Data Visualization: A Practical Approach for Absolute Beginners

Lab 2 – Visual Analysis

Overview

Beyond just storing and simple visualizations of data, there are various tools and techniques for doing analysis of your data. In this lab you'll do some basic analysis of the expense data you started in Lab 1, and then you'll look at a bigger dataset from a lemonade stand to do some basic analysis for that business.

What You'll Need

To complete the labs, you will need the following:

- A Windows, Linux, or Mac OS X computer with a web browser.
- A Microsoft account (for example a *hotmail.com*, *live.com*. or *outlook.com* account). If you do not already have a Microsoft account, sign up for one at https://signup.live.com.
- Microsoft Excel Online or Microsoft Excel for your desktop.
  - o You can get Microsoft Office Online with your Microsoft account here:
    https://products.office.com/en-us/office-online/documents-spreadsheets-presentationsoffice-online
  - o There are instructions in Module 0 for getting the trial version of Microsoft Office.

Exercise 1: Work with Data in Excel

In this exercise you will do some work on data in Excel to set it up for better analysis. If you have your worksheet from Lab 1, you can continue with that data.

Prepare your worksheet

1. You can start with your Lab 1 data, or you can copy and paste this table of data into a new worksheet in Microsoft Excel:

Place	Item	Price	Date
Grocery Store	Sandwich	$5.00	2/19/2017

Movie Theater	Ticket	$7.50	2/21/2017
Gas Station	Gasoline	$22.00	2/21/2017
Grocery Store	Candy	$6.00	2/22/2017
Gas Station	Car wash	$9.00	2/23/2017
Grocery Store	Fruit	$4.00	2/25/2017
Pizza Parlor	Pizza	$13.00	2/28/2017
Museum	Tickets	$5.00	3/1/2017
Grocery Store	Sandwich	$5.00	3/3/2017
Gas Station	Car wash	$12.00	3/3/2017
Grocery Store	Fruit	$6.00	3/4/2017
Pizza Parlor	Pizza	$11.00	3/5/2017
Grocery Store	Candy	$6.00	3/6/2017

Set up your data for better analysis

1. Select cell **A1**, and then on the **Insert** tab of the ribbon above the worksheet, click **Table**. Verify that Excel has automatically detected the data in the range **A1:D14**, and that the **My table has headers** checkbox is selected, and then click **OK**.

   Excel automatically formats the data as a table and adds drop-down buttons to the header row as shown here:

2. Click any cell to deselect the table, and then click the drop-down button for the **Place** column, and click **Filter…**
3. In the **Filter** dialog box, clear the **(Select All)** checkbox, and then select only the **Gas Station** checkbox before clicking **OK**.

   The table of data is filtered to show only the rows/records for Gas Station expenses.

4. Click the drop-down arrow for the **Price** column and click **Sort Descending**. The table of data is sorted in descending order of price, so the first row contains the data for the Gas Station expense with the highest price.

	A	B	C	D
1	Place	Item	Price	Date
4	Gas Station	Gasoline	$ 22.00	
6	Gas Station	Car wash	$ 9.00	2/23/2017
11	Gas Station	Car wash	$ 12.00	3/3/2017

5. Click the drop-down arrow for the **Place** column again and then click **Clear Filter from 'Place'**. The table now shows all the data.
6. Click the drop-down arrow for **Date** and click **Sort Ascending** to re-order the data into chronological order.

Exercise 2: Using Formulae to Explore Data in Excel

In this exercise, you will use formulae to create a derived column that extends the data.

Add Derived Columns

1. Right-Click the **Date** column header to bring up the context menu, and select Insert Columns. This inserts a new **Column1** column between the **Price** and **Date** columns as shown here:

	A	B	C	D	E	F
1	Place	Item	Price	Column	Date	
2	Grocery Store	Sandwich	$ 5.00		2/19/2017	
3	Movie Theater	Ticket	$ 7.50		✂ Cut	
4	Gas Station	Gasoline	$ 22.00		📋 Copy	
5	Grocery Store	Candy	$ 6.00		📋 Paste	
6	Gas Station	Car wash	$ 9.00			
7	Grocery Store	Fruit	$ 4.00		🔲 Insert Columns	
8	Pizza Parlor	Pizza	$ 13.00		✗ Delete Columns	
9	Museum	Tickets	$ 5.00		Clear Contents	
10	Grocery Store	Sandwich	$ 5.00			
11	Gas Station	Car wash	$ 12.00		🔲 Column Width...	
12	Grocery Store	Fruit	$ 6.00		Hide Columns	
13	Pizza Parlor	Pizza	$ 11.00		Unhide Columns	
14	Grocery Store	Candy	$ 6.00		ABC 123 Number Format	
15						

2. In cell **D1**, rename **Column1** to **Month**. Then in cell **D2**, enter the following formula:

```
=TEXT(E2, "mmmm")
```

After you enter the formula, it should be copied automatically to all the other **Month** cells in the table, and the name of the month for each record should be displayed.

