

Intro to Data Analytics

Basic Business Analytics

1	KEY METRICS ANALYSIS	1
1.1	PRACTICAL EXERCISE FOR KEY METRICS ANALYSIS.....	8
2	COMPARISON ANALYSIS.....	9
2.1	PRACTICAL EXERCISE FOR COMPARISON ANALYSIS.....	14
3	TREND ANALYSIS	15
3.1	PRACTICAL EXERCISE FOR TREND ANALYSIS.....	21
4	RANKING ANALYSIS	22
4.1	PRACTICAL EXERCISE FOR RANKING ANALYSIS	27
5	INTERACTIVE DASHBOARD	28
6	VARIANCE ANALYSIS.....	31
6.1	PRACTICAL EXERCISE FOR VARIANCE ANALYSIS	40
7	CONTRIBUTION ANALYSIS	42
7.1	PRACTICAL EXERCISE FOR CONTRIBUTION ANALYSIS	44
8	FREQUENCY ANALYSIS.....	44
8.1	PRACTICAL EXERCISE FOR FREQUENCY ANALYSIS.....	46
9	CORRELATION ANALYSIS	46
10	PARETO ANALYSIS	48

1 Key Metrics Analysis

Key Metrics or Key Performance Indicators (KPIs) are the main method that we use to measure the performance of a company, division, department or team. Key Metrics provide an overall understanding of how the business is performing. Key Metrics usually include the following types of questions:

- a) What is the Total Sales?
- b) What is the Total Profit?
- c) What is the Profit Ratio?
- d) How many transactions were there processed?

e) What was the average Sales amount?

f) What was the highest Sales amount?

g) What was the lowest Sales amount?

File to use: SalesData-v1.xlsx

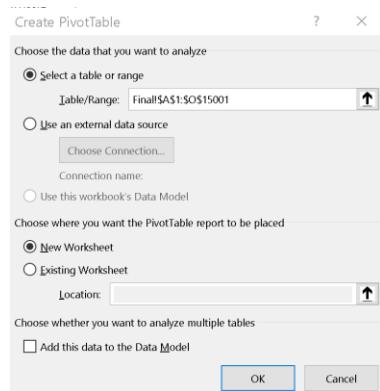
Create a Pivot table from the current range of data.

Make sure you have selected a cell in the range.

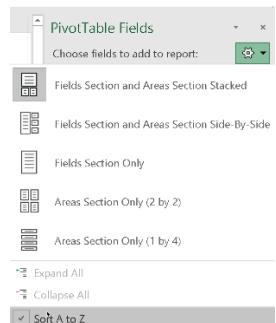
Insert -> Pivot table

The Create Pivot table should automatically detect the complete range of cells with valid values.

Make sure you create the Pivot table in a new worksheet.



You can change the order of the Pivot Table fields as they appear in the main list: it's sometimes useful to sort them in alphabetical order if you have many fields.



Select the Sales field to get the Sum of Sales, then perform a Number Format to make it easier to view.

Repeat this with the Profit field to get the Sum of Sales, with a similar Number Format as well.

Sum of Sales	Sum of Profit
55,391,760	35,589,375

Next insert a Calculated Field to obtain the Profit Ratio, and format the result accordingly

The screenshot shows two overlapping windows. On the left is the 'Insert Calculated Field' dialog box. It has fields for 'Name:' (Profit Ratio) and 'Formula:' (= Profit / Sales). Below these are dropdown menus for 'Fields:' containing various dataset columns like Order Date, Year, Order Qty, Cost of Sales, Sales, Profit, Channel, and Product Name. A blue bar highlights the 'Sales' field in this list. At the bottom are 'OK' and 'Close' buttons. On the right is the 'Format Cells' dialog box, specifically the 'Number' tab. Under 'Category:', 'Percentage' is selected, and the 'Sample' value is '64.3%'. The 'Decimal places:' dropdown is set to 1. Both dialogs are overlaid on a spreadsheet background with columns I through N visible.

Sum of Sales	Sum of Profit	Sum of Profit Ratio
55,391,760	35,589,375	64.3%

To count the number of transactions, we would typically use a unique Sales ID, which we don't have here in this dataset. We can initially use the Sales Field to start with (Drag and drop into the Values area), but this time change it to Summarize Values by Count to get the actual number of transactions (which we can also double confirm from the source dataset)

The screenshot shows a context menu open over a cell containing the value '15000'. The menu includes options like Copy, Format Cells..., Number Format..., Refresh, Sort, and Remove "Count of Sales2". Below the menu, a 'Summarize Values By' dialog box is open. It shows 'Summarize Values By' set to 'Sum' and 'Show Values As' set to 'Count'. To the right, the same table as in the previous screenshot is shown, with an additional column 'Count of Sales2' added, which contains the value '15000'.

To get the average Sales, use the Sales field again, but this time Use Summarize Values by Average, and format it accordingly.

Average of Sales2

3692.78

Copy
Format Cells...
Number Format...
Refresh
Sort >
Remove "Average of Sales2"
Summarize Values By > Sum
Show Values As > Count
Show Details > Average

Count of Sales2	Average of Sales2
15000	3,693

Add two more Sales Fields into the Values area, and then Summarize Values by Min and Max for these two fields and format them appropriately.

Max of Sales2

70212

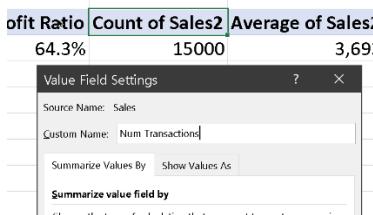
Copy
Format Cells...
Number Format...
Refresh
Sort >
Remove "Max of Sales2"
Summarize Values By > Sum
Show Values As > Count
Show Details > Average
Value Field Settings... > Max

Min of Sales3

70212

Copy
Format Cells...
Number Format...
Refresh
Sort >
Remove "Min of Sales3"
Summarize Values By > Sum
Show Values As > Count
Show Details > Average
Value Field Settings... > Min

Now that we have all our numbers, you can double click on any (or all) of the field headers to change the field names to something more appropriate if you wish.



We now have our key metrics listed out in a row:

Sum of Sales	Sum of Profit	Sum of Profit Ratio	Num Transactions	Average of Sales	Highest Sales	Lowest Sales
55,391,760	35,589,375	64.3%	15000	3,693	78312	5

You can transport the rows and columns for a better layout:

Values	
Total Sales	55,391,760
Total Profit	35,589,375
Actual Profit Ratio	64.3%
Num Transactions	15000
Average of Sales	3,693
Highest Sales	78312
Lowest Sales	5

To zoom in and get more details on the metrics based on a certain category (for e.g. country), we can filter the table to decide which subset of records to apply the various aggregation operations we have performed so far on.

For this purpose, we can use a slicer. Go to Pivot Table Analyze -> Insert Slicer, and select Country. You can then select particular countries from the slicer drop down to get the metric values only for sales records related to that particular country.

Values	
Total Sales	1,148,316
Total Profit	694,190
Actual Profit Ratio	60.5%
Num Transactions	359
Average of Sales	3,199
Highest Sales	45900
Lowest Sales	37

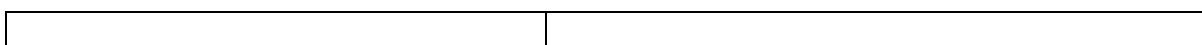
You can use the Slicer Tools options to adjust the number of columns, as well as their height and width to make them easier to work with:

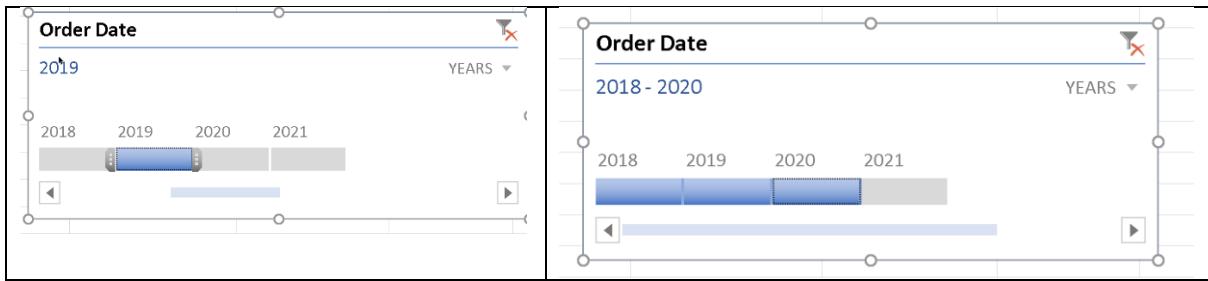
Country	Country	Country
Armenia	Australia	Bhutan
Canada	China	Denmark
France	Germany	Greece
India	Iran	Ireland
Italy	Japan	Kyrgyzstan
Malta	Pakistan	Poland
Portugal	Romania	Russia
Singapore	Slovenia	South Korea

You can use the Ctrl-Key to select multiple countries at the same time, so the aggregation operations will apply to all these countries. You can then select the clear Filter button to clear the filter and make the aggregation apply to the entire dataset.

You can now also insert a Timeline and decide which particular period (Years, Quarters, Months) to examine.

You can select a particular period or a range of periods (Shift-Click):





You can also select items from both filters to combine their filtering action together, for e.g. to perform aggregation operations on a specific country (or group of countries) over a specific period of time.

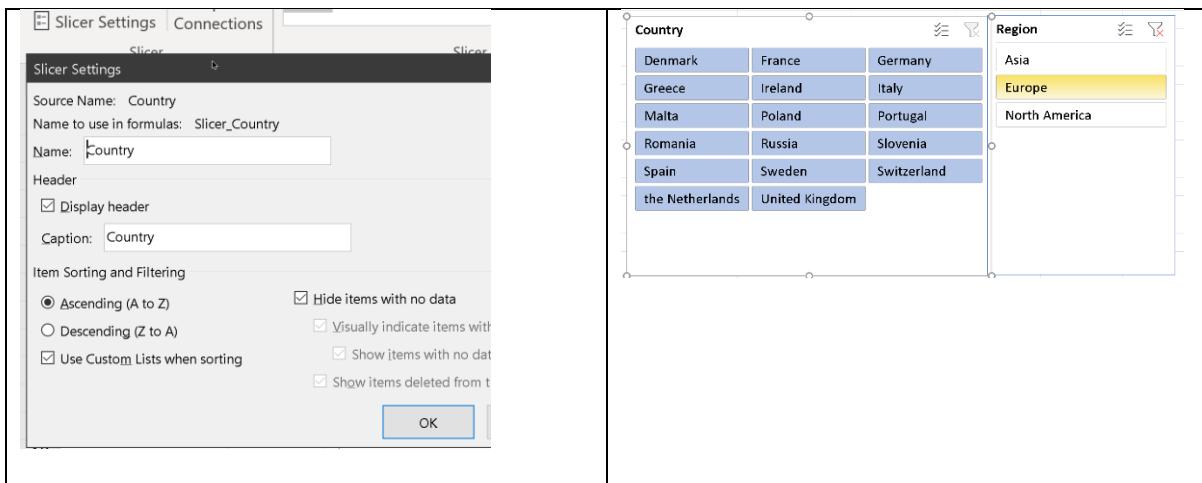
This composite image shows a 'Order Date' Slicer on the left and a 'Country' Slicer on the right. The 'Order Date' slicer is set to 'QUARTERS' and shows the period 'Q1 2018 - Q3 2019'. It includes a timeline from 2018 to 2020 with specific quarters labeled: Q1, Q2, Q3, Q4 for 2018; Q1, Q2, Q3, Q4 for 2019; and Q1, Q2 for 2020. The 'Country' slicer on the right lists various countries in three columns: Armenia, Australia, Bhutan; Canada, China, Denmark; France, Germany, Greece; India, Iran, Ireland; Italy, Japan, Kyrgyzstan; Malta, Pakistan, Poland; Portugal, Romania, Russia; and Singapore, Slovenia, South Korea. The country 'Japan' is highlighted in blue in both the timeline and the slicer.

You can introduce another slicer for Region, and notice that selection of items in either slicer for Country or Region will influence the other: based on which countries are included in which region. To see this properly in effect, you will have to clear the filter in either one or both of the slicers.

This figure displays two Slicer controls. The left Slicer is for 'Country' and the right Slicer is for 'Region'. The 'Country' slicer lists countries in three columns: Denmark, France, Germany; Greece, Ireland, Italy; Malta, Poland, Portugal; Romania, Russia, Slovenia; Spain, Sweden, Switzerland; the Netherla..., United Kingd..., Armenia; Australia, Bhutan, Canada; and China, India, Iran. The 'Region' slicer lists regions in three columns: Asia, Europe, North America; Armenia, Australia, Bhutan; China, India, Iran; Japan, Kyrgyzstan, Pakistan; Singapore, South Korea, Syria; Taiwan, Thailand, Turkmenistan; Canada, Denmark, France; Germany, Greece, Ireland; and Italy, Malta, Poland. In both cases, the region 'Europe' is highlighted in blue, indicating it is the active filter.

You can also set the Slicer Settings to only clearly show the countries within a particular region, and not just grey out the countries that are not in that region (to make the analysis even more clearer):





1.1 Practical Exercise for Key Metrics Analysis

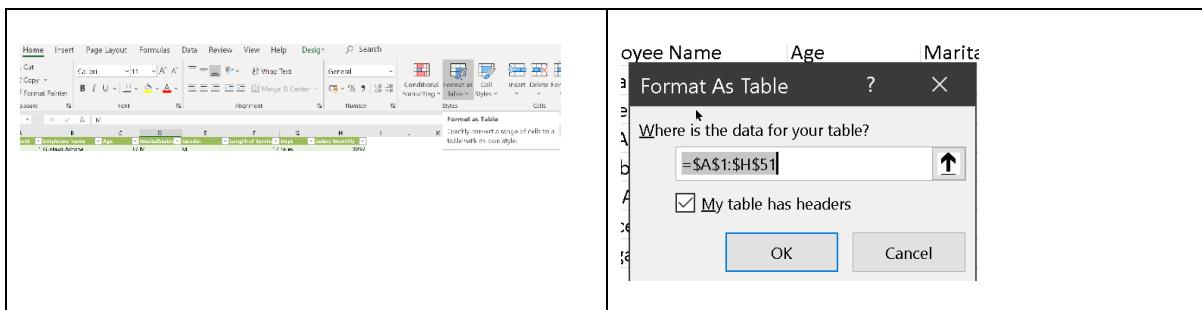
File to use: Employee-v1.xlsx

Analytical activities to perform:

1. How many employees are in John's company?
2. What is the average age of the employees?
3. What is the highest age at the company?
4. What is the lowest age at the company?
5. What is the average Length of Service?
6. What is the longest Length of Service?

