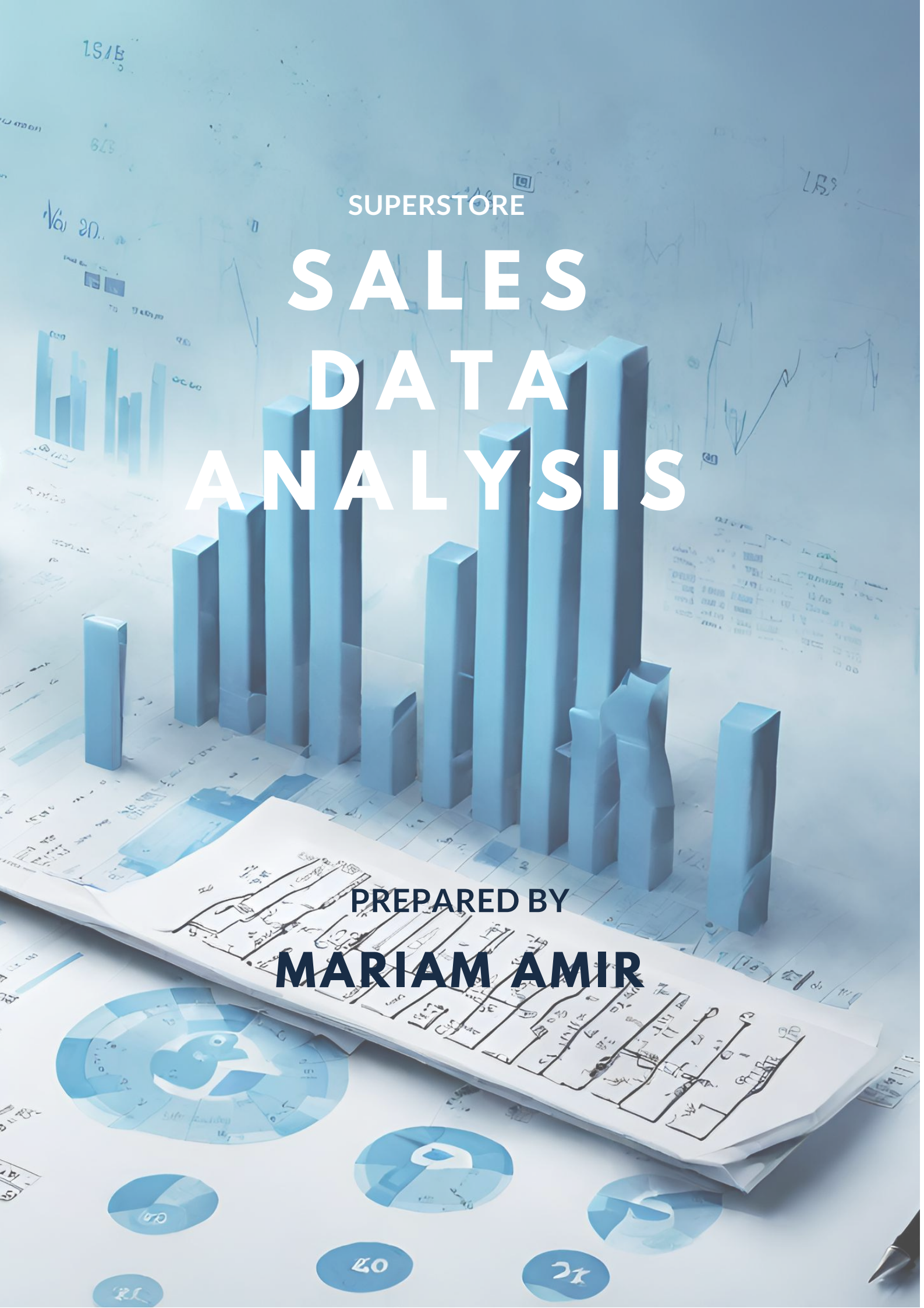


SUPERSTORE

SALES DATA ANALYSIS

PREPARED BY

MARIAM AMIR

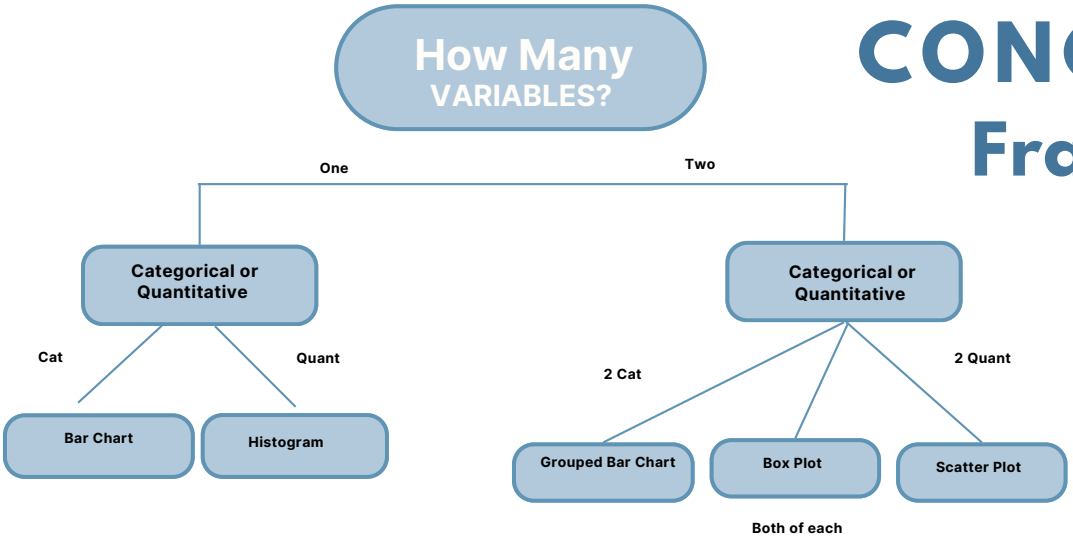
TABLE OF CONTENTS

INTRODUCTION

In this comprehensive report, our primary aim is to delve into the realm of descriptive statistics. We focus on dissecting crucial facets influencing our business operations, including order priority, shipping methods, and customer segmentation. While the report steers clear of predictive or causal analysis, its sole focus is on unraveling the intricate details and numerical insights provided by descriptive statistics.
We meticulously examine both quantitative metrics—such as unit price, sales, and shipping costs—and qualitative variables like order priority, shipping mode, and customer segments. These metrics, blending both numerical and contextual dimensions, offer a profound understanding of our business landscape.

Rather than aiming to predict or explain trends, our primary objective is to present a detailed and clear picture of these fundamental facts and specific details. This detailed view will empower us to make informed decisions and strategic choices that can steer our business toward sustained growth and success in the foreseeable future

CONCEPTUAL Framework



UNIT PRICE

The summary statistics reveal valuable insights into the distribution of unit prices within our dataset. The range of prices spans from a minimum of **$1.14** to a maximum of **$6783.02**. A quarter of the prices are below **$6.48**, while the median, representing the middle value, stands at **$20.99**, indicating that half of the prices fall below this point. The average unit price across all observations is $**108.23**, yet this mean value may be influenced by a few extremely high prices. Three-quarters of the prices lie below **$99.23**, signifying that the majority of observed prices are within this range. The presence of significantly higher prices, notably the maximum value, suggests the existence of outliers that deviate notably from the typical price range.

Figure 1

Min	Max	Mean
1.14	6783.02	108.23
1st Quartile	**Median**	**3rd Quartile**
6.48	20.99	99.23

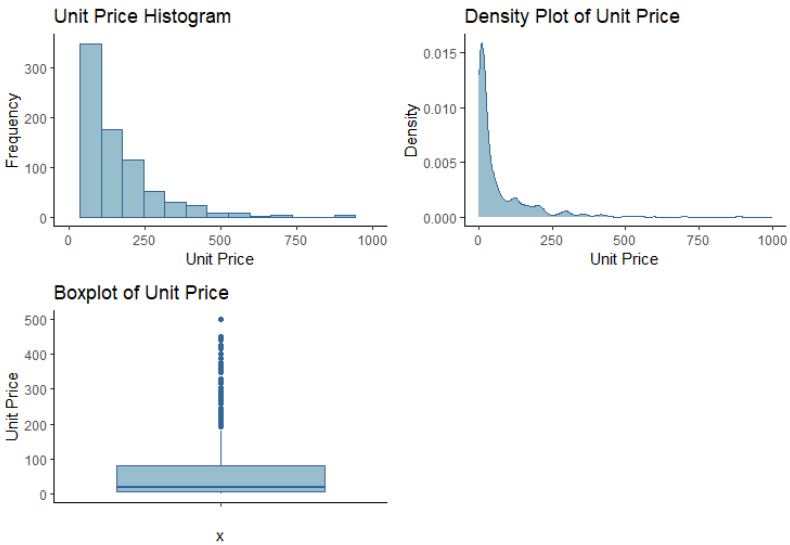


Figure 2

UNIT PRICE

The summary statistics unveil a diverse range of unit prices within our dataset. Notably, the presence of outliers becomes evident when considering the distribution. The boxplot visualization displays these outliers as data points significantly distant from the central bulk of prices, emphasizing their deviation from the typical range. Additionally, the histogram's skewed nature accentuates this disparity in pricing, with a notable concentration towards lower values and a few instances of considerably higher prices contributing to the skewness. These graphical representations, coupled with the summary statistics, collectively underscore the existence of outliers and the skewed distribution, shedding light on the variability and the presence of extreme values within our unit price data.

