Sales Performance & Marketing Effectiveness Analysis

Table of Contents

Executive Summary

This project analyzes a retail sales dataset (N = 300) to understand the drivers of sales performance and evaluate the effectiveness of marketing and operational factors.

Key findings:

- Marketing Budget shows a positive relationship with Sales Amount.
- Online Advertising Spend is highly skewed and not normally distributed.
- Payment Method is significantly associated with Sales Category.
- Advertising Method does not significantly influence Sales Category.
- Store Size has minimal association with Online Advertising Spend.

These insights provide actionable recommendations for marketing allocation and payment strategy optimization.

Business Problem

A retail company wants to answer:

1. What factors drive higher sales performance?
2. Does marketing investment significantly impact revenue?
3. Are certain advertising strategies more effective?
4. Does customer payment behavior influence sales value?

The objective is to generate statistical and strategic insights to guide decision-making.

Dataset Overview

The dataset includes:

Categorical Variables:

Region (5 major U.S. cities)

Advertising Method (Print, Radio, Social Media, TV)

Sales Category (High, Medium, Low)

Payment Method (Cash, Credit Card, Mobile Payment, Online Transfer)

Customer Type (New, Returning, VIP)

Numerical Variables:

Marketing Budget

Online Advertising Spend

Sales Amount

Store Size

Total Observations: 300

Exploratory Data Analysis (EDA) using SPSS

Exploring Variable Distributions Through Frequencies

Frequency table of Region variables

Table 1.1 Characteristics of Region variable

Region

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Chicago	49	16.3	16.3	16.3
	Houston	66	22.0	22.0	38.3
	Los Angeles	61	20.3	20.3	58.7
	New York	55	18.3	18.3	77.0
	Phoenix	69	23.0	23.0	100.0
	Total	300	100.0	100.0	

A frequency analysis was conducted to examine the distribution of respondents across different regions. The results indicated that participants were distributed across five major U.S. cities: Phoenix had the highest representation (n = 69, 23.0%), followed by Houston (n = 66, 22.0%), Los Angeles (n = 61, 20.3%), New York (n = 55, 18.3%), and Chicago (n = 49, 16.3%). The total number of participants was N = 300.

Frequency table for Advertising Method variable

Table 1.2 Characteristic of Advertising Method variable

Advertising_Method

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Print	64	21.3	21.3	21.3
	Radio	78	26.0	26.0	47.3
	Social Media	72	24.0	24.0	71.3
	TV	86	28.7	28.7	100.0
	Total	300	100.0	100.0	

A frequency analysis was performed to examine the distribution of advertising methods used by the businesses in the sample. Results showed that the most frequently used advertising method was TV advertising (n = 86, 28.7%), followed by Radio (n = 78, 26.0%), Social-Media (n = 72, 24.0%), and Print advertising (n = 64, 21.3%). The total number of observations was N = 300.

Frequency table for Sales Category variable

Table 1.3 Characteristics of Sales Category variable

Sales_Category

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	High Value	175	58.3	58.3	58.3
	Low Value	30	10.0	10.0	68.3
	Medium Value	95	31.7	31.7	100.0
	Total	300	100.0	100.0	

A frequency analysis was conducted to examine the distribution of sales transactions by value category. Results indicated that the majority of transactions were classified as High Value (n = 175, 58.3%), followed by Medium Value (n = 95, 31.7%) and Low Value (n = 30, 10.0%). The total number of observations was N = 300.

Frequency table for Payment Method variable

Table 1.4 Characteristics of Payment Method variable

Payment_Method

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Cash	70	23.3	23.3	23.3
	Credit Card	87	29.0	29.0	52.3
	Mobile Payment	77	25.7	25.7	78.0
	Online Transfer	66	22.0	22.0	100.0
	Total	300	100.0	100.0	

A frequency analysis was conducted to assess the distribution of payment methods used by customers. The most common payment method was Credit Card (n = 87, 29.0%), followed by

Mobile Payment (n = 77, 25.7%), Cash (n = 70, 23.3%), and Online Transfer (n = 66, 22.0%). The total number of observations was N = 300.

Frequency table for Customer Type variable

Table 1.5 Characteristics of Customer Type variable

Customer_Type

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	New	130	43.3	43.3	43.3
	Returning	140	46.7	46.7	90.0
	VIP	30	10.0	10.0	100.0
	Total	300	100.0	100.0	

A frequency analysis was conducted to examine the distribution of customer types in the sample. The results showed that the majority of customers were classified as Returning (n = 140, 46.7%), followed by New customers (n = 130, 43.3%) and VIP customers (n = 30, 10.0%). The total number of observations was N = 300.

Frequency table for Marketing Budget variable

Table 1.6 Characteristics of Marketing Budget variable

Statistics

Marketing_Budget

N	Valid	300
	Missing	0
Mean		14983.3543
Std. Deviation		2952.58107
Skewness		.175
Std. Error of Skewness		.141
Kurtosis		.597
Std. Error of Kurtosis		.281
Minimum		5276.20
Maximum		26558.19
Sum		4495006.30

Descriptive statistics were calculated for the marketing budget variable (N = 300). The budgets ranged from $5,276.20 to $26,558.19, with a mean of $14,983.35 (SD = $2,952.58). The

distribution of values showed slight positive skewness (skewness = 0.18, SE = 0.14) and was slightly platykurtic (kurtosis = 0.60, SE = 0.28). The total sum of the marketing budgets was $4,495,006.30, and there were no missing data.

Frequency table for Online Advertising spend variable

Table 1.7 Characteristics of Online Advertising Spend variable

Statistics

Online_Advertising_Spend

N	Valid	300
	Missing	0
Mean		7771.2608
Std. Deviation		2328.40760
Skewness		1.018
Std. Error of Skewness		.141
Kurtosis		1.672
Std. Error of Kurtosis		.281
Minimum		3573.01
Maximum		18888.76
Sum		2331378.24

Descriptive statistics were calculated for the online advertising spend variable (N = 300). The expenditures ranged from $3,573.01 to $18,888.76, with a mean of $7,771.26 (SD = $2,328.41). The distribution showed moderate positive skewness (skewness = 1.02, SE = 0.14) and was leptokurtic, indicating a relatively peaked distribution (kurtosis = 1.67, SE = 0.28). The total spending in this category was $2,331,378.24, with no missing data.

