

ANALYZE DATA WITH EXCEL POWER PIVOT: A TUTORIAL

Radhika Kaula

Management and Information Technology Department
Missouri State University
901 South National Ave.
Springfield, MO 65897
E-Mail: RadhikaKaula@missouristate.edu

Rajeev Kaula

Management and Information Technology Department
Missouri State University
901 South National Ave.
Springfield, MO 65897
Phone: 417-836-5666
E-Mail: RajeevKaula@missouristate.edu

Abstract

Excel is one of the most widely used application in businesses. It is often used as a repository of data for creating budgets, accounting, financial statements, stock analysis, forecasting, trend analysis and so on. However, educators and students often struggle to perform data analysis in a simple and efficient manner. This paper presents an extremely straightforward approach in the form of a tutorial to create related database tables and then illustrate data analysis in the form of queries that retrieve data from multiple tables even if tables don't have a direct database relationship. The tutorial utilizes Data Model and Power Pivot features for data analysis. Upon completion of this tutorial the student will understand how to use these advanced features and apply them successfully for data analytics.

Keywords: Data Model, Database, Excel, Power Pivot, Tutorial

Introduction

Excel is a powerful spreadsheet application that is widely used among business and academics (Rosenberg, 2015). Even though many organizations utilize Excel workbooks as a repository for data, due to the lack of proper database features it was often overlooked for qualitative data analysis. One basic problem was that Excel tables could not have database type relationships (Cram et.al, 2012). However, since Office 2013 Excel has included more database features in the form of Power Pivot and Data Model that have enhanced its data analytic capabilities to support decision-making (Meyer and Leanne, 2008; Ferrari & Russo, 2010; Palocsay et.al, 2010). These features besides facilitating interactive handling of large data sets (Zhimin et.al, 2014) also provide the ability to manipulate, filter, and sort millions of data without SQL skills (Barry, 2011).

As organizations increasingly emphasize business intelligence and data analytics (Chen et.al, 2012), understanding Excel's data analytics features are essential for students and professionals. Despite the widespread utilization of office software, Excel text books (Monk et.al, 2017; Cram et.al, 2012; Parsons et.al, 2010; Gaskin and Vargas, 2017; Reading & Wermers, 2017; Smart, 2017; Frye, 2017; Monk et.al, 2017) essentially focus on a fixed set of features like formatting spreadsheet, range names, budgeting, time value of money functions, what-if analysis, breakeven analysis, weighted criteria analysis, and charts. Some of the advanced features like Power Pivot, Solver, Regression, Linear programming, and Scenario analysis are covered briefly. Since data analytics in Excel is associated with the Power Pivot feature, its coverage is often very brief and limited to just one table or tables that have direct relationship with each other. But, if the tables don't have direct relationship then Excel has a limitation with respect to querying data from such tables.

The objective of this paper is to (i) provide a simple tutorial on how to create database lists of related tables through the Data Model feature in Power Pivot, and then (ii) illustrate data retrieval for analytical purposes in the form of queries that involve tables that don't have direct relationship. Due to the limitation of Excel to facilitate retrieval of data from tables that don't have direct relationship, the paper utilizes the Joiner concept (Jelen & Alexander, 2013; Jelen, 2017) to facilitate such querying.

The tutorial uses a fictitious real estate company to provide a business perspective. As Excel is learned by professionals and students of all majors, this tutorial will broaden their understanding. No prior knowledge of database is required for this tutorial. The tutorial outline will be as follows: (a) install Power Pivot plug-in, (b) create database tables with relationships, and (c) data retrieval for analytics.

Install Power Pivot

Power Pivot is an Excel add-in tool that allows addition or import of data from multiple sources into one workbook. To install Power Pivot open the File menu and select Options. In the Options window click the Add-Ins label from the list in the left navigation pane. Now the options for Add-Ins appear in the right pane of the window. Go to the Manage label at the bottom and select COM Add-Ins from the drop-down box (as shown in Figure 1) and press the Go button beside it.

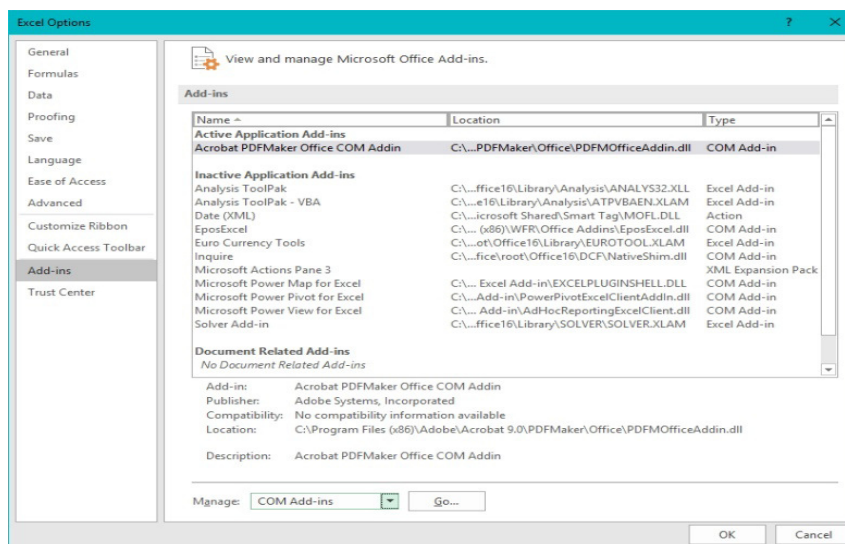


Figure 1 Select Excel Add-Ins Tools

In the ensuing COM Add-Ins window select the check box beside Microsoft Power Pivot for Excel and press the OK button (as shown in Figure 2). POWERPIVOT tab will now appear on the ribbon. You may have to restart your computer to get it installed.

