Introduction to Data Analysis Using Excel

Part I

Prepared By:

Said Fawzy

Manger of Information Center

Tendering Department
Arab Contractors

Table of Content:

Contents

Introduction:

لما كانت قوة أى مؤسسة تكمن فى سرعة إتخاذها القرار ومواجهة تغيرات السوق والمنافسة ونقل المؤسسة من مكانة إلى أخرى أفضل ععن طريق الحصول على أكبر عائد ، وتقليل التكاليف ومعرفة مواطن القوة والضعف فيها وفى البيئة التى تحيطها ، ولأن البيانات والمعلومات هى القوة الاساسية الداعمة لاتخاذ اى قرار باى مؤسسة ، ولما كانت البيانات فى حد ذاتها لا تعين على اتخاذ القرار الصحيح ، ولكن يلزم لها ان تجيب على الاسئلة المحددة والدقيقة و الدائرة فى اذهان متخذى القرار حتى يتثنى لهم الرؤية الواضحة للامور على اساس متين يعول عليه بعد إعدادها وتقديمها فى صورة تقارير ومخططات واضحة جليه.

وعليه فقد قمت بإعداد هذه الدورة كمقدمة فى تحليل البيانات عن طريق برنامج الاكسل حرصت فيه ان يكون مبسطا ومركزا ويعطى صورة للدارس الذى يرغب فى الاستمرار فى هذا المجال من معرفة اساسيات تحليل البيانات بصورة عامة وما يقدمه برنامج الاكسل بصورة خاصة.

وتوخيت ان اراجع فى بداية كل فصل من الفصول المفاهيم الاساسية والدالات التى سيحتاجها محلل البيانات حتى يفهم فكرة الفصل ، حتى اطمئن الى ان المعلومات التى سيقدمها كل فصل يفهمها الجميع بصورة صحيحة .

وقد قمت بتقسيم كل فصل الى :

- عرض مبدئى لفكرة الفصل ومثال عملى يشاهده الدارس لتطبيق الفكرة.
- اسئلة لاختبار المعرفة حتى نتأكد من فهم الدارس للنقاط التى القيت عليه.
- تدريب عملى على ملف شركة افتراضية تم جمع بياناتها خلال 6 سنوات حتى يشعر الطالب بقيمة الارقام وتحليلها وما يمثله كل رقم وكل مخطط يقوم بعمله.

واذ تكون هذه هى الخطوه الاولى فى تدريب محلل البيانات فإن البيانات المستخدمة فى التدريب قد أعدت اعدادا جيدا للتمثيل وعمل المخططات والتقارير دون صعوبة ، ولقد أرجئت طرق الحصول على البيانات واعدادها من المصادر المختلفة وعمل الربط بينها والتأكد من خلوها من الاخطاء الى التدريب المتقدم التالى .

أرجو ان تكون هذا المادة العلمية هذه وسيلة لايضاح فلسفة تحليل البيانات وتبسيط فكرة عرض وتمثيل البيانات واستخلاص النتائج منها و ان تكون بداية موفقة لكل من اراد ان يستكمل المعرفة و التدريب فى هذه المجال .

وفقنا الله الى ما يحبه ويرضاه

مهندس سعيد فوزى محمد هدى
مدير مركز المعلومات – مدير الجودة
إدارة العطاءات
المقاولون العرب

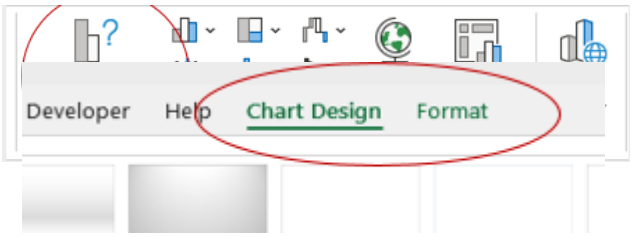
القاهرة 1 أإبريل 2023

Before We Start: Excel Chart Basics

Let us review the basics of creating charts in Excel.

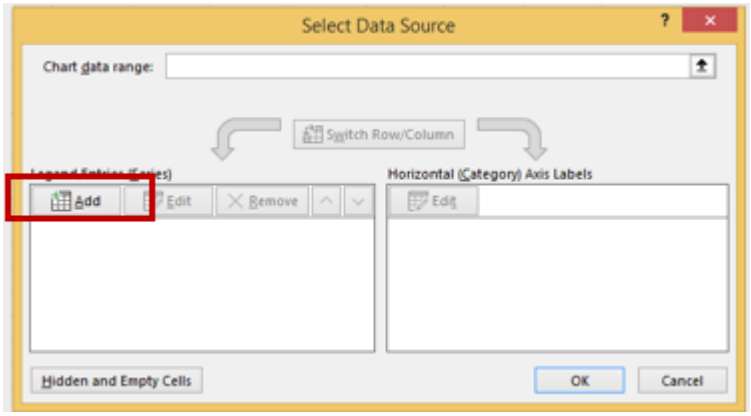
How to Insert a Chart

The easiest way is to use Insert ➔ Chart Group ➔Recommended Chart



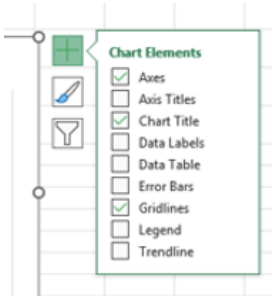
You will have two new taps when you insert chart

