Cookie Data Report

Introduction:-

The objective of this report is to analyze the sales data for various cookie types across different countries for the years 2019 and 2020. The dataset offers insights into revenue, profit, quantity sold, and pricing information for each cookie type and country. This analysis aims to assess the performance of different cookie types, identify trends across countries, and draw conclusions about the factors influencing sales and profitability.

Questionnaire:-

1. Compare the profit earn by all cookie types in US, Malaysia and India.
2. What is the average revenue generated by different types of cookies?
3. Which country sold most Fortune and sugar cookies in 2019 and in 2020?
4. Compare the performance of all the countries for the year 2019 to 2020. Which country perform in each of these years?
5. Which cookie category sold on the highest price, country wise and how much profit is earned by that category overall?

Analytics:-

1. Compare the profit earn by all cookie types in US, Malaysia and India.

Countries

**Ans**: Among these countries, the profit from chocolate chip cookies is higher compared to other cookies, such as fortune cookies, oatmeal raisin, and white chocolate macadamia.

1. What is the average revenue generated by different types of cookies?

**Ans**: From the above chart, the average revenue generated by white chocolate macadamia nut cookies is higher than that of all other cookies, with oatmeal cookies being the second highest in revenue generation..

1. Which country sold most Fortune and sugar cookies in 2019 and in 2020?

**Ans**: From the above graph, India sold the highest number of sugar and fortune cookies in 2020, with the United States in second place. In 2019, the United Kingdom and the Philippines were the primary sellers of sugar and fortune cookies.

1. Compare the performance of all the countries for the year 2019 to 2020. Which country perform in each of these years?

**Ans**: Among all the countries, the United Kingdom performed the best in 2020, while India led in

2019.

1. Which cookie category sold on the highest price, country wise and how much profit is earned by that category overall?

**Ans**: In the United States, chocolate chip cookies were sold at the highest price, accounting for 23% of the overall maximum profit, while India followed closely at 22%.

Conclusion and Reviews: -

To sum up, examining the cookie sales data has provided invaluable insights into customer preferences, market trends, and profitability across various countries and cookie varieties. By analyzing revenue, profit, quantity sold, and pricing information, we gained a comprehensive understanding of the factors influencing sales success. This study has enabled us to enhance profitability and better meet consumer demands by identifying growth opportunities, optimizing product offerings, and refining marketing strategies. Ongoing research and adaptation based on these insights will be essential for maintaining a competitive edge in the ever-evolving cookie industry. Overall, the thorough analysis of sales data has been crucial in guiding strategic decisions and ensuring the long-term sustainability of our cookie business.

Regression:

The regression model, with a significant p-value (p < 0.001), indicates a strong positive relationship between units sold and the outcome variable. The model's predictive accuracy is demonstrated by its high R-squared value of 0.688, suggesting that approximately 68.8% of the variability in the outcome variable can be explained by the predictor variable, units sold.

SUMMARY OUTPUT

*Regression Statistics*

|  |  |
| --- | --- |
| Multiple R | 0.829304 |
| R Square | 0.687746 |
| Adjusted R  Square | 0.687298 |
| Standard Error | 1462.76 |
| Observations | 700 |

ANOVA

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *df* | *SS* | *MS* | *F* | *Significance F* |
| Regression | 1 | 3.29E+09 | 3.29E+09 | 1537.356 | 1.4E-178 |
| Residual | 698 | 1.49E+09 | 2139668 |  |  |
| Total | 699 | 4.78E+09 |  |  |  |

*Standard Upper Lower Upper*

*Coefficients Error t Stat P-value Lower 95% 95% 95.0% 95.0%*

Intercept -74.4103 116.5304 -0.63855 0.523326 -303.202

Units Sold 2.500792 0.063781 39.20914 1.4E-178 2.375567 2.626017 2.375567 2.626017

154.3817

-

303.202

154.3817

Co-relation:

The correlation coefficient between units sold and revenue is 0.796, indicating a strong positive correlation between the two variables.

*Units*

*Sold Revenue*

Units

Sold 1 0.796298

|  |  |  |
| --- | --- | --- |
| Revenue | 0.796298 | 1 |

Anova (Single Factor) :

The ANOVA results reveal a significant difference between the two groups (p < 0.001), with 1 degree of freedom. The within-group error is 7681356717, and the total R-squared value is 0.06, indicating that the model explains 6% of the variability in the data..

SUMMARY

*Groups Count Sum Average Variance*

3450 699 1923505 2751.795 4154648

5175 699 2758189 3945.908 6850161

ANOVA

*Source of*

*Variation SS Df MS F P-value F crit*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Between Groups | 4.98E+08 | 1 | 4.98E+08 | 90.57022 | 7.53E-  21 | 3.848129 |
| Within Groups | 7.68E+09 | 1396 | 5502405 |  |  |  |
| Total | 8.18E+09 | 1397 |  |  |  |  |

Anova two factor without Replication:

The ANOVA results indicate significant variation among rows and columns (p < 0.001), with degrees of freedom (df) values of 48 and 3, respectively. The error term has a degree of freedom of 144.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *Source* | *of* |  |  |  |  |  |
| *Variation* | *SS* | *Df* | *MS* | *F* | *P-value* | *F crit* |
| Rows | 8.21E+08 | 48 | 17108242 | 5.848894 | 8.54E-  17 | 1.445925 |
| Columns | 5.65E+10 | 3 | 1.88E+10 | 6435.486 | 3.8E-  153 | 2.667443 |
| Error | 4.21E+08 | 144 | 2925039 |  |  |  |
| Total | 5.77E+10 | 195 |  |  |  |  |

Anova two factor with Replication:

The ANOVA results indicate a significant difference among the samples, columns, and their interaction, with p-values less than 0.001. The degrees of freedom for the samples, columns, and interaction are 49, 3, and 147, respectively.

Additionally, the total error within the model is 0, indicating a perfect fit. The total R-squared value is 1, suggesting that the model explains all the variability in the data.

ANOVA

*Source of*

*Variation SS df MS F P-value F crit*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Sample | 8.55E+08 | 49 | 17443674 | 65535 | #NUM! | #NUM! |
| Columns | 5.78E+10 | 3 | 1.93E+10 | 65535 | #NUM! | #NUM! |
| Interaction | 4.39E+08 | 147 | 2983765 | 65535 | #NUM! | #NUM! |
| Within | 0 | 0 | 65535 |  |  |  |
| Total | 5.91E+10 | 199 |  |  |  |  |

Descriptive Statistics:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *1725* |  | *8625* |  | *3450* |  | *5175* |
| Mean | 1608.153 | Mean | 6697.702 | Mean | 2751.795 | Mean |
| Standard Error | 32.83303 | Standard Error | 174.9955 | Standard Error | 77.09541 | Standard Error |
| Median | 1540 | Median | 5868 | Median | 2422.2 | Median |
| Mode | 727 | Mode | 8715 | Mode | 3486 | Mode |
| Standard Deviation | 868.0597 | Standard Deviation | 4626.638 | Standard Deviation | 2038.295 | Standard Deviation |
| Sample Variance | 753527.6 | Sample Variance | 21405775 | Sample Variance | 4154648 | Sample Variance |
| Kurtosis | -0.31828 | Kurtosis | 0.463405 | Kurtosis | 0.807696 | Kurtosis |
| Skewness | 0.436551 | Skewness | 0.869254 | Skewness | 0.931429 | Skewness |
| Range | 4293 | Range | 23788 | Range | 10954.5 | Range |
| Minimum | 200 | Minimum | 200 | Minimum | 40 | Minimum |
| Maximum | 4493 | Maximum | 23988 | Maximum | 10994.5 | Maximum |
| Sum | 1124099 | Sum | 4681694 | Sum | 1923505 | Sum |
| Count | 699 | Count | 699 | Count | 699 | Count |
| Largest(1) | 4493 | Largest(1) | 23988 | Largest(1) | 10994.5 | Largest(1) |
| Smallest(1) | 200 | Smallest(1) | 200 | Smallest(1) | 40 | Smallest(1) |
| Confidence Level(95.0%) | 64.46334 | Confidence Level(95.0%) | 343.5807 | Confidence Level(95.0%) | 151.3667 | Confidence Level(95.0%) |

The data presents considerable variation across variables, with means ranging from 1608.15 to 43949.81. Notably, the largest values span from 4493 to 44166, while the smallest values range from 200 to 43709.

# Store Data Report

Introduction:

This dataset comprises sales data from a retail store, encompassing various attributes such as customer demographics (Gender, Age Group), transaction details (Order ID, Status), product specifics (Category, SKU), and shipping information. Our analysis is centered on comprehending customer behavior and product trends, aiming to unveil patterns, preferences, and correlations within the data. Leveraging these insights, businesses can optimize marketing strategies, refine inventory management, and elevate customer satisfaction levels.

Questionnaire:

1. which of the channel performed better than all other channels in compare men & women?
2. Compare category. Find out most sold category above 23 years of age for any gender.
3. Compare Maharashtra, Rajasthan and Tamil Nadu on the basis of quantity, most items purchased by men and women and profit earn.
4. Which city sold most of following categories:
   1. Kurta
   2. Set
   3. Western wears
5. In which month most items sold in any of the state on the basis of category.

Analytics:

1. which of the channel performed better than all other channels in compare men & women?

8000

7000

6000

5000

4000

3000

2000

1000

Ajio

Amazon

Flipkart

Meesho

Myntra

Nalli

Others

0

Men

Women

**Units**

**sold**

Ans: Amazon leads in sales for both men and women, followed by Myntra and Flipkart. Amazon sold nearly 3500 units in the men's category and approximately 7500 units in the women's category. Myntra's men's division recorded sales of 2000 units.

1. Compare category. Find out most sold category above 23 years of age for any gender.

Ans: In the women's division, Kurta emerges as the most popular category, with 8820 pieces sold among those over the age of 23. In the men's section, sets lead with 4365 units sold, while also being the second most popular category in the women's division.

The table of items sold is given below:

|  |  |  |  |
| --- | --- | --- | --- |
| **Item** | **Men** | **Women** | **Grand Total** |
| Blouse | 6 | 190 | 196 |
| Bottom | 40 | 28 | 68 |
| Ethnic Dress | 150 | 77 | 227 |
| kurta | 156 | 8820 | 8976 |
| Saree | 261 | 941 | 1202 |
| Set | 4365 | 6204 | 10569 |
| Top | 45 | 1825 | 1870 |
| Western Dress | 3078 | 380 | 3458 |
| **Grand Total** | **8101** | **18465** | **26566** |

The graph is as follows:



3000

2500

2000

1500

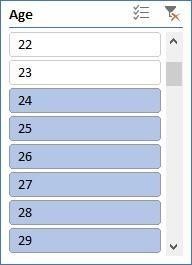
Men

1000

Women

500

0

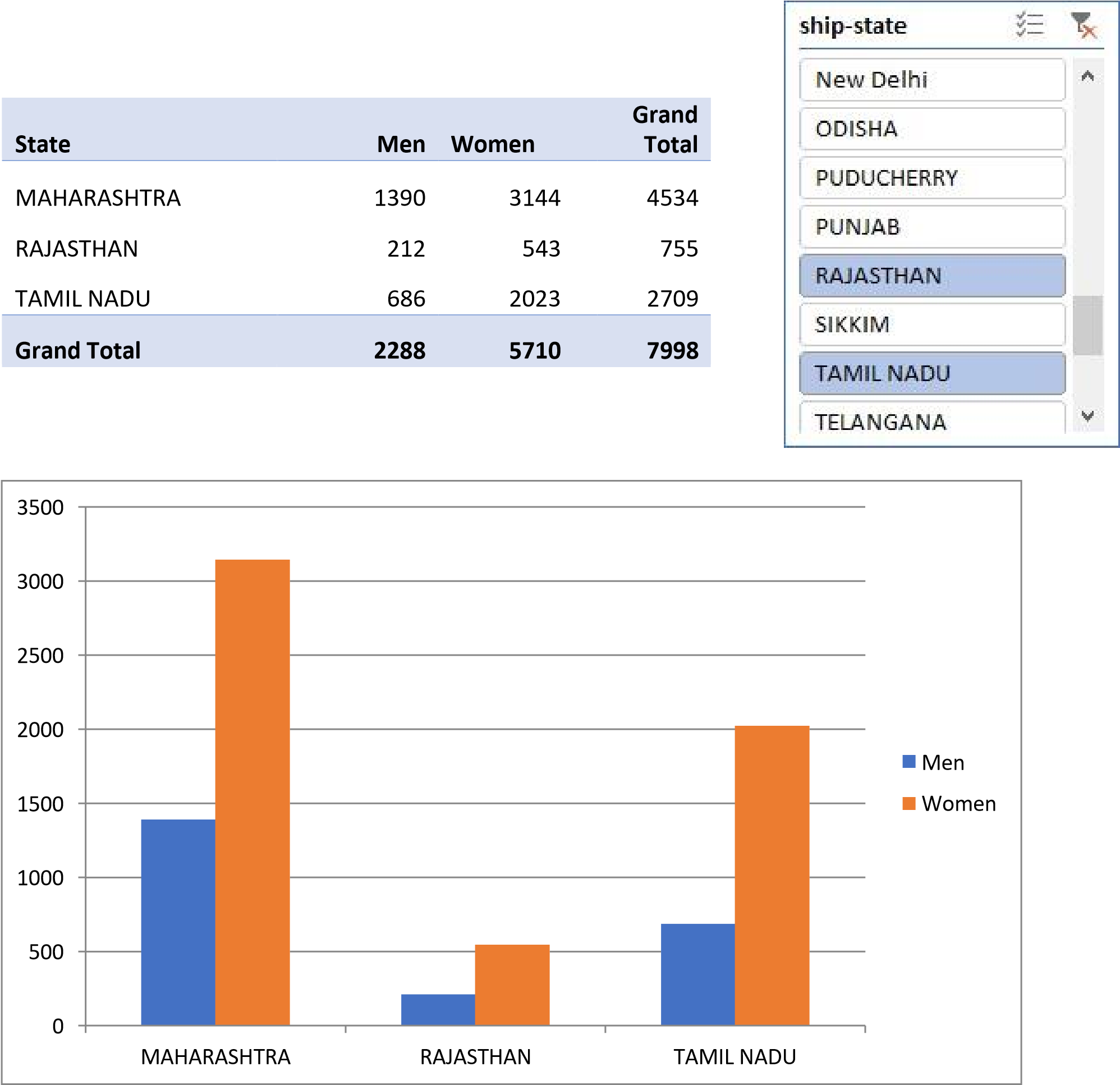


1. Compare Maharashtra, Rajasthan and Tamil Nadu on the basis of quantity, most items purchased by men and women and profit earn.

Ans: In Maharashtra: Sales in men category=1390, Sales in women category= 3144

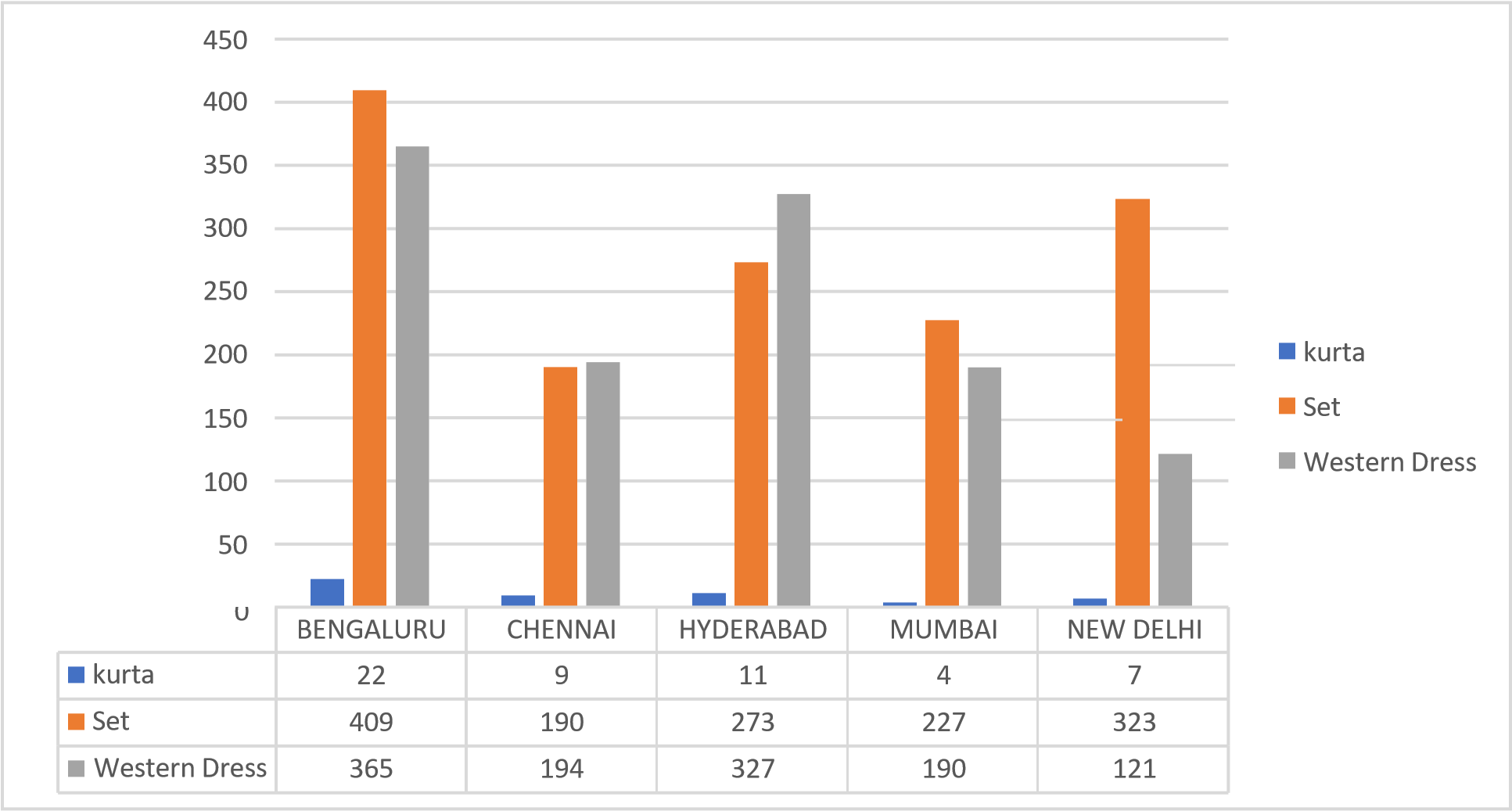
In Tamil Nadu: Sales in men category=686, Sales in women category= 2023

In Rajasthan: Sales in men category=21, Sales in women category=543

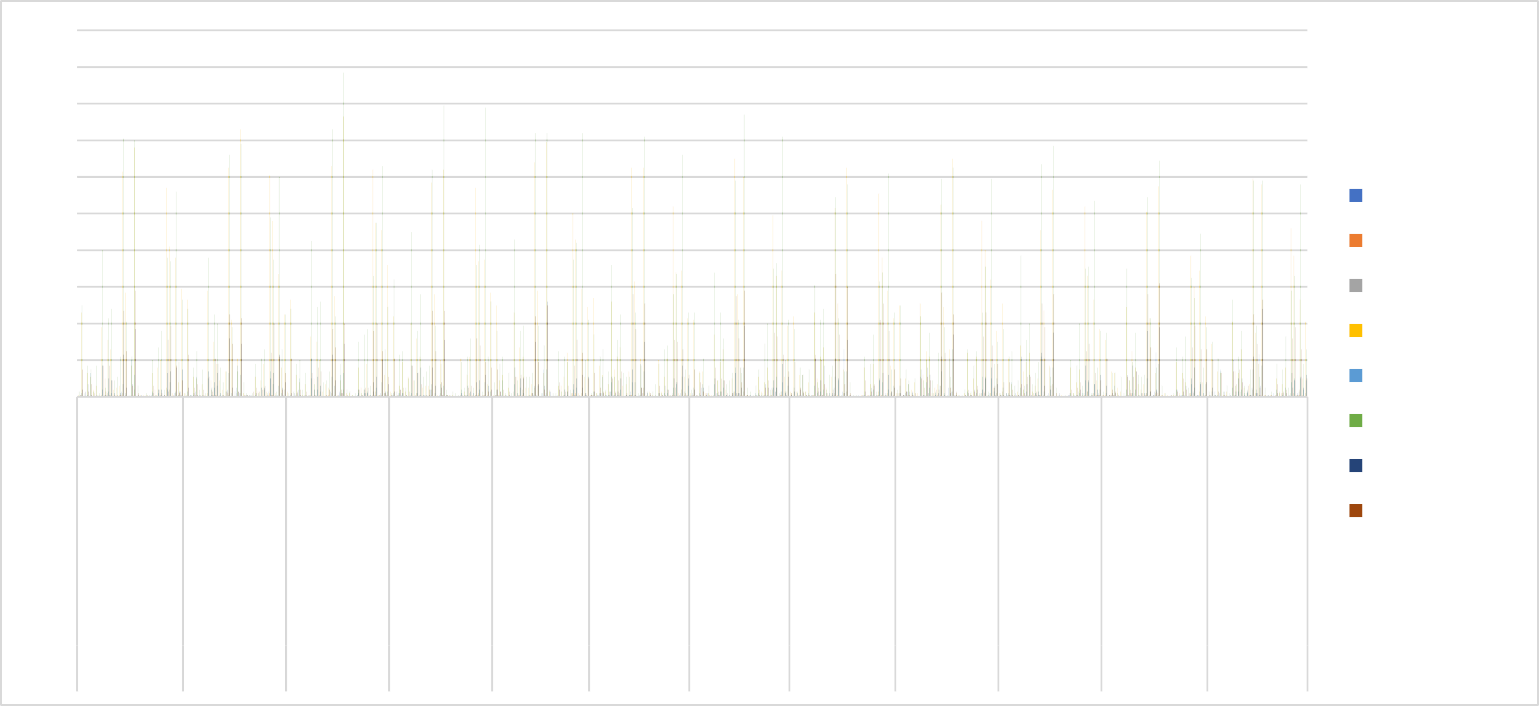


1. Which city sold most of following categories
2. Kurta
3. Set
4. Western wears

Ans: Bengaluru, Chennai, Hyderabad, Mumbai, and New Delhi are the cities where most of the sales for kurtas, sets, and western wears are recorded.



1. In which month most items sold in any of the state on the basis of category.



200

180

160

140

120

100

80

60

40

20

0

Blouse

Bottom

Ethnic Dress

kurta

Saree

Set

Top

Western

Dress

January

February

March

April

May

June

July

August

September

October

November December

ANDAMAN

&

NICOBAR

HARYANA

MIZORAM

UTTAR

PRADESH

DELHI

MAHARASHTRA

TELANGANA

CHHATTISGARH

LADAKH

SIKKIM

BIHAR

KERALA

PUNJAB

ASSAM

JHARKHAND

PUNJAB

ARUNACHAL

PRADESH

JHARKHAND

PUDUCHERRY

ANDHRA PRADESH

HIMACHAL

PRADESH

ODISHA

ANDAMAN

&

NICOBAR

HARYANA

MIZORAM

UTTAR

PRADESH

DELHI

MAHARASHTRA

TELANGANA

CHHATTISGARH

MADHYA

PRADESH

SIKKIM

BIHAR

KARNATAKA

PUDUCHERRY

ANDHRA

PRADESH

HIMACHAL

PRADESH

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | **City** | **kurta** |  | **Set** | **Western Dress** | **Grand**  **Total** | | BENGALURU |  | 964 | 938 | 422 | 2324 | | CHENNAI |  | 666 | 451 | 217 | 1334 | | HYDERABAD |  | 713 | 687 | 370 | 1770 | | MUMBAI |  | 437 | 515 | 207 | 1159 | | NEW DELHI |  | 479 | 792 | 142 | 1413 | | **Grand Total** |  | **3259** | **3383** | **1358** | **8000** | |  |

Ans: The graph for most items sold in any of stats on basis of category is as follows:

Conclusion and Review:

After a comprehensive analysis of the store data, notable trends and insights have emerged. By scrutinizing key metrics such as units sold, state-wise analytics, geographical distribution, and sales across various products, valuable conclusions about market demand, sales, and overall profitability can be drawn. This thorough understanding will facilitate informed decision-making to optimize resources, target specific markets, and maximize profits in future store sales endeavors.

