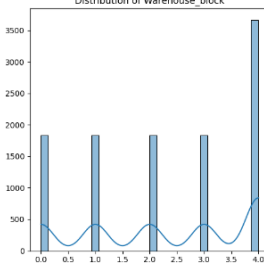
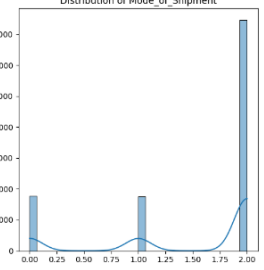
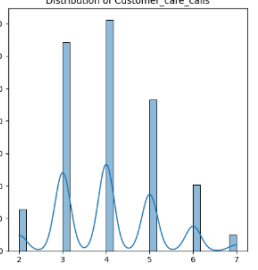
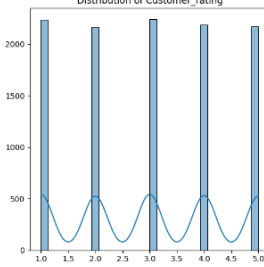
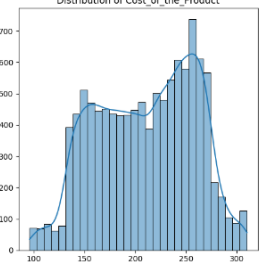
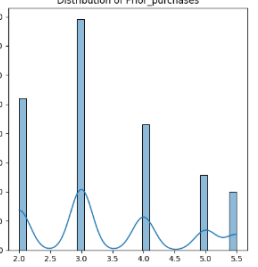


Data Collection and Preprocessing Phase

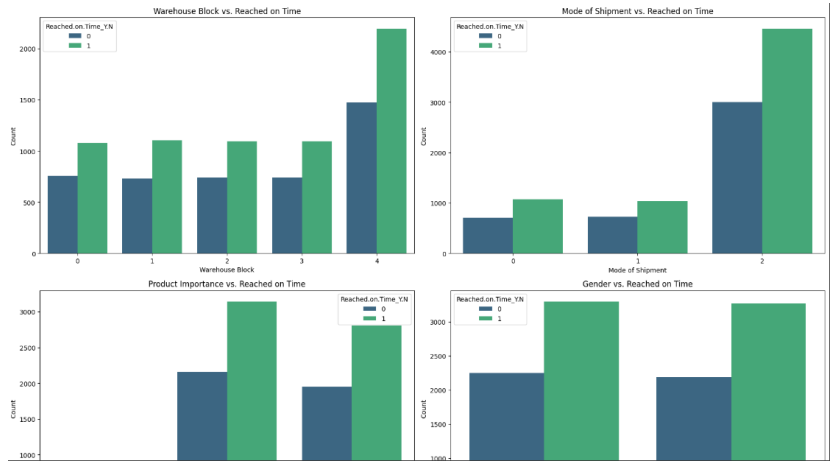
Date	25 June 2024
Team ID	SWTID1720434734
Project Title	Ecommerce Shipping Prediction
Maximum Marks	6 Marks

Data Exploration and Preprocessing Template

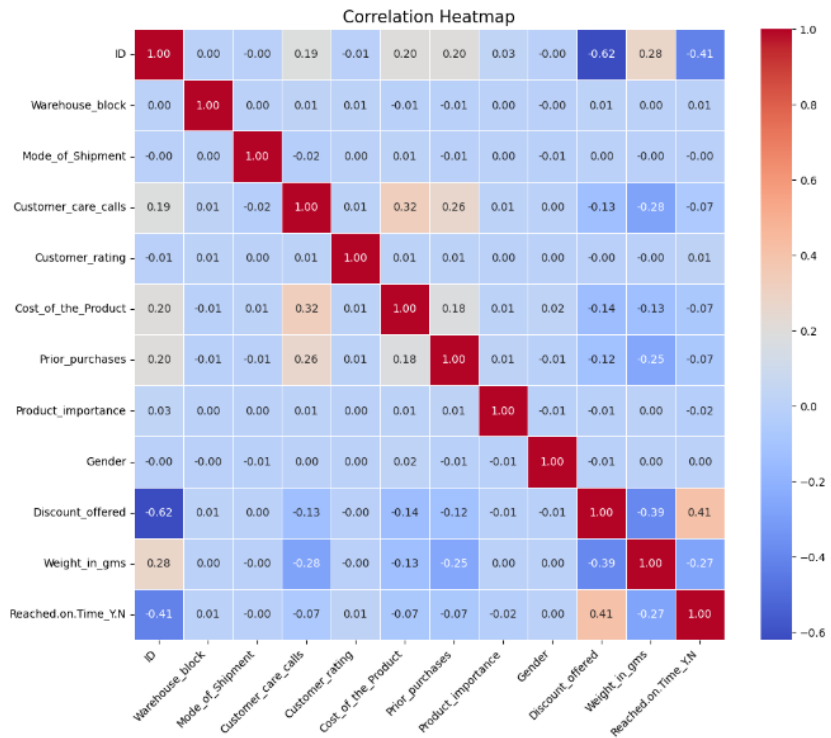
Identifies data sources, assesses quality issues like missing values and duplicates, and implements resolution plans to ensure accurate and reliable analysis.