You can add data series manually using

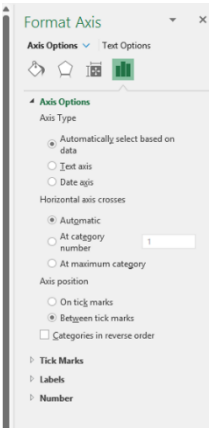


Adjusting the chart

You can add and remove chart element

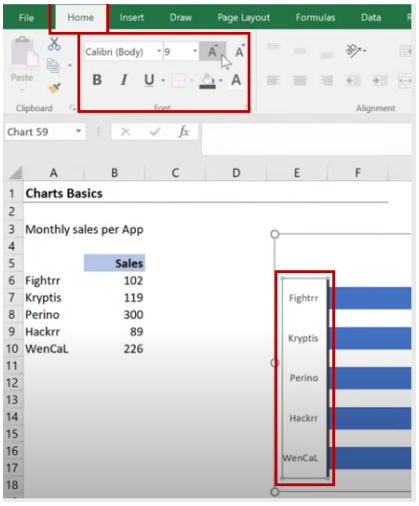


You can format any element

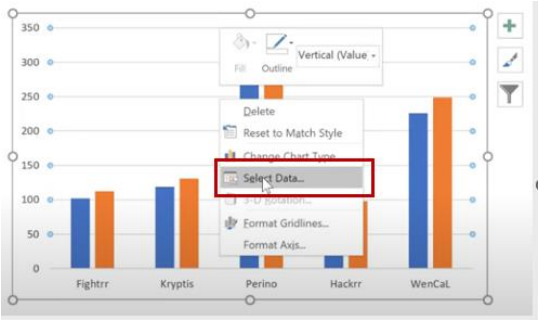


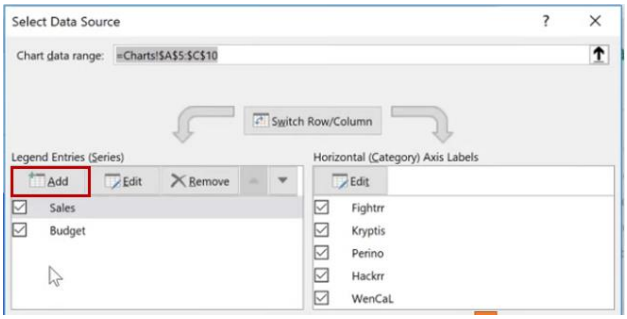
Improving Chart

Change Axis text fonts and color



Adding More Series





Lab 0: Basics of Charts

Use_Lab 00_Chart Basics_Start.xlxs from Labs Folder.

You can use the step by step available in the file

Excel Charts Basic.pptx

Step 1 Inserting a chart :

We want to visualize this table of data

1. 1-Insert a chart using Insert Recommend Charts.
2. Select a cell outside the data table , and Insert a chart Using Bar chart option.
3. Now you have an empty chart.
4. Notice you have new taps.
5. Add data to the new chart from the table.
6. Now you have another chart you made manually
7. Delete all charts you have created.
8. Create a default chart using Alt + F1
9. Change the Default chart to line chart.
10. Delete the current chart
11. Try your new default chart.

Step 2 Adjusting the chart :

1. Delete all charts you have created

2. Insert a cluster bar chart
3. Go and check data elements using the + icon.
4. Add data labels
5. Change where data labels appear.
6. Click More options
7. Right click data series and format the element.
8. Reverse the Y Access.
9. Change the Gab width between bars.
10. Select and delete the Grid line

Step 3 Improving the chart :

1. Enlarge Y Access line and increase its font and make it bold.

Step 4 Adding more series :

1. Add another column to the table Budget.
2. Make budget = 1.1 X Sales.
3. Select the Chart and expand data series.
4. Add the same series again with different name using right click the data series.
5. Change the series order.
6. Remove the new series.

Step 4 Create combination Chart :

Convert Budget to become line chart

1. Convert Budget line into Dash points
2. Change time size color of the dash
3. Add data labels whit in the bottom of columns of sales.
4. Add data labels on top for Budget
5. Add title to the chart
6. Add legend to chart
7. Delete Y axis .

Module 01: Introduction to Reporting In Excel

You are now working for Mozilla Bicycle Company. They are selling Bicycles and their components.

