

3.2. Student Handout

Power BI Data Sources: Student Handout

Overview

This handout provides a concise guide to understanding and working with data sources in Power BI. By the end of this guide, you will be familiar with connecting to various data sources, importing data, and creating data models.

What are Data Sources?

Data sources are origins from which Power BI can retrieve data to create reports. Common data sources include:

1. **Excel:** Widely used for storing business data.
 - Example: Sales data stored in an Excel spreadsheet.
 - Example: Financial reports maintained in Excel.
 - Example: Employee attendance records in Excel.
2. **CSV (Comma-Separated Values):** Text files with data separated by commas.
 - Example: Exported customer data from a CRM system.
 - Example: Product inventory lists in CSV format.
 - Example: Survey results saved as a CSV file.
3. **SQL Server:** A relational database management system.
 - Example: Customer transaction data stored in SQL Server.
 - Example: Inventory management data in SQL Server.
 - Example: Employee records in a SQL Server database.
4. **Azure:** Microsoft's cloud platform for data storage and management.
 - Example: Data stored in Azure SQL Database.
 - Example: Analytics data in Azure Data Lake.
 - Example: IoT sensor data in Azure Blob Storage.

5. **Web Sources:** Data retrieved from websites.

- Example: Stock prices from a financial website.
 - Example: Weather data from a meteorological site.
 - Example: News headlines from an RSS feed.
-

Connecting to Different Data Sources

To connect to data sources in Power BI:

1. Click on 'Get Data.'
2. Select the desired data source from the list.
3. Follow the prompts to connect and load data.

Examples:

- Connecting to an Excel file by selecting 'Excel' and browsing to the file location.
 - Connecting to a SQL Server database by selecting 'SQL Server' and entering server details.
 - Connecting to web data by selecting 'Web' and entering the URL.
-

Import vs DirectQuery

Import

- Data is loaded into Power BI and stored locally.
- Suitable for smaller datasets.
- Requires manual or scheduled refreshes for updates.

Examples:

- Importing a small Excel file for offline analysis.
- Importing a CSV file with monthly sales data.
- Importing a dataset from a local database for quick access.

DirectQuery

- Maintains a live connection to the data source.

- Suitable for large datasets.
- Always reflects the latest data without needing refreshes.

Examples:

- Using DirectQuery for a large SQL Server database.
 - Connecting to a real-time analytics database in Azure.
 - Accessing live data from a web API.
-

Best Practices for Data Import and Management

1. **Clean Your Data:** Ensure data is organized and free of unnecessary elements.

- Example: Remove empty columns from an Excel file.
- Example: Filter out irrelevant rows in a CSV file.
- Example: Standardize date formats in a SQL database.

2. **Use Import for Smaller Datasets:** Enhances performance for small data volumes.

- Example: Importing a small customer list for analysis.
- Example: Importing a quarterly financial report.
- Example: Importing a dataset with limited historical data.

3. **Use DirectQuery for Large Datasets:** Avoids loading large data volumes into Power BI.

- Example: Accessing a large transactional database.
- Example: Querying a big data source in Azure.
- Example: Connecting to a comprehensive product catalog.

4. **Set Up Scheduled Refreshes:** Keeps imported data up-to-date.

- Example: Scheduling daily refreshes for sales data.
 - Example: Setting up weekly updates for inventory data.
 - Example: Automating monthly refreshes for financial reports.
-

Introduction to Data Modeling

Data modeling involves creating relationships between tables to define how data is connected.

Examples:

- Creating a relationship between a Sales table and a Customer table using 'Customer ID.'
 - Linking a Product table to a Sales table via 'Product ID.'
 - Connecting an Orders table to a Date table using 'Order Date.'
-

Relationships Between Tables and Creating a Star Schema

A Star Schema organizes data with a central Fact Table and surrounding Dimension Tables.

Examples:

- Fact Table: Sales transactions; Dimension Tables: Customers, Products, Dates.
 - Fact Table: Orders; Dimension Tables: Suppliers, Products, Dates.
 - Fact Table: Financial transactions; Dimension Tables: Accounts, Departments, Dates.
-

Understanding Primary and Foreign Keys

- **Primary Key:** Unique identifier for each row in a table.
 - Example: 'Customer ID' in a Customer table.
 - Example: 'Product ID' in a Product table.
 - Example: 'Order ID' in an Orders table.
 - **Foreign Key:** Column in another table that refers to the primary key.
 - Example: 'Customer ID' in a Sales table linking to the Customer table.
 - Example: 'Product ID' in a Sales table linking to the Product table.
 - Example: 'Order ID' in a Shipment table linking to the Orders table.
-

Using the Query Editor for Basic Data Loading and Shaping

The Query Editor allows for data cleaning and transformation.

Examples:

- Removing unnecessary columns from a dataset.
 - Filtering rows based on specific criteria.
 - Changing data types, such as converting text to numbers.
-

Hands-On: Importing Data from Multiple Sources into Power BI

1. Open Power BI and click on 'Get Data.'
 2. Select 'Excel' and import a sample Excel file.
 3. Select 'Web' to import data from a website.
 4. Create relationships between tables in the 'Model' view.
-

Conclusion

This guide covered:

- Types of data sources supported in Power BI.
- Connecting to data sources.
- Differences between Import and DirectQuery.
- Best practices for data import and management.
- Introduction to data modeling and relationships.
- Using the Query Editor for data shaping.

By understanding these concepts, you can effectively gather, prepare, and organize data in Power BI to create comprehensive reports.