**Superstore Sales Analysis Report**

**Sumaya Mohamed**

**June 2025**

**svmaya1@outlook.com**

**Introduction:**

This project presents a comprehensive analysis of the Superstore sales dataset using SQL. The goal is to demonstrate practical data querying and analytical skills by answering key business questions related to sales performance, customer behavior and shipping efficiency. Through structured queries, data aggregation and visualisation, this analysis highlights trends and insights that can inform strategic decisions. The project reflects common tasks performed by analysts in diverse industries, showcasing proficiency in SQL as a foundational data tool.

**Data Cleaning and Preparation:**

Before conducting the analysis, the raw data underwent essential cleaning and formatting steps to ensure accuracy and consistency:

* **Data Inspection:** Initial exploration involved checking table structure, data types and completeness, including identifying and handling missing values.
* **Currency Formatting:** Dollar signs were removed from the sales columns to enable numerical operations.
* **Date Standardisation:** Order Date and Ship Date formats were converted from MM/DD/YYYY (e.g., 01/05/2017) to the ISO standard YYYY-MM-DD (e.g., 2017-05-01) for easier date comparisons and querying.
* **Data Validation:** Additional verification steps ensured no invalid or null values remained in key columns used for analysis.

The cleaning operations were performed directly on the database using SQL update statements. The full set of data cleaning queries and procedures are available in the GitHub repository under the folder **data-cleaning/.**

*For detailed cleaning scripts and execution logic, visit the repository:*<https://github.com/sum427/sum427/blob/main/data-cleaning/data_cleaning_queries.sql>

#### **Data Cleaning and Preparation:**

The example below highlights only a few affected columns from the original dataset to demonstrate key cleaning steps. The full dataset includes additional fields not shown here.

**Before Cleaning:**

| **Order ID** | **Order Date** | **Ship Date** | **Sales** |
| --- | --- | --- | --- |
| CA-2015-140795 | 02/01/2015 | 02/03/2015 | $468.90 |
| CA-2015-104269 | 03/01/2015 | 03/06/2015 | $457.57 |

**After Cleaning:**

| **Order ID** | **Order Date** | **Ship Date** | **Sales** |
| --- | --- | --- | --- |
| CA-2015-140795 | 2015-01-02 | 2015-03-02 | 468.9 |
| CA-2015-104269 | 2015-01-03 | 2015-06-03 | 457.57 |

**Insight 1: Total Sales by Product Category**

1. **Business Question:**

*Which product categories generate the most sales?*

1. **SQL Query:**

SELECT Category,

ROUND(SUM(Sales), 2) AS total\_sales

FROM Raw\_superstore\_clean

GROUP BY Category

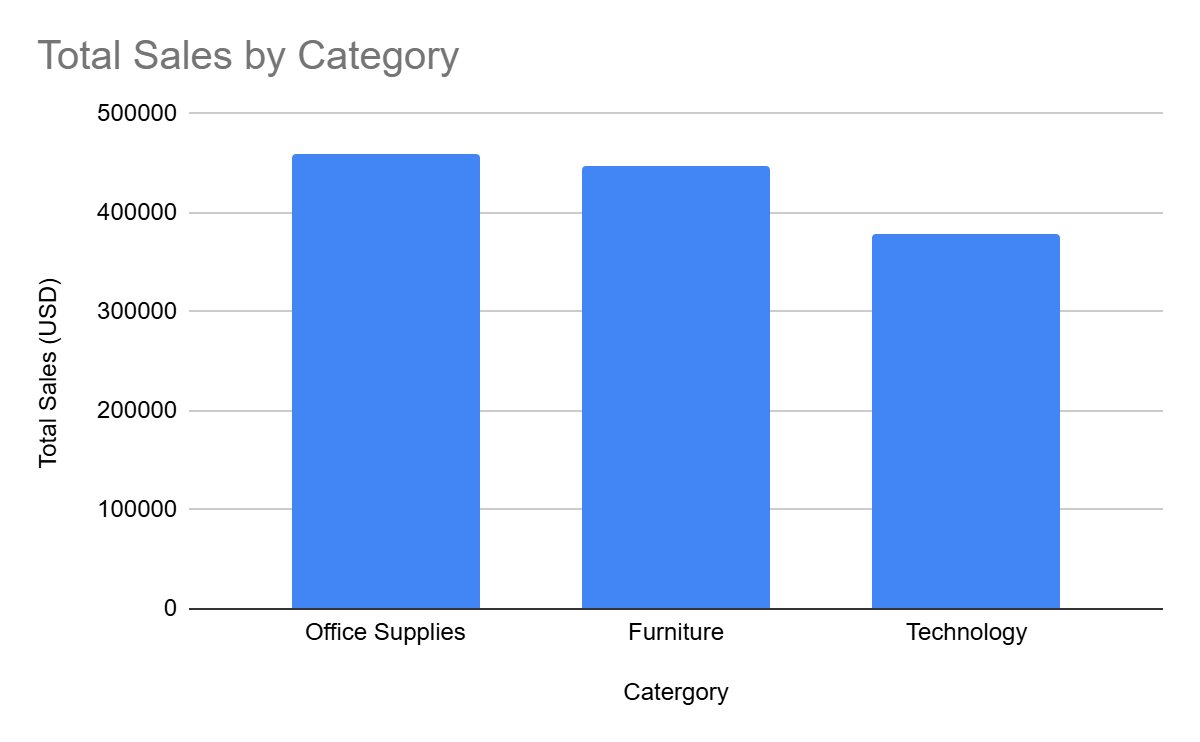
ORDER BY total\_sales DESC;

