LAB 3 – POWER BI DESKTOP

RESOURCES AND TUTORIALS FOR POWER BI DESKTOP

- Power BI Desktop basics
    - Power BI for consumers - https://docs.microsoft.com/en-us/power-bi/consumer/end-user-reading-view
    - Power BI Desktop - https://docs.microsoft.com/en-us/power-bi/desktop-what-is-desktop and https://docs.microsoft.com/en-us/power-bi/fundamentals/desktop-getting-started
    - Power BI Desktop – connect to data https://docs.microsoft.com/en-us/power-bi/desktop-quickstart-connect-to-data
- Basic data visualisation using Power BI Desktop
    - https://powerbi.microsoft.com/en-us/documentation/powerbi-desktop-getting-started/
    - https://docs.microsoft.com/en-us/learn/modules/visuals-in-power-bi/
    - https://docs.microsoft.com/en-us/power-bi/visuals/power-bi-report-add-visualizations-i
    - https://docs.microsoft.com/en-us/power-bi/visuals/power-bi-report-add-visualizations-ii
- Basic data transformation using SQL Server and Power BI Desktop
    - https://docs.microsoft.com/en-us/power-bi/desktop-shape-and-combine-data
- Interesting examples of capabilities of Power BI Desktop:
    - https://docs.microsoft.com/en-us/power-bi/desktop-tutorial-facebook-analytics
    - https://docs.microsoft.com/en-us/power-bi/connect-data/service-tutorial-use-cognitive-services

TUTORIAL – POWER BI DESKTOP – BASIC READING

*Please read the following materials about Power BI Desktop provided by Microsoft*

1. Introduction to Power BI:
    a. https://docs.microsoft.com/en-us/power-bi/fundamentals/desktop-what-is-desktop
    b. https://docs.microsoft.com/en-us/power-bi/fundamentals/desktop-getting-started
    c. https://social.technet.microsoft.com/wiki/contents/articles/53078.power-bi-import-mode-vs-directquery-mode.aspx
2. Basic data visualisation using Power BI Desktop – https://powerbi.microsoft.com/en-us/documentation/powerbi-desktop-getting-started/
3. Basic data transformation using SQL Server and Power BI Desktop – https://docs.microsoft.com/en-us/power-bi/desktop-shape-and-combine-data
4. Basic data reporting using Power BI Desktop. Additional resources - report view: https://powerbi.microsoft.com/en-us/documentation/powerbi-desktop-report-view/, visualisations: https://powerbi.microsoft.com/en-us/documentation/powerbi-service-visualization-types-for-reports-and-q-and-a/.

TUTORIAL – POWER BI DESKTOP – MODELLING AND VISUALISATION

Basic data visualisation using Power BI Desktop – https://powerbi.microsoft.com/en-us/documentation/powerbi-desktop-getting-started/

1. Create a new project in Power BI Desktop.
    a. Connect to a database server (use Get Data/SQL Server connection):
        i. Please use proper authentication method. Use DirectQuery as a connectivity mode to minimize workstation's memory usage. Use the vSalesPerson view – a very small view. If you're having troubles – look at https://docs.microsoft.com/en-us/power-bi/desktop-quickstart-connect-to-data
   b. Load the data

2. Enter the report view of the project
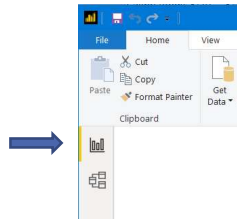
3. Create simple Client bar chart – micro tutorial
    a. From the visualisations pane select bar chart and click on it – it should appear on the report editor panel
        i. For details on different visualisations in Power BI Desktop look at:
           https://powerbi.microsoft.com/en-us/documentation/powerbi-service-visualization-types-for-reports-and-q-and-a/
        ii. You can move around and resize the element as you wish
        iii. You can configure the bar chart by selecting it on the report page and modifying its settings (below the visualisations menu)
        iv. You have three basic tabs (Visualisation pane - far right of the main window): Fields, Format and Analytics
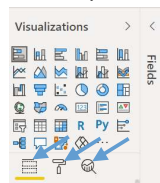
           1.
           2. In the fields tab you can define what data is displayed, in format you define how it is displayed, and in analytics you can add some predictive features (depend on the selected visualisation type)
    c. In order to select attributes you can use the Fields pane, where all available data elements are present. To select an attribute simply drag and drop elements between Fields (from available data) and Visualisation (chart settings). If you just click on a particular attribute the system (based on attribute's data type) will try to guess the best placement
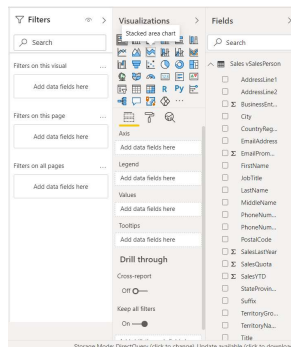
    b. You can configure the chart by selecting it (on the report page) and modifying its settings (below the Visualisations pane):

