

Data Collection and Preprocessing Phase

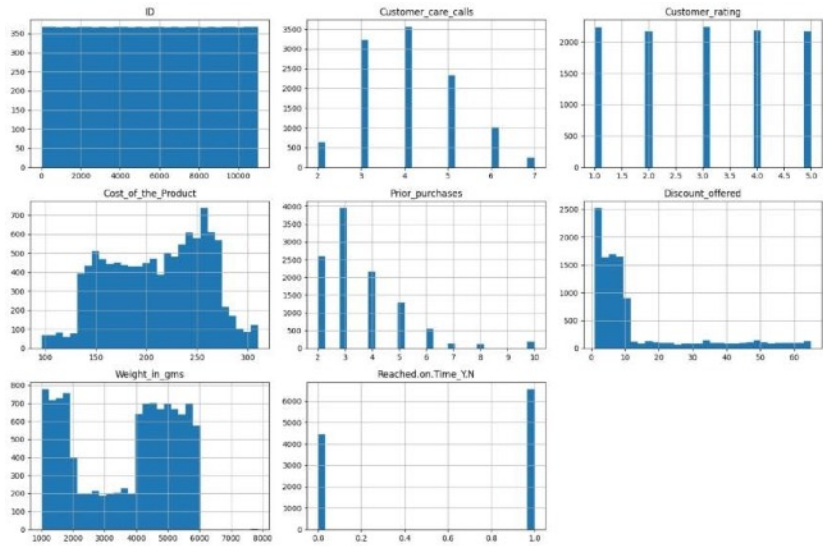
Date	15 March 2024
Team ID	SWTID1720519736
Project Title	Ecommerce Shipping Prediction Using Machine Learning
Maximum Marks	6 Marks

Data Exploration and Preprocessing Template

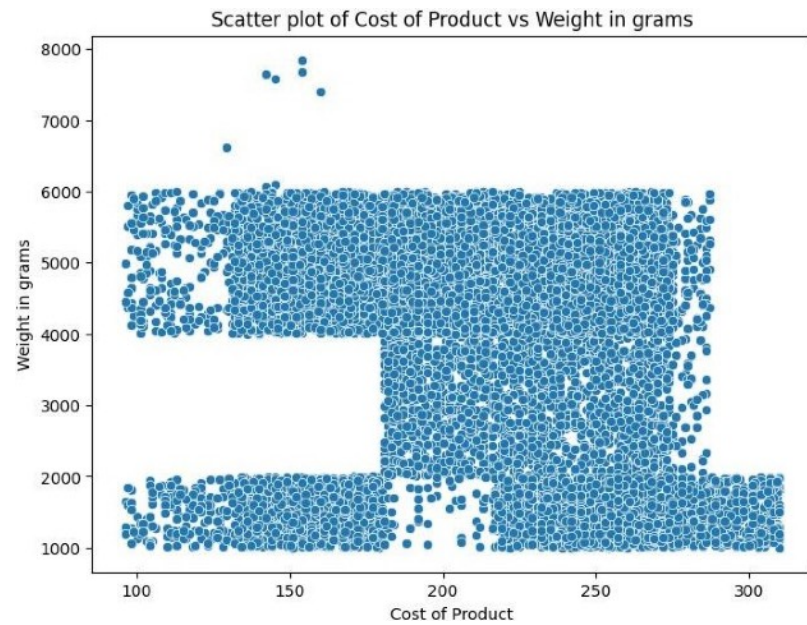
Identifies data sources, assesses quality issues like missing values and duplicates, and implements resolution plans to ensure accurate and reliable analysis.