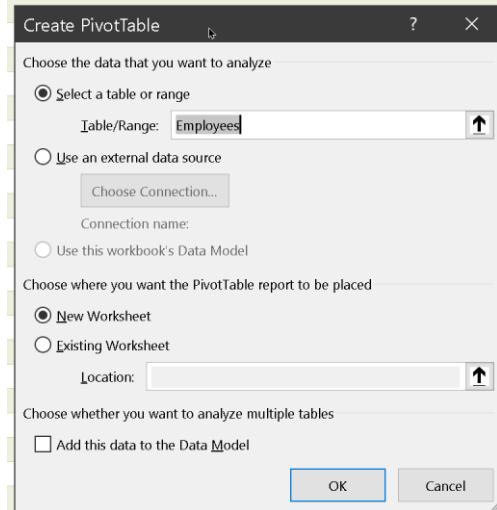
Drill down further into this information based on the following categories: Dept and Gender

You can initially format the range of cells of the original data set as a table to make it easier to work with, and to create a Pivot Table. To do this, select a cell in the data range, go to Home -> Format as Table, format it and give it an appropriate name.



EmployeeID	Employee Name	Age	MaritalStatus	Gender	Length of Service	Dept	Salary Monthly
1	Gustave Achong	37	M		17	Sales	3297
2	Catherine Abel	32	F		9	Sales	1859
3	Kim Abercrombie	54	M		11	Finance	4605
4	Humberto Acevedo	54	S	M	12	Logistics	1801
5	Markie Adams	40	M	P	20	Customer Relati	3816
6	Frances Adams	51	S	F	14	Sales	1830
7	Margaret Smith	73	S	F	0	Sales	3466
8	Carla Adams	73	M	F	7	Finance	3446
9	Jay Adams	26	M	M	13	Finance	4275
10	Ronald Adina	53	S	M	9	Logistics	4883

Once done, you can generate a Pivot Table in the usual way, but this time referencing the Table.



2 Comparison Analysis

Comparison analysis is probably one of the most popular forms of analysis you're going to do. The focus is finding the magnitude of difference between values for different fields. Visualizing your data set makes it much easier to appreciate this as opposed to directly examining the values themselves.

Sample analysis

- a) compare the sales and profit by year
- b) compare the sales by manufacturer
- c) compare the profit by product categories
- d) compare the sales by product category and by channel
- e) compare the profit by product subcategory and review that by different countries.
- f) compare the sales and profit by the countries, and review that by different product sub categories

File to use: SalesData-v2.xlsx

First, get the Sum of Sales and Profit by Years (Rows) and format the cells with an appropriate numeric format.

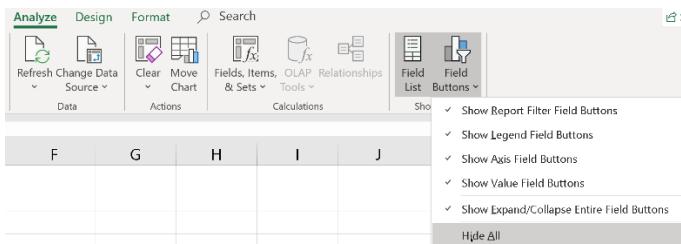
	Row Labels	Sum of Sales	Sum of Profit
2018	18,919,151	12,057,185	
2019	17,741,637	11,378,957	
2020	17,284,124	11,237,878	
2021	1,446,849	915,355	
Grand Total	55,391,760	35,589,375	

We can also sort on the Sum of Sales or Profit (either Smallest to Largest or vice versa)

The sorting can also be done via the Sort option in the Data tab

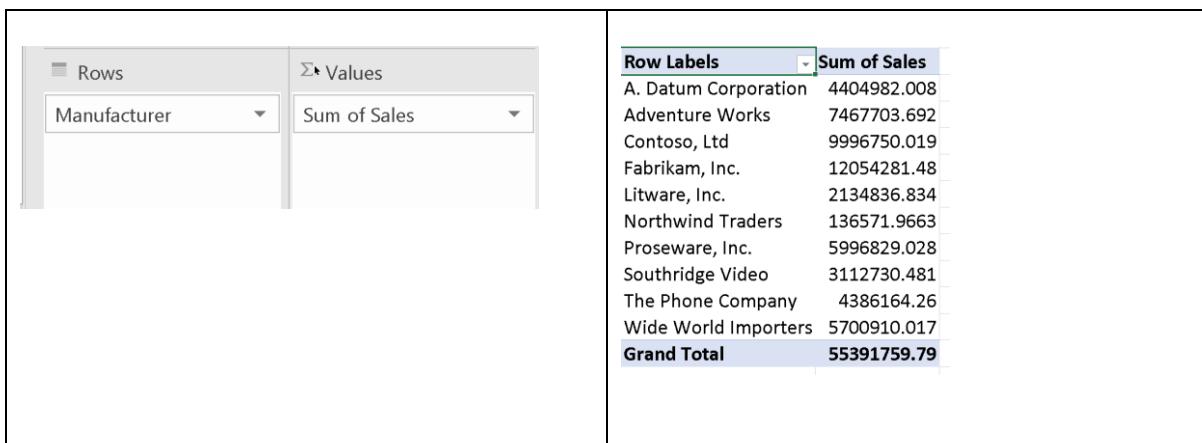
Sort in descending order for Sum of Sales, and visualize this via a Pivot Chart (choose Clustered Column)

You can format what fields to be shown in the chart, including hiding everything:

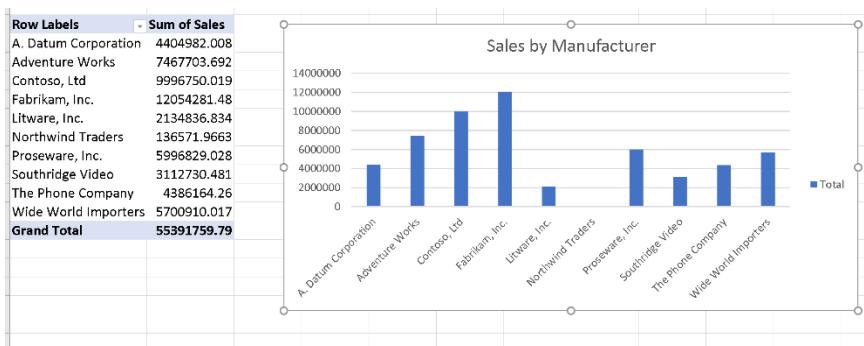


You can also change the Chart Elements appropriately, for e.g. add in Title and moving the legend to the Top.

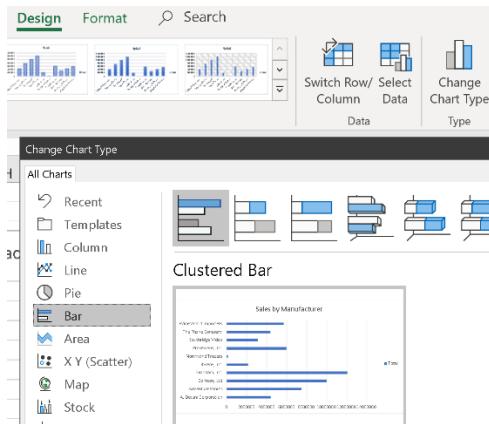
We can now (using either the current pivot table or creating a new pivot table) repeat this earlier process to get the Sum of Sales by Manufacturer (Rows) and format the cells with an appropriate numeric format.



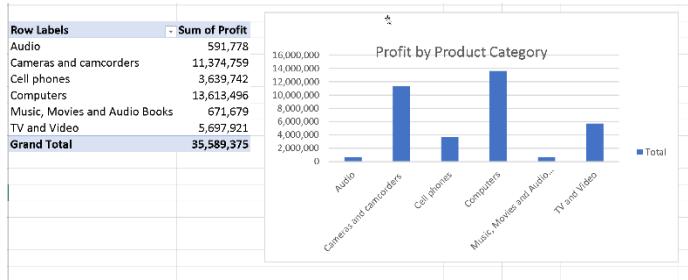
We can now insert another Clustered Column chart, and format it in the same way as we did previously.



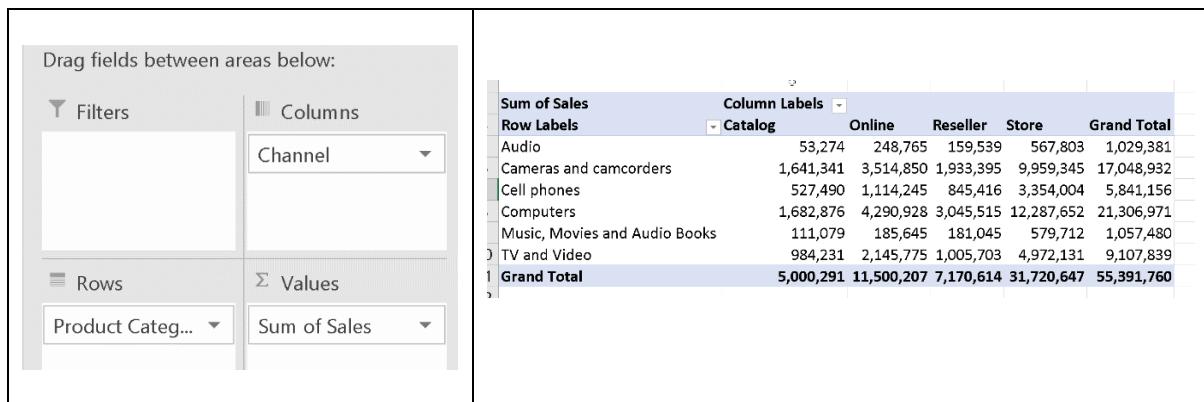
We can also opt to change the Chart Type to another type for example. Clustered Bar.



For the next analysis, we can (using either the current pivot table or creating a new pivot table) repeat this earlier process to get the Sum of Profit by Product Category (Rows) and format the cells with an appropriate numeric format, then generate the Pivot Chart (either Clustered Column or Clustered Bar).

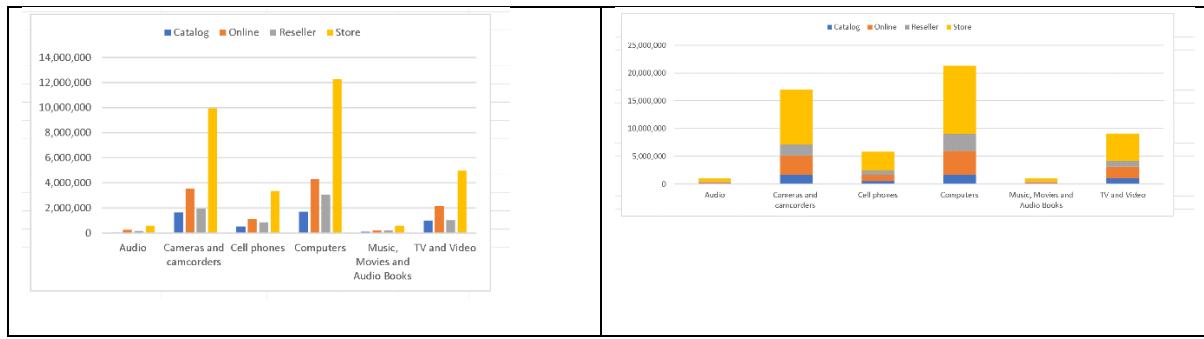


For the next analysis, we can (using either the current pivot table or creating a new pivot table) repeat this earlier process to get the Sum of Sales by Product Category (Rows) and Channel (Columns) and format the cells with an appropriate numeric format.

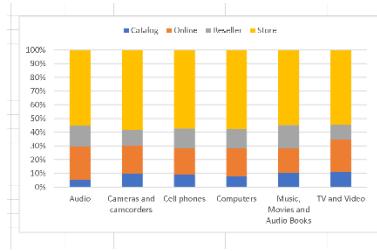


Then we could use either or both a Clustered Column chart and / or Stacked Column chart to view the differences.





We could also use the a 100% stacked column as well, which is useful for seeing how each of the different channels contribute for all products, regardless of the product total - this is useful for products whose sales totals are very small compared to other products and therefore will not come out clearly in the previous graphs.

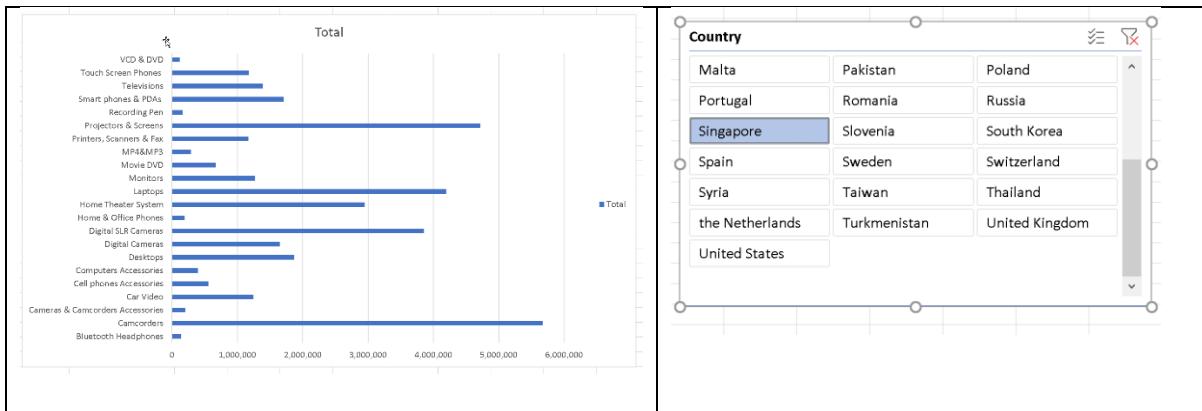


For the next analysis, we can (using either the current pivot table or creating a new pivot table) repeat this earlier process to get the Sum of Profit by Product Sub Category (Rows) and format the cells with an appropriate numeric format.

