MANIPAL INSTITUTE OF TECHNOLOGY
MANIPAL
(A constituent unit of MAHE, Manipal)

Group Number: P3

Dataset: Furniture Sales

Submitted By

PRASEEDA K	220970075
KINNARI GOHIL	220970081
SANJUBALA THOKCHOM	220970131
SUSHANTH	220970096
SHREYA	220970084
TANUSHREE DAS	220970077

Under the Guidance of

Dr. Poornima Panduranga Kundapur

Associate Professor

Department Of Data Science &

Computer Applications

Manipal Institute of Technology, Manipal-576104, India

TABLE OF CONTENTS

Chapter No.	Title	Page No.
1.	Introduction	1
2.	Variables used and Categories 2.1 Variables and Categories 2.2 Understanding categorical values 2.3 Understanding Numerical values	2
3.	Descriptive Statistics 3.1 Measure of Central Tendency 3.2 Measure of dispersion or variance 3.3 Measure of Distribution	3-4
4.	Data Visualization 4.1 Histogram 4.2 Bar Graphs 4.3 Box Plot	5-8
5.	Relationship Identified 5.1 Scatterplot 5.2 Linear Regression 5.3 Methodology 5.3.1 Scatterplot 5.3.2 Linear Regression 5.3.3 ANOVA 5.4 Discussion 5.5 Conclusion	9-13
6.	Conclusion	14
7.	References	15.

1. INTRODUCTION

With the rise of e-commerce and online shopping, it has become more important than ever for furniture retailers to accurately predict prices to remain competitive in the market.
The purpose of this project is to find the relationship between furniture price and sales. In this case study, we will explore the sample data of the provided dataset of 2000 observations.

Predicting the future impact of furniture sales can be challenging as it depends on various factors such as economic conditions, consumer trends, and industry innovations. However, it's important to note that these predictions are based on general industry trends and may not capture specific market fluctuations or unforeseen events. The future impact of furniture sales can be influenced by a wide range of factors, and market conditions may vary regionally.

We have considered sample size of 200 data on which:
There were 7 attributes in the provided dataset. The four major attributes we considered are: Rate, delivery, sale and price.

2. VARIABLES USED AND CATEGORIES

On the provided Dataset the following are the variables and its types:

Table 2.1 Variables used and categories

Variables	**Variable Type**
Furniture	Nominal
Type	Nominal
Rate	Continuous, Interval Scale
Delivery	Continuous
Sale	Discrete Variable, Nominal Scale
Price	Continuous

2.2 Understanding the Categorical Values

1. Name of Furniture:
   For e.g., Bed Side Table with Storage Shelf, Modern Zigzag TV Table, Modern Desk with side shelves, Modern TV Table, etc.
2. Type of Furniture:
   For e.g., General, Modern Home, Home Decor Centre, Rango, Flamingo, etc.
3. URL: This variable is a distinct value for each furniture.

2.3 Understanding the Numerical Values

1. Rate: range 0 - 5
2. Delivery: range 10.83 – 1236.07
3. Sale: range 0- 73%
4. Price: range 1-8500EGP

3. DESCRIPTIVE STATISTICS

Descriptive Statistics describe, show and summarize the basic features of a dataset found in a given study, presented in a presented in a summary that describes the data sample and its measurements.

Types of Descriptive Statistics are:

1. **Measures of Central Tendency:** The one that indicates where the centre of the distribution is. The most used measures of central tendency are:
   1.1. Mean
   1.2. Median
   1.3. Mode

   Using Excel Functions, we have calculated all above-mentioned measures for the attributes: rate, delivery, sale, price.

   Therefore,

   - Mean for rate, delivery, sale, price is 1.941, 125.10, 0.30, 2163.99.
   - Mode for rate, delivery, sale, price is 0, 172.14, 0, 3000.
   - Median for rate, delivery, sale, price is 0, 172.14, 0.275, 1675.