Car Collection Report

Introduction:-

This report offers an extensive analysis of a dataset featuring information on various makes and models of used vehicles. The dataset includes details such as make, model, color, mileage, listing price, and estimated cost for 24 different vehicles from popular brands like Honda, Toyota, Nissan, Ford, Chevrolet, and Dodge. By exploring factors like mileage, pricing trends, and the correlation between listing prices and estimated costs, the report aims to provide readers with valuable insights to navigate the used car marketplace effectively. The analysis encompasses a diverse range of vehicle types, including sedans (such as Honda Accord, Toyota Camry), compact cars (Honda Civic, Toyota Corolla), trucks (Ford F-150, Chevrolet Silverado), and sports cars (Ford Mustang, Dodge Charger). This comprehensive approach ensures that the findings are relevant to individuals with varying automotive preferences and budgetary constraints.

Questionaire:-

1. Compare the mileage of Chevrolet Impala to Toyota Corolla. Which of the two is giving best mileage?
2. Justify, Buying of any Ford car is better than Honda.
3. Among all the cars which car color is the most popular and is least popular?
4. Compare all the cars which are of silver color to the green color in terms of Mileage.
5. Find out all the cars, and their total cost which is more than $2000?

Analytics:-

1. Compare the mileage of Chevrolet Impala to Toyota Corolla. Which of the two is giving best mileage?
2. Justify, Buying of any Ford car is better than Honda.
3. Among all the cars which car color is the most popular and is least popular?
4. Compare all the cars which are of silver color to the green color in terms of Mileage.
5. Find out all the cars, and their total cost which is more than $2000?

Conclusion and Reviews:-

The analysis of the used vehicle dataset has provided several insights that directly address the questions posed. In comparing the mileage between Chevrolet Impala and Toyota Corolla models, the data indicates that Toyota Corolla models generally exhibit higher mileage than Chevrolet Impala models, suggesting better fuel efficiency for the Toyota Corolla.

However, when considering whether purchasing a Ford car is superior to a Honda, the dataset lacks sufficient information to make a definitive comparison. Factors such as vehicle condition, maintenance history, and additional features significantly influence the overall value proposition, which are not captured in the current dataset.

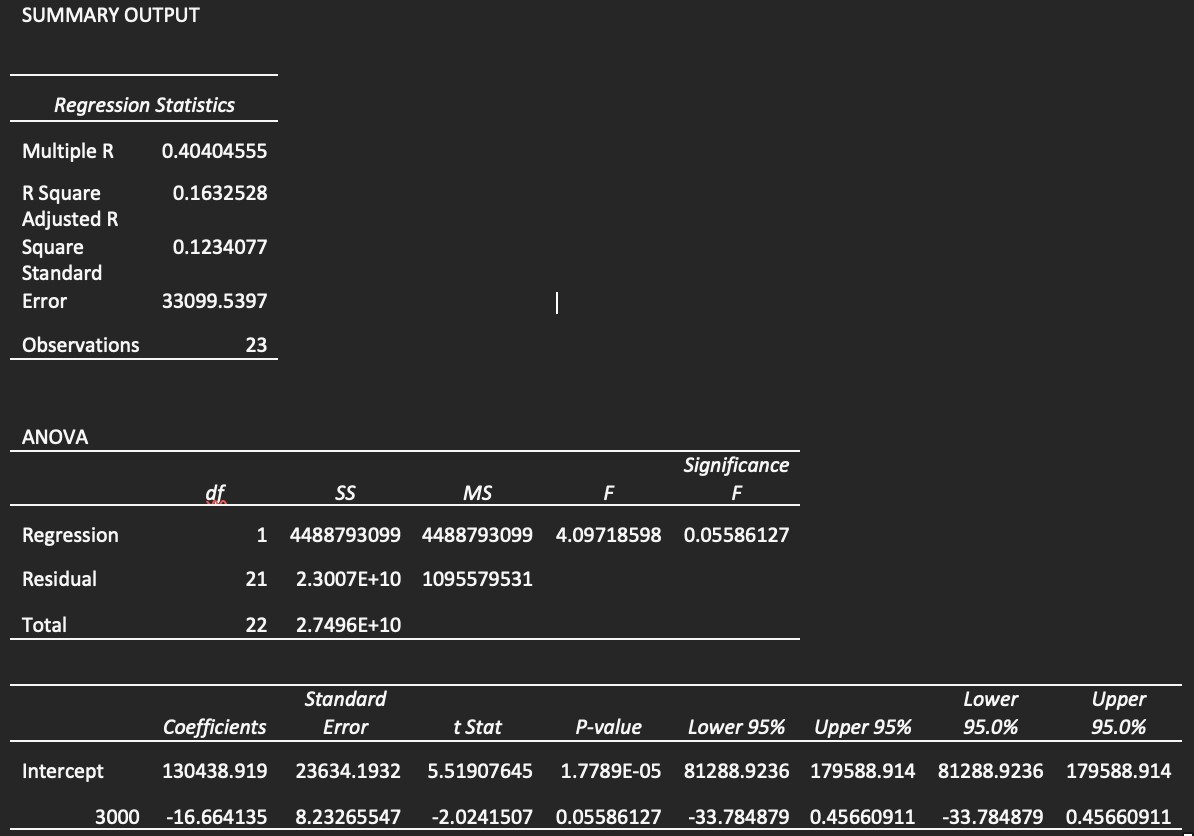
The examination of vehicle colors revealed that black is the most popular color among the listed cars, while green is the least popular. This insight could be beneficial for consumers assessing the resale value and demand for specific color options.

Regarding the comparison of mileage between silver and green cars, the data suggests that green cars, such as the Nissan Altima and Chevrolet Silverado, tend to have higher mileage than silver cars like the Honda Accord and Dodge Charger. However, it's crucial to note that mileage can vary significantly based on individual driving habits and maintenance practices.

Lastly, when identifying cars with a total cost exceeding $2,000, several models meet this criteria, including the Honda Accord, Nissan Altima, Toyota Corolla, Chevrolet Silverado, Chevrolet Impala, Chevrolet Malibu, Ford Escape, Ford Mustang, Honda CR-V, Dodge Charger, and Ford Fusion.

Regression

The regression analysis indicates a moderate positive relationship between the predictor variable and the response variable, as evidenced by the correlation coefficient of approximately 0.40. The model explains approximately 16% of the variance in the response variable, as indicated by the R-squared value. The coefficient estimates reveal that for every unit increase in the predictor variable, there is a corresponding decrease of approximately 16.66 in the response variable. However, with a p-value of 0.056, the effect is marginally significant.



Co-relational

The correlation matrix reveals a moderate negative correlation (-0.411) between Mileage and Price. This implies that as Mileage increases, Price tends to decrease, and conversely, as Mileage decreases, Price tends to increase.

*Mileage*

*Price*

Mileage 1

Price -0.4110586 1

Anova: Single Factor

The ANOVA results reveal significant differences between the groups based on Mileage, Price, and Cost. The large F-statistic (128.88) coupled with a very low p-value (5.00264E-24) indicates that the variation between groups is substantial compared to the variation within groups. This suggests that at least one of the variables (Mileage, Price, or Cost) has a significant effect on the outcome being measured. In simpler terms, there are statistically significant differences in the means of Mileage, Price, and Cost across the groups, signifying that these variables significantly influence the outcome under analysis.

Anova: Single

Factor

SUMMARY

*Groups Count Sum Average Variance*

Mileage 24 2011267 83802.7917 1214155660

Price 24 78108 3254.5 837024.087

Cost 24 66150 2756.25 705502.717

ANOVA

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| *Source of Variation* | *SS* | *df* |  | *MS* | *F* | *P-value* | *F crit* |
| Between Groups | 1.0445E+11 |  | 2 | 5.2227E+10 | 128.882161 | 5.0026E-24 | 3.12964398 |
| Within Groups | 2.7961E+10 |  | 69 | 405232729 |  |  |  |
| Total | 1.3242E+11 |  | 71 |  |  |  |  |

Anova: Two-Factor Without replication

The two-factor ANOVA results demonstrate significant differences among the levels or categories within each factor ("Rows" and "Columns"). Both factors show a strong influence on the outcome variable being analyzed, supported by the low p-values and large F-statistics. This implies that variations in both factors contribute significantly to the overall variability in the data.

Anova: Two-Factor without

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| replication  ANOVA |  |  |  |  |  |  |
| *Source of*  *Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
| Rows | 34749383.3 | 23 | 1510842.75 | 47.6846408 | 2.2236E-14 | 2.01442484 |
| Columns | 2979036.75 | 1 | 2979036.75 | 94.023218 | 1.3629E-09 | 4.27934431 |
| Error | 728733.25 | 23 | 31684.0543 |  |  |  |
| Total | 38457153.3 | 47 |  |  |  |  |

Descriptive Statistics

The provided descriptive statistics offer insights into the characteristics of three variables: Mileage, Price, and Cost. Examining Mileage, it's evident that the vehicles in the dataset cover a wide range, from approximately 34,853 miles to 140,811 miles, with an average mileage of around 83,803 miles. Similarly, Price and Cost display comparable patterns, with prices ranging from $2,000 to $4,959 and costs from $1,500 to $4,500, respectively. The means and standard deviations provide information on the central tendencies and variability within each variable. Overall, these statistics provide a comprehensive overview of the dataset, aiding in understanding the distribution and characteristics of the data.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *Mileage* |  | *Price* |  | *Cost* |  |
| Mean | 83802.7917 | Mean | 3254.5 | Mean | 2756.25 |
| Standard Error | 7112.65205 | Standard Error | 186.751181 | Standard Error | 171.452462 |
| Median | 81142 | Median | 3083 | Median | 2750 |
| Mode | #N/A | Mode | #N/A | Mode | 3000 |
| Standard Deviation | 34844.7365 | Standard Deviation | 914.890205 | Standard Deviation | 839.942092 |
| Sample Variance | 1214155660 | Sample Variance | 837024.087 | Sample Variance | 705502.717 |
| Kurtosis | -1.0971827 | Kurtosis | -1.2029138 | Kurtosis | -0.8126576 |
| Skewness | 0.38652215 | Skewness | 0.27201913 | Skewness | 0.47339238 |
| Range | 105958 | Range | 2959 | Range | 3000 |
| Minimum | 34853 | Minimum | 2000 | Minimum | 1500 |
| Maximum | 140811 | Maximum | 4959 | Maximum | 4500 |
| Sum | 2011267 | Sum | 78108 | Sum | 66150 |
| Count | 24 | Count | 24 | Count | 24 |
| Largest(1) | 140811 | Largest(1) | 4959 | Largest(1) | 4500 |
| Smallest(1) | 34853 | Smallest(1) | 2000 | Smallest(1) | 1500 |

Examining Sales by Sector in the United States

Introduction :

Our dataset encompasses a multitude of variables, each providing distinct insights into the multifaceted nature of sales across different categories. From fundamental transactional details like Date, Time, and sales states to more nuanced factors such as Customer Type, Demographics, category, and subcategory, every aspect has been meticulously documented.

