Analyzing car.csv...

Shape: (38531, 7)

Column Data Types:

manufacturer_name object

model_name object

transmission object

color object

engine_has_gas bool

body_type object

has_warranty bool

Statistical Analysis (Numerical Columns):

	manufacturer_name	model_name	transmission	color	engine_has_gas	body_type	has_warranty
count	38531	38531	38531	38531	38531	38531	38531
unique	55	1118	2	12	2	12	2
top	Volkswagen	Passat	mechanical	black	False	sedan	False
freq	4243	1423	25633	7705	37184	13011	38082

Missing Values in Dataset:

No missing values found.

Dataset Division:

Numerical Columns: []

Categorical Columns: ['manufacturer_name', 'model_name', 'transmission', 'color', 'body_type']

Categorical dataset detected. Suggested optimization: Classification methods like Decision Trees or Random Forests.

Generated Equations for Numerical Columns:

Analyzing heart.csv...

Shape: (303, 14)

Column Data Types:

age int64

sex int64

cp int64

trestbps int64

chol int64

fbs int64

restecg int64

thalach int64

exang int64

oldpeak float64

slope int64

ca int64

thal int64

target int64

Statistical Analysis (Numerical Columns):

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
count	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000
mean	54.366337	0.683168	0.966997	131.623762	246.264026	0.148515	0.528053	149.646865	0.326733	1.039604	1.399340	0.729373	2.313531	0.544554
std	9.082101	0.466011	1.032052	17.538143	51.830751	0.356198	0.525860	22.905161	0.469794	1.161075	0.616226	1.022606	0.612277	0.498835
min	29.000000	0.000000	0.000000	94.000000	126.000000	0.000000	0.000000	71.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	47.500000	0.000000	0.000000	120.000000	211.000000	0.000000	0.000000	133.500000	0.000000	0.000000	1.000000	0.000000	2.000000	0.000000
50%	55.000000	1.000000	1.000000	130.000000	240.000000	0.000000	1.000000	153.000000	0.000000	0.800000	1.000000	0.000000	2.000000	1.000000
75%	61.000000	1.000000	2.000000	140.000000	274.500000	0.000000	1.000000	166.000000	1.000000	1.600000	2.000000	1.000000	3.000000	1.000000
max	77.000000	1.000000	3.000000	200.000000	564.000000	1.000000	2.000000	202.000000	1.000000	6.200000	2.000000	4.000000	3.000000	1.000000

Missing Values in Dataset:

No missing values found.

Dataset Division:

Numerical Columns: ['age', 'sex', 'cp', 'trestbps', 'chol', 'fbs', 'restecg', 'thalach', 'exang', 'oldpeak', 'slope', 'ca', 'thal', 'target']

Categorical Columns: []

Numerical dataset detected. Suggested optimization: Regression methods like Linear or Logistic Regression.

Generated Equations for Numerical Columns:

y = m * age + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'age' is the independent variable.

y = m * sex + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'sex' is the independent variable.

y = m * cp + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'cp' is the independent variable.

y = m * trestbps + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'trestbps' is the independent variable.

y = m * chol + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'chol' is the independent variable.

y = m * fbs + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'fbs' is the independent variable.

y = m * restecg + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'restecg' is the independent variable.

y = m * thalach + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'thalach' is the independent variable.

y = m * exang + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'exang' is the independent variable.

y = m * oldpeak + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'oldpeak' is the independent variable.

y = m * slope + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'slope' is the independent variable.

y = m * ca + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'ca' is the independent variable.

y = m * thal + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'thal' is the independent variable.

y = m * target + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'target' is the independent variable.

Analyzing phishing.csv...

Shape: (11054, 32)

Column Data Types:

Index int64

UsingIP int64

LongURL int64

ShortURL int64

Symbol@ int64

Redirecting// int64

PrefixSuffix- int64

SubDomains int64

HTTPS int64

DomainRegLen int64

Favicon int64

NonStdPort int64

HTTPSDomainURL int64

RequestURL int64

AnchorURL int64

LinksInScriptTags int64

ServerFormHandler int64

InfoEmail int64

AbnormalURL int64

WebsiteForwarding int64

StatusBarCust int64

DisableRightClick int64

UsingPopupWindow int64

IframeRedirection int64

AgeofDomain int64

DNSRecording int64

WebsiteTraffic int64

PageRank int64

GoogleIndex int64

LinksPointingToPage int64

StatsReport int64

class int64

Statistical Analysis (Numerical Columns):