2. **Measures of Dispersion or Variance:** The measure of the spread of the data around the centre of the distribution is called the Measure of Dispersion. The most commonly used measures of dispersion are:
   2.1. Range
   2.2. Variance
   2.3. Standard Deviation

   Using Excel Functions, we have calculated all above-mentioned measures for the attributes: rate, delivery, sale, price.

   Therefore,

   - Range for rate, delivery, sale, price is 5, 1225.24, 73, 8499.
   - Variance for rate, delivery, sale, price is 4.644, 11190.69, -1.05156, 1.489746.
   - Standard Deviation for rate, delivery, sale, price is 2.155, 105.7861, 0.214, 1652.185.

3. **Distribution:** The Distribution of a variable is the pattern of frequencies, meaning that the set of all possible values and the frequencies associated with these values.

   Hence, the following summary table gives the complete Descriptive Statistics details for the four continuous variables i.e., rate, delivery, sale, price

Table 3.1 Descriptive Statistics

	rate	***delivery***	***sale***	***price***
Mean	1.941	125.10985	0.30745	2163.99
Median	0	172.14	0.275	1675
Mode	0	172.14	0	3000
Standard Deviation	2.155003294	105.7860804	0.214576824	1652.184757
Sample Variance	4.644039196	11190.69481	0.046043214	2729714.472
Range	5	1225.24	0.73	8499
Minimum	0	10.83	0	1
Maximum	5	1236.07	0.73	8500
Sum	388.2	25021.97	61.49	432798
Count	200	200	200	200

4. DATA VISUALIZATION

4.1 HISTOGRAMS AND NORMALIZED DISTRIBUTION OF ALL VARIABLES

A histogram is a graphical representation of data points organized into user-specified ranges. Similar in appearance to bar graph, the histogram condenses a data series into an easily interpreted visual by taking many data points and grouping them into logical ranges or bins.

