

Data Analysis Report

Analysis of: train.csv

Generated on: June 21, 2025 at 16:59

Executive Summary

Dataset Overview	
Total Records	9,800
Total Columns	18
Memory Usage	9034.93 KB
Missing Values	11

Column Information

Column Name	Data Type	Missing Values	Sample Value
Row ID	int64	0	1
Order ID	object	0	CA-2017-152156
Order Date	object	0	08/11/2017
Ship Date	object	0	11/11/2017
Ship Mode	object	0	Second Class
Customer ID	object	0	CG-12520
Customer Name	object	0	Claire Gute
Segment	object	0	Consumer
Country	object	0	United States
City	object	0	Henderson
State	object	0	Kentucky
Postal Code	float64	11	42420.0
Region	object	0	South
Product ID	object	0	FUR-BO-10001798
Category	object	0	Furniture
Sub-Category	object	0	Bookcases
Product Name	object	0	Bush Somerset Collec...
Sales	float64	0	261.96

Numeric Data Analysis

Column	Count	Mean	Std	Min	Max
--------	-------	------	-----	-----	-----

Row ID	9800	4900.50	2829.16	1.00	9800.00
Postal Code	9789	55273.32	32041.22	1040.00	99301.00
Sales	9800	230.77	626.65	0.44	22638.48

Categorical Data Analysis

Order ID

Value	Count
CA-2018-100111	14
CA-2018-157987	12
CA-2017-165330	11
US-2017-108504	11
CA-2017-105732	10
US-2016-126977	10
CA-2016-131338	10
CA-2016-132626	9
US-2016-163433	9
US-2017-114013	9

Unique values: 4922, Most frequent: CA-2018-100111

Order Date

Value	Count
05/09/2017	38
10/11/2017	35
01/12/2018	34
02/12/2018	34
02/09/2018	33
09/12/2018	33
08/12/2018	30
12/11/2018	30
09/09/2018	29
11/12/2017	28

Unique values: 1230, Most frequent: 05/09/2017

Ship Date

Value	Count
26/09/2018	34
06/12/2018	32
16/12/2016	31
06/09/2018	30
15/09/2018	30
12/12/2018	30
21/11/2018	30
08/09/2018	27
13/09/2015	27
26/09/2016	26

Unique values: 1326, Most frequent: 26/09/2018

Ship Mode

Value	Count
Standard Class	5859
Second Class	1902
First Class	1501
Same Day	538

Unique values: 4, Most frequent: Standard Class

Customer ID

Value	Count
WB-21850	35
PP-18955	34
MA-17560	34
JL-15835	33
JD-15895	32
CK-12205	32

SV-20365	32
ZC-21910	31
AP-10915	31
EP-13915	31

Unique values: 793, Most frequent: WB-21850

Customer Name

Value	Count
William Brown	35
Paul Prost	34
Matt Abelman	34
John Lee	33
Jonathan Doherty	32
Chloris Kastensmidt	32
Seth Vernon	32
Zuschuss Carroll	31
Arthur Prichep	31
Emily Phan	31

Unique values: 793, Most frequent: William Brown

Segment

Value	Count
Consumer	5101
Corporate	2953
Home Office	1746

Unique values: 3, Most frequent: Consumer

Country

Value	Count
United States	9800

Unique values: 1, Most frequent: United States

City

Value	Count
New York City	891
Los Angeles	728
Philadelphia	532
San Francisco	500
Seattle	426
Houston	374
Chicago	308
Columbus	221
San Diego	170
Springfield	161

Unique values: 529, Most frequent: New York City

State

Value	Count
California	1946
New York	1097
Texas	973
Pennsylvania	582
Washington	504
Illinois	483
Ohio	454
Florida	373
Michigan	253
North Carolina	247

Unique values: 49, Most frequent: California

Region

Value	Count
-------	-------

West	3140
East	2785
Central	2277
South	1598

Unique values: 4, Most frequent: West

Product ID

Value	Count
OFF-PA-10001970	19
TEC-AC-10003832	18
FUR-FU-10004270	16
FUR-CH-10002647	15
TEC-AC-10002049	15
TEC-AC-10003628	15
OFF-PA-10002377	14
OFF-BI-10004632	14
FUR-CH-10003774	14
OFF-BI-10001524	14

