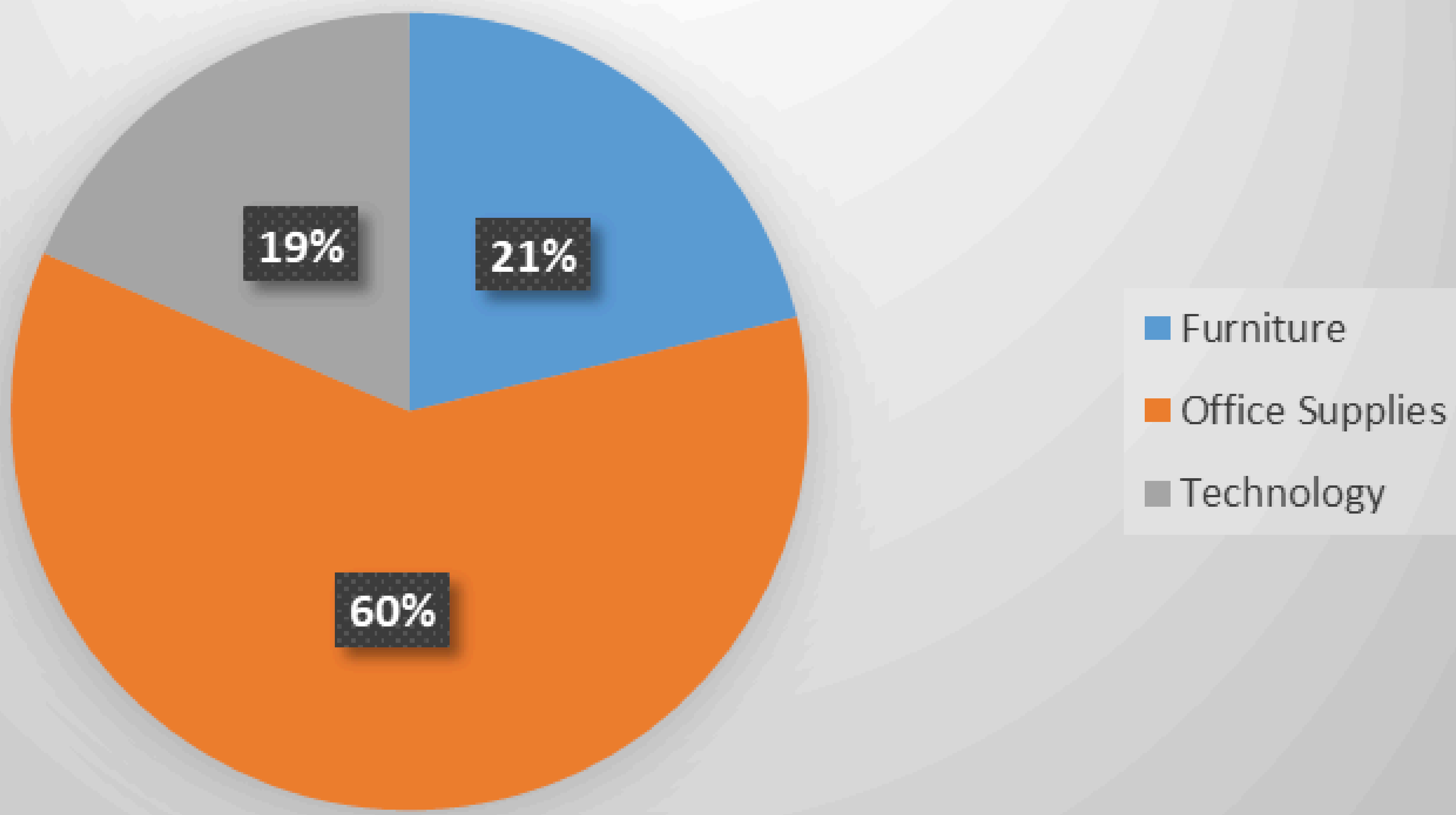


# Dashboard of Order data

Top categories in all sales



Comparing Segment and Sales

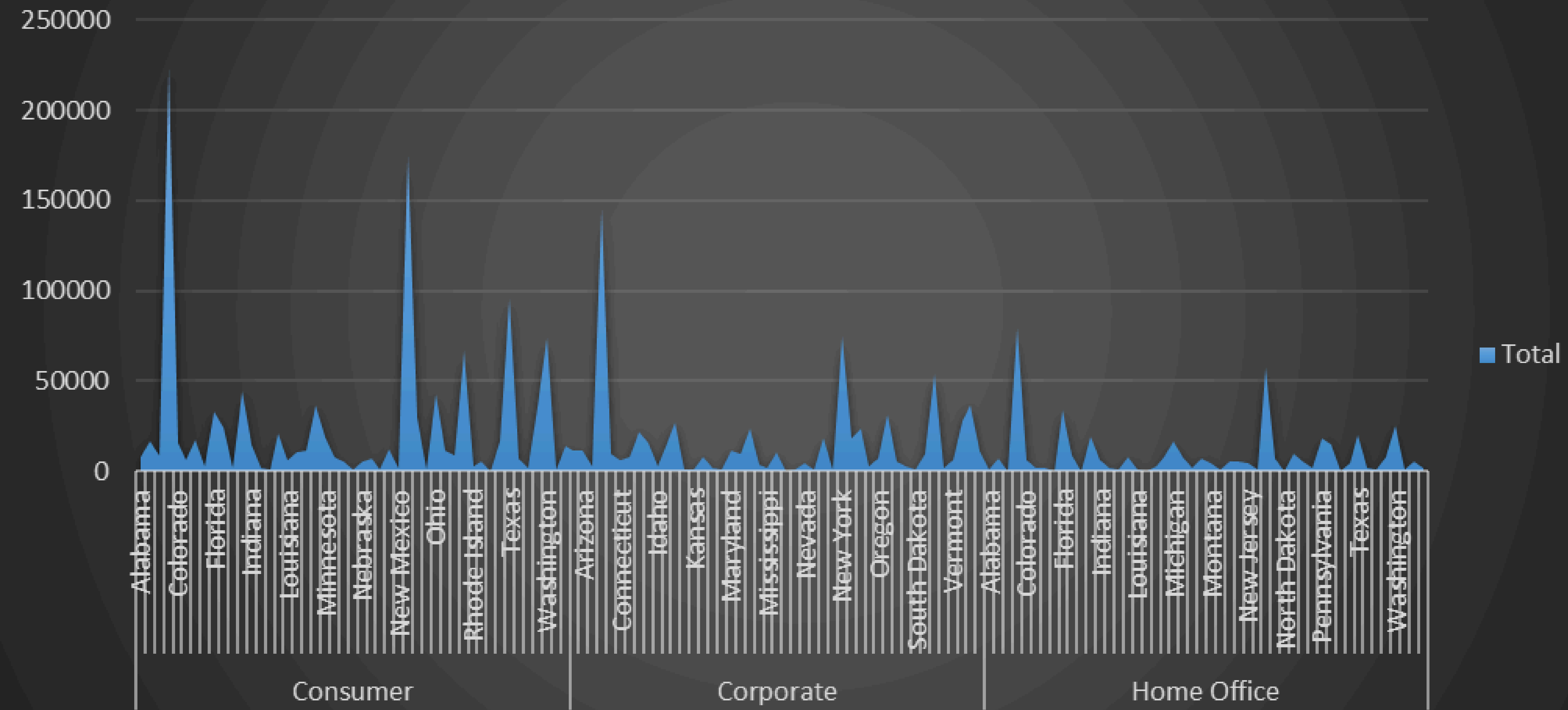
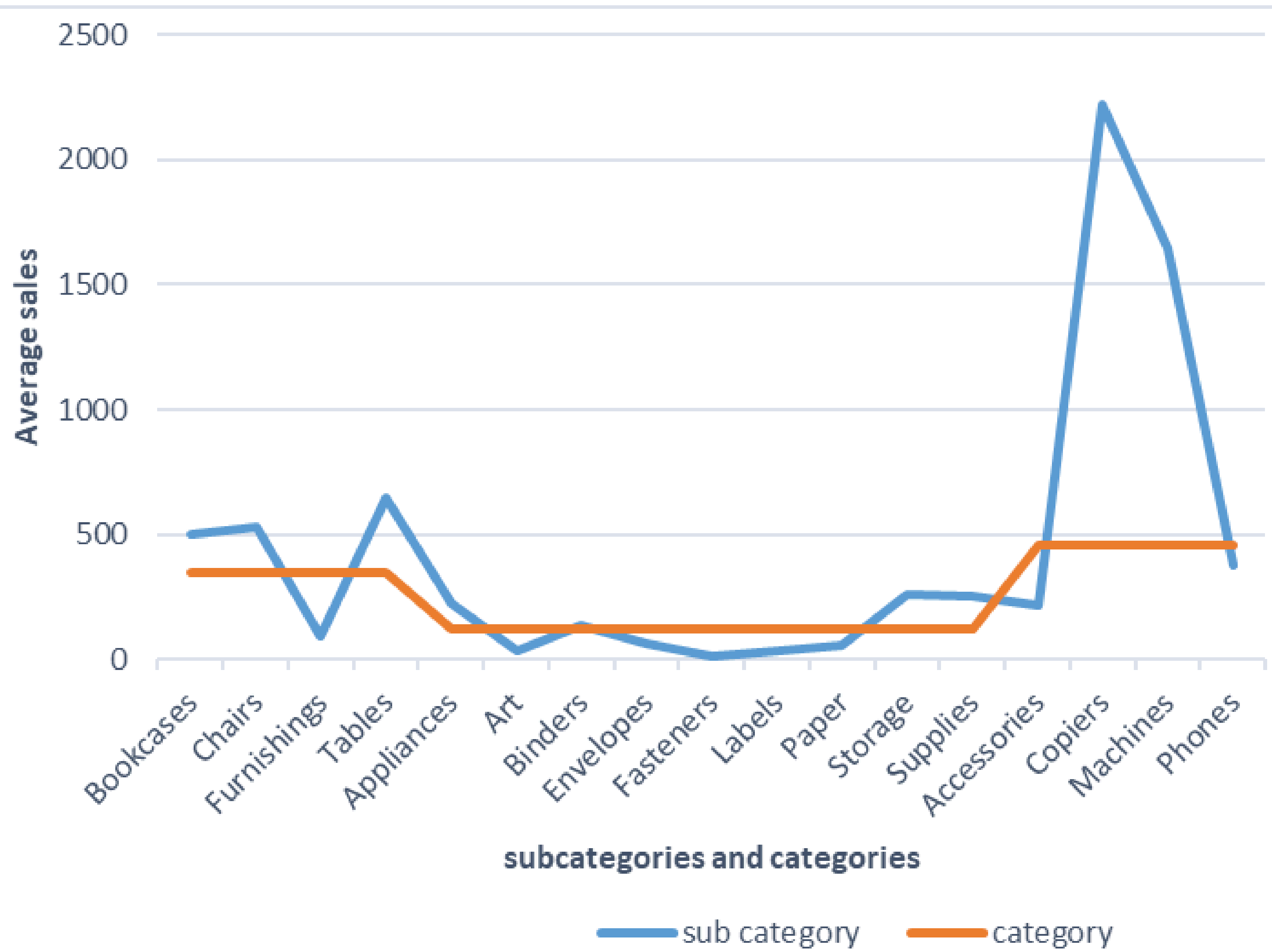
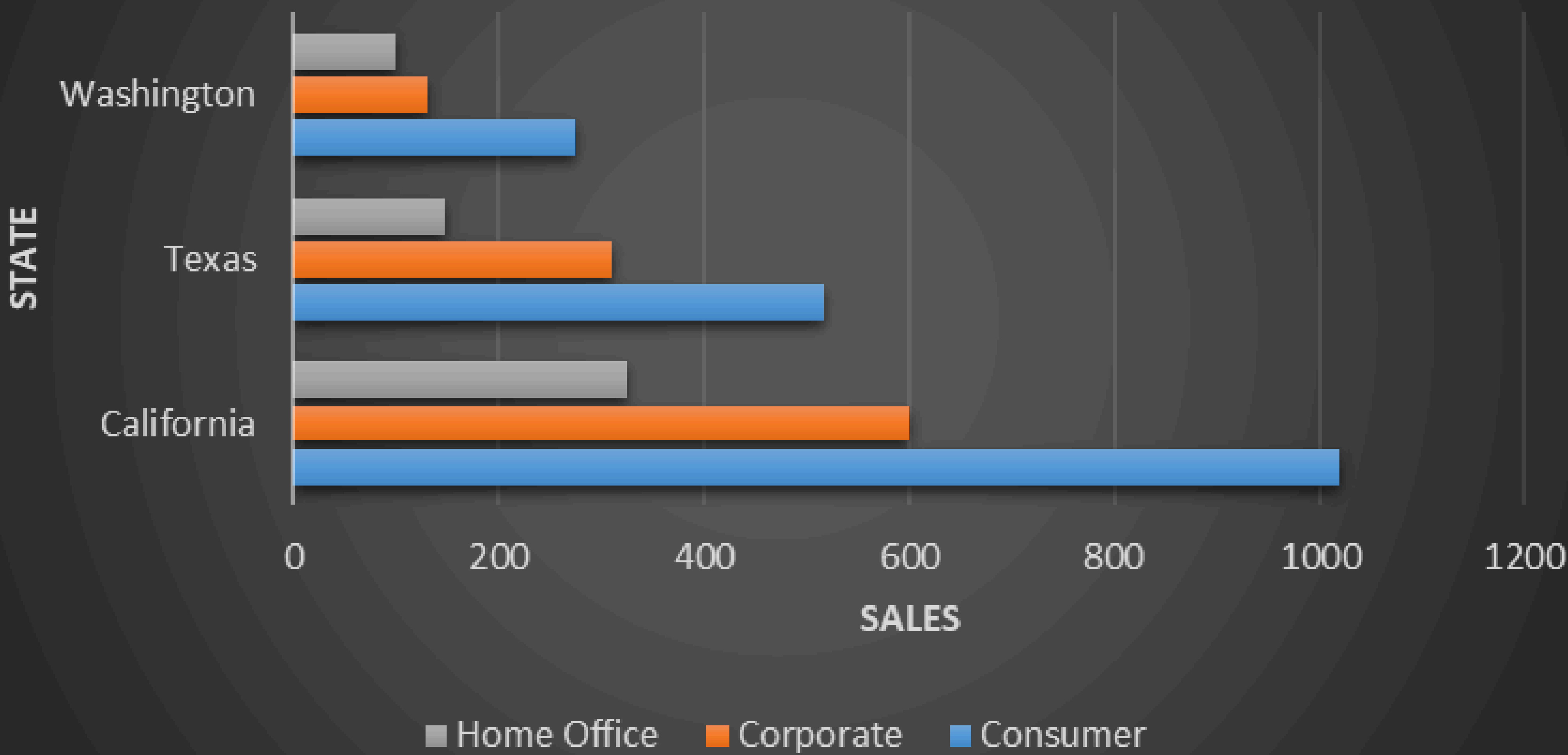
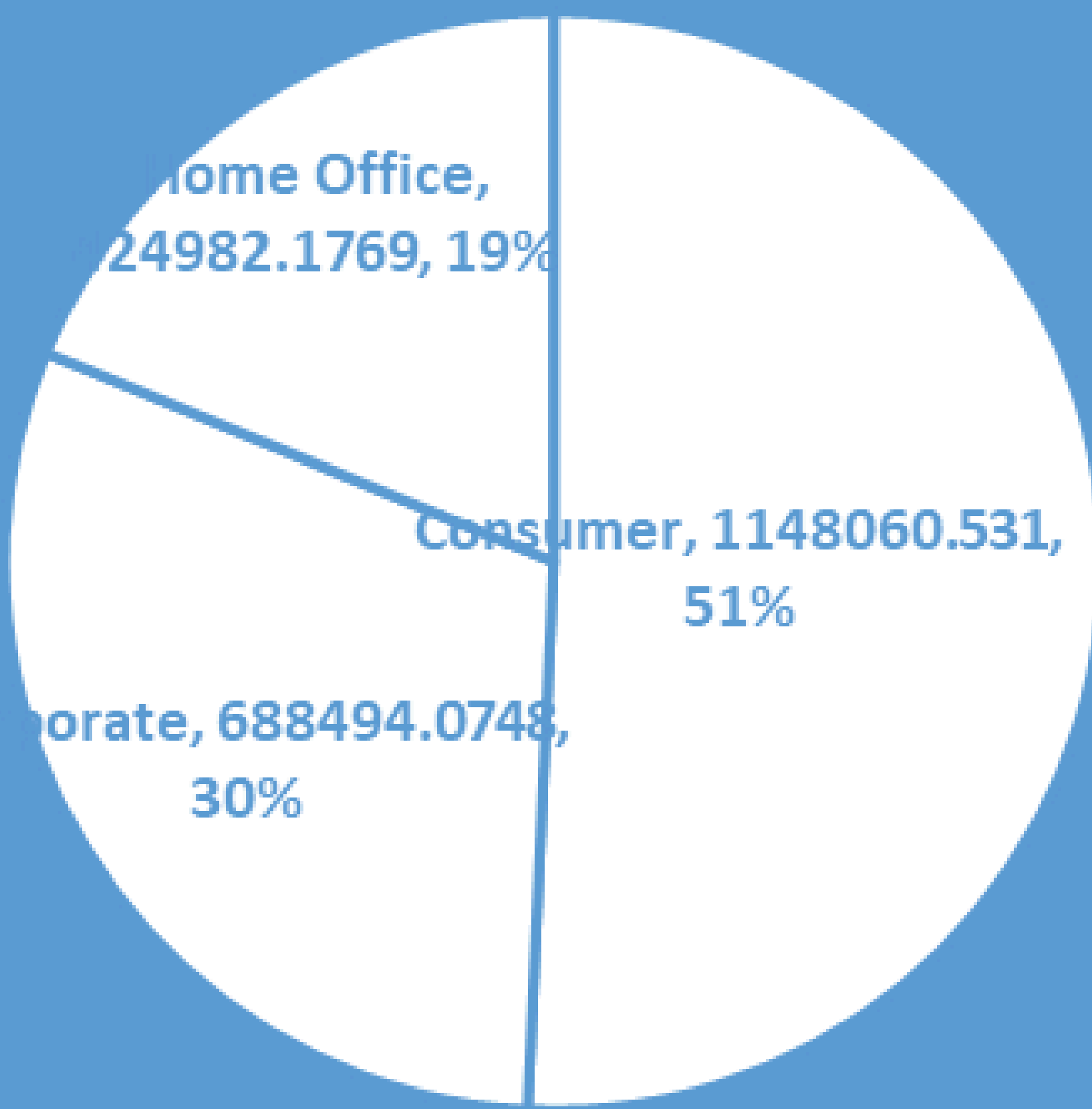


Chart Title

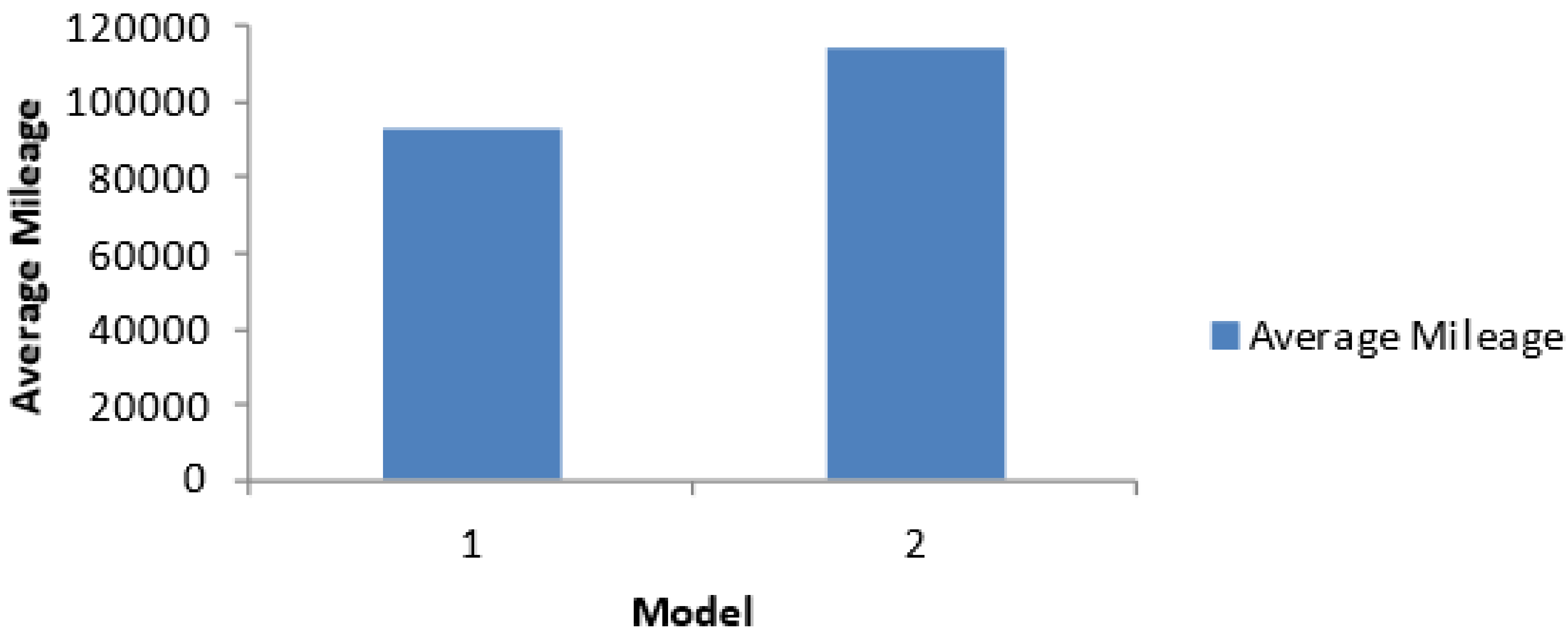


SUM OF SALES

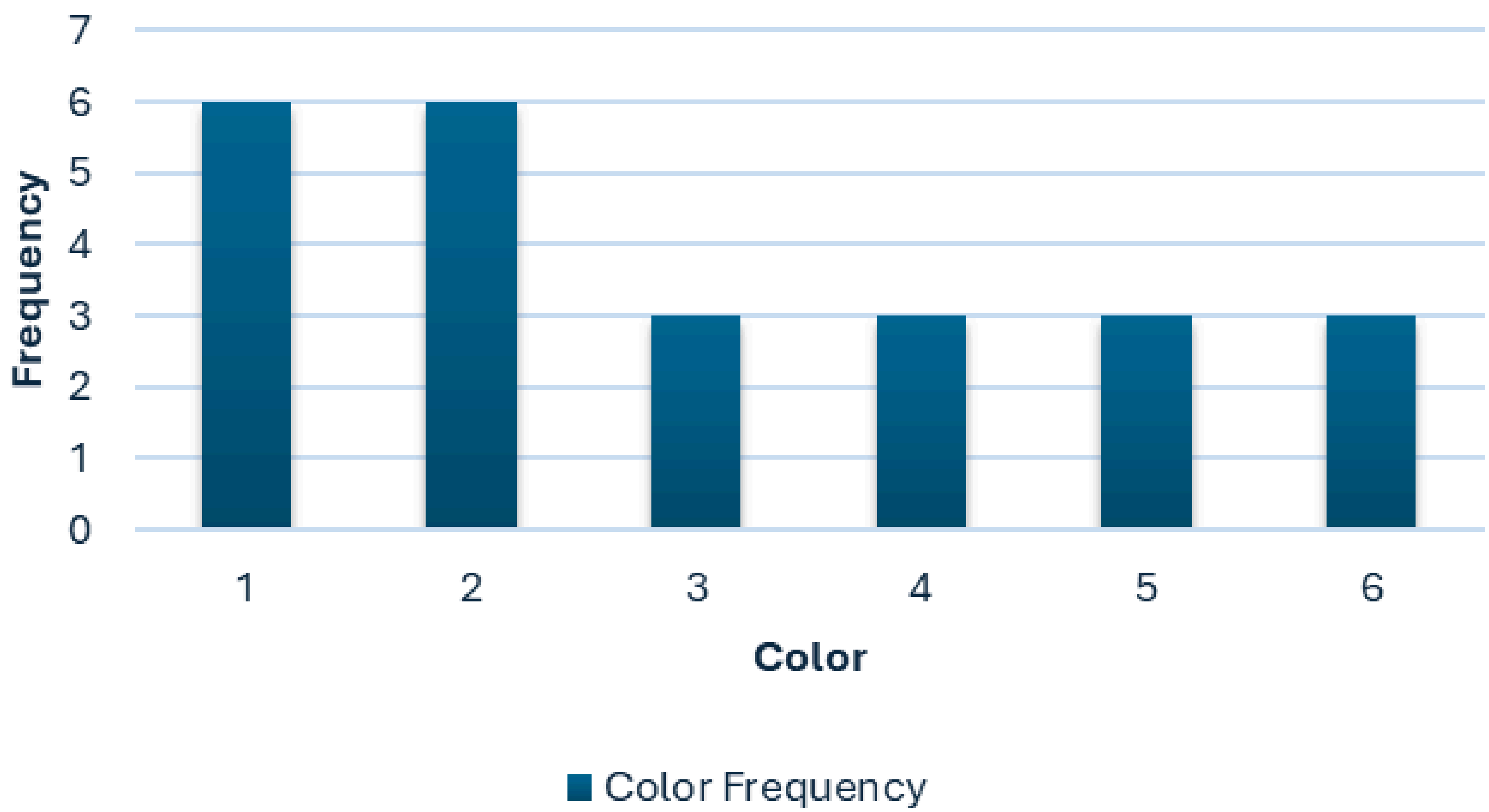


# Dashboard of Car Collection

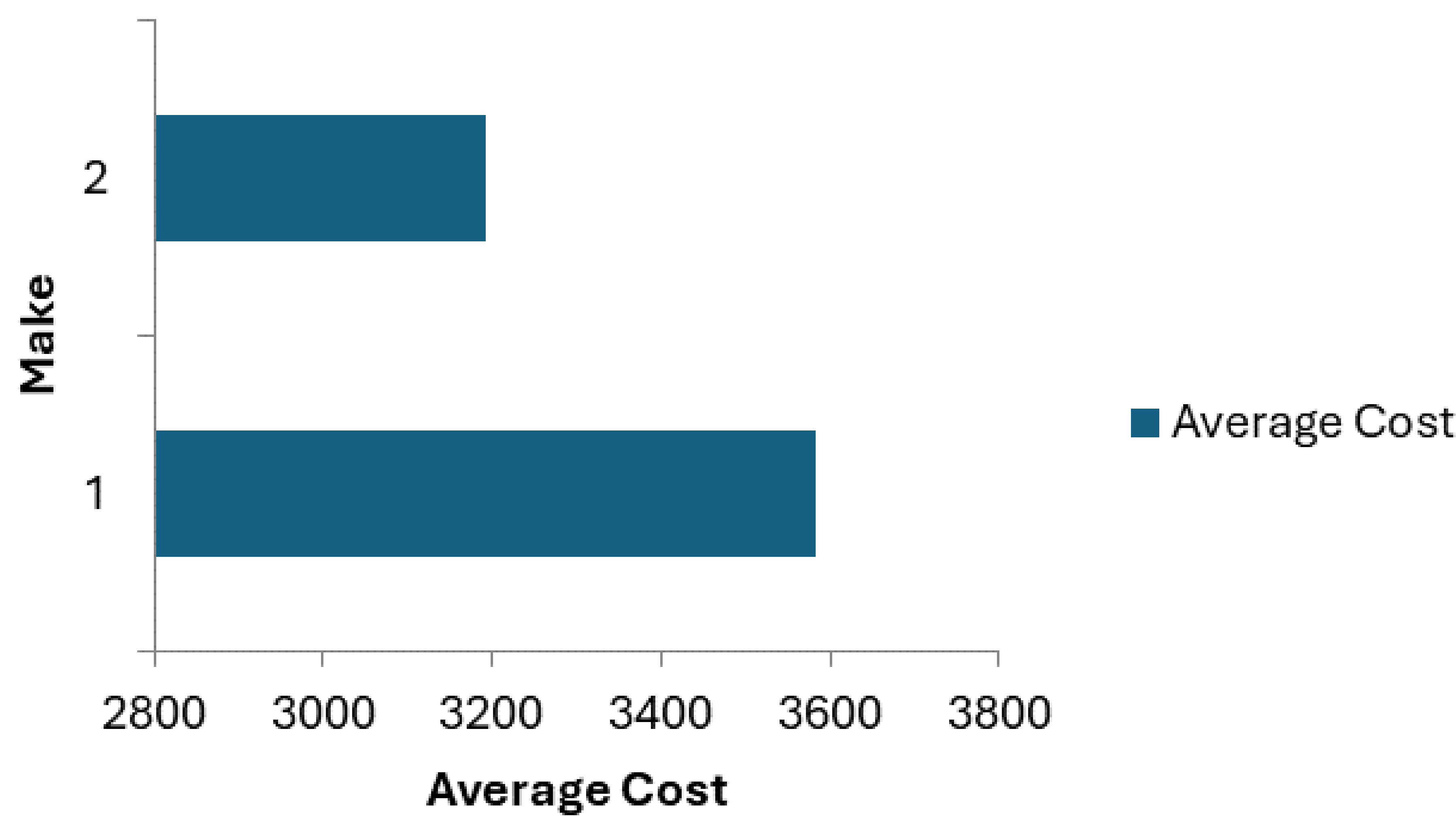
## Comparison of Mileage: Chevrolet Impala vs Toyota Corolla



