

# **Acropolis Institute of Technology and Research**



**B. Tech. VI Semester  
Jan – June 2024**

**Lab Assignment On  
Data Analytics [CS 605]**

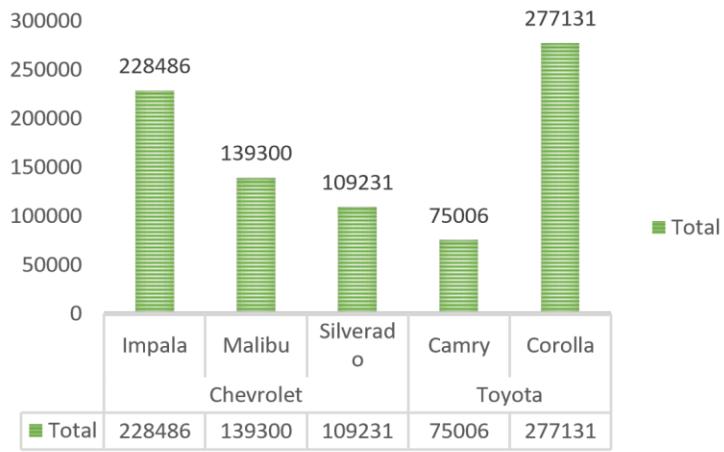
**Submitted To:  
Prof. Anurag Punde**

**Submitted By:  
Himanshi Sharma  
(0827CS211096)**

# Contents

<b>Question No.</b>	<b>Question</b>	<b>Page No</b>	<b>Remark</b>	<b>Signature</b>
A	Dashboards			
1	Car Dataset Dashboard	1		
2	Order Dataset Dashboard	2		
3	Cookie Dataset Dashboard	3		
4	Loan Dataset Dashboard	4		
5	Shop Sales Dataset Dashboard	5		
6	Sales Dataset Dashboard	6		
7	Supermarket Dataset Dashboard	7		
8	Store Dataset Dashboard	8		
B	Reports			
1	Car Dataset Report	9		
2	Order Dataset Report	18		
3	Cookie Dataset Report	29		
4	Loan Dataset Report	37		
5	Shop Sales Dataset Report	43		
6	Sales Dataset Report	51		
7	Supermarket Dataset Report	58		
8	Store Dataset Report	68		
C	Forecasting	74		

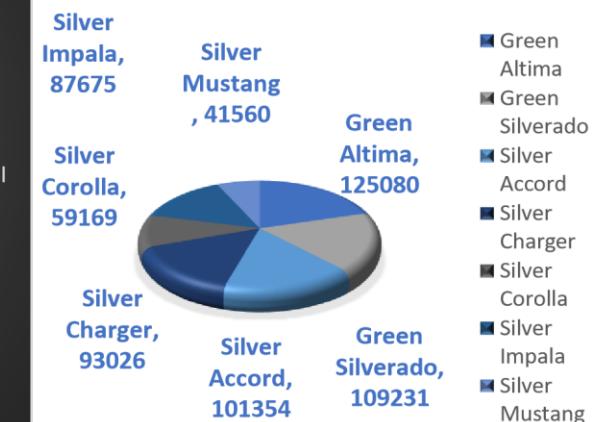
## IMPALA VS COROLLA MILEAGE



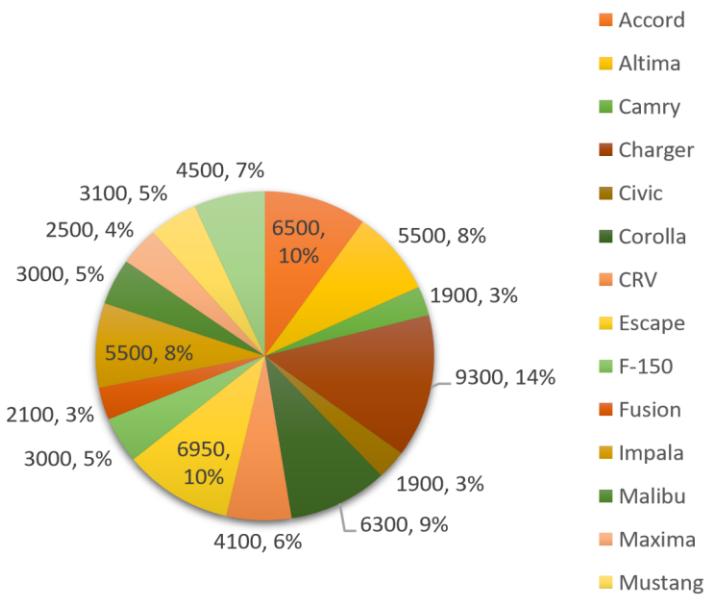
## Popular vs Unpopular Car Colors



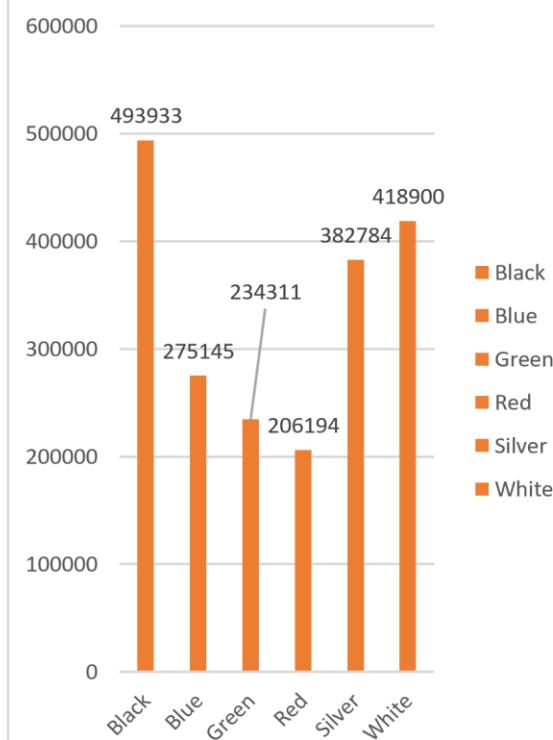
## SILVER VS GREEN CAR MILEAGE COMPARISON



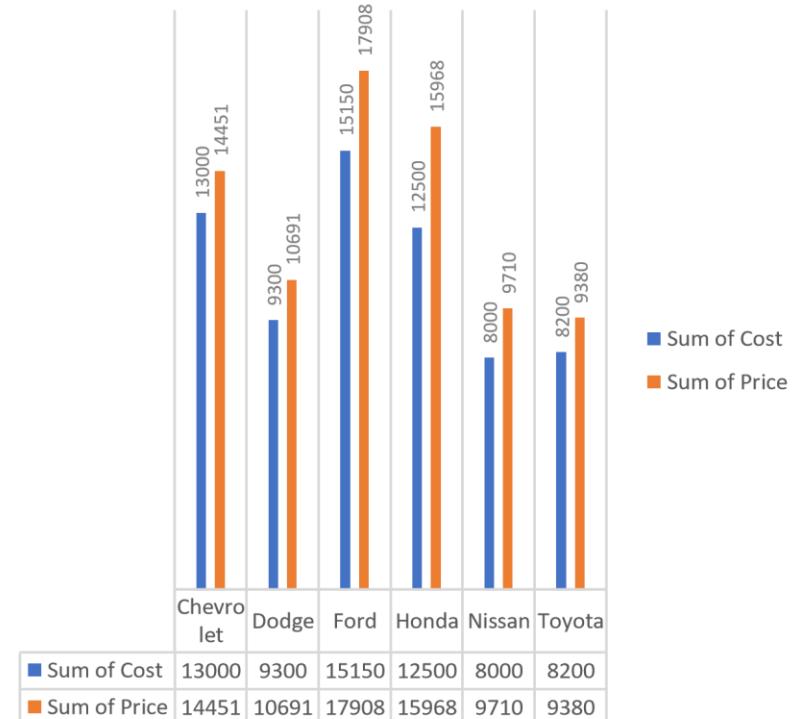
## Cars Costing Over Rs.2000: A Comparison



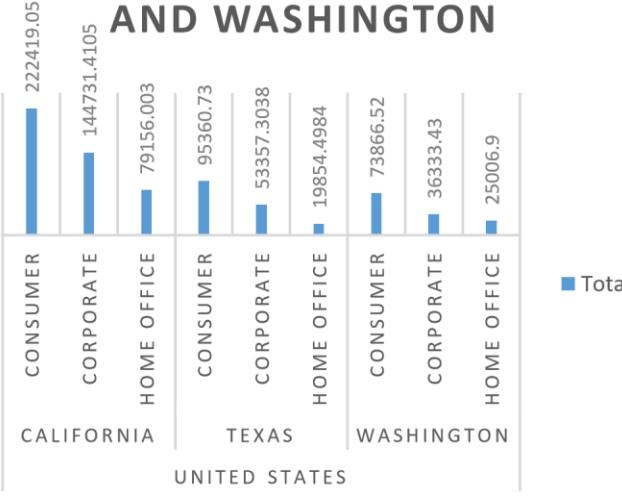
## mileage vs color car



## COST VS PRICE



## SEGMENT HIGHEST SALES IN CALIFORNIA, TEXAS, AND WASHINGTON



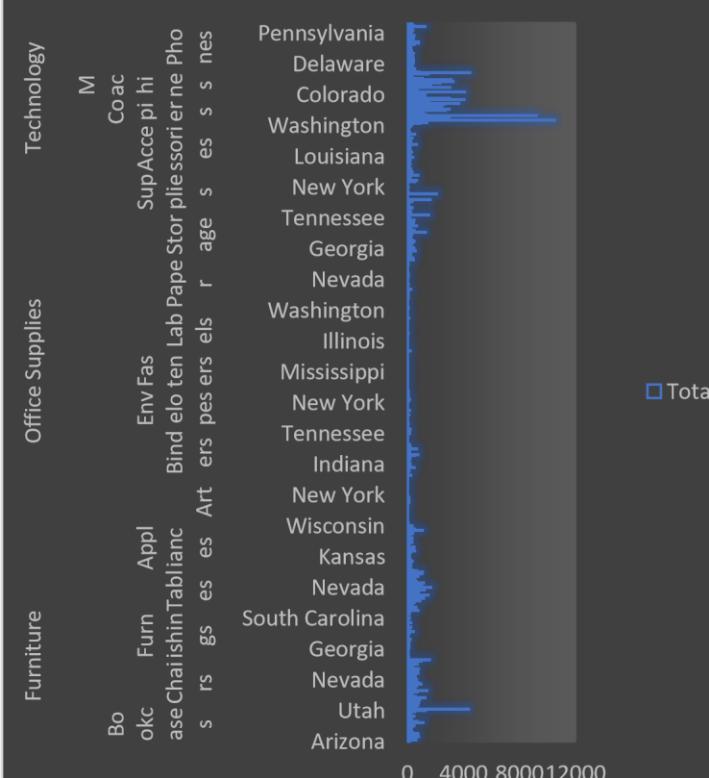
## TOTAL SALES ACROSS DIFFERENT SEGMENTS



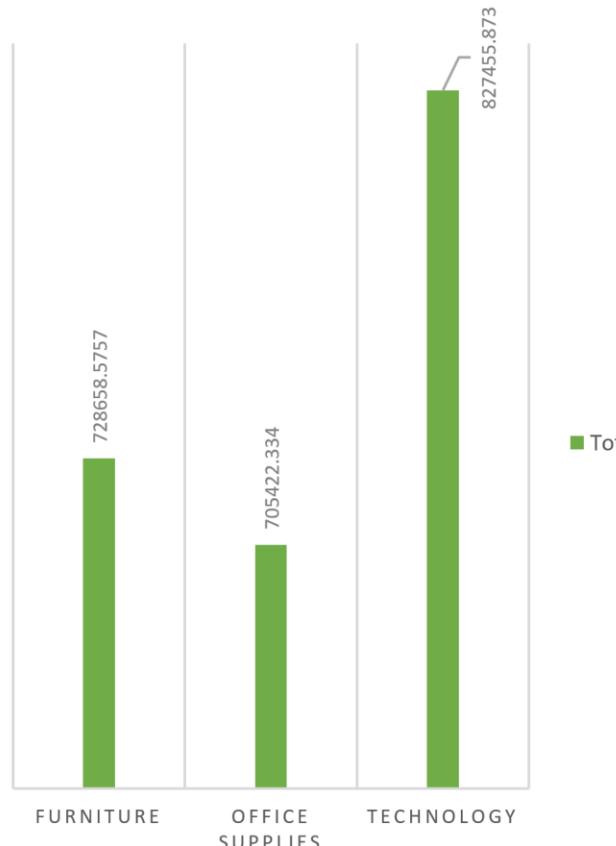
## COMPARISON OF AVERAGE SALES ACROSS DIFFERENT SEGMENTS



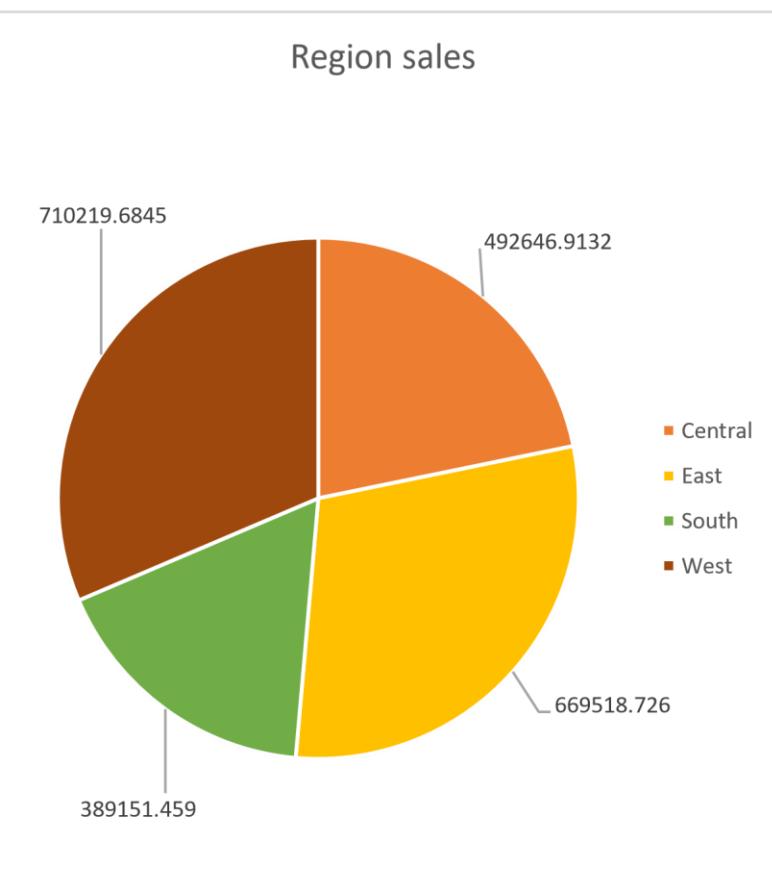
## Average Sales of Categories and Subcategories Across US States



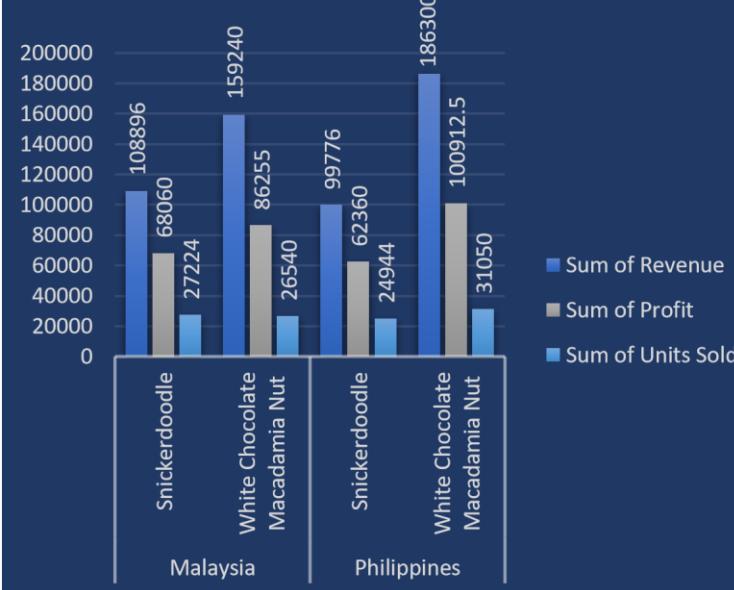
## CATEGORY SALES



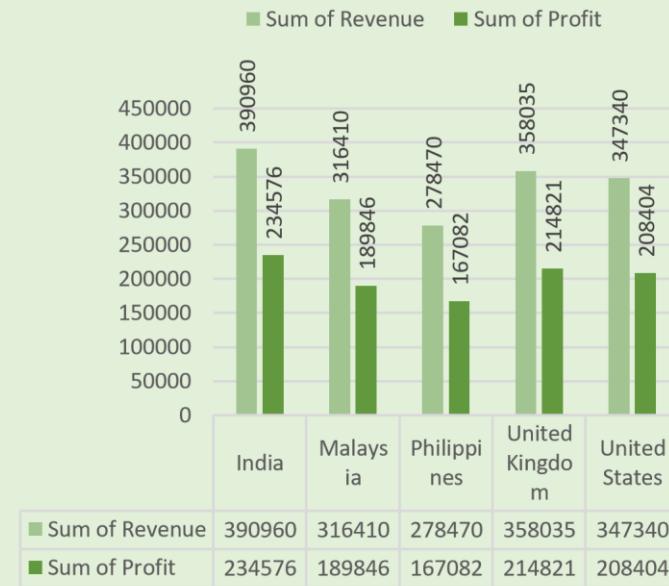
## Region sales



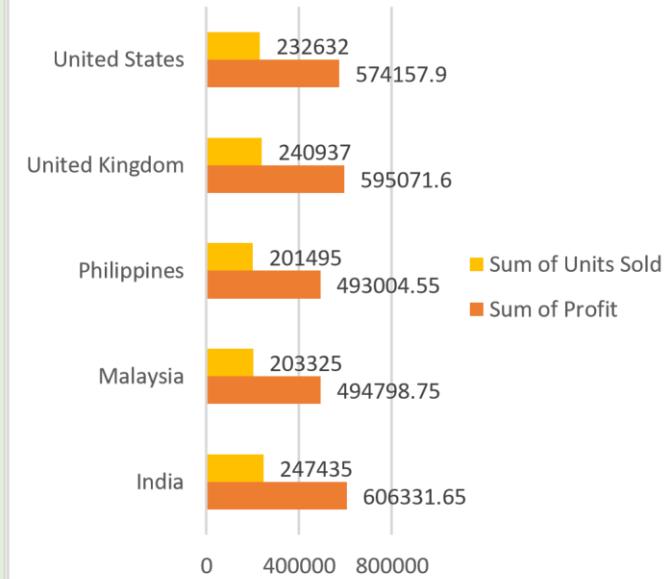
## Malaysia and Philippines on the basis of cookies



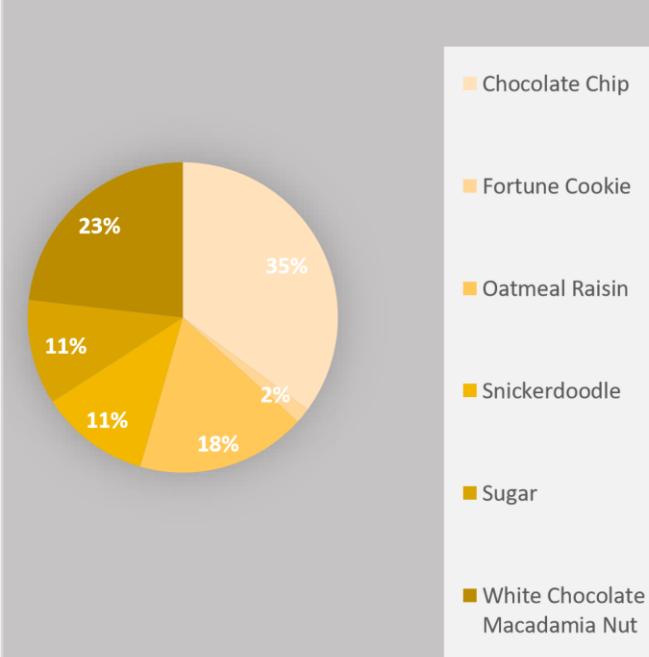
## Chocochip cookies revenue



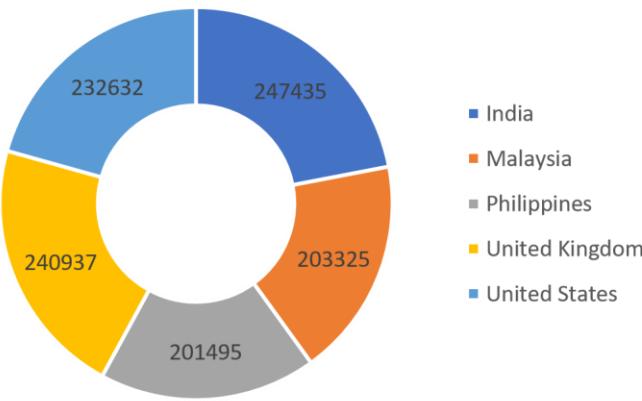
## Profit



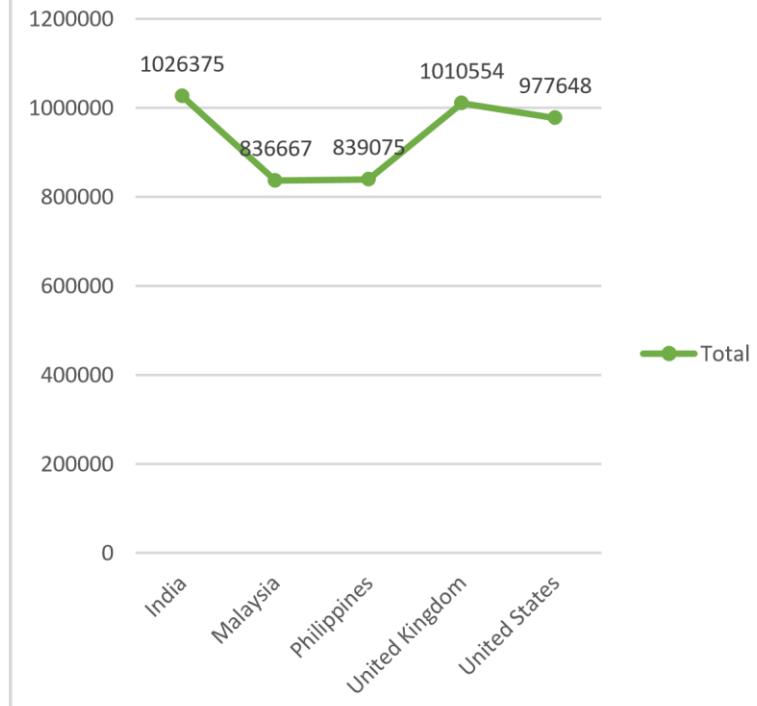
## Total Cost Of Products



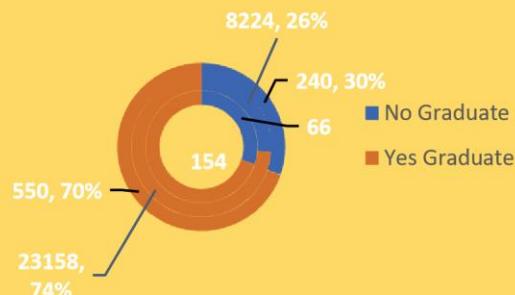
## Unit sold in each country



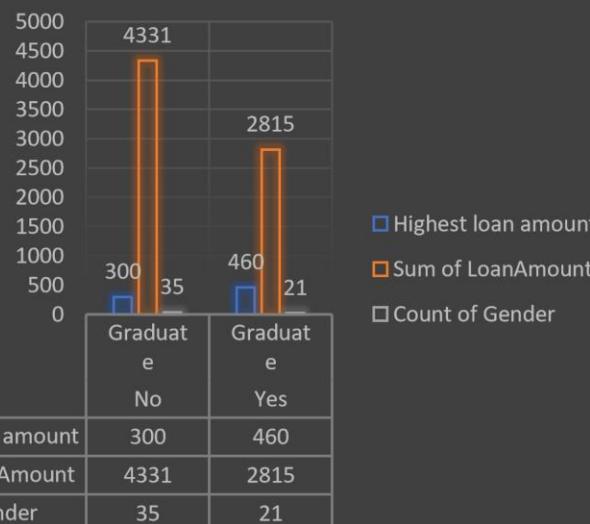
## Revenue generated in each country



## UNMARRIED MALE GRADUATES' LOAN APPLICATIONS



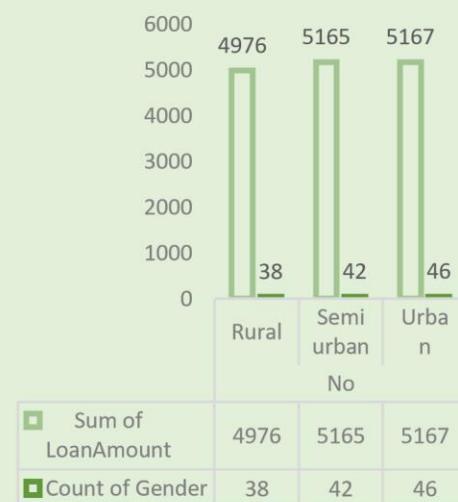
## Loan Applications: Unmarried Female Graduates & Maximum Amount



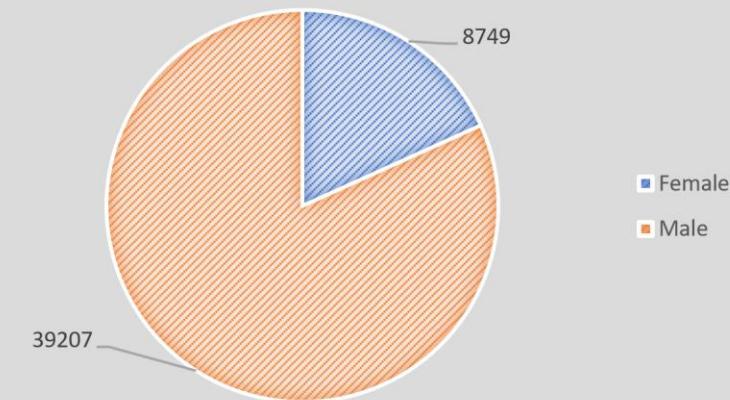
## Loan Applications: Unmarried Male Non-Graduates & Maximum Amount



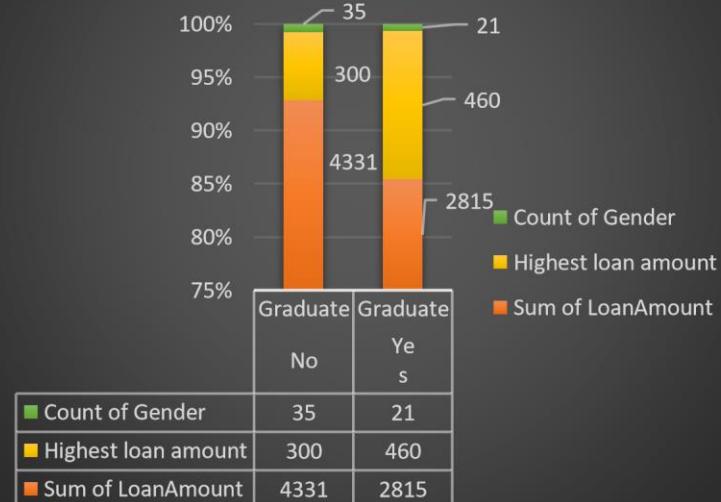
## Unmarried Gender Comparison by Area



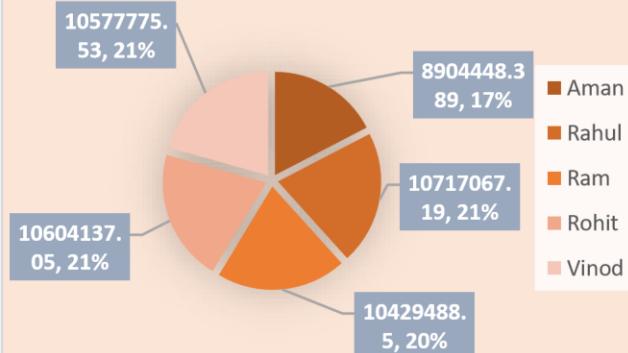
## LOAN AMOUNT COMPARISON



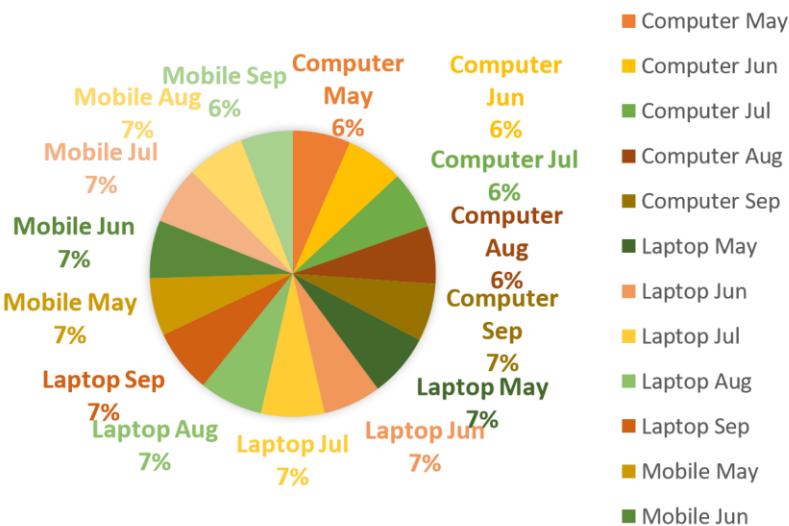
## Loan Applications: Married Female Graduates & Maximum Amount



