**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 15 March 2024 |
| Team ID | xxxxxx |
| Project Title | Forecasting Economic Prosperity: Leveraging Machine Learning For GDP Per Capita Prediction |
| Maximum Marks | 2 Marks |

**Data Collection Plan & Raw Data Sources Identification Template**

This plan ensures a comprehensive approach to data collection, leveraging reputable and diverse sources to build a robust dataset for the machine learning model, thereby enhancing the reliability and accuracy of GDP per capita predictions.

**Data Collection Plan Template**

|  |  |
| --- | --- |
| **Section** | **Description** |
| Project Overview | This machine learning project aims to forecast GDP per capita, leveraging historical GDP data and a variety of economic, social, and demographic factors. The primary objective is to develop a reliable predictive model that provides valuable insights for policymakers, investors, and businesses. |
| Data Collection Plan | Data will be collected from kaggle. |
| Raw Data Sources Identified | List the raw data sources with relevant details (as a short description). |

**Raw Data Sources Template**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source Name** | **Description** | **Location/URL** | **Format** | **Size** | **Access Permissions** |
| World Bank GDP Data | Contains historical GDP data for various countries. | https://data.worldbank.org/indicator/NY.GDP.MKTP.CD | CSV | XX GB | Public |
| Dataset 2 | Description of the data in this source. | Link of Dataset 2 | Excel | YY GB | Private (with access) |
| … | … | … | … | … | … |