Row Labels	Female	Male	Grand Total
North	$1,773.70	$6,682.60	$8,456.30
South	$2,811.70	$1,530.50	$4,342.20
East	$4,904.60	$5,339.00	$10,243.60
West	$1,403.00	$2,998.70	$4,401.70
Central	$6,439.80	$7,944.40	$14,384.20
Grand Total	**$17,332.80**	**$24,495.20**	**$41,828.00**

3. On the **Options** tab, in the **Slicer** group, click **Report Connections**.
4. In the **Report Connections** dialog box, check the option for **PivotTable2** to link the **Profession** slicer to the bottom pivot table
5. Click **OK**.



Report Connections Dialog Box

6. Repeat steps 3 and 4 to connect the **Category** slicer to the top pivot table (**PivotTable1**).



Connecting the Category Slicer

The slicers (**Profession** and **Category**) are connected to both pivot tables. Making a selection in either slicer filters the data in both pivot tables accordingly.

8.3 USING TIMELINES

Concepts

To easily display data in a pivot table for different time periods, you can use the **Timeline** option in Excel. Timelines are based on date fields, so to create a timeline the pivot table needs to include at least one date-formatted field.

A Timeline has a scroll bar that allows you to focus on a time period before selecting a specific time period in the date field to display. You can choose what level the time period is shown as - years, quarters, months, or days.

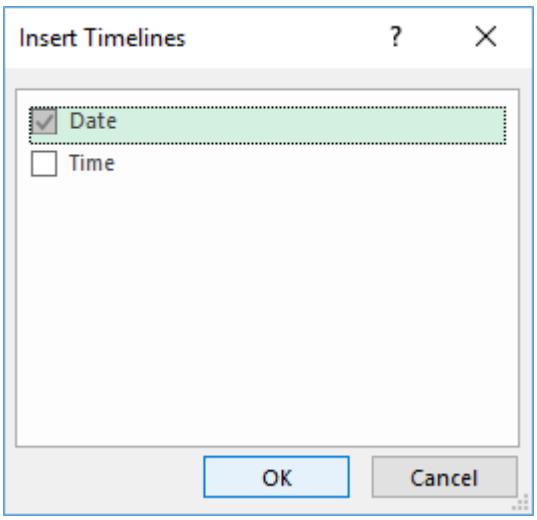
Similar to slicers, you can customise and format Timelines. You can also use a single Timeline to filter numerous pivot tables, if the pivot tables use the same data source.

Steps

To use a **Timeline** filter control in a pivot table:

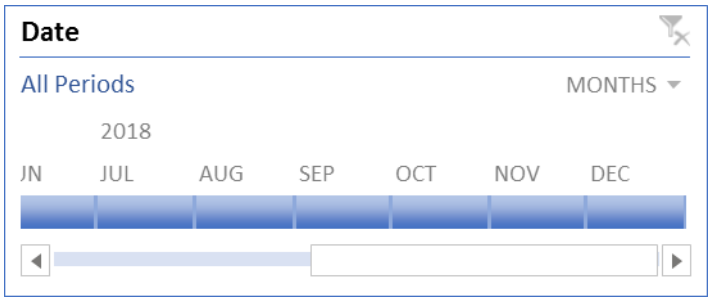
1. Open the **Timeline.xlsx** workbook and go to the **Pivot** worksheet.
2. Select any value in the first pivot table in the worksheet.
3. On the **Analyze** tab, in the **Filter** group, click **Insert Timeline**.
4. In the **Insert Timelines** dialog box select the date-formatted field on which to base the timeline – in this case, click **Date** and click **OK.**

   ***Note**: You can add multiple timelines by selecting multiple date-formatted fields.*



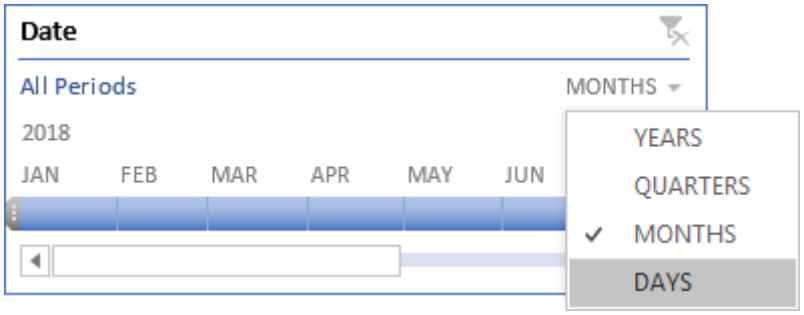
Insert Timelines Dialog Box

5. Click and drag the timeline to place it above the PivotTable.



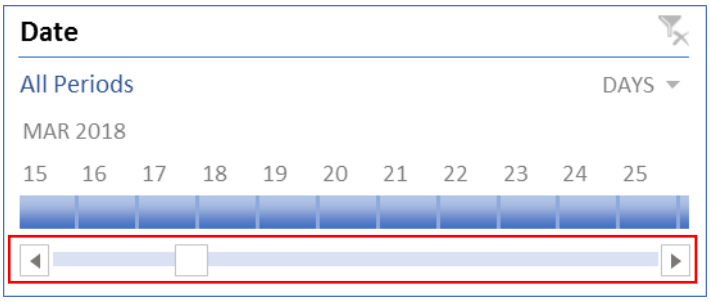
Timeline Filter

6. To change the time level, click the arrow next to the time level shown and select the new time level – in this case, **DAYS**.



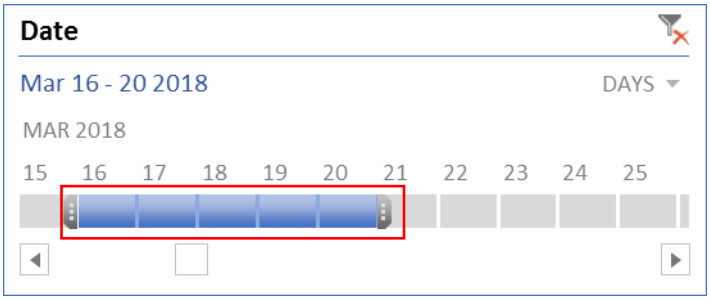
Selecting a Time Level

7. Drag the Timeline scroll bar to focus in on a particular time period - in this case, starting from the **15th March**.



Time Level Changed to Reflect Days

8. To select the time period to display, click a period tile in the timespan control and drag at either end of the tile to select the relevant tiles - in this case, from **16th** to **20th Mar**.



Adjusting the Date Range

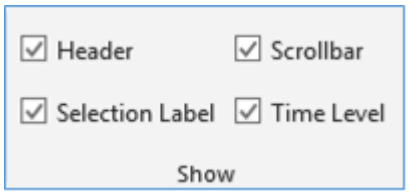
9. On the **Options** tab, in the **Timeline Styles** group, select a style for the timeline.

Applying Timeline Style

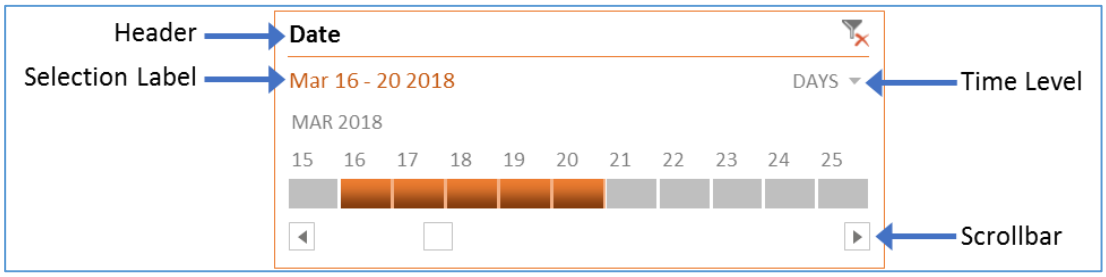
10. On the **Options** tab, in the **Size** group, change the overall size of the timeline height and width as required.

Setting Size for the Timeline

11. On the **Options** tab, in the **Show** group, set the options you want to display in the timeline.



Setting the Show Options



Timeline Elements

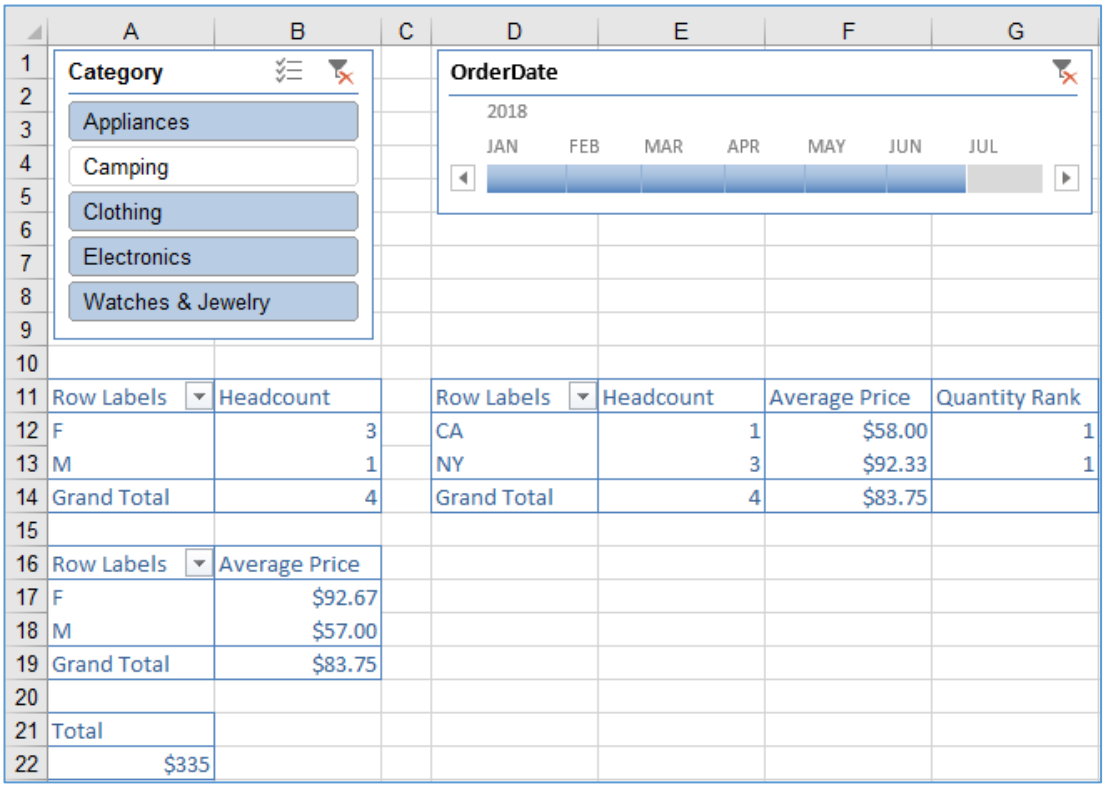
***Note:** To clear the Timeline, click the **Clear Filter** icon*  *at the top right of the timeline.*

*To delete a Timeline, select the **Timeline** and press **Delete** or right-click the Timeline and select **Remove Timeline**.*

*To use one Timeline to filter multiple pivot tables, select the **Timeline** and in the **Options** tab select **Report Connections** and select the pivot tables to include. The pivot tables need to use the same data source.*

8.4 REVIEW EXERCISE

1. Open the **Purchases.xlsx** workbook.
2. Go to the **Pivot** worksheet.
3. Add a slicer using the field **Category** for the pivot table in cell **A11:B14**.
4. Move and resize the slicer so that it fits the range **A1:B9**.
5. Link the slicer with the rest of the PivotTables in the worksheet.
6. Add a timeline using the field **OrderDate** for the PivotTable in cell **A11:B14**.
7. Change the time level to **Months**.
8. Remove the display of the **Time Level**, **Selection Label** and **Scrollbar** in the timeline.
9. Move and resize the slicer so that it fits the range **D1:G5**.
10. Link the timeline with the rest of the PivotTables in the worksheet.
11. Select all the categories except **Camping** in the **Category** slicer.
12. Set the time period from **January** to **June** in the **OrderDate** timeline.



Row Labels	Headcount
F	3
M	1
Grand Total	4

Row Labels	Headcount	Average Price	Quantity Rank
CA	1	$58.00	1
NY	3	$92.33	1
Grand Total	4	$83.75	

Row Labels	Average Price
F	$92.67
M	$57.00
Grand Total	$83.75

Total
$335

LESSON 9 –
USING PIVOT CHARTS

In this section, you will learn how to:

- Create a pivot chart from a pivot table
- Create a pivot chart from data in a table

9.1 Inserting Pivot Charts from Pivot Tables

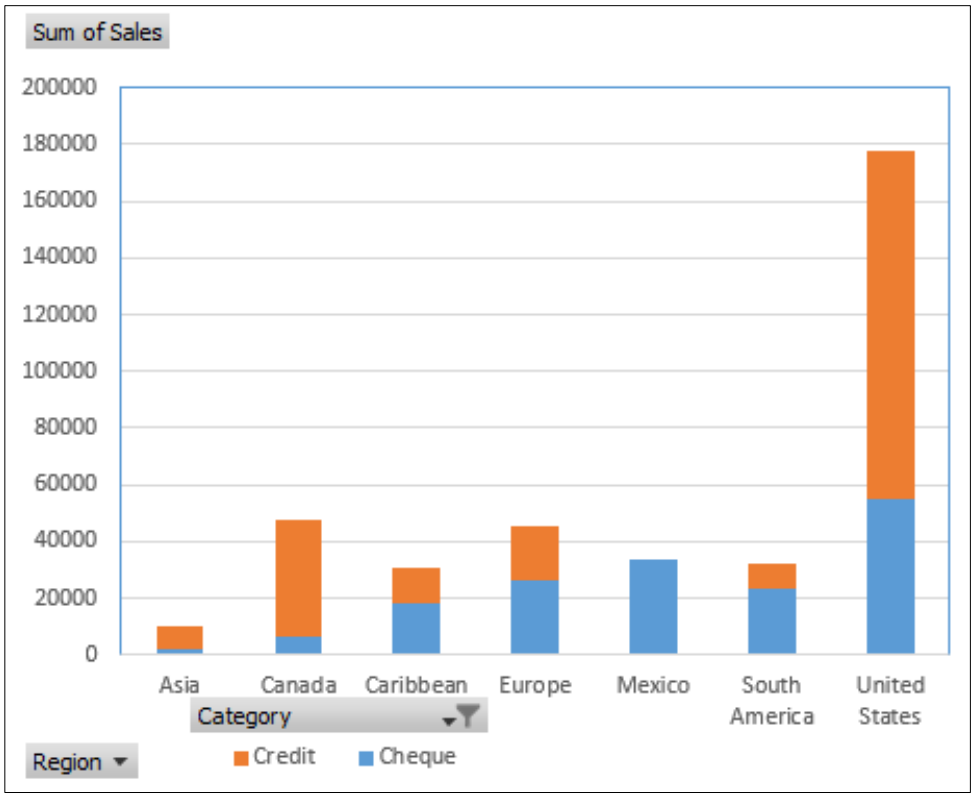
Concepts

You may want to add a visualization to an existing pivot table. Pivot charts are an effective way to create data visualizations that also allow you to display different views of your data.

You can create a pivot chart from an existing pivot table report. When you create a pivot chart from an existing pivot table report, the pivot chart reflects the view of the existing pivot table report. For example, the pivot table report below shows the totals of the cheque and credit card transactions for different regions. The corresponding pivot chart shows the same view.

Sum of Sales	Column Labels		
Row Labels	Cheque	Credit	Grand Total
Asia	1964.8	8022.6	9987.4
Canada	6193.4	41525.6	47719
Caribbean	18617.4	12503.6	31121
Europe	26204.8	19234.8	45439.6
Mexico	33836.8		33836.8
South America	23392.6	8784.2	32176.8
United States	54961.8	122605	177566.8
Grand Total	**165171.6**	**212675.8**	**377847.4**

Pivot Table for Sales



Corresponding Pivot Chart for Sales

The pivot chart uses the pivot table layout to determine the placement of fields.

- **Row fields** in the **pivot table** become **category fields** in the **pivot chart**.
- **Column fields** in the **pivot tabl**e become **series fields** in the **pivot chart**.
- **Pivot table fields** are displayed as **field buttons** in the **pivot chart**. These can be used for filtering.

In the example shown:

- The **row fields** (**Region**) in the **pivot table** are shown on the **category axis** in the **pivot chart**.
- The **column fields** (**Cheque and Credit**) in the **pivot table** are shown on the **series axis**.
- The **pivot table fields** (**Region** and **Category**) are displayed as **field buttons** in the **pivot chart**.

When you have created your pivot chart you can use it to further analyse your data. You can manipulate pivot chart data in the same way you can manipulate pivot table data. Changing the information displayed in a pivot chart automatically changes the corresponding pivot table, and vice versa. You can also add slicers to the pivot chart.

You can customize the visual display and layout of your pivot chart to suit your needs – for example, you can change the chart type, adjust the chart layout and formatting.

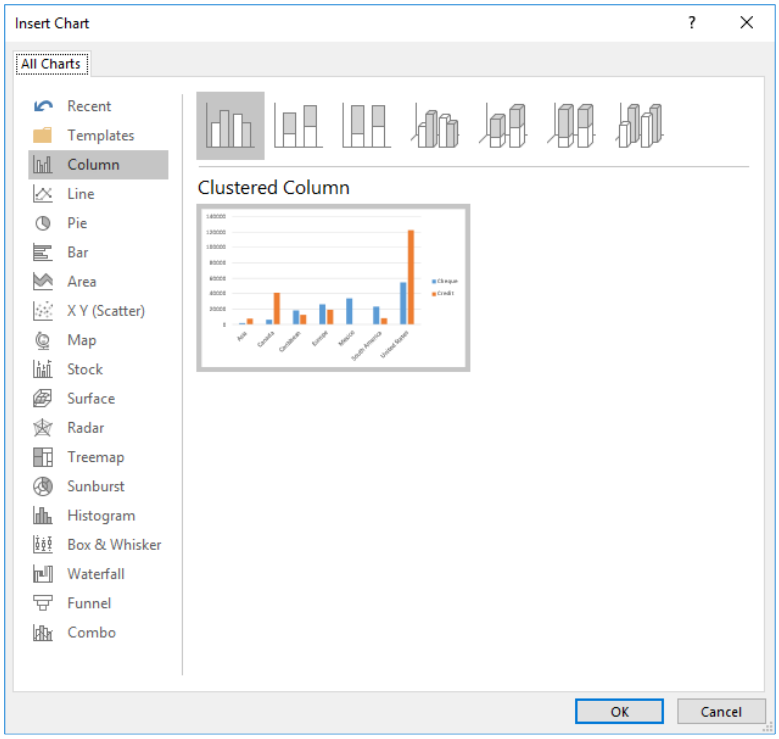
If you delete the corresponding pivot table, the pivot chart changes to an ordinary chart.

Steps

To insert a pivot chart from an existing pivot table report:

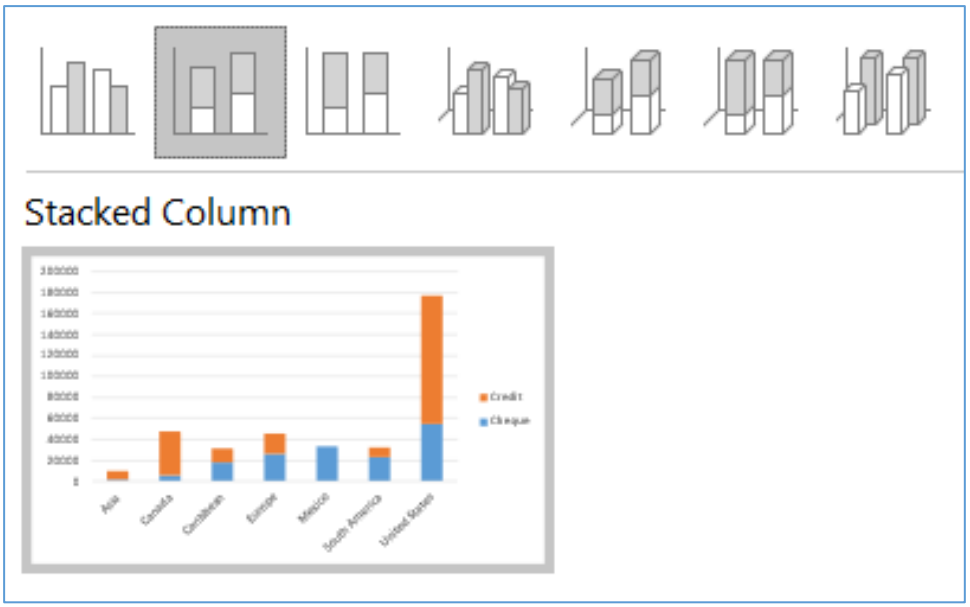
1. Open the **Regional1.xlsx** workbook.
2. Click on any cell in the pivot table.
3. On the **Analyze** tab, in the **Tools** group, click **PivotChart**.

4. In the **Insert Chart** dialog box, select the chart type – **Column** is selected be default.



Insert Chart Dialog Box

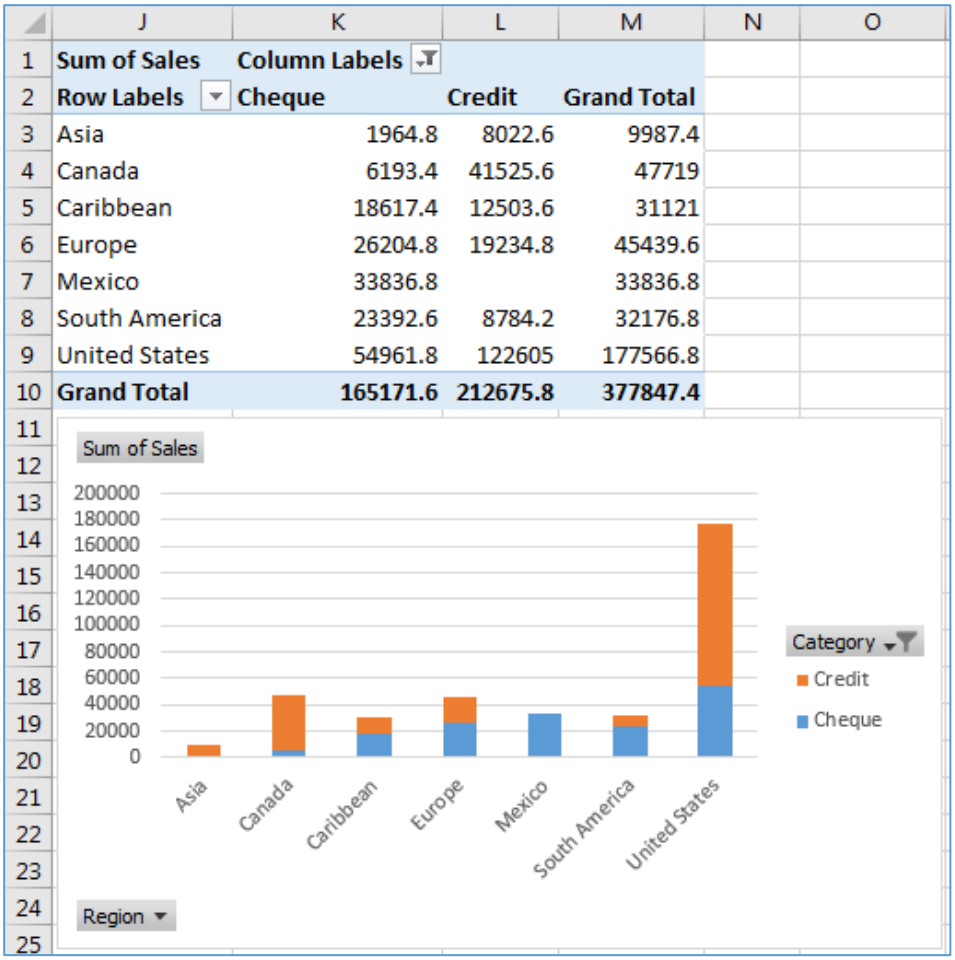
5. Select the chart variant – in this case, **Stacked Column** and click **OK**.



Stacked Column Chart Option in Excel

6. The pivot chart is placed on the same sheet as the pivot table it is based on. Move and resize the chart as necessary.

   ***Note**: To move the pivot chart to a different worksheet (or to a Chart sheet), on the **Analyze** tab in the **Actions** group, click on **Move Chart**.*



PivotChart created from an Existing PivotTable

When you have created the pivot chart you can customise the appearance and layout to suit your needs.

7. To add a chart title above the chart, click on the **PivotChart area** and select the **Chart Elements** button.

8. Select **Chart Title**, select **Above Chart** and enter the title – in this case, enter **Credit and Cheque Transactions**

9. To hide field buttons, deselect the relevant option in the **Field Buttons** drop-down menu, in the **Show/Hide** group, on the **Analyze** tab – in this case, deselect **Show Axis Field Buttons**, **Show Legend Field Buttons** and **Show Value Field Buttons**.

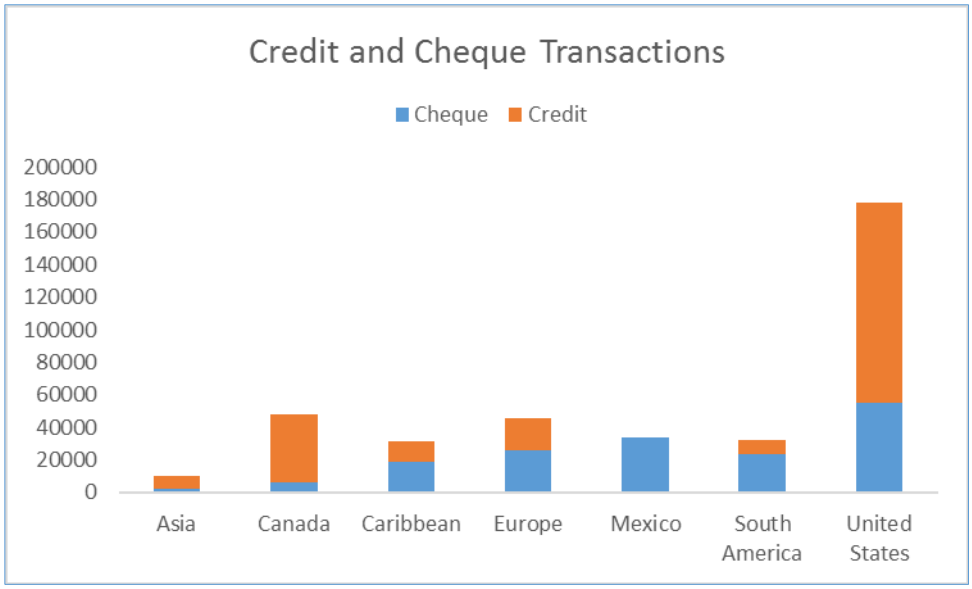
   ***Note**: Alternatively, to hide field buttons, right-click the **field button** and select the relevant option.*

10. To remove gridlines, click on the **PivotChart area** and select the **Chart Elements** button

11. Deselect **Gridlines.**

12. To move the legend below the chart title, click on the **PivotChart area** and select the **Chart Elements** button.

13. Hover over Legend and select **Top**.



Customised Pivot Chart

9.2 Creating Pivot Charts from Tables

Concepts

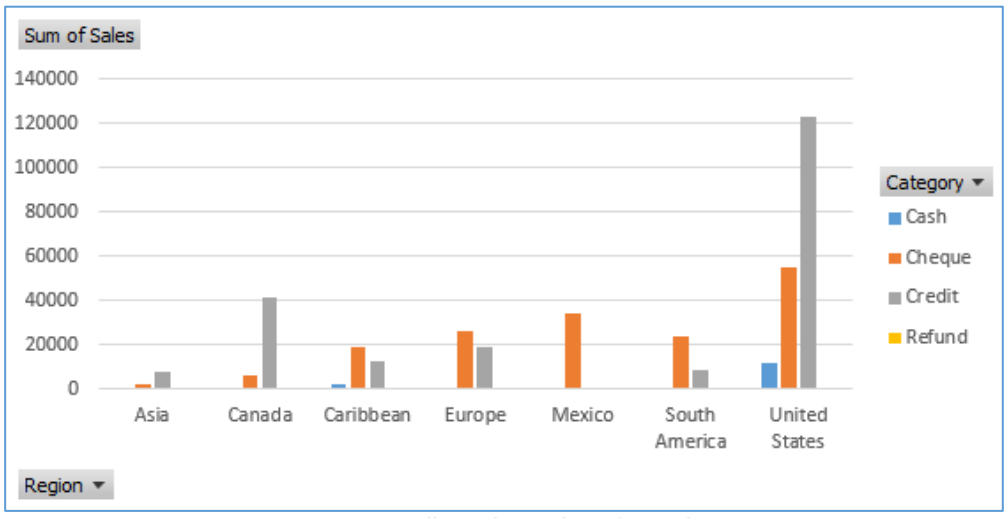
As well as creating a pivot chart from an existing pivot table report, you can create a pivot chart directly from data stored in a table.

When you create a pivot chart directly from fields in a table in Microsoft Excel, a corresponding pivot table is automatically created. A pivot chart must have a corresponding pivot table because the data used to plot the pivot chart comes from a pivot table.

In the example below, the **Region**, **Sales**, and **Category** fields in the **table** are used to create the corresponding **pivot chart**, and the matching **pivot table** is automatically created in Excel.

Account ID	Account Date	Sales Person	Region	Sales	Category
AMP3961	20/6/2018	Andrea Lee	United States	$6,498.40	Credit
SPR4030	9/5/2015	Andrea Lee	United States	$0.00	Refund
PAL3983	19/1/2017	Andrea Lee	United States	$10,138.40	Credit
ECO3964	16/7/2018	Andrea Lee	United States	$19,246.40	Credit
VIR3855	23/7/2015	Chris Thompson	United States	$1,900.40	Cash
THE3934	16/9/2017	Chris Thompson	United States	$1,804.60	Cash
BEA3865	28/10/2015	Chris Thompson	United States	$9,004.60	Credit

Data set for Sales



Corresponding Pivot Chart for Sales

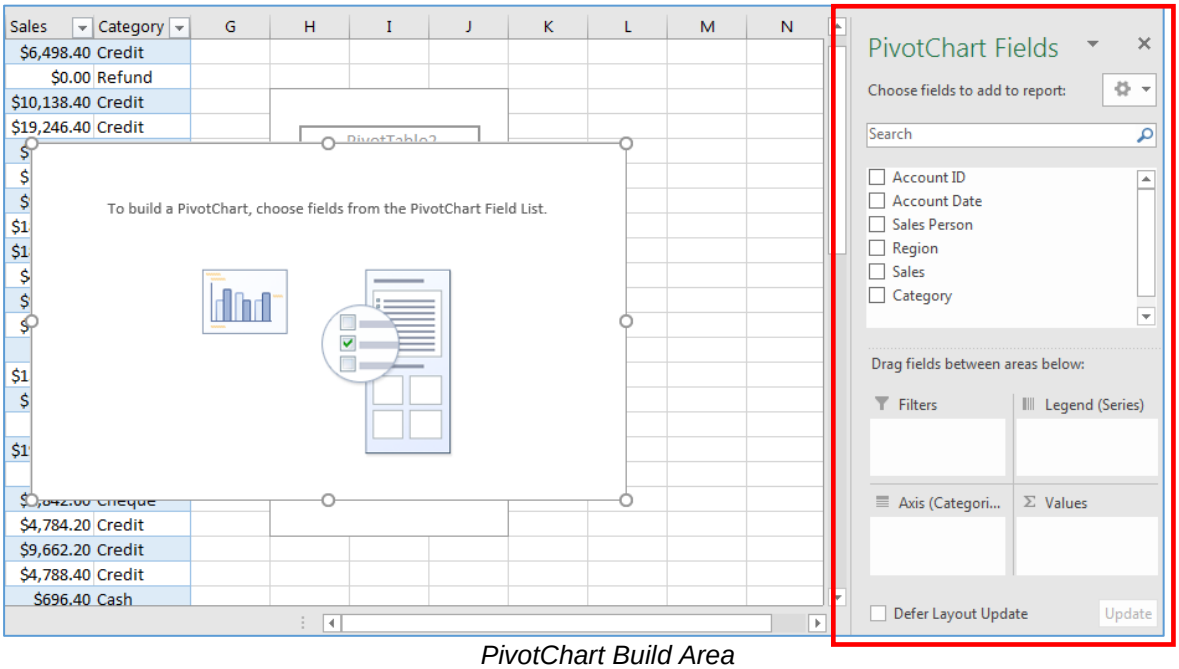
Sum of Sales	Column Labels				
Row Labels	Cash	Cheque	Credit	Refund	Grand Total
Asia		1964.8	8022.6	0	9987.4
Canada		6193.4	41525.6		47719
Caribbean	2368.6	18617.4	12503.6		33489.6
Europe	696.4	26204.8	19234.8		46136
Mexico		33836.8		0	33836.8
South America		23392.6	8784.2	0	32176.8
United States	11883.6	54961.8	122605	0	189450.4
Grand Total	**14948.6**	**165171.6**	**212675.8**	**0**	**392796**

Corresponding Pivot Table for Sales

When you create a pivot chart directly from fields in a table, a **PivotChart** build area is displayed with the **PivotChart Fields** panel.

