**University of Central Lancashire**

**UCLan Cyprus**

**School: School of Sciences**

Year of Study: 2024-2025

Assessment Title: Data Analysis for RealPOS company

**Course Title: MSc. Data Analytics**

**Module Title: Enterprise Data Management**

**School: School of Sciences**

**Data Analysis for RealPOS**

**Executive Summary**

*This report thoroughly analyzes sales and customer demographics for RealPOS, a Brazilian retail company specializing in outdoor equipment and accessories. The analysis draws on various datasets, including the company's Invoice Table, Customer Table, Basket Table, Product Table, Supplier Table, Product Origin Table, and Promotion Table. These datasets were processed using SAS Enterprise Guide, a powerful analytical tool, to uncover actionable insights. The insights derived from this analysis are not merely data points; they represent critical understandings that empower stakeholders with knowledge about RealPOS's sales performance and customer demographics. This comprehensive examination enables RealPOS to grasp the nuances of its business landscape, fostering a sense of confidence among the audience regarding the company's operations and strategies. We gained valuable insights into our customers and their purchasing behaviors through a thorough analysis. Firstly, we calculated the total number of items and the total monetary value associated with each invoice ID, providing a clear picture of sales performance. A significant portion of our customer base resides in the vibrant São Paulo (SP) area, while a slightly smaller proportion comes from the bustling Rio de Janeiro (RJ) area. This geographical distribution helps us understand where our market is strongest. Most of our customers are male and predominantly located in São Paulo. One intriguing finding is that a large segment of our clientele falls within the 'middle-aged' category. Meanwhile, we have a small number of customers under 18 years old, with very few classified as 'very young.' Our basket analysis indicates that certain products within the same categories tend to be purchased together, contributing significantly to overall revenue. Based on product ranking, eyewear is our top-selling product, followed closely by watches in second place. Behavioral analysis shows that the majority of our customers are middle-aged, with very few customers over the age of 65. The promotion analysis reveals that products sold during promotions tend to have higher sales than those sold without promotions. Additionally, our daily sales analysis indicates that total product sales, recorded by invoice date, are higher on weekends than on weekdays. Regarding supplier performance, revenue from Supplier ID 5 constitutes a significant percentage of our total revenue, while revenue from Supplier ID 3 is considerably lower. This suggests that Supplier ID 5 is our top supplier. The leading supplier is Dragon SA, a company based in the United States.*

**Table of Contents**

1. Data Pre-processing

1.1 Import Raw Data

1.2 Create Project Library

1. Sales vs. Returns
   1. Customer Age Analysis
   2. Recoding
2. Basic Analysis
   1. Invoice Total Items
   2. Invoice Total Value
   3. Region Analysis
   4. Basket Analysis
   5. Demographic Analysis
3. Advanced Analysis
   1. Product Rankings
   2. Behavioral Characteristics
   3. Promotions Analysis
   4. Daily Sales Analysis
   5. Supplier Analysis

**1. Data Pre-processing**

**1.1 Import Raw Data**

All raw files were carefully imported into SAS, creating distinct datasets for each individual file. To enhance the accuracy of the data transformation process, variables not intended for statistical analysis—such as SKU and BasketID—were explicitly configured to be read as string types. This strategic decision was made to reduce the likelihood of errors during data handling. Additionally, the datasets were organized and stored in the SASUSER library, providing a structured framework that supports efficient and comprehensive analyses in the subsequent stages of the research process.

**1.2 Create Project Library**

A project library, designated as SASMS, has been established to facilitate effective data management. This library contains all prepared datasets resulting from the raw data import phase. It is the foundational resource for all subsequent analyses, ensuring a structured approach to our data-driven initiatives.

**1.3 Sales vs. Returns**

The Invoice table was divided into two separate tables: one for Sales and another for Returns, based on the Operation variable. A bar chart illustrated the monetary values of both sales and returns, providing valuable insights into the frequency counts compared to the Operation.

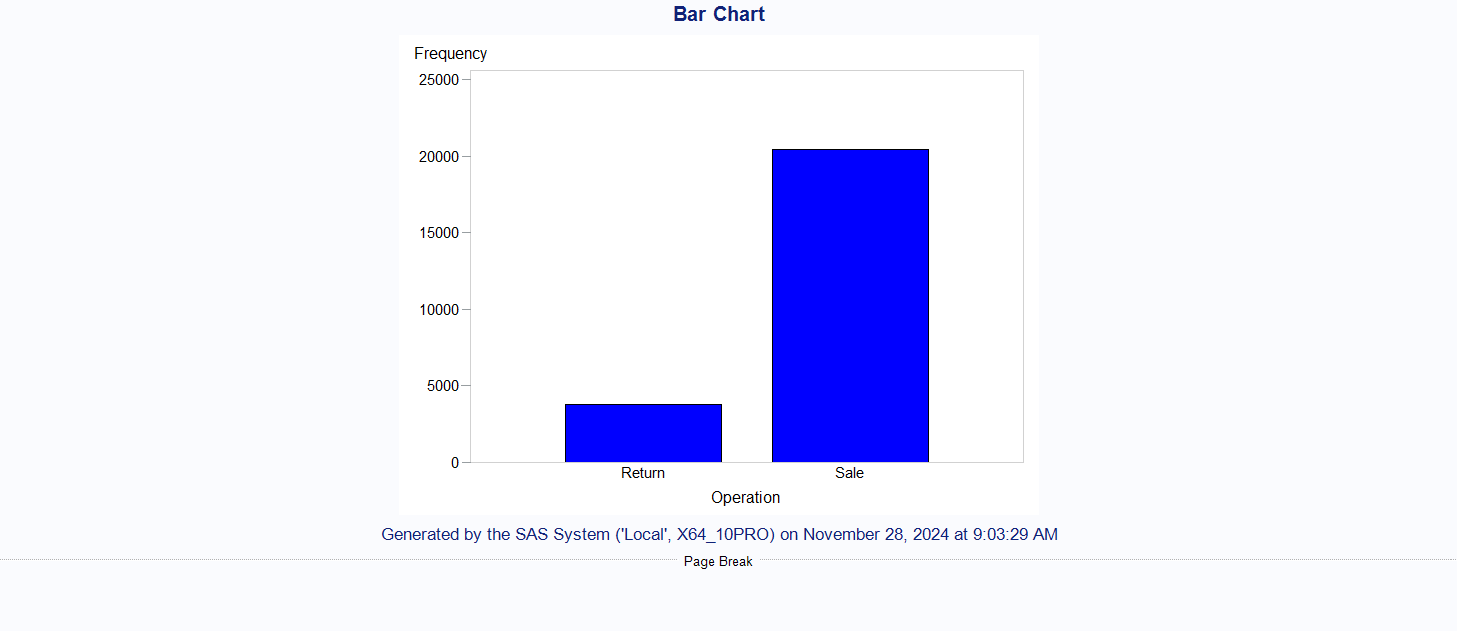


Figure 1: Frequency Count Vs. Operation

Figures 2 and 3 also show the frequency counts for sales and returns, with sales accounting for the highest percentage at 84.33%, and counts totaling 20419. Furthermore, Figures 4 present a bar chart depicting sales and returns by Invoice ID. Notably, every Invoice ID contains both sales and return values, highlighting the significance of our findings.

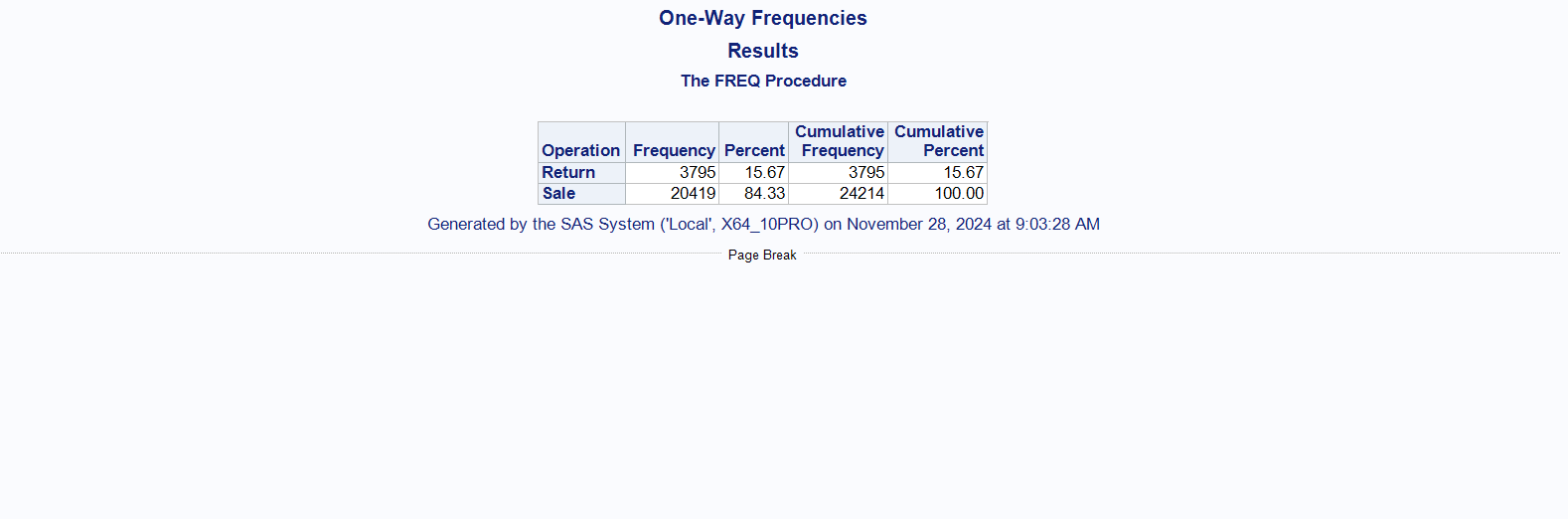


Figure 2: One way Frequency Results

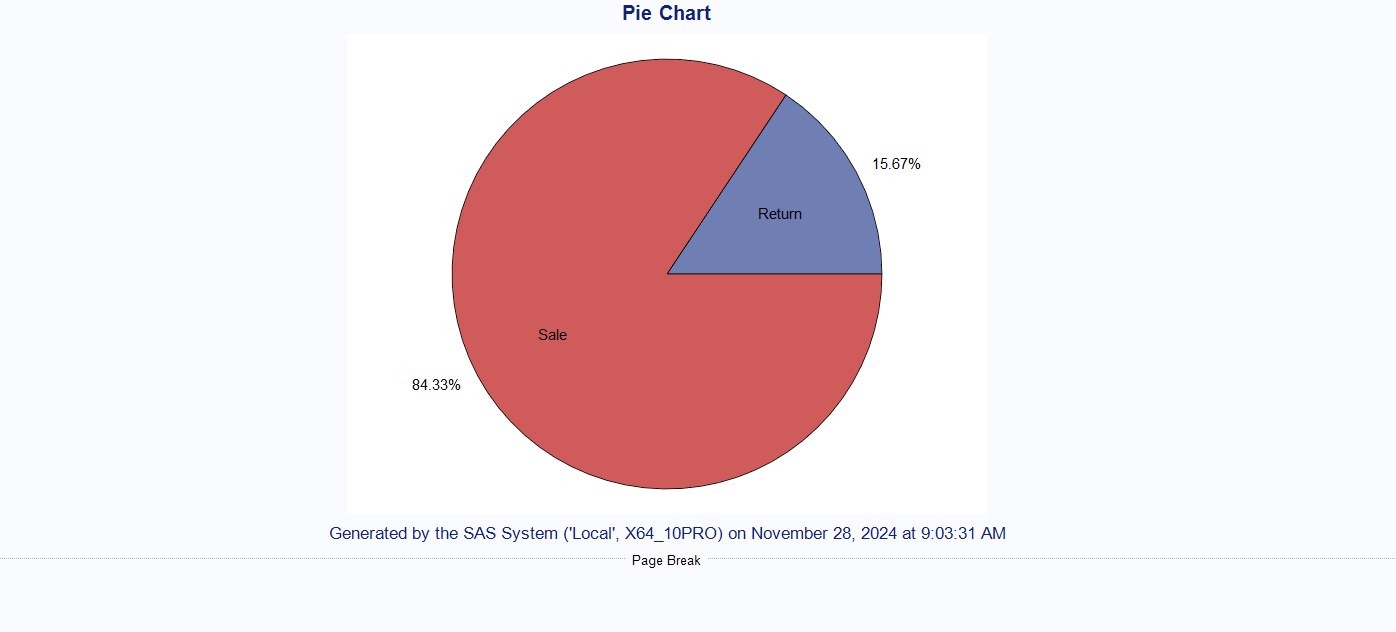


Figure 3: Percentage of sales and returns

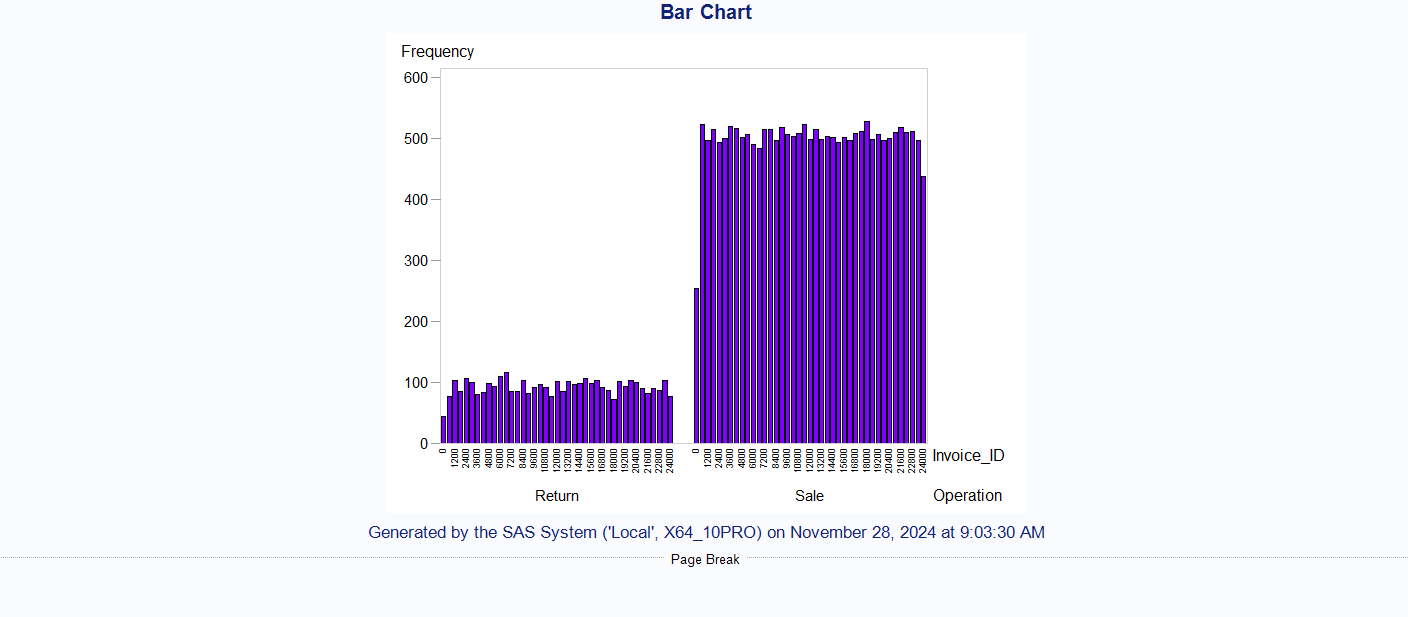


Figure 4: Frequency Count by per Invoice ID.

**1.4 Customer Age Analysis**

To enhance our understanding of customer demographics, we calculated customer ages using January 1, 2019, as the reference date and created a new integer variable labeled "Customer Age." Following this, I formatted the date of birth (DOB) for all customers and used the appropriate calculation method to obtain this information from the Customers Table. Additionally, I generated a report to identify underage customers (those under 18 years old) stored in the Customer\_Age\_under18 table to ensure compliance with GDPR. This approach enhances customer data management and underscores our commitment to data protection and regulatory compliance.

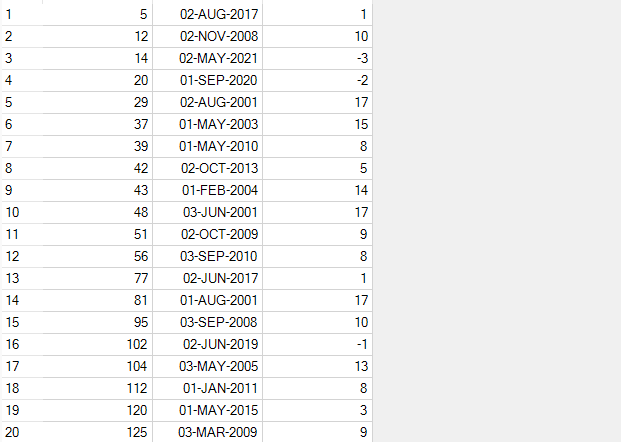


Table 1: First 20 Rows of Customer\_Age\_Under18 Data

**1.5 Recoding**

Demographic variables were recoded:

**Age Group Variable**: A new variable called Age\_Group was created based on the calculated customer age, with classifications ranging from "Under 18" to "Very Senior."



Table 2: First 20 Rows of Age Group Data

**Supplier Code Extraction**: The supplier code was extracted from the ninth digit of the SKU in the Product Table and stored in a new column labeled Supplier\_ID.

.**2. Basic Analysis**

**2.1 Invoice Total Items**

I generated a report that details the total number of items per invoice, offering insights into purchasing patterns. To create this report, I joined four tables: Customers, Invoices, Payment Methods, and Basket. I then calculated the sum of the quantity, grouping the data by Invoice ID, and stored the results in the Total Items Table.

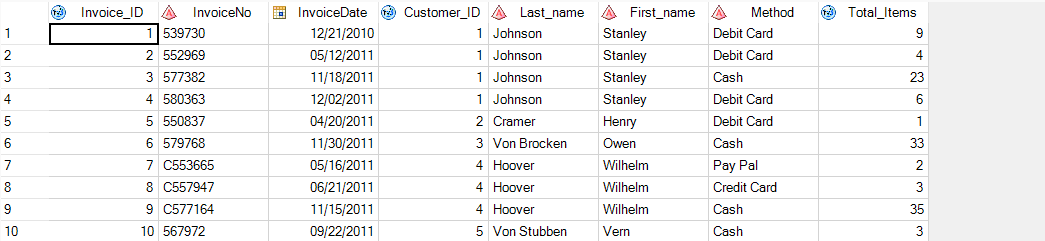


Table 3: Total Items of First 10 Invoice ID.

Figure 5 displays the total number of items linked to each invoice ID. Notably, invoice ID 2400 has the highest count, totaling 46 items. Additionally, it's important to mention that every invoice ID is associated with a minimum quantity of products, ensuring that no invoice is left without any items.

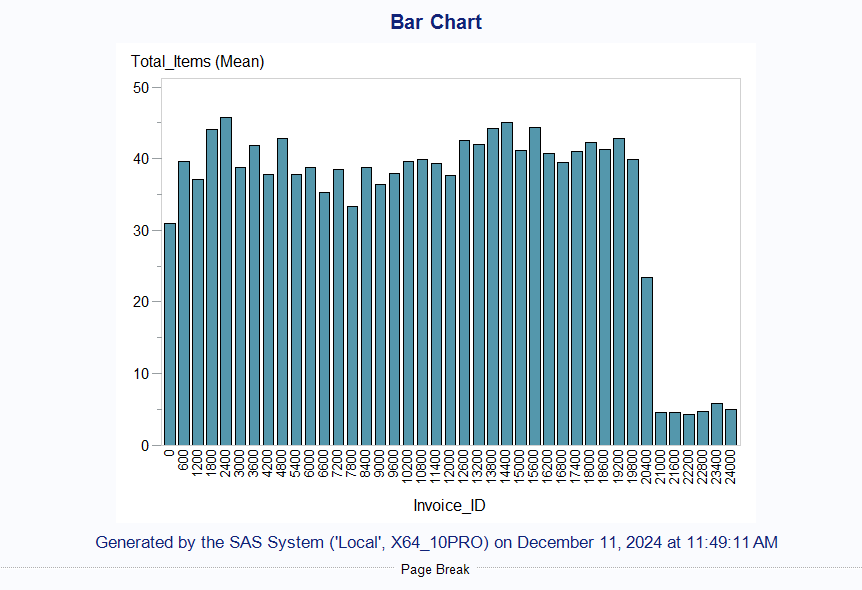


Figure 5: Total Items per Invoice ID

**2.2 Invoice Total Value**

The total value for each invoice was calculated by applying price discounts derived from promotional datasets. To arrive at this total value, I integrated four tables: Products, Invoice, Promotions, and Basket. Subsequently, I calculated the discounted prices by applying the promotions to the product prices recorded in the Discounted\_Prices table. Following this, I computed the total values using the SUM function, multiplying the quantity by the discounted prices while grouping the results by Invoice ID, as documented in the Total Values table. Finally, I joined the Customer and Payment Method tables with the Total Items table to present the Customer Name and Payment Method.

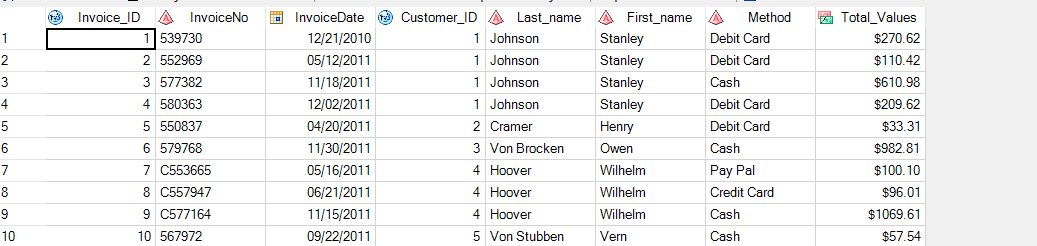


Table 4: Total Values of First 10 Invoice ID.

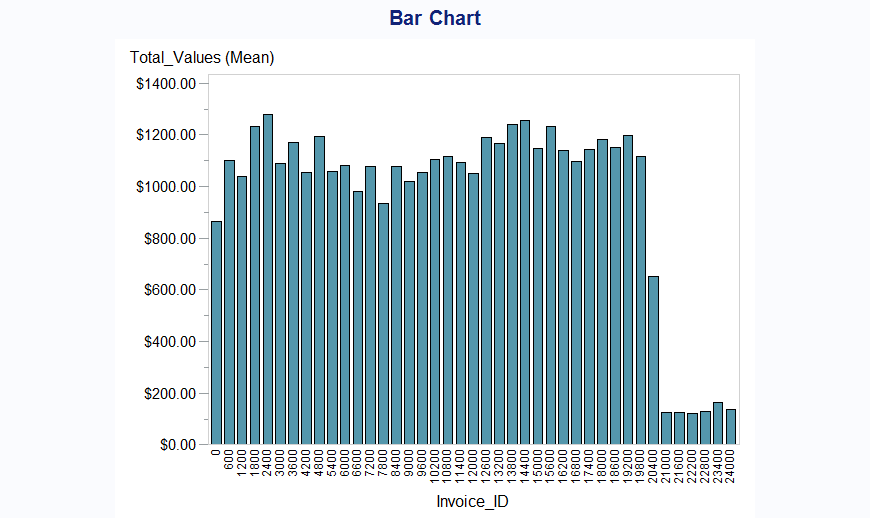


Figure 6: Total Values per Invoice ID

Figure 6 presents the total values associated with each Invoice ID, with a particular focus on Invoice ID 2400. It is important to note that the range of Invoice IDs from 21000 to 24000 reveals a limited number of values. This suggests that sales during this period have been relatively low, highlighting a potential area of concern for revenue generation within this specific range of Invoice IDs.

**2.3 Region Analysis**

To analyze the region, I combined the Total Values and Customers Table. I calculated the company's revenues by using the SUM function on the Total Values, grouped by region and gender, respectively.

Figure 7 shows a bar chart illustrating the top region based on company revenues, with SP (Sao Paulo) identified as the leading region. Figure 6 displays the percentage for the top region, where SP accounts for 40.29%.

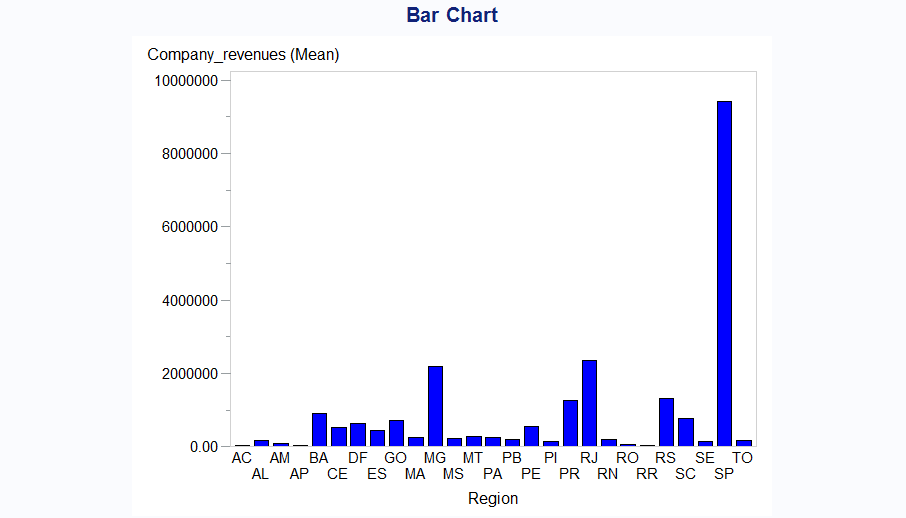


Figure 7: Company Revenues by Top Region

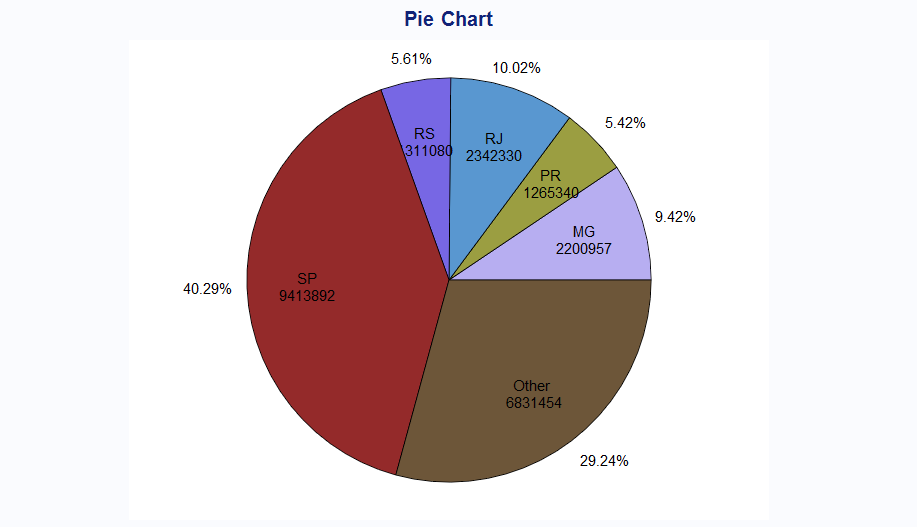


Figure 8: Proportion of Company Revenues and Top Region

Figures 9 and 10 also highlight the top region according to company revenues by gender, again with SP as the leading area. It is noteworthy that male customers generate the highest revenue, particularly in the SP area. This insight suggests potential opportunities for targeted marketing strategies in that region to further enhance engagement with this demographic.

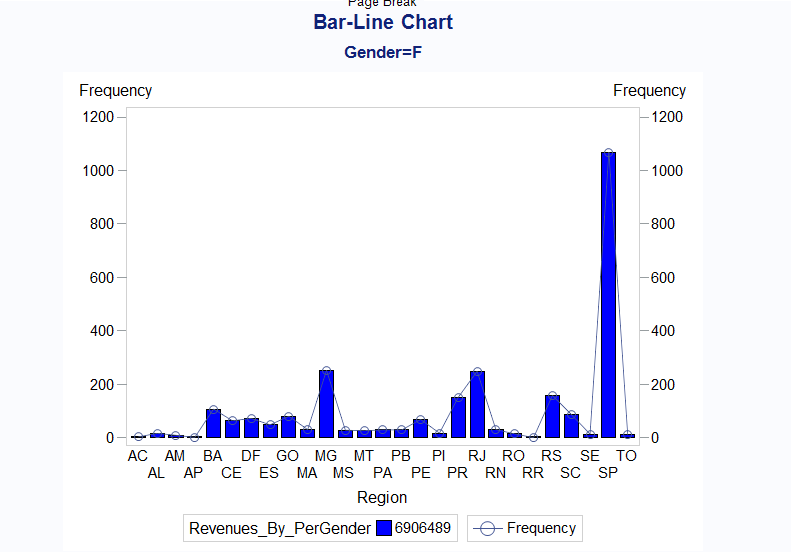
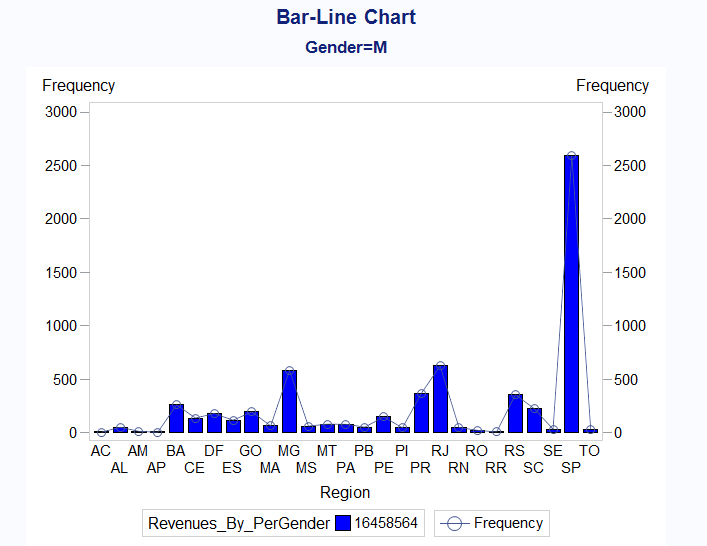


Figure 9: Revenues by Gender vs. Region

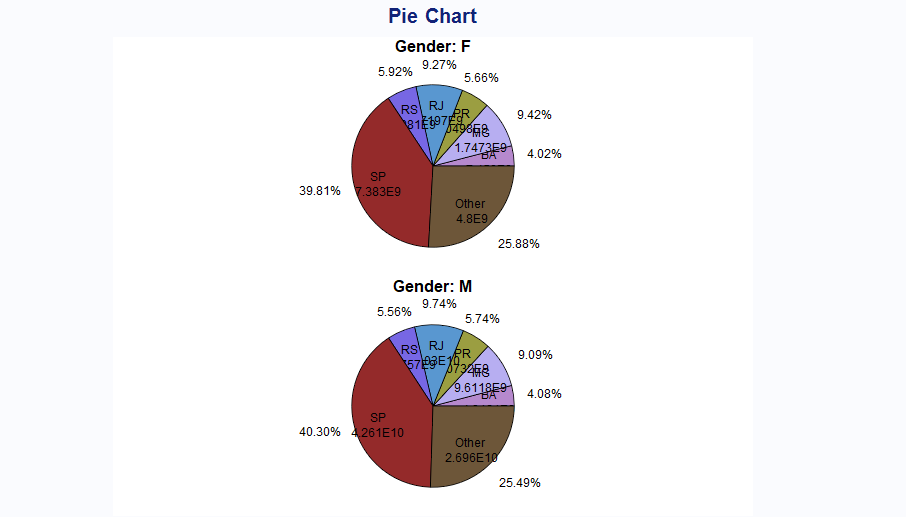
****

Figure 10: Proportion of Revenue by Gender and Region

**2.4 Basket Analysis**

In the basket analysis, I counted the total number of SKUs using the COUNT(DISTINCT) function per product, grouped by SKU from the PRODUCT\_TOTAL\_VALUES table. Then, I calculated the average basket value by dividing the total values by the SKUs. The summary statistics reveal insightful details about the products in question. Each product features a minimum of three stock-keeping units (SKUs), while certain products boast as many as 24 SKUs. This variation highlights the diverse range of items typically sold together. Moreover, there is a notable correlation between the number of SKUs and the average basket value; as the quantity of SKUs increases, so does the average value of the shopping basket, suggesting that customers tend to spend more when purchasing a greater variety of products.

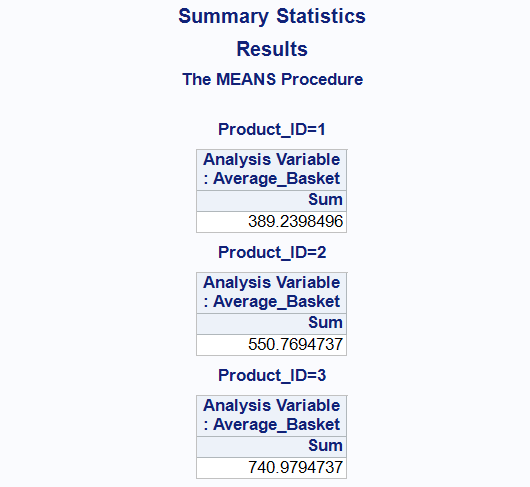


Figure 11: Average Monetary value by Product ID

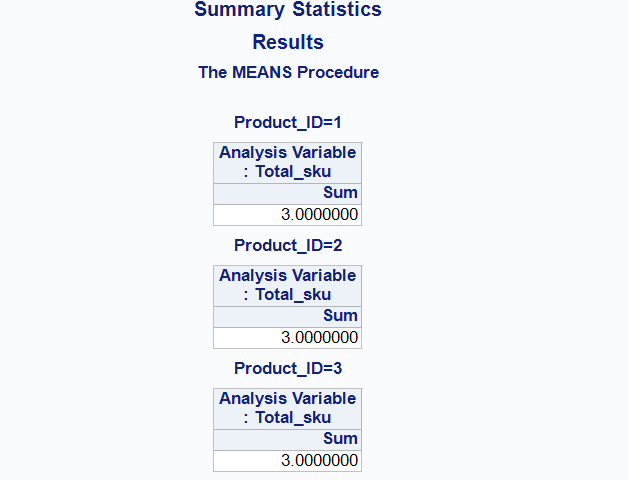
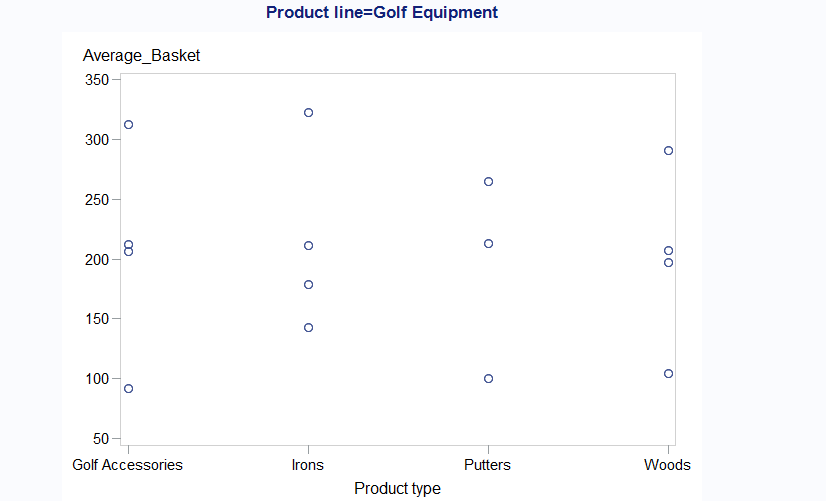


Figure 12: Average SKU by Product ID

Figure thirteen provides a detailed illustration of the relationship between Product Type and Product Line, analyzed through the lens of Average Basket value. The data reveals that Golf Equipment is significantly linked to various golf-related items, including accessories, irons, and woods, indicating a strong preference among consumers for purchasing these products together.

Similarly, the category of Personal Accessories encompasses items such as eyewear, knives, and watches, suggesting that shoppers tend to buy these complementary products in tandem. Furthermore, Camping Equipment and Mountaineering Equipment show a connection with comparable Cooking gear, Climbing accessories the idea that these related products are frequently purchased at the same time. Overall, the findings indicate that consumers often seek out and buy items that are associated with one another, enhancing their overall shopping experience.

****

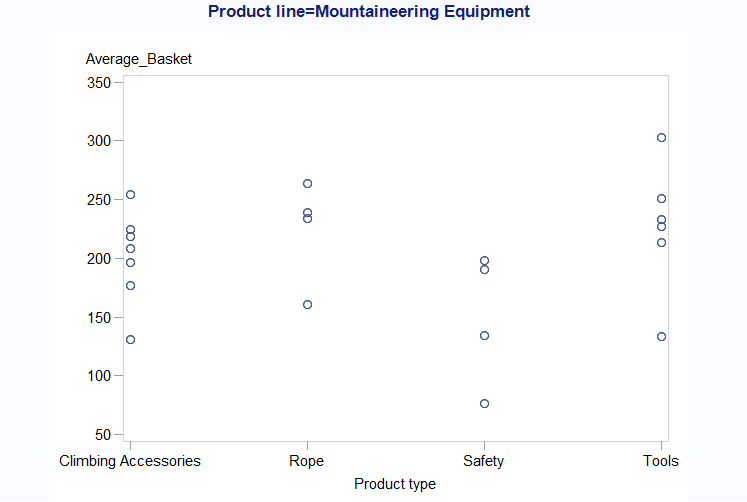
****

Figure 13: Correlation Between Product line and Type by Average Basket value

**2.5 Demographic Analysis**

In this section, We created a new age group stored in the AGE\_GROUP\_2 table, derived from the CUSTOMER\_AGE\_CUSTOMER table. A frequency table was developed based on customer age, with a maximum age range of 20-50 years. We have very few customers over the age of 70.

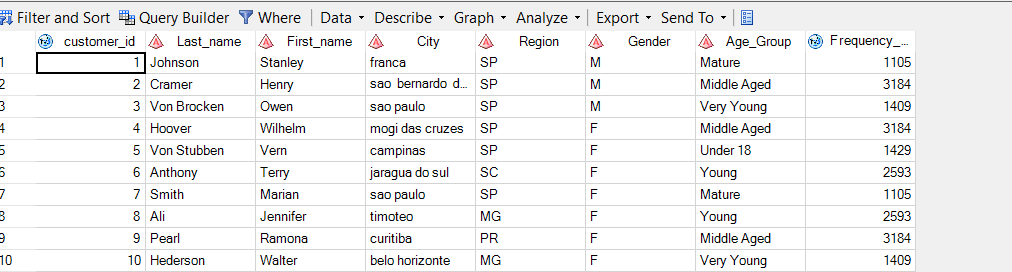


Table 5: Frequency count by Age Group.

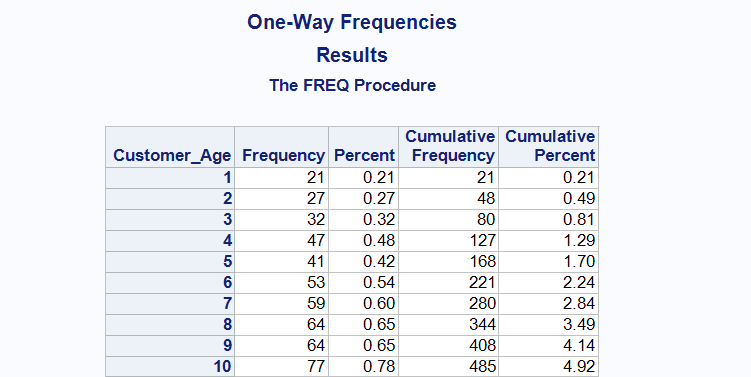


Table 6: One way frequency Table.

Analysis of Figure 14 reveals that the predominant customer demographic is the 'Middle Aged' category, representing a significant 32.29% of the overall customer base. This indicates a strong inclination towards middle-aged individuals engaging with the product or service. In stark contrast, the proportion of customers under the age of 18 is notably minimal, at just 14.49%. This suggests that younger audiences are less represented among the customer population. Furthermore, the 'Very Young' demographic shows an even smaller presence, highlighting their limited participation relative to the other age groups.

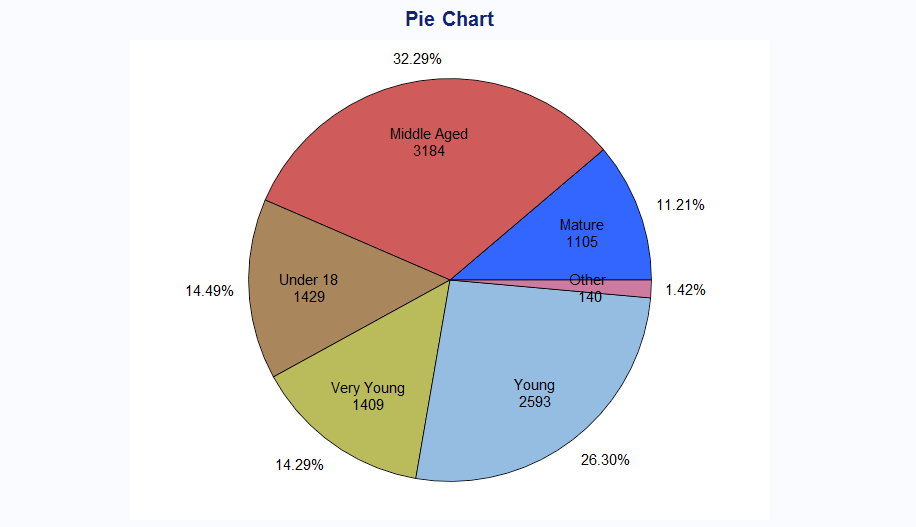


Figure 14: Percentage of frequency by Age Group.

Figure 15 indicates that the highest number of customers resides in the SP area, while the BA area has the lowest percentage at 4.08%. Additionally, male customers represent 70.34% of the total, while female customers make up about one-third described in Figure 16.

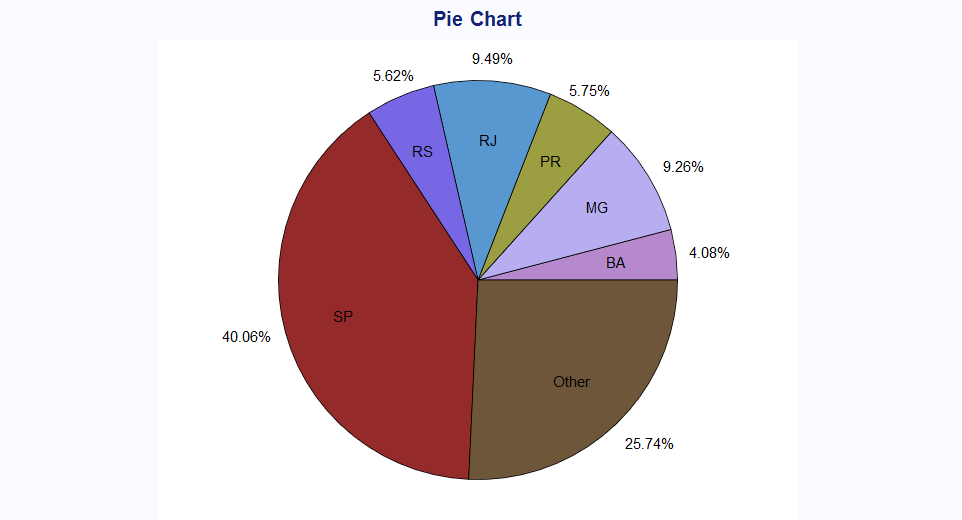


Figure 15: Percentage of Customer by region

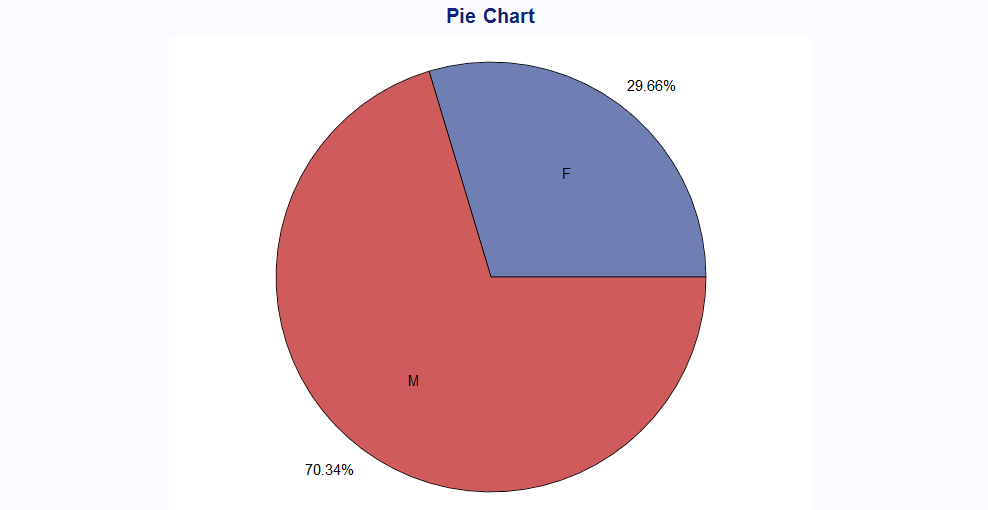


Figure 16: Percentage of Customer by Gender

**Recommendation for Basic Analysis:**

Through a thorough analysis, we can uncover a variety of valuable insights about our customers and their purchasing behaviors. Firstly, we have calculated the total number of items and the total monetary value associated with each Invoice ID, providing a clear picture of sales performance.

Our regional analysis reveals that a significant portion of our customer base resides in the vibrant São Paulo (SP) area, with a slightly smaller proportion from the bustling Rio de Janeiro (RJ) area. This geographical distribution helps us understand where our market is strongest.

Diving deeper into demographics, we find that the majority of our customers are male and predominantly located in São Paulo. One of the most intriguing findings is that a large segment of our clientele falls within the 'middle-aged' category. Meanwhile, we do have a small number of customers who are under 18 years old, and very few who can be classified as 'very young.'

Furthermore, our basket analysis indicates that certain products within the same categories tend to be purchased together, contributing significantly to overall revenue. This insight allows us to identify cross-selling opportunities. Moving forward, our advanced analysis will provide a spotlight on the top-selling products, helping us to understand market trends and customer preferences in greater detail.

**3. Advanced Analysis**

**3.1 Product Rankings**

In this analysis of product rankings, We focused on identifying the Top Product Types by calculating the total sum of product prices. This data was organized and sorted by product type, as documented in the TOP\_PRODUCT\_TYPE table. To enhance the depth of this analysis, We employed the RANK function to assign rankings to each product type based on their total sales.

The findings are depicted in Figure 17, which reveals that Eyewear is the leading product type, making up 11.35% of total sales. Watches follows as a close second in the rankings.

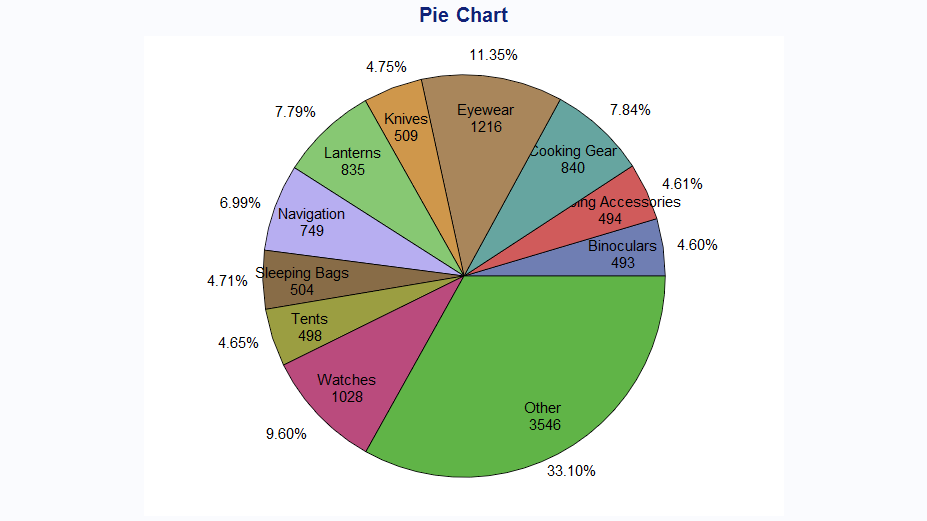


Figure 17: Percentage of Top Product Type by Product Ranking

To provide further clarity, Figure 18 illustrates the ranking system, showing that Eyewear attained the highest rank of 90, indicating its dominant position in the market.

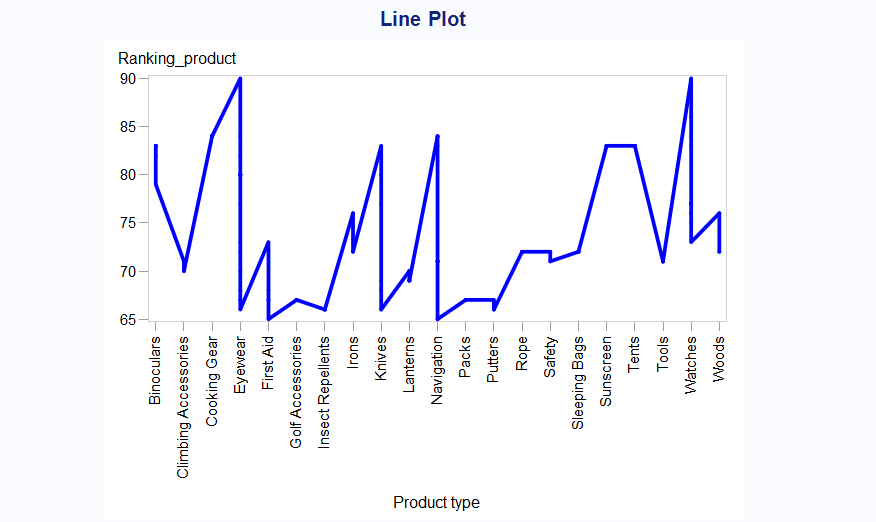


Figure 18: Top Product Type by Product Ranking

Figure 19 complements this analysis by portraying the proportion of the top product type in comparison to total sales, with Eyewear again standing out as the frontrunner. Furthermore, Figure 20 delves deeper into the data by presenting the percentage of the top product type calculated from subtotal sales, reinforcing the notion that Eyewear is the most significant product category in this analysis.

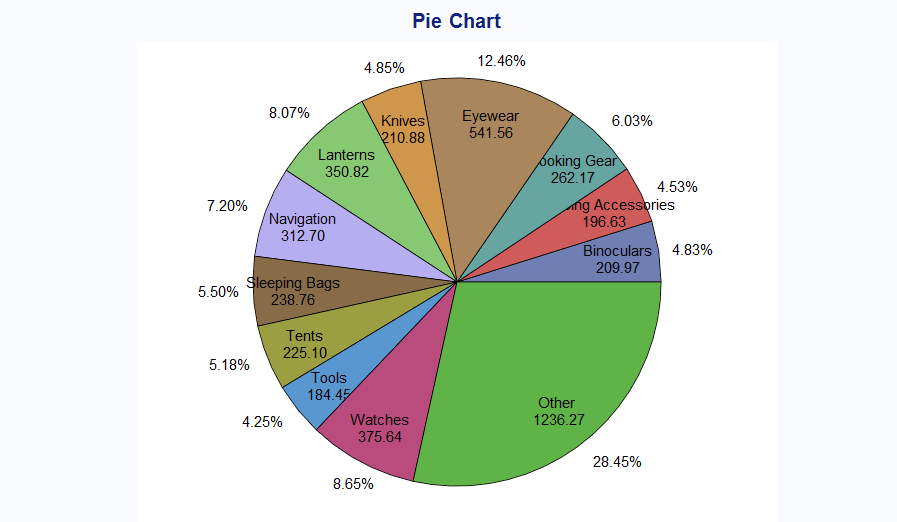


Figure 19: Proportion of Top Product Type by Total Sales

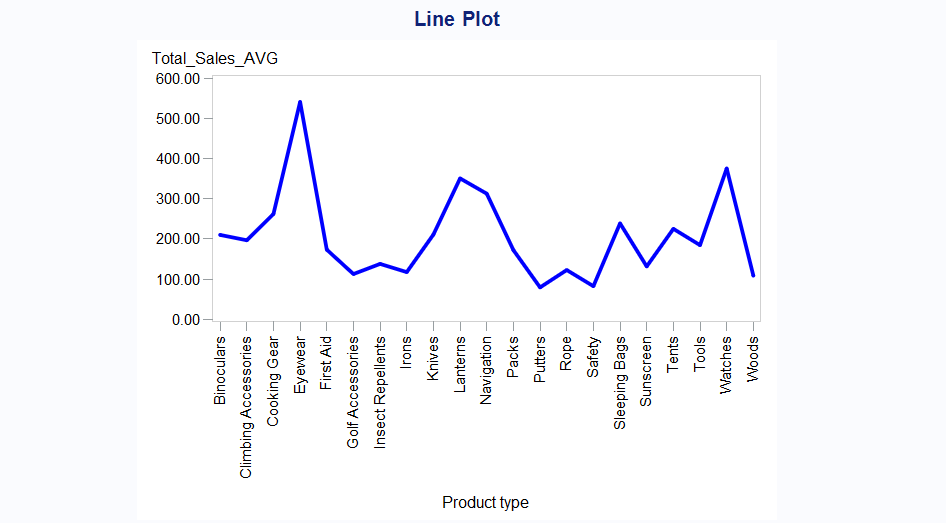


Figure 20: Top Product Type by Sub Total Sales.

Figure 21 displays the product line according to total sales, with personal accessories being the highest outliers. Therefore, we can conclude that personal accessories are the leading product line.



Figure 21: Product Line by Total Sales

Analyzing Figure 22, it becomes clear that there is a strong correlation between eyewear, watches, and personal accessories. This relationship suggests that eyewear and watches are not only popular items in their own right but also play a significant role within the broader category of personal accessories. Consequently, we have determined that eyewear and watches stand out as the most sought-after products among consumers.

****

Figure 22: Correlation between Product type and Product Line

**3.2 Behavioral Characteristics**

In this section, We created three tables: STORE\_VISIT, TOTAL\_SKU, and TOTAL\_SALES. We calculated store visits by age group using the COUNT function for invoice dates. Additionally, We employed the SUM function to determine the total SKU and total sales for each age group.

Figure 23 illustrates the proportion of store visits by age group. Most customers who visited the store were between the ages of 36 and 50, accounting for 32.86% of the total visits. In contrast, the percentage of customers aged 51 to 65 was significantly lower. The line chart indicates that while the number of customers under 18 is low, there is an upward trend in visits from customers aged 25 to 50, followed by a decline for those aged 51 to less than 75.

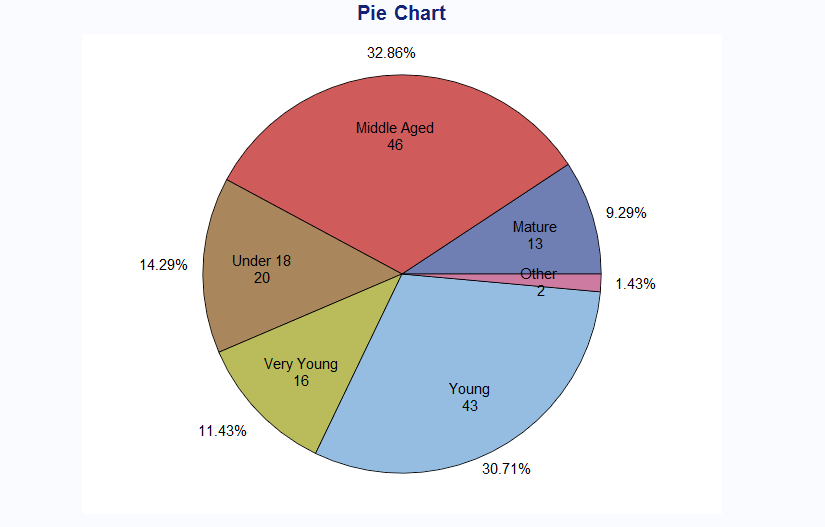


Figure 23: Proportion of Store visit by Age Group

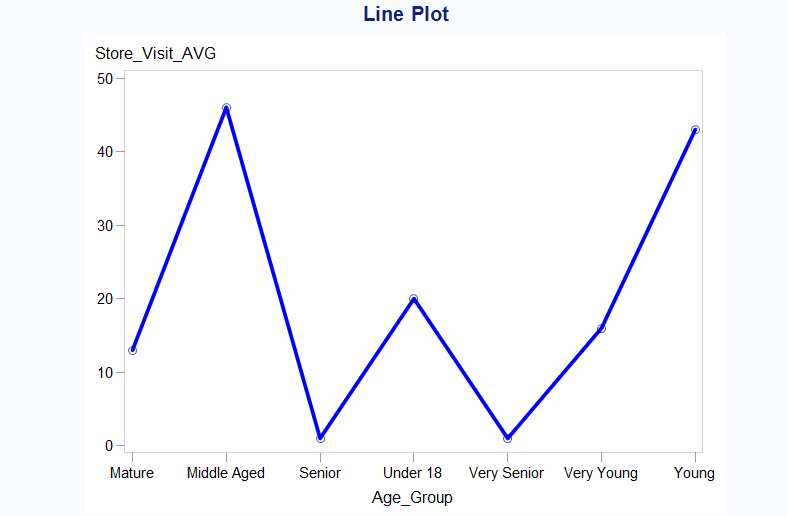


Figure 24: Store visit By Age Group

The analysis shown in Figures 26 and 27 reveals interesting trends regarding our customer demographics and sales performance. The age group of 36 to 50 years stands out significantly, showing the highest percentage of total Stock Keeping Units (SKUs) available. In contrast, the age groups of 18 to 25 and 51 to 75 have noticeably fewer SKUs. Furthermore, total sales figures are predominantly driven by customers in the 36 to 50 age bracket, underscoring their importance to our business.

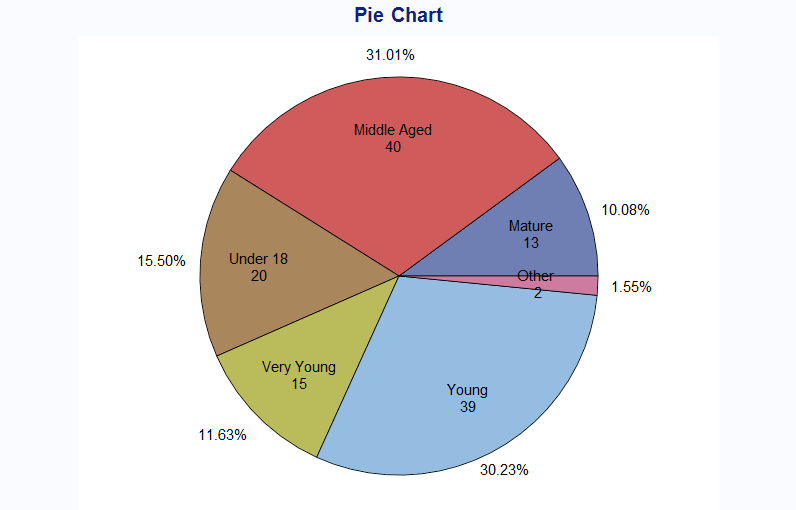


Figure 26: Percentage of Total SKU By Age Gropup

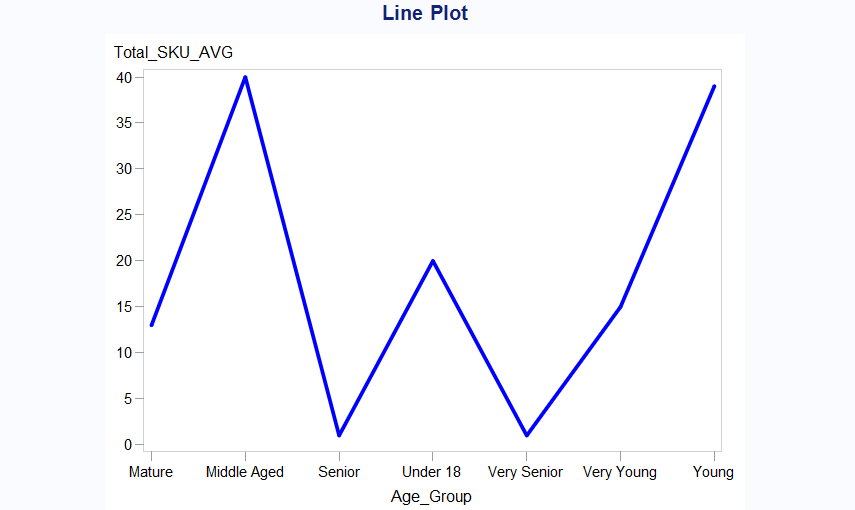


Figure 27: Total SKU vs. Age Group

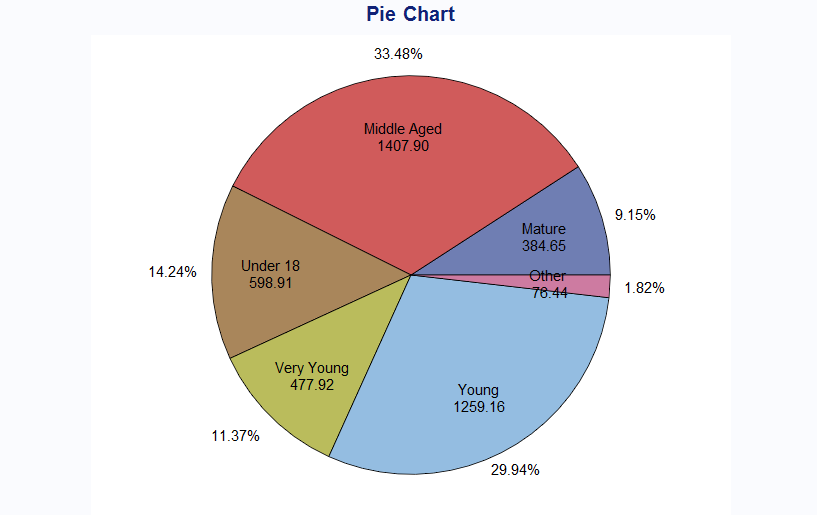


Figure 28: Percentage of Total Sales by Age Group.



Figure 29: Total prices by Age Group

Our customer base is largely composed of individuals aged 36 to 50, making it the most frequent age group among our visitors. In addition, we observe a limited number of customers are Mature to Very senior and a smaller segment in the Young to Very young. This demographic insight highlights the pivotal role that middle-aged customers play in our sales strategy.

**3.3 Promotions Analysis**

In the analysis of promotional effectiveness, I developed a set of coding variables to categorize different levels of promotion. A promotion level of 0% is classified as "no promotion," while the active promotion levels are set at 10%, 20%, and 30%. These classifications are stored in the PROMOTION\_TYPE table for further reference.

To better understand the impact of these promotions on sales, I also created the PRODUCT\_SOLD table. This table utilizes the COUNT function to quantify the number of products sold, organized by the various promotion types.

The results are visually represented in Figure 30, which depicts the proportion of products sold under promotional conditions versus those sold without any promotion. It is evident from the figure that products sold with promotions constitute a significantly higher percentage of total sales. Meanwhile, Figure 31 provides additional insights by showcasing that the average number of promoted products sold reaches an impressive 1,296 units, whereas the average for non-promoted products lags at only 432 units. This stark contrast highlights the effectiveness of promotional strategies in driving sales.

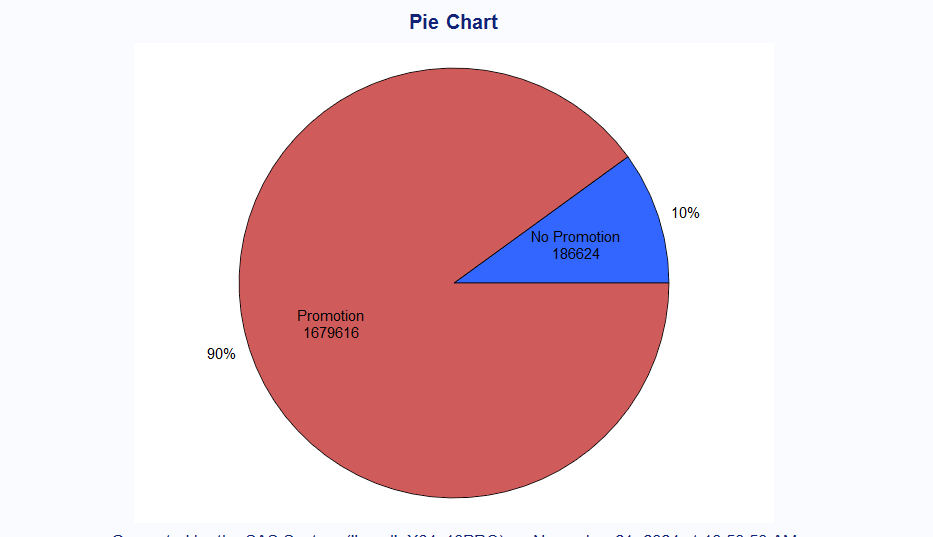


Figure 30: Proportion of Promotion and No promotion

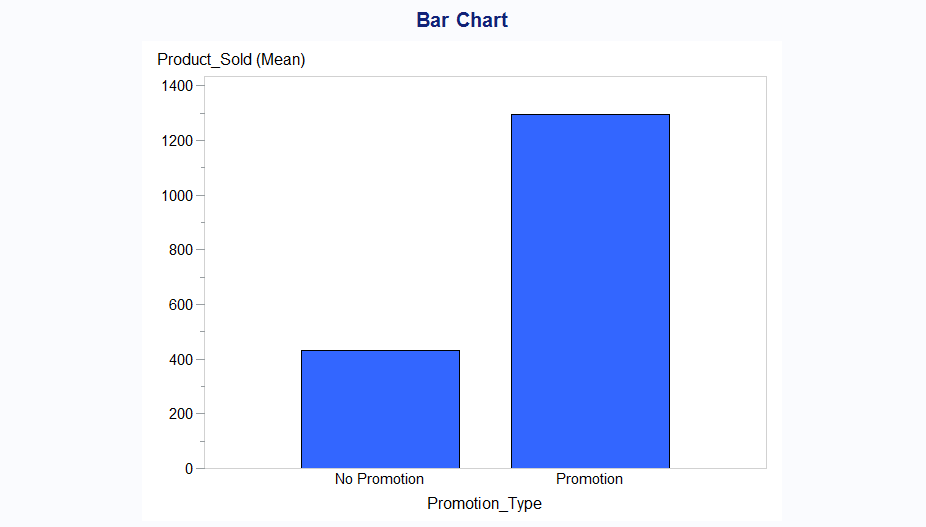


Figure 31: Products sold Vs. Promotion Type.

