Li, Tao

850001413

GDDA604

Assessment2 Project

Table of Contents

[Introduction: 2](#_Toc174625834)

[Task A –Data Transformation: 2](#_Toc174625835)

[a. Load the dataset into a DataFrame 2](#_Toc174625836)

[b. Show the first few rows of the loaded dataset 3](#_Toc174625837)

[c. Apply three (3) operations to handle missing values in a dataset 4](#_Toc174625838)

[d. Choose a column and perform the sorting technique 8](#_Toc174625839)

[e. Define a condition to filter transactions from the dataset 10](#_Toc174625840)

[f. Create a new column to derive additional information 12](#_Toc174625841)

[g. Choose the categorical column and aggregate data based on it 14](#_Toc174625842)

[Task B – Data Analysis 16](#_Toc174625843)

[a. Group the dataset based on a categorical variable and calculate summary statistics 16](#_Toc174625844)

[b. Investigate the correlations between different variables in the dataset 21](#_Toc174625845)

[c. Export a dataset to a CSV file using Python or any other similar programming tool 23](#_Toc174625846)

[d. Perform data analysis and visualization in Python to derive insights 24](#_Toc174625847)

[e. Apply inferential statistical method to quantify the relationships between variables 34](#_Toc174625848)

[Task C – Data Findings and Decision Support 36](#_Toc174625849)

[a. Analyze the results obtained from data analysis, including grouping, summarizing, investigating correlations, and applying inferential statistical methods 36](#_Toc174625850)

[b. Interpret the relationships between variables, summarize key findings, and identify significant trends or patterns 38](#_Toc174625851)

[c. Provide specific suggestions for addressing business challenges or opportunities identified in the dataset 40](#_Toc174625852)

# Introduction:

In this task, I used the dataset ‘Sales\_Samsple\_Public\_Dataset.csv’ provided by the lecturer on Canvas, which I downloaded to my local machine. The dataset contains 2,823 rows and 25 columns, capturing detailed information about orders placed by customers, and providing a comprehensive view of order transactions, customer information, product details, and sales figures. I utilized Google Colab to perform various tasks, such as transforming, analyzing, and interpreting data. This process helped generate valuable insights into business strategies for sales analysis, enabled data-driven recommendations for customer segmentation, and delivered integrated solutions for performance improvement.

# Task A –Data Transformation:

## a. Load the dataset into a DataFrame

I have imported several libraries for data analysis including the Pandas library for loading the dataset into a DataFrame. My dataset is ‘Sales\_Sample\_public\_Dataset.csv’. In the Pandas package, there is a built-in method ‘read\_csv’. I have used this method for reading the file.

|  |
| --- |
| Please find the screenshot below:  A white rectangle with red text  Description automatically generated |

## b. Show the first few rows of the loaded dataset

Through the pandas’ package, I used the built-in method ‘head()’. we can pass any number as the parameter for this method. by default, it gives the first five records. if I pass 10 as the parameter, it will display 10 records.

|  |
| --- |
| Please find the screenshot below:  A screenshot of a white screen  Description automatically generated |
| Display 10 rows:  A table with numbers and letters  Description automatically generated with medium confidence |
| Columns are:  A close-up of a computer screen  Description automatically generated |

## c. Apply three (3) operations to handle missing values in a dataset

I used the ‘isnull()’ method to generate a Boolean mask indicating where values are missing and the ‘sum()’ method to sum up the missing values. The null value in columns ‘ADDRESSLINE2’, ‘STATE’, ‘POSTALCODE’, and ‘TERRITORY.

|  |
| --- |
| Please find the screenshot below:  A screenshot of a computer screen  Description automatically generated |
| First operation:  I used ‘drop()’ method in pandas to drop the column ‘ADDRESSLINE2’ from DataFrame, ‘axis=1’ indicating that I was dropping a column. ‘inplace = True’ indicating the operation should be done directly on ‘df’. |
| Please find the screenshot below: |

|  |
| --- |
| Second operation:  I used the ‘fillna()’ method in pandas to replace the missing values ‘NaN’ in column ‘TERRITORY’ with ‘North\_America’. |
| Please find the screenshot below:  A screenshot of a computer code  Description automatically generated |
| Similarly, I used the same method to replace the missing value in column ‘STATE’ with ‘unknown’.  Please see the screenshot below:  A screenshot of a computer code  Description automatically generated |

|  |
| --- |
| Third operation:  I used the ‘mode()’ method in pandas to calculate the most frequent value in column ‘POSTALCODE’ when the ‘STATE’ is ‘CA’ and replace the missing values ‘NaN’ in column ‘POSTALCODE’ with using the ‘mode’ method. |
| Please find the screenshot below:  A screenshot of a computer program  Description automatically generated |

|  |
| --- |
| After performing the imputation, I found the ‘POSTALCODE’ column would not be used in the next tasks, so I decided to drop the column.  I checked again after handling the missing values. The data is cleaned. |
| Please see the screenshot below:  A screenshot of a computer  Description automatically generated |

