**Data Analyst Report – AdventureWorks Sales Performance**

**Objective**

This project delivers a comprehensive business intelligence workflow for analyzing sales performance using the AdventureWorks dataset provided by Microsoft. The focus is on data integrity, transformation, and deriving actionable business insights.

**Key Data Issues Identified**

* **Invalid Foreign Keys**: The CustomerKey and ResellerKey columns included placeholder values (-1) which did not correspond to any valid records in their respective dimension tables.
* **Missing Attributes**: For these placeholder records, essential fields such as city, country, and name were absent.
* **Improper Date Formats**: Date fields like OrderDateKey, DueDateKey, and ShipDateKey were stored as integers in YYYYMMDD format, making them incompatible with temporal filtering and aggregation.

**Data Cleaning Approach**

To preserve the completeness of the dataset while ensuring analytical integrity:

* Placeholder keys were replaced with dummy values: CustomerKey = 10999, ResellerKey = 0.
* Associated missing fields were populated with "NA" as a clear placeholder.
* Integer-based date fields were converted to proper date formats using SQL and Python.
* Cleaned tables were created separately for continued analysis, ensuring the original data remained intact for traceability.

All cleaning steps were implemented in:

* AdventureWorks\_Data\_Cleaning.ipynb (Python)
* 02\_AdventureWorks\_Cleaning.sql (SQL)

**Insights Derived from Sales Dashboard**

The following findings were extracted from the cleaned dataset and visualized in the accompanying dashboard:

**Sales Performance**

* Total sales amounted to **$42.9 million** for the year.
* Year-over-year sales growth was calculated at **40.6%**, indicating strong upward momentum.

**Product Breakdown**

* **Bikes** emerged as the highest-revenue product category, contributing over **83%** of total sales.
* Other notable categories included **Clothing**, **Accessories**, and **Components**.

**Trend Analysis**

* Monthly sales trends show a peak during Q3, especially in August, September, and October 2019.
* Sales demonstrated steady month-over-month growth with seasonal fluctuations.

**Reseller Insights**

* The top five resellers collectively generated over **$1.74 million**, with an average order value exceeding **$2,300**.
* Reseller sales were geographically concentrated in North America and Western Europe.

**Product Volume**

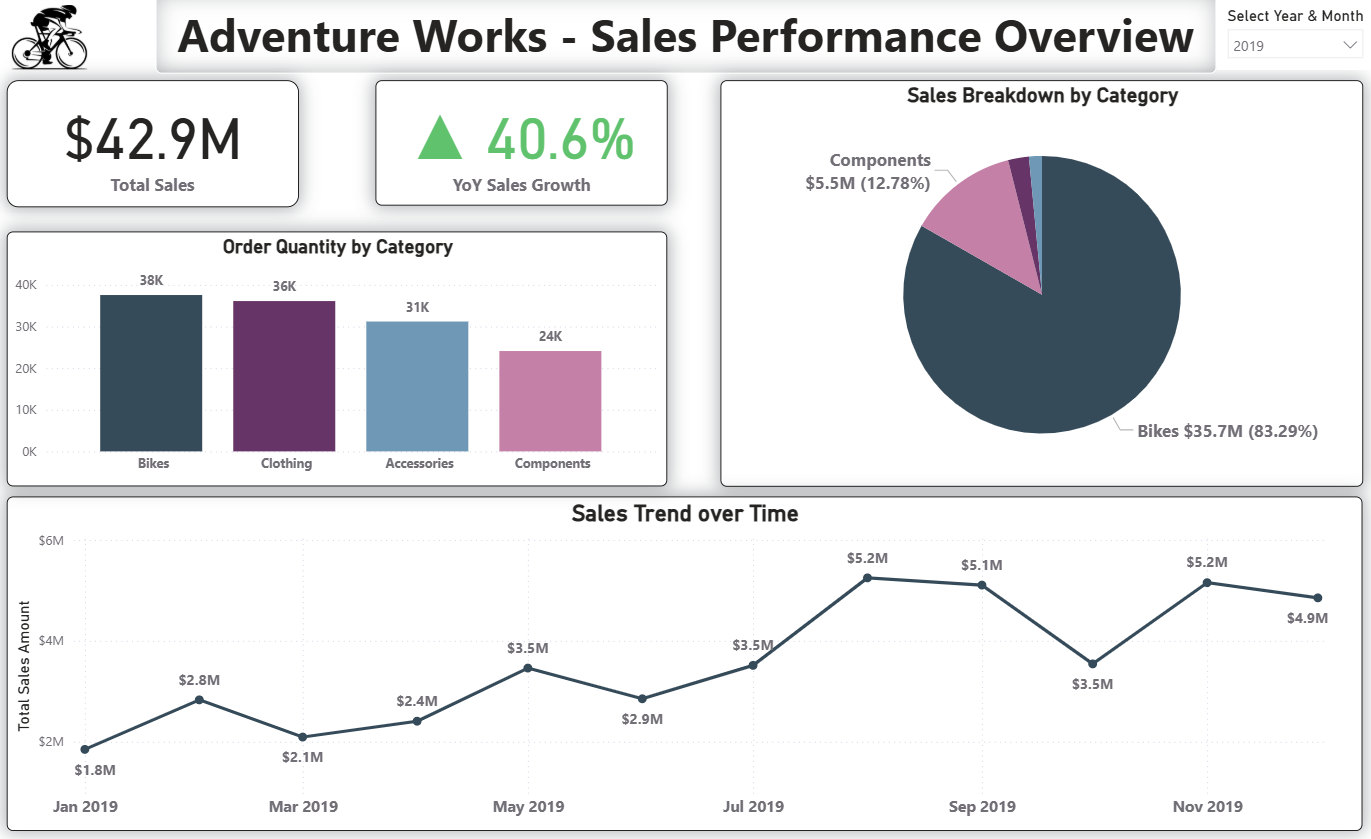
* Order quantity was highest for **Bikes** and **Clothing**, indicating strong customer demand across both high-ticket and mid-range products.
* Volume trends align with promotional or seasonal spikes seen during mid- to late-year.

