NOTE: Files in these exercises should be stored on a local folder, otherwise the completed versions of the queries will have an error.

|  |
| --- |
| Topic steps for Power Query and Power Pivot introduction |
| 1. STEPS: Simple example import data from CSV    1. Open *'01 PQ Simple Import.xlsx'*    2. Go to Data > From File > CSV    3. Choose "*PQ\_Sales.csv*" (in 'Data' folder where you saved the session files)    4. Click Transform    5. Query steps:       1. Click on the X to remove the Changed Type step       2. Filter Column1 to remove blanks, and text values other than 'Order Num'       3. Press Home > Use first row as headers       4. Check for Changed Type step. If it's missing, then select all columns using Shift+click, and press Transform > Detect Data Type       5. Review column header summary to verify all columns have 0% errors.    6. Click Close & Load > Close & Load to…    7. Choose Table, Existing Worksheet, specify cell D5.    8. Click OK. 2. STEPS: data cleanup techniques – IMDB data    1. Open *'02 PQ Data Cleanup.xlsx*'    2. Click on IMBD\_data sheet.    3. Select any cell in the IMDB data and press Data > From Table/Range    4. Review the data and steps that appear in PQ Editor.    5. Remove the "Changed Type" step, if needed.    6. Separate Rank from Title – option 1 - column from example       1. Select Column1, and go to Add Column > Column from Example.       2. In the new column (Column8), type '1' in the row for the first movie. See the suggestion for other rows. Press OK to accept the suggestion and create a column with just the movie ranks.       3. Press Column from Example again. This time, type the name of the movie on the first row and see if suggestions for other rows are shown. If not, type the name of the movie on subsequent rows until a suggestion appears and looks good. Press OK.       4. Double-click the column title for the new columns and rename them to 'Rank' and 'Title'.    7. Separate Rank from Title – option 2 - Split Column       1. Select Column1 and go to Home > Split Column > By Delimiter       2. Choose Custom and type '. ' (dot space)       3. Split at left most delimiter       4. Press OK    8. Delete the Changed Type step    9. Split the Rating from number of Votes       1. Select the column that has the score and votes, like "9.3 (3M)".       2. Choose Home > Split Column > By Delimiter       3. Choose Custom and type ' (' (space open parenthesis)    10. Delete the Changed Type step.    11. There are 2 columns with rank and 2 with title – remove the ones from the right-most columns by choosing Home > Select Columns. Uncheck the columns that aren't needed ("Column5" and last 2 columns).    12. Rename the columns as appropriate        1. Click on column header and press F2 or double-click        2. Type appropriate name ("Rank", "Title", "Score", "Votes", "Year", "RunTime", "MPAA")        3. Press Enter        4. Repeat for all columns    13. Replace ")" in the Votes column        1. Select the column and choose Home > Replace values.        2. Type ) and replace with nothing    14. Split Votes column into number and 'M' or 'K'        1. Select column and choose Split Column > By Number of Characters        2. Number of characters = 1        3. Choose Once, as far right as possible        4. Press OK    15. Delete the Changed Type step    16. Multiply votes by magnitude        1. Change the Votes column to Decimal number.        2. Go to Add Column > Conditional column.        3. Type "VoteSort" as column name        4. If Votes.2 equals M then Select a column > Votes.1        5. Else Select a column > Votes.1        6. Press OK and then look in the formula bar. Type "/1000" in the Else part, like "…each if [Votes.2] = "M" then [Votes.1] else [Votes.1]/1000)".  This will divide the number of votes for rows with K (thousands), so all rows will be in millions.    17. Re-order columns        1. Select VoteSort and drag it next to Votes.2.        2. Type in formula to move "Year" next to "Title" |
|  |