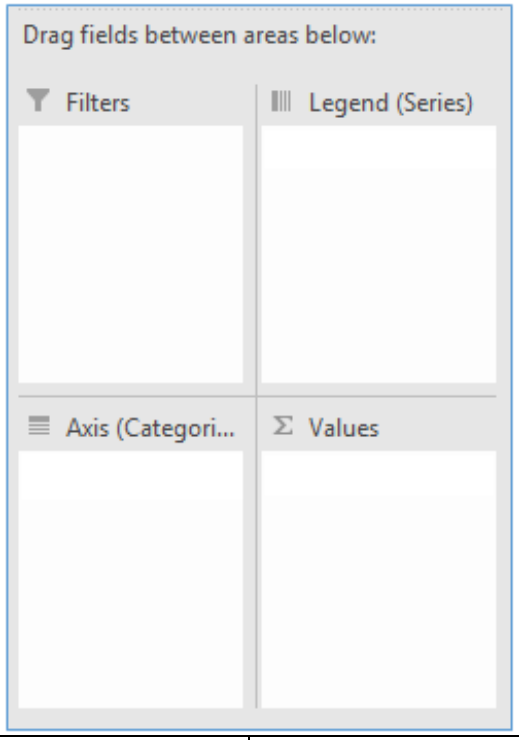
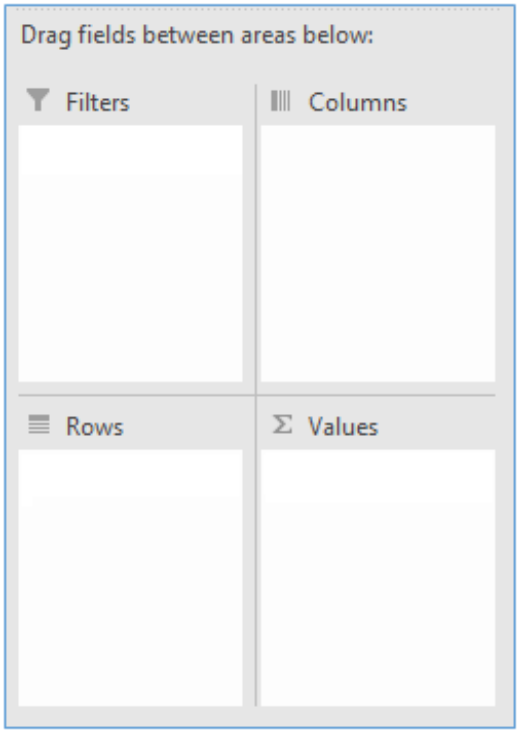
The **PivotChart Fields** pane corresponds to the **PivotTable Fields** panel.

You build the pivot chart in the same way you would a pivot table.



PivotChart Build Area

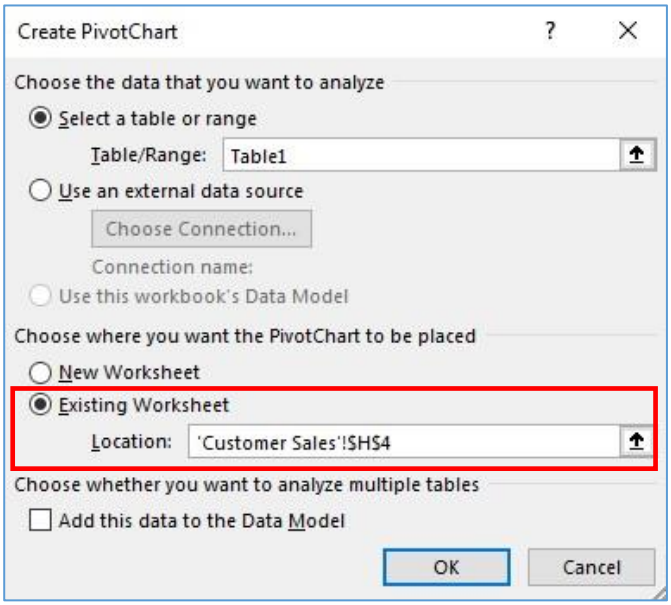
The table below describes how the elements in a **PivotChart** correspond to the elements in a **PivotTable**.

<table>
<tr><th colspan="2">PivotChart Areas</th><th colspan="2">PivotTable Areas</th></tr>
<tr><td colspan="2"> 
</td><td colspan="2"> 
</td></tr>
<tr><td>Filters: Use to filter the entire chart based on the selected item in the report filter.</td><td>Legend (Series): Use to display fields in the legend of the chart and as data series.</td><td>Filters: Use to filter the entire report based on the selected item in the report filter.</td><td>Columns: Use to display fields as columns at the top of the report.</td></tr>
<tr><td>Axis (Categories): Use to display fields along the horizontal (category) axis in the chart.</td><td>Values: Use to display summary data for the data series.</td><td>Rows: Use to display fields as rows on the side of the report.</td><td>Values: Use to display summary data.</td></tr>
</table>

Steps

To insert a pivot chart using data set fields in a table:

1. Open the **Regional2.xlsx** workbook.
2. Select any cell in the table.
3. On the **Insert** tab, in the **Charts** group, click **PivotChart**.
4. In the **Create PivotChart** dialog box, select the option **Existing Worksheet**.
5. Click in the **Location** box and select cell **H4**



Create PivotChart Dialog Box

***Note**: Clicking a cell in the table previously means the **Select a table or range** option is selected by default.*

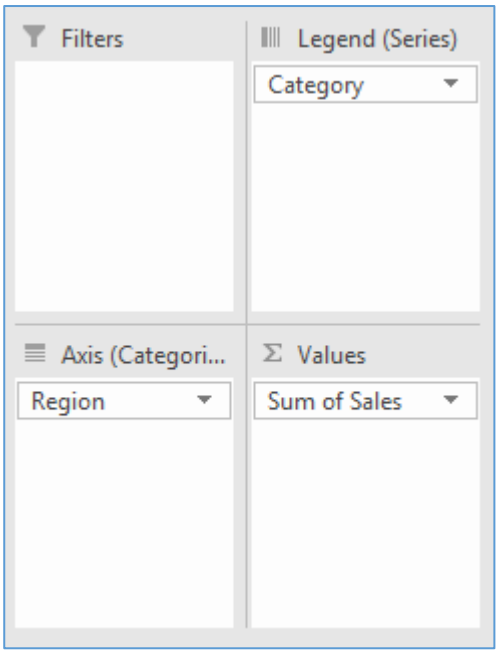
6. Click **OK**.

7. In the **PivotChart Fields** pane, drag and drop the fields to the respective areas as follows:

    a) To show each **Category** as a **data series**, drag and drop **Category** to **Legend (Series)**

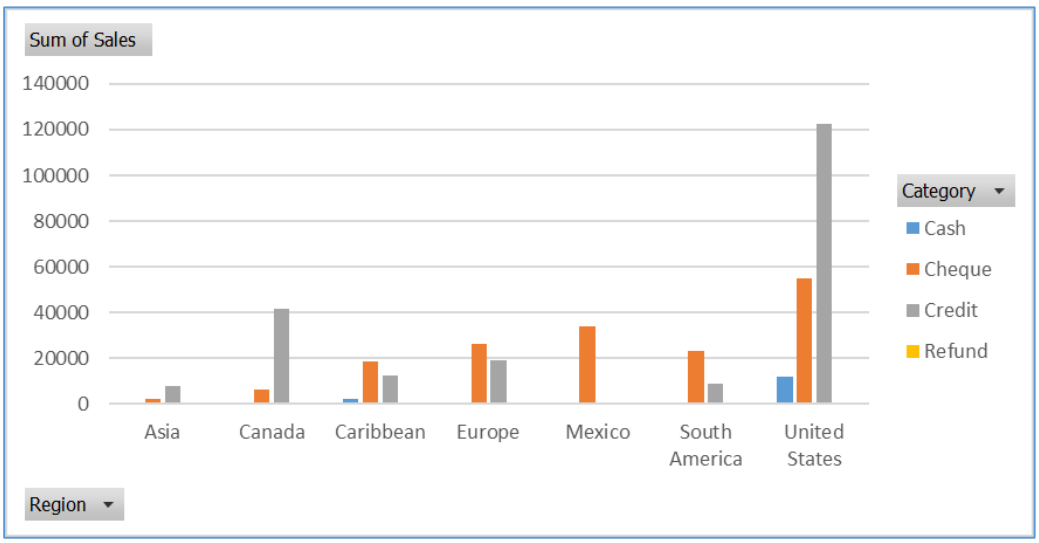
    b) To show each **Region** as a **data category** on the horizontal axis, drag and drop **Region** to **Axis (Categories)**

    c) To show the **Sales summary values** for each **category** in each **region**, drag and drop **Sales** to **Values**.



Fields added to PivotChart Areas

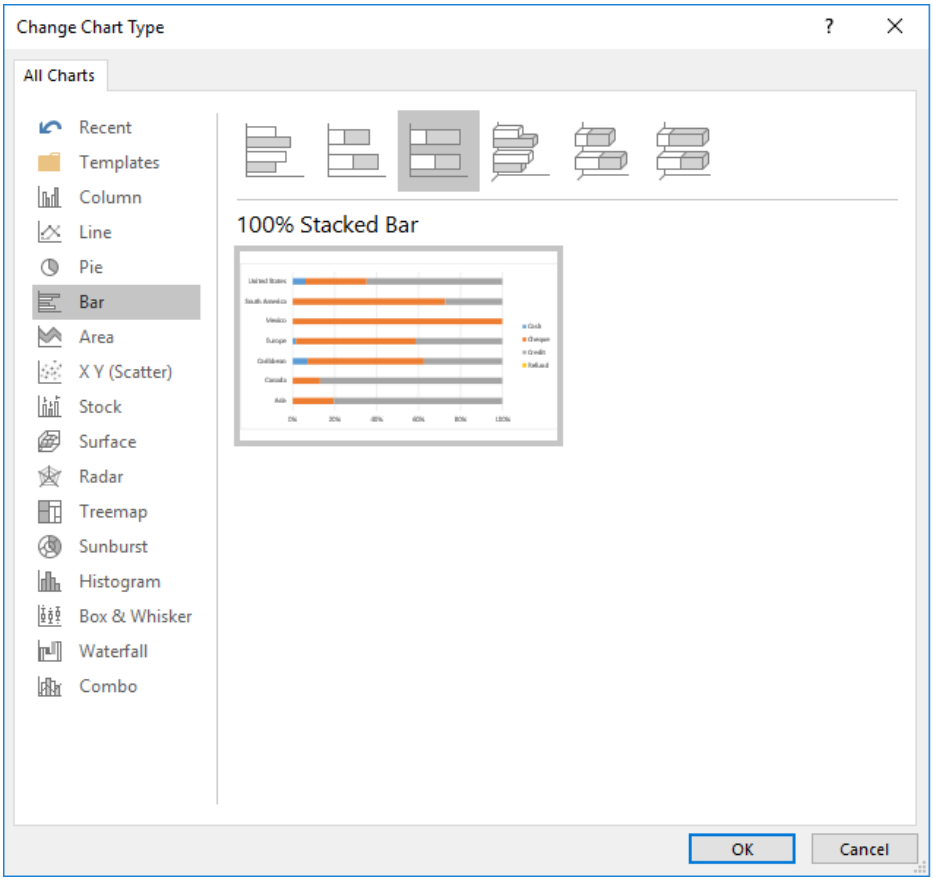
The pivot chart is created using a default chart type (a clustered column chart).



Default Chart Type

8. To change the chart type, on the **Design** tab, in the **Type** group, click **Change Chart Type**.

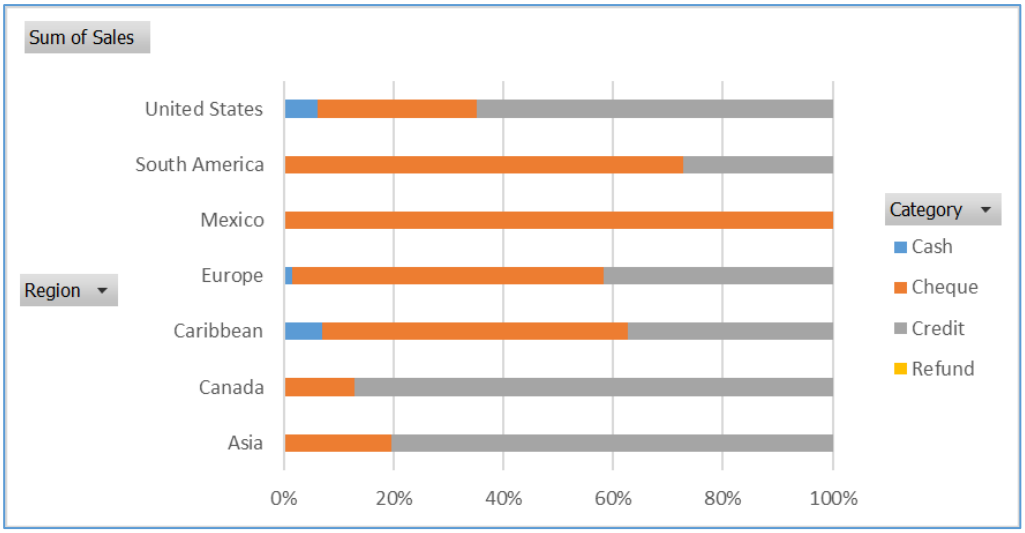
9. In the **Change Chart Type** dialog box, select the chart type, for example, select **Bar** in the chart type and **100% Stacked Bar** in the chart variant



Changing the Pivot Chart Type

10. Click **OK**.

11. The chart type is changed to a **100% stacked bar chart**. Move and resize the chart as necessary.

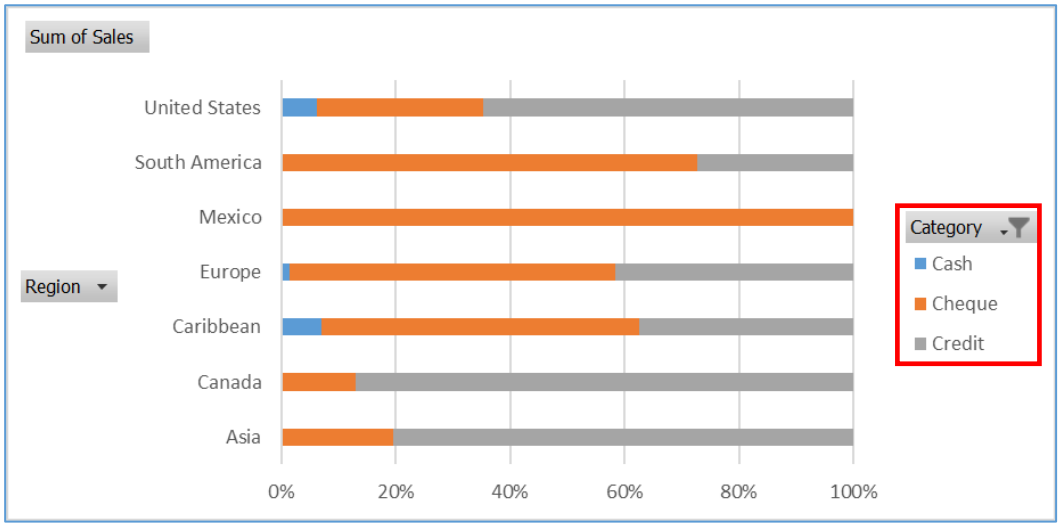


100% Stacked Bar Chart

12. You can change the information that is displayed in the pivot chart. For example, the **Legend field** button (**Category** in this example) allows you to choose the data series information to display and the **Axis** field button (**Region** in this example) allows you to choose the axis information to display.

    In this case, to exclude the **Refund** data series, select the **Category** field button, deselect **Refund** and click **OK**.

    The Refund data series is no longer displayed in the legend.

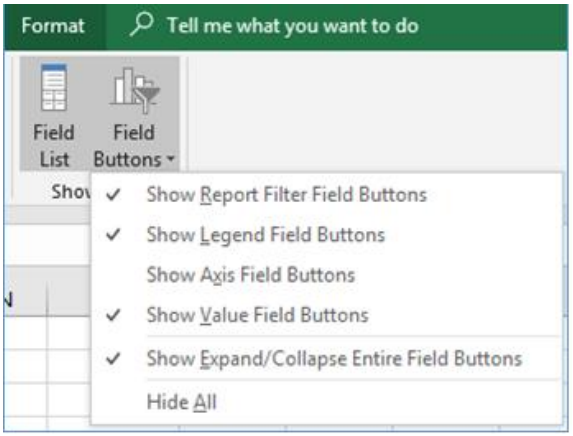


Filtered Pivot Chart

13. You can customize the layout of the chart. To add a chart title, click on the **PivotChart** area and select the **Chart Elements** button.

14. Select **Chart Title** and enter the chart title **Popular Payment Methods** above the chart.

15. To hide field buttons, deselect the relevant option in the **Field Buttons** drop-down menu, in the **Show/Hide** group, on the **Analyze** tab – in this case, deselect **Show Axis Field Buttons** to remove the **Region** field button.



Field Buttons Control for Pivot Chart

***Note**: Alternatively, to hide field buttons, right-click the field button and select the relevant option.*

16. To insert a slicer for the pivot chart, click **Insert Slicer** in the **Filter** group in the **Analyze** tab.

17. Select **Region**, to add a slicer using the **Region** field.

18. Click **OK**.

19. To format the slicer to display the buttons in **2** columns, set the **Columns** field to **2** in the **Buttons** group in the **Options** tab.

20. Move the slicer to the right of the pivot chart and resize it to show all the button labels.

21. You can customize the appearance of the pivot chart. To change the style of the chart, click on the **PivotChart** area and select the **Chart Styles** button.

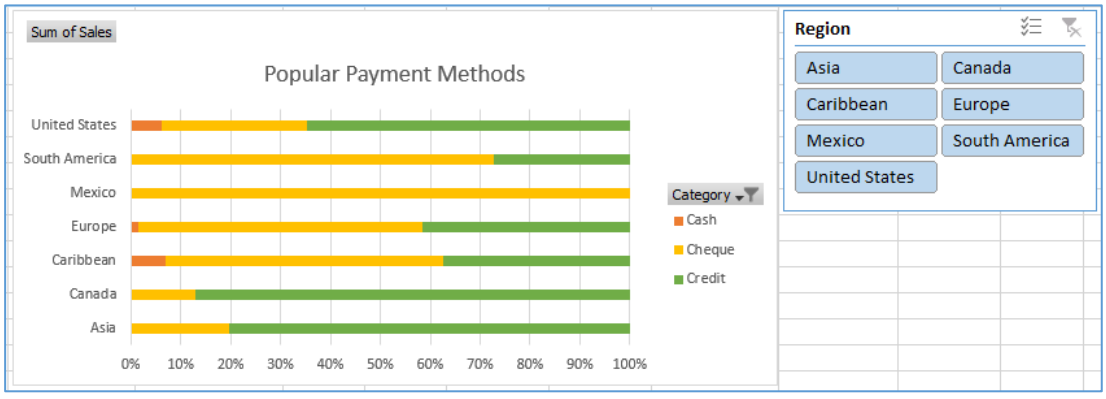
22. Select a new style – in this case, select **Style 9**.

23. To change the colour of the chart, click on the **PivotChart** area and select the **Chart Styles** button.

24. Select **Color**.

25. Select a new colour – in this case, select **Colorful Palette 3**.

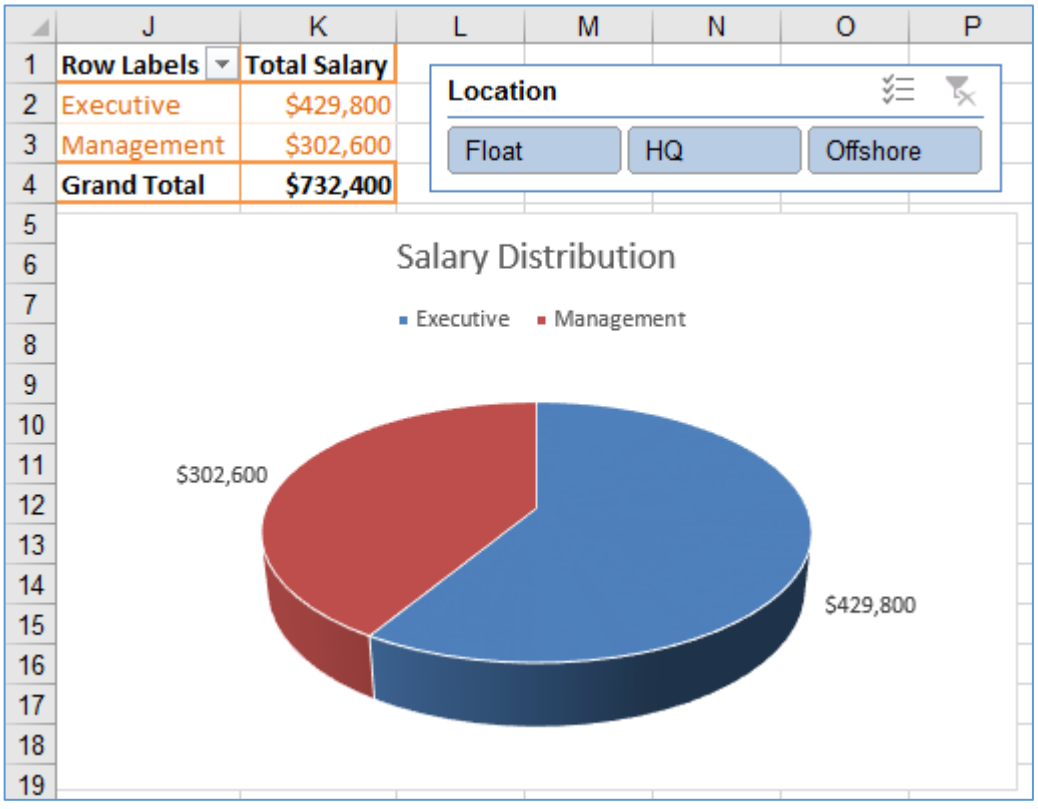
    The information displayed in the pivot chart, as well as the layout and appearance have been updated and a slicer has been added.



Customized Pivot Chart with Slicer

9.3 REVIEW EXERCISE

1. Open the **Personnel.xlsx** workbook.
2. Insert a **3D Pie** pivot chart for the existing pivot table in the workbook.
3. Move the pivot chart and position it below the pivot table.
4. Remove the field buttons in the pivot chart.
5. Modify the chart title to reflect **Salary Distribution**.
6. Add the data label to the outside end of each pie chart slice.
7. Change the legend position to the top.
8. Add a slicer using the field **Location** for the pivot chart.
9. Set the number of columns for the slicer to **3**.
10. Move and resize the slicer so that it fits the range **L1:P3**.



11. Create a **Clustered Column** pivot chart directly from the data set fields in the table and present it in a new worksheet. The chart should count the number of personnel for each 10000 range of salary. To do this:

    a. Add the fields **Salary** to the **Axis** area and **Staff ID** to the **Values** area.

    b. In the corresponding pivot table, group the Salary from **10000** to **50000** by every **10000**.

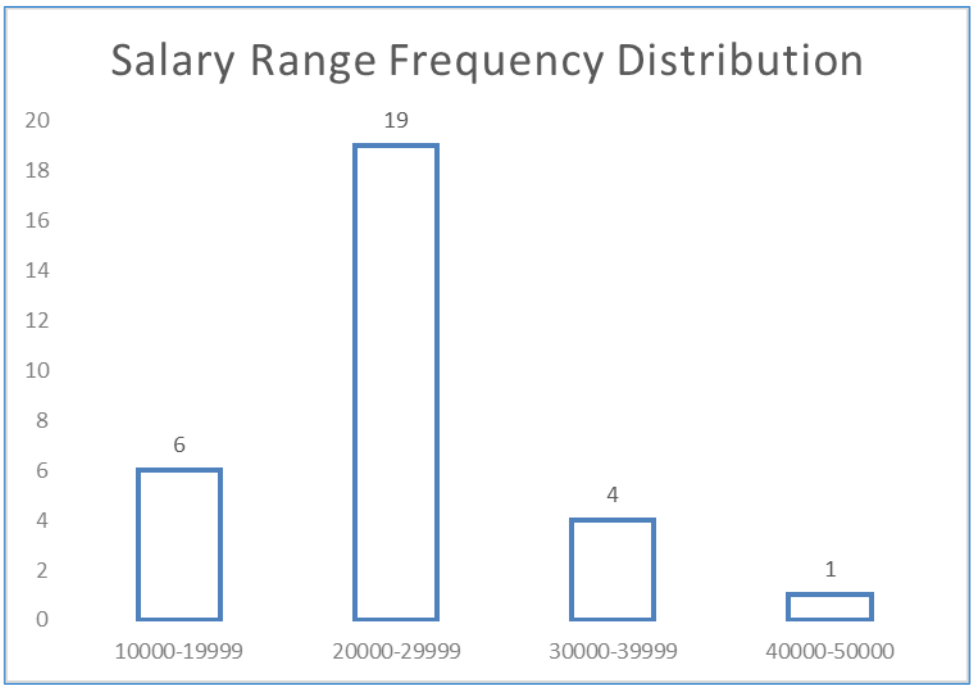
12. Remove the field buttons in the pivot chart.

13. Modify the chart title to reflect **Salary Range Frequency Distribution**.

14. Remove the legend and gridlines.

15. Add data labels to the outside end of the data series.

16. Apply **Style 11** to the pivot chart.



LESSON 10 –
DATA VISUALIZATION TOOLS

In this section, you will learn how to:

- Understand the concept of data visualization using reports and dashboards.
- Outline common visualizations.
- Recognise common data visualization tools and their functions.
- Setup tools for data visualizations.
- Import data from a spreadsheet into a data visualization tool.

10.1 Key Features of Data Visualization Tools

Concepts

Data visualization provides a visual way to access data by displaying trends and patterns within data in a visual form. Data visualization makes it simpler and more efficient to analyse and understand textual and numerical data; it saves time; it provides insights that contribute to business intelligence; and it helps with decision making. It can be used in the earlier stages of data analytics to explore and make sense of the data, as well as for reporting purposes.

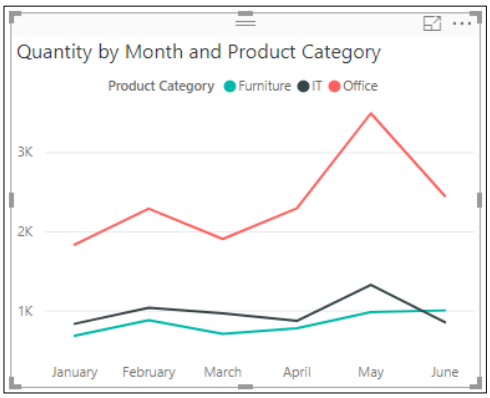
There are lots of specialized data visualization tools designed to help businesses easily perform sophisticated data analytics.

Key Features

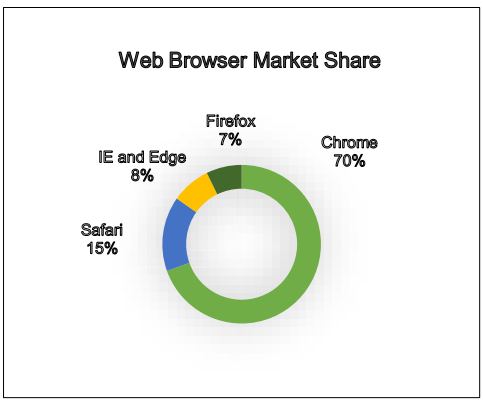
Key features of specialized data visualization tools include connecting to a wide range of **data sources** including spreadsheets and databases, as well as non-traditional sources such as web data and images. The tools allow you to create **data sets** (collections of data) based on one source of data or based on data from multiple sources combined and filtered. For example, you can create a simple data set based on a single table from a spreadsheet or you can create a data set which is a combination of data from a website table, a spreadsheet table, and an online table showing results of a social media marketing campaign.

These powerful tools allow you to explore data sets and find and present information and insights from the data in a way that suits your business needs through the creation of **data visualizations**. There are a wide range of data visualizations provided that can go beyond traditional forms. Part of your data analysis will include determining which types of visualizations provide the best insights and business intelligence for your specific needs. Common examples of data visualizations include:

- **Charts (column, bar, line, pie)** - Useful for displaying relationships among numeric data by category or over time periods.

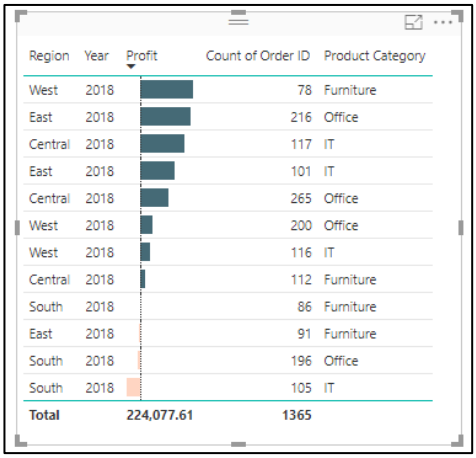


Line Chart



Donut Chart

- **Tables** - Useful for presenting and comparing detailed data and exact values.

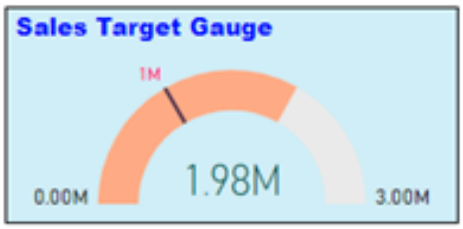


Region	Year	Profit	Count of Order ID	Product Category
West	2018		78	Furniture
East	2018		216	Office
Central	2018		117	IT
East	2018		101	IT
Central	2018		265	Office
West	2018		200	Office
West	2018		116	IT
Central	2018		112	Furniture
South	2018		86	Furniture
East	2018		91	Furniture
South	2018		196	Office
South	2018		105	IT
Total		224,077.61	1365	

Product Sales Table

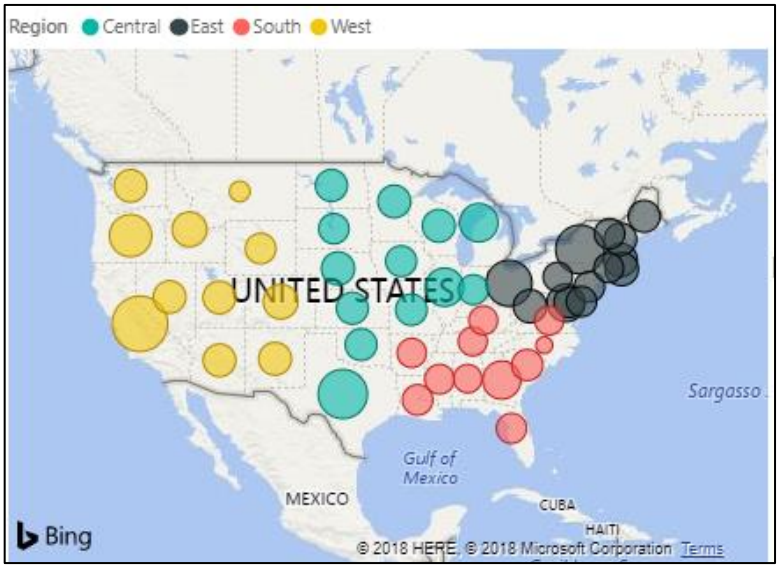
- **Key Performance Indicators** (KPIs and gauge charts) - Useful for providing key indicators for highlighting critical data and measuring performance using predefined conditions.

Sales Target KPI



Sales Target Gauge Chart

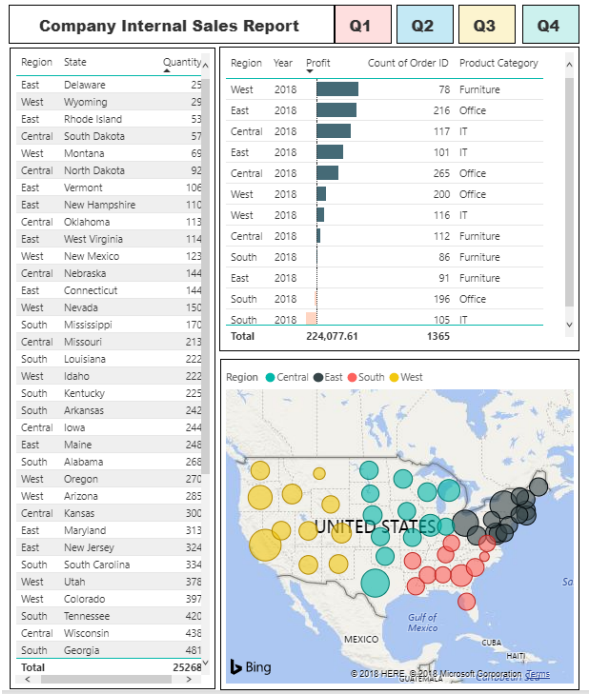
- **Maps** - Useful for presenting the distribution of data across geographic locations.



Sales Volume Map

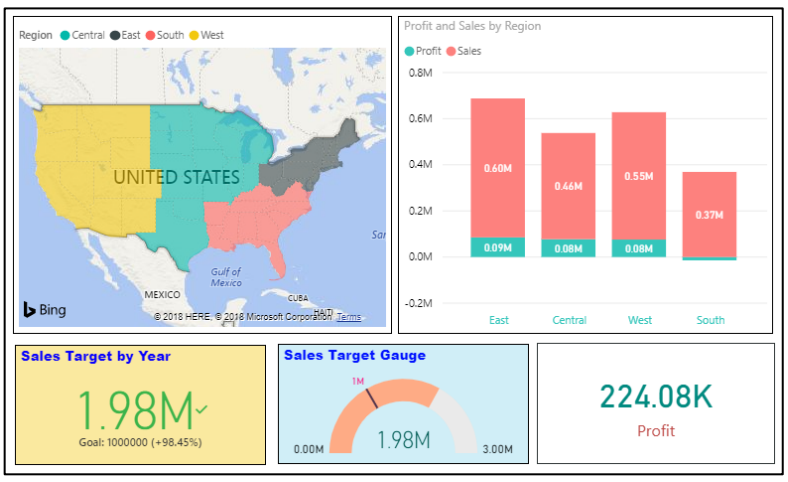
Data visualization tools go beyond presenting information in visual forms and also provide **interactive capabilities** that allow you to filter and drill into the data in visualizations for querying and analysis.

Data visualization tools also enable production of insightful visual **reports**. A basic report includes visualizations and results from the data analytic process. A report can consist of a single page with one or more visualizations or it can consist of multiple pages of visualizations. The aim of reports is to provide key insights from your data that can be used to achieve business goals, such as summary information, trends, and patterns in a visual format.