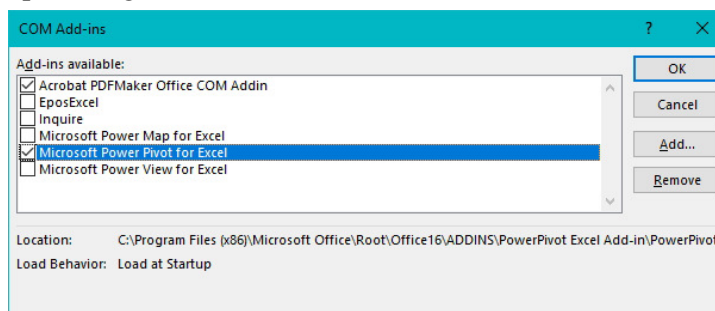


Figure 2: Select Power Pivot Add-In for Install

Create Database Tables with Relationships

A database is a collection of tables consisting of rows and columns on any information a business wishes to store (Kinser et.al, 2017). As the tutorial is based on the database requirements of a real estate company, for simplicity purposes the tutorial will have three database tables in the form of agents, listings, and customers.

Agents table will have information on real estate agents. Listings table will store information on properties that are listed with the real estate company. Customers table will store information about customers who have shown interest in buying properties available with the company. The Agents table will have database relationship with Listings and Customers tables. The steps to create the three tables are described now.

Create Agents table:

1. Open a blank workbook. Type the name of the company “Real Estate Company” in cell A1.
2. From A2:E2, enter the following names in each cell: First Name, Last Name, Email, Phone, and Agent ID.
3. Enter information in cells A3:E5 as shown in Figure 3.

	A	B	C	D	E
1	Real Estate Company				
2	First Name	Last Name	Email	Phone	Agent ID
3	John	Davis	jd123@gmail.com	417-123-4567	JD12
4	Mary	Jones	mj234@gmail.com	417-555-6565	MJ23
5	Katherine	David	kd587@gmail.com	417-622-2929	KD58

Figure 3: Agents Table Contents

4. Now we will transform the entered data into a database table. Select cells A2:E5 and go to Insert tab and click Table in the Tables group. Create Table dialog box will open as shown in Figure 4. In the Create Table dialog box, it is important to check “My table has headers” box. This will allow the column names entered in row 2 to be used as table headers.



Figure 4 Reference Agents Table contents as Table

5. In Excel table a header is referred as a field. Each field should have a unique and meaningful name. Further, a table name must not have any space in between the words. If two or more words need to be joined, an underscore should be used to conjoin the words, eg. Agents_Info.
6. As a database table is a collection of records, the information entered in rows 3 to 5 represent records.
7. A database table will have one primary key field. A primary key will have a unique value in each row. Primary keys provide unique identification to each record in the table. For the Agents table, Agent ID field will serve as the primary key field.
8. Once the table is created, it should be saved with a specific name. So, while the table is selected, in the Design tab of the Table Tools, enter “Agents” as a table name in the Properties group as shown by arrow in Figure 5.

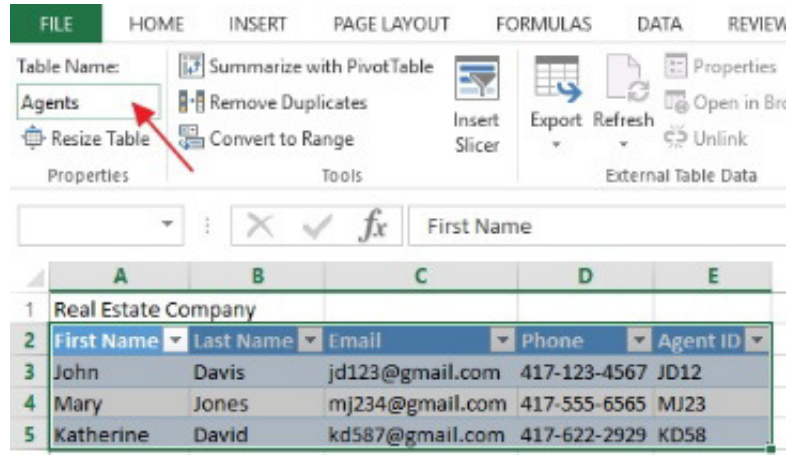


Figure 5 Create Agents Table

- The Agents table will now be added to the Power Pivot Data Model. So, click anywhere in the table and go to the Power Pivot tab and click “Add to Data Model” in Tables group. Agents table is now added to the Data Model as shown in Figure 6.

