Chapter 5

# Applying the Concepts: Excel Instructions

## Select Variables

1. Open Suppliers.csv in Excel.
2. What country is recorded for the tenth observation in the **Suppliers** table?
   1. Find the column header for “Country” in Column G.
   2. Find the observation in Cell G11 for the 10th observation since the first row contains the headers.
3. What country is recorded as the first observation in the **Suppliers** table?
   1. Find the observation in Cell G2 for the 1st observation since the first row contains the headers.
4. What country is recorded as the last observation in the **Suppliers** table?
   1. Find the observation in Cell G30 for the last observation.

## Select Distinct Values of Variables

1. Open Suppliers.csv in Excel.
2. Save a copy called Suppliers-distinct countries.csv.
3. How many distinct countries are in the **Suppliers** table?
   1. Select the column whose header is “Country” in Column G.
   2. From the Excel ribbon at the top, go to ‘Data’ → ‘Remove Duplicates’.
   3. Keep the default option ‘Expand the selection’ and click ‘Remove Duplicates…’
   4. Click the ‘Unselect All’ button.
   5. Click the “Country” check box.
   6. Click ‘OK’.
   7. You will get an information box that 13 duplicate values were found. Click OK.
   8. Select column G again. At the bottom of the screen, you will see “Count: 17”.
   9. Since the first row has the “Countries” header, subtract 17 rows - 1 header row = 16 distinct countries.
4. Which of the following are countries that Zahava’s suppliers are in?
   1. Select all of Column G.
   2. Press Ctrl + F on your keyboard (or Cmd + F on a Mac). This will open the 'Find and Replace' dialog box.
   3. Enter each country name as listed in the dialog box's 'Find what' field.

## Filter Text Variables

1. Open Suppliers.csv in Excel.
2. How many suppliers are in France?
   1. Verify the first row of Column G is "Country".
   2. Highlight all of Column G.
   3. From the Excel ribbon at the top, go to ‘Data’ → ‘Filter’. This action will enable a dropdown arrow next to the header in Column G.
   4. Click the dropdown arrow in Column G's "Country" column header. A list of all unique values in the column will appear.
   5. Uncheck the box next to 'Select All' option to clear all selections.
   6. Scroll through the list and find "France". Check the box next to "France".
   7. Click 'OK' or press 'Enter'.
   8. After applying the filter, only the rows where the "Country" column equals "France" will be displayed.
   9. Select Column G again. At the bottom of the screen, you will see “Count: 4.”
   10. Since the first row has the “Countries” header, subtract 4 rows - 1 header row = 3 countries.
3. How many suppliers are in the USA?
   1. After completing the steps in #2 to turn on the filter in Column G, you can again click the dropdown arrow in the "Country" column header in Column G.
   2. A list of all unique values in the column will appear.
   3. Uncheck the 'Select All' option to clear all selections.
   4. Scroll through the list and find "USA". Check the box next to "USA".
   5. Click 'OK' or press 'Enter'.
   6. After applying the filter, only the rows where the "Country" column equals "USA" will be displayed.
   7. Select column G again. At the bottom of the screen, you will see “Count: 5.”
   8. Since the first row has the “Countries” header, subtract 5 rows - 1 header row = 4 countries.

## Filter with the Not-Equal Operator

1. Open Suppliers.csv in Excel.
2. How many suppliers are outside the USA?
   1. Verify the first row of Column G is "Country".
   2. Select all of Column G.
   3. From the Excel ribbon, go to ‘Data’ → ‘Filter’. This action will enable a dropdown arrow in the header of Column G.
   4. Click the dropdown arrow in Column G's "Country" column header.
   5. A list of all unique values in the column will appear.
   6. Be sure the 'Select All' option is checked to keep all selections.
   7. Scroll through the list and find "USA". Uncheck the box next to "USA".
   8. Click 'OK' or press 'Enter'.
   9. After applying the filter, only the rows where the "Country" column is not equal to "USA" will be displayed.
   10. Select column G again. At the bottom of the screen, you will see “Count: 26.”
   11. Since the first row has the “Countries” header, subtract 26 rows - 1 header row = 25 countries.
3. How many suppliers are outside France?
   1. Verify the first row of Column G is "Country."
   2. Click the first cell of Column G to select the entire column.
   3. Go to the 'Data' tab on the Excel ribbon at the top of the window.
   4. Click on the 'Filter' button. This action will enable dropdown arrows on each column header
   5. Click the dropdown arrow in the "Country" column header in Column G. A list of all unique values in the column will appear.
   6. Be sure the 'Select All' option is checked to keep all selections.
   7. Scroll through the list and find "France". Uncheck the box next to "France".
   8. Click 'OK' or press 'Enter'.
   9. After applying the filter, only the rows where the "Country" column is not equal to "France" will be displayed.
   10. Select Column G again. At the bottom of the screen, you will see “Count: 27.”
   11. Since the first row has the “Countries” header, subtract 27 rows - 1 header row = 26 countries.