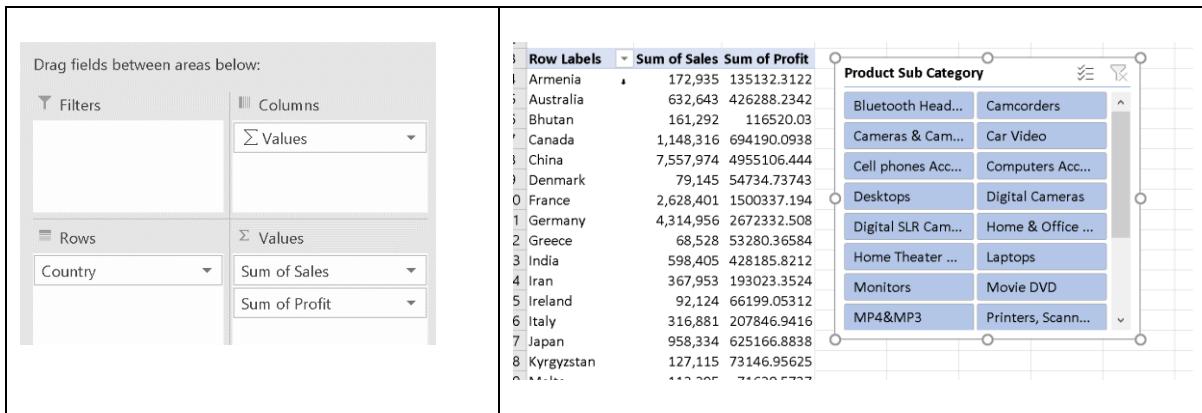
Axis (Categories)	Σ Values
Product Sub Category	Sum of Profit
Bluetooth Headphones	142,887
Camcorders	5,670,098
Cameras & Camcorders Accessories	202,945
Car Video	1,243,547
Cell phones Accessories	558,835
Computers Accessories	395,441
Desktops	1,870,700
Digital Cameras	1,647,076
Digital SLR Cameras	3,854,640
Home & Office Phones	198,426
Home Theater System	2,946,634
Laptops	4,193,350
Monitors	1,268,828
Movie DVD	671,679
MP4&MP3	288,293
Printers, Scanners & Fax	1,167,342
Projectors & Screens	4,717,835
Recording Pen	160,597
Smart phones & PDAs	1,708,954
Televisions	1,388,547
Touch Screen Phones	1,173,528
VCD & DVD	118,793
Grand Total	35,389,375

Then we generate a standard Clustered Bar Chart as we have done before, and now we can also choose to use a slicer based on the Country field for the pivot table which will also dynamically affect the content of the chart which is linked to the same table.





For the next analysis, we can (using either the current pivot table or creating a new pivot table) repeat this earlier process to get the Sum of Sales and Profit by Country (Rows) and format the cells with an appropriate numeric format. Then use Product Subcategory as a slicer



2.1 Practical exercise for comparison Analysis

File to use: File to use: SalesData-v3.xlsx

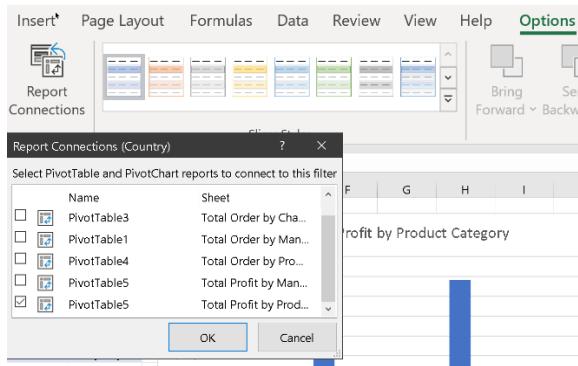
Analytical activities to perform:

1. What is the Total Order Quantity by Manufacturer?
2. What is the average Profit by Brand
3. Create a column graph displaying Total Order Quantity sold by Channel - review by Country
4. Create a bar graph displaying the Total Order quantity sold by Product Category - review by Region

Create two pivot graphs as per the instructions below. Filter both graphs by the Country and Year fields using a slicer.

1. Create a column graph displaying Total Profit by Product Category
2. Create a bar graph displaying Total Profit by Manufacturer

Note: While creating slicers, you can link a single slicer to multiple pivot tables on the same worksheet (if you decide to have more than one), using Report Connections in the Slicer Options.



3 Trend Analysis

Trend analysis is all about understanding how data changes over time: is it moving up, is it moving down, is it staying stable or is it volatile. This can be accomplished mainly via trendlines. We also want to study these changes over different time frames: for e.g. changes over years, quarters, months, days. Again, this is best understood with some form of visualization.

Activities to be undertaken:

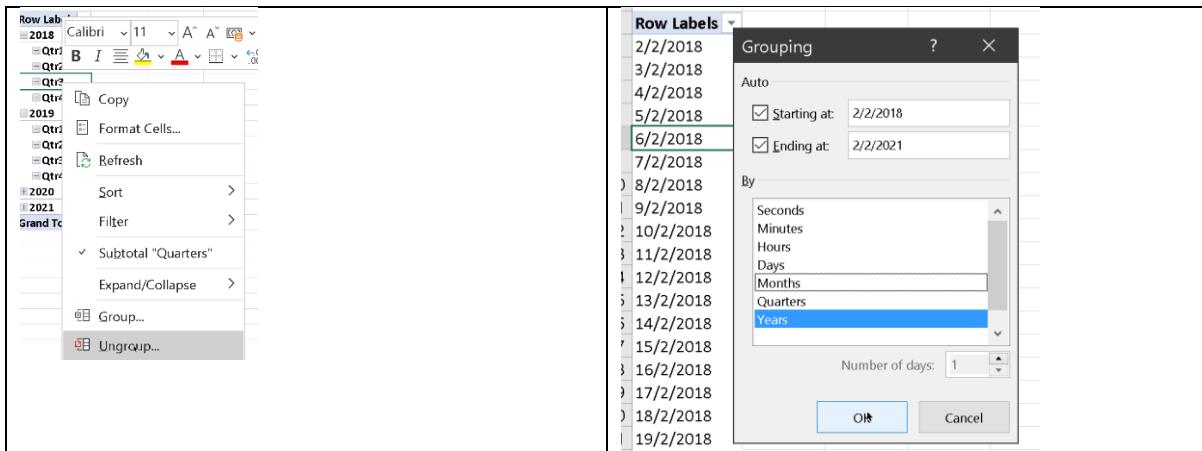
- Understand the trend for sales by year and month
- Understand the trend for sales by year and month by different product categories
- Develop a seasonality graph displaying the sales by month

File to use: SalesData-v4.xlsx

Adding Order Date to the Rows area results in Excel automatically creating a hierarchy involving Years and Quarters.

You can ungroup the hierarchy first before creating your own custom grouping





We can then continue to get Sum of Sales for each year

Row Labels	Sum of Sales
2018	18,919,151
2019	17,741,637
2020	17,284,124
2021	1,446,849
Grand Total	55,391,760

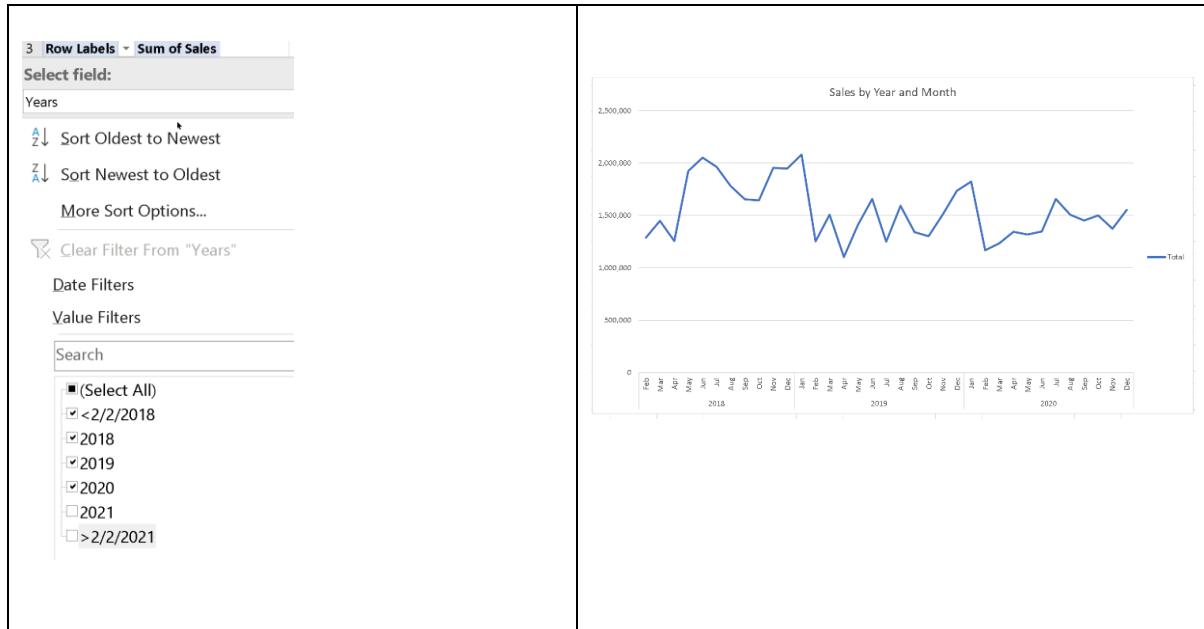
We can then alternatively change to another custom grouping (years / months)

Row Labels	Sum of Sales
2018	18,919,151
Feb	1,289,552
Mar	1,448,250
Apr	1,256,708
May	1,923,673
Jun	2,051,973
Jul	1,965,917
Aug	1,781,769
Sep	1,654,111
Oct	1,643,812
Nov	1,955,526
Dec	1,947,858
2019	17,741,637
Jan	2,081,908

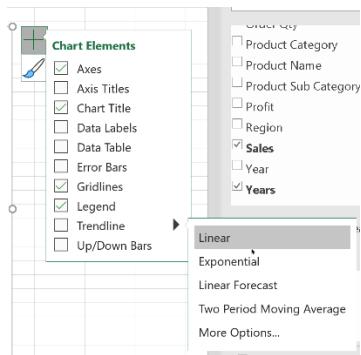
We will use a Pivot Chart to generate a Line Chart.



We can also filter out the first 2 months of 2021 to avoid issues with insufficient data at that point which will affect the trend analysis.

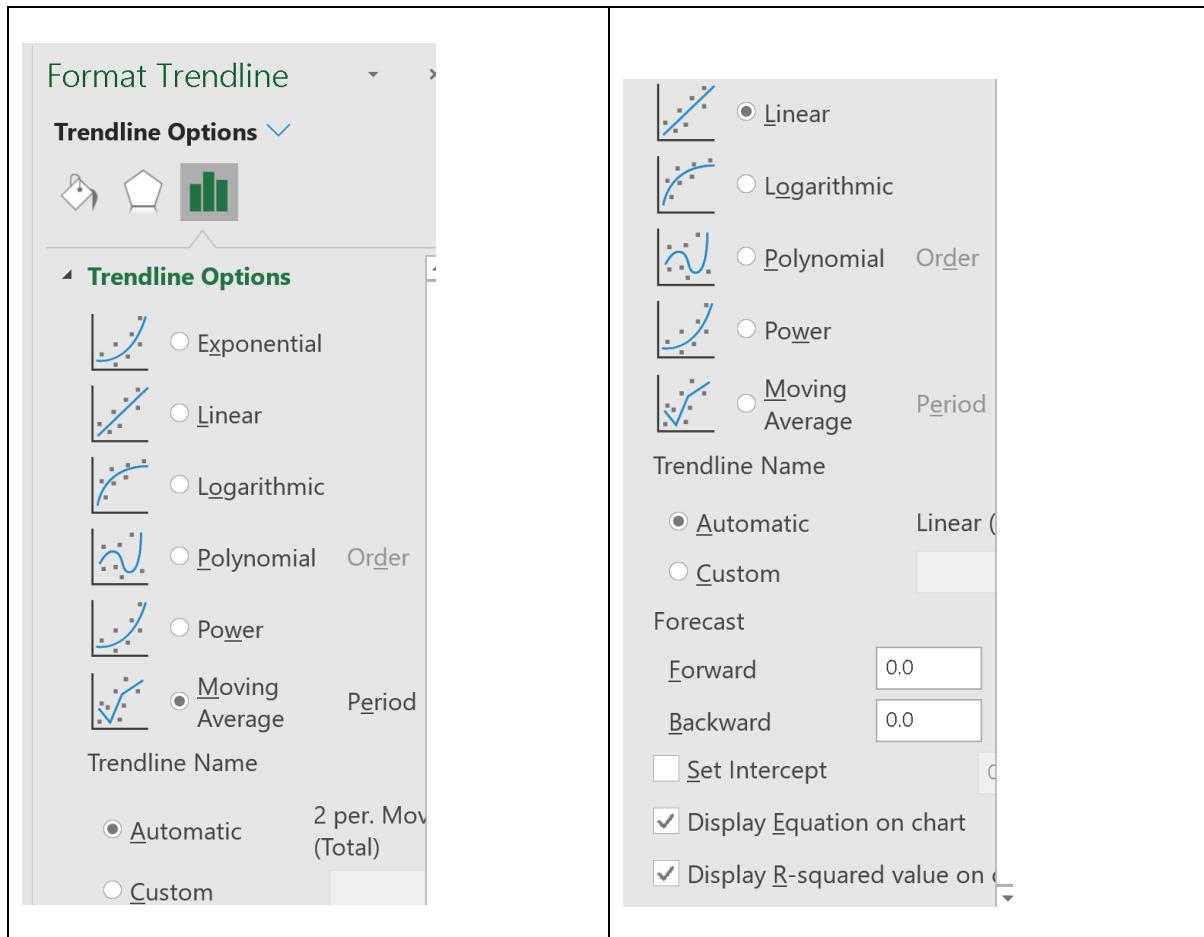


With a more proper graph, you can examine the various trend line options possible:



The first option is the default option which is linear. The rest (exponential, linear forecast) are quite similar, but not exactly the same. They are both using different algorithms to calculate what the trend line is. The two period moving average is basically taking two periods and it averages that value between the two periods. It tends to smooth the lines. It's not a trend line as such but tracking the shape of the sales lines.

In the format trend line, there are a variety of options such as polynomial, exponential, etc.

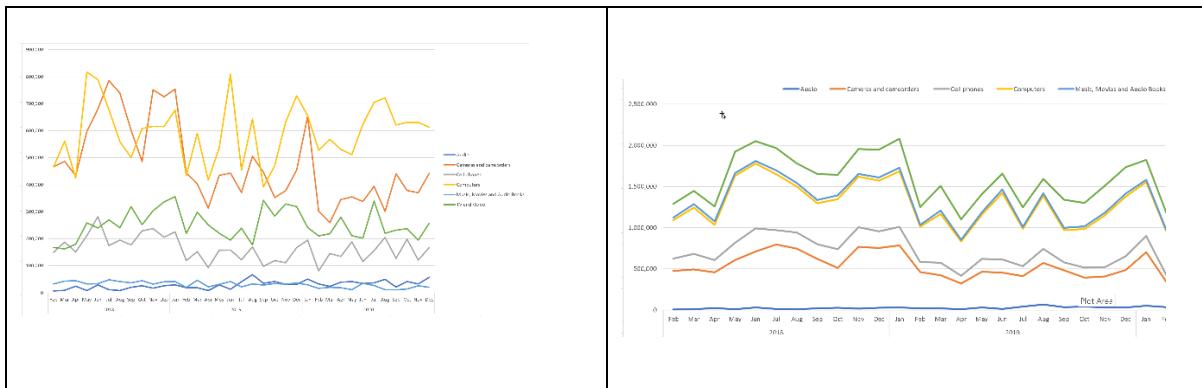


For a linear equation, you can display the trend line equation on the chart and calculate the R-squared, which is a measure of how accurate the predictions would be from the trend line equation. This is typically ranging from 0 to 1 (perfect accuracy) and is used in quantifying accuracy of multiple linear regression models.

