**Data Collection and Preprocessing Phase**

| Date | June 2024 |
| --- | --- |
| Team ID | 740291 |
| Project Title | Ecommerce Shipping Prediction using Machine Learning |
| 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 | Aims to provide accurate delivery estimates by considering factors like distance, traffic, and weather, offering real-time updates, integrating with ecommerce platforms. |
| Data Collection Plan | Data will be collected from order databases, traffic reports, weather forecasts, GPS data, and ecommerce platforms |
| Raw Data Sources Identified | Gathering data from order histories, delivery logs, traffic reports, weather forecasts, and ecommerce platform integrations. |

**Raw Data Sources Template**

| **Source Name** | **Description** | **Location/URL** | **Format** | **Access Permissions** |
| --- | --- | --- | --- | --- |
| Dataset 1 | In this project we have used .csv data. This data is downloaded from kaggle.com. | <https://www.kaggle.com/datasets/prachi13/customer-analytics?select=Train.csv> | CSV | Public |
| Dataset 2 | Access Google Drive with a Google account (for personal use) or Google Workspace account | <https://drive.google.com/drive/folders/1y5bcRIudq2kdWchMyU9g_BfVIWtWXu2S> | CSV | Private (with access) |