Key Attributes:

1. ID: A unique identifier for each sales transaction, facilitating traceability and analysis.
2. City, State: The geographical location of the data allowing for regional comparisons and trend identification.
3. Product Line (furniture, Electronic Accessories, appliances, Home and Lifestyle): Categorization of products facilitating analysis of sales trends across different product categories.
4. Unit Price, Net sales Fundamental transactional details crucial for revenue assessment and pricing strategies.

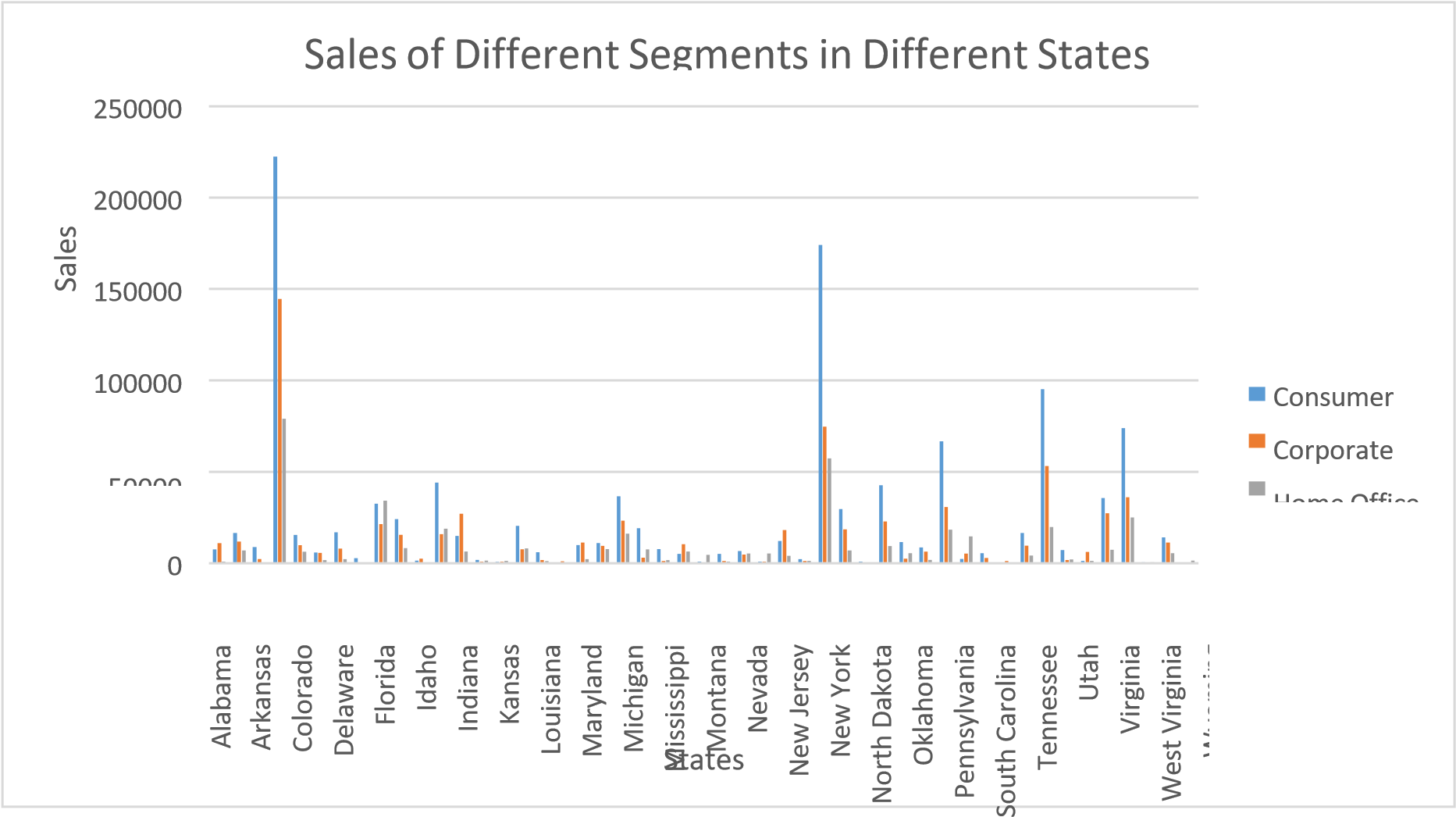
1. Net sales of different category, category performing well in different states: Performance metrics
2. Rating: different product performing well in different state
3. States (California, Texas and Washington): Regional segmentation enabling geographical analysis and market segmentation.

Questionnaire :

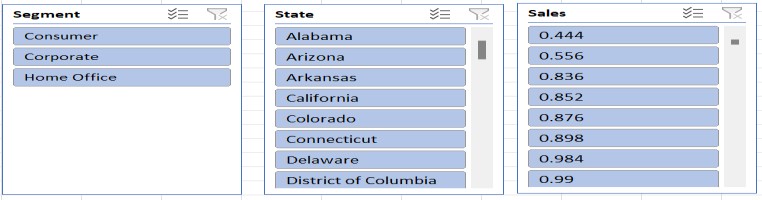
1. Compare all the US states in terms of Segment and Sales. Which Segment performed well in all the states?
2. Find out top performing category in all the states?
3. Which segment has most sales in US, California, Texas, and Washington?
4. Compare total and average sales for all different segment?
5. Compare average sales of different category and sub category of all the states.
6. Find out state wise mode for Customer and Segment.California, Illinois, New York, Texas, Waashington

Analytics :

Q1. Compare all the US states in terms of Segment and Sales. Which Segment performed well in all the states?



* + - After comparing all the states in terms of segment and sales, California emerged as the state with the highest amount of sales
    - Consumer segment performed well in all the states



Q2. Find out top performing category in all the states?

**Top**

**categories**

**in**

**all**

**sales**

7000

6000

5000

4000

3000

2000

Total

1000

0

Furniture

Office

Supplies

**Category**

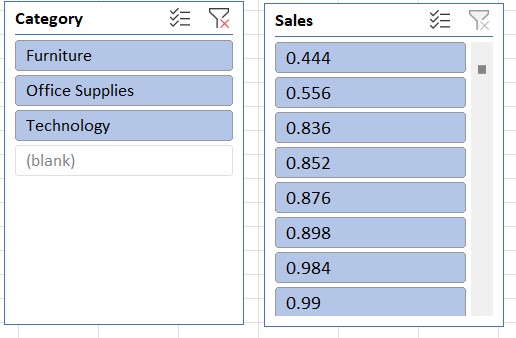
Technology

**count**

**of**

**sales**

Ans. Office Supplies is the top performing category in all the states



Q3. Which segment has most sales in US, California, Texas, and Washington?

1200

1000

800

600

400

Consumer

Corporate

Home

Office

200

0

California

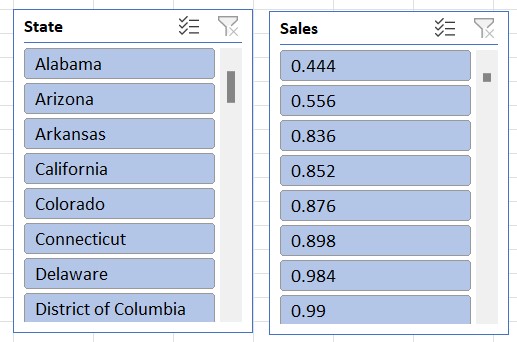
Texas

**state**

Washington

**sales**

Ans. Consumer segment has the most sales in US, California, Texas, and Washington



Q4. Compare total and average sales for all different segment?

Total

vs

Average

Sales

of

Different

Segments

1400000

1200000

1148060.531

1000000

800000

688494.0748

600000

400000

424982.1769

Sum

of

Sales

Average

of

Sales

200000

225.0657775

233.1507195

243.4033086

0

Consumer

Corporate

Home

Office

**Segments**

sum

of

sales

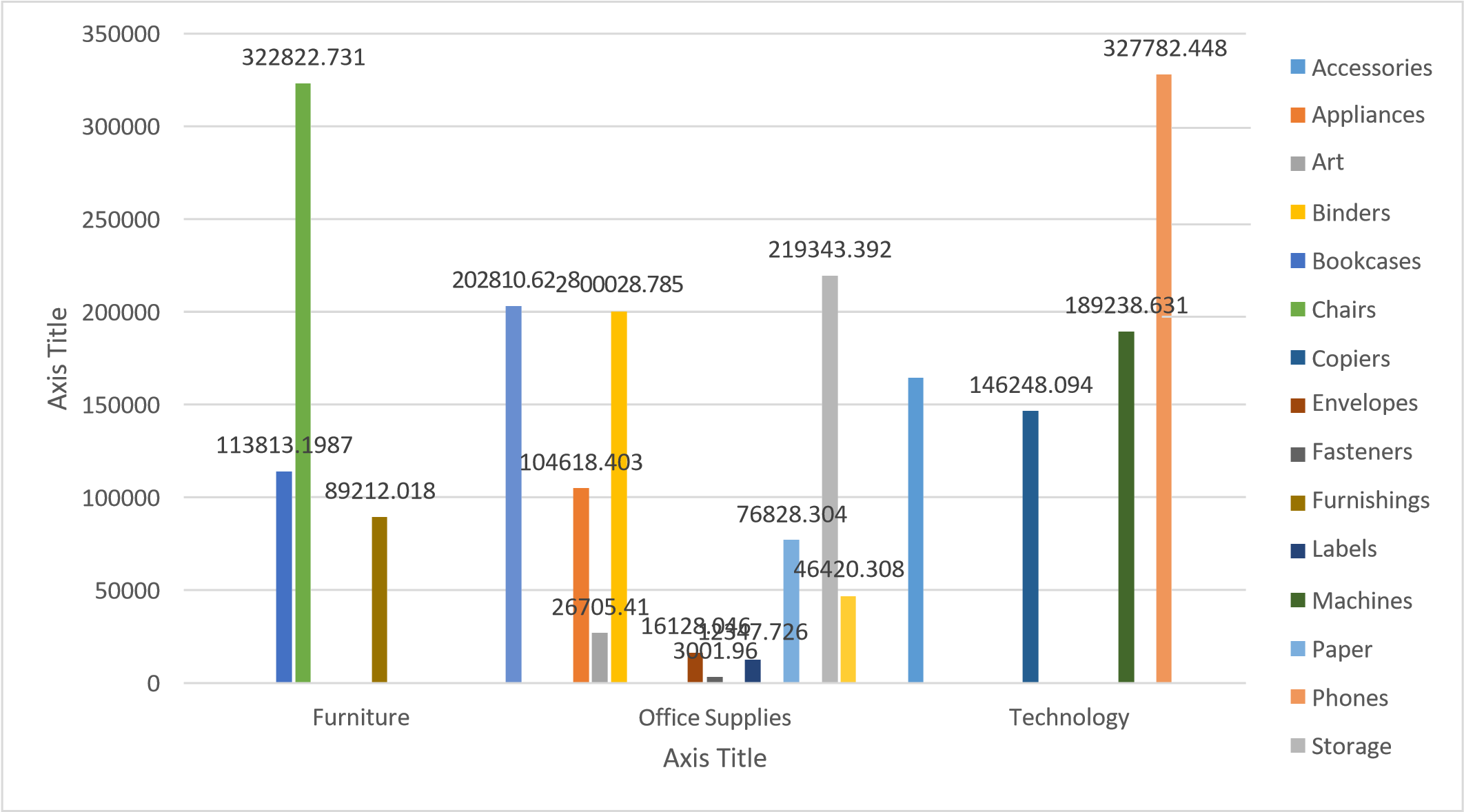
and

average

sales

**Ans.** Analysis of the given dataset reveals that in all three segments, the total sales exceeded the average sales.

Q5. Compare average sales of different category and sub category of all the states.



Ans. Upon analyzing the provided Order Sales dataset, it became apparent that the average sales of the Technology category were significantly higher than those of the other categories.