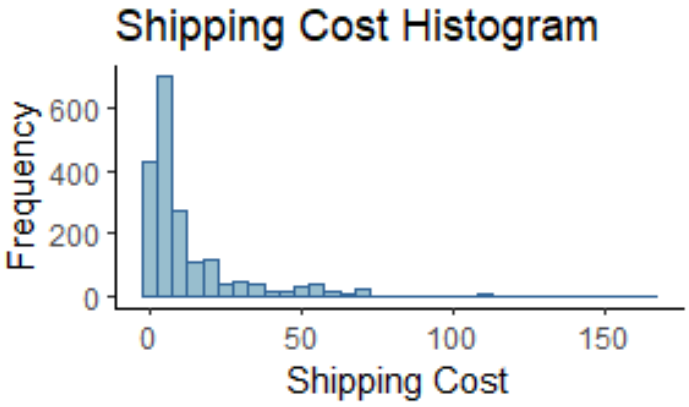
SHIPPING COST

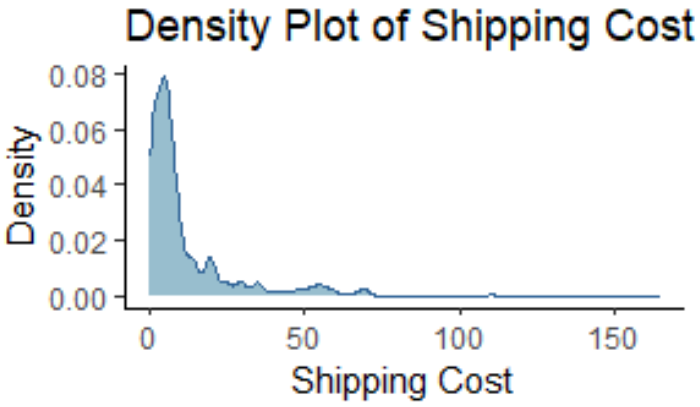
Min	Max	Mean
0.49	164.73	12.81
1st Quartile	**Median**	**3rd Quartile**
3.04	6.07	14.00

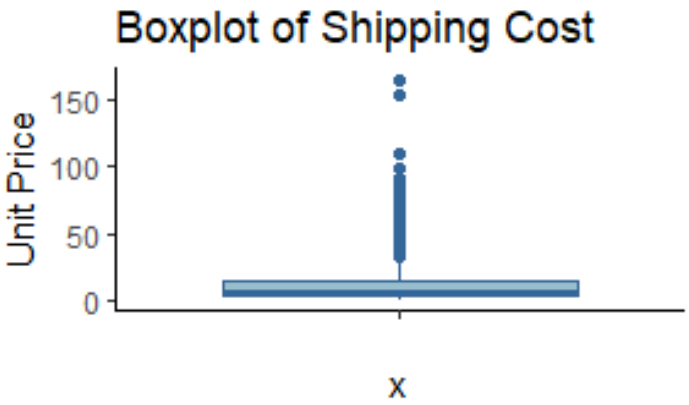
Figure 3

The summary statistics for shipping costs highlight a varied range within the dataset. The minimum cost for shipping stands at **$0.49**, with **25%** of the costs falling below **$3.04** (1st quartile) and **50%** below **$6.07** (median). The mean shipping cost, at **$12.81**, suggests an average value influenced by relatively higher costs. Moreover, three-quarters of the observed shipping costs lie below **$14.00** (3rd quartile), indicating a concentration within this range. However, the maximum shipping cost observed is notably higher at **$164.73**, signifying the presence of potential outliers impacting the overall distribution. The histogram displaying positive skewness emphasizes a concentration of lower shipping costs with a many instances of considerably higher costs (could be seen clearly from the boxplot), further supporting the presence of outliers and the right-leaning nature of the distribution

Figure 4







SALES

The summary statistics for the variable indicate a wide range in the data. The minimum value stands at **$2.25**, while the median—a middle value—settles at **$200.64**. However, the mean of **$971.67** suggests a considerable influence of extremely high values on the average. Additionally, the interquartile range (IQR) spans from **$58.67** (1st quartile) to **$772.56** (3rd quartile), indicating where the middle **50%** of the data lies. The maximum value of **$45737.33** signifies the presence of substantial outliers impacting the overall distribution of this variable.

