

## Data Collection and Preprocessing Phase

Date	17 June 2025
Team ID	178047
Project Title	Unemployed Insurance Beneficiary Forecasting
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 Template

Section	Description
Project Overview	This project aims to forecast the number of individuals who will apply for and receive unemployment insurance benefits over time. By analyzing historical records of beneficiaries, benefit amounts, and related demographic and regional data, the model will help government agencies and policymakers anticipate demand and allocate resources more efficiently.
Data Collection Plan	The primary dataset is sourced from the New York State government's open data portal ( <a href="https://data.world/data-ny-gov/xbjp-8sra">https://data.world/data-ny-gov/xbjp-8sra</a> ). Data is downloaded in CSV format and read using the pandas read_csv() function. The dataset contains monthly records of unemployment insurance claims, including region, county, beneficiary counts, and benefit amounts. Other formats such as Excel, TXT, or JSON can also be supported if needed.

Raw Data Sources Identified	The main raw data source is the CSV file “insurance_unemployed_data.csv” from the NY State open data portal. This file includes the following columns: year, month, region, county, beneficiaries, benefit_amounts_dollars, and yearmonth. The dataset contains 13,760 rows with no missing or duplicate values, as verified using pandas methods (df.info(), df.isna().sum(), df.duplicated().sum()). This ensures a clean and reliable basis for further preprocessing and modeling.
-----------------------------	--

### Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Dataset 1	Monthly records of unemployment insurance beneficiaries, benefit amounts, regions, and counties in New York State.	<a href="https://data.world/data-ny-gov/xbjp-8sra">https://data.world/data-ny-gov/xbjp-8sra</a>	CSV	~2 MB	Public