A Report with Tabular Data and Charts

As well as reports data visualization tools allow you to combine multiple visualizations to display key information and metrics in **dashboards**. Dashboards provide consolidated visual displays allowing information to be monitored at a glance on a single screen. This information is updated in real-time as the underlying data changes. The aim of dashboards is to provide insights into your business data that can be used to achieve business goals. Dashboards tend to be built to visualise and organise data for the purpose of monitoring changes in real-time.



A Dashboard with Multiple Charts and Visual Elements

Data visualization tools also provide the ability to **easily share** the business intelligence information produced in dashboards and reports. And they have the ability to support **collaboration**.

10.2 DATA VISUALIZATION TOOLS SETUP

There are many data visualization tools available. Some commonly used tools include **Microsoft Power BI** and **Tableau**:

Microsoft Power BI
Microsoft Power BI provides powerful data analytics with user-friendly interfaces. It consists of a desktop application called **Power BI Desktop**, an online service called **Power BI service**, and mobile **Power BI apps**.

Power BI Desktop is free to download and install. You can use it to connect to many types of data sources, explore your data using visualizations, and design and create reports using visualizations.

***Note**: In order to share content created in Power BI Desktop, you will need to publish it to **Power BI service**.*

URL: https://powerbi.microsoft.com/

Tableau Public
Tableau Public is a free web-based platform used to create data visualizations, which are uploaded to the **Tableau Public** website. With a large collection of data connectors and visualizations with intuitive design, it is one of the leading tools in data visualization. Each visual comes with shareable code that allows you to tweet it, share it on Facebook, or embed it in your organization's website. There is also a paid version.

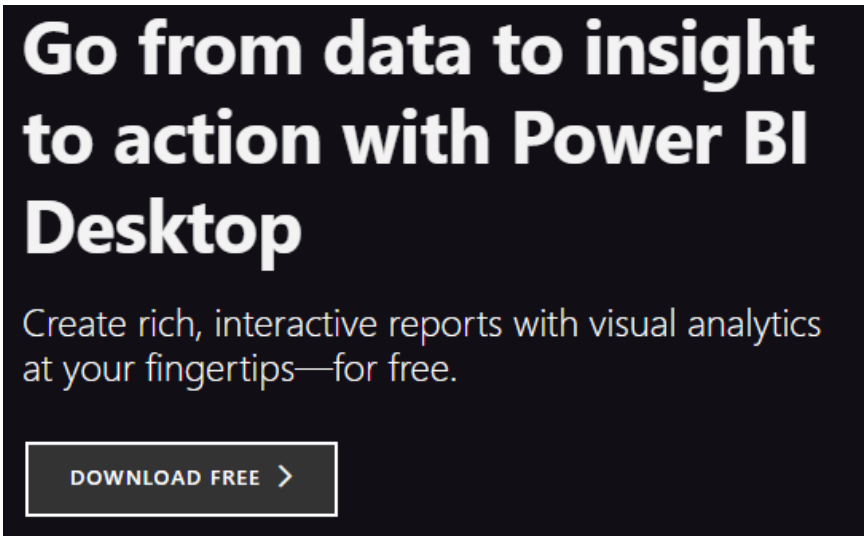
URL: https://public.tableau.com/

Steps

To download **Microsoft Power BI Desktop** to your device:

1. Go to the **Microsoft Power BI Desktop** download page: https://powerbi.microsoft.com/en-us/desktop.

2. Click on **DOWNLOAD FREE** to begin the process of downloading.



Download Page for Power BI Desktop

3. When **Power BI Desktop** is downloaded, run the installation file.

4. Follow the instructions in the setup wizard to install the application.

***Note:** Alternatively, you can download Power BI Desktop as an app by going to the Microsoft Store for Power BI Desktop and clicking **Get** or **Install** to begin the process of downloading.*

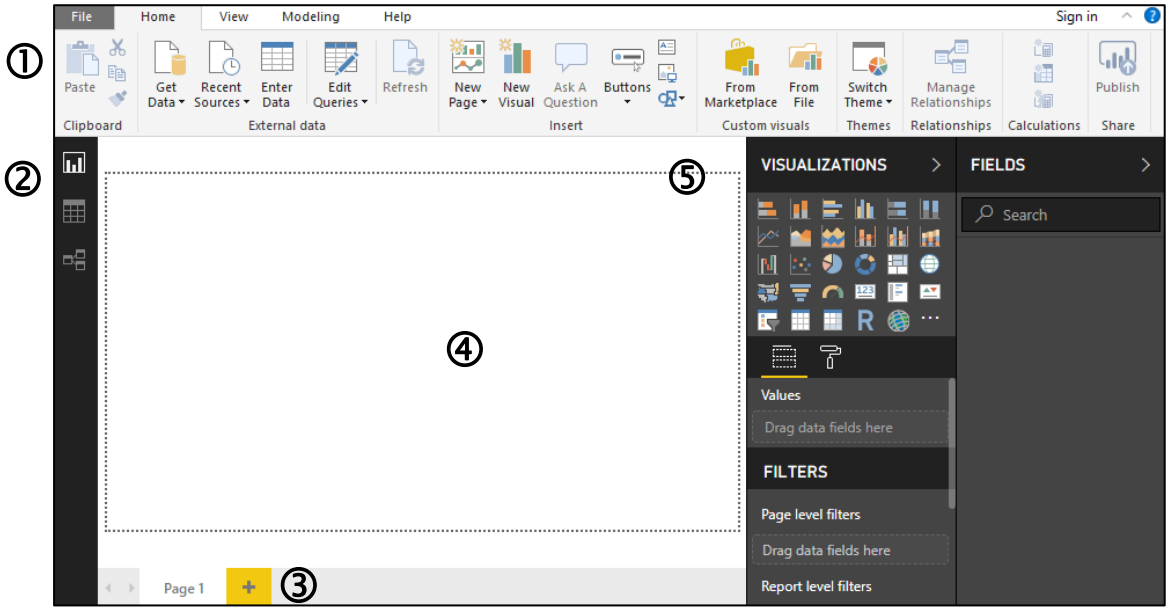
Power BI Desktop in Microsoft Store

10.3 VISUALIZATION TOOL ENVIRONMENT

Concepts

When you have installed the **Microsoft Power BI Desktop** application, you can launch the application by double-clicking the **Power BI Desktop** app icon on your desktop. This opens the **Microsoft Power BI Desktop** environment where you can get data, create visualizations, build reports and publish reports to **Power BI service**.

The main parts of the **Microsoft Power BI Desktop** environment are illustrated below:



Power BI Desktop Screen Interface

① Ribbon

Similar to **Microsoft Excel** and other **Microsoft Office** applications, the **Ribbon** displays the toolbars associated with reports and visualizations.

② Views

There are three main views as outlined below. You can change the view by selecting the relevant icon.


The **Report View**, shown above, is where you build reports. The **Report View** brings the Power BI screen to the report canvas page.


The **Data View** is where you inspect, explore, and understand the underlying data in your Power BI Desktop model.

 The Relationship View provides an overview on the relationships between the fields in different data sets.

③ Pages tab

Each tab at the bottom of the canvas represents a page in a report. The tab lets you select or add a report page.

④ Report canvas

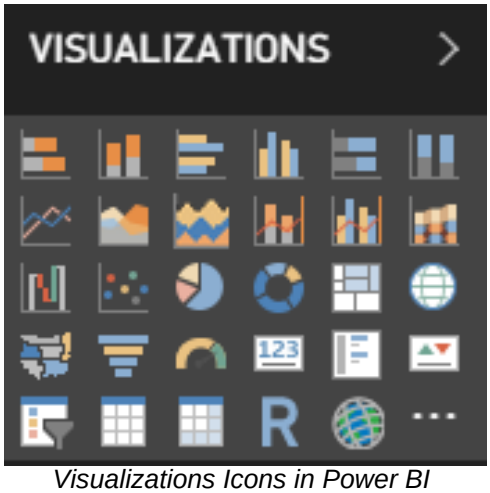
The report canvas is where visualizations are displayed. You use it to display and arrange visualizations and contents of reports.

⑤ **Visualizations**, **Filters** and **Fields** panes

You use the **Visualizations, Filters** and **Fields** panes to create visualizations.

In the **Visualizations** and **Filters** panes you select the visualizations and determine the display of visualizations – for example you can set the colour and format for the visualizations. And you can select the filters for the visualizations.

You can quickly create visualizations by selecting the icons under the **Visualizations** pane.



Visualizations Icons in Power BI

In the **Fields** pane, you select the data to be included in visualizations by selecting from the list of fields.

10.4 Importing Data Sets from Spreadsheets

Concepts

Microsoft Power BI enables you to connect to data from multiple different sources. One of the most common sources is **Microsoft Excel**.

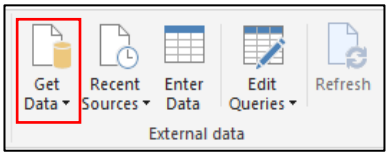
You can import a data set from a spreadsheet in **Microsoft Excel** into **Power BI** and save it in a **Power BI file**.

You can then use the data set in the **Power BI file** to create visualizations and build reports.

Steps

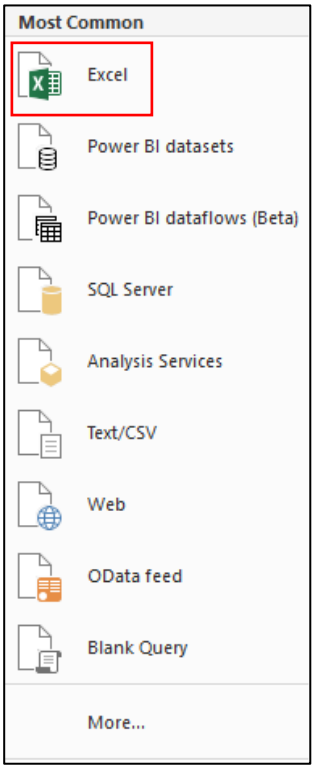
To import a data set from a spreadsheet into **Power BI Desktop:**

1. Double-click on the **Power BI Desktop** app icon to launch the application.
2. On the **Home** tab, in the **External Data** group, click **Get Data**.



The External Data group

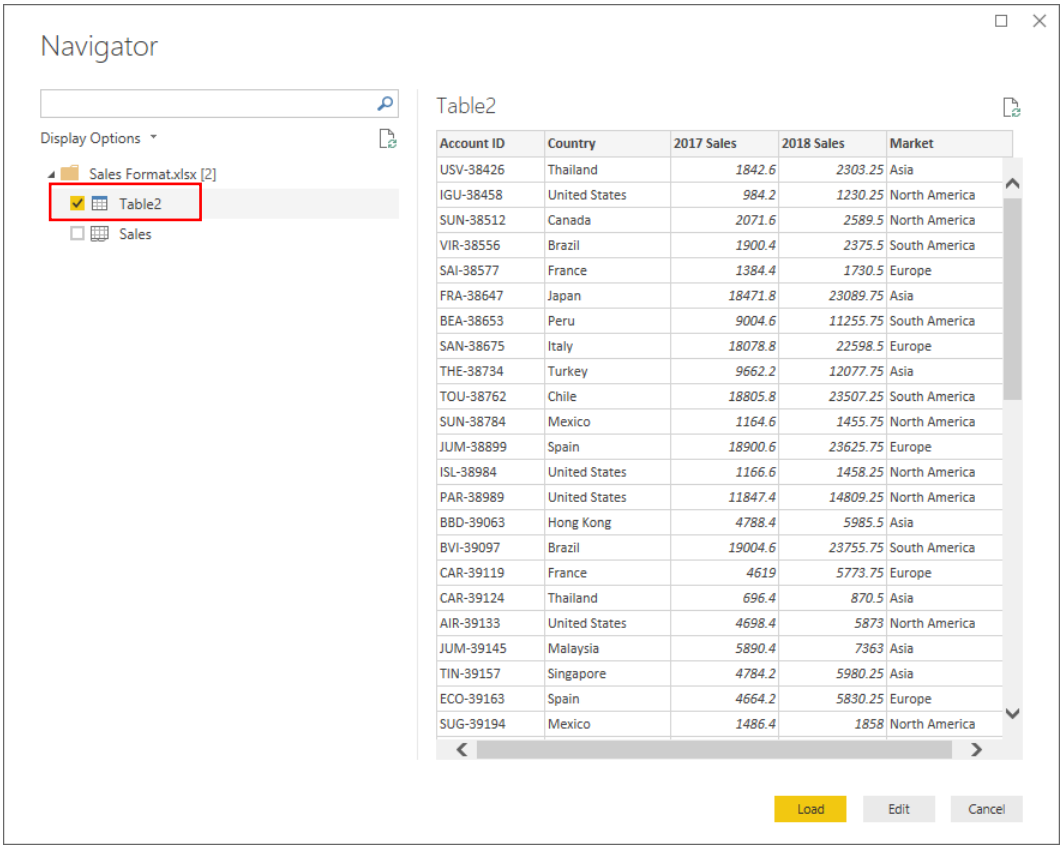
3. Select the option **Excel** in the **Get Data** menu.



The Get Data Menu

4. Locate and select the **Sales Format.xlsx** file and click **Open**.

5. The Navigator window is displayed. Check the **Table2** option under **Display Options**.

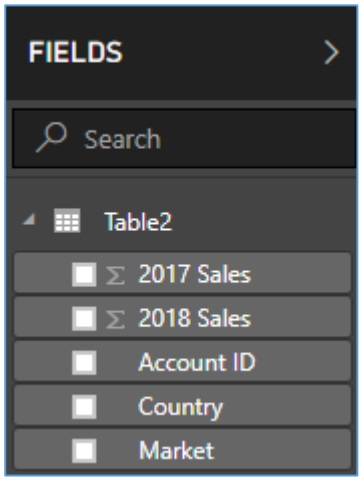


Account ID	Country	2017 Sales	2018 Sales	Market
USV-38426	Thailand	1842.6	2303.25	Asia
IGU-38458	United States	984.2	1230.25	North America
SUN-38512	Canada	2071.6	2589.5	North America
VIR-38556	Brazil	1900.4	2375.5	South America
SAI-38577	France	1384.4	1730.5	Europe
FRA-38647	Japan	18471.8	23089.75	Asia
BEA-38653	Peru	9004.6	11255.75	South America
SAN-38675	Italy	18078.8	22598.5	Europe
THE-38734	Turkey	9662.2	12077.75	Asia
TOU-38762	Chile	18805.8	23507.25	South America
SUN-38784	Mexico	1164.6	1455.75	North America
JUM-38899	Spain	18900.6	23625.75	Europe
ISL-38984	United States	1166.6	1458.25	North America
PAR-38989	United States	11847.4	14809.25	North America
BBD-39063	Hong Kong	4788.4	5985.5	Asia
BVI-39097	Brazil	19004.6	23755.75	South America
CAR-39119	France	4619	5773.75	Europe
CAR-39124	Thailand	696.4	870.5	Asia
AIR-39133	United States	4698.4	5873	North America
JUM-39145	Malaysia	5890.4	7363	Asia
TIN-39157	Singapore	4784.2	5980.25	Asia
ECO-39163	Spain	4664.2	5830.25	Europe
SUG-39194	Mexico	1486.4	1858	North America

The Navigator Window

6. Click **Load**.

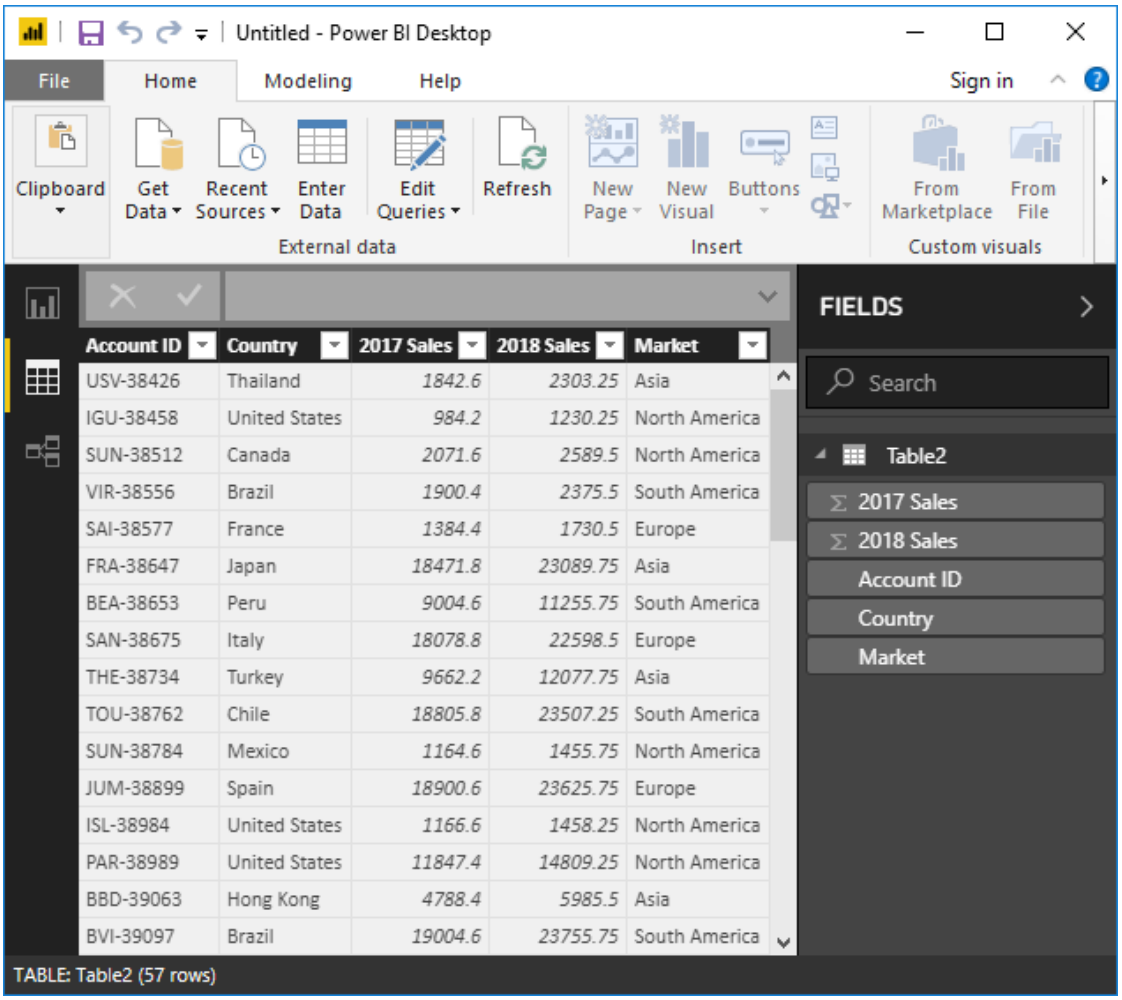
   In the report View, the **Fields** pane displays all the column headings in the data set.



The Fields Pane

7. Click the **Data View** icon on the left.

   In the Data View, the data set content is displayed.



Account ID	Country	2017 Sales	2018 Sales	Market
USV-38426	Thailand	1842.6	2303.25	Asia
IGU-38458	United States	984.2	1230.25	North America
SUN-38512	Canada	2071.6	2589.5	North America
VIR-38556	Brazil	1900.4	2375.5	South America
SAI-38577	France	1384.4	1730.5	Europe
FRA-38647	Japan	18471.8	23089.75	Asia
BEA-38653	Peru	9004.6	11255.75	South America
SAN-38675	Italy	18078.8	22598.5	Europe
THE-38734	Turkey	9662.2	12077.75	Asia
TOU-38762	Chile	18805.8	23507.25	South America
SUN-38784	Mexico	1164.6	1455.75	North America
JUM-38899	Spain	18900.6	23625.75	Europe
ISL-38984	United States	1166.6	1458.25	North America
PAR-38989	United States	11847.4	14809.25	North America
BBD-39063	Hong Kong	4788.4	5985.5	Asia
BVI-39097	Brazil	19004.6	23755.75	South America

The Data View

8. Click **File** and choose **Save**.

9. The **Save As** dialog box is displayed. Type **Report** in the **File name** box and click **Save**.

   The data will be saved as a **Power BI** file (**.pbix**).

10.5 REVIEW EXERCISE

1. Which of the following consists of one or more pages with one or more visualizations?
   a) Chart b) Summary
   c) Map d) Report

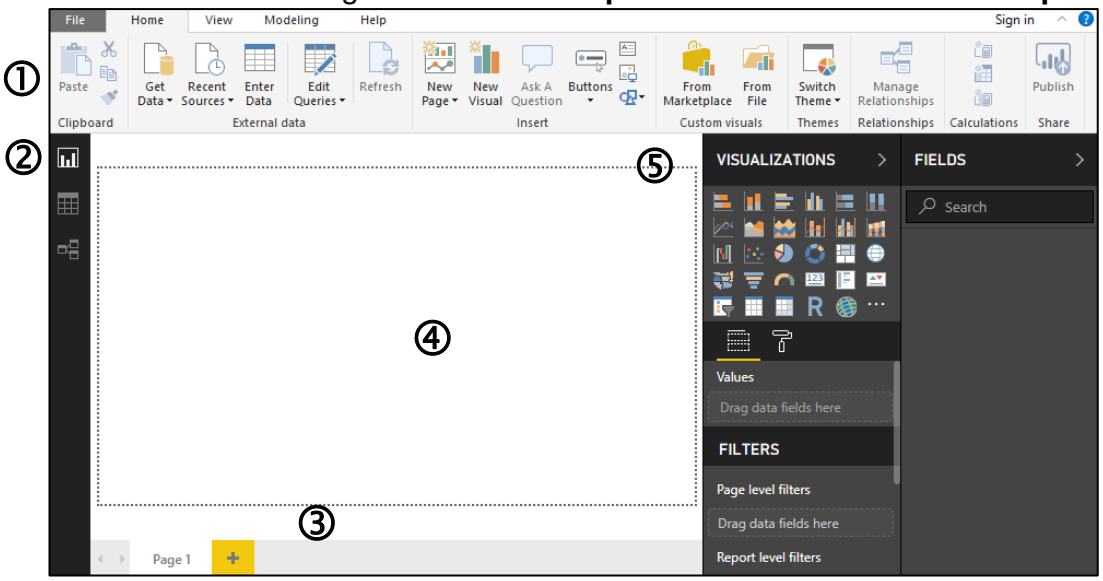
2. Which of the following consists of a single screen that provides key metrics at a glance?
   a) Dashboard b) Report
   c) KPI d) Summary

3. Which of the following is the main purpose for creating an analytical report?
   a) To create precision graphics for mechanical design.
   b) To solely represent data in rows and columns.
   c) To create, publish and share graphical insights.
   d) To store, organise and display data in a database format.

4. Which one of the following is a common type of data visualization used for measuring progress against predefined conditions?
   a) Table b) Map
   c) Line chart d) KPI

5. Which of the following is a key feature of data visualization tools?
   a) Presenting data in tabular format in spreadsheets.
   b) Publishing and sharing business intelligence.
   c) Presenting and sharing text-based reports.
   d) Sharing customer's contact details.

6. Which of the following accesses the **Report view** in **Power BI Desktop**?



7. Open **Power BI Desktop,** import the **Sales** data set in the **Sales** worksheet from the file **Sales overview.xlsx** and save it as **Sales overview imported.pbxi**.

LESSON 11 –
CREATING BASIC DATA VISUALIZATIONS

In this section, you will learn how to:

- Create table visualizations
- Create chart visualizations
- Enhance visualizations using conditional formatting
- Enhance visualizations using data bars
- Enhance visualizations using visual level filters
- Create visualizations using maps

11.1 CREATING TABLE VISUALIZATIONS

Concepts

Creating data visualizations based on a data set is one of the key functions of data visualization tools.

You can use visualizations to explore your data. And you can combine visualizations to create meaningful and insightful reports.

There are many types of visualizations to choose from. As part of your data analysis, you will determine which visualizations are appropriate for your needs.

One of the basic types of visualizations available is a **table**.

A table is useful for comparing detailed quantitative data. In a table, related data is laid out in a logical grid of columns and rows.

Steps

To create a **Table** visualization in a report:

1. Open **Power BI Desktop** and click **File** – **Open.**

2. Select the relevant **Power BI** file – in this case, select **TableChart.pbxi** and click **Open**.

3. In **Report** view, click on the **Table** icon in the **Visualizations** pane.

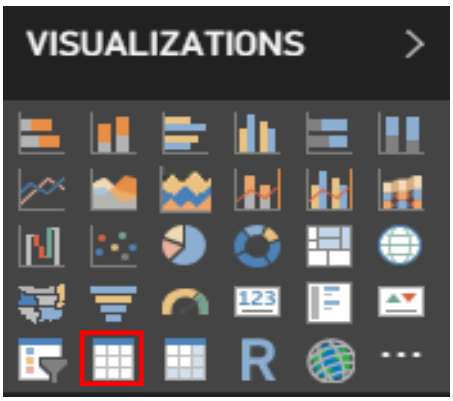


Table Visualizations Icon

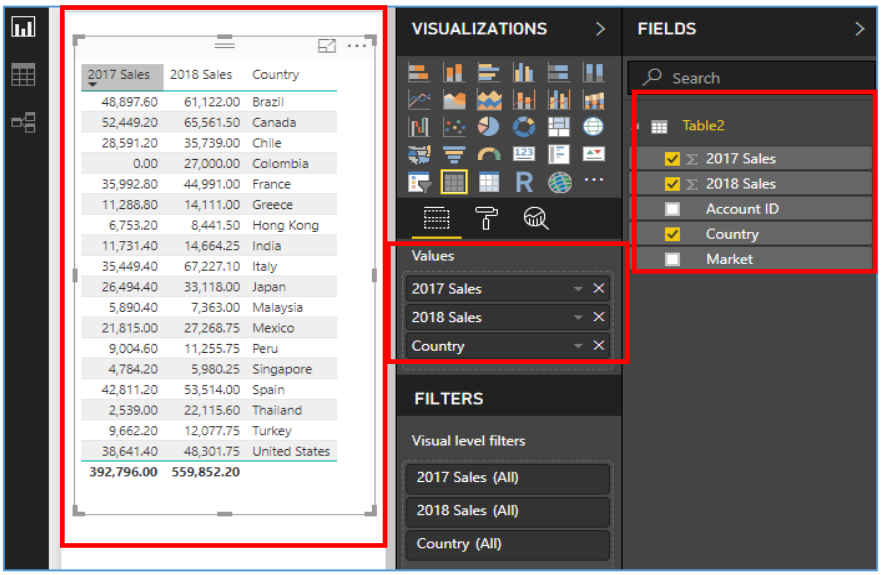
4. Select the fields to add to the table – in this case, select the **2017 Sales, 2018 Sales** and **Country** fields under **Table2** in the **Fields** pane.

   All the fields are automatically selected as **Values** under the **Visualizations** pane.

   ***Note**: Alternatively, drag and drop the fields to the **Values** area under **Visualizations** pane.*

A table comparing the sales performance between 2017 and 2018 is created.

***Note**: You may need to size the table to view all the information.*



Adding Fields for the Table

5. Select the table and click the **Format** icon  below the visualization icons in the **Visualizations** pane to show the formatting options for the table.

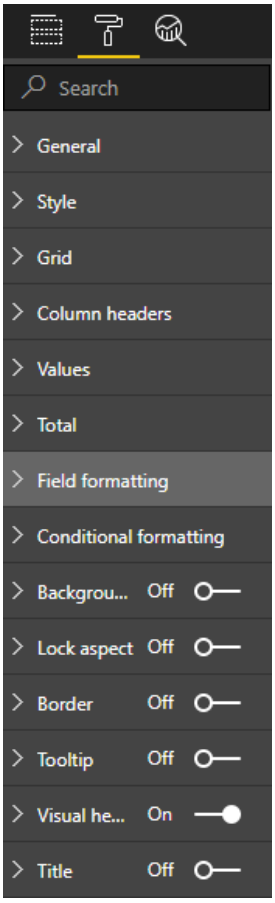
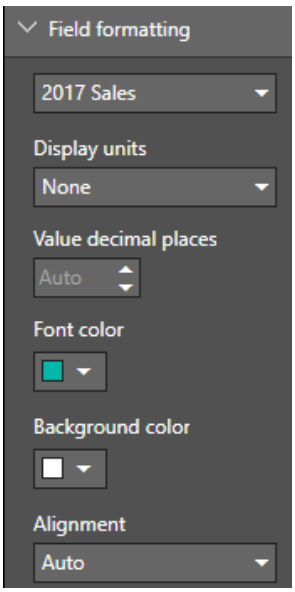


Table Formatting Options

6. Click **Field formatting** to expand the options and set the **font color** of the values in **2017 Sales** to **green** and its **background color** to **white**.



Field Formatting Options

7. Click the down arrow next to **2017 Sales** to change the field to **2018 Sales**. Set the **background color** of **2018 Sales** to **yellow**.

2017 Sales	2018 Sales	Country
48,897.60	61,122.00	Brazil
52,449.20	65,561.50	Canada
28,591.20	35,739.00	Chile
0.00	27,000.00	Colombia
35,992.80	44,991.00	France
11,288.80	14,111.00	Greece
6,753.20	8,441.50	Hong Kong
11,731.40	14,664.25	India
35,449.40	67,227.10	Italy
26,494.40	33,118.00	Japan
5,890.40	7,363.00	Malaysia
21,815.00	27,268.75	Mexico
9,004.60	11,255.75	Peru
4,784.20	5,980.25	Singapore
42,811.20	53,514.00	Spain
2,539.00	22,115.60	Thailand
9,662.20	12,077.75	Turkey
38,641.40	48,301.75	United States
392,796.00	**559,852.20**	

Formatted Sales Table

8. Save the file as **TableChart Added** Power BI file and close the file.

11.2 CREATING CHART VISUALIZATIONS

Concepts

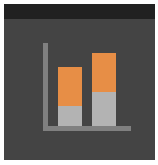
Charts are one of the most commonly used visualizations in data analytics. They reveal relationships such as similarities, differences, and trends among multiple values. Microsoft Power BI has a range of visualization options from basic charts to more advanced visuals. Some popular types of charts include:

- **Line chart**: This is the most popular chart type and is used to present a continuous data set. It is suited for trend-based visualizations of data over time, with the emphasis on the continuation or the flow of the values (a trend).

- **Pie/donut chart**: A pie chart typically represents numbers in percentages, used to visualize a part to whole relationship or a composition. Similarly, a **donut** shows the relationship of parts to a whole, allowing for space in the centre for a label or icon.

- **Bar/column chart**:
  - A bar chart is a **horizontal** column chart. It is best for multiple categories, especially those with lengthy labels or for displaying a data set with negative numbers.
  - A **column chart**, on the other hand, displays data in **vertical** columns.

To facilitate comparison of multiple variables across categories of data, **Power BI** includes **Stacked bar** and **Stacked column** charts, as well as **Clustered bar** and **Clustered column** charts.

			
Stacked bar (can be used to create simple bar chart)	**Stacked column** (can be used to create simple column chart)	**Clustered bar**	**Clustered column**

When deciding which chart type to use consider the following:

- The **purpose of the chart:** Is the purpose for comparison, composition, distribution or relationship?

- **The categories of data to be presented:** Are there single or multiple categories? Are they over a time period or among groups?

- **The data points to be presented:** How many are there for each category?

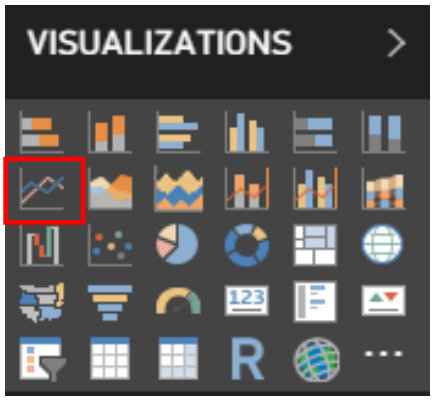
When you create your chart, you can then modify how it looks using the formatting options in **Power BI**.

Steps

To create a **Line chart** visualization - in this case, to plot the sales and profit values by month to visualise a trend:

1. Open the relevant **Power BI file** in **Power BI Desktop** – in this case, select **Sales Chart.pbxi**.

2. In **Report** view, select the **Line chart** icon in the **Visualizations** pane.

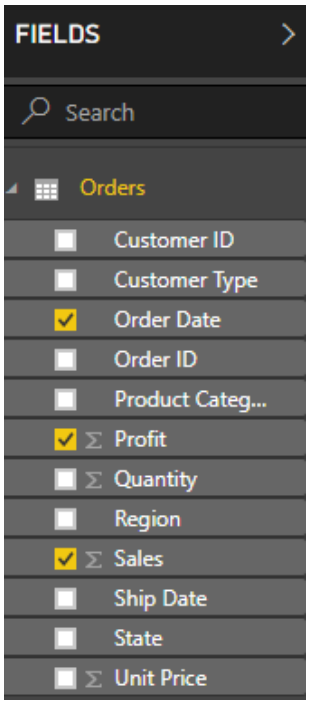


Line Chart Icon

3. Select the fields to add to the chart – in this case, select the **Order Date**, **Sales**, and **Profit** fields under the **Orders** table in the **Fields** pane.

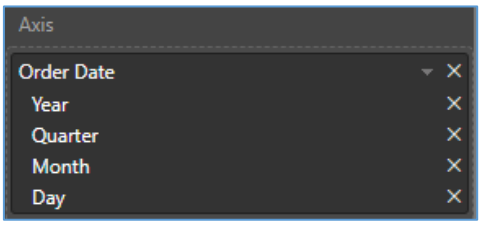
   ***Note**: Alternatively, drag and drop the fields to the respective options under **Visualizations** pane.*

   The **Order Date** field is automatically selected as **Axis**, with **Sales** and **Profit** automatically selected as **Values** under the **Visualizations** pane.