Frequency table for Sales Amount variable

Table 1.8 Characteristics Sales Amount variable

Statistics

Sales_Amount

N	Valid	300
	Missing	0
Mean		38329.7402
Std. Deviation		7041.14214
Skewness		-.125

Std. Error of Skewness	.141
Kurtosis	.017
Std. Error of Kurtosis	.281
Minimum	17574.17
Maximum	56818.50
Sum	11498922.05

Descriptive statistics were computed for the sales amount variable (N = 300). Sales ranged from $17,574.17 to $56,818.50, with a mean of $38,329.74 (SD = $7,041.14). The distribution of sales was approximately normal, with negligible skewness (skewness = –0.13, SE = 0.14) and nearly zero kurtosis (kurtosis = 0.02, SE = 0.28), suggesting a symmetric and mesokurtic distribution. The total sales amount across all observations was $11,498,922.05, with no missing data.

Frequency table for Store Size variable

Table 1.9 Characteristics Store Size variable

Statistics

Store_Size

N	Valid	300
	Missing	0
Mean		3053.3013
Std. Deviation		996.67647
Skewness		-.122
Std. Error of Skewness		.141
Kurtosis		-.173
Std. Error of Kurtosis		.281
Minimum		503.11
Maximum		5832.38
Sum		915990.39

Descriptive statistics were computed for the store size variable (N = 300). Store sizes ranged from 503.11 sq. ft. to 5,832.38 sq. ft., with a mean of 3,053.30 sq. ft. (SD = 996.68). The distribution was approximately normal, exhibiting slight negative skewness (skewness = –0.12, SE = 0.14) and slightly platykurtic shape (kurtosis = –0.17, SE = 0.28). The total store area across all observations was 915,990.39 sq. ft., with no missing data.

Normality Test for Numerical Variables

Normality Test for Marketing Budget variable

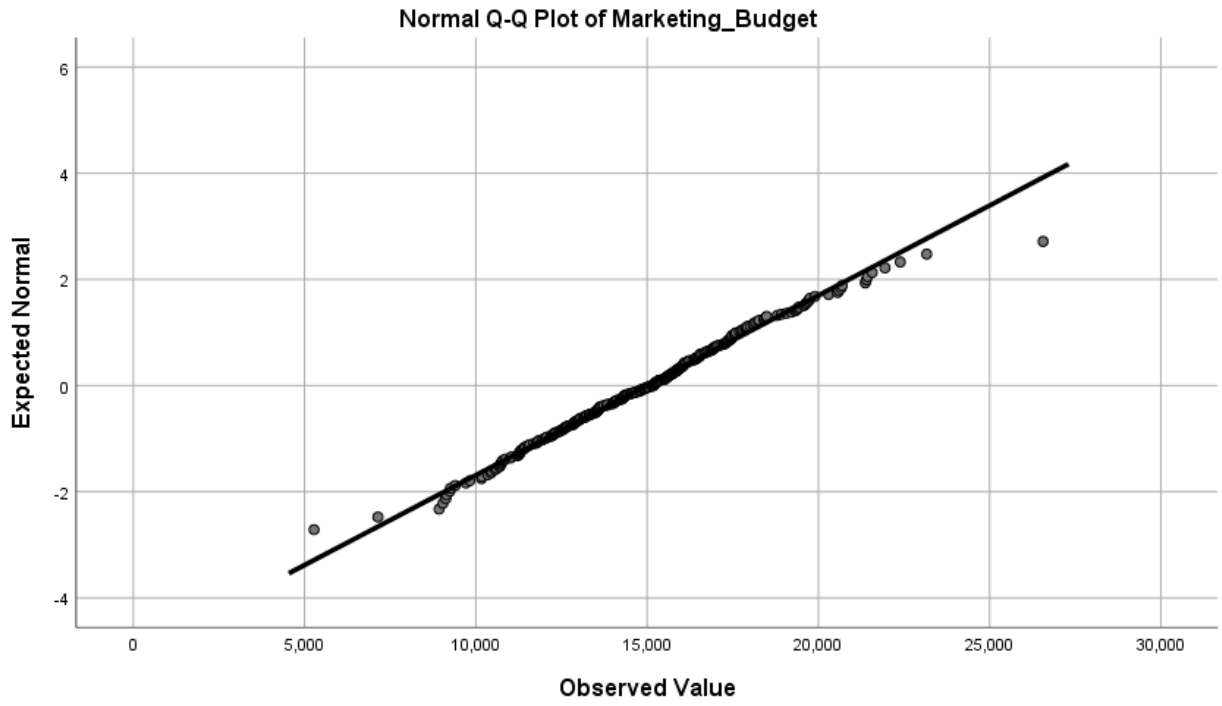
Table 2.1 Normality test result for Marketing Budget

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Marketing_Budget	.031	300	.200*	.994	300	.306

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction



A test of normality was conducted for the Marketing Budget variable. The Kolmogorov–Smirnov test indicated no significant deviation from normality, $D(300) = 0.031$, $p = .200$. Similarly, the Shapiro–Wilk test also confirmed the data were normally distributed, $W(300) = 0.994$, $p = .306$. Therefore, the assumption of normality was met.

