

## Data Collection and Preprocessing Phase

Date	24 June 2025
Team ID	SWUID20250176209
Project Title	Machine Learning Approach for Employee Performance 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	This machine learning project focuses on predicting garment worker productivity using shop-floor and departmental data. By leveraging features like SMV, overtime, incentives, idle time, and team information, the goal is to build a model that accurately forecasts actual productivity. This supports more effective HR decision-making, workforce planning, and operational efficiency in a manufacturing environment.
Data Collection Plan	<ul style="list-style-type: none"> <li>● Search for datasets related to employee productivity and performance evaluation in manufacturing sectors.</li> <li>● Prioritize datasets containing both numerical and categorical features like team, department, SMV, and productivity scores.</li> </ul>
Raw Data Sources Identified	The raw data for this project comes from <b>Kaggle's Garments Worker Productivity Dataset</b> . It contains key fields like department, team, targeted and actual productivity, SMV, overtime, and incentives, offering a solid foundation for building a machine learning model.

### Raw Data Sources Report:

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle Dataset	Contains garment worker data including productivity, incentives, and departmental info.	<a href="https://www.kaggle.com/datasets/utkarshsarbahi/productivity-prediction-of-garment-employees">https://www.kaggle.com/datasets/utkarshsarbahi/productivity-prediction-of-garment-employees</a>	CSV	93 kB	Public