Figure 5

Min	Max	Mean
2.25	45737.33	971.67
1st Quartile	**Median**	**3rd Quartile**
58.67	200.64	772.56

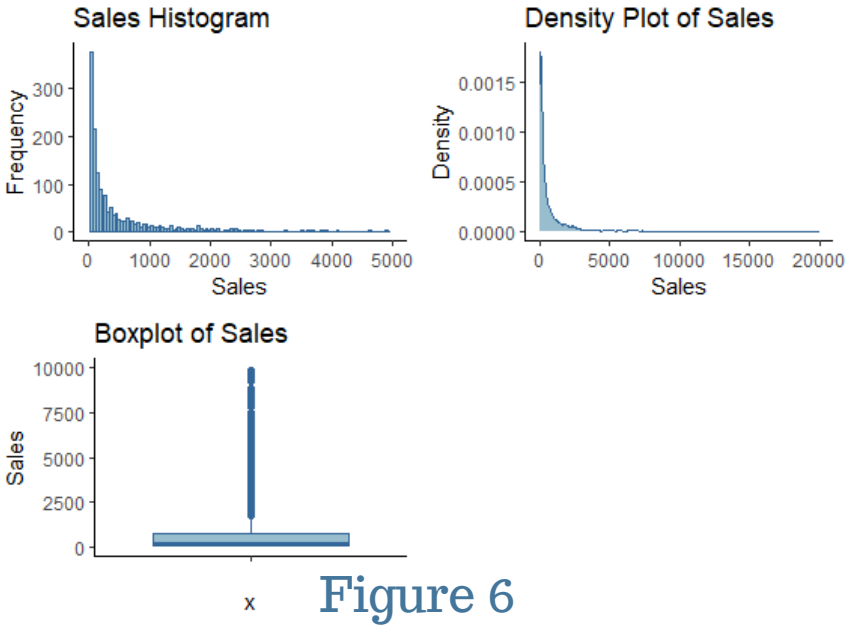


Figure 6

SALES

The histogram and boxplot visually depict the distribution and characteristics of the data. The histogram showcases a heavily right-skewed distribution, indicating a clustering of values towards the lower end and a tail extending towards extremely high values. This skewness aligns with the summary statistics, emphasizing a concentration of values at the lower end with a long tail of exceptionally high values. The boxplot reinforces these observations, highlighting numerous data points well beyond the upper whisker, representing excessive outliers that significantly extend the upper range of the distribution. These excessive outliers contribute to a widening gap between the maximum value and the rest of the data points, accentuating the right-skewed nature and the considerable influence of extreme values on the distribution.

RELATIONSHIPS & INSIGHTS

What are the degrees of association or correlation observed between Unit Price, Sales, and Shipping Cost in the dataset, and how do these correlations inform us about the interplay between these quantitative variables within the broader context of the dataset?"