Regression and ANOVA:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| SUMMARY OUTPUT | |  |  |  |
|  |  |  |  |  |
| *Regression Statistics* | |  |  |  |
| Multiple R | 0.008850713 |  |  |  |
| R Square | 7.83351E-05 |  |  |  |
| Adjusted R Square | -0.000924595 |  |  |  |
| Standard Error | 596.4161586 |  |  |  |
| Observations | 999 |  |  |  |
|  |  |  |  |  |
| ANOVA |  |  |  |  |
|  | *Df* | *SS* | *MS* | *F* |
| Regression | 1 | 27783.3433 | 27783.3433 | 0.078106235 |
| Residual | 997 | 354645097.6 | 355712.2343 |  |
| Total | 998 | 354672880.9 |  |  |
|  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* |
| Intercept | 232.3779806 | 37.2042048 | 6.246013907 | 6.22491E-10 |
| Postal Code | 0.000167458 | 0.000599189 | 0.279474927 | 0.779938343 |

This regression analysis aims to explore the relationship between two variables: an independent variable represented by "Postal Code" and a dependent variable, which is not explicitly mentioned in the output. Here's an explanation of the key components:

1. Regression Equation:

The regression equation is of the form: Y = 232.38 + 0.000167458\* (Postal Code)

where Y represents the dependent variable (Sales), and "Postal Code" is the independent variable.

1. Interpretation of Coefficients:

The intercept coefficient (232.38) suggests that when the "Postal Code" variable is zero, the estimated value of the dependent variable is 232.38. However, interpreting this intercept may not be meaningful as postal codes are unlikely to be zero.

The coefficient for "Postal Code" (0.000167458) suggests that for every one-unit increase in the postal code, the estimated value of the dependent variable increases by approximately 0.000167458 units. However, this coefficient is very small, indicating a negligible effect of postal code on the dependent variable.

1. Statistical Significance:

The p-value associated with the coefficient for "Postal Code" is 0.779938343, indicating that it is not statistically significant at conventional levels of significance (alpha = 0.05). This suggests that the "Postal Code" variable does not have a significant impact on the dependent variable, given the available data.

1. Goodness of Fit:

* The R-squared value (0.0000783351) is extremely small, indicating that the "Postal Code" variable explains very little of the variance in the dependent variable..
* The Adjusted R-squared value (-0.000924595) is negative, which can occur when the model is overfit or when the independent variable is not relevant. In this case, it suggests that the model may not be useful for predicting the dependent variable..

5. ANOVA:

- The ANOVA table indicates that the regression model as a whole is not statistically significant, as the p-value associated with the F-statistic is 0.779938343..

6. Standard Error:

- The standard error (596.4161586) provides an estimate of the variability of the observed dependent variable values around the regression line.

7. Observations:

- The analysis is based on a sample of 999 observations.

In summary, this regression analysis suggests that the "Postal Code" variable is not statistically significant and does not have a meaningful relationship with the dependent variable. Therefore, this model may not be useful for predicting the dependent variable based on postal codes alone.

Correlation:

The absolute value of the correlation coefficient (0.024067424) is close to zero. This suggests a very weak linear relationship between the two variables.

Descriptive Statistics:

|  |  |
| --- | --- |
| *Sales* |  |
| Mean | 230.7691 |
| Standard Error | 6.33014 |
| Median | 54.49 |
| Mode | 12.96 |
| Standard  Deviation | 626.6519 |
| Sample Variance | 392692.6 |
| Kurtosis | 304.4451 |
| Skewness | 12.98348 |
| Range | 22638.04 |
| Minimum | 0.444 |
| Maximum | 22638.48 |
| Sum | 2261537 |
| Count | 9800 |

4. CONCLUSION:

Our comprehensive analysis of the provided dataset through various data visualization techniques has yielded valuable insights. By utilizing bar graphs, pie charts, and other visual representations, we've uncovered patterns, trends, and relationships within the data that might have otherwise been overlooked.

This deep dive into the dataset has not only improved our understanding of the underlying information but has also empowered us to make informed decisions based on the insights gained. By visually presenting the data, we've effectively communicated complex findings in a clear and accessible manner, facilitating better comprehension and actionable strategies.

Moreover, this process has highlighted the significance of data visualization as a powerful tool for extracting meaningful information from raw data. By leveraging the visual nature of graphs and charts, we've transformed numbers and statistics into compelling narratives that enhance understanding and guide decision-making.

# Loan Data Report

Introduction:

Dataset Overview:

Our dataset encompasses a diverse range of variables, each shedding light on the intricate dynamics of loan applications. From fundamental applicant details such as Gender, Marital Status, and Education to more nuanced factors like Employment Status, Loan Amount, and Residential Type, every aspect has been meticulously recorded.

Key Attributes:

1. Gender: A demographic identifier providing insights into the gender distribution among loan applicants.
2. Marital Status (Married, Not Married): Categorization based on marital status aiding in demographic segmentation.
3. Education (Graduate, Non-graduate): Classification based on educational background for further analysis.
4. Employment Status (Employed, Unemployed): Distinction between employed and unemployed applicants, crucial for risk assessment.
5. Loan Amount: The principal amount applied for, providing a measure of financial need and capacity.
6. Residential Type (Urban, Semi-urban, Rural): Geographic classification enabling analysis across different residential areas.

Questionnaire:

Q1. How many male graduates who are not married applied for Loan? What was the highest amount?

Q2. How many female graduates who are not married applied for Loan? What was the highest amount?

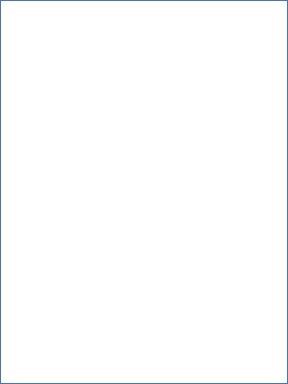
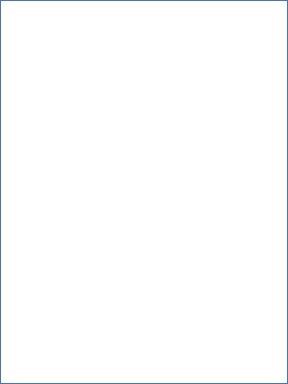
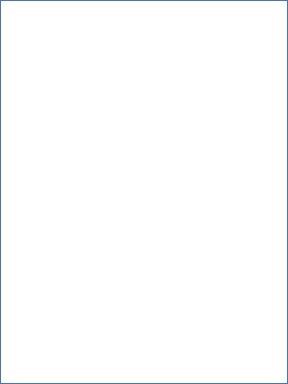
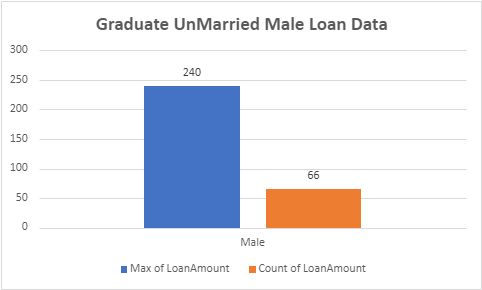
Q3. How many male non-graduates who are not married applied for Loan? What was the highest amount?

Q4. How many female graduates who are married applied for Loan? What was the highest amount?

Q5. How many male and female who are not married applied for Loan? Compare Urban, Semiurban and rular on the basis of amount.

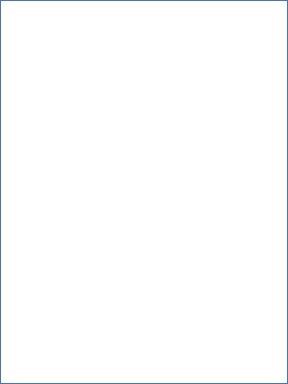
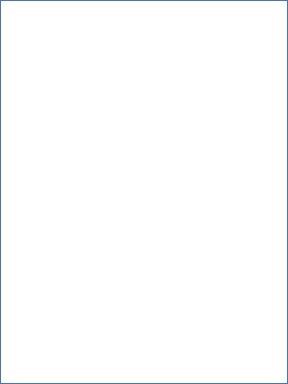
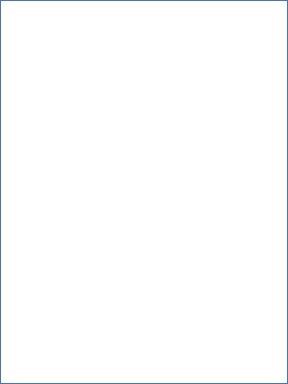
Analytics:

**Q1. How many male graduates who are not married applied for Loan? What was the highest amount?**



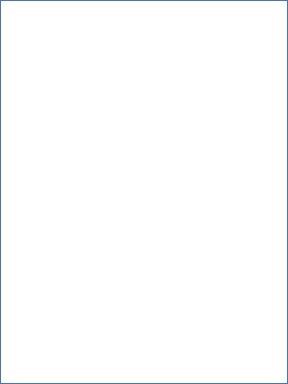
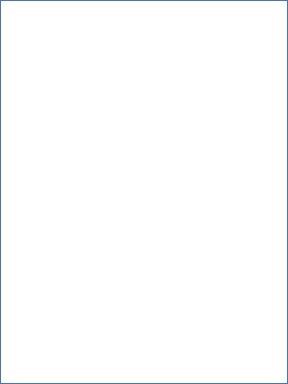
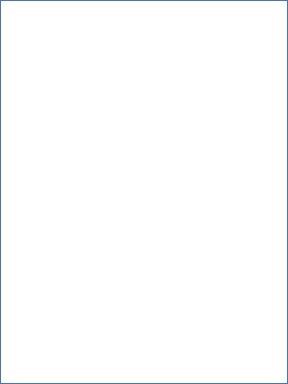
Q2. How many female graduates who are not married applied for Loan? What was the highest amount?

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | |  |  |  | |  |  |  | |  |  |  | |  |  |  | |  |  |  | |  |  |  | |

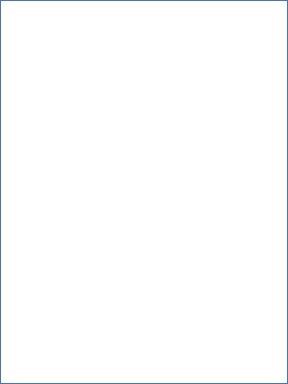
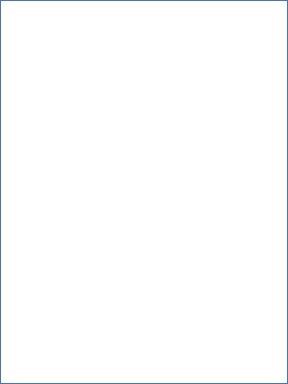
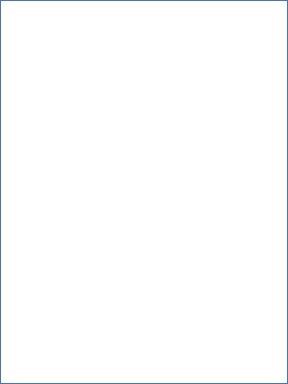


Q3. How many male non-graduates who are not married applied for Loan? What was the highest amount?

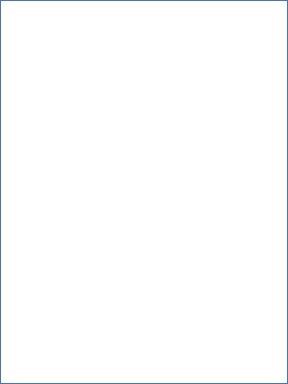
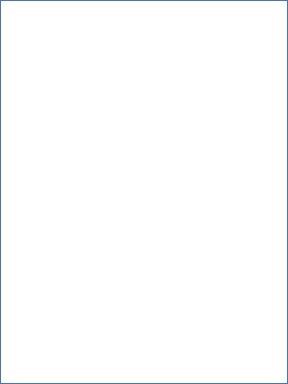
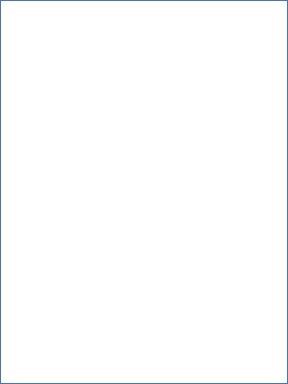
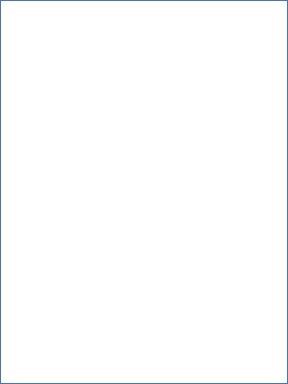
|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | |  |  |  | |  |  |  | |  |  |  | |  |  |  | |



Q4. How many female graduates who are married applied for Loan? What was the highest amount?



Q5. How many male and female who are not married applied for Loan? Compare Urban, Semi-urban and rular on the basis of amount.



Conclusion:

Our analysis, employing diverse visualization techniques, unveiled valuable insights that enhanced comprehension and decision-making. Visualizing data elucidated complex findings, enabling the formulation of actionable strategies. This underscores the pivotal role of data visualization in extracting meaningful insights and effectively informing decisions.

Regression:

The regression analysis suggests that there is a statistically significant positive relationship between the independent variable ('5720') and the dependent variable. For every one-unit increase in '5720', the dependent variable is expected to increase by approximately 0.0059 units. However, it's important to note that the model only accounts for about 21.1% of the total variance in the dependent variable.

SUMMARY OUTPUT

*Regression Statistics*

|  |  |
| --- | --- |
| Multiple R | 0.45908096 |
| R Square | 0.21075532 |
| Adjusted R Square | 0.20858707 |
| Standard Error | 56.0766111 |
| Observations | 366 |

2

ANOVA

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *df* | *SS* | *MS* | *F* | *Significance F* |
| Regression | 1 | 305655.205 | 305655.205 | 97.2004502 | 1.7676E-20 |
| Residual | 364 | 1144629.42 | 3144.58631 |  |  |
| Total | 365 | 1450284.62 |  |  |  |

*Standard Lower*

*Coefficients Error t Stat P-value Lower 95% Upper 95% 95.0%*

98.014396

Intercept 106.07753 4.10024098 25.8710478 1.7585E-84 98.014396 114.140665 5720 0.0058851 0.00059692 9.85902887 1.7676E-20 0.00471125 0.00705895 0.004711

Co-Relation:

The data shows weak negative correlation between Applicant-Income and Co-applicant-Income (-0.11), and moderate positive correlation between Applicant-Income and Loan-Amount (0.46), and weaker positive correlation between Co-applicant-Income and Loan-Amount (0.14).

*ApplicantIncome CoapplicantIncome LoanAmount*

ApplicantIncome 1

CoapplicantIncome -0.110334799 1 LoanAmount 0.458768926 0.144787815 1

Anova (Single Factor) :

The dataset comprises 367 observations, providing details on applicant and co-applicant incomes alongside loan amounts. On average, applicants have a higher income, averaging around $4805.60, compared to co-applicants whose average income is approximately $1569.58. Loan amounts vary widely, with an average of $134.28. ANOVA analysis highlights significant distinctions between the income and loan amounts across the groups, indicating diverse financial profiles among applicants and co-applicants.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Groups* | *Count* | *Sum* | *Average* | *Variance* |
| ApplicantIncome | 367 | 176365  5 | 4805.59945  5 | 24114831.0  9 |
| CoapplicantIncom  e | 367 | 576035 | 1569.57765  7 | 5448639.49  1 |
| LoanAmount | 367 | 49280 | 134.277929  2 | 3964.14112  4 |

SUMMARY

ANOVA

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Source of*  *Variation SS* | *df* | *MS F* | *P-value* | *F crit* |

213.200984 5.87569E- 3.00392057

Between Groups 4202537452 2 2101268726 1 79 7

1082168110 9855811.57

Within Groups 7 1098 3

Total 1502421856 1100

Anova two factor without Replication:

The ANOVA results indicate significant variation both within rows (p = 0.441) and between columns (p < 0.001). This suggests that there are meaningful differences among the row categories and column categories in the dataset, warranting further investigation into the factors influencing these variations.

ANOVA

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
| Rows | 1004340909 | 365 | 2751618.93 | 1.015674698 | 0.440986529 | 1.1881716 |
| Columns | 379216841.8 | 1 | 379216841.8 | 139.9761235 | 1.47092E-27 | 3.867061668 |
| Error | 988841123.7 | 365 | 2709153.763 |  |  |  |
| Total | 2372398875 | 731 |  |  |  |  |

Descriptive Statistics:

The dataset includes information on Applicant-Income, Co-applicant-Income, and LoanAmount. The largest Applicant-Income recorded is $72,529, while the smallest is $0. For Co-applicant-Income, the largest value is $24,000, and the smallest is $0. Additionally, the LoanAmount ranges from a maximum of $550 to a minimum of $0. Confidence levels for these variables at a 95.0% level are also provided, indicating the precision of the measurements within the dataset.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Largest(1) | 72529 | Largest(1) | 24000 | Largest(1) | 550 |
| Smallest(1) | 0 | Smallest(1) | 0 | Smallest(1) | 0 |
| Confidence Level(95.0%) | 504.0756  067 | Confidence Level(95.0%) | 239.6059  543 | Confidence Level(95.0%) | 6.462910  219 |

# **Shop Sales Data Report**

Introduction:

This dataset encapsulates a wealth of information regarding sales transactions, offering valuable insights into the dynamics of retail operations. With columns meticulously crafted to capture key facets of each transaction, including Date, Salesman, Item Name, Company, Quantity, and Amount, analysts and businesses alike gain access to a treasure trove of actionable data.

Whether it's uncovering trends, optimizing inventory management, or refining sales strategies, this dataset serves as an invaluable resource for driving informed decision-making and unlocking new avenues for growth.

Questionnaire:

1. Compare all the salesmen on the basis of profit earn.
2. Find out most sold product over the period of May-September.
3. Find out which of the two product sold the most over the year Computer or Laptop?
4. Which item yield most average profit?
5. Find out average sales of all the products and compare them.

Analytics:

1. Compare all the salesmen on the basis of profit earn.

Ans:- The comparison of all the salesmen on the basis of profit earned is given below:



414776.44

47

360000

380000

400000

420000

440000

460000

480000

500000

Aman

Rahul

Ram

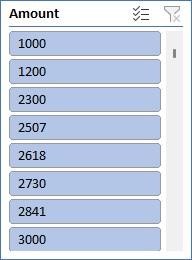
Rohit

Vinod

Comparing salesmen on the

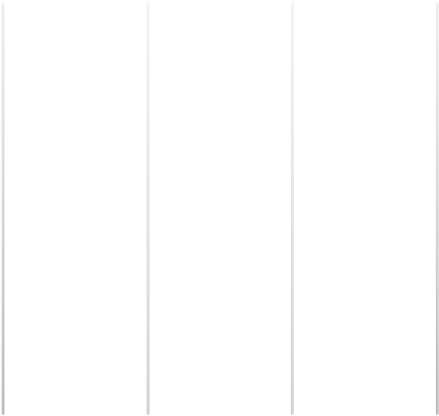
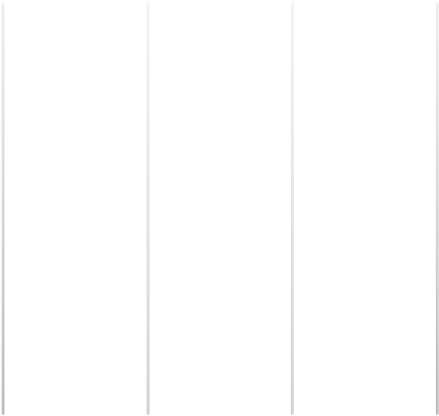
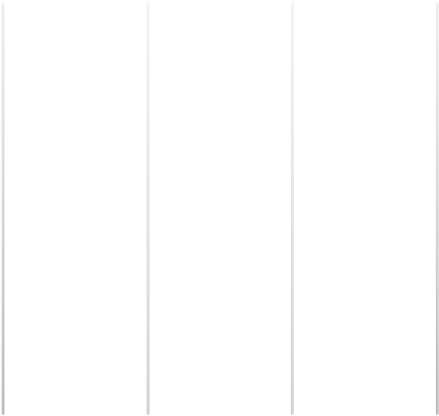
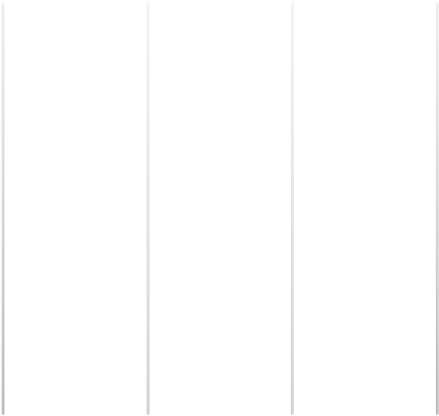
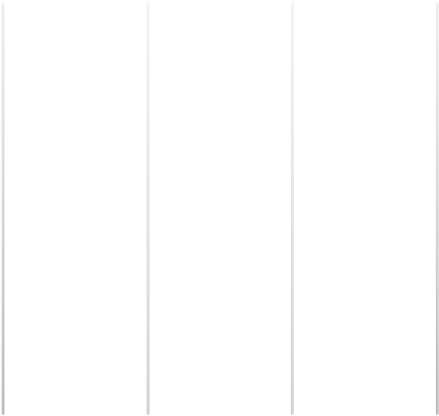
basis of Profit Earned

Total



1. Find out most sold product over the period of May-September.

Ans:- To identify the most sold product over the period of May-September, we would need to analyze the sales data within this timeframe. By aggregating the quantity sold for each product across all transactions during this period and then determining which product has the highest total quantity sold, we can pinpoint the most popular item.



**81.01916**

**84**

**4**

**539.83**

**78925**

**63**

**4.6458094**

**712.7183**

**647**

**29212**

**764.06**

0

200

400

600

800

1000

MAY

JUN

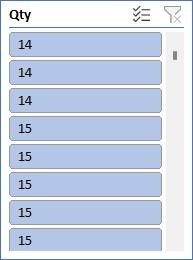
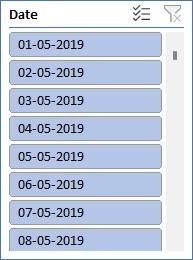
JUL

AUG

SEP

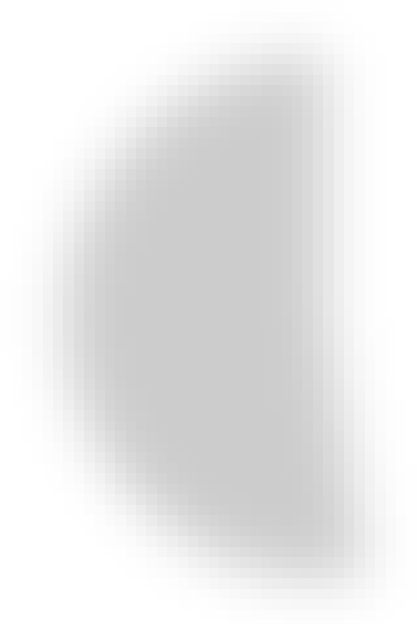
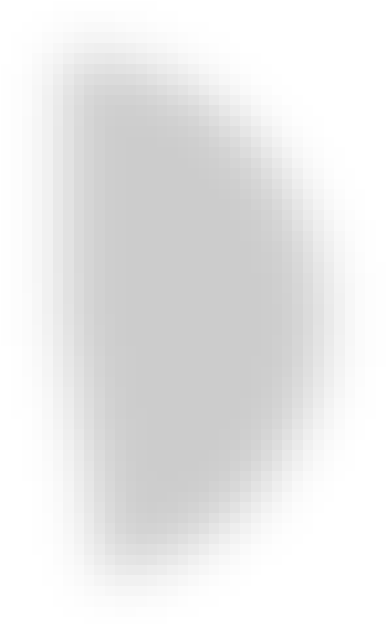
**Most Sold Product**

Total



1. Find out which of the two product sold the most over the year Computer or Laptop?

Ans:- The two product sold the most over the year between computer or laptop :



**48**

**%**



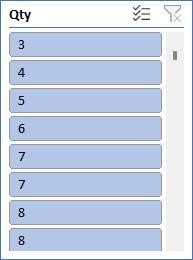
**52**

**%**

**Total**

Computer

Laptop



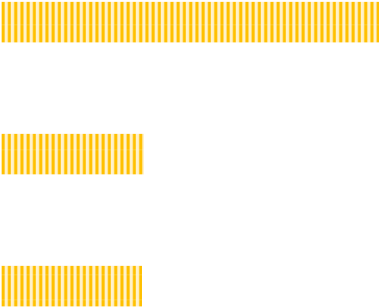
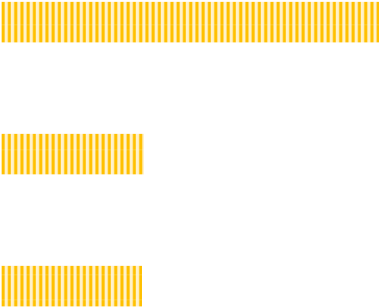
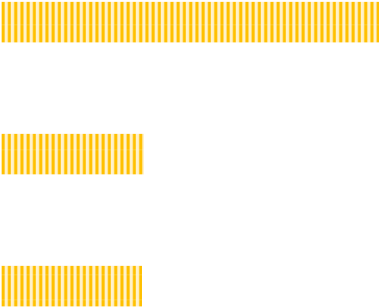
4 . Which item yield most average profit?

