Submitted Date  
January 12, 2025

**Project**

Bicycle Store Database

**Jawad Ahmad**

# Bicycle Store Data Warehouse Report

# **Problem Statement:**

The bicycle store faces challenges in managing and analyzing its sales, customer, and product data spread across different systems. This results in inefficiencies in tracking sales trends, identifying top-performing products, and optimizing customer engagement strategies.

# **Objective:**

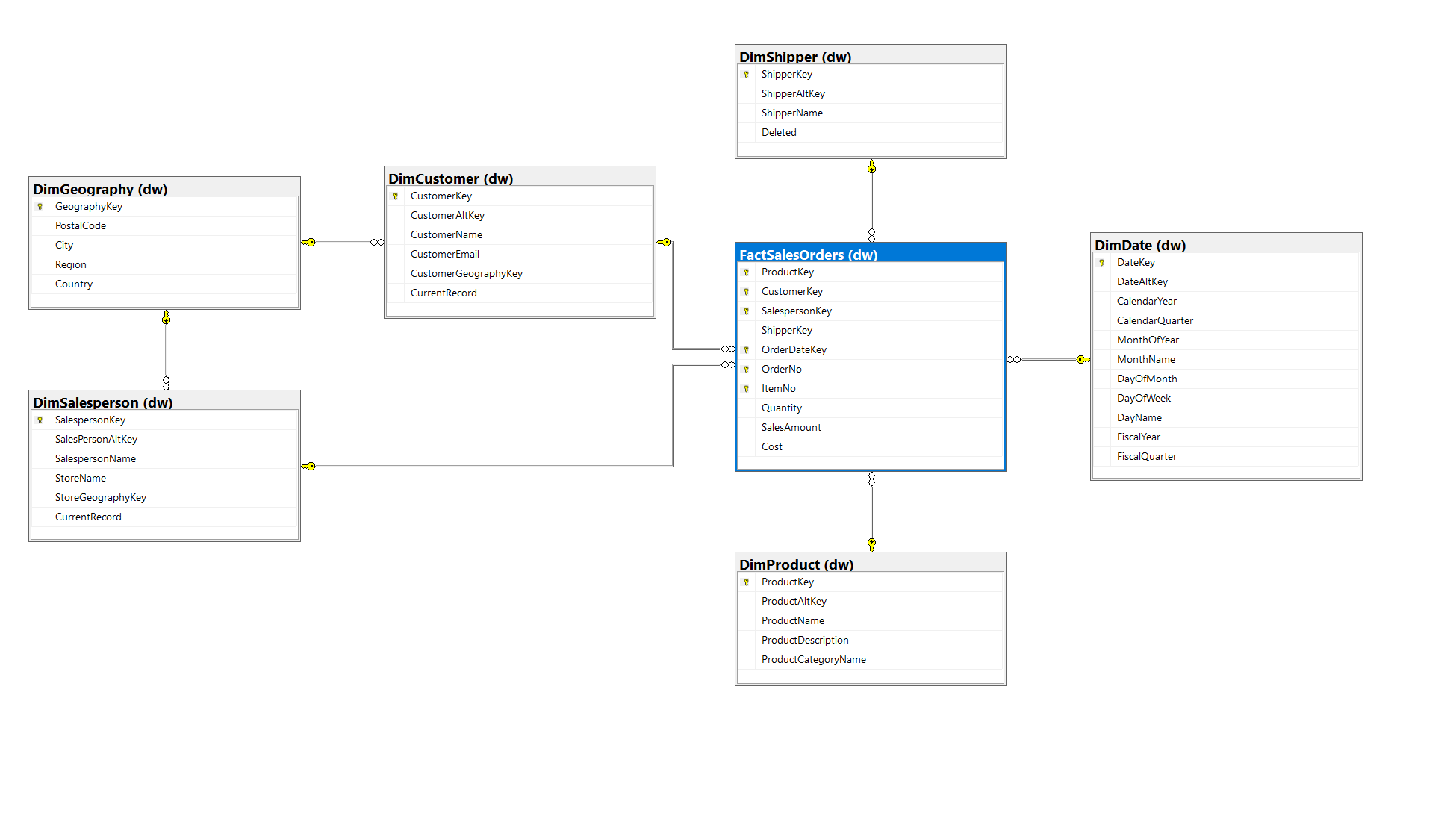
The goal of the data warehouse is to consolidate and integrate data into a unified platform, enabling better decision-making through:

* Analysis of sales trends over time.
* Identification of high-demand products and underperforming categories.
* Monitoring of salesperson performance and customer behavior.
* Geographic analysis of sales distribution.