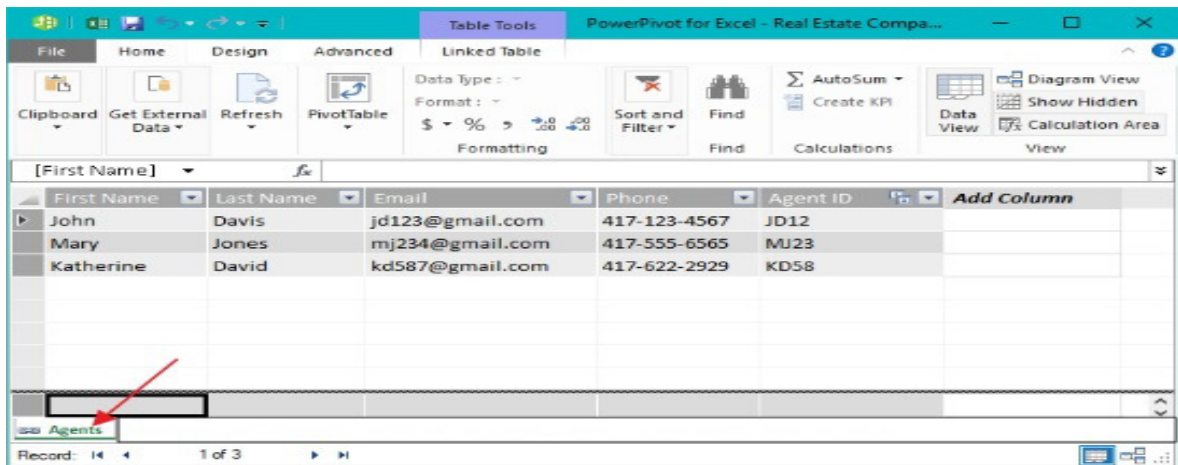


Figure 6 Add Agents Table to Power Pivot Data Model

Create Listings Table

- Create another sheet and enter the cell entries A2:K10 as shown in Figure 7 to create the Listings table similar to the way the Agents table was created. The Listings table will have the following fields: Listing ID, Agent ID, Type, List Date, Year Built, Area, Bedroom, Bathroom, Garage, Pool, and Price. The table will be populated with eight rows as reflected in range A3:K10. “Listing ID” will be the primary key of the Listings table.
- Save the table as “Listings” and then add to the Power Pivot Data Model.

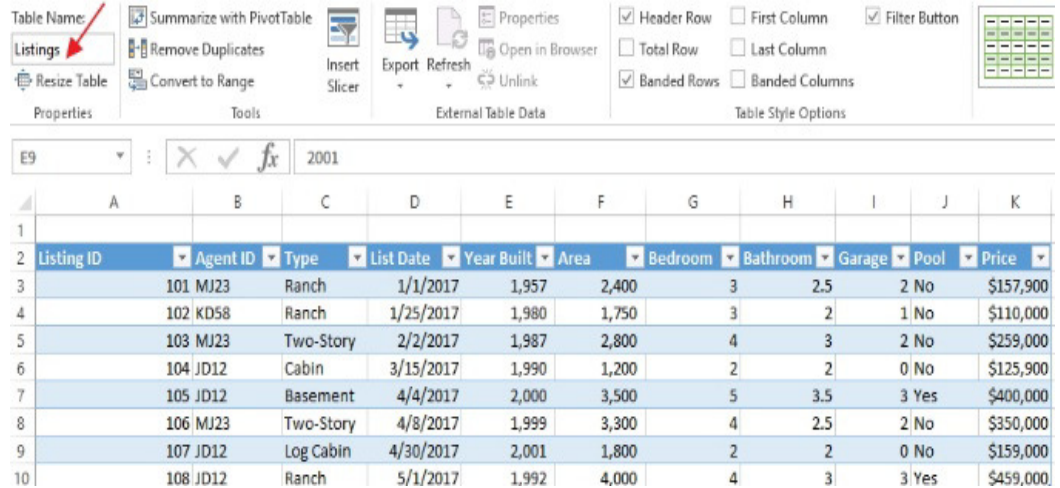


Table Name: Listings

Options: Summarize with PivotTable, Remove Duplicates, Convert to Range, Insert Slicer, Export, Refresh, Open in Browser, Unlink, Properties, Filter Button, First Column, Last Column, Total Row, Banded Columns, Header Row, Banded Rows

Listing ID	Agent ID	Type	List Date	Year Built	Area	Bedroom	Bathroom	Garage	Pool	Price
101	MJ23	Ranch	1/1/2017	1,957	2,400	3	2.5	2	No	\$157,900
102	KD58	Ranch	1/25/2017	1,980	1,750	3	2	1	No	\$110,000
103	MJ23	Two-Story	2/2/2017	1,987	2,800	4	3	2	No	\$259,000
104	JD12	Cabin	3/15/2017	1,990	1,200	2	2	0	No	\$125,900
105	JD12	Basement	4/4/2017	2,000	3,500	5	3.5	3	Yes	\$400,000
106	MJ23	Two-Story	4/8/2017	1,999	3,300	4	2.5	2	No	\$350,000
107	JD12	Log Cabin	4/30/2017	2,001	1,800	2	2	0	No	\$159,000
108	JD12	Ranch	5/1/2017	1,992	4,000	4	3	3	Yes	\$459,000

Figure 7 Listings Table

Create Customers table:

3. Create another sheet and enter the cell entries A2:H8 as shown in Figure 8 to create Customers table similar to how Agents and Listings tables were created. The Customers table will have the following fields: Customer ID, First Name, Last Name, Email, Agent ID, Phone, House Type, and Price Limit. The table will be populated with six rows as reflected in range A3:H8. “Customer ID” will be the primary key of the table.
4. Save the table as “Customers” and then add to the Power Pivot Data Model.