The next step is to refine this further to understand the trend for sales by year and month by different product categories. We can simply copy the existing pivot table to a new worksheet and work from there.

Filters		Columns		Sum of Sales									
		Product Category		Column Labels									
		Row Labels		Sum of Sales									
Years	Order Date	Audio	Cameras and camcorders	Cell phones	Computers	Music, Movies and Audio Books	TV and Video	Grand Total					
2018		182,424	6,750,412	2,196,698	6,635,135	426,434	2,728,048	18,919,151					
Feb		6,504	456,640	149,413	466,647	32,341	167,806	1,289,552					
Mar		8,254	486,074	187,716	561,066	42,948	162,191	1,448,250					
Apr		24,332	433,028	150,856	424,247	44,795	179,451	1,256,708					
May		7,940	597,172	210,988	816,645	32,167	258,762	1,923,673					
Jun		29,136	678,810	281,364	789,193	32,827	239,643	2,051,973					

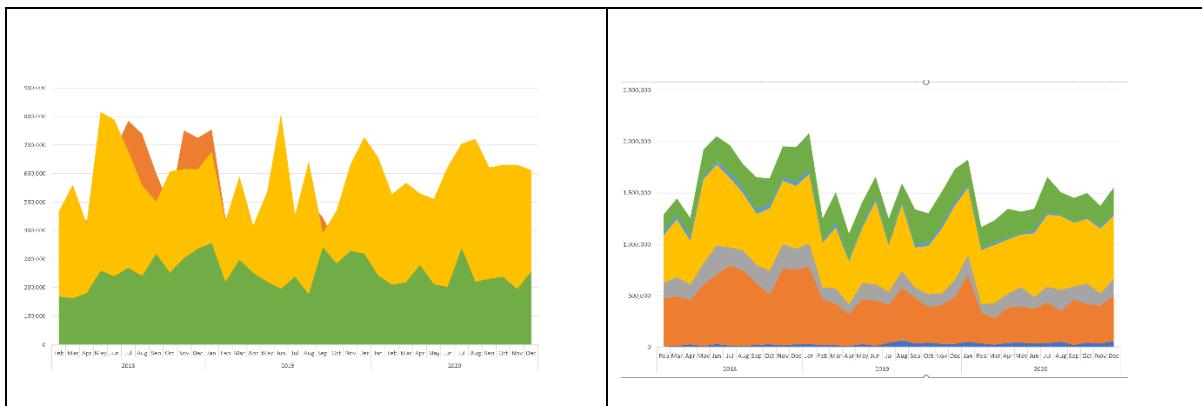
We can add a PivotChart that is a line chart and stacked line chart to get an initial view of this data.



You can select which particular product category to project a trend line on if you wish.



We can also change this to either a area graph and stacked area graph to get a better view of the contributions of each product category trending over time.



You could also take your Product Category out of the Columns and utilize it as a slicer instead, which then allows you to view the trends for each of these categories across the different chart types.



Row Labels ▾ Sum of Sales	
2018	182,424
Feb	6,504
Mar	8,254
Apr	24,332
May	7,940
Jun	29,136
Jul	10,904
Aug	7,446
Sep	19,870
Oct	25,737
Nov	16,790
Dec	25,510
2019	359,809
Jan	29,214
Feb	18,169

The next thing we will look is a seasonality graph, which we can create from another copy of the pivot table on a new worksheet. For seasonality, we typically want to take into account the years which have data for all the months, and filter out those that do not.

Row Labels ▾ Sum of Sales	
3 Row Labels ▾ Sum of Sales	
Select field:	
Years	
A↓ Sort Oldest to Newest	
Z↓ Sort Newest to Oldest	
More Sort Options...	
Clear Filter From "Years"	
Date Filters	>
Value Filters	>
Search	🔍
✓ (Select All)	
□ <2/2/2018	
□ 2018	
✓ 2019	
✓ 2020	
□ 2021	
□ >2/2/2021	
Row Labels ▾ Sum of Sales	
2019	17,741,637
Jan	2,081,908
Feb	1,252,298
Mar	1,507,751
Apr	1,103,451
May	1,408,131
Jun	1,657,980
Jul	1,248,653
Aug	1,593,694
Sep	1,340,433
Oct	1,300,579
Nov	1,511,809
Dec	1,734,950
2020	17,284,124
Jan	1,824,856
Feb	1,168,697
Mar	1,234,561
Apr	1,345,859
May	1,318,429
Jun	1,347,578
Jul	1,656,039
Aug	1,507,838
Sep	1,452,020
Oct	1,499,587
Nov	1,374,350
Dec	1,554,311
Grand Total	35,025,760

Next change custom grouping to Month. This will combine data from all the years concerned (2019, 2020).

--	--

The screenshot shows two windows side-by-side. On the left is a 'Grouping' dialog box for a 'Sum of Sales' measure. It has 'Starting at' set to 2/2/2018 and 'Ending at' set to 2/2/2021. The 'By' dropdown is set to 'Months'. On the right is a table of monthly sales data from January 2019 to December 2020, with a total of 55,391,760.

Month	Sum of Sales
Jan	5,315,113
Feb	3,749,046
Mar	4,190,561
Apr	3,706,018
May	4,650,233
Jun	5,057,531
Jul	4,870,610
Aug	4,883,301
Sep	4,446,563
Oct	4,443,978
Nov	4,841,686
Dec	5,237,119
Grand Total	55,391,760

Then we can generate a line graph to demonstrate the Seasonality.



Seasonality is really useful to understand so that we can understand where resource / sales demand peaks in specific periods of a year and therefore can significantly help with resource allocation planning. Its important to make sure you have a complete set of data for the entire time duration you are interested to perform analysis on (for e.g. all the months of a year, rather than partial months) because that will skew the visualization of the graph.

3.1 Practical exercise for Trend Analysis

Analytical activities to perform:

- Create a line graph displaying the Profit by Year and Quarter
- Add a polynomial trend line to the line graph
- Add a Slicer for Product Category
- Create an area graph displaying the Profit by Year and Quarter for the Product Category
- Add a Slicer for Country

File to use: SalesData-v5.xlsx

For first solution a), we can specifically target the quarters which are outliers in the dataset (suggesting that the data might be incomplete) and remove them using a Date filter with a Between.

The screenshot shows the Power BI interface. On the left, the 'Date Filters' pane is open, showing various filter options like 'Between...', 'Today', and 'Yesterday'. A 'Between...' filter is selected, and its configuration dialog is shown on the right. The dialog title is 'Date Filter (Order Date)' and it contains the condition 'is between 01/04/2018 and 31/12/2020'. Below this, there is a checkbox for 'Whole Days' which is unchecked. At the bottom are 'OK' and 'Cancel' buttons. To the left of the filters, a table is displayed with data grouped by year and quarter, showing the sum of profit. The data is as follows:

	Sum of Profit
2018	10,266,628
Qtr2	3,352,068
Qtr3	3,431,388
Qtr4	3,483,171
2019	11,378,957
Qtr1	3,172,056
Qtr2	2,541,105
Qtr3	2,731,591
Qtr4	2,934,205
2020	11,237,878
Qtr1	2,705,240
Qtr2	2,588,481
Qtr3	3,037,253
Qtr4	2,906,905
Grand Total	32,883,463

4 Ranking Analysis

Ranking analysis is about understanding the order of your items. Typically you would like the item with the highest value to be ranked number one and the item with the lowest value to have the lowest rank. So the easiest way to actually do a ranking is to sort your items.

Alternative forms of ranking including taking a selected number of items at the top and bottom of a sorted list. So you may want to see only say, the top ten items or the bottom 20 items.

There are different ways to compute rankings and use them to affect the chronological order of how data is normally displayed (for e.g. months of a year).

Analytical activities to be performed:

- Display top 15 countries by sales
- Display the top five product subcategories by sales
- Display the products that contribute the top 20% of profit
- Display the rankings of the profit by year and month - review by Product Category
- Display the rankings for sales and profit by different countries

File to use: SalesData-v6.xlsx

The easiest way to do a ranking is a sort of the field of interest (Sum of Sales for e.g.)

The screenshot shows the Power BI Data View ribbon. A context menu is open over a table of data, specifically over the 'Sum of Sales' column. The menu includes options like 'Copy', 'Format Cells...', 'Number Format...', 'Refresh', 'Sort', 'Remove "Sum of Sales"', 'Summarize Values By', and 'More Sort Options...'. The 'Sort' option is highlighted.

The Value Filters for a given Row Label field provides common options such as getting the top / bottom X items (rather than the entire list)

The screenshot shows two dialog boxes. On the left is the 'Filter' dialog for the 'Country' field, showing various filter options like 'Sort A to Z', 'Value Filters', and a list of countries. On the right is the 'Top 10 Filter (Country)' dialog, which allows setting the top number of items (15), the sort order ('Items by Sum of Sales'), and buttons for 'OK' and 'Cancel'.

Row Labels	Sum of Sales
United States	31,635,241
China	7,557,974
Germany	4,314,956
France	2,628,401
United Kingdom	1,324,267
Canada	1,148,316
Japan	958,334
Australia	632,643
India	598,405
Russia	434,240
Turkmenistan	412,370
Iran	367,953
Syria	329,671
Italy	316,881
Pakistan	296,571
Grand Total	52,956,223

You can also further view rankings by Product Category (or any other suitable field) by adding an appropriate slicer for that field.

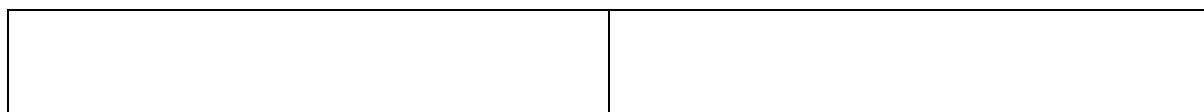
For the next analysis, we can repeat the use of the Value Filters for a given Row Label field to get the top 5. We can also add in another slicer for a suitable Field if we wish to.

Row Labels	Sum of Sales
Bluetooth Headphones	277,660
Camcorders	8,456,121
Cameras & Camcorders Accessories	321,829
Car Video	2,093,144
Cell phones Accessories	825,948
Computers Accessories	709,927
Desktops	3,491,787
Digital Cameras	2,539,792
Digital SLR Cameras	5,731,190
Home & Office Phones	319,645
Home Theater System	4,737,968
Laptops	6,403,174
Monitors	1,868,936
Movie DVD	1,057,480
MP4&MP3	459,961
Printers, Scanners & Fax	1,796,972
Projectors & Screens	7,036,175
Recording Pen	291,760
Smart phones & PDAs	2,822,295
Televisions	2,064,342
Touch Screen Phones	1,873,268
VCD & DVD	212,386
Grand Total	55,391,760

Note that you can clear the filters you have applied to the row / column labels in your Pivot table, and this is a good practice if you are going to apply multiple filters successfully.

Row Labels	Sum of Sales
Bluetooth Headphones	277,660
Camcorders	8,456,121
Cameras & Camcorders Accessories	321,829
Car Video	2,093,144
Cell phones Accessories	825,948
Computers Accessories	709,927
Desktops	3,491,787
Digital Cameras	2,539,792
Digital SLR Cameras	5,731,190
Home & Office Phones	319,645
Home Theater System	4,737,968
Laptops	6,403,174
Monitors	1,868,936
Movie DVD	1,057,480
MP4&MP3	459,961
Printers, Scanners & Fax	1,796,972
Projectors & Screens	7,036,175
Recording Pen	291,760
Smart phones & PDAs	2,822,295
Televisions	2,064,342
Touch Screen Phones	1,873,268
VCD & DVD	212,386
Grand Total	55,391,760

For the next analysis, we are going to filter on the top 20% instead.



By highlighting all the values in the column for Sum of Profit, we can determine that 45 products contribute the top 20% of the profit, which is a form of Pareto Analysis.

If we were to reverse this instead to find the bottom 20% of the list (in terms of overall profit), we would get slightly over 1000 products in that sub-list.

The idea would be that we would focus marketing efforts on the 45 products that contribute the top 20% of the profit, rather than all the other 1000 products

For the next analysis, we are interested in the rankings of the profit by year and month, so we need to perform the appropriate custom grouping.

To sort the months in terms of their profit total (rather than natural chronological order), we use a technique of adding in an extra sum of profit column, and performing a ranking on that instead.

Row Labels - Sum of Profit Sum of Profit2

	2018	12057184.99	12057184.99
Feb	898613.6353	898613.6353	
Mar	891943.5437	891943.5437	
Apr	869075.9176	869075.9176	
May	1209903.159	1209903.159	
Jun	1273089.103	1273089.103	
Jul	1236349.967	1236349.967	
Aug	1097089.286	1097089.286	
Sep	1097949.234	1097949.234	
Oct	1068177.876	1068177.876	
Nov	1270078.95	1270078.95	
Dec	1144914.316	1144914.316	
2019	11378957.22	11378957.22	
Jan	1324613.406	1324613.406	
Feb	849994.1921	849994.1921	
Mar	997446.6405	997446.6405	
Apr	559962.0165	559962.0165	
May	906164.75	906164.75	
Jun	1074977.856	1074977.856	
Jul	863281.9316	863281.9316	
Aug	994217.3148	994217.3148	
Sep	874039.7594	874039.7594	
Oct	850376.4436	850376.4436	
Nov	1001256.298	1001256.298	
Dec	1082770.009	1082770.009	
2020	11237978.22	11237978.22	
Jan	1169711.977	1169711.977	
Feb	698784.0464	698784.0464	
Mar	898613.6353	898613.6353	
Apr	872985.1686	872985.1686	
May	827143.4547	827143.4547	
Jun	888369.8336	888369.8336	
Jul	1115165.876	1115165.876	
Aug	988010.1517	988010.1517	

Contextual menu for the value 1273089.103:

- No Calculation
- % of Grand Total
- % of Column Total
- % of Row Total
- % Of...
- % of Parent Row Total
- % of Parent Column Total
- % of Parent Total...
- Difference From...
- % Difference From...
- Running Total In...
- % Running Total In...
- Rank Smallest to Largest...
- Rank Largest to Smallest...