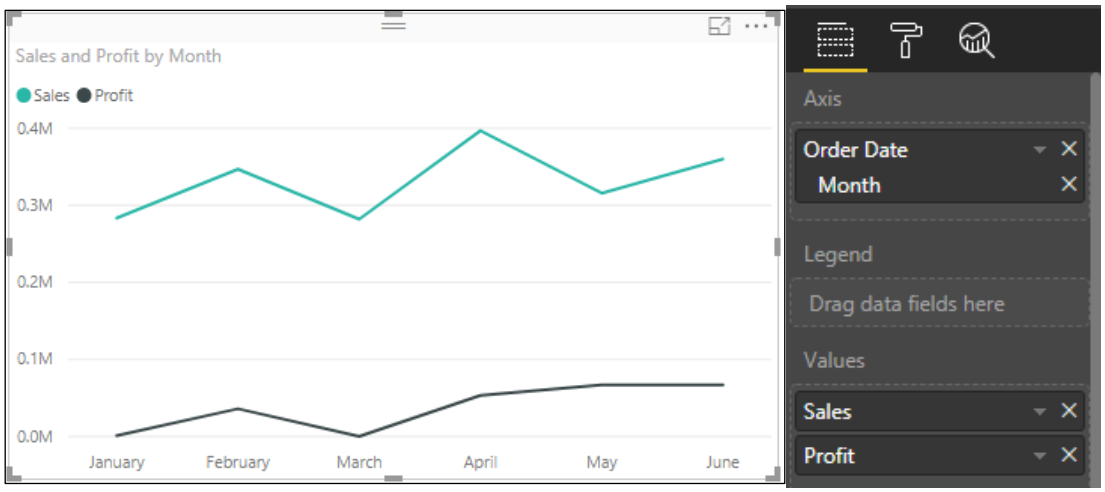
Fields displayed under the Fields pane

4. Select the time scale to display – in this case, to display the **Order Date** by **month**, deselect **Year**, **Quarter**, **Day** listed under **Order Date** by clicking the **x symbols**.



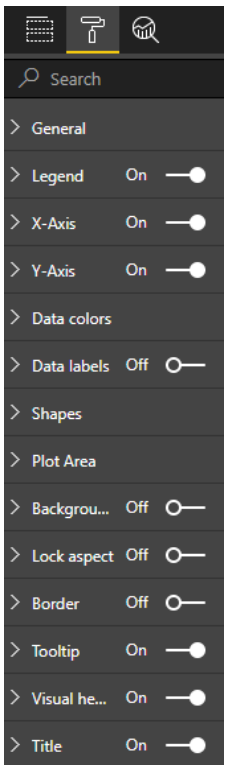
Order Date Options

5. Size the line chart to display all information on it clearly.



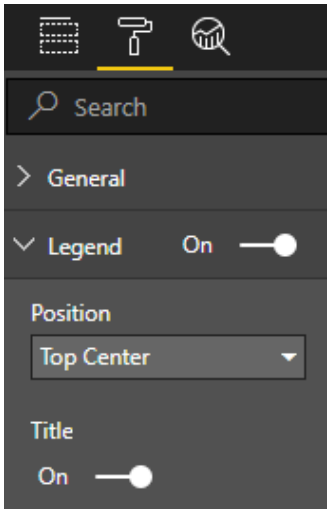
Line Chart for Sales and Profit by Month

6. Click the **Format** icon below the visualization icons in the **Visualizations** pane to show the formatting options for the line chart.



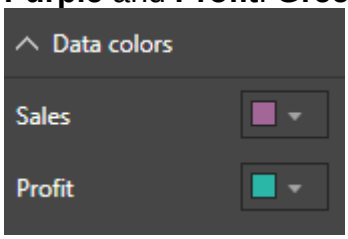
Line Chart Formatting Options

7. Click **Legend** to expand the options and set **Position** to **Top Center**.



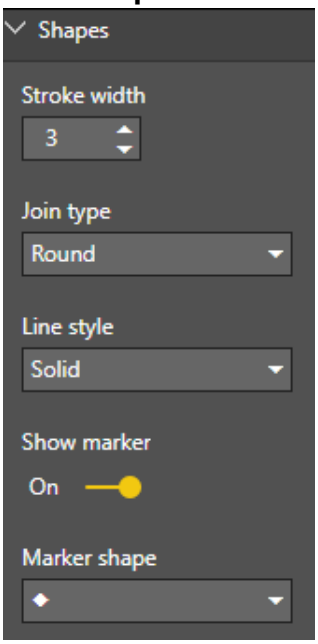
Legend Options

8. Expand the **Data colors** section and set the colours for the data series – in this case, set **Sales**: **Purple** and **Profit**: **Green**.



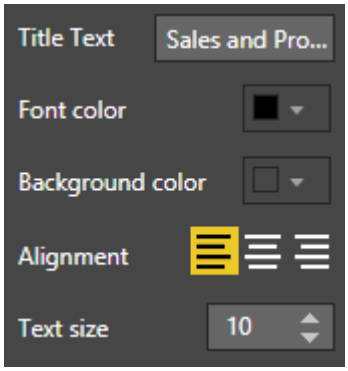
Data Colors formatting menu

9. Expand the **Shapes** section and set the following options: **Stroke width**: **3, Show marker**: **On,** and **Marker shape: Diamond**.



Shapes formatting menu

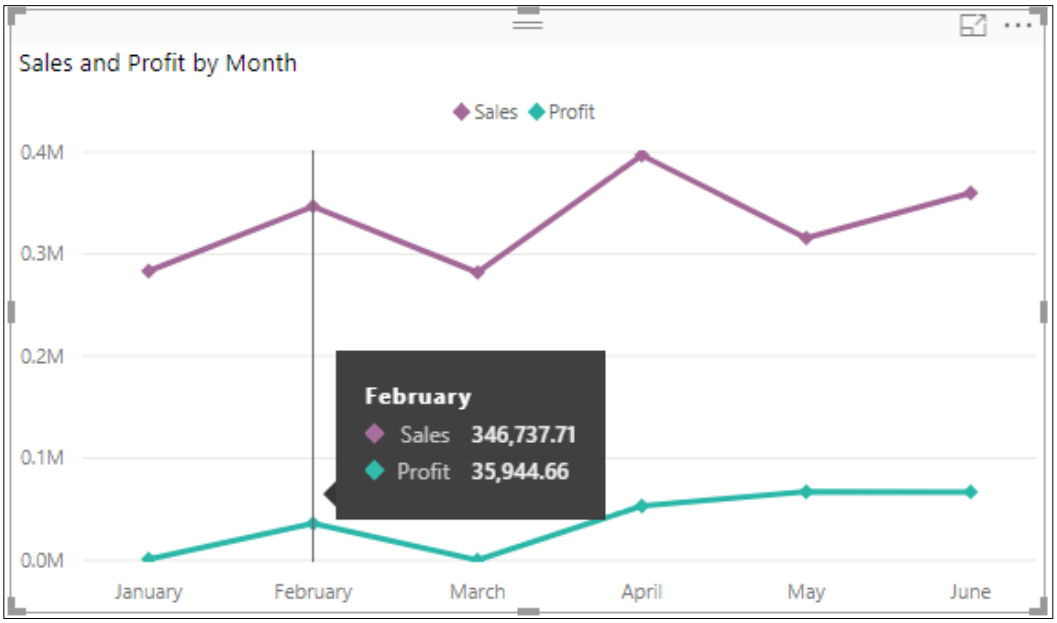
10. Expand the **Title** section and set the following options: **Font colour: Black** and **Text size: 10.**



Shapes formatting menu

11. The formatted line chart will be updated on the report canvas.

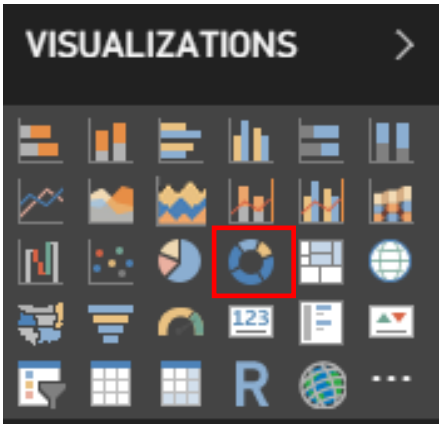
    Hover over a data point on the line chart to highlight the details and drop line.



Highlighting a Data Point

To create a **Donut chart** visualization - in this case, to show the sales distribution percentage for the various regions:

1. In **Report** view in **Power BI Desktop**, with **Sales Chart.pbxi** still open, click on an area outside the **Line Chart** in the **Report Canvas** area and select the **Donut chart** icon in the **Visualizations** pane.



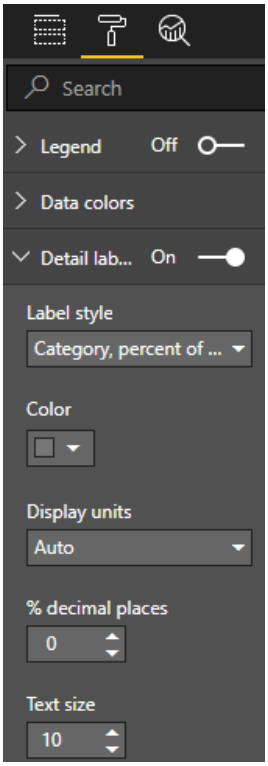
Donut Chart Icon

2. Select the fields to add to the chart – in this case, select the **Region** and **Sales** fields under the **Orders** table in the **Fields** pane.

   The **Region** field is automatically selected as **Legend**, with **Sales** automatically selected to appear as **Values** under the **Visualizations** pane.

3. Click the **Format** icon below the visualization icons in the **Visualizations** pane to display the formatting options for **Donut chart**.

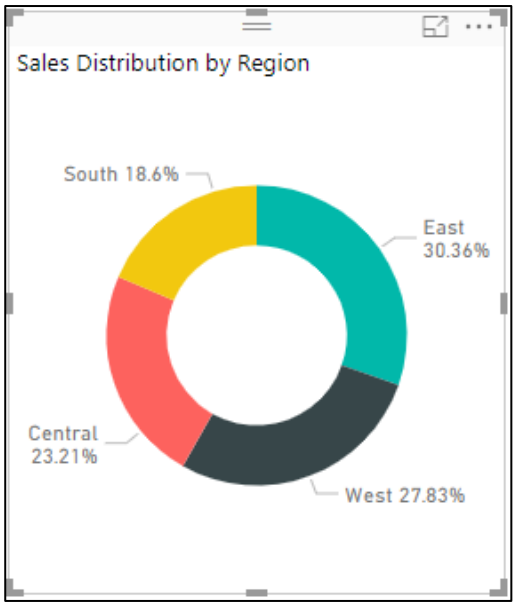
4. Expand the **Detail labels** section and set **Label style** to **Category, percent of total** and **Text size** to **10**.



Formatting Options for Donut chart

5. Expand the **Title** section and set the **Title Text** to **Sales Distribution by Region, Font colour** to **Black** and **Text size** to **10.**

6. Position the formatted donut chart to the right of the line chart and resize it as necessary.

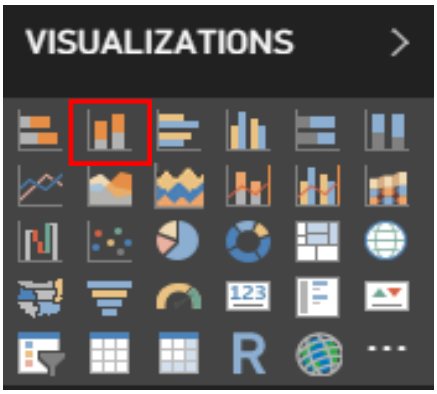


Formatted Donut Chart

Note: *To create a pie chart, select the* ***Pie chart*** *icon in the Visualizations pane. You can also change the donut chart to a pie chart by selecting the donut chart and then clicking the* ***Pie chart*** *icon in the Visualizations pane.*

To create a **Stacked column chart** visualization – in this case, to compare the sales and profit figures across the regions:

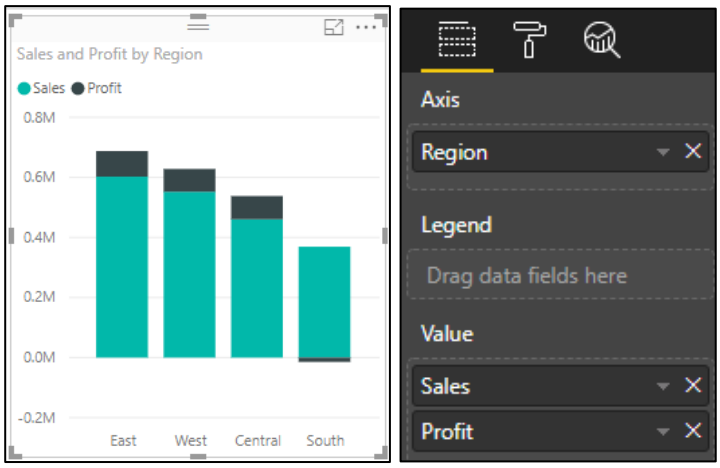
1. In **Report** view in **Power BI Desktop**, with **Sales Chart.pbxi** still open, click on an area outside the charts in the **Report Canvas** area and select the **Stacked column chart** icon in the **Visualizations** pane.



Stacked Column Chart Icon

2. Select the fields to add to the chart – in this case, select the **Region, Sales** and **Profit** fields under the **Orders** table in the **Fields** pane.

   The **Region** field is automatically selected to appear under **Axis**, with **Sales** and **Profit** automatically selected to appear under **Value** under the **Visualizations** pane.



Stacked Column Chart for Sales and Profit by Region

3. To format the chart, click the **Format** icon below the visualization icons in the **Visualizations** pane.

4. Expand the **Legend** section and set the **Position** to **Top Center** and **Text size** to **10.**

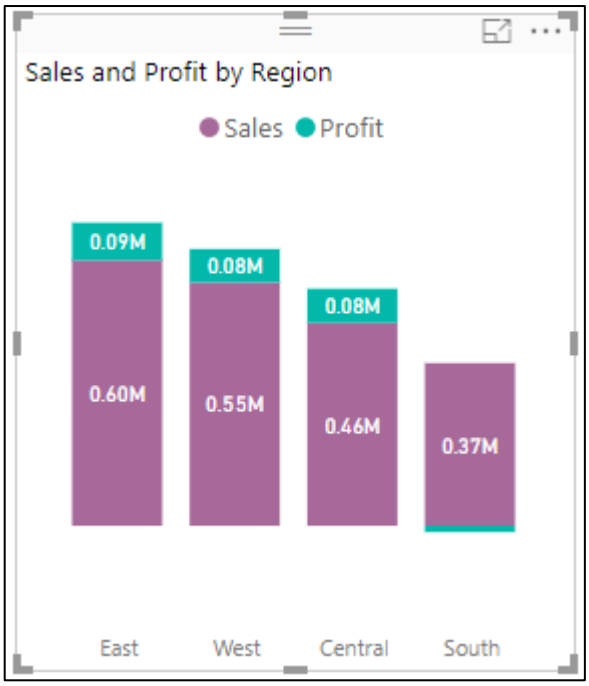
5. Slide the **Y-Axis** option to **Off** to turn off the y-axis display.

6. Expand the Data colours section and modify the series colour **Sales** to **Purple** and **Profit** to **Green.**

7. Slide the **Data labels** option to **On** to turn on the data labels.

8. Expand the **Title** section, set the **Font colour** to **Black** and **Text size** to **10**.

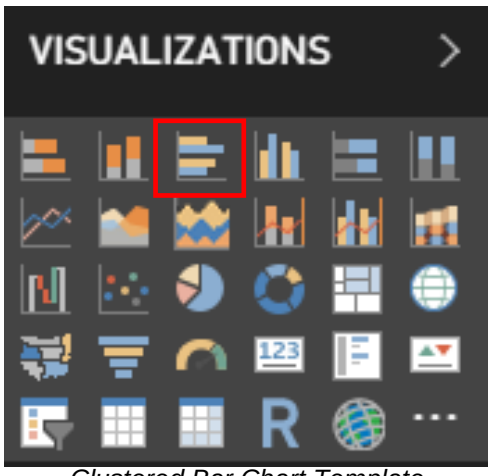
9. Position the stacked column chart below the line chart and resize as necessary.



Formatted Stacked Column Chart

To create a **clustered bar chart** – in this case, to compare sales figures across four customer types:

1. In **Report** view in **Power BI Desktop**, with **Sales Chart.pbxi** still open click on an area outside the stacked column chart in the **Report Canvas** area and select the **Clustered bar chart** icon in the **Visualizations** pane.

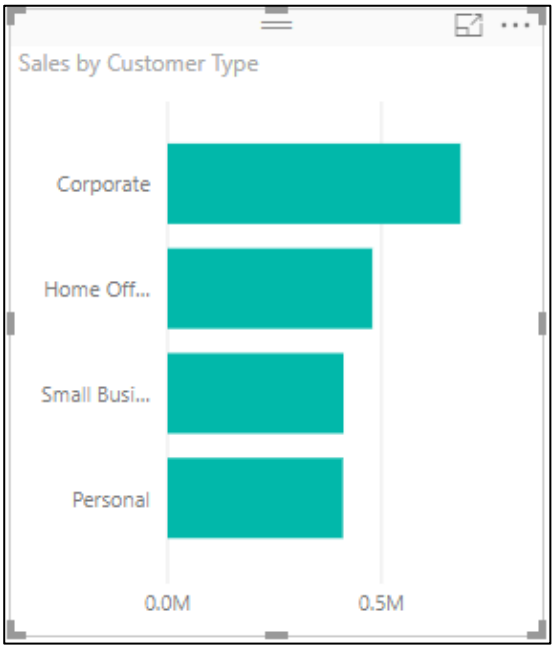


Clustered Bar Chart Template

2. Add fields to the chart – in this case, select the **Customer Type** and **Sales** fields under the **Orders** table in the **Fields** pane.

   The **Customer Type** field is automatically selected as **Axis**, with **Sales** automatically selected as **Value** under the **Visualizations** pane.

   A **Clustered bar** chart displaying **Sales by Customer Type** is created.



Clustered Bar Chart Visualization

3. Click the **Format** icon below the visualization icons in the **Visualizations** pane.

4. Slide the **X-Axis** option to **Off**.

5. Expand the **Data colours** section and set the data colour for **Corporate** to **Purple**.

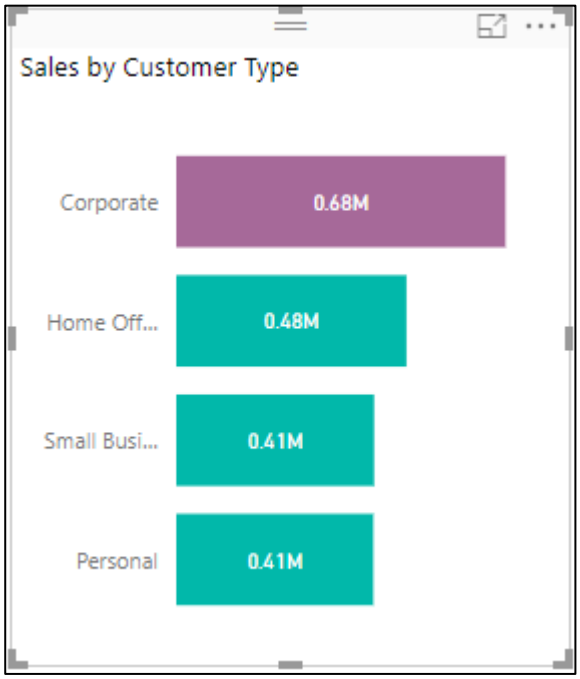
6. Slide the **Data labels** option to **On**.

7. Expand the **Data labels** section and set **Position** to **Inside Center**.

8. Expand the **Title** section and set the following options **Font colour to Black and Text size to 10.**

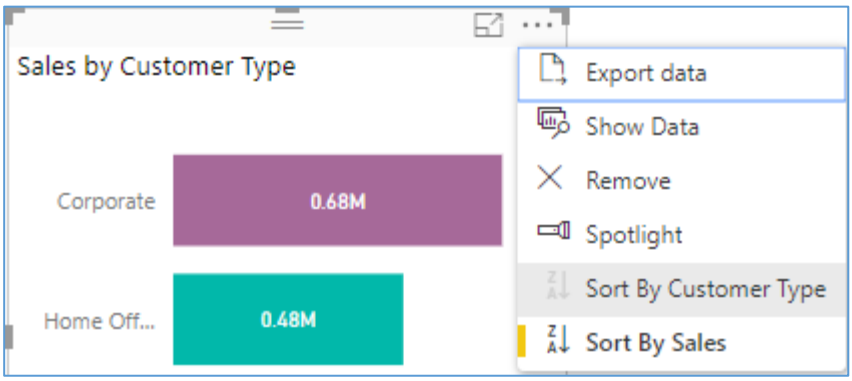
9. Position the clustered bar chart to the right of the stacked column chart and resize as necessary.

   The data series are by default sorted by the values in descending order.



Formatted Clustered Bar Chart

10. To sort the data series by customer type, select the clustered bar chart visualization, click **More options** (the ellipses) in the top right of the visualization and select **Sort by Customer Type**.

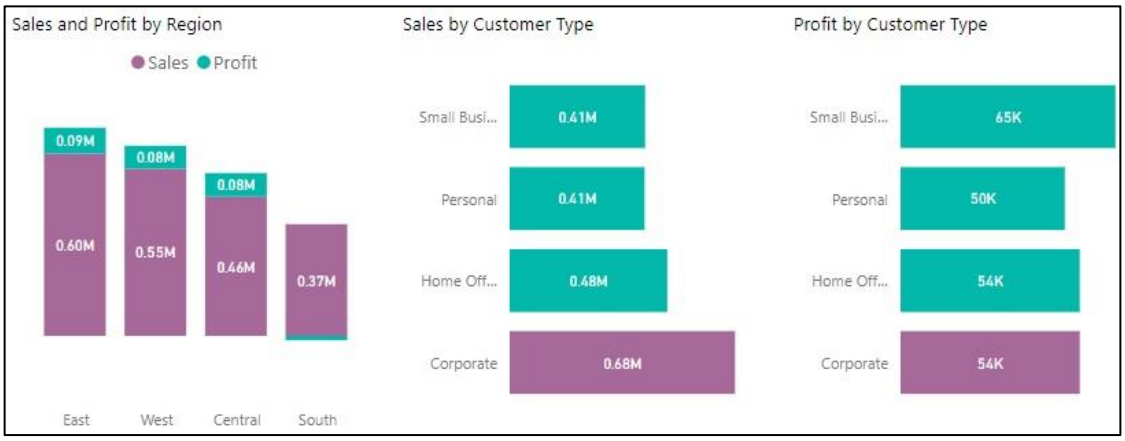


Changing the Sort Order for the Data Series

11. Create a copy of the clustered bar chart visualization and position it to the right of the copied clustered bar chart.

12. Uncheck **Sales** and check **Profit** to change the field under **Value** to **Profit**. A **Clustered bar** chart displaying **Profit by Customer Type** is created.

13. Align the three chart visualizations in a row with equal spacing in between.



Aligned Chart Visualizations

14. Save the file as **SalesChart Added** Power BI file and close the file.

11.3 ENHANCING VISUALIZATIONS USING CONDITIONAL FORMATTING

Concepts

After the creation of visualizations, you may want to enhance the visuals to note key summary features and highlight important statistical characteristics. For example, you may want to show **maximum** and **minimum** values or **above average** and **below average** values.

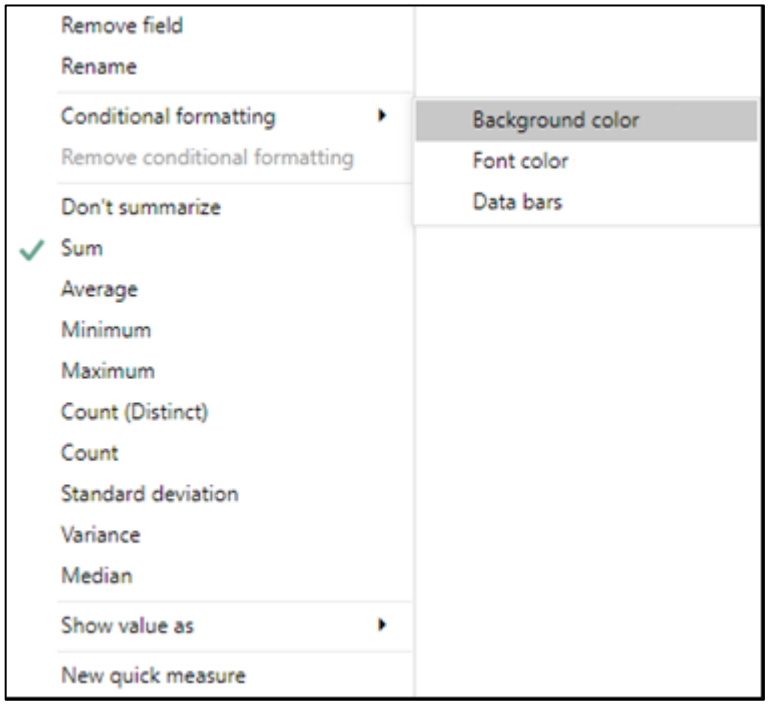
In **Power BI**, visualizations can be enhanced using **Conditional formatting**.

Some conditional formatting options, for example, for a table visualization, include setting the **background color** and **font color** for cells based on cell values or rules and showing cell values with **data bars**.

Steps

To set conditional formatting by **Background colour** on values in a table visualization:

1. Open the **Sales Chart Enhanced** Power BI file in **Power BI Desktop**.
2. In **Report** view, click on the **table** visualization to select it.
3. Go to **2017 Sales** within the **Values** area in the **Visualizations** pane and click the field's down arrow.
4. Select **Conditional Formatting** and click **Background color** from the contextual menu.



Selecting the Conditional Formatting Type

5. Specify colours for the lowest value and the highest value. In this case, set the colour yellow in the **Minimum** area and the colour purple in the **Maximum** area

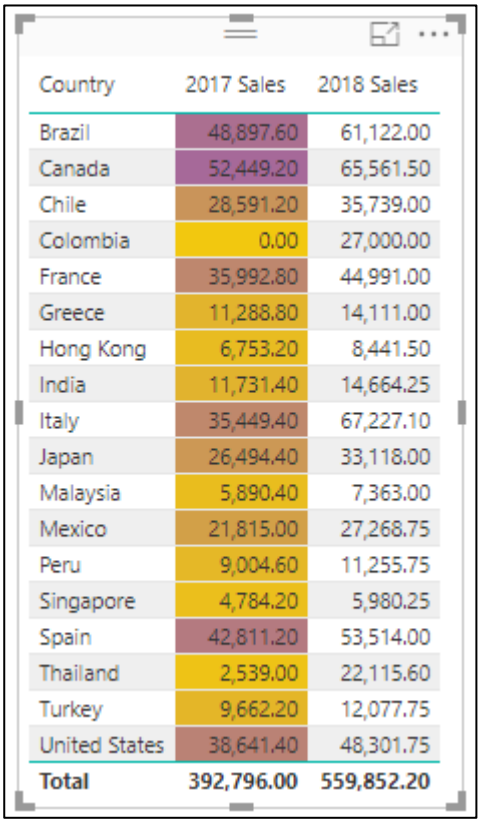
Specifying the Background Colours for the High/Low Values

***Note**: If you select the **Diverging** box, you can configure an optional **Center** value as well.*

*Depending on the Power BI Desktop version, you can also specify colours for specific value ranges by choosing the **Rules Format by** option. In the **Rules** area, set one or more value ranges and related colours by entering the relevant **If value condition**, the relevant **and value condition** and a related **colour**. Table cells with values in the specified ranges will be filled with the specified colour.*

6. Click **OK**.

7. Resize the table to ensure all the values are visible.



Country	2017 Sales	2018 Sales
Brazil	48,897.60	61,122.00
Canada	52,449.20	65,561.50
Chile	28,591.20	35,739.00
Colombia	0.00	27,000.00
France	35,992.80	44,991.00
Greece	11,288.80	14,111.00
Hong Kong	6,753.20	8,441.50
India	11,731.40	14,664.25
Italy	35,449.40	67,227.10
Japan	26,494.40	33,118.00
Malaysia	5,890.40	7,363.00
Mexico	21,815.00	27,268.75
Peru	9,004.60	11,255.75
Singapore	4,784.20	5,980.25
Spain	42,811.20	53,514.00
Thailand	2,539.00	22,115.60
Turkey	9,662.20	12,077.75
United States	38,641.40	48,301.75
Total	**392,796.00**	**559,852.20**

Table with High/Low Background Formatting

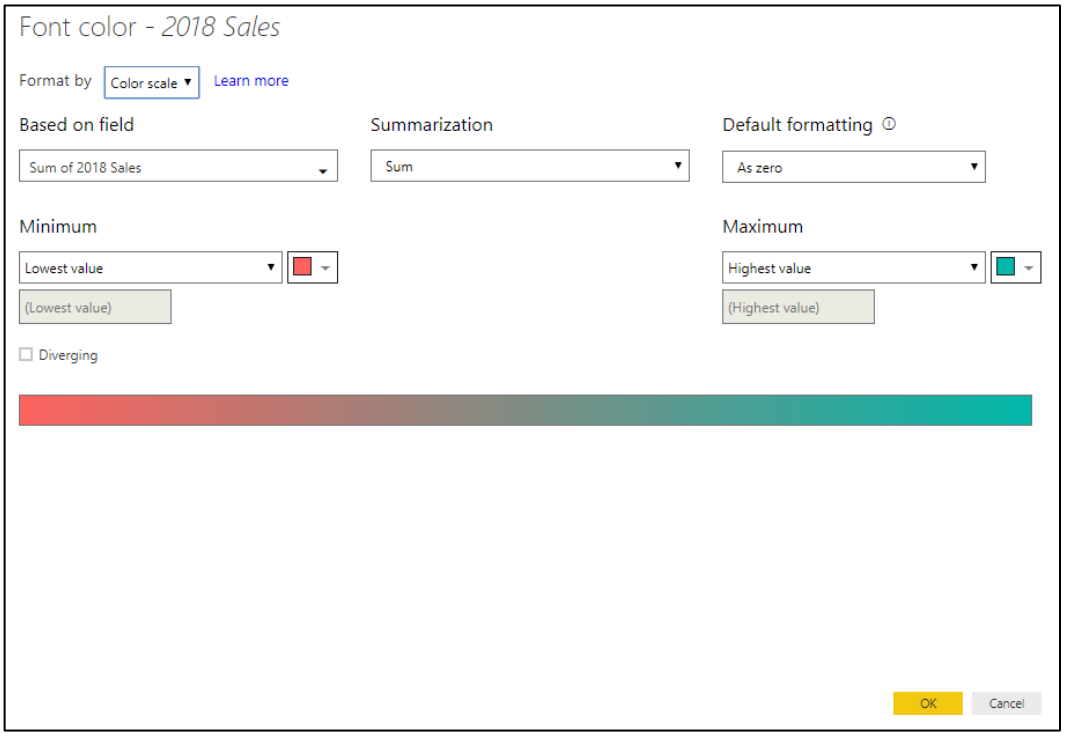
8. To edit the applied conditional formatting, select the field's down arrow, select **Conditional Formatting** and click **Background color** from the contextual menu and edit as required.

9. To remove the conditional formatting, select the field's down arrow, select **Remove Conditional Formatting**. Select **All** or select the specific formatting type to remove.

To set conditional formatting by **Font color** on values in a table visualization:

1. Go to **2018 Sales** within the **Values** area in the **Visualizations** pane, click the field's down arrow and select **Conditional Formatting**.

2. Click **Font color** from the contextual menu.



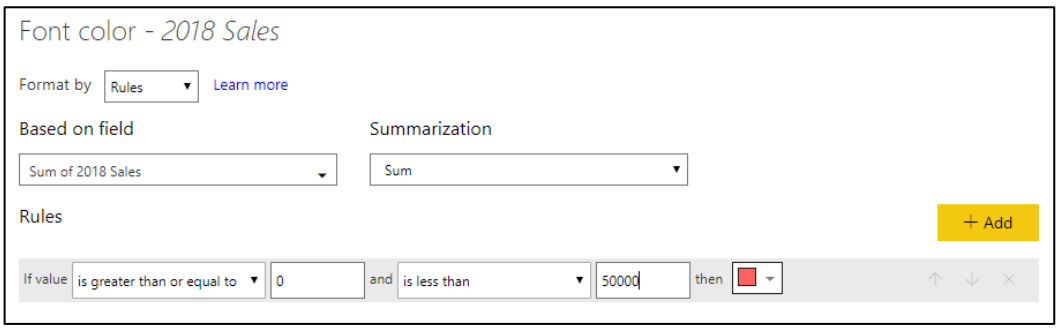
The Font Color Window

3. Click the down arrow next to **Color scale** and set the **Format by** to **Rules** option.

   ***Note**: You can use the **Color scale** option to specify colours for the minimum value and the maximum value.*

4. When you select the **Format by Rules** option, you can specify one or more value ranges, each with a set colour. In each value range, set an **If value condition**, an **and value condition**, and a **colour**.