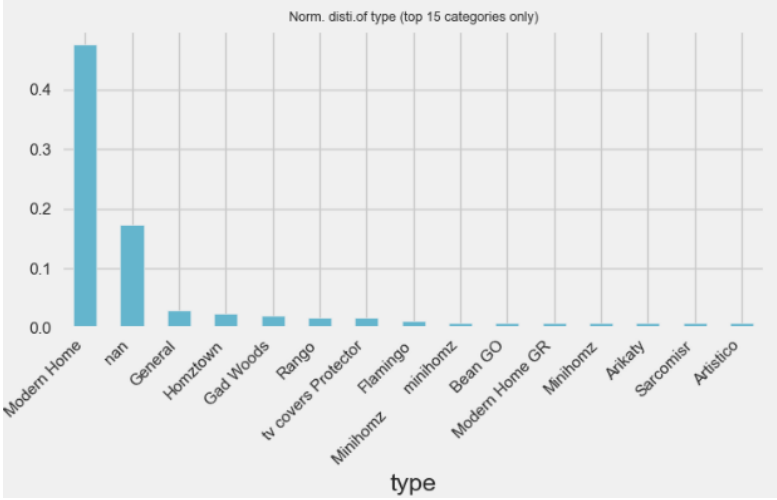

Figure 4.1.1 Histograms for Normalized distribution of type

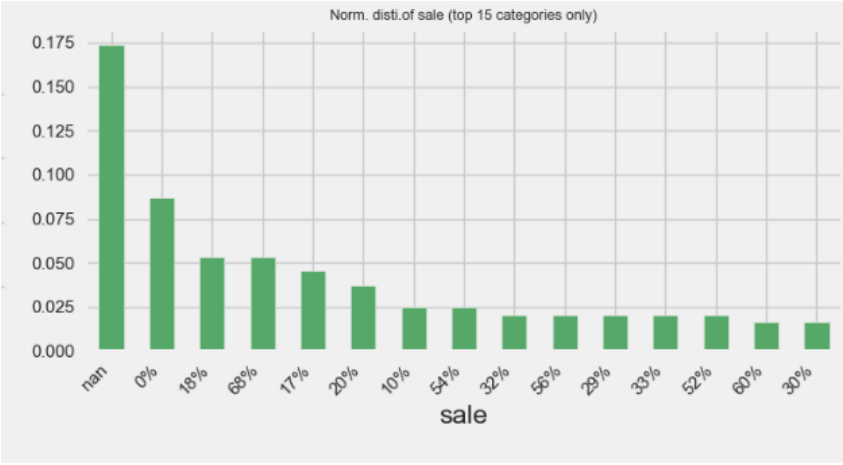

Figure 4.1.2 Histograms for Normalized distribution of sale

4.2 BAR CHART FOR EACH CONTINUOUS BY EACH CATEGORICAL VARIABLE

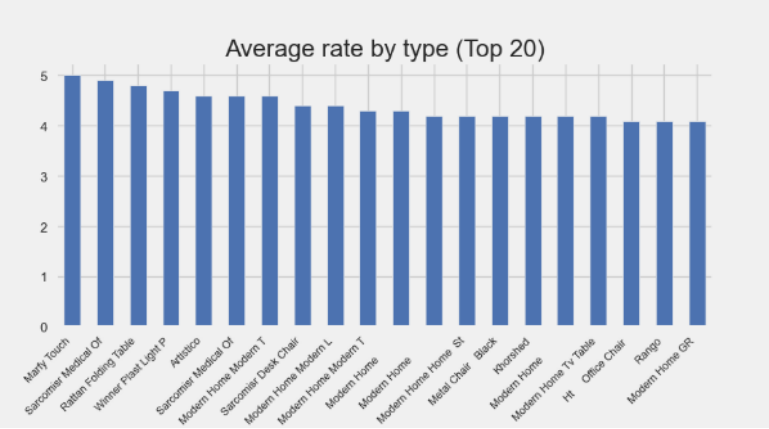

Figure 4.2.1 Bar Chart for Avg. rate by type.

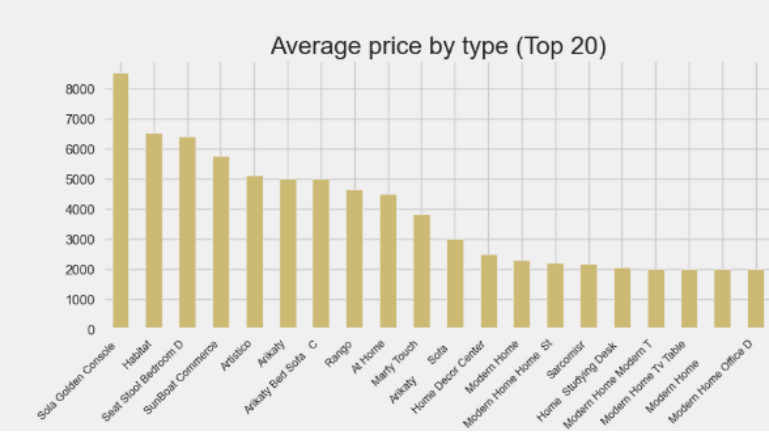

Figure 4.2.2 Bar chart for Avg. price by type.

Figure 4.2.3 Bar Chart for Avg. price by sale.

