Exploratory

Data Analysis

A part of Assessment 2 -GD604 Data Collection and Analysis

New Zealand School of Education College (NZSE)

Prepared by: Aju Peter

**Date:** 05-03-2024

Contents:

**Task A –Data Transformation**

1. Load the dataset into a DataFrame
2. Show the first few rows of the loaded dataset
3. Apply three (3) operations to handle missing values in a dataset
4. Choose a column and perform the sorting technique
5. Define a condition to filter transactions from the dataset
6. Create a new column to derive additional information
7. Choose the categorical column and aggregate data based on it

**Task B – Data Analysis**

1. Group the dataset based on a categorical variable and calculate summary statistics
2. Investigate the correlations between different variables in the dataset
3. Export a dataset to a CSV file using Python or any other similar programming tool
4. Perform data analysis and visualization in Excel, Python or any other similar programming tool to derive insights
5. Apply inferential statistical method to quantify the relationships between variables

**Task C – Data Findings and Decision Support**

1. Analyse the results obtained from data analysis, including grouping, summarizing, investigating correlations, and applying inferential statistical methods
2. Interpret the relationships between variables, summarize key findings, and identify significant trends or patterns
3. Provide specific suggestions for addressing business challenges or opportunities identified in the dataset

**Task A –Data Transformation:**

1. **Load the dataset into a DataFrame.**

Scenario:

As a data analyst, I have been tasked with investigating the sales decline in online. For the analysis, I've chosen to use Python software and the Jupyter Notebook IDE. Firstly, I have successfully imported the necessary libraries,

* Pandas: It is a library for data manipulation and analysis library for Python.
* NumPy: It is a library for statistical analysis in Python.
* Matplotlib: It is a library for plotting 2Dimensional graphs.
* Seaborn: It is a library for attractive and informative statistical graphics.
* Warnings: It is a library used to ignore.

Later, I imported the dataset named ‘Sales\_Sample\_Public\_Dataset’ with help of ‘pd.read\_csv()’ function. However, upon loading the dataset, I encountered a datatype error. To resolve this issue, I utilized the ‘encoding’ parameter with the value ‘latin1’. (see figure 1)

(**UnicodeDecodeError**: 'utf-8' codec can't decode byte 0x84 in position 5363: invalid start byte)



Figure 1

1. Show the first few rows of the loaded dataset.

I have successfully created a dataframe name ‘sales’ and displayed few rows. The data include 25 attributes and 2823 records. (see figure 2)

The attributes are ORDERLINENUMBER, SALES, ORDERDATE, STATUS, QTR\_ID, MONTH\_ID, YEAR\_ID, PRODUCTLINE, MSRP, PRODUCTCODE, CUSTOMERNAME, PHONE, ADDRESSLINE1, ADDRESSLINE2, CITY, STATE, POSTALCODE, COUNTRY, TERRITORY, CONTACTLASTNAME, CONTACTFIRSTNAME, and DEALSIZE)

A screenshot of a data report

Description automatically generated

Figure 2

1. **Data cleaning**

I have performed multiple operations to clean the data for further analysis.

1. Identified Missing Values by using the “isnull( ). sum( )” method to find any null values present in the dataset. I found below missing values in columns.

* ADDRESSLINE2: 2521
* STATE: 1486
* POSTALCODE: 76
* TERRITORY: 1074

1. I have found that few postal codes are missing, to fill the missing data I have collected the postal code from ‘Google’ and updated the postal code by using ‘data mapping’ function. (see figure 3)

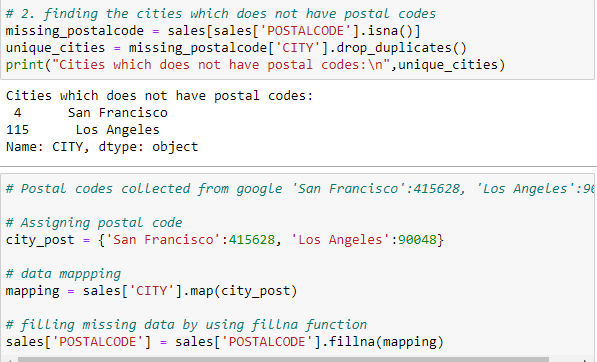


Figure 3

1. I have dropped unwanted columns which is not necessary for this analysis with the help of ‘drop’ function. (see figure 4)

* ADDRESSLINE2: 2521
* STATE: 1486
* TERRITORY: 1074

A computer code with red text

Description automatically generated

Figure 4

1. Found ‘ORDERDATE’ column in object type, I have converted the entire column into date format by using ‘pd.to\_datetime’ function. (see figure 5)

A white background with black text

Description automatically generated

Figure 5

1. Finally I have ensured there is no duplicate records in the dataframe by checking ‘duplicated( )’ function.
2. Displaying few rows of cleaned data (see figure 6)

A table with numbers and letters

Description automatically generated

Figure 6

1. **Data Sorting**

I have performed sorting techniques to retrieve some quick information about the dataset.

1. To retrieve the top 5 orders with the highest sales value, first filter the DataFrame by the 'STATUS' column for 'shipped', then use the sort\_values() function on the 'SALES' column in descending order, and finally select the top 5 rows. (see figure 7)

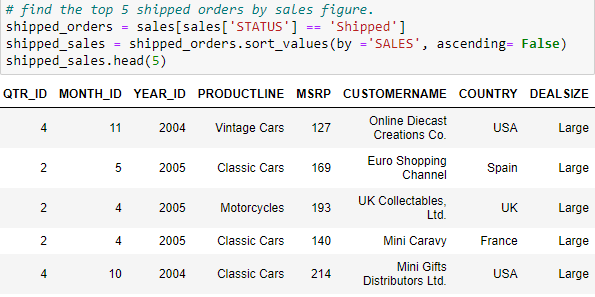


Figure 7

1. To retrieve the top 5 cancelled orders with their sales figures, first filter the DataFrame by the 'STATUS' column for 'Cancelled', then use the sort\_values() function on the 'SALES' column in descending order, and finally select the top 5 rows. (see figure 8)



Figure 8

1. To identify the top 10 orders based on their sales figures, I used the sort\_values() function on the 'SALES' column in descending order and created a new DataFrame named top\_10\_sales that includes essential details such as 'ORDERNUMBER', 'CUSTOMERNAME', 'PRODUCTLINE', 'SALES', 'COUNTRY', and 'STATUS'. This analysis allows us to easily identify top customers and understand their preferences in terms of product lines. (see figure 9)

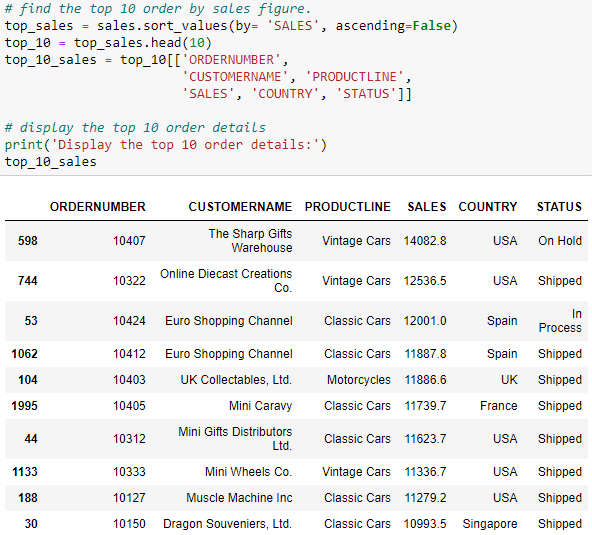


Figure 9

1. **Data filtering**

I have performed filtering techniques to retrieve some quick information about the dataset.

1. I have applied a filter to the dataset to isolate orders associated with the 'Planes' product line, thereby revealing the details of all orders related to planes. (see figure 10)

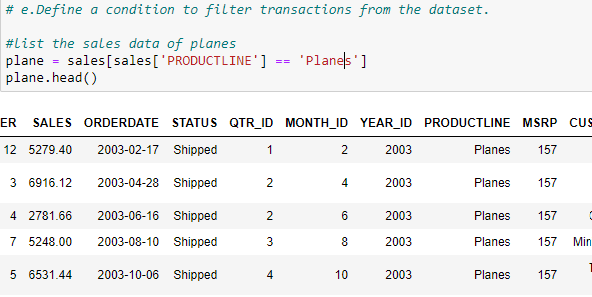


Figure 10

1. I filtered the dataset to identify orders valued at less than $500, uncovering a single order from 'La Rochelle Gifts' in 2005. (see figure 11)

A screenshot of a computer

Description automatically generated

Figure 11

1. I have applied a filter to the dataset to isolate the motorcycle sales in 2005(see figure 12)

A screenshot of a computer

Description automatically generated

Figure 12

1. **Create a new column.**

While examining the dataset, I realized that it lacks a column for the total sales order value, which is crucial for obtaining quick insights. Since each sales order can include multiple items with the sales value recorded at the item level, this absence complicates the analysis of total sales figures for unique orders. To address this gap and facilitate a more straightforward assessment of sales performance, I decided to add a new column named ‘Total Order Value’. This addition aims to aggregate the sales values at the order level, thereby simplifying the process of determining the sales figure for each unique sales order. (see figure 13)

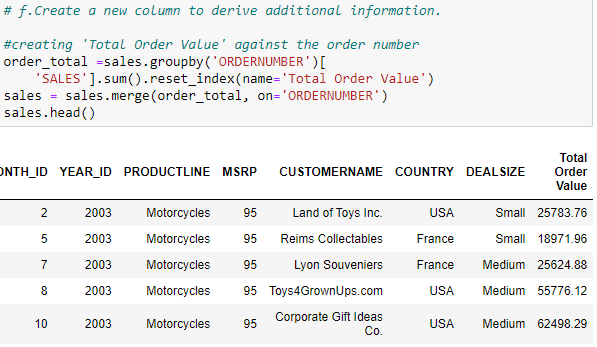


Figure 13

Additionally, to gain a clearer understanding of each unique order and its specifics, I filtered the dataset to focus on key columns. The columns I selected for this refined view include 'ORDERNUMBER', 'ORDERDATE', 'STATUS', 'CUSTOMERNAME', 'COUNTRY', and the newly added 'Total Order Value'. This filtration was implemented to concentrate on the essential details of each order, ensuring a more efficient and focused analysis of the data, particularly concerning the aspects that directly impact sales figures and customer engagement. (see figure 14)

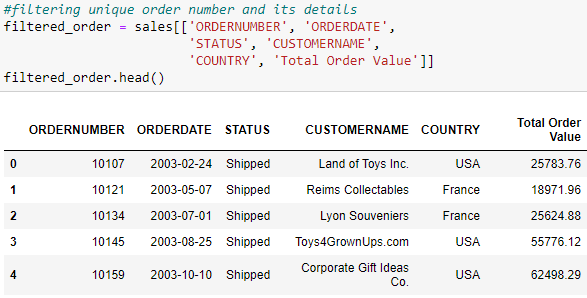


Figure 14

1. **Aggregate data**

I conducted an aggregation of the dataset based on order status to gain quick insights into the orders' statuses, including shipped, in process, disputed, on hold, resolved, and cancelled. To retrieve these details, I used the groupby() function on the 'STATUS' column, then applied the agg({}) function to perform the analysis on the 'SALES' column. (see figure 15)

A screenshot of a computer

Description automatically generated

Figure 15

Similarly, I conducted an aggregation of the dataset based on product lines to gain quick insights into the performance of the products, including Classic Cars, Motorcycles, Planes, Ships, Trains, Trucks and Buses, and Vintage Cars. To retrieve these details, I used the groupby() function on the 'PRODUCTLINE' column, then applied the agg({}) function to perform the analysis on the 'SALES' column. (see figure 16)

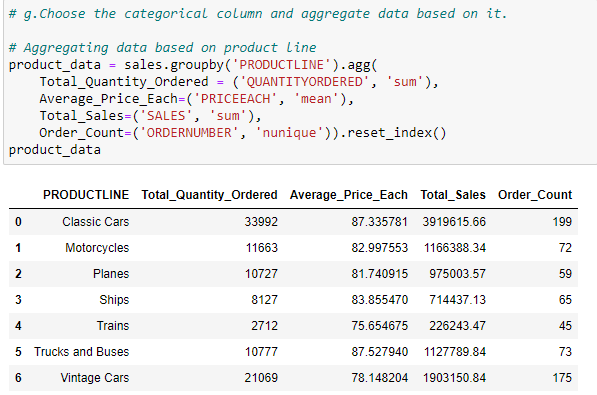


Figure 16

**Task B – Data Analysis**

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

We have 3 main categorical variable presents in our dataset, 1. Order Status, 2. Product line, 3. Country

1. For Order Status, I used the groupby() function to group the categorical variable 'STATUS' along with the column ‘Total Order Value’, and then generated summary statistics with the help of the ‘describe()’ function. (see figure 17)

A screenshot of a data

Description automatically generated

Figure 17

2. For Product line, I used the groupby() function to group the categorical variable 'PRODUCTLINE' along with the column ‘Total Order Value’, and then generated summary statistics with the help of the ‘describe()’ function. (see figure 18)

A screenshot of a data

Description automatically generated

Figure 18

3. For Country, I used the groupby() function to group the categorical variable 'COUNTRY' along with the column ‘Total Order Value’, and then generated summary statistics with the help of the ‘describe()’ function. (see figure 19)

A screenshot of a computer

Description automatically generated

Figure 19

1. **Correlations between different variables**

I have performed a correlation analysis between variables using the .corr( ) function, resulting in a table that shows the relationships among them (see table 1). Given the table contains numerous variables, it's challenging to quickly discern the strength of these relationships. To address this, I decided to visualize the correlation matrix using a heatmap from the Seaborn library. This visualization method effectively represents the relationships between variables, making it easier to identify strong correlations.

Here's an analysis based on the correlation between different variables. (see figure 20)

* SALES and PRICEEACH: 0.657841 suggests a strong positive relationship, indicating that as the price of each item increases, the sales value tends to increase as well.
* SALES and QUANTITYORDERED: 0.551426 also indicates a positive relationship, meaning higher quantities ordered tend to result in higher sales values.
* SALES and MSRP: 0.635239 is another strong positive correlation, suggesting that items with higher Manufacturer's Suggested Retail Price (MSRP) tend to generate higher sales.
* DEALSIZE and SALES: -0.862814 shows a strong negative relationship, indicating that higher sales values are associated with smaller deal sizes.

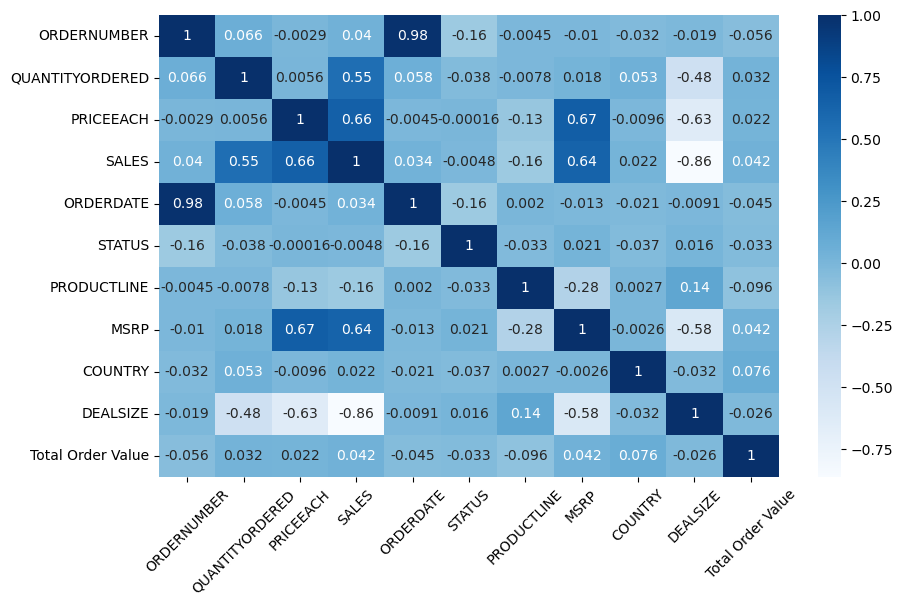


Figure 20

Table 1

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | ORDERNUMBER | QUANTITYORDERED | PRICEEACH | SALES | ORDERDATE | STATUS | PRODUCTLINE | MSRP | COUNTRY | DEALSIZE | Total Order Value |
| ORDERNUMBER | 1 | 0.06554 | -0.0029 | 0.03992 | 0.98305 | -0.159 | -0.0045 | -0.0103 | -0.0315 | -0.0185 | -0.056 |
| QUANTITYORDERED | 0.06554 | 1 | 0.00556 | 0.55143 | 0.05817 | -0.038 | -0.0078 | 0.01788 | 0.0531 | -0.4764 | 0.0322 |
| PRICEEACH | -0.0029 | 0.00556 | 1 | 0.65784 | -0.0045 | -0.0002 | -0.1308 | 0.67063 | -0.0096 | -0.6304 | 0.02234 |
| SALES | 0.03992 | 0.55143 | 0.65784 | 1 | 0.03357 | -0.0048 | -0.1604 | 0.63524 | 0.0224 | -0.8628 | 0.04174 |
| ORDERDATE | 0.98305 | 0.05817 | -0.0045 | 0.03357 | 1 | -0.1623 | 0.00203 | -0.0128 | -0.0209 | -0.0091 | -0.0453 |
| STATUS | -0.159 | -0.038 | -0.0002 | -0.0048 | -0.1623 | 1 | -0.0327 | 0.02137 | -0.0375 | 0.01611 | -0.0334 |
| PRODUCTLINE | -0.0045 | -0.0078 | -0.1308 | -0.1604 | 0.00203 | -0.0327 | 1 | -0.28 | 0.0027 | 0.13654 | -0.096 |
| MSRP | -0.0103 | 0.01788 | 0.67063 | 0.63524 | -0.0128 | 0.02137 | -0.28 | 1 | -0.0026 | -0.5802 | 0.04214 |
| COUNTRY | -0.0315 | 0.0531 | -0.0096 | 0.0224 | -0.0209 | -0.0375 | 0.0027 | -0.0026 | 1 | -0.0318 | 0.0764 |
| DEALSIZE | -0.0185 | -0.4764 | -0.6304 | -0.8628 | -0.0091 | 0.01611 | 0.13654 | -0.5802 | -0.0318 | 1 | -0.0258 |
| Total Order Value | -0.056 | 0.0322 | 0.02234 | 0.04174 | -0.0453 | -0.0334 | -0.096 | 0.04214 | 0.0764 | -0.0258 | 1 |

1. I have exported the cleaned dataset to OneDrive, please find the data location below. I used ‘to\_csv’ function to save the file in ‘csv’ format. (see figure 21)

'C:\\Users\\ajupe\\OneDrive - New Zealand Skills & Education Group\\GDDA7123C\\GD604\_Data Collection and Analysis\\Assessment-2'

A white background with colorful text

Description automatically generated

Figure 21

1. **Data analysis and visualization in Python**

I have performed data analysis and visualization in Python to gain rapid insights from a dataset by importing Matplotlib and Seaborn libraries for visualization purposes. Data visualization plays a crucial role in data analysis, as it provides a summarized view of the data, making it easier to understand and interpret without the need for extensive, detailed descriptions. To extract information from this large dataset, I employed various types of graphs, including Bar charts, Pie charts, Line plots, and Scatter plots. Each of these visualization techniques helps in revealing different aspects and patterns in the data, facilitating a more thorough analysis.

1. The bar chart and scatter plot illustrate the order value by country, clearly showing that most orders come from the USA, followed by Spain and France. It's also evident that most of the products are sold to countries in the USA and Europe. This analysis reveals that we have not yet initiated business activities in African and Asia-Pacific countries. (see figure 22&23)

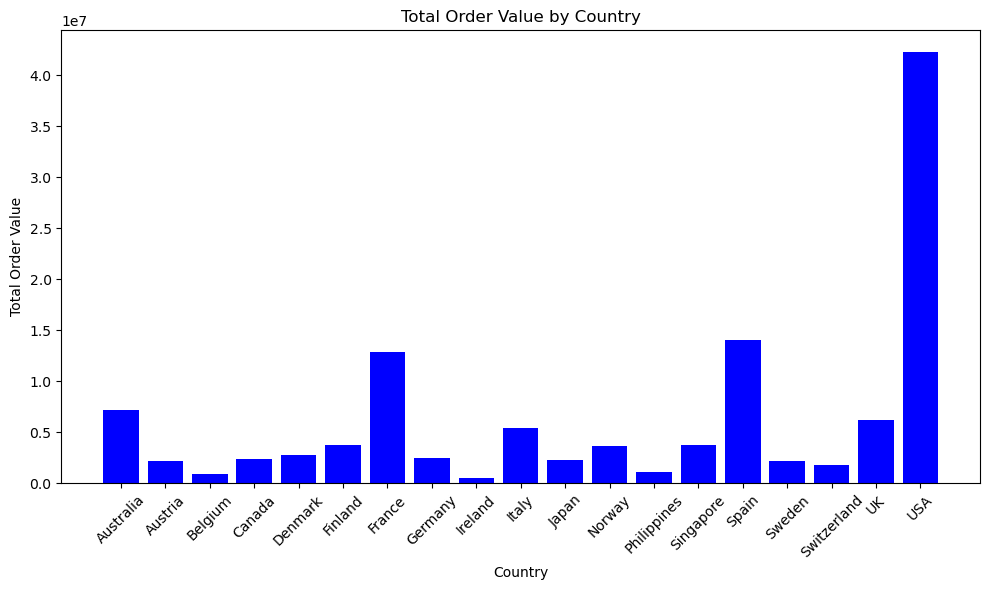


Figure 22

A graph of different countries/regions

Description automatically generated

Figure 23

1. The pie chart illustrates the distribution of product lines we manufacture and sell to customers. It reveals that we offer a total of 7 product categories: 1. Motorcycles, 2. Classic Cars, 3. Trucks and Buses, 4. Vintage Cars, 5. Trains, 6. Ships, 7. Planes. Analysis of the chart shows that Classic Cars account for 36.1% of sales, making it the largest category, followed by Vintage Cars at 19.8%, Motorcycles at 11.2%, and Planes at 10.9%. The remaining products collectively contribute less than 20% to total sales. Trains represent the smallest category, contributing only 2.9% to total sales. (see figure 24)

A pie chart with numbers and text

Description automatically generated

Figure 24

1. The bar chart displays the distribution of sales by deal size. It indicates that medium-sized orders contribute the most to overall sales, followed by small-sized orders. Large orders account for a smaller portion of total sales. Further analysis is needed to gain more insights into this trend. (see figure 25)

A graph of a bar chart

Description automatically generated

Figure 25

1. The line plot represents the monthly sales growth from January 2003 to May 2005. It clearly shows a year-over-year increase in sales, with a significant spike in orders during October and November of both 2003 and 2004. This pattern suggests the presence of seasonal trends affecting sales. (see figure 26)

A graph with blue lines and numbers

Description automatically generated

Figure 26

1. This bar chart provides essential information about order status, categorizing it into 6 groups: 1. Shipped, 2. In Process, 3. On Hold, 4. Disputed, 5. Resolved, 6. Cancelled. The graph clearly shows that the company's order management is effective, with 92.6% of the total order value being shipped. Only 2.48% of orders are cancelled, which is reasonable over three years. However, further analysis is necessary to understand the reasons behind order cancellations. (see figure 27)

A blue bar graph with white background

Description automatically generated

Figure 27

1. I have conducted a more detailed study on cancelled orders and arranged the list of cancellations month-wise. The results show that in October 2003, there were a total of 16 order cancellations, and in May and June 2004, there were 44 cancelled orders. However, the positive aspect is that we haven't identified any cancellations till May 2005, which is a good sign. These cancellations could have occurred due to several reasons, further investigation required to find the root causes. (see figure 28)

ORDERDATE

2003-10-31 16

2004-05-31 14

2004-06-30 30

A graph with red line

Description automatically generated

Figure 28

1. **Infernal statical**

1. I conducted a chi-square test to investigate the relationship between Country and Deal Size. For this analysis, I utilized the chi2\_contingency function from the scipy.stats library and generated a contingency table using pd.crosstab() for the 'COUNTRY' and 'DEALSIZE' columns. The interpretation of the Chi-square Value and P-value allowed us to determine the significance level (alpha), based on which we concluded the hypothesis test.

The results are as follows:

* Chi-square Value: 32.244780506142035
* P-value: 0.6478727539969273
* Significance Level (alpha): 0.05

Based on these results, we fail to reject the null hypothesis. This indicates that there is no significant relationship between Country and Deal Size. (see figure 29)

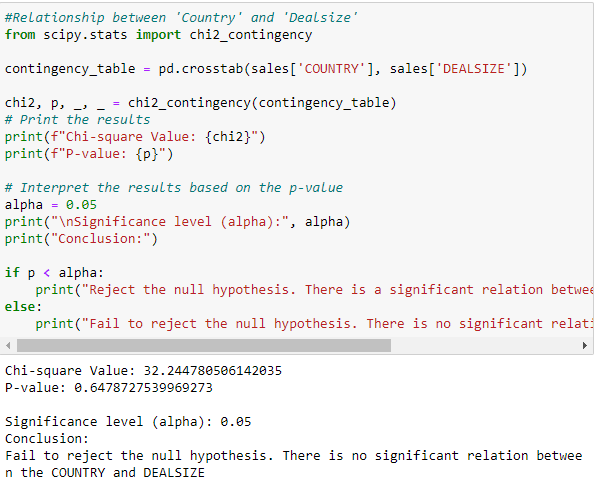


Figure 29

2. I conducted a T-statistics test to investigate the relationship between mean sales value of cancelled orders compared to the overall mean. For this analysis, I utilized the stats.ttest\_1samp( ) function from the scipy.stats. The interpretation of the T-statistics test Value and P-value allowed us to determine the significance level (alpha), based on which we concluded the hypothesis test.

* T-statistic: nan,
* P-value: nan

Based on these results, we fail to reject the null hypothesis. There is no significant difference in the mean sales value of cancelled orders compared to the overall mean. (see figure 30)

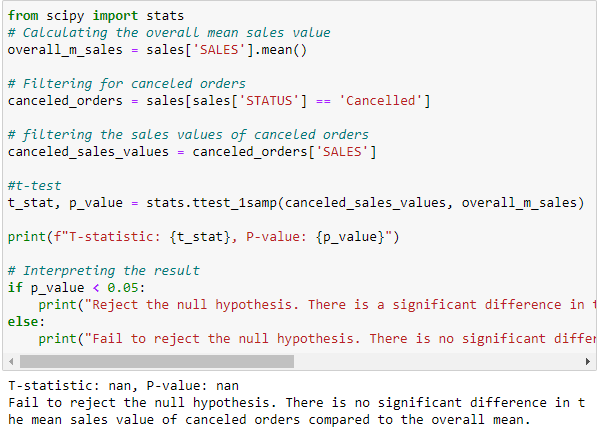


Figure 30

3. I conducted a ANNOVA test to investigate the mean sales of different product lines. For this analysis, I utilized the stats.f\_oneway(\* )function from the scipy.stats. The interpretation of the F-statistic Value and P-value allowed us to determine the significance level (alpha), based on which we concluded the hypothesis test.

* F-statistic: 25.23
* P-value: < 0.01

Based on these results, we reject the null hypothesis. This indicates that there is a significant difference in the mean sales among different product lines. (see figure 31)

A screenshot of a computer

Description automatically generated

Figure 31

**Task C – Data Findings and Decision Support**

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

Grouping based on categorical variable:

* From the analysis of summary statistics on categorical variable ’STATUS’, it reveals insights such as the total number of shipped orders being 2,617 and cancelled orders amounting to 60, indicating that cancellations are significantly low in comparison. Additionally, the analysis revealed that the average value of shipped orders is $41,736, while the average value for orders in process stands at $34,647. This information is critical as it not only highlights the efficiency and success rate of order fulfilment but also helps in understanding the financial impact of orders at different stages of the sales process. (see figure 32)

A screenshot of a data

Description automatically generated

Figure 32

* From the analysis of summary statistics on categorical variable ’PRODUCTLINE’, it reveals valuable insights such as the order count for each item, which revealed that classic cars are the most demanded items with an order count of 967, while trains have the lowest demand with order count only 77. Additionally, the analysis showed that the average order value is highest for classic cars and considerably lower for trains. This information is crucial as it not only highlights consumer preferences within our product range but also indicates potential areas for revenue optimization by focusing on high-demand, high-value items and reevaluating the strategy for lower-demand products. (see figure 33)

A screenshot of a data

Description automatically generated

Figure 33

* From the analysis of summary statistics on categorical variable ’COUNTRY’, it revealed that the highest number of orders was received from the USA, total 1,004, while the lowest order count came from Ireland, with only 16 orders. This analysis helps us identify that the USA and Europe are the major markets for our products, indicating where our customer base is most concentrated. Understanding the geographic distribution of our orders is crucial for tailoring our marketing strategies and inventory distribution to cater to the demands of these key regions more effectively. (see figure 34)

A screenshot of a computer

Description automatically generated

Figure 34

Correlation Analysis:

Here's an analysis based on the correlation between different variables.

* SALES and PRICEEACH: 0.657841 suggests a strong positive relationship, indicating that as the price of each item increases, the sales value tends to increase as well.
* SALES and QUANTITYORDERED: 0.551426 also indicates a positive relationship, meaning higher quantities ordered tend to result in higher sales values.
* SALES and MSRP: 0.635239 is another strong positive correlation, suggesting that items with higher Manufacturer's Suggested Retail Price (MSRP) tend to generate higher sales.
* DEALSIZE and SALES: -0.862814 shows a strong negative relationship, indicating that higher sales values are associated with smaller deal sizes.

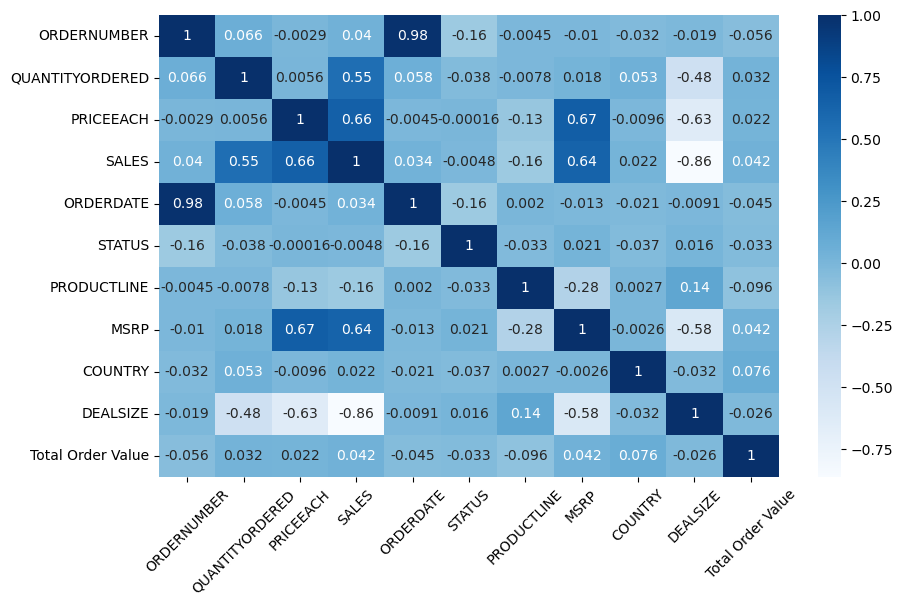
Inferential analysis summary:

* Based on the ANOVA test results, we reject the null hypothesis. This indicates that there is a significant difference in the mean sales among different product lines.
* Based on the T-test results, we fail to reject the null hypothesis. There is no significant difference in the mean sales value of cancelled orders compared to the overall mean.
* Based on the Chi2 test results, we fail to reject the null hypothesis. This indicates that there is no significant relationship between Country and Deal Size.

**b. Relationships between variables, summarize key findings, and identify significant trends or patterns.**

Correlation Analysis: Here's an analysis based on the correlation between different variables.

* SALES and PRICEEACH: 0.657841 suggests a strong positive relationship, indicating that as the price of each item increases, the sales value tends to increase as well.
* SALES and QUANTITYORDERED: 0.551426 also indicates a positive relationship, meaning higher quantities ordered tend to result in higher sales values.
* SALES and MSRP: 0.635239 is another strong positive correlation, suggesting that items with higher Manufacturer's Suggested Retail Price (MSRP) tend to generate higher sales.
* DEALSIZE and SALES: -0.862814 shows a strong negative relationship, indicating that higher sales values are associated with smaller deal sizes.



Trend and Pattern Analysis:

1. From the monthly sales growth from January 2003 to May 2005. It clearly visible the trend that a year-over-year increase in sales, with a significant spike in orders during October and November of both 2003 and 2004. This pattern suggests the presence of seasonal trends affecting sales.

A graph with blue lines and numbers

Description automatically generated

1. We can reveal a clear trend that most orders come from the USA, followed by Spain and France. It's also evident that most of the products are sold to countries in the USA and Europe.

A graph of blue rectangular bars with white text

Description automatically generated

1. Product category patterns reveals that Classic Cars account for 36.1% of sales, making it the largest category, followed by Vintage Cars at 19.8%, Motorcycles at 11.2%, and Planes at 10.9%. The remaining products collectively contribute less than 20% to total sales. Trains represent the smallest category, contributing only 2.9% to total sales.

A pie chart with numbers and text

Description automatically generated

1. Pattern in distribution of sales by deal size reveals that medium-sized orders contribute the most to overall sales, followed by small-sized orders. Large orders account for a smaller portion of total sales.

A graph of a bar chart

Description automatically generated

Summary statistics:

Summary statistics on order status revealed a high efficiency in order fulfilment, with 2,617 orders shipped and only 60 cancellations, highlighting a low cancellation rate. The average value of shipped orders at $41,736, with orders in process averaging $34,647, indicating the financial impact of orders at various stages.

Analysis by product line showed classic cars as the top demanded item with an order count of 967, while trains had the lowest demand with order count of 77. The average order value was highest for classic cars, suggesting a focus on high-demand, high-value items could optimize revenue.

Country-wise statistics indicated the USA as the leading market with 1,004 orders, and Ireland with the fewest at 16 orders, underscoring the USA and Europe as primary markets. This information is vital for refining marketing strategies and inventory distribution to meet key regional demands more effectively.

Key findings:

* The business exhibits year-over-year growth.
* The order handling team demonstrates efficiency, evidenced by a low order cancellation rate of only 2%.
* The USA emerges as a strategic and key potential market for our products.
* Small and medium deal sizes are most prevalent.
* Classic Cars and Vintage Cars are the top-selling products, highlighting their demand in the market.
* The product category of Trains shows the least popularity among our offerings.
* There exist untapped markets for our products, especially in African, Middle Eastern, and Pacific Asian countries, suggesting opportunities for expansion.

**c. Business Challenges and Suggestions**

1. Based on the analysis, it's evident that a significant portion of orders originates from the USA, followed by Spain and France. This pattern indicates that the bulk of product sales are concentrated in the USA and various European countries.

However, this heavy reliance on a single market, particularly the USA, presents potential business challenges. Such dependence is risky due to potential future uncertainties like economic recessions, pandemics, natural disasters, and increased competition from similar manufacturers.

To address these concerns, the following recommendations are proposed:

Strengthening Presence in the USA: It's crucial to maintain and amplify sales momentum in the USA through active sales engagement. Implementing sales promotions and adopting diverse marketing strategies can enhance market penetration and customer base in this region.

Expanding to New Markets: To reduce risks associated with market concentration and secure future sales growth, a strategic focus on markets with lower sales volumes is recommended. Specifically, efforts should be directed towards expanding business operations into untapped markets, especially in Pacific Asian countries, African regions, and Middle Eastern countries.

By following these recommendations, the business can diversify its market presence, reduce dependency on a single market, and position itself for sustained growth amidst global uncertainties.

2. Our product portfolio having seven distinct categories: Motorcycles, Classic Cars, Trucks and Buses, Vintage Cars, Trains, Ships, and Planes. From the analysis of sales data reveals that Classic Cars are our top-selling category, contributing 36.1% of total sales, followed by Vintage Cars at 19.8%, Motorcycles at 11.2%, and Planes at 10.9%. The remaining categories together account for less than 22% of overall sales, with Trains being the least popular, contributing a mere 2.9%.

Based on these insights, the following strategic recommendations are proposed:

Focus on Leading Categories: Priority should be given to our best-selling categories, Classic and Vintage Cars. By forecasting future sales based on historical data, we can better manage inventory levels to meet demand promptly, potentially boosting sales further.

Promotional Strategies for Mid-tier Products: For products with moderate sales, such as Planes, Ships, Trucks, and Buses, implementing promotional campaigns or offering discounts could stimulate interest and increase sales figures.

Reevaluating Low-performing Categories: A closer examination of the low sales volumes for Trains is needed to understand the root causes. Depending on the findings, it may be advised to adjust inventory levels to a demand-driven model or even consider discontinuing the production of Trains. Resources could then be reallocated towards more profitable product lines, enhancing overall business performance.