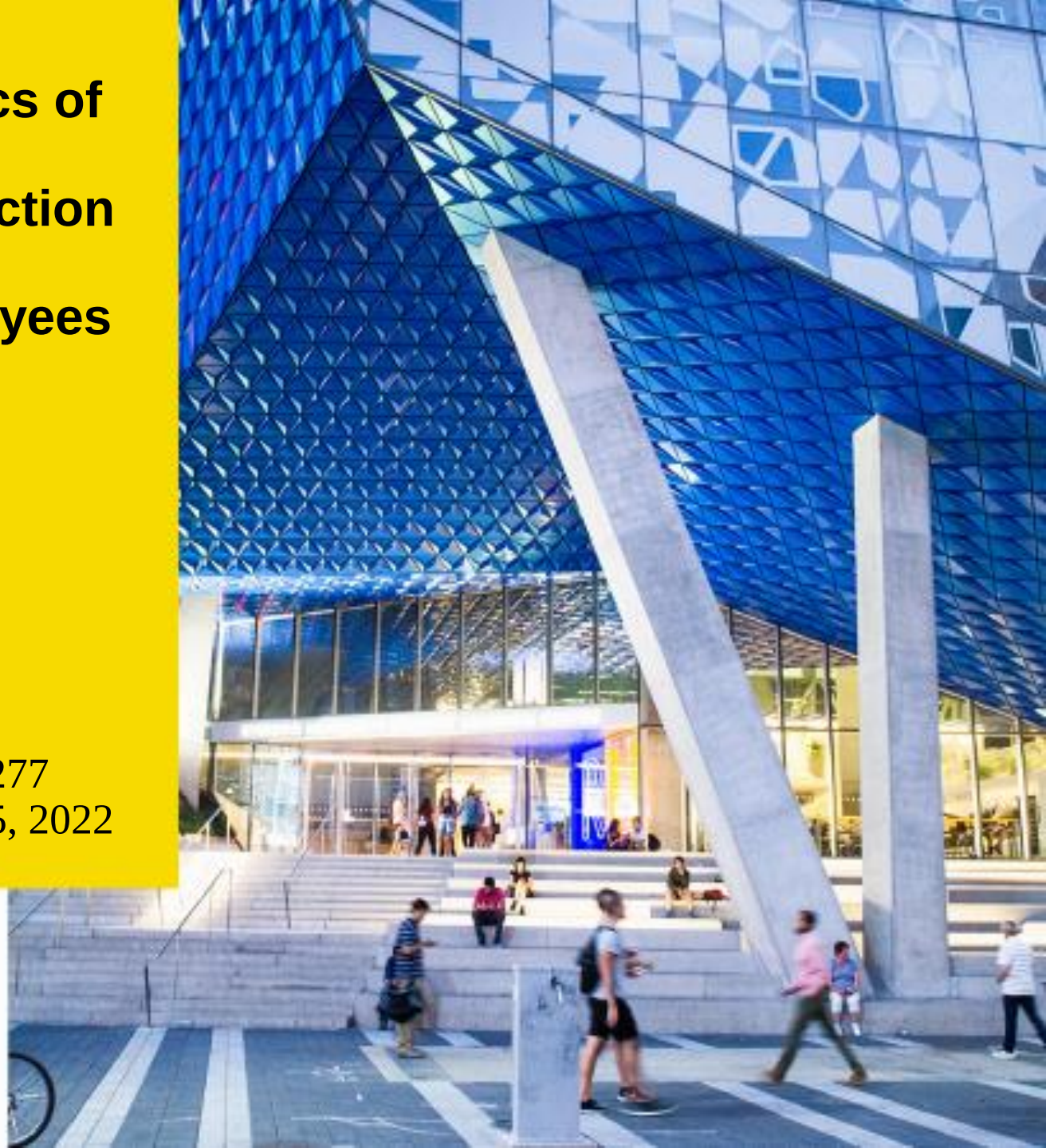
Predictive Analytics of Productivity Prediction of Garment Employees

Professor: Ceni Babaoglu
Student Name: Maryam Mortezapourdehkordi
Student Number: 501143277
Date of submission: Dec 5, 2022



Ryerson University

Introduction

Productivity of the Garment industry is very Important:

- Garment industry has significantly affected contemporary society.
- The vast number of labour and manual procedures and requires much effort
- Mainly responsible for meeting the enormous worldwide demand for clothing items.
- The profit margin of a business rises with the increase in productivity. When output levels increase, clothing production costs decrease.

The staff's high level of productivity is of paramount importance.

Goal and Questions

We are curious about the following questions:

- what factors may impact employee productivity or lead to increased productivity?
- How can work efficiency be predicted in a corporation given specific factors?
- Has time affected the efficiency or productivity of employees?

Goal and Questions

- Build models to classify and predict the productivity of garment employees for the dataset.