4.3BOX PLOT FOR EACH CONTINUOUS VARIABLE

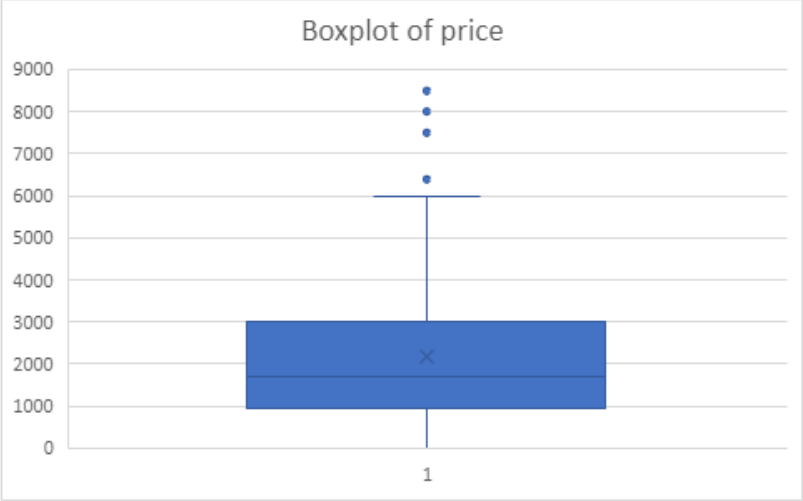

Figure 4.3.1 Box plot for price.

