# **Project Report:** Utilizing Spreadsheets for Data Analytics

## **Introduction**

The objective of this report is to provide a comprehensive analysis of the sales data from the AdventureWorks dataset, with a focus on uncovering valuable insights that can drive business decisions and strategies. AdventureWorks is a leading company in the sales industry, and understanding the patterns and trends in their sales data is crucial for maintaining and improving their market position.

I will follow the six phases of the Structured Approach to Data Analytics to ensure a systematic and thorough examination of the data. These phases include Ask, Prepare, Process, Analyze, Share, and Act. Each phase builds upon the previous one, leading to actionable insights that can significantly benefit the business.

In the "**Ask**" phase, I will define the key questions that will guide my analysis. These questions are designed to address critical aspects of the sales data, including monthly sales trends, online versus non-online sales, delivery performance, salesperson performance, geographic sales comparisons, top sales reasons, and potential correlations with external factors.

## **1. Ask**

In this phase, I will outline the specific questions that my analysis aims to answer. These questions are formulated to provide rich insights into the sales data and to address the business needs effectively. The questions are as follows:

### What is the overall monthly sales trend (rise and falls) over the specified period, and is there seasonality?

* + This question aims to identify the general trends in sales over time, highlighting any seasonal patterns that could influence business strategies.

### How do the sales trends compare between online and non-online orders over time?

* + By comparing the sales trends of online and non-online orders, I can understand the different dynamics and growth patterns in these two sales channels.

### How does delivery performance vary across different regions or sales territories?

* + Evaluating delivery performance by region or sales territory helps in identifying areas that might need improvement and ensuring customer satisfaction across all regions.

### Who are the top-performing salespersons by total sales?

* + Identifying the top-performing salespersons provides insights into successful sales strategies and helps in recognizing and rewarding high achievers.

### How do sales compare across different regions?

* + Comparing sales across different regions allows for a better understanding of market performance and potential opportunities for growth in various geographic areas.

### What are the top reasons influencing sales?

* + Analyzing the top reasons influencing sales helps in understanding customer motivations and tailoring marketing strategies to address these factors effectively.

### How does the day of the week an order was placed affect sales both online and offline?

* + Investigating correlations with day of the week an order was placed can reveal additional insights and help in aligning production and sales strategies with consumer demands.

These questions will guide the subsequent phases of my analysis, ensuring that I gather and process the relevant data to uncover actionable insights for AdventureWorks. The findings from this analysis will provide a solid foundation for making informed business decisions and optimizing sales strategies.

## **2. Prepare**

In this phase, I examined the provided datasets to understand their structure and identify any issues that need to be addressed in the Process phase. The data for this project includes four different tables: AdventureWorks.SalesOrderHeader, AdventureWorks.SalesReason, AdventureWorks.Employee, and AdventureWorks.SalesTerritory.

Here is a summary of the datasets:

* **AdventureWorks.SalesOrderHeader:** This table contains 31,465 rows and 15 columns.
* **AdventureWorks.SalesReason:** This table contains 10 rows and 4 columns.
* **AdventureWorks.Employee:** This table contains 290 rows and 6 columns.
* **AdventureWorks.SalesTerritory:** This table contains 10 rows and 4 columns.

Issues Identified:

* **Redundant Columns:** Some columns in the datasets are not necessary for the analysis and can be removed to improve performance and clarity.
* **Missing Values:** There are missing values in some columns, which need to be handled appropriately to ensure accurate analysis.
* **Incorrect Data Types:** Some columns may have incorrect data types that need to be corrected for proper analysis.

In the next phase, I will address these issues by cleaning and processing the data to make it suitable for analysis. This includes removing redundant columns, filling or handling missing values, and correcting data types. This preparation will ensure that the data is accurate and ready for the analysis phase.

