

## Data Collection and Preprocessing Phase

Date	05 August 2025
Skillwallet ID	SWUID20250186419
Project Title	Employee Performance 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 Template

Section	Description
Project Overview	This project aims to predict employee productivity using a machine learning model, as part of a Flask-based web application. The core objective is to provide a tool for stakeholders to assess and forecast worker productivity based on various features.
Data Collection Plan	The data for this project will be sourced from a single, publicly available dataset that provides comprehensive details on garment worker productivity. This dataset will be downloaded and saved locally in a CSV format for use in the project.
Raw Data Sources Identified	The primary data source is the garments_worker_productivity.csv dataset, which contains detailed information on worker productivity, including attributes such as team, targeted productivity, and various time-related metrics.

### Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
garments_worker_productivity.csv	Dataset containing information on garment worker productivity, including team, date, targeted productivity, SMV, WIP, and other relevant features.	<a href="#">link</a>	csv	95kB	Public