3. Filter the data on the Month column so you can see just the data in one month at a time. Here are the sample expenses from March:

	A	B	C	D	E
1	Place ▼	Item ▼	Price ▼	Month ▼	Date ▼
9	Museum	Tickets	$5.00	March	3/1/2017
10	Grocery Store	Sandwich	$5.00	March	3/3/2017
11	Gas Station	Car wash	$12.00	March	3/3/2017
12	Grocery Store	Fruit	$6.00	March	3/4/2017
13	Pizza Parlor	Pizza	$11.00	March	3/5/2017
14	Grocery Store	Candy	$6.00	March	3/6/2017

Exercise 3: Using Conditional Formatting to Explore Data

In this exercise, you will apply conditional formatting to data to highlight key values of interest.

Highlighting Extremes and Outliers

1. Clear any filters in your Table so you can see all the data.
2. Select cell **C2,** the first value in the Price column, and then hold the **Shift** and **Ctrl** keys and press the **Down-Arrow** key to select all the values in the **Price** column (if you are using a Mac OSX computer, hold the **Shift** and ⌘ keys, and press the **Down-Arrow** key).
3. On the **Home** tab of the ribbon, in the **Conditional Formatting** drop-down list, point to **Color Scales**, and select the **Red-White Color Scale** (with red at the top and white at the bottom). The **Price** cells are reformatted so that the highest expenses are colored an intense red, and the lowest prices are much lighter in color intensity. Scrolling through the data now, it is easier to find expenses that are particularly high or low.

|	A	B	C	D	E
1	Place ▼	Item ▼	Price ▼	Month ▼	Date ▼
2	Grocery Store	Sandwich	$5.00	February	2/19/2017
3	Movie Theater	Ticket	$7.50	February	2/21/2017
4	Gas Station	Gasoline	$22.00	February	2/21/2017
5	Grocery Store	Candy	$6.00	February	2/22/2017
6	Gas Station	Car wash	$9.00	February	2/23/2017
7	Grocery Store	Fruit	$4.00	February	2/25/2017
8	Pizza Parlor	Pizza	$13.00	February	2/28/2017
9	Museum	Tickets	$5.00	March	3/1/2017
10	Grocery Store	Sandwich	$5.00	March	3/3/2017
11	Gas Station	Car wash	$12.00	March	3/3/2017
12	Grocery Store	Fruit	$6.00	March	3/4/2017
13	Pizza Parlor	Pizza	$11.00	March	3/5/2017
14	Grocery Store	Candy	$6.00	March	3/6/2017

Exercise 4: Pivot Your Data

In this exercise, you will create a PivotTable to do deeper analysis of your expense data.