Demo 1: Creating Charts

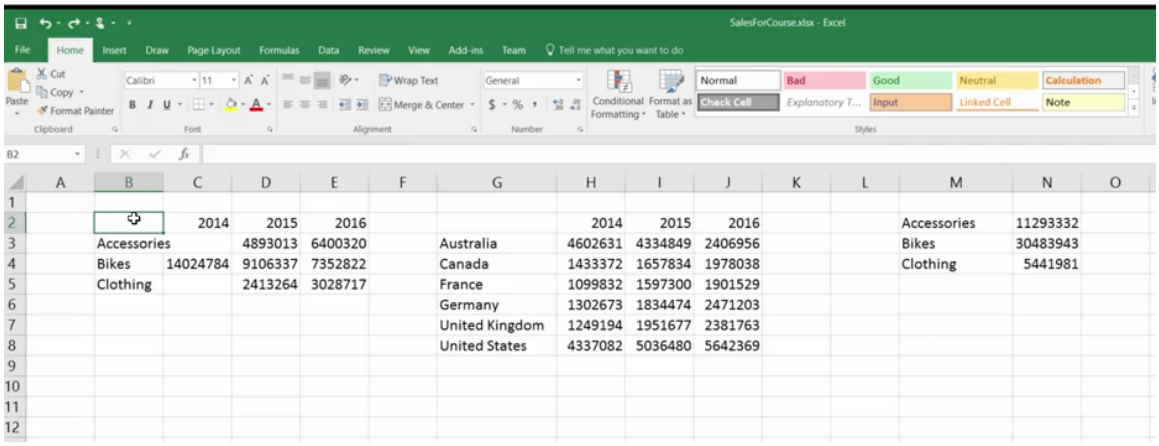
Lucy is a new Sales Manager. She wants to know information about Company sales.

You think you want to show her :

And you think it is enough to show her the performance of the company.

You contacted Jack your friend in IT and he prepared the data for you.

Jack sends you this file

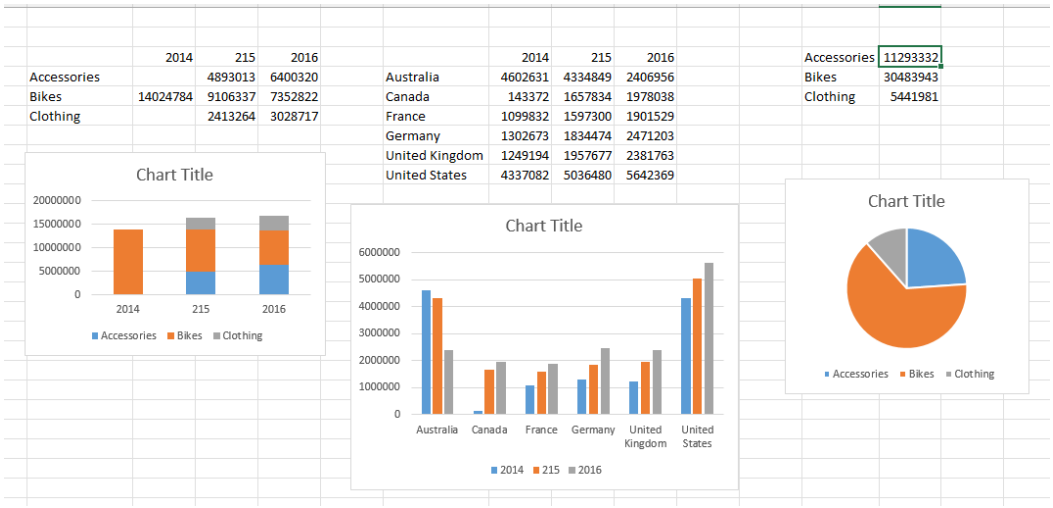


	2014	2015	2016
Accessories		4893013	6400320
Bikes	14024784	9106337	7352822
Clothing		2413264	3028717

	2014	2015	2016
Australia	4602631	4334849	2406956
Canada	1433372	1657834	1978038
France	1099832	1597300	1901529
Germany	1302673	1834474	2471203
United Kingdom	1249194	1951677	2381763
United States	4337082	5036480	5642369

Accessories	11293332
Bikes	30483943
Clothing	5441981

You created this report to Lucy



But she needs more details and many other ways to represent the data.

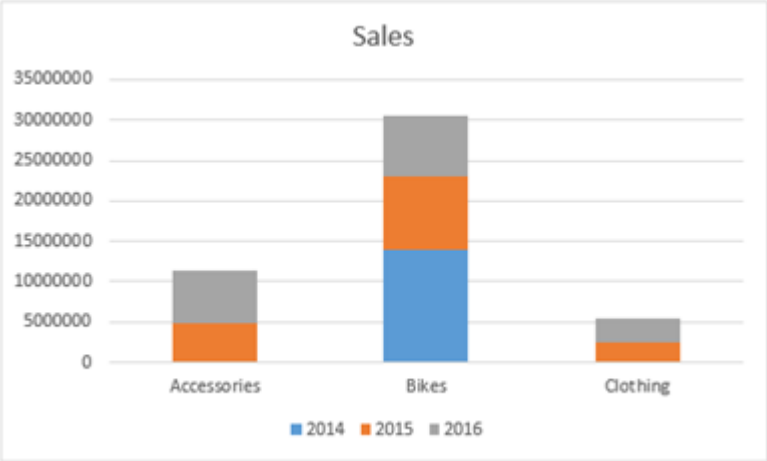
She is interested in:

- profitability,
- subcategories,
- know more about the customers.
- can we classify customers by Age groups? …. etc.

Every time you must go to Jack to give you summery of the data you want. By time Jack Got tired. Jack said I think you need something else; I will prepare for you a set of detailed data and you can create any report or summary you want. All you must do is using **Pivot tables** and **Tables**.

Knowledge Checks

Question 1



You are asked to display a trend for each product by year. What should you do?

