**8. EXPLORING DATA WITH PIVOT TABLES**

**Introduction**

A PivotTable constructed in Excel becomes a powerful tool to calculate, summarize, and analyze data, letting you see comparisons, patterns, and trends in your data.

For example, say you have a huge amount of transaction data for a shop, and you are asked to make a report along with a dashboard to show the transactions made every year for last twenty years. It would take days to complete the task manually, but with the help of PivotTable and pivot charts, it would take barely a few minutes, as they are able to powerfully analyze large volumes of data in a few clicks and minimize the chance of errors.

In this guide we will cover the following features associated with Pivot tables:

* Creating a PivotTable
* Managing Data in PivotTable
* Formulas in PivotTables
* Exploring Data with PivotChart

**Creating a PivotTable**

Before we start learning about the creation of PivotTables, it is important to understand the data format on which they are built. A PivotTable requires tabular data and treats blank spaces separately. A complete blank row is treated as a blank value, however, if there is at least one data field available in a row, then the rest of the row data field is considered blank and the row is treated just like other rows.

To understand how blank spaces are treated in a pivot table, consider these three cases:

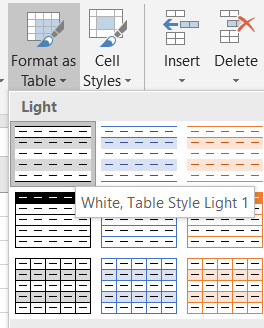
**Case 1: No Blank Values**

Consider the following table where 24 data rows are available along with four columns.

| **Month** | **Quarter** | **Shop** | **Profit in USD** |
| --- | --- | --- | --- |
| Jan | 1 | Nikodale Furniture | 2500 |
| Feb | 1 | Nikodale Furniture | 3888 |
| Mar | 1 | Nikodale Furniture | 5699 |
| Apr | 2 | Nikodale Furniture | 8875 |
| May | 2 | Nikodale Furniture | 6588 |
| Jun | 2 | Nikodale Furniture | 2233 |
| Jul | 3 | Nikodale Furniture | 6630 |
| Aug | 3 | Nikodale Furniture | 1855 |
| Sep | 3 | Nikodale Furniture | 3555 |
| Oct | 4 | Nikodale Furniture | 8795 |
| Nov | 4 | Nikodale Furniture | 1211 |
| Dec | 4 | Nikodale Furniture | 2222 |
| Jan | 1 | Samuel Arts | 6554 |
| Feb | 1 | Samuel Arts | 6899 |
| Mar | 1 | Samuel Arts | 7845 |
| Apr | 2 | Samuel Arts | 9555 |
| May | 2 | Samuel Arts | 7588 |
| Jun | 2 | Samuel Arts | 6558 |
| Jul | 3 | Samuel Arts | 5224 |
| Aug | 3 | Samuel Arts | 6554 |
| Sep | 3 | Samuel Arts | 7558 |
| Oct | 4 | Samuel Arts | 4555 |
| Nov | 4 | Samuel Arts | 3555 |
| Dec | 4 | Samuel Arts | 1555 |

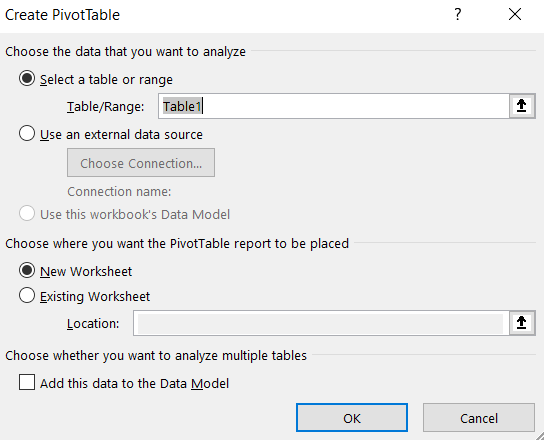
To create a PivotTable on this data, first arrange the data in the form of a table using the following hierarchy:

Menu bar > Home > Format as Table > Select a table format you like



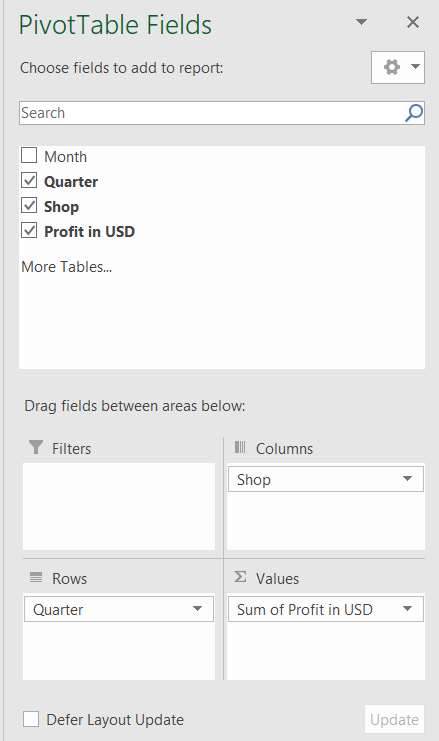
Once you get the data in a tabular form, follow this hierarchy to create a PivotTable in a new worksheet:

Menu bar > Insert > PivotTable > Press OK



This opens a new blank worksheet, and you can observe PivotTable Fields on the right side of the sheet. This box by default has two areas, Field Section and stacked Areas Section. The Field Section consists of all the column names, and the Areas Section has four sub-sections: Filters, Columns, Rows and Values. You can drag a column name from the Field Section and drop it into any of the Areas Section.

If you proceed with the following structure:



You will receive the following PivotTable:

| **Sum of Profit in USD Column Labels Row Labels** | **Nikodale Furniture** | **Samuel Arts** | **Grand Total** |
| --- | --- | --- | --- |
| 1 | 12087 | 21298 | 33385 |
| 2 | 17696 | 23701 | 41397 |
| 3 | 12040 | 19336 | 31376 |
| 4 | 12228 | 9665 | 21893 |
| Grand Total | 54051 | 74000 | 128051 |

As you can observe, there were no missing values in the original table, hence there are no blank values in the PivotTable.

**Case 2: Missing Row**

Consider a case in which the following row is missing from the table (replaced by blank values):

| **Month** | **Quarter** | **Shop** | **Profit in USD** |
| --- | --- | --- | --- |
| Sep | 3 | Nikodale Furniture | 3555 |

Now, if we try to build a PivotTable on such table, we get the following result:

| **Sum of Profit in USD Column Labels Row Labels** | **Nikodale Furniture** | **Samuel Arts** | (blank) | **Grand Total** |
| --- | --- | --- | --- | --- |
| 1 | 12087 | 21298 |  | 33385 |
| 2 | 17696 | 23701 |  | 41397 |
| 3 | 8485 | 19336 |  | 27821 |
| 4 | 12228 | 9665 |  | 21893 |
| (blank) |  |  |  |  |
| Grand Total | 50496 | 74000 |  | 124496 |

Here, blank() is considered both as a separate row and column.

**Case 3: Missing Value**

What if only few values are missing rather than a complete record? To understand the PivotTable result in such a case, let's remove the profit in the month of September from Nikodale Furniture:

| **Month** | **Quarter** | **Shop** | **Profit in USD** |
| --- | --- | --- | --- |
| Sep | 3 | Nikodale Furniture |  |

Building a PivotTable on this data results in the following:

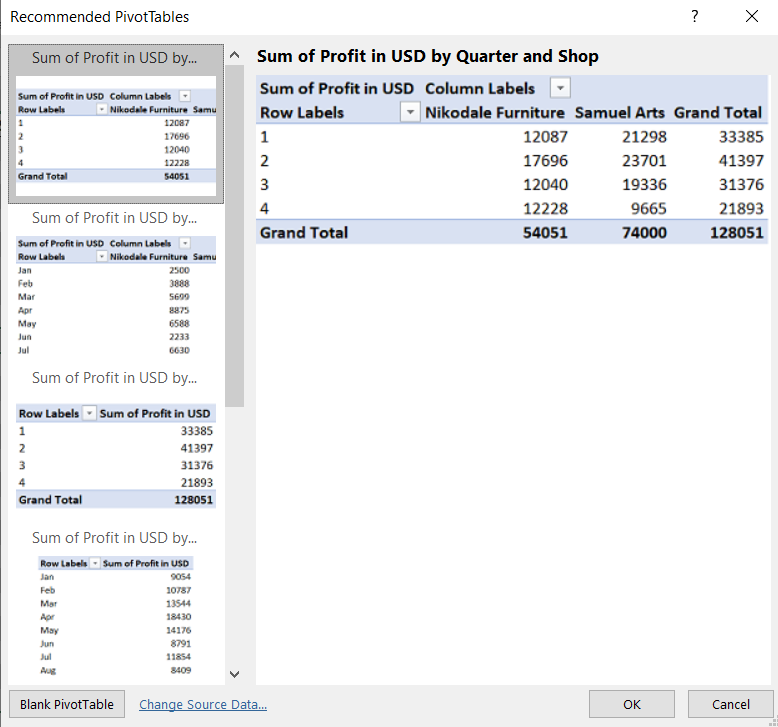
| **Sum of Profit in USD Column Labels Row Labels** | **Nikodale Furniture** | **Samuel Arts** | **Grand Total** |
| --- | --- | --- | --- |
| 1 | 12087 | 21298 | 33385 |
| 2 | 17696 | 23701 | 41397 |
| 3 | 8485 | 19336 | 31376 |
| 4 | 12228 | 9665 | 21893 |
| Grand Total | 54051 | 74000 | 128051 |

This results in a similar table as in Case 1 with a difference in the value of Quarter 3 profit for Nikoldale Furniture.

