Data Collection and Preprocessing Phase

Date	16 July 2024
Team ID	SWTID1720190389
Project Title	E-Commerce Shipping Prediction
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification Template

Elevate your data strategy with the Data Collection plan and the Raw Data Sources report, ensuring meticulous data curation and integrity for informed decision-making in every analysis and decision-making endeavor.

Data Collection Plan Template

Section	Description
Project Overview	Ecommerce shipping prediction is the process of estimating the whether the product reached on time. which is based on various factors such as the origin and destination of the package, the shipping method selected by the customer, the carrier used for shipping, and any potential delays or issues that may arise during the shipping process. Machine learning models can be used to make accurate predictions about shipping times based on historical data and real-time updates from carriers. These models may take into account factors such as weather conditions, traffic, and other external factors that can impact delivery times. Over All Ecommerce shipping prediction is an important tool for ecommerce businesses that want to provide accurate delivery estimates to their customers and improve their overall customer experience.
Data Collection Plan	Search for datasets related to E-commerce shipping in kaggle

Raw Data Sources Identified	The raw data sources for this project include datasets obtained from Kaggle, a popular platform for data science competitions and repositories. The provided sample data represents a subset of the collected information, encompassing variables such as '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'

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle	The provided sample data represents a subset of the collected information, encompassing variables such as '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'	https://www.kaggle.com/datasets/prachi13/customer-analytics?select=Train.csv	CSV	431 KB	Public