1. **Table Output:**

| **Category** | **Total Sales (USD)** |
| --- | --- |
| Office Supplies | 458615.44 |
| Furniture | 447365.65 |
| Technology | 378880.82 |

1. **Visualisation:**

**Figure 1:** Bar chart showing total sales (USD) by product category.



1. **Key Insights:**

* **Office Supplies** generate the **highest total sales** ($458,615.44), slightly ahead of Furniture and Technology.  
  Recommendation: Investigate which subcategories (e.g., paper, binders) drive this performance. Consider bundling or upselling related products to sustain momentum.
* **Furniture** follows closely in sales ($447,365.65) and may yield **higher margins per unit** due to pricing.  
  Recommendation: Prioritise **high margin furniture items** in promotions. Consider targeted marketing to corporate and home office buyers.
* Despite market potential, **Technology** sales ($378,880.82) are the **lowest** among the three categories.  
  Recommendation: Assess potential barriers, are tech items understocked, overpriced or poorly promoted? Consider bundling tech products with Office Supplies or offering training/support services to boost adoption.
* Sales are **evenly distributed**, with no single category overwhelmingly dominant.  
  Recommendation: Maintain category diversity in inventory and sales strategies but align marketing resources based on seasonality and demand elasticity in each category.

**Insight 2: Monthly sales Trend**

1. **Business Question:**

*How do total sales vary by month?*

1. **SQL Query:**

SELECT strftime('%Y-%m', "Order Date") AS order\_month,

ROUND(SUM(Sales),2) AS monthly\_sales