These three cases illustrate how missing row(s) or missing value(s) in the data can impact the corresponding PivotTable.

**Creating a Recommended Table**

When you click on the Insert tab in the Menu bar, you can see the Recommended PivotTable option adjacent to the PivotTable option. Clicking on this option pops up a new dialog box that consists of most of the common PivotTables that can be built using the provided data. For cases in which there are no missing values in the data, the recommended PivotTables are shown below:



**Managing Data in PivotTables**

So far, we have learn to create a PivotTable. Now let's learn about managing its items using this table:

| **Year** | **Quarter** | **Nikodale Furniture** | **Samuel Arts** |
| --- | --- | --- | --- |
| 2019 | 1 | 4548 | 2500 |
| 2019 | 2 | 7548 | 3888 |
| 2019 | 3 | 2154 | 5699 |
| 2019 | 4 | 8875 | 8875 |
| 2018 | 1 | 6588 | 4578 |
| 2018 | 2 | 2233 | 4221 |
| 2018 | 3 | 6630 | 6584 |
| 2018 | 4 | 1855 | 1452 |

First, we build a PivotTable out of this table keeping the following Pivot structure:

* Rows: Year followed by Quarter
* Values: Sum of Nikodale Furniture followed by Sum of Samuel Arts

This leads us to the given result:

| **Row Labels** | **Sum of Nikodale Furniture** | **Sum of Samuel Arts** |
| --- | --- | --- |
| 2018 |  |  |
| 1 | 6588 | 4578 |
| 2 | 2233 | 4221 |
| 3 | 6630 | 6584 |
| 4 | 1855 | 1452 |
| 2019 |  |  |
| 1 | 4548 | 2500 |
| 2 | 7548 | 3888 |
| 3 | 2154 | 5699 |
| 4 | 8875 | 8875 |
| Grand Total | 40431 | 37797 |

**Subtotal of Group**

As you can observe, the PivotTable has two subsections (2018 and 2019) with a final Grand Total. To get the total of each section separately, we can follow these steps:

Menu bar > Design > Subtotal > Choose any option (here, Show all Subtotals at the Bottom of Group)

| **Row Labels** | **Sum of Nikodale Furniture** | **Sum of Samuel Arts** |
| --- | --- | --- |
| 2018 |  |  |
| 1 | 6588 | 4578 |
| 2 | 2233 | 4221 |
| 3 | 6630 | 6584 |
| 4 | 1855 | 1452 |
| 2018 Total | 17306 | 16835 |
| 2019 |  |  |
| 1 | 4548 | 2500 |
| 2 | 7548 | 3888 |
| 3 | 2154 | 5699 |
| 4 | 8875 | 8875 |
| 2019 Total | 23125 | 20962 |
| Grand Total | 40431 | 37797 |

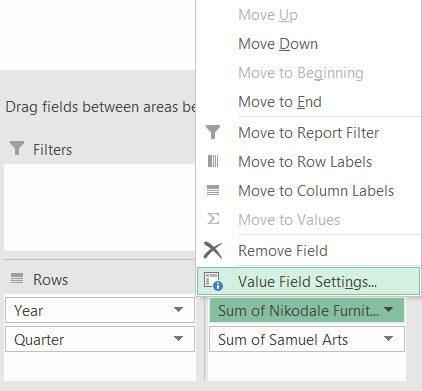
You can also control the display of Grand Total from the same menu.

**Controlling Field Settings**

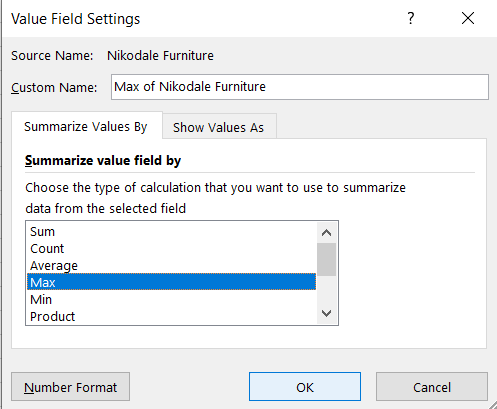
You may have observed that when we drop a field into the Values section, it automatically calls the Sum function. However, we can control what function needs to be implemented on a particular value. Let's try to change the following:

* Nikodale Furniture: From Sum to Maximum value
* Samuel Arts: From Sum to Average value

To accomplish this change, click on Sum of Nikodale Furniture under the Values section and select the Value Field Settings... option.



Now, click on Max and press OK.