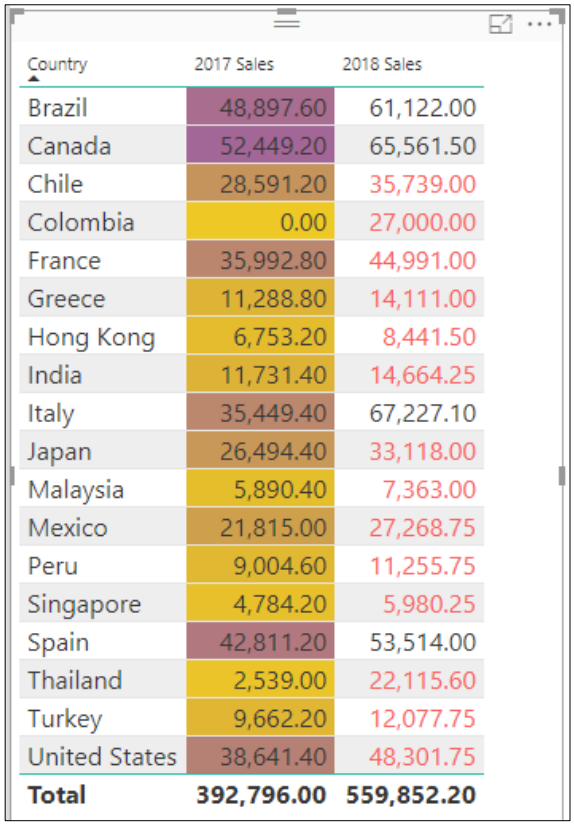
   In this case, set the **If value** to **is greater than or equal** to **0** and **is less than 50000** in the **Rules** area:



Specifying Rules

   ***Note**: You can add more rules by clicking the **Add** button.*

5. Confirm the rules by clicking **OK**.



Country	2017 Sales	2018 Sales
Brazil	48,897.60	61,122.00
Canada	52,449.20	65,561.50
Chile	28,591.20	35,739.00
Colombia	0.00	27,000.00
France	35,992.80	44,991.00
Greece	11,288.80	14,111.00
Hong Kong	6,753.20	8,441.50
India	11,731.40	14,664.25
Italy	35,449.40	67,227.10
Japan	26,494.40	33,118.00
Malaysia	5,890.40	7,363.00
Mexico	21,815.00	27,268.75
Peru	9,004.60	11,255.75
Singapore	4,784.20	5,980.25
Spain	42,811.20	53,514.00
Thailand	2,539.00	22,115.60
Turkey	9,662.20	12,077.75
United States	38,641.40	48,301.75
Total	**392,796.00**	**559,852.20**

Table with Formatting by Font colours

Values are highlighted in the table to bring attention to the **range** and **variation** within the data set.

6. To edit the applied conditional formatting, select the field's down arrow, select **Conditional Formatting** and click **Font color** from the contextual menu and edit as required.

7. To remove the conditional formatting, select the field's down arrow, select **Remove Conditional Formatting**. Select **All** or select **Font color**.

To apply and edit **Data bars** on values in a table visualization:

1. In **Report** view in **Power BI Desktop**, with **Sales Chart Enhanced** still open, click on an area outside the existing table in the report canvas. Select the **Table** icon in the **Visualizations** pane to create a new table.

2. Select the fields to add to the table – in this case, select the **Market** and **2018 Sales** fields under the **Table2** table in the **Fields** pane.

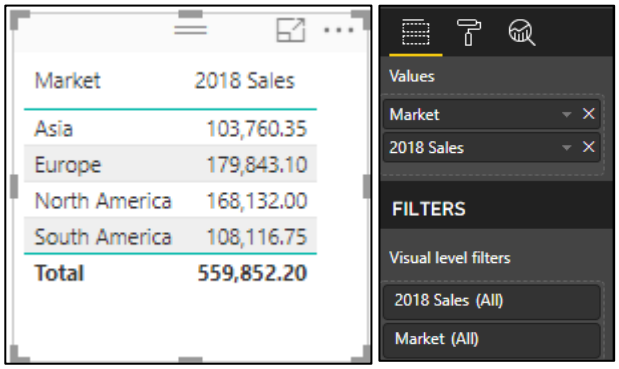
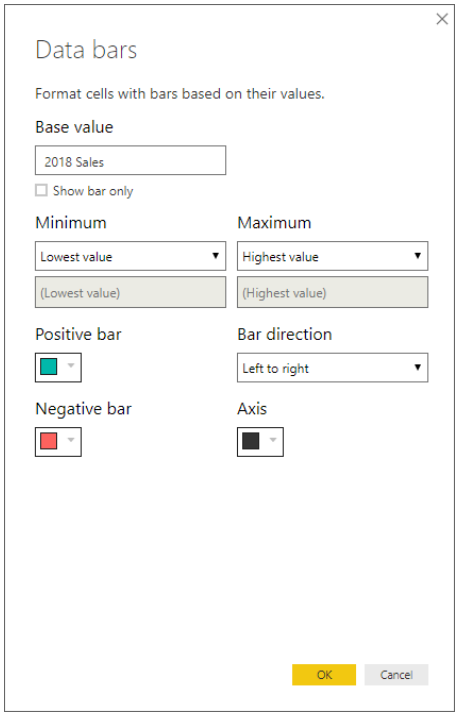


Table Visualizations for Market and 2018 Sales

3. Go to **2018 Sales** within the **Values** area in the **Visualizations** pane, click the field's down arrow and select **Conditional Formatting**.

4. Click **Data bars** from the contextual menu.

   The **Data bars** window is displayed.

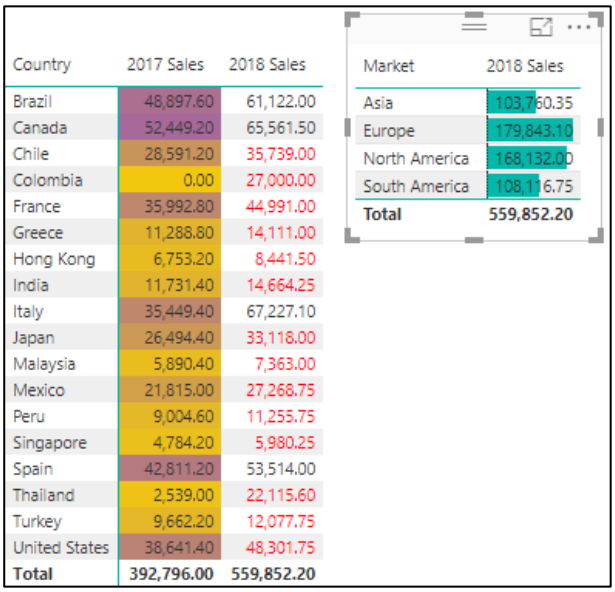


The Data Bars Window

5. The default range for the data bars is the **Lowest value** in the data set for **Minimum** and **Highest value** for **Maximum**. To specify a set value, select the **Minimum** type down arrow and select **Number**. Type **0** in the **Value** box.

6. Click the **Positive bar** down arrow and select the required colour (for example, green) and click **OK**.

7. Place the table visualization to the right of the first table and resize if necessary.



Country	2017 Sales	2018 Sales
Brazil	48,897.60	61,122.00
Canada	52,449.20	65,561.50
Chile	28,591.20	35,739.00
Colombia	0.00	27,000.00
France	35,992.80	44,991.00
Greece	11,288.80	14,111.00
Hong Kong	6,753.20	8,441.50
India	11,731.40	14,664.25
Italy	35,449.40	67,227.10
Japan	26,494.40	33,118.00
Malaysia	5,890.40	7,363.00
Mexico	21,815.00	27,268.75
Peru	9,004.60	11,255.75
Singapore	4,784.20	5,980.25
Spain	42,811.20	53,514.00
Thailand	2,539.00	22,115.60
Turkey	9,662.20	12,077.75
United States	38,641.40	48,301.75
Total	**392,796.00**	**559,852.20**

Market	2018 Sales
Asia	103,760.35
Europe	179,843.10
North America	168,132.00
South America	108,116.75
Total	**559,852.20**

Table with Data Bars

The data bars enhanced the table visualization to highlight the market share of 2018 sales.

Note: *To edit data bars, select the down arrow next to the relevant field, select* ***Conditional Formatting*** *and click* ***Data bars*** *and edit as required.*

To remove data bars, select the down arrow next to the relevant field, select ***Remove Conditional Formatting*** *and select* ***Data bars****.*

9. Save the file as **Sales Chart Formatted** Power BI file and close the file.

11.4 Enhancing Visualizations Using Visual Level Filters

Concepts

To visually inspect specific sections of a data set, you can apply filters to visualizations so that the visualizations only display the required information.

In **Power BI**, you can specify the items to display within a given visualization using **Visual Level Filters**.

Steps

To apply visual level filters on a table visualization:

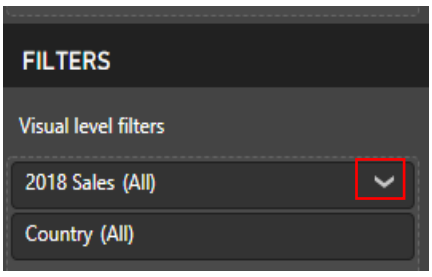
1. Open **Visual level filters.pbxi** in **Power BI Desktop**.

2. In **Report** view, click on the **table** visualization to view the **2018 Sales** listed by **Country**.

2018 Sales	Country
61,122.00	Brazil
65,561.50	Canada
35,739.00	Chile
27,000.00	Colombia
44,991.00	France
14,111.00	Greece
8,441.50	Hong Kong
14,664.25	India
67,227.10	Italy
33,118.00	Japan
7,363.00	Malaysia
27,268.75	Mexico
559,852.20	

2018 Sales Table Visualization

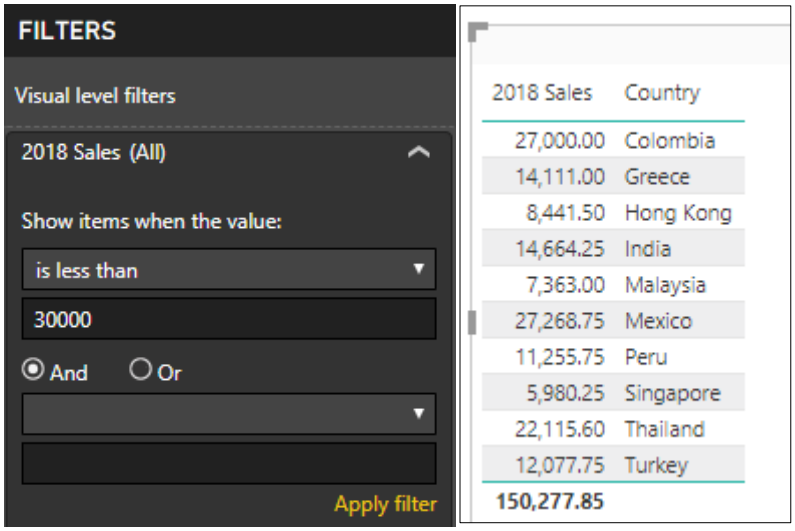
3. Under **Visual level filters** in the **Filters** pane, the value fields **2018 Sales** and **Country** are automatically included. Click on the down arrow next to **2018 Sales** under **Visual level filters** to show the visual level filter options for the **Sales 2018** field.



Visual level filter fields

For numerical values, the visual level filter options allow for display of values by setting a rule (for example, is less than or equal to) and specifying a value.

4. To show only sales volume of less than 30000 in the table visualisation, enter the value under **Show items when the value** and the option **is less than** and press **Return**.

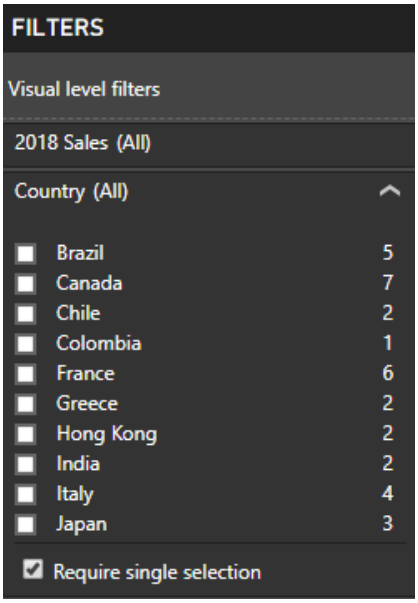


Setting a value to display selected items for 2018 Sales

5. Click on the down arrow next to **Country** under **Visual level filters** to show the visual level filter options for the **Country** field.

   For qualitative values, the items to display can be selected in a list displayed under Basic filtering.

6. To display only 2018 sales for one country at each application of the visual level filter, select **Require single selection**. This allows the sales value only for a selected country to be shown.

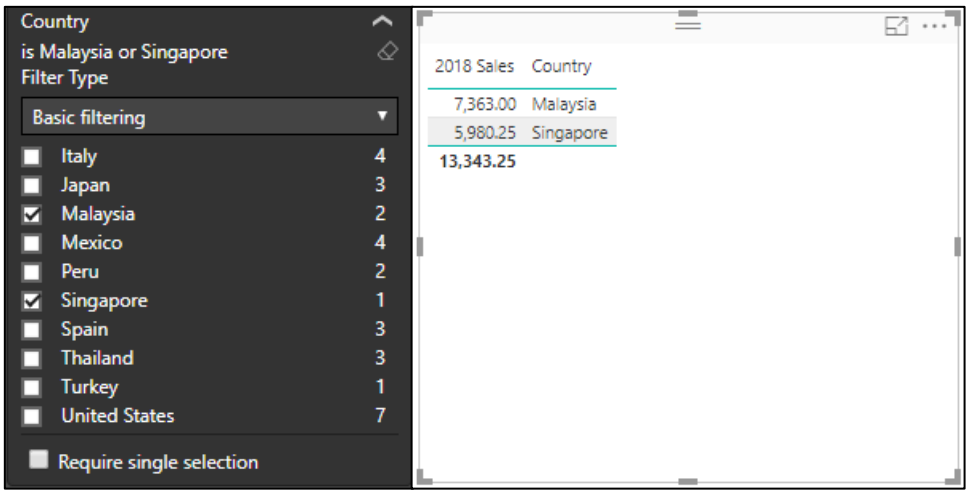


Basic Filtering to display single selection for a field

7. To show multiple selected countries in the table visualisation, uncheck the option **Require single selection**.

8. Scroll down the list and select the countries – in this case, select **Singapore and Malaysia.**

   The change in display for the visualisation is activated.



Basic Filtering to display selected countries

9. Save the file as **Visual Levels Added** Power BI file and close it.

   The setting for the **Visual level filters** is saved with the file.

11.5 CREATING VISUALIZATIONS USING MAPS

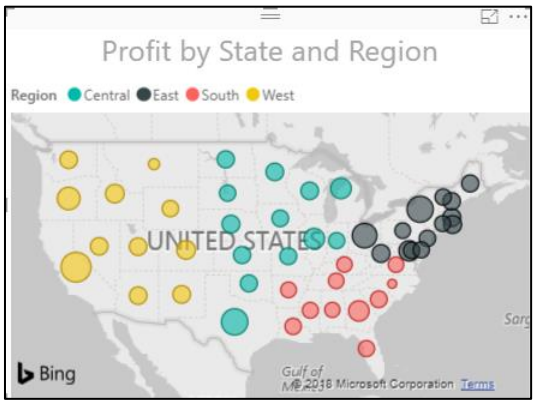
Concepts

Visualization tools can display quantitative information on a map to give an overview of data distribution across geographic locations. Data on a map is often displayed as points of colour, various shades or sized bubbles.

By providing a geographical context for numbers, quantitative maps can provide an insight into data sets by visually highlighting the best and worst performing areas, trends, and outliers.

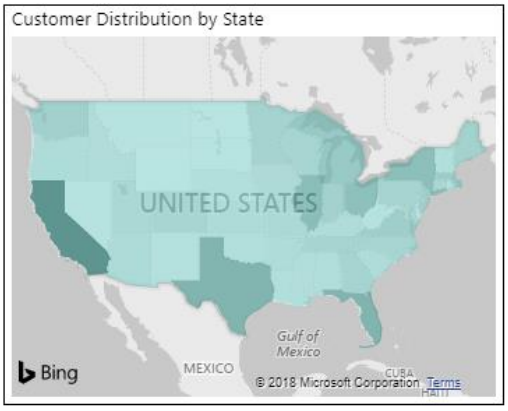
Two commonly used map types in Power BI are:

- **The Map (Bubble map) visualization**: A bubble is placed over a geographic point and the size of the bubble corresponds to the size of data. For example, the **Map** below shows the profit volume by state and region in the United States.



Map (Bubble Map) in Power BI

- **The Filled map (choropleth) visualization:** Shading is used to display how values differ in proportion across geographical locations. These relative differences are displayed with shading that ranges from light (less-frequent/lower values) to dark (more-frequent/higher values). This gives an overview of the distribution across geographic locations. For example, the **Filled Map** below shows customer distribution by state in the United States.



Filled Map (choropleth) in Power BI

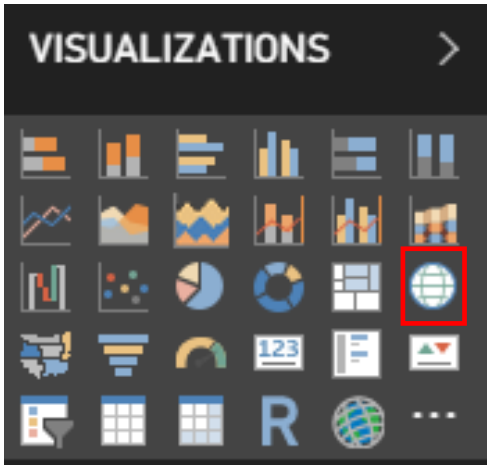
Power BI is integrated with **Bing** to provide map coordinates based on the location field in the data set.

Steps

To create a visualization using a **Map** – in this case a map chart of sales volume by state:

1. Open the relevant **Power BI** file in **Power BI Desktop** – in this case **MapChart.pbix**.

2. In **Report** view, select the **Map** icon in the **Visualizations** pane.



Map Template

3. Select the fields to add to the map – in this case, select the **State** and **Sales** fields under the **Orders** table in the **Fields** pane.

   The **State** field is automatically selected as **Location**, with **Sales** automatically selected as **Size** under the **Visualizations** pane.

   A map visualization is created.

Map Visualization with State as location and Sales as Size

4. Click the **Format** icon below the visualization icons in the **Visualizations** pane.

5. Expand the **Map styles** section and change the **Theme** to **Grayscale**.

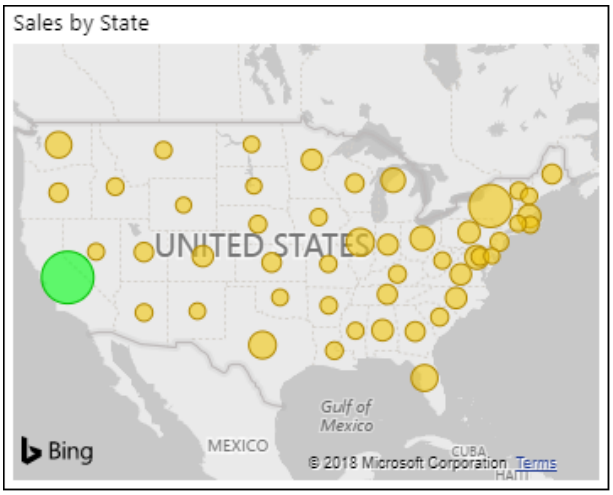
6. Slide the **Border** option to **On**.

7. Expand the **Title** section and set the **Font colour** to **Black** and **Text size** to **10**.

8. Expand the **Data colours** section and change the **Default colour** to **Yellow**.

9. To highlight the sales for a state (for example **California**) on the map, slide the **Show all** option to **On** and change the colour for **California** to **Green**.

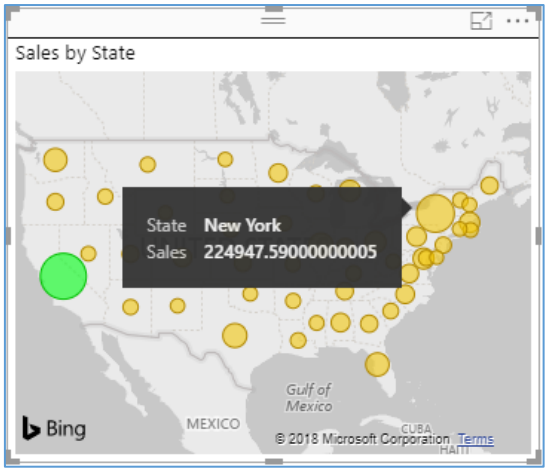
10. Expand the **Bubbles** section and set the **Size** to **5%**.



A Map chart of Sales by State and Sales highlighted for California

The map now shows the sales volume by state in proportion to the size of its bubble with a special highlight for California to facilitate quick comparison of sales volumes across the states.

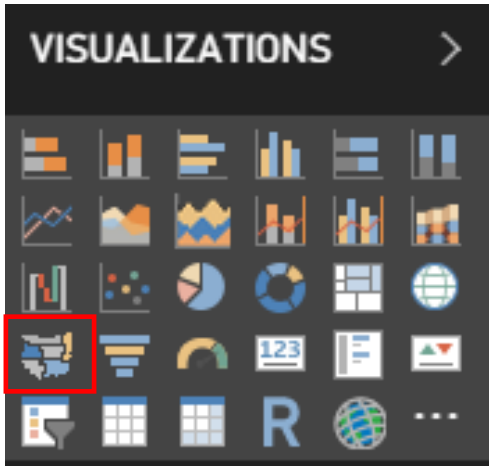
11. Hover over a state to highlight the plotted value.



Highlighting a Bubble on the Map

To create a visualization using a **Filled Map** - in this case a **Filled Map** of customer distribution by state:

1. In **Report** view in **Power BI Desktop**, click on an area outside the newly created bubble map in the **Report Canvas** area and select the **Filled map** icon in the Visualizations pane.



Filled Map Template

2. Select the fields to add to the map – in this case, select **State** and **Customer ID** fields under **Orders** table in the **Fields** pane.

   The **State** field is automatically selected as **Location**, with **Customer ID** automatically added to the Legend area under the **Visualizations** pane.

3. Drag and drop **Customer ID** to the **Color saturation** area.

4. Click the **Count of Customer ID** field's drop down list in the **Color saturation** area and select the function **Count (Distinct)**.

Filled Map Visualization

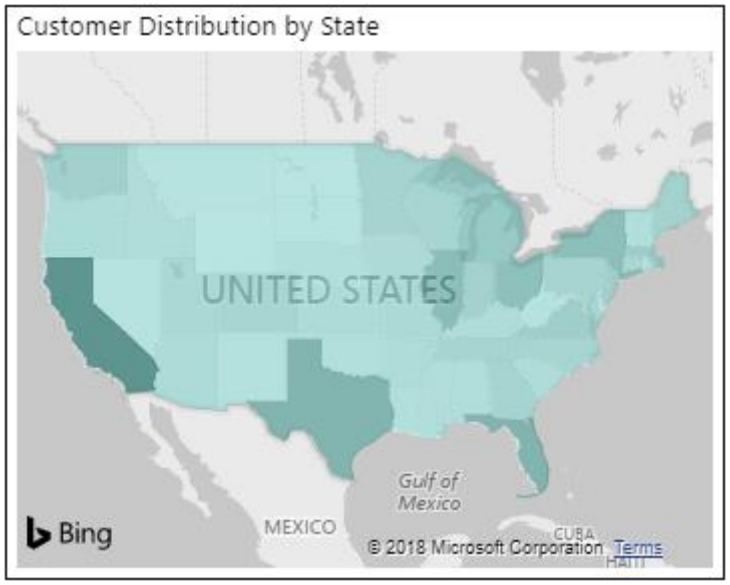
5. Click the **Format** icon below the visualization icons in the **Visualizations** pane.

6. Expand the **Map styles** section and change the **Theme** to **Grayscale**.

7. Expand the **Title** section and set the **Title Text** to **Customer Distribution by State, Font colour** to **Black, and Text size** to **10.**

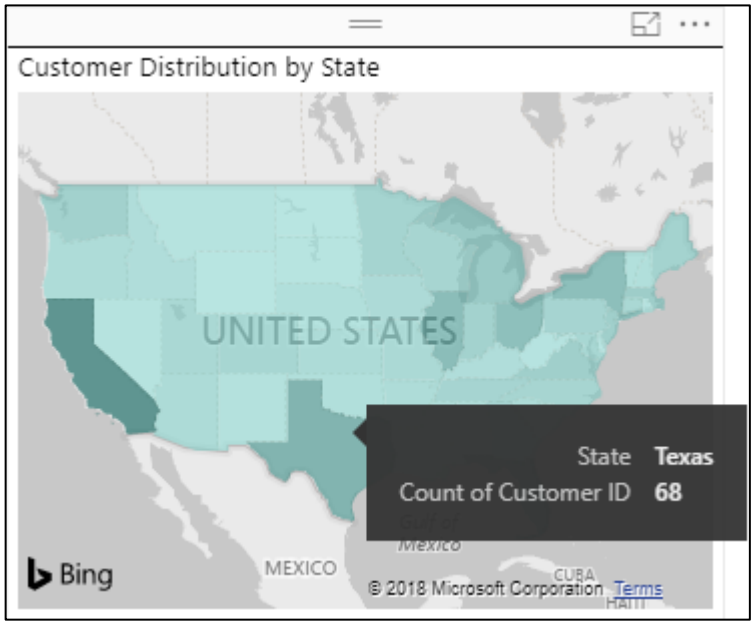
8. Slide the **Border** option to **On**.

9. Move the filled map to beneath the bubble map and resize the filled map as necessary.



Filled Map showing Customer Distribution by State

10. Hover over a state to highlight the plotted value.



Highlighting Filled Map Area

Note: *Clicking on a location in a filled map activates the effect of cross-filtering the other visualizations on the report page and vice versa. Click on the map background to clear the filter. To set a suitable view on the map, zoom in and out the map area by using the mouse wheel control.*

The Filled map facilitates an overview on the count of customers by state using concentration of colours.

11. Save the file as **MapChart Added** Power BI file and close it.

11.6 REVIEW EXERCISE

1. Which of the following is typically used to present quantitative data for comparison and inspection of details?
   a) Pie chart b) Line chart
   c) Table d) Map

2. Open the **Chartstory Power BI file** in **Power BI Desktop** and add a **border** to the **stacked column chart** for **sales and profit by month**.

3. Change the **data colours** for **Sales** to **blue**.

4. Create a **clustered column chart** for **Sales** by **month** and **region**.

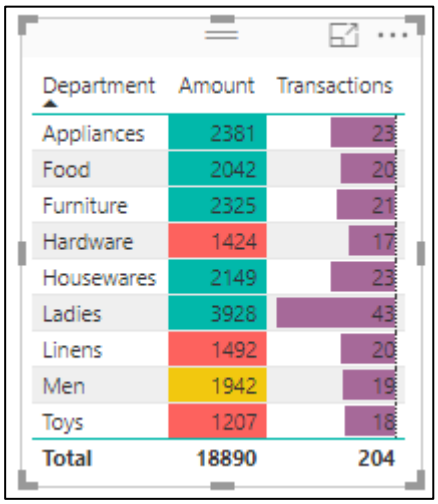
5. Place the **clustered column chart** below the **stacked column chart**.

6. Save the file as **Chartstory Added** Power BI file and close it.

7. Open the **PivotReport Power BI file** in **Power BI Desktop**

8. Use the **Background colour** conditional formatting for the **Amount column** and format all values according to the following range:
   - **Red** fill for values **between 0** and **below 1500**.
   - **Yellow** fill for values **between 1500** and **2000**.
   - **Green** fill for values **above 2000**.

9. Apply the **Data bars** conditional formatting for the **Transactions** column:
   - Set the **bar col**our to **purple**.
   - Change the **bar direction** to **Right to left**.



Department	Amount	Transactions
Appliances	2381	23
Food	2042	20
Furniture	2325	21
Hardware	1424	17
Housewares	2149	23
Ladies	3928	43
Linens	1492	20
Men	1942	19
Toys	1207	18
Total	**18890**	**204**

10. Save the file as **PivotReport Added** Power BI file and close it.

LESSON 12 –
CREATING ADDITIONAL DATA VISUALIZATIONS

In this section, you will learn how to:

- Create KPI and gauge chart visualizations to measure progress
- Create card visualizations
- Create matrix visualizations
- Add interactivity using slicers

12.1 Creating Visualizations to Measure Progress

Concepts

In **Microsoft Power BI** you can create visualizations to measure and visually communicate progress towards a set goal or target.

The two main types of visualizations for measuring progress are:

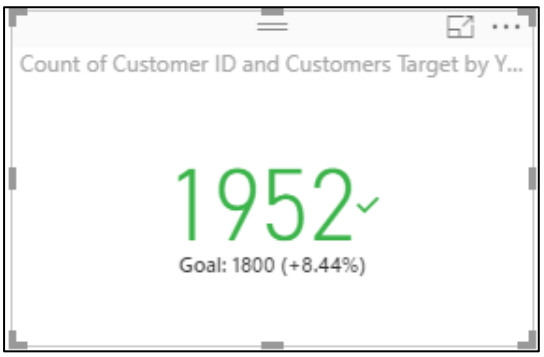
- The KPI (key performance indicator) visualization
- The Gauge visualization

The KPI (key performance indicator) visualization

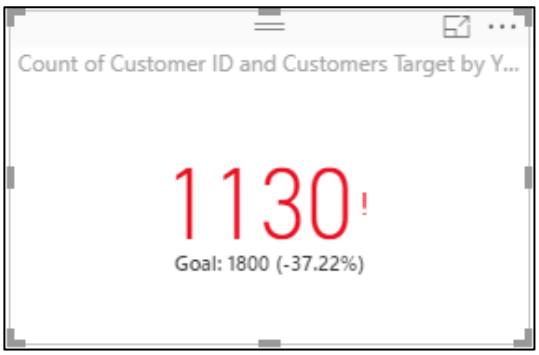
This displays the Base measure (current status), the Goal to be achieved and the distance from the goal as a percentage.

For example, the Base measure may be the current number of customers and the Goal may be the number of customers you'd like to achieve this year. The Base is evaluated against the Goal.

- If the Base measure exceeds the Goal, this is a positive KPI shown in green font with a tick.
- If the Base measure isn't meeting the Goal, this is a negative KPI shown in red font with an exclamation mark.
- If the base measure matches the goal, the font is yellow.



Positive KPI – Base Measure Exceeds Goal



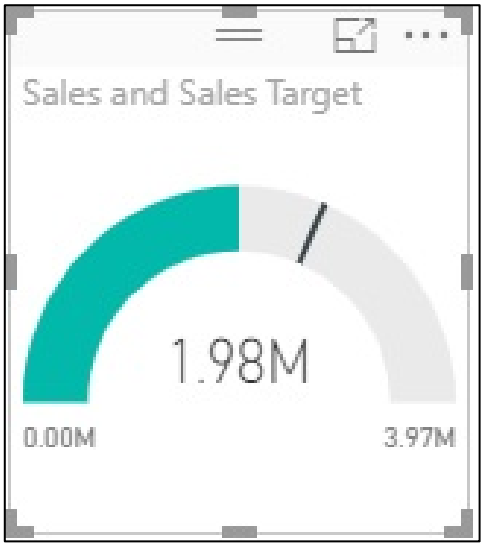
Negative KPI – Goal Not Yet Met

The Gauge visualization

This is also known as a gauge chart. This has a circular arc measuring progress toward a goal/target. All possible values are spread evenly along the arc, from the minimum (left-most value) to the maximum (right-most value).

The goal/target is represented by a marked line on the gauge.

Progress (current status) toward the goal is represented by the shading. And a single value shown in bold inside the arc represents the progress.



Gauge Chart – Goal Not Yet Met

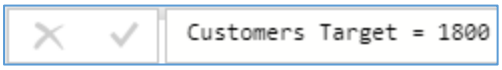
Steps

To create a **KPI** visualization – in this case, to create a **Customer's KPI** (to show the number of existing customers metric against the targeted number of customers for the year) and an **Orders KPI** (to show the number of orders metric against the targeted number of orders for the year):

1. Open the **KPI Power BI file** in **Power BI Desktop**.

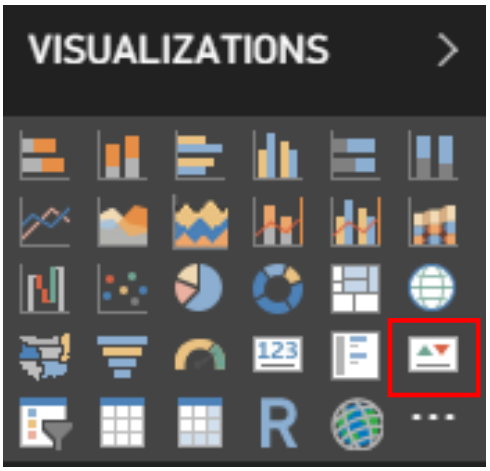
2. To set the target or goal against which to measure progress, click **New Measure** in the **Calculations** group on the **Home** tab.

3. In the **Formula bar** that appears at the top of the canvas, type **Customers Target = 1800**. Press **Enter** to add the measure to the fields list in the **Fields** pane.



Creating a New Measure

4. In the **Report** view, click on a blank area in the **Report Canvas** area and select **KPI** icon in the Visualizations pane.



KPI Template

5. Drag and drop the field **Customer ID** from the **Fields** pane into the **Indicator** area (base) of the **Visualizations** pane.

   The indicator controls the base value's display units and decimal places.

6. Drag and drop the field **Order Date** from the **Fields** pane into the **Trend axis** area of the **Visualizations** pane.

   The trend axis is displayed as the background of the KPI visualization.

7. Click the **Order Date** drop down list in the **Trend axis** area and select **Date Hierarchy**.

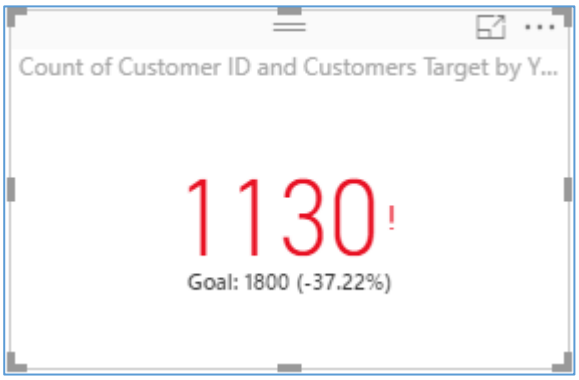
8. Drag and drop the measure **Customers Target** from the **Fields** pane into the **Target goals** area (target) of the **Visualizations** pane.

   The visualization displays the goal and the distance from the goal as a percentage.