Now the months are shown in correct chronological order, but we also are able to see their ranking as well (in terms of which months had the highest sales)

Row Labels - Sum of Profit Sum of Profit2

	2018	12057184.99
Feb	898613.6353	9
Mar	891943.5437	10
Apr	869075.9176	11
May	1209903.159	4
Jun	1273089.103	1
Jul	1236349.967	3
Aug	1097089.286	7
Sep	1097949.234	6
Oct	1068177.876	8
Nov	1270078.95	2
Dec	1144914.316	5
2019	11378957.22	

The final analysis looks at the rankings for sales and profit by different countries. We can use the same technique that we did previously.

Row Labels - Sum of Sales Sum of Sales2 Sum of Profit Sum of Profit2

	Armenia	172934.653	172934.653	135,132	135132.3122
Australia	632642.5105	632642.5105	426,288	426288.2342	
Bhutan	161291.606	161291.606	116,520	116520.03	
Canada	1148316.451	1148316.451	694,190	694190.0938	
China	7557973.796	7557973.796	4,955,106	4955106.444	
Denmark	79145.2295	79145.2295	54,735	54734.73743	
France	2628400.624	2628400.624	1,500,337	1500337.194	
Germany	4314956.37	4314956.37	2,672,333	2672332.508	
Greece	68527.9594	68527.9594	53,280	53280.36584	
India	598404.6095	598404.6095	428,186	428185.8712	

Row Labels - Sum of Sales Sum of Sales2 Sum of Profit Sum of Profit2

	Armenia	172934.653	172934.653	135,132	135132.3122
Australia	632642.5105	632642.5105	426,288	426288.2342	
Bhutan	161291.606	161291.606	116,520	116520.03	
Canada	1148316.451	1148316.451	694,190	694190.0938	
China	7557973.796	7557973.796	4,955,106	4955106.444	
Denmark	79145.2295	79145.2295	54,735	54734.73743	
France	2628400.624	2628400.624	1,500,337	1500337.194	
Germany	4314956.37	4314956.37	2,672,333	2672332.508	
Greece	68527.9594	68527.9594	53,280	53280.36584	
India	598404.6095	598404.6095	428,186	428185.8712	

The screenshot shows two separate context menus from a PivotTable in Excel. The left menu is for the field 'Sum of Sales' and the right menu is for the field 'Sum of Profit'. Both menus include options such as 'No Calculation', 'Summarize Values By', and 'Show Values As'. The 'Show Values As' option is highlighted in both cases.

Row Labels	Sum of Sales	Sum of Profit
Armenia	172934.653	135,132
Australia	632642.5105	426,288
Bhutan	161291.606	116,520
Canada	1148316.451	694,190
China	7557973.796	4,955,106
Denmark	79145.2295	54,735
France	2628400.624	1,500,337
Germany	4314956.37	2,672,333
Greece	68527.9594	53,280
India	598404.6095	428,186
Iran	367953.336	193,023
...

We can repeat this again for the sum of Profit to get rankings for both Sales and Profit, which allows use to comparatively compare across these 2 categories. This may be useful for us to identify anomalies where the sales ranking might be very high but the profit ranking is much lower or vice versa.

Row Labels	Total Sales	Sales Ranking	Total Profit	Profit Ranking
Armenia	172934.653	19	135,132	18
Australia	632642.5105	8	426,288	9
Bhutan	161291.606	20	116,520	20
Canada	1148316.451	6	694,190	6
China	7557973.796	2	4,955,106	2
Denmark	79145.2295	32	54,735	33
France	2628400.624	4	1,500,337	4
Germany	4314956.37	3	2,672,333	3
Greece	68527.9594	34	53,280	34
India	598404.6095	9	428,186	8
Iran	367953.336	12	193,023	16
...

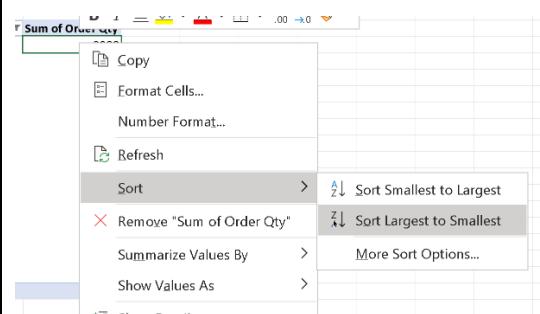
4.1 Practical Exercise for Ranking Analysis

Analytical activities to perform:

- Identify the Top 15 Products by Order Quantity sold, display the total Sales and Profit and add a Slicer for Country
- Create a Ranking for Product Name based on Total Sales and Profit, showing the ranking for both Sales and Profit separately in 2 columns

File to use: SalesData-v7.xlsx

For a), note that after the initial filter for top 15, you can further perform a sort to get the sorted according to sum of order quantity, since the default sorting order is the values in the row label (which are alphabetical values)



The screenshot shows a Microsoft Excel PivotTable with a context menu open over a cell. The menu path 'Sort' -> 'Sort Largest to Smallest' is highlighted. The PivotTable displays a list of products and their total order quantity.

Row Labels	Sum of Order Qty
Contoso In-Line Coupler E180 Silver	6780
Headphone Adapter for Contoso Phone E130 Silver	4680
Contoso Rubberized Snap-On Cover Hard Case Cell Phone Protector E160 Silver	4580
Contoso In-Line Coupler E180 White	4060
Contoso Rubberized Skin BlackBerry E100 Silver	3560
Cigarette Lighter Adapter for Contoso Phones E110 White	3520
Contoso Rubberized Snap-On Cover Hard Case Cell Phone Protector E160 Pink	3280
Contoso Touch Stylus Pen E150 Black	3060
Cigarette Lighter Adapter for Contoso Phones E110 Red	2900
Contoso Original K1m Li-Ion Standard Battery E170 Black	2860
Contoso In-Line Coupler E180 Black	2720
Contoso Rubberized Snap-On Cover Hard Case Cell Phone Protector E160 White	2700
Contoso Touch Stylus Pen E150 Red	2600
Contoso Rubberized Skin BlackBerry E100 Black	2560
Contoso Original K1m Li-Ion Standard Battery E170 Silver	2480
Grand Total	52340

For b) , you can also sort again on the separate ranking columns if you wish for Ranking on Sales and Ranking on Profit. This allows you to see correlation between ranking on sales and profit for various items.

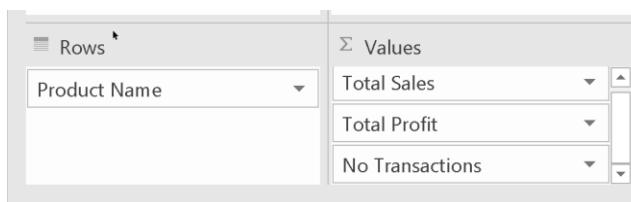
5 Interactive dashboard

Interactive dashboard involves bringing the different forms of analysis together into worksheet, which can help to provide an overview of what is happening in the company. Here we bring together the key metrics, comparison analysis, trend analysis and ranking analysis. Slicers and filters can also be added into to provide additional interactivity for the dashboard.

File to use: SalesData-v8.xlsx

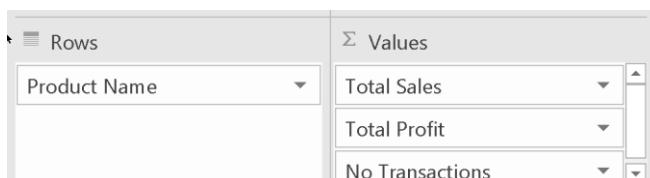
Start off with the basic metrics: Main dashboard - v1

Copy that metrics pivot table into a new pivot table, with Product Name against Total Sales, Total Profit and No. of Trans.



Use Value Filters to get the top 10 by Total Sales, and then further sort it according to Largest to Smallest on the Total Sales.

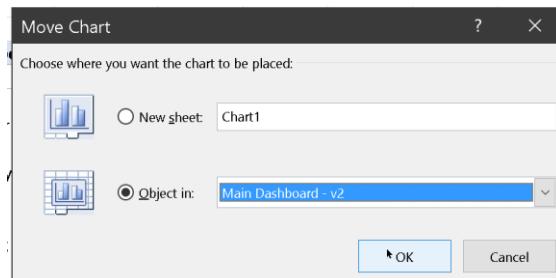
Next, we can create another Pivot Table with these fields: Main dashboard - final



Then, we create a line graph for Year / Month against sum of Sales in a separate tab: Year-Month, remembering to put a limit on the date ranges to omit outliers in the dataset

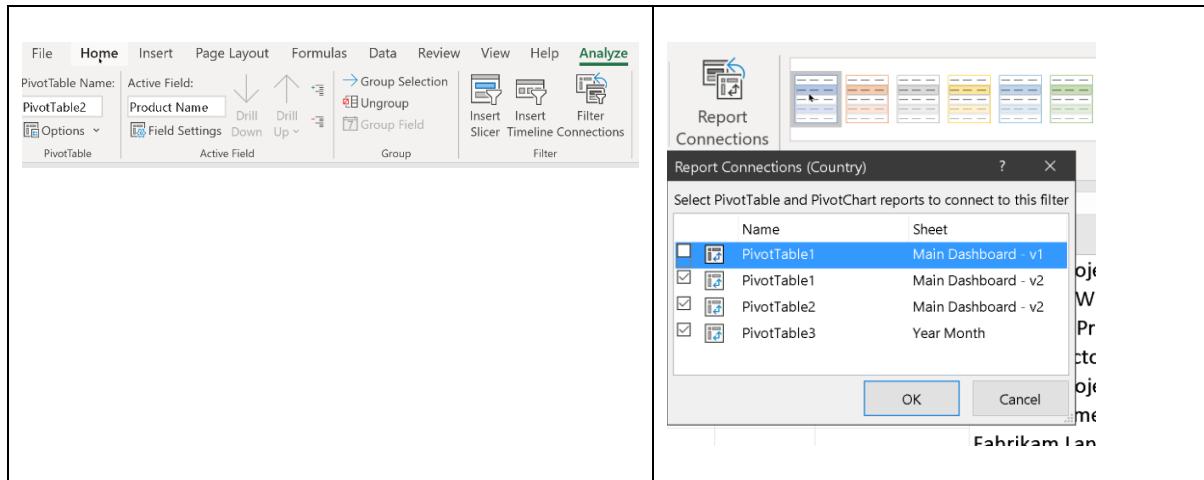


Use the move Chart option to move it back to the original Dashboard worksheet (Tab-v2)

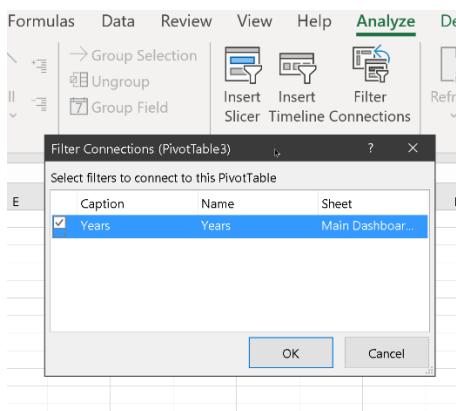


Returning back to the main Dashboard, we can add in an initial slicers linked initially to the Key metrics table. Adding in Slicer for Years, and then set the Slicer Settings to hide items with no data.

You can then link this Slicer to all the other existing Pivot Tables (including those on a different Worksheet) through the Report Connections button. You can get the names of a given Pivot Table by selecting it and going to the Analyze Tab.



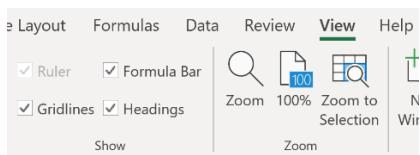
Now, selecting items in the Slicer will change the items in all linked Pivot Tables. You can also select a Pivot Table to determine the Slicer it is linked to with: Analyze -> Filter Connections



We can add another new Slicer for Country and repeat the process of linking them to all existing Pivot Tables.

Now selecting items on both Slicers (Country and Years) will combine their effects simultaneously on all PivotTables simultaneously - this is the key to dynamic interactivity on the dashboard.

For slightly better viewing and presentation as a dashboard, once you are done configuring all the key slicers and Pivot Tables, you can turn off the Gridlines.



6 Variance Analysis

This is about how data varies from other data. Some common applications that you might come across is, for example, you may have an actual value versus what was the budgeted / targeted value. You may also want to compare how data compares against another item. So you may want to use an item as a benchmark or as we've seen with trend analysis, you may have a different data for months and you may want to compare how data compared to, say, a previous month. So there's quite a lot of common applications that we could use for this.

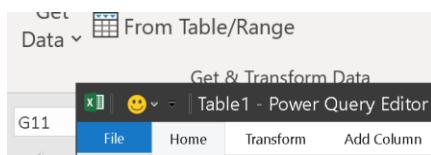
Activities:

- Calculate the variance between the actual product subcategory sales for 2019 and the budget amount
- Calculate the variance between the desktops, product sub category and the other product sub categories
- Calculate the variance between the month sales amount and the previous month for 2019.
- Calculate the percentage variance from the previous month sales as a line chart

File to use: Sales-Data-Budget-v1.xlsx

Here we are going to use Power Query to combine data from two different Pivot Tables to enable us to perform this comparison.

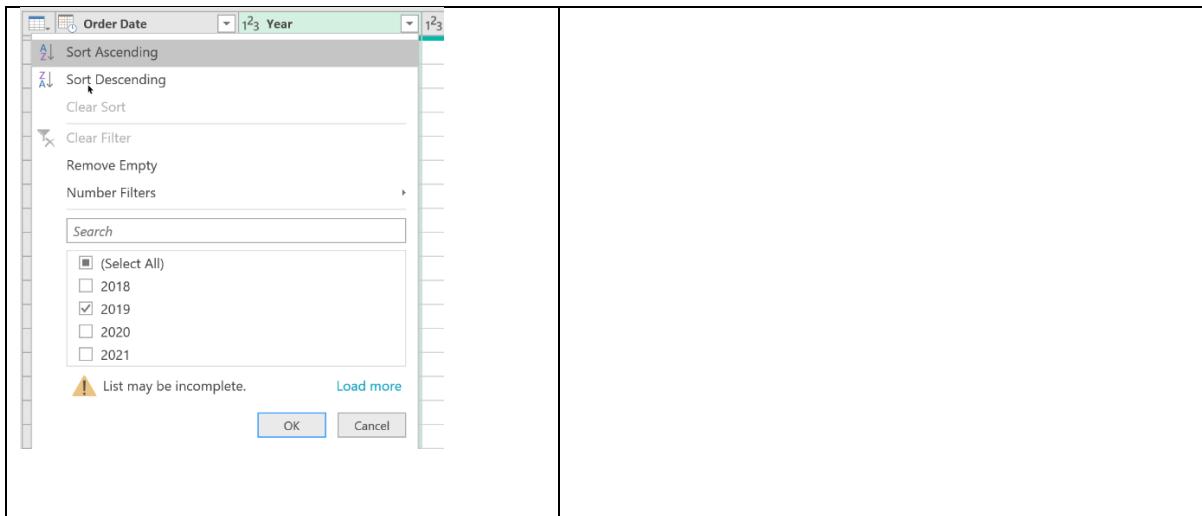
In Data worksheet, select From Table Range -> which opens up the Power Query Editor



We can change the name of the new Table to be generated from the Power Query and also apply some new steps.

