Introduction:

# Cookie Data Analysis

The purpose of this report is to analyze the sales data of various cookie types across different countries for the years 2019 and 2020. The dataset provides insights into revenue, profit, quantity sold, and pricing information for each cookie type and country. Through this analysis, we aim to understand the performance of different cookie types, identify trends across countries, and draw conclusions regarding 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.



**comparison between cookie**

**types in US, Malaysia and**

**India**

**33%**

**37%**

**30%**

India Malaysia

United States

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



**Average revenue generated by different**

**types of cookies**

**23%**

**22%**

**4%**

**12%**

**22%**

**17%**

Chocolate Chip Fortune Cookie Oatmeal Raisin Snickerdoodle Sugar

White Chocolate Macadamia Nut

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

35000

30000

25000

20000

15000

2019

2020

10000

5000

0

Fortune Sugar Fortune Sugar Fortune Sugar Fortune Sugar Fortune Sugar

Cookie Cookie Cookie Cookie Cookie

India

Malaysia

Philippines United Kingdom United States

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



900000

800000

700000

600000

500000

400000

2019

300000

2020

200000

100000

0

India

Malaysia

Philippines

United

Kingdom

United States

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



**Chocolate Chip**

**23%**

**22%**

**16%**

**21%**

**18%**

India Malaysia Philippines

United Kingdom

United States

## Conclusion and Reviews:

In conclusion, the analysis of cookie sales data has provided invaluable insights into market trends, consumer preferences, and profitability across various countries and cookie types. By examining revenue, profit, quantity sold, and pricing information, we gained a comprehensive understanding of the factors driving sales performance. This analysis has enabled us to identify opportunities for growth, optimize product offerings, and refine marketing strategies to better meet customer needs and maximize profitability. Moving forward, continued analysis and adaptation based on these insights will be essential for maintaining a competitive edge in the dynamic cookie market. Overall, the thorough examination of sales data has proven instrumental in informing strategic decision-making and ensuring the long-term success 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 supported 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* | *Significanc*  *e F* |
| Regression | 1 | 3.29E+09 | 3.29E+0  9 | 1537.35  6 | 1.4E-178 |
| Residual | 698 | 1.49E+09 | 2139668 |  |  |
| Total | 699 | 4.78E+09 |  |  |  |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | *Coefficient*  *s* | *Standard*  *Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper*  *95%* | *Lower*  *95.0%* | *Upper*  *95.0%* |
| Intercept | -74.4103 | 116.5304 | -0.63855 | 0.52332  6 | -303.202 | 154.381  7 | -303.202 | 154.381  7 |
| Units Sold | 2.500792 | 0.063781 | 39.20914 | 1.4E-  178 | 2.375567 | 2.62601  7 | 2.37556  7 | 2.62601  7 |

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 indicate 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, suggesting that the model explains 6% of the variability in the data.

SUMMARY

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| *Groups* |  | *Count* | *Sum* | *Average* | *Variance* |  |  |
| 3450 |  | 699 | 1923505 | 2751.795 | 4154648 |  |  |
| 5175  ANOVA |  | 699 | 2758189 | 3945.908 | 6850161 |  |  |
| *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 reveal 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.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ANOVA |  |  |  |  |  |  |
| *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 show that there is 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.

Furthermore, 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:

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.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *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%) |

# Store Dataset Report

## Introduction:

This dataset encompasses sales data from a retail store, featuring a range of attributes including customer demographics (Gender, Age Group), transaction details (Order ID, Status), product specifics (Category, SKU), and shipping information. With a focus on understanding customer behaviour and product trends, our analysis aims to uncover patterns, preferences, and correlations within the data. By leveraging these insights, businesses can optimize marketing efforts, enhance inventory management, and improve customer satisfaction.

## 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? Ans: Amazon leads in the sales in both men and women category followed by Myntra and Flipkart. Amazon sold almost 3500 units in men category and almost 7500 units in women category. Myntra sold 2000 units in men section.



Women

Men

Others

4000

Amazon

Flipkart Meesho Myntra

Nalli

6000

5000

Ajio

7000

8000

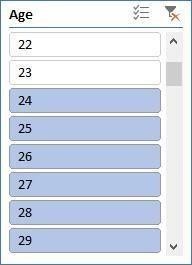
1. Compare category. Find out most sold category above 23 years of age for any gender.

Ans: In the above 23 years of age group Kurta is most sold category in women section with 8820 units sold. Set is most sold category in men section with 4365 units sold also set is the second most sold category in women section.