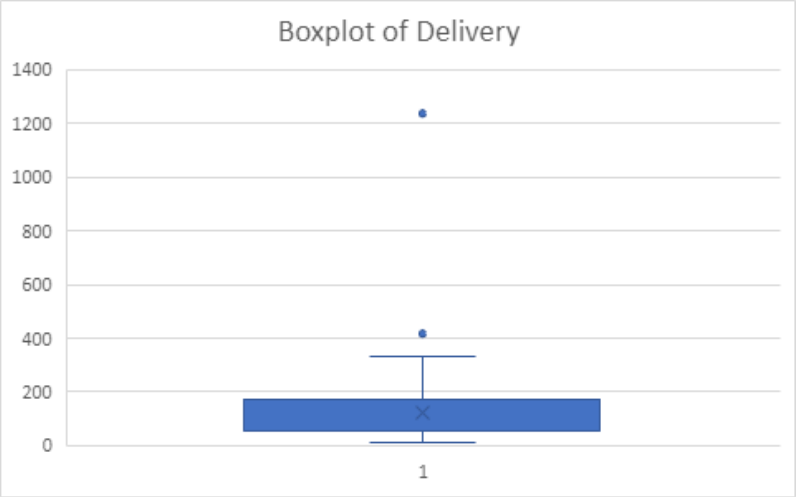

Figure 4.3.2 Box plot for delivery

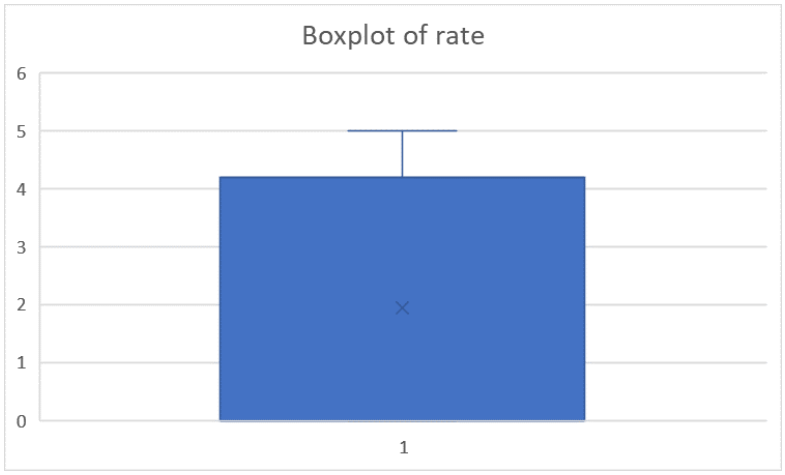

Figure 4.3.3 Box plot for rate

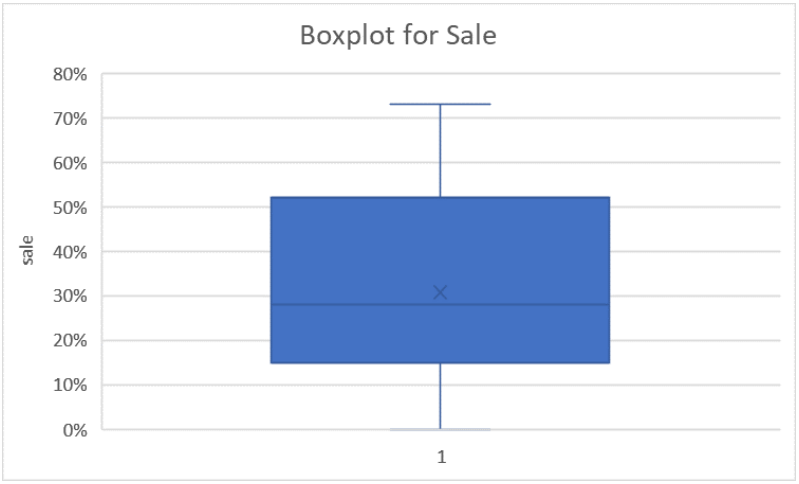

Figure 4.3.4 Box plot for sale

5. RELATIONSHIP IDENTIFIED

5.1 Scatterplot

Scatter plots are the graphs that present the relationship between two variables in a data-set. It represents data points on a two-dimensional plane or on a **Cartesian system**. The independent variable or attribute is plotted on the X-axis, while the dependent variable is plotted on the Y-axis.