## **3. Process**

In the Process phase, I focused on cleaning and preparing the data for analysis. The majority of the data cleaning was carried out on the AdventureWorks.SalesOrderHeader dataset. Here are the specific steps I took:

### Date Columns Conversion

* + I converted all date columns across the four datasets to the date data type for simplicity and clarity, as they were initially in date-time format with the time set to 00:00:00.

### AdventureWorks.SalesOrderHeader Data Cleaning

* + Retained "SalesPersonID" Column: I recognized the importance of the "SalesPersonID" column for identifying top-performing salespersons and decided to retain it despite the high percentage of missing values (87.9%). I filled the missing values with a placeholder "No SalesPerson" to retain the column without misleading data.
  + Handled Missing Values in "SalesReasonID" Column: The "SalesReasonID" column had 26.9% missing values. I created a pivot table to get unique values and their distributions. I used “-1” to replace the missing values in the "SalesReasonID" column and gave it a category name of “Unknown”. This will further help me in my analysis to better understand how sales reason influences sales.
  + Added "SalesReason" Column: I used VLOOKUP to fetch the actual category names from the AdventureWorks.SalesReason table and created a new column in the AdventureWorks.SalesOrderHeader table named 'SalesReason'.
  + Added "Country Name" Column: I used VLOOKUP to fetch the Country name from the AdventureWorks.SalesTerritory table using the "TerritoryID" column in the AdventureWorks.SalesOrderHeader table as the search\_key.

### Calculated Shipping Duration

* + I calculated the number of days taken to ship an item after the order is placed using the DATEDIFF function, passing the OrderDate and ShipDate as arguments to the function.

### Linked SalesPersonID to Employee Details

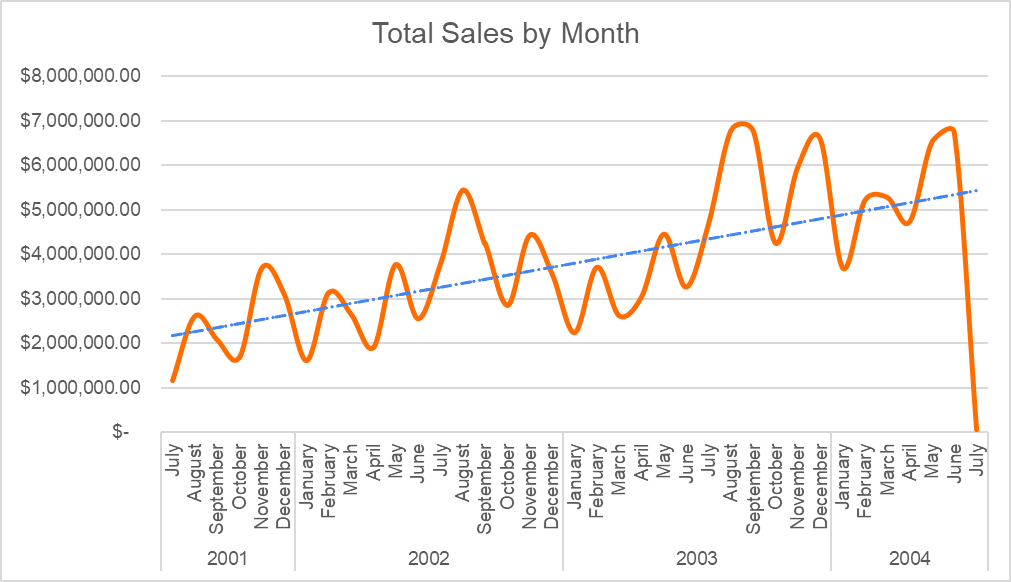
* + I used the JOIN and TRIM functions to merge the Title, FirstName, MiddleName, and LastName columns in the Employee table into a FullName column.
  + I created a SalesPerson column in the AdventureWorks.SalesOrderHeader table using the FullName column from the Employee table. I then used the IFERROR and VLOOKUP functions to lookup and attach the name of employees (sales person) each order using the salespersonID and EmployeeID.