## Filter Using a Wildcard

1. Open Products.csv in Excel.
2. How many observations start with “C”?
   1. Verify the first row of Column B is "ProductName."
   2. Select all of Column B.
   3. From the Excel ribbon at the top, go to 'Data' → ‘Filter’. This will enable a dropdown arrow on the header of ColumnB.
   4. Click the dropdown arrow in Column B's "ProductName" column header.
   5. Look for ‘Text Filters’ → ‘Begins With…’
   6. Type “C” into the text box.
   7. Click ‘OK’.
   8. Select column B again. At the bottom of the screen, you will see “Count: 10.”
   9. Since the first row has the “ProductName” header, subtract 10 rows - 1 header row = 9 product names.
3. How many product names end with “e”?
   1. First, clear your filter from #2. Select the dropdown in Column B and click ‘Clear Filter from “ProductName”’.
   2. Click the dropdown arrow again, and click ‘Text Filters’ → ‘Ends With…’
   3. Type “e” into the text box.
   4. Click ‘OK’.
   5. Select column B again. At the bottom of the screen, you will see “Count: 18”.
   6. Since the first row has the “ProductName” header, subtract 18 rows - 1 header row = 17 product names.
4. How many product units have “48” at the beginning of the description?
   1. First, clear your filter from #3. Select the dropdown in Column B and click ‘Clear Filter from “ProductName”’.
   2. Verify the first row of Column E is "Unit".
   3. Select all of Column E.
   4. From the Excel ribbon, go to ‘Data' → ‘Filter’. This will enable a dropdown arrow in the header of Column E.
   5. Click the 'Filter' button, enabling dropdown arrows on each column header.
   6. Ensure no filters are set for other columns.
   7. Click the dropdown arrow in Column E's "Unit" column header.
   8. Select ‘Text Filters’ → ‘‘Begins With…’
   9. Type “48” into the text box.
   10. Click OK.
   11. Select column E again. At the bottom of the screen, you will see “Count: 4.”
   12. Since the first row has the “Unit” header, subtract 4 rows - 1 header row = 3 product names.

## Filter a Quantitative Variable

1. Open OrderDetails.csv in Excel.
2. How many orders have a quantity that is 45 or more?
   1. Verify that the header of Column D reads "Quantity".
   2. Select all of Column D.
   3. From the Excel ribbon at the top, go to ‘Data’ → ‘Filter’. This will enable a dropdown arrow on the header of Column D.
   4. Ensure no filters are set for other columns.
   5. Click the dropdown arrow in Column D's "Quantity" column header.
   6. Look for ‘Number Filters’ → ‘Greater Than Or Equal To…’
   7. Type 45 into the text box.
   8. Select Column D again. At the bottom of the screen, you will see “Count: 73.”
   9. Since the first row has the “Quantity” header, subtract 73 rows - 1 header row = 72 orders.
3. How many orders have a quantity that is more than 10?
   1. Click the dropdown arrow in Column D's "Quantity" column header.
   2. Look for ‘Number Filters’ → ‘Greater Than…’
   3. Type 10 into the text box.
   4. Select column D again. At the bottom of the screen, you will see “Count: 387”.
   5. Since the first row has the “Quantity” header, subtract 387 rows - 1 header row = 386 orders.