Unique values: 1861, Most frequent: OFF-PA-10001970

Category

Value	Count
Office Supplies	5909
Furniture	2078
Technology	1813

Unique values: 3, Most frequent: Office Supplies

Sub-Category

Value	Count
Binders	1492
Paper	1338

Furnishings	931
Phones	876
Storage	832
Art	785
Accessories	756
Chairs	607
Appliances	459
Labels	357

Unique values: 17, Most frequent: Binders

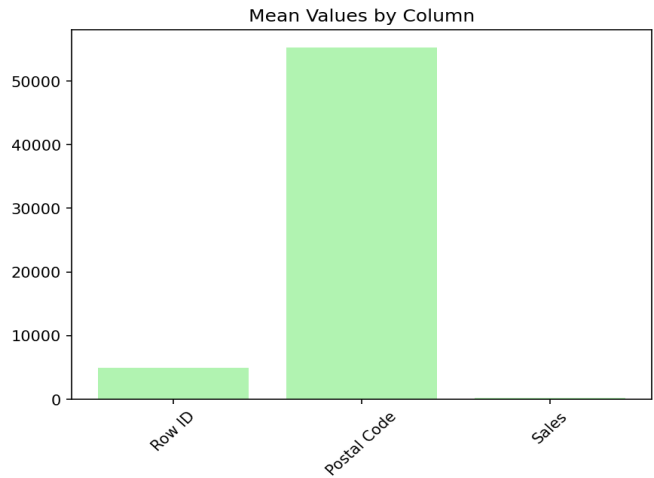
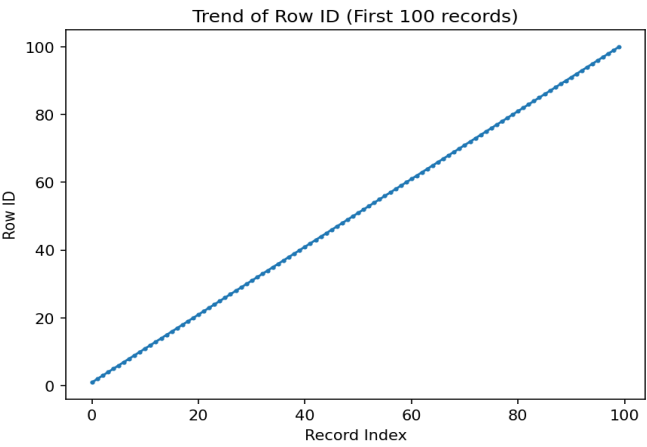
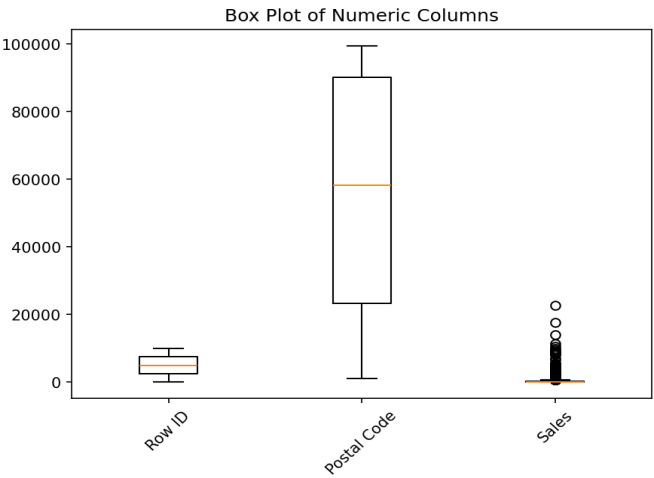
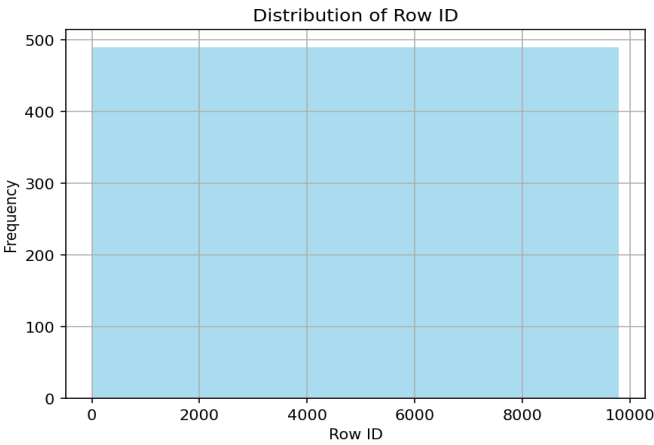
Product Name