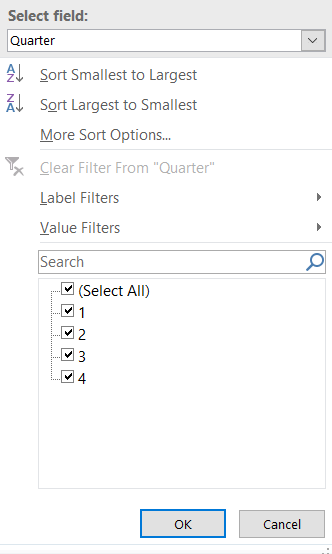


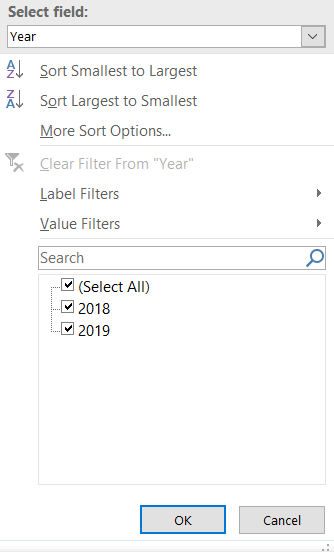
Perform a similar operation with Samuel Arts, but this time select Average. Once you've done both operations, you will receive the following PivotTable:

| **Row Labels** | **Max of Nikodale Furniture** | **Average of Samuel Arts** |
| --- | --- | --- |
| 2018 |  |  |
| 1 | 6588 | 4578 |
| 2 | 2233 | 4221 |
| 3 | 6630 | 6584 |
| 4 | 1855 | 1452 |
| 2018 Total | 6630 | 4208.75 |
| 2019 |  |  |
| 1 | 4548 | 2500 |
| 2 | 7548 | 3888 |
| 3 | 2154 | 5699 |
| 4 | 8875 | 8875 |
| 2019 Total | 8875 | 5240.5 |
| Grand Total | 8875 | 4724.625 |

**Filtering**

Just as we can apply a filter on a regular table in Excel, we can also apply a similar filter on a PivotTable. For instance, consider the table above. If you click on any year's quarter cell (1, 2, 3, or 4) and then click on the dropdown button next to the Row Labels column, you'll find an option to select all or specific quarters. Similarly, if you click on a cell with year value 2018 or 2019, the dropdown values change from quarter to year. The images below represent the dropdown box in each case:





Instead of clicking on a cell to select a field (Quarter or Year), you can also choose them right from the Select Field option available in the dropdown box.

The PivotTable below shows the data for only quarters 1 and 2 in 2018:

| **Row Labels** | **Max of Nikodale Furniture** | **Average of Samuel Arts** |
| --- | --- | --- |
| 2018 |  |  |
| 1 | 6588 | 4578 |
| 2 | 2233 | 4221 |
| 2018 Total | 6588 | 4399.5 |
| Grand Total | 6588 | 4399.5 |

**Layout Transformation**

In the above PivotTables, you may have observed that there is no separate boundary to distinguish a group from another (here, 2018 data from 2019 data). Plus, the year and quarter appear in a single column. PivotTables incude options under the Design tab named Report Layout and Blank Rows to tackle these issues. Let's learn them step by step.

**Adding and Removing Blank Rows in a PivotTable**

Consider a PivotTable with two groups as shown below:

| **Row Labels** | **Sum of Nikodale Furniture** | **Sum of Samuel Arts** |
| --- | --- | --- |
| 2018 |  |  |
| 1 | 6588 | 4578 |
| 2 | 2233 | 4221 |
| 3 | 6630 | 6584 |
| 4 | 1855 | 1452 |
| 2019 |  |  |
| 1 | 4548 | 2500 |
| 2 | 7548 | 3888 |
| 3 | 2154 | 5699 |
| 4 | 8875 | 8875 |
| Grand Total | 40431 | 37797 |

Notice that we can improve the visual of this table by adding a blank row below 2019 year row. To achieve this, follow these steps:

Menu bar > Design > Blank Rows > Insert Blank Line After Each Item

This operation results in the following PivotTable:

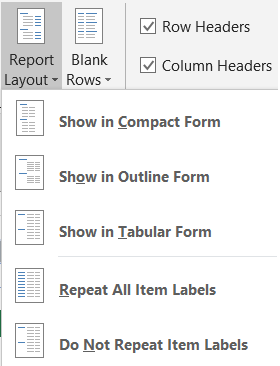
| **Row Labels** | **Sum of Nikodale Furniture** | **Sum of Samuel Arts** |
| --- | --- | --- |
| 2018 |  |  |
| 1 | 6588 | 4578 |
| 2 | 2233 | 4221 |
| 3 | 6630 | 6584 |
| 4 | 1855 | 1452 |
|  |  |  |
| 2019 |  |  |
| 1 | 4548 | 2500 |
| 2 | 7548 | 3888 |
| 3 | 2154 | 5699 |
| 4 | 8875 | 8875 |
| Grand Total | 40431 | 37797 |