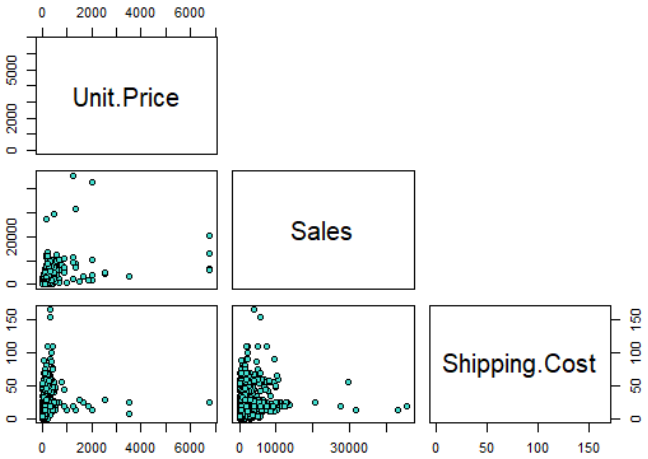


Figure 7

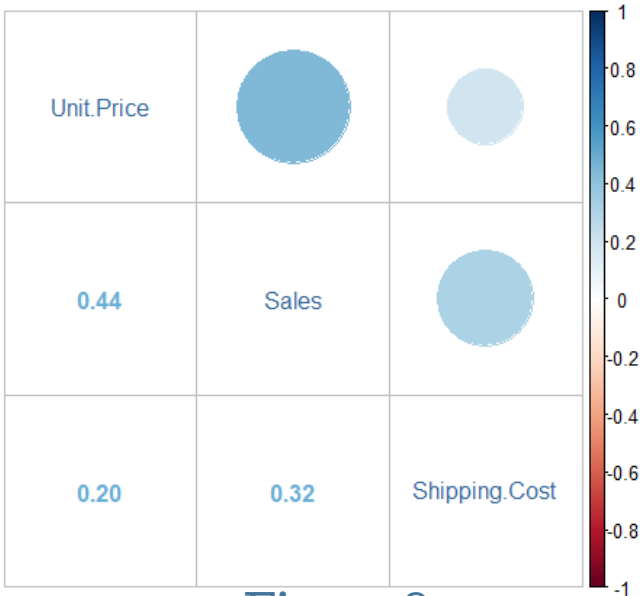


Figure 8

The results you've presented appear to be a correlation matrix between three variables: Unit Price, Sales, and Shipping Cost. Each cell in the matrix represents the correlation coefficient between the respective pair of variables. Here's the interpretation:

1. **Unit Price & Sales**:
   - There's a moderate positive correlation **(0.44)** between Unit Price and Sales.
   - It suggests that as the Unit Price increases, Sales tend to increase as well, but not strongly.
2. **Sales & Shipping Cost**:
   - There's a moderate positive correlation **(0.32)** between Sales and Shipping Cost.
   - This indicates that there's some tendency for higher Sales to be associated with higher Shipping Costs, but again, the correlation is not very strong.
3. **Unit Price & Shipping Cost**:
   - There's a weaker positive correlation **(0.20)** between Unit Price and Shipping Cost.
   - This suggests a slight tendency for higher Unit Prices to be associated with higher Shipping Costs, but the correlation is relatively weak compared to the other relationships mentioned.

In summary, based on these correlation coefficients:

- There is a moderate positive correlation between Unit Price and Sales.
- Both Sales and Shipping Cost show a moderate positive correlation.
- Unit Price and Shipping Cost have a weaker positive correlation compared to the other relationships.

However, it's crucial to note that these interpretations might not accurately represent the broader relationships due to the evident concentration of data points in specific areas, as observed in the scatter plot. This concentration significantly influences the calculated correlation coefficients and might potentially lead to misinterpretations. Therefore, further analysis is essential, including the exploration of outliers and the implementation of appropriate methodologies to address their impact on the relationships observed.

CATEGORICAL GLIMPSE

ORDER PRIORITY

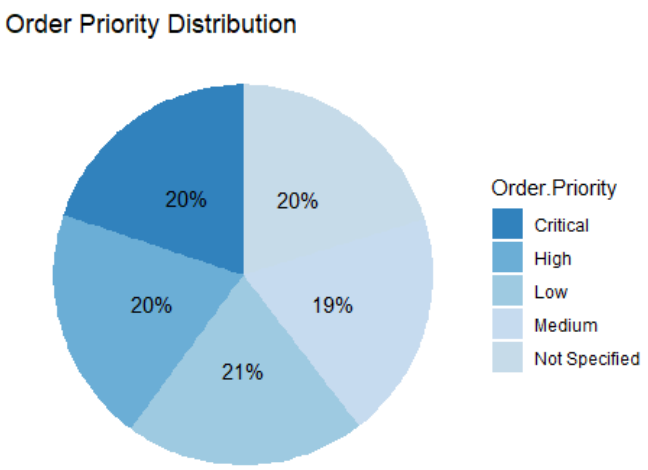


Figure 9

In this pie chart representing the distribution of order priorities, each category is depicted as approximately **20%** of the whole chart. The categories, namely '**critical,**' '**high,**' '**low,**' '**medium,**' **and** '**not specified,**' all contribute relatively equally to the distribution, each representing an approximate **fifth** of the entire dataset. This balanced distribution across these categories implies a relatively even distribution of order priorities, suggesting a consistent and diversified spread of different priority levels within the dataset. Such a distribution could indicate a well-managed variety of order priority types, with no single priority level significantly dominating the dataset,

SHIP MODE

The Bar Chart highlights the distribution of Shipping Mode across various shipping methods. Among the observed Methods, 'Regular Air' emerges as the most commonly chosen shipping method, recording the highest frequency count of **1421.** Conversely, 'Express Air' exhibits a moderately lower frequency count of **238,** indicating a notably lower but still substantial usage compared to 'Regular Air.' 'Delivery Truck' stands with the lowest frequency count of **270**.