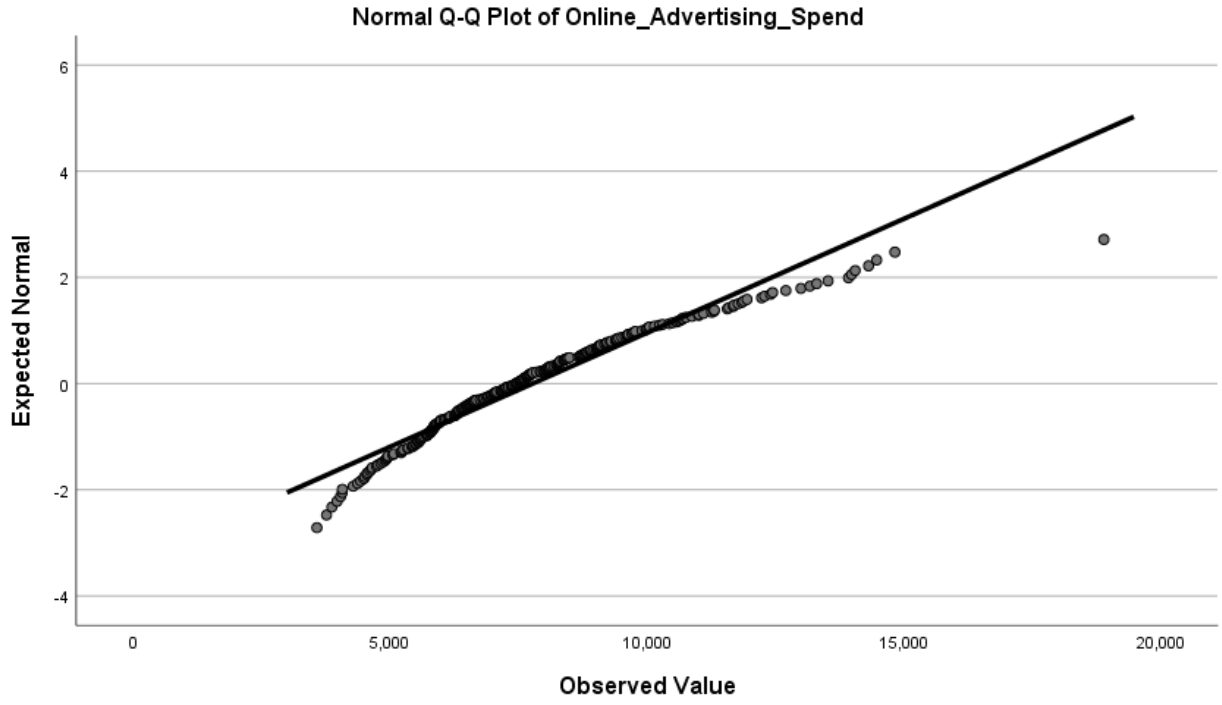
Normality Test for Online Advertising Spend variable

Table 2.2 Normality test result for Online Advertising Spend

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Online_Advertising_Spend	.083	300	.000	.946	300	.000

a. Lilliefors Significance Correction



A test of normality was conducted for the Online Advertising Spend variable. The Kolmogorov–Smirnov test showed a significant deviation from normality, $D(300) = 0.083$, $p<.001$, and the Shapiro–Wilk test also indicated non-normality, $W(300) = 0.946$,$p<.001$. These results suggest that the data for online advertising spend are not normally distributed.

Normality Test for Sales Amount variable

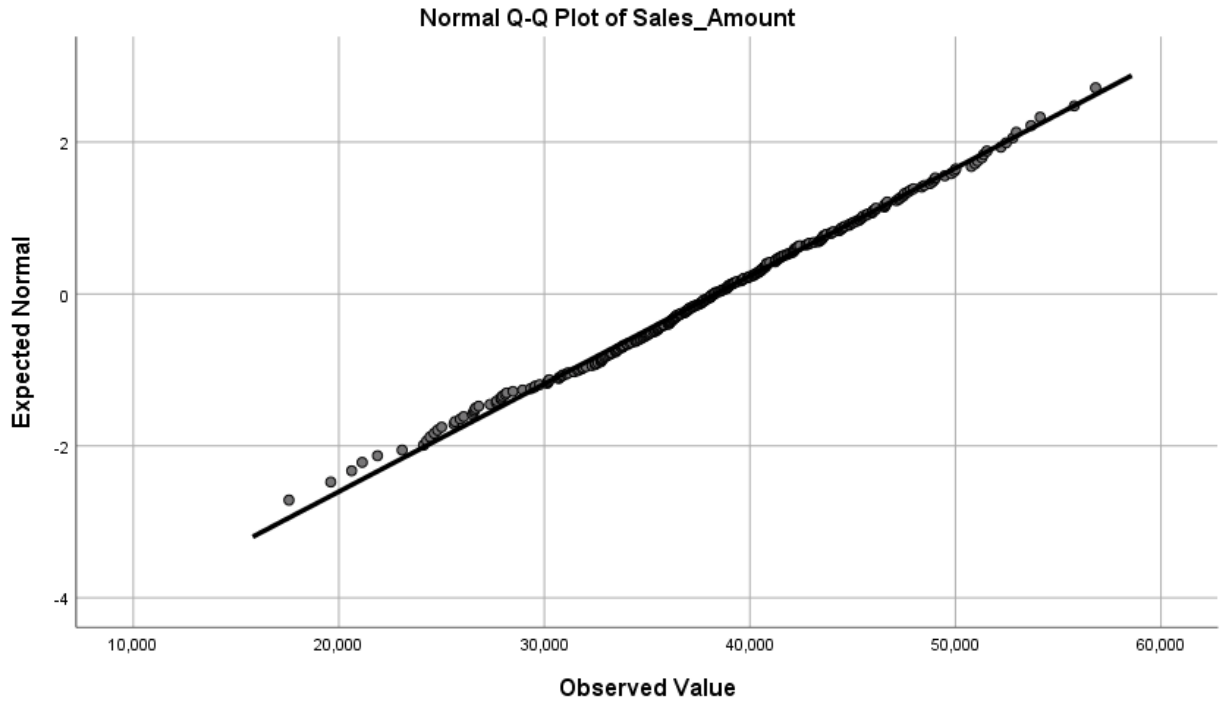
Table 2.3 Normality test result for Sales Amount

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Sales_Amount	.031	300	.200*	.997	300	.762

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction



A test of normality was conducted for the Sales Amount variable. The Kolmogorov–Smirnov test indicated no significant deviation from normality, $D(300)=0.031, p=.200$. Similarly, the Shapiro–Wilk test also showed that the data were normally distributed, $W(300)=0.997, p=.762$. These results suggest that the assumption of normality is met for sales amount.