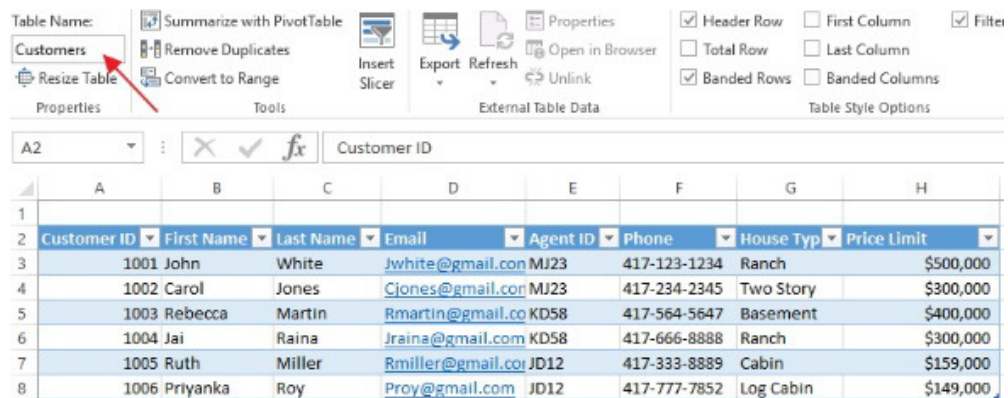


Table Name: Customers

Options: Summarize with PivotTable, Remove Duplicates, Convert to Range, Insert Slicer, Export, Refresh, Open in Browser, Unlink, Properties, Filter, First Column, Last Column, Total Row, Banded Columns, Header Row, Banded Rows

Customer ID	First Name	Last Name	Email	Agent ID	Phone	House Typ	Price Limit
1001	John	White	Jwhite@gmail.com	MJ23	417-123-1234	Ranch	\$500,000
1002	Carol	Jones	Cjones@gmail.com	MJ23	417-234-2345	Two Story	\$300,000
1003	Rebecca	Martin	Rmartin@gmail.com	KD58	417-564-5647	Basement	\$400,000
1004	Jai	Raina	Jraina@gmail.com	KD58	417-666-8888	Ranch	\$300,000
1005	Ruth	Miller	Rmiller@gmail.com	JD12	417-333-8889	Cabin	\$159,000
1006	Priyanka	Roy	Proy@gmail.com	JD12	417-777-7852	Log Cabin	\$149,000

Figure 8 Customers Table

Create Database Relationships

To access information from multiple tables, database relationships are necessary between the tables. A database relationship is the linking of one row of one table with one or more row of another table. This linking is done through a field called foreign key. A foreign key field in one table will have values that must match the primary key field values of another table. There are three types of database relationship:

one to one, one to many, and many to many. Of the three database relationships, one to many is the most common type of relationship. In this relationship the foreign key field will be in the table that has many rows linked to one row of another table's primary key field value.

In this tutorial, since an agent may have one or more customers, Agents table has one to many relationship with Customers table. As Customers table may have many rows or records associated with one record of Agents table, the foreign key field will be in Customers table. The "Agents ID" field created earlier in the Customers table will now be the foreign key field.

Similarly, an agent may have one or more listings. Agents table now has one to many relationship with Listings table. As Listings table may have many rows or records associated with one record of Agent table, the foreign key field will be in Listings table. The "Agent ID" field created earlier in the Listings table will now be the foreign key field.

Database relationships are also created through the Data Model feature in the Power Pivot tool in the form of joined tables. This feature is essential in retrieval of the data from two or more tables.

Steps to set up Database Relationship between Agents and Listings table

The following steps will create database relationships. First the relationship between Agents and Listings tables will be setup followed by relationship between Agents and Customers tables.

1. In the Power Pivot tab, click Manage under Data Model group. In the Design tab, click Create Relationship under Relationships group as shown by red arrow in Figure 9.



Figure 9 Invoke Create Relationship in Power Pivot Data Model

2. Create Relationships dialog box will open. Select Listings table first from the Table drop down menu and Agent ID field under Column drop down menu. Then select Agents table and Agent ID field. And click create button as shown in Figure 10. The table that has a primary key should be selected as the second table.

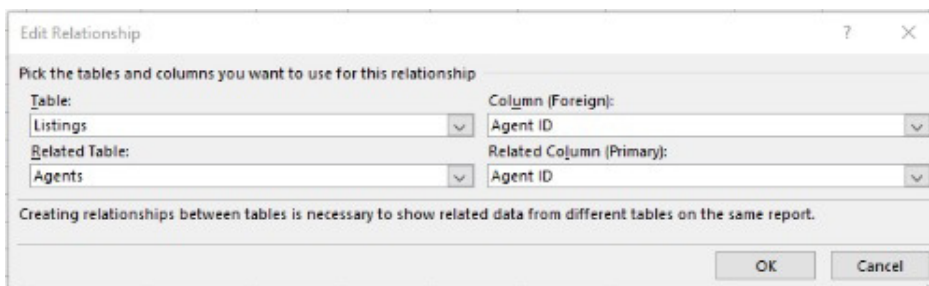


Figure 10 Create Relationship between Listings and Agent tables

- Next to set up relationship between Customers and Agents table, click Manage Relationships under the Relationships group and click Create button as shown in Figure 11. Previous relationship between Agents and Listings table is shown as active. Keep in mind that once a relationship is set up, it can be deleted or made inactive if it is no longer needed.