Positive KPI – Target Exceeds Goal

   The indicator will be displayed in green with a check icon if the target exceeds the goal.

9. Click the **Count of Customer ID** field in the **Indicator** area of the **Visualizations** pane and change the aggregation to **Count (Distinct)**.

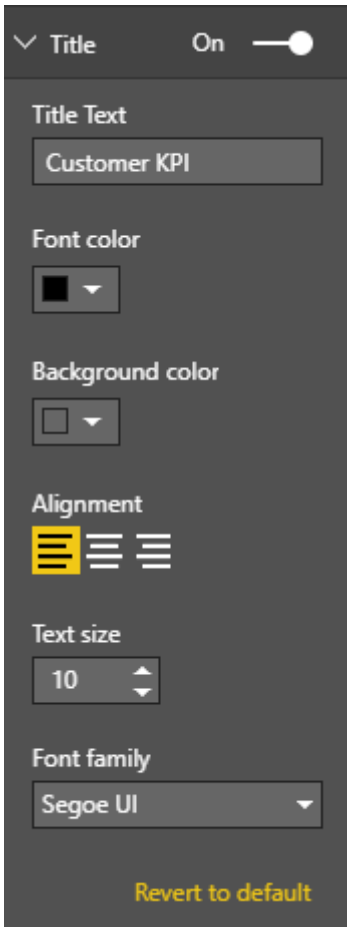


Negative KPI – Target Not Yet Met

   The indicator will be displayed in red with an exclamation icon if the target is not yet met. If the indicator matches the goal, the indicator will be displayed in yellow.

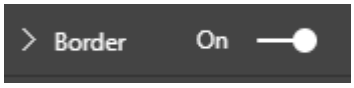
10. Click the **Format** icon below the visualization icons in the **Visualizations** pane.

11. Expand the **Title** section and set the following options:
    a. **Title Text**: **Customers KPI**
    b. **Font colour**: **Black**
    c. **Text size**: **10**



Options for KPI Title

12. Slide the **Border** option to **On**.



Options for KPI Title

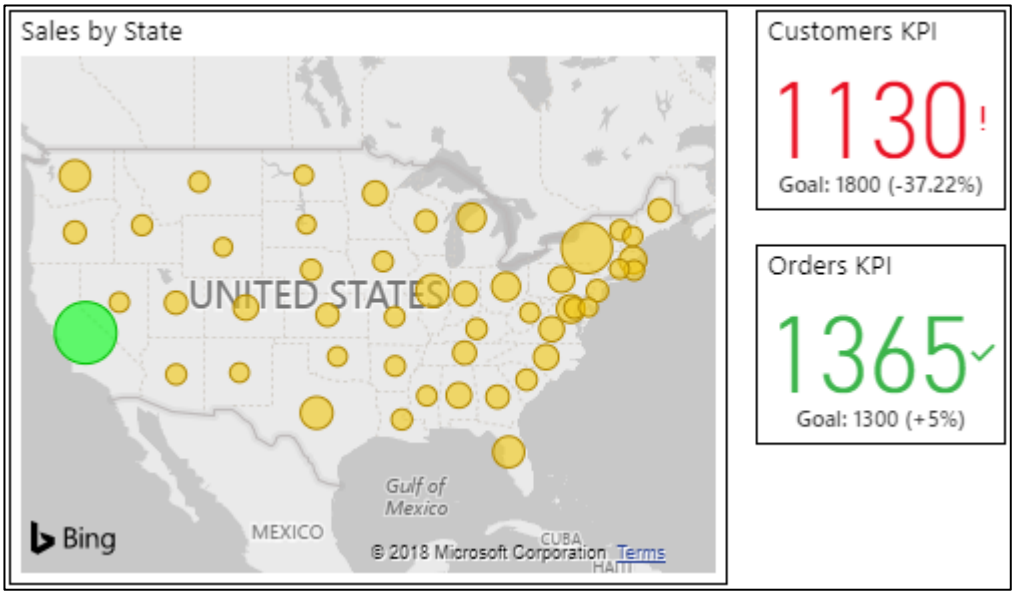
13. Position the KPI visual to the right of the bubble map and resize it as necessary.

Formatted Customers KPI

14. To create a **KPI** visualization to show the **number of orders metric** against the **targeted number of orders for the year**, first **create a copy of the Customers KPI** and **position it below the Customers KPI**.

15. To set the target or goal against which to measure progress, create a new measure called **Orders Target** and assign the value **1300** to it.

16. Select the second KPI and set the following options in the **Visualizations** pane:
    a. **Indicator**: **Order ID**
    b. **Aggregation method: Count (Distinct)**
    c. **Trend Axis: Order Date**
    d. **Order Date Hierarchy: Year**
    e. **Target goals**: **Orders Target**
    f. **Title: Orders KPI**

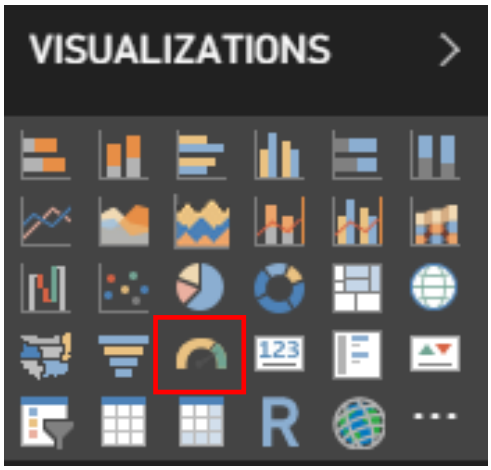


17. Save the file as **KPI Added** Power BI file and the KPI visualizations are saved with the file.

To create a gauge chart - in this case to display the **sales achieved against a target sales figure**:

1. Open the **KPI Solution Power BI file** in **Power BI Desktop**.

2. Select the **Gauge** icon in the **Visualizations** pane.

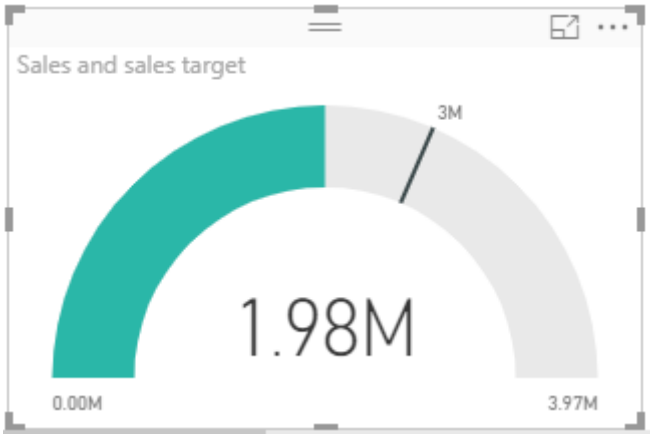


Gauge Template

3. To set the target or goal against which to measure progress, create a new measure called **Sales Target** and assign the value **2500000** to it.

4. Drag and drop the **Sales** field from the **Fields** pane into the **Value** area of the **Visualizations** pane.

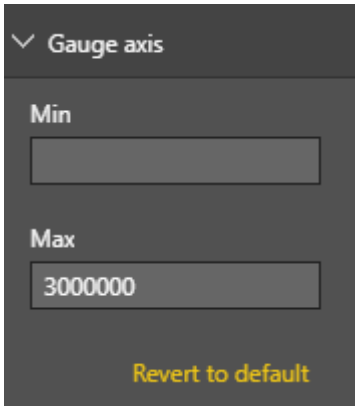
5. Drag and drop the fields **Sales Target** from the **Fields** pane into the **Target value** area of the **Visualizations** pane.



Gauge Visualization

6. Click the **Format** icon below the visualization icons in the **Visualizations** pane.

7. Expand the **Gauge axis** section and set the **Max** option to **3000000**.



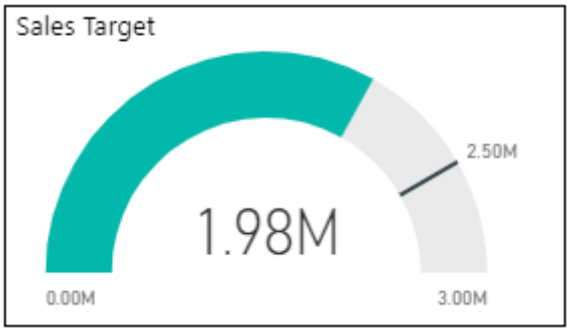
Gauge axis setting

8. Slide the **Border** option to **On**.

9. Expand the **Title** section and set the **Title Text** to **Sales Target, Font colour** to **Black, and Text size** to **10.**

10. Expand the **Target** section, ensure the option is set to **On** and set the **value decimal places** as **2**.

11. Position the gauge below the **Orders KPI** and resize as necessary.



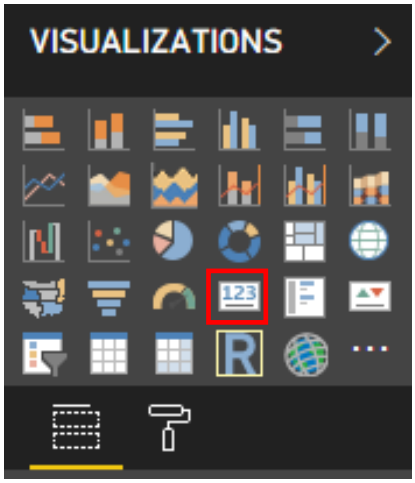
Formatted Gauge Visualization

12. Save the file as **Gauge added** Power BI file and the visualizations with gauge chart is saved with the file.

12.2 CREATING CARD VISUALIZATIONS

Concepts

Sometimes you may want to highlight a single number in a report - for example, total sales, market share year over year, or total number of customers. **Microsoft Power BI** offers **Card** as a type of data highlighting for the purposes of reporting.



Card Visualization Icon

Steps

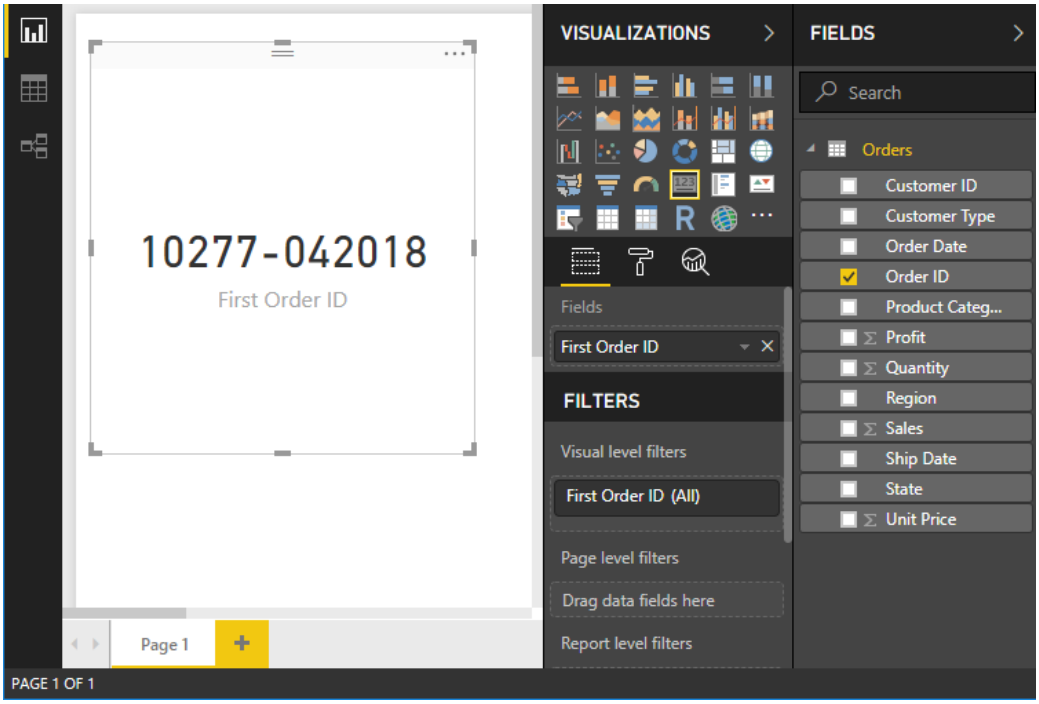
To create a card – in this case to create **three cards** to display the **number of orders**, the **number of customers** and the **total sales** in a report:

1. Open the **Orders Card Power BI file** in **Power BI Desktop**.

   The report is connected to the **Orders List.xlsx** workbook, which contains order details for the first half of the year 2018.

2. In the **Report** view, select the **Card** icon in the **Visualizations** pane.

3. Select the **Order ID** field under the **Orders** table in the **Fields** pane.



Adding a Field for the Card Visualization

The first value of the **Order ID** field is displayed by default as the field consists of text values. The aggregation method for the field needs to be changed to count the number of unique orders.

4. Click the **First Order ID** field's down arrow in the **Fields** area and select **Count (Distinct)**.

The card displays the number of unique orders in the data set.

Card Visualization with Count Aggregation

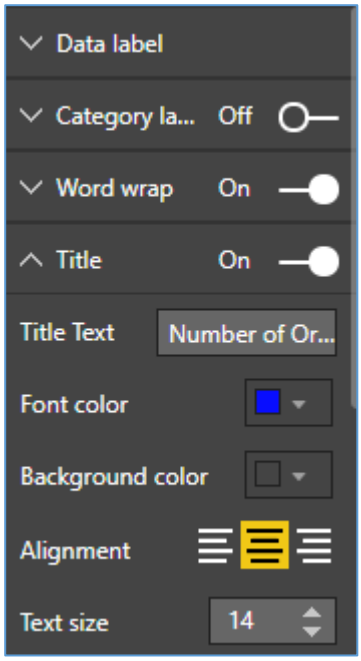
5. To format the card visualization, click the **Format** icon below the visualization icons in the **Visualizations** pane.

6. Slide the **Category** label option to **Off**.

7. Slide the **Border** and **Title** options to **On**.

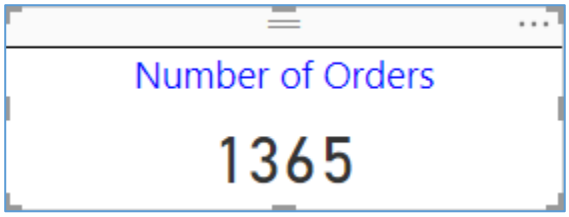
8. Expand the **Title** section and type **Number of Orders** in the **Title Text** box.

9. Select a **blue font** colour, **center align** the **title** and set the **font size** to **14**.



Formatting the Card Visualization

10. Adjust the height of the card visualization so that the title and the value are displayed without too much white space.



Formatted Card Visualization

11. In the **Data label** section, set the **Text size** to **28**.

12. To add additional cards, copy the **number or orders** visualization and paste **two** sets of it on the right.

13. Edit the two new card visualizations to display the **Number of Customers** and **Total Sales** as follows:

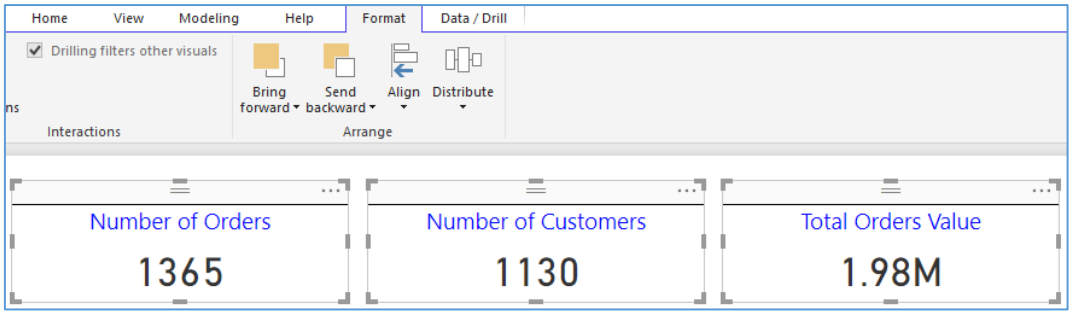
	Number of Customers	**Total Sales**
Field	Customer ID	Sales
Aggregation Method	Count (Distinct)	Sum
Title	Number of Customers	Total Orders Value

Fields with numerical values are displayed as summed value and large numbers are displayed as truncated values, e.g. 1.98M. You can use the **Display units** and **Value decimal places** settings in the **Data label** section to format them.

14. Press **CTRL** and click to select each card.

15. On the **Format** tab, choose **Align** in the **Arrange** group, and choose **Align Top** to align the cards to align the cards to in a single row to the card at the highest level.

16. If necessary, use **Distribute** in the **Arrange** group to distribute the cards into equal spacing in between them.



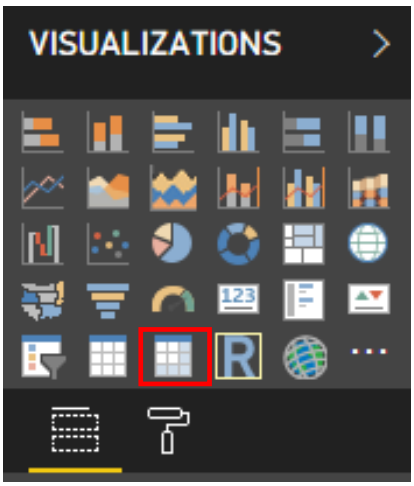
Align and Distribute Commands

17. Save the report as **Orders Card Added** Power BI file.

12.3 CREATING MATRIX VISUALIZATIONS

Concepts

Sometimes you may want to communicate summary information in a report. **Microsoft Power BI** offers Matrix as a way of organising key values for the purposes of reporting.



Matrix Icon

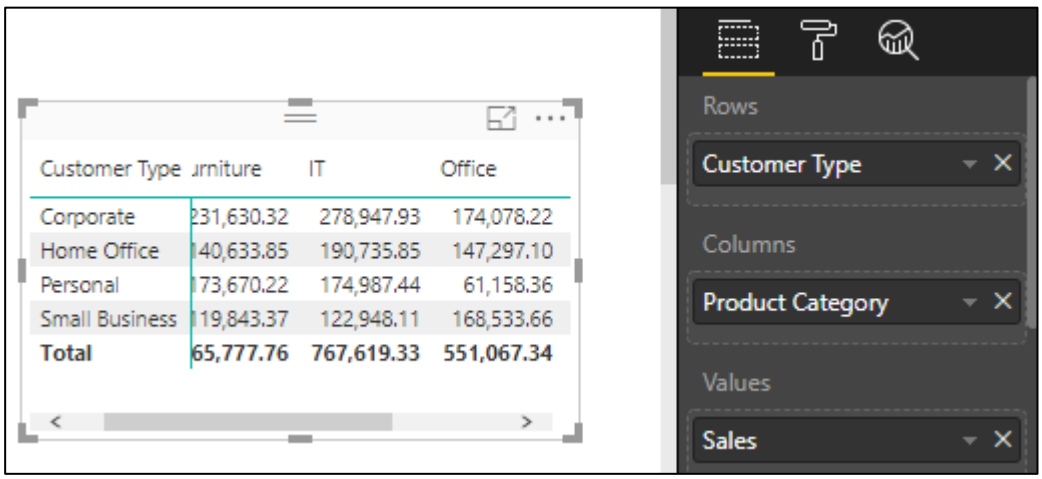
Steps

To create a matrix visualisation – in this case to display the total sales for each customer type and product category:

1. Open the **Orders Card Solution Power BI file** in **Power BI Desktop**.

2. Select the **Matrix** icon in the Visualizations pane.

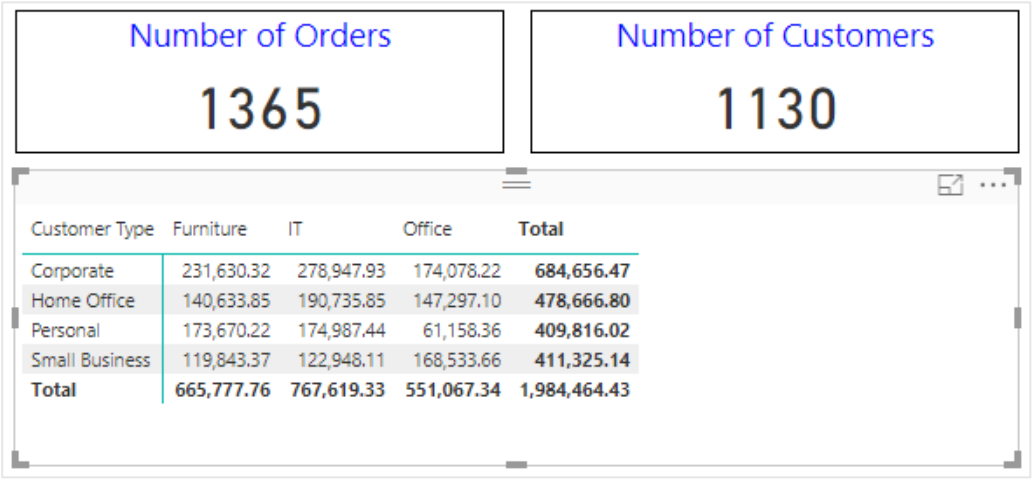
3. Drag and drop the following fields to the respective areas:
    a. **CustomerType** to **Rows**
    b. **Product Category** to **Columns**
    c. **Sales** to **Values**



Adding Fields to the Matrix Visualization

4. Position the matrix visualization below the first card.

5. Adjust the width of the matrix so that it spans approximately the width of the first two cards.



Moving and Resizing the Matrix Visualization

6. Click the **Format** icon below the visualization icons in the **Visualizations** pane.

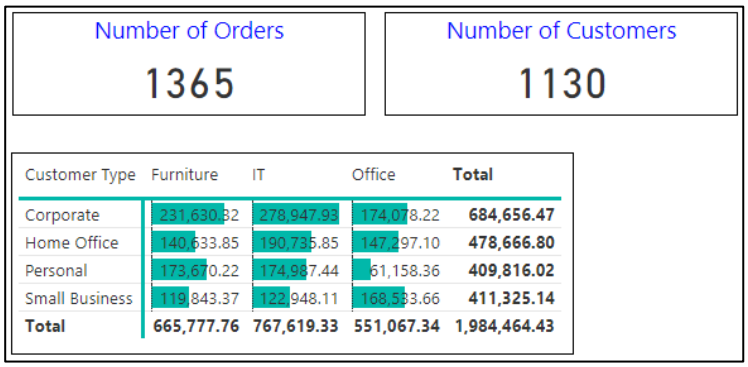
7. Expand the **Matrix style** section and set the style to **Minimal**.

8. Expand the **Grid** section and set the following options:
   a. **Row padding**: **2**
   b. **Outline weight**: **3**
   c. **Text size**: **10**

9. Expand the **Conditional formatting** section and slide the **Data bars** option to **On**.

10. Slide the **Border** option to **On**.

11. Adjust the width and height of the matrix so that all the values are visible, without scrollbars.



Formatted Matrix Visualization

12. Save the report as **Matrix Added** Power BI file.

12.4 Adding Interactivity Using Slicers

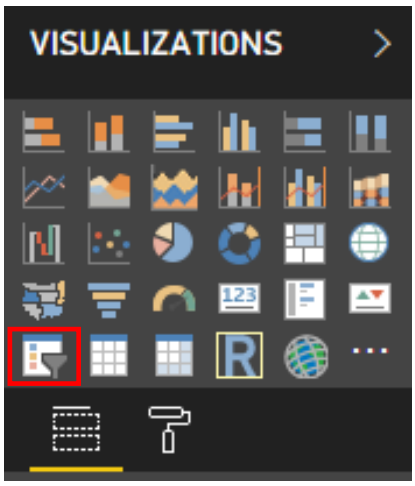
Concepts

A report can include interactive features that allow users to view different aspects of the data and derive other insights from the report. A slicer provides such interactivity features through the display of filters on the report.

Microsoft Power BI offers slicers, with similar functions as that in Microsoft Excel Pivot Tables.

Slicers can be inserted in two ways:

- Select the **Slicer** icon in the **Visualizations** pane and then select the data field and drag it to the **Fields** box in the **Visualizations** pane.



Slicer Template

- Drag the data field from the **Fields** pane into the report canvas area, and then select the **Slicer** icon in the **Visualizations** panel to turn the visualization into a slicer.

Power BI allows for different types of slicers, with different effects and options. Some examples of slicers include field slicers and the date slicers.

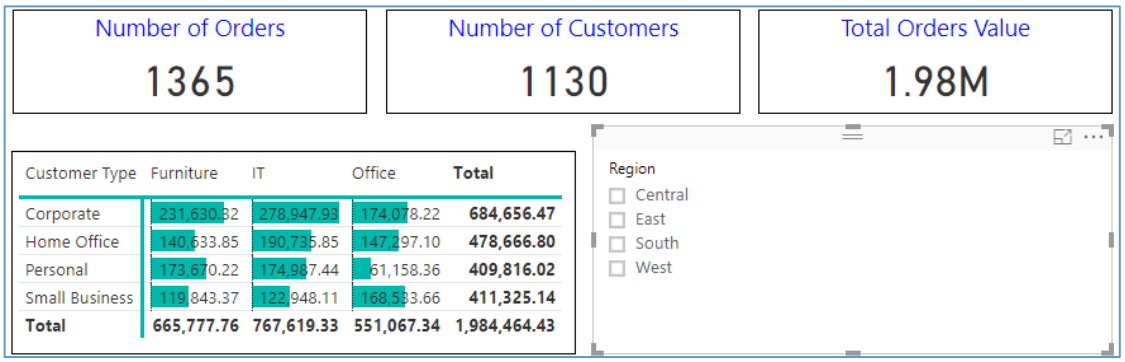
Steps

To insert a slicer - in this case for the **Region** field in a sales summary report:

1. Open the **Sales Summary Power BI file** in **Power BI Desktop**.
2. Select the **Slicer** icon in the **Visualizations** pane.
3. Drag and drop the **Region** field from the **Fields** pane into the **Field** area of the **Visualizations** pane.

The slicer is displayed as a list with options displayed with check boxes. Create more focused reports by putting slicers next to important visuals.

4. Position the slicer visualization next to the matrix visualization and resize the slicer as necessary.



Inserting the Slicer

Note: *List slicer items are sorted in ascending alphanumerical order by default. To reverse the sort order to descending, select the ellipses (...) in the top right corner of the slicer and choose* ***Sort descending*** *in the drop-down list.*

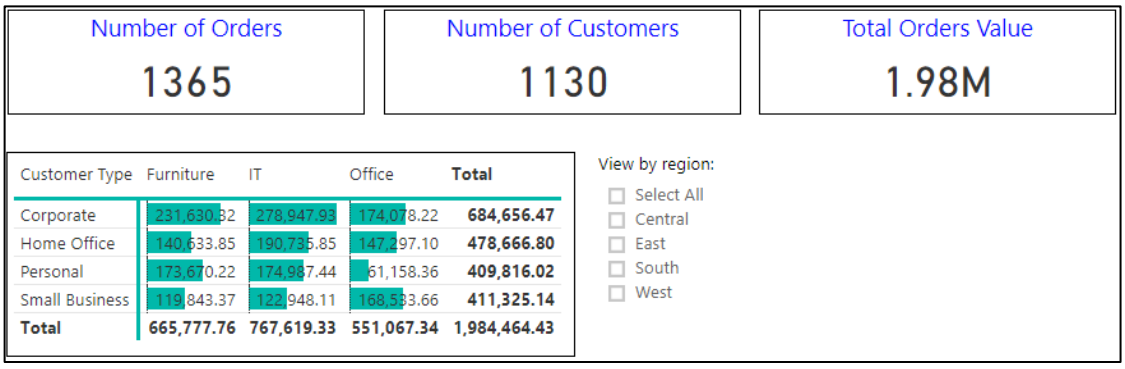
5. Click the **Format** icon below the visualization icons in the **Visualizations** pane.

6. Expand the **Selection Controls** section and slide the **Show "Select all"** option to **On**.

7. Slide the **Slicer header** option to **Off**.

8. Slide the **Title** option to **On**, expand the section and set the following options:
   a. **Title Text**: **View by region:**
   b. **Font color**: **Black**
   c. **Text size**: **10**

9. Resize the slicer as necessary.



Formatted List Slicer

To insert a date slider in the report:

1. Drag and drop the **Order Date** field from the **Fields** pane into the report canvas area.

2. With the new visualization selected, select the **Slicer** icon in the **Visualizations** pane to convert the new visualization to a slicer.

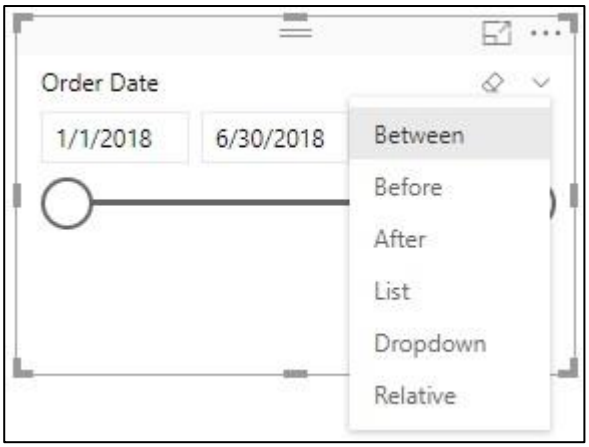
   Date fields are automatically displayed as a slider control with the date range populated. Numeric and date/time data types produce range slider slicers by default.

3. Position the slicer next to the **Region** slicer and resize as required.

   ***Note**: The slider resizes with the slicer size, but it disappears and the dates are cut off if the slicer is too small.*

Inserting a Date Slider

***Note**: The date data field produces a **Between** range slider slicer type by default. To change the slicer type, with the slicer selected, hover over the upper-right area of the slicer, click the drop down arrow that appears, and choose one of the other options, such as **Before** or **After**.*



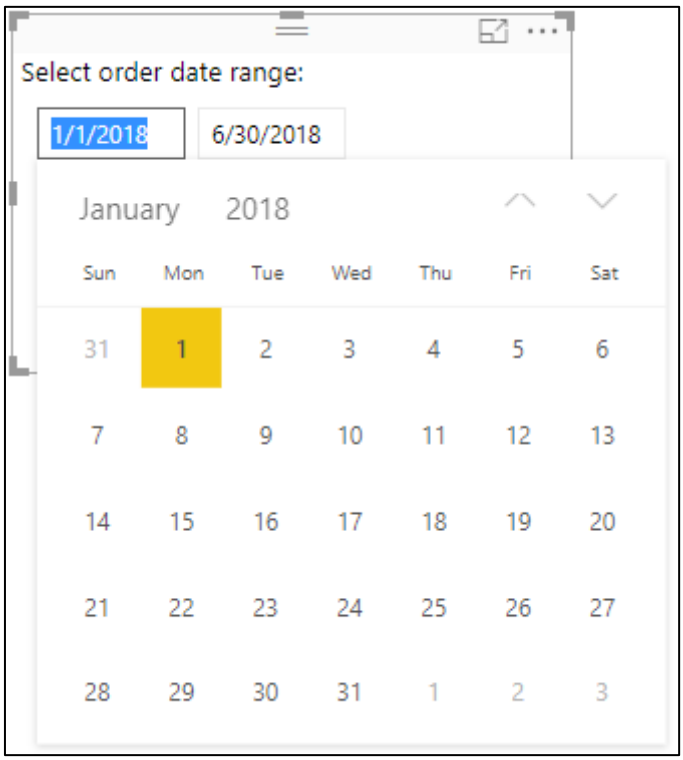
Selecting Slicer Type

4. Click the **Format** icon below the visualization icons in the **Visualizations** pane.

5. Slide the **Slicer Header** option to **Off**.

6. Slide the **Title** option to **On**, expand the section and set the following options:
   a. **Title Text**: **Select order date range:** 1 January to 30 June 2018
   b. **Font color**: **Black**
   c. **Text size**: **10**

   Select different date ranges with the slider, or select a date field to type in a value or pop up a calendar for more precise selection.



Clicking in the Date Field to Pop Up a Calendar

7. Save the report as **Orders Report Final** Power BI file.

12.5 REVIEW EXERCISE

1. Which of the following is a type of visualization that can be used to measure a current value against a target goal?
   a) Visual Slicer
   b) Card visualization
   c) Matrix visualization
   d) Gauge visualization

2. Open the **CreateSummary Power BI file** in **Power BI Desktop**.

   The file shows a table and a map visualization of 2018 sales figures for Brazil, Chile and Columbia.

3. On the **Report view** create a summary of the sales performance for **Brazil**, **Chile** and **Columbia** by adding **three card** visualizations to display the **Number of Orders, Number of Customers** and **Total Sales** for 2018 using the following fields and aggregation methods:

	Number of Orders	**Number of Customers**	**Total Sales 2018**
Field	2018 Sales	Account ID	2018 Sales
Aggregation Method	Count (Distinct)	Count (Distinct)	Sum

4. Set the **text size** of the **data labels** to **28**

5. Turn the **Category label** to **off**

6. Turn the **Border** to on.

7. Enter the **titles** for the respective cards as outlined below:

	Number of Orders	**Number of Customers**	**Total Sales 2018**
Field	2018 Sales	Account ID	2018 Sales
Aggregation Method	Count (Distinct)	Count (Distinct)	Sum
Title	Number of Orders	Number of Customers	Total Sales 2018

8. **Centre align** the **titles** and set their **text size** to **18**.

9. Position the three cards on the **top** of the report page.

10. Align the three cards in a single straight row with equal spacing in between.

11. Add a **slicer** for all **three cards** to show the information for **Brazil**, **Chile** and **Columbia**.

LESSON 13 –
PUBLISHING AND SHARING

In this section, you will learn to:

- Publish a report from Power BI Desktop to the Power BI service
- Share a report from the Power BI service using a link
- Share a report from the Power BI service to the Web
- Create a dashboard in the Power BI service
- Share a dashboard in the Power BI service

13.1 CREATING AND PREPARING TO SHARE REPORTS

Concepts

In **Power BI Desktop** you can connect to data sources and build reports based on your data set by adding appropriate visualizations in **Report View** over one or more pages.