Create a PivotTable

1. Clear any filters in your Table so you can see all the data.

2. In the Home tab, click PivotTable. In the Create PivotTable screen, select your table as the source data for your pivot table. Your table might be called Table1.
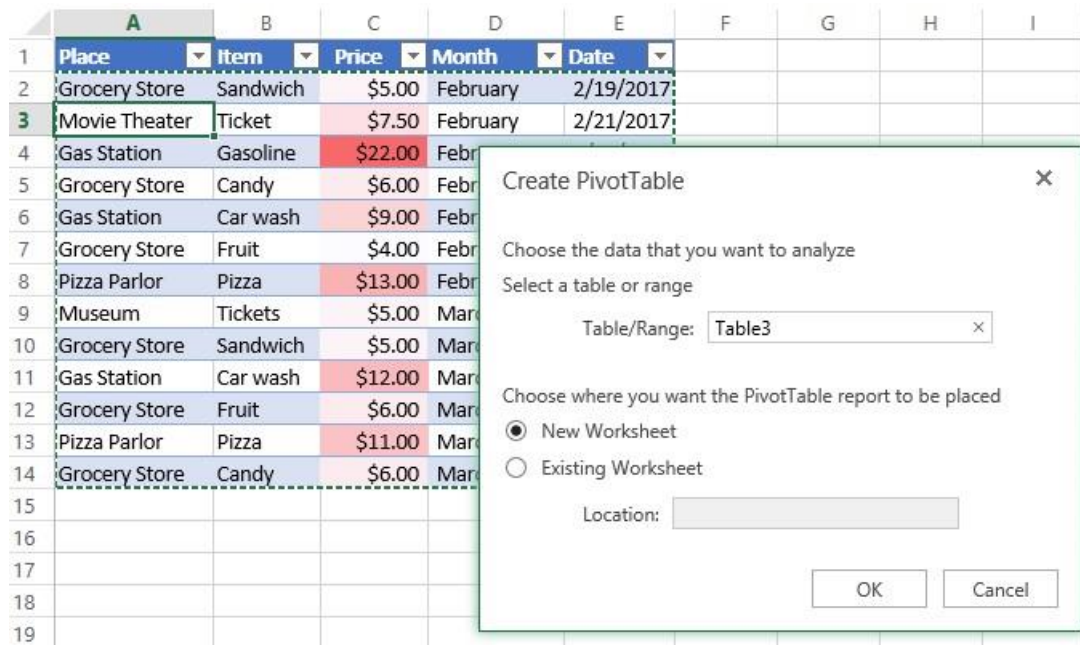
A new worksheet will be created with your PivotTable, and Excel will set focus on that new worksheet.
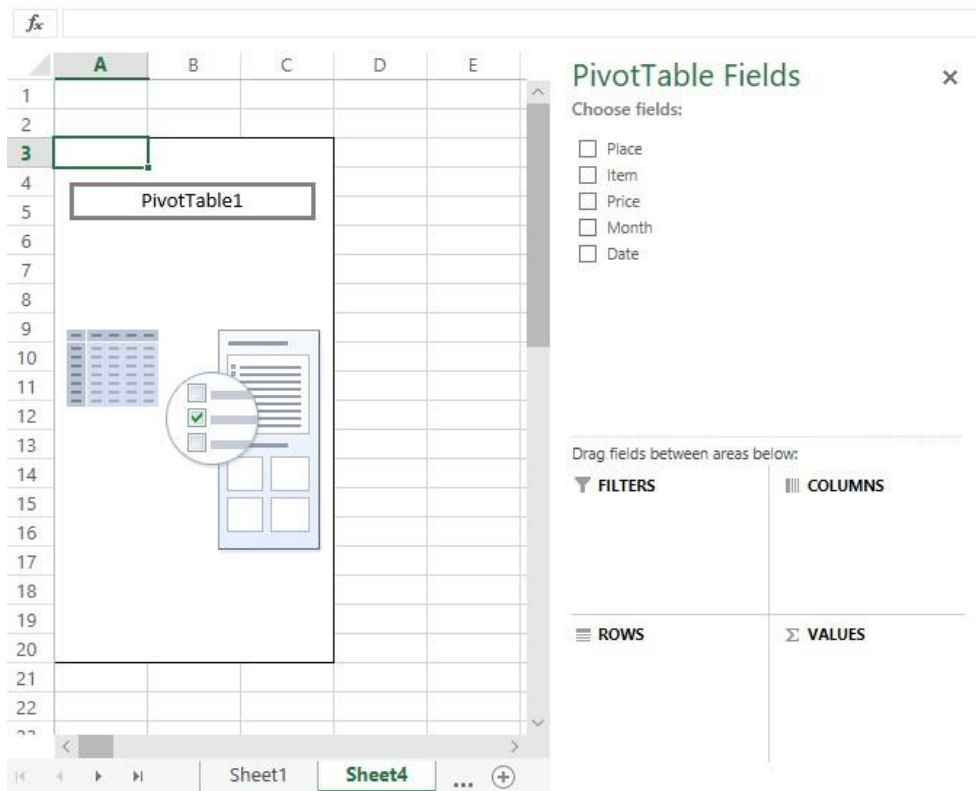
3. Explore the empty PivotTable.

There is a region in the worksheet where the empty PivotTable shows no data.

There is a PivotTable Fields section on the right side of the window that is displaying the column names from your table.

There are four boxes below the PivotTable Fields called Filters, Columns, Rows, and Values.

Display Data in the PivotTable

1. To display data and begin doing analysis you need to select Fields and place them in one of the Field Areas of the PivotTable.

   a. Click the checkbox next to **Price**. This will automatically add the Sum of all of the Prices to the Values area of the PivotTable.

   b. Click the checkbox next to **Place**. This will add the Place items in consolidated groups to the Rows area of the PivotTable.

   Notice how the PivotTable is now showing you the Sum of the prices for every row of a place. Your PivotTable will resemble the following:

Row Labels	Sum of Price
Gas Station	43
Grocery Store	32
Movie Theater	7.5
Museum	5
Pizza Parlor	24
Grand Total	**111.5**

   c. Click the checkbox next to **Item**. This will add the Item categories below the Place so you can see the Sum of prices for each Place.

Row Labels	Sum of Price
⊟ **Gas Station**	**43**
Car wash	21
Gasoline	22
⊟ **Grocery Store**	**32**
Candy	12
Fruit	10
Sandwich	10
⊟ **Movie Theater**	**7.5**
Ticket	7.5
⊟ **Museum**	**5**
Tickets	5
⊟ **Pizza Parlor**	**24**
Pizza	24
Grand Total	**111.5**