Each task must be performed individually and with full awareness. GenAI tools can be consulted to aid comprehension, but your primary goal is to develop your own problem-solving skills through individual effort, i.e., use it to enhance your understanding of solutions, not to generate them directly.

CREATE A SHORT REPORT WITH ALL SOLUTIONS

Task 1 is part of Laboratory 0 and serves as a quick recap and verification of your SQL skills. Tasks 2-5 and all Tutorials 1-3 are part of laboratory assignment 1 and are due next week.

LAB 1 – FINAL WARMUP WEEK 2

CONTENT

- Final introduction to AdventureWorks database, after this lab you are on your own (you can and should use documentation)
- Relational model – design basics
- SQL reminder – querying basics
- Pivot tables in MS Excel
- Analytical extensions to SQL

ADDITIONAL RESOURCES

- General querying information – T-SQL querying:
  - http://technet.microsoft.com/en-us/library/ms181080(v=sql.105).aspx
  - SELECT Examples (Transact-SQL) - SQL Server | Microsoft Docs
    - https://docs.microsoft.com/en-us/sql/t-sql/queries/select-examples-transact-sql
  - TSQL Tutorial - Learn Transact SQL language.
    - https://www.tsql.info
- Pivot Tables – Basic resources:
  - https://support.office.com/en-us/article/Create-a-PivotTable-to-analyze-worksheet-data-a9a84538-bfe9-40a9-a8e9-f99134456576 ;
  - https://www.gcflearnfree.org/excel2016/intro-to-pivottables/1/

TASK 1 – SQL QUERIES

SQL warmup task. Please do not use the build-in query editor – focus on pure SQL statements! Prepare adequate SQL queries to:

1. Provide information about the global sales amount (money), number of orders and volume (items sold) of the AdventureWorks business.
   a. An example of the query result is shown in Table 2.1:

Table 2.1. Result structure for task 2.1

Sales Amount	Volume	Number of orders
…	…	…

2. Provide information about the sales amount, volume, and number of orders in individual years of operation of the business.
   a. An example of the query result is shown in Table 2.2:

Table 2.2. Result excerpt for task 2.2

Year	Sales Amount	Volume	Number of orders
…	…	…	…

3. Prepare a SQL query that provides top 5 customers with the highest number of orders, try using the customer name (it might be tricky).
   a. An example of the query result is shown in Table 2.3.:

Table 2.3. Result excerpt for task 2.3

CustomerID	Last name, name	Number of orders
…	…	…

4. Prepare a SQL query that provides the names of all individual customers with the total sum of purchases (use SalesOrderHeader.SubTotal) greater than 1500USD – sorted (descending) by the total sales amount.
    a. An example of the query result is shown in Table 2.4.:

Table 2.4. Result excerpt for task 2.4

CustomerID	Last name, name	SalesAmount
…	…	…

5. Prepare a query that provides information about average price, total sales amount, and total volume in individual product categories of the AdventureWorks business.
    a. An example of the query result is shown in Table 2.5.:

Table 2.5. Result excerpt for task 2.5

CategoryID	Category name	Average price	Total Sales Amount	Total Volume
…	…	…	…	…

6. Display all subcategories which average price is higher than the average price of all categories.
    a. An example of the query result is shown in Table 2.6.:

Table 2.6. Result excerpt for task 2.6

SubcategoryID	Subcategory Name	Average price	Average price (over all categories)
…	…	…	…

7. Select sales territory (name) with sales in May 2013 higher than the average monthly sales per sales territory.
    a. An example of the query result is shown in Table 2.7.:

Table 2.7. Result excerpt for task 2.7

SalesTerritoryID	Sales Territory Name	Sales (May 2013)	Average monthly sales (per territory)
…	…	…	…

8. Create a list of sales territories (ids are enough) with an average number of orders (both real value and the largest integer less than the value) made by customers who have more than 10 orders in general (use CTE)
    a. An example of the query result is shown in Table 2.8.:

Table 2.8. Result excerpt for task 2.8

TerritoryID	Average number of orders	Average number of orders (INT)
…	…	…

9. (*) Show monthly sales amount by each sales territory in year 2013 and calculate the difference with the previous month (use 0 for 12/2012) to identify trends.
    a. An example of the query result is shown in Table 2.4.:

Table 2.9. Result excerpt for task 2.9

TerritoryID	Sales Territory Name	Mnt Sales Amt	Diff to prev
…	…	…	…

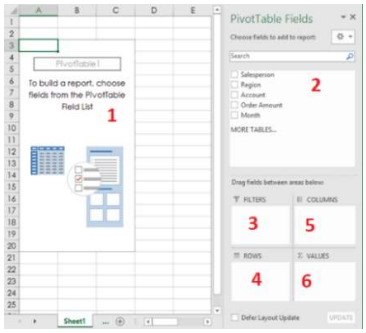
All results should be additionally stored in a single text .sql file. After completing the task, double check your approach and please upload the file to ePortal. Note that the task allows for submission of multiple files.

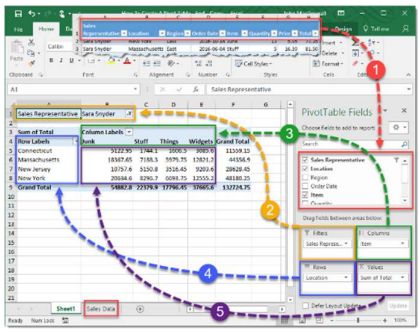
TUTORIAL 1 – PIVOT TABLES IN MS EXCEL

Pivot tables – basics. Pivot table allows for a quick, easy, and interactive analysis of multidimensional data. You can define different attributes on both axes (rows and columns), with different granularity, and the values would

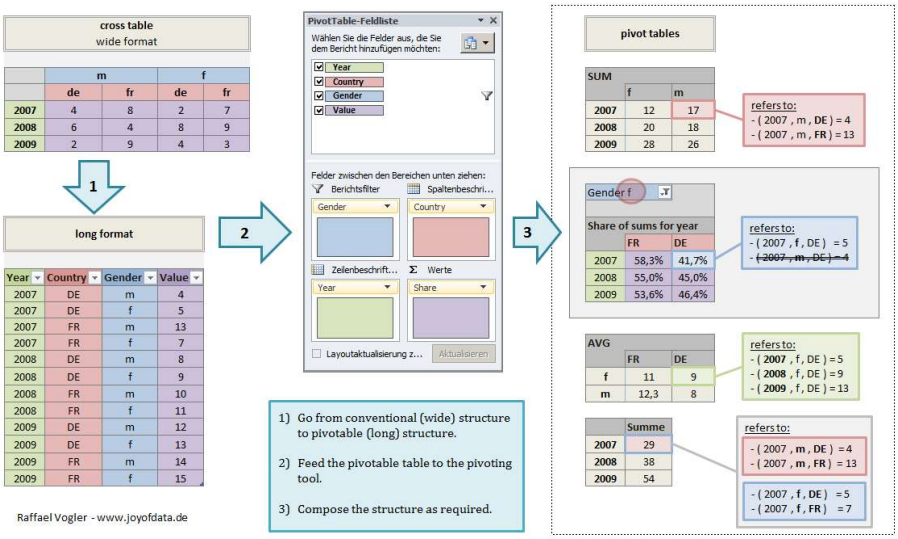
represent appropriate aggregates. Moreover, you can organise each axis in hierarchies allowing for even more customised analysis. If the provided data requires more than two dimensions you can utilise filters and work with more dimensions, although remember that the preview is always a 2D slice (filter) of the data.

1. Create a new worksheet in MS Excel and connect to [Production].[Product] table from AdventureWorks database
    a. Use Data --> New Query --> From Database --> From SQL Server option.
    b. In the database tab provide Server Name (database engine) and optionally database (AdventureWorks or leave blank).
        i. https://support.office.com/en-gb/article/Connect-a-SQL-Server-database-to-your-workbook-Power-Query-22c39d8d-5b60-4d7e-9d4b-ce6680d43bad
    c. Load the data.
2. Select all the imported data and use Insert --> Pivot Table option.
    a. Select New Worksheet to create the pivot table in new worksheet.
        i. https://www.gcflearnfree.org/excel2016/intro-to-pivottables/1/
        ii. https://www.gcflearnfree.org/excel2016/doing-more-with-pivottables/1/
        iii. https://support.office.com/en-us/article/create-a-pivottable-to-analyze-worksheet-data-a9a84538-bfe9-40a9-a8e9-f99134456576?ui=en-US&rs=en-US&ad=US
    b. In pivot table view you can utilise additional panes – pivot table view (1), available attribute pane (2), filters pane (3), rows pane (4), columns pane (5) and values pane (6).
3. Create a simple pivot table.
    a. Drag and drop attribute [DaysToManufacture] from available attribute pane (2) into rows pane (4)
    b. Drag and drop attribute [Color] from available attribute pane (2) into columns pane (5)
    c. Drag and drop attribute [ListPrice] from available attribute pane (2) into values pane (6)
    d. What information is available in the pivot table view (1)?