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

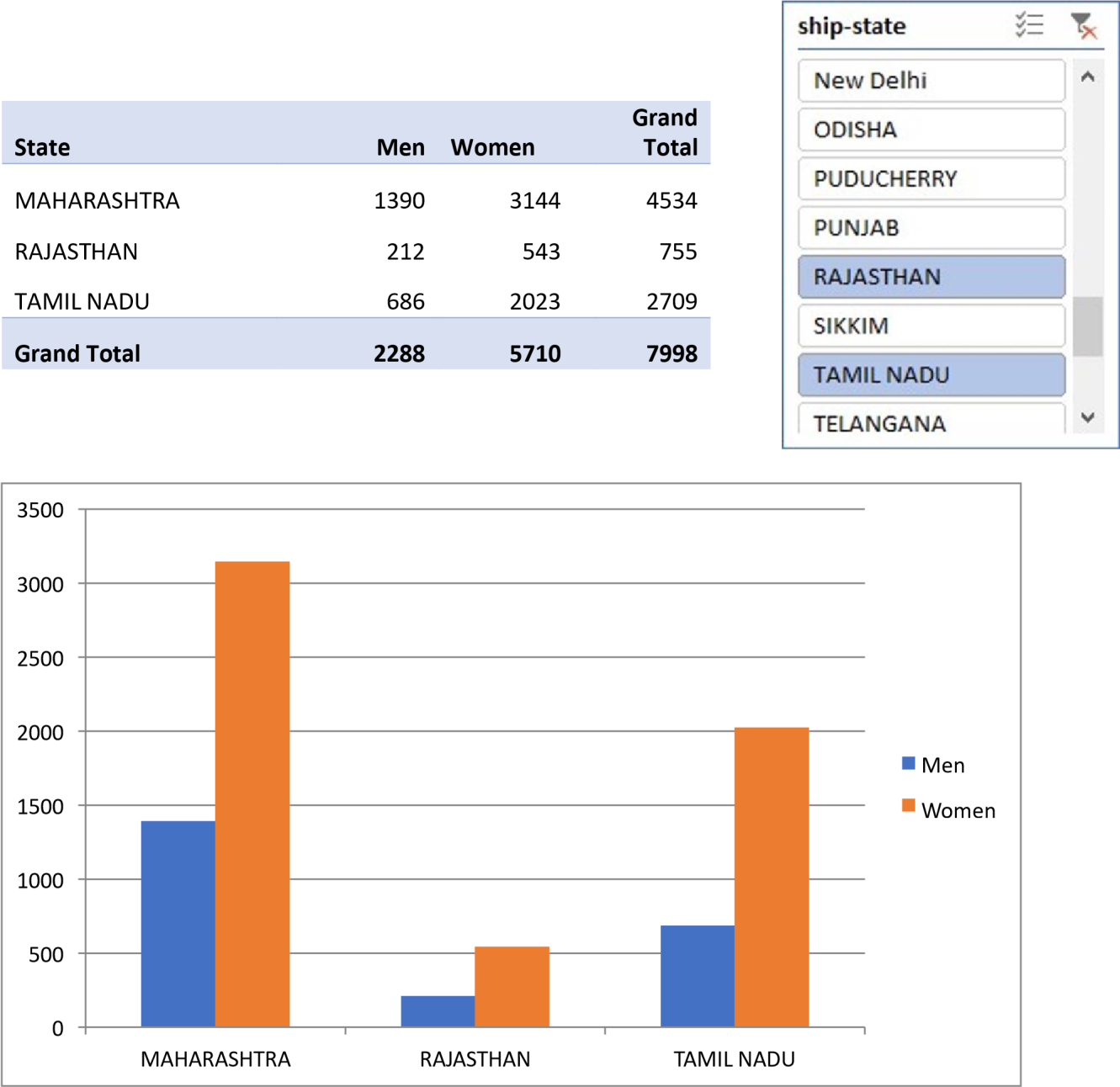
500

.

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

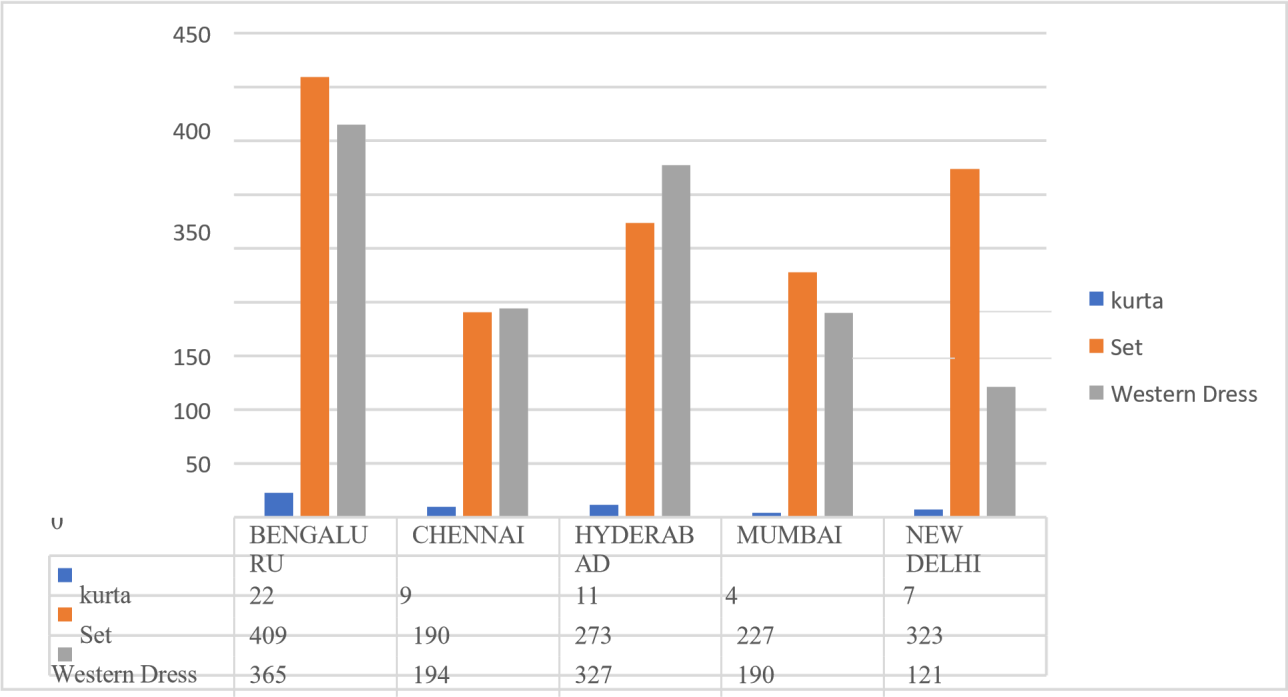


.

Which city sold most of following categories

* 1. Kurta
  2. Set
  3. Western wears

Ans: Bengaluru, Chennai, Hyderabad, Mumbai and New Delhi are the cities sold most of kurtas, Sets and western wears.



.



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **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** |

1. In which month most items sold in any of the state on the basis of category. Ans: The graph for most items sold in any of stats on basis of category is as follows:



200

180

160

140

Blouse

Bottom

120

Ethnic Dress

100 kurta

80

Saree

Set

January February March April

May

June

July August September October NovemberDecember

## Conclusion and Review:

After thorough analysis of the store data, it is evident that there are notable trends and insights to be gleaned. By examining key metrics such as units sold, state wise analytics, geographic, and sales across different stats and products, we can draw valuable conclusions about market demand, sales and overall profitability. This comprehensive understanding will enable informed decisionmaking to optimize resources, target specific markets, and maximize profits in future store sales endeavours.

# Exploring Car Dataset

## Introduction:

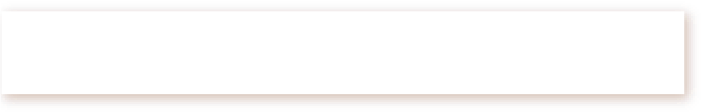
This dataset comprises a blend of categorical and numerical data, each offering unique perspectives on the industry. Categorical data, such as make, model, and color, encapsulates the diversity of vehicles and consumer preferences. Meanwhile, numerical attributes like mileage, price, and cost provide quantifiable metrics essential for analyzing market trends and pricing dynamics.

## Questionnaires:

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:

Q1. Compare the mileage of Chevrolet Impala to Toyota Corolla. Which of the two is giving best mileage? Ans. Toyota Corolla gives better mileage than Chevrolet Impala.



**Corolla vs Chevrolet Mileage**

Chevrolet

228,486

Total

Toyota

277,131

0

50,000

100,000

150,000

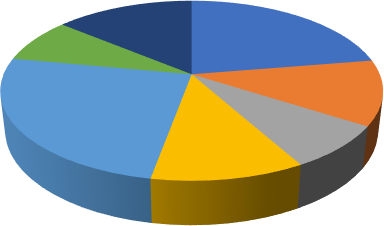
200,000

250,000

300,000

Q2. Justify, Buying of any Ford car is better than Honda.

Ans. Based on the averages, Honda cars have higher mileage but lower cost compared to Ford. Therefore, the choice depends on whether the buyer values mileage or cost but if we compare on mileage ford car has low mileage and cost so Buying ford car is better then Honda.



Ford car Vs Honda Car Comparision

Ford Escape Ford F-150 Ford Fusion Ford Mustang Honda Accord Honda Civic Honda CRV

Q3. Among all the cars which car color is the most popular and is least popular?

Ans. Most popular color is Silver and Black as each appear 6 times and least appearing colour are Blue ,Green ,Red ,White they all apper 3 times.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 0 | Black | Blue | Green | Red | Silver | White |
| Total | 6 | 3 | 3 | 3 | 6 | 3 |

Q4. Compare all the cars which are of silver color to the green color in terms of Mileage.



**Color count**

7

6

6

6

5

4

3

3

3

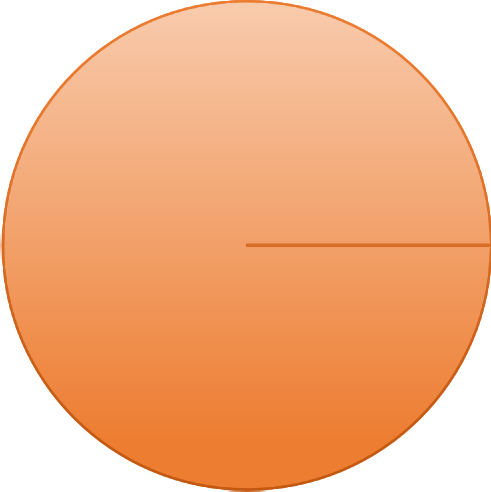
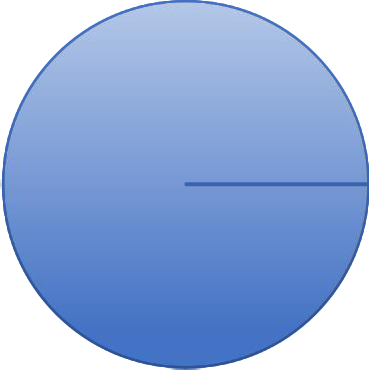
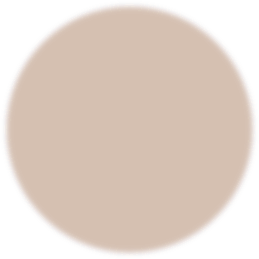
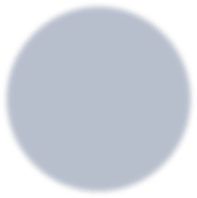
3

3

2

1

Ans. Silver color car millage is more than green color car milage if we compare there average.



**Green Vs Silver Color Car Mileage Comparision**

100%

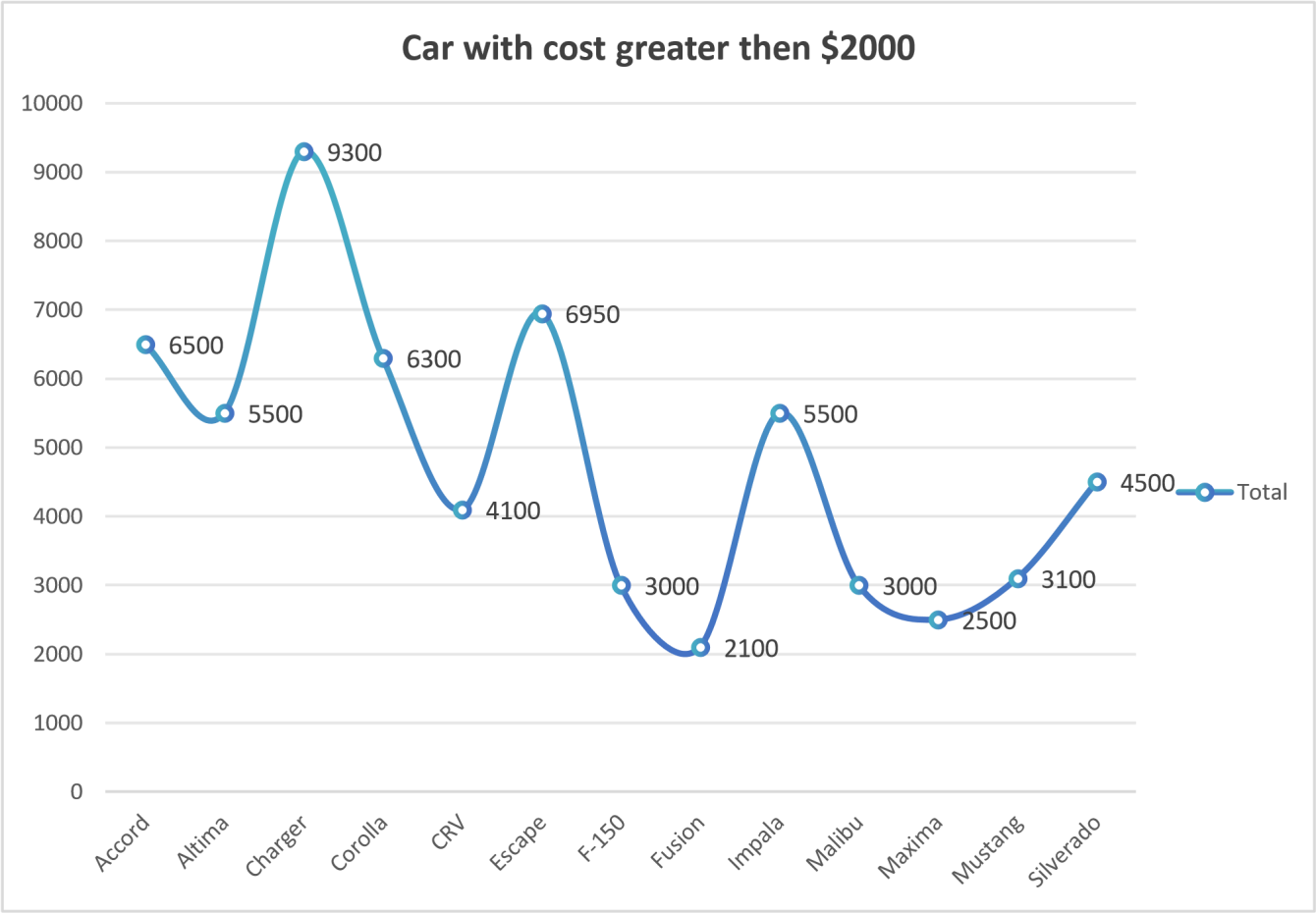
100%

Total

Q5. Find out all the cars, and their total cost which is more than $2000?

Ans. All the car mention below cost is more than $2000

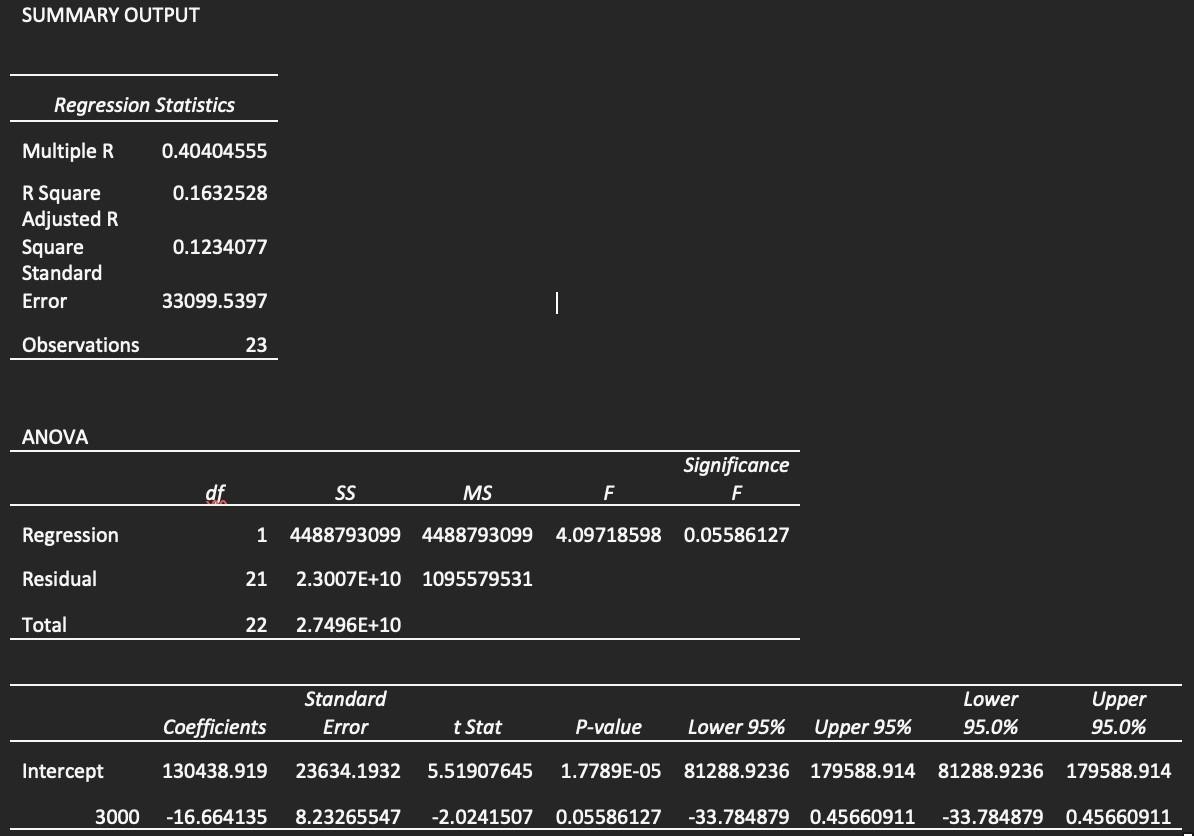
Accord, Altima, Charger, Corolla, CRV, EscapeF-150, Fusion, Impala, Malibu, Maxima, Mustang, Silverado



## Regression:

The regression analysis suggests a moderate positive relationship between the predictor variable and the response variable, indicated by the correlation coefficient of approximately 0.40. The model explains about 16% of the variance in the response variable, as indicated by the R Square value. The coefficient estimates show that for every unit increase in the predictor variable, there is a corresponding decrease of approximately

16.66 in the response variable, with a p-value of 0.056, indicating a marginally significant effect.



## Co-relational:

The correlation matrix indicates a moderate negative correlation (-0.411) between Mileage and Price. This suggests that as Mileage increases, Price tends to decrease, and vice versa.

*Mileage Price*

Mileage 1

Price -0.4110586 1

## ANOVA:

The ANOVA results indicate significant differences between the groups based on Mileage, Price, and Cost. The F-statistic is large (128.88), with a very low p-value (5.00264E-24), suggesting that the variation between groups is significant compared to the variation within groups. This implies 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, indicating that these variables play a significant role in influencing the outcome being analyzed.

### ANOVA: Single Factor

SUMMARY

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | *Count* | *Sum* | *Average* | *Variance* | *Groups* |  |
| Mileage | 24 | 2011267 | 83802.7917 | 1214155660 |  |  |
| Price | 24 | 78108 | 3254.5 | 837024.087 |  |  |
| Cost  ANOVA | 24 | 66150 | 2756.25 | 705502.717 |  |  |
| *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 indicate significant differences among the levels or categories within each factor ("Rows" and "Columns"). Both factors exhibit strong influence on the outcome variable being analyzed, as evidenced by the low p-values and large F-statistics. This suggests 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 outline the characteristics of three variables: Mileage, Price, and Cost. Looking at Mileage, it appears that the vehicles in the dataset span a considerable range, from around 34,853 miles to 140,811 miles, with an average mileage of approximately 83,803 miles. Price and Cost exhibit similar trends, with prices ranging from

$2,000 to $4,959 and costs from $1,500 to $4,500, respectively. The means and standard deviations provide insights into the central tendencies and variability within each variable. Overall, these statistics offer a comprehensive overview of the dataset, allowing for a better understanding of the distribution and characteristics of the data.

*Mileage Price Cost*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Mean  Standard Error | 83802.7917  7112.65205 | Mean  Standard Error | 3254.5  186.751181 | Mean  Standard Error | 2756.25  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 |

## Conclusion and Reviews:

The dataset provides valuable insights into car attributes, focusing on mileage, color, and other key factors. Here's a simple conclusion based on the data:

Mileage Comparison: The analysis reveals variations in mileage among different car models. Toyota Corolla

generally offers better mileage compared to Chevrolet Impala.

Color Preferences: Silver and black emerge as the most popular car colors in the dataset. Blue, green, red, and white are among the least popular color choices.

Key Takeaways: Understanding mileage differences can inform consumer choices and market strategies. Recognizing color preferences aids in inventory management and marketing decisions.

# Exploring sales on different states of US

## Introduction:

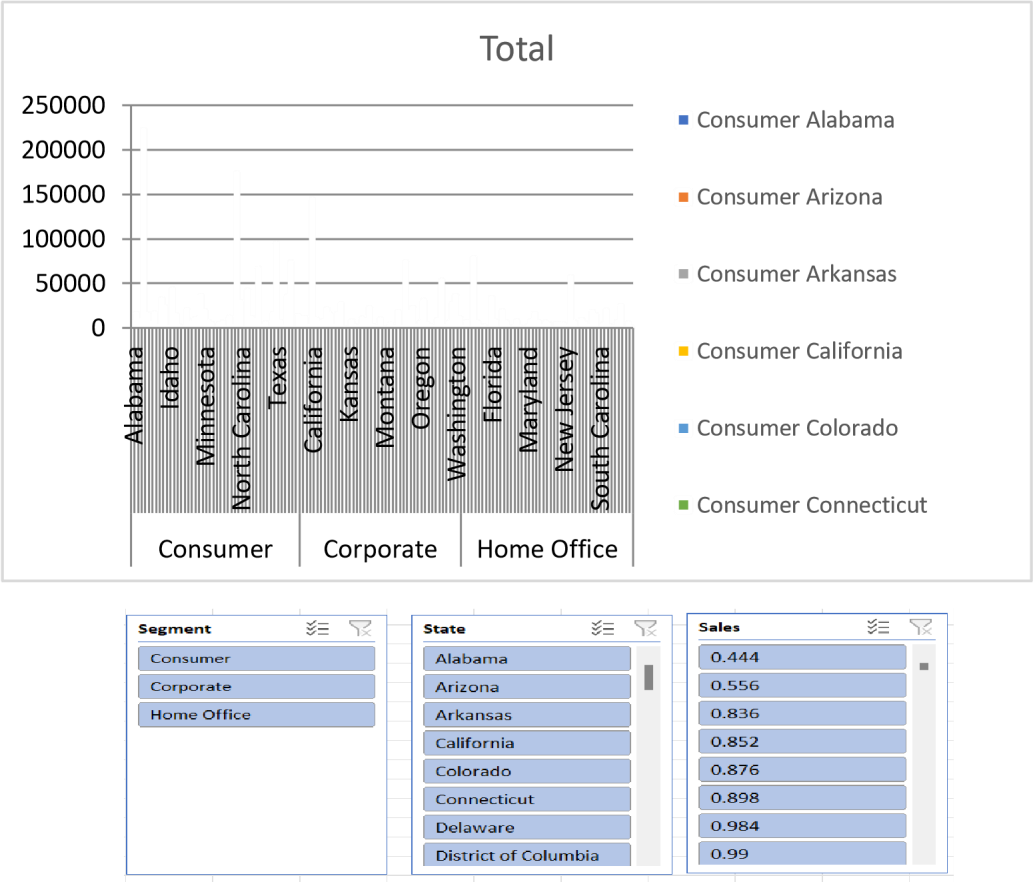
Our dataset comprises a plethora of variables, each offering unique insights into the multifaceted nature of different category sales. From fundamental transactional details such as Date, Time, sales, states to more nuanced factors like Customer Type, Demographics, category and sub category, every facet has been meticulously documented.

## 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.

## Analytics :

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



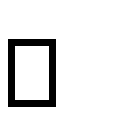


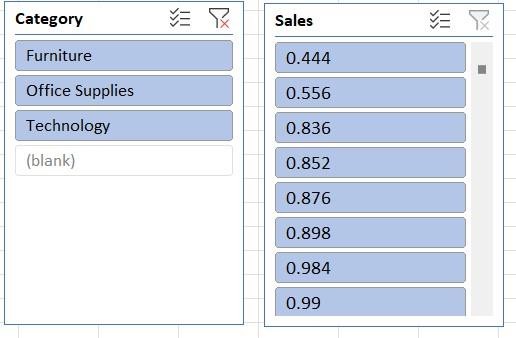
Q2. Find out top performing category in all the states?

**Total**

Furniture

Office Supplies Technology

Office Supplies is the top performing category in all the states

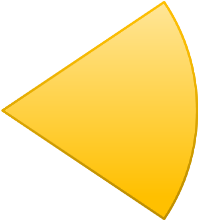
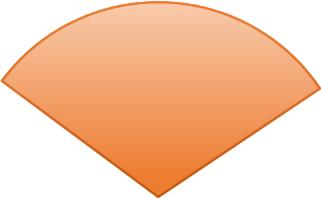
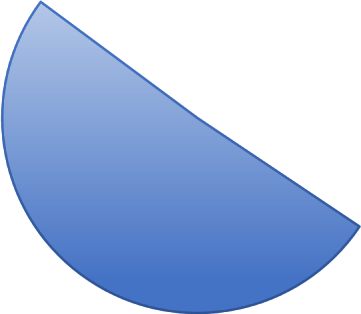
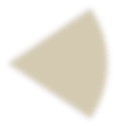
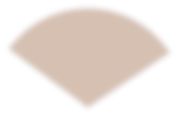
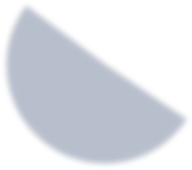


Q3. Which segment has most sales in US, California, Texas, and Washington?



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 2000  **Chart Title**  1000 | | | | | |
| 0 | | California | Texas | Washington |  |
|  | Consumer | 1020 | 516 | 275 |
| Corporate | 601 | 310 | 130 |
| Home Office | 325 | 147 | 99 |

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



688494.0748,

30%

424982.1769,

19%

Consumer

424982.1769,

19%

Corporate

Home Office

1148060.531,

51%

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



2500

2000

**Chart Title**

1500

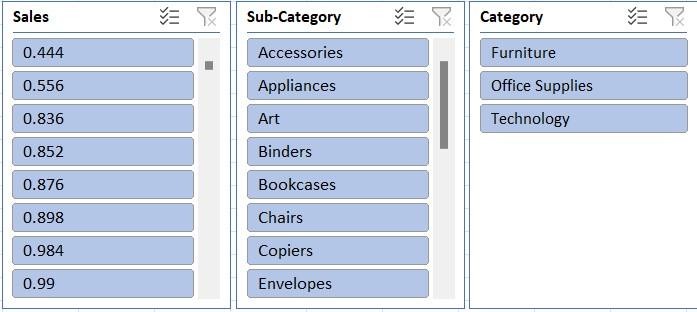
1000

500

0

sub category

category



## Conclusion and Review:

Our comprehensive analysis of the provided dataset through various data visualization techniques has yielded valuable insights. Through the creation of bar graphs, pie charts, and other visual representations, we've been able to discern patterns, trends, and relationships within the data that might have otherwise remained obscured.

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

Introduction:

# Loan Dataset Report

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.

## Questionnaire:

1. How many male graduates who are not married applied for Loan? What was the highest amount?
2. How many female graduates who are not married applied for Loan? What was the highest amount?
3. How many male non-graduates who are not married applied for Loan? What was the highest amount?
4. How many female graduates who are married applied for Loan? What was the highest amount?
5. How many male and female who are not married applied for Loan? Compare Urban, Semi-urban 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?





**Graduate UnMarried Male Loan Data**

66

Male

Count of LoanAmount

240

Max of LoanAmount

0

50

100

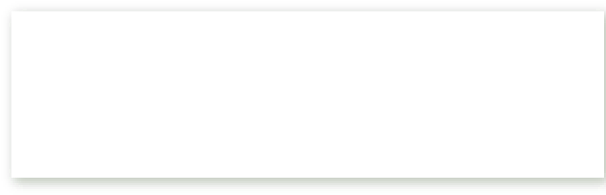
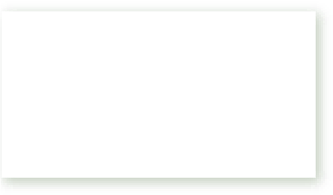
150

200

250

300

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



**Graduate UnMarried Female Loan Data**

Female

300

35

84%

86%

88%

90%

92%

94%

96%

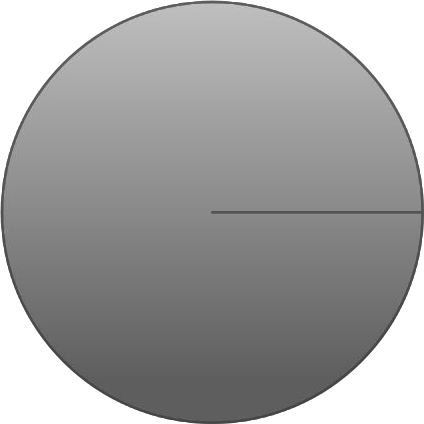
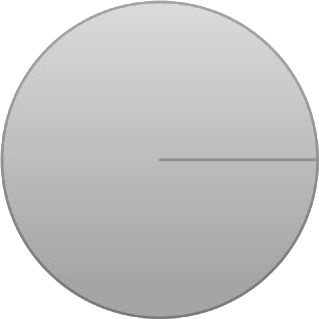
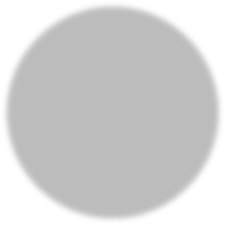
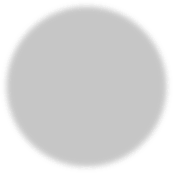
98%

100%

Max of LoanAmount

Count of LoanAmount

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



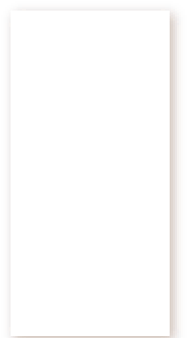
**NonGraduate UnMarried Male Loan Data**

199, 100%

199, 100%

Male

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



**Graduate Married Female Loan Data**

500

450

400

350

300

250

200

150

100

50

0

460

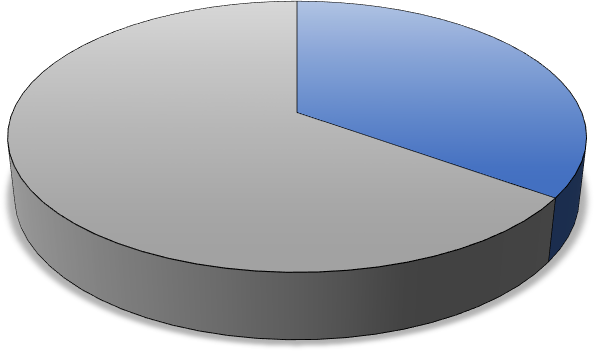
21

Max of LoanAmount

Count of Gender

|  |  |
| --- | --- |
|  | Female |
| Max of LoanAmount | 460 |
| Count of Gender | 21 |

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



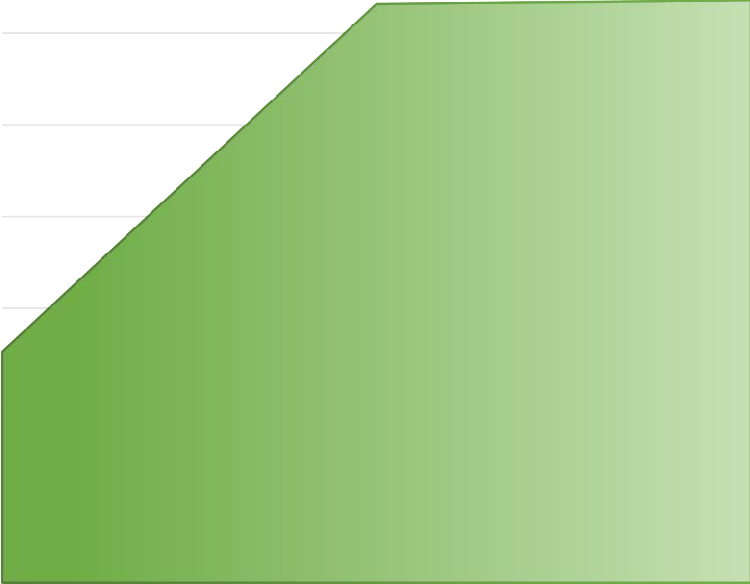
**UNMARRIED MALE AND FEMALE lOAN**

Female Male

44

82

**U**



Urban

Semiurban

Axis Title

4850

Rural

4950

4976

4900

Total

5167

5165

5000

5050

5100

5150

5200

**COMPARISION OF LOAN AMOUNT OF BASIS OF**

**PROPERTY AREA**

## Conclusion and Review:

Our analysis, using varied visualization techniques, revealed valuable insights, enhancing comprehension and decision-making. Visualizing data clarified complex findings, facilitating actionable strategies. This highlights the pivotal role of data visualization in extracting meaningful insights and informing decisions effectively.

## 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 |

## 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 |  |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Lowe Upper 95%*  *95.0* |
| Intercept | 106.07753 | 4.10024098 | 25.8710478 | 1.7585E- | 98.014396 | 114.140665 |
|  |  |  |  | 84 |  | 98.01 |
| 5720 | 0.0058851 | 0.00059692 | 9.85902887 | 1.7676E- | 0.00471125 | 0.00705895 |
|  |  |  |  | 20 |  | 0.004 |

Co-Relation :

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

*ApplicantIncome CoapplicantIncome LoanAmount*

ApplicantIncome 1

CoapplicantIncome -0.110334799 1

LoanAmount 0.458768926 0.144787815 1

### ANOVA : (Single Factor )

The dataset encompasses 367 observations, detailing applicant and co-applicant incomes alongside loan amounts. On average, applicants possess a higher income, averaging around

$4805.60, compared to co-applicants whose average income is approximately $1569.58. Loan amounts vary widely, averaging $134.28. ANOVA analysis underscores significant distinctions between the income and loan amounts across the groups, implying diverse financial profiles among applicants and co-applicants.

SUMMARY

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Groups* | *Count* | *Sum* | *Average* | *Variance* |
| ApplicantIncome | 367 | 176365  5 | 4805.5994  55 | 24114831.  09 |
| CoapplicantInco me | 367 | 576035 | 1569.5776  57 | 5448639.4  91 |
| LoanAmount | 367 | 49280 | 134.27792  92 | 3964.1411  24 |

ANOVA

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

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 420253745 |  | 210126872 | 213.20098 | 5.87569E- | 3.0039205 |
| Between Groups | 2 | 2 | 6 | 41 | 79 | 77 |
|  | 108216811 |  | 9855811.5 |  | | |
| Within Groups | 07 | 1098 | 73 |
|  | 150242185 |  |  |
| Total | 6 | 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.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *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 |  |  |  |  |

ANOVA

## 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 Coapplicant-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) |
| Smallest(1) | 0 | Smallest(1) | 0 | Smallest(1) |

Confidence Level(95.0%) 504.0756067 Confidence Level(95.0%) 239.6059543 Confidence Level(9

# Shop Sales Data

## Introduction :

This dataset encapsulates a wealth of information regarding sales transactions, providing 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 decisionmaking and unlocking new avenues for growth.

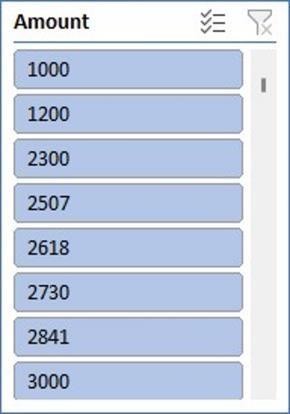
## Questionaries :

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:



Comparing salesmen on the

basis of Profit Earned

500000

480000

460000

440000

420000

400000

380000

360000

414776.44

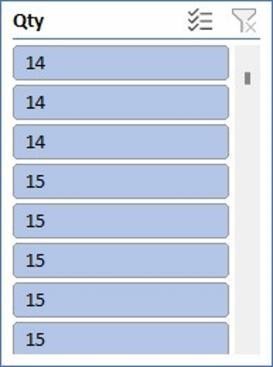
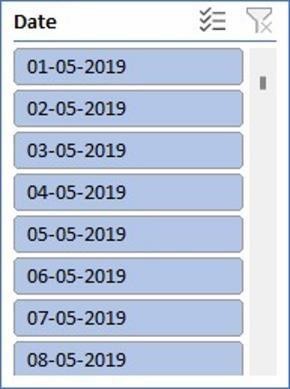
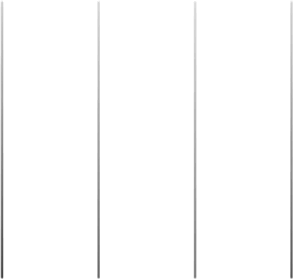
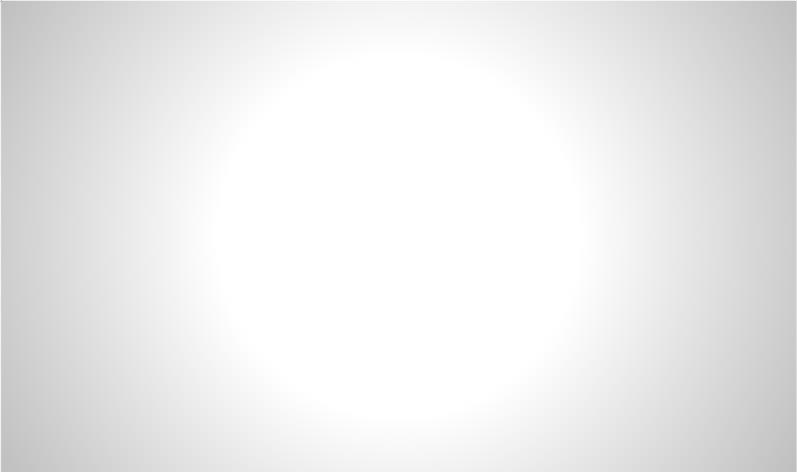
47

Total

Aman Rahul Ram Rohit Vinod

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.



**Most Sold Product**

SEP

AUG JUL JUN

MAY

0

200

400

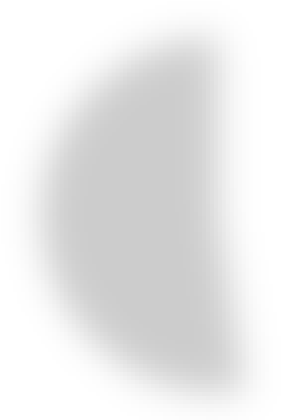
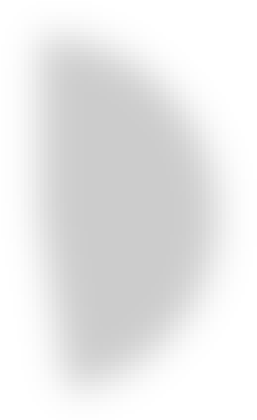
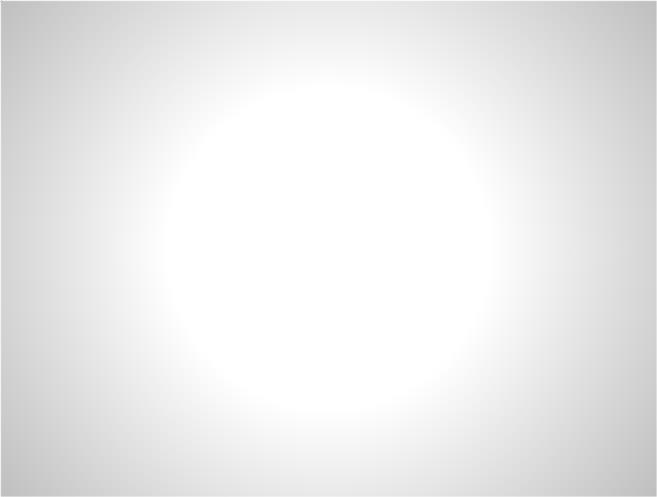
600

800 1000

Total

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **764.0629212** | | | | |
|  | | | | |
| **712.7183647** | | | |  |
|  | | | | |
| **634.6458094** | | |  | |
|  | | | | |
| **539.8378925** | |  | | |
|  | | | | |
| **481.0191684** |  | | | |

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 :



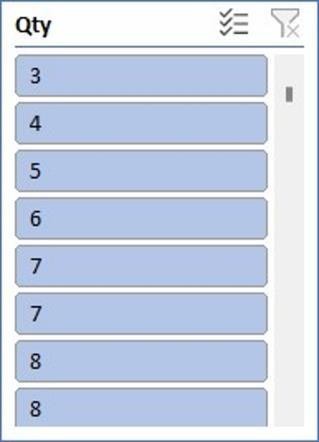
**Total**

**52%**

**48%**

Computer

Laptop



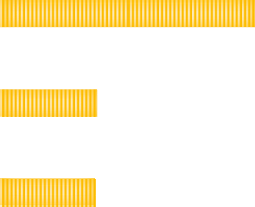
1. Which item yield most average profit?

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

|  |  |
| --- | --- |
| **MOST AVERAGE PROFIT**  Total  COMPUTER LAPTOP MOBILE |  |

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

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



**AVERAGE SALES**

Mobile

Laptop

Total

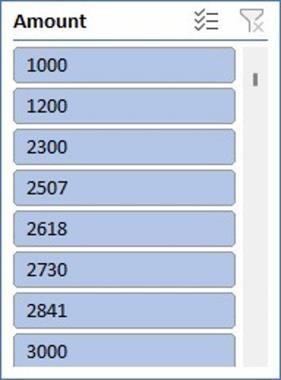
Computer

6600

6800

7000

7200



## Conclusion and Review :

|  |
| --- |
| The shop sales dataset offers 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 provides comprehensive information on sales transactions. |
| It allows for various analyses, but could benefit from additional variables for deeper |
| insights.Overall, it's a valuable resource for understanding sales dynamics and informing  business decisions. |

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

Regression:

SUMMARY OUTPUT

*Regression Statistics*

|  |  |
| --- | --- |
| Multiple R | 0.812617 |
| R Square | 0.660347 |
| Adjusted R Square | 0.629469 |
| Standard Error | 1215.119 |
| Observations | 13 |

## ANOVA

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | *SS* | | *MS* | | *F* | |  | |
| *df*  Regression 1 | | | 31576697 | | 31576697 | | 21.38598 | | *Significance F*  0.000753 | |
| Residual 11 | | | 16241653 | | 14776514 | |  | |  | |
| Total 12 | | | 47818350 | |  | |  | |  | |
|  | | |  | |  | |  | |  | |
| *Coefficients* | | | *Standard Error* | | *t Stat* | | *P-value* | | *Lower 95% Upper 95%* | |
| Intercept | 244.7062 | 754.0557 | | 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

2 #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 |  |  |
|  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |
| 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 |

Introduction:

# Sales Data Samples

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 analysing product performance and customer behaviour, this dataset provides a rich source of information essential for strategic decisionmaking and operational optimization in today's competitive landscape.

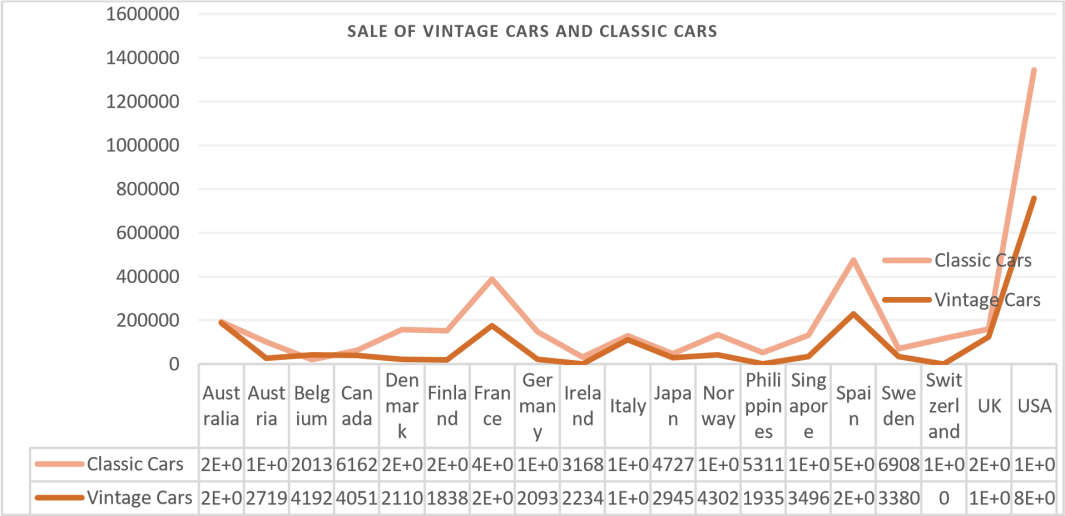
## Questionaries:

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.