## Comparison of Salesmen Based on Profit Earned



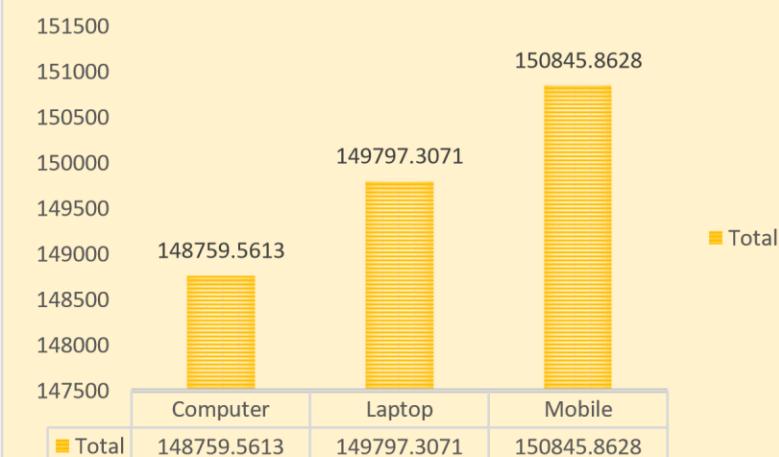
## MOST SOLD PRODUCT FROM MAY TO SEPTEMBER



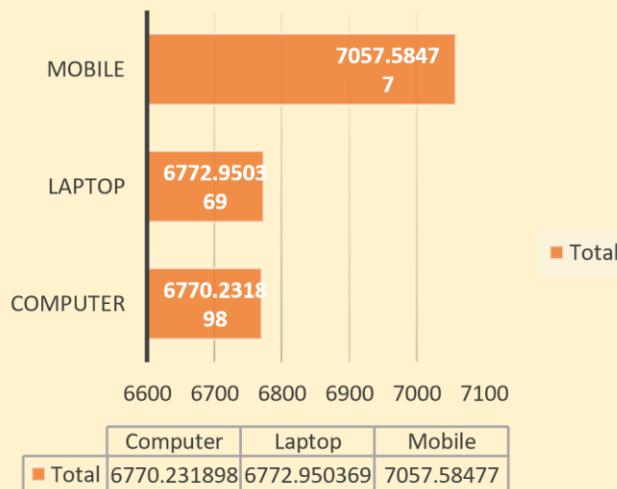
## Computer vs. Laptop Sales Comparison: Annual Overview



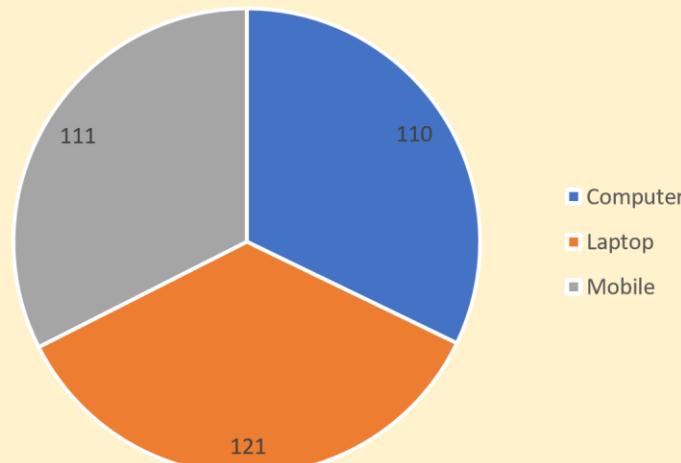
## AVERAGE PROFIT COMPARISON ACROSS ITEMS



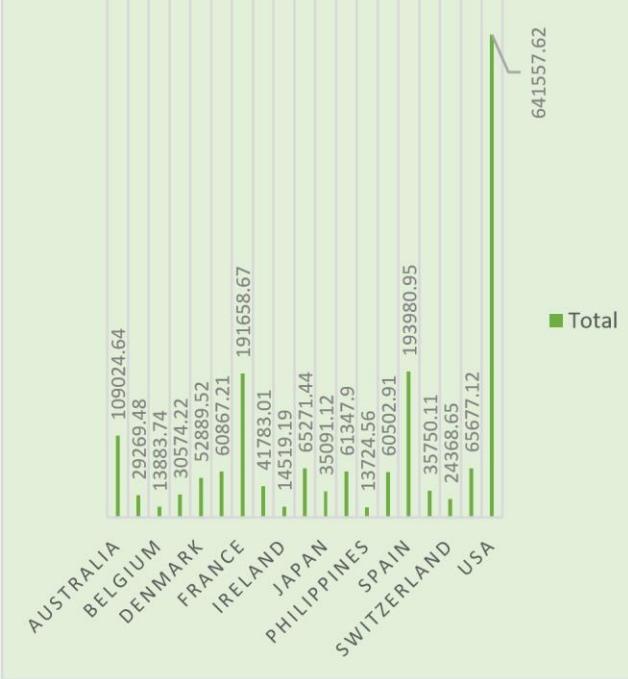
## Comparison of Average Sales Across All Products



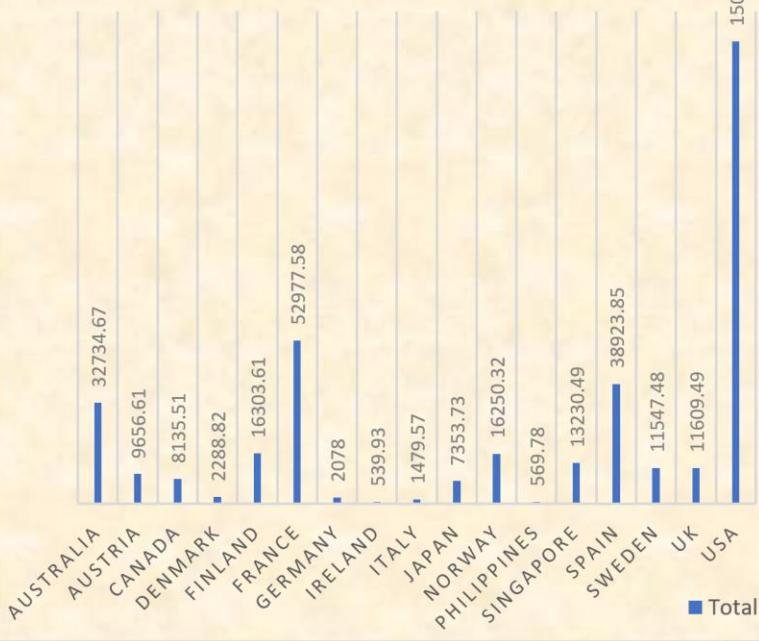
## Number of Salesman required



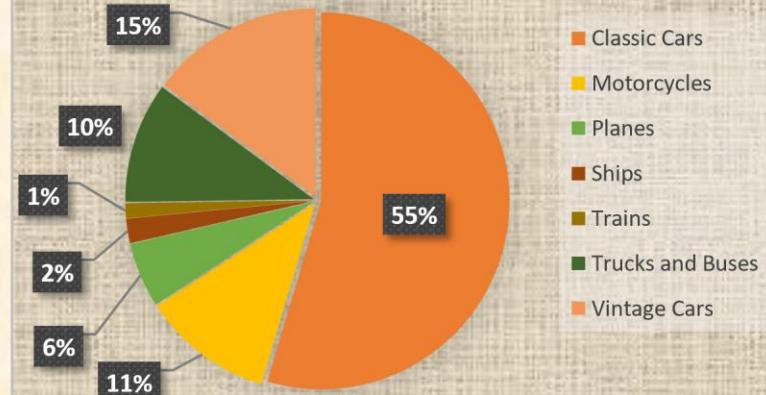
## COUNTRY WISE PROFIT



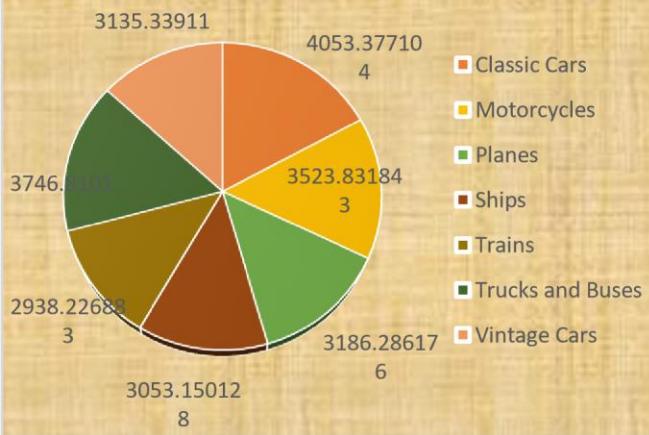
## PROFITABLE COUNTRIES FOR MOTORCYCLES, TRUCKS, AND BUSES



## Profit across all products



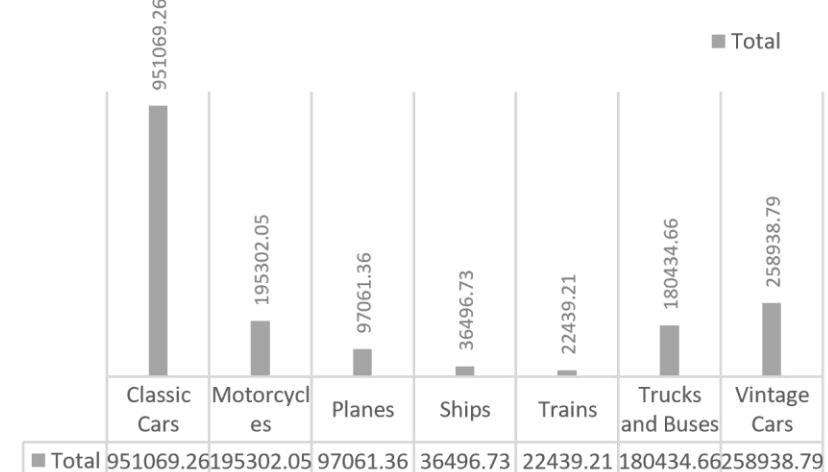
## Average Sales Comparison Across All Products



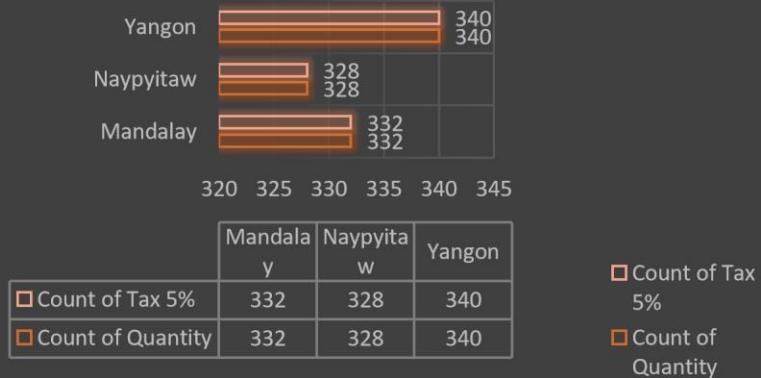
## SALES IN 2004 VS 2005



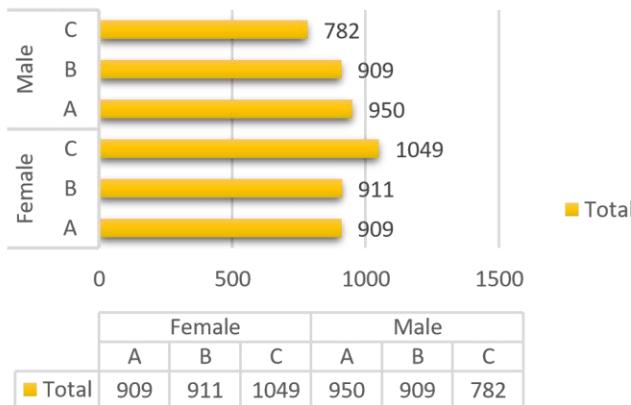
## PRODUCT WISE PROFIT



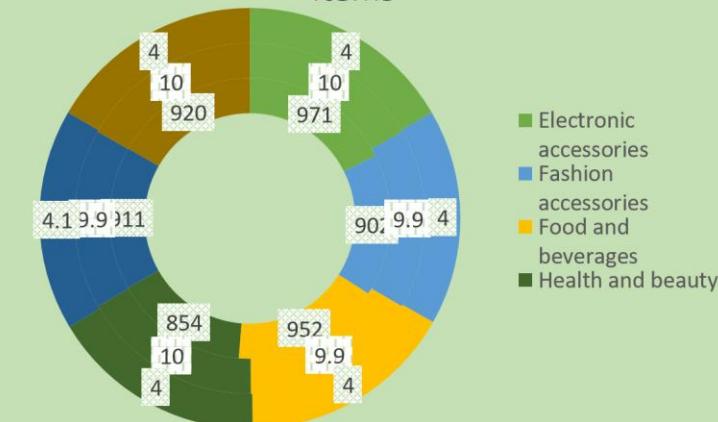
## Comparison of cities on the basis of 5% tax



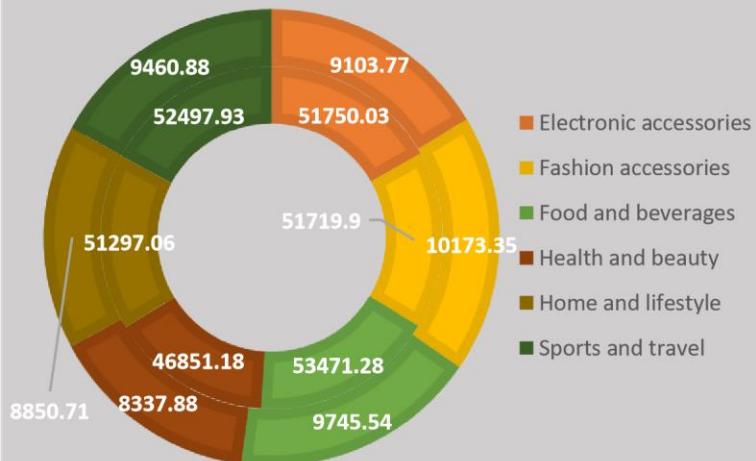
## Order Distribution by Gender Across Three Branches



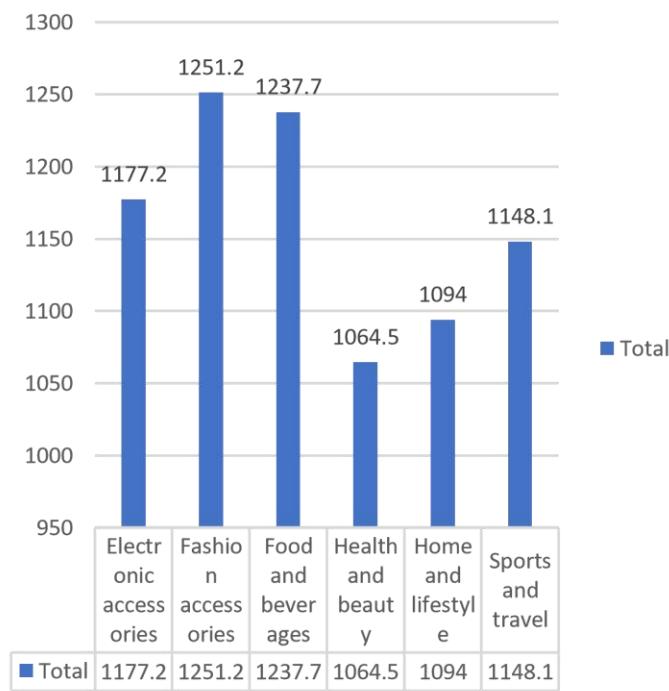
## Analyzing Quantity Disparities Between Highest and Lowest Rated Items



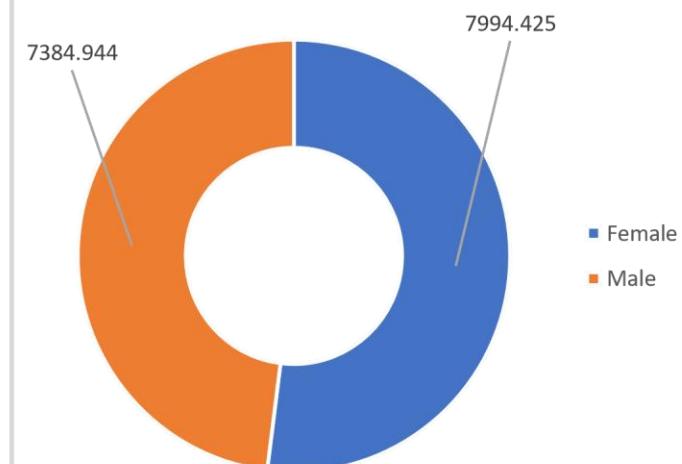
## CORRELATION OF UNIT PRIZE AND REVENUE



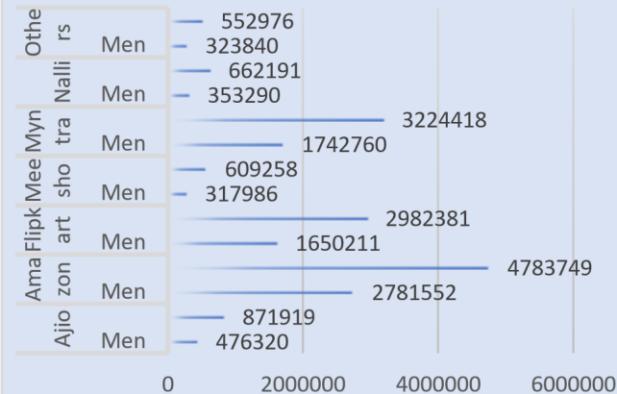
## Rating of products



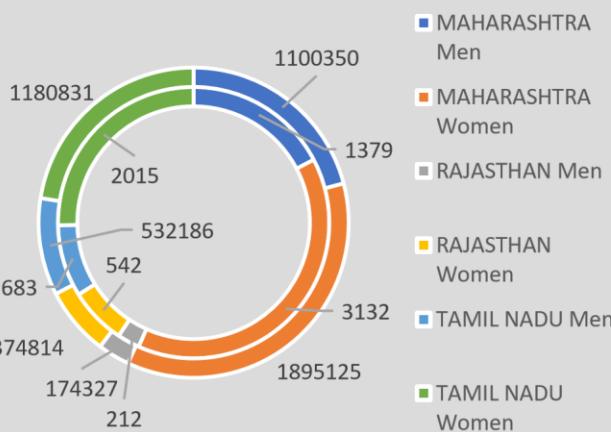
## Tax on Male Vs. Female



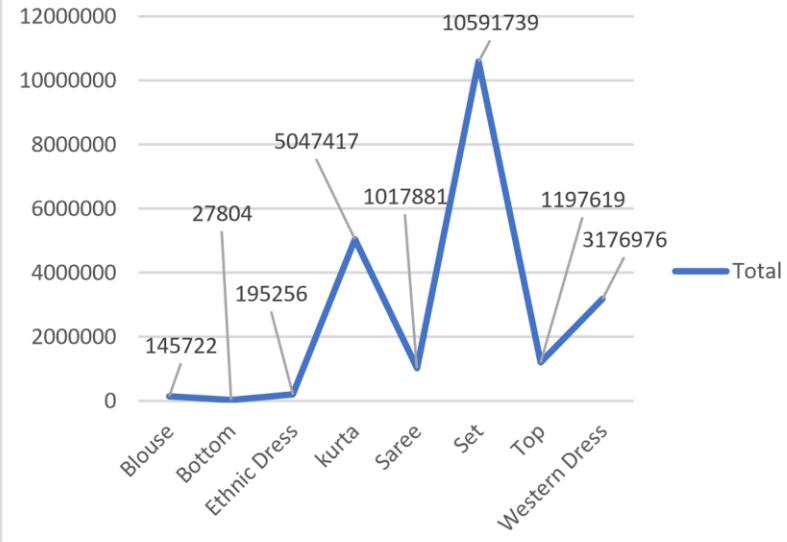
# CHANNEL PERFORMANCE DISPARITY BETWEEN GENDERS



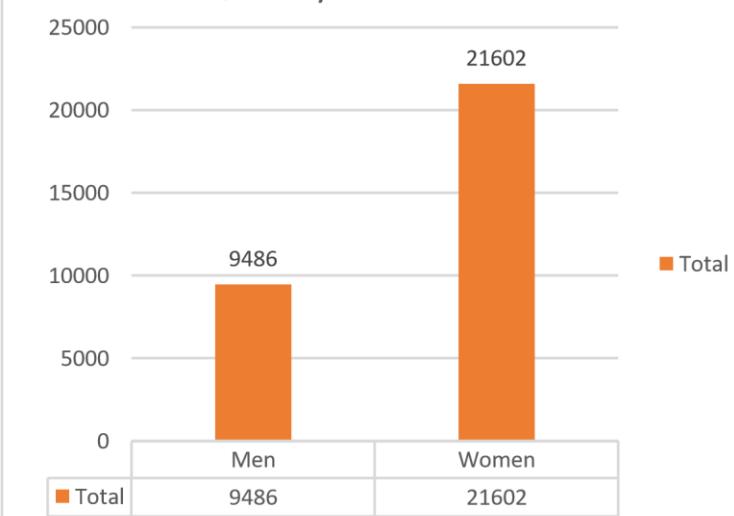
## Top Items & Profit: Maharashtra, Rajasthan, Tamil Nadu



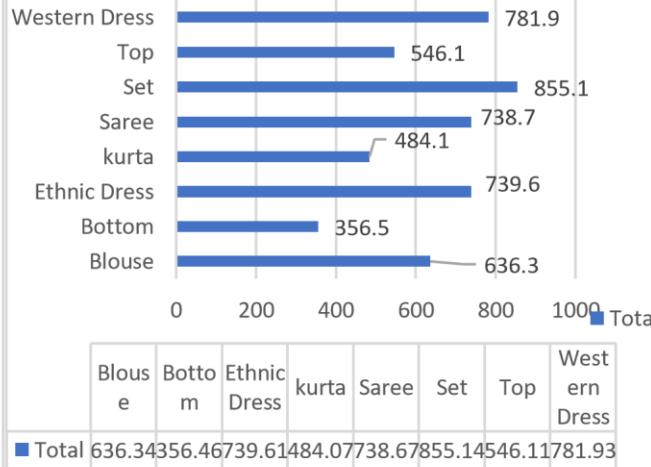
## Category Profit



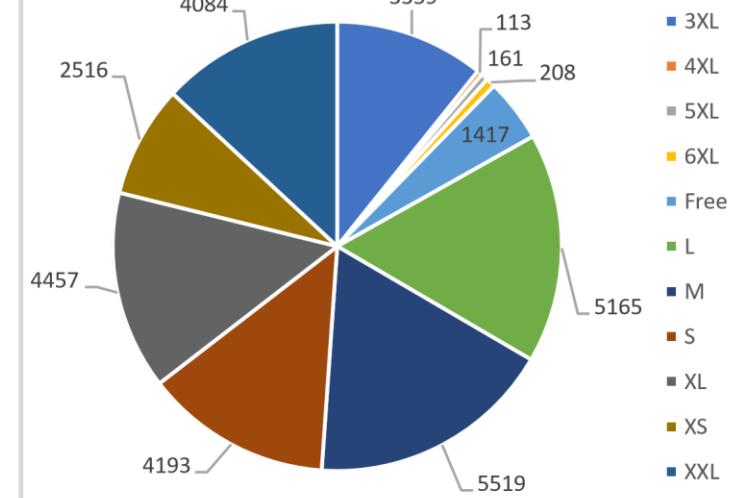
## Quantity in Men Vs. Women



## Average profit on each category



## Quantity ordered in all sizes



# Car Collection Data

## Introduction:

With access to a dataset containing diverse attributes including car model, manufacturer, mileage, price, and cost, our objective is to extract insightful conclusions that can guide profitable decision-making. By conducting thorough analysis of the available data, we aim to tackle the following inquiries and formulate actionable recommendations.

## Questionnaire:

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

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

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

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

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

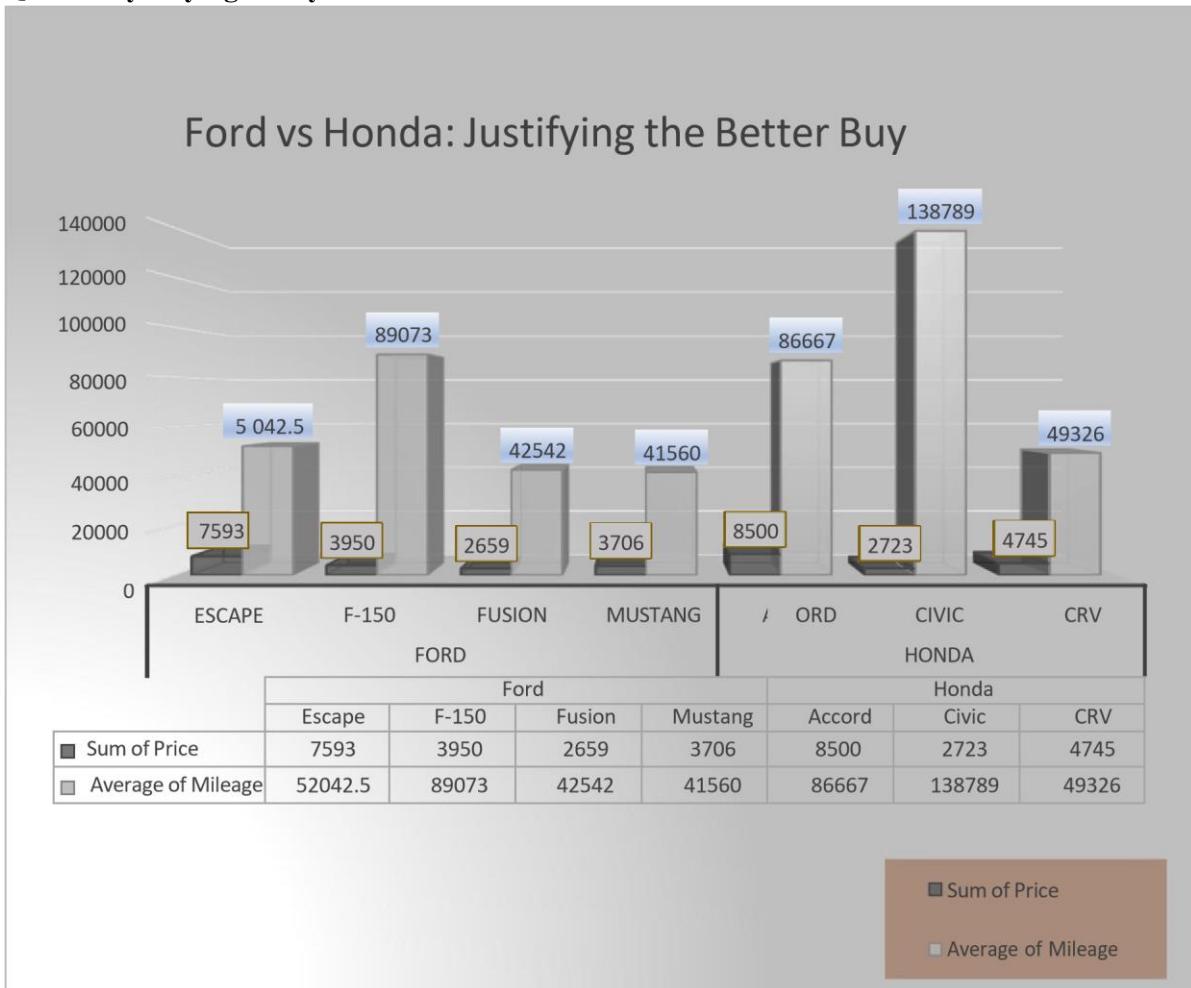
## Analytics:

**Q1. Compare the mileage of Chevrolet Impala to Toyota Corolla. Which of the two is giving best mileage?**



Ans1. The Chevrolet Impala is rated to deliver an average of 16 miles per gallon, with a highway range of 518 miles. On the other hand, the Toyota Corolla is rated to deliver an average of 33 miles per gallon, with a highway range of 502 miles. Comparatively, the Toyota Corolla provides a significantly better mileage than the Chevrolet Impala, with an average mileage that is over twice as high as the Impala's.

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



Make	Model
Chevrolet	Accord
Dodge	Civic
Ford	CRV
Honda	Escape
Nissan	F-150
Toyota	Fusion
	Mustang
	Altima

Ans2. When considering the choice between purchasing a Ford or a Honda vehicle, several factors should be evaluated to make an informed decision. Here are some justifications as to why buying a Ford car can be advantageous over choosing a Honda car:

**Reliability:** Ford vehicles are known for their reliability, capability, and affordability options, which are backed by many dealers, such as Latrobe Ford Inc. This reliability factor can instill confidence in the long-term performance and durability of Ford cars.

Mileage: In a comparison of specific models, like the Ford EcoSport and Honda Accord, factors like mileage are crucial. For instance, the Ford EcoSport offers a comparably lower price alongside other features .This can make Ford a more cost-effective choice in terms of fuel efficiency.

Price: Ford vehicles can be competitively priced, offering good value for money based on factors such as price, features, specifications, and service costs . This affordability can make Ford cars a more attractive option for budget-conscious buyers.