- Change the chart type to pie chart.
- Modify the chart to have the product categories as Legend and years as horizontal axis.
- Transpose the table and create the chart from scratch.
- Add summary for each year and insert a new column chart.

Question 2

You have collected the following data and are considering using a pie chart to display it.

	2014	2015	2016
Australia	4602631	4334849	2406956
Canada	1433372	1657834	1978038
France	1099832	1597300	1901529
Germany	1302673	1834474	2471203
United Kingdom	1249194	1951677	2381763
United States	4337082	5036480	5642369

What is a problem that you are likely to encounter by using a pie chart?

- You will have to put the legend below or above the pie because the county names are too long to appear next to the appropriate slice
- The differences in the data are too small between Canada, Germany, and the United Kingdom for the pie chart to have value.
- The pie chart will only show one of the year's data.
- The numbers collected in the data exceed those that can be used effectively in a pie chart.

Question 3

You have collected the following data and want to use a chart to display it.

	2014	2015	2016
Australia	4602631	4334849	2406956
Canada	1433372	1657834	1978038
France	1099832	1597300	1901529
Germany	1302673	1834474	2471203
United Kingdom	1249194	1951677	2381763
United States	4337082	5036480	5642369

Which chart should you use to display it?

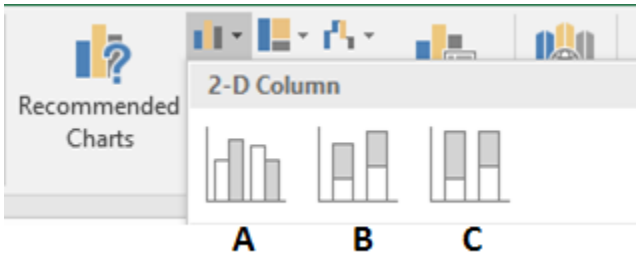
- Histogram
- Line chart.
- Pie chart
- Scatter plot

Question 4

You are asked to display the percentage of sales by category for each year in a column chart.

	2014	2015	2016
Accessories		4,893,013	6,400,320
Bikes	14,024,784	9,106,337	7,352,822
Clothing		2,413,264	3,028,717

Which of the following chart types should you use?



- A
- A and B
- B and C
- C

Lab 01:

Scenario:

Use Data in file Lab1Start.xlsx

You are the new marketing manager of an established Bicycle company. The company sells bicycles and accessories, such as clothing and other accessories to bikers in six countries.

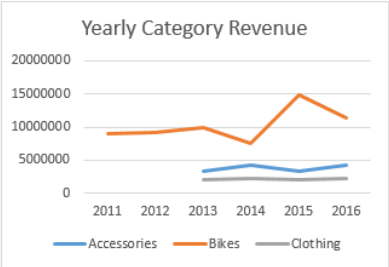
The company has just hired Lucy as its new Sales manager. You are tasked to introduce Lucy to the company, its product portfolio, and its sales performance since 2011. To do this, you have asked Jack, the IT manager to prepare some data for you.

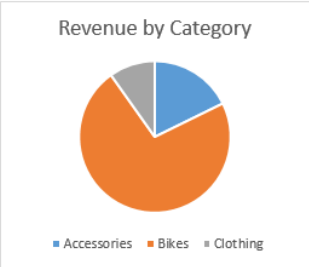
Now, it's your job to present this data in a compelling manner.

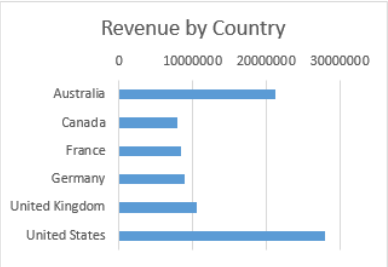
Step 1:

The first thing you'd like to do is to present the data graphically. You do this by creating charts for the three groups of data you received from Jack, your IT manager. If you haven't done so, give the chart a descriptive title.

Your first solution should look like this (numbers may not be the same).







Step 2:

Once you have created the three charts above, let's customize them a little bit more.

For the "Yearly Category Revenue", display the total for each year as a separate line.

1- Create a total line in the range B7 to H7.
2- Select the **Yearly Category Revenue** chart.
3- The data source for the chart is highlighted.
4- Drag the highlighted selection to include the total row (B7 to H7)

Question 1

What does the Yearly total sales look like?

- Increasing year over year.
- Similar to the yearly bike's revenue
- Decreasing year over year.
- Similar to the yearly accessories' revenue

Step 3:

For the "Revenue by Category", display the percentage for each component.

1. Select the **Revenue by Category** chart.
2. Select the pie in the chart, right-click and add data labels.
3. Select **Percentage** in the **Label Options**.

Question 2

What percentage of the total revenue comes from the **Bikes** category?

- 10%.
- 18%
- 36%.
- 72%.