To remove the blank row, follow these steps:

Menu bar > Design > Blank Rows > Remove Blank Line After Each Item

**Changing Table Layout**

The Report Layout option under the Design tab has the following options:



By default, the PivotTable comes with the first option, Show in Compact Form. Let's use another format, Show in Outline Form, to separate Year from Quarter, which results in the following PivotTable:

| **Year** | **Quarter** | **Sum of Nikodale Furniture** | **Sum of Samuel Arts** |
| --- | --- | --- | --- |
| 2018 |  | 17306 | 16835 |
|  | 1 | 6588 | 4578 |
|  | 2 | 2233 | 4221 |
|  | 3 | 6630 | 6584 |
|  | 4 | 1855 | 1452 |
| 2019 |  | 23125 | 20962 |
|  | 1 | 4548 | 2500 |
|  | 2 | 7548 | 3888 |
|  | 3 | 2154 | 5699 |
|  | 4 | 8875 | 8875 |
| Grand Total |  | 40431 | 37797 |

Now you can test rest of the options and observe the difference in the table.

**Formulas in PivotTables**

In this section, we will learn about two formulas:

* GETPIVOTDATA
* Calculated Field

**GETPIVOTDATA**

Sometimes you start with a PivotTable structure, but the structure may change depending upon the requirements. Therefore, a function named GETPIVOTDATA is suggested in scenarios where you want to keep the information regardless of changes in the PivotTable structure.

For instance, consider the following PivotTable:

| **Row Labels** | **Max of Nikodale Furniture** | **Average of Samuel Arts** |
| --- | --- | --- |
| 2018 |  |  |
| 1 | 6588 | 4578 |
| 3 | 6630 | 6584 |
| 2018 Total | 6588 | 5581 |
| Grand Total | 6588 | 5581 |

Here, if you want to keep the maximum profit value ($6630) of Nikodale Furniture in Quarter 3 of year 2018 in a separate cell regardless of changes in the PivotTable structure, then proceed with the following steps:

* Click on a new cell and write =.
* Next, click on the cell with value 6630. This will result in the following formula in the new cell, which you have selected: =GETPIVOTDATA("Max of Nikodale Furniture",$A$3,"Year",2018,"Quarter",3).
* Press Enter, which leaves you with the value 6630 in the cell with the formula.

So far, we are able to retrieve the value using the formula. Now let's change the structure of the table by adding Quarter 2 into the table, which changes the location of the cell with value 6630.

| **Row Labels** | **Max of Nikodale Furniture** | **Average of Samuel Arts** |
| --- | --- | --- |
| 2018 |  |  |
| 1 | 6588 | 4578 |
| 2 | 2233 | 4221 |
| 3 | 6630 | 6584 |
| 2018 Total | 6588 | 5127.67 |
| Grand Total | 6588 | 5127.67 |

As you can see, the value has remain unchanged!

**Calculated Field**

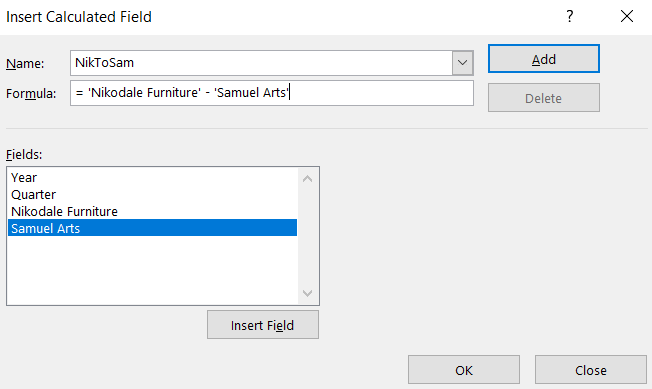
Just as we can add a new column to an Excel table by connecting them with a formula, in a similar fashion we can add a new field to the PivotTable. Consider the PivotTable below:

| **Row Labels** | **Max of Nikodale Furniture** | **Max of Samuel Arts** |
| --- | --- | --- |
| 2018 |  |  |
| 1 | 6588 | 4578 |
| 2 | 2233 | 4221 |
| 3 | 6630 | 6584 |
| 4 | 1855 | 1452 |
| 2018 Total | 6630 | 6584 |
| 2019 |  |  |
| 1 | 4548 | 2500 |
| 2 | 7548 | 3888 |
| 3 | 2154 | 5699 |
| 4 | 8875 | 8875 |
| 2019 Total | 8875 | 8875 |
| Grand Total | 8875 | 8875 |

So to add a new column that represents the difference between columns Max of Nikodale Furniture and Max of Samuel Arts, use the following steps:

Menu bar > Analyze > Fields, Items & Sets > Calculated Fields...