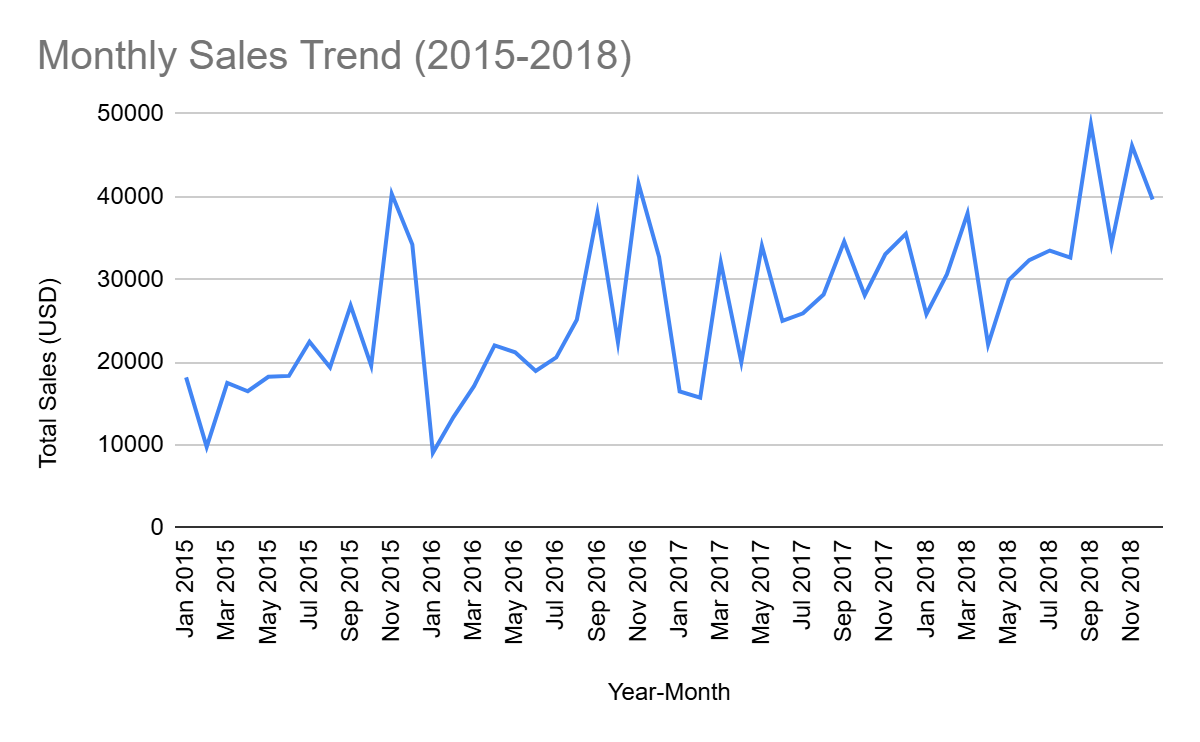
FROM Raw\_superstore\_clean

GROUP BY order\_month

ORDER BY order\_month;

1. **Visualisation:**

**Figure 2:** Line Graph showing Monthly Sales Trend (2015-2018)



1. **Key Insights:**

* **Seasonal Demand Planning**

Sales peak consistently in **November and December**.  
Recommendation: Increase inventory, marketing efforts and workforce capacity in Q4 to capture holiday driven demand. Consider launching targeted promotions in late Q3 to build momentum.

* **February Slowdowns**Insight: February is repeatedly the **lowest performing month**.  
  Recommendation: Use this period for internal operations (e.g., training, system upgrades). Consider introducing flash sales or loyalty perks to boost off season engagement.
* There is a clear **year over year upward trend**, particularly strong in **2018**.  
  Recommendation: Evaluate what drove 2018’s success (product mix, marketing, regional factors) and replicate those strategies. Consider expanding sales efforts or product lines in growth driving segments.
* **Promotional Timing**Insight: March and September also show notable sales spikes.  
  Recommendation: Consider these months for launching **quarterly campaigns**, product refreshes, or reengagement efforts.
* Overall Sales data show a clear **cyclical pattern**.

**Insight 3: Top 10 Customers by Sales**

1. **Business Question:**

*Which customers contribute the most to our sales revenue?*

1. **SQL Query:**

SELECT "Customer Name",

ROUND(SUM(Sales), 2) AS total\_sales

FROM Raw\_superstore\_clean

GROUP BY "Customer Name"