## Car Color Popularity



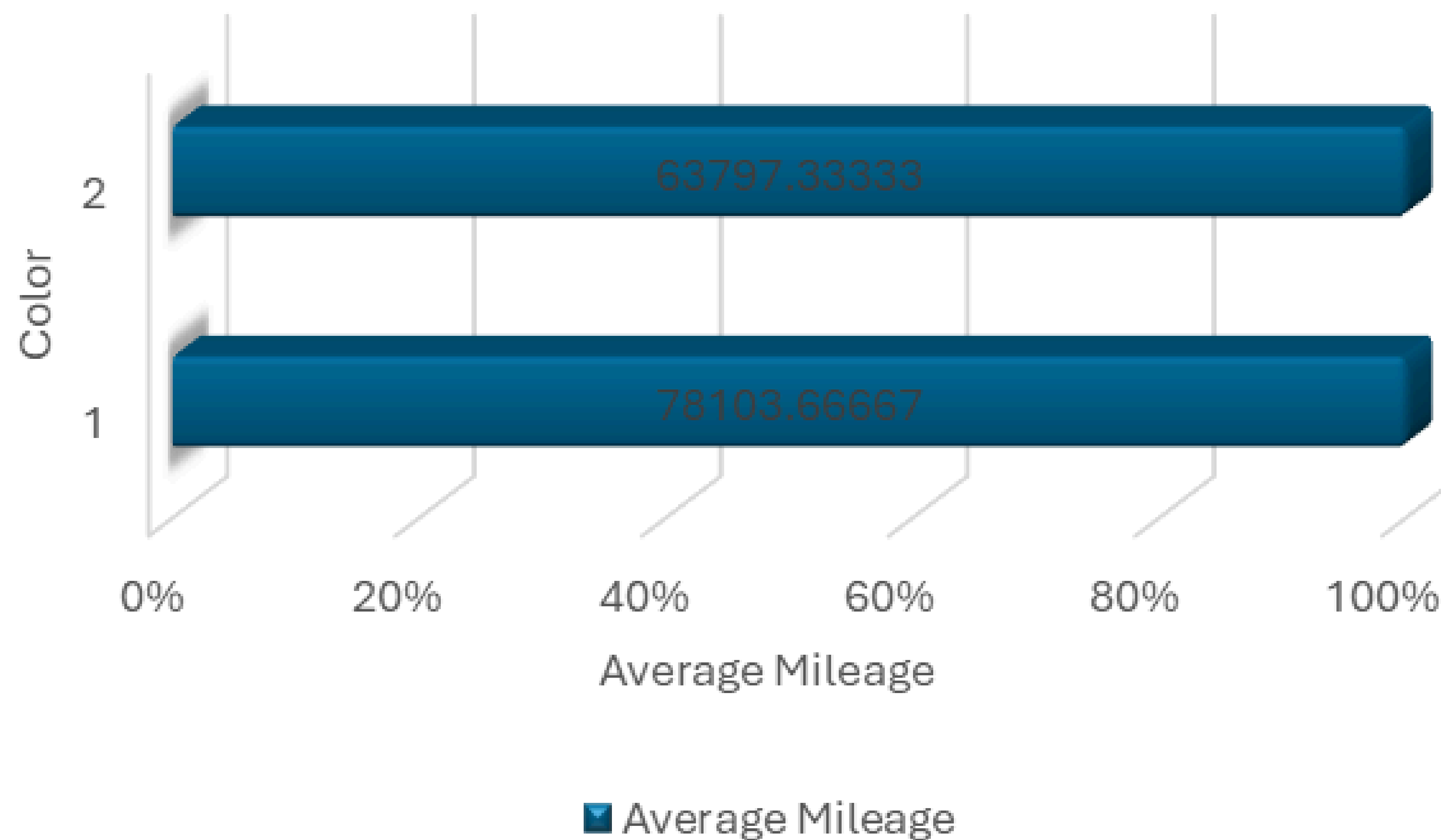
## Average Cost Comparison



## CARS WITH TOTAL COST ABOVE \$2000

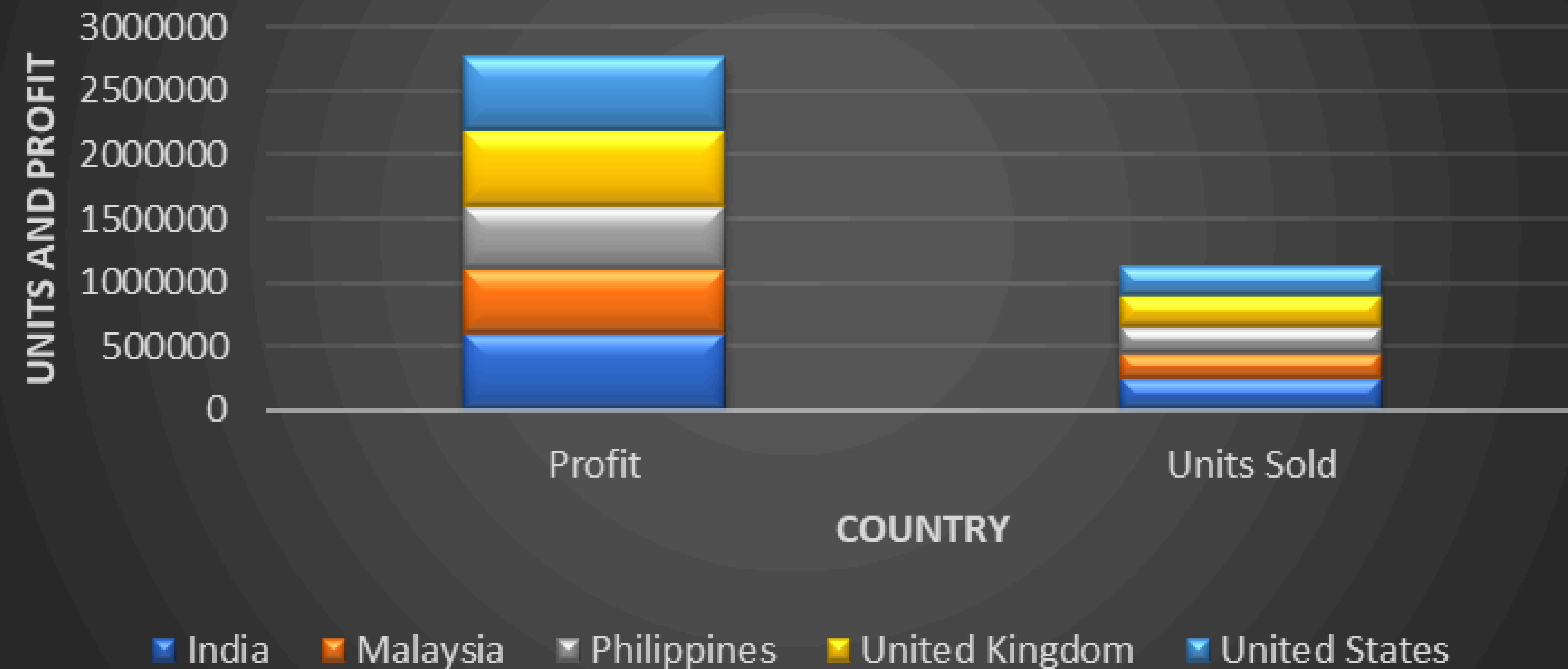


## Mileage Comparison: Silver vs Green Cars

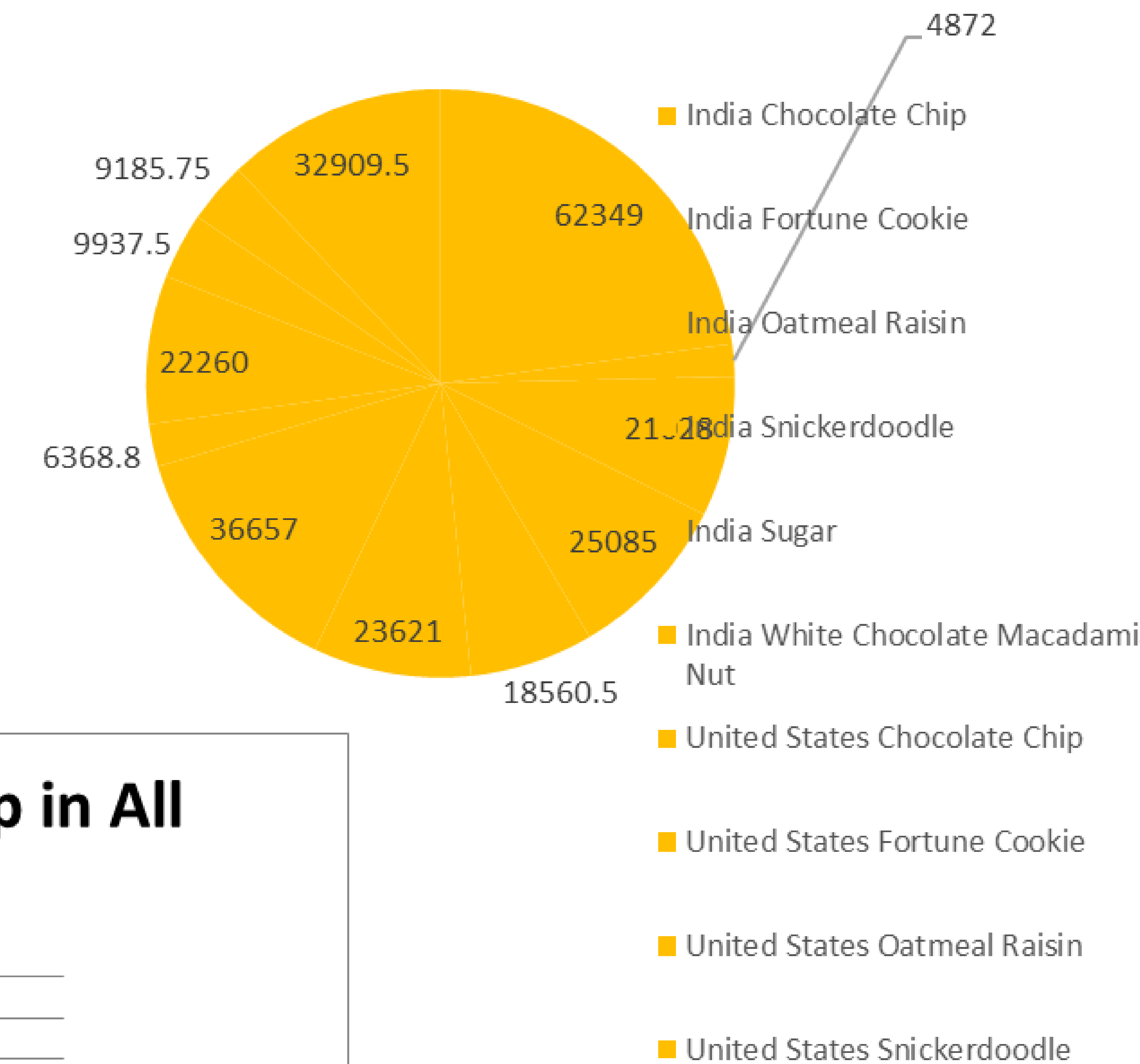


# Dashboard of Cookie data

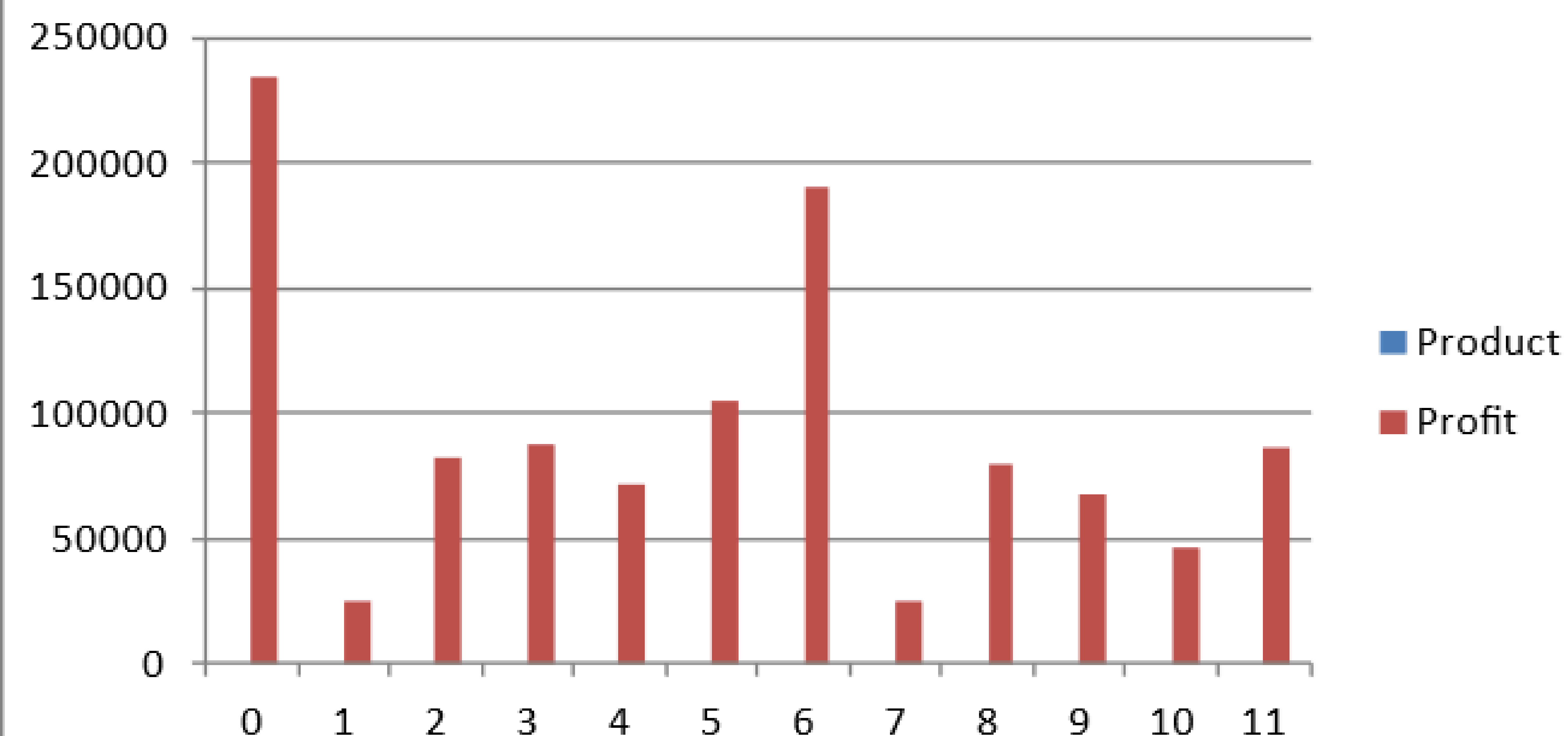
## Performance of Countries on the basis of Units Sold



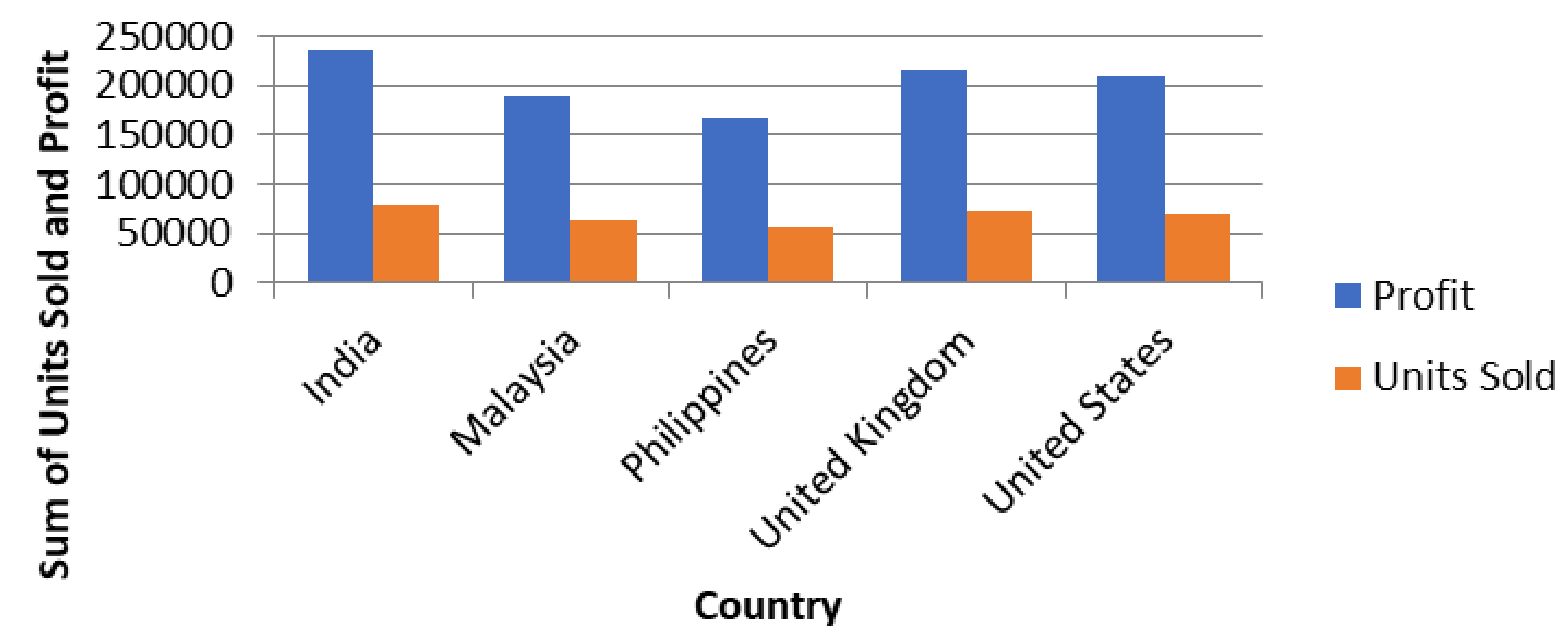
## Performance of all cookies in India and US in 2019