i. You have three basic tabs – Fields, Format and Analytics. In the fields tab you define which attributes are displayed, in format you define how the visualisation is displayed, and in analytics you can add some predictive features to the visualisation (only available to selected visualisations)

c. You can further add filters using the Filters pane

   a. Ref: https://docs.microsoft.com/en-us/power-bi/power-bi-report-add-filter
   b. Please note there are different types of filters – Advanced, Basic and TopN – and different contexts for the filters – visual, page and pages.

4. Prepare a simple bar chart that displays how many sales representatives are in different cities.

   d. Select IDs as values and change aggregation function to Count
   e. Select City as Axis

5. Add filter using the *Filters* pane

   f. Drag and drop Country attribute to filters related to this particular visual, select basic filtering and choose one of the available countries
   g. Modify the filter to select the country with most representatives (use TopN)

6. Add additional visualisation to the report and create a simple filled map.

   h. Prepare a simple filled map chart that displays the information how many different representatives are in different countries.

To the available dataset add an additional table that contains basic order information (just the header) – to do so you need to enter the Power Query Editor, define connection to external data source and define proper data model. Use the project from the previous task.

7. To the dataset add additional table that contains basic order information (just the header)

   i. Use Get Data and attach Sales.SalesOrderHeader table – load data.

8. Prepare a chart to display total sales amount generated by each sales representative (use Last name).

   j. Is the data correct? If yes – prove it; If not – describe why.

9. To define proper model:

   k. Go to the Model view of the project.
      i. https://docs.microsoft.com/en-us/power-bi/desktop-relationship-view
      ii. Note that in the Model view you can also specify storage models for tables, define aggregations, hide columns, organize columns into folders, manage columns formatting and define sorting.
   l. Alternatively you can manage relationships using Manage Relationship in the ribbon (under Home tab)
      i. https://docs.microsoft.com/en-us/power-bi/desktop-create-and-manage-relationships
   m. Set up proper connection between orders and sales reps – drag and drop SalesPersonID (SalesOrderHeader table) attribute onto BussinessEntityID (SalesPerson view). Define proper relationship.

10. Prepare a chart to display total sales amount generated by each sales representative (use Last name).

    n. Is everything ok with data now?
    o. Filter out sales that are not associated with any of the sales representatives.

11. Add additional visualisation to the report and create simple map.

    a. Prepare a filled map chart that displays countries that the sales reps are located (add total sales as tooltip).
       i. If nothing is displayed on the map, make sure to change the *Data Category* (under column tools of current selected column) of the country attribute to *Country*
    b. Change visualisation to a simple map chart that displays total sales (use size) for different countries that the sales representatives are located (use location).

    c. Modify the simple map chart to display total sales (use size) for different cities that the sales representatives are located (use location).

        i. Again, use proper *Data Category*

12. Save project.

TUTORIAL – POWER BI DESKTOP – BASICS OF DATA PROCESSING

Basic data transformation using SQL Server and Power BI Desktop – https://docs.microsoft.com/en-us/power-bi/desktop-shape-and-combine-data

We will now define some basic data transformation tasks, on the dataset from the previous task, to prepare the data for handling specific user information needs. In particular, we will prepare data for basic analysis of sales business process from shipment and billing location (business entities) perspectives and sales person (business entity) perspective. As a result, you need to prepare three Queries (based on connections to tables in AdventureWorks database) – SalesData (with information about sales, in the context of shipment, billing location and salesperson), SalesPersonDIM (with information about sales people) and LocationDIM (information about different locations). The latter requires a new connection to the Person.Address table).

