Case Study: Retail Sales & Customer Insights

# **Problem Statement**

Adventure Works is a bike manufacturer also dealing in components, accessories and clothing related to biking. It generates business through online (retail) and reseller (B2B) sales channels. Currently, the company is struggling with competitive pressures in online sales in some geographies, where it is losing out to the competition, and certain sections of the customer base. Brian Welcker, VP - Sales, needs to tackle these issues but with a data-driven approach. However, he is experiencing challenges in extracting actionable insights from the company’s disparate data sources. The company uses multiple systems for inventory management, customer relationship management (CRM), and point-of-sale (POS) transactions. However, these systems are not integrated, resulting in fragmented data that is difficult to analyze.

Specifically, Adventure Works faces the following issues:

* **Lack of a Centralized Data System:** Data is spread across multiple platforms such as POS, CRM, eCommerce, and inventory, making it difficult to gather insights from different sources simultaneously.
* **Inconsistent Reporting and Delayed Insights:** Employees rely on different reports, which often lead to conflicting information and delays in decision-making.
* **Sales and Revenue Analysis:** The company struggles to track sales trends, profitability, and performance by product or location, making it harder to optimize pricing, inventory, and promotions.
* **Customer Insights:** Adventure Works does not have a clear understanding of its customers' purchasing behavior, loyalty, or demographics, making targeted marketing and customer retention strategies challenging.

Adventure Works needs a solution to centralize its data, improve reporting, and analyze sales and gain customer insights to make informed business decisions. This solution will leverage SQL for data management, Python for data processing, and Power BI for visualization.

# Solution Overview

The solution involves implementing a **Data Warehouse** that consolidates data from all business systems (POS, CRM, eCommerce, and Inventory). Using **SQL** for data management, **Python** for ETL (Extract, Transform, Load) processes, and **PowerBI** for reporting and visualization, Adventure Works will be able to perform comprehensive sales and revenue analysis and gain deeper insights into customer behavior.

## Data Warehousing using SQL Server:

The Data Warehouse (DW) will act as a central repository for all business data, structured in a star schema to support reporting and analysis. The ETL pipeline will be automated using **Python**, and **SQL** will be used for querying the data warehouse.

### Data Sources:

* + **Point-of-Sale (POS) System:** Sales data such as products sold, quantities, prices, discounts, and payment methods.
  + **Customer Relationship Management (CRM) System:** Customer data including demographics, purchase history, and loyalty program information.
  + **eCommerce Platform:** Online transaction data, browsing behavior, abandoned carts, and customer interactions.
  + **Inventory Management System:** Stock data, sales patterns, product restocks, and stock-outs.

### ETL Process:

* + **Extract:** Data will be extracted from all relevant systems.
  + **Transform:** Using **Python**, data will be cleaned, standardized, and transformed into a common format. For example, product names will be standardized, currencies will be converted, and sales transactions will be categorized.
  + **Load:** Transformed data will be loaded into a **SQL-based Data Warehouse** (e.g., MySQL)

### Data Warehouse Schema (Star Schema):

* + Fact Tables:
    - **Sales Fact Table:** Records of transactions, including sales amount, quantities, discounts, and revenue.
    - **Inventory Fact Table:** Information about products in stock, restock quantities, and sales velocity.
  + Dimension Tables:
    - **Customer Dimension:** Customer details such as name, age, location, and loyalty status.
    - **Product Dimension:** Information about products, categories, and pricing.
    - **Time Dimension:** Date, week, month, and year.
    - **Store Location Dimension:** Information about physical store locations.

## ETL Process using Python:

Using **Python** and libraries like **Pandas** and **SQLAlchemy**, we can automate the ETL process.

### Extract Data from Multiple Sources

### Data Transformation

### Load Data into the Data Warehouse

## Sales Analysis and Customer Insights using Power BI:

Once the data is loaded into the data warehouse, the Power BI report can connect with it to extract the required insights. The following points pertain only to online sales of the company.

### Worst performing sales territories

Create a visual showing Total Sales and Avg Order Value by Sales Territory to help identify the where the company needs to improve penetration.

### Customer Segmentation based on Occupation and Annual Income

Show Avg Sales per Customer by Occupation and Annual Income of Customers so as to identify customer groups who are below average in Sales per Customer.

### Customer age-group to target for increasing sales

Create a visual to help Brian identify an age-group (25-30, 30-35, and so on) under the age of 50 to target. Add the Age column in the data model, not in Power Query.