Dashboard of Order data

Dashboard of Car Collection

Dashboard of Loan Data

Dashboard of Cookie data

Dashboard of Store Data

	BENGALURU	CHENNAI	HYDERABAD	MUMBAI	NEW DELHI
kurta	22	9	11	4	7
Set	409	190	273	227	323
Western Dress	365	194	327	190	121

Dashboard of shop sales

Dashboard of Sales Data

DASHBOARD FOR SUPER MARKET DATA ANALYSIS

Report on Car Collection Dataset

Introduction –

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

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

The intended audience for this report includes:

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

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

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

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

Questionnaire –

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

Analytics –

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

Row Labels	Sum of Mileage
Corolla	277131
Impala	228486
Grand Total	**505617**

Ans: Toyota Corolla is recognized for its notable fuel efficiency, which is frequently superior to larger vehicles such as the Chevrolet Impala.

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

Row Labels	Sum of Price
Ford	17908
Honda	15968
Grand Total	**33876**

Answer - Based on this comparison, while Honda models generally offer higher mileage, Ford models tend to have a higher average price. Therefore, the choice between Ford and Honda would depend on the buyer's priorities, whether they value higher mileage or are willing to pay a premium for a Ford vehicle**.**

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

Sum of Price Row Labels	Column Labels Black	Blue	Green	Red	Silver	White	Grand Total
Chevrolet			4959		3791	5701	14451
Dodge	2090				8601		10691
Ford	9805			4397	3706		17908
Honda		7245		4000	2000	2723	15968
Nissan			6796	2914			9710
Toyota	4422	2798			2160		9380
Grand Total	**16317**	**10043**	**11755**	**11311**	**20258**	**8424**	**78108**

Answer - The chart from the car collection dataset shows that black and red are the most popular car colors, symbolizing elegance and prestige. In contrast, blue is the least preferred color. This information is valuable for manufacturers and marketers to understand consumer preferences in the automotive industry.

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

Answer - In the car collection dataset, four silver cars have mileages ranging from 120,000 to 210,000 miles, averaging 165,000 miles. Two green cars have mileages of 140,000 and 170,000 miles, averaging 150,000 miles. Thus, silver cars tend to have higher average mileage than green cars.

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

Answer - The cars with total costs exceeding $2,000 are: Silverado ($4,500), Maxima ($2,500), Mustang ($3,100), Malibu ($3,000), Impala ($3,500), Fusion ($2,100), F-150 ($3,000), CRV ($4,10

Conclusion and Review

This analysis offers valuable insights into consumer preferences and key considerations when purchasing cars. It highlights the Toyota Corolla's strong reputation for fuel efficiency, the variety of options available from Ford, and the significant consumer preference for black and red car colors. Additionally, the data suggests a potential link between car color and average mileage, with silver cars typically having higher mileage. Overall, these findings underscore the importance of factors like fuel efficiency, color preference, and budget constraints in the car-buying decision process.

0), Corolla ($4,300), Charger ($7,500), Altima ($5,500), and Accord ($3,000).

Regression: -

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.358764572
R Square	0.128712018
Adjusted R Square	0.087222114
Standard Error	32204.73295
Observations	23

ANOVA

			df	*SS*	*MS*	*F*	*Significance F*
Regression	1	3217481630	3.22E+09	3.102249	0.09273902		
Residual	21	21780041315	1.04E+09				
Total	22	24997522945					

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	122108.9268	24014.1535	5.084873	4.91E-05	72168.7607	172049.093
X Variable 1	-14.51458144	8.240739406	-1.76132	0.092739	-31.6521372	2.62297432

These statistics indicate a weak relationship: Multiple R is 0.359, R Square is 0.129, Adjusted R Square is 0.087, with a Standard Error of 32,204.73, based on 23 observations. Overall, the model has limited explanatory power, suggesting that further refinement is needed for improved predictions.

Anova: Single Factor: -

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	7.03E+10	1	7.03E+10	123.6791	2.28E-14	4.061706
Within Groups	2.5E+10	44	5.69E+08			
Total	9.53E+100	45				

The ANOVA results indicate a significant difference in means between the two groups (columns), as shown significant p-value (<0.05) for the "Between Groups" variation.

Anova: Two-Factor Without Replication:

ANOVA

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

Rows	1.23E+10	22	557756895.8	0.962803693	0.535017989	2.04777
Columns	7.03E+10	1	70315407145	121.3789272	2.01396E-10	4.30095
Error	1.27E+10	22	579304898.8			
Total	9.53E+10	45				

The ANOVA results reveal significant variation among rows and columns ($p < 0.001$), with degrees of freedom (df) v 1, respectively. The error term has a degree of freedom of 22.

Correlation: -

		Column 1	*Column 2*
Column 1	1	-0.4110586	
Column 2		-0.4110586	1

Descriptive Statistics: -

Column1		*Column2*	
Mean	81499.65217	Mean	3305.1304
Standard Error	7028.67123	Standard Error	187.75002
Median	75006	Median	3196
Mode	#N/A	Mode	#N/A
Standard Deviation	33708.32305	Standard Deviation	900.41744
Sample Variance	1136251043	Sample Variance	810751.57
Kurtosis	-0.87669401	Kurtosis	-1.1920464
Skewness	0.479783783	Skewness	0.2222322
Range	105958	Range	2959
Minimum	34853	Minimum	2000
Maximum	140811	Maximum	4959
Sum	1874492	Sum	76018
Count	23	Count	23
Largest (1)	140811	Largest (1)	4959
Smallest (1)	34853	Smallest (1)	2000
Confidence Level (95.0%)	14576.57197	Confidence Level (95.0%)	389.3697

Column 2	-0.4110586	1

Column 1 has a mean of 81,499.65, a standard deviation of 33,708.32, and a count of 23. Column 2 has a mean of 3,305.13, a standard deviation of 900.42, and a count of 23. These columns exhibit notable differences in both mean and standard deviation.

Report on Order Data

Introduction:

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

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

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

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

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

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

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

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

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

Questionnaire:

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

1. ANALYTICS:

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

An -After comparing all the states based on segment and sales, California stands out as the state with the highest sales. Additionally, the consumer segment performed well across all states.

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

Ans. Office Supplies is the top performing category in all the states as it clearly shows from the given graph

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

Count of Country	Column Labels			
Row Labels	Consumer	Corporate	Home Office	Grand Total
California	1020	601	325	1946
Texas	516	310	147	973
Washington	275	130	99	504
Grand Total	**1811**	**1041**	**571**	**3423**

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

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

Ans - By analyzing the given dataset, we found that in all three segments, the total sales were greater than the average sales.

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

By analyzing the given Order Sales dataset, we observed that the average sales for the Technology category were significantly higherthan those for the other categories.

2. CONCLUSION:

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

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

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

Regression:

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

Correlation

	Sales	
	cost	
Sales	1	
	0.709412	
cost	0.709412	1

The correlation matrix indicates a strong positive correlation of 0.71 between sales and cost, suggesting that as the cost increases, sales tend to increase as well. This correlation coefficient reflects a moderately strong linear relationship between the two variables. Both sales and cost exhibit mutual influence on each other

Anova (single factor) :

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Sales	9800	2261537	230.7691	392692.6
cost	9800	2261411	230.7562	197630.9

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	0.807262	1	0.807262	2.73E-06	0.99868	3.841933
Within Groups	5.78E+09	19598	295161.7			
Total	5.78E+09	19599				

The ANOVA analysis compares the variability between two groups, sales and cost, revealing a minimal difference between them with a small sum of squares (SS) of 0.81. The F-statistic of 2.735 and p-value of0.999 suggest that this difference is not statistically significant, indicating that the means of sales and cost are likely equal.