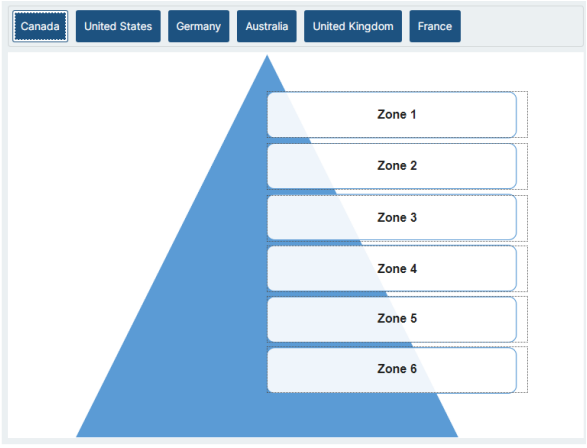
Step 4:

For the **Revenue by Country**, sort the countries from the one with highest revenue to lowest revenue.

1. Select the data source for the **Revenue by Country** chart.
2. Sort the data by **Revenue** from **Largest to Smallest**.
3. Select the **Revenue by Country** chart.
4. Select the Y axis, the axis that shows the country name.
5. Select **Categories in reverse order** in the **Axis options**.

Question 3

Rank the Countries from the highest to lowest revenue. (Highest in Zone 1, Lowest in Zone 6)



Module 02: Excel Tables

Jack prepared us an excel file with the data needed. He said try to use the Excel tables. So what is it ?

Before we start:

Functions you should know:

- IF
- Nested IF
- IFS
- RIGHT
- TEXT
  See Extra Demo Folder

Demo 02 : Excel Tables

Knowledge Checks:

From Quiz folder in Folder Mod 02 use this file :
SalesForCourse_quizz_table.xlsx.

Question 1

What is the difference between a table and a range in Excel? Select four that apply.

- In tables, you can add many types of totals for each column without writing any formulas..
- A range can be converted to a table, but a table cannot be converted to a range.
- Tables are formatted with alternate colors by default.
- Formulas in tables refer to other columns by name and not by regular Excel reference.
- Formulas in table columns are automatically applied to new rows.

Question 2

If you filter the table to show only sales in United Kingdom, what will be calculated in the Total row?

- The total of all rows.
- The totals will recalculate according to the filtered rows.
- The totals will recalculate according to the unfiltered rows.
- The totals will recalculate only when you reopen the file.

Question 3

You are only interested in the data for male customers in France. After setting the appropriate filters, answer the following questions.

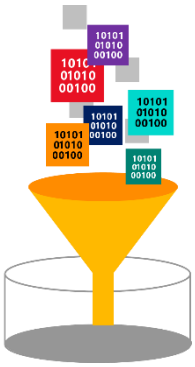
a. How many rows are displayed?
b. What is the total revenue for the displayed rows? Only provide whole number. Do not provide symbols or any decimals places.
c. What is the average revenue? Only provide numbers, no currency symbols, and report to two decimal points.

Lab 02: Creating Excel Table

Use Data in file Lab2Start.xlsx

Scenario

While your first attempt to show the company's performance to Lucy was not bad, clearly, she has a lot more requirements than what you provided. She wants to know more about the year over year sales, sliced into different categories, sub-categories, and countries. She also wants to see additional information such as customer demographics.



Jack has provided you with a different data source. This time the data has more than one hundred thousand rows.

Before you can create additional reports to Lucy, first you need to prepare the data.

Step 1:

The first thing you need to do is to convert the data into an Excel table.

Once you do that, you can add total row, filter the data, and the total will reflect the total only for the filtered data. Let's try this. Add a total row for the table and use the Sum aggregation to show the total of the **Revenue** column and then filter the data only for **United States**.

Question 1

What is the total revenue for all the sales in the **United States**?

- 101990988
- 27975547
- 24606895
- 12494351

Step 2:

Now, you need to add several columns, derived from existing columns in the data.

First, let's add a "Month" column. Insert a new column to the left of the **Customer ID** column, and use formula to derive the month of sales from the **Date** column. HINT: Use the Text() function (=TEXT([@Date],"MM")

Question 2

What is the total revenue for all the sales in the month of **December**?

- 9086931
- 8801313
- 6809175
- 6494188

Step 3:

Next, let's add an "Age Group" column. Insert a new column to the left of the **Customer Gender**, and use formula to derive the age group from the **Customer Age** column. Let's group the customers based on the following criteria:

  - Youth (<25)
  - Young Adults (25-34)
  - Adults (35-64)
  - Seniors (>64)

HINT: Use the nested IF() functions. Alternatively, you can use the IFS() function if it is available in your version (2016 + updates from O365)

Question 3

What is the total revenue for all the sales for **Young Adults** Age Group?

- 56053751
- 29918498
- 30655614
- 5556163

Step 4:

Now, let's add a "Frame Size" column. Insert a new column to the left of the **Order Quantity**, and use formula to derive the frame size of a bicycle from the last two characters of the **Product** column, when the **Product Category** is **Bikes**. Otherwise, leave it blank.
HINT: Use the IF() and RIGHT() function.

Question 4

What is the total revenue for all the bikes with frame size **62** for the customer age group **Seniors**?

- 22914
- 125893
- 159102
- 12452

Step 5:

Last but not least, let's add a "Profit" column. Insert a new column to the right of the **Revenue**, and use formula to derive the Profit from both the **Revenue** and **Cost** columns. Show the total for the **Profit** column. Use the Sum aggregation in the total row of the table, for the **Profit** column.
HINT: **Profit** is **Revenue** less **Cost**.