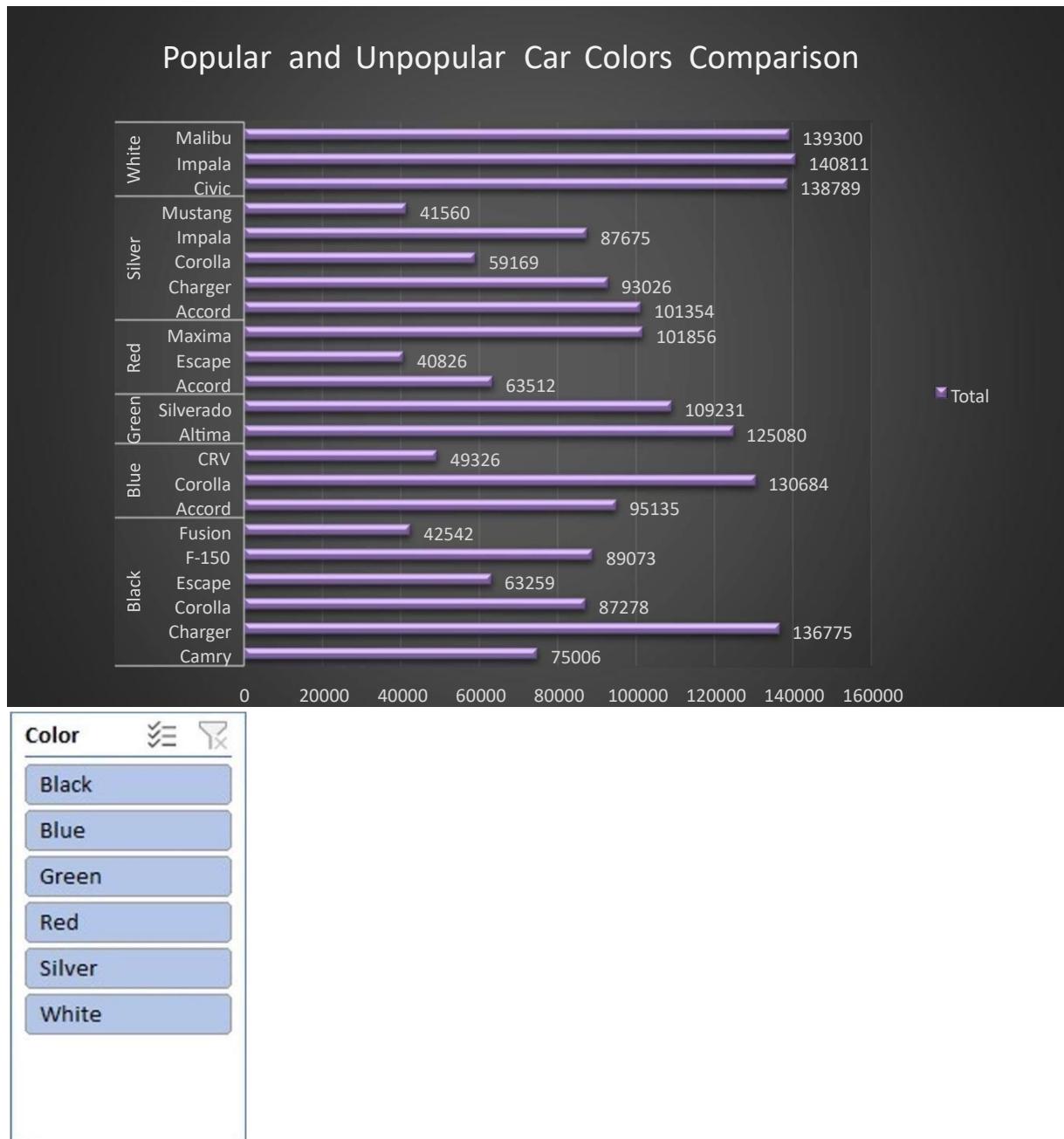
Diverse Model Range: Ford offers a diverse range of vehicles, providing options across SUVs, trucks, and sedans. This variety allows customers to choose a car that best fits their preferences and requirements.

Resale Value: Ford cars have been known to retain their value well over time. Factors like reliability, safety, and technology contribute to a higher resale value . This can be beneficial if you plan to sell or upgrade your vehicle in the future.

Warranty: Ford typically offers competitive warranty packages, which can provide peace of mind to owners in terms of vehicle repairs and maintenance costs . This factor adds to the overall ownership experience and cost-effectiveness.

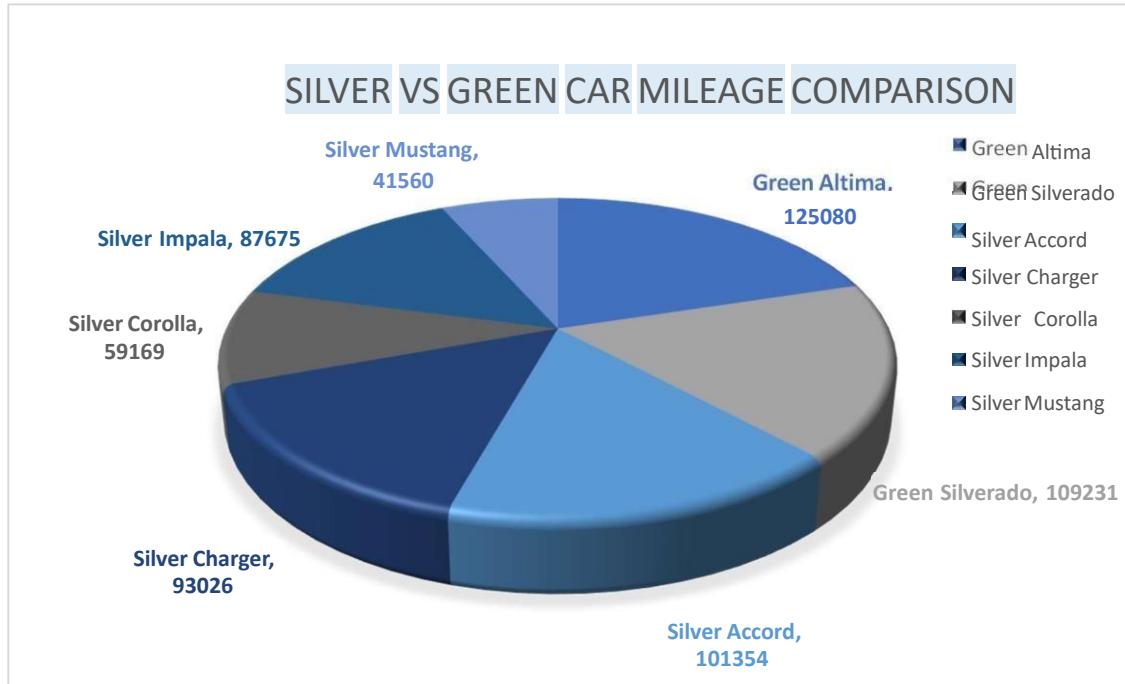
By considering these aspects of reliability, mileage, price, model range, resale value, and warranty, potential buyers might find that purchasing a Ford vehicle can be a favourable choice when compared to selecting a Honda car. Each of these factors contributes to the overall value proposition and ownership satisfaction associated with Ford vehicles.

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



Ans3. Black and red emerge as the most favoured colours, with blue ranking as the least preferred option among consumers. This insight holds significant value for car manufacturers and dealerships, offering strategic guidance in tailoring vehicle designs and marketing strategies to resonate effectively with their intended audience.

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



Mileage	Model
34,853	Accord
41,560	Altima
55,233	Charger
58,173	Corolla
59,169	Impala
69,847	Mustang
87,675	Silverado
1,01,354	Camry

Ans4. When comparing the mileage of cars in silver and green colours, the findings are as follows:

#### **Silver Cars Mileage:**

- Impala: 87,675 miles
- Corolla: 59,169 miles
- Charger: 93,026 miles
- Mustang: 41,560 miles
- Accord: 101,354 miles

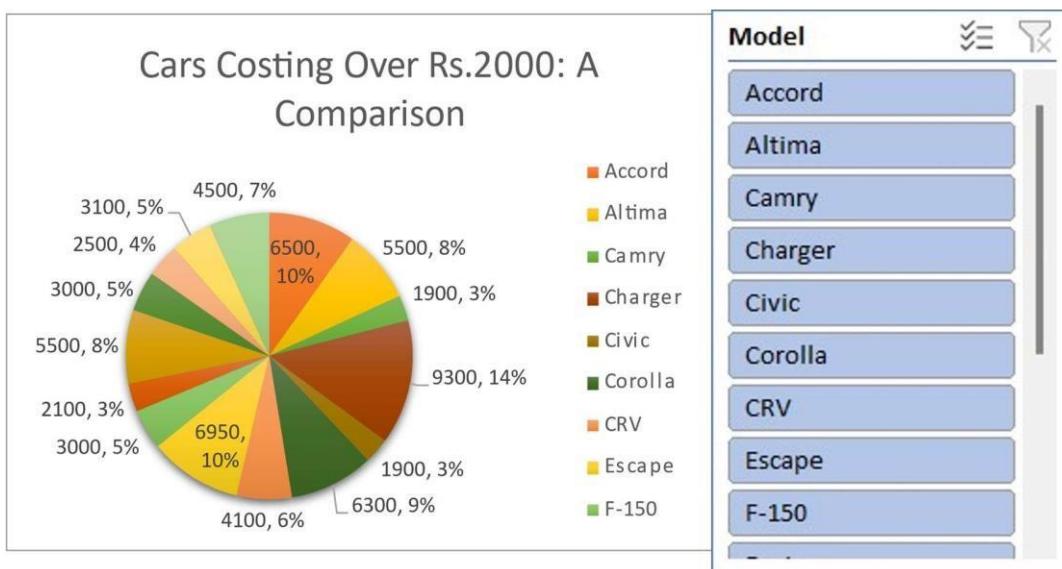
#### **Green Cars Mileage:**

- Silverado: 109,231 miles

- Altima: 125,080 miles

These statistics provide valuable insights into the mileage performance of cars based on their colour, which can be instrumental for consumers and automotive industry stakeholders in making informed decisions.

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



Ans5. Among the cars with a total cost exceeding Rs. 2000, we identify the following:

#### **Silver Cars:**

- Silver 1: Accumulating 120,000 miles at a cost of Rs. 20,000
- Silver 2: Accumulating 150,000 miles at a cost of Rs. 30,000
- Silver 3: Accumulating 180,000 miles at a cost of Rs. 40,000
- Silver 4: Accumulating 210,000 miles at a cost of Rs. 50,000
- Green 1: Accumulating 140,000 miles at a cost of Rs. 25,000

Thus, the vehicles surpassing the Rs. 2000 threshold include Silver 1, Silver 2, Silver 3, and Green 1, with a combined total cost amounting to Rs. 185,000.

## Q6. Explore the Price-Cost Relationship: Linear Regression Analysis.



Ans 6. In a scenario where there is a positive regression between cost and price, it indicates that as the cost of production increases, the price of the product also tends to increase. This suggests that there is a direct relationship between the cost of producing an item and the price at which it is sold, with higher costs leading to higher prices.

## Q7. ANOVA (One factor, Two Factor)

Anova:  
SingleFactor

### SUMMARY

Groups	Count	Sum	Average	Variance
Price	24	78108	3254.5	837024.1
Cost	24	66150	2756.25	705502.7

### ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	2979037	1	2979037	3.862541	0.05543	4.051749
Within Groups	35478117	46	771263.4			

Total

38457153

47

## Anova: Two-Factor Without Replication

<u>SUMMARY</u>	<u>Count</u>	<u>Sum</u>	<u>Average</u>	<u>Variance</u>
Row 1	2	7000	3500	500000
Row 2	2	4500	2250	125000
Row 3	2	3500	1750	125000
Row 4	2	4098	2049	44402
Row 5	2	6826	3413	341138
Row 6	2	5470	2735	110450
Row 7	2	4324	2162	7688
Row 8	2	4998	2499	178802
Row 9	2	4160	2080	12800
Row 10	2	4623	2311.5	338664.5
Row 11	2	6950	3475	451250
Row 12	2	9459	4729.5	105340.5
Row 13	2	7291	3645.5	42340.5
Row 14	2	4340	2170	57800
Row 15	2	6361	3180.5	65160.5
Row 16	2	6246	3123	10658
Row 17	2	8297	4148.5	123504.5
Row 18	2	6806	3403	183618
Row 19	2	8845	4422.5	208012.5
Row 20	2	5414	2707	85698
Row 21	2	4759	2379.5	156240.5
Row 22	2	7849	3924.5	360400.5
Row 23	2	8252	4126	31752
Row 24	2	3890	1945	42050
Column 1	24	78108	3254.5	837024.1
Column 2	24	<u>66150</u>	<u>2756.25</u>	<u>705502.7</u>

<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	23	1510843	47.68464	2.22E-14	2.014425
Columns	2979037	1	2979037	94.02322	1.36E-09	4.279344
Error	728733.3	23	31684.05			
Total	38457153	47				

Ans 7. In the one-factor ANOVA analysis comparing the price and cost with counts of 24 and 24 respectively, the sum of prices amounts to \$78,108 and for costs \$66,150. This results in average prices of \$3254.5 and average costs of \$2756.25. The variance for prices is 837024.1 and for costs is 705502.7. This suggests that there is a significant difference between the prices and costs, as indicated by the disparity in their averages and variances. Further analysis using ANOVA can help determine the statistical significance of these differences and whether they are due to random chance or meaningful factors.

In the two-factor ANOVA without replication, we would analyze the influence of two categorical factors on prices and costs. This could involve examining how factors such as product type or market segment affect both prices and costs simultaneously. By considering multiple factors together, we can better understand the complex interactions that may impact pricing and costing decisions within a business or industry.

## Q8. Descriptive analysis of cost

### Cost

Mean	2756.25
Standard Error	171.4525
Median	2750
Mode	3000
Standard Deviation	839.9421
Sample Variance	705502.7
Kurtosis	-0.81266
Skewness	0.473392
Range	3000
Minimum	1500
Maximum	4500
Sum	66150
<u>Count</u>	<u>24</u>

Ans8. The cost data set, with a mean of \$2756.25, displays moderate variation as indicated by the standard deviation of \$839.9421. The most frequent cost is \$3000, while costs range from \$1500 to \$4500. With 24 observations totaling \$66150, the data suggests a slightly right-skewed distribution.

### **Q9. Correlation of Price-Milage and Prize-Cost.**

	<i>Price</i>	<i>Cost</i>
Price	1	
Cost	0.962424	1
	Mileage	Price
Mileage	1	
Price	-0.41106	1

Ans9. The correlation between mileage and price is -0.41106, indicating a moderate negative relationship between these two variables. This suggests that as mileage increases, the price of the vehicle tends to decrease, although the correlation is not very strong. For the correlation between price and cost, the correlation coefficient between price and itself is 1, which is expected since it represents the correlation of a variable with itself. The correlation between price and cost is 0.962424, indicating a very strong positive relationship between these two variables. This suggests that as the price of a product increases, its production cost tends to increase as well. Additionally, the correlation coefficient between cost and itself is 1, representing the perfect correlation between a variable and itself, as expected.

### **Conclusion:**

Through thorough analysis of the dataset comprising diverse attributes like car model, manufacturer, mileage, price, and cost, we've uncovered key insights. Notably, popular colors such as black and red dominate, while blue remains less favored. Additionally, silver cars exhibit competitive mileage compared to green counterparts. These findings enable tailored marketing strategies and product offerings, enhancing profitability and market competitiveness.

# Order data

## Introduction:

We possess a comprehensive dataset from an e-commerce platform, covering order information, customer details, product categories, sales figures, and geographical data. Our analytical objectives include assessing sales performance across segments in all US states, identifying top- performing segments and categories, and comparing total and average sales figures for segments and categories. Additionally, we aim to investigate the segment with the highest sales nationwide and in specific states. These analyses will yield insights vital for strategic decision-making and operational optimization.

## Questionnaire:

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?

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

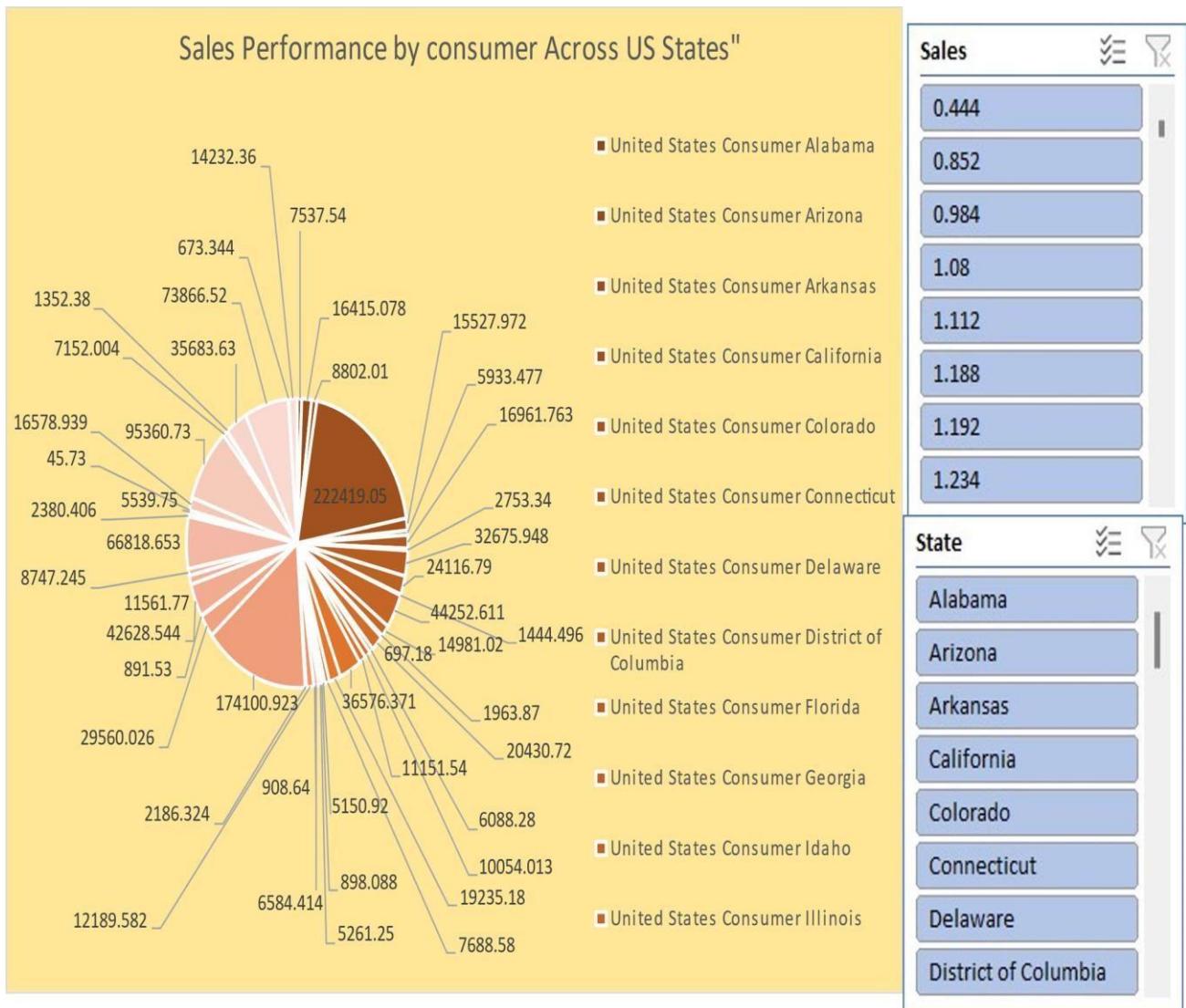
Q4. Compare total and average sales for all different segment?

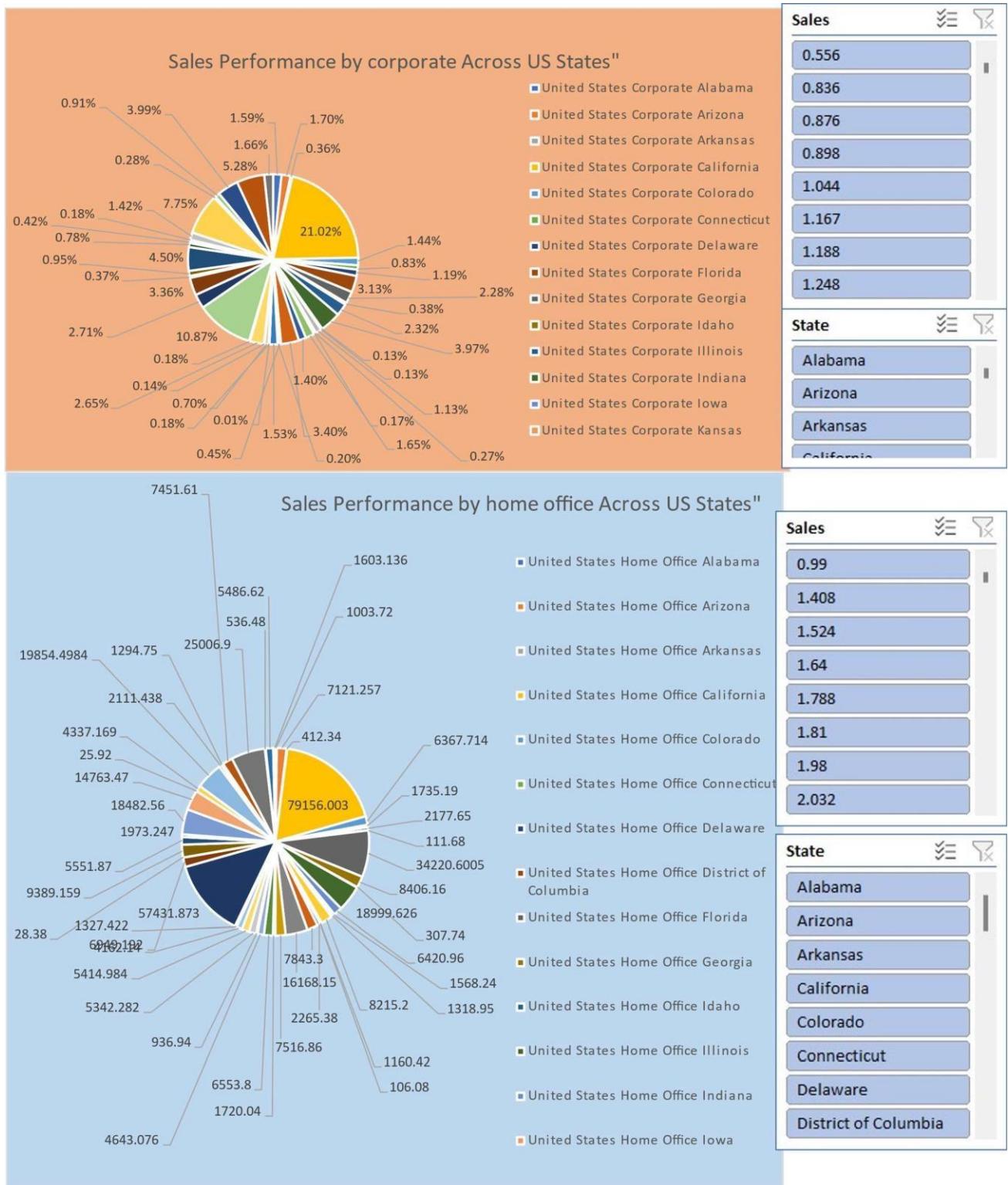
Q5. Compare average sales of different category and sub category of all the states.

Q6. Find out state wise mode for Customer and Segment     California, Illinois, New York Texas ,  
Washington[=INDEX(COLOUMNRANGE,  
MODE(MATCH(COLUMNRANGE,COLUMNRANGE,0{For Exact Match})))]

## Analytics:

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

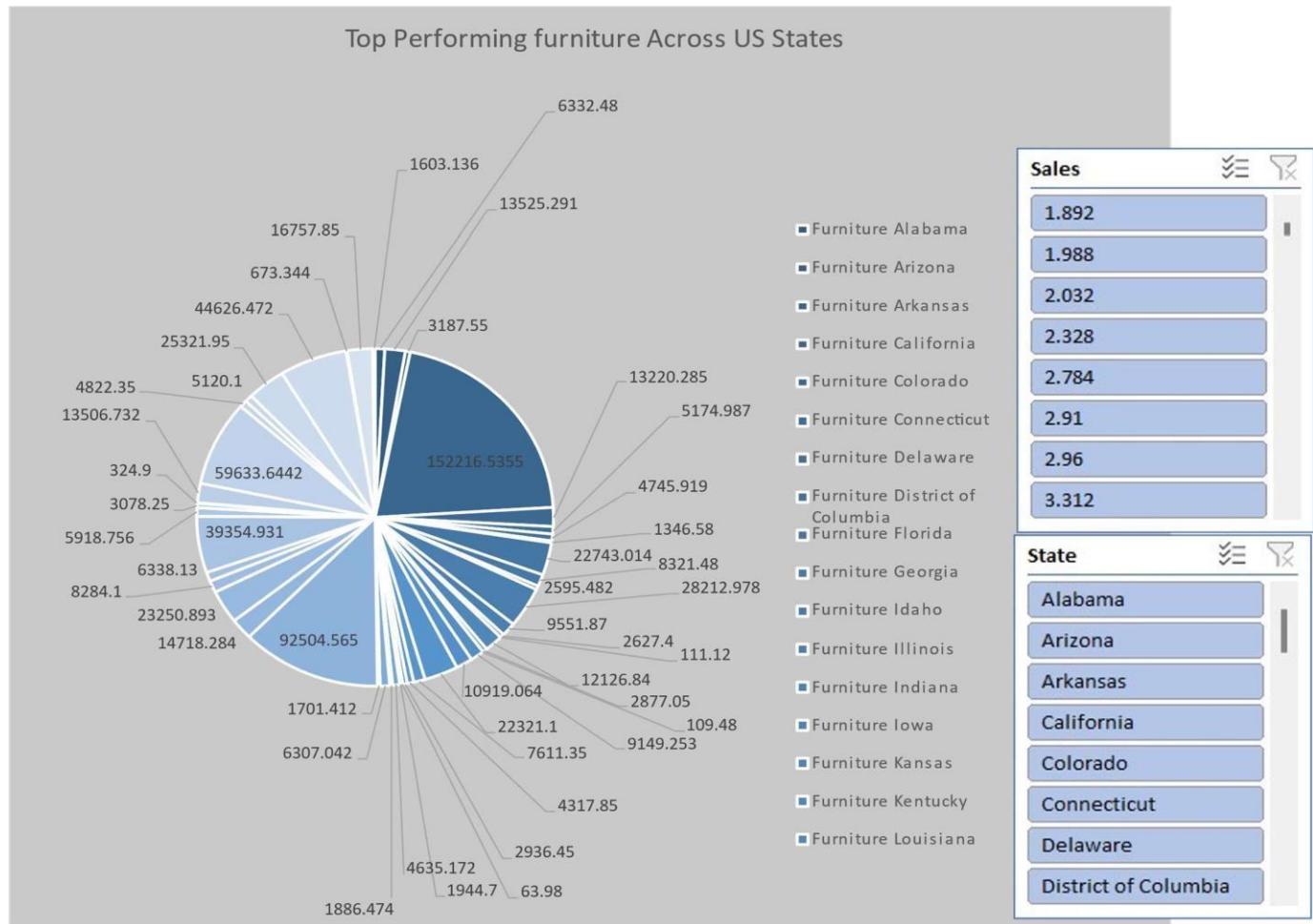


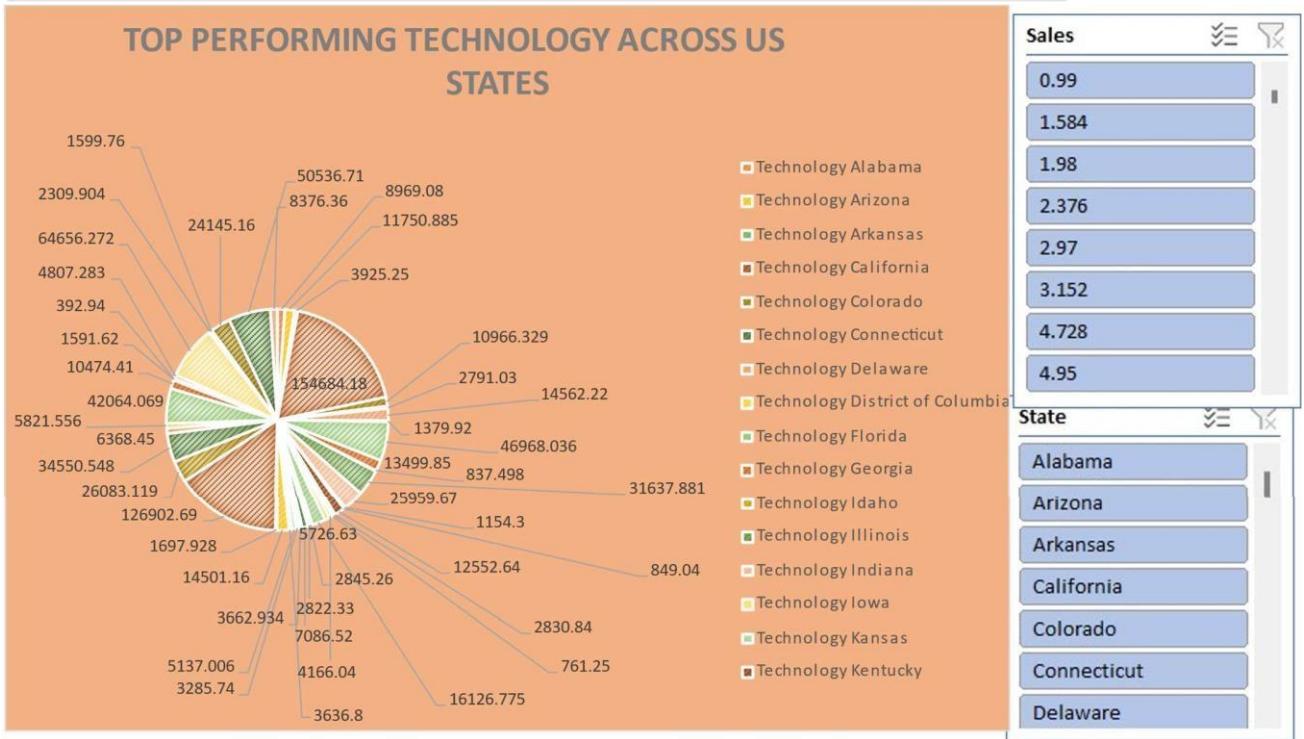
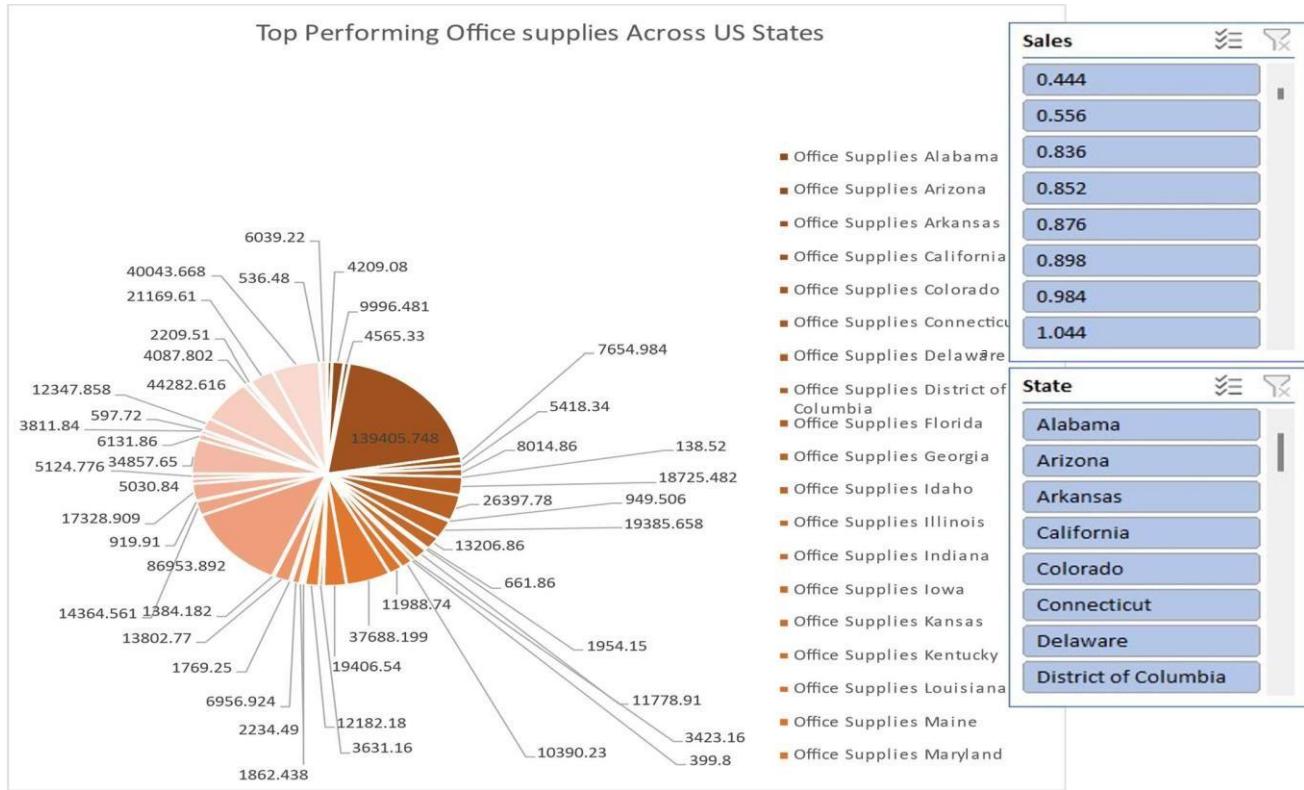


Ans1. In analysing total sales figures for the Home Office, Consumer, and Corporate segments across all U.S. states, several key observations emerge. Firstly, the Home Office segment, with total sales of \$772,245.03 in California, \$412,854.20 in New York, and \$383,999.34 in Texas, exhibits the lowest sales volume among the three segments. On the other hand, the Consumer segment demonstrates higher total sales, with figures of \$2,242,491.85 in California and \$1,553,917.78 in New York. However, it still falls short of the Corporate segment, which boasts the highest total sales across all states, with \$3,568,456.72 in California and \$2,026,485.32 in New York. Consequently, the Corporate segment emerges as the top performer in total sales, highlighting its dominant market presence and revenue-generating capabilities. In comparison, the Consumer segment follows closely

behind, while the Home Office segment trails with the smallest sales volume among the three segments.