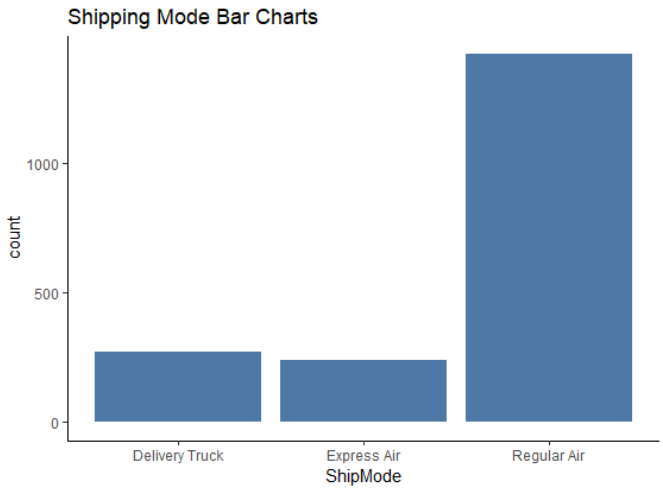


Figure 10

Customer.Segment	Frequency
Consumer	407
Corporate	671
Home Office	468
Small Business	383

Figure 11

The **"Corporate"** segment stands out with the highest frequency count of **671**, indicating that **671** orders or customers belong to this category. Following closely behind is the **"Home Office"** segment, which boasts a frequency count of **468**, showcasing a substantial presence within this customer category. Meanwhile, the **"Consumer"** category comprises **407** orders or customers, maintaining a noteworthy but relatively lower count compared to the previous segments. Lastly, the "**Small Business**" category demonstrates a frequency count of **383**, representing a sizeable yet slightly smaller portion of customers within this segment.

RELATIONSHIPS & INSIGHTS

Is there a particular customer segment that consistently demonstrates a higher inclination towards prioritizing their orders compared to others?

The Bar chart makes it clear that corporate orders are most frequent in the "Critical Orders" category. At first glance, it seems like corporates receive special attention. However, a closer look to the **Customer Segment Distribution Pie Chart** shows that corporate orders actually dominate across all priority levels shown in the chart. This might mean that corporate orders make up a large percentage, **35%,** of all orders placed. So, while they stand out in the "Critical Orders" category, they also have a strong presence across all priority levels, showing their significant role in the overall order distribution

Figure 12

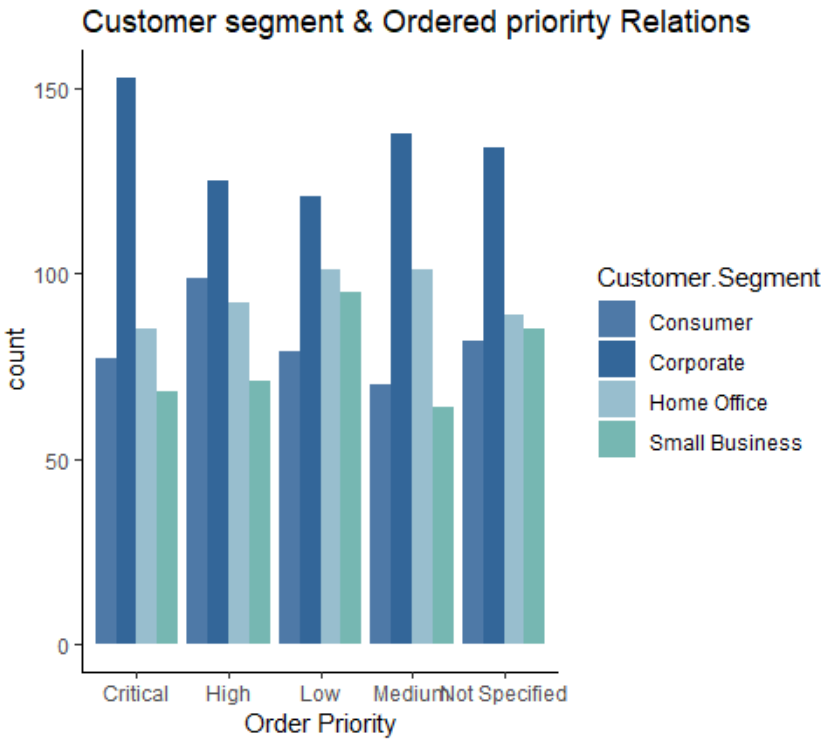
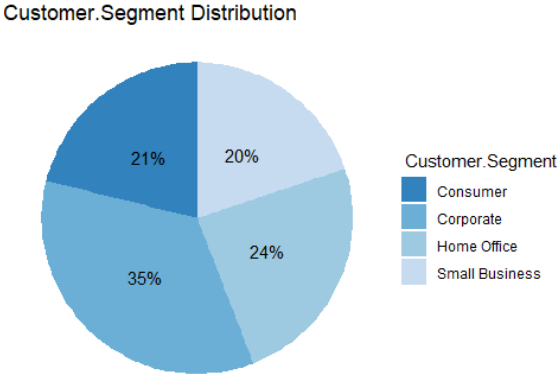