**Schema Design:**

The data warehouse is designed using a star schema for efficient querying and reporting. The schema includes:

* **Fact Table:** FactSalesOrders (stores transaction-level data such as sales amount, quantity, and costs).
  + **Dimension Tables:**
* DimProduct: Contains product-related details (name, category, description).
* DimCustomer: Stores customer information (name, email, geography).
* DimGeography: Includes location details (city, region, country).
* DimSalesperson: Tracks salesperson performance.
* DimShipper: Details shipping providers.
* DimDate: Facilitates time-based analysis.

**ETL Design:**

The ETL (Extract, Transform, Load) process is designed as follows:

* Extract: Data is extracted from operational systems (e.g., sales records, product inventory, customer database).
* Transform: Data is cleaned, standardized, and transformed to align with the star schema structure:
* Standardizing product names and categories.
* Mapping customer locations to geography dimensions.
* Aggregating sales data for the fact table.
* Load: Transformed data is loaded into the corresponding fact and dimension tables

**SQL Scripts:**

Here are the key purposes for designing these queries to evaluate and enhance our Business Intelligence capabilities:

**Product Revenue Analysis:**

Identifies top-performing products and categories for inventory optimization and category expansion.

**Customer Value Analysis:**

Tracks customer spending for segmentation, targeted marketing, and engagement strategies.

**Salesperson Performance Dashboard:**

Evaluates individual and store sales performance for resource allocation and performance reviews.

**Shipping Performance Metrics:**

Analyzes shipping provider efficiency for cost optimization and better contract negotiations.

**Regional Sales Distribution:**

Maps regional sales to guide market expansion and allocate marketing budgets.

**Monthly Performance Dashboard:**

Tracks monthly trends for seasonal planning and performance monitoring.

**High-Value Customer Analysis:**

Identifies top customers for VIP retention and personalized marketing.

**Category Profit Analysis:**

Assesses category profitability to optimize product mix and pricing.

**Price Optimization Opportunities:**

Highlights high-volume, low-revenue products to adjust pricing strategies and promotions.

**Implementation Process:**

* **Star Schema Design:**

The schema follows a star schema approach, centralizing the transactional data in the FactSalesOrders table, linked to dimension tables for descriptive attributes.

* The fact table captures sales metrics (SalesAmount, Quantity, Cost) and connects to related dimensions (DimProduct, DimCustomer, DimDate, etc.) through foreign keys.
* **Data Integration and Transformation:**