Section	Description																																																																																																																					
Data Overview	<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><th>Prior_purchases</th><th>Product_importance</th><th>Gender</th><th>Discount_offered</th><th>Weight_in_gm</th><th>Reached_on_Time_Y/N</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><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>5000.00000</td><td>2.333394</td><td>1.516865</td><td>4.054459</td><td>2.986545</td><td>210.196835</td><td>3.421629</td><td>1.348031</td><td>0.495863</td><td>8.590963</td><td>3634.016729</td><td>0.596691</td></tr><tr><td>std</td><td>3175.28214</td><td>1.480726</td><td>0.756594</td><td>1.141490</td><td>1.413603</td><td>48.063272</td><td>1.136903</td><td>0.631434</td><td>0.500006</td><td>6.095461</td><td>1635.377251</td><td>0.490584</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><td>2.000000</td><td>0.000000</td><td>0.000000</td><td>1.000000</td><td>1001.000000</td><td>0.000000</td></tr><tr><td>5%</td><td>2750.00000</td><td>1.000000</td><td>1.000000</td><td>3.000000</td><td>2.000000</td><td>169.000000</td><td>3.000000</td><td>1.000000</td><td>0.000000</td><td>4.000000</td><td>1828.000000</td><td>0.000000</td></tr><tr><td>95%</td><td>5000.00000</td><td>3.000000</td><td>2.000000</td><td>4.000000</td><td>3.000000</td><td>214.000000</td><td>3.000000</td><td>1.000000</td><td>0.000000</td><td>7.000000</td><td>4149.000000</td><td>1.000000</td></tr><tr><td>75%</td><td>8249.50000</td><td>4.000000</td><td>2.000000</td><td>5.000000</td><td>4.000000</td><td>251.000000</td><td>4.000000</td><td>2.000000</td><td>1.000000</td><td>10.000000</td><td>5050.000000</td><td>1.000000</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><td>5.000000</td><td>2.000000</td><td>1.000000</td><td>19.000000</td><td>7846.000000</td><td>1.000000</td></tr></table>		ID	Warehouse_block	Mode_of_Shipment	Customer_care_calls	Customer_rating	Cost_of_the_Product	Prior_purchases	Product_importance	Gender	Discount_offered	Weight_in_gm	Reached_on_Time_Y/N	count	10999.00000	10999.00000	10999.00000	10999.00000	10999.00000	10999.00000	10999.00000	10999.00000	10999.00000	10999.00000	10999.00000	10999.00000	mean	5000.00000	2.333394	1.516865	4.054459	2.986545	210.196835	3.421629	1.348031	0.495863	8.590963	3634.016729	0.596691	std	3175.28214	1.480726	0.756594	1.141490	1.413603	48.063272	1.136903	0.631434	0.500006	6.095461	1635.377251	0.490584	min	1.00000	0.000000	0.000000	2.000000	1.000000	96.000000	2.000000	0.000000	0.000000	1.000000	1001.000000	0.000000	5%	2750.00000	1.000000	1.000000	3.000000	2.000000	169.000000	3.000000	1.000000	0.000000	4.000000	1828.000000	0.000000	95%	5000.00000	3.000000	2.000000	4.000000	3.000000	214.000000	3.000000	1.000000	0.000000	7.000000	4149.000000	1.000000	75%	8249.50000	4.000000	2.000000	5.000000	4.000000	251.000000	4.000000	2.000000	1.000000	10.000000	5050.000000	1.000000	max	10999.00000	4.000000	2.000000	7.000000	5.000000	310.000000	5.000000	2.000000	1.000000	19.000000	7846.000000	1.000000
	ID	Warehouse_block	Mode_of_Shipment	Customer_care_calls	Customer_rating	Cost_of_the_Product	Prior_purchases	Product_importance	Gender	Discount_offered	Weight_in_gm	Reached_on_Time_Y/N																																																																																																										
count	10999.00000	10999.00000	10999.00000	10999.00000	10999.00000	10999.00000	10999.00000	10999.00000	10999.00000	10999.00000	10999.00000	10999.00000																																																																																																										
mean	5000.00000	2.333394	1.516865	4.054459	2.986545	210.196835	3.421629	1.348031	0.495863	8.590963	3634.016729	0.596691																																																																																																										
std	3175.28214	1.480726	0.756594	1.141490	1.413603	48.063272	1.136903	0.631434	0.500006	6.095461	1635.377251	0.490584																																																																																																										
min	1.00000	0.000000	0.000000	2.000000	1.000000	96.000000	2.000000	0.000000	0.000000	1.000000	1001.000000	0.000000																																																																																																										
5%	2750.00000	1.000000	1.000000	3.000000	2.000000	169.000000	3.000000	1.000000	0.000000	4.000000	1828.000000	0.000000																																																																																																										
95%	5000.00000	3.000000	2.000000	4.000000	3.000000	214.000000	3.000000	1.000000	0.000000	7.000000	4149.000000	1.000000																																																																																																										
75%	8249.50000	4.000000	2.000000	5.000000	4.000000	251.000000	4.000000	2.000000	1.000000	10.000000	5050.000000	1.000000																																																																																																										
max	10999.00000	4.000000	2.000000	7.000000	5.000000	310.000000	5.000000	2.000000	1.000000	19.000000	7846.000000	1.000000																																																																																																										
Univariate Analysis	<div><div>Distribution of Warehouse_block</div><div>Distribution of Mode_of_Shipment</div><div>Distribution of Customer_care_calls</div><div>Distribution of Customer_rating</div><div>Distribution of Cost_of_the_Product</div><div>Distribution of Prior_purchases</div></div>																																																																																																																					

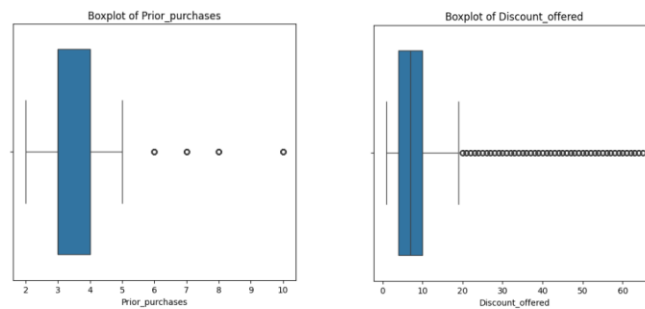
Bivariate Analysis



Multivariate Analysis



Outliers and Anomalies



Data Preprocessing Code Screenshots

Loading Data

```
data = pd.read_csv('content/Train.csv')
data.head()
```

	ID	Warehouse_block	Mode_of_Shipment	Customer_care_calls	Customer_rating	Cost_of_the_Product	Prior_purchases	Product_importance	Gender	Discount_offered	Weight_in_gms	Reached.on.Time_Y.N
0	1	D	Flight	4	2	177	3	low	F	44	1233	1
1	2	F	Flight	4	5	216	2	low	M	59	3588	1
2	3	A	Flight	2	2	183	4	low	M	48	3374	1
3	4	B	Flight	3	3	176	4	medium	M	10	1177	1
4	5	C	Flight	2	2	184	3	medium	F	45	2484	1

Handling Missing Data

```
# Check for missing values
print(data.isnull().sum())
```

```
ID                                0
Warehouse_block                   0
Mode_of_Shipment                  0
Customer_care_calls               0
Customer_rating                   0
Cost_of_the_Product               0
Prior_purchases                   0
Product_importance                0
Gender                            0
Discount_offered                  0
Weight_in_gms                     0
Reached.on.Time_Y.N              0
dtype: int64
```

Data Transformation

```
[ ] categorical_cols = data.select_dtypes(include=['object']).columns
print("Categorical columns:", categorical_cols)

Categorical columns: Index(['Warehouse_block', 'Mode_of_Shipment', 'Product_importance', 'Gender'], dtype='object')

[ ] label_encoder = LabelEncoder()

# Apply LabelEncoder to each categorical column
for col in categorical_cols:
    data[col] = label_encoder.fit_transform(data[col])
```

Feature Engineering

Attached the codes in final submission.

Save Processed Data

-