ORDER BY total\_sales DESC

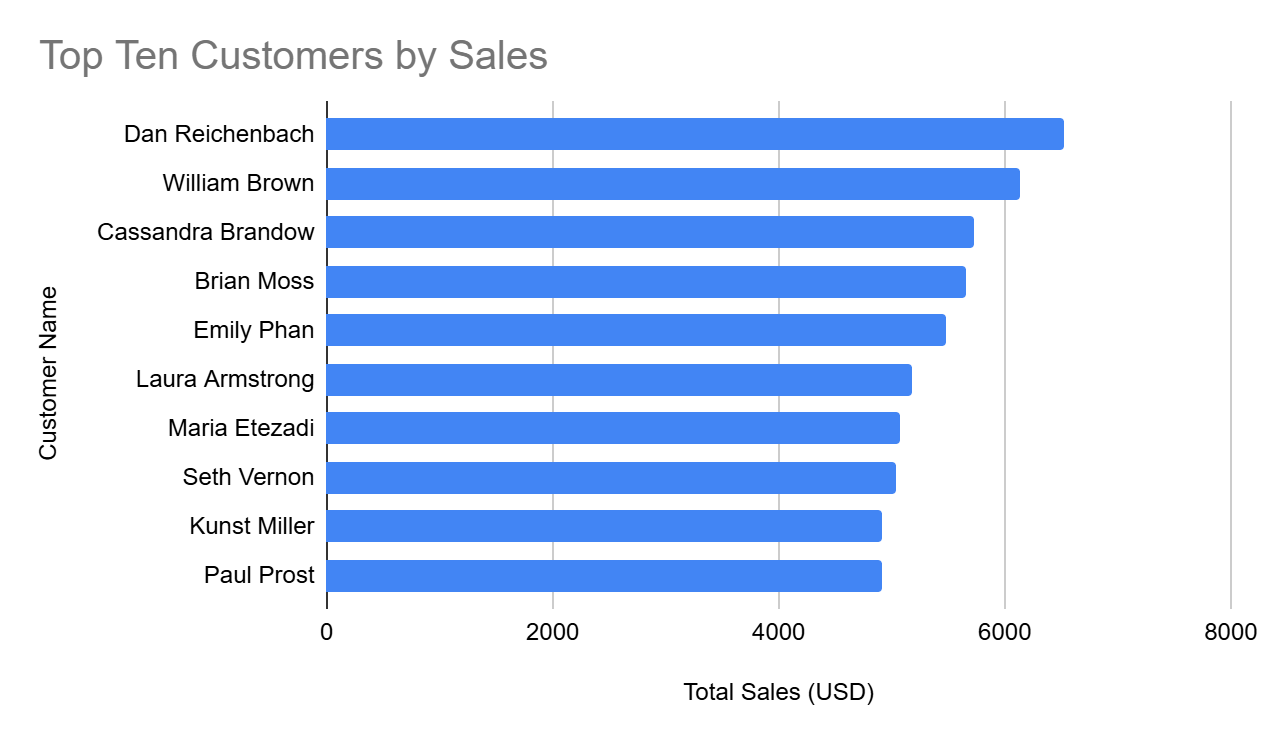
LIMIT 10;

1. **Table Output**

| **Customer Name** | **Total Sales (USD)** |
| --- | --- |
| Dan Reichenbach | 6528.03 |
| William Brown | 6129.71 |
| Cassandra Brandow | 5730.68 |
| Brian Moss | 5662.04 |
| Emily Phan | 5478.06 |
| Laura Armstrong | 5171.33 |
| Maria Etezadi | 5070.98 |
| Seth Vernon | 5033.74 |
| Kunst Miller | 4909.47 |
| Paul Prost | 4905.79 |

1. **Visualisation**

**Figure 3:** Top 10 Customers by Sales



1. **Key Insights:**

*Based on Figure 3, the distribution of total sales among the top 10 customers reveals a notable concentration of revenue in a few key accounts.*

* **Dan Reichenbach leads** as the top customer with $6,528.03 in total sales, closely followed by William Brown and Cassandra Brandow.
* The sales distribution is **relatively even** across the top 10 customers, with only a ~$1,600 difference between the highest and tenth ranked.
* This even spread suggests **a broad base of moderately high value customers**, reducing over reliance on a single client.
* Maintaining this balanced portfolio is beneficial for stability however **identifying potential for upselling** within this group could further boost revenue.
* Consider engaging these top customers with **personalised offers, feedback loops or loyalty incentives** to encourage continued spending and retention.

**Insight 4: Total Sales by Customer Segment**

1. **Business Question**

Which customer segment generates the highest sales revenue?

