**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 04 JUNE 2024 |
| Team ID | SWTID1720260935 |
| Project Title | **Ecommerce Shipping Prediction Using Machine Learning** |
| Maximum Marks | 6 Marks |

**Data Exploration and Preprocessing Template:**

|  |  |
| --- | --- |
| **Section** | **Description** |
| **Data Overview** | * **Internal:** * Order ID, product specifications, client information, shipment method, and delivery time are all hiical ostorrder data. * Product catalog data (product weight, dimensions)   • **External** (potential):  • Real-time carrier data (shipping rates, transit times)  • Weather data (location-based, impacting delivery times)  • Holiday calendars (potential delays) |
| **Univariate Analysis** | **Delivery Time (target variable):**   * Mean: 9-10 days * Median: 6-7 days (deliveries tend to be faster than the average) * Minimum: 4 days * Maximum: 10 days (shows a range of delivery times) |
| **Bivariate Analysis** | We expect a positive correlation, meaning locations further away (higher distance) will tend to have longer delivery times. This helps identify factors influencing delivery times. |
| **Multivariate Analysis** | The traditional way of shipping heavy products long distances may take longer. |
| Outliers and Anomalies | Expedited shipping |
| **Data Preprocessing Code Screenshots** | |
| Loading Data |  |
| Handling Missing Data |  |
| Data Transformation |  |
| Feature Engineering |  |
| Save Processed Data |  |