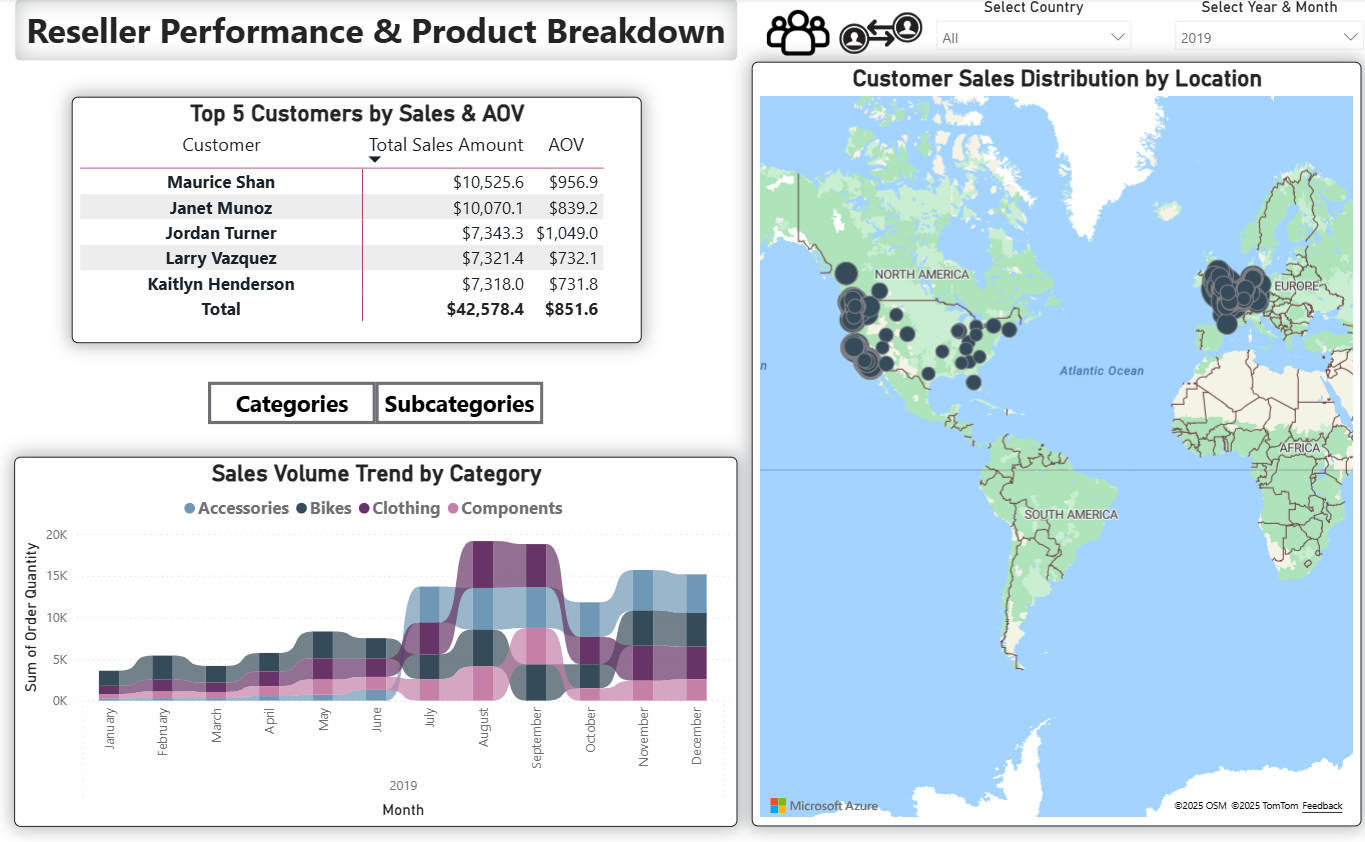
**Recommendations**

To prevent data quality issues in future pipelines:

* Collaborate with the **Sales team** to validate customer and reseller entries before ingestion.
* Engage the **IT or Data Engineering teams** to implement integrity constraints and validation checks at the database level.
* Establish data quality dashboards to monitor anomalies like invalid keys and missing critical attributes in real time.

**Source and Supporting Files**

* **Raw Dataset**: Microsoft AdventureWorks Sales Sample  
  [AdventureWorks Sales.xlsx](https://github.com/microsoft/powerbi-desktop-samples/blob/main/AdventureWorks%20Sales%20Sample/AdventureWorks%20Sales.xlsx?utm_source=chatgpt.com)
* **Cleaned Dataset**: AdventureWorks Sales\_Cleaned.xlsx
* **Dashboards**:
  + Dashboard\_Page1.png (Sales Overview 2019)
  + A screenshot of a computer screen

    AI-generated content may be incorrect.Dashboard\_Page2.png (Top 5 Resellers & Product Breakdown 2019)
  + Dashboard\_Page2.1.png (Top 5 Customers & Product Breakdown)