1. **SQL Query:**

SELECT Segment,

ROUND(SUM(Sales), 2) AS total\_sales

FROM Raw\_superstore\_clean

GROUP BY Segment

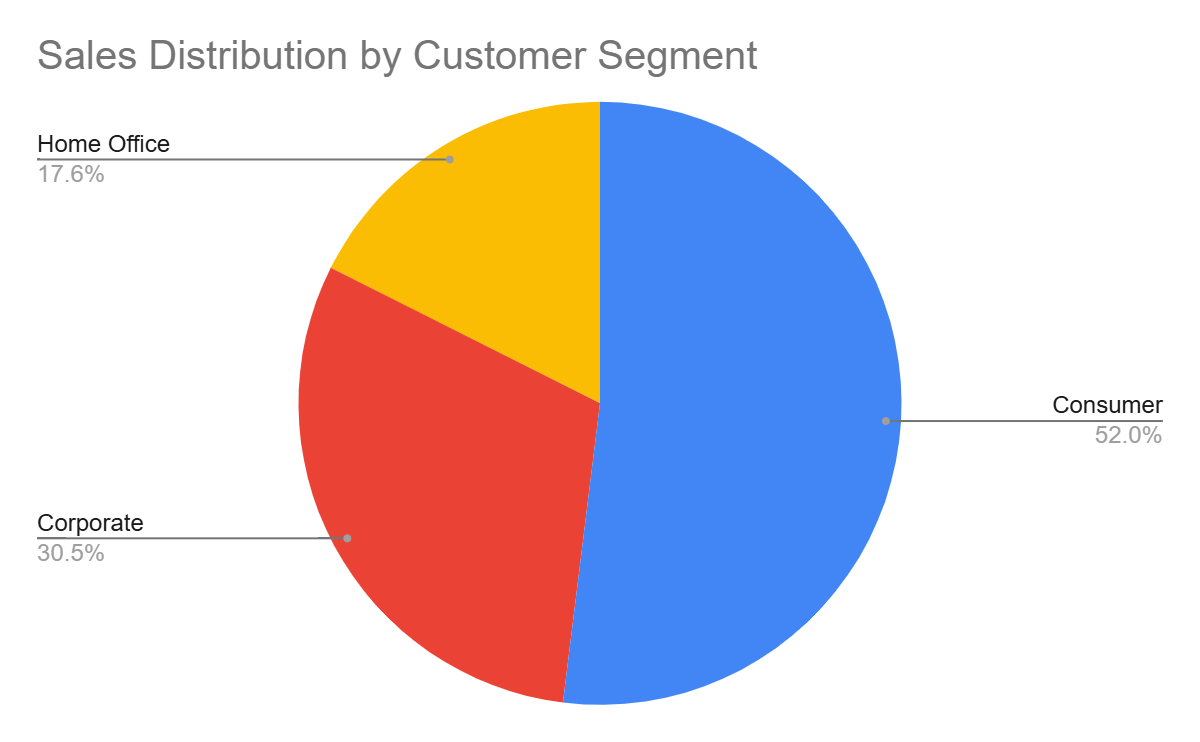
ORDER BY total\_sales DESC;

1. **Table Output:**

| **Segment** | **Total Sales (USD)** |
| --- | --- |
| Consumer | 667666.36 |
| Corporate | 391342.66 |
| Home Office | 225852.89 |

1. **Visualisation:**

**Figure 4:** Pie chart showing Sales distribution by Customer Segment (%)



1. **Key Insights:**

Refer to Figure 4 (Pie Chart), which illustrates the sales distribution across the three customer segments.

* **Consumer segment dominates**, contributing approximately **52%** of total sales ($667,666.36), indicating the business has strong traction in the B2C space.
* **Corporate customers** make up about **30%** of sales ($391,342.66), highlighting the importance of maintaining relationships in the B2B space as well.
* **Home Office segment** accounts for roughly **18%** ($225,852.89), the smallest share but still represents a **significant niche** with potential for tailored outreach.
* The segment breakdown suggests a **diversified customer base** which helps mitigate risk and offers room for targeted growth strategies.

**Insight 5: Shipping Delay Analysis**

1. **Business Question**

*What is the average shipping delay for different ship modes, and how does delivery speed vary?*