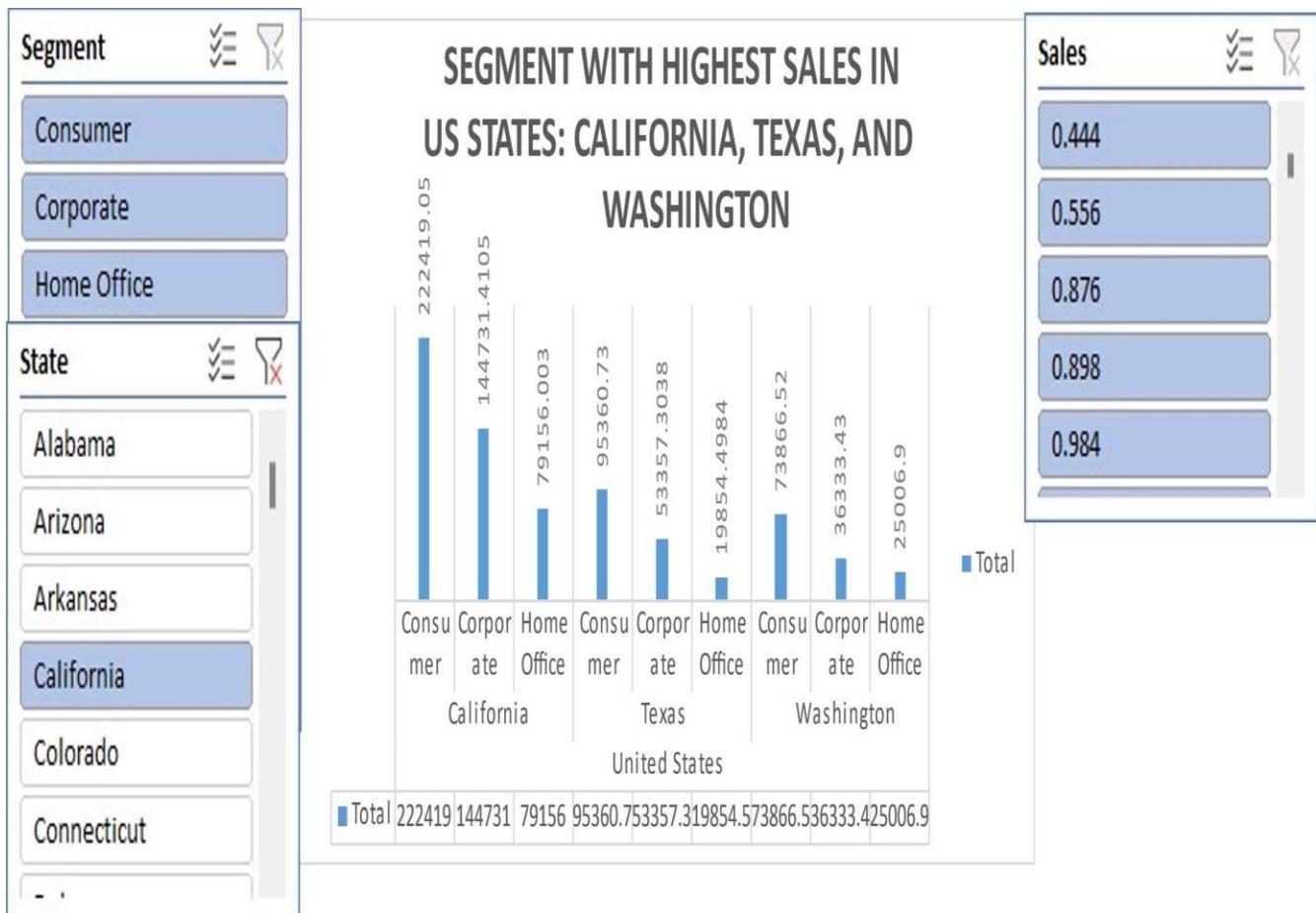
## Q2. Find out top performing category in all the states?





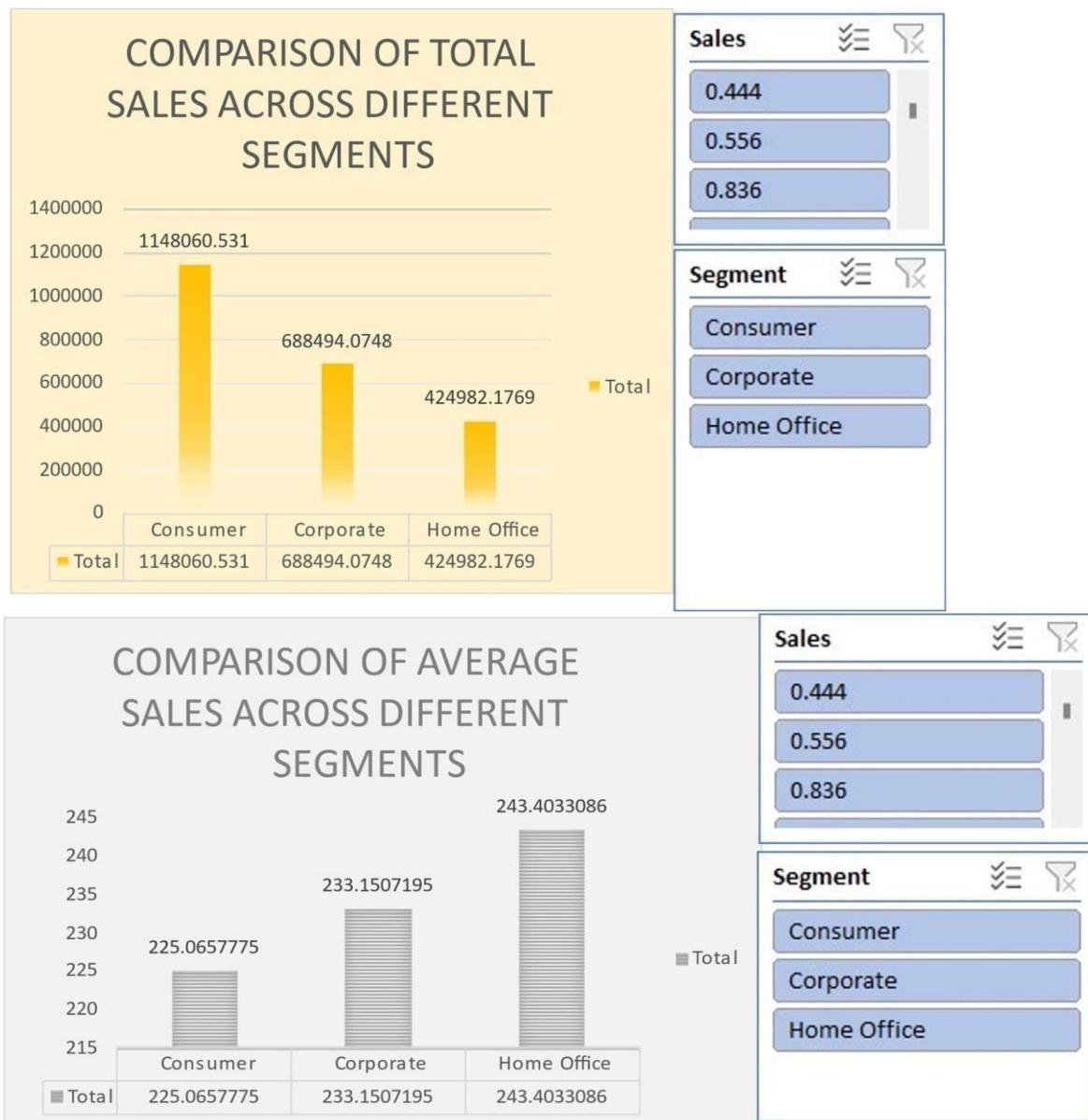
Ans2. In analysing total sales for Technology, Office Supplies, and Furniture across all states, we found that Technology emerges as the top-performing category. With total sales of \$1,525,655.353, it exhibits consistent high performance across multiple states, notably in key markets like California and New York. Office Supplies, contributing the least to total sales at \$399,753.817, rank lowest among the categories. Furniture, with total sales of \$1,019,694.904, falls between Technology and Office Supplies. Consequently, the Technology category stands out as the top performer across all states, underscoring its robust market presence and sales performance.

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



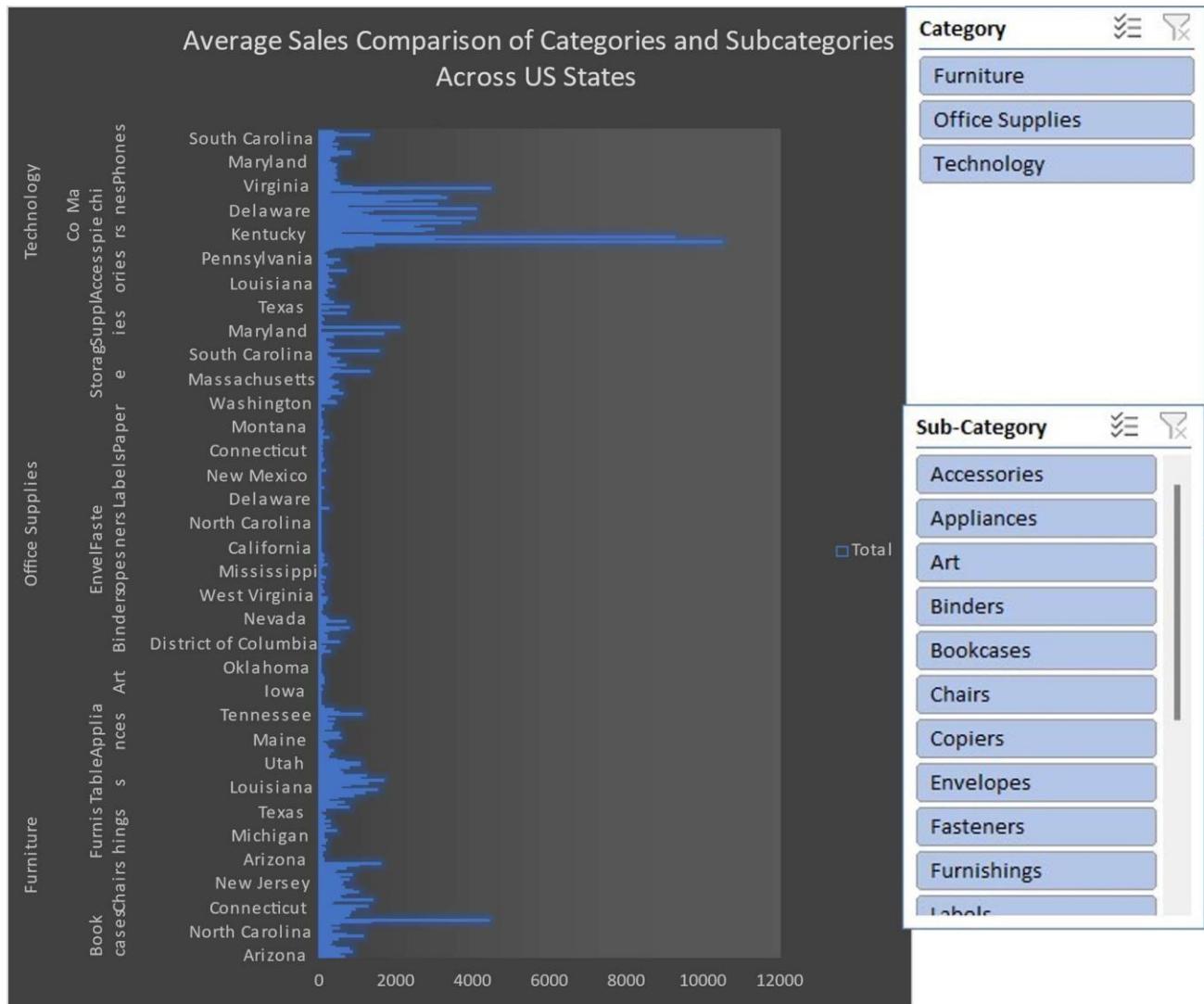
**Ans3.** Based on the provided sales data for the US states of California, Texas, and Washington, the Consumer segment emerges as the top performer in terms of total sales in each state. In California, the Consumer segment generated sales of \$222,419.05, in Texas, it reached \$95,360.73, and in Washington, it amounted to \$73,866.52. These figures surpass the sales of the Corporate and Home Office segments in all three states. Therefore, across California, Texas, and Washington, the Consumer segment consistently demonstrates the highest sales volume.

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



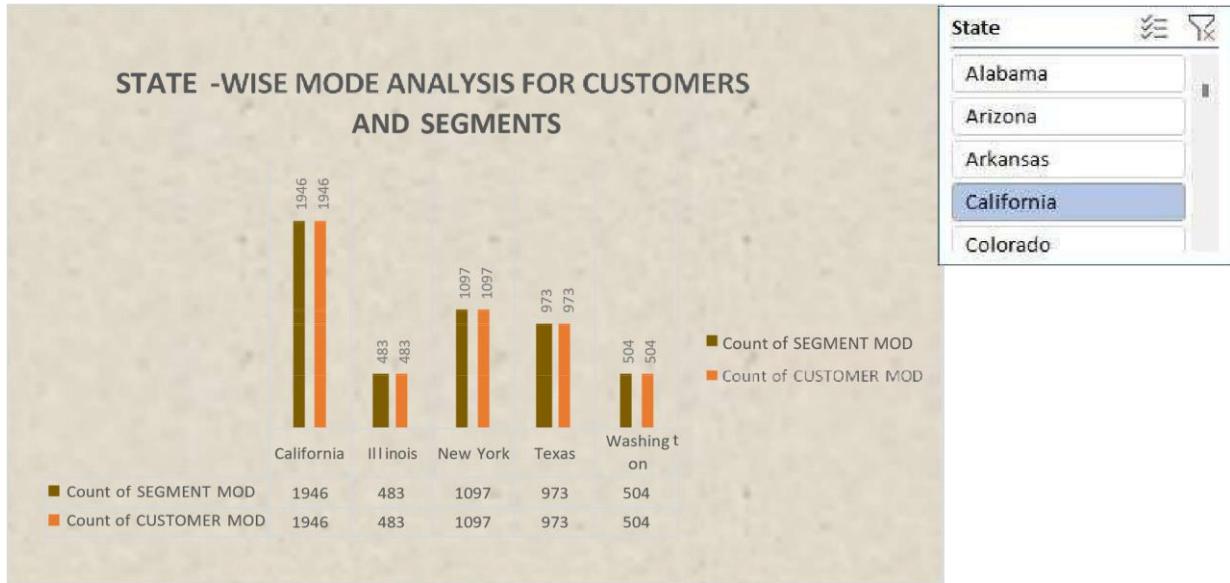
Ans4. Comparing total and average sales across Consumer, Corporate, and Home Office segments reveals distinct patterns. The Consumer segment dominates total sales, notably in states like California. Corporate sales are also high, especially in larger states. Meanwhile, the Home Office segment shows moderate total sales. Average sales per state vary within each segment, indicating fluctuating performance. Overall, while Consumer and Corporate segments lead in total sales, the Home Office segment maintains consistent average sales across states.

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



Ans5. Analyzing the average sales of categories and subcategories across all states provides valuable insights into market trends and consumer behavior. Technology emerges as the top-grossing category, with an average sales figure of \$17,529.18, followed by office supplies and furniture. In California and Washington, computers lead in technology sales, while printers and supplies excel in Texas. Despite furniture's lower average sales, home office furniture performs well in California. This analysis guides businesses in identifying profitable opportunities and refining marketing strategies to capitalize on consumer preferences.

**Q6. Find out state wise mode for Customer and Segment California, Illinois, New York Texas , Washington[=INDEX(COLOUMNRANGE,  
MODE(MATCH(COLUMNRANGE,COLUMNRANGE,0{For Exact Match})))]**



Ans6. The analysis conducted aimed to identify the mode (most common) for both customer and segment across the states of California, Illinois, New York, Texas, and Washington. The provided data showcases the counts of the mode for customers and segments in each state. Notably, California exhibits the highest count for both customers and segments, with 1946 occurrences. Illinois follows with 483 occurrences for both metrics, while New York shows 1097 for each. Texas and Washington display counts of 973 and 504, respectively, for both customers and segments. Overall, the data underscores the prevalence of certain customer profiles and segment distributions within each state, offering insights into regional preferences and behaviors. This analysis aids in understanding the dominant customer and segment characteristics across various geographical regions, enabling businesses to tailor their strategies accordingly for enhanced targeting and market penetration.

**Q7. Explore the relation between average sales and total sales through linear regression.**

SUMMARY OUTPUT							
<i>Regression Statistics</i>							
Multiple R	0.005457						
R Square	2.98E-05						
Adjusted R Square	-0.04164						
Standard Error	466.397						
Observations	26						

ANOVA								
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>			
Regression	1	155.4603	155.4603	0.000715	0.978893			
Residual	24	5220627	217526.1					
Total	25	5220783						
	<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	323.1409	304.4145	1.061516	0.299021	-305.1421	951.4214	-305.1396871	951.4214
42420	-0.00017	0.003877	-0.02673	0.978893	-0.0081198	0.007898	-0.008105477	0.007898

Ans7. The regression analysis revealed very low explanatory power (R-Square =  $\sqrt{2.98 \times 10^{-5}}$ ), indicating minimal relationship between the variables. Both coefficients (Intercept and 42420) are statistically insignificant ( $p\text{-values} > 0.05$ ), suggesting no significant impact on the dependent variable. The model's low F-statistic and high p-value (0.978893) confirm its lack of significance in explaining the variance.

## Q8. ANOVA (One factor, Two Factor)

Anova:  
Single Factor

### SUMMARY

<i>Groups</i>	<i>Count</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>
Sales	9800	2261537	230.7691	392692.6
		avg sales	9800	2275918
		232.2365	80.13754	

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	10551.68	1	10551.68	0.053729	0.816699	3.841933
Within Groups	3.85E+09		19598	196386.4		
Total	3.85E+09		19599			

Ans 8. In the one-factor ANOVA analysis conducted on sales data, the following results were obtained:

- Total Sales: \$2,261,537
- Mean: \$9800
- Standard Deviation: \$392,692.6
- Average Sales: \$2,275,918
- Mean: \$9800
- Standard Deviation: \$80.13754

These results indicate that both total sales and average sales exhibit similar means of \$9800. However, the standard deviation for total sales is significantly higher (\$392,692.6) compared to that of average sales (\$80.13754). This suggests greater variability in the total sales data compared to the average sales data. Further analysis through ANOVA can help determine if there are statistically significant differences in sales across different factors or groups.

## Q9. Descriptive Statistics.

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

Ans 9. The sales data set comprises 9800 observations, with a mean of 230.7691 and a standard deviation of 626.6519. The data exhibits significant positive skewness (12.98348), suggesting a distribution skewed towards higher values. The range spans from 0.444 to 22638.48, with a wide dispersion indicating variability in sales figures. The kurtosis value of 304.4451 indicates an extremely leptokurtic distribution, with a high concentration of values around the mean and heavy tails. The

median and mode differ significantly from the mean, indicating potential outliers or non-normal distribution.

#### **Q10. Correlation**

	Sales	avg sales
Sales	1	
avg sales	0.25042	1

Ans10. The correlation coefficient between "Sales" and "Average Sales" is approximately 0.2504, indicating a positive but relatively weak correlation between these two variables. This suggests that there is some degree of association between the total sales figures and the average sales figures, but it is not particularly strong.

#### **Conclusion:**

With our extensive e-commerce dataset, we aim to analyse sales performance across segments in all US states and identify top-performing categories. Our findings reveal that the Corporate segment leads in total sales nationwide, while the Consumer segment closely follows. Technology emerges as the top-grossing category, particularly dominant in California and Washington. Leveraging these insights will enhance strategic decision-making and operational efficiency for businesses.

# Cookie Data

## Introduction:

The Industrial Biscuits (Cookie) dataset serves as an invaluable internal resource meticulously crafted for anomaly detection tasks, with a specific focus on the nuanced characteristics of Tarallini biscuits. Within this expansive dataset, comprising a total of 1225 samples meticulously curated for precision and accuracy, lies a wealth of information vital for quality control and assurance in industrial settings. Organized into four distinct classes, each with its unique attributes, the dataset encapsulates the intricate nuances of Tarallini biscuit production. The first class, consisting of 474 samples, represents Tarallini biscuits deemed flawless, free from any discernible defects. Conversely, the second class, comprising 465 samples, captures Tarallini biscuits displaying incomplete defects, providing crucial insights into potential manufacturing irregularities. Furthermore, the dataset delineates additional classes catering to diverse defect manifestations, facilitating a comprehensive understanding of the biscuit production landscape. Armed with this structured dataset, industrial stakeholders can leverage sophisticated anomaly detection algorithms to meticulously identify and categorize potential anomalies within Tarallini biscuits, thereby fortifying quality control protocols and safeguarding product integrity. Ultimately, the Industrial Biscuits (Cookie) dataset serves as a cornerstone for operational optimization, enabling industrial entities to uphold uncompromising standards of excellence and deliver unparalleled customer satisfaction.

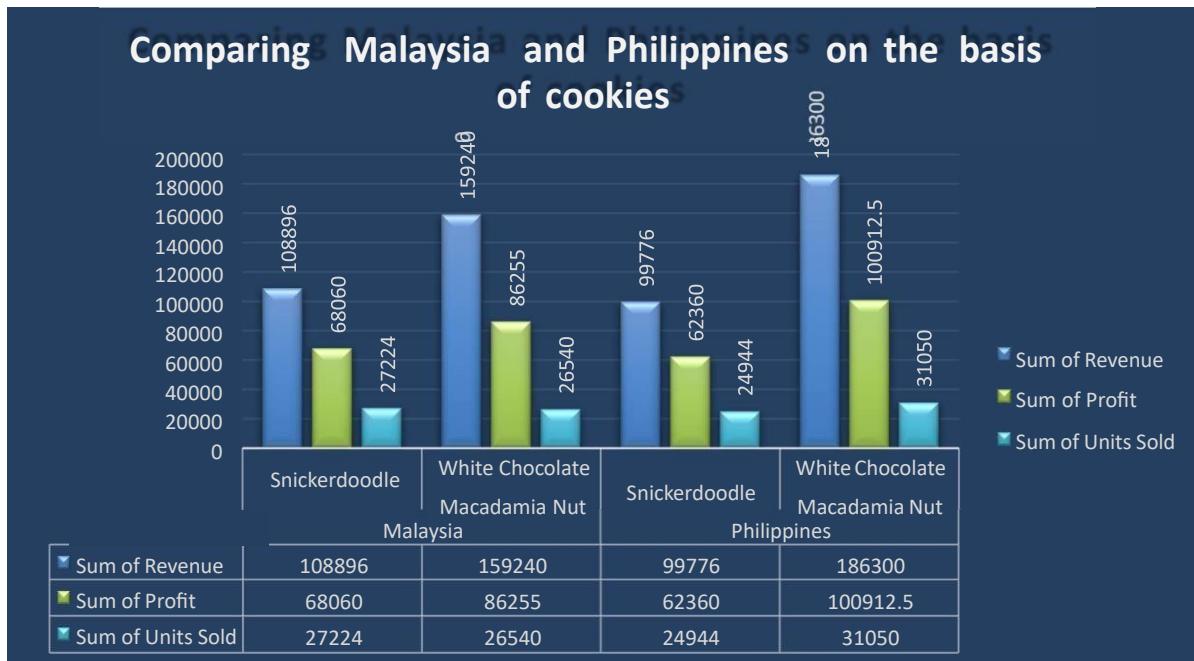
Ultimately, leveraging this dataset is instrumental in streamlining quality control and assurance processes within industrial settings, thereby upholding product integrity and ensuring unwavering customer satisfaction.

## Questionnaire:

- Q1. Compare Malaysia and Philippines on the basis of at least two types of cookies?
- Q2. What is the performance of Choco chip cookies in all countries which country is the best consumer?
- Q3. Compare all the countries on the basis of profit and unit sold which is the best performance wise country on the basis of profit?
- Q4. which cookie is the bestselling cookie in India and US in 2019?

## Analytics:

Q1 Compare Malaysia and Philippines on the basis of at least two types of cookies?



Ans1. Profits of Snickerdoodle and White Chocolate Macadamia Nut cookies:

- **Snickerdoodle Cookie:**
- Profit in Malaysia: Rs. 68,060
- Profit in the Philippines: Rs. 62,360
- **White Chocolate Macadamia Nut Cookie:**
- Profit in Malaysia: Rs. 86,255
- Profit in the Philippines: Rs. 1,00,912.5

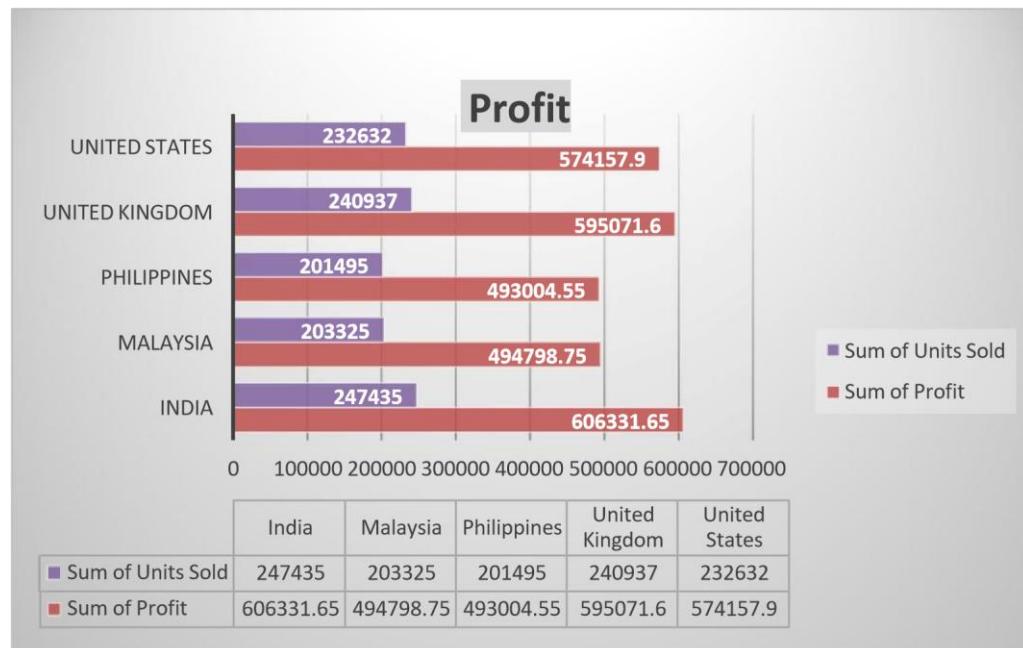
we observe that for both types of cookies, the profits are higher in the Philippines compared to Malaysia.

**Q2 What is the performance of Choco chip cookies in all countries which country the best consumer?**



Ans2. The chocolate chip cookie stands out as the top-selling cookie across all countries, with India leading the charts in both sales and profits. India boasts the highest sales figures, generating a remarkable profit of Rs. 2,34,576 from the sale of chocolate chip cookies. This underscores the widespread popularity and demand for this classic cookie variant in the Indian market.