Table (1): Two-Way Contingency Table for Consumer Segement & Order Priority

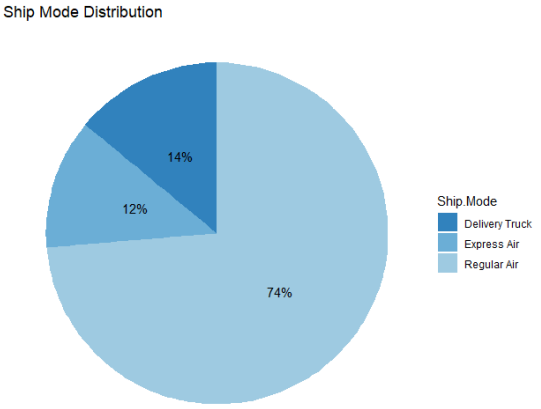
	Critical	High	Low	Medium	Not Specified	Sum
Consumer	77	99	79	70	82	407
Corporate	153	125	121	138	134	671
Home Office	85	92	101	101	89	468
Small Business	68	71	95	64	85	383
Sum	383	387	396	373	390	1929



RELATIONSHIPS & INSIGHTS

Is there a strategic link between order priority levels and the shipping modes adopted for prompt fulfillment?

It seems like there might not be a strong connection between the choice of shipping mode and order priority. The fact that "Regular Air" is used for **74%** of shipments and "Regular Air" is consistently the top choice across all order priorities suggests that the shipping method might not change much based on how urgent an order is.



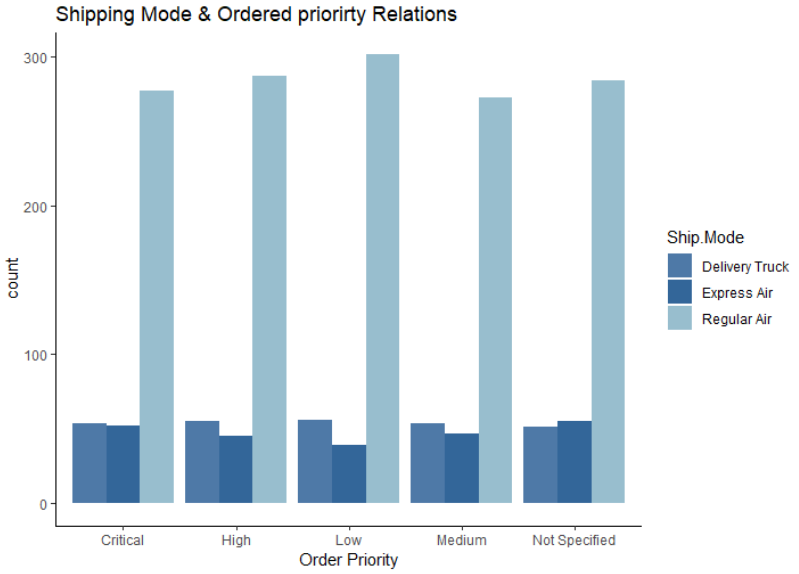


Table (2): Two-Way Contingency Table for Ship Mode & Order Priority

	Critical	High	Low	Medium	Not Specified	Sum
Delivery Truck	54	55	56	54	51	270
Express Air	52	45	39	47	55	238
Regular Air	277	287	301	272	284	1421
Sum	383	387	396	373	390	1929

Figure 13

Do particular customer segments drive the choice of shipping modes to maintain efficient delivery timelines?

It seems like there might not be a strong connection between the choice of shipping mode and Customer Segment. The fact that "Regular Air" is used for **74%** of shipments and "Regular Air" is consistently the top choice across all customer segments suggests that the shipping method might not change much based on the customer ordering the order.