1. **SQL Query**

SELECT "Ship Mode", ROUND(AVG(julianday("Ship Date") - julianday("Order

Date")),0) AS AV\_SHIPPING\_DAYS

FROM Raw\_superstore\_clean

WHERE "Order Date" IS NOT NULL AND "Ship Date" IS NOT NULL

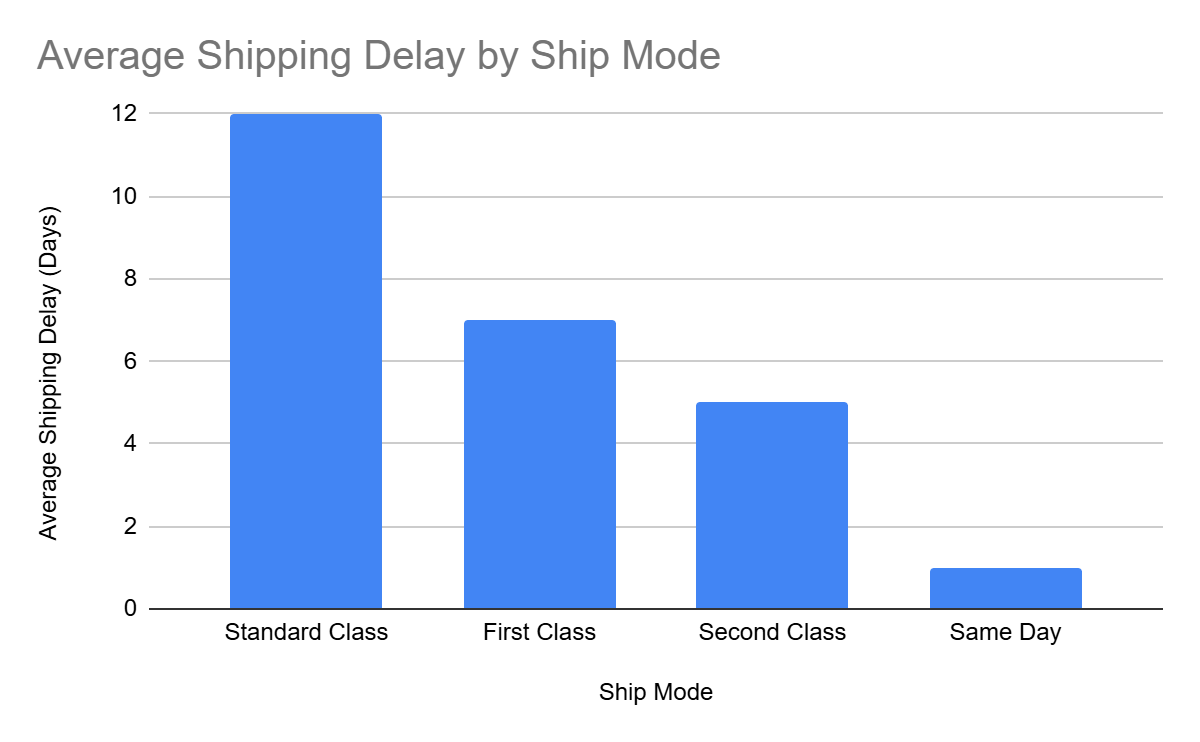
GROUP BY "Ship Mode" ORDER BY AV\_SHIPPING\_DAYS DESC;

1. **Table Output**

| **Ship Mode** | **Average Shipping Delay (Days)** |
| --- | --- |
| Standard Class | 12 |
| First Class | 7 |
| Second Class | 5 |
| Same Day | 1 |

1. **Visualisation:**

**Figure 5:** Bar chart showing Average Shipping by Ship Mode (Days)



1. **Key Insights:**

Refer to Figure 5: Bar chart showing average shipping time by mode (in days).

* **Standard Class shipping** has the longest average delay at **12 days**, potentially impacting customer satisfaction for standard deliveries.
* **First Class and Second Class** offer faster delivery, averaging **7 and 5 days** respectively, making them preferable options for time sensitive orders.
* **Same Day shipping** performs optimally with an average delay of just **1 day**, aligning with high customer service expectations.
* The clear variance in delays suggests opportunities to **encourage premium shipping options** or **improve logistics for Standard Class** to boost overall delivery performance.

**Conclusion:**

* **Product Category Focus**: Office Supplies and Furniture lead in total sales but Furniture likely has a higher **margin per unit potential**, warranting **promotional focus** or **targeted upselling** to maximise profitability.
* **Sales Trends Over Time**: There is a **clear upward sales trend** from 2015 to 2018, with **consistent seasonal peaks** in **November and December,** a period likely influenced by holiday and end of year demand. February consistently marks the **lowest sales**, suggesting opportunities for seasonal promotions to boost performance.
* **Customer Concentration**: Revenue is **heavily concentrated among a small group of high value customers**. Maintaining relationships with top clients like Sean Miller and Tamara Chand is critical. There is potential to **implement loyalty initiatives or tailored campaigns** to retain and grow this base.
* **Customer Segments**: The **Consumer segment dominates sales (over 50%)**, highlighting the importance of **B2C strategies**. However, Corporate and Home Office clients still account for a **significant share (combined ~49%)**, suggesting B2B opportunities should not be neglected.
* **Shipping Efficiency**: **Standard Class shipping** has the highest average delay (12 days) which may negatively impact customer satisfaction. **Same Day and First Class shipping** significantly outperform in speed, presenting an opportunity to **optimise delivery logistics or upsell faster shipping tiers** for key customers.

**Limitations and Future Work**

**Limitations:**

* The dataset only covers sales data up to December 2018, so insights may not reflect more recent trends or impacts (e.g., COVID-19).
* Some columns had missing or inconsistent data that required cleaning, which might have introduced minor inaccuracies.
* The analysis focused mainly on sales and shipping data; other dimensions such as customer demographics or product returns were not explored.

**Future Work:**

* Expand analysis to include customer segmentation and product level profitability.
* Integrate additional external data sources (e.g., market trends, competitor data) to enrich insights.
* Automate data cleaning and reporting using Python or workflow tools to improve efficiency and repeatability.
* Build interactive dashboards (e.g., in Power BI) to allow stakeholders dynamic exploration of sales trends.

**Appendix:**

**A1. Monthly Sales Data Table**

| **Year-Month** | **Total Sales (USD)** |
| --- | --- |
| Jan 2015 | 18145.55 |
| Feb 2015 | 9746.59 |
| Mar 2015 | 17486.38 |
| Apr 2015 | 16467.3 |
| May 2015 | 18210.55 |
| Jun 2015 | 18322.58 |
| Jul 2015 | 22449.53 |
| Aug 2015 | 19360.64 |
| Sep 2015 | 26832.46 |
| Oct 2015 | 19568.1 |
| Nov 2015 | 40329.03 |
| Dec 2015 | 34209.83 |
| Jan 2016 | 9026.17 |
| Feb 2016 | 13344.2 |
| Mar 2016 | 17083.19 |
| Apr 2016 | 22003.36 |
| May 2016 | 21179.98 |
| Jun 2016 | 18921.38 |
| Jul 2016 | 20553.67 |
| Aug 2016 | 25116.16 |
| Sep 2016 | 38006.13 |
| Oct 2016 | 22393.36 |
| Nov 2016 | 41546.96 |
| Dec 2016 | 32697.84 |
| Jan 2017 | 16447.71 |
| Feb 2017 | 15687.64 |
| Mar 2017 | 32035.26 |
| Apr 2017 | 19984.32 |
| May 2017 | 34055.02 |
| Jun 2017 | 24964.7 |
| Jul 2017 | 25894.48 |
| Aug 2017 | 28156.94 |
| Sep 2017 | 34567.86 |
| Oct 2017 | 28078.94 |
| Nov 2017 | 33020.53 |
| Dec 2017 | 35494.46 |
| Jan 2018 | 25778.47 |
| Feb 2018 | 30610.06 |
| Mar 2018 | 37969.11 |
| Apr 2018 | 22103.28 |
| May 2018 | 29898.89 |
| Jun 2018 | 32316.71 |
| Jul 2018 | 33476.36 |
| Aug 2018 | 32613.16 |
| Sep 2018 | 48714.65 |
| Oct 2018 | 34217.07 |
| Nov 2018 | 46128.19 |
| Dec 2018 | 39647.16 |

### **A2. GitHub Repository**

All project files including raw data, cleaning scripts, SQL queries and reports are available here:  
<https://github.com/sum427/sum427>

### **A3. Technical Details**

* Database platform: SQLite
* Data cleaning: SQL UPDATE statements for currency formatting and date standardisation
* Data size: ~9,800 records × 21 columns
* Visualisations: Created in Excel based on query outputs
* Query execution: Optimised for performance on dataset size