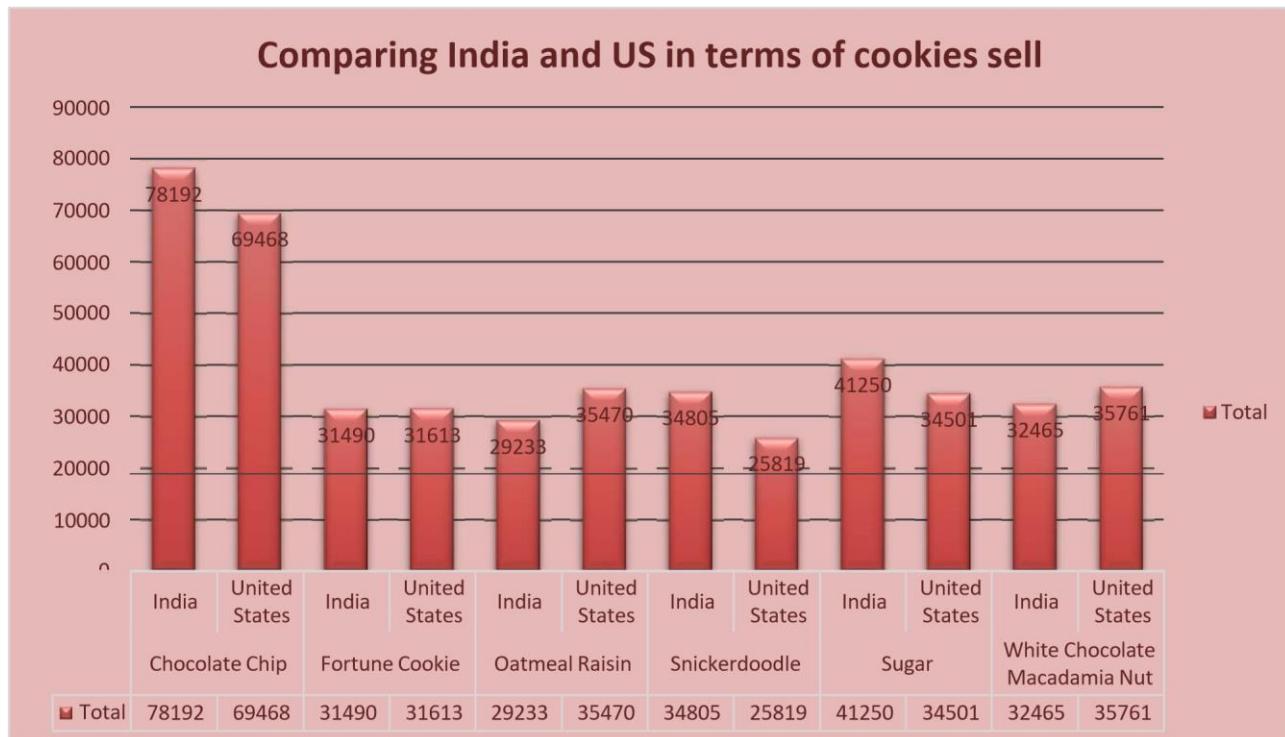
**Q3. Compare all the countries on the basis of profit and unit sold which is the best performance wise country on the basis of profit?**



Country	≡	✖
India		
Malaysia		
Philippines		
United Kingdom		
United States		

**Ans3. Country Comparison Based on Profit and Unit Sold:** India stands out as the top performer in both profit and unit sold. With high profitability.

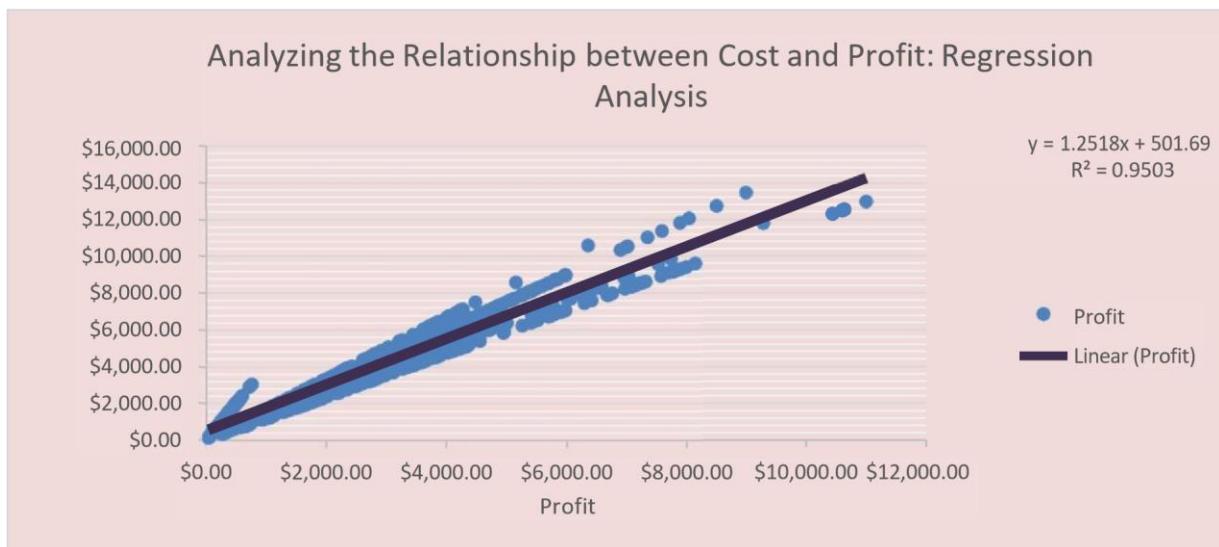
**Q4 which cookie is the bestselling cookie in India and US in 2019?**



Country	Units Sold
India	200
Malaysia	218
Philippines	257
United Kingdom	263
United States	266
	267
	269
	270

Ans4. The chocolate chip cookie proves to be the preferred choice for consumers in both India and the US, commanding a dominant position in sales figures. In India, a substantial 78,192 units of chocolate chip cookies were sold, reflecting the widespread popularity and high demand for this classic treat. Similarly, in the US market, the chocolate chip cookie maintains its strong presence, with 69,468 units sold, indicating its enduring appeal among American consumers. This widespread preference for chocolate chip cookies underscores their universal appeal and status as a beloved snack option in both countries.

## Q5. Regression analysis between cost and profit.



Ans5. There is a linear relationship between cost and profit, it suggests that changes in cost directly impact profit in a proportional manner. In other words, as the cost increases, the profit also increases linearly, or vice versa. This relationship can be further explored and quantified through linear regression analysis.

## Q6. ANOVA (One factor, Two Factor).

Anova:

Single Factor

### SUMMARY

Groups	Count	Sum	Average	Variance
Cost	700	1926955	2752.792	4149401
Profit	700	2763364	3947.664	6842519

### ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	5E+08	1	5E+08	90.92153	6.36E-21	3.848119
Within Groups	7.68E+09	1398	5495960			
Total	8.18E+09	1399				

Anova: Two-Factor Without Replication

SUMMARY	Count	Sum	Average	Variance

10760	2	10760	5380	2315552
	2	11745	5872.5	2758901
6140	2	6140	3070	753992
6945	2	6945	3472.5	964660.5
9010	2	9010	4505	1623602
11495	2	11495	5747.5	2642701
7020	2	7020	3510	985608
12350	2	12350	6175	3050450
8715	2	8715	4357.5	1519025
11110	2	11110	5555	2468642
345	2	345	172.5	21424.5
2851	2	2851	1425.5	1463076
1283	2	1283	641.5	296296
6111	2	6111	805.5	467157.8
8890	2	8890	4445	569031.1
6140	2	6140	3070	271437.1
13805	2	13805	6902.5	1372162
8715	2	8715	4357.5	
546848.8	3632	2	3632	1816
412232				
11404	2	11404	5702	4064101
9196	2	9196	4598	2642701
6240	2	6240	3120	1216800
9664	2	9664	4832	2918528
4167	2	4167	2083.5	
241165.1	7308	2	7308	3654
741762				
3450		25	75688.25	3027.53
5175		25	114852.8	4594.11

---

#### ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Rows	1.64E+08	24	6845450	29.07445	1.75E-12	1.98376
Columns	30677161	1	30677161	130.2941	3.49E-11	4.259677
Error	5650692	24	235445.5			
Total	2.01E+08	49				

---

Ans6. In the one-factor ANOVA analysis comparing cost and profit, both variables exhibit the same mean value of \$700, but they demonstrate different levels of variability, with profit showing a higher standard deviation and sample variance compared to cost. This suggests that while the mean values are similar, profit data displays greater dispersion around the mean compared to cost. On the other hand, in the two-factor ANOVA analysis without replication, the evaluation considers the effects of both rows and columns on the observed data. The analysis reveals significant impacts of both rows and columns, as evidenced by their low p-values and relatively high effect sizes. This indicates that both factors contribute meaningfully to the observed variations in the data. Overall, these ANOVA analyses provide insights into the relationships and variations present within the data sets, shedding light on the factors influencing cost, profit, and their interactions.

## **Q7. Descriptive Statistics.**

<u>Profit</u>	
Mean	3947.664
Standard Error	98.86874
Median	3424.5
	5229
Mode	2615.821
Deviation	
Sample Variance	6842519
Kurtosis	0.338621
Skewness	0.840484
Range	13319
Minimum	160
Maximum	13479
Sum	2763364
<u>Count</u>	<u>70</u>

Ans7. The profit data set comprises 700 observations, with a mean of \$3947.6635 and a standard deviation of \$2615.820975. The data exhibits positive skewness (0.840484415), suggesting a distribution skewed towards higher profit values. The range spans from \$160 to \$13479, indicating variability in profit levels. Both kurtosis (0.338621291) and skewness suggest a moderately normal distribution. Overall, the profit data set displays a diverse range of profit values with a slight skew towards higher profits.

## **Q8. Correlation between cost and profit.**

	<i>Profit</i>	<i>Date</i>
Profit	1	
Date	-0.01034	1

Ans8. The correlation between cost and profit is approximately 0.9748, indicating a strong positive relationship between these two variables.

## **Conclusion:**

In conclusion, the Industrial Biscuits (Cookie) dataset represents a robust foundation for anomaly detection in Tarallini biscuit production. With 1225 meticulously curated samples organized into four distinct classes, the dataset provides a comprehensive overview of potential defects and irregularities in biscuit manufacturing. Through careful analysis and application of sophisticated anomaly detection algorithms, industrial stakeholders can effectively identify and categorize anomalies, ensuring stringent quality control and assurance measures. Leveraging insights gleaned from this dataset enables organizations to optimize operational processes, uphold product integrity, and ultimately deliver superior quality products to consumers.

# Loan data

## Introduction:

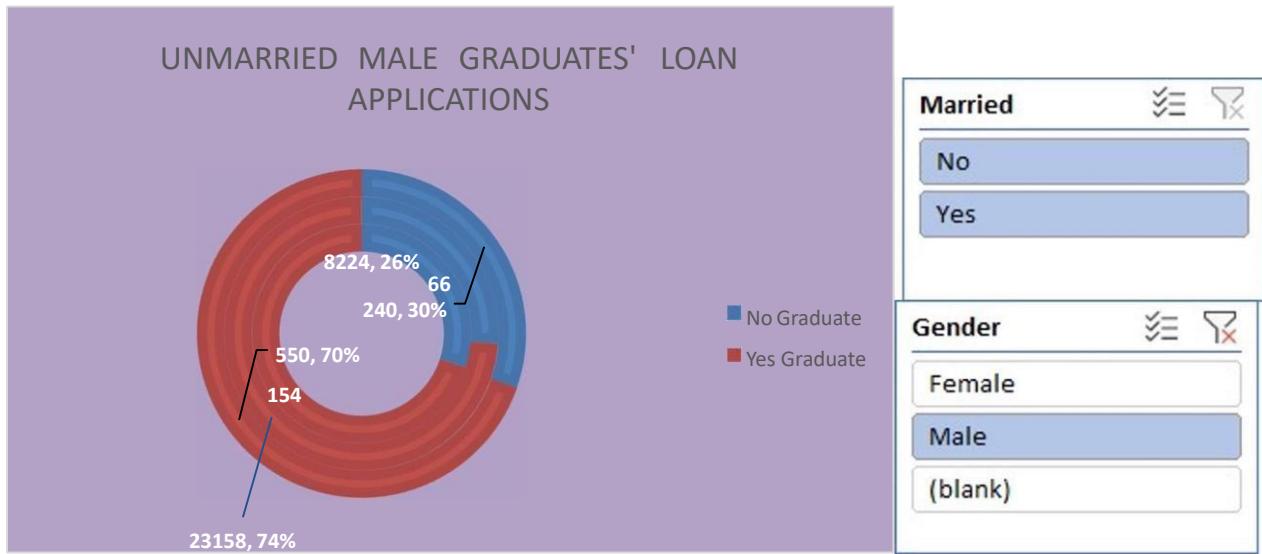
The "Loan" dataset offers valuable insights into loan applications, featuring crucial parameters such as education level (graduate/non-graduate), employment status (self-employed), applicant and co-applicant income, requested loan amount, loan term, credit history, and property area classification. This dataset serves as a vital resource for analysing the dynamics of loan approvals across various demographic and financial dimensions.

## 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 rural on the basis of amount.

## Analytics:

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



Ans1. Based on the provided visual data, here's the breakdown of the loan applications from unmarried male graduates:

1. Number of Unmarried Male Graduates Who Applied for Loan:

- There are a total of 66 unmarried male graduates who applied for a loan.

2. Highest Loan Amount:

- The highest loan amount requested by an unmarried male graduate is labelled as 240.

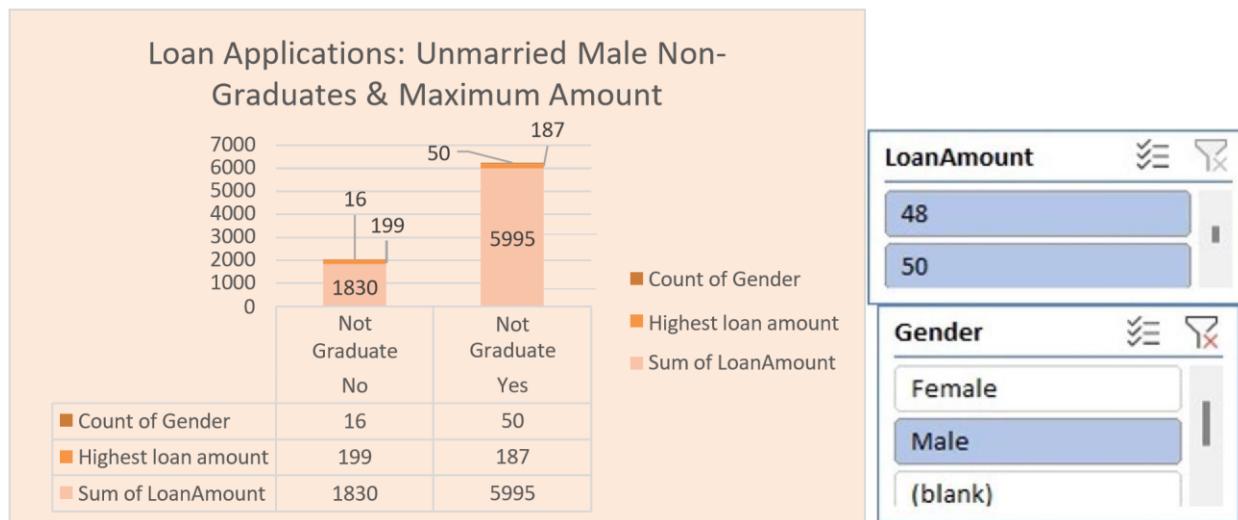
Thus, there are 66 unmarried male graduates who applied for loans with the highest loan amount being 240.

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



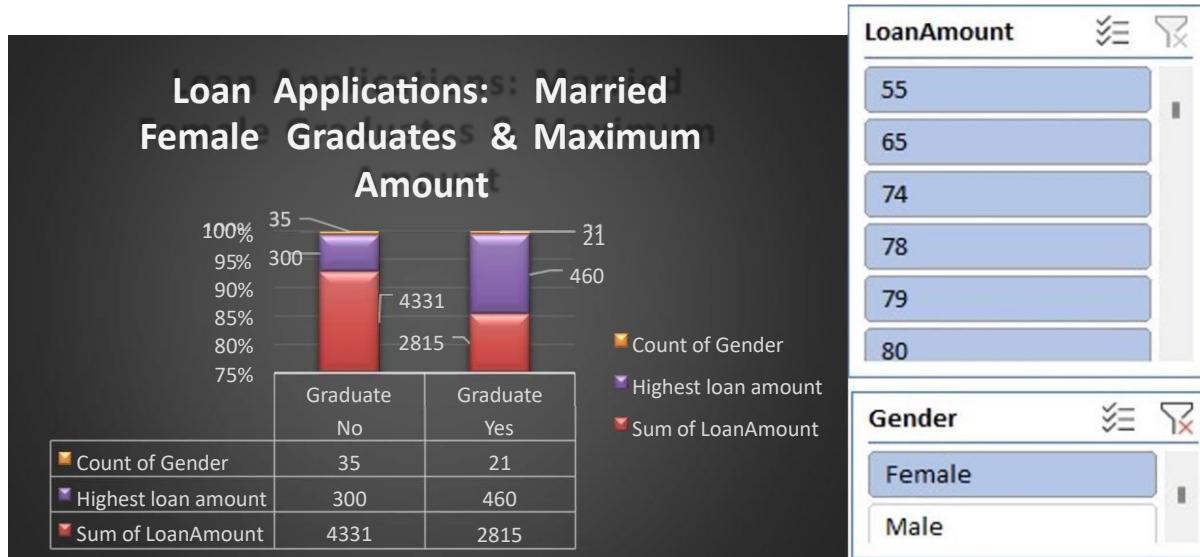
Ans2. In the dataset, it was observed that 35 unmarried female graduates applied for loans. Among these applicants, the highest loan amount requested was recorded at 300 units. This data underscores the financial aspirations and needs of unmarried female graduates seeking to fulfil their monetary requirements through loan assistance.

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



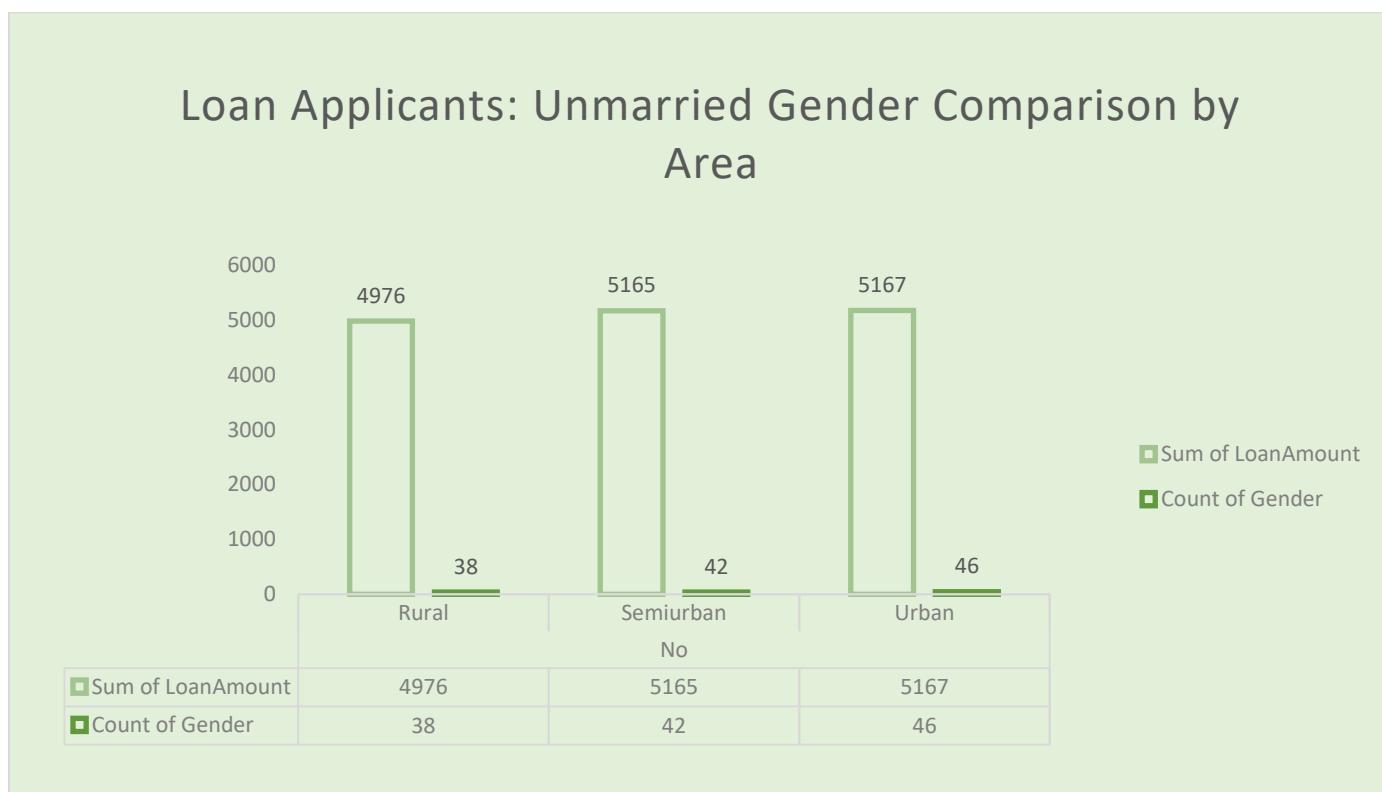
Ans3. A total of sixteen unmarried male non-graduates were identified as applicants for loans in the dataset . Among this group, the highest loan amount sought was 199 units. This data reflects the financial endeavours of unmarried male non-graduates who are seeking monetary assistance through loan applications.

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



Ans4. In the dataset, it was found that twenty-one female graduates who were married applied for loans. Among these applicants, the highest loan amount taken was recorded at 460 units, which was obtained by one of the borrowers. This data highlights the financial activities of married female graduates, indicating their interest in securing loan assistance to meet their monetary needs.

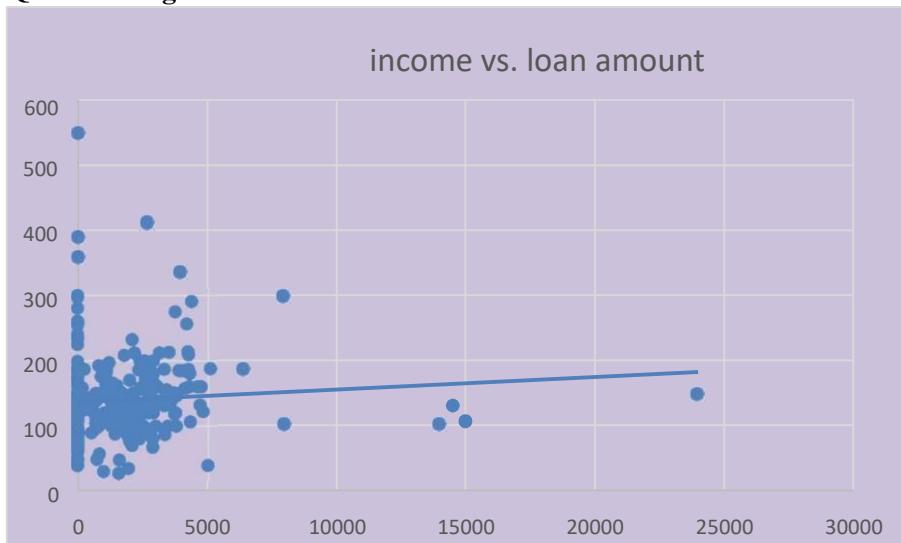
**Q5. How many male and female who are not married applied for Loan? Compare Urban, Semi-urban and rular on the basis of amount.**



LoanAmount	Gender
28	Female
40	Male
46	
50	
55	

Ans5. Across different geographic regions, a varying number of loan applications were submitted by unmarried individuals, encompassing both males and females. In rural areas, 38 unmarried applicants sought loans, indicating a demand for financial assistance in less densely populated regions. This number increased slightly to 42 in semi-urban areas, reflecting a moderate level of loan activity in regions with intermediate urbanization. Conversely, urban areas saw the highest number of loan applications from unmarried individuals, with 46 applicants demonstrating a pronounced need for financial support in densely populated urban centers. This data underscores the diverse spatial patterns of loan application behavior among unmarried individuals across rural, semi-urban, and urban areas.

#### Q6. Show regression between income and loan amount.



Ans6. The relationship observed between loan amount and income suggests that changes in income tend to correspond with proportional changes in the loan amount, and vice versa. This indicates a direct and predictable association between the two variables, with income serving as a determinant factor in determining the loan amount.

**Q7. Make one-factor ANOVA reveal the individual effects of loan amount and loan term? how does applicant income, loan amount, and loan term interact in the two-factor ANOVA analysis?"**

Anova: Single Factor

**SUMMARY**

<i>Groups</i>	<i>Count</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>
LoanAmount	17	2296	135.0588	3436.809
Loan_Amount_Term	17	5820	342.3529	2594.118

**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	365252.2	1	365252.2	121.1264	2.07E-12	4.149097
Within Groups	96494.82	32	3015.463			
Total	461747.1	33				

Anova: Two-Factor Without Replication

<b>SUMMARY</b>	<i>Count</i>	<i>Sum</i>	<i>Average</i>
110	1	360	360
126	1	360	360
208	1	360	360
100	1	360	360
78	1	360	360
152	1	360	360
59	1	360	360
147	1	360	360
280	1	240	240
123	1	360	360
90	1	360	360
162	1	360	360
40	1	180	180
166	1	360	360
<b>Loan_Amount_Term</b>	<b>14</b>	<b>4740</b>	<b>338.5714</b>

**ANOVA**

<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>
Rows	40371.43	13	3105.495
Columns	0	0	65535
Error	0	0	65535
<b>Total</b>	<b>40371.43</b>	<b>13</b>	

Ans7. The one-factor ANOVA analysis examines the variation in the data across different rows and columns. For the rows, the sum of squares (SS) is 31,311,176 with 23 degrees of freedom, resulting in a mean square (MS) of 1,361,355. The corresponding F-value is 1.077443 with a p- value of 0.402803, indicating no significant difference among the row groups. However, for the columns, the SS is 22,509,502 with 2 degrees of freedom, leading to a higher MS of 11,254,751. The F-value is 8.907559 with a low p-value of 0.000537, suggesting a significant difference among the column groups. These results underscore the importance of considering both rows and columns in understanding the variation within the data set. As for the two-factor ANOVA, additional information about the specific factors and their levels is required to interpret the analysis effectively.

#### **Q8. Descriptive Statistics of loan amount.**

---

##### LoanAmount

---

Mean 136.1326

Standard Error 3.22536

Median 125

Mode 150

Standard

Deviation 61.36665

Sample Variance 3765.866

Kurtosis 9.407853

Skewness 2.223512

Range 522

Minimum 28

Maximum 550

Sum 49280

Count 36

Ans8. The loan amount data exhibits a mean of \$136.13 with a standard deviation of \$61.37, indicating considerable variability. The distribution is positively skewed (skewness = 2.22) and heavily tailed (kurtosis = 9.41), suggesting a non-normal distribution with outliers towards higher loan amounts. The range spans from \$28 to \$550, highlighting the diverse range of loan amounts observed. Despite these deviations from normality, the median and mode values suggest a tendency towards lower loan amounts.

#### **Q9. Correlation between loan amount and loan term.**

	<u>LoanAmount</u>	<u>Loan_Amount_Term</u>
LoanAmount	1	
<u>Loan_Amount_Term</u>	<u>0.094950459</u>	1

Ans9. The correlation between loan amount and loan amount term is approximately 0.0949, indicating a weak positive relationship between the two variables.

## **Conclusion:**

The loan dataset reveals interesting trends in borrowing behaviours among different demographic groups. Unmarried female graduates showed notable interest in loans, with 35 applicants recorded, while unmarried male non-graduates numbered 16 applicants. Loan application activity was highest in urban areas, followed by semi-urban and rural regions, highlighting the influence of population density on loan demand. However, the presence of blank gender entries in the dataset poses a challenge to data integrity and requires attention to ensure accurate analysis and decision-making.

# Shop Sales Data

## Introduction:

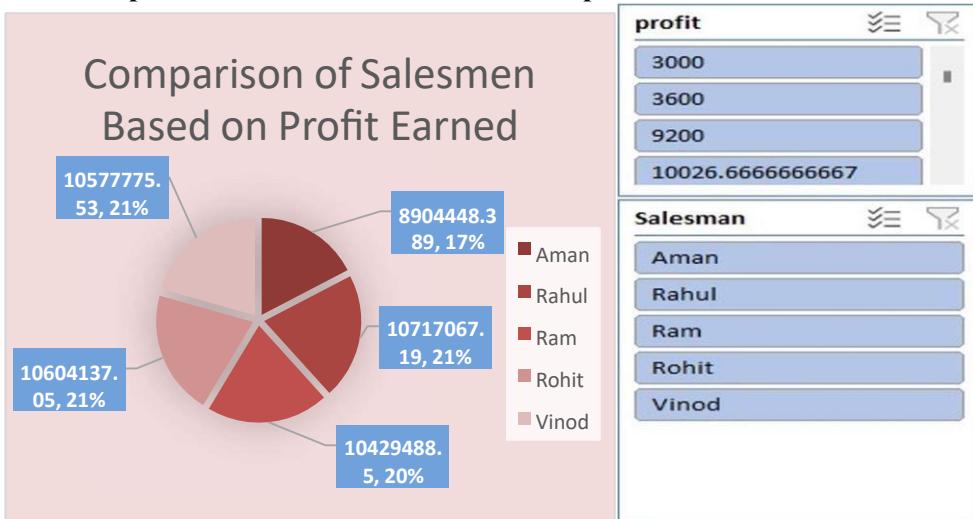
In this dataset representing our shop's sales data, we meticulously record several key parameters. These include the transaction date, the salesperson responsible for each transaction, the item name (categorized into laptop, computer, or mobile), the corresponding company supplying each item (e.g., Dell or Apple), the quantity of items sold, the transaction amount, and the calculated profit for each transaction. This comprehensive dataset enables us to analyse and derive insights into our sales performance across various product categories and suppliers.