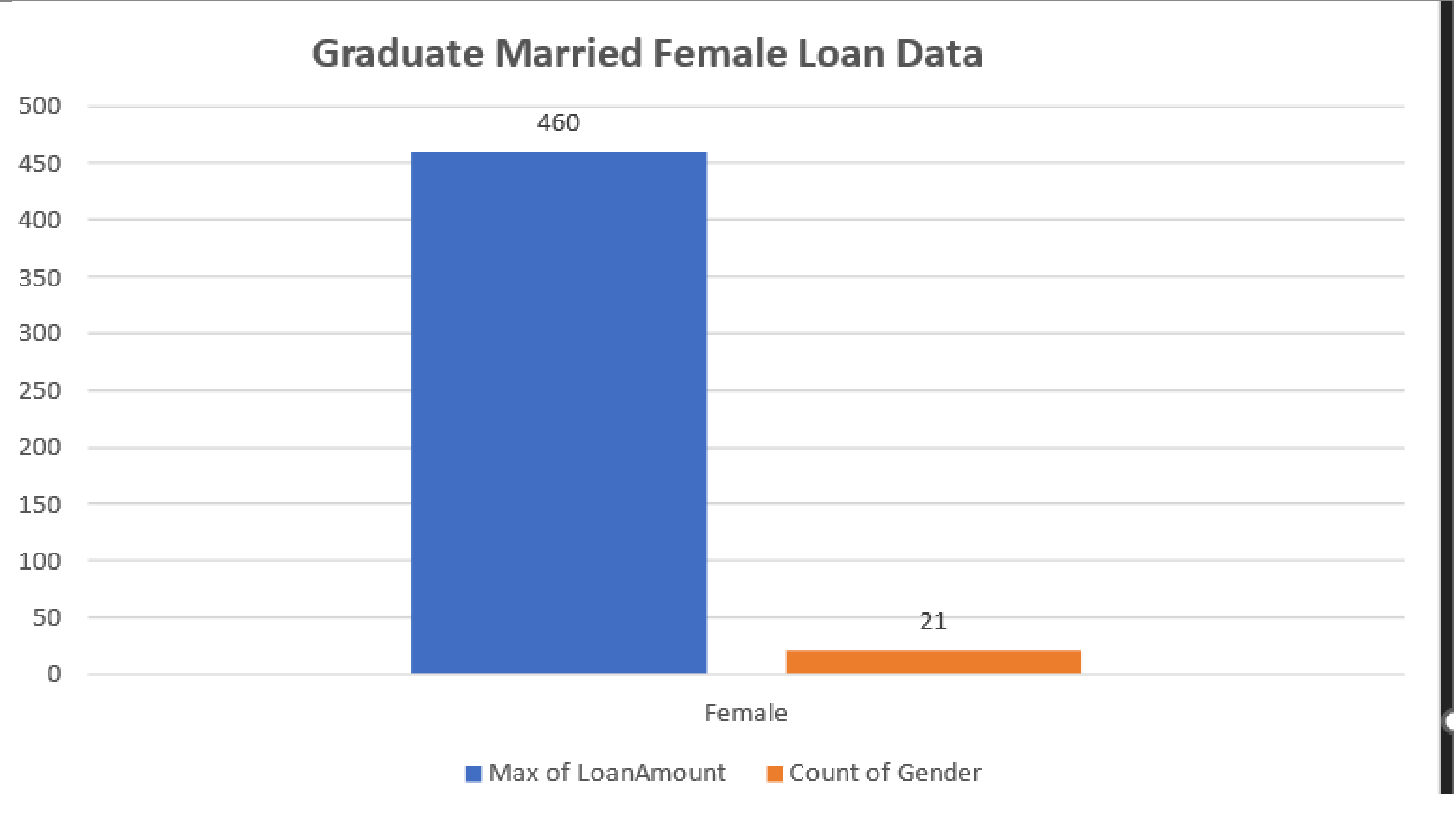
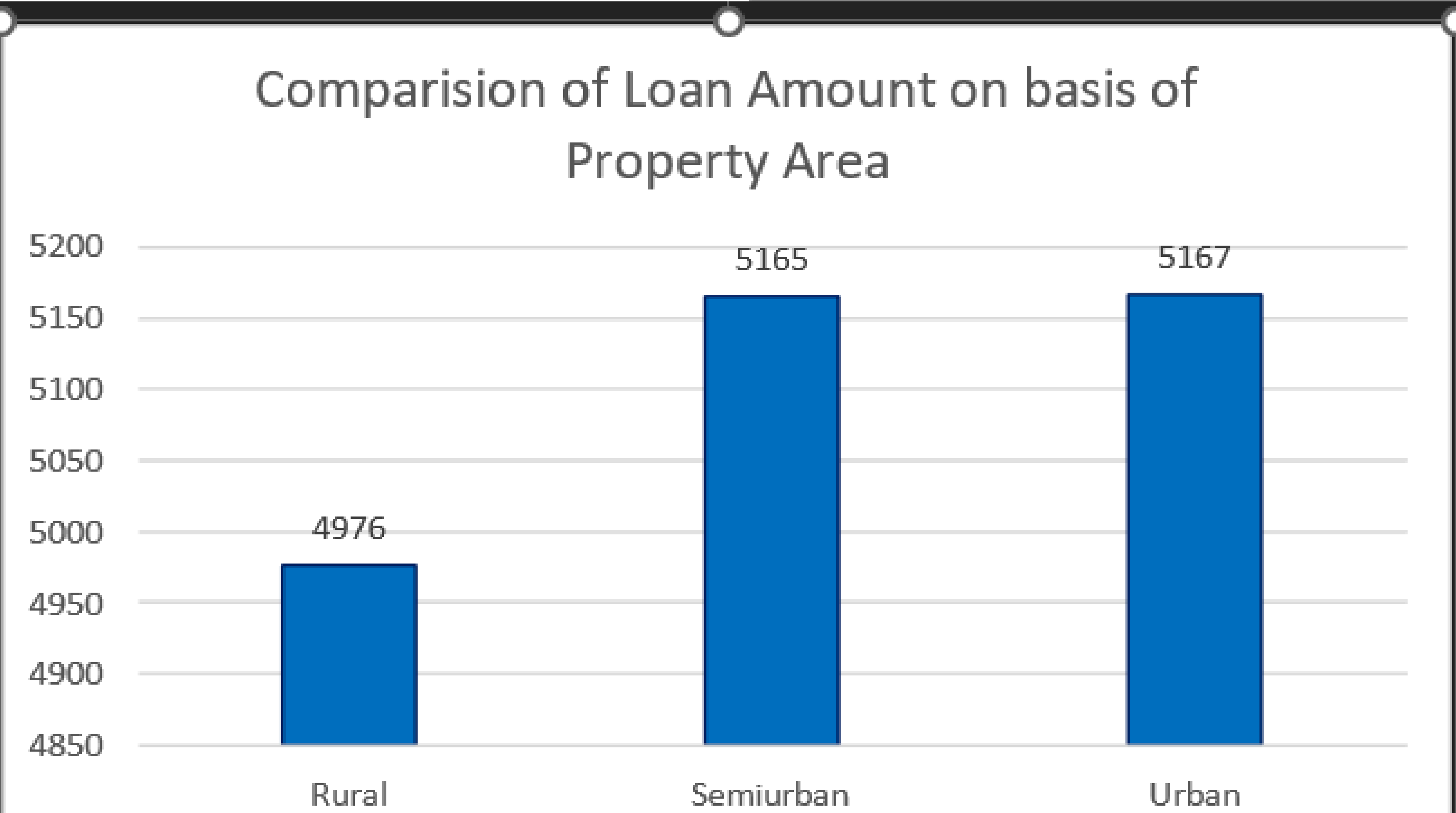
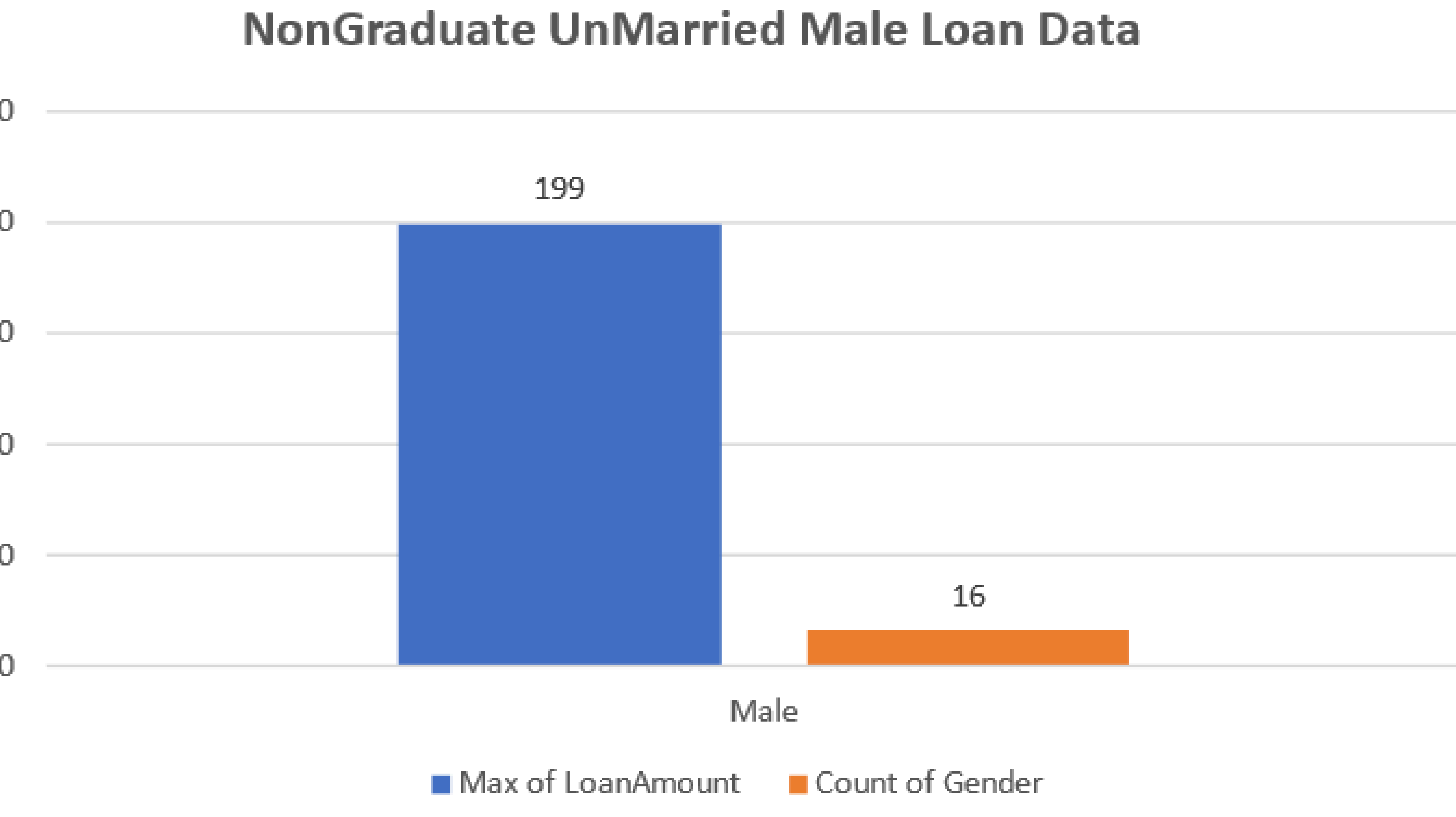
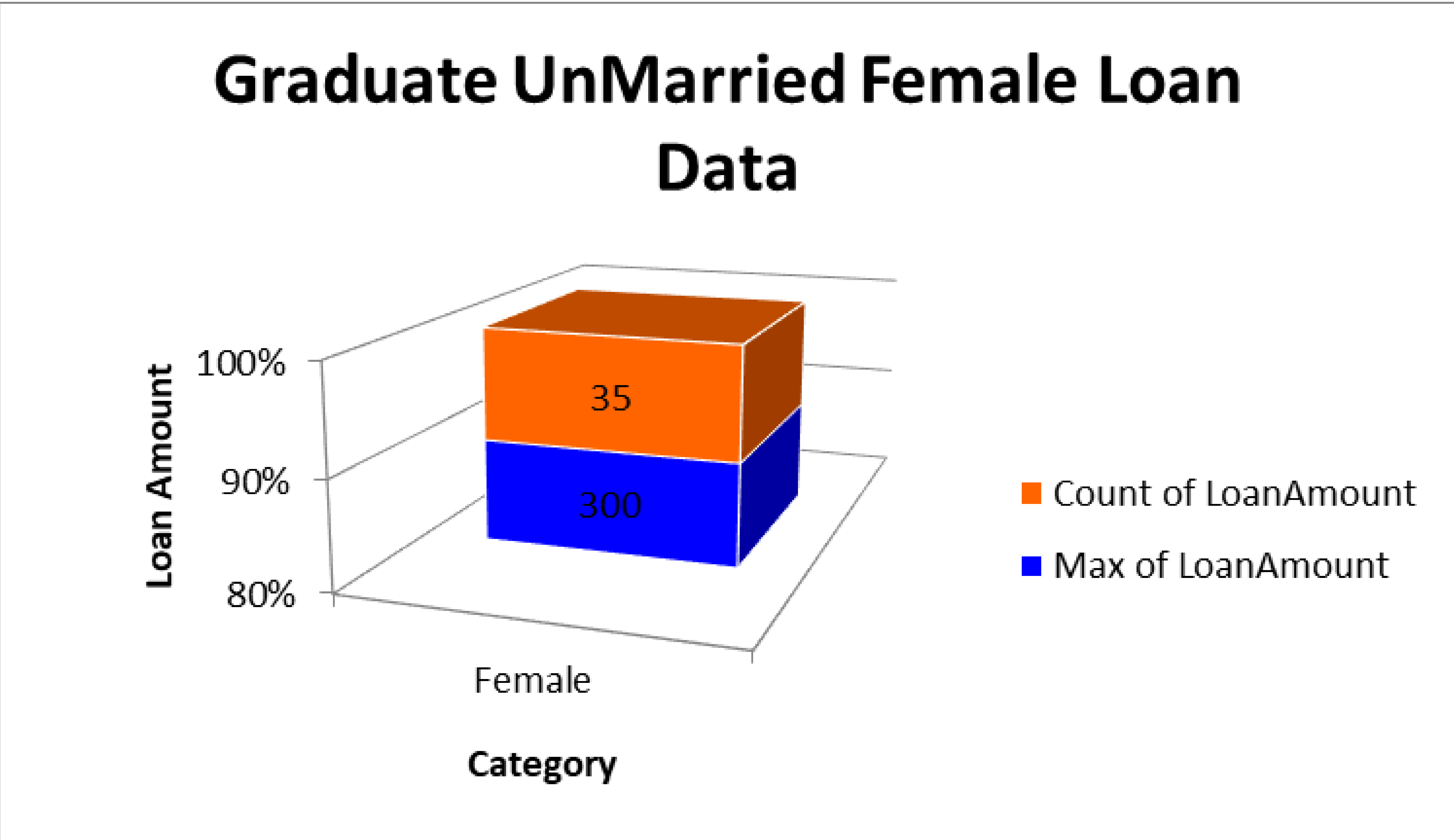
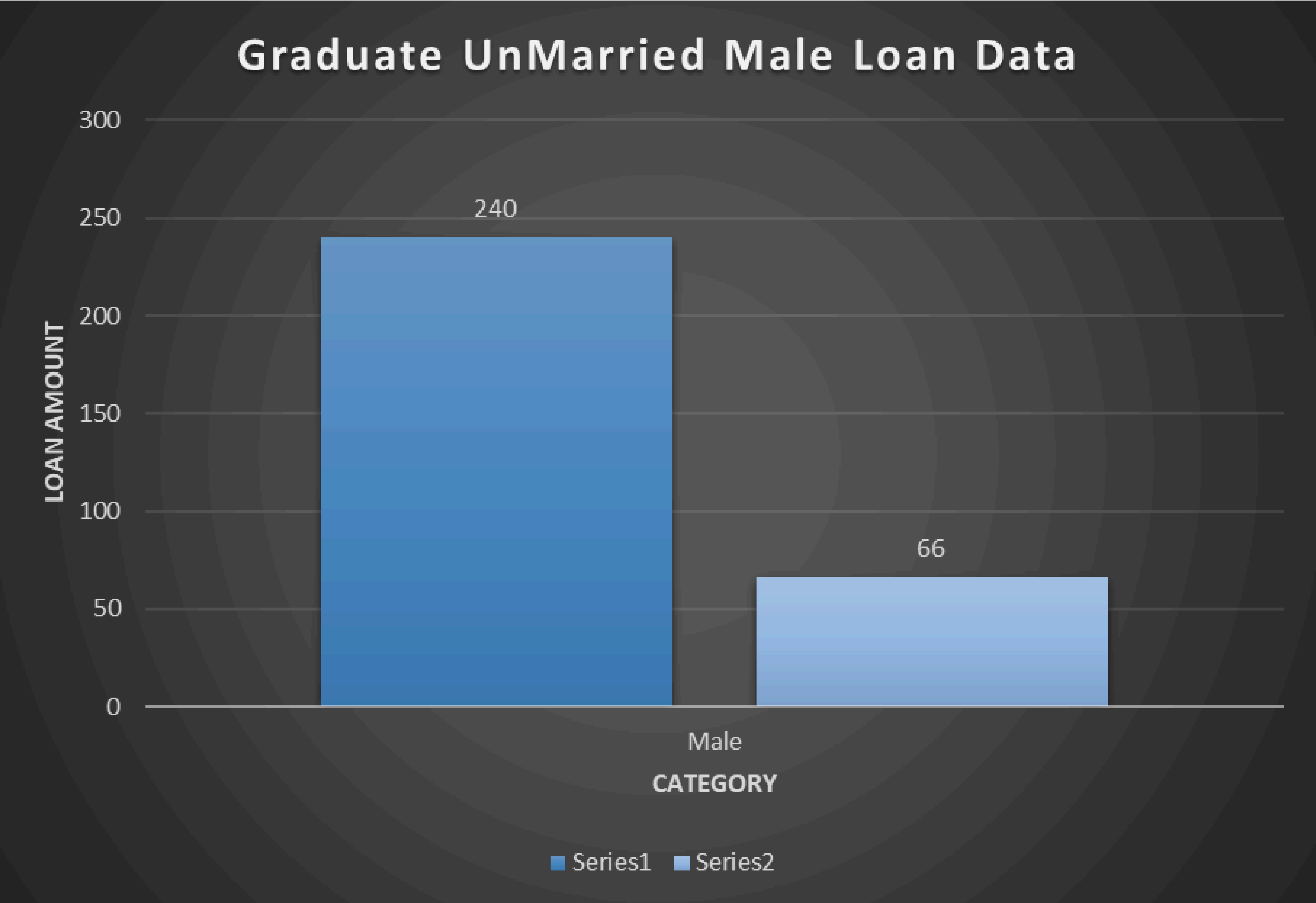
## Chart for Profit by Cookie Types



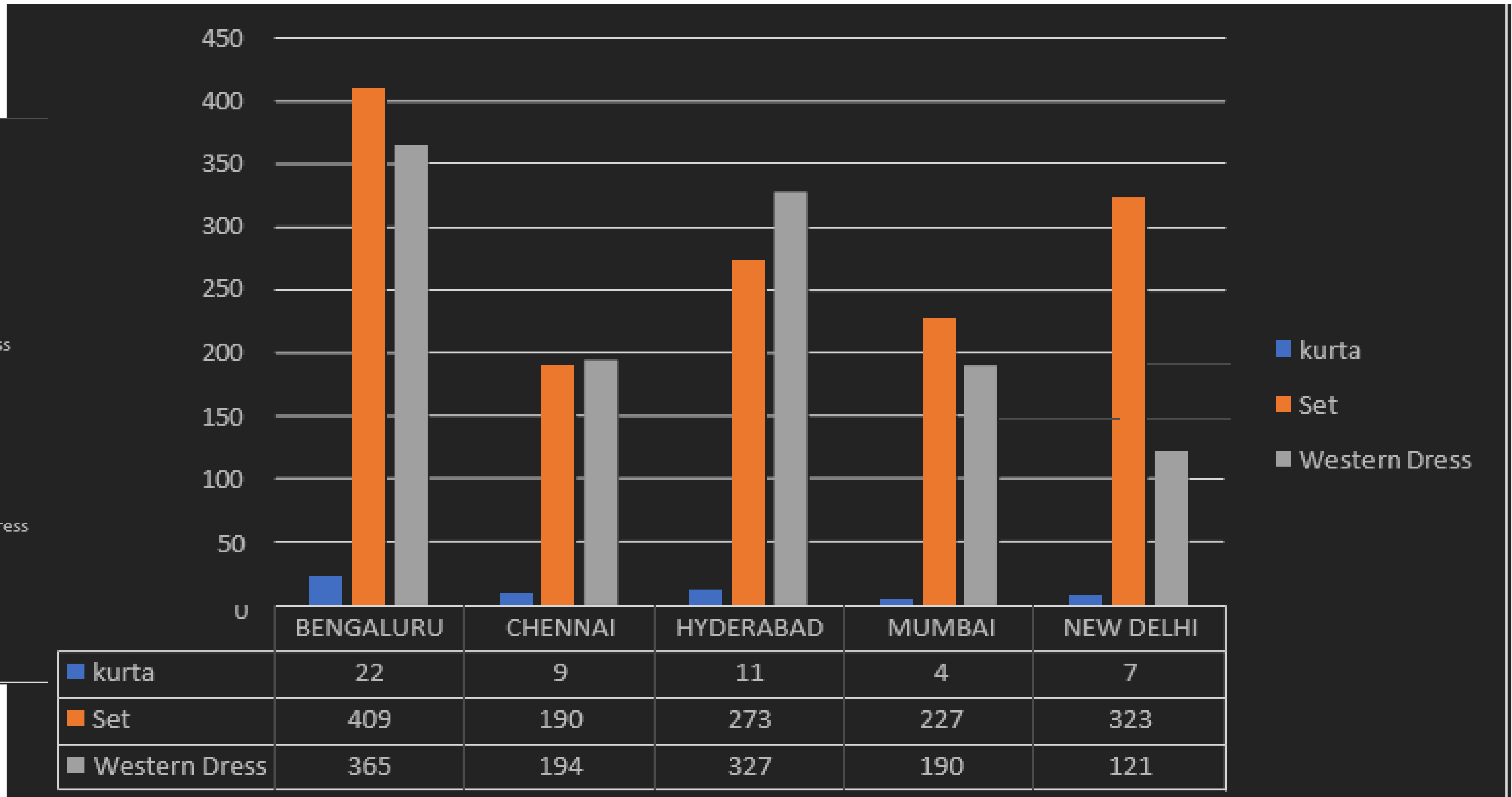
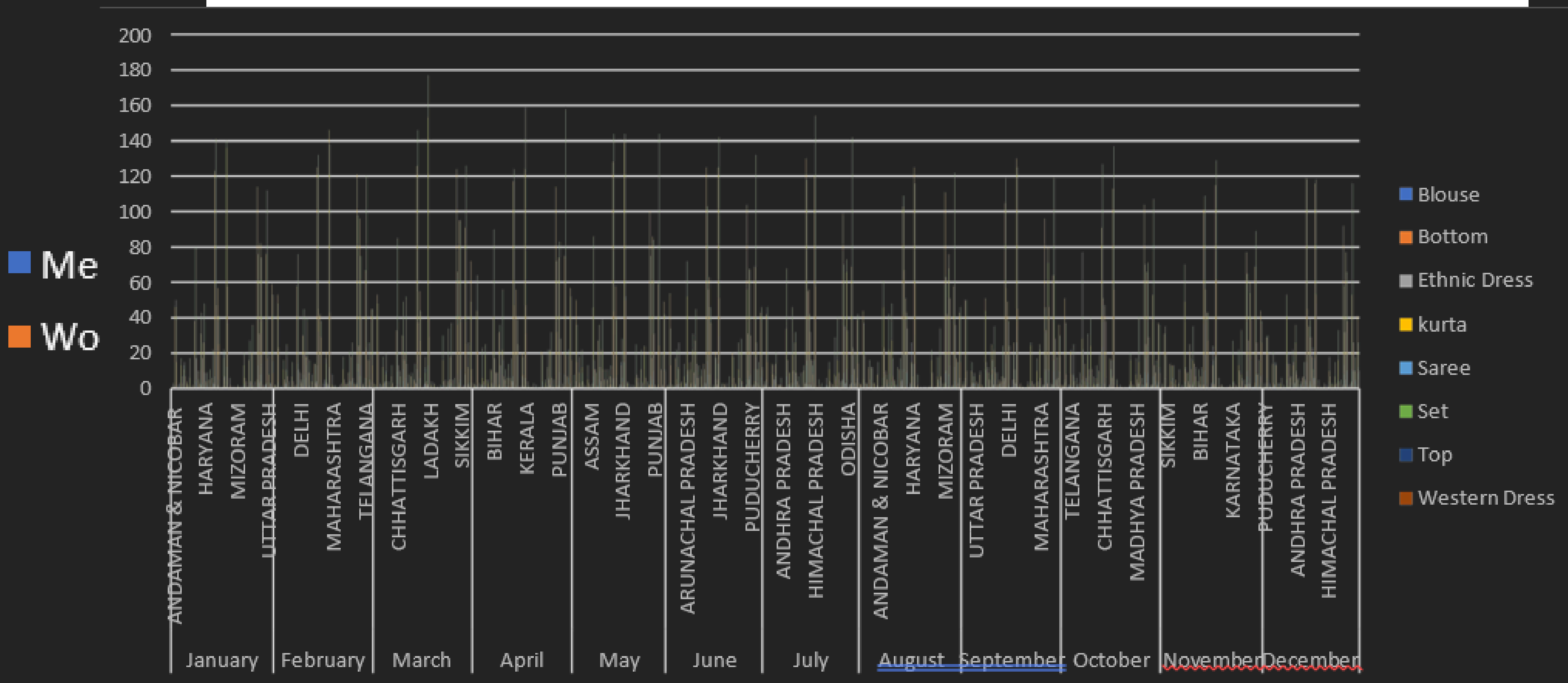
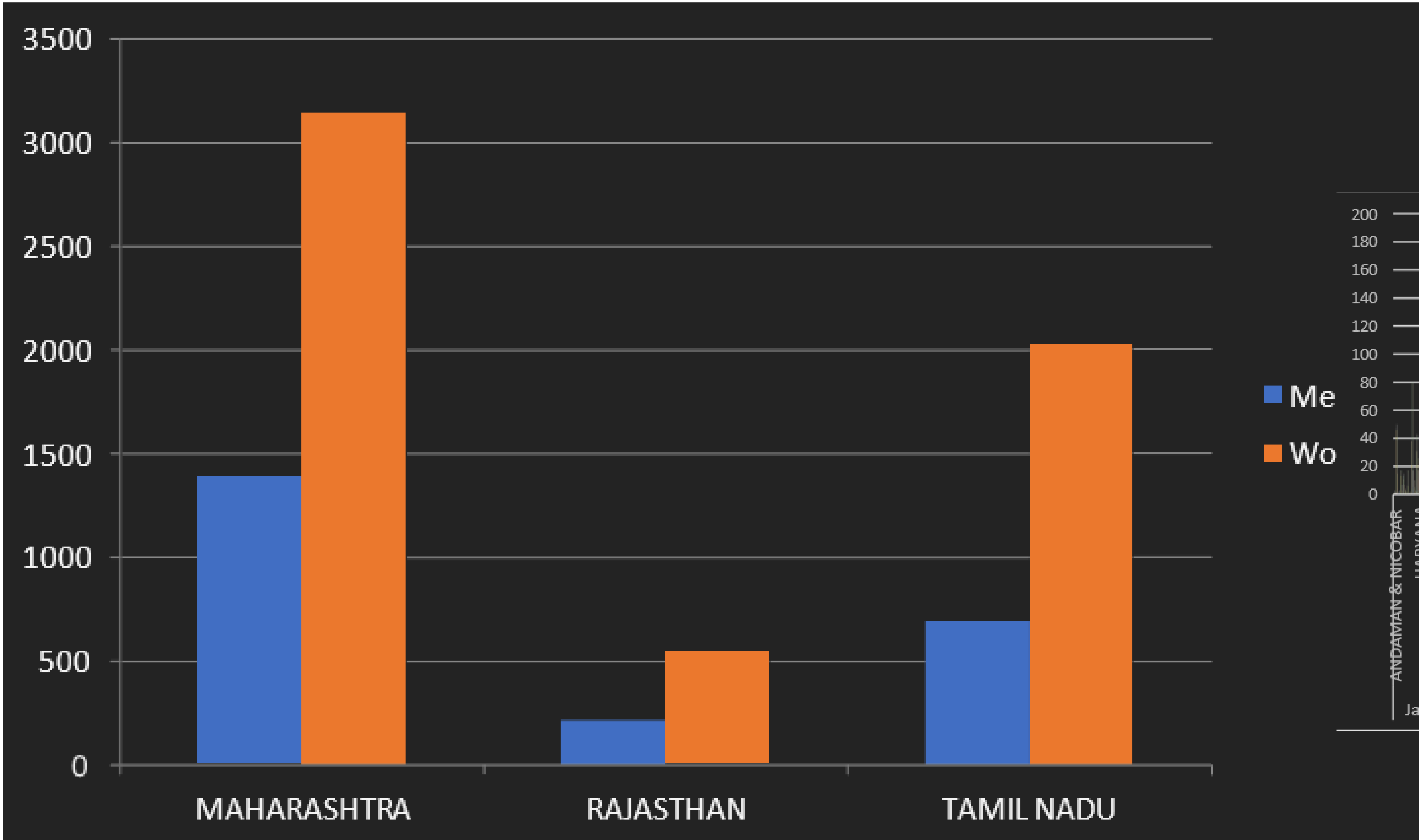
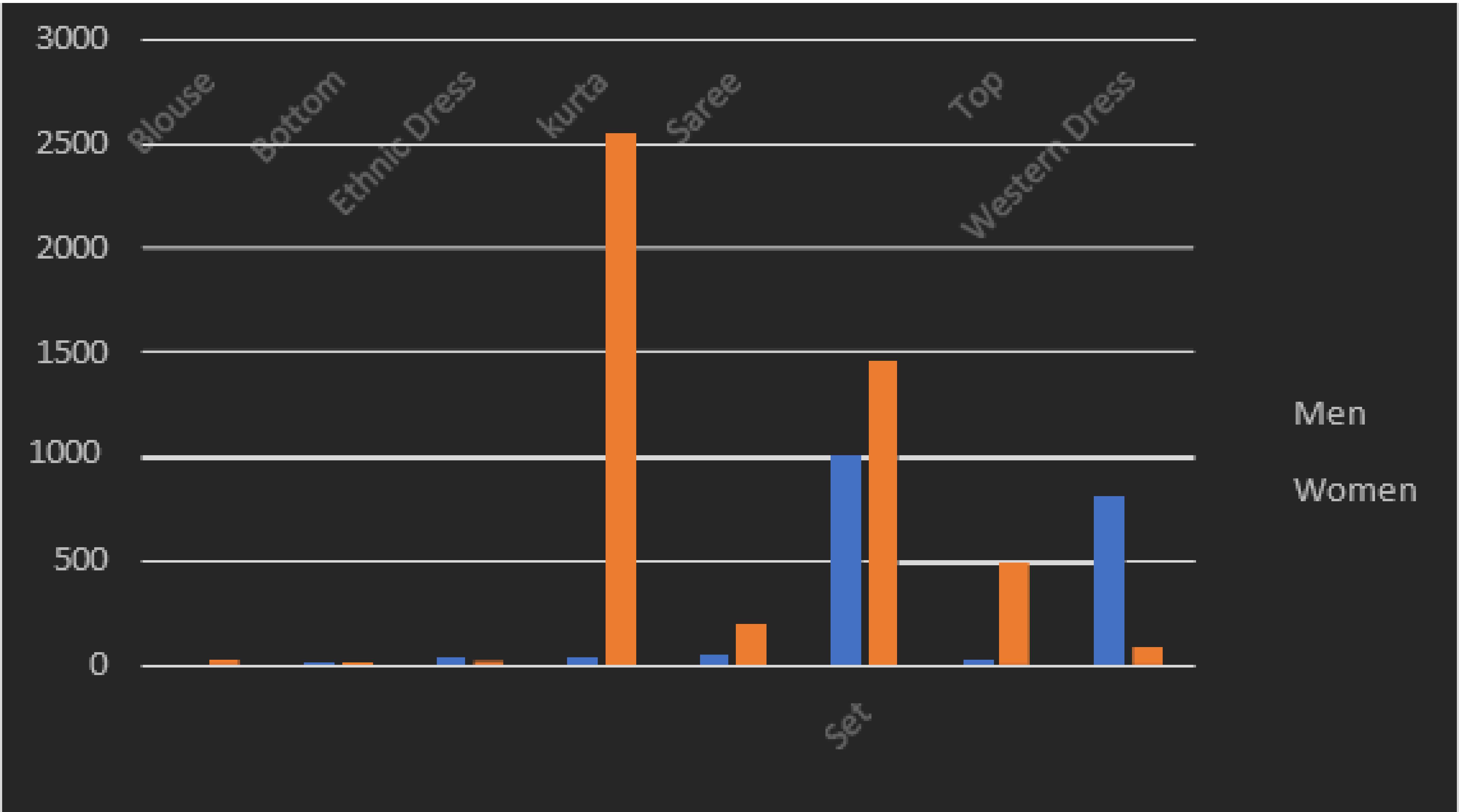
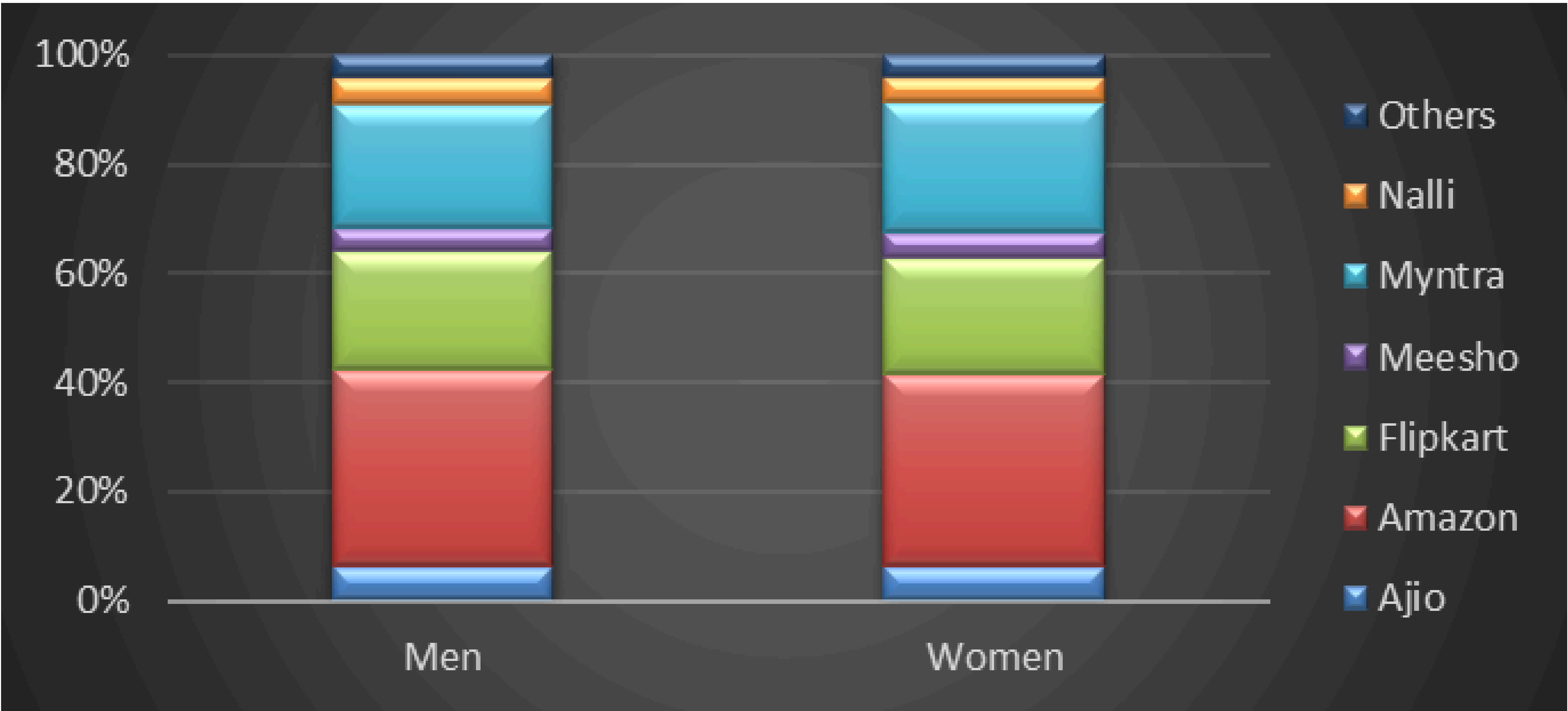
## Performance of Chocolate Chip in All Countries



