Final Project on Data Analysis using Power BI

Guidelines:

1. Data Importing: import all files to the Power BI (hint: don't forget to check the file type to know how you will import the data).

   Solution:
   All files were in CSV format
   Get Data → Text /CSV
2. Data Transformation: before loading the data check if it needs any kind of cleaning or transformation.

   Solution:
   Perform data profiling, and address any data inconsistencies, missing values, null values, incorrect data types or repeated columns as needed.

   After uploading the file, choose "Transform" to clean the data in the files respectively

   - Customer file = Table.RenameColumns(#"Removed Columns",{{"AddressLine1", "AddressLine"}}) / Close&Apply.

   - InternetSale file = Table.RemoveColumns(#"Changed Type",{"CarrierTrackingNumber", "CustomerPONumber", "UnitPriceDiscountPct", "DiscountAmount", "TotalProductCost", "SalesAmount", "ExtendedAmount"}) / Close&Apply.

   - Product file = Table.RemoveColumns(#"Changed Type",{"EndDate"}) / Close&Apply

   - ProductCategory file = Table.TransformColumnTypes(#"Promoted Headers",{{"ProductCategoryKey", Int64.Type}, {"ProductCategoryAlternateKey", Int64.Type}, {"EnglishProductCategoryName", type text}, {"SpanishProductCategoryName", type text}, {"FrenchProductCategoryName", type text}}) / Close&Apply.

   - ProductSubCategory file = Table.TransformColumnTypes(#"Promoted Headers",{{"ProductSubcategoryKey", Int64.Type}, {"ProductSubcategoryAlternateKey", Int64.Type}, {"EnglishProductSubcategoryName", type text}, {"SpanishProductSubcategoryName", type text}, {"FrenchProductSubcategoryName", type text}, {"ProductCategoryKey", Int64.Type}}) / Close&Apply.
   - SalesTerritory file = Table.TransformColumnTypes(#"Promoted

Headers",{{"SalesTerritoryKey", Int64.Type}, {"SalesTerritoryAlternateKey", Int64.Type}, {"SalesTerritoryRegion", type text}, {"SalesTerritoryCountry", type text}, {"SalesTerritoryGroup", type text}}) / Close&Apply.

- Date file = Table.TransformColumnTypes(#"Promoted Headers",{{"DateKey", type date}, {"FullDateAlternateKey", type date}, {"DayNumberOfWeek", Int64.Type}, {"EnglishDayNameOfWeek", type text}, {"SpanishDayNameOfWeek", type text}, {"FrenchDayNameOfWeek", type text}, {"DayNumberOfMonth", Int64.Type}, {"DayNumberOfYear", Int64.Type}, {"WeekNumberOfYear", Int64.Type}, {"EnglishMonthName", type text}, {"SpanishMonthName", type text}, {"FrenchMonthName", type text}, {"MonthNumberOfYear", Int64.Type}, {"CalendarQuarter", Int64.Type}, {"CalendarYear", Int64.Type}, {"CalendarSemester", Int64.Type}, {"FiscalQuarter", Int64.Type}, {"FiscalYear", Int64.Type}, {"FiscalSemester", Int64.Type}}) / Close&Apply.

3. Data Exploring: Go to the table view: explore the data and identify which is the fact table also check each column to know what it is all about.

Solution:
InternetSale.csv (Potential Fact Table)

Based on the provided information, the InternetSale.csv file is likely the fact table. It contains columns related to sales transactions, such as order dates, order numbers, quantities, prices, and financial information.

Product.csv (dimension table)

The Product.csv file appears to contain information about the products sold. It includes columns such as product names, numbers, attributes (color, size, weight), cost, pricing, and category/subcategory associations.

ProductCategory.csv (dimension table)

The ProductCategory.csv file provides information about the different product categories. It includes columns for category keys and names in multiple languages.

ProductSubcategory.csv (dimension table)

The ProductSubcategory.csv file contains information about the subcategories of the products. It includes columns for subcategory keys, names in multiple languages, and the associated product category key.

Date.csv (dimension table)

The Date.csv file contains information about dates and their various attributes. It includes columns such as date keys, full dates, day of the week, month, quarter, year, and fiscal information. This file is a dimension table.

Customer.csv (dimension table)

The Customer.csv file contains information about the customers. It includes columns such as customer keys, names, personal details, contact information, demographic data, and occupation. This file is a dimension table.

SalesTerritory.csv (dimension table)

The SalesTerritory.csv file appears to provide information about different sales territories. It includes columns such as territory keys, alternate keys, region, country, and territory groups. This file is likely a dimension table that can be used to provide additional context and analysis when combined with other tables.

Based on the available information, the InternetSale.csv file appears to be the fact table that captures sales transactions. The other files (Product.csv, ProductCategory.csv, ProductSubcategory.csv, Date.csv, and Customer.csv) are dimension tables that provide additional information and context for analysis and reporting on the sales transactions.

4. Data Modelling: Go to the model view: check the automatic joins and create the necessary relationships if not identified by the Power BI.

Solution:
Internet Sale file is the fact sheet where all other files are connected to it in "star schema"

The tables based on their common keys or identifiers. in the given files, potential relationships could be between the following columns.

InternetSale.csv Relationship with Date.csv
PK: DateKey (Date.csv)
FK: OrderDateKey (InternetSale.csv)

InternetSale.csv Relationship with Product.csv
PK: ProductKey (Product.csv)
FK: ProductKey (InternetSale.csv)

InternetSale.csv Relationship with Customer.csv

PK: CustomerKey (Customer.csv)
FK: CustomerKey (InternetSale.csv)

InternetSale.csv Relationship with SalesTerritory.csv
PK: SalesTerritoryKey (SalesTerritory.csv)
FK: SalesTerritoryKey (InternetSale.csv)

ProductSubcategory.csv Relationship with ProductCategory.csv
PK: ProductCategoryKey (ProductCategory.csv)
FK: ProductCategoryKey (ProductSubcategory.csv)

Product.csv Relationship with ProductSubcategory.csv
PK: ProductSubcategoryKey (ProductSubcategory.csv)
FK: ProductSubcategoryKey (Product.csv)

5. Before start visualizing: go back to the table view to explore the data, identify at least 3-5 business questions as if you were a business owner.

   Solution:
   ***What is high seasons sales?**
   ***What is the highest product category in sales?**
   ***What is sales distribution & profit margin ratio by market?**
   ***What is profits & growth rate?**
   ***What is the highest products with taxes?**
   ***What is popular trend?**
   ***Markets analysis by country.**
   ***Customer analysis by Gender & Age.**

6. Write your insights and recommendations as if you were the CEO of the store (3 at least).

   Solution:
   ***More marketing activities needed production costs to be decreased in some countries with Freight costs.**
   ***More Markets needs to be approached to open new potentials.**
   ***Sales and Profit Margins are improving.**
   ***Customers with younger age group need to be approached.**

You must use few of the DAX functions that you have learnt during (use them to create at least 1-2 columns and few measures)

Solution:
Age , Age(bins) , Number Of customer , OrderSales, Profit ,Profit margin ratio, Total Sales, Growth rate.