Normality Test for Sales Amount variable

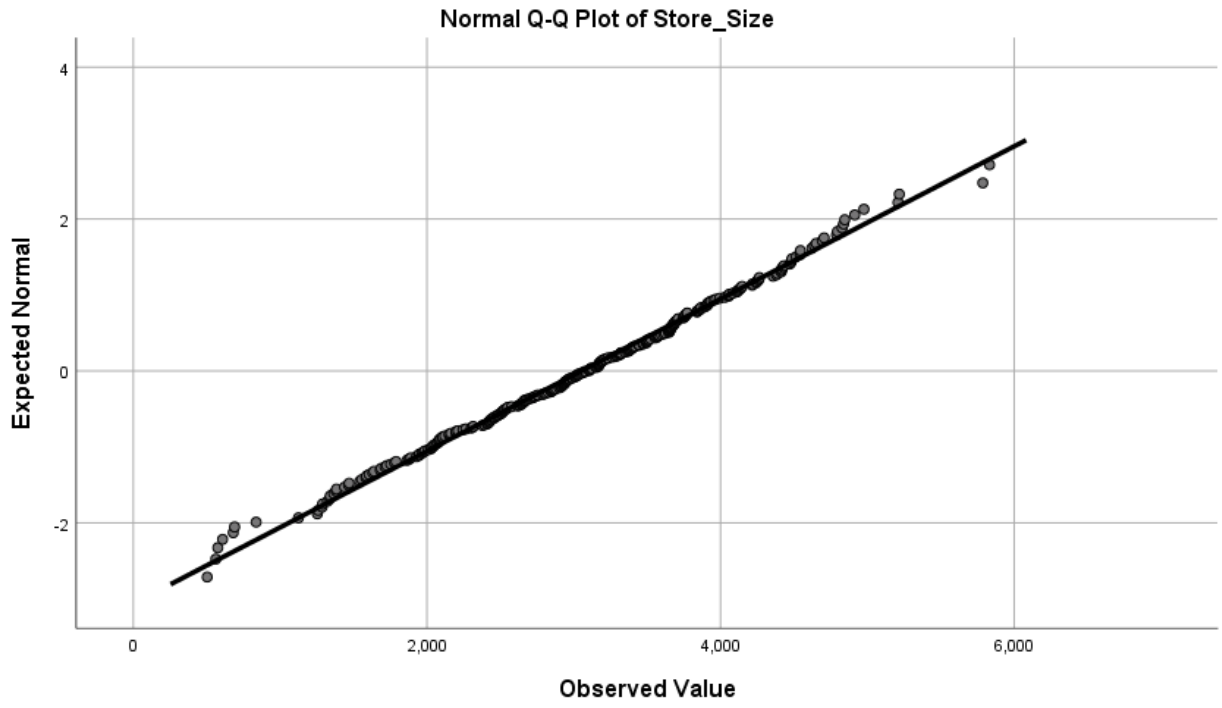
Table 2.3 Normality test result for Sales Amount

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Store_Size	.031	300	.200*	.995	300	.486

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction



A normality test was conducted for the Store Size variable. The Kolmogorov–Smirnov test indicated no significant deviation from normality, D(300)=0.031,p=.200. The Shapiro–Wilk test also supported normality, W(300)=0.995,p=.486. These findings suggest that the distribution of store size is approximately normal, and the assumption of normality is met.

Evaluating Association Between Categorical Variables

Chi-Square Tests of Association between Advertising Method and Sales Category

Table 3.1 Cell Count between Variables

Advertising_Method * Sales_Category Crosstabulation

Count

		Sales_Category			
		High Value	Low Value	Medium Value	Total
Advertising_Method	Print	34	6	24	64
	Radio	43	10	25	78
	Social Media	42	9	21	72
	TV	56	5	25	86
Total		175	30	95	300

Table 3.2 Test results

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	4.732[a]	6	.579
Likelihood Ratio	4.875	6	.560
N of Valid Cases	300		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 6.40.

A Pearson chi-square test was conducted to examine the association between the two categorical variables (Advertising Method and Sales Category). The test indicated no statistically significant relationship, $X^2(6, \mathrm{N} = 300) = 4.73, \mathrm{p} = .579$. The likelihood ratio chi-square test also confirmed this result, $X^2(6, \mathrm{N} = 300) = 4.88, \mathrm{p} = .560$. All expected cell counts were above 5, satisfying the assumption of the test (minimum expected count = 6.40).

Chi-Square Tests of Association between Sales Category and Payment Method

Table 3.3 Cell Count between Variables

Sales_Category * Payment_Method Crosstabulation

Count

		Payment_Method				
		Cash	Credit Card	Mobile Payment	Online Transfer	Total
Sales_Category	High Value	41	43	53	38	175
	Low Value	11	6	3	10	30
	Medium Value	18	38	21	18	95
Total		70	87	77	66	300

Table 3.4 Test results

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	16.008[a]	6	.014
Likelihood Ratio	16.063	6	.013
N of Valid Cases	300		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 6.60.

A Pearson chi-square test was conducted to assess the association between two categorical variables (Sales Category and Payment Method). The result was statistically significant, indicating a relationship between the variables, $\chi 2(6, N=300) = 16.01$, $p = .014$. The likelihood ratio chi-square test also supported this result, $\chi 2(6, N=300) = 16.06$, $p = .013$. Assumptions for the test were met, as no cells had expected counts less than 5 (minimum expected count = 6.60).

Chi-Square Tests of Association between Customer Type and Payment Method

Table 3.5 Cell Count between Variables

Customer_Type * Payment_Method Crosstabulation

Count

		Payment_Method				
		Cash	Credit Card	Mobile Payment	Online Transfer	Total
Customer_Type	New	28	41	32	29	130
	Returning	33	39	39	29	140
	VIP	9	7	6	8	30
Total		70	87	77	66	300

Table 3.6 Test results

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	2.539[a]	6	.864
Likelihood Ratio	2.521	6	.866
N of Valid Cases	300		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 6.60.

A Chi-square test of independence was performed to examine the relationship between the two categorical variables (Customer Type and Payment Method). The test results indicated no significant association, $\chi^2(6, N = 300) = 2.54$, $p = .864$ (Pearson Chi-Square). The likelihood ratio test similarly showed no significance, $\chi^2(6, N = 300) = 2.52$, $p = .866$.

All expected cell counts exceeded 5, meeting the assumptions required for the test.

Chi-Square Tests of Association between Advertising Method and Region

Table 3.7 Cell Count between Variables

Advertising_Method * Region Crosstabulation

Count

		Region					
		Chicago	Houston	Los Angeles	New York	Phoenix	Total
Advertising_Method	Print	7	13	15	14	15	64
	Radio	15	21	14	17	11	78
	Social Media	16	12	19	8	17	72
	TV	11	20	13	16	26	86
Total		49	66	61	55	69	300

Table 3.8 Test results

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	16.294[a]	12	.178
Likelihood Ratio	16.887	12	.154
N of Valid Cases	300		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 10.45.

