MES College of Engineering Pune-01 Department of Computer Engineering

|  |  |
| --- | --- |
| **Name of Student:** | **Class:** |
| **Semester/Year:** | **Roll No:** |
| **Date of Performance:** | **Date of Submission:** |
| **Examined By:** | **Experiment No: Mini Project** |

# ASSIGNMENT – Mini Project

# AIM: Mini Project: A BI report must be prepared outlining the following steps:

# a) Problem definition, identifying which data mining task is needed.

# b) Identify and use a standard data mining dataset available for the problem.

**TITLE:** Performing Data Mining Tasks on Adventure Works 2019 Dataset**.**

# OBJECTIVES:

# To analyze Adventure Works Data Warehouse.

# To perform various mining tasks on the data.

# To find insights regarding the data.

# REQUIREMENTS:

* Data: AdventureWorksDW2019.bak
* Tools: Power BI, Excel, SQL Server, SQL Server Management Studio.
* Data Mining Tasks: Summarization, Frequent Item Set, Data Cleaning.

**THEORY:**

Adventure Works 2019 Data Warehouse is a sample database that serves as a comprehensive data source for demonstrating data warehousing concepts, architectures, and design patterns. It has a wide range of features that make it an ideal choice for learning and practicing data warehousing concepts. Some of the key features of Adventure Works 2019 Data Warehouse are:

1. Comprehensive Data: The database contains a vast amount of data that covers different domains like sales, marketing, finance, production, and human resources. It has various fact and dimension tables that are interconnected to provide a complete picture of the organization's operations.
2. Easy to Understand: Adventure Works 2019 Data Warehouse is designed to be easy to understand and use. It has a well-structured schema and sample data that makes it simple for users to navigate and query the database.
3. Flexible Design: The database has a flexible design that allows users to add new dimensions and facts easily. It also supports different data warehousing architectures like star and snowflake schemas.
4. Realistic Data: The data in Adventure Works 2019 Data Warehouse is realistic and reflects the data of a real organization. It is based on a fictitious bicycle manufacturing company called Adventure Works Cycles and is modeled after real-world scenarios.
5. Support for OLAP: The database supports Online Analytical Processing (OLAP) and can be used with popular OLAP tools like Microsoft Excel and Power BI.

We want to get the following insights about the data:

* Internet Sales Report.
* Selling of products of client over time.
* Filtering Sales persons.
* Total Budget.

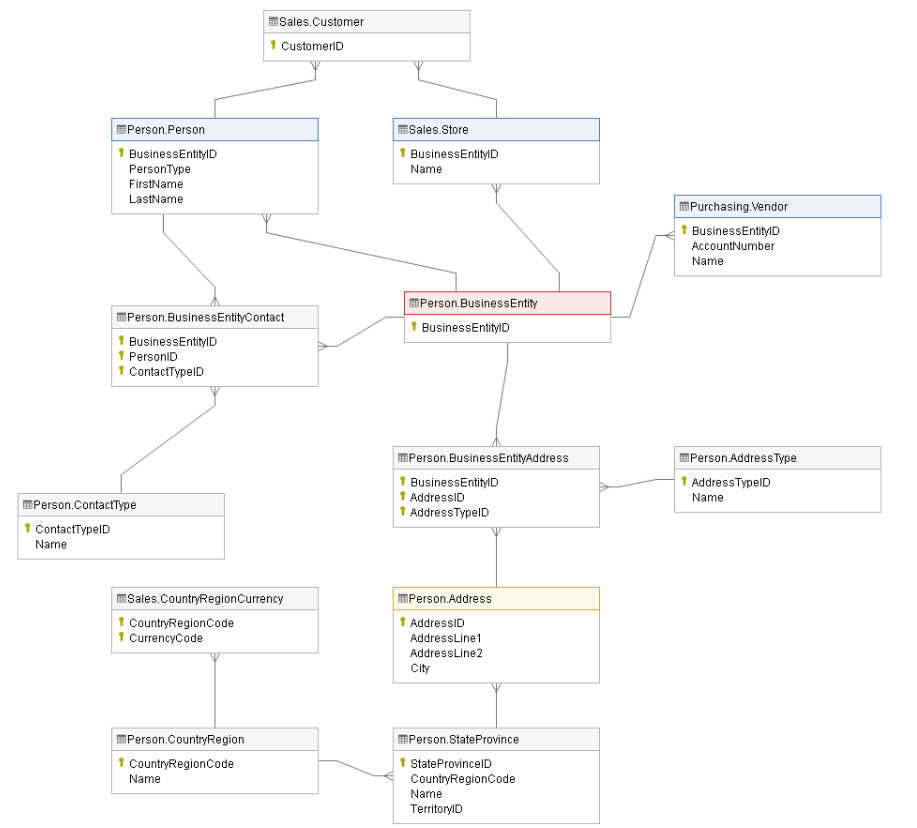
This the problem definition and the following data mining tasks need to be performed on the data. Overall, Adventure Works 2019 Data Warehouse is an excellent resource for learning and practicing data warehousing concepts. Its features make it an ideal choice for students, researchers, and professionals who want to explore data warehousing and business intelligence.

SQL Server is a relational database management system (RDBMS) developed by Microsoft. It is used to store and manage data for a variety of applications, ranging from small single-user databases to large enterprise-level systems. SQL Server supports SQL (Structured Query Language) which is used to manage and manipulate the data stored in the database. SQL Server also includes tools for managing, monitoring, and securing the database, as well as business intelligence and analytics tools for data analysis and reporting. It is commonly used in industries such as finance, healthcare, manufacturing, and e-commerce. It was used in this project to clean, update and perform SQL queries on the data.

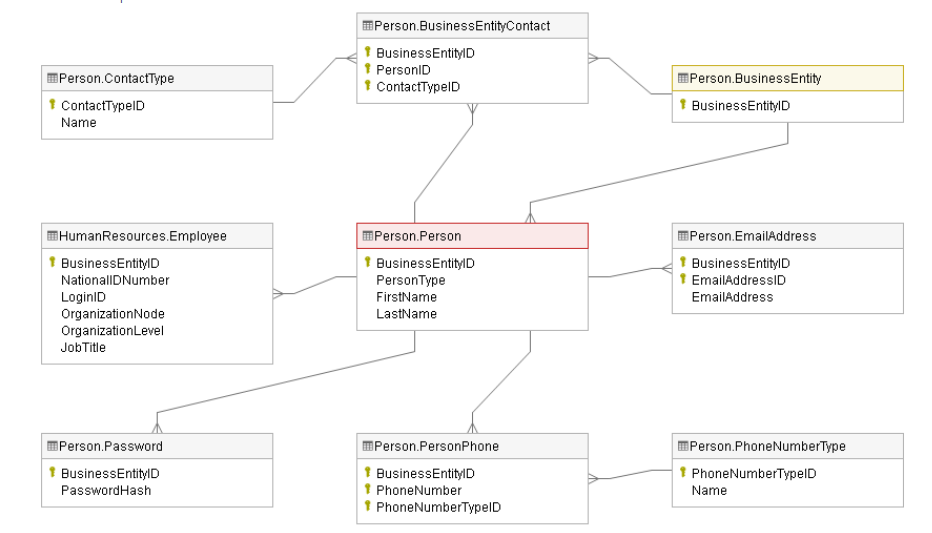
Power BI is a data visualization and business intelligence tool developed by Microsoft. It allows users to connect to a wide range of data sources, including SQL Server, Excel spreadsheets, and cloud-based sources such as Azure and SharePoint. With Power BI, users can create interactive dashboards and reports that provide insights into their data. The tool includes a wide range of data visualization options, including charts, maps, and gauges. Users can also create custom visuals using the Power BI Developer Tools.

Power BI also includes features for data modeling and transformation, such as the ability to merge and shape data from multiple sources. Users can create relationships between tables, define calculated columns and measures, and apply filters and slicers to their data.

Power BI is available in several editions, including a free version and a Pro version with additional features such as collaboration and sharing capabilities. It can be accessed via a web browser, a mobile app, or a desktop application. Power BI was used to visualize and perform data mining process such as summarization, frequent item sets and more.

****

1. Business Entities

****

2. People Entities

**DATA SET DESCRIPTION:**

The Adventure Works 2019 data warehouse has several tables with millions of rows of data, including:

# FactInternetSales: 60,398,185 rows

# FactResellerSales: 72,398,808 rows

# DimProduct: 36,306 rows

# DimDate: 30,536 rows

# DimCustomer: 18,484 rows

# DimEmployee: 29,146 rows

# The size of the data warehouse can vary depending on how many tables are used and the amount of historical data being stored. Its file size is around 100 mb. Overall, the Adventure Works 2019 data warehouse is considered a large data warehouse.

# OUTPUT (SCREENSHOTS OF IMPLEMENTATION):

# [This slide contains the following visuals: textbox ,slicer ,slicer ,slicer ,slicer ,slicer ,slicer ,BBED0A4D4D8D44CE93CF2DAFADEE466C ,pieChart ,lineChart ,Sales by Top 10 Customer ,Sales by Top 10 Product Name ,map. Please refer to the notes on this slide for details](https://app.powerbi.com/groups/me/reports/ca88fc5d-1e1d-45dc-96e6-2ca52aafe118/?pbi_source=PowerPoint)

# CONCLUSION:

# Thus we created a report outlining the problem definition, performing the mining task and used Adventure Works 2019 as our dataset.