

Data Collection and Preprocessing Phase

Date	15 July 2024
Team ID	740054
Project Title	Doctors Annual Salary Prediction
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 endeavor.

Data Collection Plan:

Section	Description
Project Overview	The objective of this project is to predict the annual salaries of doctors based on various factors such as experience, specialization, location, and type of employer. Accurate salary predictions can help in understanding the financial landscape of the medical profession, assist hospitals and clinics in budgeting, and provide valuable insights for doctors making career decisions.

Data Collection Plan	<ul style="list-style-type: none"> • Surveys and interviews with doctors. • Existing datasets from medical associations, job portals, and salary websites.
Raw Data Sources Identified	The raw data sources for this project include datasets obtained from Kaggle & UCI, the popular platforms for data science competitions and repositories. The provided sample data represents a subset of the collected information, encompassing variables such as speciality, overall satisfaction, satisfied income for machine learning analysis.

Raw Data Sources Report:

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle Dataset	The dataset comprises doctors details (specialist, overall satisfaction, satisfied Income).	https://data.world/aik/u-sdoctors-pay-2016/workspace/file?filename=USDoctors%27-Pay-2016.xlsx	xlsx	17.04 KB	Public

