

## Case Study: Data-Driven Retail Sales Optimization

---

### Problem Statement

Adventure Works, a prominent bike manufacturer, also specializes in biking components, accessories, and apparel. The company operates through two primary sales channels: direct online retail and business-to-business (B2B) reseller sales. However, it is currently facing significant challenges in maintaining its competitive edge in specific geographic regions and customer segments.

Brian Welcker, Vice President of Sales, aims to address these challenges using a data-driven approach. However, the lack of a centralized and integrated data infrastructure has made it difficult to extract actionable insights. The company relies on disparate systems for inventory management, customer relationship management (CRM), and point-of-sale (POS) transactions, which results in fragmented and inconsistent data.

Key challenges include:

1. **Lack of Centralized Data** – Data is dispersed across multiple platforms (POS, CRM, eCommerce, and inventory systems), making it difficult to consolidate insights effectively.
2. **Inconsistent Reporting & Delayed Insights** – Different teams rely on varied reporting methods, leading to conflicting information and delays in decision-making.
3. **Sales & Revenue Analysis Challenges** – The company struggles to track sales trends, profitability, and product performance by location, affecting pricing, inventory, and promotional strategies.
4. **Limited Customer Insights** – There is a lack of comprehensive understanding of customer demographics, purchasing behavior, and loyalty trends, making it difficult to implement targeted marketing and retention strategies.

To overcome these issues, Adventure Works requires a centralized data solution that enhances reporting capabilities, provides deeper sales analysis, and delivers actionable customer insights. This solution will be implemented using SQL for data management, Python for data processing, and Power BI for data visualization.

---

### Solution Overview

The proposed solution involves implementing a **Data Warehouse** that integrates all business data sources (POS, CRM, eCommerce, and inventory management systems).

The solution will use **SQL** for database management, **Python** for ETL (Extract, Transform, Load) processes, and **Power BI** for interactive reporting and analytics.

### **Data Warehousing with SQL Server**

A **centralized Data Warehouse (DW)** will be created to consolidate all relevant business data. The warehouse will be structured using a **star schema**, facilitating efficient reporting and analysis. The **ETL process** will be automated using Python, and SQL will be utilized for querying the data warehouse.

### **Data Sources**

- **Point-of-Sale (POS) System** – Includes transaction details such as products sold, quantities, pricing, discounts, and payment methods.
- **Customer Relationship Management (CRM) System** – Captures customer demographics, purchase history, and loyalty program details.
- **eCommerce Platform** – Provides insights into online transactions, browsing behavior, abandoned carts, and customer interactions.
- **Inventory Management System** – Tracks stock levels, sales patterns, product restocks, and stock-out data.

### **ETL Process (Extract, Transform, Load)**

1. **Extract** – Data is collected from various business systems.
2. **Transform** – Python is used to clean, standardize, and format the data. This includes:
  - Standardizing product names.
  - Converting currencies.
  - Categorizing sales transactions.
3. **Load** – The transformed data is stored in a **SQL-based Data Warehouse** (e.g., MySQL, SQL Server).

### **Data Warehouse Schema (Star Schema)**

#### **Fact Tables:**

- **Fact Internet Sales** – Contains transaction records, sales amounts, quantities, discounts, and revenue.
- **Fact Reseller Sales** – Tracks product stock levels, restock quantities, and sales velocity.

#### **Dimension Tables:**

- **Customer Dimension** – Stores customer details such as name, age, location, and loyalty status.
  - **Product Dimension** – Includes product details, categories, and pricing information.
  - **Store Location Dimension** – Captures information about physical store locations.
  - **Reseller Information** – Reseller information such as Reseller Name, Business Type
- 

## ETL Process Implementation Using Python

To automate the ETL workflow, Python will be used in combination with libraries such as **Pandas** and **SQLAlchemy**. The steps include:

- Extracting data from multiple sources.
  - Transforming and cleaning the data to ensure consistency.
  - Loading the processed data into the SQL Data Warehouse.
- 

## Sales Analysis & Customer Insights Using Power BI

Once the data is centralized in the Data Warehouse, **Power BI** will be used to generate insightful visualizations and reports. These reports will help in decision-making by providing key insights, including:

1. **Identifying Underperforming Sales Territories**
  - A visual representation of **Total Sales and Average Order Value by Sales Territory** will be created.
  - This will help pinpoint regions where sales performance is below expectations and require strategic intervention.
2. **Customer Segmentation Based on Occupation & Income**
  - An analysis of **Average Sales per Customer by Occupation and Annual Income** will be conducted.
  - This will identify customer segments with below-average spending, allowing for targeted marketing strategies.
3. **Targeting the Right Age Group for Sales Growth**

- A visualization will be developed to help identify which **age group (e.g., 25-30, 30-35, etc.) under 50 years old** should be targeted for marketing efforts.
- The **Age column** will be incorporated into the data model to enable better segmentation and forecasting.