## Filter Missing Values

1. Open OrderDetails.csv in Excel.
2. How many orders have no order identification number (OrderID)?
   1. Verify the first row of Column B is labeled "OrderID".
   2. Highlight all of Column B.
   3. From the Excel ribbon at the top, go to ‘Data’ → ‘Filter’. This will enable a dropdown arrow on the header of Column B.
   4. Click the dropdown arrow in Column B's "OrderID" column header.
   5. Uncheck the 'Select All' option to clear all selections.
   6. Scroll to the bottom of the list to find “(Blanks)”. Check the box next to "(Blanks)". Since "(Blanks)" is not in the menu, it means there are no missing data.

## Filter Multiple Conditions Simultaneously

1. Open OrderDetails.csv in Excel.
2. How many observations have an Order ID of 10293 and a quantity of at least ten?
   1. Verify the first row of Column B is labeled “OrderID” and Column D is labeled "Quantity".
   2. Select Columns B, C, and D simultaneously.
   3. From the Excel ribbon at the top, go to ‘Data’ → ‘Filter’. This will enable dropdown arrows on each of the column headers for Columns B-D.
   4. Click the dropdown arrow in Column B's "OrderID" column header.
   5. Click ‘Number Filters’ then ‘Equals…’
   6. In the text box, type “10293” (omit the quotation marks).
   7. Click 'OK' or press 'Enter'.
   8. Click the dropdown arrow in Column D's "Quantity" column header.
   9. Click 'Number Filters' then ‘Greater Than Or Equal To…’
   10. In the text box, type “10” (omit the quotation marks)
   11. Click 'OK' or press 'Enter.'
   12. After applying the filters, you will see two observations.
3. How many observations have an Order ID of 10294 and a quantity greater than fifteen?
   1. Click the dropdown arrow in Column B's "OrderID" column header.
   2. Click 'Number Filters' then ‘Equals…’
   3. In the text box, type “10294” (omit the quotation marks)
   4. Click 'OK' or press 'Enter'.
   5. Click the dropdown arrow in Column D's "Quantity" column header.
   6. Click 'Number Filters' then ‘Greater Than…’
   7. In the text box, type “15” (omit the quotation marks)
   8. Click 'OK' or press 'Enter'.
   9. After applying the filters, you will see two observations.

## 

## Order by a Variable

1. Open Suppliers.csv in Excel.
2. Who is the contact person for the supplier that appears second when the list of suppliers is arranged alphabetically by supplier name?
   1. Verify the first row of Column B is "SupplierName."
   2. Select all of Column B.
   3. From the Excel ribbon at the top, go to 'Data' → ‘Sort A to Z’.
   4. Choose “Expand the selection” in the pop-up window.
   5. Click the ‘Sort’ button to close the pop-up window.
   6. The list is now sorted alphabetically, from A to Z; you may look for the contact name of the second observation (Cell C3).
3. Who is the contact person for the supplier that appears second when the list of suppliers is arranged in descending order by supplier name?
   1. Verify the first row of Column B is "SupplierName."
   2. Select all of Column B.
   3. From the Excel ribbon at the top, go to 'Data' → ‘Sort Z to A’.
   4. Choose “Expand the selection” in the pop-up window.
   5. Click the ‘Sort’ button to close the pop-up window.
   6. The list is now sorted descending; you may look for the contact name of the second observation (Cell C3).

## Summarize a Quantitative Variable

1. Open OrderDetails.csv in Excel.
2. What is the average quantity customers ordered (one decimal place)?
   1. Verify the first row of Column D is labeled "Quantity".
   2. In any empty cell (e.g., cell F1), type:

=AVERAGE(D:D)

* 1. What is the maximum quantity customers ordered (no decimal places)?
  2. In any empty cell (e.g., cell F2), type:

=MAX(D:D)

* 1. What is the minimum quantity customers ordered (no decimal places)?
  2. In any empty cell (e.g., cell F3), type:

=MIN(D:D)

