AdventureWorks 2019 OLTP Analysis Report

# Introduction

AdventureWorks is a fictional bicycle and outdoor equipment company used by Microsoft for demonstrating database, business intelligence, and machine learning solutions. In this project, the AdventureWorks 2019 OLTP dataset was analyzed to uncover key insights into product performance, customer behavior, and regional sales patterns. The objective was to assist the company in making data-driven decisions to enhance sales strategies, target valuable customer segments, and improve forecasting accuracy.

# Data Overview

* Data consists of 71 table, and we used 25.
* Max number of columns is 43.
* Max number of records is 121,317.

# Problem Statement

AdventureWorks faces challenges in optimizing product sales, retaining customers, and analyzing regional sales performance. With a large product catalog and a diverse customer base, there is a need to identify underperforming products, revenue-driving customer segments, and sales trends across regions. Without clear visibility into these areas, strategic planning is hindered.

# Analysis Methodology

The project followed a structured pipeline including:

1. Data Cleaning – Fixed date formats and ensured numeric sales values.

* Duplicates: No duplicates in the data.
* Nulls: 11 table has nulls and was randomized using Python.

2. Exploratory Data Analysis – Conducted using SQL to understand product and customer patterns.

* Created ERD Diagram
* Concluded KPIs and solved it using SQL Functions like with, order by, group by, join.

3. Data Visualization – Created interactive dashboards in Tableau & Power Bi for key business metrics.

* Sales Dashboards
* Customer Analysis Dashboard
* Product Analysis Dashboard
* HR Dashboard
* Supply Chain Dashboard

4. Data Preprocessing – Analyze distribution of Monthly Sales, visualize distribution of numerical columns, handling outliers in monthly sales data (python)

5. Model Selection – ARIMA, Prophet, and Moving Average models were selected based on data properties.

6. Model Building – Models were trained and evaluated using an 80/20 train-test split.

# Key Insights

• Monthly sales data showed strong seasonality and increasing trends, with notable spikes in early 2014.

• Several products were identified as underperforming and could benefit from targeted marketing.

• Valuable customer segments were discovered based on order frequency and total sales.

• Regional sales analysis revealed imbalances that can inform targeted sales strategies.

# Sales Forecasting Solution

Three forecasting models were used:  
• Moving Average: Provides a simple baseline.  
• ARIMA (2,1,0): Captures trends and autocorrelations.  
• Prophet: Handles seasonality and outliers effectively.  
These models were used to generate 12-month forecasts to aid inventory planning and revenue goal setting.

# Recommendations

* Increase number of employees in shifts 2 and 3.
* Change your marketing strategy for the low 3 products.
* Do some discounts or marketing campaigns to lower the at-risk customer percentage.
* Check why Germany has the lowest total sales.

# Deliverables

• Cleaned Dataset (CSV format)

• Source Code (Python and SQL)

• Tableau & Power Bi Dashboards

• Project Documentation

# Team Members and Roles

• Manar Salem – SQL Analysis

• Aya Yasser – SQL Analysis

• Mostafa El-Taweel – Python Cleaning and Modeling

* Omnia Badway – Python Cleaning and Modeling

• Jana Mohamed – Tableau Visualization

• Laila Said – Tableau Visualization and Project Lead