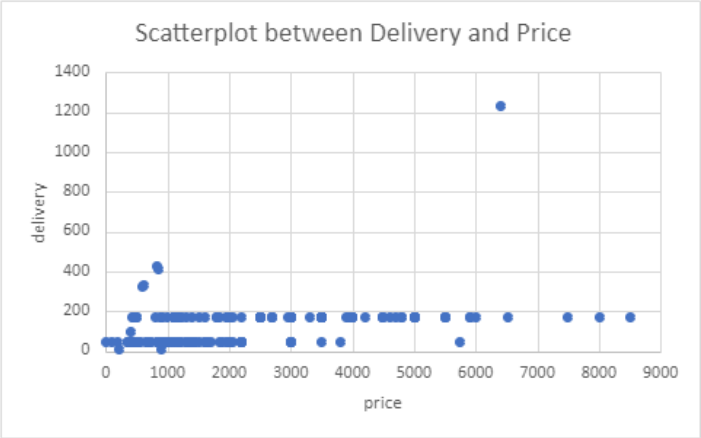

Figure 5.1.1 Scatterplot for delivery and price

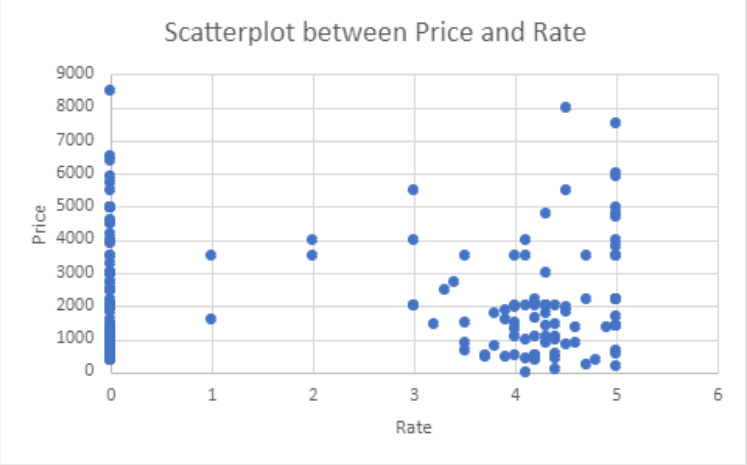

Figure 5.1.2 Scatterplot for price and rate

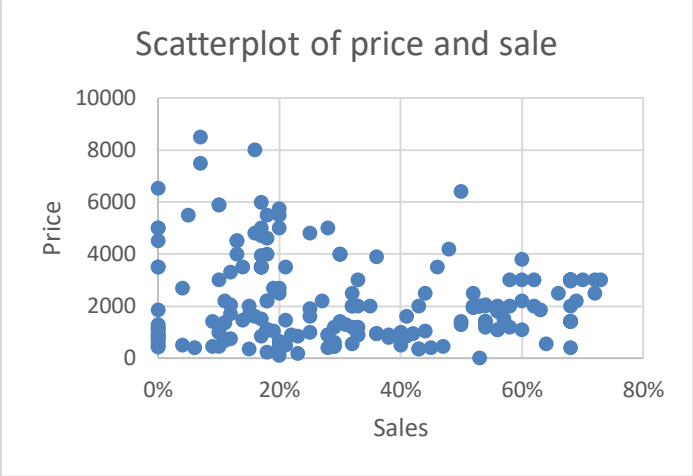

Figure 5.1.3 Scatterplot for price and sale

5.2 Simple Linear Regression

Simple linear regression is used to model the relationship between two continuous variables. Often, the objective is to predict the value of response variable based on the value of an predictor variable.

- Dependent variable: Price
- Independent variable: Sale

Table 5.2.1 Regression Statistics

Regression Statistics	
Multiple R	0.104663047
R Square	0.010954353
Adjusted R Square	0.005933817
Standard Error	1651.257641
Observations	200

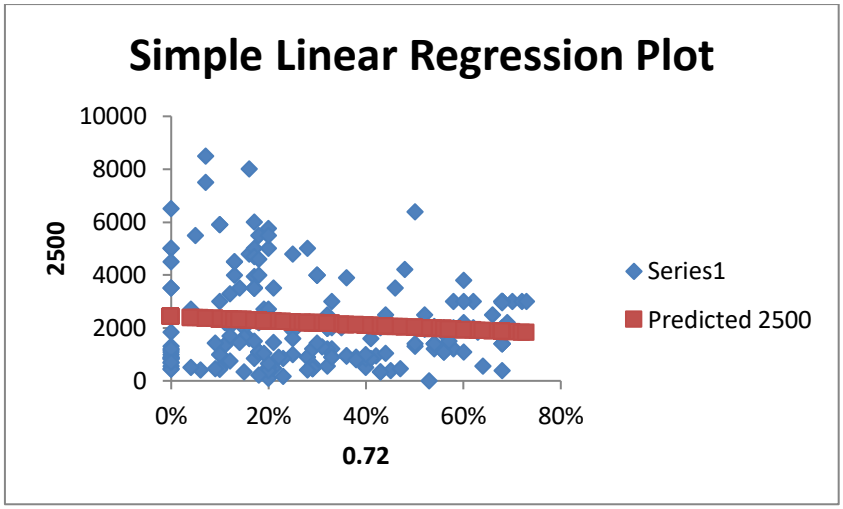

Figure 5.2.2 Scatterplot for price and sale

5.3 Methodology

5.3.1 Scatterplot:

The relationship between price and sale in a furniture dataset can be analysed using a scatterplot. The scatterplot can show whether there is a positive, negative, or no correlation between the two variables.

If there is a positive correlation, this means that as the price of the furniture increases, the number of units sold also increases. This can indicate that customers perceive the furniture as having a higher value or quality, and are willing to pay more for it.

If there is a negative correlation, this means that as the price of the furniture increases, the number of units sold decreases. This can indicate that the price is too high for customers, and they are not willing to pay that much for the furniture.

If there is no correlation, this means that there is no relationship between the two variables. In this case, price and sale are not related to each other, and changes in one variable do not affect the other variable.

By analysing the relationship between price and sale in a furniture dataset, we can say that there is a negative correlation between sale and price. As sale increases price decreases.

Correlation between sale & price:

r = -0.1017

Which means there is a negative correlation, as sale increases, price decreases.

5.3.2 Simple Linear Regression

Use the regression equation to make predictions about the price of a piece of furniture based on its sale. In the above table, we have used the equation, Y=a+bx to predict the price of a new piece of furniture based on its sale, or to compare the prices of different pieces of furniture based on their sales.

5.3.3 ANOVA: Single Factor

ANOVA stands for Analysis of Variance, which is a statistical technique used to analyse the differences between two or more groups or treatments. It involves comparing the variation within groups to the variation between groups to determine whether the differences between groups are significant.

