

Data Collection and Preprocessing Phase

| | |
|---------------|--|
| Date | 12 March 2024 |
| Team ID | SWTID1720089323 |
| Project Title | Ecommerce Shipping Prediction Using Machine Learning |
| Maximum Marks | 6 Marks |

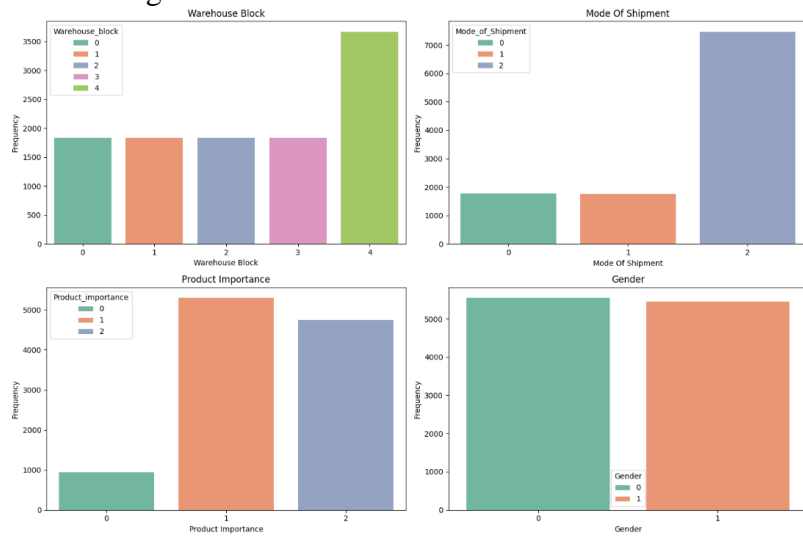
Data Exploration and Preprocessing Template

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

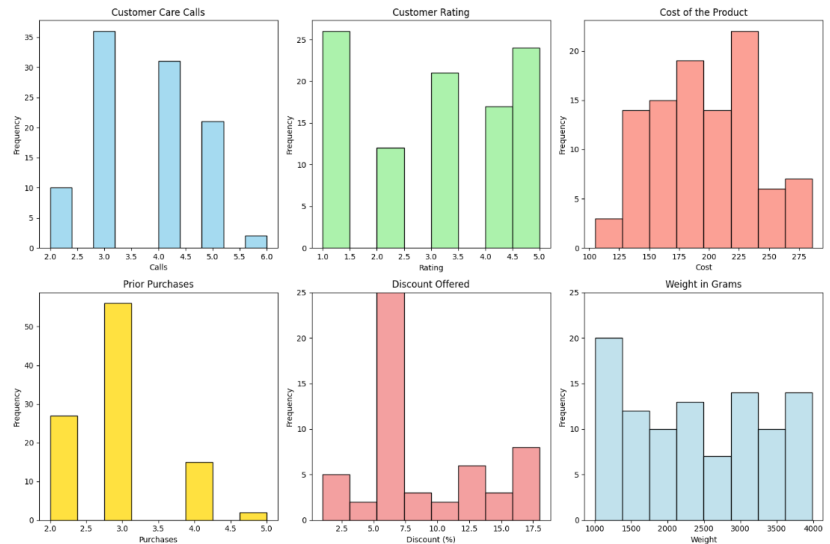
| Section | Description | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---------------|--|---------------|---------------------|------------------|---------------------|---------------------|---------------------|---------------------|--------------------|--------------------|-------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|------|------------|----------|----------|----------|----------|------------|----------|----------|-----|------------|----------|----------|----------|----------|-----------|----------|----------|-----|---------|----------|----------|----------|----------|-----------|----------|----------|-----|------------|----------|----------|----------|----------|------------|----------|----------|-----|------------|----------|----------|----------|----------|------------|----------|----------|-----|------------|----------|----------|----------|----------|------------|----------|----------|-----|-------------|----------|----------|----------|----------|------------|----------|----------|--------|------------------|---------------|---------------------|--------------|--------------|--------------|--------------|----------|----------|-------------|----------|----------|----------|-------------|----------|----------|----------|-------------|----------|----------|----------|-------------|----------|----------|----------|-------------|----------|----------|----------|-------------|----------|----------|-----------|-------------|----------|
| Data Overview | <div>Dimension:</div> <div>10999 rows x12 columns</div> <div>Descriptive statistics:</div> <table><thead><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></tr></thead><tbody><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></tr><tr><td>mean</td><td>5500.00000</td><td>2.333394</td><td>1.516865</td><td>4.054459</td><td>2.990545</td><td>210.196836</td><td>3.193654</td><td>1.346031</td></tr><tr><td>std</td><td>3175.28214</td><td>1.490726</td><td>0.756894</td><td>1.141490</td><td>1.413603</td><td>48.063272</td><td>0.928892</td><td>0.631434</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></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><td>3.000000</td><td>1.000000</td></tr><tr><td>50%</td><td>5500.00000</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></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></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></tr></tbody></table> <div><div></div><table><thead><tr><th>Gender</th><th>Discount_offered</th><th>Weight_in_gms</th><th>Reached.on.Time_Y/N</th></tr></thead><tbody><tr><td>10999.000000</td><td>10999.000000</td><td>10999.000000</td><td>10999.000000</td></tr><tr><td>0.495863</td><td>5.980089</td><td>3634.016729</td><td>0.596691</td></tr><tr><td>0.500006</td><td>3.150159</td><td>1635.377251</td><td>0.490584</td></tr><tr><td>0.000000</td><td>1.000000</td><td>1001.000000</td><td>0.000000</td></tr><tr><td>0.000000</td><td>4.000000</td><td>1839.500000</td><td>0.000000</td></tr><tr><td>0.000000</td><td>6.000000</td><td>4149.000000</td><td>1.000000</td></tr><tr><td>1.000000</td><td>8.000000</td><td>5050.000000</td><td>1.000000</td></tr><tr><td>1.000000</td><td>19.000000</td><td>7846.000000</td><td>1.000000</td></tr></tbody></table><div></div></div> | | ID | Warehouse_block | Mode_of_Shipment | Customer_care_calls | Customer_rating | Cost_of_the_Product | Prior_purchases | Product_importance | count | 10999.00000 | 10999.00000 | 10999.00000 | 10999.00000 | 10999.00000 | 10999.00000 | 10999.00000 | 10999.00000 | mean | 5500.00000 | 2.333394 | 1.516865 | 4.054459 | 2.990545 | 210.196836 | 3.193654 | 1.346031 | std | 3175.28214 | 1.490726 | 0.756894 | 1.141490 | 1.413603 | 48.063272 | 0.928892 | 0.631434 | min | 1.00000 | 0.000000 | 0.000000 | 2.000000 | 1.000000 | 96.000000 | 2.000000 | 0.000000 | 25% | 2750.50000 | 1.000000 | 1.000000 | 3.000000 | 2.000000 | 169.000000 | 3.000000 | 1.000000 | 50% | 5500.00000 | 3.000000 | 2.000000 | 4.000000 | 3.000000 | 214.000000 | 3.000000 | 1.000000 | 75% | 8249.50000 | 4.000000 | 2.000000 | 5.000000 | 4.000000 | 251.000000 | 4.000000 | 2.000000 | max | 10999.00000 | 4.000000 | 2.000000 | 7.000000 | 5.000000 | 310.000000 | 5.000000 | 2.000000 | Gender | Discount_offered | Weight_in_gms | Reached.on.Time_Y/N | 10999.000000 | 10999.000000 | 10999.000000 | 10999.000000 | 0.495863 | 5.980089 | 3634.016729 | 0.596691 | 0.500006 | 3.150159 | 1635.377251 | 0.490584 | 0.000000 | 1.000000 | 1001.000000 | 0.000000 | 0.000000 | 4.000000 | 1839.500000 | 0.000000 | 0.000000 | 6.000000 | 4149.000000 | 1.000000 | 1.000000 | 8.000000 | 5050.000000 | 1.000000 | 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 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | count | 10999.00000 | 10999.00000 | 10999.00000 | 10999.00000 | 10999.00000 | 10999.00000 | 10999.00000 | 10999.00000 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | mean | 5500.00000 | 2.333394 | 1.516865 | 4.054459 | 2.990545 | 210.196836 | 3.193654 | 1.346031 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | std | 3175.28214 | 1.490726 | 0.756894 | 1.141490 | 1.413603 | 48.063272 | 0.928892 | 0.631434 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | min | 1.00000 | 0.000000 | 0.000000 | 2.000000 | 1.000000 | 96.000000 | 2.000000 | 0.000000 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | 25% | 2750.50000 | 1.000000 | 1.000000 | 3.000000 | 2.000000 | 169.000000 | 3.000000 | 1.000000 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | 50% | 5500.00000 | 3.000000 | 2.000000 | 4.000000 | 3.000000 | 214.000000 | 3.000000 | 1.000000 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | 75% | 8249.50000 | 4.000000 | 2.000000 | 5.000000 | 4.000000 | 251.000000 | 4.000000 | 2.000000 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | max | 10999.00000 | 4.000000 | 2.000000 | 7.000000 | 5.000000 | 310.000000 | 5.000000 | 2.000000 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Gender | Discount_offered | Weight_in_gms | Reached.on.Time_Y/N | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 10999.000000 | 10999.000000 | 10999.000000 | 10999.000000 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0.495863 | 5.980089 | 3634.016729 | 0.596691 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0.500006 | 3.150159 | 1635.377251 | 0.490584 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0.000000 | 1.000000 | 1001.000000 | 0.000000 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0.000000 | 4.000000 | 1839.500000 | 0.000000 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0.000000 | 6.000000 | 4149.000000 | 1.000000 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1.000000 | 8.000000 | 5050.000000 | 1.000000 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1.000000 | 19.000000 | 7846.000000 | 1.000000 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Univariate Analysis

