LAB ASSIGNMENT 3 - TASKS

As results, please provide at least a set of .pbix files or a single .pbix file (in situation with conflicting models, consider introducing different naming conventions). Please use annotations to provide additional comments to the visualisations.

TASK 1 – POWER BI DESKTOP – BASICS OF DATA VISUALISATION

Connect to AdventureWorks database and create a very basic visualisation of sales representatives' data:

1. Prepare a simple report page:
    a. Use vSalesPerson view.
    b. Include a bar chart that displays how many sales representatives are in different cities.
    c. Add a filter to the visualisation to focus on the country with most representatives (use TopN).
    d. Add additional visualisation to the report and create a simple filled map that displays the information how many different representatives are in different countries.
2. Prepare an additional report page:
    a. Add needed tables from AdventureWorks database to present sales information.
    b. Include a chart to display total sales amount generated by each sales representative (use Last name).
        i. Is the data correct? If yes – prove it; If not – describe why.
    c. Add additional visualisation to the report and prepare a filled map chart that displays countries that the sales reps are located (add total sales as tooltip).
    d. Add additional visualisation – a simple map chart that displays total sales (use size) for different cities that the sales representatives are located (use location).
3. Save the project, upload as the solution of Task 1 and present to the teacher.

TASK 2 – POWER BI DESKTOP – BASICS OF DATA MODELLING

Please prepare the data from AdventureWorks database to handle simple sales data analysis. At first, we will focus on basic analysis of sales business process from shipment and billing location (business entities) perspectives. Please plan your approach, in terms on what data and how this data is extracted. During this task you do not have to focus on data transformation part.

First read the entire task and plan your approach.

1. Connect to AdventureWorks database, select proper tables, and prepare an adequate data model for a simple sales data reporting. We are focusing solely on sales measurements (sales values, quantity, etc.) in two perspectives – shipment and billing location.
    a. Please use the method of separating data connection data from the queries.
2. Prepare a report page. Add a bar chart and a matrix to display the following information:
    a. Count orders by different "shipment countries" and Count orders by different "billing countries"
        i. Is it possible?

Consider updating the model to support the following set of user requests.

3. Prepare an additional report page to visualise sales measurements from billing location data. Your task is to select proper visualisations and present the following information:
    a. Present sales values and quantity for different billing locations
    b. Present Top 10 billing cities (sales value wise)
    c. Present Top 10 billing cities (sales quantity wise)
    d. Present Top 10 billing cities (sales number wise)
    e. Present sales values and quantity for top 1 country (sales number wise)
4. Save the project, upload as results of Task2 and present your solution to the teacher.

TASK 3 – POWER BI DESKTOP – BASICS OF DATA MODELLING

Please extend the data model from Task 2 to include information about sales representatives associated with the sales. Consider a situation where decisions made by the manager require evaluation of sales reps performance for different business locations (assuming that customer billing location is the main operating location of the customer).

1. Prepare an additional report page to visualise sales measurements from salesperson and billing location perspectives. Your task is to select proper visualisations and present the following information:
    a. Present sales values and quantity for different salespersons and different billing location. Consider proper formatting and ordering of data elements.
    b. Identify and present billing countries for top 3 salespersons (sales value wise)
2. Prepare additional charts that allow basic analysis of this data.
    a. Use table visualisation with conditional formatting to show salesperson total sales amount.
        i. https://docs.microsoft.com/en-us/power-bi/create-reports/desktop-conditional-table-formatting
    b. Use bar visualisation to show total sales amount by different years – sort the data by sales amount
        i. https://docs.microsoft.com/en-us/power-bi/create-reports/desktop-sort-by-column
    c. Use matrix visualisation to show total sales by salesperson and year
        i. https://docs.microsoft.com/en-us/power-bi/visuals/desktop-matrix-visual
    d. Use filter – page level – to filter data based on year (order date)
        i. https://docs.microsoft.com/en-us/power-bi/create-reports/power-bi-report-add-filter
2. Prepare a set of additional 3 visualisations – for each, name the provided information, justify your selection of message (how they support the decisions) and presentation (why this type of visualisation and why such formatting).
3. Save the project, upload as results of Task3 and present your solution to the teacher.