By performing these steps, I ensured that the data was clean, accurate, and ready for the analysis phase. This preparation allows for more reliable and meaningful insights in the subsequent stages of the project.

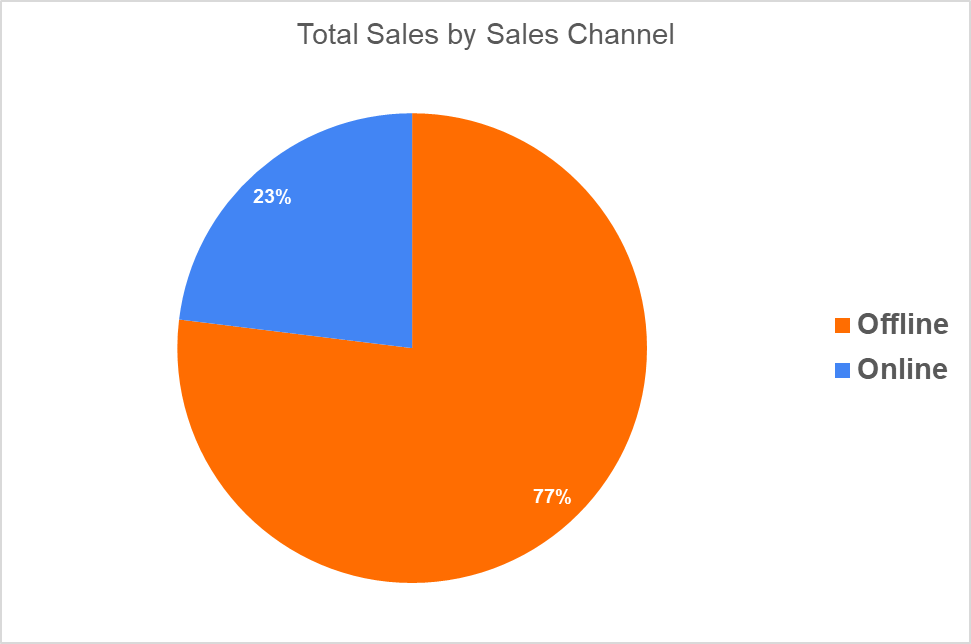
## **4. Analyze**

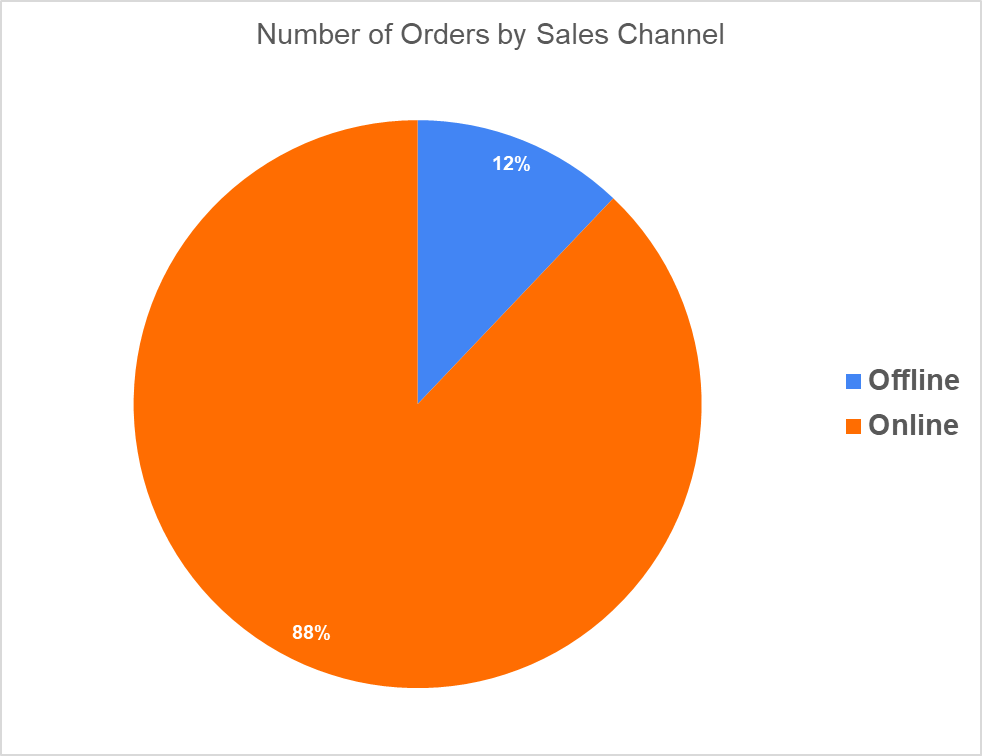
In the Analyze phase, I performed the analysis to answer the questions and gain meaningful insights that drive business decisions. Here’s a detailed breakdown of my analysis for each question:

1. **Overall Monthly Sales Trend (Rise and Falls) Over Time:** To examine the overall monthly sales trend over time, I started by creating two new columns from the 'OrderDate' column: 'OrderMonth' and 'OrderYear'. This allowed me to properly visualize sales by month for each year. I then created a pivot table to aggregate sales data by these new columns and used the pivot table to generate a line plot, which effectively illustrated the rise and fall in sales over time.

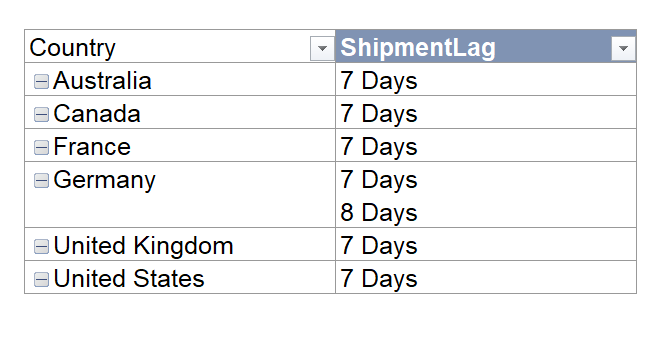


1. **Sales Trends Comparison Between Online and Non-Online Orders:** Next, to compare sales trends between online and non-online orders, I created a new column called 'SalesChannel' using the values from the 'OnlineOrderFlag' column, where 1 represents online orders and 0 represents offline orders. I then created a pivot table to summarize the total number of orders and total sales by sales channel. To visualize this data, I created two pie charts—one showing the distribution of orders and the other showing the distribution of sales revenue.