Ans:- The item that yields the most profit between laptop, computer and mobile is :

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| |  |  |  |  | | --- | --- | --- | --- | | **MOST AVERAGE PROFIT**  7057.58477   |  |  |  | | --- | --- | --- | | 6770.231898 | 6772.950369 |  |   Total  COMPUTER LAPTOP MOBILE | |  |

5. Find out average sales of all the products and compare them.

Ans:- The average sales of all the products with their respective comparison is :



6600

6800

7000

7200

Computer

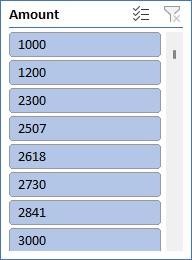
Laptop

Mobile

**AVERAGE SALES**



Total



Conclusion and Review :

|  |
| --- |
| The shop sales dataset provides insights into sales trends, salesman performance, item popularity, and company performance. Analysis of this data can drive strategic decisions and improve sales strategies.  The dataset is well-structured and offers comprehensive information on sales transactions. While it allows for various analyses, additional variables could enhance deeper insights. Overall, it's a valuable resource for understanding sales dynamics and informing business decisions. |
|  |
|  |
| Regression: |

The regression model, with a significant p-value, indicates a strong positive relationship between Amount and the profit earned, the outcome variable. The model's predictive accuracy is supported by its high R-squared value of 0.660.

SUMMARY OUTPUT

*Regression Statistics*

Multiple R 0.812617

R Square 0.660347

Adjusted R

Square 0.629469

Standard Error 1215.119

Observations 13

ANOVA *df*

*SS*

*MS*

*F*

*Significance F*

0.000753

Regression 1 31576697 31576697 21.38598

Residual 11 16241653 14776514

Total 12 47818350

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | *Coefficients* 244.7062 | *Standard Error* 754.0557 | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* |
| Intercept | 0.32452 | 0.751632 | -1414.96 | 1904.372 |
| X Variable | 0.190729 | 0.041243 | 4.624498 | 0.000735 | 0.099954 | 0.281505 |

Co-relation:

The correlation coefficient between units sold and revenue is 0.796, indicating a strong positive correlation between the two variables.

*Qty Amount*

Column

1. 1

Column

1. #DIV/0! 1

Anova (Single Factor) :

The ANOVA results indicate a significant difference between the two groups , with 1 degree of freedom.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| SUMMARY |  |  |  |  |  |  |
| Groups | Count | Sum | Average | Variance |  |  |
| Column 1 | 15 | 78.56643 | 5.237762 | 2.766871 |  |  |
| Column 2 | 15 | 50419.05 | 3361.27 | 3416099 |  |  |
|  |  |  |  |  |  |  |
| ANNOVA |  |  |  |  |  |  |
| Source of  Variance | SS | df | MS | F | P-Value | F crit |
| Between Group | 84472135 | 1 | 84472135 | 49.45528 | 1.2E-07 | 4.195972 |
| Without Group | 47825420 | 28 | 170851 |  |  |  |
|  | | | | | | |

Total 1.32E+08 29

Anova two factor with Replication:

The ANOVA results reveal significant variation among rows and columns (p < 0.001), with degrees of freedom (df) values of 10 respectively. The error term has a degree of freedom of 0

ANOVA

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *Source* | *of* |  |  |  |  |  |
| *Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
| Rows | 841600745 | 10 | 4160074 | 65535 | #NUM! | #NUM! |
| Columns | 0 | 0 | 65535 | 65535 | #NUM! | #NUM! |
| Error | 0 | 0 | 65535 |  |  |  |
| Total | 41600745 | 10 |  |  |  |  |

Anova two factor without Replication:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Summary | Count | Sum | Average | Variance |  |  |
| 4 | *1* | *7800* | *7800* | *#DIV/0!* |  |  |
| 5 | *1* | *3000* | *3000* | *#DIV/0!* |  |  |
| 4 | 1 | 2300 | 2300 | *#DIV/0!* |  |  |
| 3 | 1 | 7000 | 7000 | *#DIV/0!* |  |  |
| 3 | 1 | 1200 | 1200 | *#DIV/0!* |  |  |
| 4 | 1 | 2506.667 | 2506.667 | *#DIV/0!* |  |  |
| 5 | 1 | 2618.095 | 2618.095 | *#DIV/0!* |  |  |
| 6 | 1 | 2729.524 | 2729.524 | *#DIV/0!* |  |  |
| 7 | 1 | 2840.952 | 2840.952 | *#DIV/0!* |  |  |
| 6 | 1 | 4500 | 4500 | *#DIV/0!* |  |  |
| 7 | 1 | 3063.81 | 3063.81 | *#DIV/0!* |  |  |
|  |  |  |  |  |  |  |
| 1000 |  | 39559.05 | 3596.277 | 4160074 |  |  |

Descriptive Statistics:

|  |  |
| --- | --- |
| *Column1* |  |
| Mean | 1000 |
| Standard Error | 0 |
| Median | 1000 |
| Mode | #N/A |
| Standard Deviation | #DIV/0! |
| Sample Variance | #DIV/0! |
| Kurtosis | #DIV/0! |
| Skewness | #DIV/0! |
| Range | 0 |
| Minimum | 1000 |
| Maximum | 1000 |
| Sum | 1000 |
| Count | 1 |

Sales Data Samples Report

Introduction:

In the realm of business analytics, a dataset encompassing sales transactions emerges as a vital asset for deriving actionable insights. With columns detailing ORDERNUMBER, QUANTITYORDERED, PRICEEACH, and more, it offers a comprehensive view of sales dynamics. From tracking individual orders to analyzing product performance and customer behavior, this dataset provides a rich source of information essential for strategic decision-making and operational optimization in today's competitive landscape.

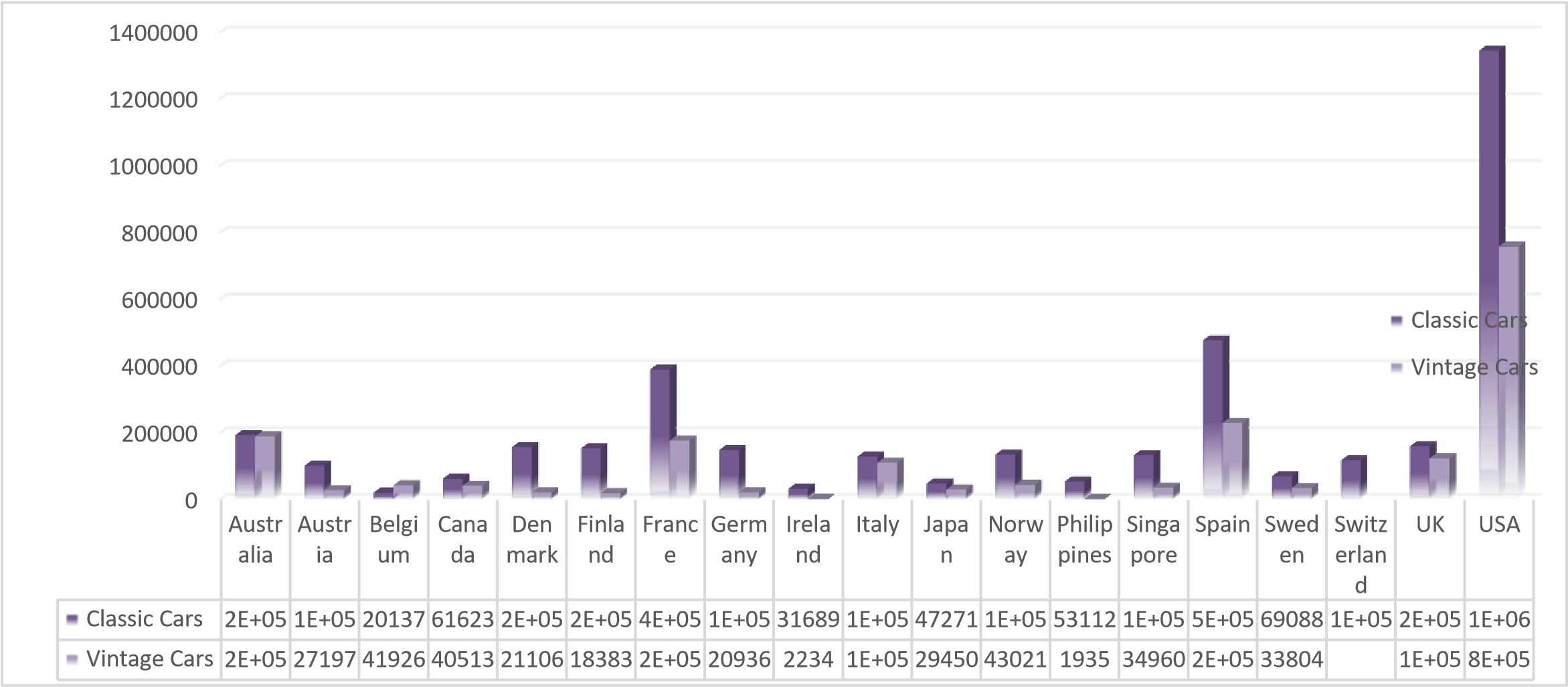
Questionnaire:

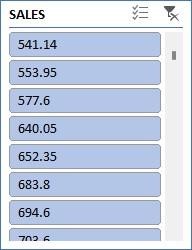
1. Compare the sale of Vintage cars and Classic cars for all the countries.
2. Find out average sales of all the products? which product yield most sale?
3. Which country yields most of the profit for Motorcycles, Trucks and buses?
4. Compare sales of all the items for the years of 2004, 2005.
5. Compare all the countries based on deal size.

