



**DATA SCIENCE PROJECT REPORT  
ON  
PRODUCTIVITY PREDICTION OF GARMENTS  
EMPLOYEES DATASET**

**GROUP MEMBERS:**

**Hafsa Baig (20k-1683)  
Maryam Rahim (20k-1700)**

**INSTRUCTOR:**

**Dr. Noman Durrani**

**SECTION: BSE-6A**

## INTRODUCTION:

This report provides an analysis of the Garments Worker Productivity dataset. The dataset contains information about the productivity of workers in a garments factory, including various attributes such as date, day of the week etc.

## UNDERSTANDING THE DATA:

The dataset consists of the following attributes:

**date** : The date of the observation

**day** : The day of the week (Sunday to Saturday)

**quarter** : The quarter of the year (1 to 4)

**department** : The department in which the worker is employed (e.g., finishing, finishing1, finishing2, ...)

**team** : The team number in which the worker is employed (1 to 12)

**targeted\_productivity** : The targeted productivity for the worker on a given day

**smv** : The standard minute value (SMV) for the task being performed by the worker

**wip** : The work in progress (WIP) for the task being performed by the worker

**over\_time** : The overtime hours worked by the worker on a given day

**incentive** : The incentive given to the worker on a given day

**idle\_time** : The idle time for the worker on a given day

**idle\_men** : The number of idle workers on a given day

**no\_of\_workers** : The number of workers in the team on a given day

**actual\_productivity** : The actual productivity achieved by the worker on a given day.

## DATA SUMMARY:

The dataset consists of a total of 1197 rows and 15 attributes.

The datatype of each attributes are as follows:

**date** object  
**quarter** object  
**department** object  
**day** object  
**team** int64  
**targeted\_productivity** float64  
**smv** float64  
**wip** float64  
**over\_time** int64  
**incentive** int64  
**idle\_time** float64  
**idle\_men** int64  
**no\_of\_style\_change** int64  
**no\_of\_workers** float64  
**actual\_productivity** float64  
**dtype:** object

**Head:** Display first few records of the dataset.

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

**Tail:** Display the last few records of the dataset.

```
----- TAIL -----
```

	date	quarter	department	day	team	targeted_productivity	smv	wip	over_time	incentive	idle_time	idle_men	no_o
1192	3/11/2015	Quarter2	finishing	Wednesday	10	0.75	2.9	NaN	960	0	0.0	0	
1193	3/11/2015	Quarter2	finishing	Wednesday	8	0.70	3.9	NaN	960	0	0.0	0	
1194	3/11/2015	Quarter2	finishing	Wednesday	7	0.65	3.9	NaN	960	0	0.0	0	
1195	3/11/2015	Quarter2	finishing	Wednesday	9	0.75	2.9	NaN	1800	0	0.0	0	
1196	3/11/2015	Quarter2	finishing	Wednesday	6	0.70	2.9	NaN	720	0	0.0	0	

**Null Values:** This dataset contains some null values in the ‘wip’ attribute, which we have replaced with mean value.

```
----- NULL VALUES -----
```

```
date                0
quarter             0
department          0
day                0
team               0
targeted_productivity 0
smv                0
wip                506
over_time           0
incentive           0
idle_time           0
idle_men            0
no_of_style_change  0
no_of_workers       0
actual_productivity 0
dtvpe: int64
```

Additional Information about the dataset is as follows:

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1197 entries, 0 to 1196
Data columns (total 15 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   date                                  1197 non-null   object
1   quarter                              1197 non-null   object
2   department                           1197 non-null   object
3   day                                   1197 non-null   object
4   team                                  1197 non-null   int64
5   targeted_productivity                1197 non-null   float64
6   smv                                   1197 non-null   float64
7   wip                                   691 non-null    float64
8   over_time                            1197 non-null   int64
9   incentive                            1197 non-null   int64
10  idle_time                            1197 non-null   float64
11  idle_men                             1197 non-null   int64
12  no_of_style_change                   1197 non-null   int64
13  no_of_workers                        1197 non-null   float64

```

```

10  idle_time                            1197 non-null   float64
11  idle_men                             1197 non-null   int64
12  no_of_style_change                   1197 non-null   int64
13  no_of_workers                        1197 non-null   float64
14  actual_productivity                  1197 non-null   float64
dtypes: float64(6), int64(5), object(4)
memory usage: 140.4+ KB
None