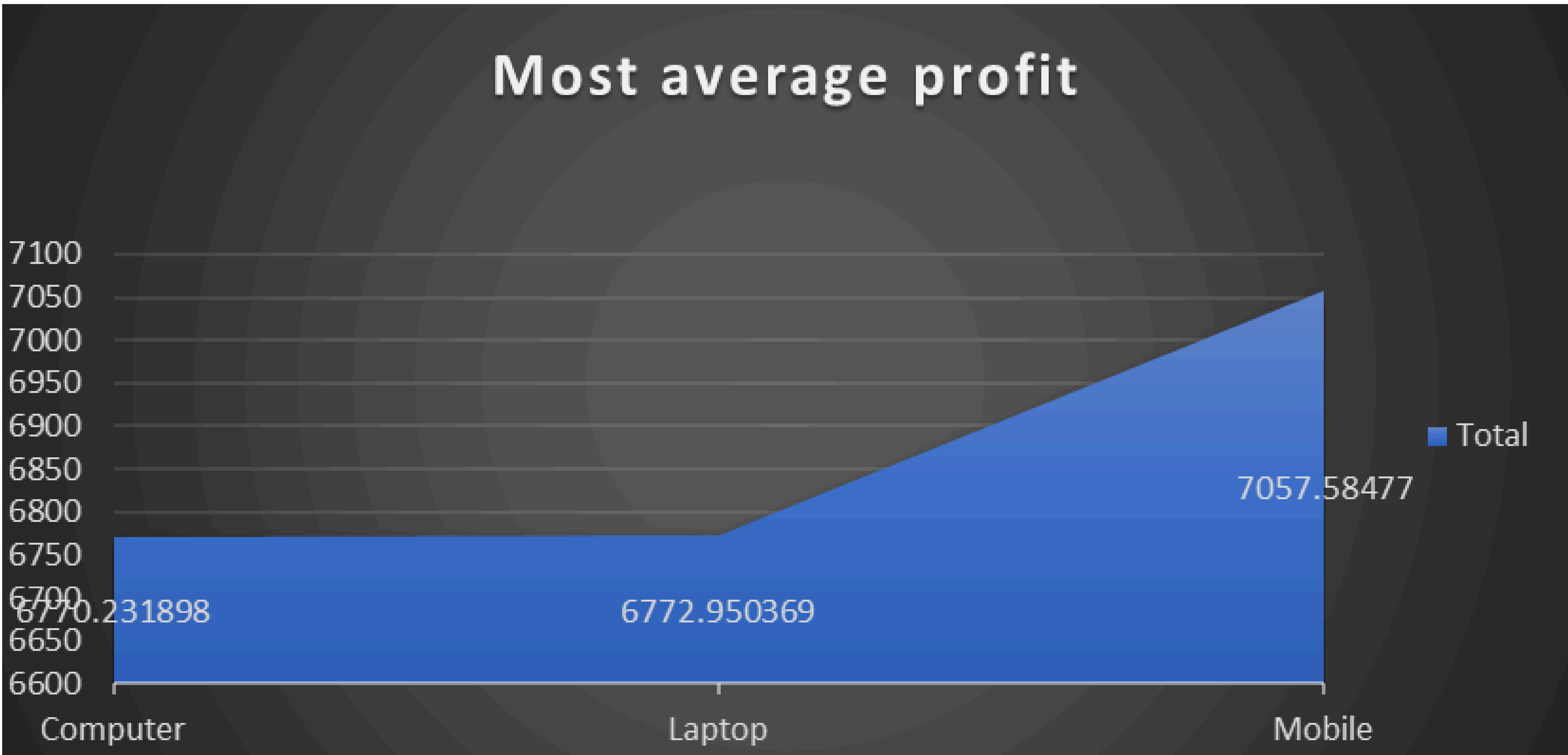
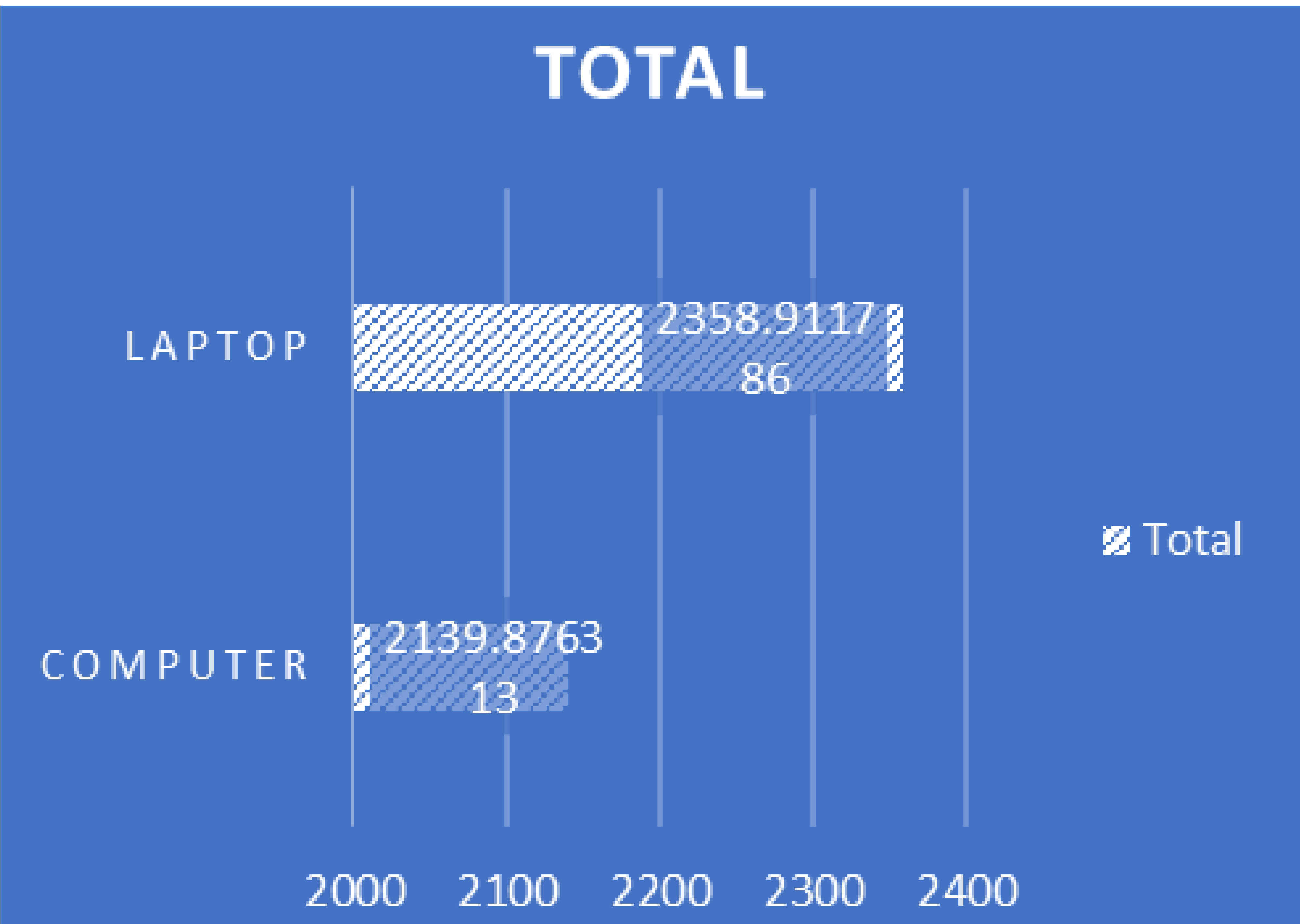
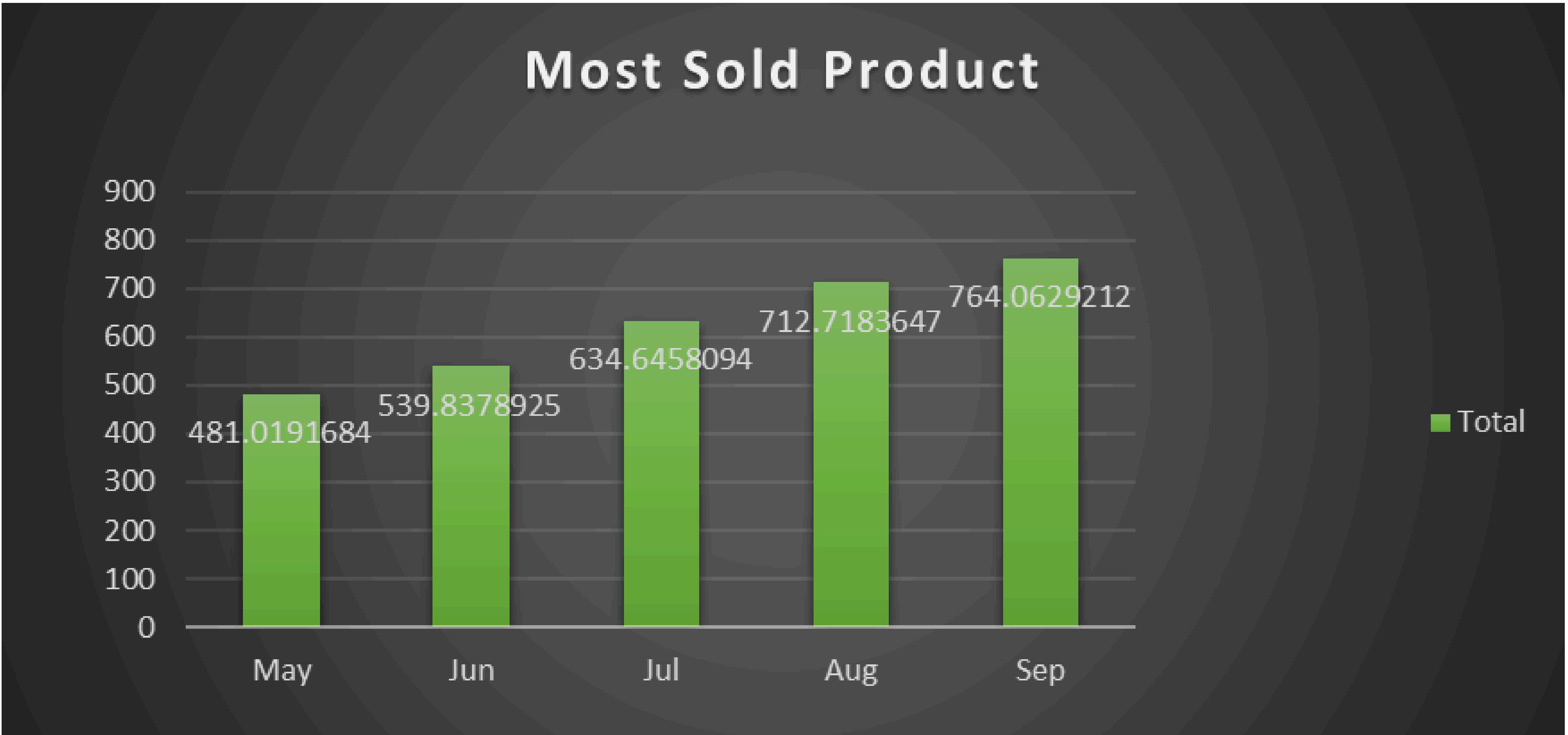
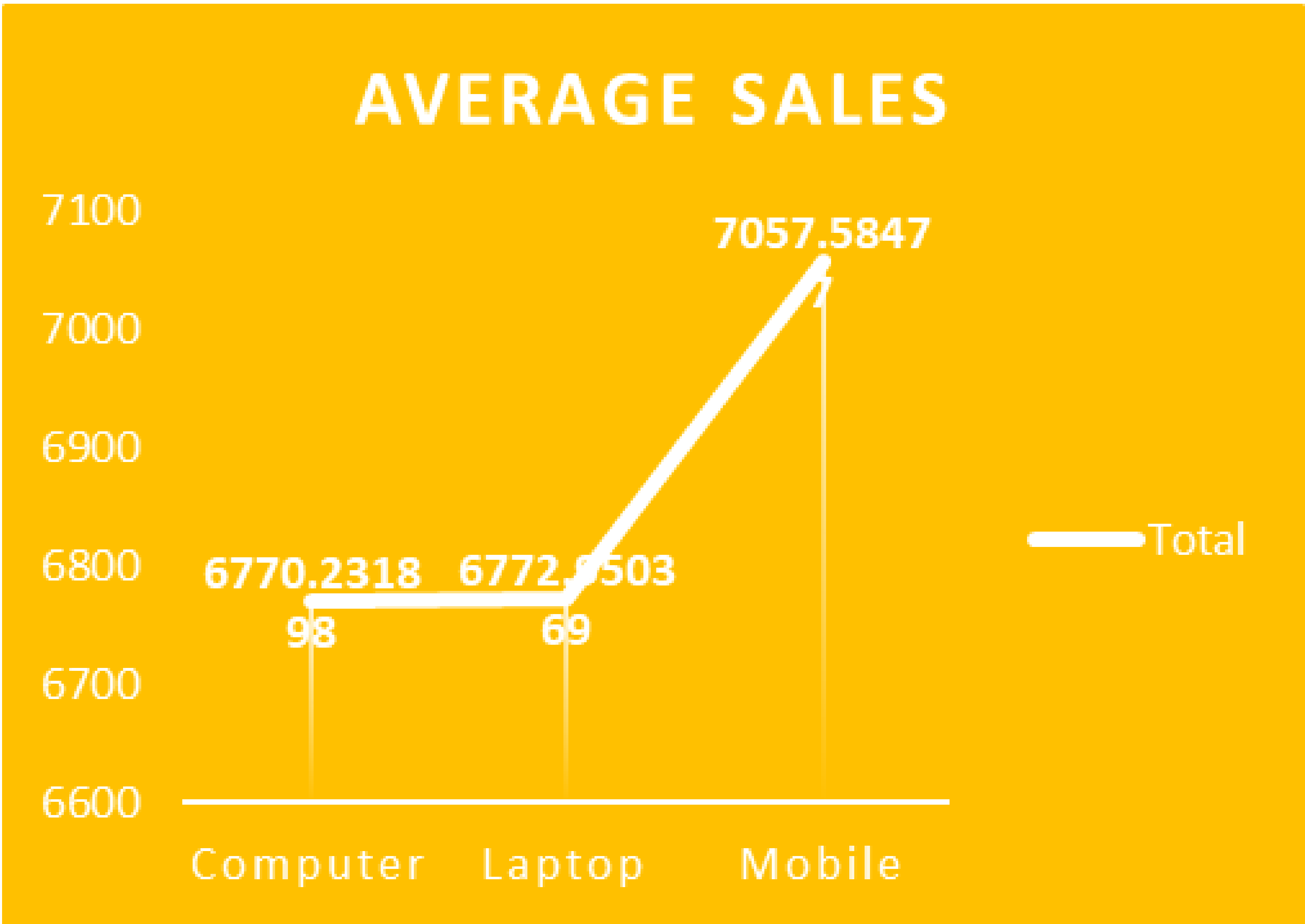
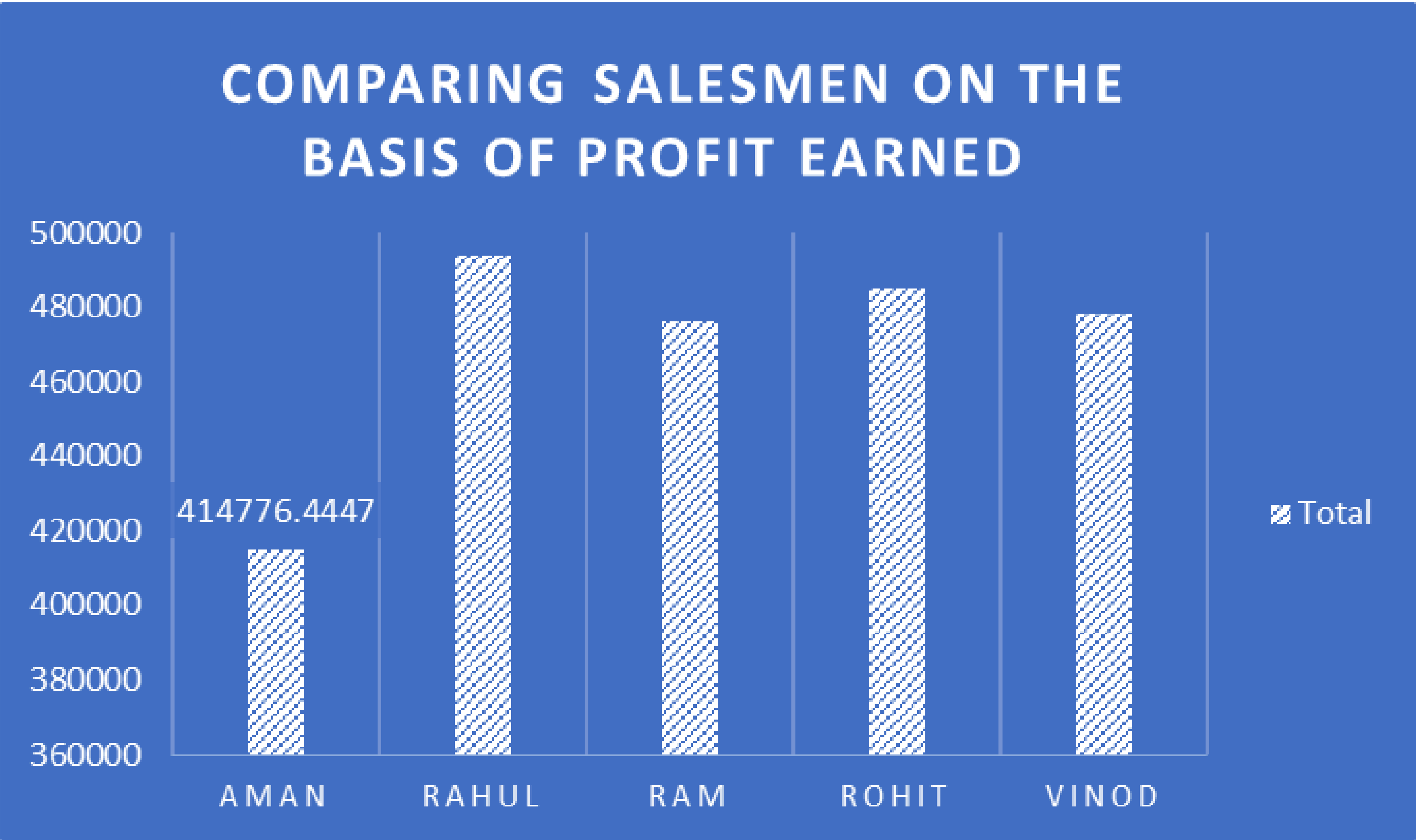
# Dashboard of Loan Data



# Dashboard of Store Data

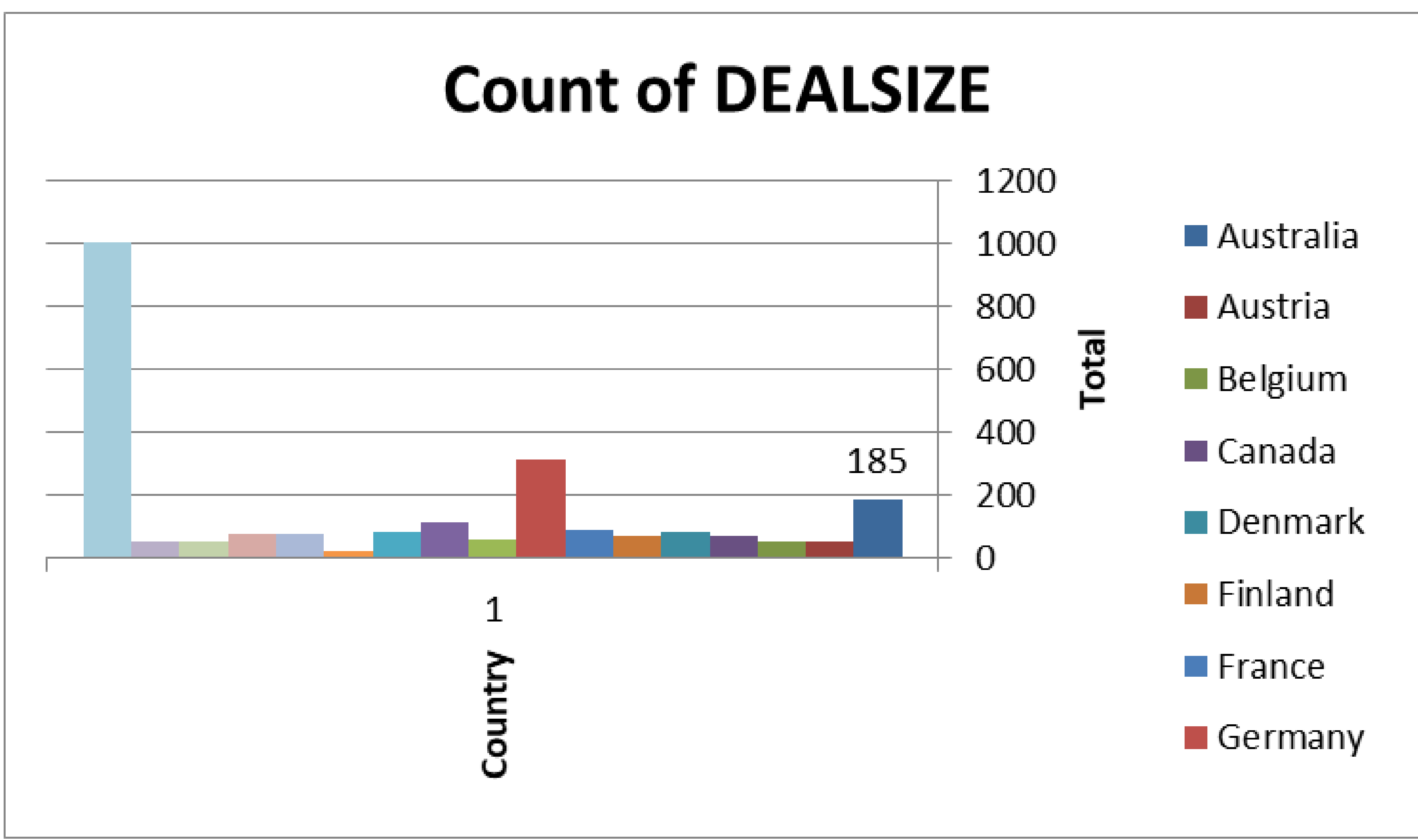
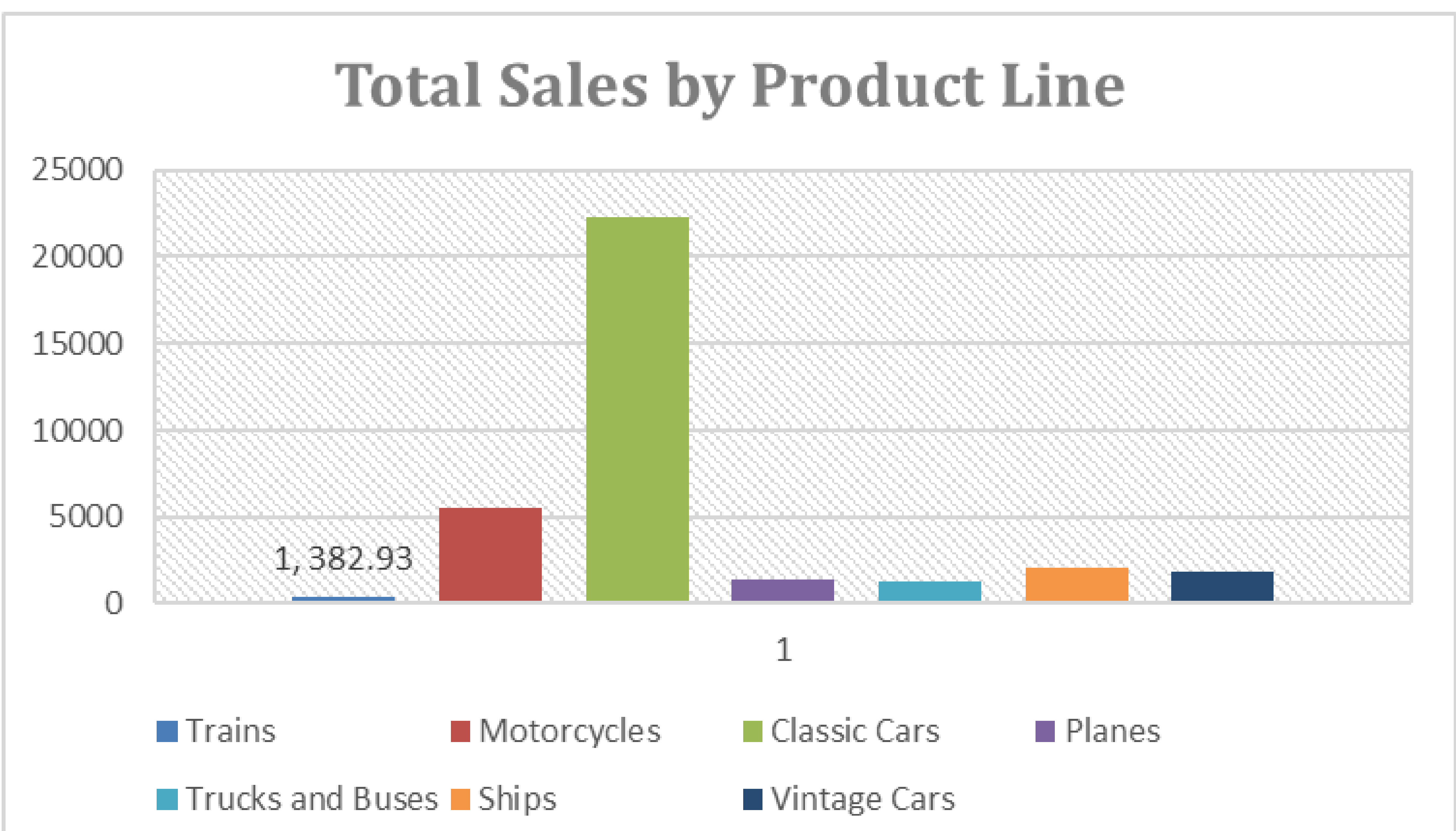
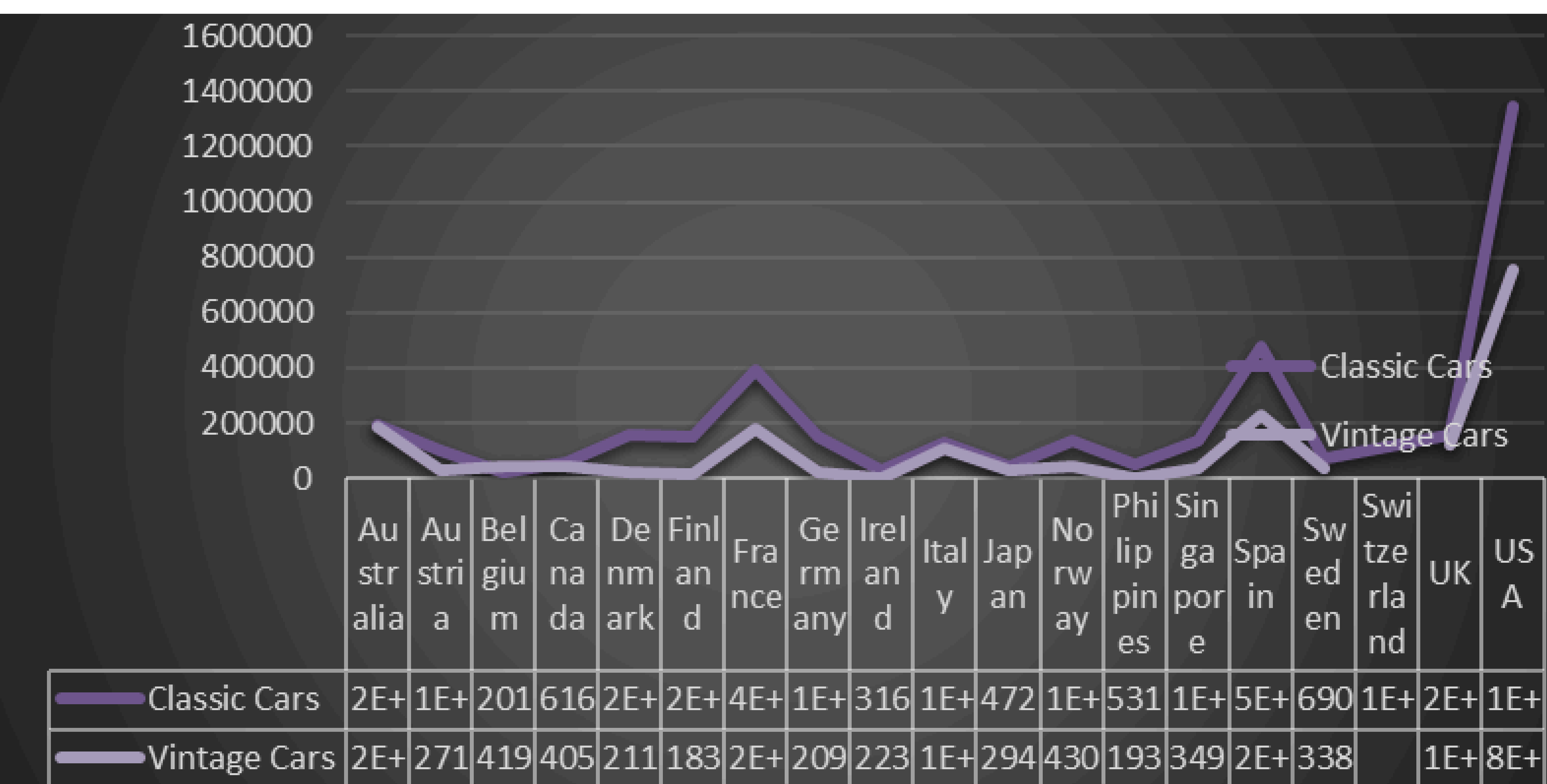


