**Data Flows vs Wrangling Data Flows in Azure Data Factory**

**Part A – Mapping Data Flow Lab**

**Step 1 – Create Linked Services**

* In ADF → Manage → Linked Services.
* Create:
  + **Azure Blob Storage** (point to labdata).
  + **Azure SQL Database** (point to your Sales DB).

**Step 2 – Create Datasets**

* Blob CSV → OrdersDataset.
* Blob JSON → ProductsDataset.
* SQL Table → SalesFactDataset.

**Step 3 – Build Mapping Data Flow**

1. Go to **Author → Data Flows → New Mapping Data Flow**.
2. **Add Sources**:
   * Source 1: Orders (CSV).
   * Source 2: Products (JSON).
3. **Join Transformation**:
   * Join Orders and Products on ProductID.
   * Output: OrderID, ProductName, CustomerID, Quantity, Price, OrderDate.
4. **Derived Column Transformation**:
   * Add column SalesAmount = Quantity \* Price.
   * Rename Quantity → Qty.
5. **Select Transformation**:
   * Keep: OrderID, ProductName, CustomerID, Qty, SalesAmount, OrderDate.
6. **Sink Transformation**:
   * Write to SalesFact SQL table.
   * Set to “Truncate table” before insert (for testing).

**Step 4 – Debug & Validate**

* Use **Debug mode**.
* Preview output → ensure SalesAmount calculated correctly.
* Publish and Trigger pipeline.
* Query SQL DB:
* SELECT \* FROM SalesFact;

**Part B – Wrangling Data Flow Lab**

**Step 1 – Create Wrangling Data Flow**

1. Go to **Author → Data Flows → New Wrangling Data Flow**.
2. Choose **OrdersDataset** as the source.

**Step 2 – Clean and Shape Data (Power Query)**

* In the Power Query editor:
  1. Remove duplicates based on OrderID.
  2. Split OrderDate into Year and Month columns.
  3. Filter out rows where Quantity <= 0.
  4. Merge (join) with Customer SQL table on CustomerID.
  5. Rename columns for readability (CustomerName, Region).

**Step 3 – Sink**

* Set sink to a new SQL table: CleanedOrders.
* Schema:
* CREATE TABLE CleanedOrders (
* OrderID INT,
* CustomerID VARCHAR(20),
* CustomerName VARCHAR(50),
* Region VARCHAR(50),
* Qty INT,
* Year INT,
* Month INT
* );

**Step 4 – Debug & Validate**

* Run **Data Preview** in Power Query.
* Apply transformations → Publish.
* Trigger pipeline.
* Query SQL DB:
* SELECT \* FROM CleanedOrders;

**Lab Deliverables**

1. **SalesFact table** populated with enriched data using Mapping Data Flows.
2. **CleanedOrders table** created using Wrangling Data Flowsz