Start with Filter on Year, to only obtain data from 2019

	<p>APPLIED STEPS</p> <ul style="list-style-type: none"> Source Changed Type Filtered Rows
--	---



Next, we click on Group By command button to perform an aggregation operation based on the Product Sub Category for the Sales field.

	A ^b c Product Sub Category	1.2 Total Sales
1	Digital SLR Cameras	1724927.085
2	Smart phones & PDAs	915181.8
3	MP4&MP3	193363.3281
4	Recording Pen	85907.385
5	Bluetooth Headphones	80537.898
6	Movie DVD	383248.0182
7	Digital Cameras	848300.952
8	Monitors	640450.88
9	Home & Office Phones	96184.0212
10	Televisions	658051.0361
11	Camcorders	2637074.9
12	Printers, Scanners & Fax	527153.05
13	Desktops	1234013.057
14	Cell phones Accessories	128601.06
15	Computers Accessories	223895.1711
16	Laptops	1811082.764
17	Car Video	712095.8
18	Projectors & Screens	2347458.44
19	Touch Screen Phones	549477.03
20	Cameras & Camcorders Accesso...	88163.7959
21	VCD & DVD	70797.224
22	Home Theater System	1785672.124

When done, click Close and Load. You should see a new table resulting from the execution of the Power Query appearing in a new worksheet.

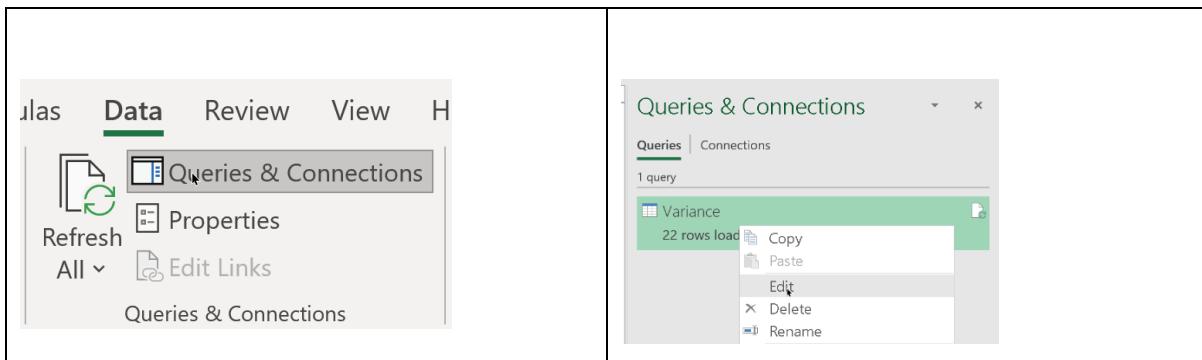
Table Name: Variance

Summarize with Pivot tab
Remove Duplicates
Resize Table
Convert to Range
Properties
Tools

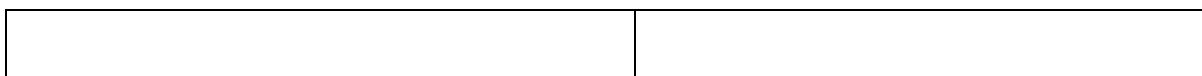
A1 ... X ✓ fx

	A	B
1	Product Sub Category	Total Sales
2	Digital SLR Cameras	1724927.085
3	Smart phones & PDAs	915181.8
4	MP4&MP3	193363.3281
5	Recording Pen	85907.385
6	Bluetooth Headphones	80537.898
7	Movie DVD	383248.0182
8	Digital Cameras	848300.952
9	Monitors	640450.88
10	Home & Office Phones	96184.0212
11	Televisions	658051.0361
12	Camcorders	2637074.9
13	Printers, Scanners & Fax	527153.05
14	Desktops	1234013.058
15	Cell phones Accessories	128601.06
16	Computers Accessories	223895.1711
17	Laptops	1811082.764
18	Car Video	712095.8
19	Projectors & Screens	2347458.44
20	Touch Screen Phones	549477.03
21	Cameras & Camcorders Accessories	88163.7959
22	VCD & DVD	70797.224
23	Home Theater System	1785672.124

Return back to Power Query Editor by editing the Query in the Queries and Connections Pane, accessible from the Data tab



Returning back to the Budget worksheet, highlight the entire range and select From Table/Range which now creates a new Power Query.



The screenshot shows the Power Query Editor interface. On the left, a 'Create Table' dialog box is open, prompting for the range of data (=\$A\$1:\$B\$23) and indicating that 'My table has headers' is checked. On the right, the 'Table3' query is displayed in the Power Query Editor, showing a transformed list of products and their budgets. The table has two columns: 'Product Sub Category' and 'Budget'. The data includes items like Bluetooth Headphones (100,000), Camcorders (2,200,000), and VCD & DVD (90,000). The formula bar at the top shows the query definition: = Table.TransformColumnTypes(Source, {{"Product Sub Category", Text}, {"Budget", Number}}).

Product Sub Category	Budget
Bluetooth Headphones	100,000
Camcorders	2,200,000
Cameras & Camcorders Accessories	90,000
Car Video	650,000
Cell phones Accessories	110,000
Computers Accessories	240,000
Desktops	1,400,000
Digital Cameras	650,000
Digital SLR Cameras	1,900,000
Home & Office Phones	105,000
Home Theater System	1,800,000
Laptops	2,100,000
Monitors	800,000
Movie DVD	340,000
MP4&MP3	150,000
Printers, Scanners & Fax	700,000
Projectors & Screens	2,100,000
Recording Pen	120,000
Smart phones & PDAs	840,000
Televisions	700,000
Touch Screen Phones	700,000
VCD & DVD	90,000

Rename the new query to Budget.

The screenshot shows the 'Query Settings' dialog for the 'Budget' query. Under 'PROPERTIES', the name is set to 'Budget'. Under 'APPLIED STEPS', there is a step named 'Changed Type'. In the main area, the 'Queries [2]' pane shows two queries: 'Variance' and 'Budget', where 'Budget' is highlighted.

Select Variance query and select Merge Queries.

The screenshot shows the 'Queries [2]' pane with 'Variance' selected. In the ribbon, the 'Merge Queries' button is highlighted. A tooltip for 'Merge Queries' states: 'Merge this query with another query in this workbook.'

Select the table to merge into (Budget) and the fields that are common between these two tables that will be used to guide the merge operation.

The screenshot shows the 'Merge' dialog box in Power BI. It displays two tables: 'Variance' and 'Budget'. The 'Variance' table contains data for Product Sub Categories like Digital SLR Cameras, Smart phones & PDAs, MP4&MP3, Recording Pen, and Bluetooth Headphones, with their respective Total Sales values. The 'Budget' table contains data for Product Sub Categories like Bluetooth Headphones, Camcorders, Cameras & Camcorders Accessories, Car Video, and Cell phones Accessories, with their respective Budget values. A 'Join Kind' dropdown is set to 'Left Outer (all from first, matching from second)'. Below it is a checkbox for 'Use fuzzy matching to perform the merge'. A note indicates that the selection matches 22 of 22 rows from the first table. At the bottom are 'OK' and 'Cancel' buttons.

In the new merged column, we can select the columns to bring over from the other table that was merged into. In this case, we are only interested in the Budget.

The screenshot shows the 'ExpandTableColumn' dialog box. On the left, under 'Search Columns to Expand', there is a list of columns: '1.2 Total Sales' (selected), 'Budget' (checked), and others like 'Product Sub Category', '1.2 Total Sales', and 'Budget.1'. Below this is a checkbox for 'Use original column name as prefix'. At the bottom are 'OK' and 'Cancel' buttons. On the right, the expanded table shows the merged data with an additional 'Budget.1' column containing the budget values for each row.

To find the difference between the total sales and the budgeted sales (variance), we can select both columns (with Total Sales being selected first), then in the Add Column tab, select Subtract. This produces a new column with the default name of Subtraction that represents the subtraction of the values of Budget from Total Sales.

The screenshot shows the Power BI Data Editor interface. On the left, the ribbon has 'Add Column' selected under 'View'. The formula bar at the top shows the formula: `=Table.AddColumn(#"Merchandise", "Subtraction", each [Total Sales] - [Budget.1], type number)`. The main area displays a table with columns 'Product Sub Category', 'Total Sales', and 'Budget.1'. A context menu is open over the 'Total Sales' column, with the 'Subtract' option highlighted. The table on the right shows the final data with an additional column 'Subtraction'.

Product Sub Category	Total Sales	Budget.1	Subtraction
Digital SLR Cameras	1724927.085	1900000	-175072.915
Bluetooth Headphones	80537.898	100000	-19462.102
Smart phones & PDAs	915181.8	840000	75181.8
Camcorders	2637074.9	2200000	437074.9
MP4&MP3	193363.3281	150000	43363.3281
Cameras & Camcorders Accessories	88163.7959	90000	-1836.2041
Recording Pen	85907.385	120000	-34092.615
Car Video	712095.8	650000	62095.8
Cell phones Accessories	128601.06	110000	18601.06
Movie DVD	383248.0182	340000	43248.0182
Computers Accessories	223895.1711	240000	-16104.8289
Digital Cameras	848300.952	650000	198300.952
Desktops	1234013.057	1400000	-165986.9425
Monitors	640450.88	800000	-159540.12
Home & Office Phones	96184.0212	105000	-8815.9788
Televisions	658051.0361	700000	-41948.9639
Home Theater System	1785672.124	1800000	-14227.876
Printers, Scanners & Fax	527153.05	700000	-172846.95
Laptops	1811082.764	2100000	-288917.236
Projectors & Screens	2347458.44	2100000	247458.44
Touch Screen Phones	549477.03	700000	150322.37
VCD & DVD	70797.224	90000	-19202.776

Select Home -> Close and Load, you should now that the Variance table has been updated.

Product Sub Category	Total Sales	Budget.1	Subtraction
Digital SLR Cameras	1724927.085	1900000	-175072.915
Bluetooth Headphones	80537.898	100000	-19462.102
Smart phones & PDAs	915181.8	840000	75181.8
Camcorders	2637074.9	2200000	437074.9
MP4&MP3	193363.3281	150000	43363.3281
Cameras & Camcorders Accessories	88163.7959	90000	-1836.2041
Recording Pen	85907.385	120000	-34092.615

For the next activity, we want to calculate the variance between the desktops, product sub category and the other product sub categories

Create a new pivot table from the Data Original and name its containing worksheet 2nd Variance - New. Use the following fields:

The screenshot shows the Power BI Pivot Table settings. The 'Rows' section is set to 'Product Sub Category'. The 'Values' section is set to 'Sum of Sales'.

For the Sum of Sales column, we can identify a particular value for a given item (say Desktops) and then calculate how all the other values for all the other items are different from it.

The screenshot shows a blank Power BI report canvas with two empty rectangular boxes for placing visualizations.

Digital SLR Cameras	5,731	Calibri	11
Home & Office Phones	315	B	I
Home Theater System	6,737,000	Copy	A
Laptops	6,403,174	Format Cells...	A
Monitors	1,868	Number Format...	A
Movie DVD	1,057	Refresh	A
MP3&MP3	455	Sort	A
Printers, Scanners & Fax	1,796	Remove "Sum of Sales"	A
Projectors & Screens	7,036	Summarize Values By	A
Recording Pen	293	Show Values As	A
Smart phones & PDAs	3,832		
Televisions	7,064		
Touch Screen Phones	1,873		
VCD & DVD	212		
Grand Total	55,391		

Desktops	3,491,787	Show Values As (Sum of Sales)	?
digital cameras	2,539,792	X	OK
digital SLR cameras	5,731,190		Cancel
home & office phones	319,645		
home theater system	4,737,968		
laptops	6,403,174		
monitors	1,868,936		
movie DVD	1,057,480		
mp3&mp3	459,961		

The difference in the individual values of each item and the Desktops value is then shown:

Row Labels	Sum of Sales
Bluetooth Headphones	-3,214,127
Camcorders	4,964,334
Cameras & Camcorders Accessories	-3,169,958
Car Video	-1,398,643
Cell phones Accessories	-2,665,839
Computers Accessories	-2,781,860
Desktops	

Another possible alternative for showing variance include Showing Values as a % of

Digital Cameras	2,239,407	No Calculation	
Digital SLR Cameras	-3,17	% of Grand Total	
Home & Office Phones	1,24	% of Column Total	
Home Theater System	2,91	% of Row Total	
Laptops	-1,62	% Of...	
Monitors	-2,45	% of Parent Row Total	
Movie DVD	-3,03	% of Parent Column Total	
MP3&MP3	-1,66	% of Parent Total...	
Printers, Scanners & Fax	3,54	Difference From...	
Projectors & Screens	3,21		
Recording Pen	-9,21		
Smart phones & PDAs	-1,42		
Televisions	-1,61		
Touch Screen Phones	-1,61		
VCD & DVD	-9,21		
Grand Total	100.00%		

Desktops	3,491,787	Show Values As (Sum of Sales)	?
digital cameras	2,539,792	X	OK
digital SLR cameras	5,731,190		Cancel
home & office phones	319,645		
home theater system	4,737,968		
laptops	6,403,174		
monitors	1,868,936		
movie DVD	1,057,480		
mp3&mp3	459,961		

Which again produces another similar result:

Row Labels	Sum of Sales
Bluetooth Headphones	7.95%
Camcorders	242.17%
Cameras & Camcorders Accessories	9.22%
Car Video	59.94%
Cell phones Accessories	23.65%
Computers Accessories	20.33%
Desktops	100.00%

To remove this modified computation:

100.00%	100.00%	No Calculation	
72.74%	72.74%	% of Grand Total	
9.15%	9.15%	% of Column Total	
183.38%	183.38%	% of Row Total	
53.52%	53.52%	% Of...	
30.28%	30.28%	% of Parent Row Total	
13.17%	13.17%	% of Parent Column Total	
51.46%	51.46%	% of Parent Total...	
20.15%	20.15%	Difference From...	
8.36%	8.36%		
80.83%	80.83%		
59.12%	59.12%		
53.65%	53.65%		
6.08%	6.08%		

Another common analytical activity is to calculate the variance between the various months for a given field for a given year (for e.g. variance between the month sales amount and the previous month for 2019).