| 1. STEPS – Data quality, merge queries – BoxOfficeData    1. Go to BoxOfficeData tab and press Data > From Table/Range.    2. Remove Changed Type step.    3. Press Home > Use First Row as Headers    4. Review the Column Quality indicators (valid, error, empty) at the top of columns. If you don't see it, go to View > Column quality.    5. There may be errors in the Rank column. Hover your mouse over it, then click the … in the flyout and choose Keep Errors.    6. Click on one of the error values to see the error detail at the bottom of the window.    7. Notice that "Rank" is a value that can't be converted to number. Remove the Keep Errors step.    8. Go to the Promoted Headers step and filter out "Rank" from the Rank column.    9. Check Column quality again to see that errors have been removed.    10. Merge queries – (to see movies on both lists)        1. Click Home > Merge Queries        2. Choose the IMDBTop250 query in the drop-down.        3. Select the Title column in both tables.        4. Choose Join Kind = Inner        5. Click OK.        6. Click on the Expand button at the top of the IMDBTop250 column.        7. Choose which columns to expand – Rank and Score, keep the checkbox to Use original column name as prefix        8. Click OK        9. Select the IMDB Rank column and move it next to the Rank column so you can see both together.        10. Click the data type button in the column header of IMDB Rank and choose Whole Number.        11. Review the Column Profile (distinct and unique count) for Rank columns. More distinct (number of different values) than unique (only appearing once) means that at least one value is duplicated in the list.        12. It's because Lion King has 2 release years in the IMDB list.            1. Go back to the Merge Queries step, and click on the gear icon.            2. Hold Ctrl key and click on the Year column of both tables.            3. Click OK.            4. Review the Column Profile again.        13. Click the Sort button in the Rank column and Sort Ascending.        14. Click Close & Load > Load To…        15. Choose Table, Existing worksheet, navigate to 'PQ3'!$D$5 |
|  |
| 1. STEPS – Reconcile 2 lists    1. Open "*03 PQ reconciliation.xlsx*"    2. Go to the Data tab.    3. Select a cell in the bank transactions table.    4. Go to Data > From Table/Range    5. Press Close & Load > Load To…, and choose Only Create Connection.    6. Press OK.    7. Select a cell in the MIS transactions table.    8. Go to Data > From Table/Range    9. Press Home > Merge Queries. *(interim step 01)*    10. Select the "tblBank" query in the drop-down list.    11. Select both columns in both tables (use Ctrl+click).    12. Choose Join Type = Full Outer (all rows from both) and press OK. *(interim step 02)*    13. Click the 'expand' button at the top of the tblBank column.    14. Choose all fields and Use original column name as prefix. Press OK. *(interim step 03)*    15. Right-click on the 'tblMIS' query and choose Duplicate.    16. Select the tblMIS query and rename to 'Only in MIS'.    17. Filter the tblBank.Date column to 'null' only. This will show rows that are only in MIS.  *(interim step 04)*    18. Select the new query that was duplicated and rename to 'Only in Bank'    19. Filter the Date col to 'null' only, to show rows that are only in Bank. *(interim step 05)*    20. In both queries, remove the columns that show only 'null' values. Use the Home > Choose Columns to do this.    21. In both queries, click the data type button in the date column and choose Date, and in the Amount column, choose Currency. *(interim steps 06, 07)*    22. Click Close & Load and choose Only create connection.    23. Go to the Reconciliation sheet and select D7.    24. Go to Data > Queries & Connections. In the Queries pane, right-click on the 'Only in Bank' query and choose Load To…    25. Choose Table, Existing worksheet =$D$7. Press OK.    26. Select G7 and repeat similar steps for the 'Only in MIS' query. |
|  |
| 1. STEPS – Unpivot data that's pre-aggregated    1. Open *'04 PQ unpivot.xlsx'*    2. Go to the Data tab and select a cell in the table of sales data.    3. Press Data > From Table/Range    4. Remove the Changed Type step.    5. Select the Store column and press Transform > Fill > Fill Down. *Notice that it doesn't fill down. This is because the values need to say 'null'.*    6. Remove the Fill Down step.    7. Press Transform > Replace Values. Type 'null' in the Replace With field and press OK.    8. Do the Fill Down step again.    9. Select the Store and Item columns (hold shift and click)    10. Choose Transform > Unpivot Columns drop-down > Unpivot Other Columns    11. Rename Attribute column to Date and Value to Amount    12. We want to convert the Date column to dates, but it won't work with the format of mmm-yy.        1. Select the Date column and choose Transform > Replace Values.        2. Type '-2' and replace it with '-202' which will create 4-digit years.    13. Press the data type button for the Date column and choose Date.    14. Choose File > Close & Load to…    15. Choose Table, Existing worksheet, 'PQ1'!$D$5 and press OK.    16. Go to the PQ1 sheet and select a cell within the data.    17. Choose Insert > PivotTable |