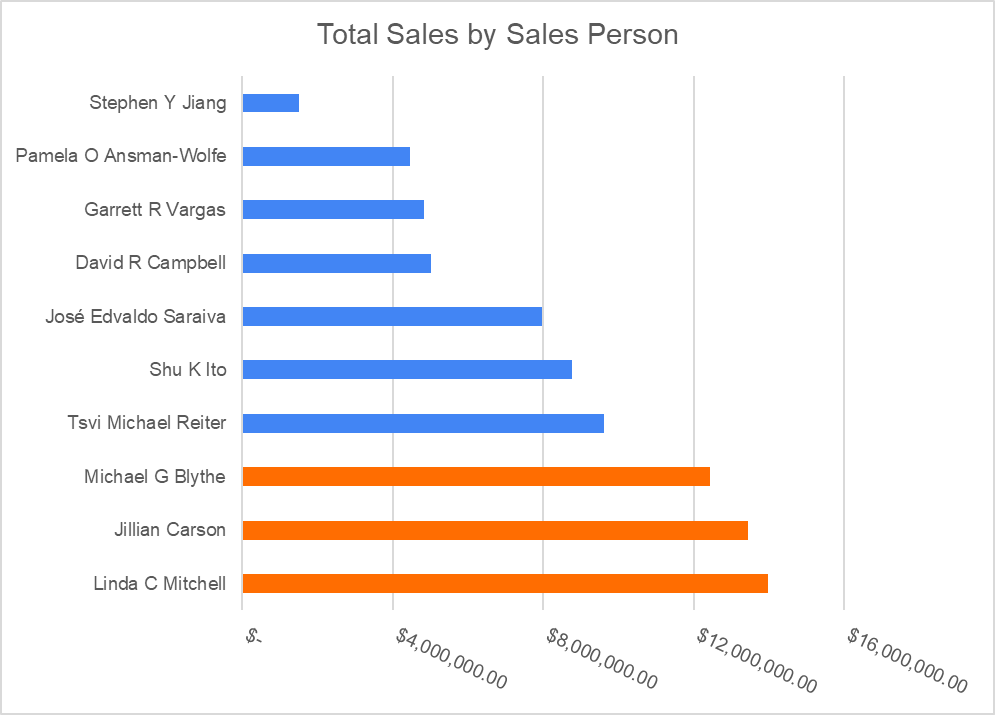




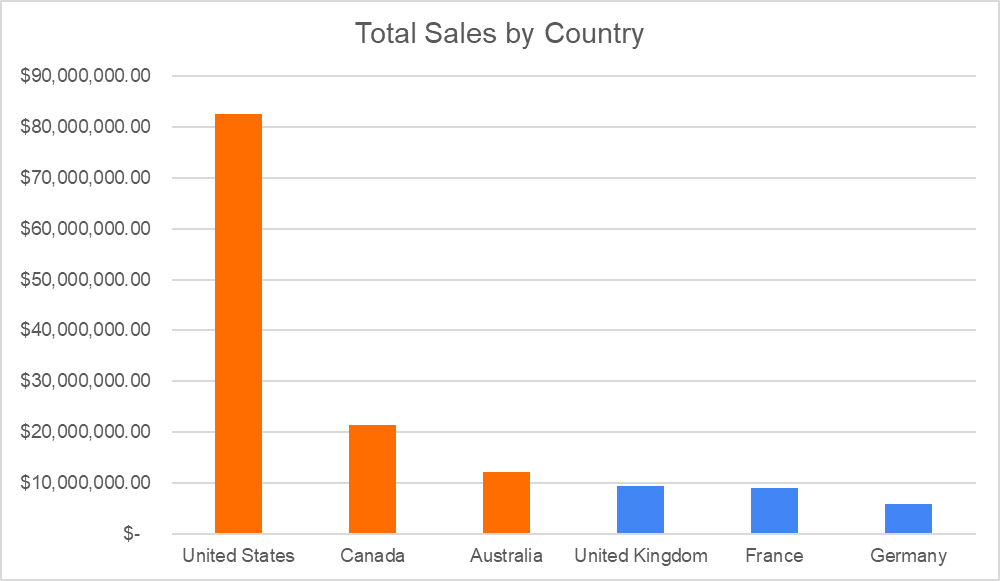
1. **Delivery Performance Across Different Sales Countries:** To analyze how delivery performance varies across different sales countries, I calculated the number of days between the 'OrderDate' and 'ShipDate' by creating a new column called 'ShipmentLag' using the DATEDIFF function. I then created a pivot table to show the average shipment lag for each territory, providing a clear view of delivery performance across regions.