In a new pivot table, configure as below:

The image shows a Microsoft Excel pivot table configuration. On the left, the 'Rows' section contains 'Years' and 'Order Date'. The 'Values' section contains 'Sum of Sales'. On the right, a 'Grouping' dialog box is open, showing 'Starting at: 2/2/2018' and 'Ending at: 2/2/2021'. Under 'By', 'Months' is selected. A 'Number of days:' dropdown is set to 1. The 'OK' button is highlighted.

Further filter to 2019

The image shows a Microsoft Excel pivot table with 'Row Labels' set to 'Sum of Sales'. The 'Values' section displays monthly sales for the year 2019. The data is as follows:

	Row Labels	Sum of Sales
1	2019	17,741,637
2	Jan	2,081,908
3	Feb	1,252,298
4	Mar	1,507,751
5	Apr	1,103,451
6	May	1,408,131
7	Jun	1,657,980
8	Jul	1,248,653
9	Aug	1,593,694
10	Sep	1,340,433
11	Oct	1,300,579
12	Nov	1,511,809
13	Dec	1,734,950
14	Grand Total	17,741,637

Add in another Sum of Sales column, and then use that new column to compute the difference between successive months using the Show as option again:

A large empty rectangular box for adding another 'Sum of Sales' column.

Row Labels ▾ Sum of Sales Sum of Sales2

	Sum of Sales	Sum of Sales2
=2019	17,741,637	17741636.82
Jan	2,081,908	2,081,908.359
Feb	1,252,298	1,252,297.631
Mar	1,507,751	1,507,751.055
Apr	1,103,451	1,103,450.735
May	1,408,131	1,408,131.43
Jun	1,657,980	1,657,980.084
Jul	1,248,653	1,248,652.914
Aug	1,593,694	1,593,693.618
Sep	1,340,433	1,340,432.531
Oct	1,300,579	1,300,579.196

Which will give us:

Row Labels ▾ Sum of Sales Sum of Sales2		
2019	17,741,637	
Jan	2,081,908	
Feb	1,252,298	-829,611
Mar	1,507,751	255,453
Apr	1,103,451	-404,300
May	1,408,131	304,681
Jun	1,657,980	249,849
Jul	1,248,653	-409,327
Aug	1,593,694	345,041
Sep	1,340,433	-253,261
Oct	1,300,579	-39,853
Nov	1,511,809	211,230
Dec	1,734,950	223,141
Grand Total	17,741,637	

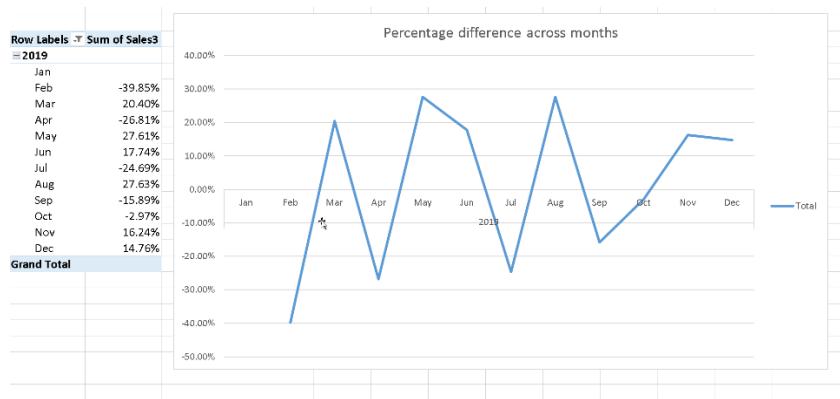
You can repeat this demonstration of difference using percentage instead of actual value difference, again using a new Sum of Sales column and a different option for Show Values As.

Row Labels ▾ Sum of Sales Sum of Sales2 Sum of Sales3

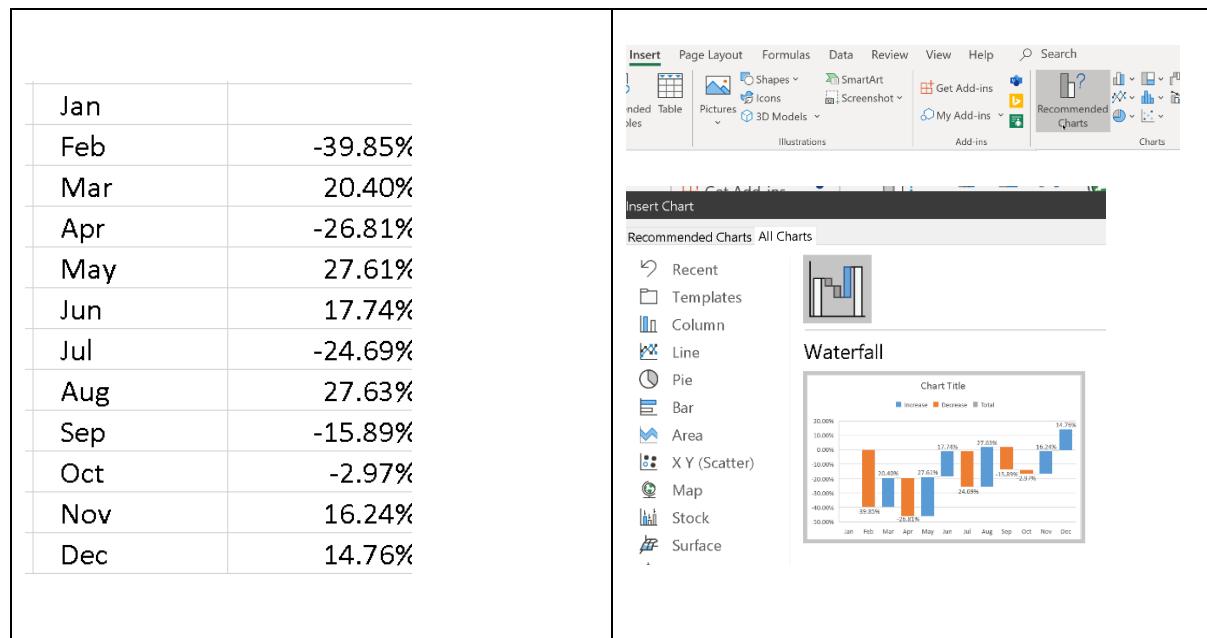
	Sum of Sales	Sum of Sales2	Sum of Sales3
=2019	17,741,637	17741636.82	17741636.82
Jan	2,081,908	2,081,908	2,081,908
Feb	1,252,298	-829,611	-39.85%
Mar	1,507,751	255,453	20.40%
Apr	1,103,451	-404,300	-26.81%
May	1,408,131	304,681	27.61%
Jun	1,657,980	249,849	17.74%
Jul	1,248,653	-409,327	-24.69%
Aug	1,593,694	345,041	27.63%
Sep	1,340,433	-253,261	-15.89%
Oct	1,300,579	-39,853	-2.97%
Nov	1,511,809	211,230	16.24%
Dec	1,734,950	223,141	14.76%
Grand Total	17,741,637		

The last thing is to calculate the percentage variance from the previous month sales as a line chart

Generally when creating a chart from a pivot table, only include the columns in the pivot table that you want displayed in the chart.



An alternative chart option where values rise and fall over consecutive months (or time period) within a given time duration is the Waterfall chart. However, this chart option cannot be directly generated from inside a pivot table, so you will need to copy the relevant contents of the pivot table as a range (instead of a pivot table) to a new section in the worksheet and insert the Waterfall chart using the standard Insert option for a normal range / table.



6.1 Practical Exercise for Variance Analysis

Analytical activities to perform:

- Display the Profit by Year and Month, calculate the difference from the Previous Month. Use a data bar (conditional formatting) to display the difference in a visual format. Also calculate the percentage of difference from the Previous Month
- Finally, display the Ranking by Month and add in a Slicer for country, so we can visualize the specific data for a given country or combination of countries.

File to use: SalesData-v9.xlsx

For the data bar (conditional formatting), you can create another Sum of Profit column, then repeat the standard way to calculate the difference from the Previous Month. Then use the standard Conditional Formatting for Data bars, using any particular

Sum of Profit3	
	-6,670
	-22,868
	340,827
	63,186
	-36,739
	-139,261
	860
	-29,771
	201,901
	-125,165

You can also set a new formatting rule to remove the numbers in the cells with the data bar to make the data bar visualization much clearer.

New Formatting Rule

Apply Rule To: =\$D\$6:\$D\$15

Selected cells
 All cells showing "Sum of Profit3" values
 All cells showing "Sum of Profit3" values for "Order Date"

Select a Rule Type:

- Format all cells based on their values
- Format only cells that contain
- Format only top or bottom ranked values
- Format only values that are above or below average
- Use a formula to determine which cells to format

Edit the Rule Description:

Format all cells based on their values:

Format Style: Data Bar Show Bar Only

Minimum: Automatic Maximum: Automatic

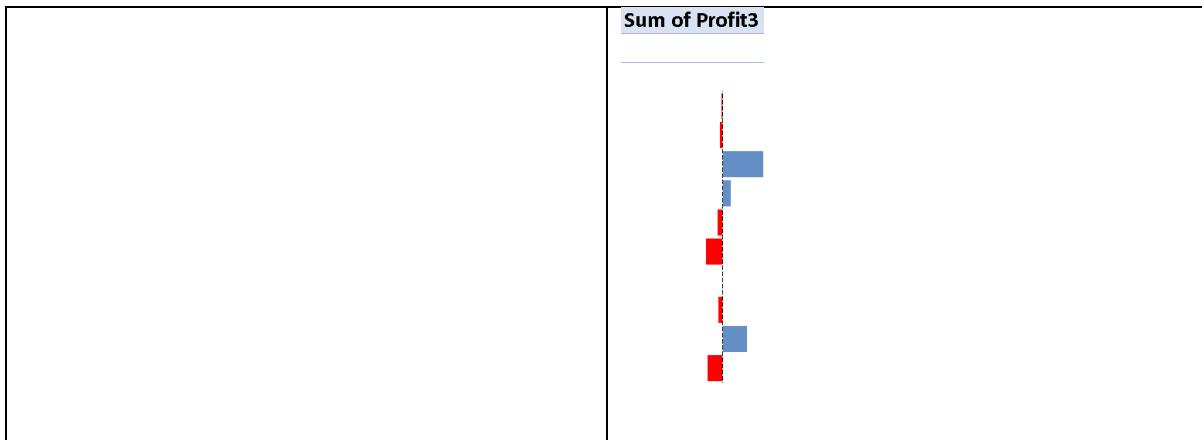
Type: Automatic Value: (Automatic) (Automatic)

Bar Appearance:

Fill: Solid Fill Color: Blue Border: No Border Color: Black

Negative Value and Axis... Bar Direction: Context

Preview:



For displaying ranking by month, again use a new Sum of Profit field and select Show Values As -> Rank Largest to Smallest.

7 Contribution Analysis

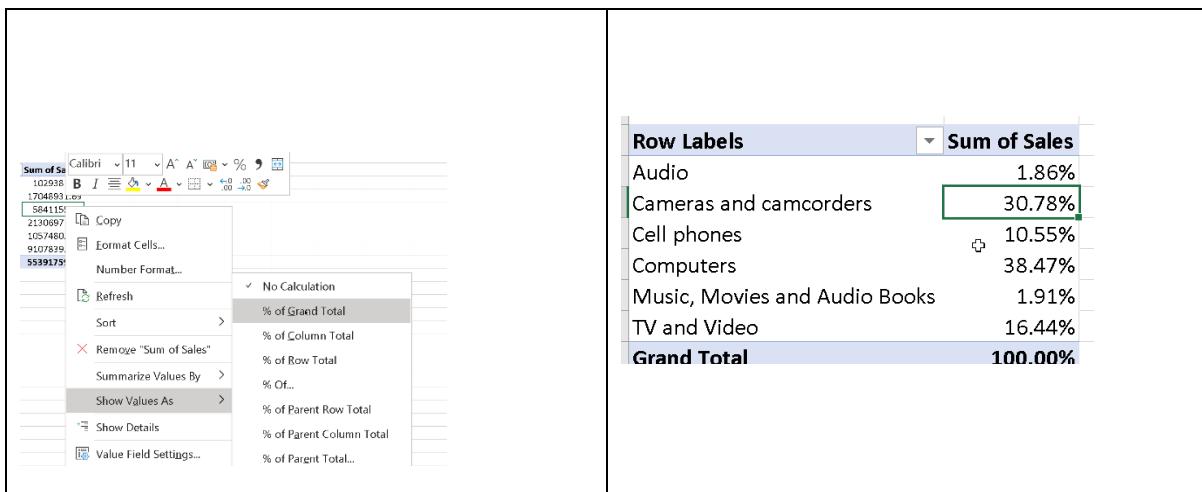
Contribution analysis is all about understanding percentage contributions from various units to the whole. Pie graphs and 100% stacked column graph are also often used when showing percentage contributions.

Analytic activities to be performed:

- What is the percentage contribution of sales by product category?
- What is the percentage contribution of sales by product category and channel?
- Create a pie graph displaying the percentage contribution of profit by product category, review this by different countries.
- Create a 100% column graph displaying sales by product category, review this by region.

File to use: SalesData-v10.xlsx

We can again use the Show Values As option to get the contribution of each product category to the overall sales total.





When we incorporate channel in addition to product category, we can again start to generate % of Grand Total

	Catalog	Online	Reseller	Store	Grand Total
Audio	0.10%	0.45%	0.29%	1.03%	1.86%
Cameras and camcorders	2.96%	6.35%	3.49%	17.98%	30.78%
Cell phones	0.95%	2.01%	1.53%	6.06%	10.55%
Computers	3.04%	7.75%	5.50%	22.18%	38.47%
Music, Movies and Audio Books	0.20%	0.34%	0.33%	1.05%	1.91%
TV and Video	1.78%	3.87%	1.82%	8.98%	16.44%
Grand Total	9.03%	20.76%	12.95%	57.27%	100.00%

We may also wish to see for each of the different channels their percentage contribution to the column total

	Catalog	Online	Reseller	Store	Grand Total
Audio	1.07%	2.16%	2.22%	1.79%	1.86%
Cameras and camcorders	32.82%	30.56%	26.96%	31.40%	30.78%
Cell phones	10.55%	9.69%	11.79%	10.57%	10.55%
Computers	33.66%	37.31%	42.47%	38.74%	38.47%
Music, Movies and Audio Books	2.22%	1.61%	2.52%	1.83%	1.91%
TV and Video	19.68%	18.66%	14.03%	15.67%	16.44%
Grand Total	100.00%	100.00%	100.00%	100.00%	100.00%

Conversely, we may also wish to see for each of the different product categories their percentage contribution to the row total

	Catalog	Online	Reseller	Store	Grand Total
Audio	5.18%	24.17%	15.50%	55.16%	100.00%
Cameras and camcorders	9.63%	20.62%	11.34%	58.42%	100.00%
Cell phones	9.03%	19.08%	14.47%	57.42%	100.00%
Computers	7.90%	20.14%	14.29%	57.67%	100.00%
Music, Movies and Audio Books	10.50%	17.56%	17.12%	54.82%	100.00%
TV and Video	10.81%	23.56%	11.04%	54.59%	100.00%
Grand Total	9.03%	20.76%	12.95%	57.27%	100.00%

When we are using a pie graph, 2 things to keep in mind:

- a) Pie graph is not really designed to show many items - you only want to see a few items
- b) It should always show percentage contribution.

