1.2. Student Handout

Student Handout: Introduction to Dataflows in Power BI

Overview

This handout provides a concise guide to understanding and using dataflows in Power BI. It covers the basics of dataflows, their benefits, differences between dataflows and datasets, and a step-by-step guide to creating a dataflow.

What are Dataflows?

Dataflows in Power BI are tools for collecting, cleaning, and transforming data from various sources before using it in reports and dashboards. They serve as a self-service ETL (Extract, Transform, Load) solution.

Examples:

1. **Sales Data Preparation**: Extract sales data from multiple sources, clean it, and transform it for analysis.
2. **Customer Data Integration**: Combine customer data from CRM and support systems into a unified format.
3. **Financial Data Consolidation**: Aggregate financial data from different departments for a comprehensive financial report.

Key Benefits for Self-Service ETL

Dataflows empower business users to prepare data independently, offering several advantages:

1. **Reusability**: Create once, use across multiple reports.
2. **Centralized Data Preparation**: Maintain consistency in data logic.
3. **Scalability**: Handle large datasets and complex transformations.
4. **Integration**: Connect to both cloud and on-premises data sources.

Examples:

1. **Marketing Campaign Analysis**: Reuse dataflows for different campaign reports.
2. **Inventory Management**: Centralize data preparation for inventory reports.
3. **Employee Performance Tracking**: Scale dataflows to handle large employee datasets.

Dataflows vs Datasets: When to Use Which

- **Dataflows**: Used for data preparation (ETL).
- **Datasets**: Used for ready-to-use data in reports.

When to Use Dataflows:

- Need to clean and transform data from multiple sources.
- Reuse data preparation logic across reports.

When to Use Datasets:

- Data is already clean and ready for reporting.
- No complex transformations required.

Examples:

1. **Dataflows**: Transform raw sales data from multiple regions.
2. **Datasets**: Use pre-cleaned financial data for quarterly reports.
3. **Dataflows**: Integrate and clean customer feedback from surveys.

Creating a Dataflow in Power BI Service

Step-by-Step Guide:

1. **Access Power BI Service**: Log in and navigate to your workspace.
2. **Create a New Dataflow**: Select "New" and then "Dataflow."
3. **Connect to a Data Source**: Choose from cloud-based or on-premises sources.
4. **Define Entities**: Select and rename tables or data structures.
5. **Use Power Query**: Transform and clean data using Power Query.
6. **Merge, Append, and Aggregate Data**: Combine data from multiple sources.

7. **Save and Refresh the Dataflow**: Save your work and set a refresh schedule.

Examples:

1. **Connect to Azure SQL Database**: Extract and transform sales data.
2. **Define Customer Entities**: Select customer tables for analysis.
3. **Use Power Query**: Remove duplicates and filter rows in product data.

Data Preparation in Dataflows

Use Power Query to clean and transform data:

- **Remove duplicates**.
- **Filter rows**.
- **Change data types**.
- **Split columns**.

Examples:

1. **Remove Duplicates**: Clean duplicate entries in customer data.
2. **Filter Rows**: Exclude non-relevant transactions from sales data.
3. **Change Data Types**: Convert date columns to the correct format.

Merging, Appending, and Aggregating Data

- **Merging**: Join tables based on common columns.
- **Appending**: Stack tables on top of each other.
- **Aggregating**: Summarize data for analysis.

Examples:

1. **Merging**: Combine order and customer tables using customer ID.
2. **Appending**: Add sales data from different quarters.
3. **Aggregating**: Calculate total sales per region.

Hands-On: Creating a Dataflow for Machine Learning

1. **Create a Dataflow**: Connect to sales and customer tables.
2. **Clean Data**: Remove missing values and filter columns.
3. **Merge and Aggregate**: Prepare data for machine learning models.

Examples:

1. **Sales and Customer Data**: Prepare for churn prediction models.
2. **Product and Supplier Data**: Integrate for supply chain analysis.
3. **Employee and Performance Data**: Aggregate for HR analytics.

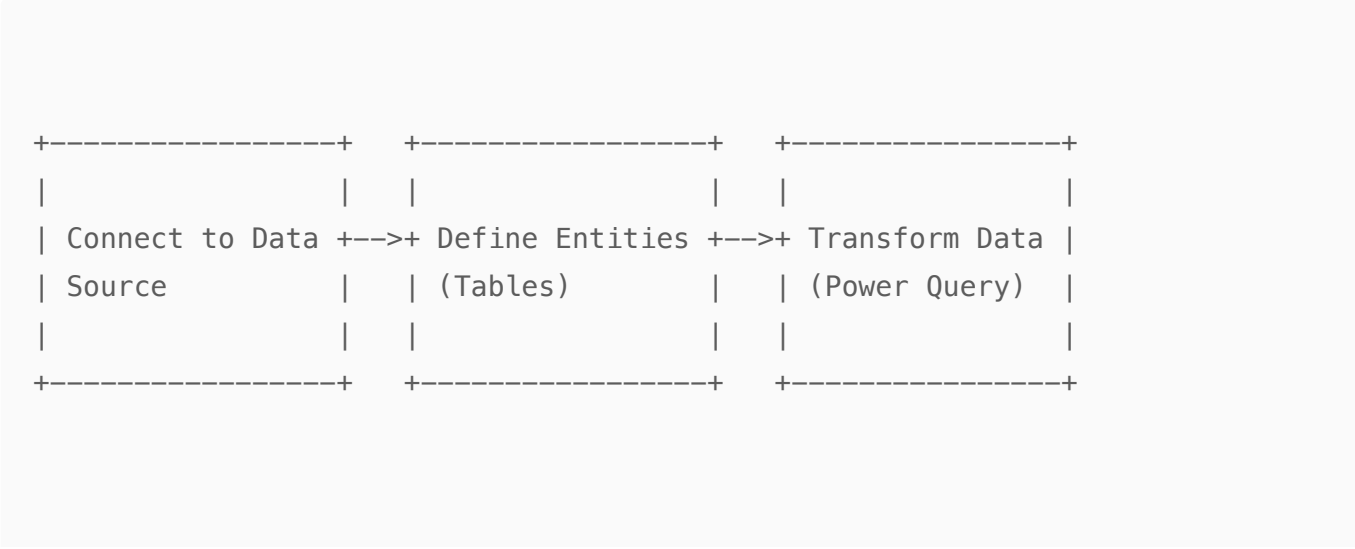
Conclusion

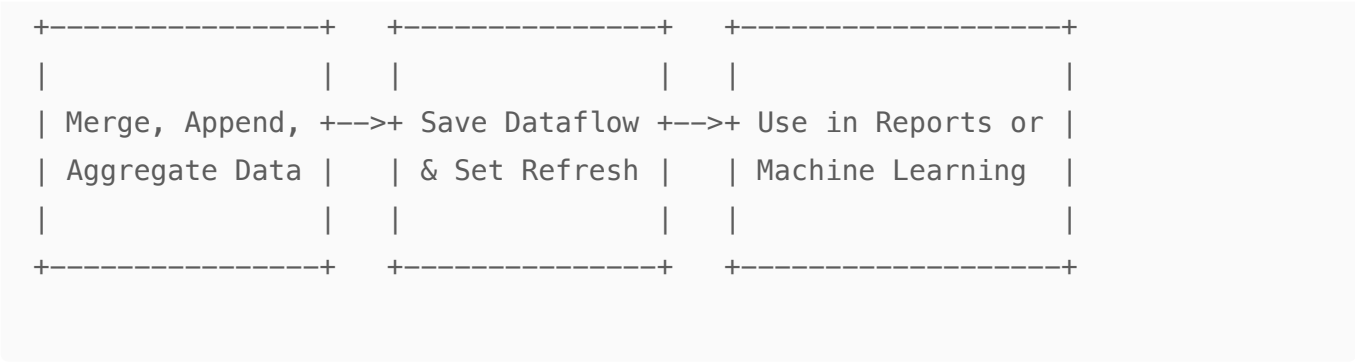
Dataflows in Power BI are essential for self-service ETL, enabling users to prepare data for reports, dashboards, and machine learning models.

Key Takeaways:

- Dataflows are for data preparation; datasets are for reporting.
- Connect to both cloud and on-premises sources.
- Use Power Query for data transformation.
- Merge, append, and aggregate data as needed.

Diagram: Dataflow Creation Process





Time Required to Read

This handout should take approximately **10-12 minutes** to read through, depending on your pace.

Feel free to reach out with any questions or for further clarification on any of the topics covered!