Table (3): Two-Way Contingency Table for Ship Mode & Customer Segment

	Consumer	Corporate	Home Office	Small Business	Sum
Delivery Truck	53	93	60	64	270
Express Air	49	88	54	47	238
Regular Air	305	490	354	272	1421
Sum	407	671	468	383	1929

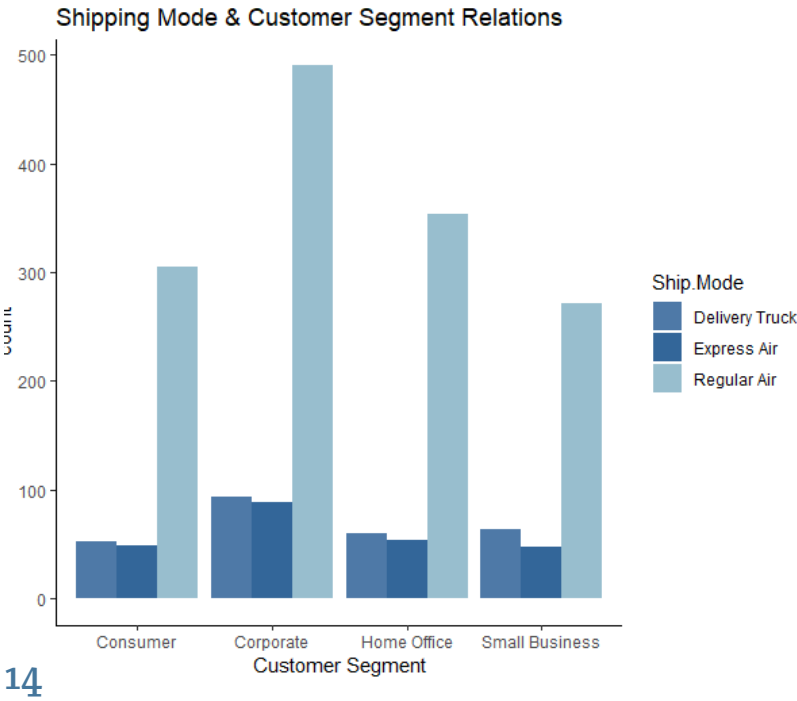


Figure 14

APPENDIX

Figure 15: Ship Mode Frequency table

Ship Mode	Frequency
Delivery Truck	270
Express Air	238
Regular Air	1421

Figure 16: Order Priority Frequency table

Order.Priority	Frequency
Critical	383
High	387
Low	396
Medium	373
Not Specified	390

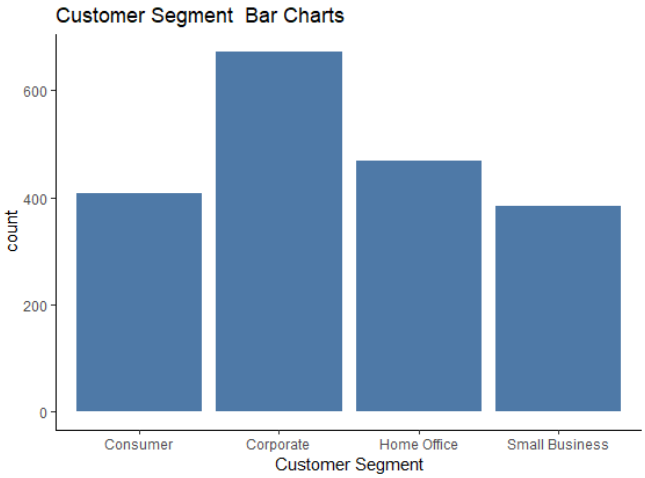


Figure 17

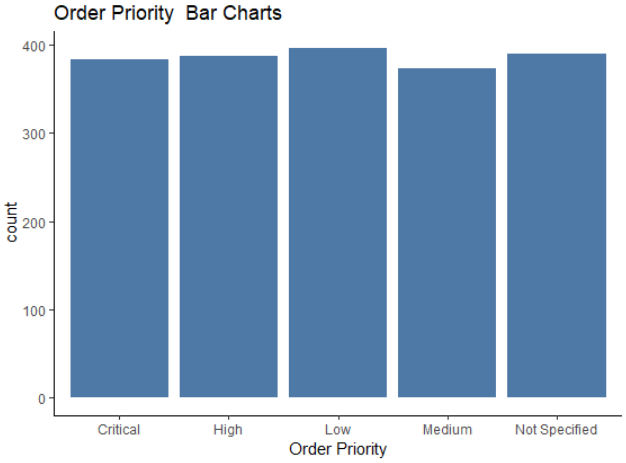


Figure 18