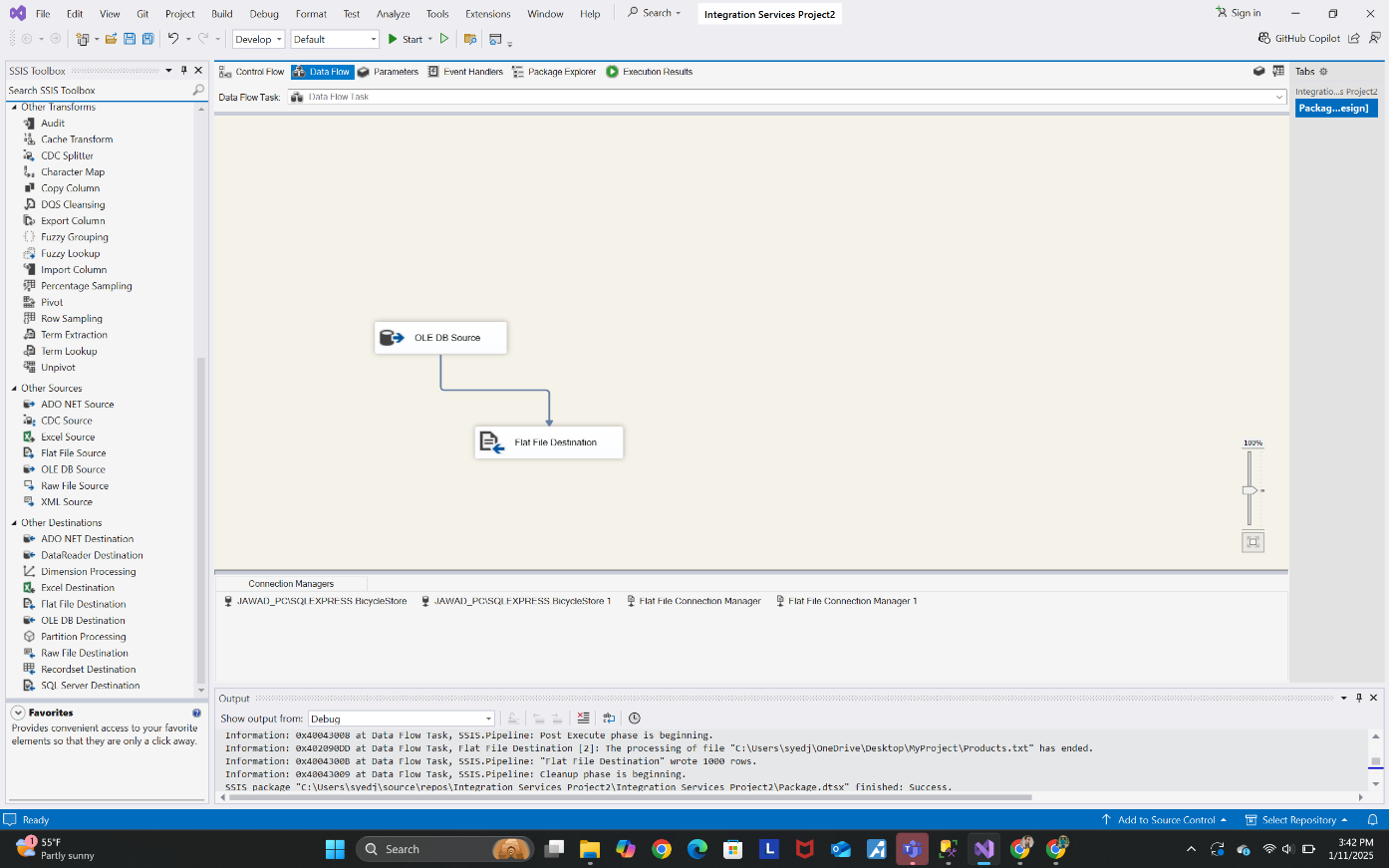
Data from disparate systems is extracted into a staging area, cleaned, deduplicated, and integrated for consistency. The transformed data is then loaded into the dw schema, populating the FactSalesOrders and dimension tables.

* **Optimization for Analytics:**
* **Temporal analysis** is supported by the DimDate table, which includes attributes such as fiscal year, month, and quarter.
* **Geographic insights** are enabled by linking DimGeography to customers and salespersons.
* **Customer data** is structured for segmentation and engagement strategies.