In the Insert Calculated Field dialog box, provide a new column name and suitable formula, as shown in the image below:



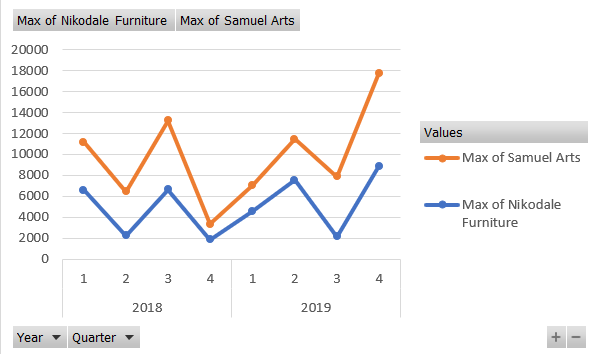
This results in the following PivotTable:

| **Row Labels** | **Max of Nikodale Furniture** | **Max of Samuel Arts** | **Sum of NikToSam** |
| --- | --- | --- | --- |
| 2018 |  |  |  |
| 1 | 6588 | 4578 | 2010 |
| 2 | 2233 | 4221 | -1988 |
| 3 | 6630 | 6584 | 46 |
| 4 | 1855 | 1452 | 403 |
| 2018 Total | 6630 | 6584 | 471 |
| 2019 |  |  |  |
| 1 | 4548 | 2500 | 2048 |
| 2 | 7548 | 3888 | 3660 |
| 3 | 2154 | 5699 | -3545 |
| 4 | 8875 | 8875 | 0 |
| 2019 Total | 8875 | 8875 | 2163 |
| Grand Total | 8875 | 8875 | 2634 |

**Exploring Data with PivotChart**

We can also visualize the data available in a PivotTable using PivotChart, provided under the Insert menu. We build a stacked line chart with markers on the following PivotTable:

| **Row Labels** | **Max of Nikodale Furniture** | **Max of Samuel Arts** |
| --- | --- | --- |
| 2018 |  |  |
| 1 | 6588 | 4578 |
| 2 | 2233 | 4221 |
| 3 | 6630 | 6584 |
| 4 | 1855 | 1452 |
| 2018 Total | 6630 | 6584 |
| 2019 |  |  |
| 1 | 4548 | 2500 |
| 2 | 7548 | 3888 |
| 3 | 2154 | 5699 |
| 4 | 8875 | 8875 |
| 2019 Total | 8875 | 8875 |
| Grand Total | 8875 | 8875 |



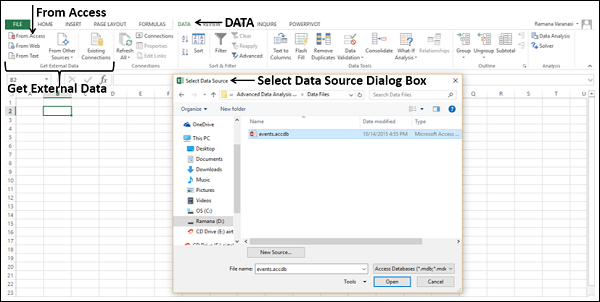
From the above chart, it is easy to infer that the profit of Samuel Arts is always higher than the profit earned by Nikodale Furniture in all the quarters of 2018 and 2019.

ADDITIONAL OPTIONS

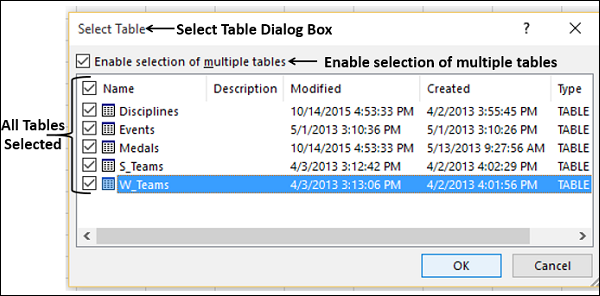
Creating a PivotTable to analyze External Data

To create a PivotTable to analyze external data −

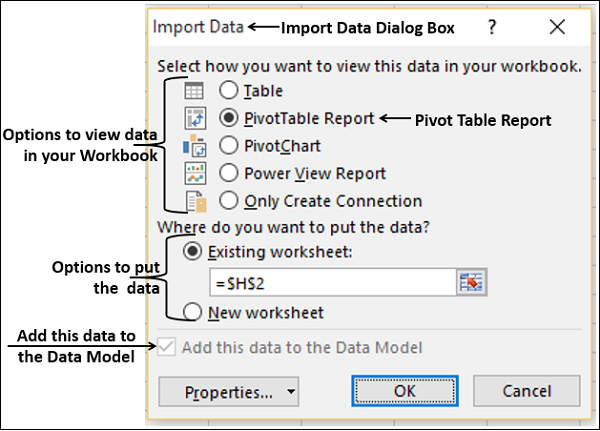
* Open a new blank workbook in Excel.
* Click the DATA tab on the Ribbon.
* Click From Access in the Get External Data group. The Select Data Source dialog box appears.
* Select the Access database file.