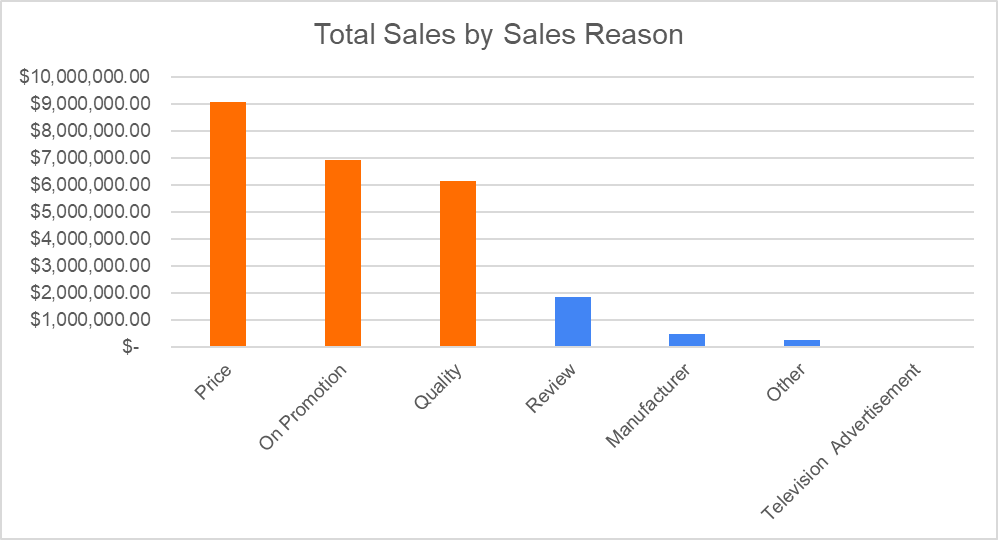
1. **Top-Performing Salespersons by Total Sales:** For identifying the top-performing salespersons by total sales, I created a pivot table that displayed the total sales revenue generated by each salesperson. I then plotted this data using a bar plot, which highlighted the performance of each salesperson in terms of sales volume.



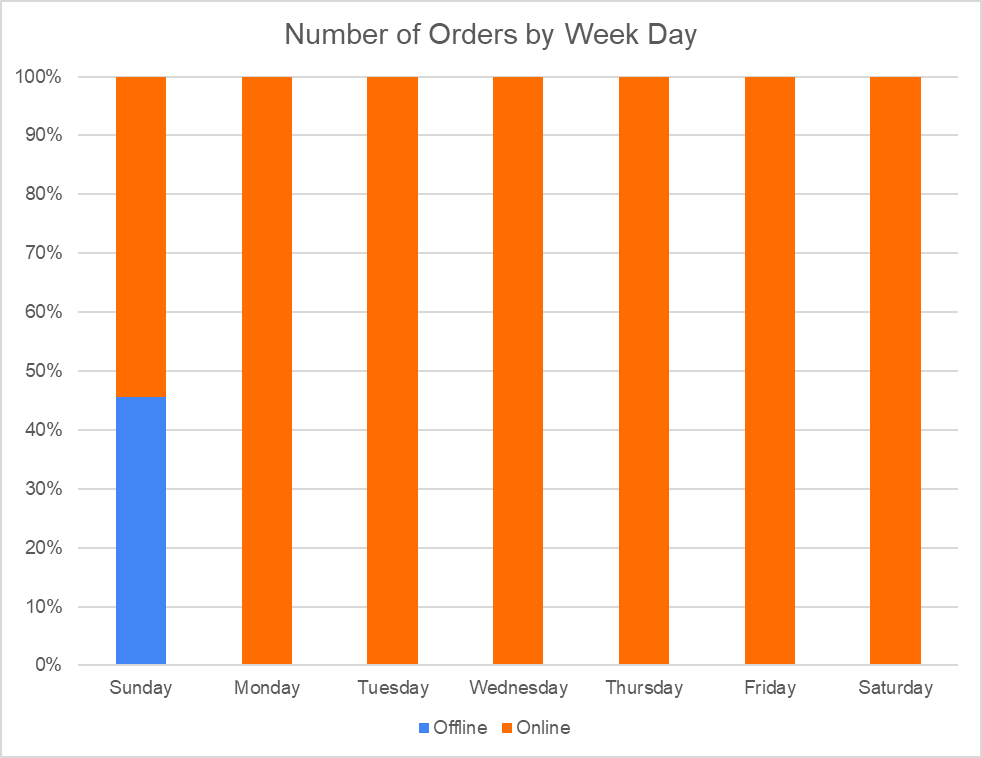
1. **Sales Comparison Across Different Countries:** To compare sales across the different countries, I created a pivot table to show the sum of sales for each territory. I visualized this data with a column chart, which provided a clear comparison of sales performance across various countries.