# Dashboard of shop sales

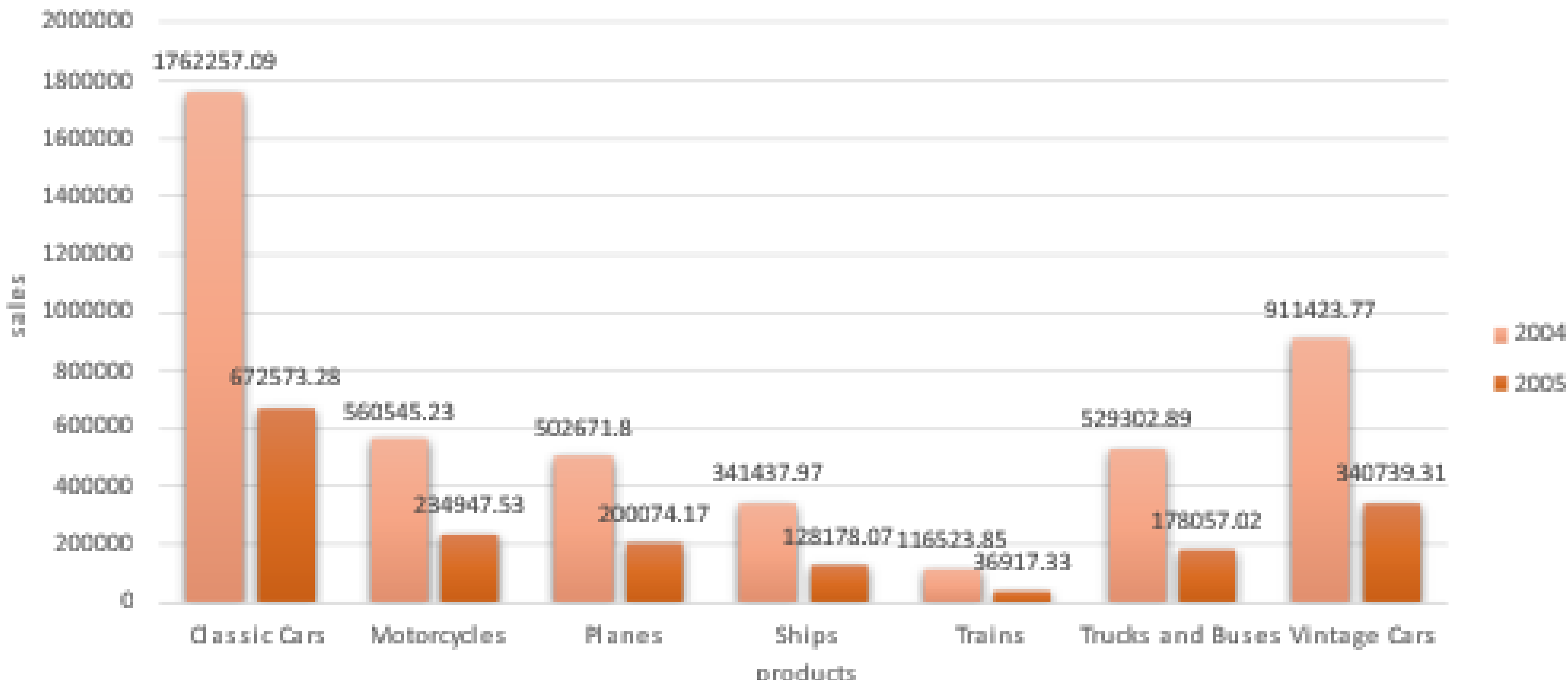




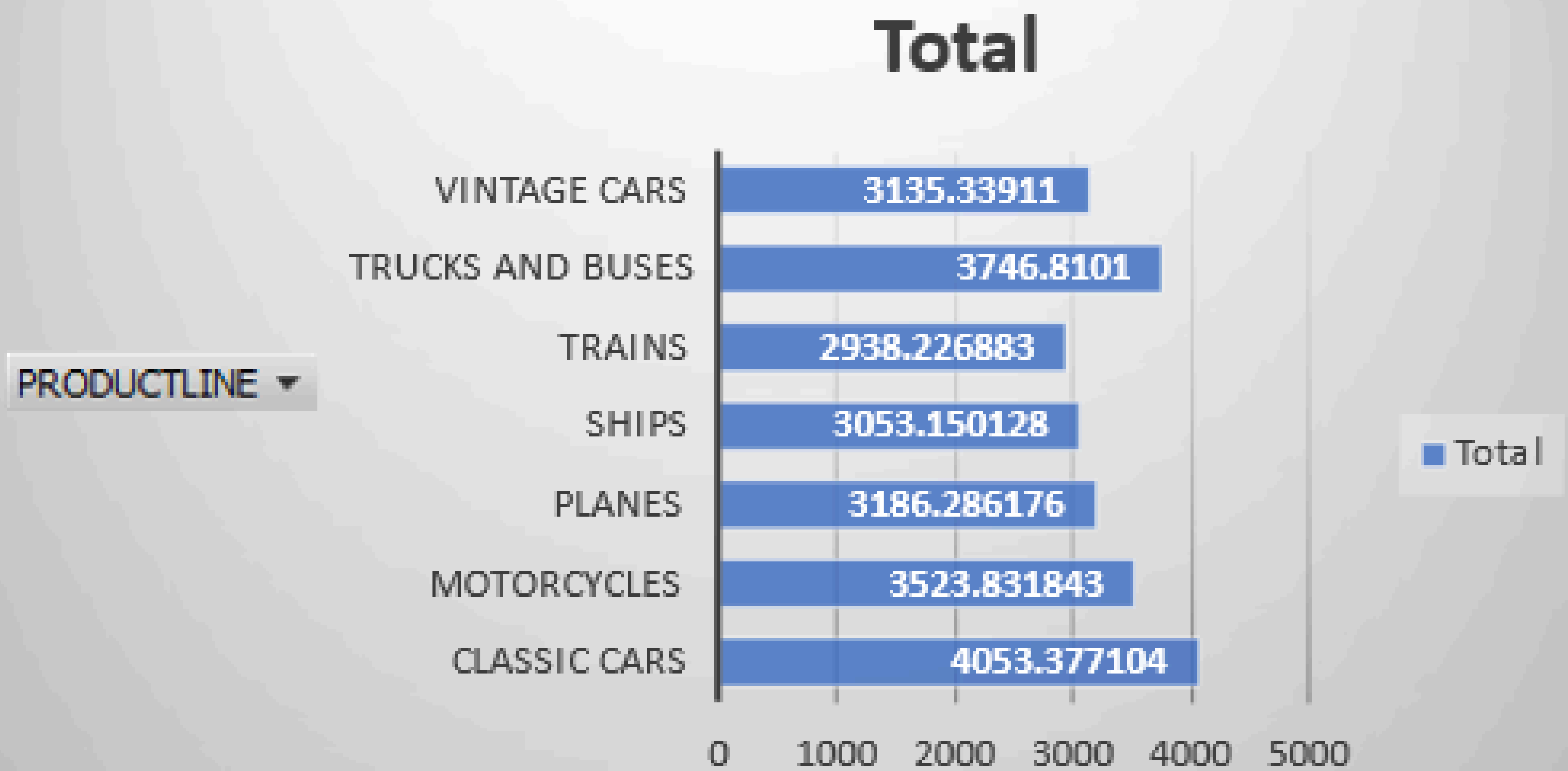
# Dashboard of Sales Data



Sales from 2004&2005



Average of SALES



# DASHBOARD FOR SUPER MARKET DATA ANALYSIS





# Report on Car Collection Dataset

## Introduction –

In this report, we delve into a comprehensive analysis of the Car Collection dataset, which encompasses a variety of vehicles spanning several makes, models, colors, and economic factors such as mileage, price, and cost. The dataset provides a snapshot of key attributes that influence consumer preferences and economic decisions within the automotive market.

The primary objective of this analysis is to derive actionable insights that could significantly impact stakeholders ranging from automotive dealers and manufacturers to individual consumers and automotive industry analysts. By comparing different car brands like Chevrolet, Toyota, Ford, and Honda, and assessing attributes like mileage and cost effectiveness, this report aims to outline which cars offer the best value for money and which models stand out in specific categories such as fuel efficiency and popular aesthetic choices (e.g., color).

## The intended audience for this report includes:

**Car Dealerships and Retailers:** To better understand market trends and consumer preferences to optimize their inventory and marketing strategies.

**Potential Car Buyers:** To provide a clear, data-driven insight into which cars offer the best value, considering various metrics such as mileage, cost, and overall price.

**Automotive Industry Analysts:** To gain an in-depth view of the competitive landscape within the car market, which can help in forecasting trends and advising clients in the automotive sector.

By exploring these dimensions, the report not only aids in making informed purchasing decisions but also assists manufacturers and dealerships in aligning their strategies with consumer trends. The scope of this project covers a detailed analysis of car performance in economic terms, popularity trends among consumers, and a comparative assessment that highlights the best options available in the car market today.

## Questionnaire –

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?

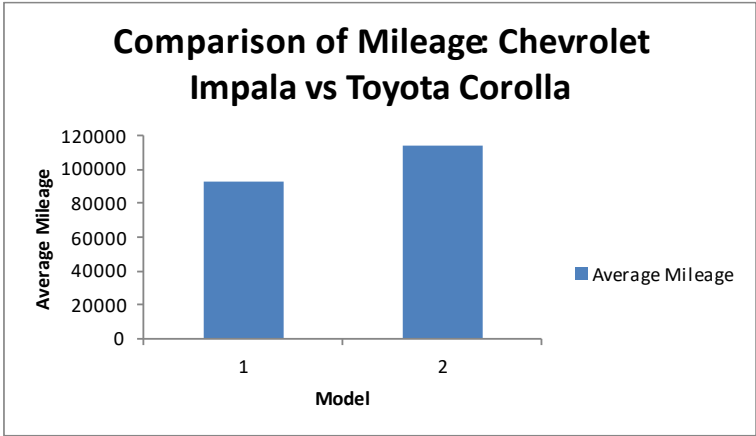
## Analytics –

1. Compare the mileage of Chevrolet Impala to Toyota Corolla. Which of the two is giving best mileage?

Answer - The average mileage of a Chevrolet Impala is 114,243 miles.

The average mileage of a Toyota Corolla is 92,377 miles.

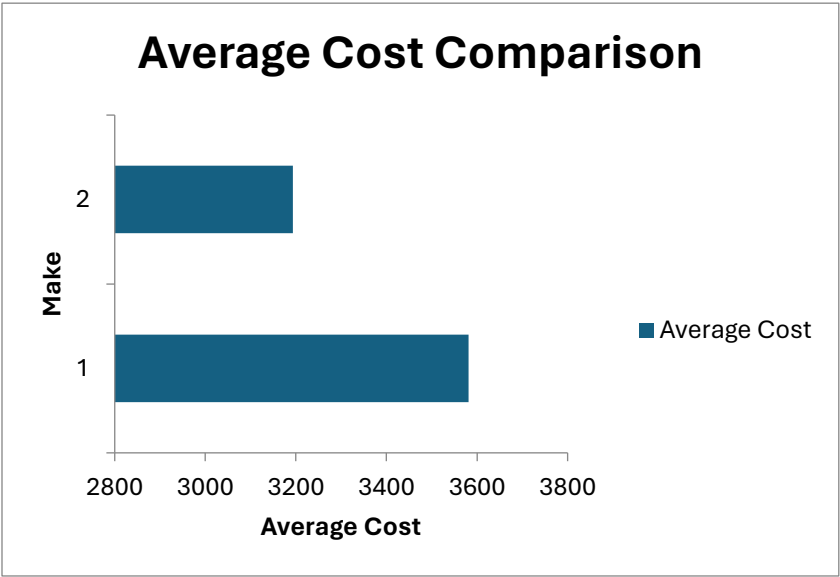
Based on the average mileage, the Chevrolet Impala provides better mileage than the Toyota Corolla.



Row Labels	Sum of Mileage
Corolla	92377
Impala	114243
Grand Total	206620

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

Answer - Ford cars are more expensive both in terms of cost and price compared to Honda cars, which might indicate better features or quality, depending on the context (such as additional options, better performance or durability). However, whether buying a Ford is "better" could depend on specific needs and preferences including factors like reliability, maintenance costs, resale value, etc.

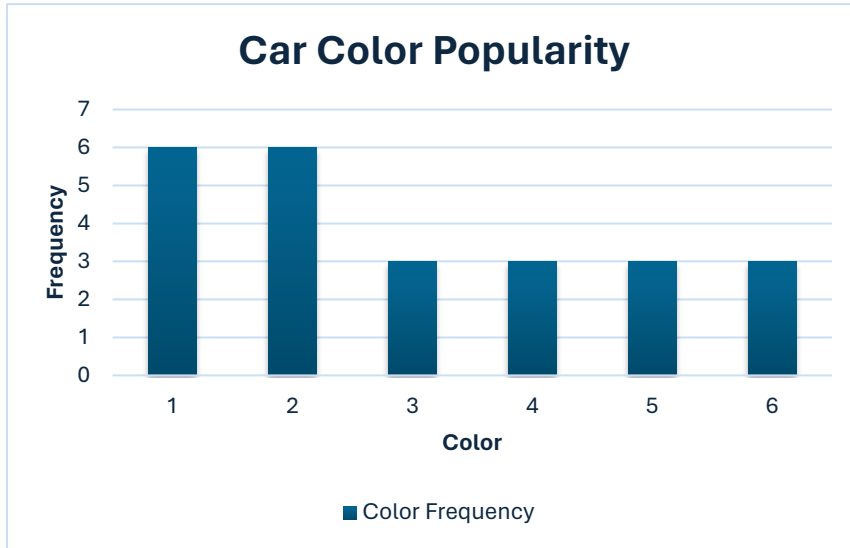


Make	Cost	Price
Ford	3030	3581.6
Honda	2500	3193.6
Sum	5530	6775.2

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

Answer - The most popular car color is Silver, with 6 cars listed in that color.

The least popular car color is White, with only 3 cars listed in that color.



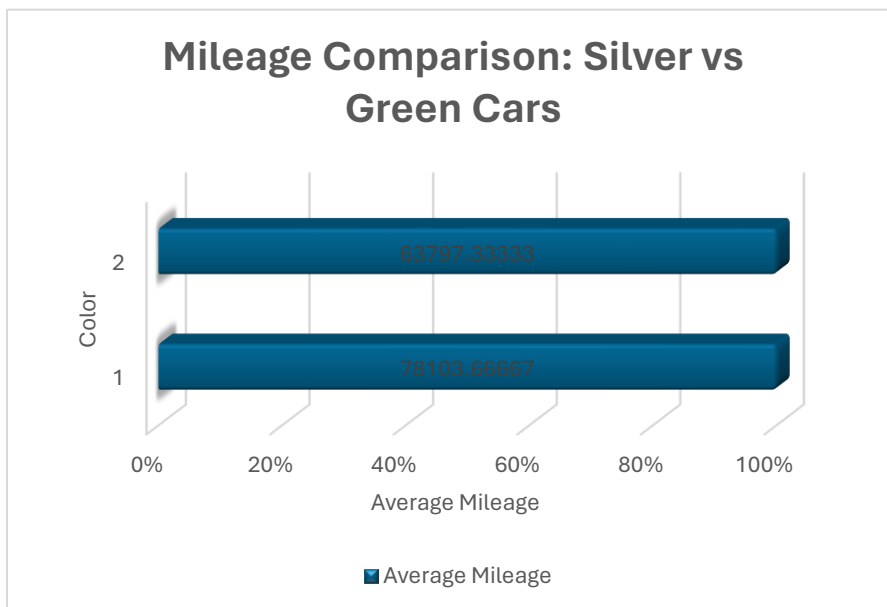
Color	Frequency
Silver	6
Black	6
Red	3
Blue	3
Green	3
White	3

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

Answer - The average mileage for cars that are Green is 78,104 miles.

The average mileage for cars that are Silver is 63,797 miles.

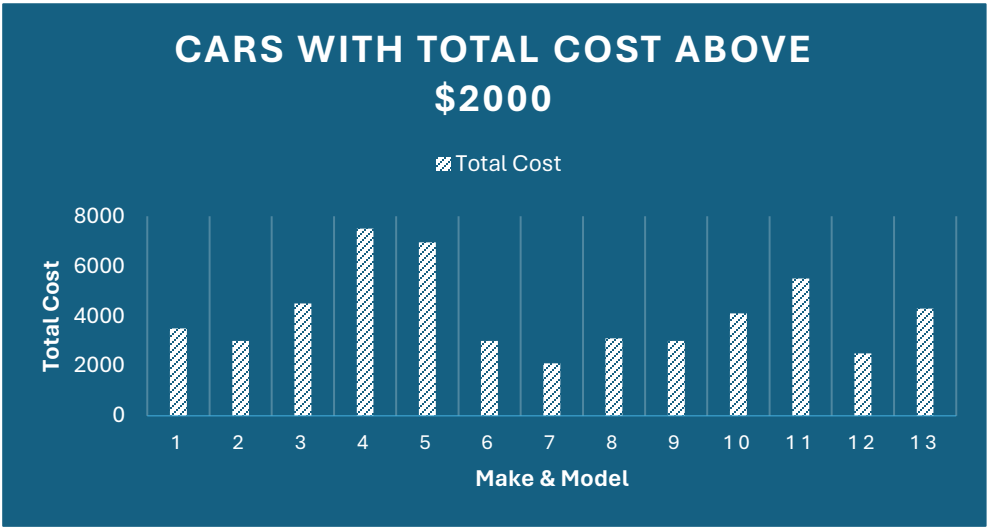
Based on the average mileage, cars that are Green have higher mileage than those that are Silver.



Color	Mileage
Green	78103.6667
Silver	63797.3333
Sum	141901

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

Answer - These results cover all the cars in your dataset with costs above \$2,000, providing a clear view of the more expensive options in your collection.



Make	Model	Cost
Chevrolet	Impala	3500
Chevrolet	Malibu	3000
Chevrolet	Silverado	4500
Dodge	Charger	7500
Ford	Escape	6950
Ford	F-150	3000
Ford	Fusion	2100
Ford	Mustang	3100
Honda	Accord	3000
Honda	CRV	4100
Nissan	Altima	5500
Nissan	Maxima	2500
Toyota	Corolla	4300
Sum	0	53050

Make
Chevrolet
Dodge
Ford
Honda
Nissan
Sum
Toyota

Cost
2100
2500
3000
3100
3500
4100
4300
4500

Model
F-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.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.40404555
R Square	0.1632528
Adjusted R Square	0.1234077
Standard Error	33099.5397
Observations	23

ANOVA					
	df	SS	MS	F	Significance F
Regression	1	4488793099	4488793099	4.09718598	0.05586127
Residual	21	2.3007E+10	1095579531		
Total	22	2.7496E+10			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	130438.919	23634.1932	5.51907645	1.7789E-05	81288.9236	179588.914	81288.9236	179588.914
3000	-16.664135	8.23265547	-2.0241507	0.05586127	-33.784879	0.45660911	-33.784879	0.45660911

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.

	<i>Mileage</i>	<i>Price</i>
Mileage	1	
Price	-0.4110586	1

Anova: Single Factor

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

<i>Groups</i>	<i>Count</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>
Mileage	24	2011267	83802.7917	1214155660
Price	24	78108	3254.5	837024.087
Cost	24	66150	2756.25	705502.717

ANOVA						
<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
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							
<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>	
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.

<i>Mileage</i>		<i>Price</i>		<i>Cost</i>	
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

## Conclusion/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.



# Report on Cookie Data

## Introduction :

In our dataset, we analyze six distinct types of cookies: Chocolate Chip, Fortune Cookie, Sugar, Oatmeal Raisin, Snickerdoodle, and White Chocolate Macadamia Nut. This dataset provides comprehensive insights including unit sales, costs, revenue, and profit margins across different countries and time periods.

The purpose of this report goes beyond merely presenting numbers; it aims to delve into consumer preferences, pricing strategies, and geographical popularity. By exploring these factors, we provide valuable information that can help businesses understand market dynamics and customer behavior in the cookie industry. Get ready to dive deep into the data and discover key trends that could influence your business strategies.

## Questionaries :

- 1 . Compare Malaysia and Philippines on the bases of two types of Cookies
2. What is the performance of Choco Chips Cookies in all Country Which Competes the best.
3. Compare all the countries on the bases of profit and unit sold, which is the best performance country on the basis of profit.
4. which Cookie is the best Selling Cookie in India and US in year 2019,

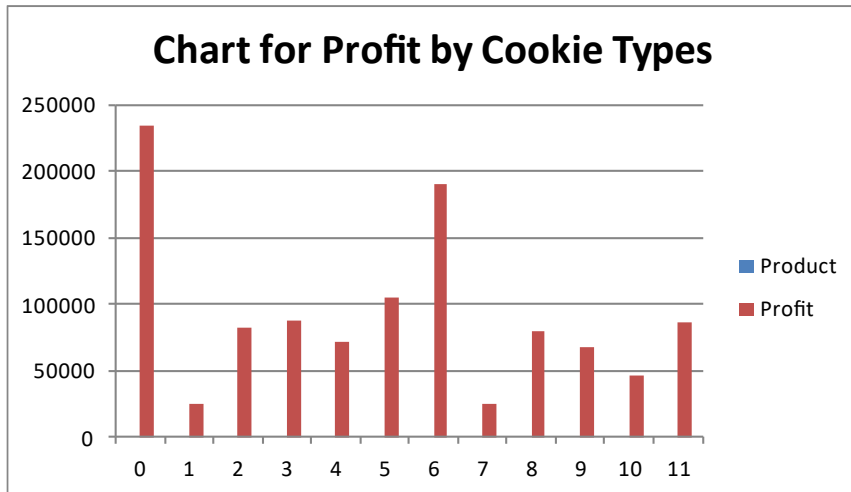
## Analytics :

- 1 . Compare Malaysia and Philippines on the bases of two types of Cookies.