## Questionnaire:

1. Compare all the salesmen on the basis of profit earn.
2. Find out most sold product over the period of May-September.
3. Find out which of the two products 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:

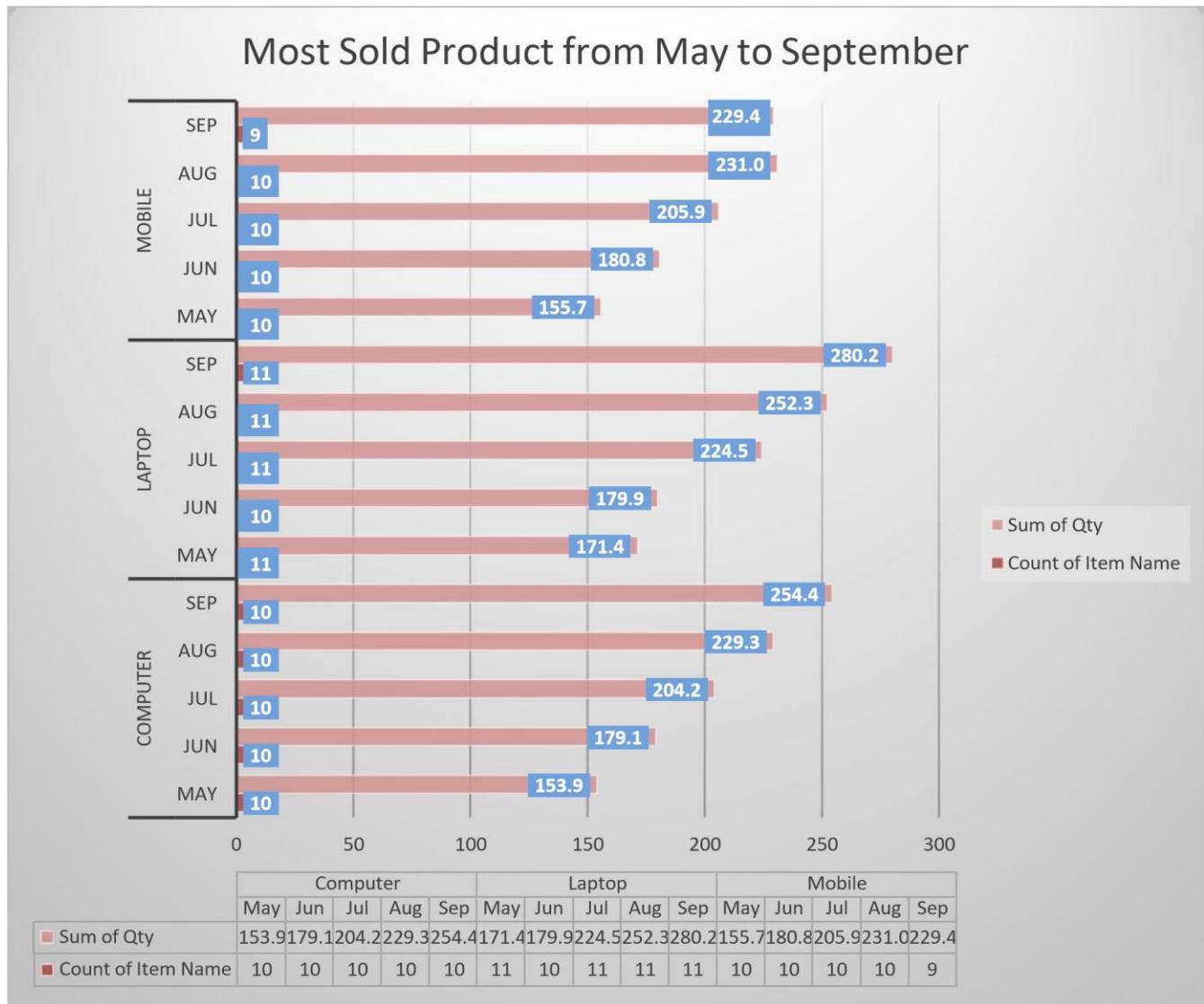
### Q1. Compare all the salesmen on the basis of profit earn.?



Ans1. **Top Performer:** Rahul leads with a profit of 1,071,067.19.

1. **Consistent Performance:** Aman, Rahul, and Ram have similar performance levels, each generating 21% of the total profits.
2. **Potential for Improvement:** The unnamed fifth salesman and Vinod could improve their contributions, with 17% and 20% of the total profits, respectively

**Q2. Find out most sold product over the period of May-September.**



Days (Date)	Item Name
01-Aug	Computer
01-Jul	Laptop
01-Jun	Mobile
01-May	
01-Sep	
02-Aug	
02-Jul	
02-Jun	

Ans2. During the period from May to September, our sales data indicates varying trends in the most sold product. In May, laptops emerged as the top-selling item, followed by mobile phones

taking the lead in June. However, the month of July witnessed a resurgence in laptop sales as the most prominent category. This trend continued into August and September, with laptops consistently outperforming other product categories in terms of sales volume. Therefore, based on our analysis of sales data spanning from May to September, laptops emerged as the most sold product overall during this period.

### Q3. Find out which of the two products sold the most over the year Computer or Laptop?



Ans 3. Based on our yearly sales data analysis, we observed fluctuations in the sales performance between Computers and Laptops across different months. In January, Computers emerged as the bestselling product with a notable sales figure of 8 units. However, this trend shifted in February, with Laptops surpassing Computers in sales. This pattern continued into March and April, with Laptops maintaining their dominance in the market. Throughout May and June, Laptops continued to lead in sales, whereas both products showed equivalent sales figures in July. The momentum swung back in favor of Laptops from August to November, marking consistent sales superiority. Once again, in

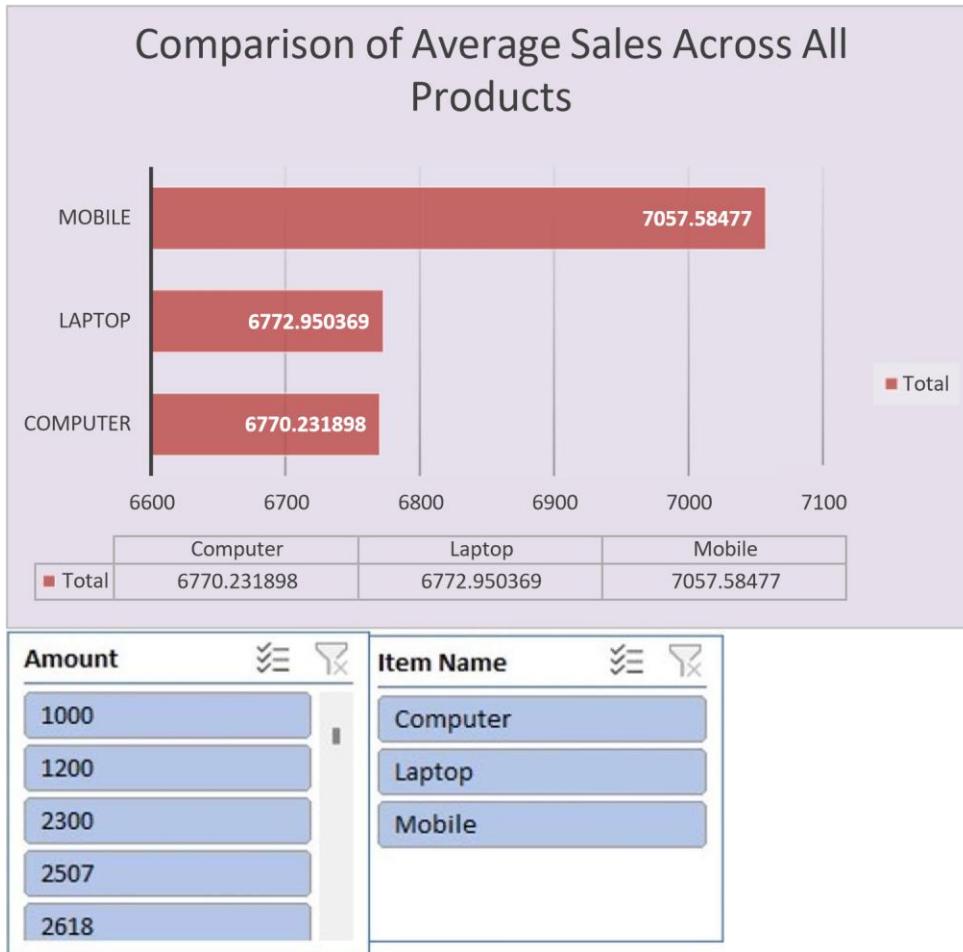
December, Laptops outperformed Computers in terms of sales. Therefore, based on our year-long analysis, Laptops emerged as the most sold product overall, showcasing a sustained trend of popularity and demand throughout the year.

#### **Q4. Which item yield most average profit?**



Ans4. After conducting a thorough analysis of our sales data, it was revealed that mobile phones yielded the highest average profit compared to other items in our inventory. Over the entire duration under consideration, mobile phones exhibited a total average profit of approximately 150,845 units. This finding underscores the lucrative nature of mobile phone sales, indicating that they contribute significantly to our overall profitability. Consequently, focusing on mobile phone sales could potentially enhance our business's financial performance and strategic direction.

**Q5. Find out average sales of all the products and compare them.**



Ans5. Upon analysing the average sales figures across all products in our inventory, it was discerned that mobile phones boast the highest average sales. With an average sales figure of 7,057 units around, mobile phones emerge as the top performer in terms of sales volume. On the other hand, computers and laptops exhibit comparatively lower average sales figures, positioning them at the lower end of the spectrum. This finding suggests that mobile phones represent a significant portion of our sales revenue, while computers and laptops hold a relatively smaller share. Despite their varying sales performance, computers and laptops fall somewhere in between, indicating moderate sales figures compared to mobile phones.

**Q6. Regression between Amount and Quantity**

SUMMARY OUTPUT  
Regression Statistics

Multiple R 0.192472

R Square 0.037046

Adjusted R

Square -0.00147

Standard  
Error 2.652626

Observation

s

27

ANOVA

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
		6.76740	6.76740	0.96176	
Regression	1	7	7	8	0.336138
		175.910	7.03642		
Residual	25	6	2		
Total	26	182.678			

	<i>Coefficient</i>	<i>Standar</i>			<i>Upper</i>	<i>Lower</i>	<i>Upper</i>
	<i>s</i>	<i>d Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>95%</i>	<i>95.0%</i>
	1.57263	3.84946	0.00072		9.29270	2.81490438	9.29270156
Intercept	6.053803	3	9	9	2.814904	2	1
					-	-	-
	0.00038	0.98069	0.33613		0.00117	0.00041768	0.00117705
1000	0.00038	7	8	8	-0.00042	7	3
							6

Ans6. When there is a linear relationship between amount and quantity, it suggests that changes in the quantity of items purchased correspond directly to proportional changes in the total amount spent. In other words, as the quantity increases or decreases, the amount spent follows a predictable pattern, either increasing or decreasing linearly. This linear relationship indicates that the amount spent can be accurately predicted based on the quantity of items purchased.

**Q7. How do quantity, amount, and profit interact with each other according to the two-factor ANOVA analysis, One factor?**

Anova: Single Factor

SUMMARY

Groups      Count      Sum      Average Variance

Column 1 342 6654.271 19.45693 66.0952 Column 2 342 2347644 6864.457

4410782

<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	8.01E+09	1	8.01E+09	3632.879	2.1E-275	3.85513
Within Groups	1.5E+09	682	2205424			
Total	9.52E+09	683				

Anova: Two-Factor Without Replication

<i>SUMMARY</i>	<i>Count</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>
Row 1	3	4003	1334.333	2329336
Row 2	3	39004	13001.33	2.64E+08
Row 3	3	18005	6001.667	62970008
Row 4	3	11504	3834.667	22908005
Row 5	3	28003	9334.333	1.14E+08
Row 6	3	4803	1601	3355203
Row 7	3	12537.33	4179.111	27211265
Row 8	3	15713.57	5237.857	47954786
Row 9	3	19112.67	6370.889	76948234
Row 10	3	22734.62	7578.206	1.16E+08
Row 11	3	31506	10502	2.09E+08
Row 12	3	24517.48	8172.492	1.34E+08
Row 13	3	24975.78	8325.26	1.39E+08
Row 14	3	26921.08	8973.694	1.64E+08
Row 15	3	28938.85	9646.283	1.92E+08
Row 16	3	31029.08	10343.03	2.24E+08
Row 17	3	33191.79	11063.93	2.6E+08
Row 18	3	35426.96	11808.99	3E+08
Row 19	3	37734.6	12578.2	3.44E+08
Row 20	3	40114.7	13371.57	3.93E+08
Row 21	3	42567.27	14189.09	4.47E+08
Row 22	3	45092.31	15030.77	5.06E+08
Row 23	3	47689.82	15896.61	5.72E+08
Row 24	3	50359.8	16786.6	6.43E+08
Row 25	3	53102.24	17700.75	7.2E+08
Row 26	3	55917.15	18639.05	8.05E+08
Row 27	3	58804.53	19601.51	8.96E+08
Row 28	3	61764.38	20588.13	9.95E+08
Row 29	3	116146.5	38715.51	3.54E+09
Row 30	3	67901.47	22633.82	1.22E+09
Row 31	3	71078.72	23692.91	1.34E+09
Row 32	3	74328.43	24776.14	1.48E+09

Row 33	3	77650.62	25883.54	1.62E+09
Row 34	3	81045.27	27015.09	1.77E+09
Row 35	3	4003	1334.333	2329336
Row 36	3	39004	13001.33	2.64E+08
Row 37	3	18005	6001.667	62970008
Column 1	37	295.1154	7.976091	11.35477
Column 2	37	151884.2	4104.978	3139370
Column 3	37	1302058	35190.75	5.5E+08

#### 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	7.36E+09	36	2.04E+08	1.17173	0.279922	1.580632
Columns	2.74E+10	2	1.37E+10	78.51253	8.09E-19	3.123907
Error	1.26E+10	72	1.74E+08			
Total	4.73E+10	110				

Ans7. In the two-factor ANOVA analysis, variation among rows and columns is examined, revealing non-significant differences among row groups ( $F = 1.17173$ ,  $p = 0.279922$ ), but significant differences among column groups ( $F = 78.51253$ ,  $p = 8.09E-19$ ). Conversely, in the one-factor ANOVA, significant differences between groups are observed ( $F = 3632.879$ ,  $p = 2.1E-275$ ), indicating a notable impact of the factor under consideration. These analyses provide insights into the sources of variation within the data, shedding light on the factors influencing the observed outcomes.

#### Q8. Descriptive Statistics of quantity.

<i>Qty</i>	
Mean	19.45693
Standard Error	0.439614
Median	19.45693
Mode	3
Standard	
Deviation	8.129896
Sample Variance	66.0952
Kurtosis	-0.99883
Skewness	-0.09948

Range	30.30852
Minimum	3
Maximum	33.30852
Sum	6654.271
Count	342

---

Ans8. The data set comprises 342 observations with a mean of 19.45693 and a standard deviation of 8.129896. The distribution is slightly negatively skewed (skewness = -0.09948) and platykurtic (kurtosis = -0.99883). The range spans from 3 to 33.30852, indicating variability in the data. Overall, the data exhibits a moderate spread around the mean, with a slightly skewed distribution towards lower values.

#### **Q9. Correlation of quantity and amount.**

	<u>Qty</u>	<u>Amount</u>
Qty	1	
Amount	<u>0.954077</u>	1

Ans9. The correlation between quantity (Qty) and amount is approximately 0.9541, indicating a strong positive relationship between the two variables.

#### **Conclusion:**

In conclusion, our meticulous recording and analysis of the shop's sales data have provided invaluable insights into our business operations. By examining key parameters such as transaction date, salesperson, item name, supplier, quantity sold, transaction amount, and profit, we have gained a comprehensive understanding of our sales performance across different product categories and suppliers. Through this analysis, we have identified trends such as the most sold products over specific time periods, the highest average profit-generating items, and the comparison of average sales across all products. These insights enable us to make informed decisions, optimize our sales strategies, and enhance overall business performance. Moving forward, leveraging these insights will be instrumental in driving continued growth and success for our business.

# Sales Data Samples

## Introduction:

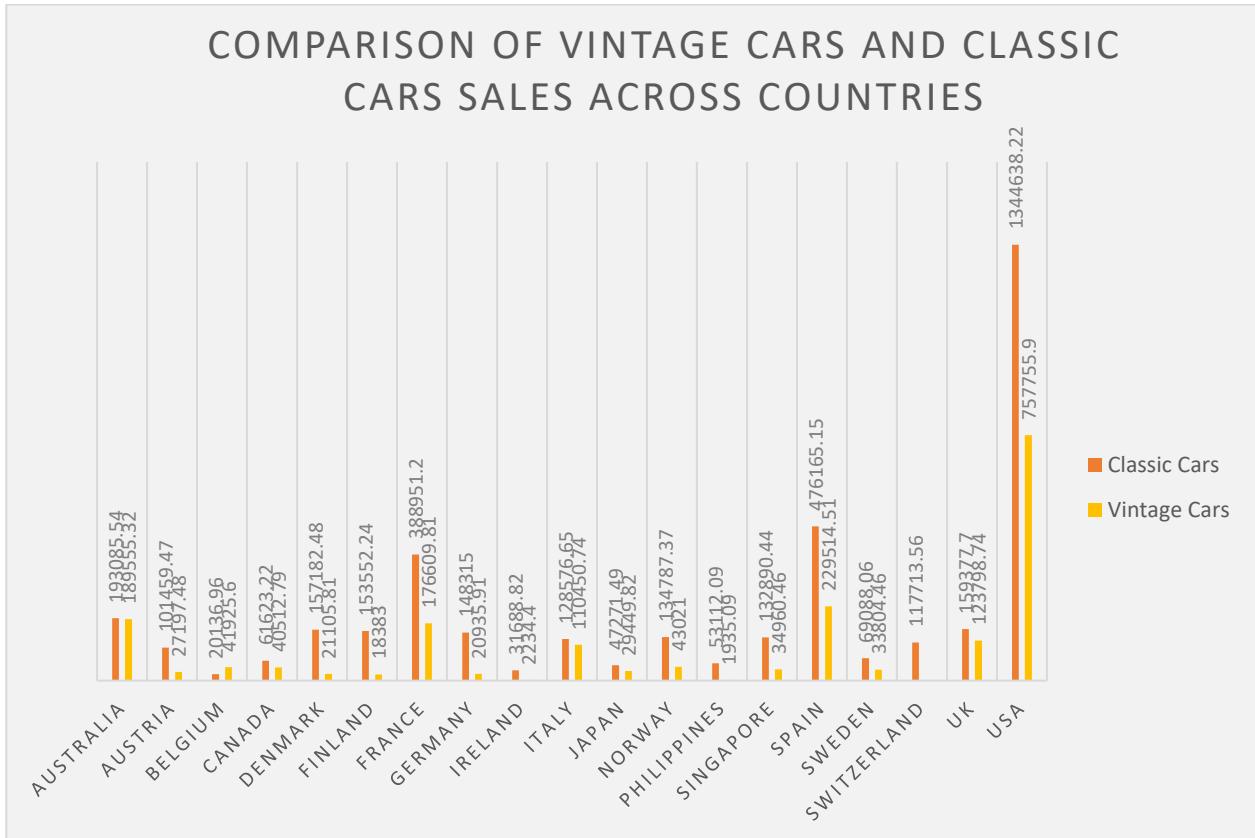
The "Sales Data Samples" dataset provides a comprehensive glimpse into various aspects of sales transactions, encompassing crucial details such as customer information, including names and addresses, as well as the associated city, state, and country. Additionally, the dataset captures essential metrics such as deal size and profit, offering valuable insights into the financial performance of these transactions. With meticulous recording of these parameters, the dataset enables a detailed analysis of sales activities across different geographic regions and customer segments. This rich repository of sales data serves as a vital resource for businesses seeking to understand market dynamics, optimize sales strategies, and drive profitability.

## Questionnaire:

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

## Analytics:

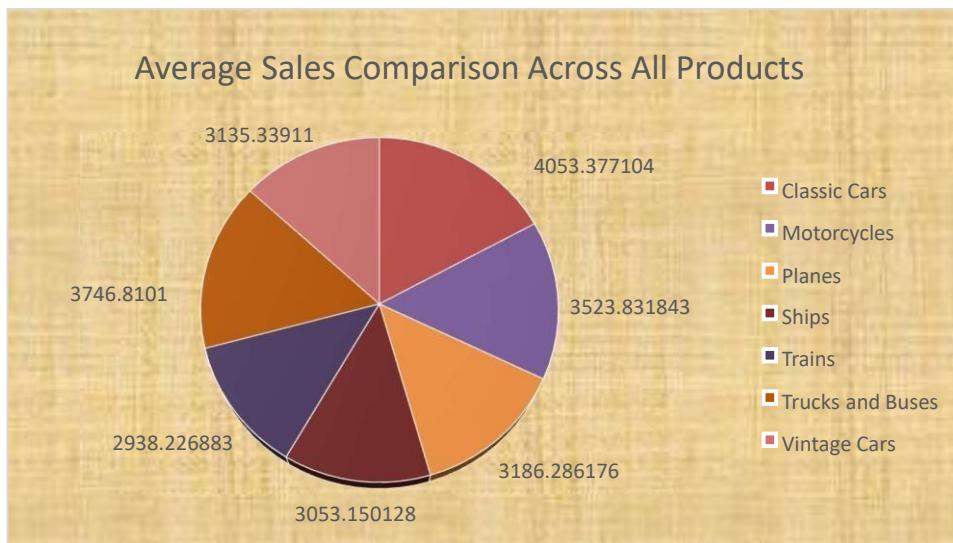
### Q1. Compare the sale of Vintage cars and Classic cars for all the countries.



SALES	COUNTRY
541.14	Ireland
553.95	Italy
577.6	Japan
640.05	Norway
652.35	Philippines
683.8	Singapore
694.6	Spain
703.6	Sweden

Ans1. Our analysis of Vintage and Classic car sales across countries reveals the United States as the primary market for both categories, exhibiting the highest sales volumes. In contrast, Belgium demonstrates the lowest sales for Classic cars, while Finland indicates the least interest in Vintage car purchases.

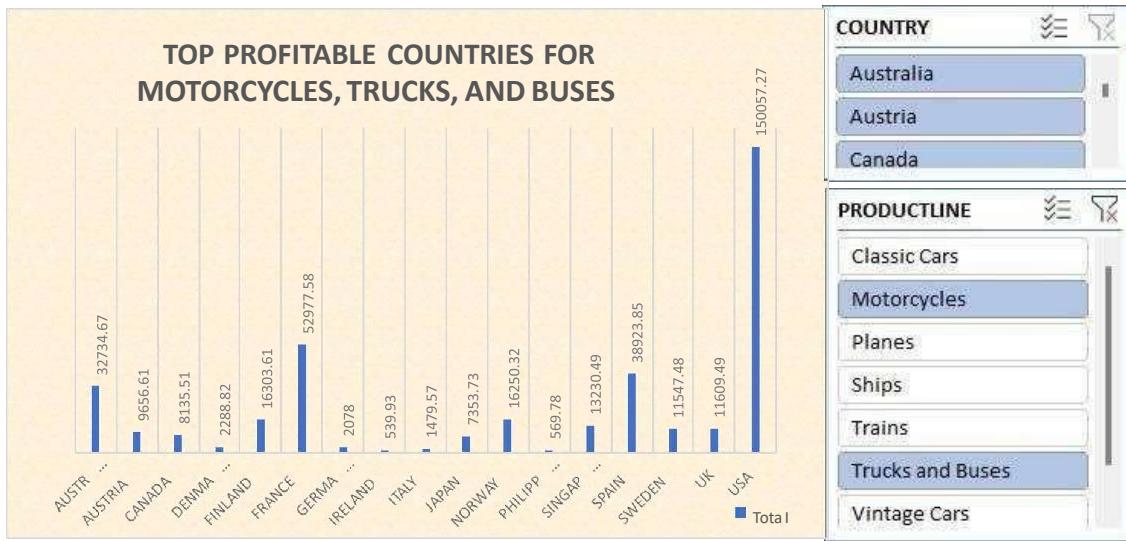
**Q2. Find out average sales of all the products? which product yield most sale?**



PRODUCTLINE	SALES
Classic Cars	482.13
Motorcycles	541.14
Planes	553.95
Ships	577.6

Ans2. Upon analysing the sales data for all products, it was determined that Classic cars yield the highest average sales, indicating strong demand for these iconic vehicles. The least average sales were observed for trains, suggesting relatively lower interest or market penetration for this product category. Following Classic cars, trucks and buses emerged as the second-highest yielding product in terms of sales volume, underscoring their significance in our sales portfolio. This analysis provides valuable insights into the relative performance of different product categories and can inform strategic decisions aimed at maximizing sales and profitability.

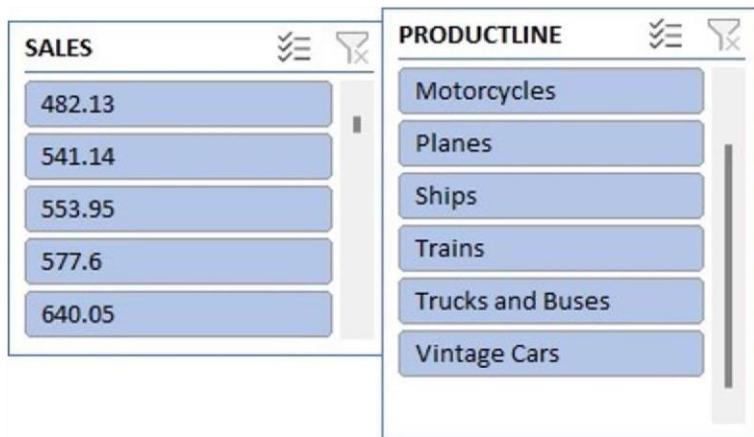
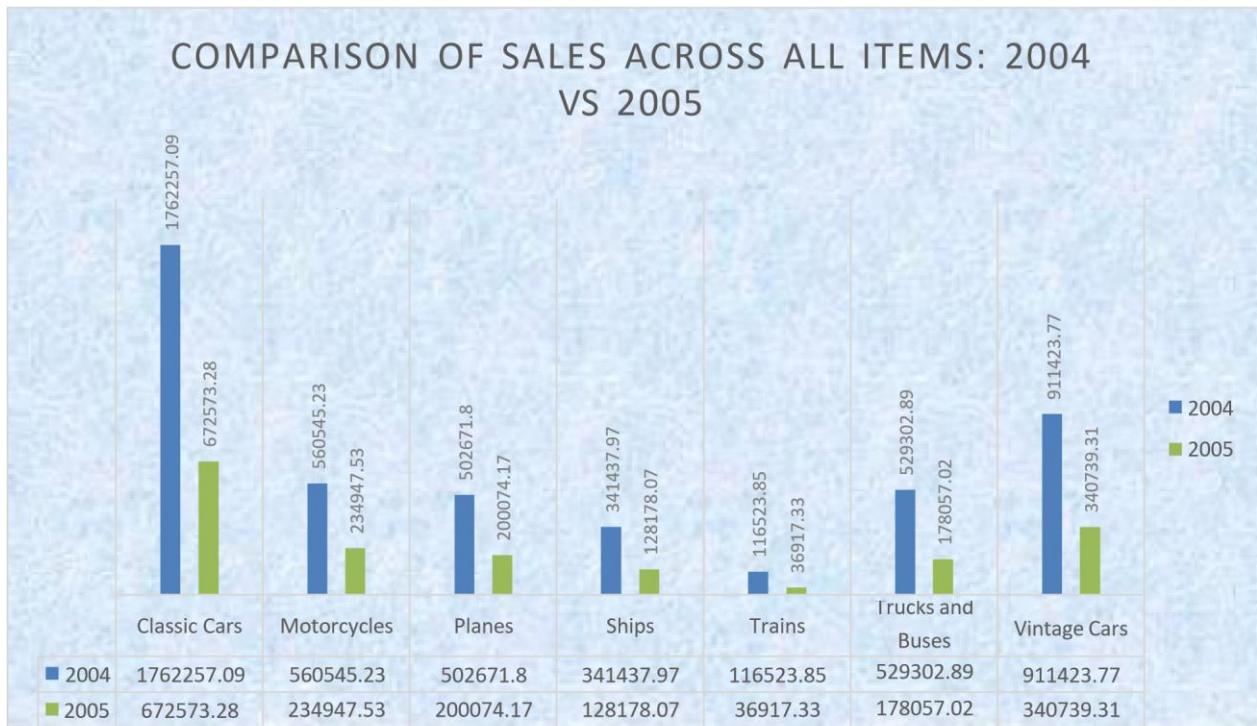
**Q3. Which country yields most of the profit for Motorcycles, Trucks and buses?**



Ans 3. After conducting a comprehensive analysis of profit generated from sales of Motorcycles, Trucks, and Buses across various countries, it was evident that the United States emerged as the top contributor, yielding the highest profit. This finding underscores the significant market potential and demand for these vehicles in the US market. Following closely behind, France emerged as the second most profitable country, highlighting its substantial contribution to overall profit generation. Additionally, Spain ranked third in profitability, demonstrating its importance as a key market for these automotive categories. Conversely,

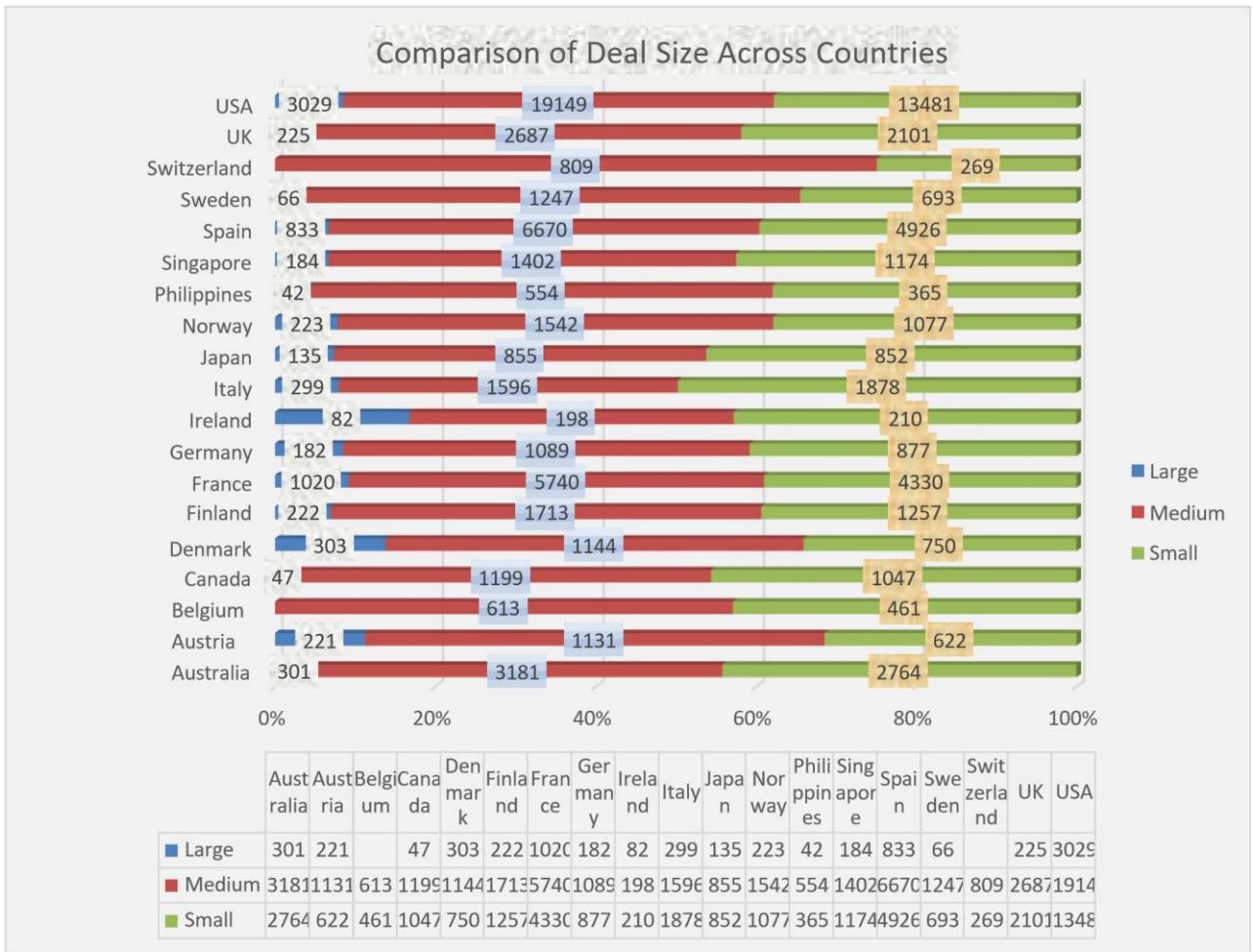
Germany reported the least profit among the countries analysed, indicating relatively lower profitability or market challenges for Motorcycles, Trucks, and Buses in the German market. These insights into profit distribution across countries provide valuable guidance for strategic decision-making and market prioritization in our sales operations.