Index UsingIP LongURL ShortURL Symbol@ Redirecting// PrefixSuffix-
SubDomains HTTPS DomainRegLen Favicon NonStdPort HTTPSDomainURL
RequestURL AnchorURL LinksInScriptTags ServerFormHandler InfoEmail AbnormalURL
WebsiteForwarding StatusBarCust DisableRightClick UsingPopupWindow IframeRedirection
AgeofDomain DNSRecording WebsiteTraffic PageRank GoogleIndex LinksPointingToPage
StatsReport class

count 11054.000000 11054.000000 11054.000000 11054.000000 11054.000000
11054.000000 11054.000000 11054.000000 11054.000000 11054.000000 11054.000000
11054.000000 11054.000000 11054.000000 11054.000000 11054.000000
11054.000000 11054.000000 11054.000000 11054.000000 11054.000000 11054.000000
11054.000000 11054.000000 11054.000000 11054.000000 11054.000000
11054.000000 11054.000000 11054.000000 11054.000000 11054.000000
11054.000000 11054.000000 11054.000000 11054.000000 11054.000000

mean 5526.500000 0.313914 -0.633345 0.738737 0.700561 0.741632
-0.734938 0.064049 0.251040 -0.336711 0.628551 0.728243 0.675231
0.186720 -0.076443 -0.118238 -0.595712 0.635788 0.705446 0.115705
0.762077 0.913877 0.613353 0.816899 0.061335 0.377239
0.287407 -0.483626 0.721549 0.343948 0.719739 0.113986

std 3191.159272 0.949495 0.765973 0.674024 0.713625 0.670837 0.678165

0.817492 0.911856 0.941651 0.777804 0.685350 0.737640 0.982458
0.715116 0.763933 0.759168 0.771899 0.708796 0.319885 0.647516
0.406009 0.789845 0.576807 0.998162 0.926158 0.827680
0.875314 0.692395 0.569936 0.694276 0.993527
min 0.000000 -1.000000 -1.000000 -1.000000 -1.000000 -1.000000 -1.000000
-1.000000 -1.000000 -1.000000 -1.000000 -1.000000 -1.000000 -1.000000
-1.000000 -1.000000 -1.000000 -1.000000 -1.000000 0.000000 -1.000000
-1.000000 -1.000000 -1.000000 -1.000000 -1.000000 -1.000000
-1.000000 -1.000000 -1.000000 -1.000000 -1.000000
25% 2763.250000 -1.000000 -1.000000 1.000000 1.000000 1.000000
-1.000000 -1.000000 -1.000000 -1.000000 1.000000 1.000000 1.000000
-1.000000 -1.000000 -1.000000 -1.000000 1.000000 1.000000 0.000000
1.000000 1.000000 1.000000 1.000000 -1.000000 -1.000000
0.000000 -1.000000 1.000000 0.000000 1.000000 -1.000000
50% 5526.500000 1.000000 -1.000000 1.000000 1.000000 1.000000
-1.000000 0.000000 1.000000 -1.000000 1.000000 1.000000 1.000000
1.000000 0.000000 0.000000 -1.000000 1.000000 1.000000 0.000000
1.000000 1.000000 1.000000 1.000000 1.000000 1.000000
1.000000 -1.000000 1.000000 0.000000 1.000000 1.000000
75% 8289.750000 1.000000 -1.000000 1.000000 1.000000 1.000000
-1.000000 1.000000 1.000000 1.000000 1.000000 1.000000 1.000000
1.000000 0.000000 0.000000 -1.000000 1.000000 1.000000 0.000000
1.000000 1.000000 1.000000 1.000000 1.000000 1.000000
1.000000 1.000000 1.000000 1.000000 1.000000 1.000000
max 11053.000000 1.000000 1.000000 1.000000 1.000000 1.000000
1.000000 1.000000 1.000000 1.000000 1.000000 1.000000 1.000000

1.000000 1.000000 1.000000 1.000000 1.000000 1.000000 1.000000

1.000000 1.000000 1.000000 1.000000 1.000000 1.000000

1.000000 1.000000 1.000000 1.000000 1.000000 1.000000

Missing Values in Dataset:

No missing values found.

Dataset Division:

Numerical Columns: ['Index', 'UsingIP', 'LongURL', 'ShortURL', 'Symbol@', 'Redirecting//', 'PrefixSuffix-', 'SubDomains', 'HTTPS', 'DomainRegLen', 'Favicon', 'NonStdPort', 'HTTPSDomainURL', 'RequestURL', 'AnchorURL', 'LinksInScriptTags', 'ServerFormHandler', 'InfoEmail', 'AbnormalURL', 'WebsiteForwarding', 'StatusBarCust', 'DisableRightClick', 'UsingPopupWindow', 'IframeRedirection', 'AgeofDomain', 'DNSRecording', 'WebsiteTraffic', 'PageRank', 'GoogleIndex', 'LinksPointingToPage', 'StatsReport', 'class']

Categorical Columns: []

Numerical dataset detected. Suggested optimization: Regression methods like Linear or Logistic Regression.