Ans:-The comparsion of Malaysia and Philippines on bases of Chocolate chip and White Chocolate Macadmia nut is given below:-

Profit	Product
25192	Chocolate Chip
25403.2	Fortune Cookie
45725.75	Oatmeal Raisin
68060	Snickerdoodle
72187.5	Sugar
79508.8	White Chocolate Maca...
81852.4	
86255	

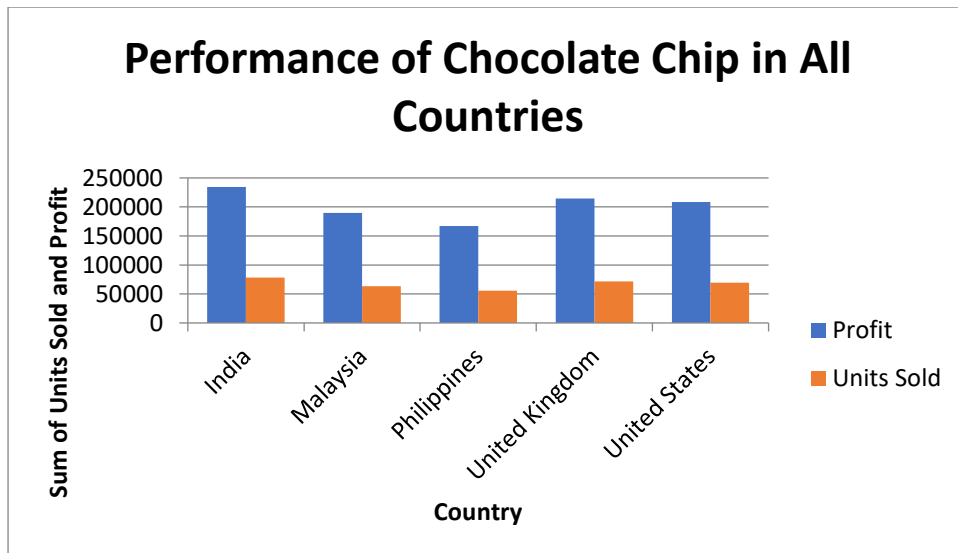
Country	Product	Profit
India	Chocolate Chip	234576
	Fortune Cookie	25192
	Oatmeal Raisin	81852.4
	Snickerdoodle	87012.5
	Sugar	72187.5
Malaysia	White Chocolate Macadamia Nut	105511.3
	Chocolate Chip	189846
	Fortune Cookie	25403.2
	Oatmeal Raisin	79508.8
	Snickerdoodle	68060
	Sugar	45725.75
	White Chocolate Macadamia Nut	86255



2. What is the performance of Choco Chips Cookies in all Country Which Competes the best.

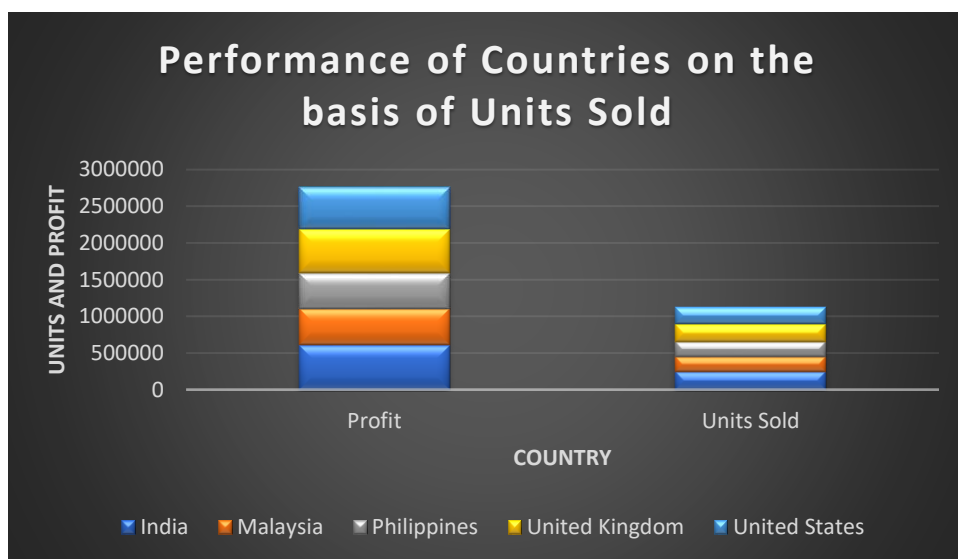
Ans:- India stands out as the foremost consumer of Choco chips worldwide, primarily due to its exceptional profitability and record-breaking sales figures. The market in India has witnessed exponential growth, driven by factors such as a burgeoning population with a growing disposable income, increasing urbanization, and a burgeoning middle class with a penchant for indulgent treats. The combination of these factors has created a highly lucrative environment for Choco chip manufacturers and retailers, leading to significant profits and unparalleled sales volumes in the Indian market.

Country	Profit	Units Sold
India	234576	78192
Malaysia	189846	63282
Philippines	167082	55694
United Kingdom	214821	71607
United States	208404	69468
Sum	1014729	338243



3. Compare all the countries on the bases of profit and unit sold, which is the best performance country on the basis of profit.

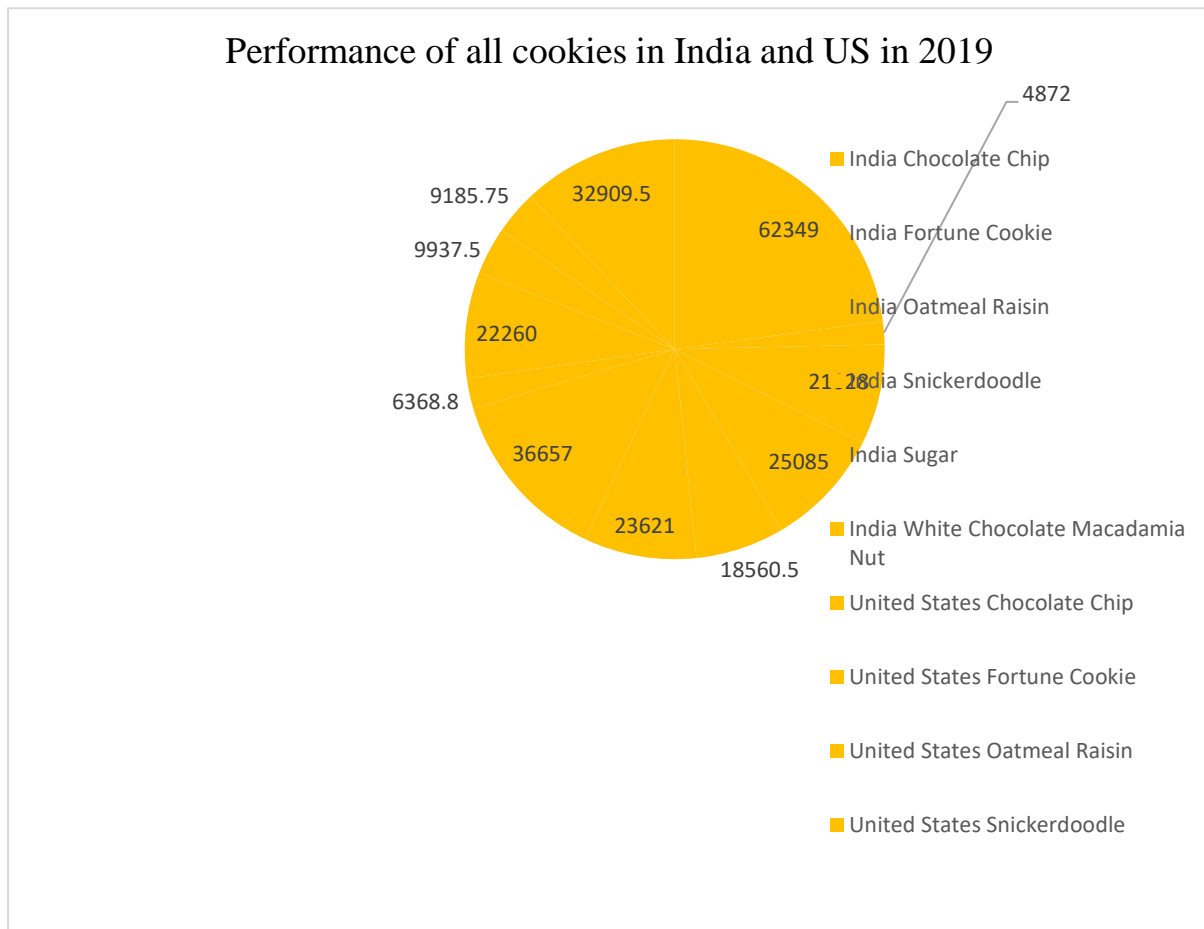
Ans:- India stands out as the leading performer globally when it comes to both profit generation and units sold in the Choco chip market.



Country	Profit	Units Sold
India	606331.7	247435
Malaysia	494798.8	203325
Philippines	493004.6	201495
United Kingdom	595071.6	240937
United States	574157.9	232632
Sum	2763364	1125824

4 .which Cookie is the best Selling Cookie in India and US in year 2019,

Ans:- In the year 2019, chocolate chip cookies emerged as the top-selling cookie in both India and the United States.



## Conclusion and Review :

After thorough analysis of the cookie sales data, it is evident that there are notable trends and insights to be gleaned. By examining key metrics such as units sold, revenue, cost, and profit across different countries and products, we can draw valuable conclusions about market demand, pricing strategies, and overall profitability. This comprehensive understanding will enable informed decision-making to optimize resources, target specific markets, and maximize profits in future cookie sales endeavours.

## 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

<i>Regression Statistics</i>	
Multiple R	0.829304
R Square	0.687746
Adjusted R Square	0.687298
Standard Error	1462.76
Observations	700

ANOVA					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	1	3.29E+09	3.29E+09	1537.356	1.4E-178
Residual	698	1.49E+09	2139668		
Total	699	4.78E+09			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>
Intercept	-74.4103	116.5304	-0.63855	0.523326	-303.202	154.3817	-303.202	154.3817
Units Sold	2.500792	0.063781	39.20914	1.4E-178	2.375567	2.626017	2.375567	2.626017

## Co-relation:

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

	<i>Units Sold</i>	<i>Revenue</i>
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				
<i>Groups</i>	<i>Count</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>

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 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%)	

# Report on Loan Data

## 1. INTRODUCTION:

Our dataset offers a detailed look at the variables influencing loan applications. It includes basic applicant information like Gender, Marital Status, Education level, as well as other critical factors like Employment Status, the amount of loan requested, and the type of area the applicant resides in.

1. Gender: This data point helps us understand the breakdown of applicants by gender.
2. Marital Status (Married, Not Married): This helps us differentiate applicants based on whether they are married, which can be valuable for demographic studies.
3. Education (Graduate, Non-graduate): This tells us the highest education level achieved by applicants, which can be a significant factor in financial decision-making.
4. Employment Status (Employed, Unemployed): This important distinction categorizes applicants by their job situation, a factor that's often considered in loan approval processes.
5. Loan Amount: This is the amount of money the applicant is seeking to borrow and can give us insights into their financial needs and repayment capability.
6. Residential Type (Urban, Semi-urban, Rural): The living environment of the applicant, which can affect loan conditions and availability.

Each of these attributes helps paint a fuller picture of loan application trends and behaviors.

## 2. 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, Semi-urban and rural on the basis of amount.

## 3. ANALYTICS:

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

**Gender**

Female

Male

(blank)

**Married**

No

Yes

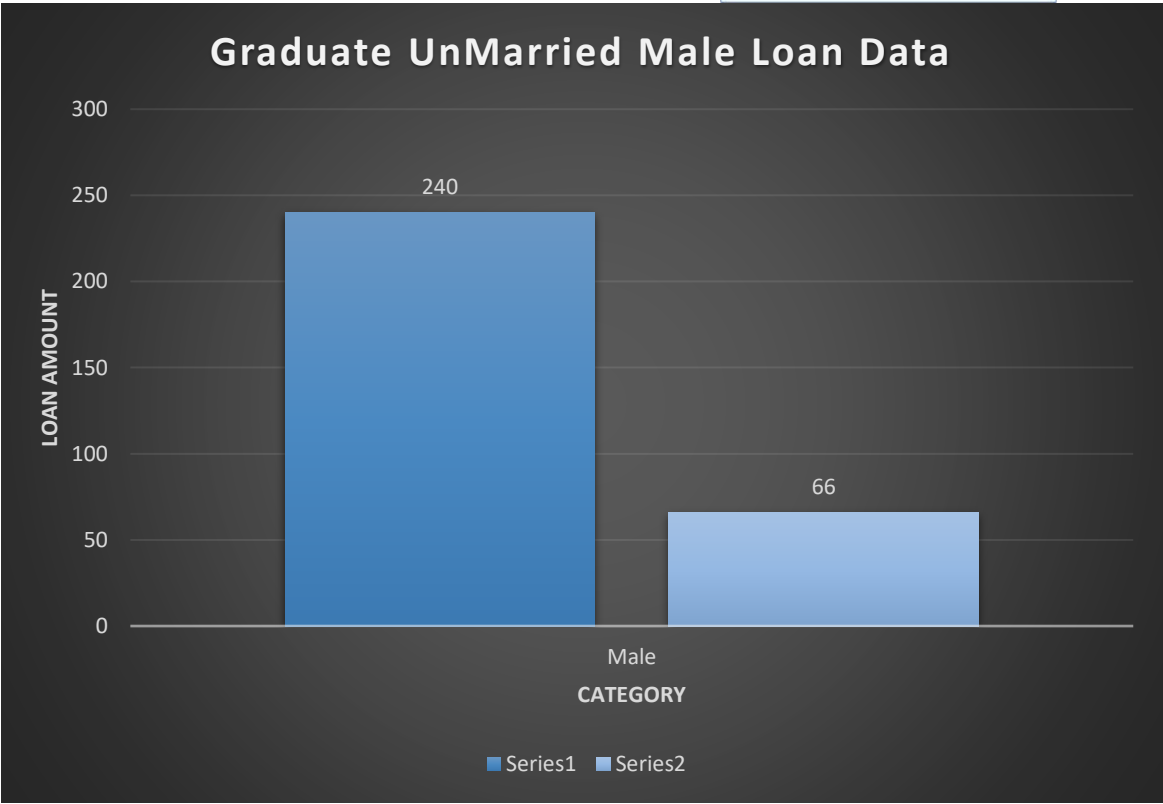
(blank)

**Education**

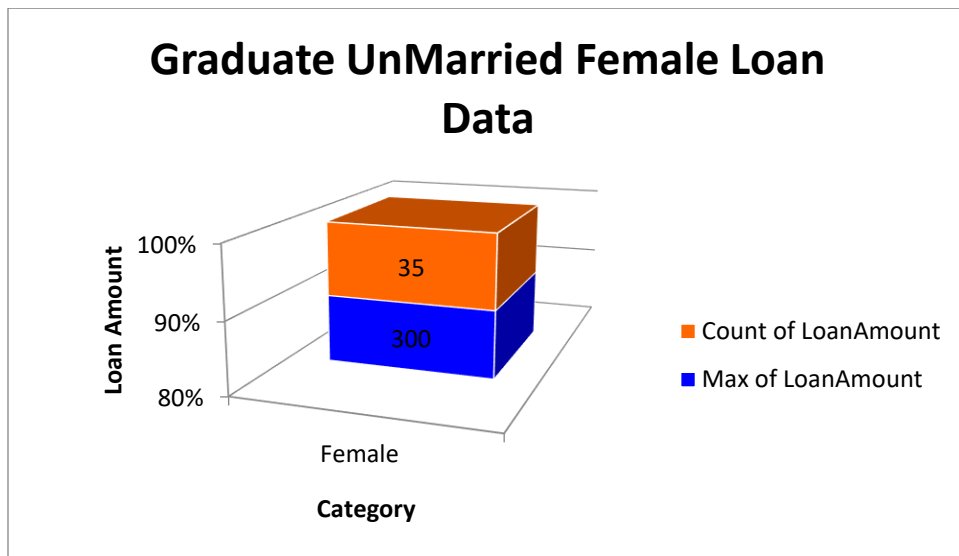
Graduate

Not Graduate

(blank)



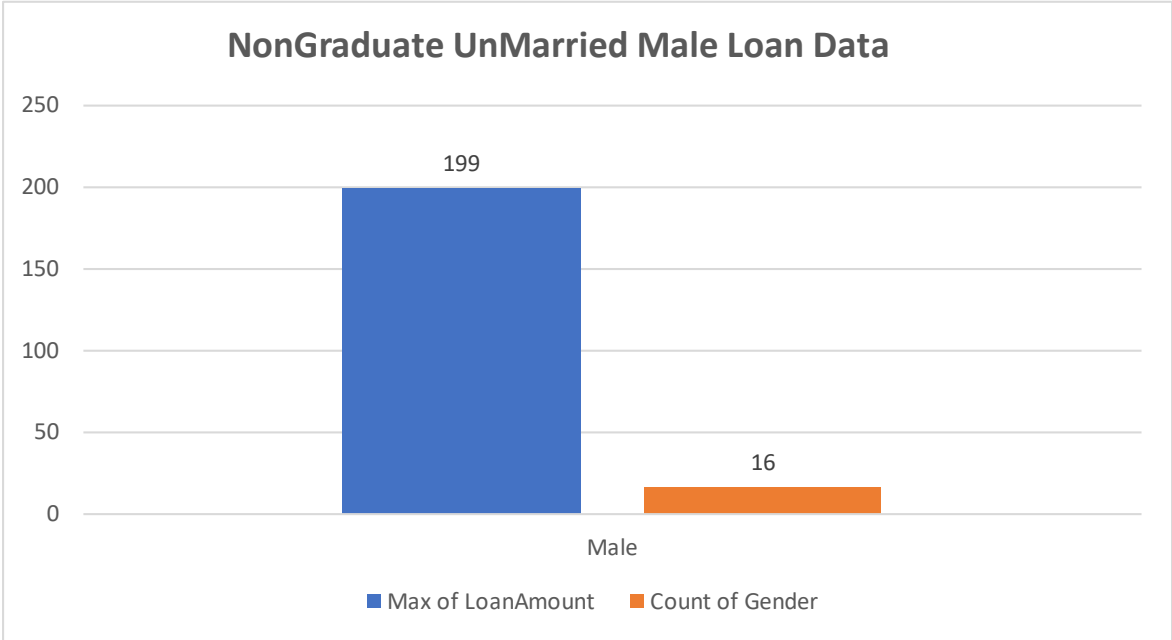
Q2. How many female graduates who are not married applied for Loan? What was the highest amount?



Category	Max of LoanAmount	Count of LoanAmount
Male	240	66

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





**Gender** ✓≡ ⚙️

Male

(blank)

**Married** ✓≡ ⚙️

No

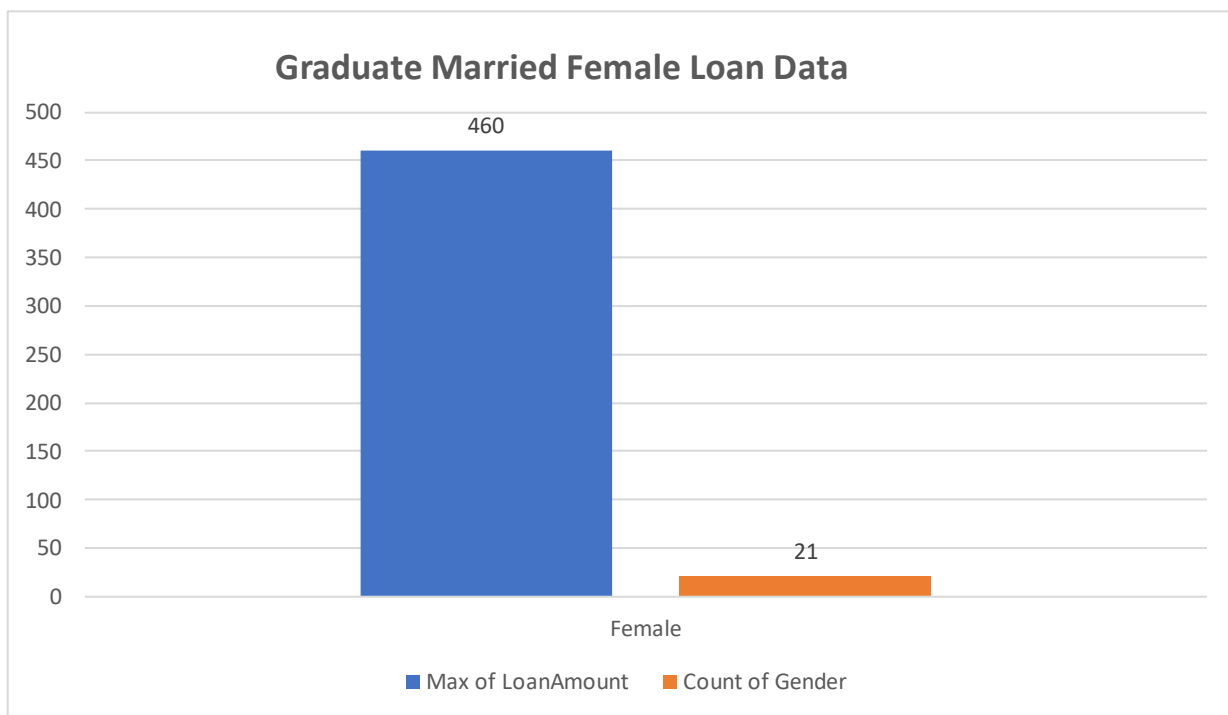
(blank)

**Education** ✓≡ ⚙️

Not Graduate

(blank)

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



**Gender** ✓☰🗑️

Female

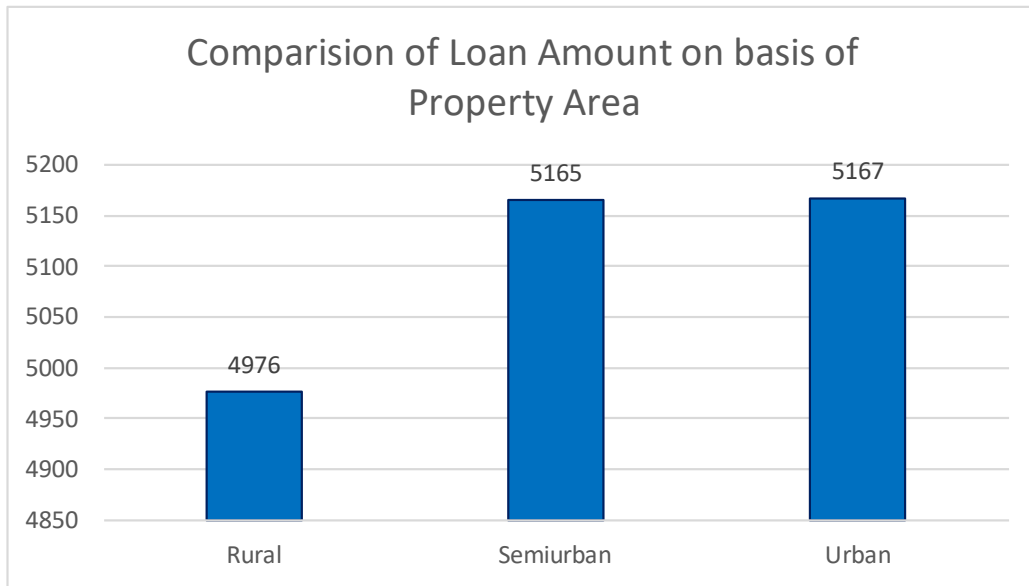
**Married** ✓☰🗑️

Yes

**Education** ✓☰🗑️

Graduate

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



**Property\_Area** ✓☰✗

Rural

Semiurban

Urban

(blank)

**LoanAmount** ✓☰✗

28

40

46

50

55

59

60

64

65

#### 4. CONCLUSION:

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.

## 5. 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

<i>Regression Statistics</i>	
Multiple R	0.45908096
R Square	0.21075532
Adjusted R Square	0.20858707
Standard Error	56.0766111
Observations	366

### ANOVA

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	1	305655.205	305655.205	97.2004502	1.7676E-20
Residual	364	1144629.42	3144.58631		
Total	365	1450284.62			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>
Intercept	106.07753	4.10024098	25.8710478	1.7585E-84	98.014396	114.140665	98.014396
5720	0.0058851	0.00059692	9.85902887	1.7676E-20	0.00471125	0.00705895	0.00471125

## 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).

	<i>ApplicantIncome</i>	<i>CoapplicantIncome</i>	<i>LoanAmount</i>
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

<i>Groups</i>	<i>Count</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>
ApplicantIncome	367	176365	4805.59945	24114831.0
CoapplicantIncome	367	576035	1569.57765	5448639.49
LoanAmount	367	49280	134.277929	3964.14112

### ANOVA

<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
Between Groups	4202537452	2	2101268726	213.2009841	5.87569E-79	3.003920577
Within Groups	1082168110	1098	9855811.573			
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

<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
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 Loan-Amount. 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 Loan-Amount 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%)	504.0756067	Confidence Level(95.0%)	239.6059543	Confidence Level(95.0%)	6.46



# Report on Order Data

## Introduction:

Our dataset is rich with varied variables that give us an in-depth view into the sales landscape across different categories. It records everything from basic sales transactions, including dates, times, and amounts, to more specific details like the types of customers, demographic profiles, and product classifications. Every detail is captured with precision.

ID: Each sale is given a unique code which helps in tracking and detailed examination.

City, State: This provides a spatial perspective, useful for comparing regions and spotting location-based trends.

Product Line (such as furniture, electronics, home goods): This helps sort sales data by product type, making it easier to spot which items are selling and where.

Unit Price, Net Sales: These are key financial details that contribute to understanding revenue streams and shaping pricing approaches.

Net Sales by Category and State: This serves as a measure of success, showing which products do well in which locations.

Rating: Reveals customer satisfaction and product performance in various states.

States (including California, Texas, Washington): Offers a breakdown by region, crucial for targeted market analysis and strategy development.

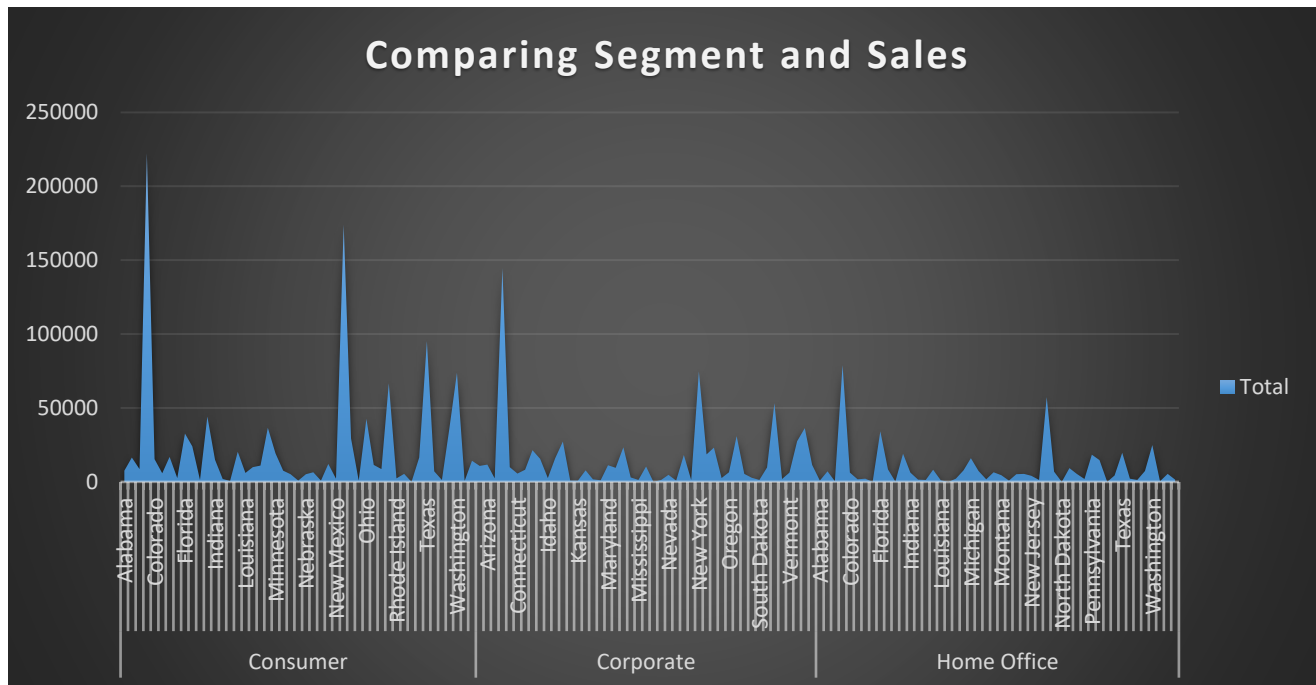
In essence, this dataset functions as a comprehensive tool for dissecting sales patterns and making informed business decisions.