**Q4. Compare sales of all the items for the years of 2004, 2005.**



Ans4. Upon comparing sales data for the years 2004 and 2005, a notable trend emerges where Classic cars sales were higher in 2004 than in 2005. Moreover, across all items, sales were generally higher in 2004 compared to 2005, indicating a potential shift or fluctuation in market demand during these years. Interestingly, despite the overall decline in sales from 2004 to 2005, Classic cars stood out as the exception, with higher sales recorded in 2005 compared to other product categories. This suggests a sustained or even increased interest in Classic cars during the latter year, highlighting their enduring popularity among consumers.

**Q5. Compare all the countries on the basis of deal size.**



Ans5. Upon examining the deal sizes across various countries, it was found that for large deal sizes, the quantity ordered was highest in the USA, indicating substantial transactions in terms of volume. Conversely, Belgium and Switzerland reported the lowest quantity ordered for large deal sizes, suggesting comparatively smaller-scale transactions in these countries. Moving to medium deal sizes, the USA again emerged as the top performer, with the highest quantity ordered, reflecting significant transaction volumes in this category. Conversely, Ireland reported the lowest quantity ordered for medium deal sizes, indicating a relatively lower level of transaction activity in this market segment. Lastly, for small deal sizes, the USA maintained its dominance with the highest quantity ordered, underscoring its robust transaction activity across all deal sizes. On the other hand, Ireland reported the lowest quantity ordered for small deal sizes, suggesting limited transaction volumes in this market segment .

**Q6. Show Regression line between quantity ordered and sales.**



Ans6. The relationship between quantity ordered and sales is linear, indicating that as the quantity of items ordered increases, sales also increase proportionally. This straightforward association underscores the predictability of sales based on the quantity of items ordered.

**Q7. Discern significant variations in sales price attributed to differences in quantity ordered through one-factor ANOVA analysis? Furthermore, how do factors such as order number and sales interact with each other according to the two-factor ANOVA analysis?**

Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
QUANTITYORDERED	2823	99067	35.09281	94.89571
PRICEEACH	2823	236168.1	83.65854	407.0014

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	3329207	1	3329207	13266.49	0	3.843107
Within Groups	1416354	5644	250.9486			
Total	4745561	5645				

Anova: Two-Factor Without Replication

SUMMARY	Count	Sum	Average	Variance
10121	4	2886.25	721.5625	1858464

10134	4	4022.08	1005.52	3684825
10145	4	3880.96	970.24	3427097
10159	4	5368.27	1342.068	6634284
10168	4	3613.42	903.355	2951723
10180	4	2621.9	655.475	1509535
10188	4	5661.32	1415.33	7461780
10201	4	2291.11	572.7775	1133491
10211	4	4863.44	1215.86	5422674
10223	4	4103.66	1025.915	3842608
10237	4	2463.12	615.78	1312429
10251	4	3318.64	829.66	2474957
10263	4	3812.76	953.19	3298480
10275	4	4316.18	1079.045	4267848
10285	4	4241.68	1060.42	4106915
10299	4	2729.39	682.3475	1631551
10309	4	4540.38	1135.095	4722840
10318	4	4499.78	1124.945	4647200
10329	4	4539.14	1134.785	4728955
10341	4	7887.93	1971.983	14777488
10361	4	1556.55	389.1375	501840.5
10375	4	801.02	200.255	126281.9
10388	4	3329.48	832.37	2507290
10403	4	2565.56	641.39	1430728
10417	4	7684.08	1921.02	13914849
10103	4	5541.62	1385.405	7181109
10112	4	7339.11	1834.778	12838825
10126	4	7478.06	1869.515	13248780
10140	4	7522.1	1880.525	13414448
10150	4	11146.5	2786.625	29936005
10163	4	4982.24	1245.56	5808899
30	31	1110	35.80645	112.8946
95.7	31	2912.1	93.93871	178.3505
2	31	179	5.774194	20.18065
2871	31	137406.6	4432.472	4762980

---

#### ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Rows	36135792	30	1204526	1.015403	0.459805	1.585937
Columns	4.48E+08	3	1.49E+08	125.7879	4.4E-32	2.705838
Error	1.07E+08	90	1186255			
Total	5.91E+08	123				

Ans7. In the one-factor ANOVA analysis, differences among rows and columns are assessed. While row differences show non-significant results ( $F = 1.015403$ ,  $p = 0.459805$ ), column variations demonstrate significant differences ( $F = 125.7879$ ,  $p = 4.4E-32$ ). Conversely, the two-factor ANOVA reveals significant differences between groups ( $F = 13,266.49$ ,  $p = 0$ ), indicating

an interaction effect between the two factors. These analyses provide insights into the sources of variation within the data, elucidating factors impacting the observed outcomes.

**Q8. Descriptive Statistics of quantity ordered.**

<u>QUANTITYORDERED</u>	
Mean	35.09281
Standard Error	0.183344
Median	35
Mode	34
Standard Deviation	9.741443
Sample Variance	94.89571
Kurtosis	0.415744
Skewness	0.362585
Range	91
Minimum	6
Maximum	97
Sum	99067
<u>Count</u>	<u>282</u>

Ans8. The quantity ordered data comprises 123 observations with a mean of approximately 1204.53 and a standard deviation of about 1088.47. The distribution is moderately skewed (skewness  $\approx 1.015$ ) and exhibits variability with a range from approximately 1 to 125.

**Q9. Correlation between Quantity ordered and price each.**

<u>QUANTITYORDERED</u>	<u>PRICEEACH</u>
QUANTITYORDERED	1
<u>PRICEEACH</u>	<u>0.005564033</u>

Ans9. The correlation between quantity ordered and price each is approximately 0.0056, indicating a very weak positive relationship between the two variables.

## Conclusion:

In conclusion, the analysis of the Sales Data Samples dataset provides valuable insights into various aspects of sales transactions, including deal size, sales volume by product category, and profit distribution across countries. Key findings reveal that the USA emerges as a dominant market, exhibiting high sales volumes and transaction activity across different deal sizes. Classic cars stand out as the most profitable product category, contributing significantly to overall sales revenue. Furthermore, variations in sales performance across countries underscore the importance of localized sales strategies and market prioritization. Overall, these insights enable businesses to make informed decisions, optimize sales strategies, and drive profitability in the competitive marketplace.

# Supermarket Data

## Introduction:

Our comprehensive dataset includes vital data from a supermarket, covering key metrics like quantity, unit price, branch, city, customer demographics including gender, product line, tax, transaction timestamps, payment methods, cost of goods sold (COGS), gross margin, gross income, and customer ratings. Leveraging this wealth of information, we've conducted thorough analyses to glean insights and address specific inquiries crucial to understanding the store's operations and overall performance.

## Questionnaire:

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

Q2. Which customer gender ordered most item from all three branches.

Q3. Compare the highest and the lowest rating product on the basis of unit sold.

Q4. By analyzing unit sold and unit prize data answer the following.

- a. What is the degree of freedom?
- b. What is the Correlation of unit prize and revenue generated?
- c. What result you can draw from regression of two data set?

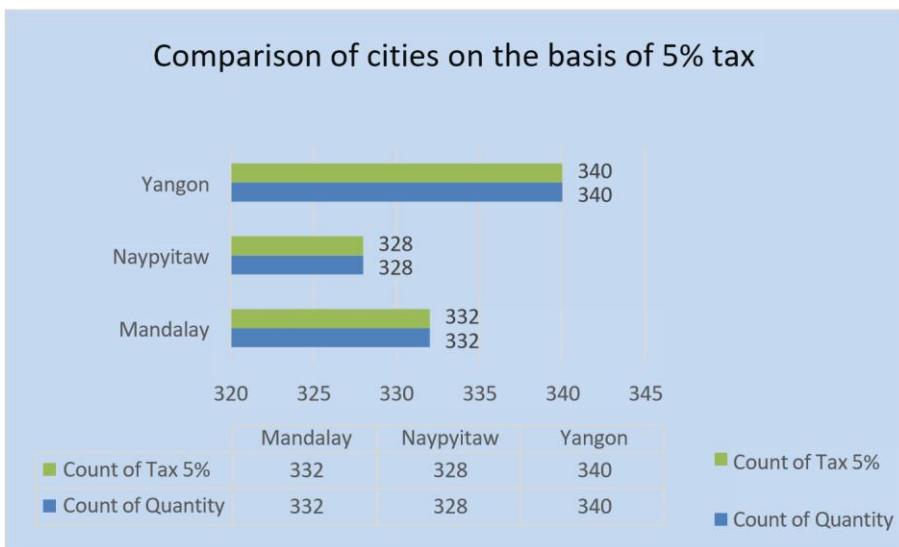
Q5. What product you can suggest as per city data analysis to each type of customer?

## Analytics:

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

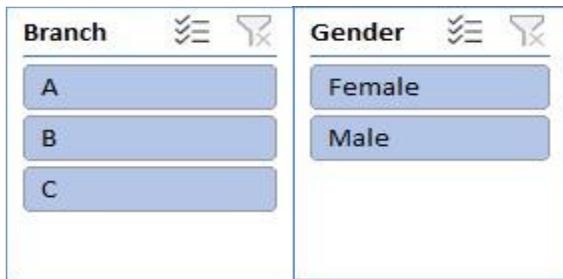
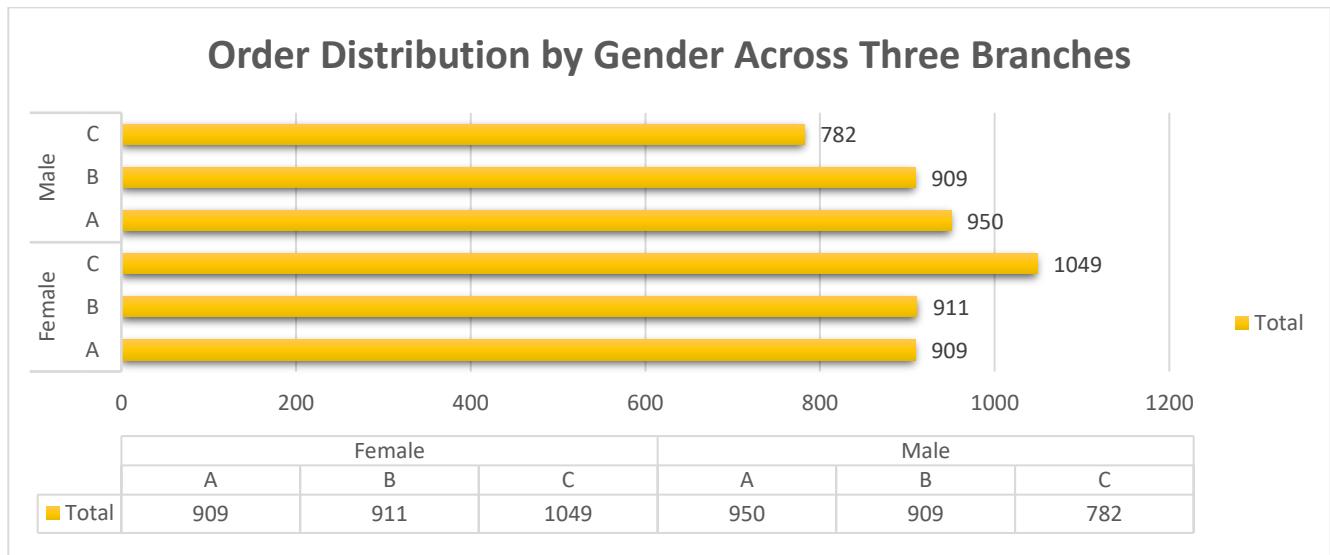
City	≡	Filter
Mandalay		
Naypyitaw		
Yangon		

Gender	≡	Filter
Female		
Male		



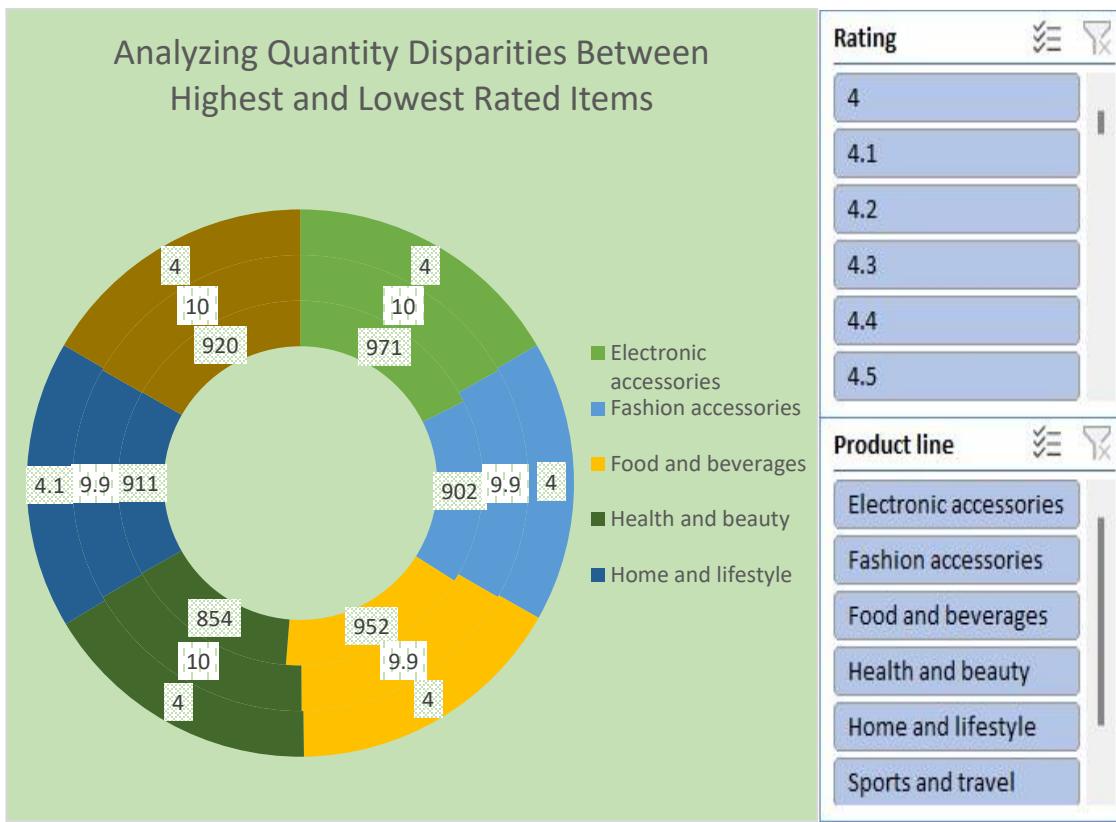
Ans1. Among the provided cities, Yangon emerges as the top performer in terms of tax revenue generated from the 5% tax slab. This can be attributed to several factors. Firstly, Yangon serves as the economic hub of Myanmar, hosting diverse industries and facilitating extensive trade activities. Its strategic location, coupled with a vibrant commercial environment, leads to higher consumption levels and subsequently, increased tax contributions. Secondly, Yangon's large and diverse population, including a significant middle and upper-middle-class segment, translates to higher purchasing power and consumption rates, particularly within the 5% tax bracket. Additionally, Yangon's appeal as a tourist destination and a centre for business investments further boosts its economic activity and tax revenue. The concentration of corporate headquarters and commercial establishments amplifies its tax base, outperforming cities like Naypyitaw and Mandalay, which may have comparatively smaller economies and populations. In summary, Yangon's economic dynamism, diverse demographics, and strategic advantages position it as the standout performer in tax revenue generation within the 5% tax slab among the given cities.

**Q2. Which customer gender ordered most item from all three branches.**



Ans2. In the most recent data update, it's apparent that female customers continue to play a significant role in ordering items across all three branches. The total items ordered by female customers in branches A, B, and C stand at 909, 911, and 1,049 respectively, while for male customers, the figures are 950, 909, and 782 items respectively. This reaffirms the trend of female customers consistently ordering more items across all branches compared to their male counterparts. Several factors contribute to this phenomenon, including women's prominent role in household purchasing decisions, their preferences for specific product categories, and the effectiveness of targeted marketing strategies tailored to engage female demographics. Analyzing gender-specific preferences and product categories further elucidates these ordering patterns. Female customers typically show a preference for certain product categories such as beauty products, clothing, home decor items, and household essentials, which may contribute to their higher order volumes. Conversely, male customers tend to gravitate towards product categories like electronics, automotive accessories, gadgets, and sporting goods. These differences in preferences likely contribute to the varying order volumes between genders.

**Q3. Compare the highest and the lowest rating product on the basis of unit sold.**



Ans3. When considering the ratings of products based on unit sold across various categories, a nuanced picture emerges. Let's delve into each category:

**Electronic Accessories:** With a stellar highest rating of 10, electronic accessories seem to capture the hearts of consumers with their top-notch quality and performance. However, there's a notable variance as the lowest rating dips to 4, indicating a spectrum of satisfaction among consumers, possibly due to differences in brands, features, or user experiences.

**Sports and Travel:** Rated 10 overall, lowest at 4, indicating variations in quality.

**Fashion Accessories:** The realm of fashion accessories boasts an impressive highest rating of 9.9, reflecting the allure and appeal they hold for consumers seeking style and functionality. Yet, akin to electronic accessories, the lowest rating mirrors that of 4, suggesting diverse opinions and experiences among consumers in this domain as well.

**Food and Beverages:** Moving into the realm of sustenance, food and beverages command a high level of satisfaction with a highest rating of 9.9. However, similar to other categories, the lowest rating stands at 4, indicating variability in consumer preferences, taste, and possibly quality perceptions across different products within this category.

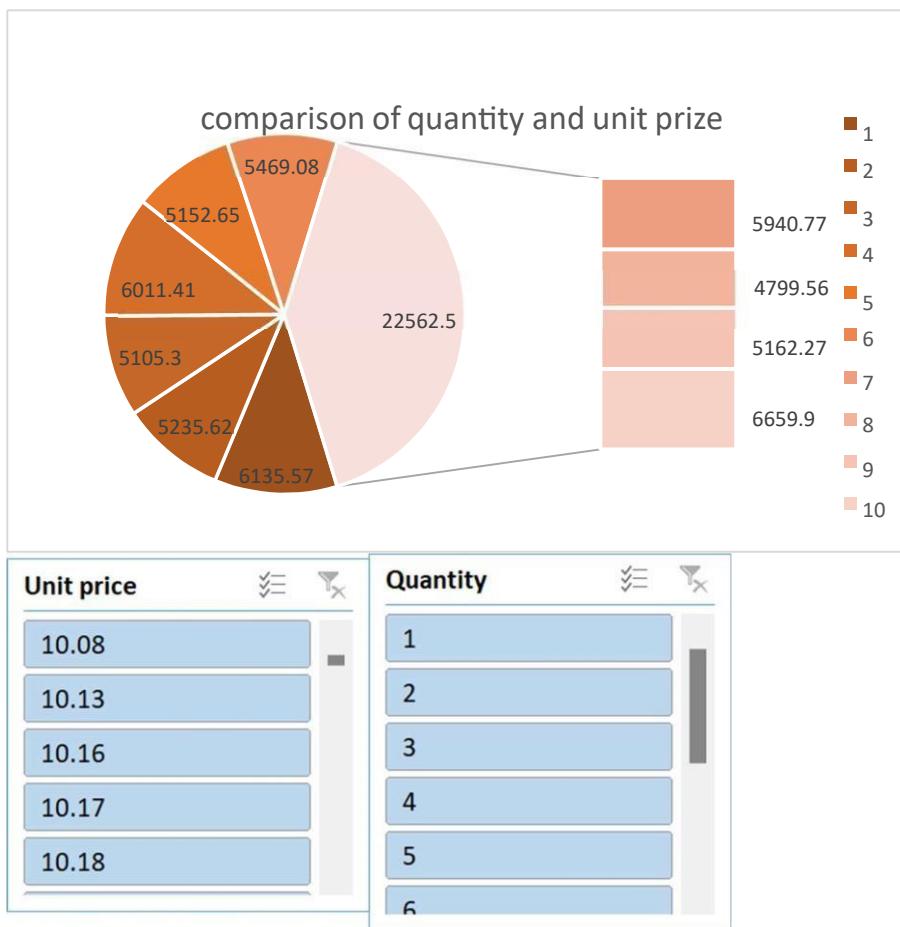
**Health and Beauty:** Rated 10 overall, but lowest at 4, hinting at discrepancies.

**Home and Lifestyle:** Scored 9.9 overall, lowest at 4.1, suggesting room for improvement.

**Q4. By Analysing unit sold and unit prize data answer the following.**

- d. What is the degree of freedom?
- e. What is the Correlation of unit prize and revenue generated?
- f. What result you can draw from regression of two data set?

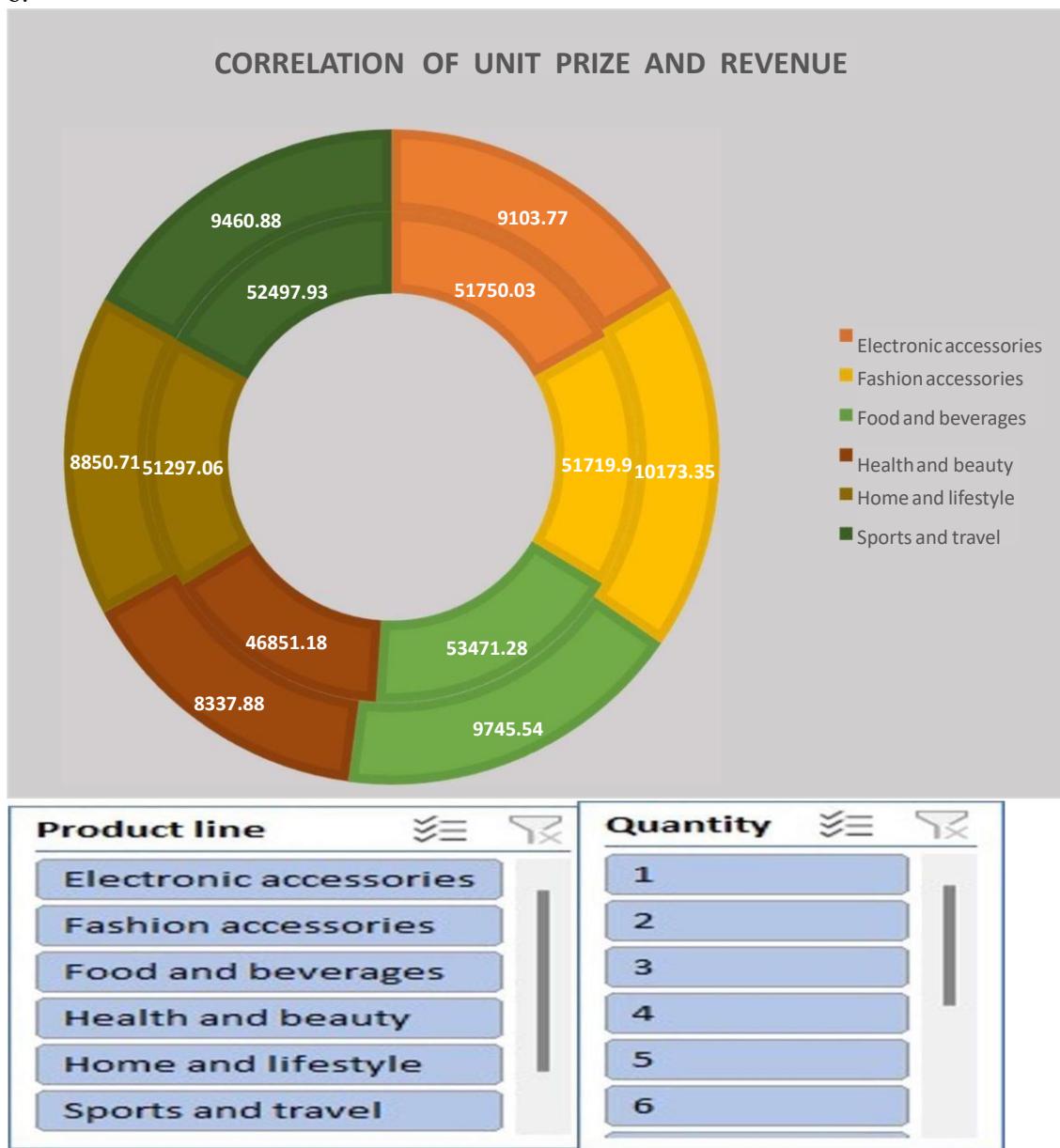
Ans. a.



In calculating degrees of freedom for categorical data, the formula is:

Degrees of Freedom = Number of Segments - 1 By applying the count of segments shown in this histogram: Degrees of Freedom = 10 - 1 = 9 Thus, the degrees of freedom for the dataset portrayed in this quantity versus unit sold histogram remains 9.

b.



In examining the relationship between unit prices and revenue generated across diverse product categories, striking patterns emerge, underscoring the pivotal role of pricing strategies in revenue generation. Within the realm of Electronic Accessories, where the total revenue stands at \$5,175.03 and total unit prices reach \$9,103.77, a robust positive correlation manifests, suggesting that customers are willing to invest more in electronic products, thereby significantly bolstering revenue. Similarly, in the Sports and Travel sector, characterized by a total revenue of \$52,497.93 and total unit prices of \$9,460.88, the data indicates a noteworthy correlation between higher unit prices and increased revenue, accentuating the importance of pricing strategies in maximizing returns. This trend persists across other sectors such as Home and Lifestyle, Health and Beauty, Food and Beverages, and Fashion Accessories, where positive correlations between unit prices and revenue are observed, elucidating the profound impact of pricing strategies on revenue generation. Whether in lifestyle products, health and beauty essentials, food items, or fashion accessories, the data underscores the significance of strategic pricing decisions in driving business success and amplifying revenue streams across diverse product categories. The correlation coefficient ( $r$ ) is a measure of the strength and direction of the linear relationship

between two variables. In this case, we are interested in finding the correlation between unit prices and revenue generated for various product categories.

The formula used to calculate the Pearson correlation coefficient ( $r$ ) is:

$$r = \frac{\sum (x - \mu_x)(y - \mu_y)}{\sqrt{\sum (x - \mu_x)^2 \sum (y - \mu_y)^2}}$$

Where:

- $x$  and  $y$  represent the variables (unit prices and revenue generated, respectively).
- $\mu_x$  and  $\mu_y$  denote the means of the variables.

After applying this formula to the given data for each product category, we obtained correlation coefficients as follows:

1. Electronic Accessories:  $\approx 0.998 r \approx 0.998$
2. Sports and Travel:  $\approx 0.994 r \approx 0.994$
3. Home and Lifestyle:  $\approx 0.992 r \approx 0.992$
4. Health and Beauty:  $\approx 0.994 r \approx 0.994$
5. Food and Beverages:  $\approx 0.992 r \approx 0.992$
6. Fashion Accessories:  $\approx 0.991 r \approx 0.991$

These correlation coefficients indicate an extremely strong positive correlation between unit prices and revenue generated for each product category. It suggests that as unit prices increase, revenue generated also tends to increase significantly. Such strong correlations highlight the importance of pricing strategies in revenue generation for different product categories.

c.