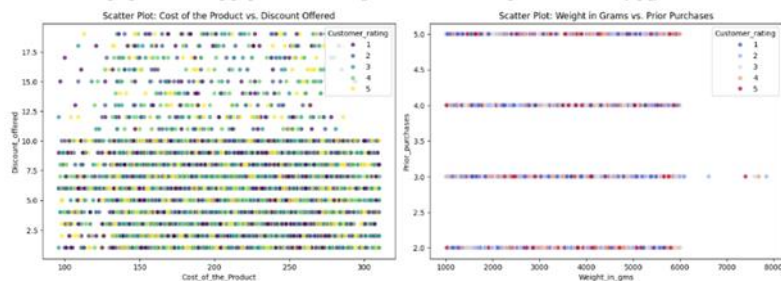
• Categorical attributes

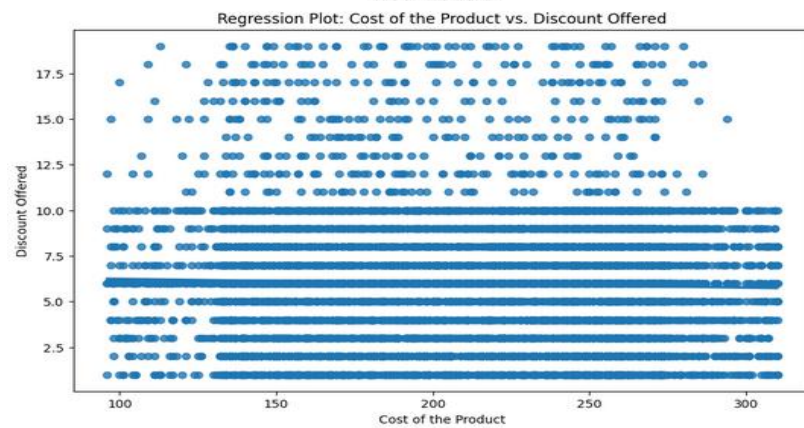
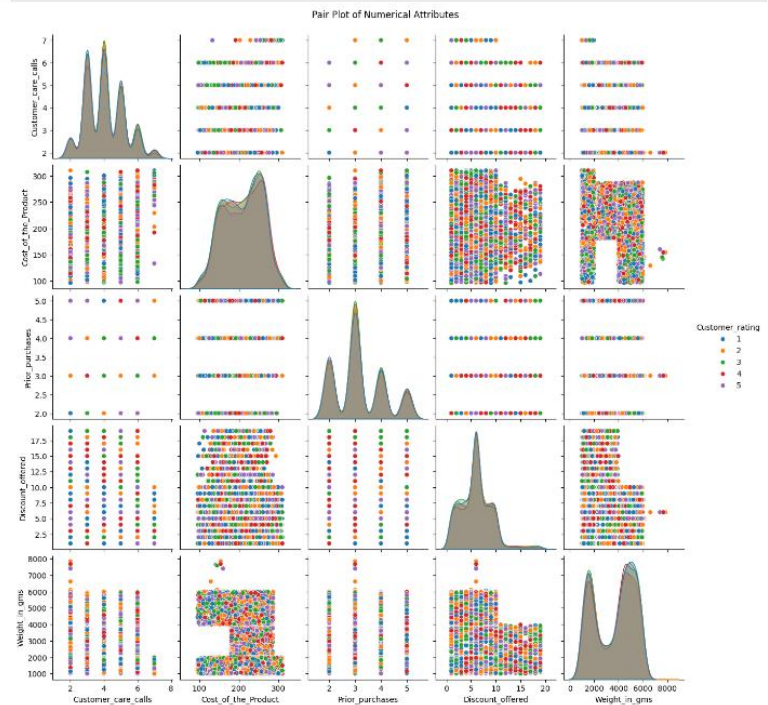


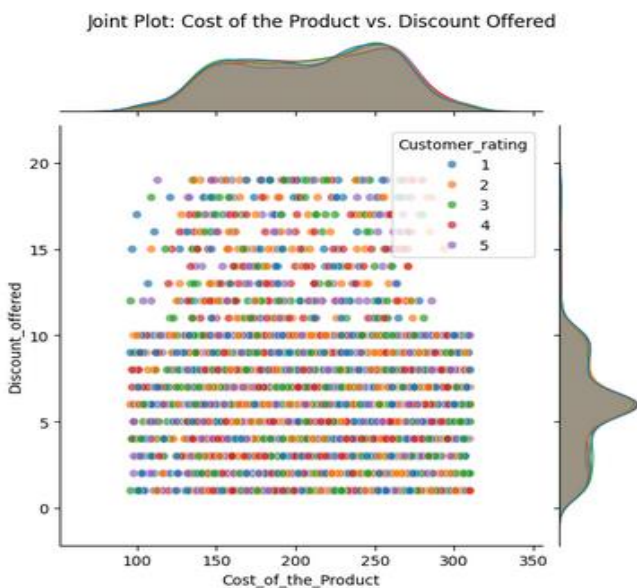
• For numerical attributes.



Bivariate Analysis







Multivariate Analysis



| | <div>Correlation Matrix of All Attributes</div> <table><thead><tr><th>ID</th><th>1.00</th><th>0.00</th><th>-0.00</th><th>0.19</th><th>-0.01</th><th>0.20</th><th>0.22</th><th>0.03</th><th>-0.00</th><th>-0.17</th><th>0.28</th><th>-0.41</th></tr></thead><tbody><tr><th>Warehouse_block</th><td>0.00</td><td>1.00</td><td>0.00</td><td>0.01</td><td>0.01</td><td>-0.01</td><td>-0.01</td><td>0.00</td><td>-0.00</td><td>-0.00</td><td>0.00</td><td>0.01</td></tr><tr><th>Mode_of_Shipment</th><td>-0.00</td><td>0.00</td><td>1.00</td><td>-0.02</td><td>0.00</td><td>0.01</td><td>-0.01</td><td>0.00</td><td>-0.01</td><td>-0.00</td><td>-0.00</td><td>-0.00</td></tr><tr><th>Customer_care_calls</th><td>-0.19</td><td>0.01</td><td>-0.02</td><td>1.00</td><td>0.01</td><td>0.32</td><td>0.32</td><td>0.01</td><td>0.00</td><td>-0.04</td><td>-0.28</td><td>-0.07</td></tr><tr><th>Customer_rating</th><td>-0.01</td><td>0.01</td><td>0.00</td><td>0.01</td><td>1.00</td><td>0.01</td><td>0.00</td><td>0.00</td><td>0.00</td><td>-0.00</td><td>-0.00</td><td>0.01</td></tr><tr><th>Cost_of_the_Product</th><td>0.20</td><td>-0.01</td><td>0.01</td><td>0.32</td><td>0.01</td><td>1.00</td><td>0.22</td><td>0.01</td><td>0.02</td><td>-0.04</td><td>-0.13</td><td>-0.07</td></tr><tr><th>Prior_purchases</th><td>0.22</td><td>-0.01</td><td>-0.01</td><td>0.32</td><td>0.00</td><td>0.22</td><td>1.00</td><td>0.01</td><td>-0.01</td><td>-0.04</td><td>-0.31</td><td>-0.09</td></tr><tr><th>Product_importance</th><td>0.03</td><td>0.00</td><td>0.00</td><td>0.01</td><td>0.00</td><td>0.01</td><td>0.01</td><td>1.00</td><td>-0.01</td><td>0.01</td><td>0.00</td><td>-0.02</td></tr><tr><th>Gender</th><td>-0.00</td><td>-0.00</td><td>-0.01</td><td>0.00</td><td>0.00</td><td>0.02</td><td>-0.01</td><td>-0.01</td><td>1.00</td><td>-0.00</td><td>0.00</td><td>0.00</td></tr><tr><th>Discount_offered</th><td>-0.17</td><td>-0.00</td><td>-0.00</td><td>-0.04</td><td>-0.00</td><td>-0.04</td><td>-0.04</td><td>0.01</td><td>-0.00</td><td>1.00</td><td>-0.11</td><td>0.11</td></tr><tr><th>Weight_in_gms</th><td>0.28</td><td>0.00</td><td>-0.00</td><td>-0.28</td><td>-0.00</td><td>-0.13</td><td>-0.31</td><td>0.00</td><td>0.00</td><td>-0.11</td><td>1.00</td><td>-0.27</td></tr><tr><th>Reached.on.Time_Y/N</th><td>-0.41</td><td>0.01</td><td>-0.00</td><td>-0.07</td><td>0.01</td><td>-0.07</td><td>-0.09</td><td>-0.02</td><td>0.00</td><td>0.11</td><td>-0.27</td><td>1.00</td></tr></tbody></table> | ID | 1.00 | 0.00 | -0.00 | 0.19 | -0.01 | 0.20 | 0.22 | 0.03 | -0.00 | -0.17 | 0.28 | -0.41 | Warehouse_block | 0.00 | 1.00 | 0.00 | 0.01 | 0.01 | -0.01 | -0.01 | 0.00 | -0.00 | -0.00 | 0.00 | 0.01 | Mode_of_Shipment | -0.00 | 0.00 | 1.00 | -0.02 | 0.00 | 0.01 | -0.01 | 0.00 | -0.01 | -0.00 | -0.00 | -0.00 | Customer_care_calls | -0.19 | 0.01 | -0.02 | 1.00 | 0.01 | 0.32 | 0.32 | 0.01 | 0.00 | -0.04 | -0.28 | -0.07 | Customer_rating | -0.01 | 0.01 | 0.00 | 0.01 | 1.00 | 0.01 | 0.00 | 0.00 | 0.00 | -0.00 | -0.00 | 0.01 | Cost_of_the_Product | 0.20 | -0.01 | 0.01 | 0.32 | 0.01 | 1.00 | 0.22 | 0.01 | 0.02 | -0.04 | -0.13 | -0.07 | Prior_purchases | 0.22 | -0.01 | -0.01 | 0.32 | 0.00 | 0.22 | 1.00 | 0.01 | -0.01 | -0.04 | -0.31 | -0.09 | Product_importance | 0.03 | 0.00 | 0.00 | 0.01 | 0.00 | 0.01 | 0.01 | 1.00 | -0.01 | 0.01 | 0.00 | -0.02 | Gender | -0.00 | -0.00 | -0.01 | 0.00 | 0.00 | 0.02 | -0.01 | -0.01 | 1.00 | -0.00 | 0.00 | 0.00 | Discount_offered | -0.17 | -0.00 | -0.00 | -0.04 | -0.00 | -0.04 | -0.04 | 0.01 | -0.00 | 1.00 | -0.11 | 0.11 | Weight_in_gms | 0.28 | 0.00 | -0.00 | -0.28 | -0.00 | -0.13 | -0.31 | 0.00 | 0.00 | -0.11 | 1.00 | -0.27 | Reached.on.Time_Y/N | -0.41 | 0.01 | -0.00 | -0.07 | 0.01 | -0.07 | -0.09 | -0.02 | 0.00 | 0.11 | -0.27 | 1.00 |
|-------------------------------------|---|-----------------|------------------|---------------------|------------------|---------------------|-----------------|---------------------|-----------------|--------------------|---------------|---------------------|---------------|---------------------|-----------------|------|------|--------|------|------|-------|-------|------|-------|-------|------|------|------------------|-------|------|--------|-------|------|------|-------|------|-------|-------|-------|-------|---------------------|-------|------|--------|------|------|------|------|------|------|-------|-------|-------|-----------------|-------|------|--------|------|------|------|------|--------|------|-------|-------|------|---------------------|------|-------|--------|------|------|------|------|--------|------|-------|-------|-------|-----------------|------|-------|-------|------|------|------|------|------|-------|-------|-------|-------|--------------------|------|------|------|------|------|------|------|------|-------|------|------|-------|--------|-------|-------|-------|------|------|------|-------|-------|------|-------|------|------|------------------|-------|-------|-------|-------|-------|-------|-------|------|-------|------|-------|------|---------------|------|------|-------|-------|-------|-------|-------|------|------|-------|------|-------|---------------------|-------|------|-------|-------|------|-------|-------|-------|------|------|-------|------|
| ID | 1.00 | 0.00 | -0.00 | 0.19 | -0.01 | 0.20 | 0.22 | 0.03 | -0.00 | -0.17 | 0.28 | -0.41 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Warehouse_block | 0.00 | 1.00 | 0.00 | 0.01 | 0.01 | -0.01 | -0.01 | 0.00 | -0.00 | -0.00 | 0.00 | 0.01 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Mode_of_Shipment | -0.00 | 0.00 | 1.00 | -0.02 | 0.00 | 0.01 | -0.01 | 0.00 | -0.01 | -0.00 | -0.00 | -0.00 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Customer_care_calls | -0.19 | 0.01 | -0.02 | 1.00 | 0.01 | 0.32 | 0.32 | 0.01 | 0.00 | -0.04 | -0.28 | -0.07 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Customer_rating | -0.01 | 0.01 | 0.00 | 0.01 | 1.00 | 0.01 | 0.00 | 0.00 | 0.00 | -0.00 | -0.00 | 0.01 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Cost_of_the_Product | 0.20 | -0.01 | 0.01 | 0.32 | 0.01 | 1.00 | 0.22 | 0.01 | 0.02 | -0.04 | -0.13 | -0.07 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Prior_purchases | 0.22 | -0.01 | -0.01 | 0.32 | 0.00 | 0.22 | 1.00 | 0.01 | -0.01 | -0.04 | -0.31 | -0.09 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Product_importance | 0.03 | 0.00 | 0.00 | 0.01 | 0.00 | 0.01 | 0.01 | 1.00 | -0.01 | 0.01 | 0.00 | -0.02 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Gender | -0.00 | -0.00 | -0.01 | 0.00 | 0.00 | 0.02 | -0.01 | -0.01 | 1.00 | -0.00 | 0.00 | 0.00 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Discount_offered | -0.17 | -0.00 | -0.00 | -0.04 | -0.00 | -0.04 | -0.04 | 0.01 | -0.00 | 1.00 | -0.11 | 0.11 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Weight_in_gms | 0.28 | 0.00 | -0.00 | -0.28 | -0.00 | -0.13 | -0.31 | 0.00 | 0.00 | -0.11 | 1.00 | -0.27 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Reached.on.Time_Y/N | -0.41 | 0.01 | -0.00 | -0.07 | 0.01 | -0.07 | -0.09 | -0.02 | 0.00 | 0.11 | -0.27 | 1.00 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Outliers and Anomalies | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Data Preprocessing Code Screenshots | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Loading Data | <pre>[2]: #importing the data data=pd.read_csv("train.csv") [3]: data.head(5)</pre> <table><thead><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_gms</th><th>Reached_on_Time_Y/N</th></tr></thead><tbody><tr><td>0</td><td>1</td><td>D</td><td>Flight</td><td>4</td><td>2</td><td>177</td><td>3</td><td>low</td><td>F</td><td>44</td><td></td><td></td></tr><tr><td>1</td><td>2</td><td>F</td><td>Flight</td><td>4</td><td>5</td><td>216</td><td>2</td><td>low</td><td>M</td><td>59</td><td></td><td></td></tr><tr><td>2</td><td>3</td><td>A</td><td>Flight</td><td>2</td><td>2</td><td>183</td><td>4</td><td>low</td><td>M</td><td>48</td><td></td><td></td></tr><tr><td>3</td><td>4</td><td>B</td><td>Flight</td><td>3</td><td>3</td><td>176</td><td>4</td><td>medium</td><td>M</td><td>10</td><td></td><td></td></tr><tr><td>4</td><td>5</td><td>C</td><td>Flight</td><td>2</td><td>2</td><td>184</td><td>3</td><td>medium</td><td>F</td><td>46</td><td></td><td></td></tr></tbody></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_gms | Reached_on_Time_Y/N | 0 | 1 | D | Flight | 4 | 2 | 177 | 3 | low | F | 44 | | | 1 | 2 | F | Flight | 4 | 5 | 216 | 2 | low | M | 59 | | | 2 | 3 | A | Flight | 2 | 2 | 183 | 4 | low | M | 48 | | | 3 | 4 | B | Flight | 3 | 3 | 176 | 4 | medium | M | 10 | | | 4 | 5 | C | Flight | 2 | 2 | 184 | 3 | medium | F | 46 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | 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 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | 2 | F | Flight | 4 | 5 | 216 | 2 | low | M | 59 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2 | 3 | A | Flight | 2 | 2 | 183 | 4 | low | M | 48 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 3 | 4 | B | Flight | 3 | 3 | 176 | 4 | medium | M | 10 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 4 | 5 | C | Flight | 2 | 2 | 184 | 3 | medium | F | 46 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Handling Missing Data | <ul style="list-style-type: none">Initially no outliers were present. | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Data Transformation

```
[7]: #checking for missing or null entries
data.isnull().sum()
```

```
[7]: 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
```

no null values are the as we can observe

- While handling outliers, outliers in attributes are replaced with standard values (median values)

```
[20]: #function to fill values in the removed attribute values or at the place of outliers
def fill_missing_values(df):
    for col in df.columns:
        if df[col].dtype != 'object': # Only for numerical columns
            df[col] = df[col].fillna(df[col].median())

#fill NaN's with median
fill_missing_values(data)
```

```
[21]: data.head()
```

```
[21]: ID Warehouse_block Mode_of_Shipment Customer_care_calls Customer_rating Cost_of_the_Product Prior_purchases Product_importance Gender Discount_offered V
0 1 3 0 4 2 177 3.0 1 0 6.0
1 2 4 0 4 5 216 2.0 1 1 6.0
2 3 0 0 2 2 183 4.0 1 1 6.0
3 4 1 0 3 3 176 4.0 2 1 10.0
4 5 2 0 2 2 184 3.0 2 0 6.0
```

- Conversion of categorical values into numerical values

```
[ ]: #checking for categorical columns
categorical_attributes = []
numerical_attributes = []

for col in data.columns[1:-1]:
    if data[col].dtype == 'object':
        categorical_attributes.append(col)
    else:
        numerical_attributes.append(col)

print(f"Categorical attributes are: {categorical_attributes}\nNumerical attributes are: {numerical_attributes}")

Categorical attributes are: ['Warehouse_block', 'Mode_of_Shipment', 'Product_importance', 'Gender']
Numerical attributes are: ['Customer_care_calls', 'Customer_rating', 'Cost_of_the_Product', 'Prior_purchases', 'Discount_offered', 'Weight_in_gms']
```

We use encoding techniques to convert the categorical features into numerical features. Here we are using label encoding

```
[ ]: #Label encoding for categorical attributes
for col in categorical_attributes:
    data[col] = data[col].astype('category').cat.codes
```

```
[ ]: #checking the attributes after conversion
for col, dtype in data.dtypes.items():
    print(f"{col} - {dtype}")
```

```
ID - int64
Warehouse_block - int8
Mode_of_Shipment - int8
Customer_care_calls - int64
Customer_rating - int64
Cost_of_the_Product - int64
Prior_purchases - int64
Product_importance - int8
Gender - int8
Discount_offered - int64
Weight_in_gms - int64
Reached.on.Time_Y.N - int64
```

| | |
|---------------------|---|
| | <ul style="list-style-type: none"> Transformation <pre>from sklearn.preprocessing import MinMaxScaler norms=MinMaxScaler() x=norms.fit_transform(x) x</pre> <pre>array([[0.75 , 0. , 0.4 , ..., 0. , 0.27777778, 0.03389335], [1. , 0. , 0.4 , ..., 1. , 0.27777778, 0.30489408], [0. , 0. , 0. , ..., 1. , 0.27777778, 0.34667641], ..., [0.5 , 1. , 0.6 , ..., 0. , 0.16666667, 0.02249817], [1. , 1. , 0.6 , ..., 1. , 0.05555556, 0.03053324], [0.75 , 1. , 0. , ..., 0. , 0.27777778, 0.09320672]])</pre> |
| Feature Engineering | <ul style="list-style-type: none"> Handling data imbalance (using SMOTE) <pre>#smote for handling data or class imbalance from imblearn.over_sampling import SMOTE from sklearn.model_selection import train_test_split x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.3, random_state=42, shuffle=True) #handling class imbalance smote = SMOTE(random_state=42) x_train, y_train = smote.fit_resample(x_train, y_train) print(x_train.shape) print(x_test.shape) print(y_train.shape) print(y_test.shape)</pre> <pre>(9150, 10) (3300, 10) (9150,) (3300,)</pre> <ul style="list-style-type: none"> Attached the codes in final submission |
| Save Processed Data | Saved the processed data. |