|  |
| 1. STEPS – Import from folder    1. Open *'05 PP from folder\_step1.xlsx'*    2. Go to the PQ1 sheet.    3. Go to Data > Get Data > From File > From Folder    4. Navigate to the folder 'SalesCSVs' and click Open.    5. You should see a list of files in the preview window.    6. Click Transform Data    7. Click the filter button on the Extension column and filter to only '.csv'.    8. Click the filter button on the Name column and go to Text Filters > Begins With.    9. In the dialog that appears, type 'InstrumentSales'    10. Press OK. *(interim step 01)(05 PP from folder\_step2.xlsx)*    11. Click the Combine Files button in the header of the Content column.      * 1. In the Combine Files dialog, make sure you see one of the sales reports in the data preview. If not, choose one of the sales reports in the Sample File drop-down.   2. Press OK. *(05 PP from folder\_step3.xlsx)*   3. In the queries list, choose 'Transform Sample File'.      * 1. On the Home tab, choose Remove Rows > Remove blank rows.      * 1. Click the Filter button for Column1 and uncheck <End of Report>, Reporting Year, and Store, since these rows are not needed. Click OK.      * 1. Click Use First Row as Headers      * 1. Select the SalesCSVs query and notice that the changes you just made are reflected there. *(05 PP from folder\_step4.xlsx)*   2. In the SalesCSVs query, select the first column which has the file name, and rename the column to 'Store'      * 1. Go to the Tranform tab and choose Extract > Text After Delimiter      * 1. In the dialog, type \_ (underscore) as the delimiter, and choose 'From the end of the input' in the advanced options. Press OK.      * 1. Then choose Transform > Replace Values.      * 1. Find '.csv' and replace with nothing. This will leave just the Store name in the column.     *(05 PP from folder\_step5.xlsx)*   * 1. In the SalesCSVs query, set the data types as appropriate on the columns by clicking the data type button in the column headers. An easy way is to select all columns (Ctrl+A or Shift+click the column headers), then press Transform > Detect Data Type.      * 1. Go to the columns that have currency values and click the data type button in the column and choose 'Currency'      * 1. Close the Power Query editor and Keep changes.   2. Select cell D5 on the PQ1 sheet   3. Go to Data > Queries and Connections to open the queries pane.   4. Right-click on the SalesCSVs query and choose Load To…   5. In the Load To dialog, choose Table, Existing Worksheet, $D$5 and 'Add this data to the Data Model'. Press OK.     *(05 PP from folder\_step6.xlsx)* |
|  |
| 1. STEPS – Power Pivot – add tables to Data model    1. Open *'06 PP Power Pivot Data Model\_step1.xlsx'*. Enable content.    2. Go to the Products sheet, and select a cell in the table.    3. Go to the Data tab of the ribbon and press 'Manage Data Model'. You may be prompted to enable the Data Analysis add-ins. Choose Enable.      * 1. The Power Pivot data model window will appear. Notice it has a tab for the SalesCSVs table. Close the window.   2. With a cell in the table on the Products sheet selected, go to the Power Pivot tab of the ribbon and press 'Add to Data Model'      * 1. Notice the table is added to the data model. Close the data model window.   2. Go to the Employees sheet and add the table to the data model as in the previous steps. *(06 PP Power Pivot Data Model\_step2.xlsx)*   3. Go to the PP1 sheet and select cell D5. Go to Insert > PivotTable > From Data Model.      * 1. In the PivotTable Fields pane that appears:      1. Expand Employees and check the box next to Name (should go to the Rows)      2. Expand SalesCSVs and check Amt Invoiced (should go to the Values)      * + 1. Why do all employees have the same amount? Because Employees and sales data are in separate tables and we need to establish a relationship between the tables.     2. In the PivotTable Fields pane, there should be a warning about this     3. Click CREATE…, and in the dialog that appears, choose the Data Model table SalesCSVs as the Table with Salesperson as the column. For Related Table, choose Data Model Table: Employees with related column: EmpID. Click OK.      * + 1. The data in the PivotTable should update to seem more realistic. Each employee should have a different Sum of Amt.   *(06 PP Power Pivot Data Model\_step3.xlsx)*   * 1. We don't need the sales table to appear in the PQ1 worksheet, so let's remove it.      1. Select a cell in the table and choose Table > Unlink.      * + 1. Select the entire table and press Delete on your keyboard.     2. Go back to the PP1 sheet and select a cell in the PivotTable. Notice that the list of tables doesn't have any duplicates now, but the data is still available in the data model.   1. Go back to the PivotTable and make the following changes:      1. Expand the ItemPrices table and drag the ItemName field down to the Rows area.      2. Again, a warning will appear about create a relationship. Don't click in it.      3. Go to the Power Pivot tab of the ribbon and choose Manage.      4. In the Power Pivot window, go to Home and press Diagram View      * + 1. Resize the tables so you can see all the fields. Notice there is already a relationship between SalesCSVs and Employees.     2. In the ItemPrices table, click on ItemID and drag/drop it on the Item field in the SalesCSVs table to create a relationship.     After you let go of the mouse, it will show the relationship line.   * + 1. Close the Power Pivot window and then return to the PivotTable. It should show more reasonable values.     *(06 PP Power Pivot Data Model\_step4.xlsx)*   * 1. Create a measure:      1. Go to the Power Pivot Tab and press Measures > New Measure      2. Table name = SalesCSVs      3. Measure name= AmountInvoicedSum      4. Formula =SUM([Amt Invoiced])      5. Formatting – choose Currency with 0 decimal places.      6. Press OK.      * 1. Remove 'Sum of Amount Invoiced' from the PivotTable by dragging it away or clicking it and choosing Remove Field.      1. *PP Power Pivot Data Model\_step5.xlsx)*    1. Add a calendar table to your data model - Create queries in Power Query to get the earliest and latest date from your data.       1. Open PQ editor by pressing Alt+F12.       2. Right-click the SalesCSVs query and choose Reference.      * + 1. Rename the new query 'DateEarliest'     2. Click Choose Columns and pick only the Date column. Click OK.     3. Right-click on the query and choose Duplicate.     4. Rename the new query 'DateLatest'     5. In the DateEarliest query, go to Transform > Date > Earliest.      * + 1. In the DateLatest query, go to Transform > Date > Latest.     2. Create a new blank query (Home > New Source > Other > Blank Query.      * + 1. Rename the new query 'Calendar'     2. In the formula bar, type the following:   = {Number.From(DateEarliest)..Number.From(DateLatest)}   * + 1. Click List Tools > To Table and click OK in the dialog.      * + 1. Rename the column to Date.     2. Change the data type to Date.     3. Add other columns to the calendar        1. Select the Date column.        2. Go to the Add Column tab and click Date > Year > Year.      * + - 1. Repeat similar steps to add the following columns.          1. > Month > Month Name          2. > Quarter > Quarter of Year          3. > Day > Name of Day     1. Click Close & Load To… and choose Only Create Connection. Click OK.     2. Right-click the Calendar query in the Queries & Connections pane and choose Load To…, choose 'Add this data to the Data Model'.      * + 1. Click Power Pivot > Manage to open the data model.     2. Click the Diagram View button     3. Click and drag the Date field from the Calendar table to the Date field of the sales table to create the relationship.     (It should appear as below afterward)     * + 1. Close the Power Pivot window. *(06 PP Power Pivot Data Model\_step6.xlsx)*   1. Add a calculated column called 'Workday' to show 1 for workdays, 0 for non-work days.      1. Open the Power Pivot window and go to the Calendar table (tab).      2. Double-click the 'Add column' header and type 'Workday'      3. In the formula bar, type the formula:  =IF(WEEKDAY('Calendar'[Date],2)>5,0,1)     All values in the column should show 1 (workdays) or 0 (non workdays).   * 1. Close the Power Pivot window.   2. Add a measure to calculate the count of work days.      1. Go to Power Pivot > Measures > New Measure      2. Name it 'CountWorkDays'      3. Type the formula =SUM('Calendar'[Workday])      4. Choose Number with zero decimals      * 1. Add a measure to calculate amount invoiced per day.      1. Power Pivot > Measures > New Measure      2. Name it AmountPerDay      3. Formula =[AmountInvoicedSUM]/[CountOfDays]      4. Choose Currency with no decimals      * + 1. Click OK and see that the PivotTable now includes the new data. *(06 PP Power Pivot Data Model\_step7.xlsx)* |
|  |
|  |