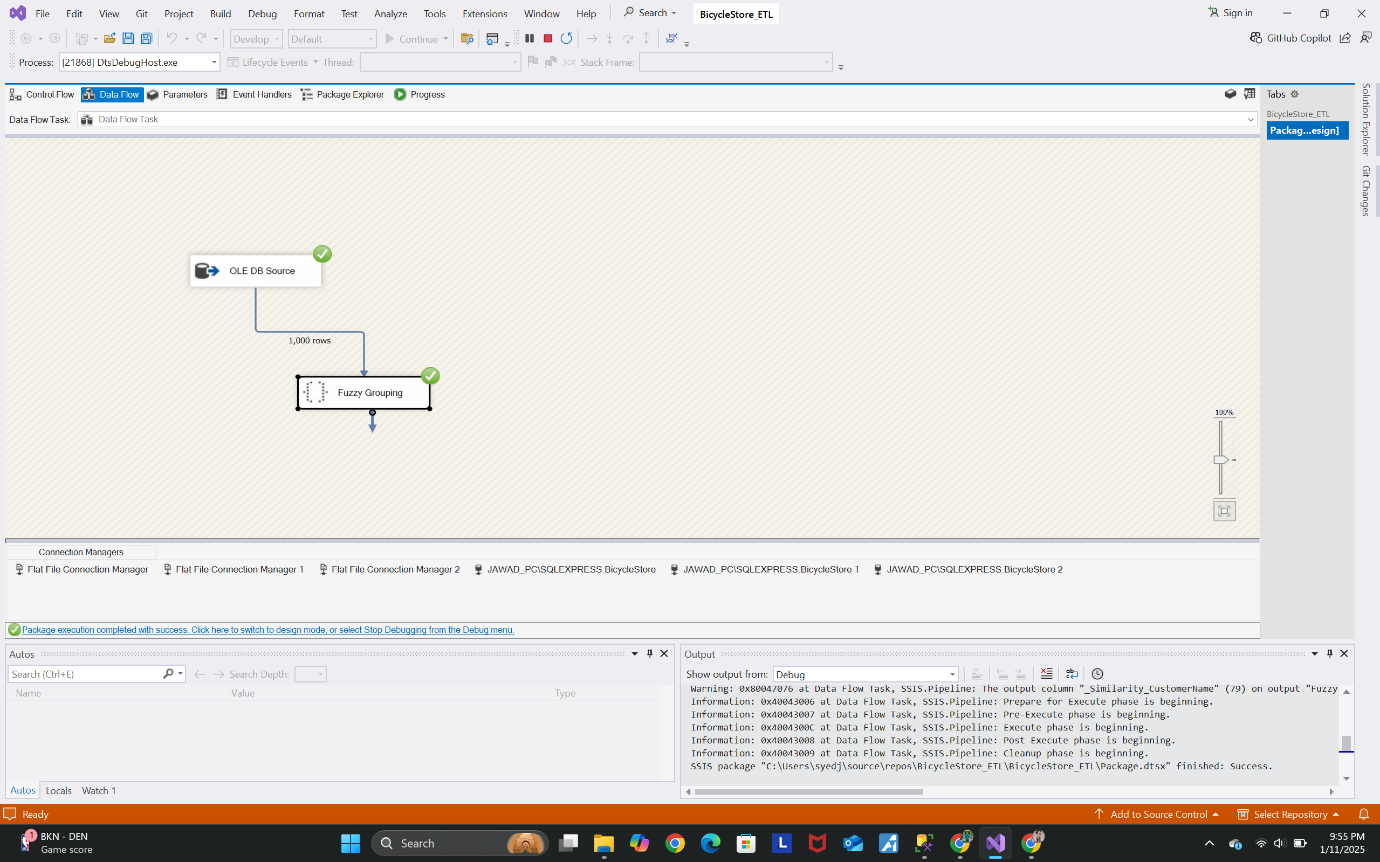
**SSIS TOOL:**

We perform ETL using SSIS tool.