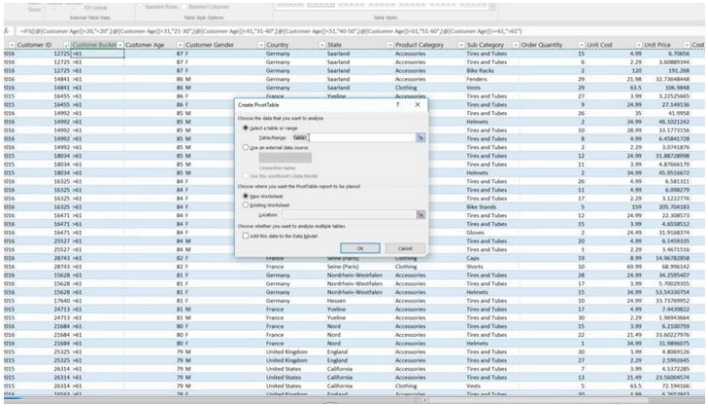
Question 5

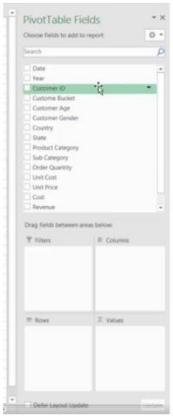
What is the total profit for **United States** sales in the month of **October 2015**, for customer age group **Adults**?

- 111823
- 142036
- 148903
- 138419

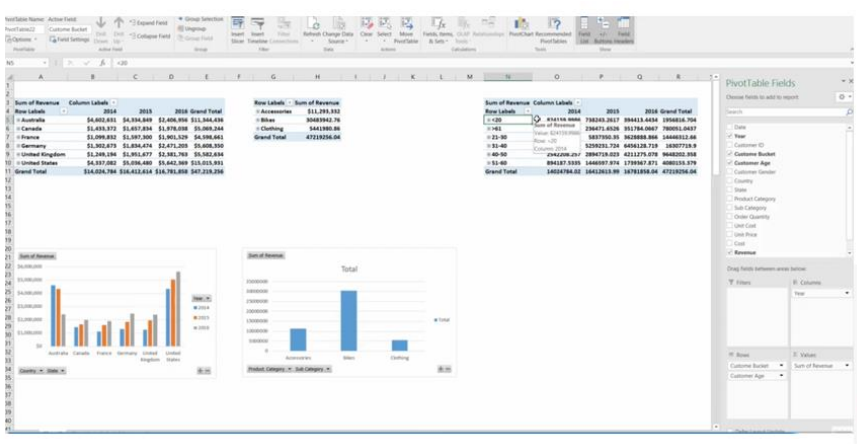
Module 03: Basic Pivot Tables

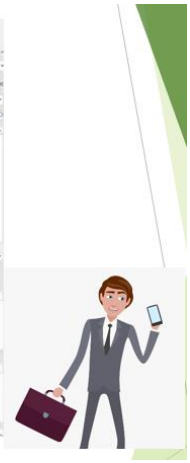
Demo 3A: Basics of Pivot Tables





Demo 3B: Creating Dashboard using Pivot table



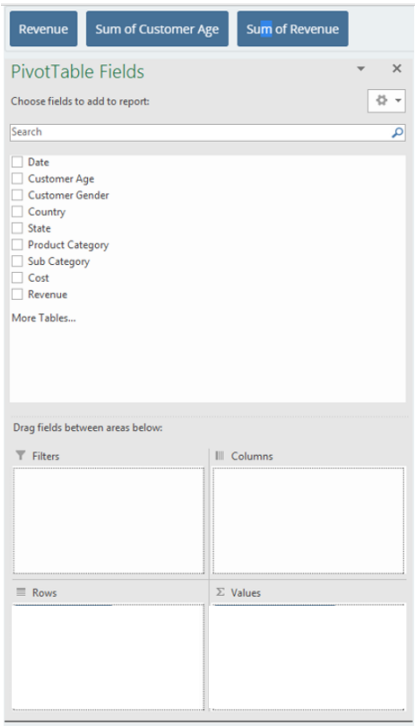


Knowledge Checks

Question 1

PROBLEM

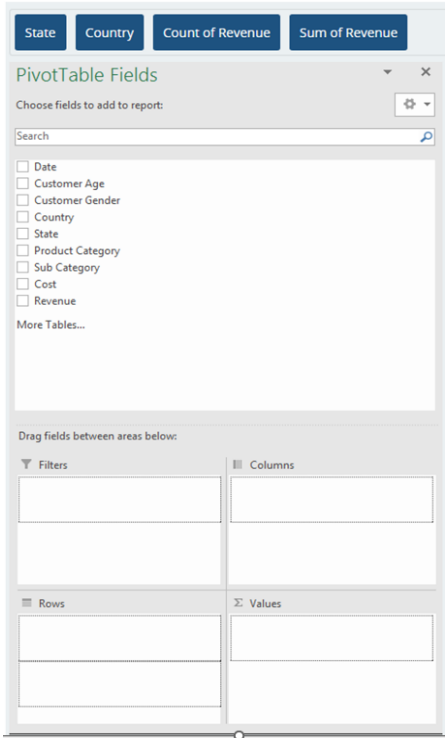
- You have the following Pivot Table fields.
- You need to calculate the number of sales by age.
- Drag the appropriate field names and aggregated field names to the appropriate Pivot Table Areas (Filters, Rows, Columns, or Values) to perform the task



Question 2

PROBLEM

- You have the following Pivot Table fields.
- Your manager asks you to build a pivot table that shows revenue by state and country.
- How would you arrange the field names and aggregated field names to the appropriate Pivot Table Areas (Filters, Rows, Columns, or Values)?



