DAV Project Report

# Project Title: Productivity Analysis and Forecasting in Apparel Manufacturing

Submitted By:  
- Shehzana Bibi – Reg No: BSCS-2022-44  
- Sarmad Sultan – Reg No: BSCS-2022-15

Submitted To: Sir Bilal Hydral

Course: Data Analysis and Visualization (DAV)

## Objective of the Project

The aim of this project is to perform detailed exploratory data analysis and predictive modeling on an apparel manufacturing dataset. The dataset contains information on multiple factors that affect productivity on a daily basis such as SMV (Standard Minute Value), overtime, idle time, number of workers, and more. Using this data, we attempt to draw insights and forecast productivity trends using regression and, where possible, time series techniques.

## Dataset Overview

- The dataset includes 1197 records across 15 columns.  
- The main target variable is actual productivity.  
- Important features include:  
 - team  
 - smv  
 - wip  
 - over\_time  
 - idle\_time  
 - no\_of\_workers  
 - targeted\_productivity  
 - quarter, department, and day (categorical variables)

## Preprocessing Steps

- Missing values (especially in wip) were handled through imputation.  
- Categorical variables such as day, quarter, and department were converted to numerical format using one-hot encoding.  
- Features were scaled where appropriate to normalize their ranges.  
- Data was split into training and testing sets for model evaluation.

## Linear Regression Modeling

We implemented a Linear Regression model to predict actual\_productivity.  
  
- R² Score: ~0.36  
- RMSE (Root Mean Squared Error): Indicated moderate error margin.  
  
Although the model is simple, it still captured some variability in productivity. However, the modest R² score suggests the need for a more complex or nonlinear model or additional features.  
  
Insert Plot: Scatter plot of actual vs. predicted productivity

## Correlation and Feature Importance

We analyzed correlation among numerical features to identify key predictors:  
- Targeted productivity and SMV showed notable influence on actual productivity.  
- Idle time and over time also negatively impacted productivity.  
- No of workers generally had a positive correlation with productivity.  
  
Insert Plot: Correlation heatmap

## Time Series Analysis

We attempted to apply seasonal decomposition using the seasonal\_decompose function from statsmodels.  
  
Findings:  
- Due to limited data at the monthly aggregation level, time series decomposition could not be fully applied. It requires at least two full cycles (e.g., 24 months) for meaningful trend and seasonality extraction.  
- However, we generated a monthly average plot which hinted at small fluctuations in productivity over months.  
  
Insert Plot: Line plot showing monthly average productivity

## Key Insights & Conclusions

- Targeted Productivity vs. Actual: In most cases, the actual productivity did not reach the targeted levels, indicating systemic inefficiencies or overestimation.  
- SMV and Idle Time: Higher SMV and idle times correlate with decreased actual productivity, highlighting bottlenecks or mismanagement in task assignments.  
- Team & Department Performance: Some teams and departments consistently performed better than others, suggesting uneven skill distribution or workflow optimization.  
- Predictive Modeling: Linear regression showed potential but with limited accuracy, suggesting a need for ensemble models or neural networks for better forecasting.  
- Time Series Limitations: While initial attempts were made to analyze trends over time, the dataset was insufficient in size for robust temporal modeling.

## Future Recommendations

- Data Collection: A longer and more complete time series would allow for deeper forecasting using ARIMA or LSTM models.  
- Feature Engineering: Adding more features such as weather, shift timings, or employee data could enhance predictive power.  
- Model Enhancement: Trying models like Random Forests, XGBoost, or SVR could improve prediction performance.

## Instructions for Visualization

Wherever mentioned to 'Insert Plot', you can insert corresponding visualizations generated in the Jupyter Notebook:  
- Scatter plot for linear regression predictions  
- Correlation heatmap  
- Monthly line plot for productivity