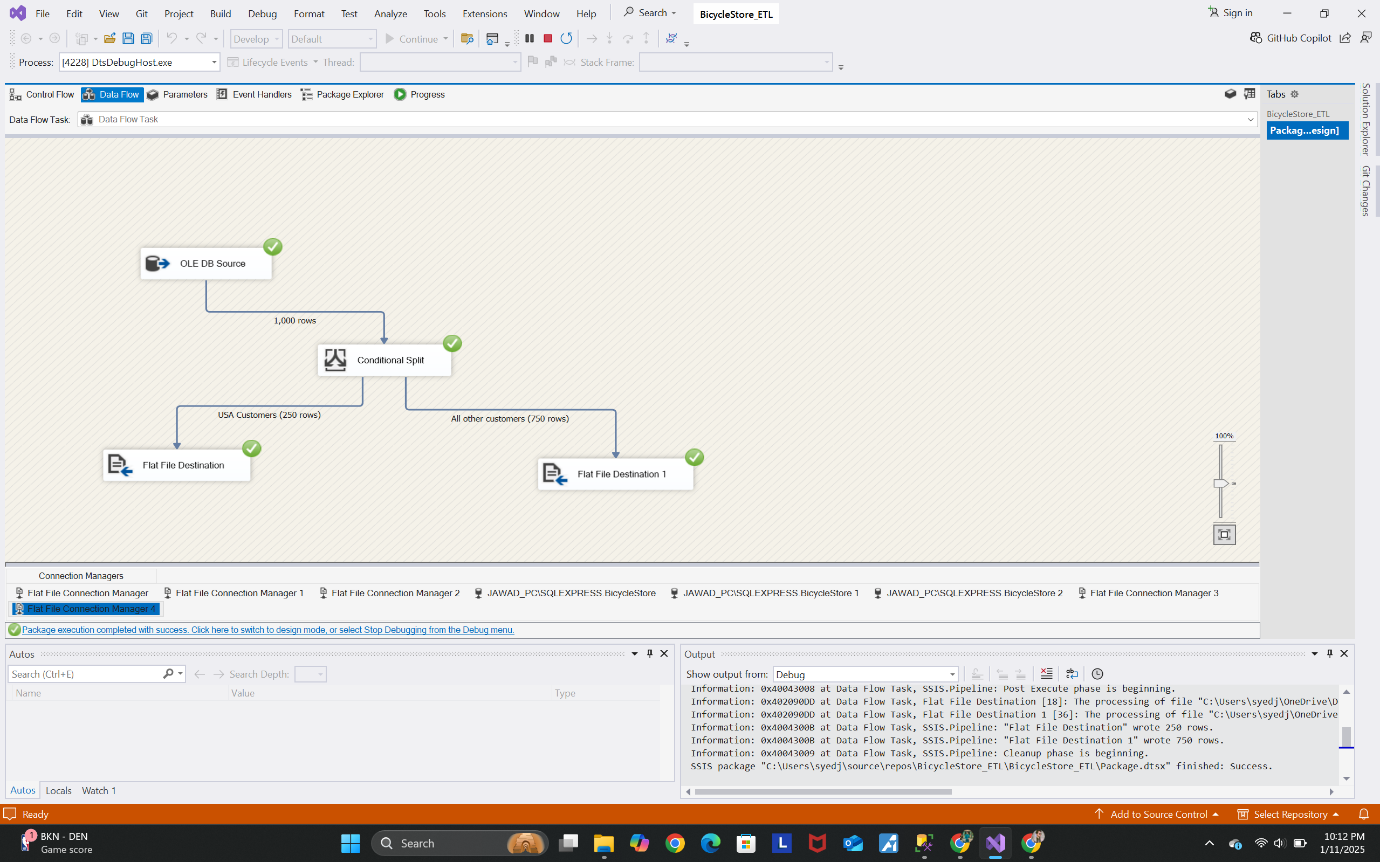
* **Data Flow Task**
* Simple OLE DB Source to Flat File Destination
* Basic data extraction pattern



* **Data Transformation**
* Uses Fuzzy Grouping for data cleansing/matching(applied to customer data mainly).
* Connected to BicycleStore database



* **Data Routing:**
* Implements Conditional Split for customer segregation
* Separates USA customers (250 rows) from others (750 rows)
* Routes to different flat file destinations



## **Insights Derived:**

* **Top Products:** Identified best-selling products and categories driving revenue.
* **Geographic Trends:** Highlighted regions with the highest sales, enabling targeted marketing campaigns.
* **Salesperson Performance:** Evaluated top-performing salespeople and identified areas for improvement.
* **Seasonal Trends:** Analyzed sales trends across months and quarters to optimize inventory and promotions.

**Visualization:**

We visualize the above insigts using power bi given following:

