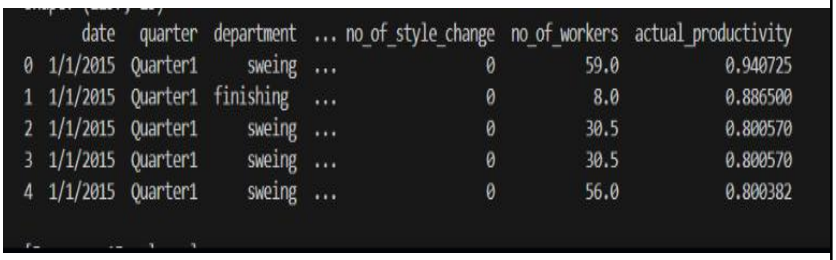
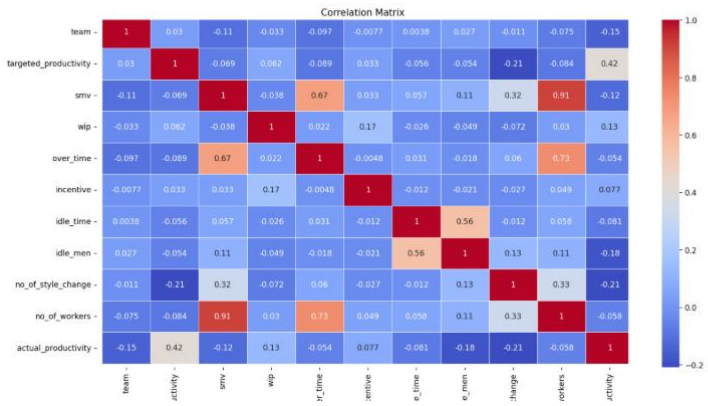
Data Collection and Preprocessing Phase

Date	26 June 2025
Team ID	NONE
Project Title	Employee Performance Prediction using Machine Learning

Data Exploration and Preprocessing Report

The dataset variables are analyzed to identify patterns and relationships affecting employee productivity. Using Python, preprocessing tasks such as handling missing values, label encoding categorical features, and exploratory data analysis (EDA) are performed. Data cleaning ensures high-quality input for machine learning models, while correlation analysis and visualization help uncover key drivers of performance. This phase forms a solid foundation for accurate productivity prediction using algorithms like Random Forest and XGBoost

Section	Description
Data Overview	Dimension: 1197 rows × 15 columns Descriptive statistics: 
Univariate Analysis	Univariate analysis was performed to understand the distribution of individual features in the dataset. Key observations include: actual_productivity (Target Variable): The distribution is approximately normal with a slight right skew, centered around 0.77, indicating most employees perform close to average. over_time: Highly right-skewed — many employees work moderate overtime, while a few work significantly longer hours. idle_time: Most employees have low idle time, with peaks at 0 and 60 minutes, suggesting scheduled breaks or downtime. department: "sweing" department has the highest employee count, followed by "finishing". quarter: Q2 and Q4 show higher activity, likely due to seasonal demand.

	These insights were visualized using histograms and bar plots in visualization.py to guide preprocessing and feature engineering decisions.
	

Outliers and Anomalies	
Data Preprocessing Code Screenshots	
Loading Data	(base) PS C:\Users\vansh\OneDrive\Desktop\employee_performance_ml> & Linear Regression R² Score: 0.1681682566306545 Random Forest R² Score: 0.44671974539154335 XGBoost R² Score: 0.3538597397101051 Best model saved to model/best_model.pkl (base) PS C:\Users\vansh\OneDrive\Desktop\employee_performance_ml>

Feature Engineering