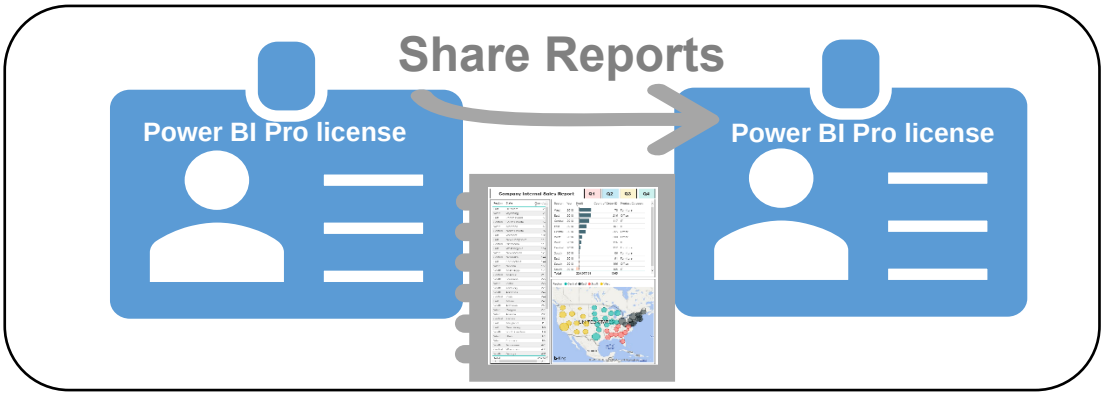
Alternatively, you can connect to data sources and create reports based on your data set in **Power BI service** - the **Power BI** online platform.

Once you have created and saved your report you may want to share it with others. To do this, you need to use **Power BI service**.

Before you can share reports from **Power BI service** you need to set up the appropriate access. To access **Power BI service** as an individual user, there are two main licensing options:

- A **Power BI Pro license** – this enables you to access all content and capabilities of the Power BI service including connecting to data sets, creating reports, and sharing content and collaborating with others. To share and collaborate, the other users will also need to have Power BI Pro licenses.

- A **Power BI free license** – this allows you to access some of the Power BI service capabilities, but it doesn't allow you to share content or collaborate.



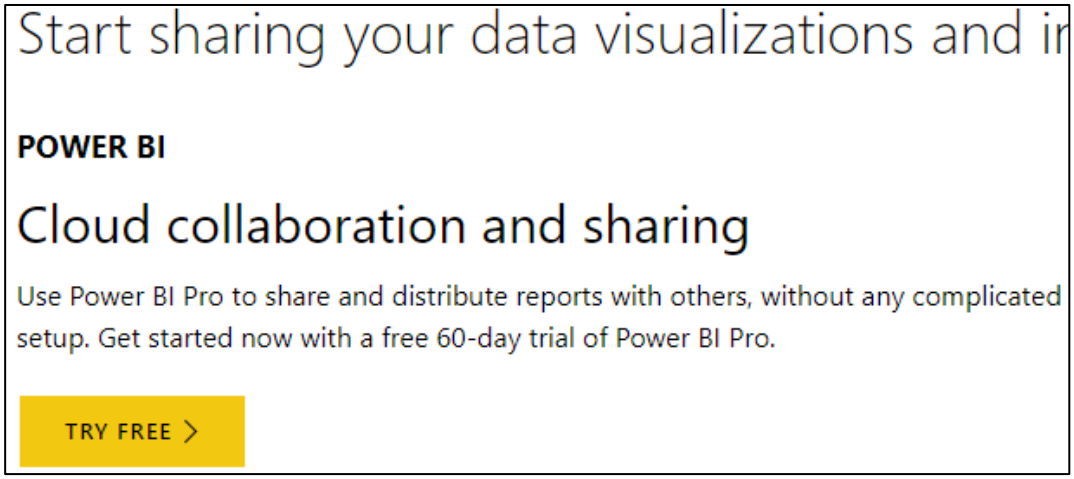
As an individual you can purchase a **Power BI Pro license**, or you can sign up for a **Power BI Pro trial license**. With a trial license, you are automatically moved to a **Power BI free license** when the trial expires. If you are in an organisation that uses Office 365, you should check with your IT administrator to find out your licensing arrangements.

Steps

To set up **Power BI service** – in this case to sign up for a free 60-day trial of the **Power BI service** using a **Power BI Pro license**:

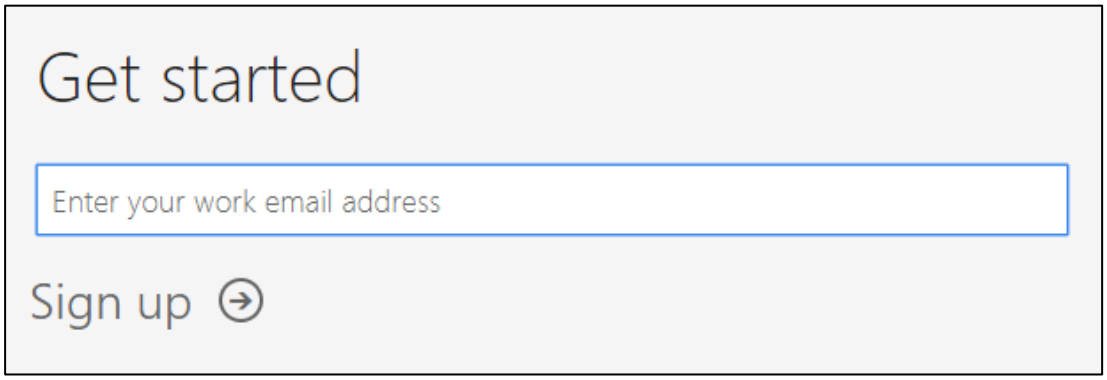
1. Go to **Power BI signup page**: **https://powerbi.microsoft.com/en-us/get-started/**.

2. On the get started page, select **Try Free >** under **Power BI**.



Signing Up for a Free Power BI Account

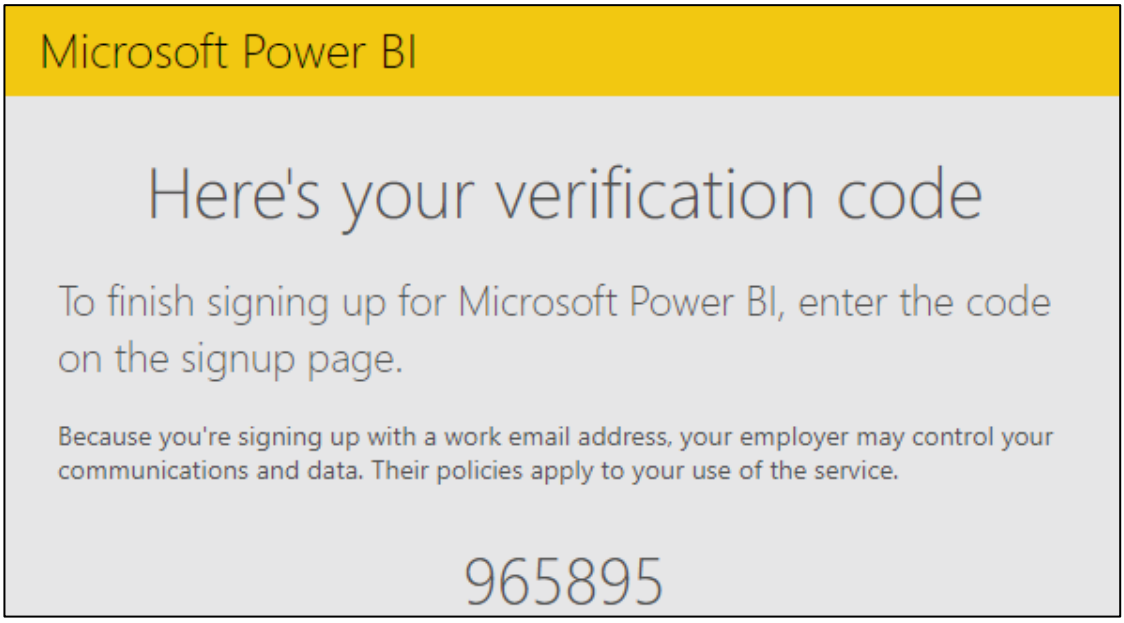
3. Type a work email address (not personal email) in the text box provided, and then click **Sign up**.



Enter a Working Email Address

4. Fill in the required information such as name, password and country.

5. Check your email for a link or verification code email from **Power BI**.



Verification Code

6. Select the link within the email to verify your email address or key in the verification code and click **Start**.

7. If prompted to invite more people, click **Skip**.

   You will be taken to **https://app.powerbi.com** and you can begin using **Power BI service**.

13.2 PUBLISHING AND SHARING REPORTS

Concepts

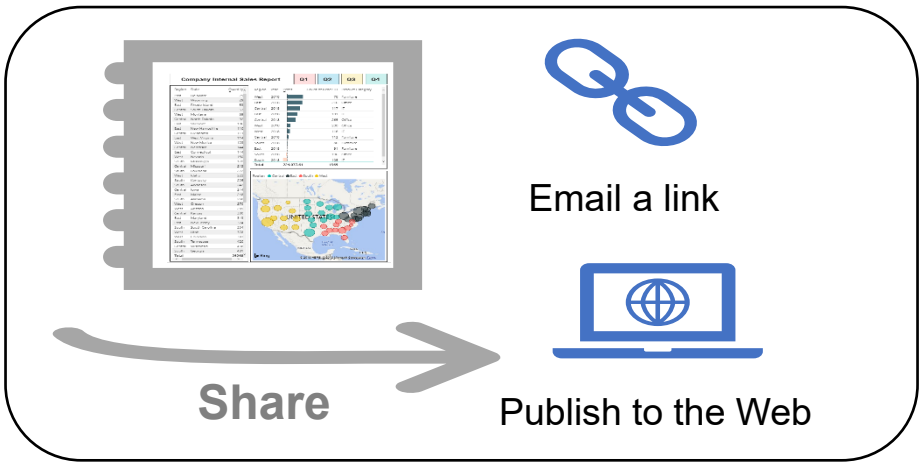
You can share reports created in **Power BI Desktop** by first publishing them to **Power BI service**. When you publish a **Power BI Desktop** file to **Power BI service**, the data set and any reports you created in Report view are published to a workspace in **Power BI service**. A workspace is a dedicated space used to collaborate and share content with others. You can also share reports you created in **Power BI service**.

Two options for sharing a report from a workspace in **Power BI service** are:

- **Emailing a link**: You can share a report with a few people by sending an email with a link to the report. This can be done from within Power BI service or by copying the link into another messaging app. You and your recipients will need to have a Power BI Pro license to do this.

  Recipients of the report can view and interact with it, but they can't edit it. Users within your organisation can share the report with others within your organisation, if granted the appropriate permission, but users outside your organisation won't be able to share it.

- **Publishing to the Web**: You can share a report publicly by publishing it to the web – for example you can embed it in blog posts, websites, through emails or social media. In this case, you and your audience don't need a Power BI Pro license.



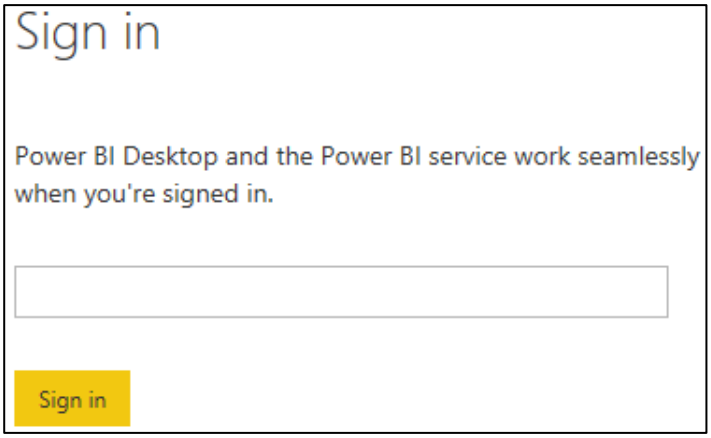
Ways to share reports

***Note:** You can also allow someone to edit content if you both have Power BI Pro licenses – for example by adding them to the relevant workspace in Power BI service.*

Steps

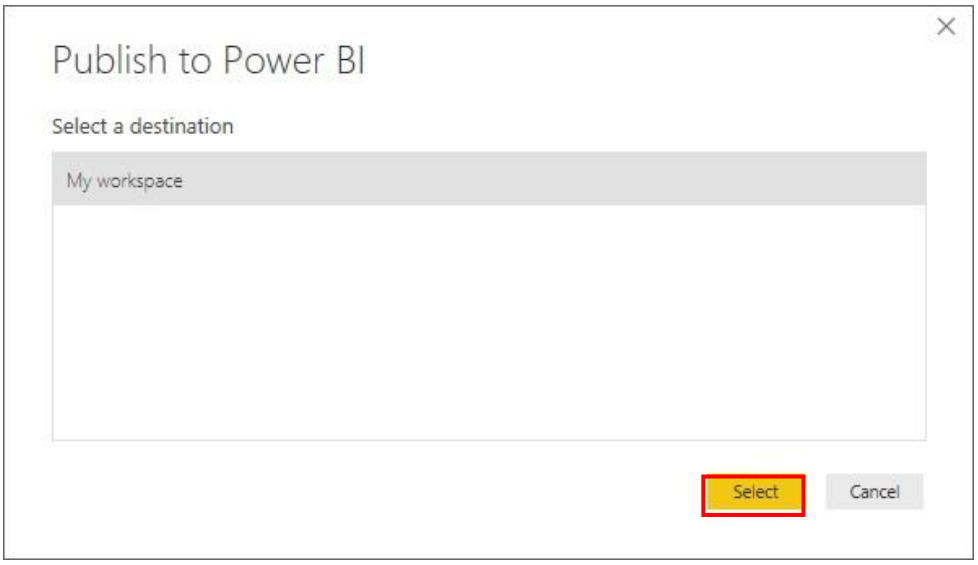
To publish a report created in **Power BI Desktop** to the **Power BI service**:

1. In **Power BI Desktop**, open the report you have created – in this case, **Orders Final - Solution** Power BI file.
2. On the **File** tab, select **Publish**.
3. Select **Publish to Power BI**.
4. Sign in to your **Power BI** account.



Signing In to Power BI Account

5. In the **Publish to Power BI** window, select **My workspace** and click **Select**.

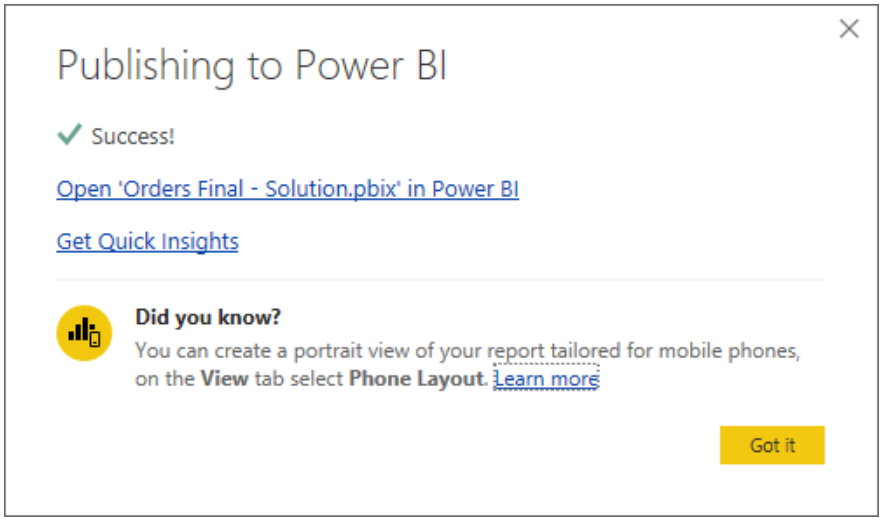


Selecting a Workspace to Publish to Power BI Window

***Note**: Multiple app workspaces can be created by **Power BI Pro** users in the **Power BI service** to organise related content.*

6. Once the publishing is completed, you receive a link to your report. Click the link **Open 'Orders Final – Solution.pbix' in Power BI** to open the report in your **Power BI** site.

   Alternatively, click **Got It** to continue working in **Power BI Desktop**.



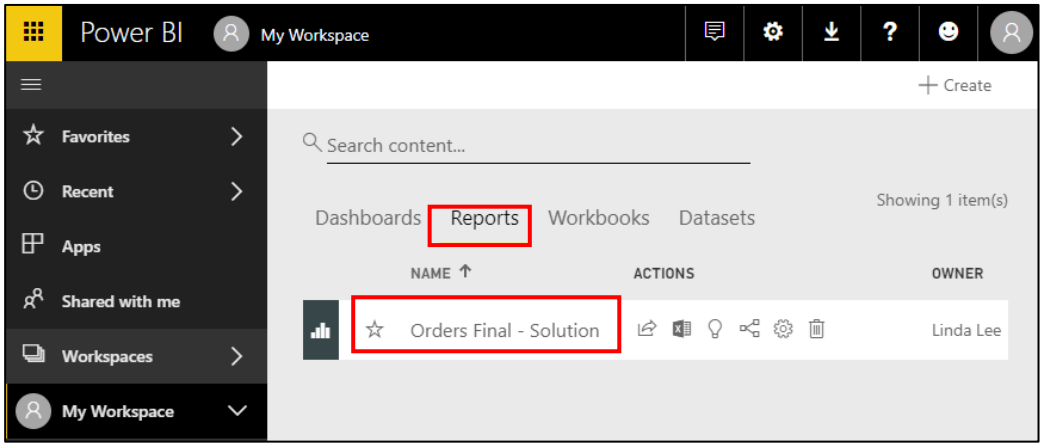
Publishing Completed

To view a report in the **Power BI service**:

1. If not already in the **Power BI service**, go to **https://app.powerbi.com** and sign in using your **Power BI** account.

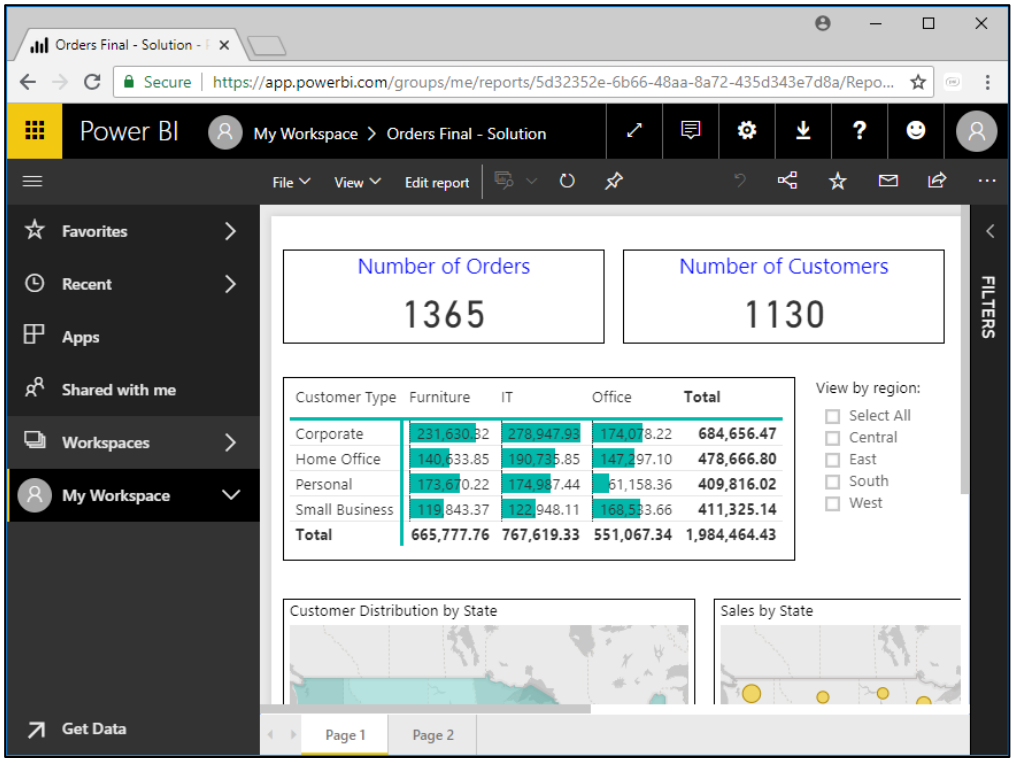
2. Select **My Workspace** on the left menu pane and click the heading **Reports**.

3. The published report will be listed in the window. Click on the report name to view it.



Published Report List

The report is displayed.

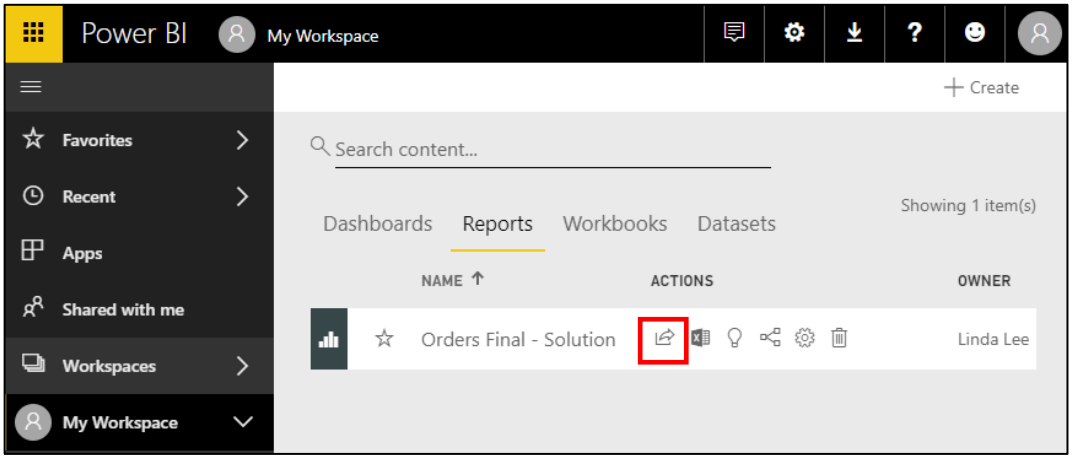


Report in the Power BI service

To share a report from the **Power BI service** using a link:

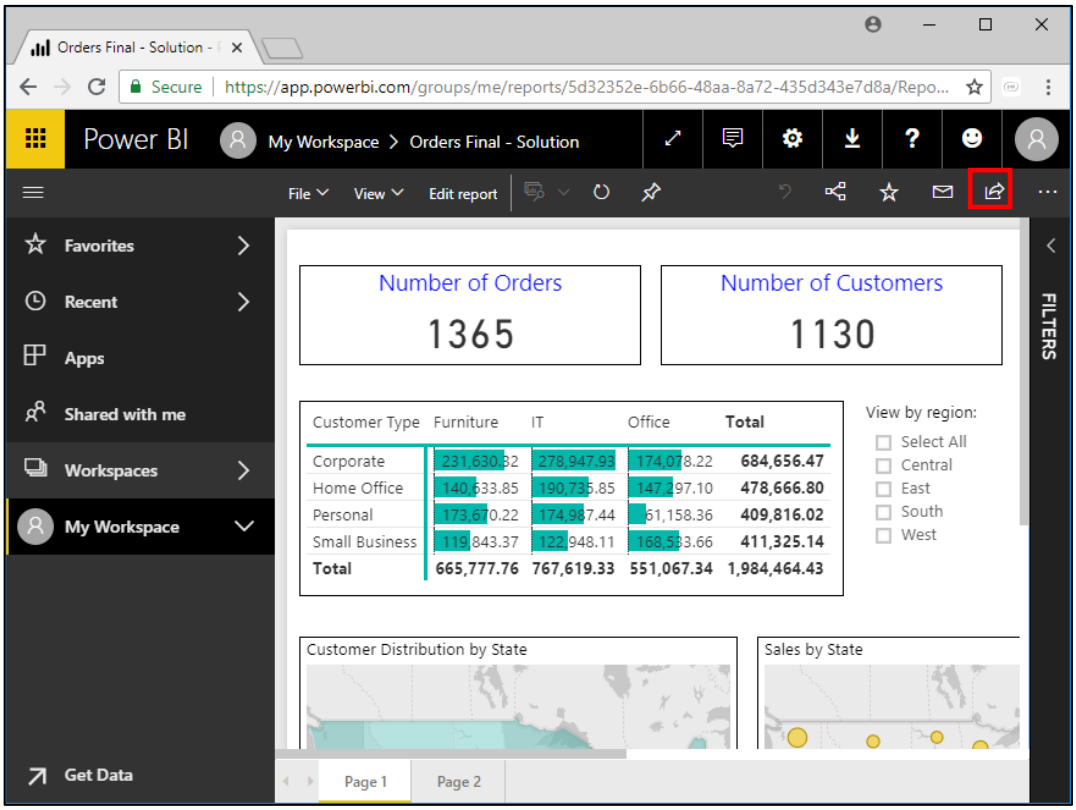
1. If not already in the **Power BI service**, go to **https://app.powerbi.com** and sign in using your **Power BI** account.

2. To share the report, select **My Workspace**, click **Reports** and click the **Share** icon  .



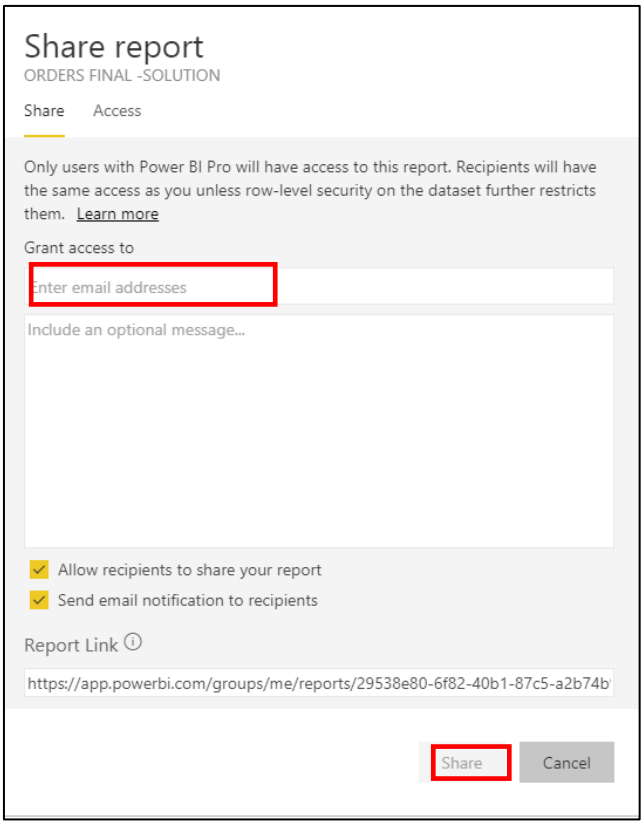
Sharing the Report using a link

***Note**: Alternatively, you can share the report from within the **Report** view by clicking on the **Share** icon.*



Sharing the Report from within the Report view

3. Enter the email addresses of other users in the **Share report** window and click **Share**.



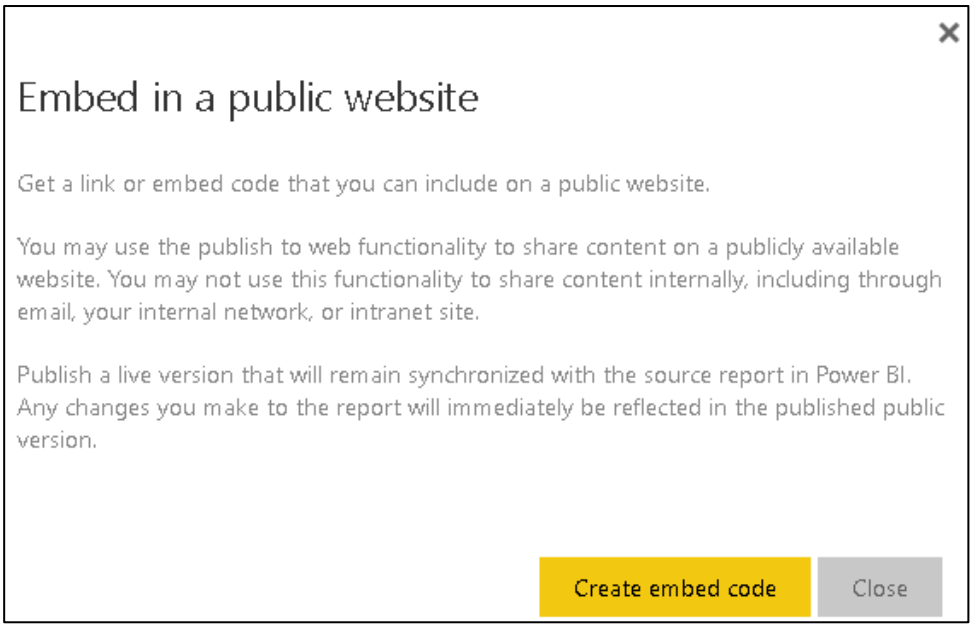
Sharing Options in Share report

***Note**: Alternatively, copy the URL provided under **Report Link** and send via other messaging apps to other users.*

4. To exit the report, click the **Profile** icon  at the top right of the browser window and select **Sign Out**.

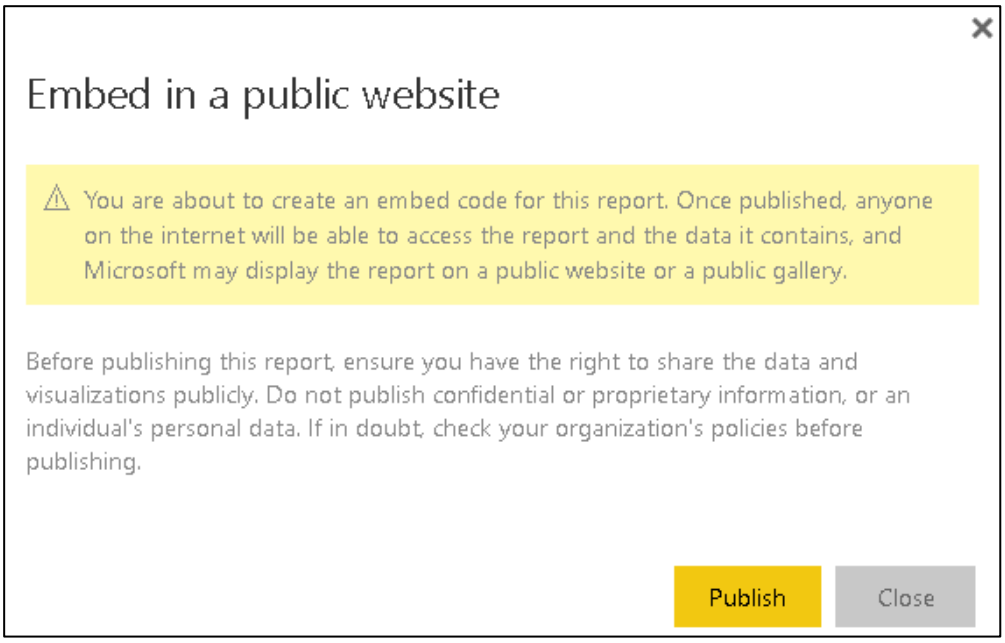
To publish a report from the **Power BI service** to the **Web**:

1. If not already in the **Power BI service**, go to **https://app.powerbi.com** and sign in using your **Power BI** account.
2. In the **My Workspace** section, click on the heading **Reports**.
3. Select the **Orders Final - Solution** report to open it.
4. Select **File** menu and choose **Publish to web**.
5. In the **Embed in a public website window**, click **Create embed code**.



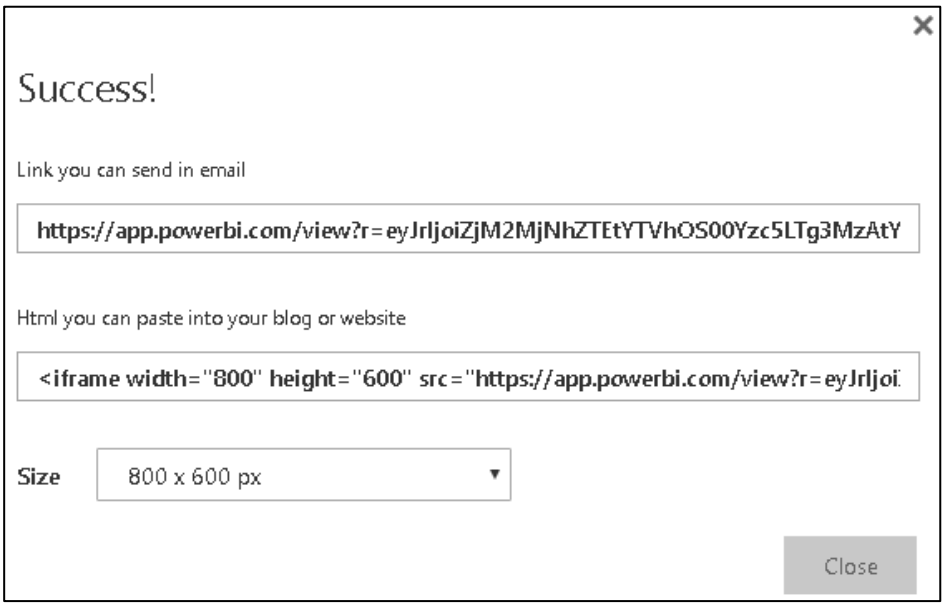
Embed in a Public Website

6. Review the warning shown in the window and confirm that the data is okay to embed in a public website and click **Publish**.



Warning Message on Publishing

7. A window appears that provides a link that can be sent in email or HTML that can be pasted directly into your web page or blog.



