PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

Business Case

Comprehensive Sales and Customer Analysis for AdventureWorks using R Shiny

AdventureWorks Dataset:

Reference and Background Information

AdventureWorks is a fictitious multinational manufacturing company that produces and sells metal and composite bicycles to North American, European, and Asian commercial markets. The dataset for AdventureWorks is designed as a comprehensive sample database, commonly used for demonstrating database design, querying techniques, and business intelligence practices. It is especially popular in the SQL Server community for training and educational purposes.

Components:

- **Sales Data**
  Includes detailed records of all sales transactions, customer information, and product details. It is a rich source for analyzing sales trends, customer behavior, and product performance.
- **Product Inventory**
  Covers the inventory details of products, including stock levels, warehouse information, and product categories. This part is crucial for supply chain analysis and inventory management.
- **Employee Records**
  Contains information about the company's employees, their organizational structure, and their sales performance. Useful for HR analytics and sales force performance analysis.
- **Purchasing and Supplier Information**

Details about the company's suppliers and purchasing orders, offering insights into the procurement process and supplier management.

Resources:

- **GitHub Repository**

The dataset can be downloaded from the GitHub repository. This repository provides the necessary scripts to set up the AdventureWorks database in a MySQL environment.

The database is relativly common sample database, so you can find a lot of resources online.

Schema - AdventureWorks – Data Dictionary (dataedo.com)

Requirements

- ✓ **MySQL Local Instance Setup**
  - Install a local MySQL instance.
    - MySQL :: MySQL 8.3 Reference Manual :: 2 Installing MySQL
- ✓ Download and install the AdventureWorks2019 database from GitHub.
- ✓ **Database-Centric Application**
- ✓ All operations should be conducted through the MySQL database connection.
- ✓ External data sources or offline data **are not permitted. Only exception is to use caching mechanism.**
- ✓ **No need to go in business details, keep it simple. Your ability to work with complex database schema,implementation of good practices of reactive coding and good insights will be assesed the most.**

Areas of Investigation

Exploratory data analysis

Primary Question

Calculate measures such as mean, median, mode, standard deviation, variance, coefficient of variation, and skewness to describe the distributions of sales amounts, customer demographics, product category and product price?

Tables

Sales.SalesOrderHeader | Sales.SalesOrderDetail | Sales.Customer | Person.Person | Production.ProductCategory | Production.Product

Export Features

Visualize the above-mentioned variables in appropriate plots.

Sales Trends by Product and Region

Primary Question

What are the sales trends across different territories for various product categories?

Tables

SalesOrderHeader | SalesOrderDetail | Product | SalesTerritory

Export Feature

PDF export of the sales trend report.

Inventory Management Efficiency by Product and Region

Primary Question

How effective is company's inventory management in meeting sales demands? Invetory management is company's ability to keep relevant items in their inventory on time and in right quantities

Tables

Product | ProductInventory | Location | ProductCategory

Export Feature

No export, focus on interactive dashboard insights.

Supplier Efficiency and Sales Performance

Primary Question

How do supplier efficiency and product availability correlate with overall sales performance?

Tables

Product | ProductVendor | Vendor | PurchaseOrderHeader | PurchaseOrderDetail | Employee | Person

Export Feature

PDF export of comprehensive supplier-product analysis report.

Sales prediction

Primary Question

Build regression models to predict sales based on product features and customer demographics. Explain the results to a scientist with little or no knowledge in statistics.

Tables

Sales.SalesOrderHeader | Sales.SalesOrderDetail | Sales.Customer | Production.Product | Person.Person

Export Features

In your shiny application allow choosing different dependent and independent variables.

Application Features to Have

i. **Dynamic SQL Queries with Complex Joins**
   a. The application will dynamically generate and execute SQL queries, especially for the multi-join analysis.
ii. **Interactive Data Exploration****
   a. Users can interact with the application to filter and view data in various formats.
iii. **Sophisticated Data Visualizations**
   a. Advanced graphical representations of data will be provided for clearer understanding.
iv. **Data Export Options**
   a. Functionality for exporting data as PDF or Excel files based on user requirements.
v. **All in single shiny**

**** you are free to add any filters and present any extra insights**

that you might find relevant for the users. The application should be user friendly and adaptive.

The code for the R Shiny application must be maintained in a GitHub repository (public/private as per your preference) and be shared with us. It is adviced to work with git as you would do normally during collaboration, so meaningful commits, pull requests, etc etc.(not just a single big push)