## Summarize a Quantitative Variable by a Categorical Variable

1. Open OrderDetails.csv in Excel.
2. What is the mean quantity for Product ID 28?
   1. Verify that the first row of Column B is labeled “OrderID”, the first row of Column C is labeled “ProductID”, and the first row of Column D is labeled "Quantity".
   2. Select Columns B through D simultaneously.
   3. From the Excel ribbon at the top, select ‘Insert’ → ‘Pivot Table’.
   4. Choose ‘From Table/Range’.
   5. Accept the Table/Range of “OrderDetails!$B:$D” and the default of “New Worksheet”; do not change any other options.
   6. Select ‘OK’.
   7. You will see a new Pivot Table sheet.
   8. Drag “ProductID” from the right-hand menu under ‘PivotTable Fields’ to the box below called “Rows”. This indicates the variable by which you want to group your data.
   9. Drag “Quantity” from the right-hand menu under ‘PivotTable Fields’ to the box below called “Values.” This indicates the variable you wish to take an average of.
   10. In the same “Values” box on the lower-right side, choose from the ‘Sum of Quantity’ drop-down the ‘Value Field Settings…’ option.
   11. Choose ‘Average’.
   12. Click ‘OK’.
   13. Navigate to Product ID 28 to find the mean quantity.
3. What is the mean quantity for Product ID 44?
   1. Find the answer in the results of #2. Navigate to Product ID 44.
4. What is the mean quantity for Order ID 10276?
   1. Ensure you have completed Steps 2(a)-(f). If you’ve finished #2, drag “ProductID” and “Average of Quantity” back to the main window under ‘PivotTable Fields’.
   2. Repeat the steps in #2 except replace step (g) with the following:

Drag OrderID from the right-hand menu under “PivotTable Fields” to the box below called “Rows.” This indicates the variable by which you want to group your data.

## Create a New Variable

1. Open Products.csv in Excel.
2. What is the Price\_in\_Turkish\_Lira for the last observation in the data set?
   1. Verify the first row of Column F is labeled "Price."
   2. In the header of Column G, Cell G1, type “Price\_in\_Turkish\_Lira”.
   3. In cell G2, type:

=F2 \* 30

* 1. Select Cell G2.
  2. Move your cursor to the lower-right corner of cell G2 until it changes to a small black cross (the fill handle).
  3. Click and drag the fill handle down along Column G to the last row where you want the formula to be applied (in this case, row 78).
  4. Check the cells to ensure the formula has been correctly copied and automatically adjusted for each row.
  5. Check the Price\_in\_Turkish\_Lira value in the last row, 78.

## Merge Two Tables

Note: for this exercise, you will need to use .xlsx files, not .csv files. You can download these with the rest of your data sets.

1. Open Products.xlsx in Excel.
2. Also open Categories.xlsx in Excel.
3. What is the category name of the product called “Chartreuse verte”?
   1. Start in Products.csv.
   2. Verify the first row of Column B is labeled "ProductName."
   3. In Cell H1, type “Category\_Name”.
   4. In Cell H2, type:

=VLOOKUP(D2,[Categories.xlsx]Sheet1!$A:$C,2,FALSE)  
Note: Excel can be finicky with copying/pasting. So, we give the logic below to type rather than paste the “=VLookup” command. You can complete it by typing and clicking on the correct cells. VLookup has four parameters, each separated by commas:

* + 1. The first parameter identifies Cell D2 as CategoryID from the **Products** table.
    2. The second parameter (e.g., “[Categories.xlsx]Sheet1!$A:$C”) identifies Column A in the open Categories.csv file as the place to find the CategoryID in the **Categories** table while looking for associated information like CategoryName in Columns A through C.
    3. The third parameter (e.g., “2”) returns the CategoryName as the 2nd column of the range from step (ii) above.
    4. The fourth parameter (e.g., “FALSE”) tells the VLOOKUP function to look for an exact match of the lookup value.
    5. If an exact match is found, VLOOKUP returns the value.
    6. If no exact match is found, VLOOKUP will return an error (e.g., #N/A).
  1. Click on cell H2 to select it.
  2. Move your cursor to the lower-right corner of cell H2 until it changes to a small black cross (the fill handle).
  3. Press the fill handle's left mouse button (or the primary mouse button on a Mac).
  4. Drag the fill handle down along Column H to the last row where you want the formula to be applied (in this case, row 78)
  5. Release the mouse button to copy the formula from H2 into all selected cells in Column H.
  6. Check the cells to ensure the formula has been correctly copied and automatically adjusted for each row.
  7. Find the category name by searching for 'Chartreuse verte' in the ProductName column (B) using the Find function, typically accessed through the keyboard shortcut for 'Find' (often Ctrl-F on Windows or Cmd-F on Mac). Then look up the category name in Column H.

1. Answer the other questions using the output of #3 to find the categories of “Valkoinen suklaa” and “Røgede sild”.