After including the Pie graph, use Quick Layout in the Design tab, to select a suitable way to show the categories and the percentage of contribution. You can also customize it appropriately and then add in a slicer for country.

For the 100% column graph, we can also configure using Show Values by % Row Total, so that the figures in the pivot table match the column heights in the graph.

7.1 Practical Exercise for Contribution Analysis

Analytical activities to perform:

- a) Create a Table displaying the Profit value and Profit percentage by Country
- b) Create a 100% Stacked Column graph displaying the percentage Profit by Product Category by Year
- c) Create a pie graph displaying the Profit by Region
- d) Create a Slicer for Product Category

File to use: SalesData-v11.xlsx

8 Frequency Analysis

In frequency analysis, we're interested in creating groupings and then understand how many occurrences happen within those groupings. A good example is in HR data. So let's say, for example, you've got a number of people in your company and what you want to know is how many people are between the ages of 20 and 30 or 30 and 40, 40 and 50, maybe older than 50. Or you may want to understand how many orders make a profit between specific values. So how many orders make a profit between 0 and \$500 or \$500 to \$1000? 1000? \$1,500.

Analytical activities:

- a) How many employees are aged between 20 to 30, 30, 40, 40 to 50 and over 50.
- b) Breakdown of the number of employees by the different genders: percentage breakdown and also what is the breakdown by the departments

File to use: Employee-v12.xlsx

We can initially start off in the Pivot Table by getting the count of employees at a specific age.

Rows	Σ Values	
Age	Count of Emplo...	

Row Labels	Count of Employee Name
26	1
27	1
28	1
29	1
30	1
31	2
32	4
33	1

Next using the Grouping functionality on the Row Column,

Row Labels	Count of Employee Name
26	
27	
28	
29	
30	
31	
32	
33	
34	
35	

We can also create a simple column chart to visualize this.

We can extend on this to include gender as a column label and view the breakdown across the different age ranges as a percentage.

Count of Employee Name	Column Labels	F	M	Grand Total
Row Labels				
20-29		2.00%	6.00%	8.00%
30-39		10.00%	16.00%	26.00%
40-50		12.00%	22.00%	34.00%
>50		18.00%	14.00%	32.00%
Grand Total		42.00%	58.00%	100.00%

We can also easily extend this to the case of gender across depts:

Count of Employee Name	Column Labels	F	M	Grand Total
Row Labels				
Finance		6.00%	16.00%	22.00%
Human Resource		6.00%	6.00%	12.00%
Logistics		4.00%	12.00%	16.00%
Production		16.00%	12.00%	28.00%
Sales		10.00%	12.00%	22.00%
Grand Total		42.00%	58.00%	100.00%

8.1 Practical Exercise for Frequency Analysis

Analytical activities to perform:

- a) Calculate the number of employees who earn salaries -
 - 0 to 1000
 - 1000 to 2000
 - 2000 to 3000
 - 3000 to 4000
 - Over 4000
- b) Understand the salary groupings by Gender. Add Gender to Columns
- c) Create a Stacked Column graph to display the results and add Department as a Slicer

File to use: Employee-v13.xlsx

9 Correlation Analysis

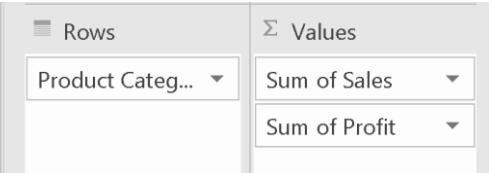
Correlation refers to the relationship between two variables and whether the rise in value of one variable coincides with the rise in values (or vice versa, drop of values) in another variable. Example of typical correlated variables that we want to analyze might be sales and profit (which are usually positively correlated - in that the rise of values in one coincides with the rise of values in the other).

Here we will perform a correlation analysis:

- a) Correlate the sales and profit by your different product categories
- b) Correlate your sales and profit by the product subcategories

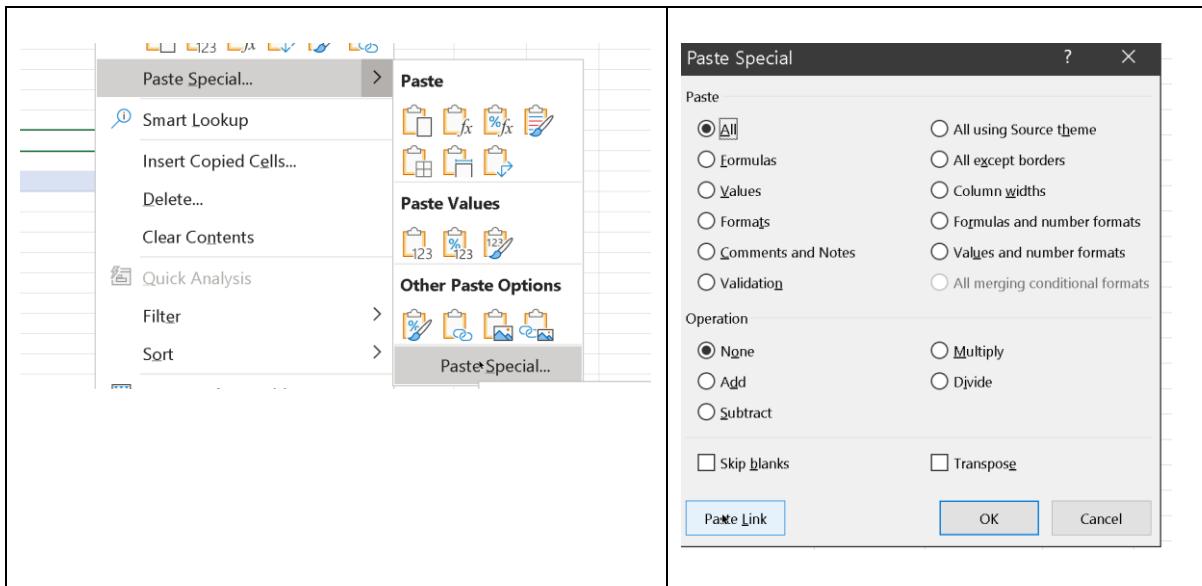
File to use: SalesData-v12.xlsx

We first include the relevant fields in our pivot table:

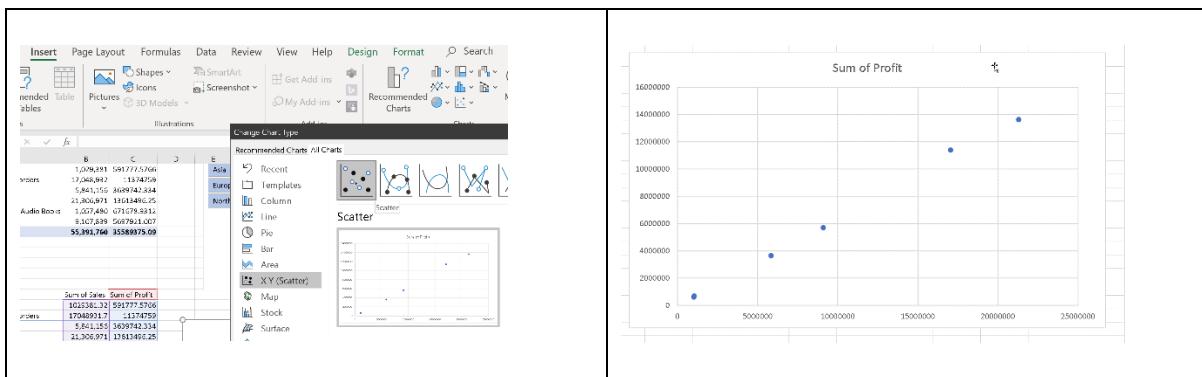


Row Labels	Sum of Sales	Sum of Profit
Audio	1,029,381	591777.5766
Cameras and camcorders	17,048,932	11374759
Cell phones	5,841,156	3639742.334
Computers	21,306,971	13613496.25
Music, Movies and Audio Books	1,057,480	671678.9312
TV and Video	9,107,839	5697921.007
Grand Total	55,391,760	35589375.09

Then we would ideally plot a scatter plot to be able to visualize this correlation properly, however again, we cannot directly generate scatter plots from a Pivot table, so first we must copy the content as a link to a different part of the worksheet and then generate a normal graph from it.

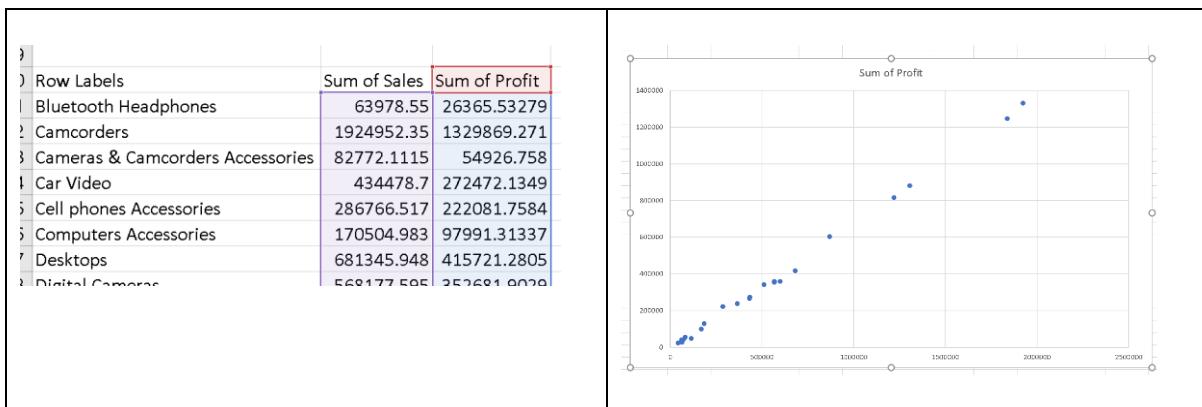


It is important / useful to paste it as a link so that if you were to apply a slicer to the original pivot table the linked normal range would also be affected equally by the slicer selection. Then you can finally use select the linked normal range and plot that as a normal scatter plot.



The fact that both values are rising at the same time indicate high correlation.

We repeat exactly the same operation for a new pivot table, but this time involving product subcategories rather than product category. Again, we can see that there is a nice correlation between sales and profit for this particular field.



10 Pareto Analysis

The Pareto principle (also known as the 80/20 rule) states that for many outcomes, roughly 80% of consequences come from 20% of causes. In business data analytics, you may find that a small amount of something actually contributes a large amount of result. Now this becomes really important when you're doing analysis because you may want to understand what the small amount is that can contribute to a good result. For e.g. if you find that 20% of your customers contribute to 80% of your profit, then it would make much more sense to focus on this 20% of customers and get a much better return.

Here we will use Pareto analysis to understand the percentage of products that contribute 80% of the profit.

File to use: SalesData-v13.xlsx

First, we do a simple sort of the product Name based on their sum of profit.

Rows	Σ Values	Row Labels	Sum of Profit
Product Name	Sum of Profit	Contoso Projector 1080p X980 Black	343,349
		Proseware Projector 1080p LCD86 White	337,808
		Contoso Projector 1080p X980 White	337,366
		Proseware Projector 1080p LCD86 Black	277,263
		Proseware Projector 1080p DLP86 Silver	268,987
		Proseware Projector 1080p DLP86 Black	212,056
		Fabrikam Independent Filmmaker 1" 25mm X400 Blue	208,870
		WAN Projector 1080p X980 Blue	204,771

We then add in a second Sum of Profit, and do a Show as % of Column Total.

Rows	Σ Values	Row Labels	Sum of Profit	Show as % of Column Total
Product Name	Sum of Profit	Contoso Projector 1080p X980 Black	343,349	0.96%
	Sum of Profit2	Proseware Projector 1080p LCD86 White	337,808	0.95%
		Contoso Projector 1080p X980 White	337,366	0.95%
		Proseware Projector 1080p LCD86 Black	277,263	0.78%

To find the products that contribute 80% of the overall profit, we will need to perform a running total. We add in another Sum of Profit column, and then do a Show as % Running Total

Rows	Σ Values	Row Labels	Sum of Profit	Show as % Running Total
		Contoso Projector 1080p X980 Black	343,349	0.96%
		Proseware Projector 1080p LCD86 White	337,808	0.95%
		Contoso Projector 1080p X980 White	337,366	0.95%
		Proseware Projector 1080p LCD86 Black	277,263	0.78%
		Proseware Projector 1080p DLP86 Silver	268,987	0.76%
		Proseware Projector 1080p DLP86 Black	212,056	0.60%

Sum of Profit	
337,366	0.03% 337366
337,366	0.03% 337366
277,263	0.78% 277263
268,687	0.75% 268687
213,056	0.60% 213056
208,870	0.59% 208870
185,111	0.54% 185111
185,393	0.53% 185393
170,378	0.48% 170378
165,978	0.48% 165978
165,787	0.47% 165787
168,576	0.47% 168576
166,046	0.47% 166046
160,702	0.45% 160702
157,707	0.44% 157707
157,357	0.44% 157357
152,403	0.43% 152403
149,322	0.42% 149322
147,914	0.42% 147914
145,694	0.41% 145694
145,694	0.41% 145694
145,694	0.41% 145694
136,948	0.38% 136948
135,910	0.38% 135910
132,873	0.38% 132873
120,087	0.36% 1200870724
128,914	0.36% 1289143734
123,514	0.35% 1235140928
123,105	0.35% 1231055423

Going down this column, we can identify that there are 552 items that are contributing 80% of the overall profit.

Fabrikam SLR Camera 35" M358 Grey 19,935 0.06% 80.03%

Some basic calculation will reveal to us that 33% of the product range contributes to 80% of the profit, and we can therefore focus more on the marketing / manufacturing of these items.