Experiment 09: Integrating Multiple Data Sources in a BI Tool and Creating Unified Reports for Actionable Insights

Title:
To integrate multiple data sources in a BI tool and create unified reports for actionable insights.

Objective:

- To understand how multiple data sources can be connected in a BI tool.
- To create unified dashboards and reports using data blending/joining features.
- To generate actionable insights by combining data from different sources.

Software/Tools Required:

- Tool: Tableau Public
- Datasets:
    - Sample Superstore Dataset (CSV)
    - People/Returns Dataset (Excel or CSV)
    - Optional: Google Sheets or Database connection
- System: Windows/Linux/Mac with internet access

Theory Overview:
Business Intelligence (BI) Tools:
Business Intelligence (BI) tools provide organizations with the capability to connect, analyze, and visualize data for better decision-making. In real-world enterprises, data rarely exists in a single file or database. Instead, it is spread across multiple sources such as Excel sheets, relational databases, cloud storage (Google Drive, OneDrive), CRM systems (Salesforce), ERP systems (SAP), and web services. BI tools like Tableau, Power BI, and Google Data Studio allow users to bring all these sources together into one integrated environment.

Data Integration in BI:
Data integration is the process of combining structured and semi-structured data from different systems into a single, unified view. It ensures that end-users can work with consistent, accurate, and holistic information. BI integration supports:

- **ETL (Extract, Transform, Load):** Extract data from different sources, transform it into a common format, and load it into the BI tool.
- **ELT (Extract, Load, Transform):** Modern approach where raw data is loaded first, and transformations happen within the BI tool.

Data Blending vs. Data Joining:

- **Joining:**
    - Used when datasets share a **common key** (e.g., Customer ID, Order ID).
    - Example: Joining Sales Orders table with Customer Details table to get complete information.
- **Blending:**

- Used when datasets come from **different data sources** and do not share an exact key.
- Example: Combining **Sales Data (Excel)** with **Target Data (Google Sheets)** to compare actual sales with goals.

Benefits of Integrating Multiple Data Sources in BI:

1. **Single Source of Truth:** Eliminates data silos by consolidating scattered datasets into a unified report.
2. **Improved Decision-Making:** Provides a holistic view of performance across departments (Sales, HR, Finance, Marketing).
3. **Cross-Functional Insights:** Enables correlation of different domains, e.g., Sales vs. Returns, Marketing Spend vs. Revenue Growth.
4. **Increased Data Accuracy:** Reduces redundancy and ensures consistency in reporting.
5. **Time Efficiency:** Automates data refreshes across sources, saving manual reporting effort.

Challenges in Multi-Source Data Integration:

- **Data Format Inconsistency:** Different sources may use varied formats (CSV, SQL tables, JSON APIs).
- **Duplicate or Missing Data:** Ensuring data quality while merging.
- **Performance Issues:** Large datasets from multiple systems may slow down dashboards.
- **Data Governance & Security:** Access control and compliance must be maintained when pulling from multiple sources.

BI Tools for Data Integration (Example: Tableau Public):

- **Multiple Connections:** Tableau Public allows importing from Excel, text files, Google Sheets, and databases.
- **Joins and Relationships:** Define how different tables connect using primary and foreign keys.
- **Data Blending:** Merge multiple datasets from different sources into a single view.
- **Unified Dashboards:** Combine KPIs from various domains into one interactive report, enabling organizations to track performance at a glance.

Procedure (Using Tableau Public):

1. **Install Tableau Public**
   - Download and install from https://public.tableau.com.
2. **Connect Multiple Data Sources**
   - Load the **Sample Superstore** dataset (CSV).
   - Add another data source: **People Dataset (Excel)** or **Returns Dataset (CSV)**.
3. **Data Integration**
   - Use **Joins** when datasets share a common field (e.g., Customer ID, Order ID).
   - Use **Data Blending** when combining unrelated sources (e.g., Sales data + Google Sheets with Targets).
4. **Create Unified Reports**

- Build a combined dashboard including:
  - Sales by Region (Superstore)
  - Returned Orders (Returns dataset)
  - Sales by Manager (People dataset)