Question 3

You have a pivot that is based on an Excel table. You need to add new rows to the Excel table and want to be able to see the values for these new rows in the PivotTable. What should you do?

- Create a new pivot table.
- Add rows anywhere in the table, and it will automatically update the pivot to include the new values.
- Add rows anywhere in the table, and refresh the pivot table.
- You must add the new rows only after the last row and refresh the pivot table.

Use Data in file SalesForCourse_quizz_table.xlsx to answer

Question 4

Using the data provided, create a pivot with “Country “as rows, “Month” as columns, and “Revenue” as values. Drag “Years” to the filter area. What does the value 286,779 in the first summarized field of the pivot table (uppermost left value) represent?

- The total for January in France for the first year (2015).
- he total for January in France for the last year (2016).
- The total for January in all countries.
- The aggregated value for January in France in 2015 and 2016

Question 5…Numeric input

Create a new PivotTable with “Customer Gender” as rows, “Product Category” as columns, and “Revenue” as values.

a. How much revenue was raised through the purchase of bikes by women? Provide a whole number without currency symbols or decimal places.

b. How much revenue was raised through clothing purchased by men? Provide a whole number without currency symbols or decimal places.

Question 6

When you summarize data using an Excel Pivot Table, each number found in a Pivot Table cell represents one or more records in the pivot table's source data and is directly related to the Column and Row fields you used in the Pivot Table. A quick way to review this underlying set of records is to use the Show Details feature - right click within the desired cell and select Show Details from the menu options. This is commonly called **Drilling down into the data** and can also be achieved by double-clicking in the cell in question.

Create a new PivotTable with "Country" as rows, "Product Category" as columns, "Year" as a filter, and "Revenue" as values. Filter by year 2015.

Drill down into the data for the cell that represents the revenue for bike sales in Germany, 2015. A new worksheet is created with a table in it. What is the content of this table?

- The table shows columns for Year, Country, Revenue, and Products
- The table contains all rows in SalesTable corresponding to 2015, Bikes and Germany
- The table includes data from SalesTable for the first month of 2015 and the first subcategory of Bikes.
- The table includes data from SalesTable for 2015 and the first subcategory of Bikes.

Question 7…Numeric input

Calculate the answers to the following questions using the data provided and PivotTables. All answers should be provided in whole numbers, without currency symbols or decimals

a. Total revenue for sales of bikes in the country that sold the most in 2016:

b. Number of sales transactions for bikes in the US in 2016:

c. Average number of clothing items per transaction purchased in the US during 2016 rounded to the nearest whole number:

d. The largest revenue registered in a single transaction in the Germany during 2015:

e. The largest revenue registered in a single transaction during 2015:

Lab 3: Creating Pivot tables

Use Data in file Lab3AStart.xlsx

Scenario:

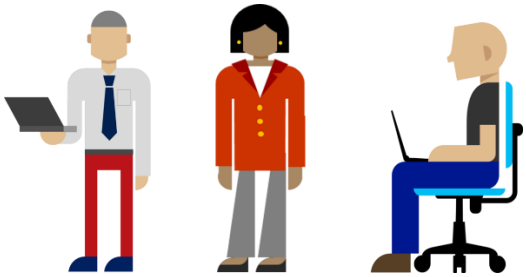
Now that you have prepared the data in an Excel table, you can start to create pivot tables to aggregate the data and create some reports.

From your conversation with Lucy, you know that she is interested in looking into the yearly sales data broken down by countries, product categories, and age groups.

Lab 3A:

Step 1

First, let's start by naming the Excel table. Name the Excel table prepared in the previous lab to **SalesTable**. From now on, every time you add a pivot table, it should be based on this SalesTable, unless mentioned otherwise.



Now, proceed to add your first pivot table. Insert a new pivot table based on the SalesTable to a new sheet. Arrange the layout so that the pivot table displays the **Product Category** and **Sub Category** in the **Rows**, **Year** in the **Columns**, and **Revenue** (Sum of) as the **Values**.

Question 1

Which year did the company start selling **Touring Bikes**?