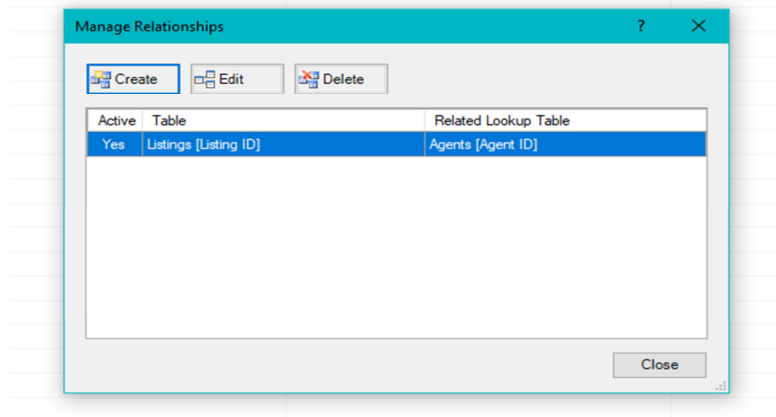


Figure 11 Manage Relationship to View, Create, and Maintain Relationships

- Like the setting of relationship between Agents and Listings table, select Customers table from the table drop down menu and pick Agent ID field first as the foreign key, and then select the Agents table and pick Agent ID field as the primary key. Now click the Create button. All three tables are now connected.

From a database perspective now individual rows can be inserted or deleted, and row values can be updated directly in their respective tables. Thereafter to ensure consistency in database relationships the “update all” feature in Power Pivot can be invoked to update the overall data model as shown in Figure 12.

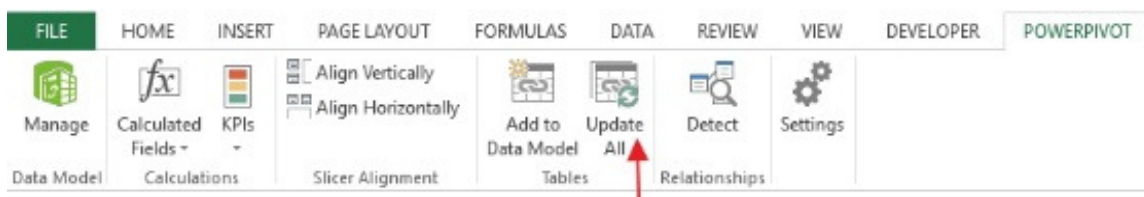


Figure 12 Update overall Data Model

Data Retrieval for Analytics

Retrieval of information for analysis from the tables is explained in two steps: (a) retrieval and analysis from one or two tables to understand the basics of querying, and then (b) retrieval and analysis from multiple tables with Joiner. To facilitate data retrieval and analysis Pivot Table feature is utilized. Pivot Table summarizes information from Excel data sets. It can provide summary of information in two dimensions by using two fields to group the records and then use a third field to summarize the

information. Since data visualization is a common way to express analytics, Pivot Charts are utilized to illustrate the results.

Retrieval from One or Two Tables

Information from Listings table can be retrieved through Pivot Table as follows:

1. In the Power Pivot tab, click Manage under Data Model group. Go to the Listings tab and click Pivot table. Create PivotTable dialog box will open as in Figure 13.

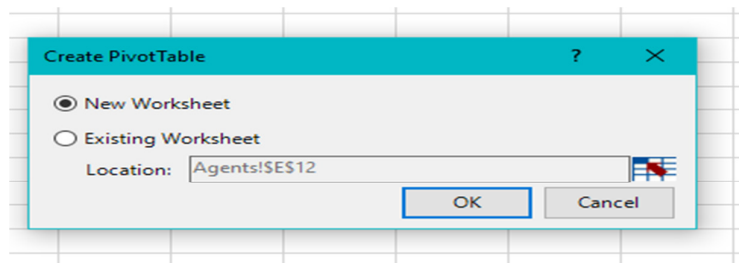


Figure 13 Create Power Pivot Table

Click New Worksheet radio button to move the table to a new worksheet and press the OK button. Placeholders for Pivot Table will be created on the left side of the worksheet and Pivot Table Fields task pane will open on the right side as shown in Figure 14.

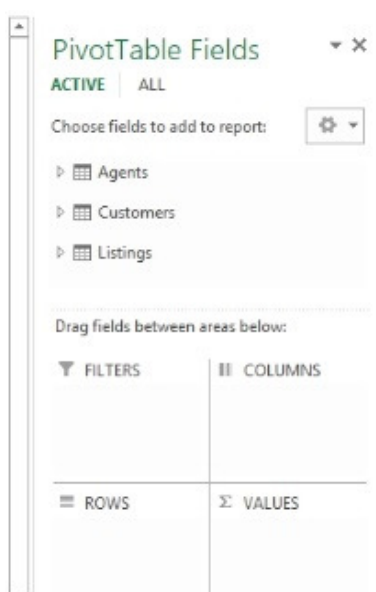


Figure 14 Pivot Table Task pane

As shown in Figure 14, all three tables in the Data Model like Agents, Customers, and Listings appear on the top. Arrow symbol next to each table can be clicked to display all the available fields of that table. Bottom

part of the task pane shows four zones: Filters, Legends, Axis, and Values. A field can be checked or dragged and dropped to a proper zone within the task pane. Fields can also be dragged and dropped within each zone of the task pane. The Columns zone is also sometimes displayed as Legend Fields, while the Rows zone appears as Axis Fields. Fields in the Values zone show summarized information using subtotals. A field in the Filters zone filters the display records.