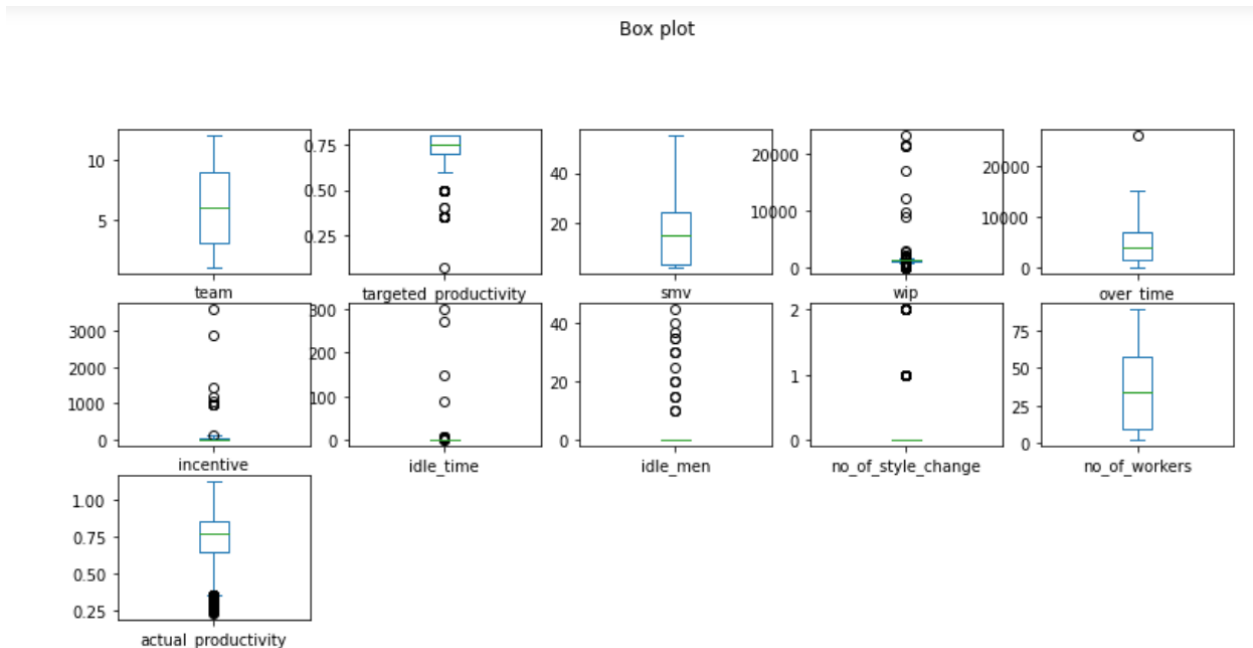
```

## DATA VISUALIZATION:

To gain further insights into the dataset, various visualizations have been created.

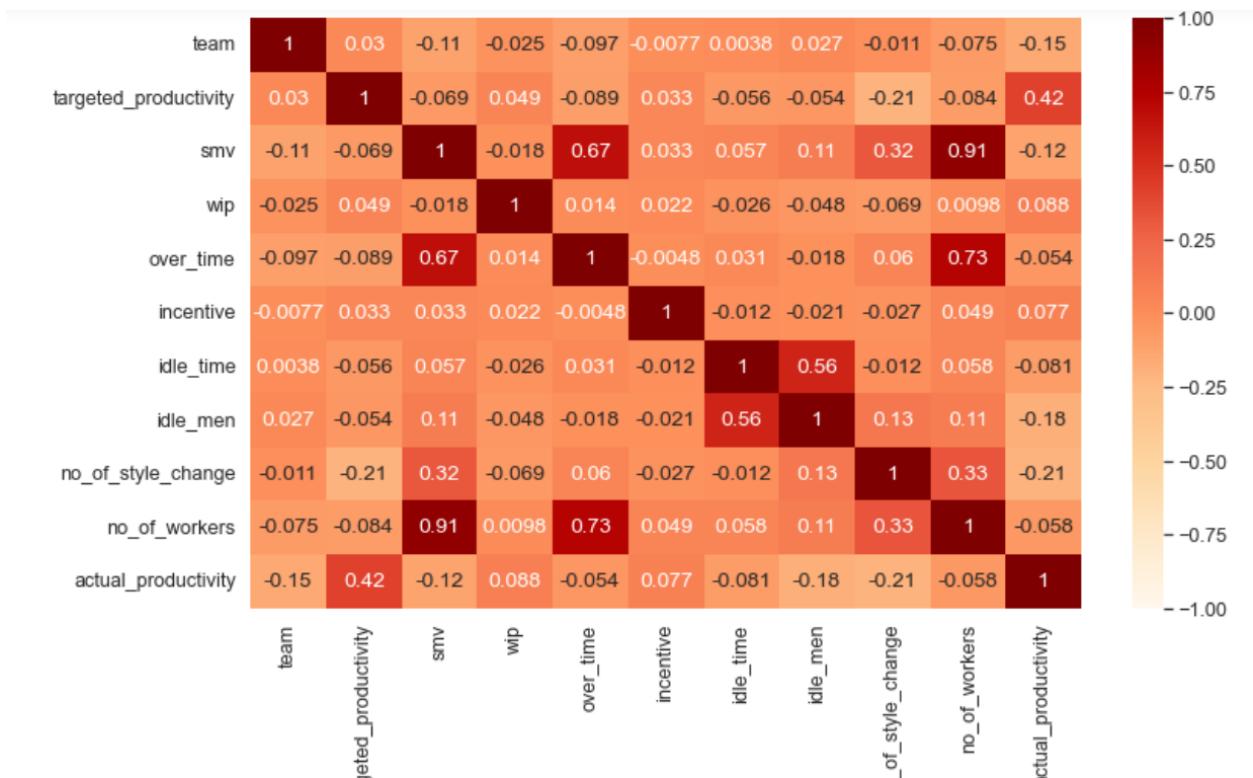
### BOX PLOT:

A box plot was created to visualize the distribution of the dataset attributes. The box plot provides information about the median, quartiles, and outliers for each attribute.



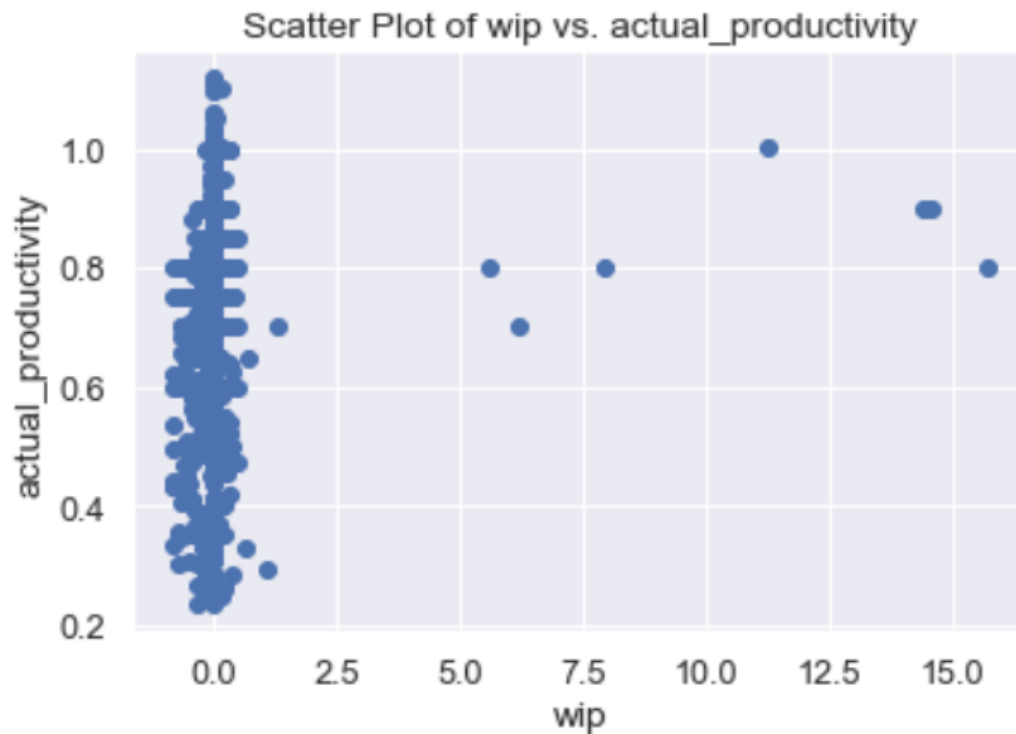
## Correlation Heat Map:

It was generated to visualize the correlation between different attributes in the dataset.



## Scatter Plot:

To visualize the relationship between 'wip' and the 'actual\_productivity'.



## Applying ML Algorithm:

To predict the actual productivity of workers, ML algorithms were applied to the dataset.

## Data Transformation:

Before applying the machine learning algorithms, the data was transformed as:

- We converted categorical values to numerical values.
- Then those numerical values were standardized using StandardScaler.

## **Linear Regression:**

We use following predictors:

Targeted\_productivity, smv, wip, over\_time, incentive, idle\_time and no\_of-workers.

The dataset was split into training and testing datasets, then a linear regression model was trained on the training set. The result of the Linear Regression algorithm are as follows:

**alpha**=0.31522097332966414

**Linear Regression accuracy:** 19.97249698668001 %

## **Other ML Algorithms:**

For each algorithm, the model was trained on the training set and evaluated using R-squared score on the testing set. The accuracy results for each algorithm are as follows:

- Support Vector Regression(SVM):  
**Accuracy:**2.1684216596857153 %
- Random Forest Regression:  
**Accuracy:**45.55589086199572 %

This concludes the analysis of the Garments Worker Productivity dataset.