A Pearson chi-square test was conducted to assess the association between two categorical variables (Advertising Method and Region). The result was not statistically significant, $\chi 2(12, N=300) = 16.29$, $p = .178$, indicating no evidence of a meaningful relationship between the variables. The likelihood ratio chi-square test yielded similar results, $\chi 2(12, N=300) = 16.89$, $p = .154$. All expected cell counts were above 5, meeting the test's assumption (minimum expected count = 10.45).

Chi-Square Tests of Association between Region and Customer Type

Table 3.9 Cell Count between Variables

Customer_Type * Region Crosstabulation

Count

		Region					
		Chicago	Houston	Los Angeles	New York	Phoenix	Total
Customer_Type	New	23	32	22	27	26	130
	Returning	22	30	30	24	34	140
	VIP	4	4	9	4	9	30
Total		49	66	61	55	69	300

Table 3.10 Test results

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	6.186[a]	8	.626
Likelihood Ratio	6.208	8	.624
N of Valid Cases	300		

a. 1 cells (6.7%) have expected count less than 5. The minimum expected count is 4.90.

A Pearson chi-square test was conducted to assess the association between two categorical variables (Region and Customer Type). The result was not statistically significant, $\chi 2(8, N=300) = 6.19$, $p = .626$, indicating no significant relationship between the variables. The likelihood ratio chi-square test produced a similar result, $\chi 2(8, N=300) = 6.21$, $p = .624$. Assumption checks revealed that 1 cell (6.7%) had an expected count less than 5, with a minimum expected count of 4.90, which may slightly affect the test's validity.

Graphs and Visual Interpretation

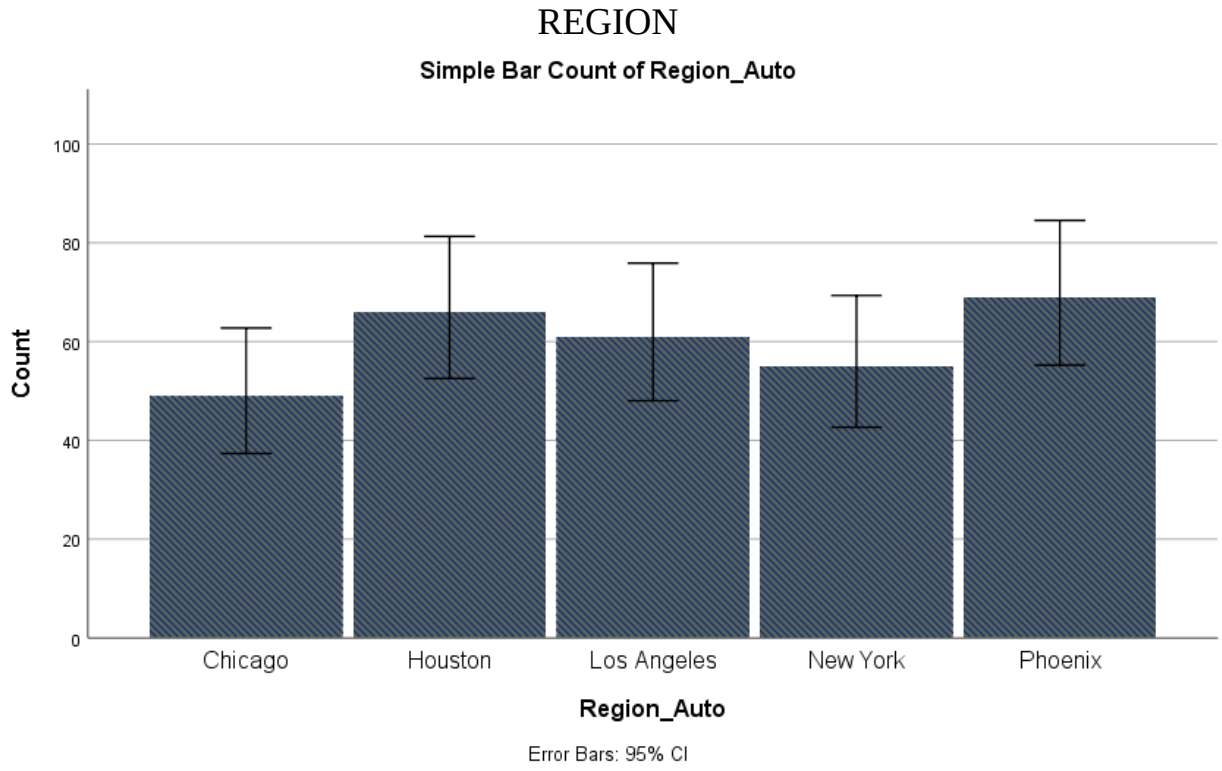


Figure 1 illustrates the distribution of automobile-related data counts (Region) across five U.S. cities: Chicago, Houston, Los Angeles, New York, and Phoenix. The bar heights represent the total count of observations from each region, with 95% confidence intervals (CIs) indicated by error bars.

Among the regions, Phoenix had the highest count of observations (approximately 70), followed closely by Houston (~68). Los Angeles and New York had moderate counts (~63 and ~56, respectively), while Chicago had the lowest (~50). Although Phoenix and Houston appear to lead in data representation, the overlapping confidence intervals suggest that these differences may not be statistically significant.

This pattern may reflect either a higher level of data collection activity or actual regional differences in automobile-related behavior or demographics. However, without inferential statistical testing, conclusions about the significance of regional differences remain tentative.