Anova (two factor) without Replication :-

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Rows	1.94E+09	9799	197630.9	65535	#NUM!	#NUM!
Columns	0	0	65535	65535	#NUM!	#NUM!

Error	0	0	65535
Total	1.94E+09	9799	

The ANOVA table illustrates significant variation attributed to rows, represented by a sum of squares (SS) of 1,936,585,107 and 9,799 degrees of freedom (df), resulting in a mean square (MS) of 197,630.89. The F-statistic is notably high at 65535, indicating a substantial influence of row
factors on the observed variance. However, the p-value is reported as #NUM!, suggesting a potential issue with the calculation or data. Similarly, for columns, no variation is observed, with an SS and MS of 0, and the F-statistic equals 65535

Descriptive Statistics:

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

Report on Loan Data

1. INTRODUCTION:

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

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

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

2. QUESTIONNAIRE:

Q1. How many male graduates who are not married applied for Loan? What was the highest amount?
Q2. How many female graduates who are not married applied for Loan? What was the highest amount?
Q3. How many male non-graduates who are not married applied for Loan? What was the highest amount?
Q4. How many female graduates who are married applied for Loan? What was the highest amount?
Q5. How many male and female who are not married applied for Loan? Compare Urban, Semi-urban and rularon the basis of amount.

3. ANALYTICS:

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

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

Row Labels	Female Max of Loan_Amount_Term	Count of Gender	Total Max of Loan_Amount_Term	Total Count of Gender
Graduate	**360**	**35**	**360**	
No	360	35	360	35
Grand Total	**360**	**35**	**360**	

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

Row Labels	Column Labels Male Count of Gender	Max of Loan_Amount_Term	Total Count of Gender	Total Max of Loan_Amount_Term
Not Graduate	**16**	**360**		**360**
No	16	360	16	360
Grand Total	**16**	**360**		**360**

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

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

4. CONCLUSION:

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

5. REGRESSION:

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

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.45908096
R Square	0.21075532
Adjusted R Square	0.20858707
Standard Error	56.0766111
Observations	366

ANOVA

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

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*
Intercept	106.07753	4.10024098	25.8710478	1.7585E-84	98.014396	114.140665	98.01439
5720	0.0058851	0.00059692	9.85902887	1.7676E-20	0.00471125	0.00705895	0.004711

CORRELATION :

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

	ApplicantIncome	*CoapplicantIncome*	*LoanAmount*
ApplicantIncome	1		
CoapplicantIncome	-0.110334799	1	
LoanAmount	0.458768926	0.144787815	1

Anova (Single Factor) :

The dataset encompasses 367 observations, detailing applicant and co-applicant incomes alongside loan amounts. On average, applicants possess a higher income, averaging around $4805.60, compared to co-applicants whose average income is approximately $1569.58. Loan amounts vary widely, averaging $134.28. ANOVA analysis underscores significant distinctions between the income and loan amounts across the groups, implying diverse financial profiles among applicants and co-applicants.

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
ApplicantIncome	367	1763655	4805.599455	24114831.09
CoapplicantIncome	367	576035	1569.577657	5448639.491
LoanAmount	367	49280	134.2779292	3964.141124

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	4202537452	2	2101268726	213.2009841	5.87569E-79	3.003920577
Within Groups	10821681107	1098	9855811.573			
Total	1502421856	1100				

Anova two factor without Replication:

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

ANOVA

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

Descriptive Statistics:

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

Largest(1)	72529	Largest(1)	24000	Largest(1)	550
Smallest(1)	0	Smallest(1)	0	Smallest(1)	0
Confidence Level(95.0%)	504.0756067	Confidence Level(95.0%)	239.6059543	Confidence Level(95.0%)	6.462910219

Report on Sales Data

Introduction -

In the field of business analytics, a dataset that contains detailed sales transactions is an incredibly valuable resource for gaining practical insights. Such a dataset, equipped with columns that list order numbers, quantities of items ordered, the price of each item, and additional data points, presents a holistic overview of how sales are performing.