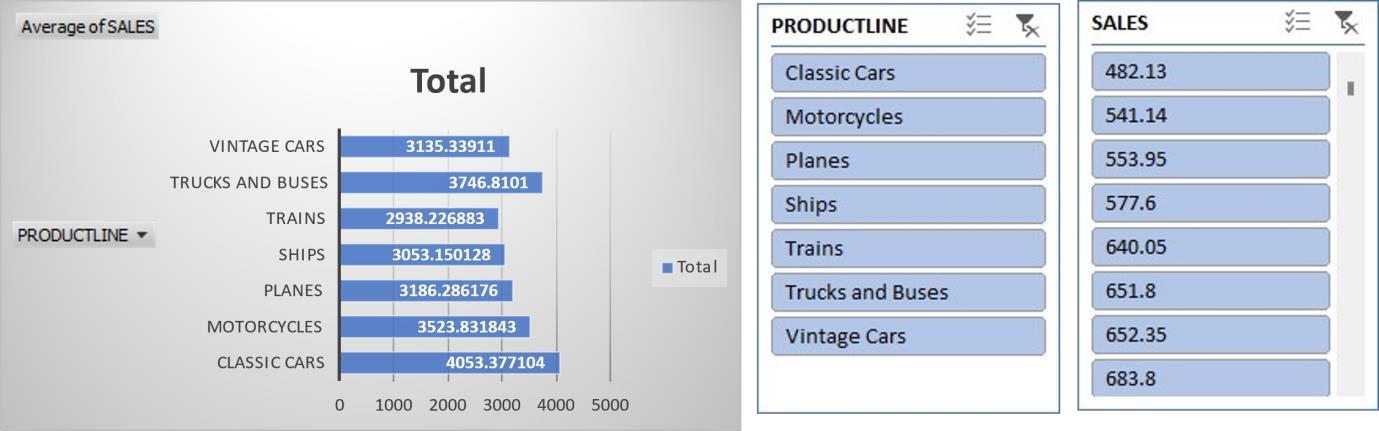
## Analytics:

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:-





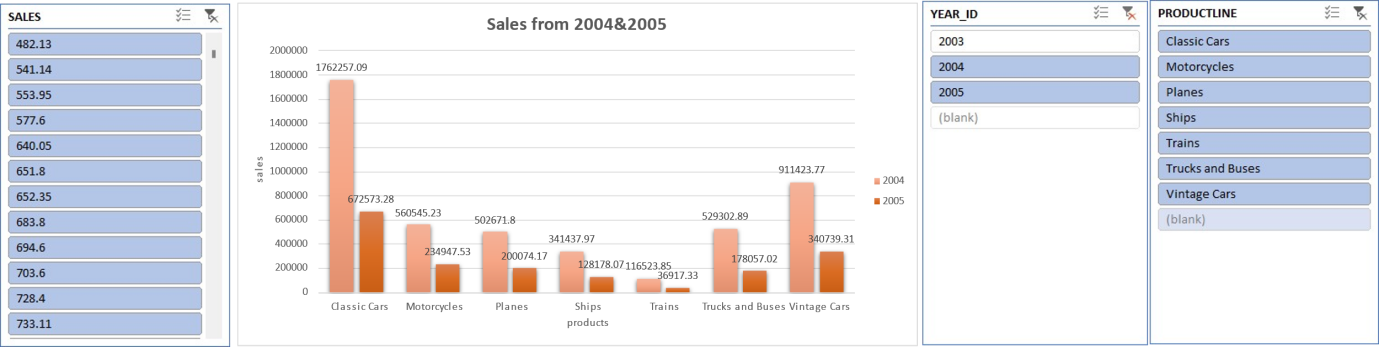
1. Find out average sales of all the products? which product yield most sale? Ans:
2. Which country yields most of the profit for Motorcycles, Trucks and buses?

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

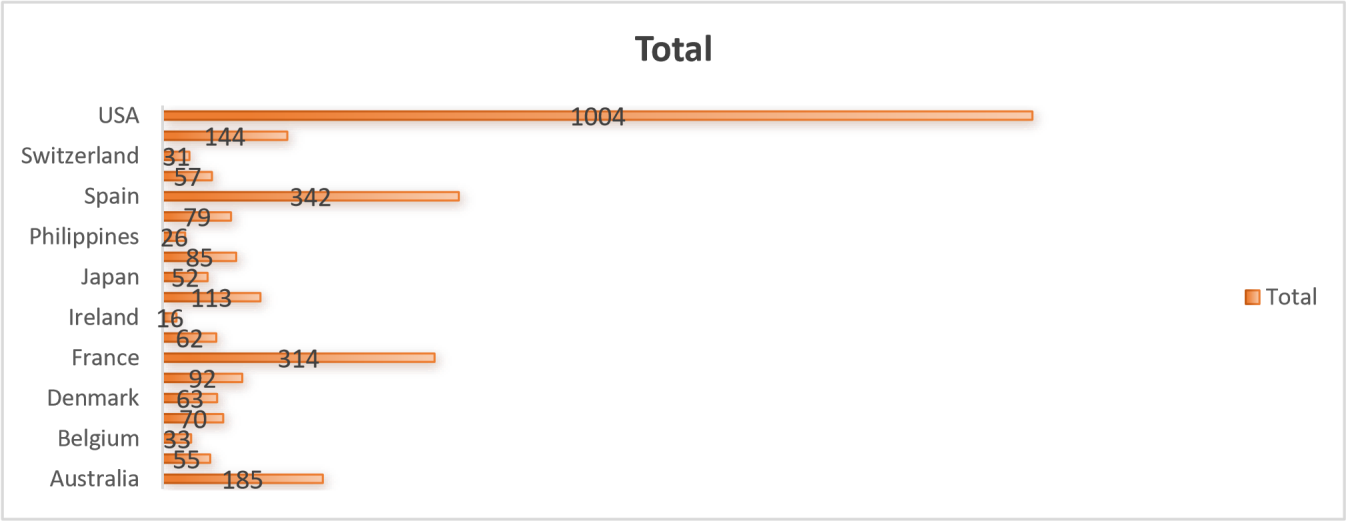


1. Compare sales of all the items for the years of 2004, 2005.

Ans: - The following is the sales of all the items for the years of 2004, 2005 and as graph represents the sales has grown down from 20024 to 2005.



1. Compare all the countries based on deal size.

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

## Conclusion and Review:

In conclusion, the analysis of the provided sales dataset offers a window into the intricacies of business operations, shedding light on 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 continues to evolve, harnessing the power of such datasets remains instrumental in staying competitive and responsive to the ever-changing demands of the market.