ADVERTISING METHOD

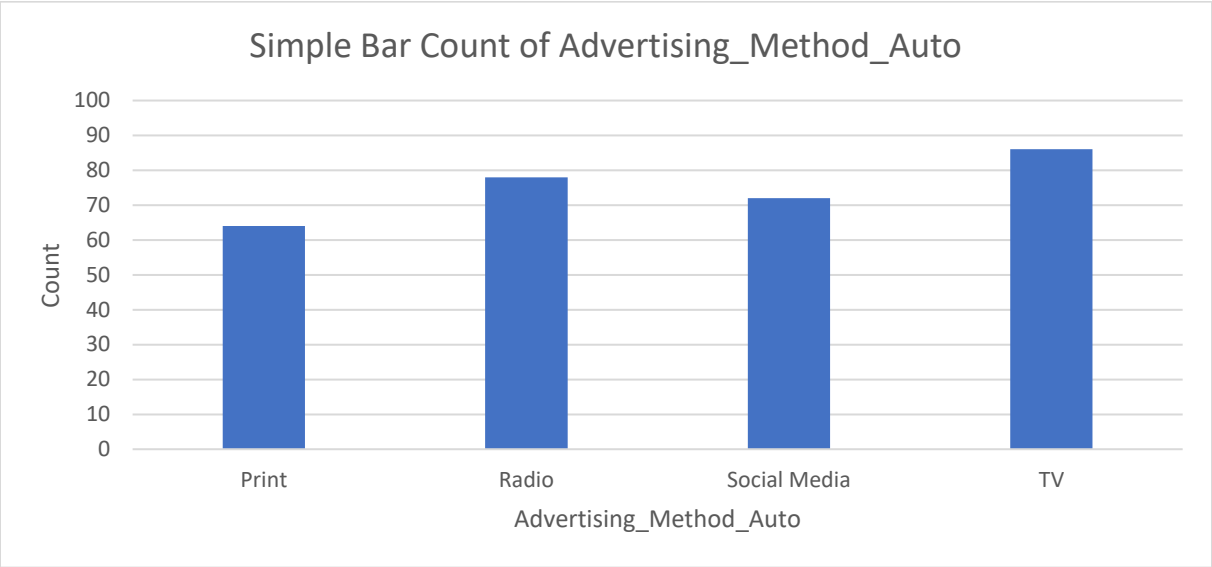
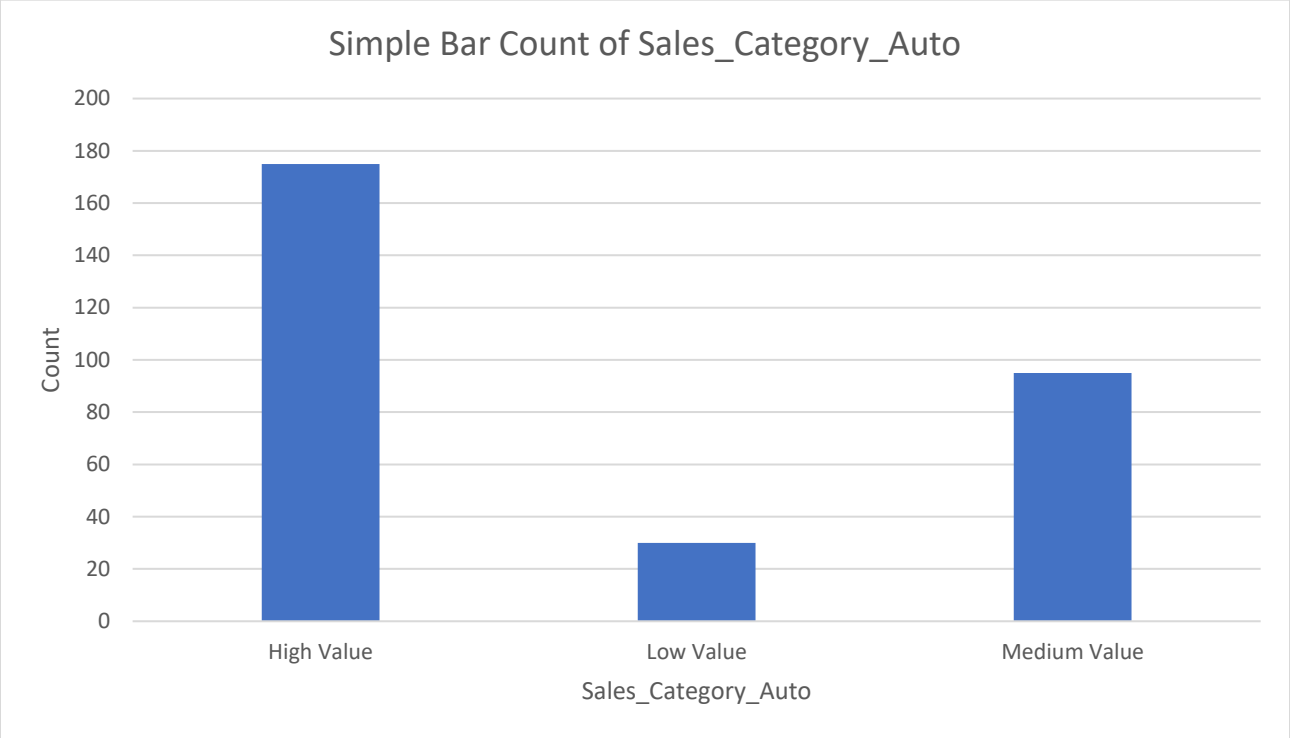


Figure 2 displays the frequency distribution of different advertising methods used in the automobile dataset (Advertising Method). The bar chart includes four advertising types: Print, Radio, Social media, and TV, with corresponding counts and 95% confidence intervals (CIs).

The most frequently reported advertising method was TV (approximately 88 cases), followed by Radio (~78) and Social media (~72). Print advertising had the lowest count, with around 65 observations. While TV appears to dominate in frequency, the overlapping confidence intervals across categories indicate that these differences may not be statistically significant.