Links and Embed Code Generated

8. Click **Close**.

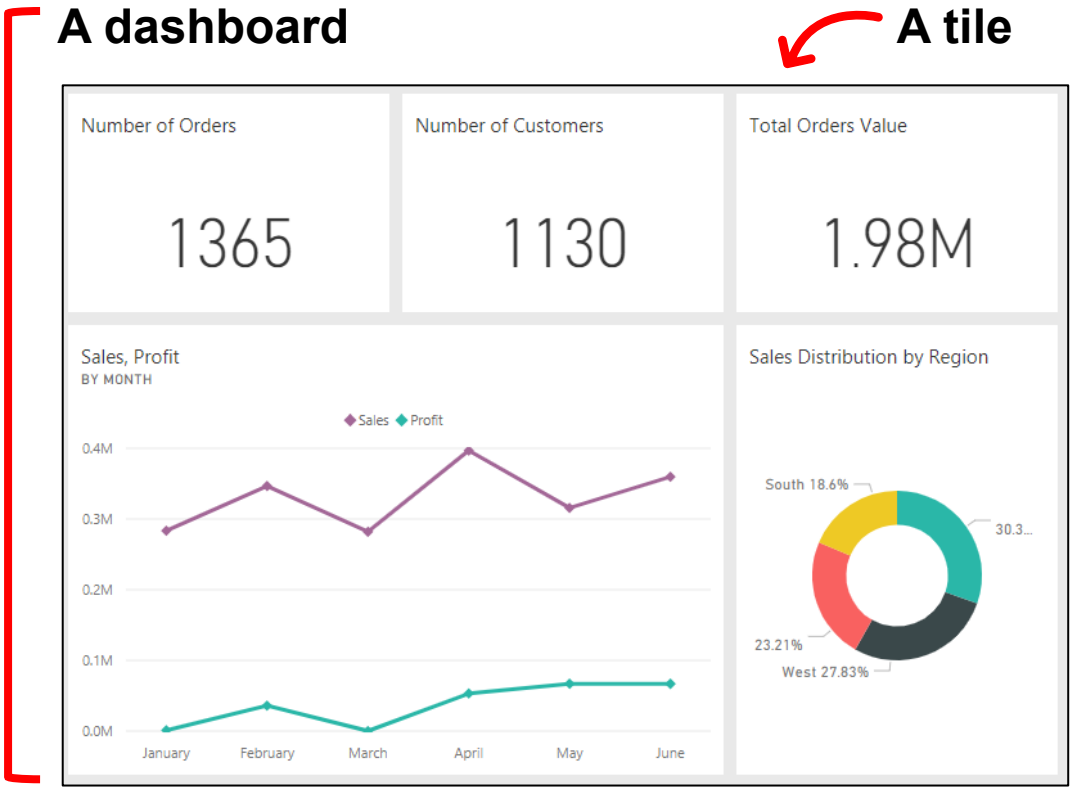
13.3 CREATING DASHBOARDS

Concepts

Dashboards are a visual display of key vital information needed to highlight one or more objectives. Information is consolidated and arranged on a single screen thus allowing information to be monitored at a glance. Dashboards provide an overview and allow organisations to monitor essential performance metrics. A dashboard is interactive and highly customisable, and the tiles update as the underlying data changes.

You can create dashboards in the **Microsoft Power BI service** but not in **Power BI Desktop**.

The visualizations on a dashboard in the **Microsoft Power BI service** are called tiles and you pin (add) them to the dashboard from reports. A tile is a snapshot of your data.



A dashboard made up of tiles

***Note**: If the original visualization used to create the tile changes, the tile does not change. For example, if you pinned a column chart from a report and then you changed the chart to a line chart, the dashboard tile continues to show a column chart. The data refreshes, but the visualization type does not.*

Steps

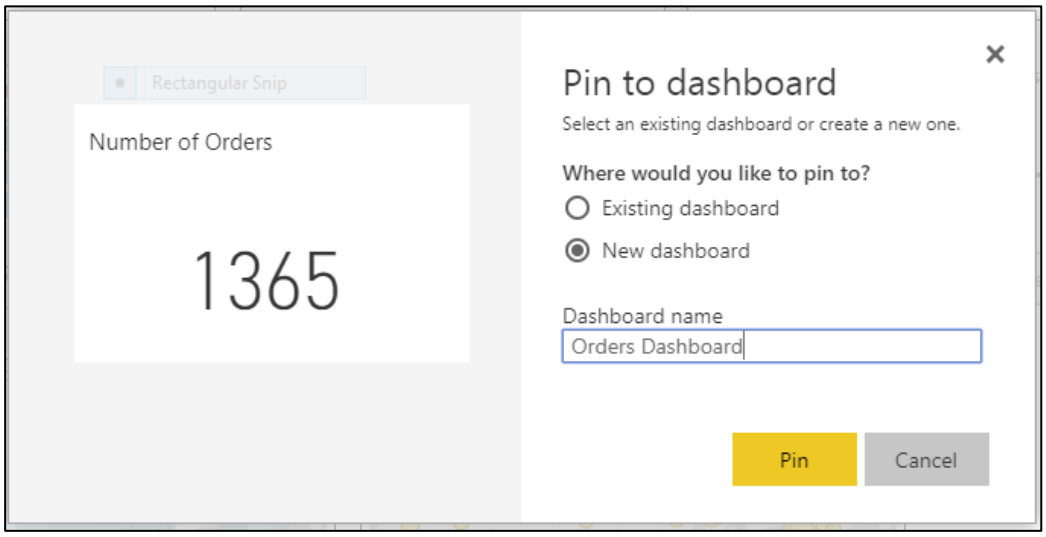
To create a dashboard in the **Power BI service** by selecting visualizations from a report:

1. If not already in the **Power BI service**, go to **https://app.powerbi.com** and sign in using your **Power BI** account.

2. Select **My Workspace** on the left menu pane and select **Orders Final - Solution** under the **Reports** heading.

3. On **Page 1**, hover over the **Number of Orders** visualization and click on the pushpin icon  .

4. On the **Pin to dashboard** window, select **New dashboard** and type **Orders Dashboard** in the text box.



Creating a New Dashboard

5. Confirm the creation of new dashboard by clicking **Pin**.

6. From page 1 of the report, repeat the steps to pin **Number of Customers** card and **Total Orders Value** card visualizations to the existing dashboard (**Orders Dashboard**).

7. From page 2 of the report, repeat the steps to pin **Sales and Profit by Month** line chart and **Sales Distribution by Region** donut chart to the existing dashboard (**Orders Dashboard**).

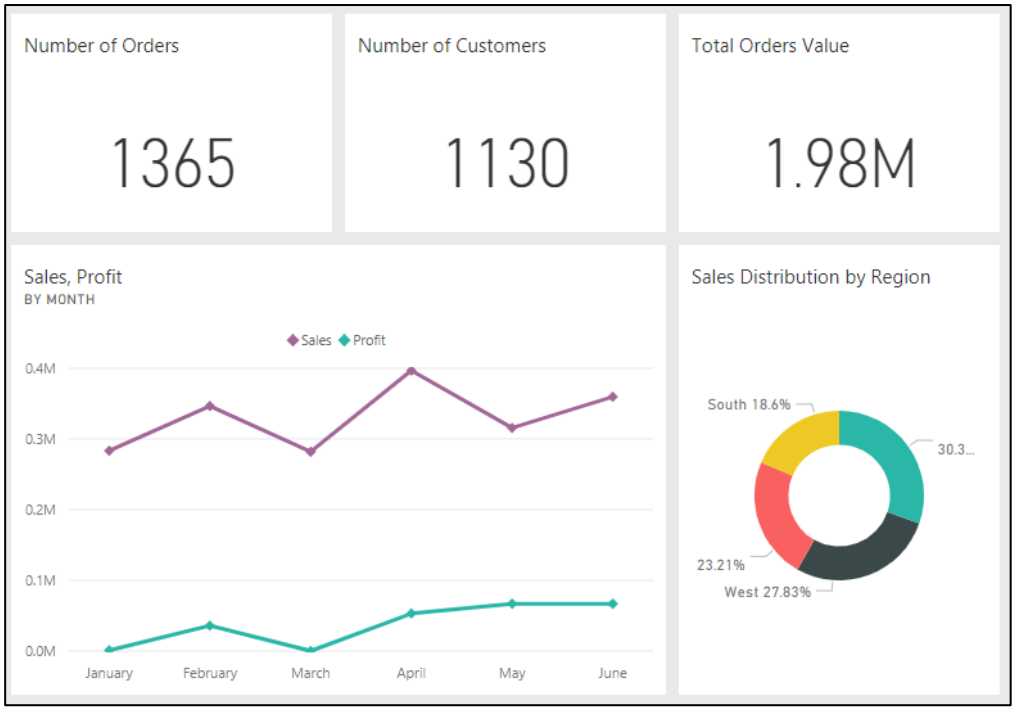
To move and resize tiles in a dashboard:

1. Select **My Workspace** on the left menu pane and select **Orders Dashboard** under the **Dashboards** heading.

2. Hover over the **Number of Orders** Card tile and click on the resize handle ( ) at the bottom right corner.

   The tiles can be resized to a variety of sizes -- from 1x1 tile units up to 5x5.

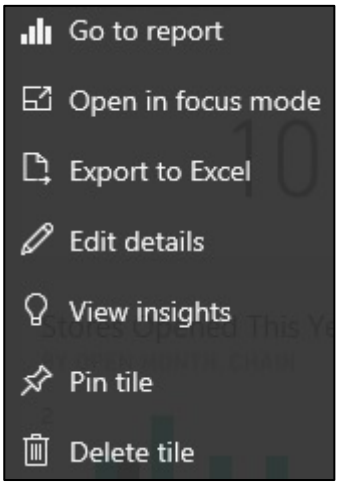
3. Repeat the steps for the rest of the tiles.



Resized Tiles

Tiles also can be renamed for more descriptive and meaningful headings.

4. Hover over the line chart and click on the **More options** icon (...) to open the tile action menu.

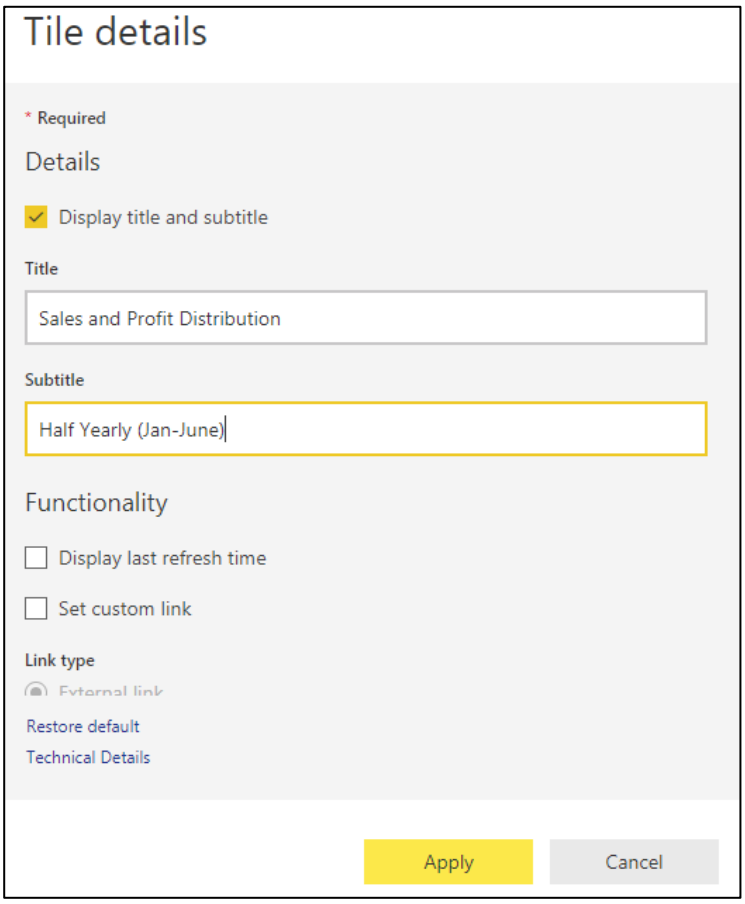


Tile Action Menu

The options available will vary by tile type.

5. Select **Edit details** ( ).

6. In the **Tile details** window, enter the **Title** as **Sales and Profit Distribution** and **Subtitle** as **Half Yearly (Jan to Jun)** and confirm the setting by clicking **Apply**.



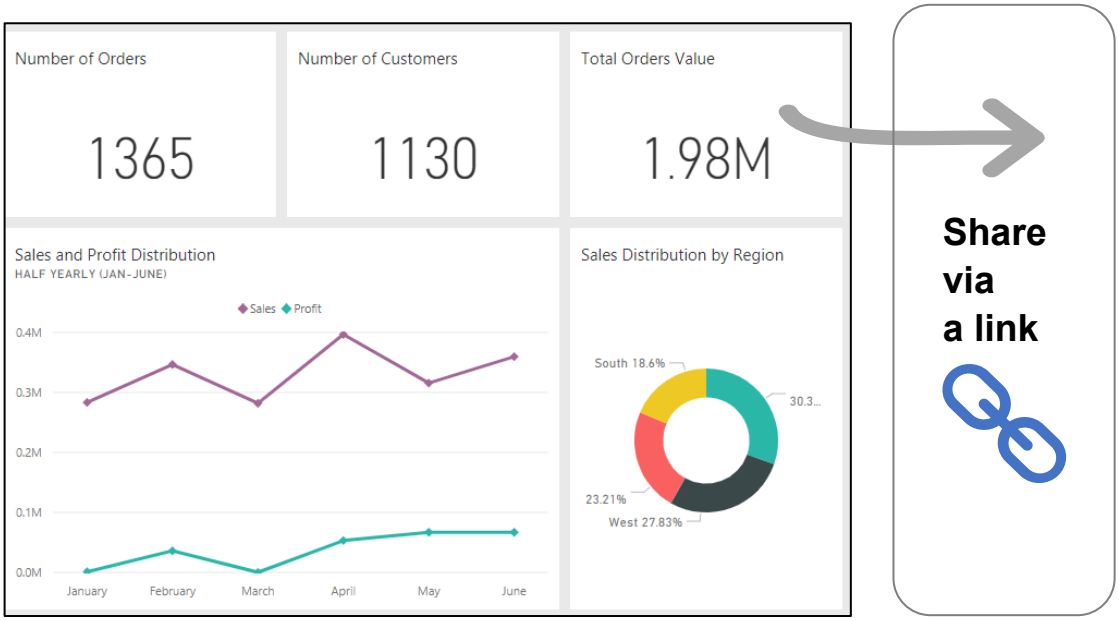
Editing Tile Titles

***Note**: To permanently remove a tile from a dashboard, select **Delete tile** ( ) from the ellipsis dropdown menu.*

13.4 SHARING DASHBOARDS

Concepts

When you create and finalise dashboards in the **Microsoft Power BI service**, you can share them with others. One option is to share them via a link.



***Note**: You and the intended recipients will need to have **Power BI Pro licenses**.*

Steps

To share a dashboard from the **Microsoft Power BI service**:

1. If not already in the **Power BI service**, go to **https://app.powerbi.com** and sign in using your **Power BI** account.
2. Select **My Workspace** on the left pane and select **Orders Dashboard** under the **Dashboards** heading.
3. Click the **Share** icon ( ) under the **Actions** column.

Selecting a Dashboard for Sharing

***Note**: Alternatively, from within an open **Dashboard** click the **Share** icon.*

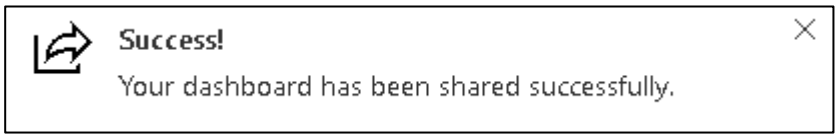
4. In the **Share Dashboard** window, type the full email addresses of the recipients, separating each with a space.

5. Check the options at the bottom part of the window as required.

Share Options

6. Click **Share**.

   Power BI sends an email invitation to the individuals, with a link to the shared content. A message indicating the dashboard has been shared will be displayed.



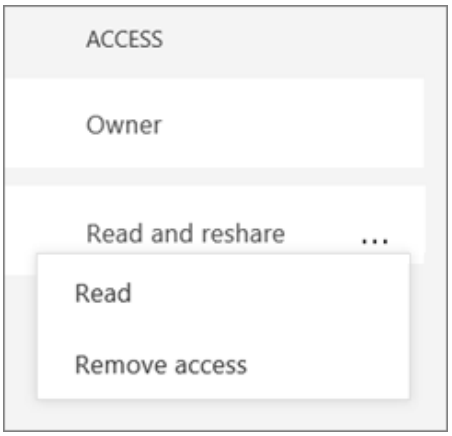
Sharing Message

To modify the access rights of recipients:

1. In the list of dashboards, or from within the dashboard, select **Share** ( ).

2. In the **Share dashboard** window, select **Access**.

3. Select the ellipsis (**...**) next to **Read and reshare** for a recipient and select:
   a. **Read** to keep the recipient from sharing with anyone else.
   b. **Remove access** to keep that recipient from seeing the shared content at all.



Change or Remove Access

4. In the **Remove access** window, decide if you want to remove access to related content, too, such as related reports and data sets.

   If you remove items with a warning icon, it is best to remove related content because it will not display properly.

5. Click **Remove access**.

   ***Note**: You can modify the access rights for recipients of a report by following similar steps for a report.*

13.5 GOOD DESIGN PRACTICE

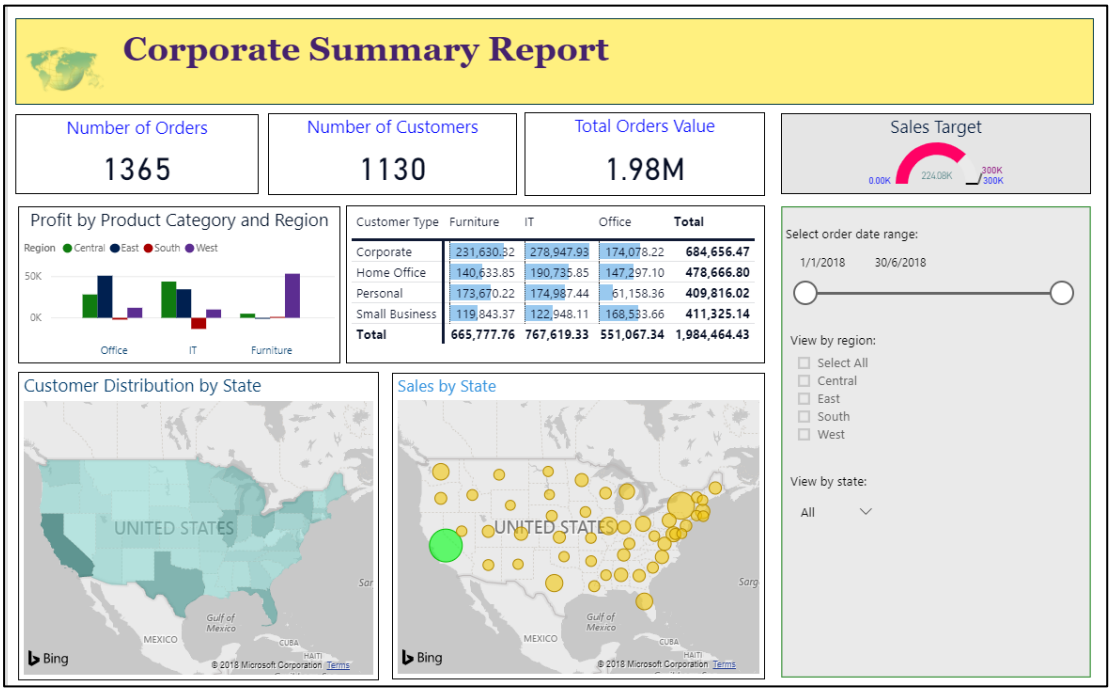
Concepts

When you are creating reports and dashboards your aim is to present key insights and actionable and useful information that can be understood at a glance. To do this, you need to determine the data that you need to gather and what type of visualisations to use. You also need to consider the best way to present the information to your audience. You will need to think about the design and layout of your reports and dashboards.

Incorporating good design practices into your reports and dashboards will make them easier to understand and therefore more effective. Good design practices simplify the visual representation of complex data, helping others understand the key insights and information more easily. Some tips for good design include:

- **Use a clean and uncluttered layout**

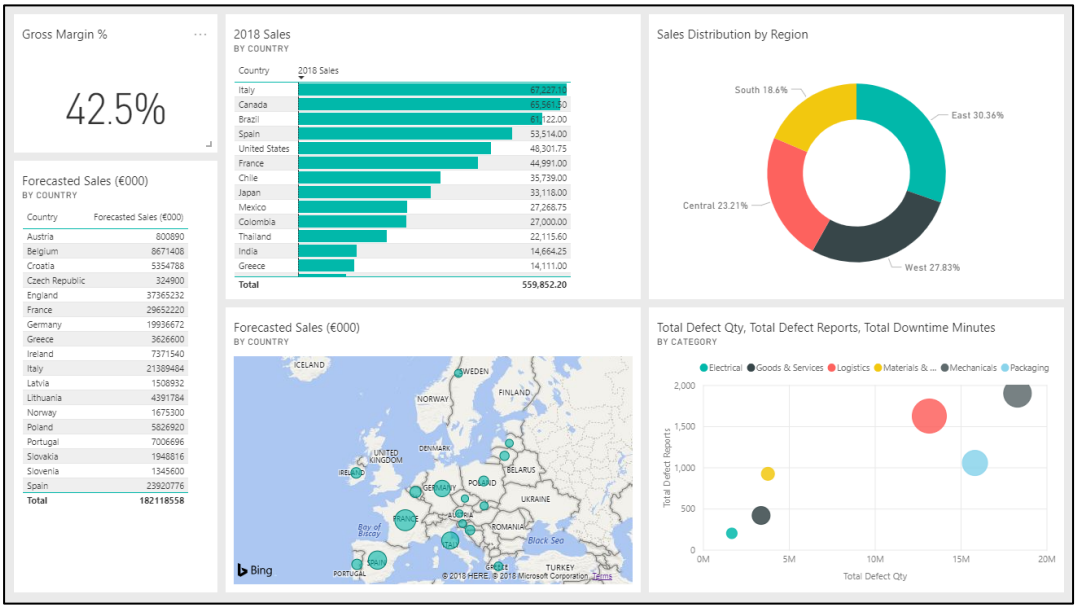
Use a clean and uncluttered layout on reports and dashboards so that the reader can see the key insights easily. Present only the most relevant data as a cluttered layout can be confusing and can distract the user from key insights.



A Clean, Clutter-Free Dashboard

- **Use descriptive titles and labels**

Use descriptive titles to help readers understand the data visualizations. Also consider adding descriptive labels using text boxes to describe the page, grouping of visuals, or to describe an individual visual.



A Dashboard with Descriptive Titles and Labels

- **Use consistent fonts and colour**

Use consistent fonts and colours to help the reader understand the information being presented. In general use one or two font types and up to three font sizes. Using different font sizes helps to create a visual hierarchy in the content. For example, use different font sizes for titles, labels, and data.

- **Use colour for emphasis and understanding**

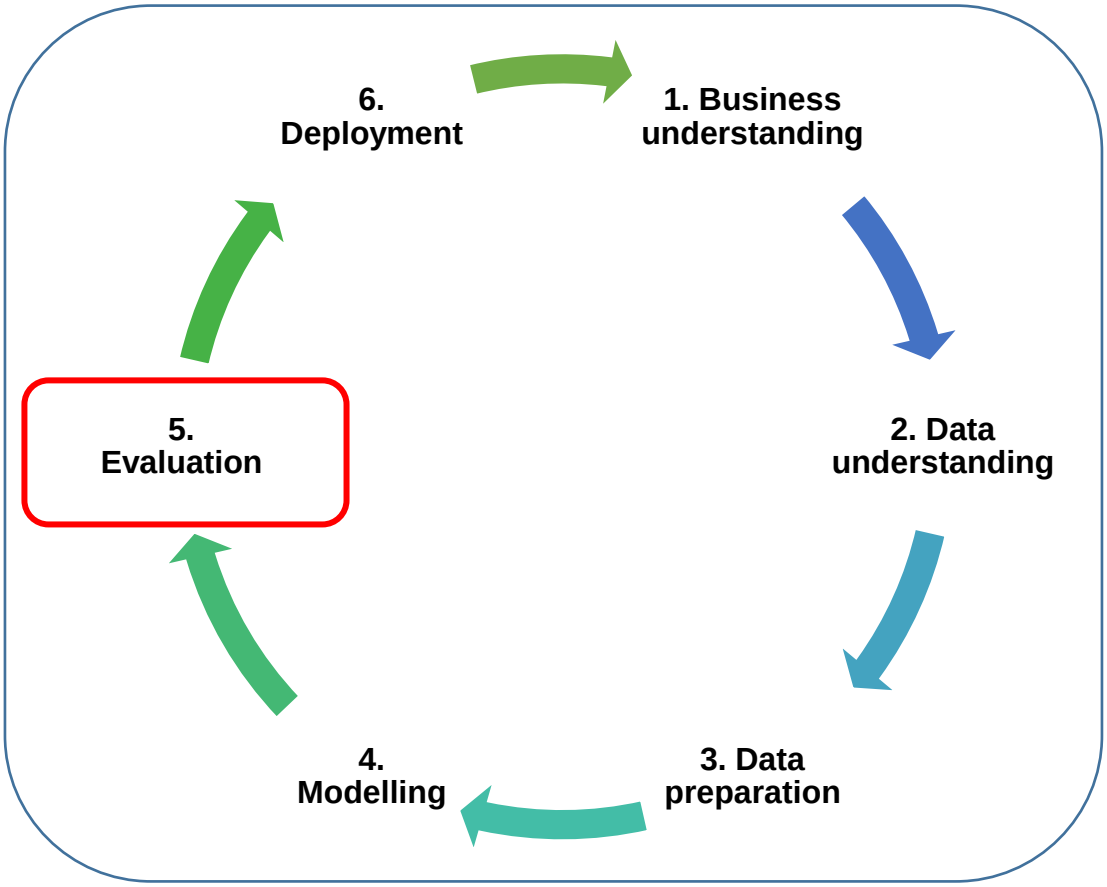
Use colour appropriately to help highlight critical data and improve the reader's understanding of the visuals. The appropriate use of colour can help the reader make connections between visuals and make data processing easier. When applying colours to reports and dashboards consider:

- Displaying most of the data using natural colours.
- Using bright or dark colours to highlight outliers or critical data.
- Using colour shades or gradients to show the spread of values across the data.
- Using colour to highlight variance around a central value. For example, use green to show positive values and red to show negative values.

13.6 EVALUATING RESULTS

Concepts

With the results of the data analytical processes easily accessible through reports and dashboards, you and your team can collaborate on the evaluation of the results in the evaluation phase of the data analysis process.



Data Analysis Phases

Some questions you might want to raise in the evaluation process:

- What insights have you gained from the process?
- Do the insights address the problem you set out to solve or the goals you set out to achieve?
- What are the practical implementation steps that can be taken to address the problem or achieve the goals?
- What might be the next data analytical phase?

If you find that you haven't solved the original problem or met the original objectives, you can revisit and review the business understanding and data understanding phases to make the most of the on-going data analysis. The key to gaining insights is not always getting more data but gathering the right data to address your needs. And the dynamic nature of business and its continuous development means that process of data analysis is always ongoing.

13.7 REVIEW EXERCISE

1. Which one of the following is used for sharing analytical reports?
   a) Power BI service.
   b) Power BI Report Canvas.
   c) Power BI Tiles.
   d) Power BI Desktop.

2. Open the **Orders Metric** Power BI file in **Power BI Desktop**.
   a) Publish the file to **My Workspace**.
   b) Share the file with another user *(**Note**: Remember, you can only share files with other Power BI Pro users)*.
   c) Sign out from the **Power BI service**.

3. Which one of the following should you avoid to achieve a clean, uncluttered layout in your reports and dashboards?
   a) Divide the visualizations over multiple pages.
   b) Use borders or white space to section the visualizations.
   c) Add decorative elements such as background images.
   d) Align and distribute the visualizations.

4. Why should you use consistent fonts and colours to design your reports and dashboards?
   a) To differentiate between the headings and actual data.
   b) To accentuate a hierarchical structure.
   c) To emphasize critical data.
   d) To show the availability of various fonts.

Ref	ICDL Task Item	Location
1.1.1	Identify the main types of data analytics: descriptive, diagnostic, predictive, prescriptive, quantitative, qualitative.	*1.1 Types of Data Analytics.*
1.1.2	Outline the business benefits of data analytics: identifies patterns/trends, improves efficiency, supports decision making, presents information effectively.	*1.2 Business Benefits.*
1.1.3	Identify the main phases of data analysis: business understanding, data understanding, data preparation, modelling, evaluation, deployment.	*1.3 Data Analysis Process.*
1.1.4	Recognise data protection considerations when analysing data like: anonymise personal data if possible, comply with applicable data protection regulations.	*1.4 Data Protection Considerations.*
1.2.1	Describe measures of central tendency of a data set: mean, median, mode.	*2.1 Summary Statistics Introduction* *2.2 Measures of Central Tendency.*
1.2.2	Calculate the central tendency value of a data set using a function: mean, median, mode.	*2.3 Calculating Central Tendency.*
1.2.3	Describe measures of variation of a data set: quartiles, variance, range.	*2.1 Summary Statistics Introduction* *2.4 Measures of Variation.*
1.2.4	Calculate the variation of a data set: quartile, variance, range.	*2.5 Calculating Variation.*
2.1.1	Import data into a spreadsheet application: .csv file, spreadsheet, website table, database table.	*3.1 Importing Data Sets Introduction* *3.2 Importing Data from Text Files.* *3.3 Importing Data from Spreadsheets.* *3.4 Importing Data from Website Tables.* *3.5 Importing Data from Database Tables.*
2.1.2	Remove duplicate data.	*4.2 Removing Duplicate Data.*
2.1.3	Validate that given values belong to a reference data set using the VLOOKUP function.	*4.3 Validating Data Using VLOOKUP.*
2.1.4	Validate that given values belong to a specified range using one or more if functions.	*4.4 Validating Data Using IF Functions.*
2.1.5	Extract values from a string using text functions: left, right, len, mid, find.	*4.5 Extracting Values Using Text Functions.*
2.2.1	Format a data set as a built-in table.	*5.1 Formatting Data Sets as Tables*
2.2.2	Insert and use table slicers.	*5.2 Using Table Slicers.*
3.1.1	Change the method of aggregation for a value: sum, average, count, minimum, maximum.	*6.2 Changing Aggregation Methods.*
3.1.2	Display multiple aggregation values.	*6.3 Displaying Multiple Aggregation Values.*
3.1.3	Display values as: % calculation, difference from specific values, running total, ranked.	*6.4 Using Built-In Calculations.*

Ref	ICDL Task Item	Location
3.2.1	Automatically, manually group data and rename groups.	*7.1 Grouping Date, Time and Numeric Data.* *7.2 Creating Custom Groups.*
3.2.2	Ungroup data.	*7.3 Ungrouping Data.*
3.3.1	Use the report filter.	*8.1 Using Report Filters.*
3.3.2	Insert and use slicers to filter single, multiple pivot tables.	*8.2 Using Slicers.*
3.3.3	Insert and filter a timeline.	*8.3 Using Timelines.*
3.4.1	Insert a pivot chart for an existing pivot table.	*9.1 Inserting Pivot Charts from Pivot Tables.*
3.4.2	Create a pivot chart from fields in a data set.	*9.2 Creating Pivot Charts from Tables.*
4.1.1	Understand the concept of data visualization using reports and dashboards. Outline common visualizations like: charts, key performance indicators (KPIs), maps.	*10.1 Key Features of Data Visualization Tools* *11.2 Creating Chart Visualizations.* *11.5 Creating Visualizations using Maps* *12.1 Creating Visualizations to Measure Progress*
4.1.2	Recognise common data visualization tools and their functions like: visualise data, publish and share business intelligence.	*10.1 Key Features of Data Visualization Tools* *10.2 Data Visualization Tools Setup.*

Ref	ICDL Task Item	Location
4.1.3	Understand good design practice in reports and dashboards like: clean and uncluttered layout, descriptive titles, consistent fonts and colour, use of colour for emphasis and understanding.	*13.5 Good Design Practice.*
4.1.4	Import a data set from a spreadsheet into a data visualization tool and save the file.	*10.4 Importing Data Sets from Spreadsheets.*
4.2.1	Create tables in a report.	*11.1 Creating Table Visualizations.*
4.2.2	Visualize data as a chart: column, bar, line, pie.	*11.2 Creating Chart Visualizations.*
4.2.3	Apply, edit font and background conditional formatting to show: high/low values, above/below average values.	*11.3 Enhancing Visualizations Using Conditional Formatting.*
4.2.4	Apply, edit data bars.	*11.3 Enhancing Visualizations Using Conditional Formatting.*
4.2.5	Apply, edit visual level filters.	*11.4 Enhancing Visualizations Using Visual Level Filters.*
4.3.1	Publish a report.	*13.1 Creating and Preparing to Share Reports.* *13.2 Publishing and Sharing Reports.*
4.3.2	Create a dashboard.	*13.3 Creating Dashboards.*
4.3.3	Share a report, dashboard using a link. Share a report to web.	*13.1 Creating and Preparing to Share Reports.* *13.2 Publishing and Sharing Reports.* *13.4 Sharing Dashboards.*

Congratulations! You have reached the end of the ICDL Data Analytics - Foundation book.

You have learned about the key skills relating to Data Analytics - Foundation including:

- Understanding the key concepts relating to the application of data analytics in business.
- Understanding and applying key statistical analysis concepts.
- Importing data into a spreadsheet and preparing it for analysis using data cleansing and filtering techniques.
- Summarising data sets using pivot tables and pivot charts.
- Understanding and applying data visualisation techniques and tools.
- Creating and sharing reports and dashboards in a data visualisation tool.

Having reached this stage of your learning, you should now be ready to undertake an ICDL certification test. For further information on taking this test, please contact your ICDL test centre.

Digital Fortress Global Services Limited
26 Shasha Road, Akowonjo Roundabout
Dopemu, Lagos State
Nigeria
icdlafrica.org