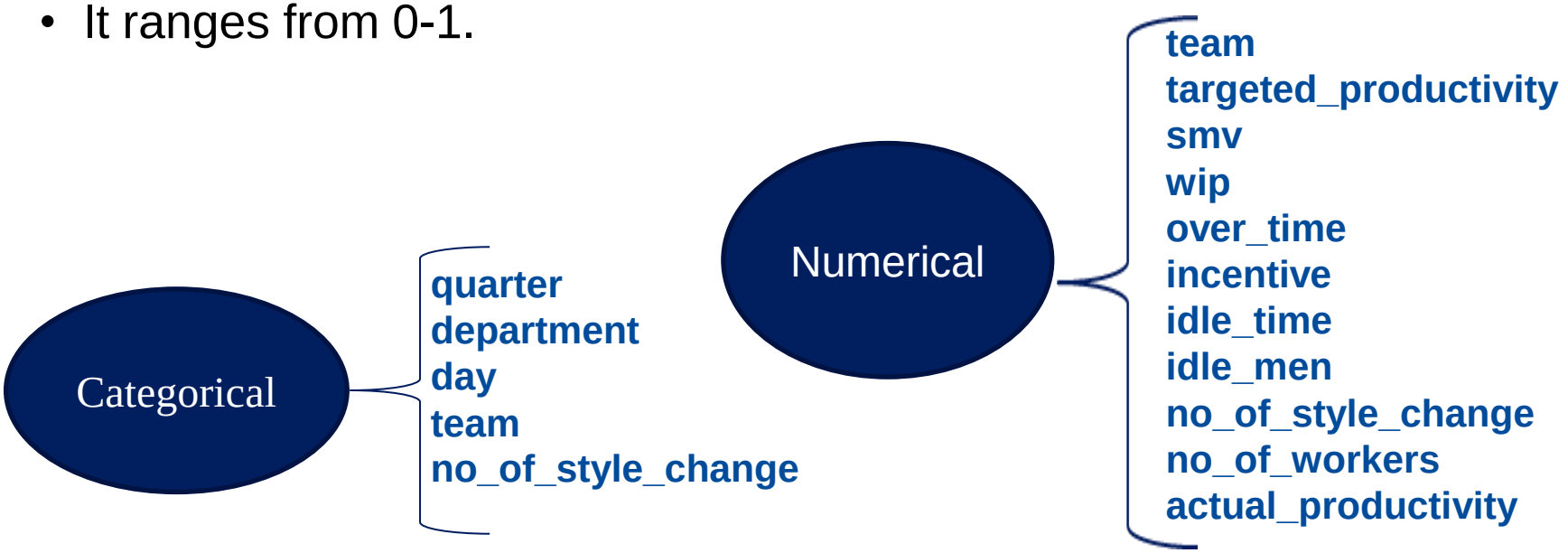
- Compare the performance of Regression Models (Linear, Lasso, Ridge, Poisson, decision tree, KNN regression, SVR, Random Forest) and boosting algorithms (xgboost, Gradient boosting).

Literature review

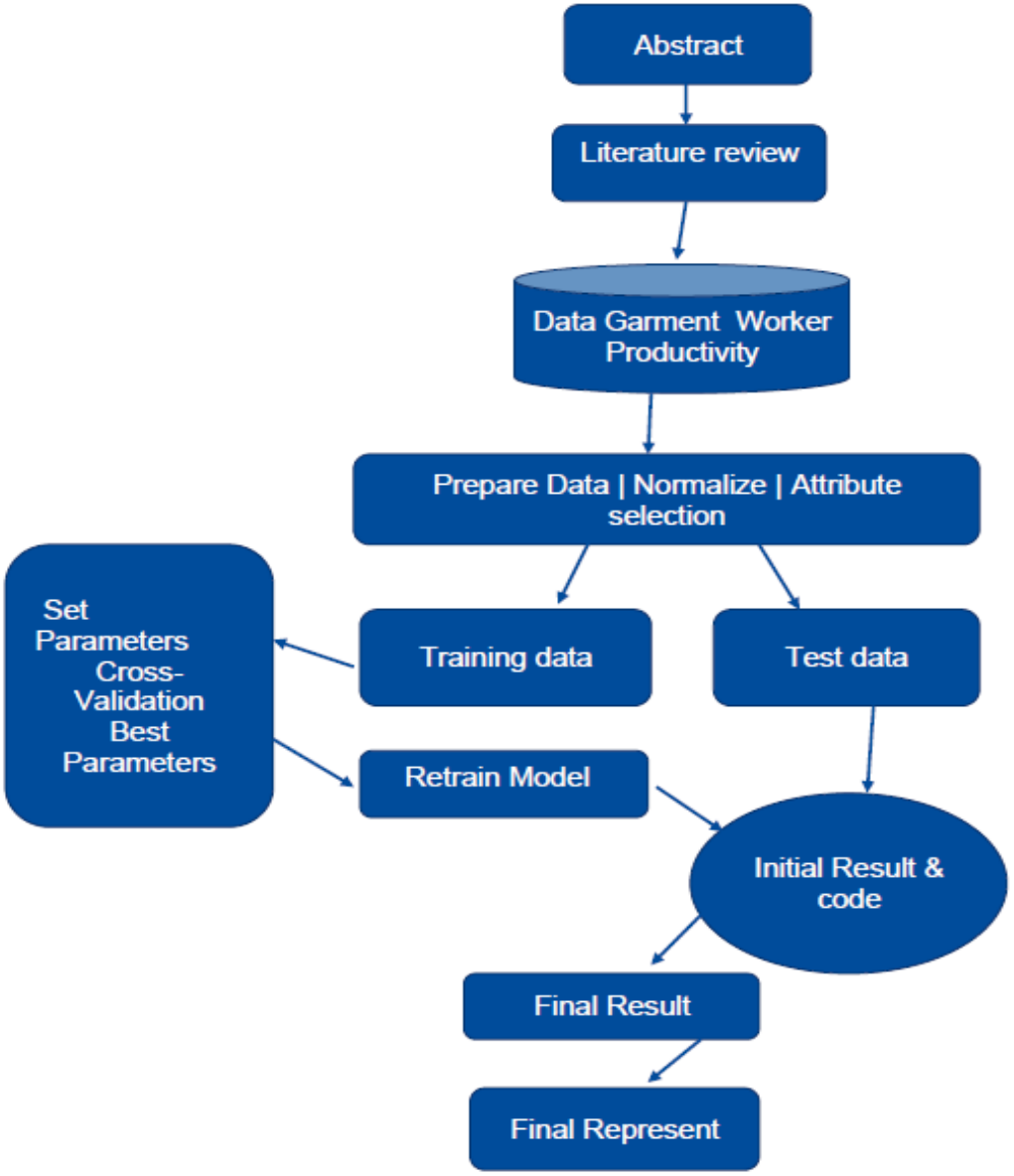
- Result some study shows in predicting employee productivity which is one of the most substantial factors in any organization. After applying classification algorithms, we obtained a Random forest has minimal values of the correlation coefficient, MAE, and RMSE, which reflect that RF is very appropriate in predicting employee productivity.
- Some results demonstrated that the gradient-boosted decision trees were the best through all the decision-tree-based methods.
- Four different ML algorithms, including support vector machine, optimized support vector machine (using a genetic algorithm), random forest, XGBoost and Deep Learning, For predicting the overall equipment effectiveness (OEE), which is a performance measurement of the manufacturing industry.
- Deep learning and random forest with cross-validation manifest the best results for predicting OEE.
- Another Study that was released focused on predicting garment employee productivity using different machine learning algorithms such as J48, RF, SVM, NB, and RBF with and without ensemble learning algorithms, including bagging and Adaboost. The best results were obtained by J48 combined with Adaboost on 20th iterations with 0.9916 accuracies, 0.0083 MAE and 0.0908 RSME. Consequently, J48 with the Adaboost algorithm was found to be the best for garment employee productivity prediction.