The basic idea of ANOVA is to divide the total variation in a set of data into two parts: the variation between groups and the variation within groups. If the variation between groups is larger than the variation within groups, it suggests that the groups are significantly different from each other. Conversely, if the variation within groups is larger than the variation between groups, it suggests that the groups are not significantly different from each other.

Table 5.3.4 ANOVA

Source of Variation	***SS***	***df***	***MS***	***F***	***P-value***	***F crit***
Between Groups	672978261	3	224326087	324.8953627	2.8893E-137	2.616202754
Within Groups	544079654.1	788	690456.4139			
Total	1217057915	791				

5.4 Discussion

The scatter plot of the furniture dataset shows a clear negative relationship between the price and sale of the furniture. As the sale of the furniture increases, the price tends to decrease as well. However, it is also clear that there is a large amount of variability in the data, with some pieces of furniture priced much higher or lower than would be expected based on their sale alone. One potential explanation for this variability is that other factors may be influencing the price of the furniture, such as the type of furniture, delivery and rate. To further explore these possibilities, additional variables could be added to the analysis, such as the material or brand of the furniture.

Table 5.4.1 Intercept and slope

	Coefficients	Standard Error
Intercept	2410.701807	204.8926518
Slope	-813.4220763	550.6778869

Simple linear regression is a powerful statistical technique that can be used to explore the relationship between two variables. In the case of the furniture dataset, simple linear regression can be used to examine the relationship between the price and sale of different pieces of furniture.

Here we can see that slope is negative and it suggests that there is an inverse relationship between the two variables, which means that as one variable increases, the other variable tends to decrease The results of the simple linear regression analysis show a statistically significant inverse relationship between the price and sale of furniture, with a regression intercept value of 2410.701807 and slope value of -813.4220763. Here we used Y=a+bx equation for regression analysis where Y is price and x is sale.

ANOVA can be used to test whether there is a statistically significant difference in the mean price of different types of furniture, such as chairs, sofas, and tables. The ANOVA results for the furniture dataset show that there is a statistically significant difference in the mean price of the different types of furniture.

5.5 Results

Overall, the scatter plot of the furniture dataset provides a useful starting point for exploring the relationship between the price and sale of furniture. By examining the patterns and variability in the data, researchers can gain valuable insights into the factors that influence the price of different pieces of furniture and develop more accurate and nuanced models of pricing in the furniture industry.

The results of the regression analysis showed that the slope coefficient for the independent variable (sale) was positive, indicating an inverse relationship between sale and price. This suggests that as the sale of the furniture increases, the price tends to decrease. The intercept coefficient was positive, indicating that even when the size of the furniture is zero, the predicted price is still greater than zero. The coefficient of determination (R-squared) was 0.010 indicating that the variation in price could be explained by the variation in sale. This suggests that while sale is an important factor in determining the price of furniture.

6.CONCLUSION

In conclusion, the furniture dataset provides valuable insights into the factors that affect the price of furniture. The dataset includes information about the type, size, and price of various pieces of furniture, and has been used to explore the relationships between these variables using various statistical techniques.

The analysis of the furniture dataset using scatterplots and simple linear regression has shown that there is an inverse relationship between the sale of furniture and its price, indicating that higher the discount furniture tends to be priced lower. However, this relationship is not particularly strong, suggesting that other factors, such as quality, design, and materials, are also important in determining the price of furniture.

The insights gained from this analysis can be used by manufacturers, retailers, and consumers to make more informed decisions about the pricing and purchase of furniture.

7.REFRENCES

1. Myatt, G. J., & Johnson, W. P. (2009). *Making sense of data II: a practical guide to data visualization, advanced data mining methods, and applications.* Hoboken, N.J., John Wiley & Sons.
2. Kothari, C. and Garg, G., 2014. *Research methodology Methods and Techniques*. 3rd Ed. New Delhi: New Age International (P) Ltd., p.63.