## 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.

## 1. ANALYTICS:

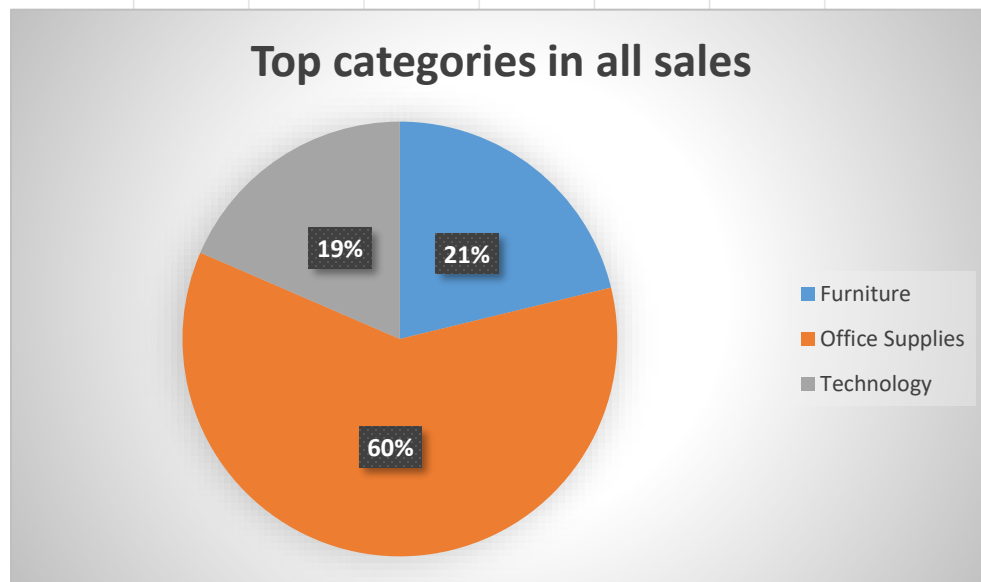
**Q1.** Compare all the US states in terms of Segment and Sales. Which Segment performed well in all the states?



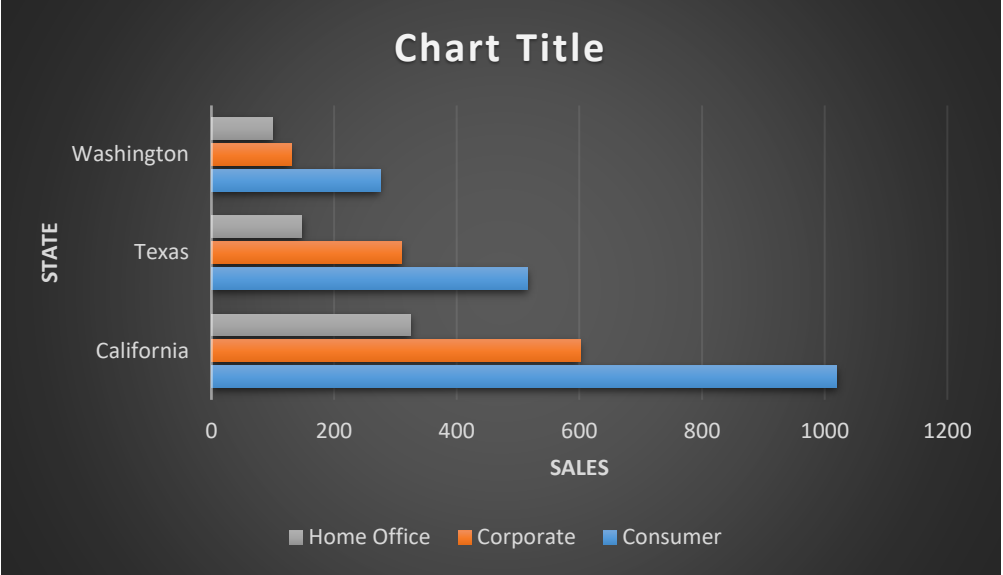
Segment	State	Sales
Consumer	Alabama	0.444
Corporate	Arizona	0.556
Home Office	Arkansas	0.836
	California	0.852
	Colorado	0.876
	Connecticut	0.898
	Delaware	0.984
	District of Columbia	0.99

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

Category	Sales
Furniture	0.444
Office Supplies	0.556
Technology	0.836
(blank)	0.852
	0.876
	0.898
	0.984
	0.99



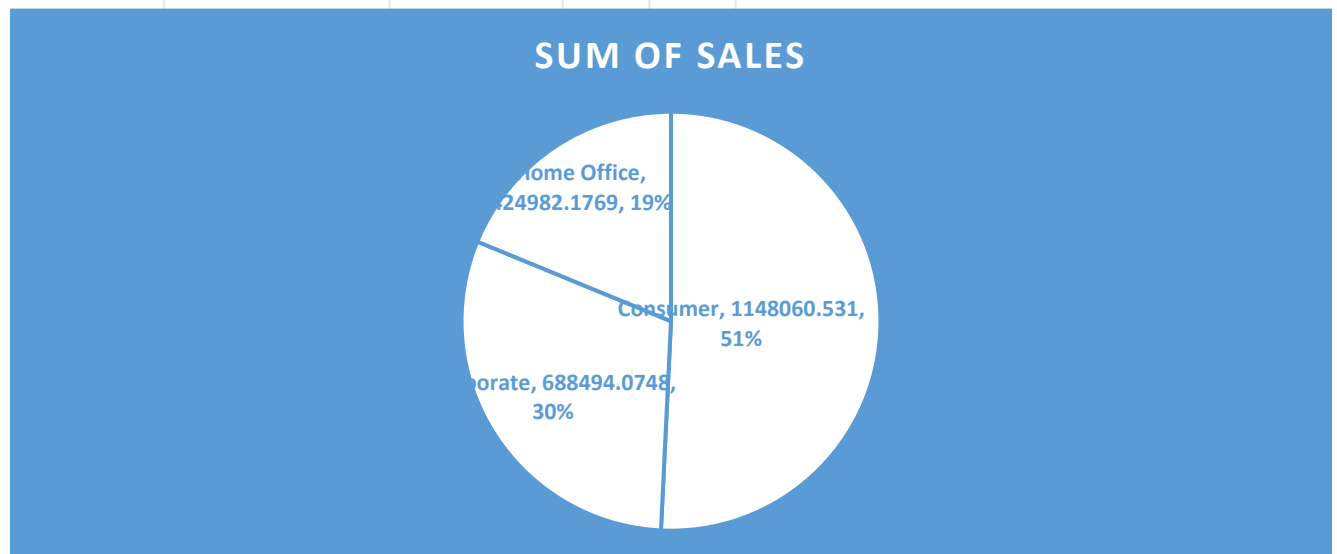
**Q3.** Which segment has most sales in US, California, Texas, and Washington?



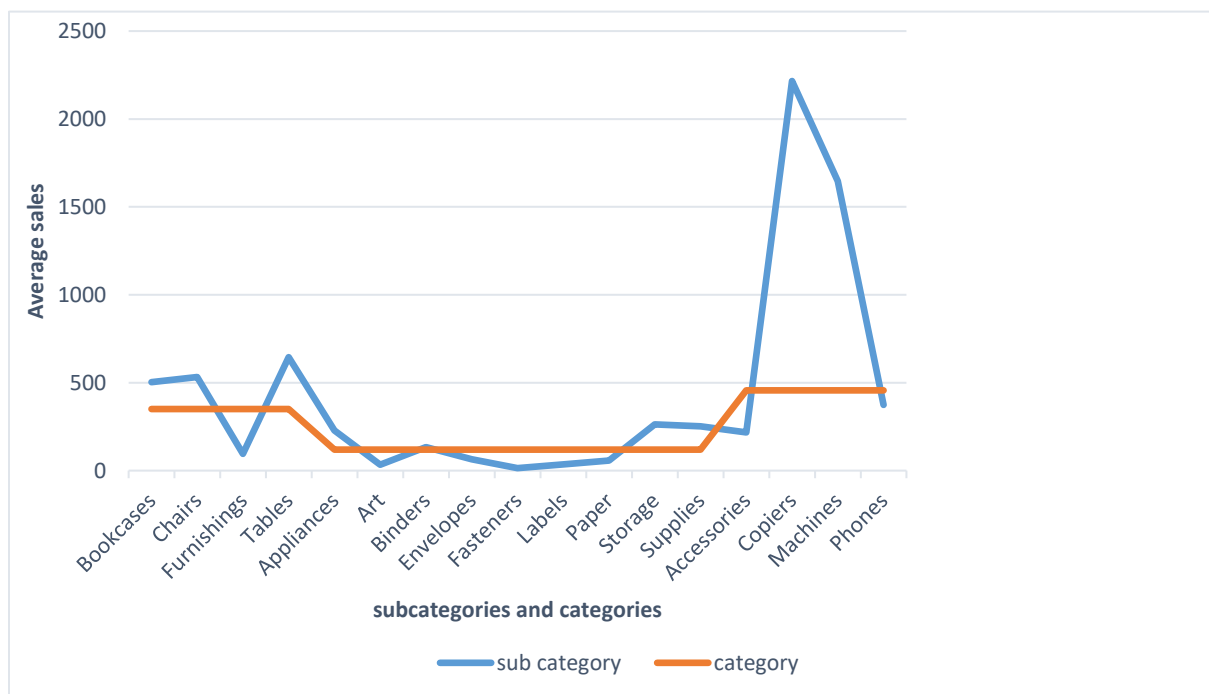
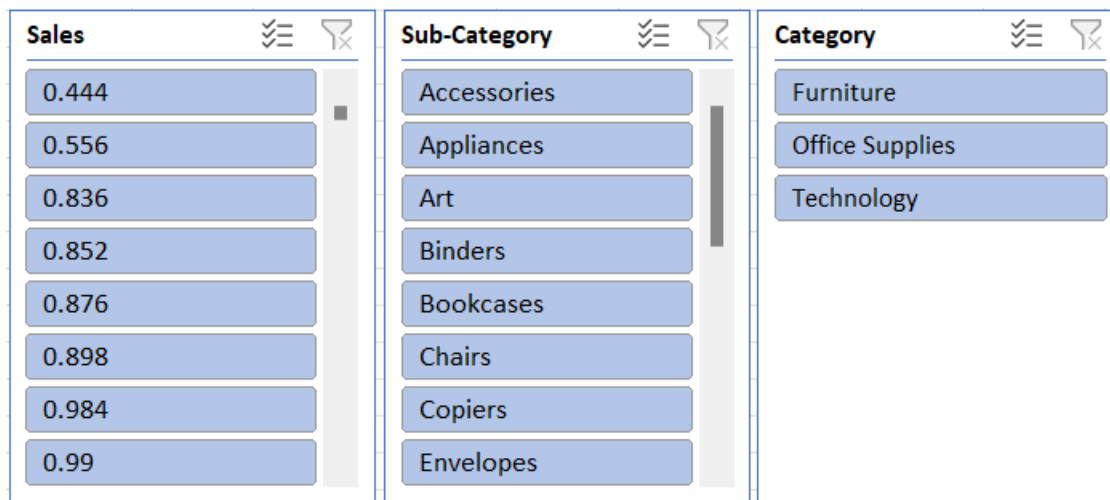
State	Sales
Alabama	0.444
Arizona	0.556
Arkansas	0.836
California	0.852
Colorado	0.876
Connecticut	0.898
Delaware	0.984
District of Columbia	0.99

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

Segment	Sales
Consumer	0.444
Corporate	0.556
Home Office	0.836
blank	0.852
	0.876
	0.898
	0.984
	0.99



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



## 2. CONCLUSION:

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.

Furthermore, this process has underscored the importance of data visualization as a powerful tool for extracting meaningful information from raw data. By harnessing the visual nature of graphs and charts, we've transformed numbers and statistics into compelling narratives that drive understanding and inform decision-making.

# Supermarket Sales Data Report

## Introduction:

## Dataset Overview:

Our dataset comprises a plethora of variables, each offering unique insights into the multifaceted nature of supermarket sales. From fundamental transactional details such as Invoice ID, Date, Time, and Payment Method to more nuanced factors like Branch Location, Customer Type, Gender Demographics, Product Line, and Product Ratings, every facet has been meticulously documented.

## Key Attributes:

1. Invoice ID: A unique identifier for each sales transaction, facilitating traceability and analysis.
2. Branch (A, B, C): The geographical location of the supermarket branch, allowing for regional comparisons and trend identification.
3. Customer Type (Normal, Member): Distinguishing between regular customers and members, offering insight into loyalty and engagement levels.
4. Gender (Male, Female): Demographic segmentation aiding in understanding purchasing preferences and patterns.
5. Product Line (Fashion Accessories, Electronic Accessories, Food and Beverages, Health and Beauty, Home and Lifestyle, Sports and Travel): Categorization of products facilitating analysis of sales trends across different product categories.
6. Unit Price, Quantity, Tax (5%): Fundamental transactional details crucial for revenue assessment and pricing strategies.
7. Payment Method (Credit Card, Cash, E-wallet): Reflecting evolving payment preferences and trends in consumer behavior.
8. Gross Margin Percentage, Gross Income, COGS: Performance metrics illuminating profitability and operational efficiency.
9. Rating (1 to 10): Customer feedback providing a qualitative assessment of product satisfaction and service quality.
10. City (Yangon, Mandalay, Naypyitaw): Regional segmentation enabling geographical analysis and market segmentation.

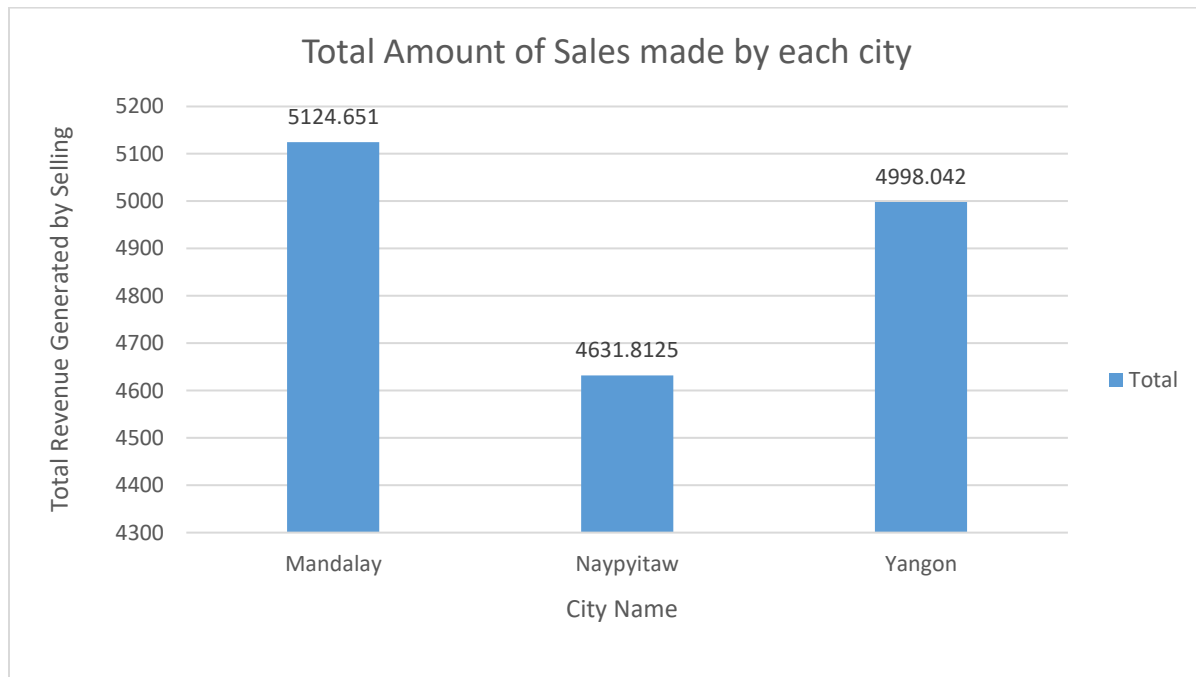


## Questionnaire:

- Q1. Which of the given cities having tax 5% slab performed better than all the others?
- Q2. Which customer gender ordered most items from all the three branches?
- Q3. Compare highest and lowest rating products on the basis of units sold.
- Q4. Analyzing units sold and unit price data answer the following sub questions
- What is the degree of freedom?
  - Co-relation of Unit price and revenue generated
  - What result you can draw from regression of the two data
- Q5. What product will you suggest as per the city data analysis to each type of customer

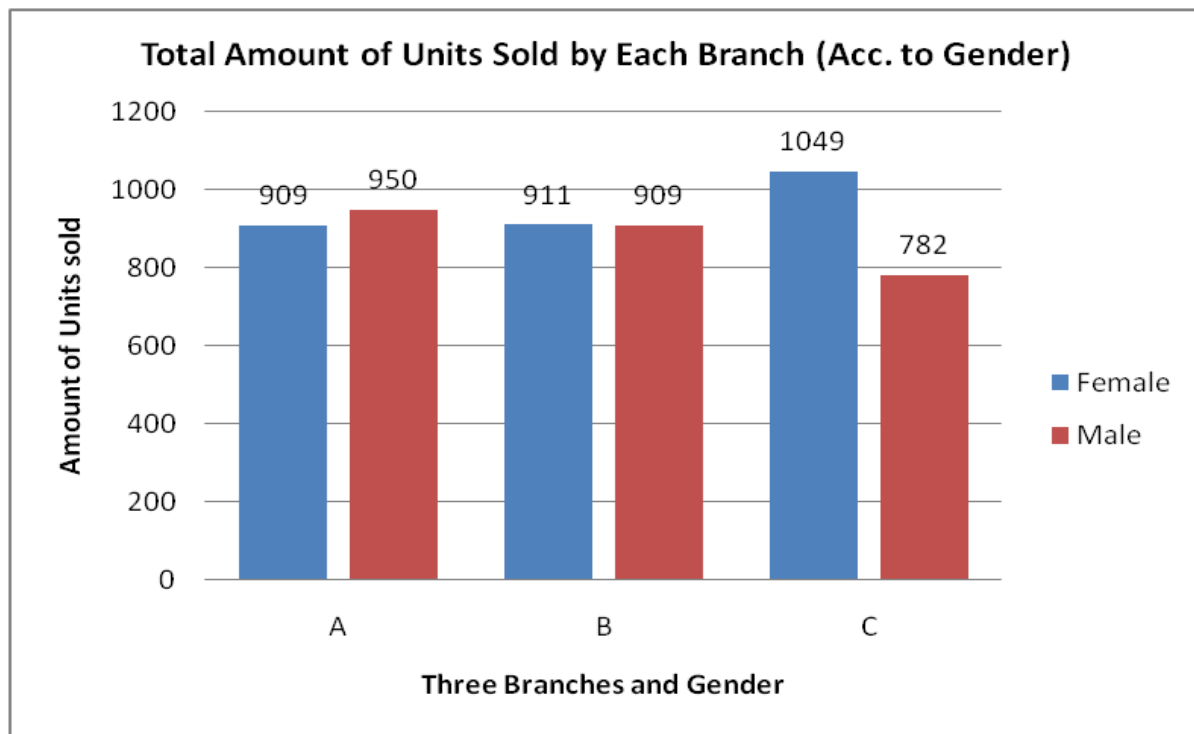
## Analytics:

**Q1.** Which of the given cities having tax 5% slab performed better than all the others?



Based on the data analyzed, the city that outperformed all is **Mandalay**. This conclusion is drawn from superior performance in total sales/revenue generation compared to the other cities in the same tax slab of 5%.

**Q2.** Which customer gender ordered most items from all the three branches?

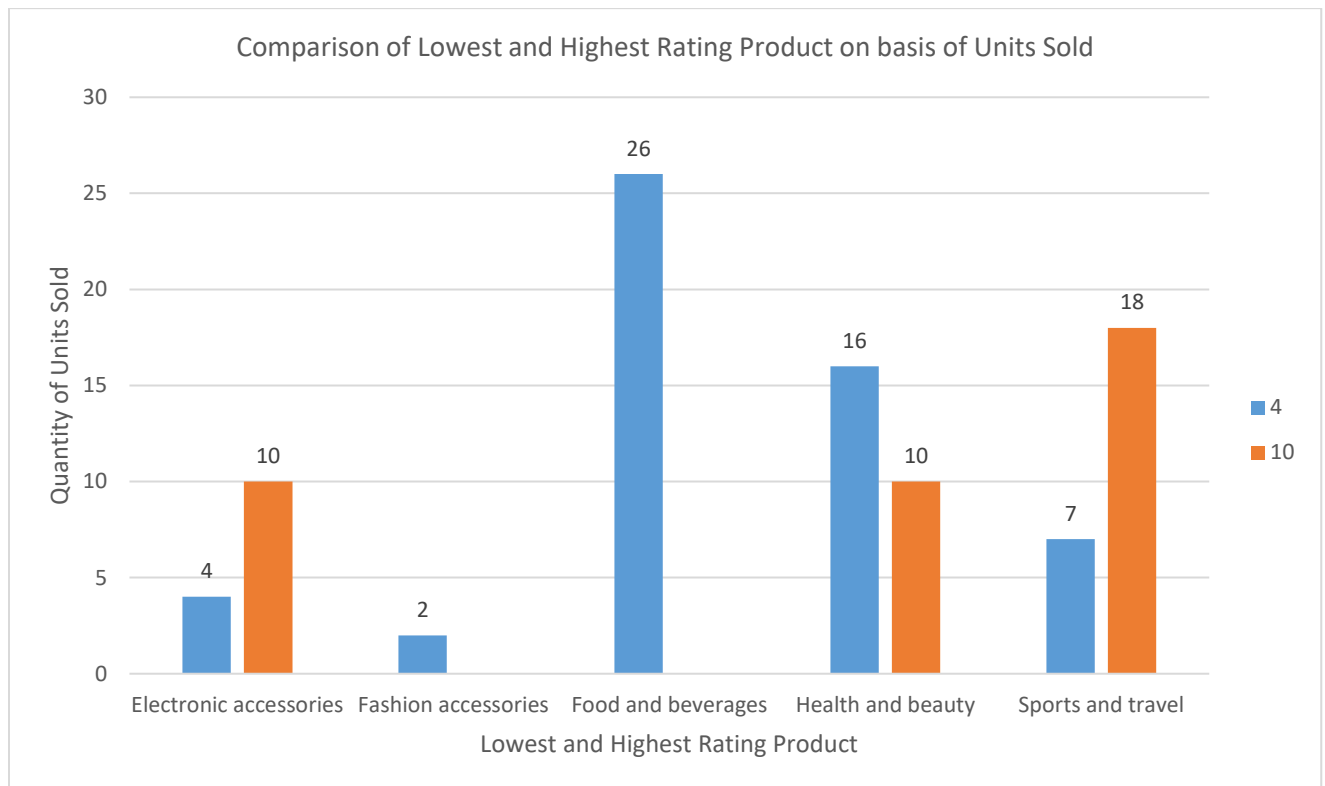


Quantity	Gender	Branch
1	Female	A
2	Male	B
3		C
4		
5		
6		
7		
8		

Answer. Our analysis of the Supermarket Sales Data revealed the following:

- At Branch A, females placed the highest number of orders.
- Branch B saw higher number of orders placed by Females
- Meanwhile, at Branch C, males placed the most orders.

**Q3.** Compare highest and lowest rating products on the basis of units sold.



Answer. Upon analyzing the Supermarket Sales Data, we discovered that product ratings ranged from a minimum of 4 to a maximum of 10.

- Electronic Accessories with higher ratings garnered more customer purchases, indicating a preference for quality in this category.
- Fashion accessories and food and beverages mainly comprised lower-rated products in customer purchases.
- Health and beauty products also leaned towards lower-rated items in customer preferences.
- However, in the Sports and Travel category, customers showed a tendency to purchase higher-rated products.

**Q4.** Analyzing units sold and unit price data answer the following sub questions

- What is the degree of freedom?
- Co-relation of Unit price and revenue generated
- What result you can draw from regression of the two data

SUMMARY OUTPUT						
<i>Regression Statistics</i>						
Multiple R	0.010777564					
R Square	0.000116156					
Adjusted R Square	-0.000885732					
Standard Error	2.924724997					
Observations	1000					
<i>ANOVA</i>						
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>	
Regression	1	0.9917274	0.991727	0.115937	0.733555221	
Residual	998	8536.908273	8.554016			
Total	999	8537.9				
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	5.443794599	0.215314544	25.28299	2.1E-109	5.021273429	5.86631577
Unit price	0.001189202	0.003492565	0.340495	0.733555	-0.005664411	0.008042815

Answer:

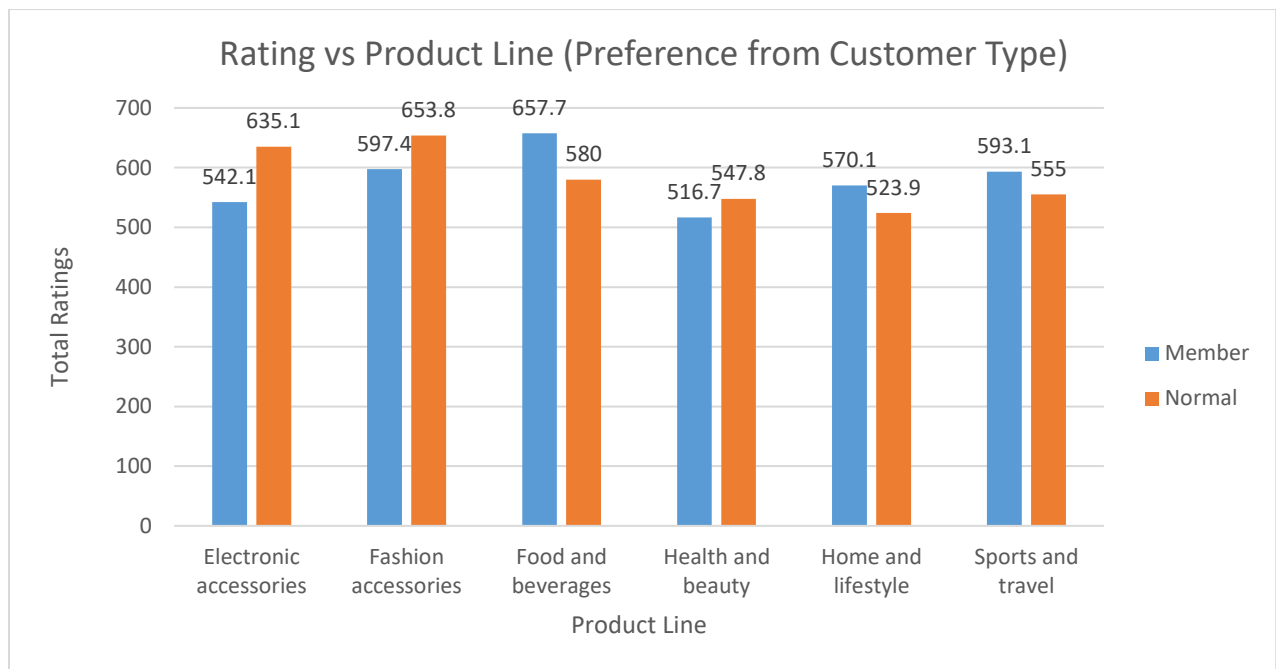
- The degree of freedom of the analyzed data is 1.
- The correlation between unit price and generated revenue was found to be 0.63392, indicating a moderate positive relationship. The analysis focused on the columns of unit price and total revenue, employing the CORREL function.
- Upon examining the regression results, we aimed to discern the relationship between quantity and unit price, exploring how customers' purchasing quantity correlates with the unit price of a product.

However, from the regression analysis, it's evident that the observed trend lacks consistency. The expected outcomes derived from the trend deviate significantly from the actual outcomes.

With a degree of freedom of 1, the trendline equation stands as

Quantity = 0.0012x + 5.4438. Despite this equation, the coefficient of determination (R<sup>2</sup>) is merely 0.0001, highlighting the inconsistency in customer buying patterns solely based on unit price.

**Q5.** What product will you suggest as per the city data analysis to each type of customer



Rating	Customer type	Product line
4	Member	Electronic accessories
4.1	Normal	Fashion accessories
4.2		Food and beverages
4.3		Health and beauty
4.4		Home and lifestyle
4.5		Sports and travel
4.6		
4.7		

Answer. As per the city Data Analysis, **Food and Beverages** will be a good option for **Member** type customer and **Fashion Accessories** for **Normal** type of customers.

