Data Collection and Preprocessing Phase

Date	6th July 2024
Team ID	739719
Project Title	Garment Workers Productivity Predictions
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification Template

Data Collection Plan: Outline objectives, data sources (surveys, CRM, social media, e-commerce, government, academic), data types, frequency, volume, methodology, and security.
Raw Data Sources: Identify sources, relevant variables, and access methods, noting potential data quality issues and resolution plans.

Data Collection Plan Template

Section	Description
Project Overview	Provide a brief overview of the project, its objectives, and the importance of data collection.
Data Collection Plan	From Dashboard source we collected the data

Raw Data Sources Identified	**1. Customer Surveys** • **Description**: Surveys conducted among the customer base to gather information on demographics, preferences, and feedback. • **Relevant Variables**: Age, gender, location, product preferences, satisfaction ratings. **2. Social Media** • **Description**: Data collected from major social media platforms, including user-generated content and engagement metrics. • **Relevant Variables**: Post content, likes, shares, comments, follower count, sentiment analysis. **3.E-commerce Platforms** • **Description**: Data from online sales platforms, detailing customer transactions and product reviews. • **Relevant Variables**: Purchase history, product ratings, customer reviews, transaction timestamps. **4.Government Databases** • **Description**: Publicly available datasets from government agencies, providing demographic and economic data. • **Relevant Variables**: Population demographics, labor statistics, income levels, economic indicators.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Dataset 1	Productivity Prediction of Garment Employees contains data to predict garment worker productivity. It includes various features like date, department, team number, targeted productivity, actual productivity, incentive, number of workers, and idle time. This dataset is designed to help in understanding and improving the productivity levels within garment manufacturing processes.	https://www.kaggle.com/datasets/ishadss/productivity-prediction-of-garment-employees	CSV	104 KB	Public