1. Create new project in Power BI Desktop
2. Create new connection to
    a. vSalesPerson table - Use Import connection mode – as we want to further modify the data and process it in-memory
    b. Person.Address table - Use Import mode.
    c. SalesOrderHeader table - Use DirectQuery connection mode – as the table contains many rows, and wouldn't like to process this data in-memory
3. Enter Query Editor (Home / Transform Data in the ribbon menu).
    a. Read the short description available at https://docs.microsoft.com/en-us/power-bi/desktop-getting-started#shape-and-combine-data
    b. Verify that all queries are already available there – vSalesPerson, Address and SalesOrderHeader.
4. Prepare SalesData - in order to do so we need to focus on Sales.OrderHeader query:
    a. Turn on data profiling options – go to View and tick Column Quality, Column Distribution, Column Profile (in Data Preview) options. You should now be able to get basic information about the profile of data stored in each particular column.
    b. Remove all columns besides: order id, subtotal, freight, tax, totaldue, status, territoryid, customerid, shiptoaddressid, billtoaddressid, salespersonid, orderdate
    c. Introduce separate values representing day, month, year of order
        i. Use Add Column / Date
    d. Introduce a new calculated attribute – calculated as the ratio of freight and subtotal
        i. Use Add Column / Custom Column
5. Prepare – LocationDIM - in order to do so we need to focus on Person.Address query:
    a. Remove all columns besides addressid, addresslines, city, stateprovinceid
    b. Define integrated addressline column
        i. Use Transform / Merge Columns
    c. Before continuing add connection to some missing tables: Person.StateProvince, Sales.Territory and Person.CountryRegion
    d. Merge data with Person.StateProvince table (use Home / Merge Queries).
        i. Use Home / Merge Queries
        ii. Extend data to include information about name, stateprovincecode, territoryid and countryregioncode.
    e. Merge data with Sales.Territory table and include information about territory name.

        f. Merge data with Person.CountryRegion to get country name.
6. Prepare – SalesPersonDIM (based on vSalesPerson query)
        a. Introduce concatenated FullName field – with FirstName and LastName
            i. Hint! Use & operator.
7. Add proper relation between SalesPersonDIM and SalesData, and LocationDIM and SalesData

Basic calculations using SQL Server and Power BI Desktop – https://docs.microsoft.com/en-us/power-bi/desktop-tutorial-create-measures and https://docs.microsoft.com/en-us/power-bi/desktop-tutorial-create-calculated-columns

Some needed calculations on the available data can be defined at the level of a particular report, note that these calculations do not happen during the process of data transformation, but later at data consumption stage. Note that – in general – it is better to perform all data cleaning, conforming and extension tasks during data transformation stage. In general, we can introduce new measures (that describe the performance of the analysed business process; numerical) – called calculated measures (as they are calculated at the consumption stage), and new attributes (that describe some aspect of the utilised business entities; typically, textual) – called calculated columns. FYI - the language used to define calculated measures and calculated columns is called DAX (https://docs.microsoft.com/en-us/power-bi/desktop-quickstart-learn-dax-basics) – we won't be using it extensively throughout this course.

8. Within the Report view, add a measure *RatioSalesFreight* that represents the ratio of total sales amount (without tax and freight) and total shipment costs (freight).
        a. Select SalesData, in the Fields pane, and use Modelling / New Measure.
        b. Use basic calculation
            i. RatioSalesFreight = DIVIDE(SUM(MeasureR = divide(SUM('Sales SalesOrderHeader'[TotalDue]),SUM('Sales SalesOrderHeader'[Freight]))
        c. Create a table visualisation.
            i. First use only RatioSalesFreight measure – what value(s) are visualised?
            ii. Next, add SalesOrderID to the visualisation – what value(s) are visualised?
            iii. Next, change the aggregation of SalesOrderID to 'Don't summarize' – what value(s) are visualised?
            iv. Next, remove SalesOrderID from the visualisation and add SalesPersonID, change the aggreagation of SalesPersonID to 'Don't summarize' – what value(s) are visualised?
9. Add a calculated column that allows to uniquely identify a sales person – use sales person ID and last name in the following format "SalesPersonID – Last Name"
        d. Select Sales.SalesOrderHeader, in the fields pane, and use Modelling / New Column.
        e. HINT: Use RELATED function
10. Add calculated column that indicates whether an order was successfully completed (use status attribute).HINT: Use IF function.
11. Consider using grouping and binning of attribute values:
        f. Group two sales representatives together – Vargas and Valdez
            i. https://docs.microsoft.com/en-us/power-bi/desktop-grouping-and-binning
        g. Define months from Shipment date
            i. https://docs.microsoft.com/en-us/power-bi/desktop-grouping-and-binning
12. Create basic visualisations that utilise the introduced calculated measures and calculated fields.
13. Save project, upload results.