* Click the Open button. The Select Table dialog box appears, displaying the tables in the database. Access database is a relational database and the tables will be similar to Excel tables, with the exception that relationships exist among those tables.
* Check the box Enable selection of multiple tables.
* Select all the tables. Click OK.

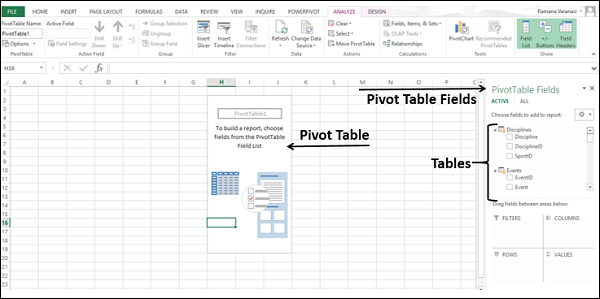


The **Import Data** dialog box appears. Select PivotTable Report. This option imports the tables into your Excel workbook and creates a PivotTable for analyzing the imported tables.



As you observe, the checkbox Add this data to the Data Model is selected and disabled, indicating that the tables will be added to the Data Model automatically.

The data will be imported and an empty PivotTable will be created. The imported tables appear in the PivotTable Fields list.



Exploring Data using PivotTable

You know how to add fields to PivotTable and drag fields across areas. Even if you are not sure of the final report that you want, you can play with the data and choose the appropriate report.

Suppose you want to have a report displaying the following −

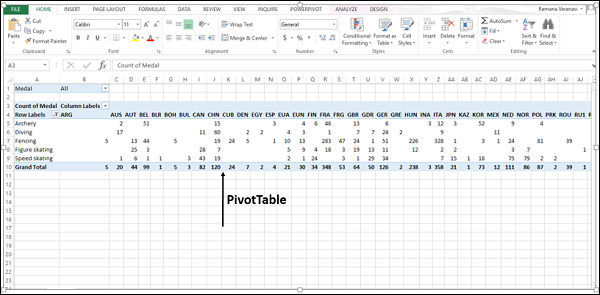
* Data for five disciplines - Archery, Diving, Fencing, Figure Skating and Speed Skating.
* Regions that scored more than 80 medals in these 5 disciplines.
* The count of medals in each of the five disciplines in each of these regions.
* Total count of medals for the five disciplines in each of these regions.

You can see how easily you can create this report in few steps.

To start with, create a PivotTable displaying the count of medals in all the regions for the selected five disciplines as follows −

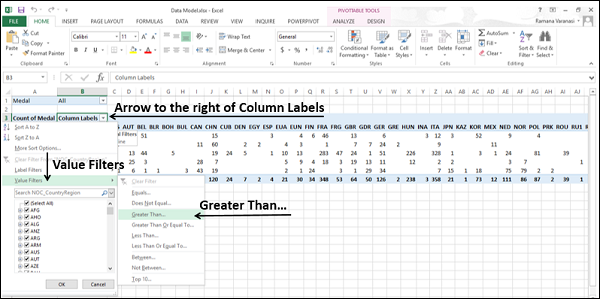
* Drag the NOC\_CountryRegion field from the Medals table to the COLUMNS area.
* Drag Discipline from the Disciplines table to the ROWS area.
* Filter Discipline to display only the five disciplines for which you wanted the report. This can be done either in the PivotTable Fields area, or from the Row Labels filter in the PivotTable itself.
* Drag Medal from the Medals table to the VALUES area.
* Drag Medal from the Medals table to the FILTERS area.

You will get the following PivotTable −

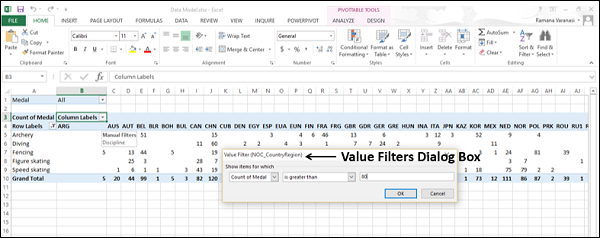


As you observe, Count of Medals is displayed for all the regions and for the five disciplines that you selected. Next, you have to fine-tune this report so that only those regions with total count of medals greater than 80 will be displayed.

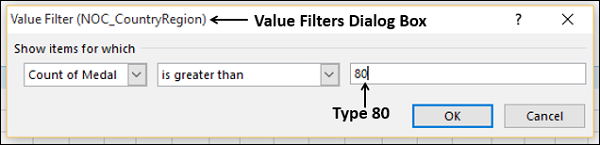
* Click the arrow button to the right of Column Labels.
* Click Value Filters in the drop-down list that appears.
* Select **Greater Than…** from the drop-down list that appears.