- For illustration purposes we will show type of the houses that were listed in April 2017 with the Listing ID and the Listing Price. To accomplish this retrieval, we place Type in the Rows zone, Listing ID in the Columns zone, List Date in Filters zone, Sum of Price for subtotals in Values zone as shown in Figure 15.

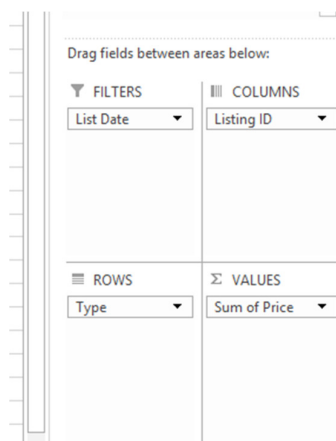


Figure 15 Tutorial Query Task Pane

Visualization with Pivot Charts:

- Go to Analyze tab and click Pivot Chart in the Tools group. Select clustered column as type of the chart. Pivot chart now illustrates the visualization of summarized information of the Pivot Table as shown in Figure 16. The chart will automatically update as Pivot Table is modified.

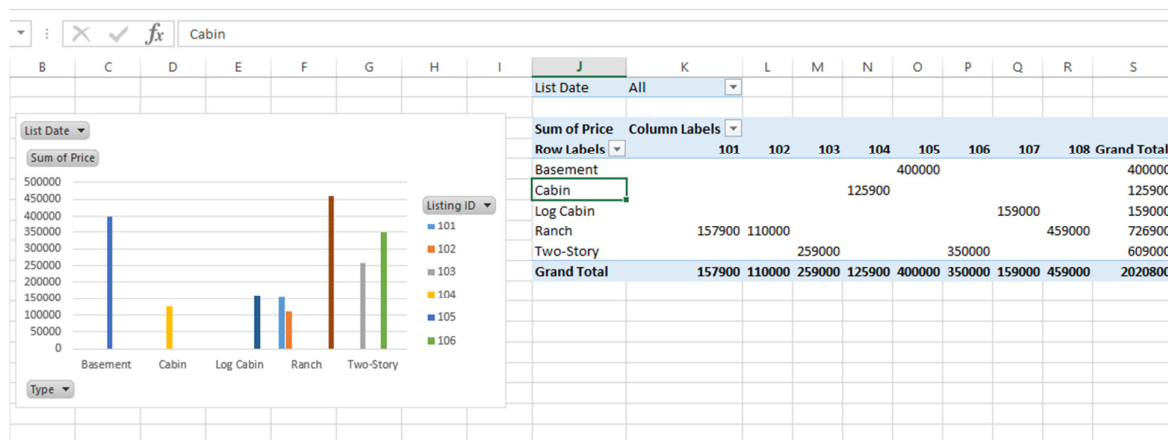


Figure 16 Tutorial Query Pivot Table and Pivot Chart Analytic display

To filter the records that were listed in April 2017 only, click the drop down arrow beside All of List Date display at the top of the Pivot Table. Once you click the All drop down arrow, check the box at the bottom to select multiple items. Now only check the dates of April month. Filter is now applied to both the table and the chart simultaneously. It will display only the listings of the month April as shown in Figure 17.