**3.4 Daily Sales Analysis**

The daily sales report provided valuable insights into purchasing behavior, particularly when comparing weekdays to weekends. It revealed a striking trend: weekends consistently showcased a notably higher number of distinct SKUs sold per invoice than their weekday counterparts.

As part of our thorough analysis of daily sales data, we carefully counted the total SKUs sold for each invoice date. The summary statistics, accompanied by Figure 22, paint a clear picture of this trend. Notably, December 17, 2010, February 25, 2011, June 21, 2011, and November 15, 2011, all recorded an impressive total of 9 SKUs sold. In stark contrast, the interval from March 6, 2011, to June 7, 2011, showed the least activity, marked by a significantly lower total number of SKUs sold.

This analysis strongly suggests that dates with a higher total of SKUs sold are linked to increased sales activity, highlighting the importance of the day of the week in influencing consumer purchasing patterns.

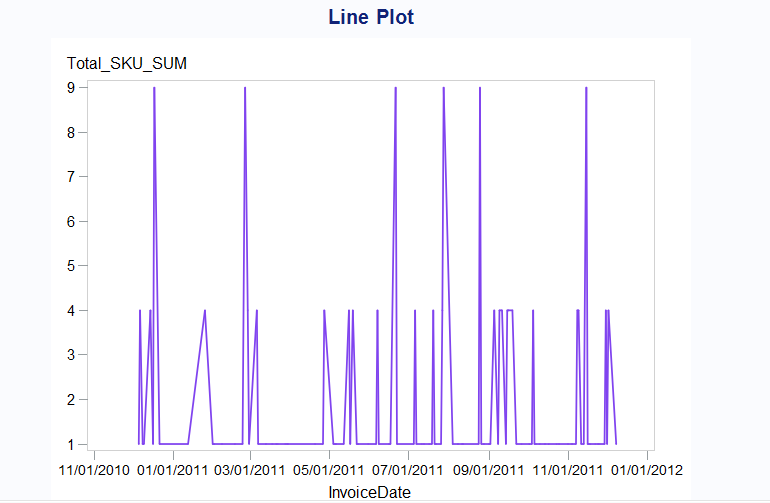
****

Figure 32: Total SKU by Invoice Date

**3.5 Supplier Analysis**

In our comprehensive supplier analysis, we meticulously calculated the total revenue associated with each supplier, organizing this data by supplier ID as documented in the Total\_Revenue table. Additionally, we assessed the overall quantity of products provided by each supplier, with this information stored in the Total\_Product table.

Figures 33 and 34 vividly illustrate that supplier ID 5 stands out with an impressive total revenue of $929.56 and offers a total of 29 products, which positions it as the supplier with the highest percentage contribution among all suppliers analyzed.

Delving deeper into the summary statistics, we identify Dragon SA as the top supplier, proudly representing a U.S. company. In contrast, Carper & Sons holds the position of the lowest supplier in terms of revenue and product count, having its operations based in Turkey. This analysis highlights the diverse landscape of our suppliers and their varying contributions to our overall performance.

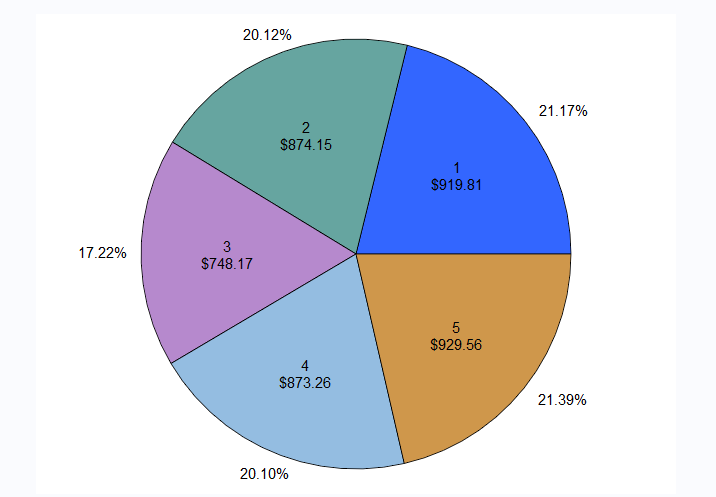
****

Figure 33: Total revenue by Supplier ID

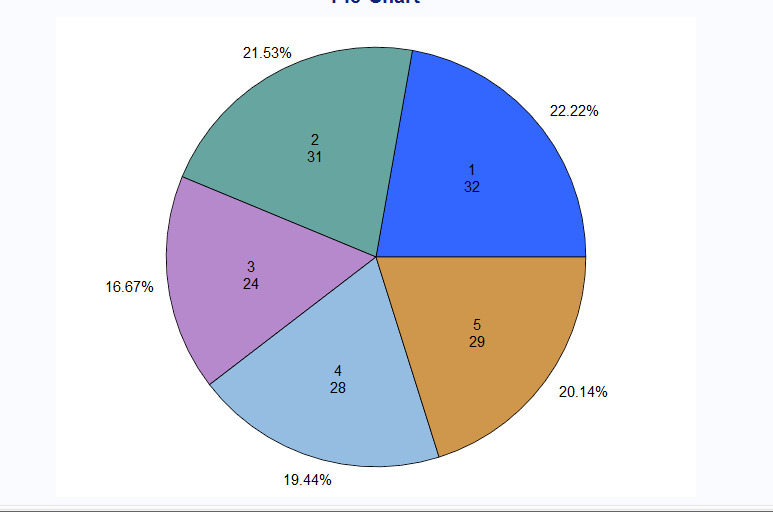
****

Figure 34: Supplier by Supplier ID

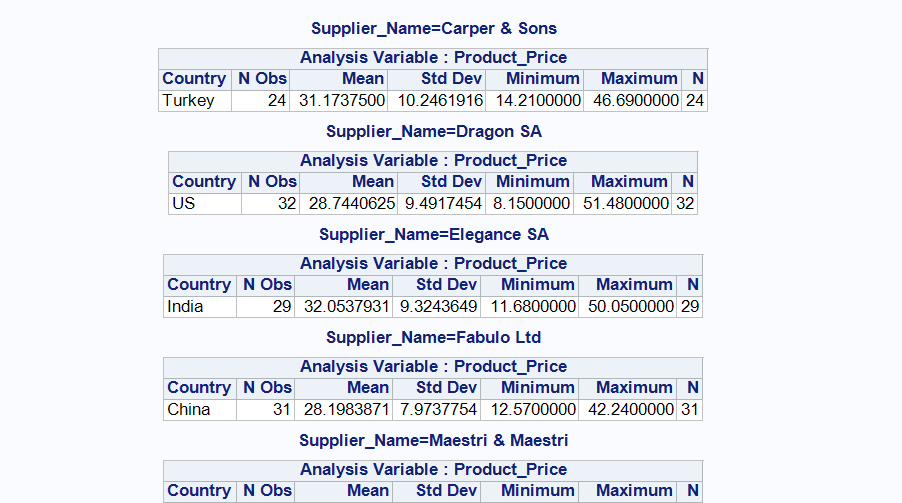
****

Figure 35: Summary statistics

**Recommendation for Advanced analysis:**

Based on the product ranking analysis, we can see that eyewear is the top-selling product, followed closely by watches in second place. The Personal Accessories product line is the most popular, showing a strong correlation with both eyewear and watches.

Behavioral analysis indicates that the majority of our customers are middle-aged, with very few customers over the age of 65. Young customers represent the second-largest demographic in terms of the percentage of total SKUs. This highlights the important role that middle-aged customers play in our sales strategy.

The promotion analysis reveals that products sold during promotions tend to have higher sales compared to those sold without promotions. Additionally, the daily sales analysis shows that total product sales, recorded by invoice date, are higher on weekends than on weekdays.

In terms of supplier performance, revenue from Supplier ID 5 constitutes a significant percentage of our total revenue, whereas revenue from Supplier ID 3 is considerably lower. This indicates that Supplier ID 5 is our top supplier. Furthermore, the leading supplier is Dragon SA, a company based in the United States.

**Results and Conclusions:**

This report analyzes sales and customer demographics for RealPOS, a Brazilian retailer of outdoor equipment and accessories. It uses datasets like the Invoice Table and Customer Table, processed with SAS Enterprise Guide to find useful insights.

These insights offer important knowledge about sales performance and customer demographics, helping RealPOS understand its business better and instilling confidence among stakeholders about the company's operations and strategies.

We gained valuable insights into customer purchasing behaviors through our analysis. We calculated total items and monetary value per Invoice ID, clarifying sales performance. Most customers are male and reside primarily in São Paulo (SP), with a smaller portion in Rio de Janeiro (RJ). A significant segment of our clientele is middle-aged, with very few under 18 or over 65. Our basket analysis shows that certain products within the same categories are frequently purchased together, with eyewear and watches as the top-selling items. Promotional sales outperform regular sales, and overall sales are higher on weekends. Supplier ID 5, our top supplier, generates a significant revenue share, while Supplier ID 3's revenue is much lower. Dragon SA, based in the United States, is our leading supplier.

**Recommendations:**

1. Focus Marketing Efforts: Tailor marketing campaigns to align with the predominant customer demographics, with particular emphasis on the "Young" and "Middle Aged" segments.
2. Enhance Promotions: Capitalize on the high levels of engagement generated by promotions to explore additional discount opportunities, ensuring that profit margins remain intact.
3. Regional Strategies: Formulate targeted regional strategies to address the substantial variations in revenue contributions across Brazil, with a focus on enhancing personal interactions within the highest-revenue regions.