4. Modify the aggregation function.
    a. Use properties of values (value pane (6)) and change aggregation to Average
    b. In the table, right-click on the value field, and then select Summarize Values By
        i. https://support.office.com/en-us/article/sum-values-in-a-pivottable-9ee73790-646a-42c9-9fc7-e1ca30096d9c
5. The general process of preparing wide format data (in our case we start from step 2) and final view of the data, along with the highlighted associations to the original data, are depicted on the figure below:



6. Add [MakeFlag] attribute above and below [Color] attribute within columns
    a. Identify what changes in the table.
    b. Remove [MakeFlag] from columns.
7. Add a filter based on the [MakeFlag] attribute
        i. https://edu.gcfglobal.org/en/excel2016/doing-more-with-pivottables/1/
    a. What information is available in the pivot table view (1)?
    b. Add a slicer (Analyze -> Insert Slicer) to the table using [ProductLine] attribute.
8. Add timeline based on [SellStartDate]
    a. Use Pivot Tables Tools -> Analyze -> Insert Timeline
        i. https://support.office.com/en-us/article/create-a-pivottable-timeline-to-filter-dates-d3956083-01be-408c-906d-6fc99d9fadfa
9. Add calculated field that represents Profit (difference between cost and price)
    a. Use Pivot Tables Tools -> Analyze -> Fields, Items & Sets -> Calculated field to calculate product's profit
        i. Is it possible?
        ii. https://www.contextures.com/excel-pivot-table-calculated-field.html
    b. Select Colors and use Pivot Tables Tools -> Analyze -> Fields, Items & Sets -> Calculated Item to define a new type of color Dark representing Black, Blue and Red products (Use Black + Blue + Red as expression)
        i. https://www.contextures.com/excelpivottablecalculateditem.htm
10. Modify the table by replacing [DaysToManufacture] with [Name].
    a. Right click on product names (within the data view) and select filter -> label filters, as a filter expression provide [*Vest*]
        i. https://www.contextures.com/excelpivottablelabelfilters.html
11. Modify the way how the calculations are performed within the table:
    a. In the table, right-click on the value field (one of the values), and select Show Values As
        i. https://support.office.com/en-us/article/show-different-calculations-in-pivottable-value-fields-014d2777-baaf-480b-a32b-98431f48bfec
    b. Try different types of calculations and analyse the results
12. Sort data within the table
    a. In the table, right-click on the value field (one of the values), and select Sort
    b. Try different types of sorting and analyse the results
13. Try changing the way how subtotals and grand totals are displayed
    a. Use Pivot Tables Tools -> Design -> Subtotals / Grand Totals

   i. https://support.office.com/en-us/article/show-or-hide-subtotals-and-totals-in-a-pivottable-fc4d8406-f230-4762-aa2f-310826f3e5e2

14. Try changing the report look
    a. Add additional attribute to rows (to introduce multiple levels of analysis)
    b. Use Pivot Tables Tools -> Design -> Report Layout / Blank Rows and Pivot Tables Styles
15. Finally add a PivotChart:
    a. You add a PivotChart by going into Analyze tab (ribbon) and selecting PivotChart (under Tools section of the ribbon), while the PivotTable. In case, you want to directly add PivotChart to data (without creating PivotTable first) you can use Insert tab (ribbon) and select PivotChart (under Charts section of the ribbon). You can also directly create a PivotChart from a database connection.
    b. Use different types of charts.
    c. After creating try modifying the table (adding, removing attributes) and see how it affects the chart.
16. After this task I assume that you already know how to perform the above-mentioned operations and you can use the presented mechanisms/techniques on external data.