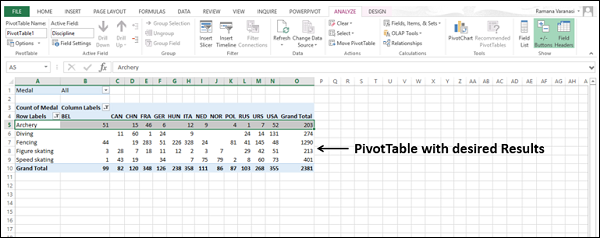
Value Filters dialog box appears.



As you observe, Count of Medals and is greater than are displayed in the boxes below **Show items for which**. Type 80 in the box next to the box containing is greater than and click OK.



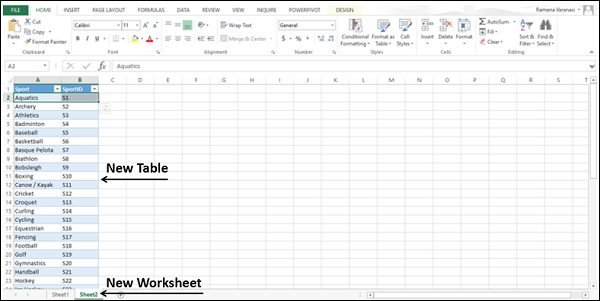
Now, the PivotTable displays only those regions with total count of medals in the selected five disciplines greater than 80.



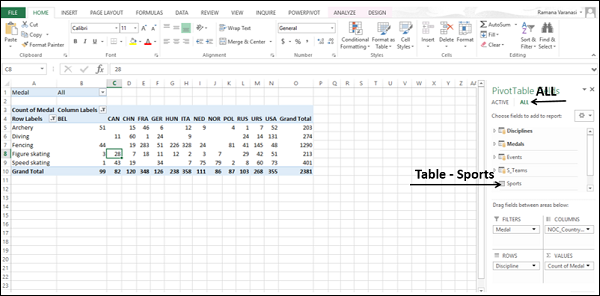
Creating a Relationship between Tables with PivotTable Fields

If you do not import the tables at the same time, if the data is from different sources, or if you add new tables to your workbook, you have to create the relationships among the tables by yourself.

Add a new worksheet with a table that contains Sport and SportID fields to your workbook.



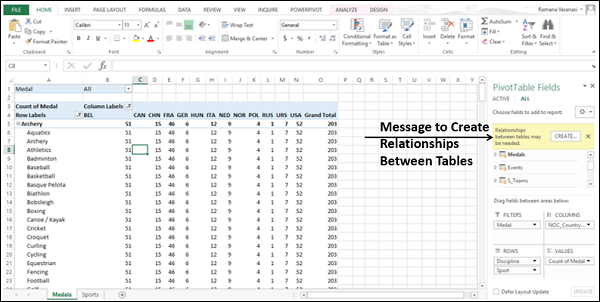
* Name the table - **Sports**.
* Click ALL in the PivotTable Fields list in the PivotTable worksheet.



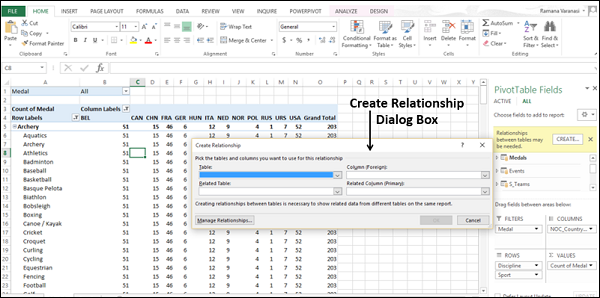
You can see that the newly added table- Sports is also visible in the PivotTable Fields list.

Next, add the field Sport also to the PivotTable as follows −

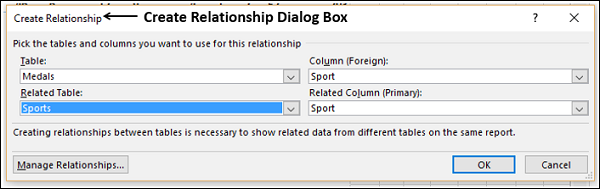
* Drag the field Sport from the table Sports to ROWS area. The Sport values appear as Row labels in the PivotTable.
* A message will appear in the PivotTable Fields list that Relationships between tables may be needed. A CREATE button appears next to the message.



Click the CREATE button. The Create Relationship dialog box appears.



* Select Medals under Table.
* Select Sport under Column.
* Select Sports under Related Table. Sport appears under related column.
* Click OK.



Drag Discipline under **Sport** in **ROWS**. This is to define the hierarchy in the PivotTable. The PivotTable displays the Sport and the corresponding group of disciplines for that sport.