Dataset

- Dataset experts and published in the UC Irvine Machine Learning Repository. Dataset has 15 features for 1197 instances.
- Eleven features are numerical
- four features are categorical.
- The dependent variable is Actual Productivity, which shows The actual % of productivity that was delivered by the workers.
- It ranges from 0-1.



Methodology



Steps to implement

Data Processing

- ✓ Collecting data
- ✓ Cleaning data
- ✓ Missing data
- ✓ Labelling raw data
- ✓ Visualizing the data.
- ✓ Encoding and Scaling

Feature selecting

- ✓ Cross-validation for split data to train and test
- ✓ Recursive feature elimination (RFE) for feature selection
- ✓ Creating an estimator, Creating an RFE object
- ✓ fitting the training data into our model are all steps that I applied for feature selection,
- ✓ prepared data for data modelling

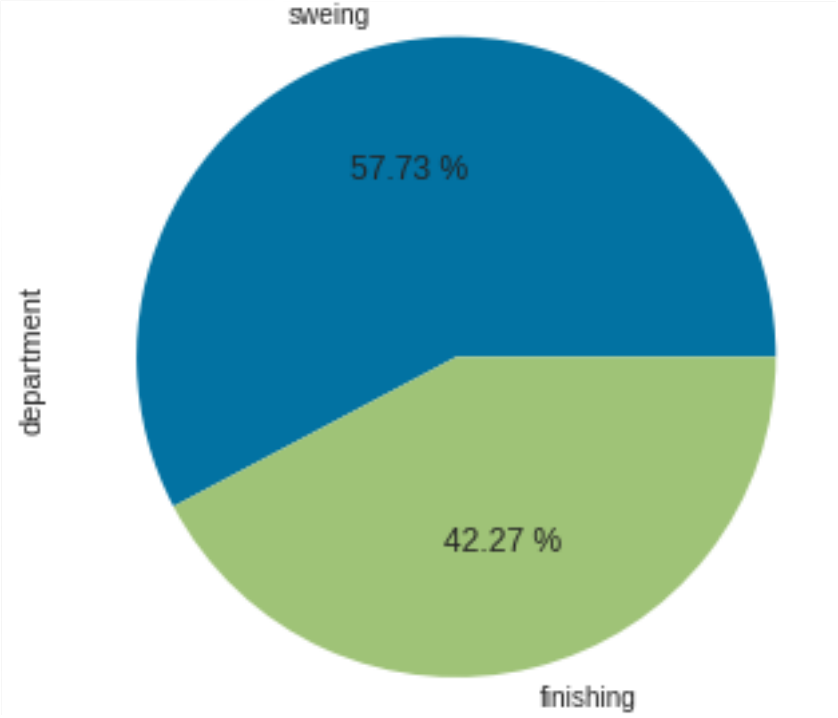
Ryerson University

Steps to implement

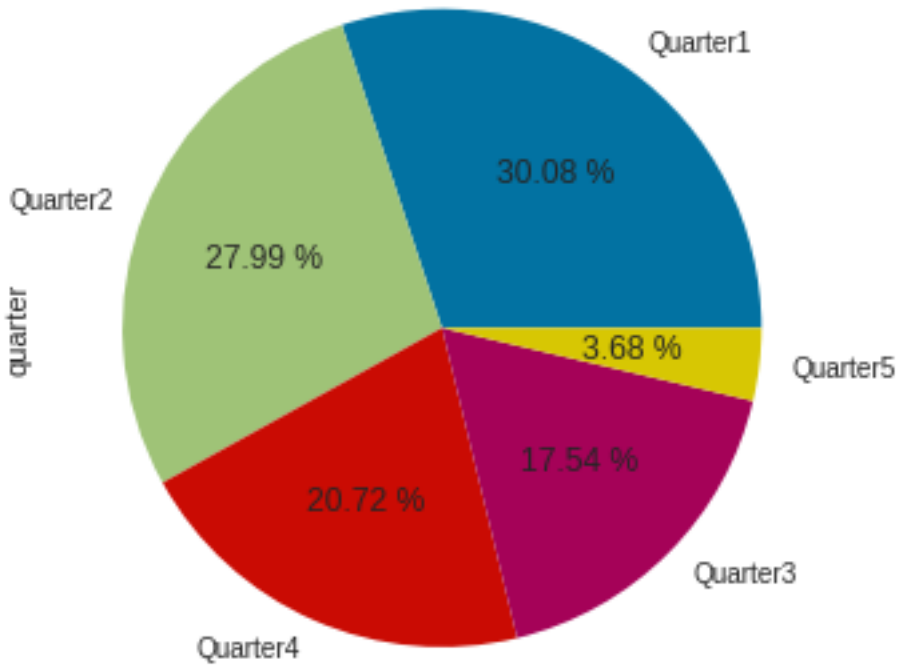
- **Model Development**

- ✓ Linear Regression
- ✓ Lasso Regression
- ✓ Ridge Regression
- ✓ Decision Trees
- ✓ Random Forest
- ✓ Support Vector Regression,
- ✓ Kneighbour Regression
- ✓ XGBoost,
- ✓ Gradient Boost.

Pie chart for understanding Categorical feature:

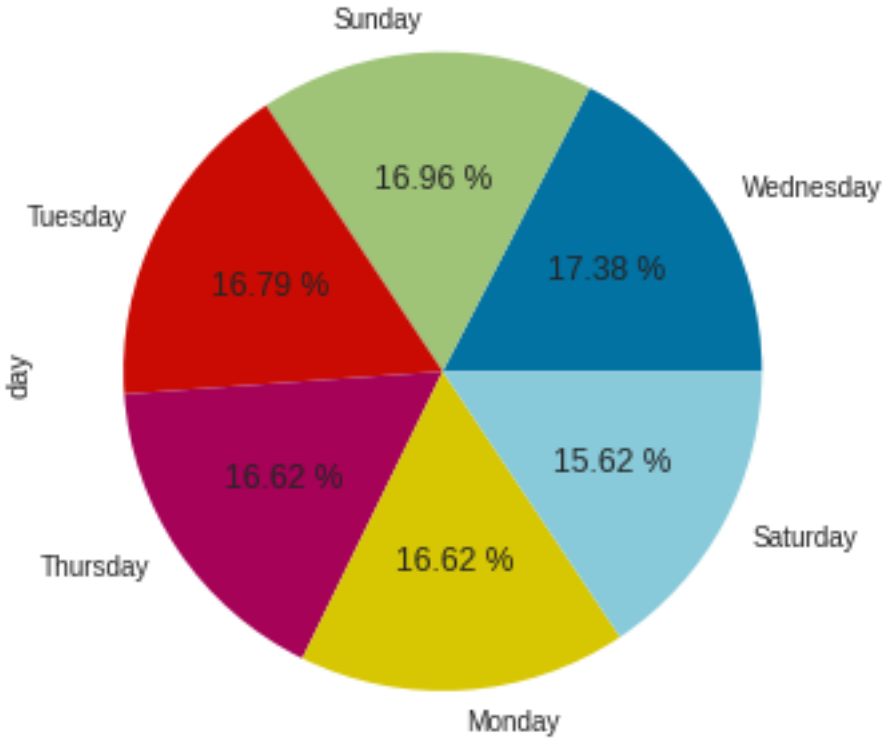


Department

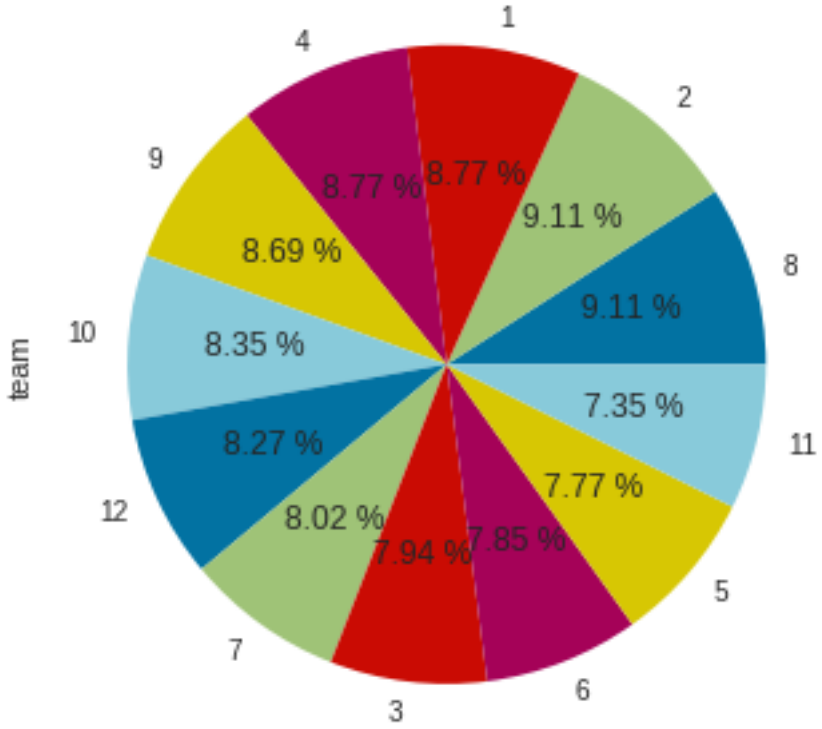


Quarter

Pie chart for understanding Categorical feature:

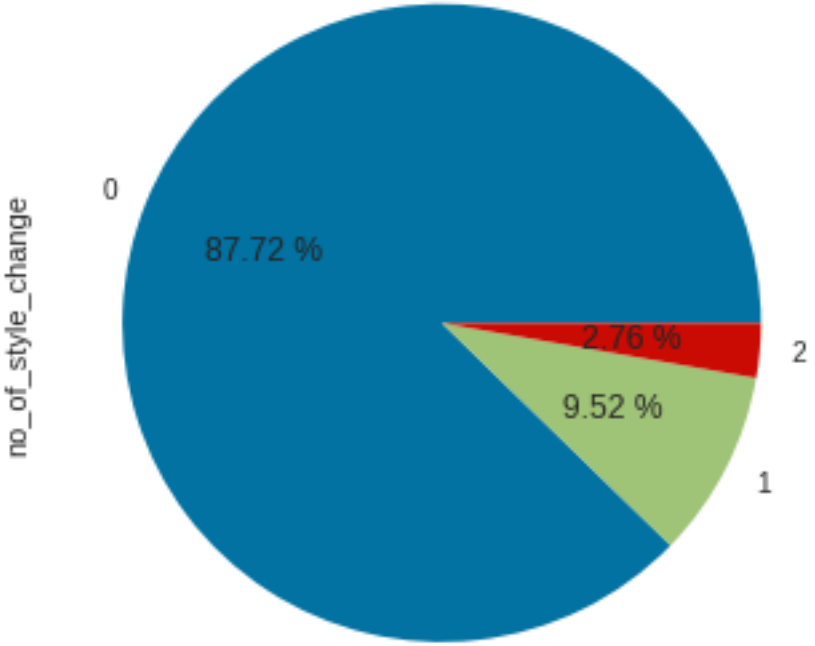


Day



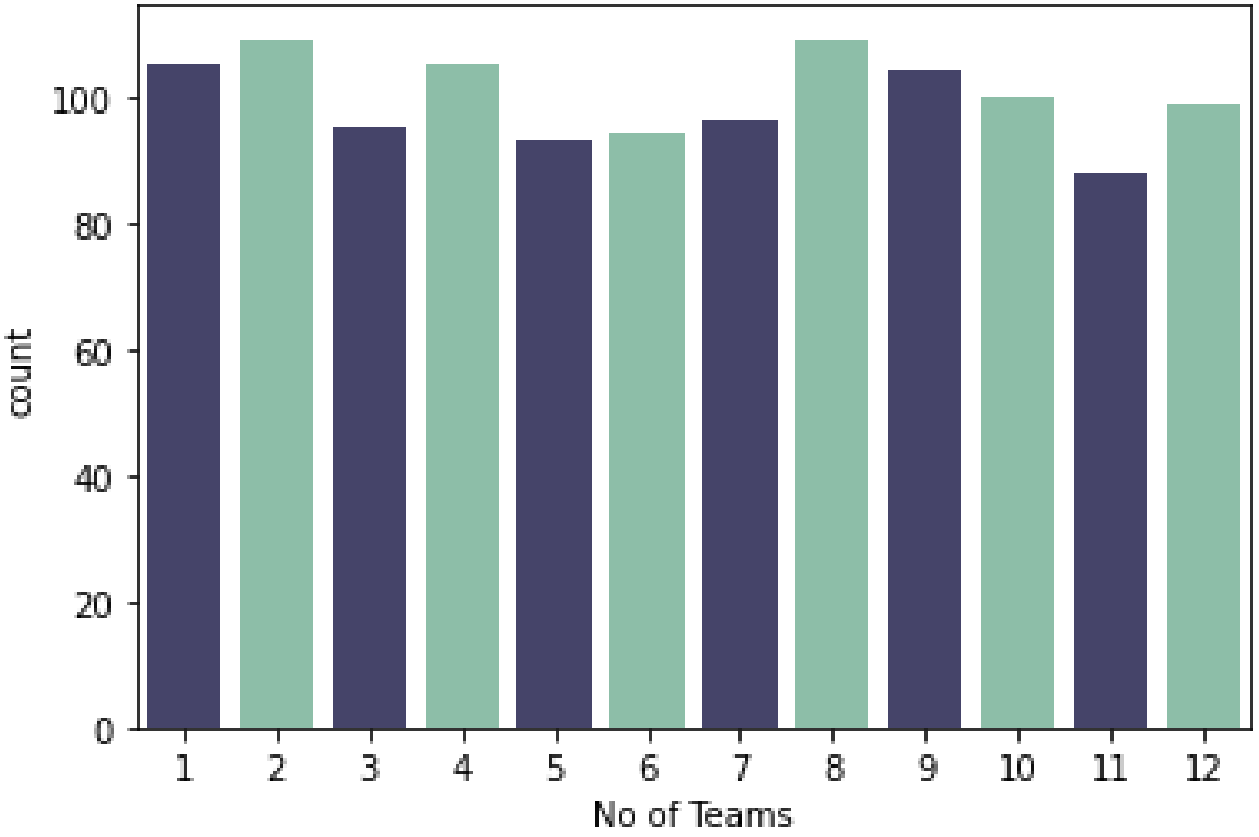
Team

Pie chart for understanding Categorical feature:



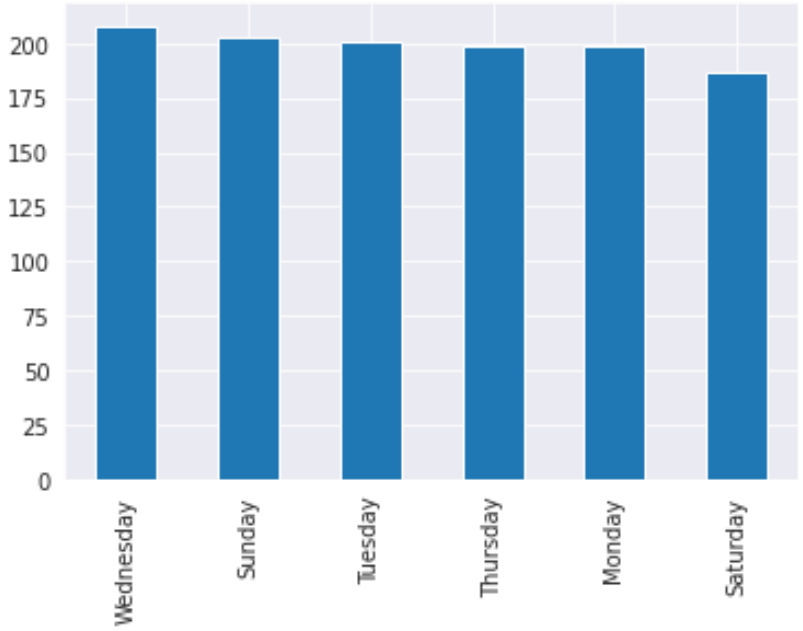
No-of-work

Bar chart for understanding Numerical feature:

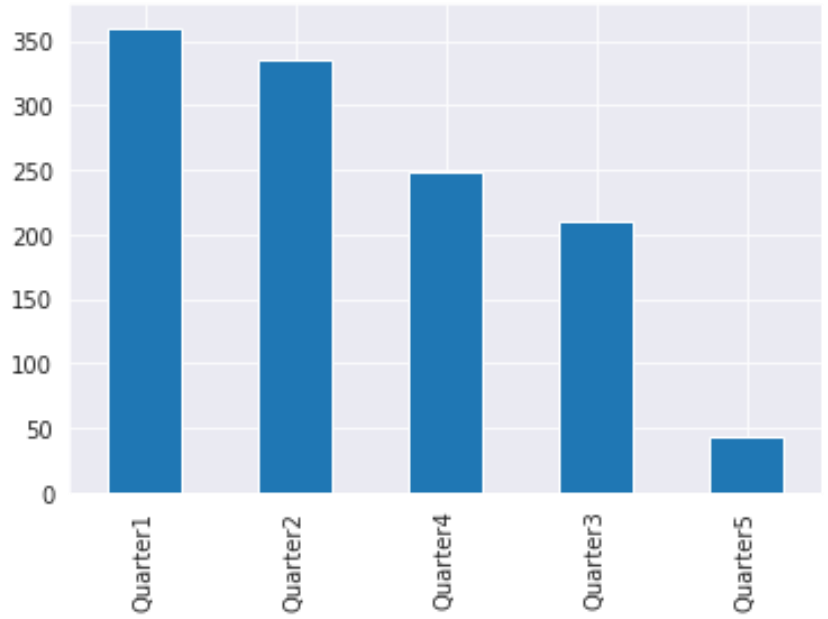


Team: There are 12 teams. Because It is highly desirable among the decision-makers in the garments industry to track, analyze and predict the productivity performance of the working teams in their factories, my analysis will be on team basis