TASK 1 – DATA VISUALISATION IN POWER BI DESKTOP USING SOURCE DATA

Let us now look at some reporting capabilities of the prepared data model and the Power BI Desktop tool (PBI for short). We plan to use the order's data to create a simple set of visualisations to aid to analysis of sales from two basic points of view – focusing on product and sales representative. In particular, you are now the BI user, and your goal is to consume the available AdventureWorks data to gather some insights about the performance of

individual sales representatives. Consider a situation, in which you are trying to identify the most prominent sales representatives and their sales focus.

Let us now try to analyse some information available in the AdventureWorks database. We will focus on sales data (based on completed orders) from sales representative and product perspectives. We would like to be able to evaluate the performance of sales using different properties of ordered products. For example, we would like to be able to create a pivot chart to support questions like: "what are the top 3 sales representatives?", "Are sales handled by sales representatives higher than sales through other channels?", "For a given sales representative, is the profit from products with color "red" higher than the profit from other products?", "For a given sales representative, does the total number of sold items depend on the weight category (light, medium, heavy) of products?", "what are top 5 popular product's subcategories, for a given sales representative?", "Does a given sales representative generate more revenue from cheaper products than from more expensive ones?".

Using the source data from AdventureWorks (connect to required tables and make sure that all relations are present in the PBI model) in Power BI Desktop generate a set of basic analytical reports (use multiple report pages). In this approach <u>do not</u> modify the source data using SQL and <u>do not</u> modify the data using Power BI Desktop data transformation capabilities – use data as-is:

1. Please visualise the data in a report to show Order('Lastname, Firstname', Year, Month, ProductCategory, SalesValue). If this is not possible, just explain what is the issue?
2. Prepare a dashboard (a single report page with arranged and interconnected collection of visualisations) to help identifying the most prominent sales representatives – focus on total sales amount, total quantity of sold products, and total discount. Use time dependencies here. If there are some issues with certain elements, please comment.
3. Prepare a dashboard (a single report page with arranged and interconnected collection of visualisations) to help identifying the most prominent product categories for a specified sales representative – focus on total sales amount, total quantity of sold products, and total discount. Use time dependencies here. If there are some issues with certain elements, please comment.

TASK 2 – WIDE TABLE APPROACH

Wide Table approach is a design approach, in which all the data from the OLTP relational schema is joined into one wide denormalized table. As a result, this approach is commonly called the "One Big Table" (OBT for short, this is also probably the more popular name of this design approach). In short, this method, characterized by its simplicity, involves storing data in a single, expansive table, offering a much-simplified data model by significantly reducing the number of tables that need to be governed and updated.

Previously used approach allowed us to analyse data in a single table, or multiple tables using some of the relationships defined between the data. Moreover, in all cases we used the data as-is — directly the way the data is defined and stored within the operational database. Now, let us try to analyse data using Pivot Tables, but first we are going to prepare it – transform to a more suitable form. To do so, we will follow a very simple, somewhat naïve, process: identifying needed data, creating dedicated view and preparing a resultant pivot table.

Using a dedicated tool (Azure Data Studio, SQL Management Studio, etc.) prepare data in a more analysis friendly structure. Before implementation, please first read the entire task.

1. First focus on identifying main tables and important attributes related to products. Look at the data dictionary and investigate attributes available in product-related tables (start analysis from Production.Product) – consider using attributes that are useful in our main goal:
    a. Do we need information about all products? Be able to justify your selection (or lack) of certain attributes.
    b. At least, use the following product details: product name, category name, subcategory name, product's catalogue price, product's color, product's weight category.
2. Next focus on identifying main tables and attributes related to sales representatives. Look at the data dictionary and investigate attributes available in order-related tables:
    a. Just focus on very basic information – include first name, last name and full name of the sales representative.