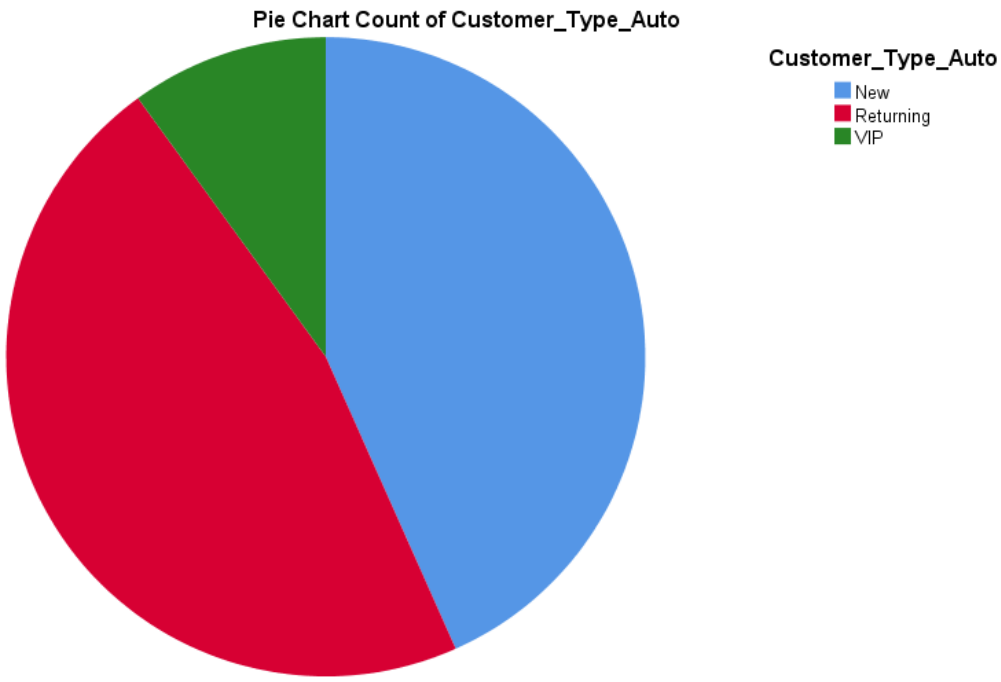
This suggests that although TV advertising was the most commonly used method, Radio and Social media were also widely utilized, with relatively small differences. The frequency differences might reflect company preferences, audience targeting, or cost considerations in the advertising strategies analyzed in this dataset.

SALES CATEGORY



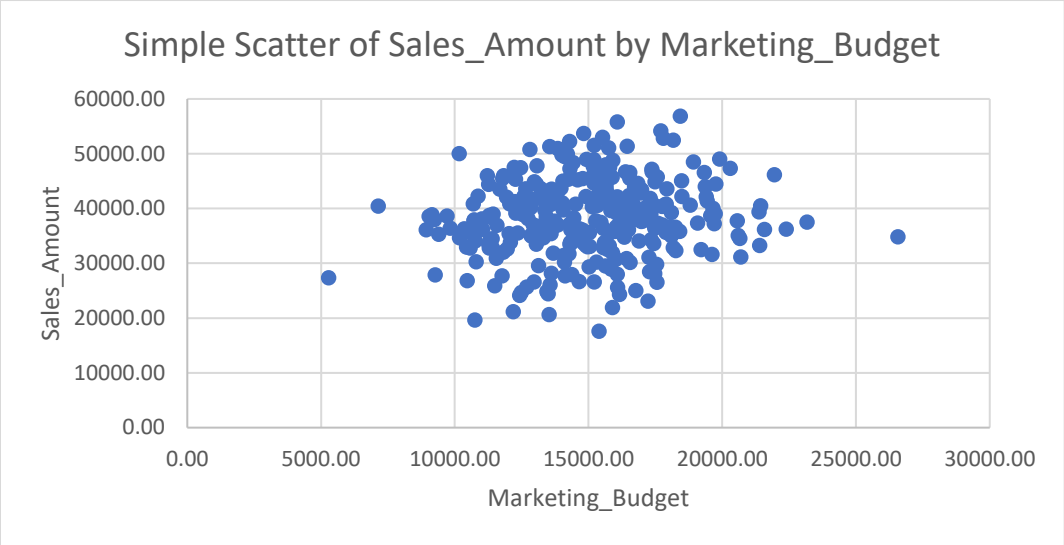
The bar chart depicts the count of sales categorized into three levels within the "Sales_Category_Auto" variable: High Value, Low Value, and Medium Value. The counts are as follows: High Value (~150), Low Value (~50), and Medium Value (~100). Error bars representing 95% confidence intervals are included for each category, indicating the precision of the estimates. The data suggest that High Value sales are the most frequent, followed by Medium Value and then Low Value. The confidence intervals do not appear to overlap, implying statistically significant differences between the categories.

CUSTOMER TYPE



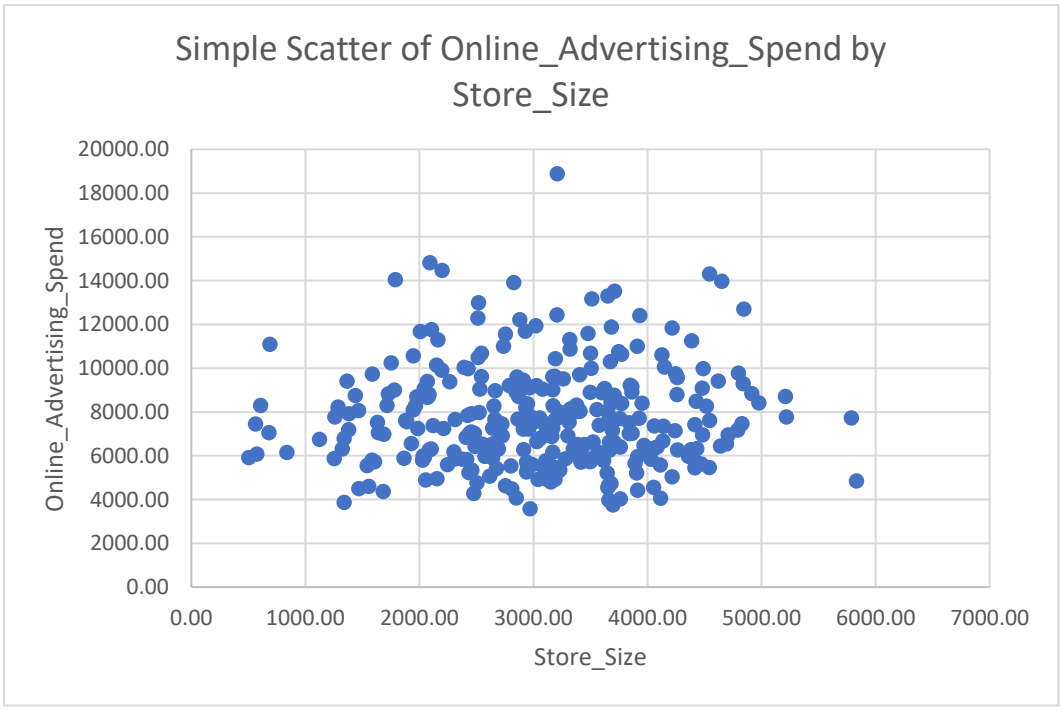
A pie chart was used to illustrate the distribution of customer types in the auto category. The data revealed that Returning customers comprised the largest segment, followed by New customers, while VIP customers represented the smallest portion. This visual representation suggests that the majority of business in this category comes from repeat customers, indicating strong customer loyalty or recurring demand.