- 2011.
- 2012
- 2013.
- 2014

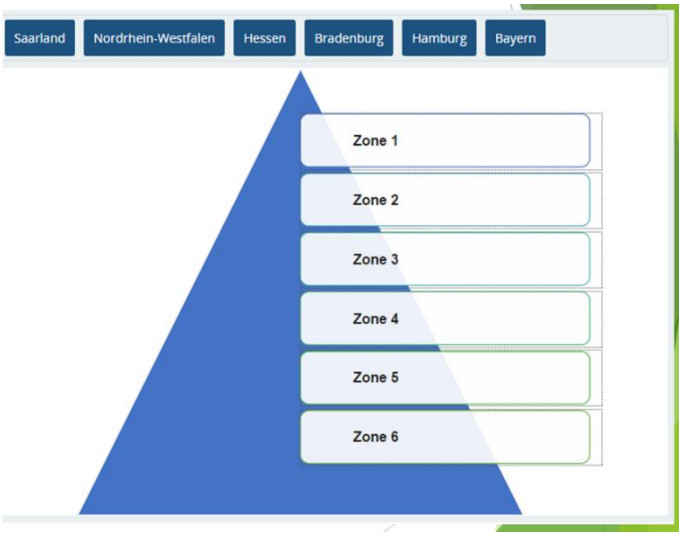
Step 2:

Insert another pivot table to the same sheet, next to the existing pivot table. Arrange the layout so that the pivot table displays the **Country** and **State** in the **Rows**, **Year** in the **Columns**, and **Revenue** (Sum of) as the **Values**. Sort the pivot table by **Sum of Revenue** so that the **Country** and **State** with the highest revenue is displayed first.

Question 2

PROBLEM

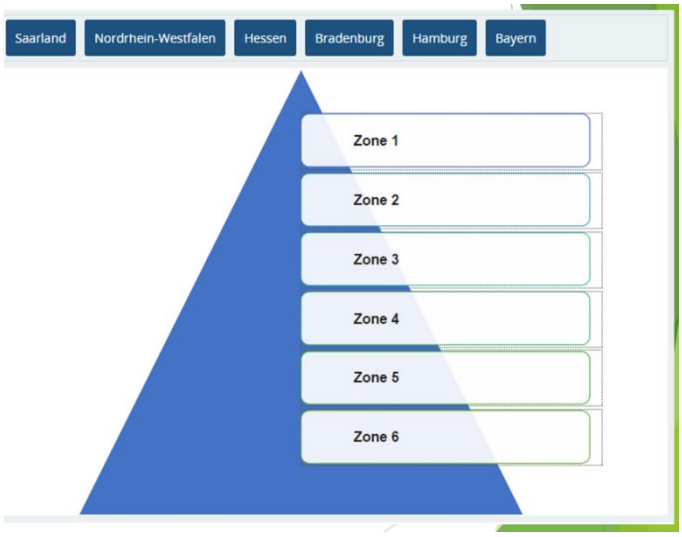
Rank the **States** for **Germany**, from the highest to lowest revenue.
Hint: Sort the **States** field in descending order by **Sum of Revenue**



Question 3

PROBLEM

What about for the year 2013?
Hint: Expand **Germany** and sort the **State** field in descending order by **Sum of Revenue** using values from the 2013 column.



Step 3:

Let's add another pivot table. This time arrange the layout so that the pivot table displays the **Frame Size** in the **Rows** and **Revenue** (Sum of) as the **Values**. Hide the rows that do not have a Frame size (blank Frame size), then sort the

pivot table by **Sum of Revenue** so that the **Frame size** with the highest revenue is displayed first.

Question 4

Which frame size sold the most?

- 40.
- 42
- 44
- 48

Step 4:

Last but not least, add another pivot table with **Age Group** as the **Rows** and **Revenue** (Sum of) as the **Values**. You will learn how to custom sort the Age Group in the next module. But for now, sort the pivot table by **Sum of Revenue** so that the **Age Group** with the highest revenue is displayed first.

Question 5

Which **Age Group** has the lowest revenue?

- Youth.
- Young Adults
- Adults
- Seniors

Lab 3B:

Step 1:

Now you can start adding some charts to the sheet.

First, add a pivot chart for the pivot table that shows yearly sales by Country. Select a **Line** chart to display the yearly trend. Make sure that the **Years** are located in the X axis, the **Revenue** in the Y axis, and the **Countries** as categories. Hint: You might find the Switch Row/Column in Select Data Source window useful. In addition, try Collapsing the Country fields in order to hide the States fields.

Question 1

In this chart, you can clearly see the sales trend for each Country.

Which country's trend is the most different when compared to the other countries?

- United States.
- Australia
- Germany.
- United Kingdom.

Step 2:

Add another pivot chart for the pivot table that shows yearly sales by Product Category. Select a **Column** chart to display the yearly sales by category so that the years are together.

Question 2

Which year does the Bikes category have the least sales?

- 2011.
- 2012
- 2013.
- 2014

Step 3:

Add another pivot chart, this time for the pivot table that shows Revenue by Age Group. Select a **Pie** chart to display the proportion of each **Age Group**. Add the **Data labels** to show percentage, formatted to two decimal points.

Question 3

In this chart, you can clearly see the proportion of sales for each **Age Group**.

What is the proportion of sales for **Young Adults**?

- 54.96%.
- 29.33%
- 35.95%
- 49.74%

Step 4:

Add another pivot chart, this time for the pivot table that shows Revenue by Frame size. Select a **Bar** chart to display the order of sales by **Frame size**. Sort the Y axis to show the **Frame size** that has the most sales on the top.
Hint: Select the **Categories in reverse order** option in the Axis Options.

Question 4

In this chart, you can clearly see the order of sales by Frame size.

Which frame size sold the least?

- 52.
- 54
- 46
- 48