1. **Top Reasons Influencing Sales:** To find the top reasons influencing sales, I created a pivot table to show the average sales associated with each sales reason. I then used a column chart to visualize this data, making it easy to see which reasons had the most significant impact on sales.



1. **Effect of Day of the Week on Sales (Online vs. Offline):** Finally, to understand how the day of the week an order was placed affects sales for both online and offline channels, I created a 'DayofWeek' column by extracting the day from the 'OrderDate' using the TEXT and DAY functions. I created a pivot table to show the number of orders made each day of the week for both online and offline sales channels. I then visualized this data using a 100% stacked column chart, which allowed me to compare order volumes by day of the week for both sales channels.



By conducting these analyses, I prepared the data to generate insights that can drive business decisions and strategies. The results from these analyses will be shared in the next phase.

## **5. Share**

In the Share phase, I present the insights derived from the analysis, providing valuable information that can inform business decisions and strategies. Here are the key insights gained from each analysis:

### Overall Monthly Sales Trend

* + The analysis revealed a general upward trend in sales over time, indicating an increase in sales from 2001 to 2004, followed by a decline starting in July 2004. This pattern suggests that the company experienced growth initially, but sales began to drop in the latter part of the analyzed period.
  + There was a noticeable seasonality in sales, with repeating patterns observed in certain months each year. Specifically, top sales were recorded in February, May, August, and November. This seasonal trend suggests that certain times of the year are more favorable for sales, possibly due to market conditions, holidays, or promotional activities. Further diagnostic analysis can be conducted to identify the factors influencing high sales in these months.

### Sales Trends Comparison Between Online and Non-Online Orders

* + The comparison between online and offline sales trends showed that more orders were placed using the online sales channel, but the revenue generated from offline sales was significantly higher. Specifically, 88% of total orders were placed online, contributing only 23% of total sales revenue, while 12% of orders placed offline accounted for 77% of total revenue.
  + This finding indicates that while online orders are more frequent, they tend to generate lower revenue per order compared to offline orders. This suggests that the company might need to explore strategies to increase the average order value for online sales or leverage the higher revenue potential of offline sales.

### Delivery Performance Across Different Sales Countries

* + The analysis of delivery performance showed that orders were generally shipped within 7 days across all countries, except for Germany, where shipping could take up to 8 days sometimes. This indicates that the company's logistics operations are fairly consistent, but there may be specific challenges or inefficiencies in the German market that need to be addressed to improve delivery performance.

### Top-Performing Salespersons by Total Sales

* + The analysis identified 'Linda C Mitchell' as the top-performing salesperson, generating $13,975,741.46 in total sales. She was followed by 'Jillian Carson', who generated $13,434,509.55, and 'Michael G Blythe', who made $12,433,502.84 in total sales.
  + Recognizing these top performers can help the company understand what strategies or practices these salespersons employ, which can be shared across the sales team to boost overall performance. Additionally, these insights can be used for incentive and recognition programs.

### Sales Comparison Across Different Countries

* + The sales comparison across countries showed that the United States generated the highest sales ( $82,425,341.28) which can be attributed to the fact that there are five territories in the United States (Northwest, Northeast, Central, Southwest, Southeast). The United States was followed by Canada ($21,501,812.46) and Australia ($12,197,515.53). Germany generated the least total sales ($5,939,763.50) among the countries analyzed.
  + These insights highlight the country with the strongest sales performance and can guide the company in allocating resources and efforts to maximize growth in high-performing areas while developing strategies to boost sales in underperforming countries like Germany.