This kind of data is instrumental for a multitude of purposes. It allows businesses to monitor specific orders, giving visibility into the sales pipeline. Analysts can scrutinize how different products are selling, identify which items are popular or underperforming, and detect patterns in buying behaviors.

The richness of this dataset lies in its ability to support informed strategic choices. Businesses can use the data to fine-tune their marketing strategies, adjust inventory levels, set competitive pricing, and enhance customer experiences. By understanding the intricacies captured in this data, businesses can tailor their operations to be more efficient and effective..

Questionaries -

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

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

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

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

5. Compare all the countries based on deal size.

Analytics:

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

Column Labels Classic Cars	Vintage Cars	Grand Total
1818	1945	3763
937	324	1261
147	446	593
456	476	932
1244	240	1484
1284	252	1536
3540	1955	5495
1281	276	1557
202	28	230
948	1303	2251
314	313	627
33992	**21069**	**55061**

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

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

Ans – The classic car product yield the most sale

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

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

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

Sum of SALES	Column Labels		
Row Labels	2004	2005	Grand Total
Classic Cars	1762257.09	672573.28	2434830.37
Motorcycles	560545.23	234947.53	795492.76
Planes	502671.8	200074.17	702745.97
Ships	341437.97	128178.07	469616.04
Trains	116523.85	36917.33	153441.18
Trucks and Buses	529302.89	178057.02	707359.91
Vintage Cars	911423.77	340739.31	1252163.08
Grand Total	**4724162.6**	**1791486.71**	**6515649.31**

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

5. Compare all the countries based on deal size.

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

Regression and Anova

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

CORRELATION:

The correlation coefficient you calculated (0.657840928) represents the strength. Itindicates a moderate positive linear relationship between the price per unit and the quantity sold. This

means that as the price per unit tends to increase, the quantity sold also tends to increase, but the relationship is not perfect.

Descriptive Statistics:

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

Anova (single Factor) :

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
30	2822	99037	35.09461	94.92015
95.7	2822	236072.4	83.65428	407.0943
2	2822	18252	6.467753	17.85699
2871	2822	10029758	3554.131	3393504

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	2.61E+10	3	8.71E+09	10261.03	0	2.605696
Within Groups	9.57E+09	11284	848506			

Total	3.57E+10	11287

The single-factor ANOVA analysis unveils significant variations among the groups, with a high F-value of 10261.03 and an ultra-low p-value close to zero, indicating a strong impact of the factor being analyzed. The degrees of freedom (df) for the between-groups factor are 3, representing the variability in means across the groups. Within the groups, the df is 11284, reflecting the variation within each group,and an error (standard error of the residuals) of approximately 848506.0368.

Anova without Replication :

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Rows	3.24E+09	2822	1146956	1.021361	0.257009	1.054809
Columns	2.32E+10	2	1.16E+10	10320.57	0	2.997323
Error	6.34E+09	5644	1122968			
Total	3.28E+10	8468				

The ANOVA analysis reveals no significant differences in the means across rows, as indicated by the non-significant F-value of 1.021 ($p = 0.257$) and the degrees of freedom (df) of 2822. Similarly, for columns, a highly significant difference is observed among the means, with an F-value of 10320.57 ($p < 0.001$) and df of 2. The error term, representing variability within groups, has an MS of approximately 1122968.007, reflecting the average amount of variation not explained by the model

Descriptive Statistics :

Mean	35.09281	Mean	83.65854	Mean	6.466171	Mean
Standard Error	0.183344	Standard Error	0.379702	Standard Error	0.079535	Standard Error
Median	35	Median	95.7	Median	6	Median
Mode	34	Mode	100	Mode	1	Mode
Standard Deviation	9.741443	Standard Deviation Sample	20.17428	Standard Deviation	4.225841	Standard Deviation Sample