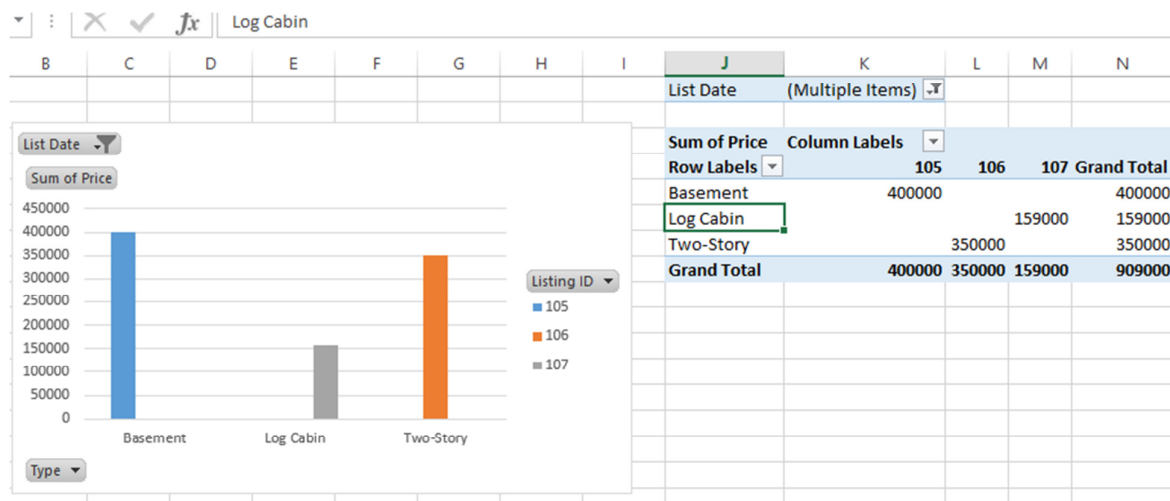


Figure 17 Tutorial Simulated Analytic Display

Both the Pivot chart and the Pivot table show that there were 3 houses that were listed in April 2017 with Listing Id 105, 106, and 107. It shows the individual price of each listing as well as the total price of all three listings which is \$909,000. Rows provide information about what type of houses are listed and at what price. For example, Basement home is listed at \$400,000 and has Listing Id 105.

If we want to add fields from another table that is directly related like Agents and Listings table, then we first select specific fields from the other table like Agents. The selected fields can be drag and dropped in appropriate zones as per the query just like a querying from a single table.

Retrieval from Multiple Tables with Joiner

Joiner concept facilitates querying from tables that don't have direct relationship with each other. For example, suppose we want to show Agent Id from Agents table and their listings with Listing Id, Type, Sum of Price from Listings table, and to filter records using House Type field from Customers table. This query involves not only getting data from Agents and Listings tables but also Customers table. Since Customers and Listings tables do not have a direct relationship with each other, the Joiner concept can provide the bridge between the two tables. The Joiner concept involves creating a new table with only one field that is common to the tables that have to be linked. This field in the Joiner type table is also its primary key. So, in the Tutorial we create a "Joiner_House_Type" table as a Joiner table. The Joiner_House_Type table will have relationship with Listings and Customers table, Once the

Joiner is created, one can query all the three tables to retrieve related information. The following steps show how to create a Joiner table and then retrieve information from all three tables.

1. Create a new table “Joiner_House_Type” on the same worksheet where Customers table was created similar to how tables were created earlier. The field name should be House Type and five rows should be populated with the information as shown in Figure 18.

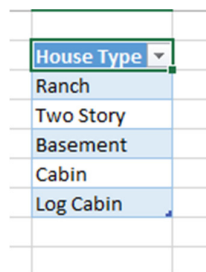


Figure 18 Joiner_House_Type Joiner Table

2. Add Joiner_House_Type table to the Data Model. Click Manage Data Model and select Pivot Table to create a new Pivot table. Go to the Data tab and in the Data Tools group select Relationships. Manage Relationships window will open with all the previous list of relationships. Now set up two relationships: Joiner_House_Type with Customers table and then Joiner_House_Type with Listings table. In the first relationship, the first table should be Customers and the foreign key should be House Type. Second table should be Joiner_House_Type and the primary key should be House Type. In the second relationship, the first table should be Listings table with Type as a foreign key and the second table should be Joiner_House_Type with House Type as a primary key. Now Joiner_House_Type is an intermediary between Customers and Listings table. This will allow us to see any listing with Listing Id, Agent Id, and the correct Price all at once. Once the Joiner is created in the Pivot Table fields that are part of Joiner table should be added from the Joiner table only.
3. We will now display Listings with Listing Id, Type of houses, and Sum of Price, Agents with Agent ID, and for Filter we will utilize House Type that customers are interested in for analyzing.
4. From the Agents table add Agent ID field to the Rows zone. From the Listings table, Listing Id to the Columns zone, Type to the Rows zone, and Price in the Value zone. In the Filter zone select House Type by adding “House Type” field from the Joiner_House_Type table as shown in Figure 19.

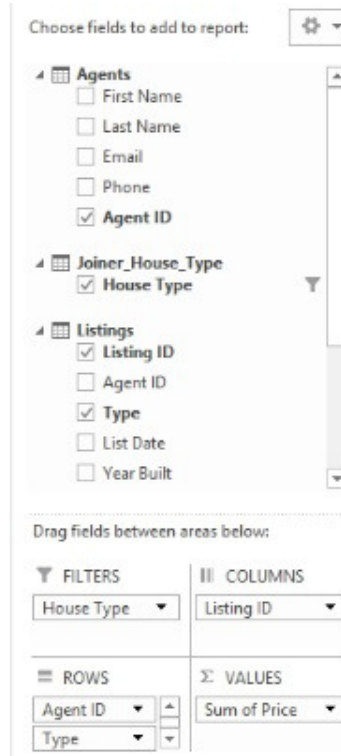


Figure 19 Tutorial Query Task Pane with Joiner Table

- From the Analyze tab of the Pivot Table, click Pivot Chart from the tools group and select Clustered column as a chart type. Figure 20 shows the Pivot Table and the Pivot Chart displaying all the listings with Listing ID, Agent ID, Price and the Type of the houses.



Figure 20 Tutorial Query Pivot Table and Pivot Chart Analytic display with Joiner Table

- If we want to display just the Ranch then from the House Type that is used as a filter, we can uncheck all the other types and display just the Ranch as shown in Figure 21.