You do not upload any results from this task. After completing the task, please consult any of your questions or comments with the teacher.

TUTORIAL 2 – RELATIONSHIPS AND SQL BASED CONNECTIONS

When connecting to MS Excel in the previous task you've already used multiple tables. However, please notice that we haven't defined the relationship between the tables. MS Excel managed to infer these from the integration consistency rules defined within the database, i.e., foreign keys. In general, when you import related tables from a relational database, Excel will automatically create those relationships and store them in the Data Model – all this is happening behind the scenes. For all other cases, you'll need to create relationships manually.

Let us now focus how such relationship can be managed manually:

1. Make sure the workbook contains at least two tables, and that each table has a column that can be mapped to a column in another table.
    a. Connect from Excel to two product tables – Product, ProductSubCategory
    b. Click Data > Relationships.
        i. If Relationships is greyed out, your workbook contains only one table.
    c. Double check that the current set of relations is correct.
2. Create a simple pivot table – use product's subcategory name in rows and count number of products as values
3. Remove current relationships
    a. Click Data > Relationships.
    b. Remove all relations.
4. Refresh pivot table – click Data > Refresh All
    a. What has changed?
5. Now recreate the relations – in order to add new relation between tables
    a. In the Manage Relationships box, click New.
    b. In the Create Relationship box, click the arrow for Table, and select a table from the list. In a one-to-many relationship, this table should be on the many side.
        i. Using our product example, you would choose the product table first, because many product subcategories are likely to occur on many products.
    c. For Column (Foreign), select the column that contains the data that is related to Related Column (Primary).

        i. For example, SubCategoryId is shared by both tables.
    d. For Related Table, select a table that has at least one column of data that is related to the table you just selected for Table.
    e. For Related Column (Primary), select a column that has unique values that match the values in the column you selected for Column.
    f. Click OK.
6. Refresh pivot table and check if the data returned to the original state

While connecting to the data source from Excel you are not only limited to available tables or views, you can also use a predefined query. In such a case, all of the pre-processing part (defined within the query) is performed by the database engine and not within Excel (not using its in-memory Power Query engine).

Let us now focus how such query-based connection can be created:

0. Connect to a SQL Server database. In the Data tab of the ribbon, click Get Data > From Database > From SQL Server Database.
1. In the Microsoft SQL Database popup window:
    a. Specify the Server and Database from where you want to import data using native database query. Database is required when you want to use SQL query.
    b. Expand the SQL Statement / Advanced options field and paste or enter your native database query, then click OK.
        i. Use a query that returns all products from Products and matching subcategories from ProductSubCategory, and matching categories from ProductCategory tables.
2. Run native database queries.
    a. If this is the first time you're connecting to this server, you'll see a prompt to select the authentication mode to connect to the database. Select an appropriate authentication mode and continue.
3. Double check that the connections is stored – go to Data > Queries & Connections; In the Queries & Connections panel you should see all defined connections.
4. Notice that, on the Excel part such a dataset is viewed as a regular table or a database view.

You do not upload any results from this task. After completing the task, please consult your results with the teacher.

TUTORIAL 3 – PIVOTING AND GROUPING SETS IN SQL

Let us focus on selected detailed querying capabilities of SQL. In particular, we are interested in accessing summary information from a strictly operational database. Let us assume that we want to assess the yearly performance of individual sales representatives working for AdventureWorks company. As such, we will focus on defining a set of basic SQL queries, which retrieve all the required information that can be further reported to the management.

The major goal of this task is to investigate, on a real example, how easy it is to run analytical processing tasks, here a particular query, on a strictly transactional database. Additional resources regarding needed T-SQL clauses are available below:

- *Overview: https://docs.microsoft.com/en-us/sql/t-sql/queries/queries*
- *T-SQL querying: http://technet.microsoft.com/en-us/library/ms181080(v=sql.105).aspx*
- *PIVOT: http://technet.microsoft.com/en-us/library/ms177410%28v=sql.105%29.aspx*
- *CASE: http://msdn.microsoft.com/en-us/library/ms181765.aspx*
- *GROUP BY ROLL UP, CUBE and GROUPING SETS: https://learn.microsoft.com/en-us/sql/t-sql/queries/select-group-by-transact-sql*
- *OVER: https://docs.microsoft.com/en-us/sql/t-sql/queries/select-over-clause-transact-sql*

TASK 2 – PIVOT TABLES IN EXCEL

Let us now try to analyse some basic information available in the AdventureWorks database directly using basic tables covering information about orders, customers, and products.