Section	Description																																																															
Data Overview	Dimension: 10999 rows X 12 Columns																																																															
	<table><tr><th></th><th>ID</th><th>Warehouse block</th><th>Mode of Shipment</th><th>Customer care calls</th><th>Customer rating</th><th>Cost of the Product</th></tr><tr><td>count</td><td>10999.00000</td><td>10999.00000</td><td>10999.00000</td><td>10999.00000</td><td>10999.00000</td><td>10999.00000</td></tr><tr><td>mean</td><td>5500.00000</td><td>1.833167</td><td>0.998454</td><td>4.054459</td><td>2.990545</td><td>210.196836</td></tr><tr><td>std</td><td>3175.28214</td><td>1.343823</td><td>0.567099</td><td>1.141490</td><td>1.413603</td><td>48.063272</td></tr><tr><td>min</td><td>1.00000</td><td>0.000000</td><td>0.000000</td><td>2.000000</td><td>1.000000</td><td>96.000000</td></tr><tr><td>25%</td><td>2750.50000</td><td>1.000000</td><td>1.000000</td><td>3.000000</td><td>2.000000</td><td>169.000000</td></tr><tr><td>50%</td><td>5500.00000</td><td>1.000000</td><td>1.000000</td><td>4.000000</td><td>3.000000</td><td>214.000000</td></tr><tr><td>75%</td><td>8249.50000</td><td>3.000000</td><td>1.000000</td><td>5.000000</td><td>4.000000</td><td>251.000000</td></tr><tr><td>max</td><td>10999.00000</td><td>4.000000</td><td>2.000000</td><td>7.000000</td><td>5.000000</td><td>310.000000</td></tr></table>		ID	Warehouse block	Mode of Shipment	Customer care calls	Customer rating	Cost of the Product	count	10999.00000	10999.00000	10999.00000	10999.00000	10999.00000	10999.00000	mean	5500.00000	1.833167	0.998454	4.054459	2.990545	210.196836	std	3175.28214	1.343823	0.567099	1.141490	1.413603	48.063272	min	1.00000	0.000000	0.000000	2.000000	1.000000	96.000000	25%	2750.50000	1.000000	1.000000	3.000000	2.000000	169.000000	50%	5500.00000	1.000000	1.000000	4.000000	3.000000	214.000000	75%	8249.50000	3.000000	1.000000	5.000000	4.000000	251.000000	max	10999.00000	4.000000	2.000000	7.000000	5.000000	310.000000
		ID	Warehouse block	Mode of Shipment	Customer care calls	Customer rating	Cost of the Product																																																									
	count	10999.00000	10999.00000	10999.00000	10999.00000	10999.00000	10999.00000																																																									
	mean	5500.00000	1.833167	0.998454	4.054459	2.990545	210.196836																																																									
	std	3175.28214	1.343823	0.567099	1.141490	1.413603	48.063272																																																									
	min	1.00000	0.000000	0.000000	2.000000	1.000000	96.000000																																																									
	25%	2750.50000	1.000000	1.000000	3.000000	2.000000	169.000000																																																									
	50%	5500.00000	1.000000	1.000000	4.000000	3.000000	214.000000																																																									
	75%	8249.50000	3.000000	1.000000	5.000000	4.000000	251.000000																																																									
max	10999.00000	4.000000	2.000000	7.000000	5.000000	310.000000																																																										

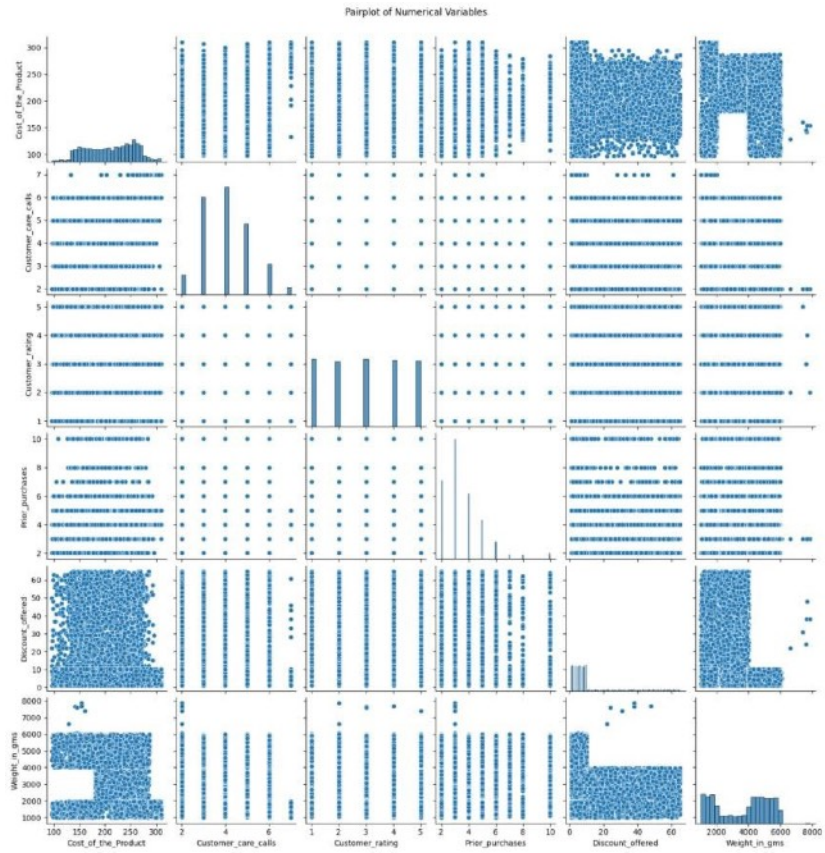
Univariate Analysis



Bivariate Analysis



Multivariate Analysis



Outliers and Anomalies

N/A

Data Preprocessing Code Screenshots

Loading Data

```
data = pd.read_csv("Train.csv")
```

✓ 0.0s Python

```
data.head()
```

Python

	ID	Warehouse block	Mode of Shipment	Customer care calls	Customer_rating	Cost of the Product	Prior purcha
0	1	D	Flight	4	2	177	
1	2	F	Flight	4	5	216	
2	3	A	Flight	2	2	183	
3	4	B	Flight	3	3	176	
4	5	C	Flight	2	2	184	

Handling Missing Data	N/A
Data Transformation	N/A
Feature Engineering	Code attached in final submission
Save Processed Data	N/A