3. Next focus on identifying main tables and attributes related to time/dates. Look at the data dictionary and investigate attributes available in order-related tables:
    a. Do we need information about all dates?
    b. Identify needed attributes representing time/dates – like month, year. Be able to justify your selection (or lack) of attributes.
    c. Note – use DATEPART (T-SQL function) to extract required information.
4. Next focus on identifying main tables and attributes related to orders (from product and time perspective). Look at the data dictionary and investigate attributes available in order-related tables:
    a. Do we need information about all orders?
    b. Identify needed performance measures – like sales amount, sales volume, profit – that can be used to analyse and measure the performance of sales process. Be able to justify your selection (or lack) of attributes.
5. Prepare a <u>single SQL query</u> which provides basic information about product's sales – covering all user needs:
    a. Note that you need to include product id, product name, product category name and product subcategory name, product color, product weight. Consider including other attributes (use results from point 1). Focus on order's date.
    b. Note that you need to include sales representative information. Consider including all required attributes (use results from point 2)
    c. Note that you need to include time/date information. Consider including all required time/date indicators (use results from point 3)
    d. Note that you need to include sales amount and order quantity. Consider including other performance measure (use results from point 4)

Let us now look at some reporting capabilities of the prepared data model and the Power BI Desktop tool (PBI for short). We plan to use the order's data to create a simple set of visualisations to aid to analysis of sales from two basic points of view – focusing on product and customer. In particular, you are now the BI user, and your goal is to consume the available data (prepared in the previous task) to gather some insights about the performance of individual sales representatives. Consider a situation, in which you are trying to identify the most prominent sales representatives and their sales focus.

Connect to the prepared data from Power BI Desktop, either using a view or directly using the prepared query. In a new report create a data visualisation for each of the following user requirements:

6. Please visualise the data in a report to show Order('Lastname, Firstname', Year, Month, ProductCategory, ProductName, SalesValue).
7. Prepare a dashboard (a single report page with arranged and interconnected collection of visualisations) to help identifying the most prominent sales representatives. Use time dependencies here.
8. Prepare a dashboard (a single report page with arranged and interconnected collection of visualisations) to help identifying the most prominent product categories. Use time dependencies here.

As the result, please use SQL queries and the Power BI Desktop / MS Excel file with all pivots.

TASK 3 – DATA MODEL IN POWER BI DESKTOP

Let us now look at some data transformation capabilities of Power BI Desktop tool (PBI for short). We plan to use the order's data, extend it to include information about products, time and sales representatives. In the next task you'll create a simple set of visualisations to aid to analysis of sales from these four basic points of view – focusing on time, product, customer, and shipment location.

Using the prepared data (views from the previous laboratory assignment) use Power BI Desktop to connect to the model and extend the model by adding information about customers.

1. Connect to raw data in AdventureWorks database from Power BI Desktop. If needed define proper relationships between PBI queries.

2. Prepare a PBI query that results in a dimension representing products. Please remember that we want to have a unique set of all products. Please include product's name, subcategory name, category name, color, weight, size and listprice.
3. Prepare a PBI query that results in a dimension representing sales representatives, do not use dedicated view in AdventureWorks database. In the resultant dataset each row should represent a unique salesperson. Please make sure that first name, last name, full name, address, city, state/province, country name attributes are present.
4. Prepare a PBI query that results in a dimension representing dates. In the resultant dataset each row should represent a unique day. Please make sure that date, day name of week, day of month, month, month name, quarter, year attributes are present.
5. Prepare a PBI query that results in a table representing sales. In the resultant dataset each row should represent a unique sales item (sold order item). Please make sure that basic performance metrics are available – total sales amount, total order quantity, discount value.
6. Define proper relationships between all the newly created PBI queries.

Let us now look at some reporting capabilities of the prepared data model and the Power BI Desktop tool (PBI for short). We plan to use the order's data to create a simple set of visualisations to aid to analysis of sales from two basic points of view – focusing on product and customer. In particular, you are now the BI user, and your goal is to consume the available data (prepared in the previous task) to gather some insights about the performance of individual sales representatives. Consider a situation, in which you are trying to identify the most prominent sales representatives and their sales focus.

Using the prepared data model in Power BI Desktop to generate a set of basic analytical reports (use multiple report pages):

7. Please visualise the data in a report to show Order('Lastname, Firstname', Year, Month, ProductCategory, ProductName, SalesValue).
8. Prepare a dashboard (a single report page with arranged and interconnected collection of visualisations) to help identifying the most prominent sales representatives. Use time dependencies here.
9. Prepare a dashboard (a single report page with arranged and interconnected collection of visualisations) to help identifying the most prominent product categories. Use time dependencies here.

TASK 4 – EXTENDING THE DATA MODEL

No implementation is required here.

(*) Could you explain (in more details) what changes are needed to the process to allow the end user to perform the following analysis: "Which product categories sale best before Christmas?".

CONCLUSIONS

In the conclusions, please compare the approaches used in task 1, task 2, and task 3.