Value	Count
Staple envelope	47
Staples	46
Easy-staple paper	44
Avery Non-Stick Binders	20
Staples in misc. colors	18
Staple remover	18
Storex Dura Pro Binders	17
KI Adjustable-Height Table	17
Staple-based wall hangings	16
Situations Contoured Folding C...	15

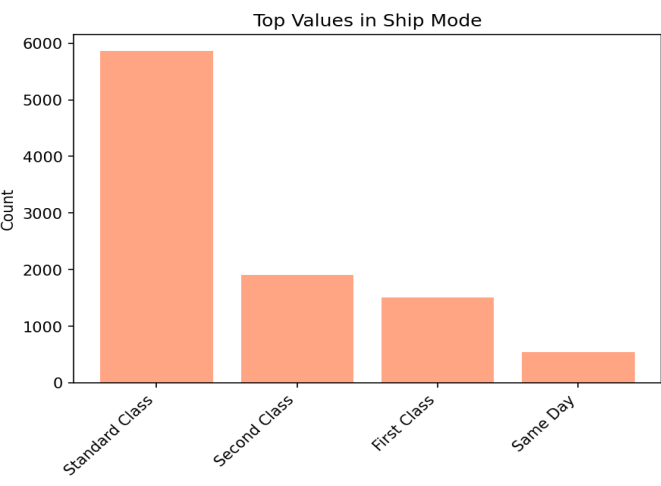
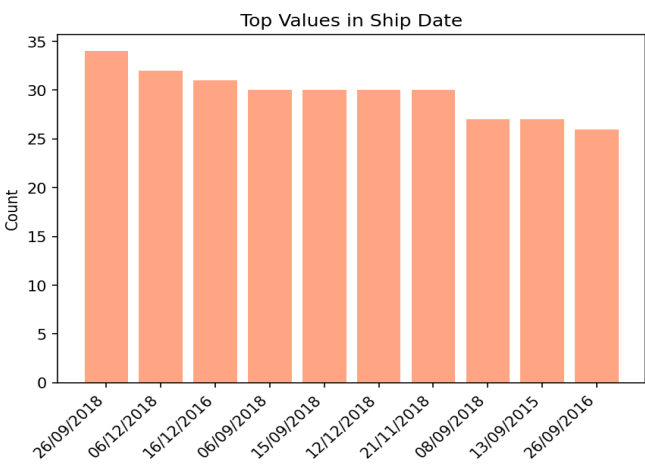
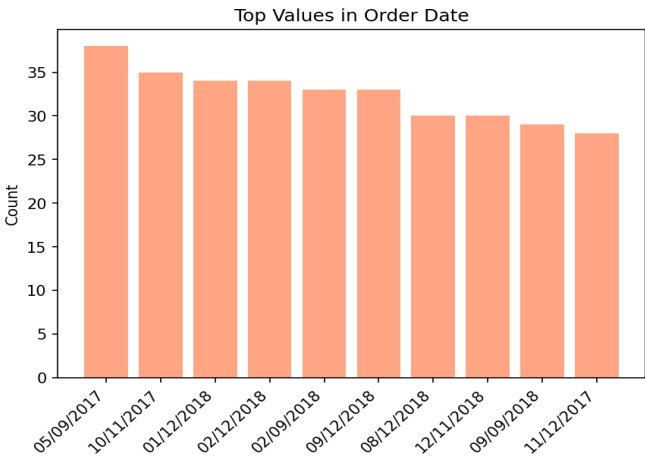
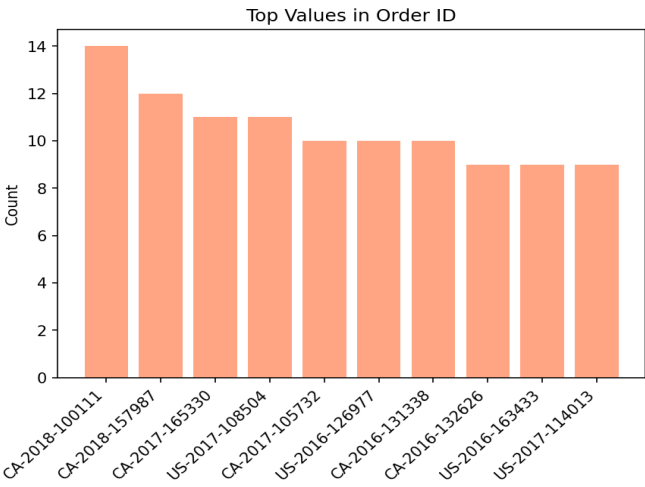
Unique values: 1849, Most frequent: Staple envelope

