

Productivity Prediction of Garment Employees

Dataset

<https://www.kaggle.com/datasets/ishadss/productivity-prediction-of-garment-employees>

Introduction

The Garment Industry is one of the key examples of the industrial globalization of this modern era. It is a highly labor-intensive industry with lots of manual processes. Satisfying the huge global demand for garment products is mostly dependent on the production and delivery performance of the employees in the garment manufacturing companies. So, it is highly desirable among the decision makers in the garments industry to track, analyze and predict the productivity performance of the working teams in their factories.

Problem definition

This dataset contains several attributes which can potentially influence the productivity of garments' employees. It also contains the actual productivity of the employees scaled in range (0-1). It can be used for predicting the productivity range (0-1) or for classification by transforming the productivity range (0-1) into different classes.

Dataset Description

The dataset contains 1197 rows and 15 columns.

1. date : Date in MM-DD-YYYY
2. day : Day of the Week
3. quarter : A portion of the month. A month was divided into four quarters (can be derived from date as well)
4. department : Associated department with the instance
5. teamno : Associated team number with the instance
6. no_of_workers : Number of workers in each team
7. no_of_style_change : Number of changes in the style of a particular product
8. targeted_productivity : Targeted productivity set by the Authority for each team for each day.
9. smv : Standard Minute Value, it is the allocated time for a task
10. wip : Work in progress. Includes the number of unfinished items for products
11. overtime : Represents the amount of overtime by each team in minutes

12. incentive : Represents the amount of financial incentive (in BDT) that enables or motivates a particular course of action.
13. idle_time : The amount of time when the production was interrupted due to several reasons
14. idle_men : The number of workers who were idle due to production interruption
15. actual_productivity : The actual % of productivity that was delivered by the workers. It ranges from 0-1.

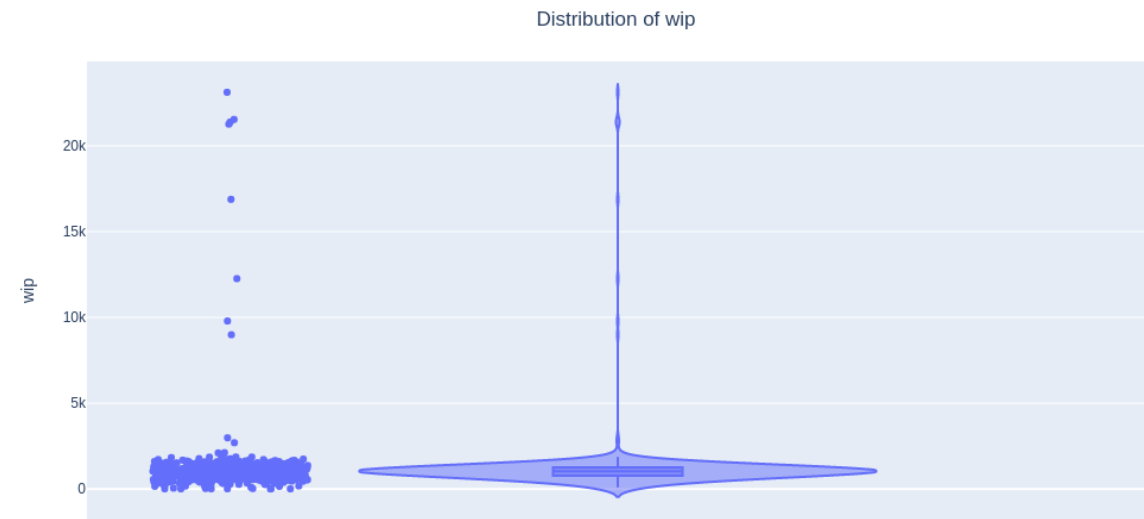
Missing data:

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

Clean data

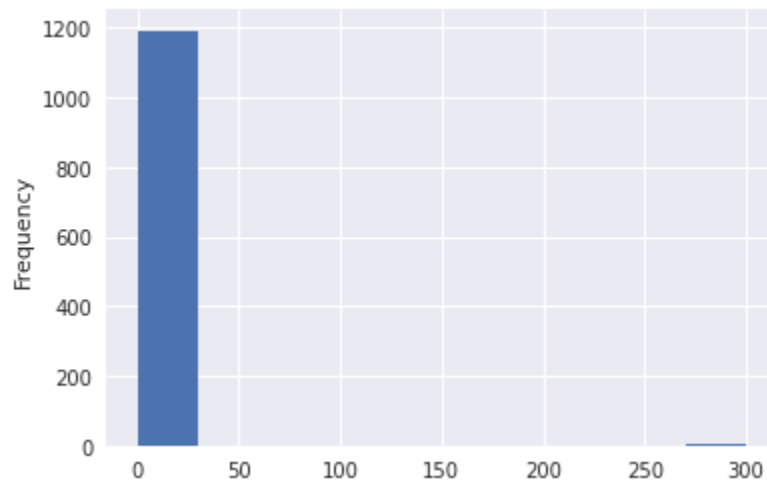
Department field contains 'sweing' in place of 'sewing' and 'finishing ' in some rows in place of 'finishing'