TASK 4 – POWER BI DESKTOP – BASICS OF DATA DISCOVERY AND PROFILING

Let us now look at some data profiling capabilities of the Power BI Desktop tool. The focus is on analysis, determination of characteristics, and evaluation of a sample external dataset – "ProductRating" (provided as two external files). The available dataset contains information about a survey run on company's website to rate products.

Your task here is to study the dataset, understand and describe the domain at hand, and asses the quality of the data. As such, you are focusing on a form of basic data discovery and data profiling. Data discovery describes processes of developing certain level of understanding of the domain and the data within the dataset, typically as the initial step before data integration and/or data analysis. During profiling, you focus on creating profiles that help analysing the content, quality, and structure of data sources. As a part of the profiling process, you discover the metadata of data sources. You use different profiles for different types of data analysis, such as a column value distribution profile, column statistic profile, primary key discovery, foreign key discovery, and data domain discovery. Data profiling enables you to assess the quality of your source data before you use it in data warehousing or other data integration scenarios. In short, data profiling analyses the content, structure, and relationships within data to uncover patterns and rules, inconsistencies, anomalies, and redundancies. Finally, data discovery and data profiling offer businesses a way to make their data clean, easily understandable, and user-friendly. A comprehensive solution should be able to be used by all members of the business.

The process of profiling data in Power BI Desktop is intuitive and very simple. In short, you just need to connect to the dataset (for instance using a Text file connection) and in the Power Query Editor you have access to basic inner visualisations that support studying the data. For instance, you gain access to a series of interactive histograms that support the evaluation and study of value distributions. Bear in mind that PBI supports only column profiling, but still support structure and content discovery and analysis.

TASK 4 - DETAILS:

Study the provided dataset and present your finding according to the following scope:

- *domain data dictionary – with information about table/sheet/location, attribute name, attribute type (high level type representation, like Numerical, Money, Text, Date, etc.), description (short description of the meaning of the attribute)*

- *quality assessment of source data – with information about table/sheet/location, attribute name, attribute type (lower-level type representation, like varchar(20), decimal(5,2), etc.), type of data (nominal, ordinal, interval, ratio, continuous), number of unique values, null ratio, quality assessment description (short description of the results of the attribute quality assessment – focusing on a column consistency assessment).*

In the resultant tables, please mark all occurrences of questionable (in terms of further usage in data analysis) attributes. Use, red marking for attributes that cannot be used further, orange marking for attributes that require some additional treatment, green marking for attributes that can be used directly, no marking means you are not sure about the attribute (use this last option sparingly). Please remember to later justify your markings and decisions.

TASK 4 - SOLUTIONS:

Use this structure for providing your solutions, please remember to present two tables:

DOMAIN DATA DICTIONARY

	Location	Attribute name	Attribute type	Description
e.g.	Data.csv	Name	String, len 10	Name of the person involved in …
1				

QUALITY ASSESMENT SHEET

	Location	Attribute name	Attribute type	Type of data	# unique values	# null values	Quality assessment
e.g.	Data.csv	Name	Varchar(15)	Nominal	100	20	Requires null value handling, special characters encoding …
1							

CONCLUSIONS

Use this section to provide insights on your methodology, i.e., the process you have utilised to solve the task, and try to identify different types (categorise the issues) of quality issues.

Save the project, upload as results of Task3 and present your solution to the teacher.

TASK 5 – POWER BI DESKTOP – BASICS OF DATA REPORTING

Let us now look at some basic reporting capabilities of the Power BI Desktop tool. We plan to use the already prepared data to create a simple set of visualisations to aid to analysis of sales and rating from different points of view – focusing on sales, salesperson and rating.

1. Use the data model from previous tasks and extend it to include information about rating. For the time being focus on extracting the data and establishing proper relations – you are not required to perform any data transformation tasks (like cleaning, standardization, etc.).
2. Next try to prepare a set of additional tables/matrices (visualisation) that could support basic decision making on sales rep performance based on the comparison of average ratings associated with sales made by different customers.
3. Save project, upload results, and present your solution to the teacher.