Data Visualizations

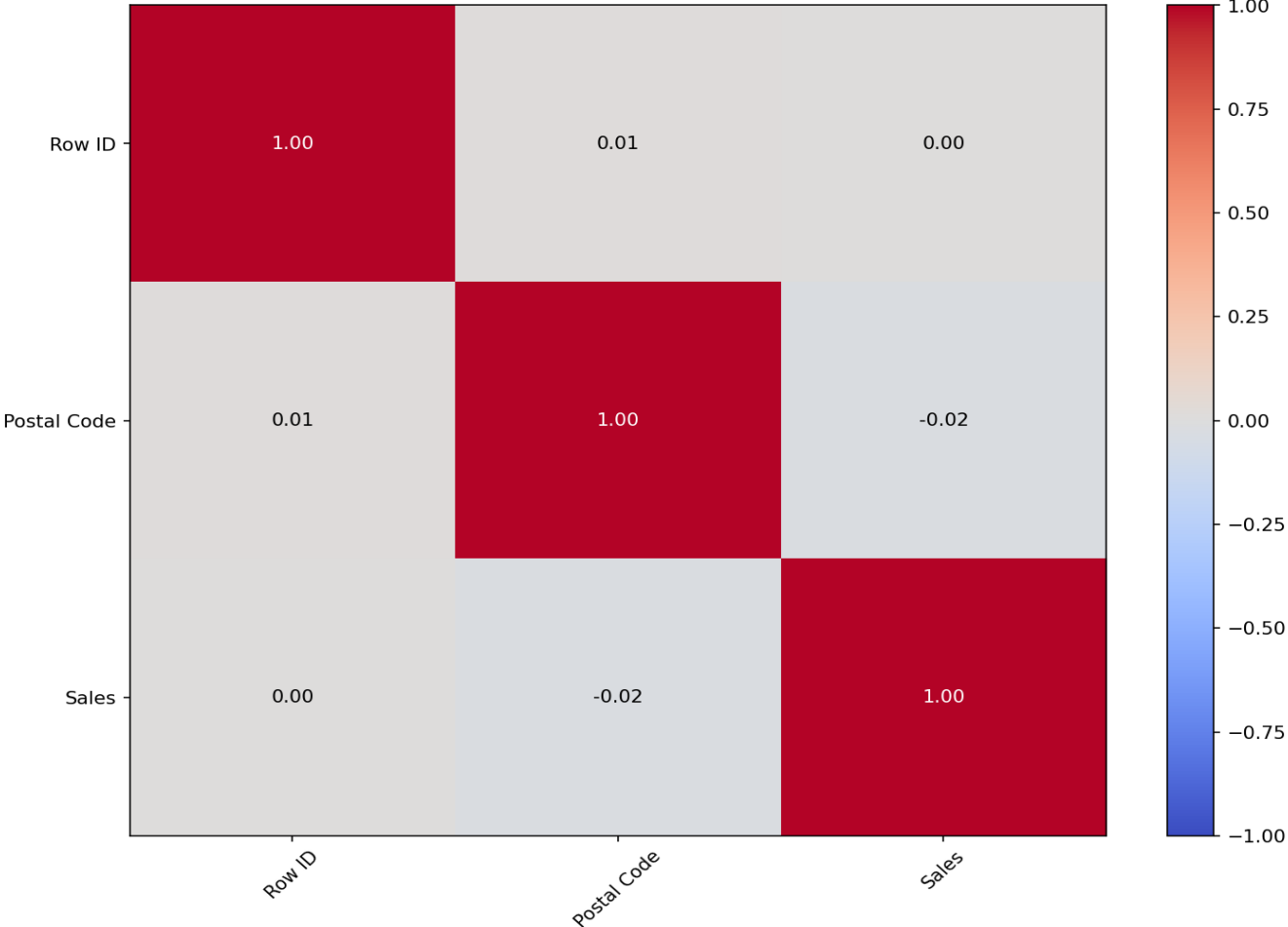
Numeric Data Analysis



Categorical Data Analysis



Correlation Matrix



Conclusion & Insights

In the column **Order ID**, the most frequent value is **CA-2018-100111**.

The column **Postal Code** has the highest number of missing values: **11**.

Among numeric columns, **Postal Code** has the highest mean value of **55273.32**.