Data Collection and Preprocessing Phase

Date	5th July 2024
Team ID	740141
Project Title	Garment Workers Productivity Predictions
Maximum Marks	6 Marks

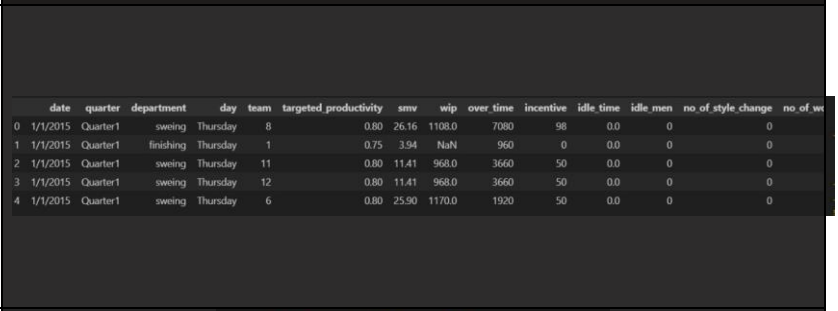
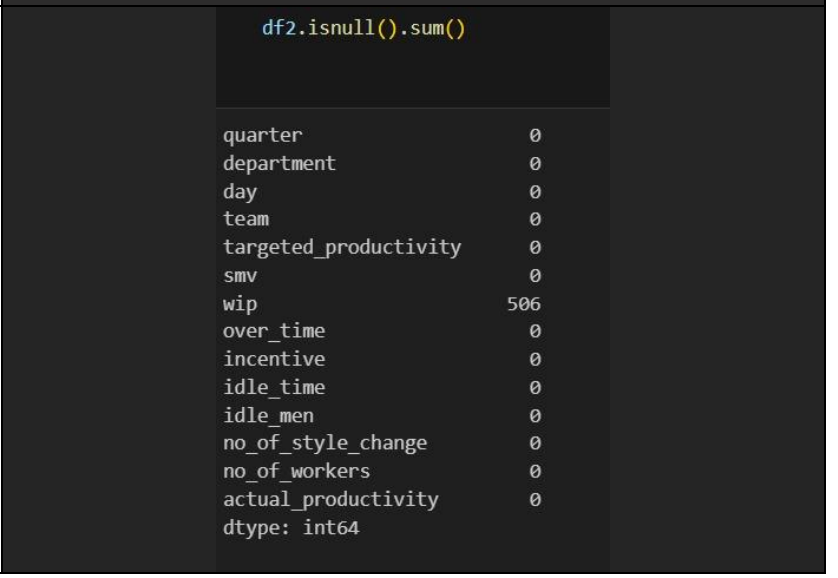
Preprocessing Template

The images will be preprocessed by Data collection,Handling missing values,Normalization,Data aggregation,Data splitting,Visualization,Data storage,Documentation.This template provides a concise yet complete approach to preprocessing a dataset in preparation for machine learning.

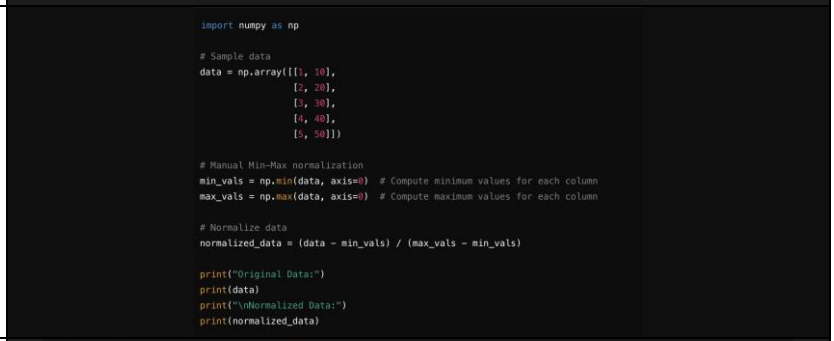
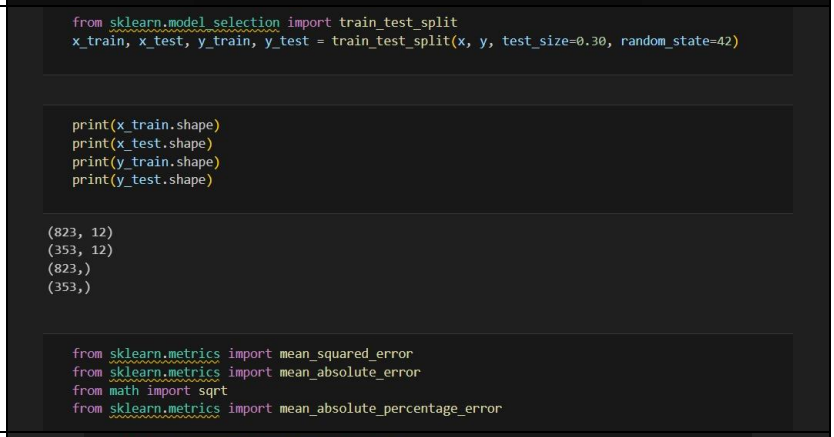
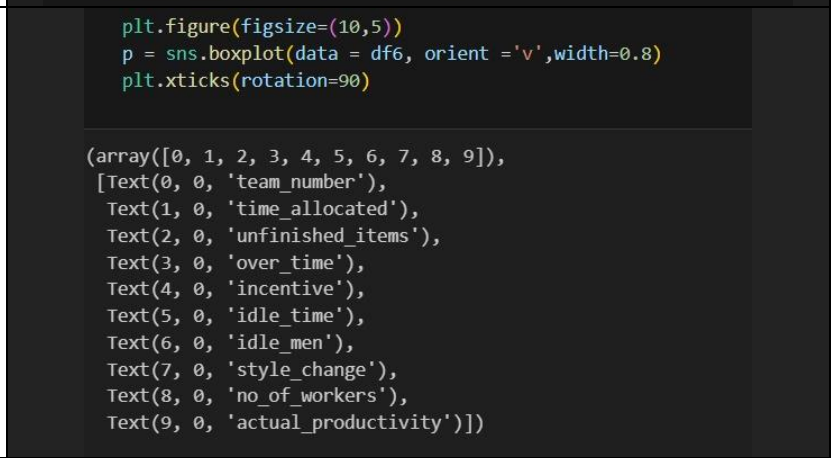
Section	**Description**
Project Overview	Brief description of the project and its goals.
Data Collection	List of data sources and Types of data collected
Handling Missing Values	Techniques used (e.g., imputation, removal).
Normalization:	Procedures for scaling data
Data Aggregation:	Summarizing data to higher-level formats.
Data Splitting	Proportion of data split and Strategy used for cross-validation

Visualization Tools:	Tools and libraries used (e.g., Matplotlib, Seaborn).
Data Storage	Systems used for storing preprocessed data (e.g., SQL, NoSQL, cloud storage).
Documentation	Detailed documentation of all preprocessing steps for reproducibility.and A comprehensive data dictionary describing all variables and their transformations

Data Preprocessing Code Screenshots

Loading Data	
Data Collection	
Handling Missing Values	

	date	quarter	department	day	team	targeted_productivity	smv	wip	over_time	incentive	idle_time	idle_men	no_of_style_change	no_of_wo
0	1/1/2015	Quarter1	sweing	Thursday	8	0.80	26.16	1108.0	7080	98	0.0	0	0	
1	1/1/2015	Quarter1	finishing	Thursday	1	0.75	3.94	NaN	960	0	0.0	0	0	
2	1/1/2015	Quarter1	sweing	Thursday	11	0.80	11.41	968.0	3660	50	0.0	0	0	
3	1/1/2015	Quarter1	sweing	Thursday	12	0.80	11.41	968.0	3660	50	0.0	0	0	
4	1/1/2015	Quarter1	sweing	Thursday	6	0.80	25.90	1170.0	1920	50	0.0	0	0	

Normalization	
Data Aggregation	
Data Splitting	
Visualization Tools:	
Data Storage	

Documentation	`df = pd.read_csv(r'C:\Users\srira\Downloads\miniProject\garments_worker_productivity.csv')` `df.head()`