

Data Collection and Preprocessing Phase

Date	15 June 2025
Team ID	SWTID1749896042
Project Title	Unemployed insurance beneficiary forecasting
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 machine learning project aims to forecast the number of unemployment insurance beneficiaries based on historical data. Using datasets containing features such as year, state, number of beneficiaries, benefit amounts, and economic indicators, the objective is to build a model that accurately predicts future beneficiary counts, thereby assisting government agencies in planning and resource allocation.
Data Collection Plan	<ul style="list-style-type: none"> ● Search for datasets related to unemployment insurance beneficiaries and benefit amounts. ● Prioritize datasets with time-series data spanning multiple years for robust forecasting. ● Collect data from official government portals and open data repositories to ensure data authenticity and reliability

Raw Data Sources Identified	The raw data source for this project is obtained from data.world, providing historical unemployment insurance statistics, including beneficiary counts and total benefit amounts paid in New York State counties from 2001 onwards, essential for time series forecasting analysis.
-----------------------------	---

Raw Data Sources Report:

Source Name	Description	Location/URL	Format	Size	Access Permissions
Data World	This dataset contains historical records of unemployment insurance beneficiaries in New York State. It includes variables such as year, county name, number of beneficiaries, and benefit amounts paid, making it suitable for trend analysis and forecasting future beneficiary counts.	https://data.world/data-ny-gov/xbjp-8sra	CSV	~3 MB	Public