**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 15 March 2024 |
| Team ID | 739850 |
| Project Title | Air Quality Index Analyzer 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 | This project aims to predict the success of Air Quality Index  Analyzer first stage landings. The objectives include collecting data on previous launches, analyzing factors influencing landing success, and building a predictive model. |
| Data Collection Plan | Data will be collected from Kaggle datasets, and public records of launch events. |
| Raw Data Sources  Identified | The raw data sources include AQI for real-time data, Kaggle for historical launch data, and publicly available datasets on launch statistics. |

**Raw Data Sources Template**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source**  **Name** | **Description** | **Location/URL** | **Format** | **Size** | **Access Permissions** |
| Kaggle  Dataset | Historical data on  Air Quality Data in  India Air Quality Index (AQI) and hourly data across stations and cities in  India | [https://www.kagg](https://www.kagg/) le.com/datasets/ro hanrao/airquality-data-inindia | CSV | 76MB | Public |