# Conclusion and Reviews

The comprehensive analysis of supermarket sales dynamics provides valuable insights into consumer behavior, operational trends, and performance metrics. Here's a summary of the findings and reviews:

## 1. City Performance:

Mandalay emerged as the top-performing city among those with a 5% tax slab. Its superior sales/revenue generation signifies a potentially lucrative market for supermarket businesses.

## 2. Gender-based Ordering:

Female customers showed a higher propensity to order items from Branch A, while males dominated in Branch C. Branch B saw equal orders from both genders. This gender-specific trend highlights the importance of targeted marketing strategies.

## 3. Rating and Units Sold:

Further analysis is needed to compare products with the highest and lowest ratings based on units sold. Understanding the correlation between product ratings and sales volume can inform inventory management and marketing decisions.

## 4. Unit Price and Revenue Relationship:

The regression analysis revealed a weak correlation ( $R^2 = 0.0001$ ) between unit price and quantity sold. This suggests that customers' purchasing decisions may not be significantly influenced by unit price alone, indicating the need for deeper insights into consumer preferences and behavior.

## 5. Product Recommendations:

Based on city data analysis, Food and Beverages are recommended for member-type customers, while Fashion Accessories are suggested for normal customers. These recommendations align with the observed preferences and purchasing patterns in respective cities.

## Reviews:

The report provides a thorough exploration of supermarket sales dynamics, covering various aspects such as city performance, gender-based ordering trends, and product recommendations.

The inclusion of regression analysis enhances the depth of insights, though further interpretation of the results could strengthen the analytical rigor.

Clear visuals, such as graphs and charts, would enhance the presentation of findings and aid in understanding complex relationships.

Overall, the report offers valuable insights for supermarket stakeholders, highlighting areas for strategic focus and improvement in marketing and operational strategies.

# Report on Shop Sales Data

## Introduction

This dataset is a rich repository of sales transaction details, serving as a vital tool for understanding the nuances of retail business activity. It is structured with carefully selected columns that record essential aspects of each sale, such as the transaction date, the sales representative involved, product names, the manufacturer, the volume of goods sold, and the total sale amount.

For analysts and businesses, this dataset is an incredibly useful asset. It allows them to tap into a deep well of information that can be used to detect sales patterns over time, adjust stock levels to better meet consumer demand, and fine-tune sales approaches for better performance.

Essentially, the dataset is not just a passive collection of past sales records—it's a springboard for strategic thinking and decision-making. It empowers businesses to make well-informed choices that can lead to improved operations, higher sales efficiency, and ultimately, growth and expansion in the market. It offers the insights needed to anticipate market shifts, cater to consumer preferences more effectively, and chart a course for future business success..

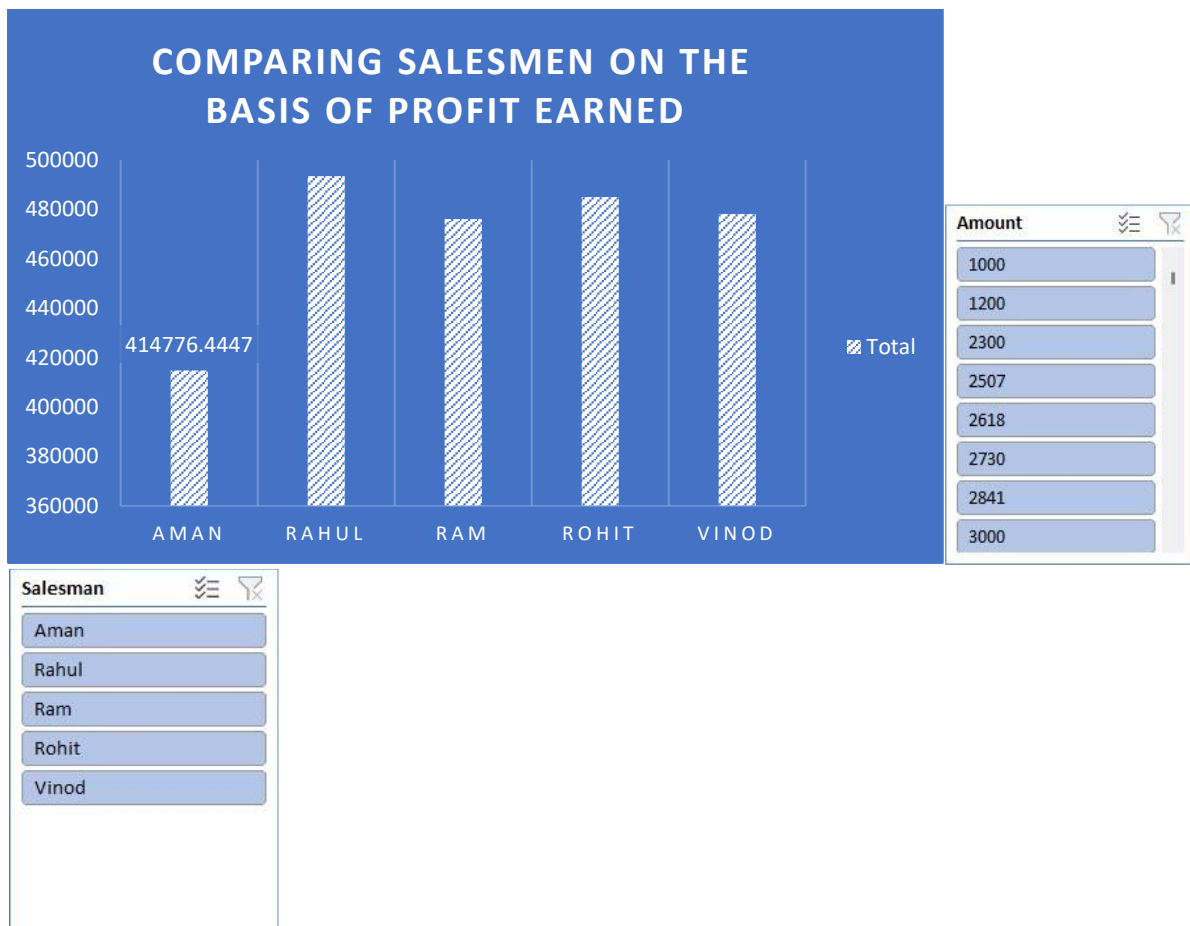
## Questionaries

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:



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

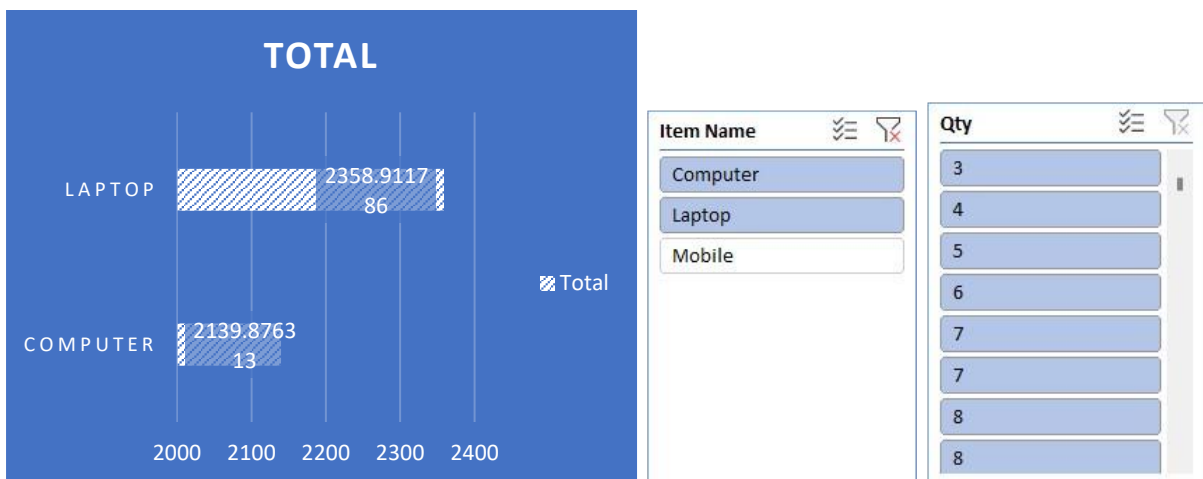
Ans:- To identify the most sold product over the period of May-September, we would need to analyze the sales data within this timeframe. By aggregating the quantity sold for each product across all transactions during this period and then determining which product has the highest total quantity sold,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





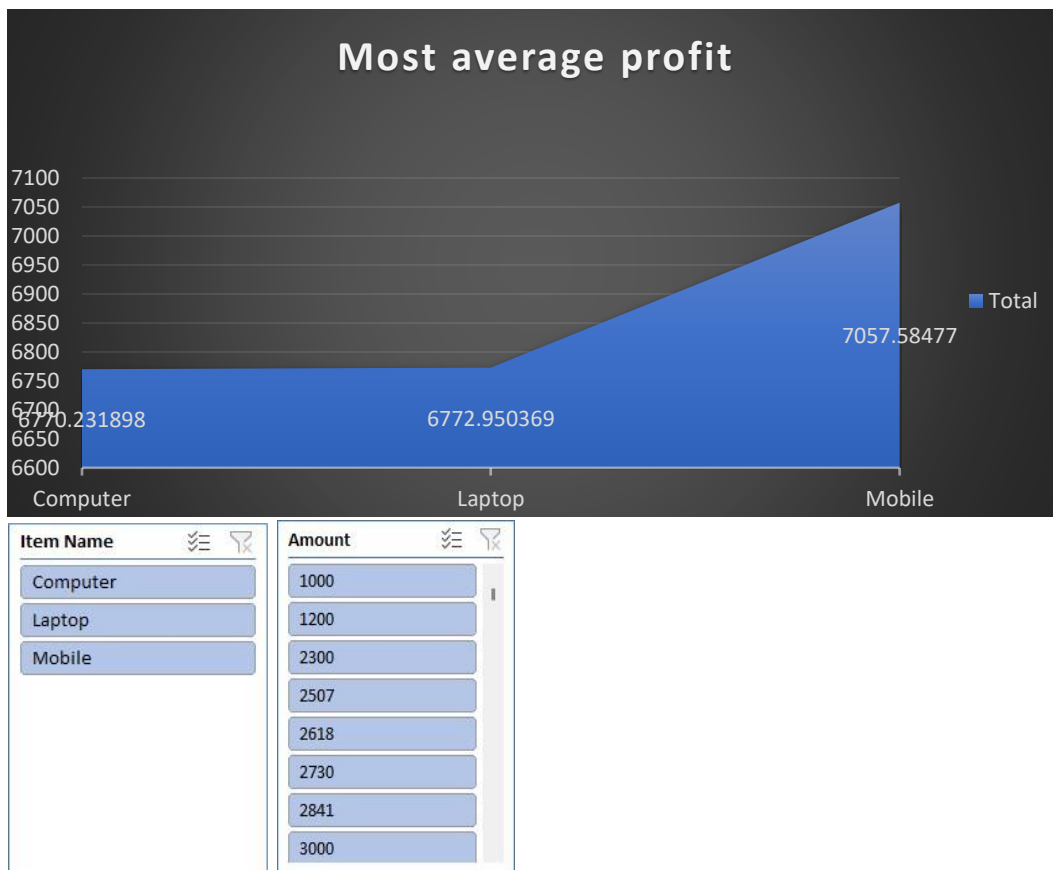
3. Find out which of the two product sold the most over the year Computer or Laptop?

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



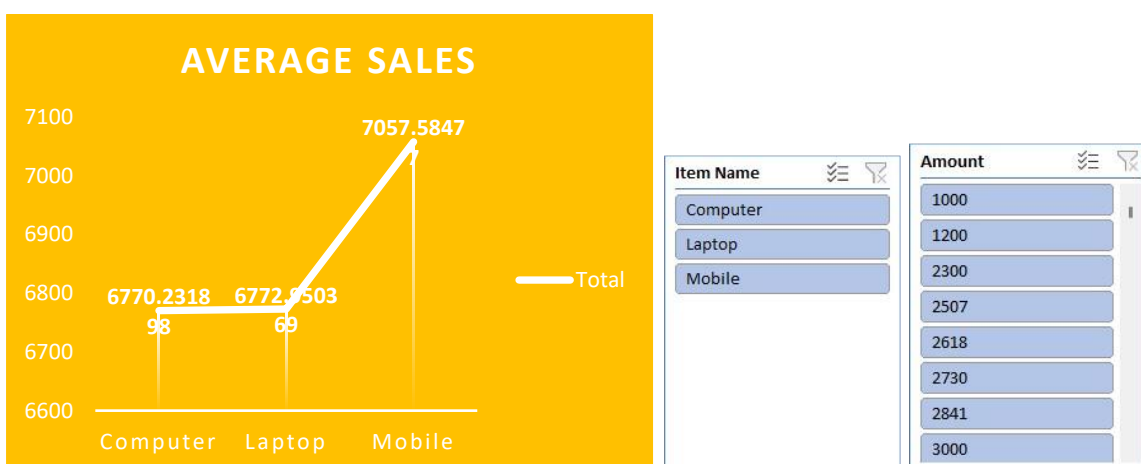
4 . Which item yield most average profit?

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



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 :



## 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.

### Regression:

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.

#### SUMMARY OUTPUT

<i>Regression Statistics</i>	
Multiple R	0.812617
R Square	0.660347
Adjusted R Square	0.629469
Standard Error	1215.119
Observations	13

ANOVA					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	1	31576697	31576697	21.38598	0.000753
Residual	11	16241653	14776514		
Total	12	47818350			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
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.

	<i>Qty</i>	<i>Amount</i>
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

#### ANNOVA

Source of Variation	SS	df	MS	F	P-Value	F crit
Between Groups	84472135	1	84472135	49.45528	1.2E-07	4.195972
Within Groups	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	41600074	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:

<i>Column1</i>	
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

# Store Dataset Report

## Introduction

dataset is essentially a snapshot of the bustling activity within a retail environment, containing detailed records that span various facets of the retail process. It's like a multi-dimensional storybook where each page tells a part of the shopper's journey and the life cycle of a product, from shelf to checkout.

At its core, the dataset captures demographic information that paints a picture of who the customers are, segmenting them by attributes such as gender and age. This demographic breakdown is crucial because it can reveal which segments are most engaged with the store and what products appeal to different groups.

Shipping information closes the loop, connecting the dots between what leaves the store and how it gets to the customer. This segment of the data can signal how efficient the store's logistics are and might even influence customer satisfaction—long shipping times or problematic deliveries can deter future sales. By analyzing this dataset, businesses can unearth patterns in purchase behaviors—like peak buying times or popular items—and cross-reference them with customer demographics to fine-tune target audiences for marketing campaigns. Product trend analysis can guide inventory management, ensuring that the right products are stocked in the right quantities. Shipping trends can inform logistics strategies, improving the speed and reliability of deliveries to boost customer satisfaction.

In short, this dataset isn't just a collection of numbers and categories; it's a dynamic tool that, when interpreted correctly, can serve as a blueprint for smarter business decisions that aim to align the retail store's operations with the desires and needs of its customer base.

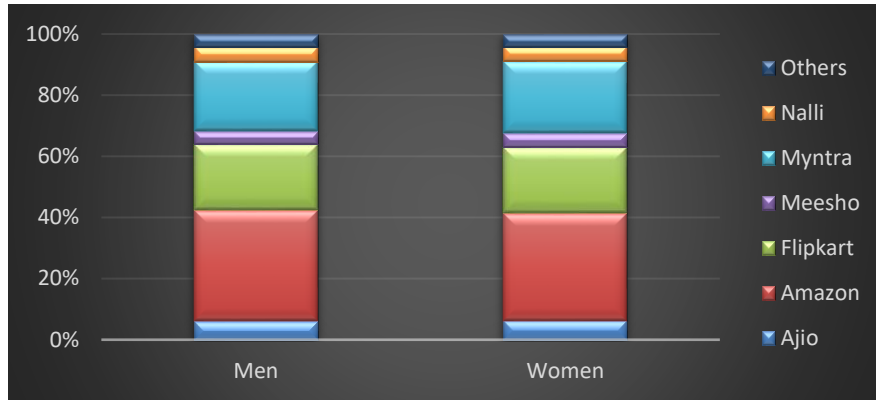
## Questionnaire:

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:
  - a. Kurta
  - b. Set
  - c. 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.



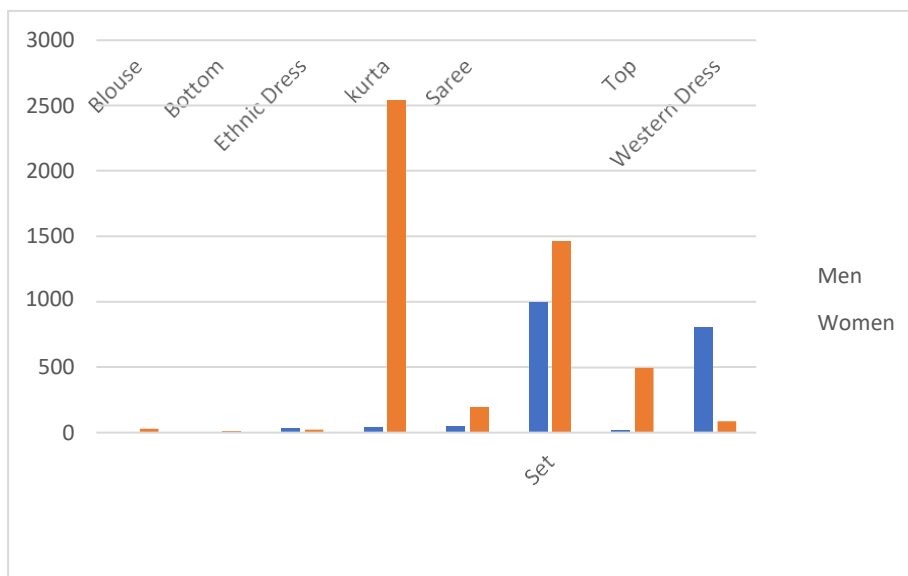
2. Compare category. Find out most sold category above 23 years of age for any gender.

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
<b>Grand Total</b>	<b>8101</b>	<b>18465</b>	<b>26566</b>

The graph is as follows:



Age

22

23

24

25

26

27

28

29

3. 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=212, Sales in women category=543

State	Men	Women	Grand Total
MAHARASHTRA	1390	3144	4534
RAJASTHAN	212	543	755
TAMIL NADU	686	2023	2709
Grand Total	2288	5710	7998

ship-state

New Delhi

ODISHA

PUDUCHERRY

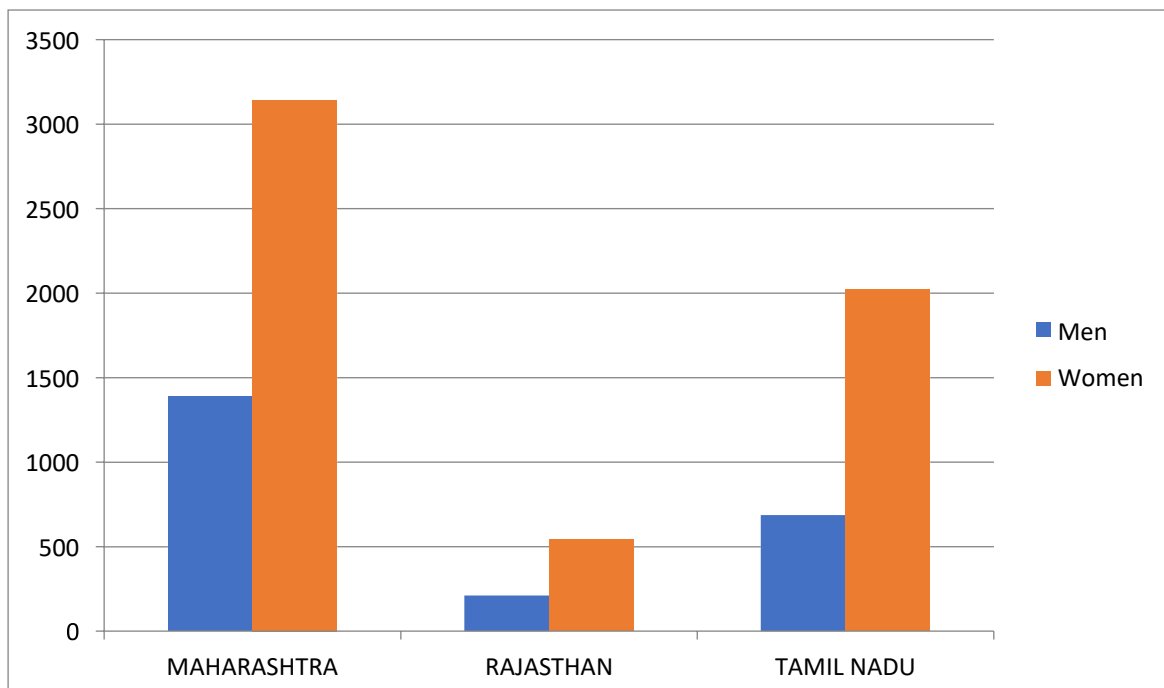
PUNJAB

RAJASTHAN

SIKKIM

TAMIL NADU

TELANGANA

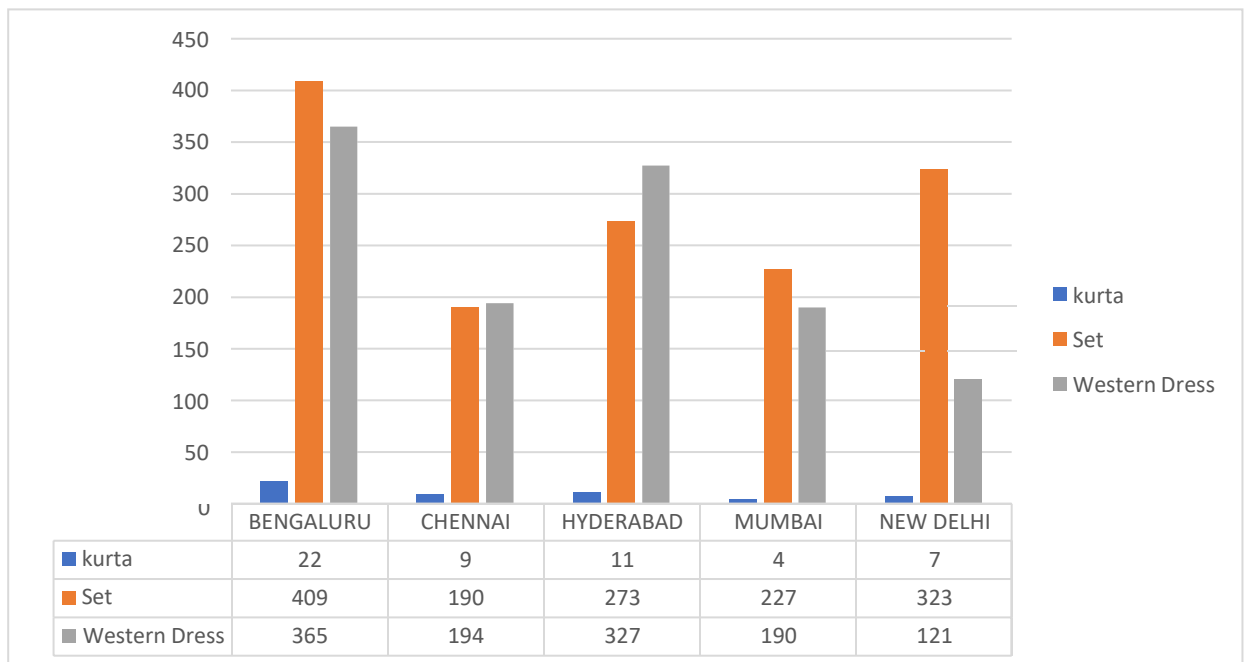




Which city sold most of following categories

- Kurta
- Set
- 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
<b>Grand Total</b>	<b>3259</b>	<b>3383</b>	<b>1358</b>	<b>8000</b>

Category

Blouse

Bottom

Ethnic Dress

kurta

Saree

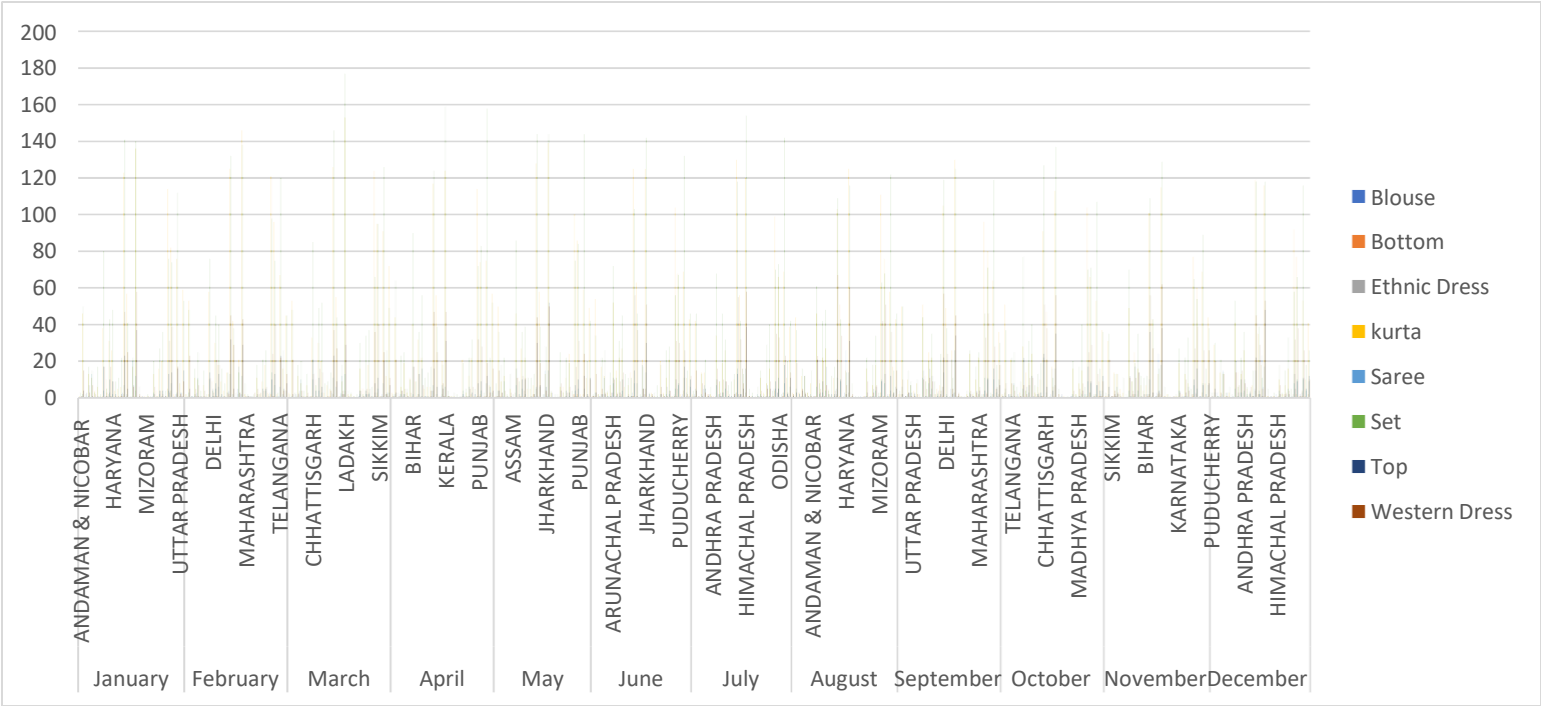
Set

Top

Western Dress

4. 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:



## 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 decision-making to optimize resources, target specific markets, and maximize profits in future store sales endeavours.