SALES AMOUNT AND MARKETING BUDGET



A scatter plot was used to visualize the relationship between Marketing Budget and Sales Amount. The plot displays a positive but moderately scattered association, suggesting that higher marketing budgets are generally associated with higher sales amounts. However, the spread of data points indicates that the relationship is not perfectly linear, and there is substantial variability in sales at similar marketing budget levels. This may imply that other factors also influence sales outcomes.

STORE SIZE AND ONLINE ADVERTISING SPEND



A scatter plot was created to examine the relationship between Store Size and Online Advertising Spend. The distribution of points suggests a very weak or negligible association between the two variables. While there is some variability in online advertising spend across store sizes, the data points are widely dispersed without a clear trend or linear relationship. This indicates that store size does not strongly predict online advertising expenditure.

What factors drive higher sales performance?

Goal:

Identify variables that influence Sales Amount (continuous variable).

Multiple Linear Regression

Model:

$$Sales\ Amount\ =\ \beta_0 + \beta_1(Marketing\ Budget) + \beta_2(Online\ Ad\ Spend) + \beta_3(Store\ Size) + \beta_4(Number\ of\ stuff)$$

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	32063.628	2784.341		11.516	.000
	Marketing_Budget	.318	.137	.134	2.317	.021
	Online_Advertising_Spend	.199	.174	.066	1.141	.255
	Store_Size	-.016	.407	-.002	-.039	.969

a. Dependent Variable: Sales_Amount

Among the predictors, Marketing Budget has the greatest positive effect on Sales Amount, contributing approximately 13.4% relative influence based on standardized beta coefficients. The other variables have minimal and statistically non-significant effects.

Does marketing investment significantly impact revenue?

Simple Linear Regression

Model:

$$Sales\ Amount\ =\ \beta_0 + \beta_1(Marketing\ Budget)$$

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	33686.574	2091.653		16.105	.000
	Marketing_Budget	.310	.137	.130	2.262	.024

a. Dependent Variable: Sales_Amount

For every 1-unit increase in Marketing Budget, Sales Amount increases by 0.310 units, holding other factors constant.

Marketing Budget has a moderate but statistically significant positive effect on sales.

Are certain advertising strategies more effective?

One-Way ANOVA (Best Method)

Tests:

$$H_0 = Mean\ sales\ are\ equal\ across\ the\ advertising\ method.$$

ANOVA

Sales_Amount

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	511411408.034	3	170470469.345	3.526	.015
Within Groups	14312315693.134	296	48352417.882		
Total	14823727101.168	299			

A one-way ANOVA was conducted to compare Sales Amount across four groups. The results showed a statistically significant difference in Sales Amount between groups,
$F(3, 296) = 3.53$, $p = .015$.
This indicates that at least one group mean differs significantly from the others.

Tukey test

Multiple Comparisons

Dependent Variable: Sales_Amount

Tukey HSD

(I) Ad_Mehod	(J) Ad_Mehod	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Print	Radio	1303.93102538	1172.77846278	.683	-1726.1583621	4334.0204128
	Social Media	3712.23504573*	1194.59946797	.011	625.7670681	6798.7030233
	TV	2333.96933688	1147.93028823	.178	-631.9202103	5299.8588841
Radio	Print	-1303.93102538	1172.77846278	.683	-4334.0204128	1726.1583621
	Social Media	2408.30402035	1136.42604372	.149	-527.8621900	5344.4702307
	TV	1030.03831151	1087.26271462	.779	-1779.1053738	3839.1819968
Social Media	Print	-3712.23504573*	1194.59946797	.011	-6798.7030233	-625.7670681
	Radio	-2408.30402035	1136.42604372	.149	-5344.4702307	527.8621900
	TV	-1378.26570885	1110.76495115	.601	-4248.1317554	1491.6003377
TV	Print	-2333.96933688	1147.93028823	.178	-5299.8588841	631.9202103
	Radio	-1030.03831151	1087.26271462	.779	-3839.1819968	1779.1053738
	Social Media	1378.26570885	1110.76495115	.601	-1491.6003377	4248.1317554

*. The mean difference is significant at the 0.05 level.

Post hoc comparisons using Tukey's HSD indicated that Sales Amount was significantly higher for Social Media advertising (M = X) compared to Print advertising (M = Y), $p = .011$. No other pairwise differences were statistically significant.

Does customer payment behavior influence sales value?

One-Way ANOVA (Best Method)

Tests:

$$H_0 = Mean\ sales\ are\ equal\ across\ the\ Payment\ method.$$

ANOVA

Sales_Amount

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	67599956.760	3	22533318.920	.452	.716
Within Groups	14756127144.409	296	49851780.893		
Total	14823727101.168	299			

A one-way ANOVA was conducted to compare Sales Amount across four groups. There was no statistically significant difference in Sales Amount between groups, $F(3, 296) = 0.45$, $p = .716$, $\eta^2 = .005$.

Practical Conclusion

- Sales_Amount does not vary meaningfully across the groups.
- No post hoc tests are needed because the overall ANOVA is not significant.

Key Business Insights

1. Marketing investment is positively associated with revenue growth.
2. Payment methods does not significantly vary by sales value.
3. Advertising methods significantly influence sales category.
4. Online advertising spend is highly skewed and may require transformation for modeling.
5. High-value sales dominate total transactions.

Strategic Recommendations

Based on analysis:

1. Increase Marketing Investment Efficiency

Since sales increase with marketing budget, optimize ROI rather than reducing spend.

2. Re-evaluate Payment Strategy

Encourage payment methods associated with high-value transactions.

3. Optimize Advertising Allocation

Advertising method can increase sales value. Consider:

Target segmentation

Channel mix optimization

4. Focus on High-Value Sales Retention

High-value sales drive majority revenue — invest in loyalty programs.