Automated EDA Report

The file has 9994 rows and 21 columns.

Column names in the file are:
['Row ID', 'Order ID', 'Order Date', 'Ship Date', 'Ship Mode', 'Customer ID', 'Customer Name',
'Segment', 'Country', 'City', 'State', 'Postal Code', 'Region', 'Product ID', 'Category',
'Sub-Category', 'Product Name', 'Sales', 'Quantity', 'Discount', 'Profit']

Data types of each column:
Row ID   int64
Order ID   object
Order Date   object
Ship Date   object
Ship Mode   object
Customer ID   object
Customer Name   object
Segment   object
Country   object
City   object
State   object
Postal Code   int64
Region   object
Product ID   object
Category   object
Sub-Category   object
Product Name   object
Sales   float64
Quantity   int64
Discount   float64
Profit   float64
dtype: object

Number of unique values in each column:
Row ID   9994
Order ID   5009
Order Date   1237
Ship Date   1334
Ship Mode   4
Customer ID   793
Customer Name   793
Segment   3
Country   1
City   531
State   49
Postal Code   631
Region   4
Product ID   1862
Category   3
Sub-Category   17
Product Name   1850
Sales   5825
Quantity   14
Discount   12
Profit   7287
dtype: int64

Number of null values in each column:
Row ID   0
Order ID   0
Order Date   0

```
Ship Date         0
Ship Mode         0
Customer ID       0
Customer Name     0
Segment           0
Country           0
City              0
State             0
Postal Code       0
Region            0
Product ID        0
Category          0
Sub-Category      0
Product Name      0
Sales             0
Quantity          0
Discount          0
Profit            0
dtype: int64
```

```
Percentage of missing values in each column:
Row ID            0.0
Order ID          0.0
Order Date        0.0
Ship Date         0.0
Ship Mode         0.0
Customer ID       0.0
Customer Name     0.0
Segment           0.0
Country           0.0
City              0.0
State             0.0
Postal Code       0.0
Region            0.0
Product ID        0.0
Category          0.0
Sub-Category      0.0
Product Name      0.0
Sales             0.0
Quantity          0.0
Discount          0.0
Profit            0.0
dtype: float64
```

Total duplicate rows in the dataset: 0

Summary statistics for numeric columns:

	Row ID	Postal Code	Sales	Quantity	Discount	Profit
count	9994.00	9994.00	9994.00	9994.00	9994.00	9994.00
mean	4997.50	55190.38	229.86	3.79	0.16	28.66
std	2885.16	32063.69	623.25	2.23	0.21	234.26
min	1.00	1040.00	0.44	1.00	0.00	-6599.98
25%	2499.25	23223.00	17.28	2.00	0.00	1.73
50%	4997.50	56430.50	54.49	3.00	0.20	8.67
75%	7495.75	90008.00	209.94	5.00	0.20	29.36
max	9994.00	99301.00	22638.48	14.00	0.80	8399.98

Correlation matrix (numeric columns only):

	Row ID	Postal Code	Sales	Quantity	Discount	Profit
Row ID	1.00	0.01	-0.00	-0.00	0.01	0.01
Postal Code	0.01	1.00	-0.02	0.01	0.06	-0.03

Automated EDA Report

```
Sales        -0.00        -0.02   1.00     0.20     -0.03    0.48
Quantity     -0.00         0.01   0.20     1.00      0.01    0.07
Discount      0.01         0.06  -0.03     0.01      1.00   -0.22
Profit        0.01        -0.03   0.48     0.07     -0.22    1.00
--------------------------------------------------------------------------------

Number of outliers detected in each numeric column (IQR method):
{'Row ID': 9994, 'Postal Code': 9994, 'Sales': 8827, 'Quantity': 9824, 'Discount': 9072, 'Profit':
8717}
--------------------------------------------------------------------------------
```