## d. Choose a column and perform the sorting technique

I used the ‘sort\_values()’ function in pandas to sort the DataFrame by the value column ‘SALES’ in ascending order, which is the default way (‘ascending=True’). If I added the ‘inplace=True’ parameter which would have modified the DataFrame in place, I could have done so if I wanted to avoid creating a new sorted DataFrame.

If I want to sort the DataFrame in descending order, I can add the ‘ascending=False’ parameter.

Through this process, I can quickly identify the categories with the highest and lowest sales orders, which is very helpful for further analysis, such as product sales trends and customer segmentation. In this case, ‘Sharp Gift Warehouse’ and “Euro Shopping Channel’ placed the largest orders based on the sorted sales data.

|  |
| --- |
| Please find the screenshot below:  DataFrame in ascending order:  A screenshot of a computer screen  Description automatically generated  A screenshot of a computer  Description automatically generated  DataFrame in descending order:  A screenshot of a document  Description automatically generated  A screenshot of a computer  Description automatically generated |

## e. Define a condition to filter transactions from the dataset

I used the ‘df[condition]’ technique leveraging Boolean indexing in pandas to filter data from the DataFrame based on a specific condition. Here I gave the condition that each element is ‘True’ if the corresponding ‘PRICEEACH’ value is less than 100, otherwise, it is excluded.

This method filtered the ‘df’ to include only rows where the ‘PRICEEACH’ value is less than 100.

With this approach, I can identify products that are popular among customers looking for more affordable options, revealing a segment of price-sensitive customers and informing targeted marketing strategies. Additionally, by analysing the sales and quantities ordered within this filtered data, I can assess the contribution of lower-priced items to overall sales volume.

|  |
| --- |
| Please find the screenshot below:  A screenshot of a computer  Description automatically generated  A close-up of a document  Description automatically generated |

## f. Create a new column to derive additional information

I created a new column ‘ORDER\_YEAR\_MONTH’ by combining ‘YEAR\_ID’ and ‘MONTH\_ID’. I used the ‘astype(str)’ function in pandas to convert the columns ‘MONTH\_ID’ and ‘YEAR\_ID’ to string format data type. I used ‘zfill(2)’ function in pandas to pad the single-digit month with a zero to perform a consistent two-digit format. At last, I used the ‘+’ operator to concatenate the string with a hyphen ‘-‘ separator in between to create the new column called ‘ORDER\_YEAR\_MONTH’.

With this new column being added, I can analyse sales patterns on a month-by-month basis, identifying peak sales periods, seasonal trends, and fluctuations. By understanding when certain products are in higher demand, I can optimize inventory levels and ensure the supply chain is prepared for peak periods. Identifying which months customers tend to place larger or smaller orders also can help to tailor promotions and discounts to drive sales, helping to refine future marketing strategies.

|  |
| --- |
| Please see the screenshot below:  A close-up of a computer screen  Description automatically generated |

|  |
| --- |
| A screenshot of a number  Description automatically generated  A close-up of a document  Description automatically generated |

## g. Choose the categorical column and aggregate data based on it

I aggregated the data on the ‘COUNTRY’ column and calculated the total sales for each country.

I used the ‘groupby()’ function in pandas by the value in the ‘COUNTRY’ column, allowing me to create groups of rows where each group corresponds to a specific country. I used the ‘agg()’ function in pandas to perform aggregation, allowing me to summarize and complement statistics on data, the function {‘SALES’: ‘sum’} specifies to apply the ‘sum()’ function to the ‘SALES’ column.

The ‘reset\_index()’ method ensures that the ‘COUNTRY’ column is a regular column instead of an index.

By summing up the sales for each country, I can easily identify which countries contribute the most to overall sales, helping in recognizing top-performing markets. Moreover, with a clear view of sales by country, I can allocate resources more effectively focusing on regions with the highest return on investment. On the other hand, if certain countries show lower sales, this might indicate potential growth opportunities, exploring some strategies to increase market share is also a choice, setting targets for countries that are lagging behind others.

In this analysis, the USA is the top-performing country, indicating that the USA is the most critical market. After the USA, Spain and France show strong sales, making them key markets in Europe. Conversely, countries like Ireland, the Philippines, and Switzerland show relatively low sales.

|  |
| --- |
| Please find the screenshot below:  A screenshot of a computer  Description automatically generated |

# Task B – Data Analysis

## a. Group the dataset based on a categorical variable and calculate summary statistics

I used the ‘groupby()’ function in pandas to group the dataset by the ‘STATUS’ column and then applied the ‘describe()’ function in pandas to calculate summary statistics for each group. This provides an extensive overview of the data across different statuses, showing count, mean, standard deviation, minimum, and other statistics for various columns.

The mean order number is consistent across status. Dates vary widely among statuses, with ‘Shipped’ and ‘Cancelled’ having more recent dates on average. The ‘Shipped’ status stands out with high counts across several metrics, suggesting it is a major contributor to overall sales and order volume. In contrast, statuses like ‘Disputed’ and ‘On Hold’ show higher values in specific metrics but with fewer counts.

|  |
| --- |
| Please find the screenshot and the result below:  A close-up of a computer code  Description automatically generated |

|  |
| --- |
| A screenshot of a computer screen  Description automatically generated  A table of numbers and a few percentages  Description automatically generated with medium confidence |

I applied the same method to the ‘productline’ colume to analyse its distribution and patterns. Classic cars, trucks, and buses have the highest average order values. Trucks and buses have the highest average quantity ordered while vintage cars have the lowest per order. The highest average price per unit is for classic cars while vintage have the lowest. The variability in pricing and order quantities indicates a diverse market with different customer preferences and pricing strategies.

|  |
| --- |
| A screenshot of a computer screen  Description automatically generated  A screenshot of a document  Description automatically generated  A screenshot of a document  Description automatically generated |

To better understand the distribution by ‘productline’, I created a histogram to visualize the results. I found that different product categories exhibit a similar pattern in both quantity order and sales. This consistent trend suggests that, regardless of the product line, customer ordering behaviour and sales performance follow a common distribution across categories.

|  |
| --- |
| Please find the screenshot below:  A screenshot of a graph  Description automatically generated |

## b. Investigate the correlations between different variables in the dataset

I first extracted the numerical features from the DataFrame using the ‘select\_dtypes()’ function in pandas and passed ‘include=[‘number’]’ to extract only numeric features. Then I used ‘corr()’ function in pandas to calculate the correlation matrix for the features. To visualize it, I used Seaborn to create a heatmap of the correlation matrix by annotating the heatmap with correlation coefficients, formatting the numbers to two decimal places, using the ‘coolwarm’ colour map, and adjusting the visual style. At last, I printed the matrix to the console to have a comprehensive view of the data features.

‘Price each’, ‘Quantity ordered’, and ‘MSRP’ have a strong positive correlation with ‘Sales’. The date-time-related features show weak correlations with most other features, indicating that temporal features have minimal impact on the numeric attributes in the dataset. Analysing how quantity ordered affects sales can help in inventory and order management. By leveraging these insights, I can enhance my understanding of feature relationships and make more informed decisions regarding business strategies.

|  |
| --- |
| Please find the screenshot and result below:  A computer screen shot of a computer code  Description automatically generated  A chart with red and blue squares  Description automatically generated  A table of numbers with numbers  Description automatically generated |

## c. Export a dataset to a CSV file using Python or any other similar programming tool

I have mounted my Google Drive to the Colab environment and defined the file path. I used the ‘to\_csv()’ function in pandas to export the dataset to a CSV file and saved it in the Google Drive file under ‘604 assessment 2’ for future use. ‘index=False’ indicated excluding the index in the CSV file.

|  |
| --- |
| Please find the screenshot below: |

|  |
| --- |
| I can confirm the dataset is successfully saved in my Google Drive folder.  A screenshot of a computer  Description automatically generated |

## d. Perform data analysis and visualization in Python to derive insights

In this task, I utilized Python to visualize the data and derive meaningful insights. By leveraging various Python libraries, such as Matplotlib and Seaborn, I created visualizations to better understand the underlying patterns and relationships within the data. This approach helped in identifying trends, correlations, and anomalies, enabling a more informed analysis and interpretation of the data.

|  |
| --- |
| * Correlation of country and product line by total sales   I called the ‘groupby()’, the ‘sum()’ and the ‘unstack()’ functions for data aggregation and preparation from the pandas library.  I called the ‘sns.heatmap()’ function to create the heatmap plot from the seaborn library and passed several arguments, including the data, the annotation, the formats, and the colour of the map. I called the ‘plt.figure()’ and ‘plt.show()’ functions from the matplotlib library to set up the figure and show the plot. |
| Please find the screenshots below:  A screenshot of a computer code  Description automatically generated  A screenshot of a data sheet  Description automatically generated |

|  |
| --- |
| * Comparison of average price each and MSRP by product line   I called the ‘groupby()’ and the ‘mean()’ functions for data aggregation and preparation of the mean value from the pandas library. I called the ‘plt.bar()’ function from the matplotlib library to create a bar plot and passed arguments to it, ‘x’ is the x-coordinates of the bars, ‘width’ is the width of the bars, ‘label’ is the label for the legend, ‘color’ is the colour of the bars, ‘alpha’ is the transparency of the bars. |
| Please find the screenshots below:  A screenshot of a computer program  Description automatically generated  A graph of blue and red bars  Description automatically generated |
| * Monthly sales trends over time   I called the ‘groupby()’ and the ‘sum()’ functions for data aggregation and preparation of the sum value from the pandas library. I called the ‘sns.lineplot()’ function from the seaborn library to create a line plot and passed arguments to it. I grouped sales data by month and year, aggregating the value by summing them. I then created a line plot to visualize the sales over time, providing insights into sales trends. |
| Please see the screenshots below:  A screenshot of a computer code  Description automatically generated  A graph with blue lines and numbers  Description automatically generated |

|  |
| --- |
| * Total sales distribution by country   I called the ‘groupby()’ and the ‘sum()’ functions for data aggregation and preparation of the sum value from the pandas library. I called the ‘sns.barplot()’ function from the seaborn library to create a bar plot and passed arguments to it. I grouped sales data by country, aggregating the value by summing them. I then created a bar plot to visualize the sales in different countries. |
| Please see the screenshot below:  A screenshot of a computer code  Description automatically generated  A graph with blue and purple bars  Description automatically generated |

|  |
| --- |
| * Total sales distribution by product line |
| Please see the screenshots below:  A screenshot of a computer code  Description automatically generated  A graph with different colored bars  Description automatically generated |
| * Total sales analysis by product line and deal size   A screen shot of a computer code  Description automatically generated  A graph of sales by product line  Description automatically generated |
| * Total sales analysis by top 10 customers and deal size   A screenshot of a computer code  Description automatically generated  A graph of sales  Description automatically generated |

|  |
| --- |
| * Relationship between price and sales   I set the size of the plot to 12 inches wide and 6 inches tall. I called the ‘sns.scatterplot()’ function from seaborn library to create a scatter plot. It plot ‘PRICEEACH’ on the x-axis and ‘SALE’ on the y-axis. The ‘alpha=0.6’ parameter adjusted the transparency of the points. Then I set up the title and labels and displayed the plot. |
| Please see the screenshot below:  A screen shot of a computer  Description automatically generated  A graph of a graph showing a number of blue dots  Description automatically generated with medium confidence |

|  |
| --- |
| * Treemap of each country’s sales in year   I imported the plotly.express as px which is used for creating interactive visualizations. I called the ‘px.treemap()’ function from the plotly express library and passed the arguments. The ‘path’ parameter specifies the hierarchy of the categories for the treemap. |
| Please see the screenshot below:  A close-up of a sign  Description automatically generated  A screenshot of a computer screen  Description automatically generated |

|  |
| --- |
| * Order status distribution in each year   I first grouped the df by ‘YEAR\_ID’ and ‘STATUS’, counting the number of order of each group. The result is stored in a new column ‘count’. I got the unique years to be prepared for creating subplots. I then created pie charts for each year by using the ‘for’ loop iterating over the unique years. |
| Please find the screenshot below:  A computer screen shot of a program code  Description automatically generated  A red circle with a blue triangle and a blue triangle  Description automatically generated |

## e. Apply inferential statistical method to quantify the relationships between variables

In this task, I performed three inferential statistical methods in the previous task, ANOVA test, Pearson correlation test, and Chi-square test.