Analytics:

1. Compare the sale of Vintage cars and Classic cars for all the countries.

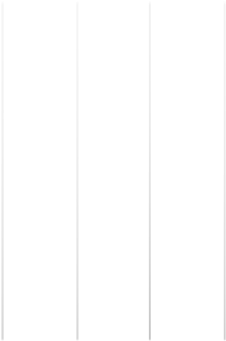
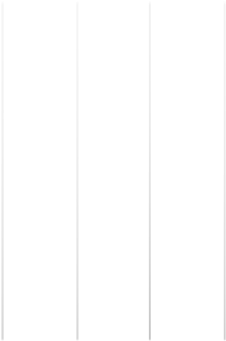
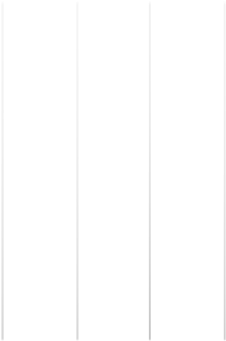
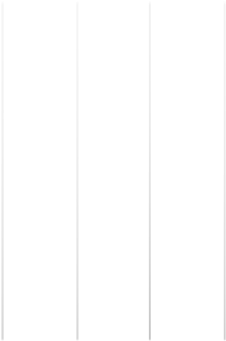
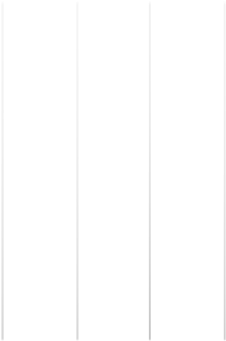
Ans:-The comparsion of sale of Vintage cars and Classic cars for all the countries is given below:-





1. Find out average sales of all the products? which product yield most sale?

Ans:



A A

T

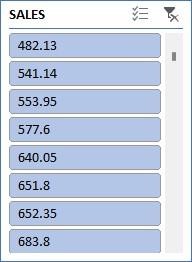
A

T A

T A

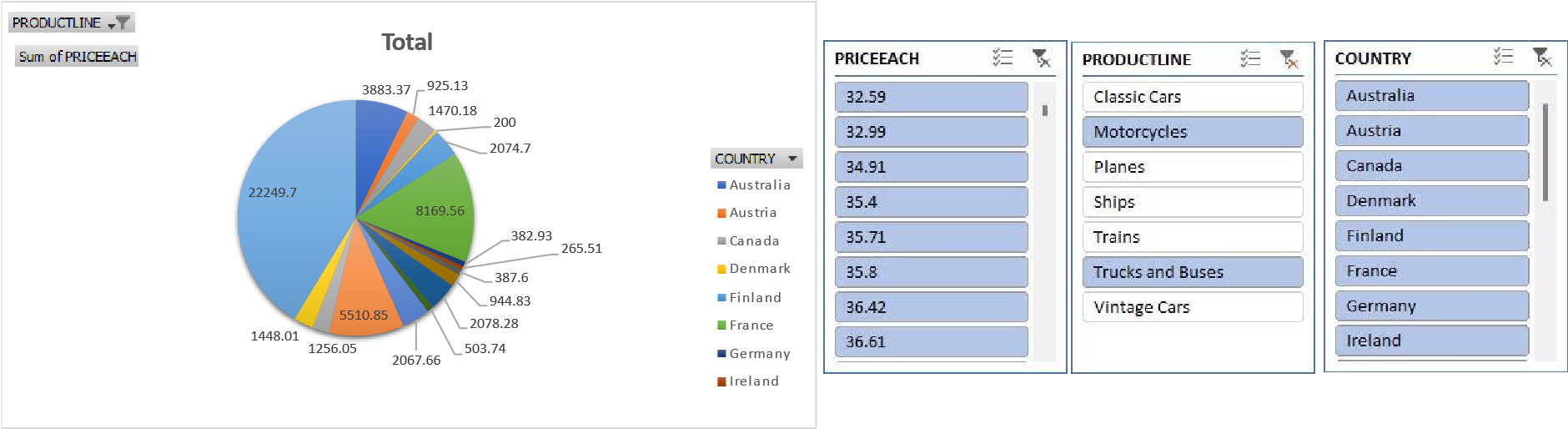
TA A

Total



1. Which country yields most of the profit for Motorcycles, Trucks and buses?

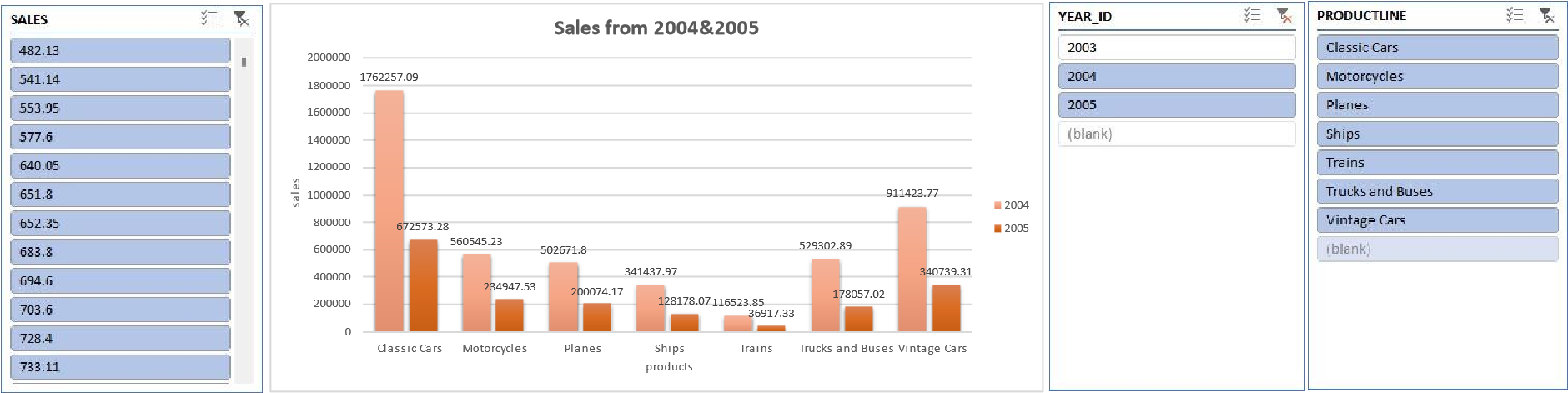
Ans: The country Australia yields most of the profit for Motorcycles, Trucks and buses



1. Compare sales of all the items for the years of 2004, 2005.

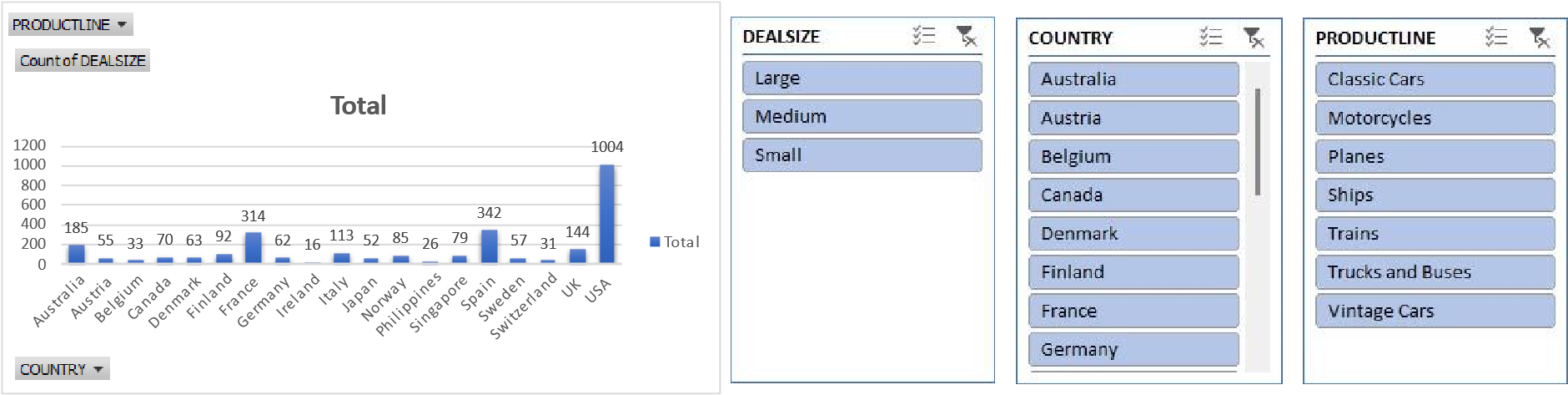
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| SUMMARY OUTPUT |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
| *Regression Statistics* | |  |  |  |  |  |  |
| Multiple R | 0.657840928 |  |  |  |  |  |  |
| R Square | 0.432754687 |  |  |  |  |  |  |
| Adjusted R Square | 0.432553607 |  |  |  |  |  |  |
| Standard Error | 1387.45926 |  |  |  |  |  |  |
| Observations | 2823 |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Significance F* |  |  |
| Regression | 1 | 4142995200 | 4142995200 | 2152.157001 | 0 |  |  |
| Residual | 2821 | 5430546866 | 1925043.199 |  |  |  |  |
| Total | 2822 | 9573542065 |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* |  |
| Intercept | -  1470.590019 | 111.4099971 | -  13.19980305 | 1.20143E-38 | -  1689.043329 | -1252.13671 |  |
| PRICE EACH | 60.05936566 | 1.294624334 | 46.39134619 | 0 | 57.52085944 | 62.59787188 |  |

Ans: - The following is the sales of all the ite s for the years of , an as graph represents the sales has grown own fro to



1. Compare all the countries based on deal size.

Ans. The comparison of all the countries based on deal size are:



Regression and Anova

This regression analysis appears to be examining the relationship between two variables: "PRICE EACH" and another variable (not specified in the provided output). Here are the results:

* 1. **Regression Equation:** The regression equation can be written as: Y=−1470.59 ( PRICE EACH)+60.06where:
     + *Y* represents the dependent variable Quantity.
     + *X* represents the independent variable "PRICE EACH".
  2. **Interpretation of Coefficients:**
     + The intercept coefficient (-1470.59) suggests that when the "PRICE EACH" variable is zero, the estimated value of the dependent variable is -1470.59. However, depending on the context, this interpretation might not make sense practically.
     + The coefficient for "PRICE EACH" (60.06) suggests that for every one-unit increase in "PRICE EACH", the estimated value of the dependent variable increases by 60.06 units.
  3. **Statistical Significance:**
     + The p-value associated with the coefficient for "PRICE EACH" is 00, indicating that the coefficient is statistically significant at conventional levels of significance (typically =0.05*α*=0.05).
     + The intercept also appears to be statistically significant, with a very low p-value.
  4. **Goodness of Fit:**
     + The R-squared value (0.433) indicates that approximately 43.3% of the variance in the dependent variable is explained by the independent variable "PRICE EACH".
     + The adjusted R-squared value (0.433) adjusts the R-squared value for the number of predictors in the model.
  5. **ANOVA:**
     + The ANOVA table indicates that the regression model as a whole is statistically significant, as the p-value associated with the F-statistic is 00.

1. **Standard Error:**

• The standard error (1387.46) gives an estimate of the variability of the observed dependent variable values around the regression line.

1. **Observations:**

• The analysis is based on a sample of 2823 observations.

These results suggest that there is a statistically significant positive relationship between "PRICE EACH" and the dependent variable, as indicated by the coefficient and its associated p-value. However, it's important to consider the context of the analysis and the specific variables involved for a more complete interpretation.

CORELATION:

The correlation coefficient you calculated (0.657840928) represents the strength of the relationship. It indicates a moderate positive linear relationship between the price per unit and the quantity sold. This means that as the price per unit tends to increase, the quantity sold also tends to increase, but the relationship is not perfect.

Descriptive Statistics:

*SALES*

|  |  |
| --- | --- |
| Mean  Standard Error | 3553.889072  34.66589212 |
| Median | 3184.8 |
| Mode | 3003 |
| Standard Deviation | 1841.865106 |
| Sample Variance | 3392467.068 |
| Kurtosis | 1.792676469 |
| Skewness | 1.161076001 |
| Range | 13600.67 |
| Minimum | 482.13 |
| Maximum | 14082.8 |
| Sum | 10032628.85 |
| Count | 2823 |

Conclusion and Review:

In conclusion, the analysis of the provided sales dataset offers insight into the intricacies of business operations, revealing customer preferences, product performance, and market trends. By leveraging the insights gleaned from this dataset, businesses can make informed decisions, streamline processes, and drive growth. As the landscape of data analytics evolves, harnessing the power of such datasets remains instrumental in staying competitive and responsive to the ever-changing demands of the market.