Outliers

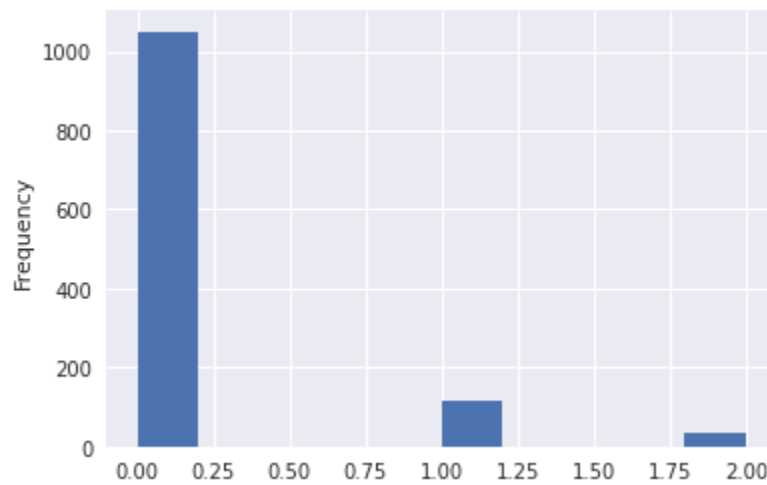


Exploring the relevant features

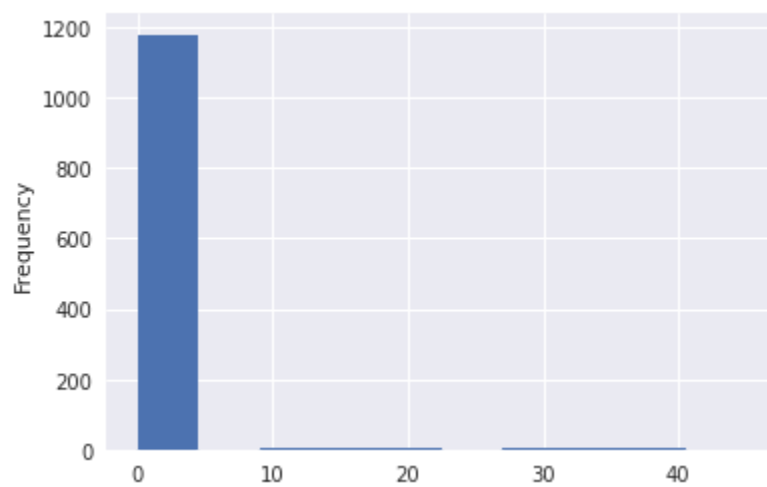
Histogram of idle_time:



Histogram of idle_men:

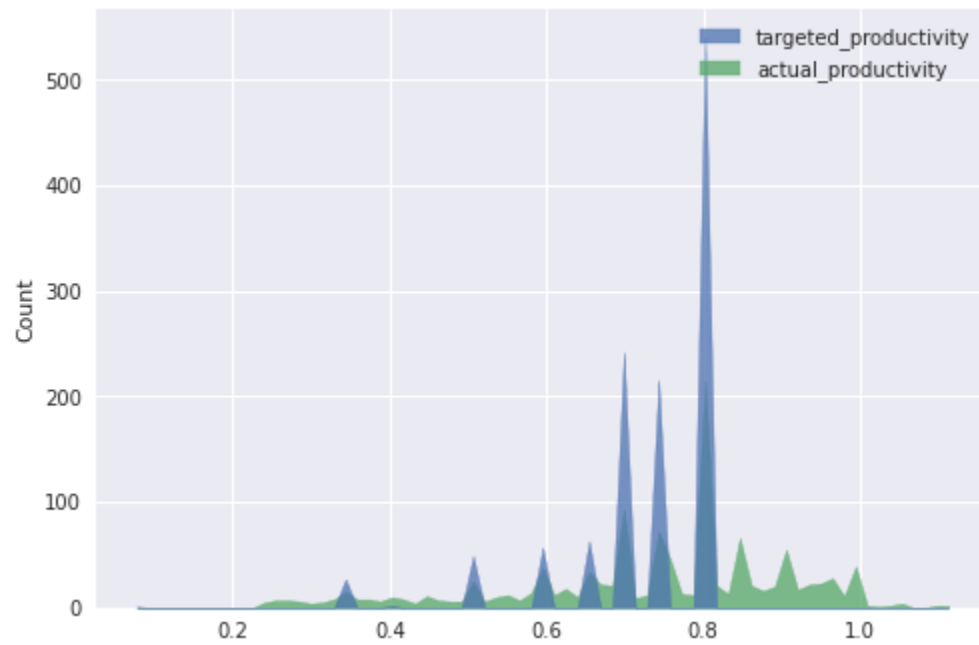


Histogram of no_of_style_change:

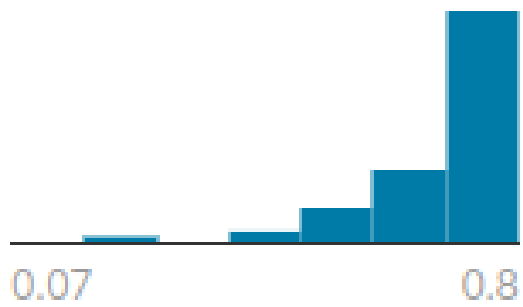


Decision : Most of the values of these three fields are Zero. We can drop these 3 columns.

Targeted vs actual productivity



Targeted productivity:



Actual productivity:

