Experiment 10
BUSINESS INTELLIGENCE LAB

Aim: Experiment to study any one BI tool such as Pentaho, Tableau and QlikView

Theory:

Definition:
Pentaho:
Pentaho is a business intelligence (BI) and data integration platform that helps organizations manage and analyze their data. It provides a suite of tools for data integration, analytics, reporting, and dashboarding. Pentaho is designed to handle both structured and unstructured data from various sources, including databases, big data platforms, and cloud services.

Key components of the Pentaho platform include:

1.   **Pentaho Data Integration (PDI)**, also known as Kettle: This component allows users to extract, transform, and load (ETL) data from various sources into a data warehouse or other target systems. PDI provides a graphical interface for designing ETL processes without requiring extensive coding.

2.   **Pentaho Business Analytics:** This component provides tools for creating reports, dashboards, and visualizations to analyze data and gain insights. Users can create interactive dashboards and reports using drag-and-drop interfaces.

3.   **Pentaho Data Mining:** This feature allows users to perform advanced analytics such as predictive modeling and clustering to uncover patterns and trends in their data.

4.   **Pentaho Big Data Integration:** Pentaho integrates with big data technologies such as Hadoop, Spark, and NoSQL databases, allowing organizations to process and analyze large volumes of data efficiently.

DATASET DESCRIPTION:
The dataset contains information about sales transactions related to various chocolate products made by different salespersons in different countries.

1. Sales Person: Names of the salespersons involved in the transactions.
2. Country: The country where the sales transaction took place.
3. Product: The type of chocolate product sold.
4. Date: The date when the transaction occurred.
5. Amount: The monetary value of the transaction.
6. Boxes Shipped: The number of boxes of the chocolate product shipped in the transaction.

Each row represents a unique sales transaction, detailing the salesperson, the country, the specific chocolate product, the date of the transaction, the amount of the transaction, and the number of boxes shipped.

The dataset includes transactions from various countries such as the UK, India, Australia, USA, Canada, and New Zealand. The products sold include various types of chocolates such as bars, cubes, bites, syrup, etc. The dataset spans transactions from January 2022 to August 2022.

TYPES OF ATTRIBUTES OUR DATASET REPRESENT:
1. Categorical Attributes:

Sales Person: Categorical attribute representing the names of the salespersons. This attribute helps in analyzing the performance of individual salespersons.

Country: Categorical attribute indicating the country where the sales transaction occurred. This attribute allows for analysis based on geographical regions.

Product: Categorical attribute specifying the type of chocolate product sold. It helps in understanding the popularity of different chocolate variants.

2. Temporal Attribute:

Date: Temporal attribute indicating the date of each transaction. This attribute enables time-based analysis such as sales trends over time, seasonal variations, etc.

3. Numerical Attributes:

Amount: Numerical attribute representing the monetary value of each transaction. It facilitates analysis related to revenue, profitability, and sales performance.

Boxes Shipped: Numerical attribute indicating the quantity of chocolate products shipped in each transaction. This attribute helps in understanding the volume of sales.

QlikView:
QlikView is a business intelligence (BI) tool developed by Qlik, a software company based in Radnor, Pennsylvania. It is designed to help businesses in analyzing, visualizing, and interpreting their data. QlikView allows users to consolidate data from multiple sources into a single application, create interactive dashboards and reports, and explore data through dynamic visualizations.

Some key features of QlikView include:

1.   Data Integration: QlikView can connect to various data sources such as databases, spreadsheets, and web services to consolidate data for analysis.

2.   In-Memory Data Processing: QlikView utilizes an in-memory data processing engine, which allows for fast data retrieval and analysis.

3.   Associative Data Model: One of the distinctive features of QlikView is its associative data model. This allows users to explore data relationships dynamically, enabling them to make discoveries and insights that might not be apparent with traditional BI tools.

4.   Interactive Visualizations: QlikView offers a range of visualization options including charts, graphs, and tables that users can interact with to explore data from different perspectives.

Tableau:
Tableau is a powerful and widely used data visualization software that allows users to create interactive and shareable dashboards, reports, and visualizations from various data sources. It enables individuals and organizations to analyze and understand their data more effectively through intuitive drag-and-drop functionality and a user-friendly interface.

Key features of Tableau include:

1.   **Data Connection:** Tableau can connect to a wide range of data sources including databases, spreadsheets, cloud services, and big data platform

2. **Data Visualization:** Users can create a variety of visualizations such as bar charts, line graphs, scatter plots, maps, and more to represent their data in a meaningful way.

3. **Dashboard Creation:** Tableau allows users to combine multiple visualizations into interactive dashboards, where users can filter, drill down, and interact with the data dynamically.

4. **Analytics and Calculations:** Tableau offers built-in analytics tools and the ability to create calculated fields, parameters, and sets to perform complex data analysis and calculations.
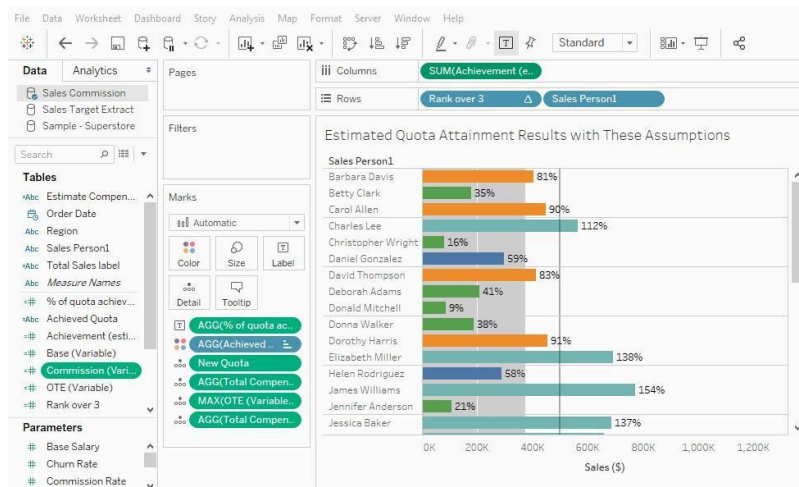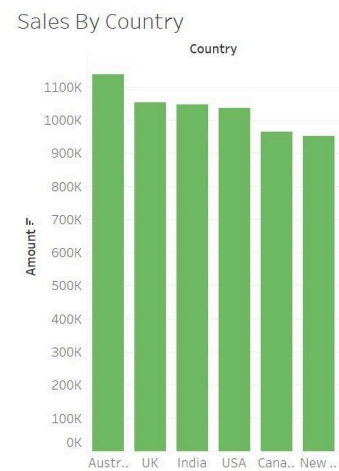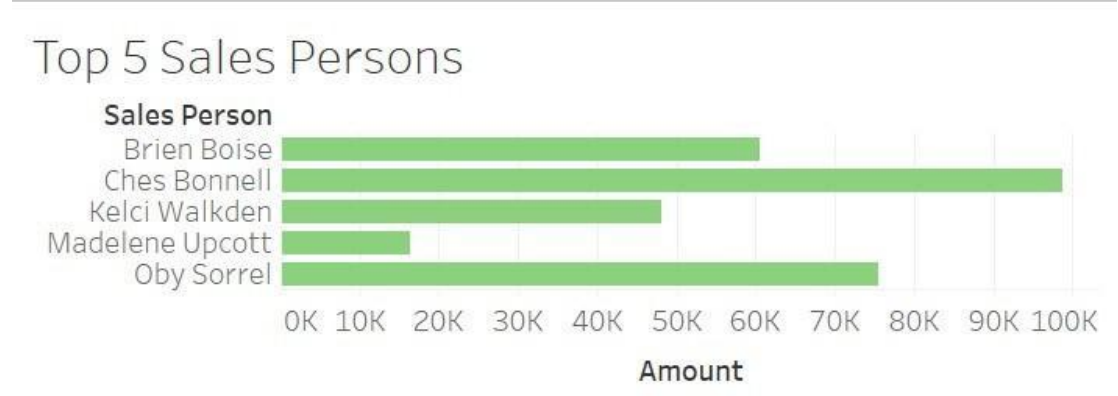
IMPLEMENTATION:

1. Load the dataset in the tool.

In this step we have loaded dataset in the form of .xlsx(excel file) in the tableau BI tool to perform further visualization and analysis.

2. Perform visualization (build Dashboard)

Sales By Country

Here we have performed the visual representation of total sales by country in sheet 1 using a bar diagram format.

Top 5 Sales Persons

Here we have performed the visual representation of top five sales persons in sheet 2 using a bar diagram format.

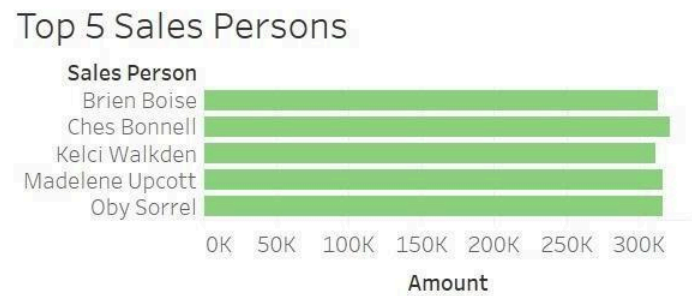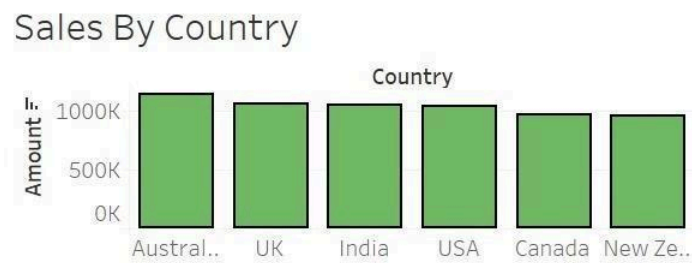
Business Summary

Amount	1,045,800
Boxes Shipped	29,470
Count of data	184

Here we have calculated the business summary by taking "Amount", "Boxes Shipped" and "Data Count" categories placed at the row with their respective values on their sides.

Dashboard:

Business Summary

Amount	6,183,625
Boxes Shipped	177,007
Count of data	1,094

Sales By Country

Top 5 Sales Persons

3.Perform data analysis.

In this part we have taken country India as a filter and accordingly found the sales persons and business summary.

3. Create a Report

The report you sent me is a business summary that includes sales by country, boxes shipped, and top performing sales people.
Sales by Country: The majority of sales come from Australia, followed by the UK, India, and the USA. Canada and New Zealand have the lowest sales out of the countries listed.
Top 5 Sales People: Brien Boise is the top salesperson, followed by Ches Bonnell, Kelci Walkden, Madelene Upcott, and Oby Sorrel.

Conclusion: Here BI tool such as Pentaho, Tableau and QlikView were studied and compared and data visualization, analysis and business summary report was made.

\