

## Data Collection and Preprocessing Phase

Date	10 July 2024
Team ID	739709
Project Title	To Predict Consumer Price Index
Maximum Marks	2 Marks

### Data Collection Plan & Raw Data Sources Identification Report:

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 endeavour.

### Data Collection Plan:

Section	Description
Project Overview	Predicting the Consumer Price Index (CPI) involves employing data-driven techniques to forecast future CPI values based on historical trends and economic indicators. The CPI serves as a vital measure of inflation, capturing changes in the prices of a basket of goods and services essential to households. This predictive process begins with collecting comprehensive historical CPI data from reliable sources, such as government agencies. Subsequently, data preprocessing is conducted to clean and prepare the data for analysis, including handling missing values and outliers. Machine learning models are then selected and trained, ranging from traditional statistical methods to advanced algorithms like neural networks or time series models. Evaluation metrics, such as Mean Absolute Error or Root Mean Squared Error, assess the model's accuracy on validation data.

Data Collection Plan	To develop a comprehensive data collection plan for predicting the Consumer Price Index (CPI), several key steps must be outlined. First, identify primary sources such as government agencies (e.g., Bureau of Labor Statistics in the US) that regularly publish CPI data. Secondary sources like economic research institutions or central banks can provide supplementary datasets on related economic indicators. Specify data requirements, including the types of CPI (e.g., overall CPI, core CPI excluding volatile items) and the frequency of updates needed (e.g., monthly, quarterly).



Raw Data Sources Identified	The raw data sources for this project include datasets are Kaggle does not typically host a specific dataset for predicting consumer price Index . Historical CPI data Economic Indicators. Consumer Price and spending data and the Commidity Prices
-----------------------------	---

--	--

### Raw Data Sources Report:

Source Name	Description	Location/URL	Format	Size	Access Permissions
-------------	-------------	--------------	--------	------	--------------------

Kaggle Dataset	The dataset comprises patient details (rural & urban)	<a href="https://www.kaggle.com/datasets/fiq423ubf/all-india-consumer-price-index-ruralurban/data">https://www.kaggle.com/datasets/fiq423ubf/all-india-consumer-price-index-ruralurban/data</a>	CSV	20 kB	Public
----------------	---	---	-----	-------	--------