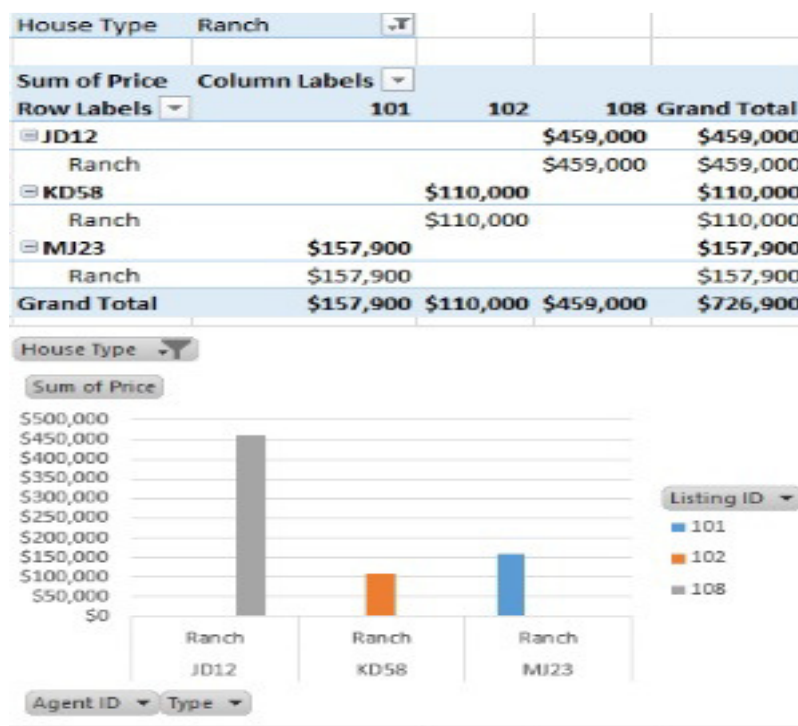


Figure 21 Tutorial Simulated Analytic Display with Joiner Table

Our results show three agents, JD12, KD58, and MJ23 who have listed Ranch with their respective Listing Id, their prices and the grand totals. This makes analyzing data from multiple tables very quick and easy.

Conclusions

Excel is an essential software for business. Many aspects of business can benefit from strong Excel knowledge. This paper outlines a tutorial based on a simple business example to demonstrate how to create a database in Excel and then query the database for analysis. The tutorial illustrates the use of Data Model, Pivot Table, and Pivot Chart features. It also outlines an approach to overcome the limitation of Excel in analyzing data from tables that don't have direct relationship through the Joiner concept.

The tutorial in the paper will allow students and professionals to grasp essential database operations with Excel in a business context. This simpler approach helps technical as well as non-technical students in understanding data modeling and database concepts in a faster way. Further, an understanding of how Power Pivot and Power Chart work with Joiner concept will enable any Excel professional to deliver data analytics and Business Intelligence solutions without prior expertise in data analytics.

References

- Chen, H., Roger, H. L. Chiang and Veda, C. Storey (2012) 'Business Intelligence And Analytics: From Big Data to Big Impact,' *MIS Quarterly*, Vol. 36, No. 4, pp. 1165-1188.
- Cram, C. and Friedrichsen, L. (2013) *Portfolio Projects for Business Analysis*, Course Technology, Cengage Learning.
- Ferrari, A. and Russo, M. (2010) *Microsoft PowerPivot for Excel 2010: Give Your Data Meaning*, Microsoft Press.
- Frye, Curtis (2017) *Microsoft Excel 2016 step by step*, Microsoft Press.
- Gaskin, S. and Vargas, A. (2017) *GO! with Microsoft Excel 2016 Comprehensive*, Pearson Publishing.
- Jelen, B. and Alexander, M. (2013) *Excel 2013 Pivot Table Data Crunching*, Que Publishing.
- Jelen, B. (2017) *POWER EXCEL with MrExcel*, Holy Macro! Books.
- Kinser, A.S., Moriarity, B.P., Kinser, E. and Kosharek, D. (2017) *Microsoft Access 2016*, Pearson Publishing.
- Meyer, Z. D. and Leanne, A.M. (2008) 'Excel as a Qualitative Data Analysis Tool,' *Field Methods*, pp. 91-112.
- Monk, E.F., Brady, J.A. and Mendelsohn, E.I. (2017) *Microsoft Access & Excel 15 Edition*, Cengage Publishing.
- Palocsay, S.W., Markham, I. S. and Markham, S. E. (2010) 'Utilizing and teaching data tools in Excel for exploratory analysis,' *Journal of Business Research*, Vol. 63, No. 2, pp. 191-206.
- Parsons, J.J., Oja, D., Ageloff, R., and Carey, P. (2010) *New Perspectives Microsoft Excel 2010*, Cengage Learning.
- Ralston, B. (2011) *PowerPivot for Business Intelligence using Excel and SharePoint*, Springer.
- Reding, E.E. and Wermers, L. (2016) *Microsoft Office 365 Excel 2016 Comprehensive*, Cengage Learning.
- Rosenberg, E. (2015) *Investopedia*, 24 March 2015. Digital.
- Smart, M. (2017) *Excel 2016 Expert Skills with Smart Method*, Smart Method Ltd.
- Zhimin, C., Narasayya, V. and Chaudhuri, S. (August 2014) 'Fast foreign-key detection in Microsoft SQL server PowerPivot for Excel,' *Proceedings of the VLDB Endowment*, pp. 1417-1428.

Reproduced with permission of copyright owner. Further reproduction prohibited without permission.