Note that while defining a connection, within MS Excel, to data stored in SQL Server you can tick the "select multiple tables" option and load data from multiple tables (table selection window). If these tables are connected through a foreign key constraint, Excel will automatically create proper relationships between the tables and understand their basic structure (though you can check/modify them under Data tab (ribbon) and Relationship option (under Data Tools section)). These multiple tables can be further used in a single PivotTable or PivotChart.

Do not load data directly into MS Excel sheets – please only define connections to SQL Server and create Pivot Tables using defined data model / queries. As such, resultant .xls file is not bloated with data (only Pivot table results are stored – not the detailed data).

Focus on using the following set of tables (not limited to and not all might be needed): Production.Product, Production.ProductSubCategory, Production.ProductCategory, Sales.SalesOrderHeader, Sales.SalesOrderDetail, Sales.Customer, Person.Person, Sales.Store, Sales.SalesTerritory, Sales.SalesPerson. Look at the data dictionary and data stored within the AdventureWorks database. ***Store each result as a separate sheet (please use meaningful sheet names)****.*

1. Prepare a pivot table that displays the number of orders per different order statuses and order types (online/instore).
2. Prepare a pivot table that displays the number of products per color and subcategory – limit to subcategories containing word "Bike" in their name.
3. Prepare a pivot table that displays the percentage of the number of orders (number of transactions) per different territories (use territory name).
4. Prepare a pivot table that displays the sum of orders (value-wise) and total volume (quantity-wise) per product category (category name is required) with a simple PivotChart (use bar chart).
    a. Add a slicer with order date and filter the data to year 2013.
5. (*) Create a pivot table that displays the number of customers per different customer types and sales location.

All results should be stored in a single MS Excel file (with all pivots in each sheet) and a text file (PDF – with all short description and answers). After completing the tasks, double check your approach and please present your results to the teacher.

TASK 3 – PIVOT

Pivoting – prepare SQL queries within a tool of your choice (like SQL Management Studio or Azure Data Studio). Please prepare for each subtask a query utilising ***Pivot operator*** *which as a result:*

1. provides information about product's subcategory, product's colour and total sales value; please do not use pivoting here – this will be our base query for the pivot.
2. provides information about total sales value for different colours of different product subcategories; please put colours on columns, products' subcategory names on rows, and total sales value as values; use the query from the previous point.
    a. please prepare a version of this query, where only subcategories from category *Bike* are presented.
3. (*) provides information about average sales subtotal amounts in years and months; please put months on columns, years on rows, and subtotal as values – do this without manually specifying all individual years.

TASK 4 – CASE

Please prepare for each subtask – query, result (or at least its excerpt). Please use the data structures created in the previous Lab Assignments. ***Use Case expression*** *– prepare SQL queries which as a result:*

1. provides different product's price categories:
    a. ListPrice < 20.00 – Inexpensive
    b. 20.00 < ListPrice < 75.00 – Regular
    c. 75 < ListPrice < 750.00 – High

   d. 750.00 < ListPrice – Expensive
2. provides information about total volume of product for different price categories and different product categories – use price categories in columns, product categories in rows, and total volume as values (0 - never sold a product from a given category); please use CASE to put years on columns

TASK 5 – GROUPING OPERATORS

*Please prepare a series of SQL queries using **Grouping operators**. For each point, please prepare a single SQL query, which as a result:*

1. provides total sales amount for different product categories along with a total value of sales for all categories.
2. provides total sales amount for each product category and color of products (include also products without specified color), for each color, total for each category and total sales amount:
   a. Please use grouping function to identify the total sales amount from the sales amount for products without specified color:
      i. https://docs.microsoft.com/en-us/sql/t-sql/functions/grouping-transact-sql
3. provides total sales amount for different products (use product's name) in different product categories and subcategories – please provide sales summaries for each category and subcategory and a total value.

REMARKS:

- *Try to prepare a short report from your work*
- *A submission without final conclusions will not be checked and results in a negative score!*
  - If you are not preparing a report, please use comments on Eportal to provide your conclusions.