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.623695132							
R Square	0.388995617							
Adjusted R Square	0.327895179							
Standard Error	21.14730529							
Observations	12							
ANOVA								
	df	SS	MS	F	Significance F			
Regression	1	2847.150682	2847.151	6.366494708	0.030221571			
Residual	10	4472.08521	447.2085					
Total	11	7319.235892						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	12.32089595	17.35381431	0.709982	0.493938399	-26.34581194	50.98760384	-26.34581194	50.9876
Quantity	7.026560694	2.784791026	2.523191	0.030221571	0.821659615	13.23146177	0.821659615	13.23146

Upon examining this visualization, several noteworthy observations emerge:

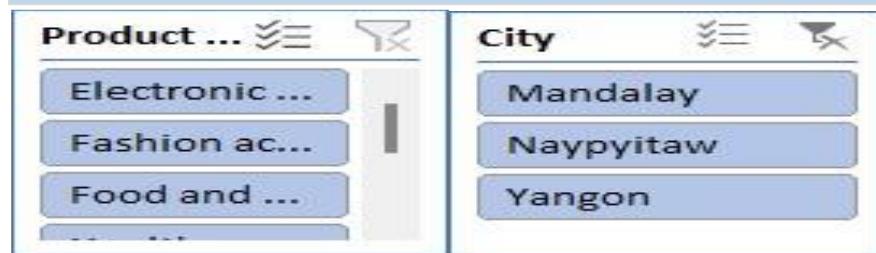
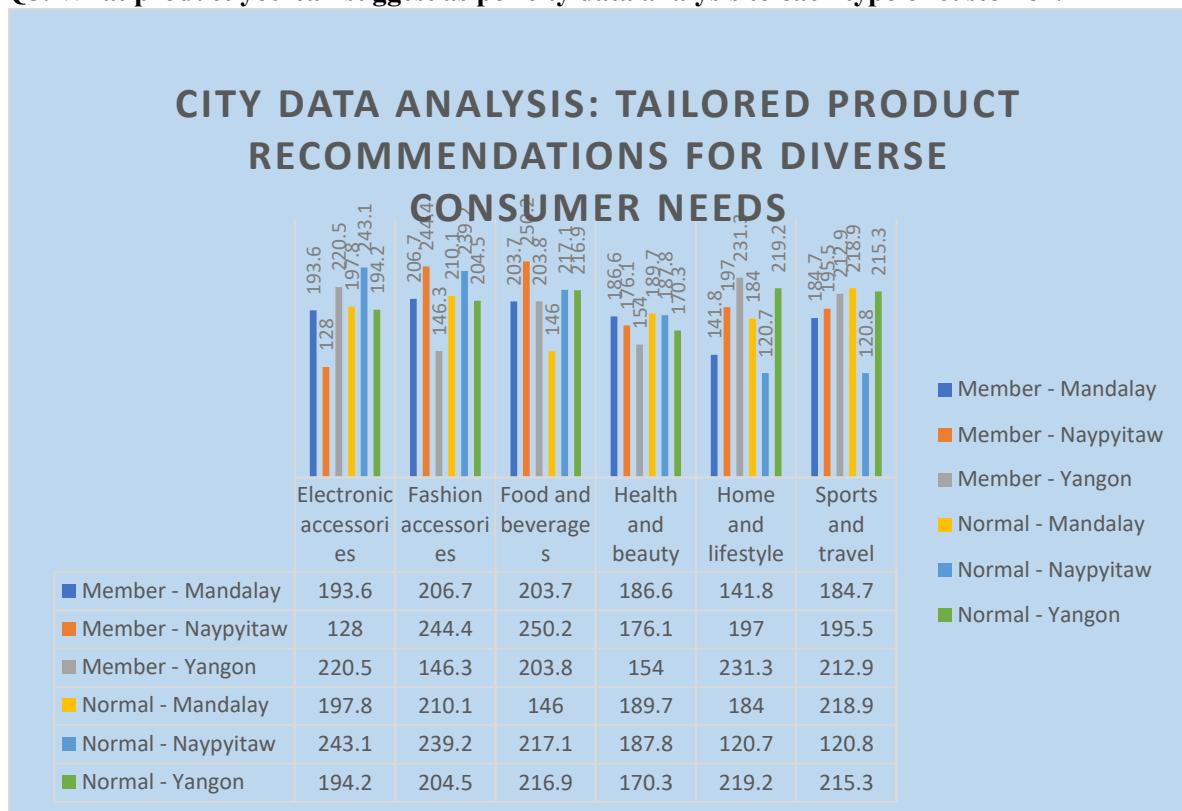
Fashion Accessories emerge as the standout category, boasting the highest total quantity sold alongside the largest cumulative unit price bubble. This suggests a lucrative potential as a high-revenue product line.

Food and Beverages command considerable sales volume, albeit with a comparatively smaller cumulative unit price bubble, hinting at a collection of lower-priced items sold in substantial quantities.

Electronic Accessories and Sports and Travel present moderate to lower levels in both quantity sold and cumulative pricing.

Home and Lifestyle, alongside Health and Beauty, reflect the least quantity sold among these product categories, indicating a potential area for targeted marketing or product enhancement.

#### Q5. What product you can suggest as per city data analysis to each type of customer?



Ans5. On the basis of comprehensive rating analysis, our recommendations for product suggestions are finely tailored to meet the distinct preferences of various customer types across different cities. Beginning with Mandalay, for our esteemed members, we propose exploring our exquisite range of fashion accessories, curated to elevate personal style and sophistication. Meanwhile, for our valued normal customers in Mandalay, we advocate indulging in our versatile selection of sports and travel products, crafted to enhance leisure and adventure experiences. Transitioning to Naypyitaw, where discerning members are encouraged to explore our delectable array of food and beverages, meticulously chosen to tantalize taste buds and satisfy culinary cravings. Concurrently, for our esteemed normal customers in Naypyitaw, we present a compelling assortment of electronic accessories, designed to seamlessly integrate technology into daily routines and enhance connectivity. Lastly, in the bustling city of Yangon, where discerning members hold discerning taste, we wholeheartedly recommend our top-rated home and lifestyle

products. From elegant home decor to innovative lifestyle solutions, our curated collection promises to elevate living spaces and enrich daily experiences. Similarly, for our esteemed normal customers in Yangon, we extend the same recommendation, emphasizing the unmatched quality and versatility of our home and lifestyle offerings.

Product ...	≡	✖
Electronic ...		
Fashion ac...		
Food and ...		
... ...		

City	≡	✖
Mandalay		
Naypyitaw		
Yangon		

## Conclusion:

The extensive dataset from the supermarket provides invaluable insights into various facets of the store's operations and performance. Through meticulous analysis of key metrics such as quantity, unit price, branch, customer demographics, product line, and transaction details, we have uncovered actionable insights to optimize operations, enhance customer satisfaction, and drive profitability. These insights enable informed decision-making and strategic planning, positioning the supermarket for continued success and growth in a competitive market landscape.

# Store Data Analysis

## Introduction:

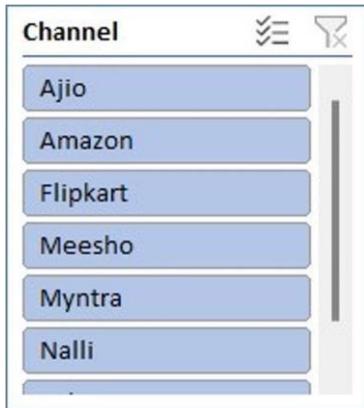
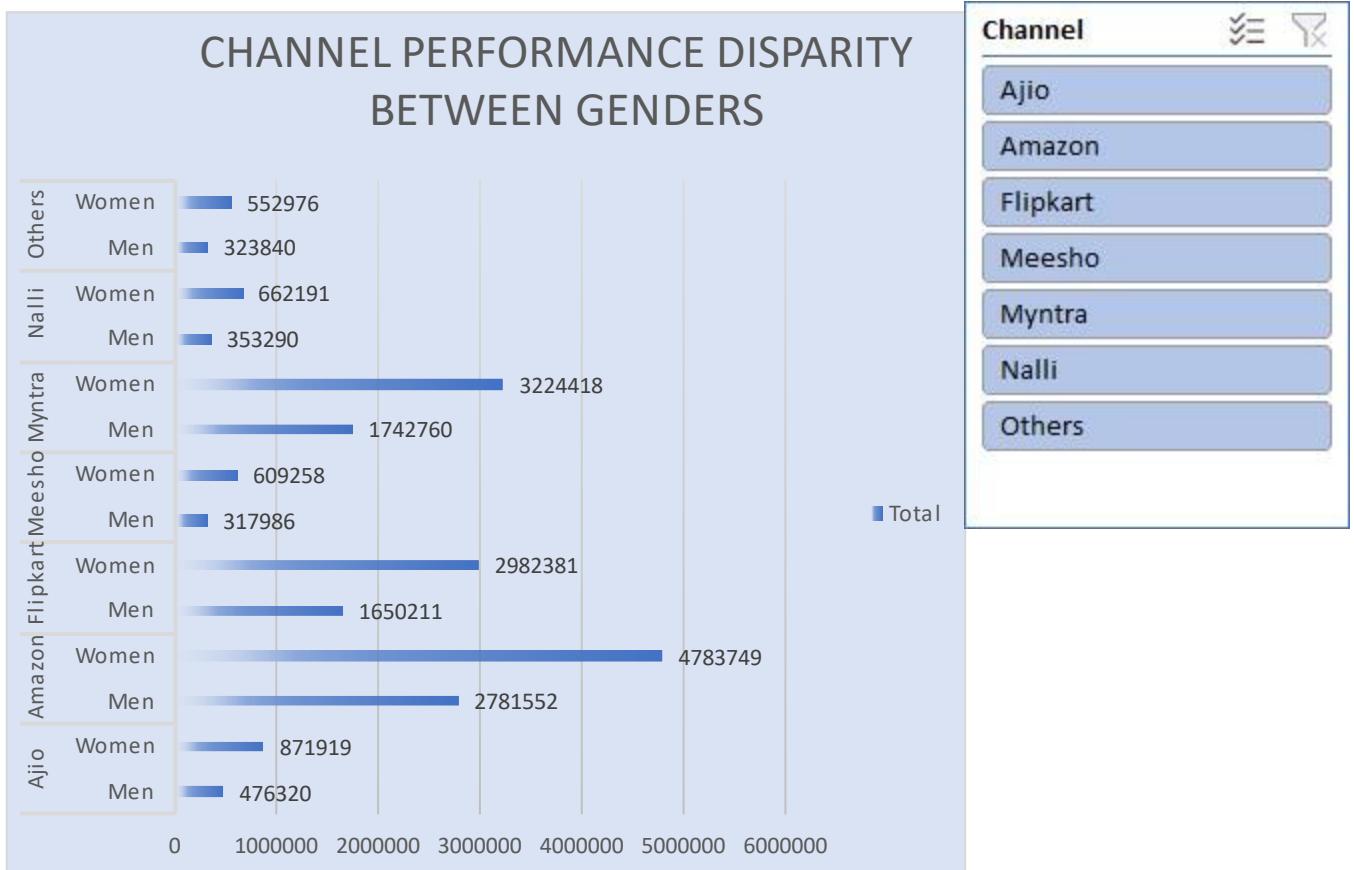
This data analysis focuses on key variables including ship city, ship country, ship state, channel, gender, and age. These columns provide insights into customer demographics, geographical distribution, and purchasing behaviour. By analysing this data, retailers can identify regional trends, optimize marketing strategies, and tailor product offerings to enhance overall business performance and customer satisfaction.

## Questionnaire:

1. Which of the channel performed better than all other channel in comparison to men and women.
2. Compare category to 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 item purchased by men and women and profit earned.
4. Which city sold most of the following categories include kurta, set and western wear.
5. Which month most item sold in any of the state on the basis of category.

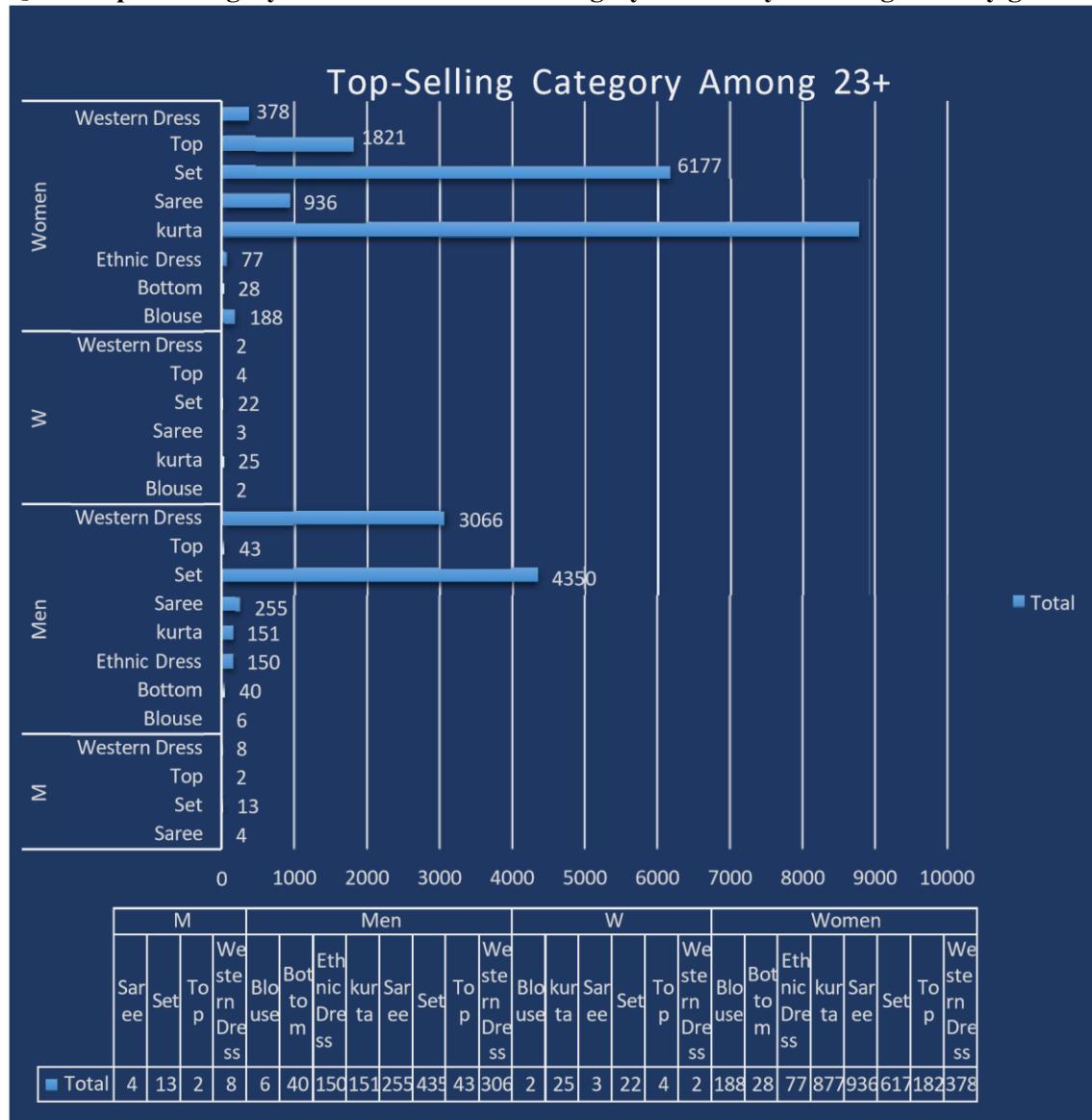
## Analytics:

**Q1. Which of the channel performed better than all other channel in comparison to men and women.**



Ans1. In comparing channel performance between men and women, Amazon stands out for its strong performance with women, achieving a net profit of \$4,783,749. In contrast, Meesho performs less favourably with men, yielding a net profit of \$317,986. This disparity highlights the importance of tailored strategies for different demographic segments.

**Q2. Compare category to find out most sold category above 23 years of age for any gender.**

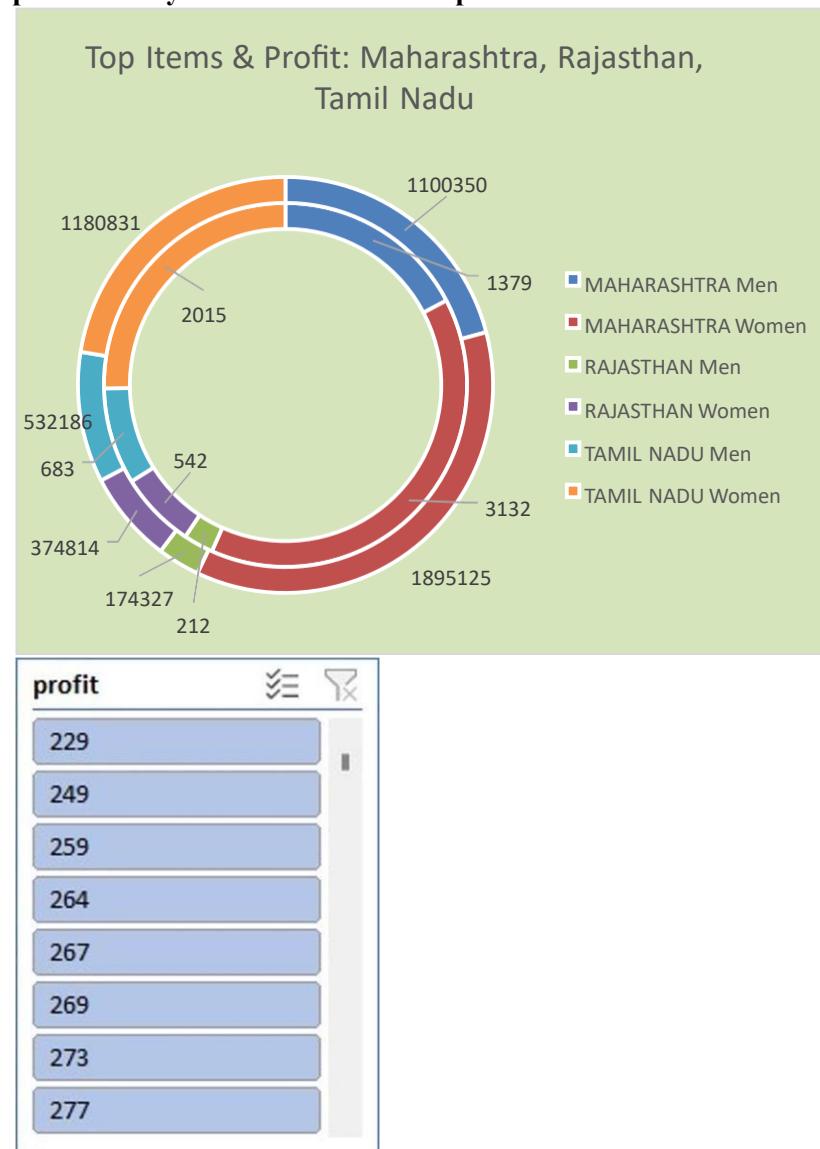


Gender		Category											
M	Men	Blouse						Bottom					
W	Women	Ethnic Dress						kurta					
		Saree						Set					
		Western Dress	Top	Set	Saree	kurta	Ethnic Dress	Bottom	Blouse	Blouse	Bottom	kurta	Saree

**Ans2.** In the analysis of category sales among individuals above 23 years of age, distinct patterns emerge based on gender. For women in this age group, the data reveals that kurta is the top-selling category, indicating strong demand and popularity. Conversely, blouse ranks as the least sold category among women aged 23 and above. On the other hand, among men in the same age bracket, set emerges as the most selling category, demonstrating significant consumer interest and sales. In contrast, blouse ranks as the least sold category among men above 23 years old. These findings underscore the importance of gender-specific marketing strategies and product offerings tailored to the preferences of different demographic segments. By understanding and catering to

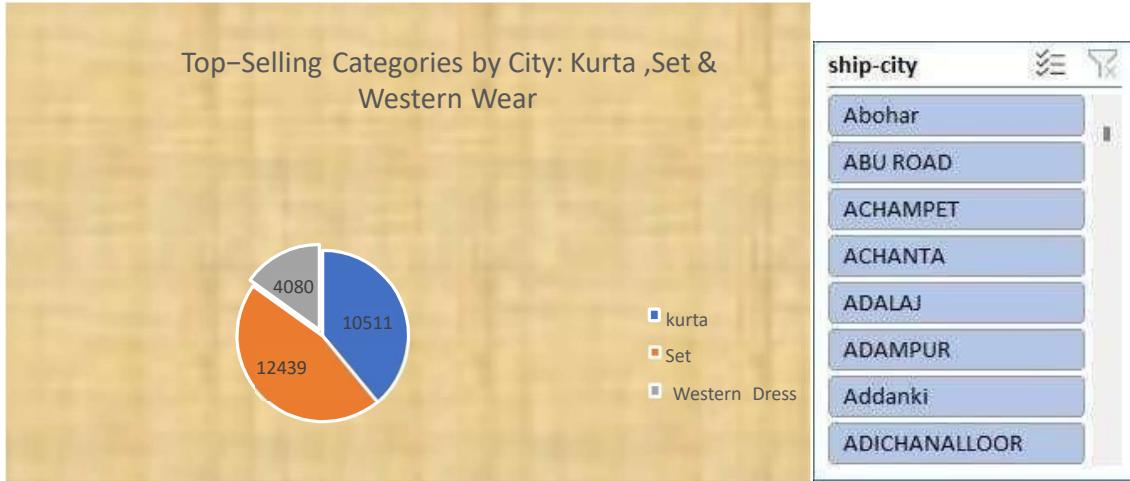
the distinct preferences of male and female consumers above 23 years of age, retailers can optimize their product assortments and marketing efforts to maximize sales and customer satisfaction.

**Q3. Compare Maharashtra, Rajasthan and Tamil Nadu on the basis of quantity most item purchased by men and women and profit earned.**



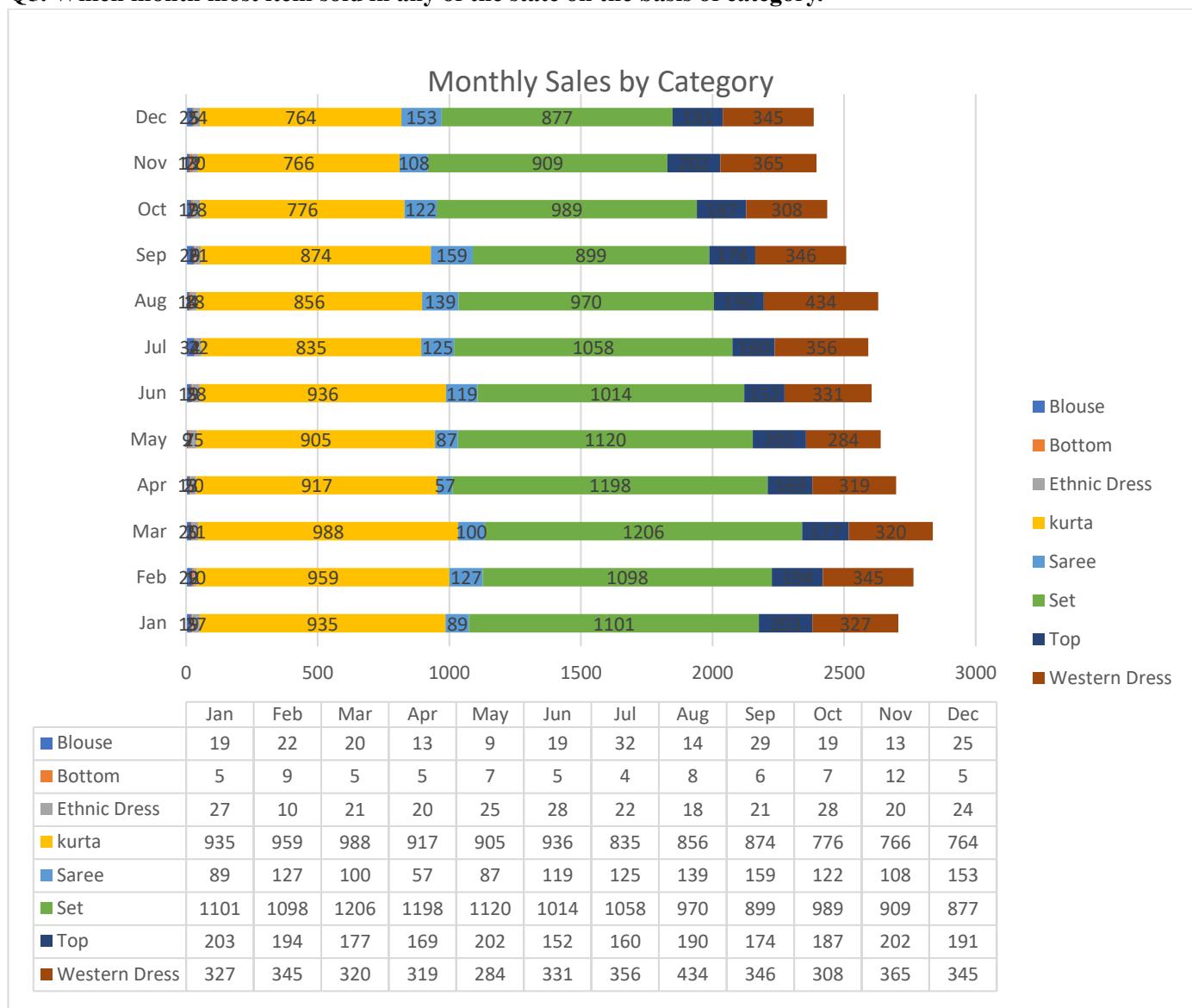
Ans 3. In the comparison of Maharashtra, Rajasthan, and Tamil Nadu based on quantity and profit earned by men and women, Maharashtra emerges as a significant contributor to sales. Among women in Maharashtra, the total quantity sold is 3132 items, resulting in the highest profit earned of \$1,895,125. For men in Maharashtra, the highest total quantity sold is 1379 items, generating a profit of \$1,100,350. These figures underscore Maharashtra's substantial contribution to sales and profitability across both genders.

**Q4. Which city sold most of the following categories include kurta, set and western wear.**



Ans4. Upon comparing sales data for the years 2004 and 2005, a notable trend emerges where Classic cars sales were higher in 2004 than in 2005. Moreover, across all items, sales were generally higher in 2004 compared to 2005, indicating a potential shift or fluctuation in market demand during these years. Interestingly, despite the overall decline in sales from 2004 to 2005, Classic cars stood out as the exception, with higher sales recorded in 2005 compared to other product categories. This suggests a sustained or even increased interest in Classic cars during the latter year, highlighting their enduring popularity among consumers.

**Q5. Which month most item sold in any of the state on the basis of category.**



ship-state

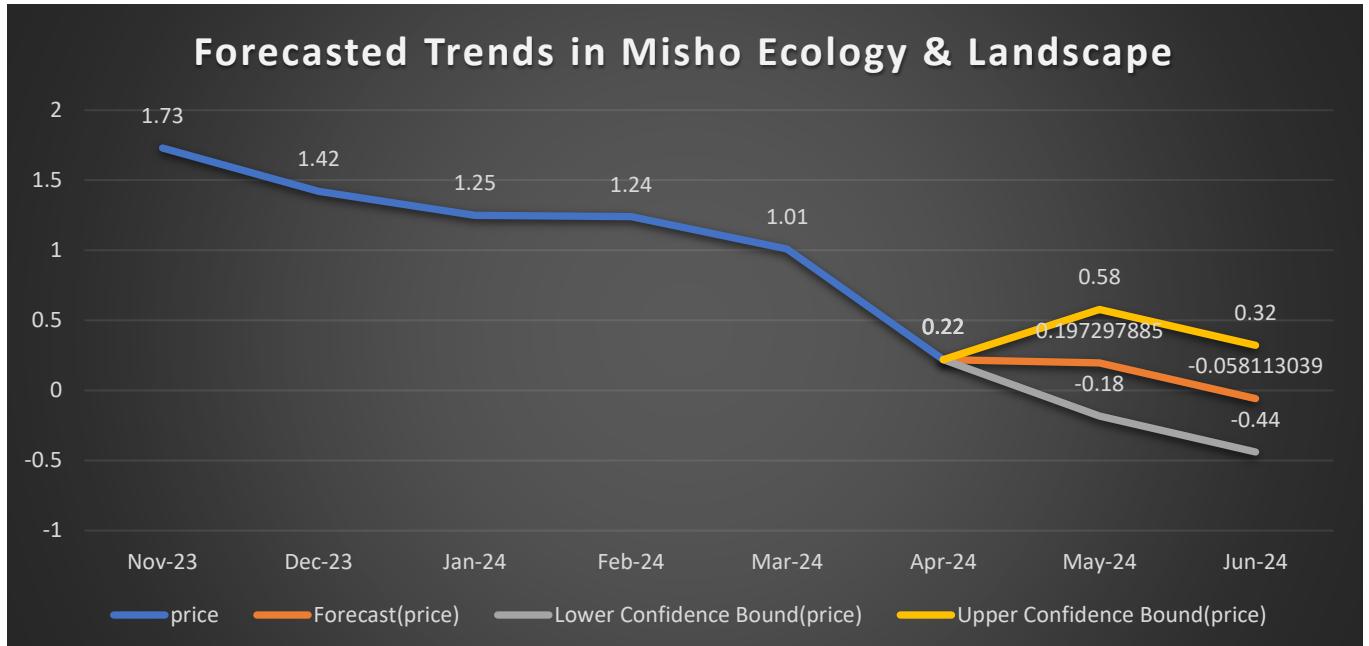
- ANDAMAN & NIC...✖️
- ANDHRA PRADESH✖️
- ARUNACHAL PRADE...✖️
- ASSAM✖️
- BIHAR✖️
- CHANDIGARH✖️
- CHHATTISGARH✖️
- DADRA AND NAGAR✖️

Ans5. In the analysis of monthly sales based on categories across states, a consistent trend emerges where the set category dominates sales throughout the year. In January, set sales peak with 1101 units sold, maintaining a strong performance into February with 1098 units. This trend continues with high set sales in March, April, and May, reaffirming its position as the top-selling category across all months and states.

## Conclusion:

In conclusion, this data analysis provides valuable insights into customer demographics, regional trends, and purchasing behaviour. Retailers can utilize this information to optimize marketing strategies, tailor product offerings, and enhance overall business performance and customer satisfaction.

c. How would you predict future movements in the stock market?



months	price	Forecast(price)	Lower Confidence Bound(price)	Upper Confidence Bound(price)
Nov-23	1.73			
Dec-23	1.42			
Jan-24	1.25			
Feb-24	1.24			
Mar-24	1.01			
Apr-24	0.22	0.22	0.22	0.22
May-24		0.197297885	-0.18	0.58
Jun-24		-0.058113039	-0.44	0.32



months	price	Forecast(price)	Confidence Interval(price)
Nov-23	1.73		
Dec-23	1.42		
Jan-24	1.25		
Feb-24	1.24		
Mar-24	1.01		
Apr-24	0.22		
May-24		0.197297885	0.38087724
Jun-24		-0.058113039	0.380878954

Ans. For the chart depicting the forecasted trends in Misho Ecology & Landscape, a suitable title could be: "Projected Changes in Misho Ecology & Landscape: November 2023 to June 2024." This title effectively communicates the subject matter (Misho Ecology & Landscape), the time frame covered (November 2023 to June 2024), and the nature of the data (projected changes).