Data Collection and Preprocessing Phase

Date	6 July 2024
Team ID	team-739757
Project Title	Medical Cost 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

Section	**Description**
Project Overview	The project aims to develop a machine learning model to predict medical costs based on historical data.
Data Collection Plan	Data will be collected from various healthcare databases and public datasets, including patient records and insurance claims.
Raw Data Sources	The raw data sources are given in the below template

Raw Data Sources Report :

Source Name	Description	Location / URL	Format	Size	Access Permissions
Medical insurance	Historical medical cost data from a public healthcare database.	https://www.kaggle.com/insurance	CSV	500 MB	Public

Patient Surveys	Patient-reported data on health status, treatments, and costs	http://surveys-patient.com	Unstructured data (e.g., JSON)	200 MB	Surveys data team only, requires patient consent and anonymization.