### Top Reasons Influencing Sales

* + The analysis of sales reasons revealed that the top three reasons influencing sales were Price ($9,057,881.53), On Promotion ($6,926,301.87), and Quality ($6,132,636.01). On the other hand, Television Advertisement had the least impact on sales, with an average of $26,722.48.
  + This suggests that competitive pricing, precise promotions, and the product quality are the most significant factors driving sales. The low impact of television advertising indicates that the company might need to reconsider its advertising strategy or evaluate other factors like the frequency and reach of these ads.

### Effect of Day of the Week on Sales (Online vs. Offline)

* + The analysis showed that while the majority (87.9%) of orders were placed online, Sundays saw the highest order placements for both online and offline sales. This insight can help the business plan and allocate resources, such as staff and server capacity, to meet the higher demand expected on Sundays.
  + Understanding the weekly sales pattern can help in optimizing operations and ensuring that the company is well-prepared to handle peak order times effectively.

By sharing these insights, the company can make informed decisions and develop strategies to enhance sales performance, improve delivery efficiency, recognize and reward top salespersons, and optimize marketing efforts. These insights provide a solid foundation for driving business growth and achieving strategic objectives.

## **6. Act**

Based on the analysis and insights obtained, several actionable recommendations can help AdventureWorks improve its business performance:

### 1. Address the Decline in Sales Post-2004

Investigate the factors contributing to the decline in sales starting in July 2004. This could involve reviewing market conditions, competition, internal changes, or external events that may have impacted sales. Implement strategies to counteract these negative influences and regain the growth trajectory.

### 2. Leverage Seasonality for Strategic Planning

Utilize the insights on seasonality to plan marketing campaigns, promotions, and inventory management. Since February, May, August, and November are peak sales months, the company should align its promotional efforts and stock levels to capitalize on these periods. Additionally, identifying the factors behind these seasonal spikes can help replicate the success in other months.

### 3. Enhance Online Sales Revenue

Although online orders are more frequent, they generate less revenue compared to offline orders. To address this, the company should explore strategies to increase the average order value for online sales. This could include offering bundled products, upselling and cross-selling, implementing loyalty programs, and improving the online shopping experience.

### 4. Improve Delivery Performance in Germany

The slightly longer delivery times in Germany suggest potential inefficiencies in the logistics process for that country. Conduct a detailed analysis to identify the root causes and implement measures to streamline operations. This could involve optimizing shipping routes, partnering with local couriers, or enhancing warehouse efficiency in Germany.

### 5. Recognize and Replicate Top Salesperson Practices

Linda C Mitchell, Jillian Carson, and Michael G Blythe have demonstrated outstanding sales performance. Recognize their achievements and analyze their sales techniques, customer interactions, and strategies. Share these best practices across the sales team through training sessions and mentorship programs to elevate the overall sales performance.

### 6. Focus on High-Performing Countries

With the United States, Canada, and Australia showing strong sales performance, allocate more resources to these countries to further boost sales. This could include targeted marketing campaigns, increased sales personnel, and tailored promotions. Simultaneously, develop strategies to improve sales in underperforming regions like Germany.

### 7. Optimize Advertising Strategies

The low impact of television advertising on sales suggests that the current approach may not be effective. Reevaluate the television advertising strategy, considering factors like ad content, timing, frequency, and target audience. Additionally, explore other advertising channels such as digital marketing, social media, and influencer partnerships that may yield better returns.

### 8. Resource Allocation Based on Weekly Sales Patterns

The finding that Sundays have the highest order placements for both online and offline sales can guide resource allocation. Ensure that adequate staff, server capacity, and inventory are available to handle the increased demand on Sundays. This proactive approach can enhance customer satisfaction and operational efficiency.

By implementing these recommendations, AdventureWorks can address identified issues, capitalize on opportunities, and enhance overall business performance. These actions, informed by data-driven insights, will help the company achieve its strategic objectives and drive sustained growth.