Generated Equations for Numerical Columns:

y = m * Index + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'Index' is the independent variable.

y = m * UsingIP + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'UsingIP' is the independent variable.

y = m * LongURL + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'LongURL' is the independent variable.

y = m * ShortURL + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'ShortURL' is the independent variable.

y = m * Symbol@ + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'Symbol@' is the independent variable.

y = m * Redirecting// + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'Redirecting//' is the independent variable.

y = m * PrefixSuffix- + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'PrefixSuffix-' is the independent variable.

y = m * SubDomains + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'SubDomains' is the independent variable.

y = m * HTTPS + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'HTTPS' is the independent variable.

y = m * DomainRegLen + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'DomainRegLen' is the independent variable.

y = m * Favicon + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'Favicon' is

the independent variable.

y = m * NonStdPort + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'NonStdPort' is the independent variable.

y = m * HTTPSDomainURL + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'HTTPSDomainURL' is the independent variable.

y = m * RequestURL + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'RequestURL' is the independent variable.

y = m * AnchorURL + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'AnchorURL' is the independent variable.

y = m * LinksInScriptTags + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'LinksInScriptTags' is the independent variable.

y = m * ServerFormHandler + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'ServerFormHandler' is the independent variable.

y = m * InfoEmail + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'InfoEmail' is the independent variable.

y = m * AbnormalURL + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'AbnormalURL' is the independent variable.

y = m * WebsiteForwarding + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'WebsiteForwarding' is the independent variable.

y = m * StatusBarCust + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'StatusBarCust' is the independent variable.

y = m * DisableRightClick + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'DisableRightClick' is the independent variable.

y = m * UsingPopupWindow + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'UsingPopupWindow' is the independent variable.

y = m * IframeRedirection + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'IframeRedirection' is the independent variable.

y = m * AgeofDomain + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'AgeofDomain' is the independent variable.

y = m * DNSRecording + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'DNSRecording' is the independent variable.

y = m * WebsiteTraffic + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'WebsiteTraffic' is the independent variable.

y = m * PageRank + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'PageRank' is the independent variable.

y = m * GoogleIndex + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'GoogleIndex' is the independent variable.

y = m * LinksPointingToPage + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'LinksPointingToPage' is the independent variable.

y = m * StatsReport + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'StatsReport' is the independent variable.

y = m * class + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'class' is the independent variable.

Analyzing winequality-red.csv...

Shape: (1599, 12)

Column Data Types:

fixed acidity float64

volatile acidity float64

citric acid float64

residual sugar float64

chlorides float64

free sulfur dioxide float64

total sulfur dioxide float64

density float64

pH float64

sulphates float64

alcohol float64

quality int64

Statistical Analysis (Numerical Columns):

fixed acidity volatile acidity citric acid residual sugar chlorides free sulfur dioxide total sulfur dioxide density pH sulphates alcohol quality

count 1599.000000 1599.000000 1599.000000 1599.000000 1599.000000 1599.000000 1599.000000 1599.000000 1599.000000 1599.000000 1599.000000

mean 8.319637 0.527821 0.270976 2.538806 0.087467 15.874922 46.467792 0.996747 3.311113 0.658149 10.422983 5.636023

std 1.741096 0.179060 0.194801 1.409928 0.047065 10.460157 32.895324 0.001887 0.154386 0.169507 1.065668 0.807569

min 4.600000 0.120000 0.000000 0.900000 0.012000 1.000000 6.000000 0.990070 2.740000 0.330000 8.400000 3.000000

25% 7.100000 0.390000 0.090000 1.900000 0.070000 7.000000 22.000000 0.995600 3.210000 0.550000 9.500000 5.000000

50% 7.900000 0.520000 0.260000 2.200000 0.079000 14.000000 38.000000 0.996750 3.310000 0.620000 10.200000 6.000000

75% 9.200000 0.640000 0.420000 2.600000 0.090000 21.000000 62.000000 0.997835 3.400000 0.730000 11.100000 6.000000

max 15.900000 1.580000 1.000000 15.500000 0.611000 72.000000 289.000000 1.003690 4.010000 2.000000 14.900000 8.000000

Missing Values in Dataset:

No missing values found.

Dataset Division:

Numerical Columns: ['fixed acidity', 'volatile acidity', 'citric acid', 'residual sugar', 'chlorides', 'free sulfur dioxide', 'total sulfur dioxide', 'density', 'pH', 'sulphates', 'alcohol', 'quality']

Categorical Columns: []

Numerical dataset detected. Suggested optimization: Regression methods like Linear or Logistic Regression.

Generated Equations for Numerical Columns:

y = m * fixed acidity + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'fixed acidity' is the independent variable.

y = m * volatile acidity + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'volatile acidity' is the independent variable.

y = m * citric acid + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'citric acid' is the independent variable.

y = m * residual sugar + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'residual sugar' is the independent variable.

y = m * chlorides + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'chlorides' is the independent variable.

y = m * free sulfur dioxide + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'free sulfur dioxide' is the independent variable.

y = m * total sulfur dioxide + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'total sulfur dioxide' is the independent variable.

y = m * density + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'density' is the independent variable.

y = m * pH + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'pH' is the independent variable.

y = m * sulphates + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'sulphates' is the independent variable.

y = m * alcohol + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'alcohol' is the independent variable.

y = m * quality + b (Linear Regression)

Reason: This equation suggests a linear relationship where 'y' is the predicted value and 'quality' is the independent variable.