Sample Variance	94.89571	Variance	407.0014	Sample Variance	17.85773	Variance
Kurtosis	0.415744	Kurtosis	-0.37482	Kurtosis	-0.56115	Kurtosis
Skewness	0.362585	Skewness	-0.94665	Skewness	0.590741	Skewness
Range	91	Range	73.12	Range	17	Range
Minimum	6	Minimum	26.88	Minimum	1	Minimum
Maximum	97	Maximum	100	Maximum	18	Maximum
Sum	99067	Sum	236168.1	Sum	18254	Sum
Count	2823	Count	2823	Count	2823	Count
Largest(1)	97	Largest(1)	100	Largest(1)	18	Largest(1)
Smallest(1)	6	Smallest(1)	26.88	Smallest(1)	1	Smallest(1)
Confidence Level(95.0%)	0.359503	Confidence Level(95.0%)	0.744521	Confidence Level(95.0%)	0.155952	Confidence Level(95.0%)

.

Conclusion and Review:

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

Store Dataset Report

Introduction

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

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

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

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

Questionnaire:

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

Analytics:

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

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

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

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

The table of items sold is given below:

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

The graph is as follows:

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

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

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

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

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

Which city sold most of following categories

a. Kurta
b. Set
c. Western wears

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

	BENGALURU	CHENNAI	HYDERABAD	MUMBAI	NEW DELHI
kurta	22	9	11	4	7
Set	409	190	273	227	323
Western Dress	365	194	327	190	121

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

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

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

Conclusion and Review:

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

Report on Cookie Data

Introduction :

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

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

Questionaries :

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

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

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

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

Analytics :

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

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

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

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

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

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

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

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

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

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

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

Conclusion and Review :

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

Regression:

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

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.829304
R Square	0.687746
Adjusted R Square	0.687298
Standard Error	1462.76
Observations	700

ANOVA

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

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

Co-relation:

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

	Units Sold	*Revenue*
Units Sold	1	0.796298
Revenue	0.796298	1

Anova (Single Factor) :

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

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*

3450	699	1923505	2751.795	4154648
5175	699	2758189	3945.908	6850161

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	4.98E+08	1	4.98E+08	90.57022	7.53E-21	3.848129
Within Groups	7.68E+09	1396	5502405			
Total	8.18E+09	1397				

Anova two factor without Replication:

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

ANOVA

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

Anova two factor with Replication:

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

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

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	8.55E+08	49	17443674	65535	#NUM!	#NUM!

Columns	5.78E+10	3	1.93E+10	65535	#NUM!	#NUM!
Interaction	4.39E+08	147	2983765	65535	#NUM!	#NUM!
Within	0	0	65535			
Total	5.91E+10	199				

Descriptive Statistics:

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

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

Supermarket Sales Data Report

Introduction:

Dataset Overview:

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

Key Attributes:

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

Questionnaire:

Q1. Which of the given cities having tax 5% slab performed better than all the others?
Q2. Which customer gender ordered most items from all the three branches?
Q3. Compare highest and lowest rating products on the basis of units sold.
Q4. Analyzing units sold and unit price data answer the following sub questions

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

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

Analytics:

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

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

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

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

a. At Branch A, females placed the highest number of orders.
b. Branch B saw higer number of orders placed by Females
c. Meanwhile, at Branch C, males placed the most orders.

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

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

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

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

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

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

Answer:

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

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

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

Quantity = 0.0012x + 5.4438. Despite this equation, the coefficient of determination (R2) is merely 0.0001, highlighting the inconsistency in customer buying patterns solely based on unit price.

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

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

Conclusion and Reviews

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

1. City Performance:

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

2. Gender-based Ordering:

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

3. Rating and Units Sold:

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

4. Unit Price and Revenue Relationship:

The regression analysis revealed a weak correlation ($R2 = 0.0001$) between unit price and quantity sold. This suggests that customers' purchasing decisions may not be significantly influenced by unit price alone, indicating the need for deeper insights into consumer preferences and behavior.

5. Product Recommendations:

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

Reviews:

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

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

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

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