|  |
| --- |
| I imported the relevant libraries. |
| Please see the screenshot below:  A black and purple text  Description automatically generated |

|  |
| --- |
| * ANOVA test   To determine if different product lines would have impacts on the sales amounts, I performed the ANOVA test.  The sum of squares for the C (product line) indicates the variation in sales attributed to different product lines, suggesting substantial differences in sales among different products. However, the P-value is 0.058, which is close to 0.05 but slightly greater than 0.05, indicating accept the null hypothesis that the difference in sales among product lines are not significant.  Hence, based on the data, the product line does not have a significant effect on sales. |
| Please see the screenshot below:  A screenshot of a computer  Description automatically generated |

|  |
| --- |
| * Pearson correlation   To better understand how promotional and bundle sales strategies might impact the sales price on the quantity ordered, I performed a Pearson correlation test. This test helps to evaluate the strength and direction of the linear relationship between the two variables, providing insights into how marketing strategies might influence pricing.  The correlation value of 0.0056 indicating there is a very weak positive correlation between ‘QUANTITYORDERED’ and ‘PRICEEACH’, meaning that as one variable changes, the other does not show a consistent directional change.  Additionally, the P-value of 0.76 is significantly greater than the common significance level of 0.05, indicating accept the null hypothesis that there is no significant relationship between the two variables ‘QUANTITYORDERED’ and ‘PRICEEACH’. |
| Please see the screenshot below:  A close-up of a code  Description automatically generated |

|  |
| --- |
| * Chi-square test   To understand if the deal size affects the order status, I performed the Chi-square test.  The P-value is 0.00016 lower than 0.05 indicating that I can reject the null hypothesis. Hence, to conclude that there is a statistically significant association between ‘DEALSIZE; and ‘STATUS’. This information can be valuable for understanding how deal size influence order fulfilment outcomes and can guide decision-making related to order management and customer service strategies. |
| Please find the screenshot below:  A screenshot of a computer code  Description automatically generated |

# Task C – Data Findings and Decision Support

## a. Analyze the results obtained from data analysis, including grouping, summarizing, investigating correlations, and applying inferential statistical methods

In this part, I focused on analysing the data from previous tasks. This included sorting, aggregating, and filtering the data in Task A, and performing grouping, correlation, visualization, and inferential statistical methods in Task B to derive meaningful insights from the analysis results.

Grouping and summarizing:

* Product line analysis:

I grouped the data by ‘product line’, found the average of ‘price each’ and ‘MSRP’, and plotted the difference in a bar plot. The bar plot showed the sales price and the manufacturer’s suggested retail price between different product categories. Classic cars are the best-selling category while trains have the lowest sales orders.

* Sales trends over time:

I grouped the data by ‘order year month’ and visualized it with a line plot, identifying the sales trends over time. This provided insights into seasonal patterns and overall sales trends. November is the peak month for sales which always outperforms. Sales data for 2005 are only available for the first five months of the year, so analysis is limited to this period.

* Sales distribution by region and variable:

I grouped the data by ‘country’ and summed the distribution of sales across different regions, showing the comparative performance of various countries. Then I applied the same method to the ‘product line’, ‘deal size’, and ‘customer name’ variables to gain a better understanding of the distribution of these variables in terms of ‘sales’. The USA is the top-performing country. After the USA, Spain and France show strong sales, making them key markets in Europe.

* Price vs Sales:

I used the scatter plot between ‘price each’ versus ‘sales’ to visualize the price range where the most sales occurred, finding the most orders are within the price range 60-80.

Investigating correlations:

* Sales distribution:

I grouped the data by ‘country’ and ‘product line’ and then summed the ‘sales’ for each combination. I used a heatmap to visualize the summed sales data, which provided a matrix showing sales distribution across different countries and product lines. The USA contributed the most in terms of sales, especially in the classic car category.

Inferential statistical methods:

I performed three inferential statistical methods in the previous task, ANOVA test, Pearson correlation test, and Chi-square test.

* ANOVA test:

The results indicate insufficient statistical evidence to claim a significant association between product line and sales. However, the P-value is close to 0.05 suggesting that there may be some potential relationship present that warrants further investigation.

* Pearson correlation test:

The results, with a P-value of 0.76, reveal no significant relationship between the quantity ordered and the price of each product.

* Chi-square test:

The results, with a P-value of 0.00016, indicate a significant association between order status and deal size, suggesting the status of the orders varies with different sizes of the deals.

## b. Interpret the relationships between variables, summarize key findings, and identify significant trends or patterns

In this section, I focused on summarizing the main findings mentioned in the task above to explore trends and possibilities.

Key findings

* Country sales:

1. The USA has the highest sales globally, with its sales volume exceeding the combined sales of the second (Spain) and third (France) highest countries. In 2004, the UK surpassed Australia in sales, becoming the fourth-highest sales country, and in the same year, sales in Germany and Japan also increased significantly. At the beginning of 2005, besides the top three countries, sales in Australia and Finland were also very high.
2. The Philippines and Ireland have the lowest sales.
3. Some countries sell products in very limited categories, such as Switzerland only sells classic cars and Denmark does not sell motorcycles.

* Product category:

Classic cars are the best-selling category, with sales twice as high as the second place, vintage cars, indicating a need to focus on high-performing product lines.

* Deal size:

The most common deal size is medium across all product lines.

* Customer:

The key customers with the highest purchasing power are Euro Shopping Channel and Mini Gifts Distributors, whose deal sizes are predominantly medium.

* Product pricing:

The relationship between the sale price and MSRP indicates that some products are always priced lower than their suggested price, suggesting that most of these products are sold through promotions.

* Order status:

Most orders have been shipped, with only a small number being cancelled or put on hold, indicating the order fulfilment process is relatively smooth.

Significant trends and patterns

* Seasonality:

Monthly sales data show a seasonal trend where November consistently outperform other months, indicating periods of high and low demands. This is crucial for inventory planning and marketing strategies.

* Customer segmentation:

Customers with large order sizes, such as Euro Shopping Channel, contribute significantly to total sales, indicating the importance of focusing on these segments.

* Geographical sales:

Sales in certain countries, particularly the USA, Spain, and France, are high, indicating key markets for the business.

* Price and quantity ordered:

There is no obvious direct relationship between the selling price and the quantity of the order.

## c. Provide specific suggestions for addressing business challenges or opportunities identified in the dataset

In this part, I paid more attention to discussing possible challenges and opportunities in the business through data analysis and giving constructive suggestions.

Challenges:

Due to market competition or distribution issues, some countries, such as the Philippines and Ireland, might decline their orders. It is essential to pay close attention to the seasonal fluctuations, especially during the peak month of November, as out-of-stock situations could negatively impact customer satisfaction and sales. Additionally, certain product categories, such as trains and ships, may face low sales due to insufficient market demand.

Opportunities:

* Strengthen key customer relationships:

To focus on improving relationships with the top customers, consider implementing personalized marketing and loyalty programs, such as exclusive offers for high-value customers like Euro Shopping Channel and Mini Gifts Distributors. Offering benefits for repeat purchases such as discounts, early access to new products or special promotions can help retain top customers and encourage them to make larger or more frequent purchases.

* Optimize inventory and marketing:

Use the seasonal trends to optimize inventory levels and marketing strategies, particularly the high sales in November, such as plan promotions based on peak sales periods to maximize revenue and perform stock-take plans to reduce out-of-stock situations. This involves forecasting demand more accurately based on historical sales data and seasonal trends.

* Geographical expansion:

Invest in distribution channels in high-performance and quick-growing countries such as the USA, Spain, and France to further leverage these markets. For countries with low sales like the Philippines and Ireland, investigate potential barriers such as market competition and distribution challenges or opportunities to boost sales. Explore opportunities by adapting strategies to local market conditions, including targeted promotions and strategic partnerships.

* Pricing strategy:

Analyze the product category with significant differences between the selling price and MSRP to understand the market value, adjusting pricing accordingly to achieve higher profit margins, considering factors such as production costs, competitive pricing, and customer willingness to pay. Offer limited-time deals or bundle pricing to attract customers and increase sales volume.

* Encourage larger deals:

Leverage bundle deals, volume discounts, and loyal programs to encourage customers to make larger purchases. This approach can increase average order size and overall revenue. Expand loyalty programs to include rewards for large purchases, encouraging customers to spend more per transaction. This can include tiered rewards based on order size or frequency of purchases.

* Invest in high-performing product lines:

Consider expanding the range or variants of high-performing products like classic cars to boost sales further. This can include introducing new models, accessories, or related products to capture additional market share and boost sales.

To conclude, implementing these recommendations can help tackle business challenges, such as declines in orders in certain countries and low orders for specific products. By leveraging data analysis to identify effective strategies, businesses can address these issues through actions like strengthening customer relationships, optimizing inventory and marketing strategies, exploring geographic expansion, improving pricing methods, encouraging larger deals, and investing in high-performance product lines. These measures are crucial for improving business performance and driving growth.

For the reference of the code, please find the GitHub link below:

<https://github.com/YiliaTao0122/604Assessment2>