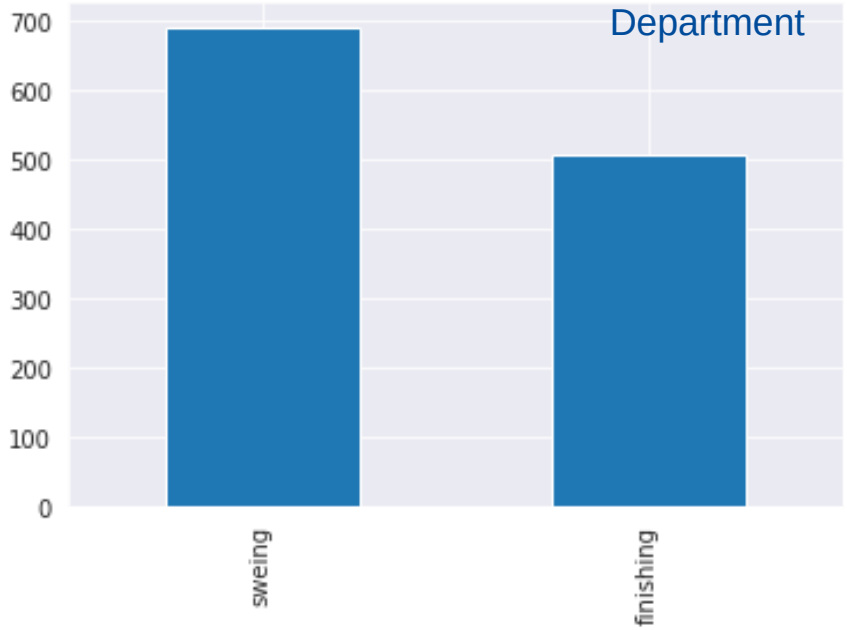
Day



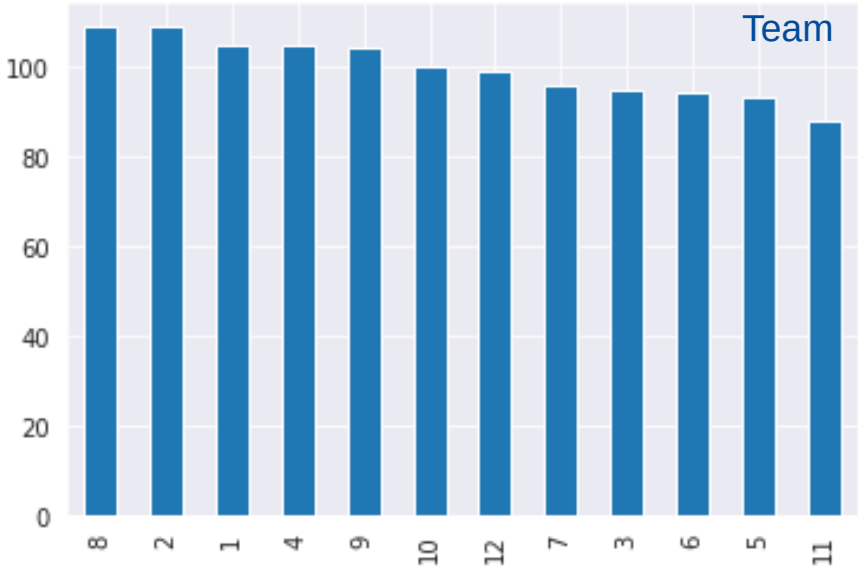
Quarter

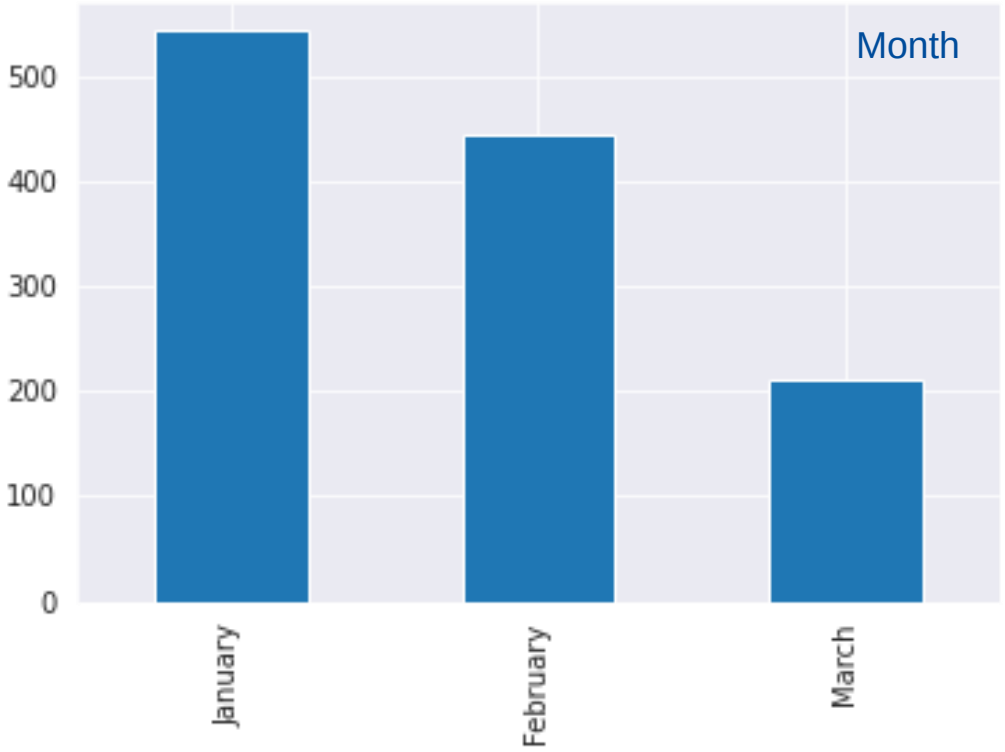


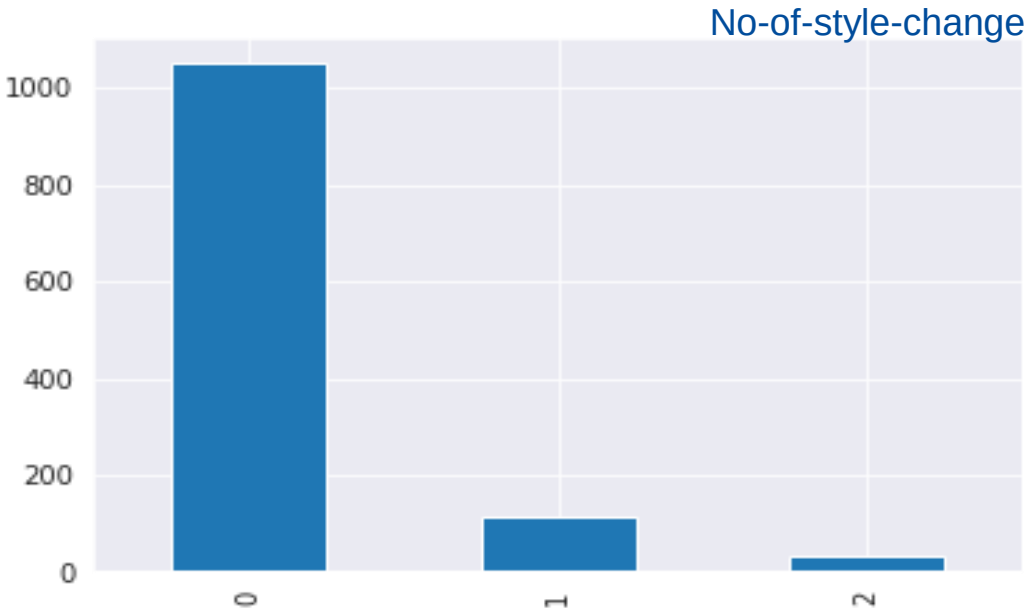
Department



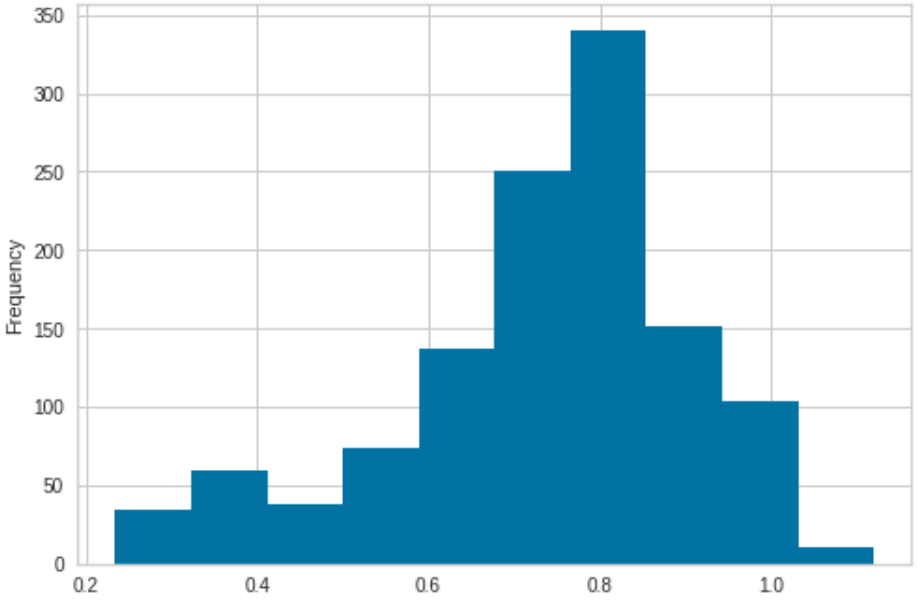
Team



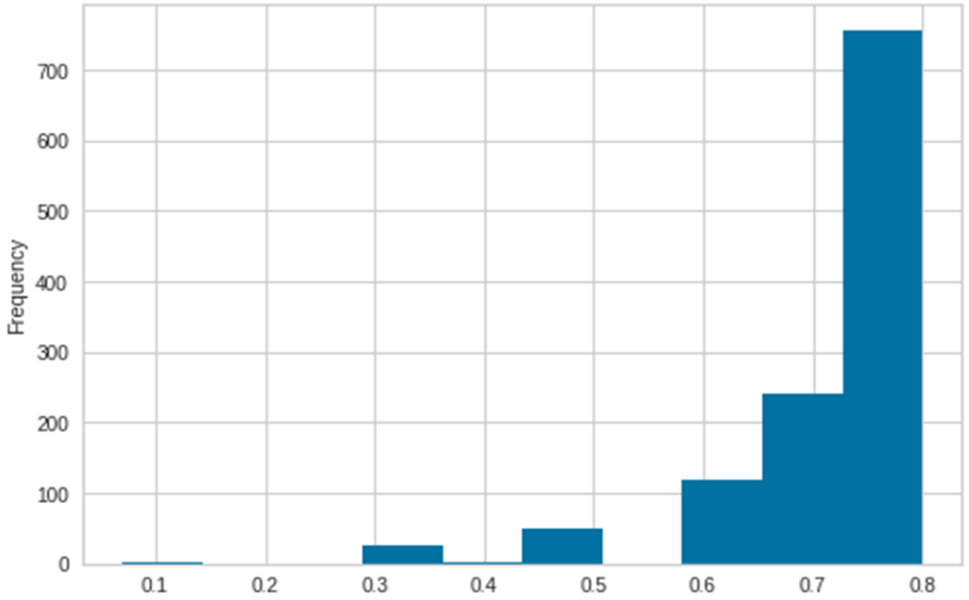

Month
500
400
300
200
100
0
January
February
March


No-of-style-change
1000
800
600
400
200
0
0
1
2

Numerical feature:



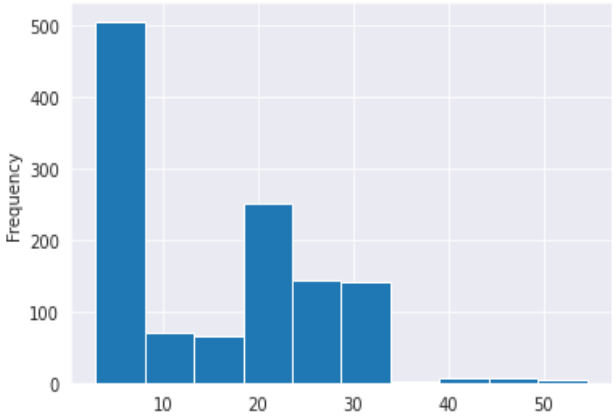
Actual productivity



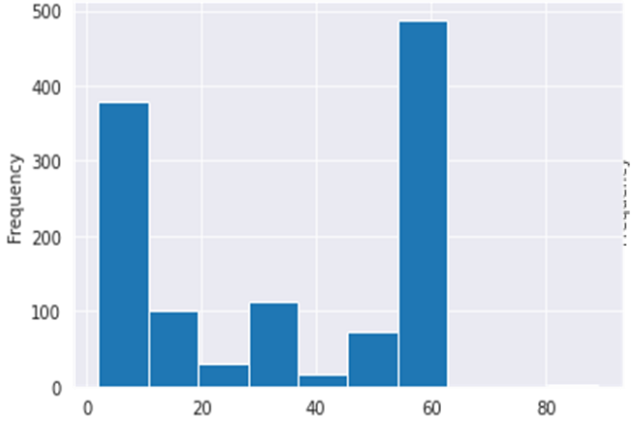
targeted_productivity

Numerical feature:

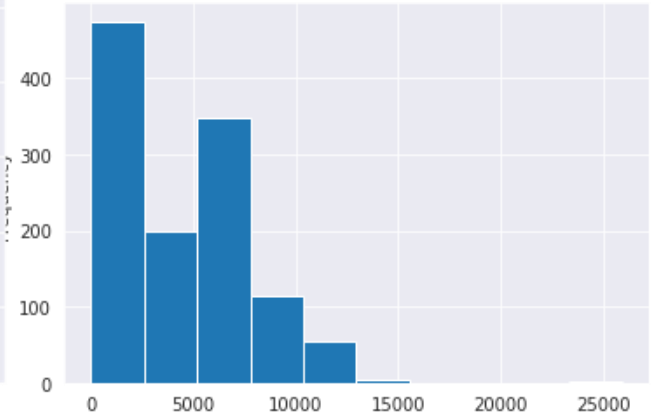
SMV

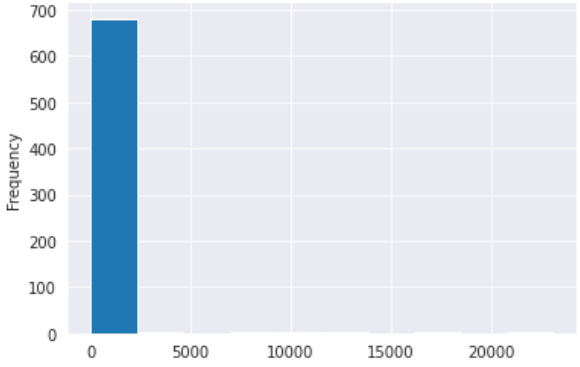


no_of_workers

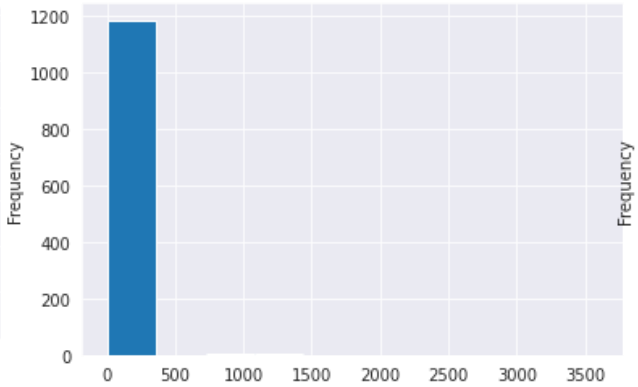


Over_time

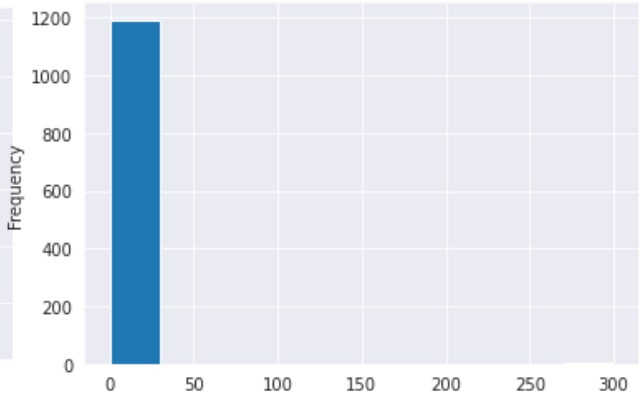




wip

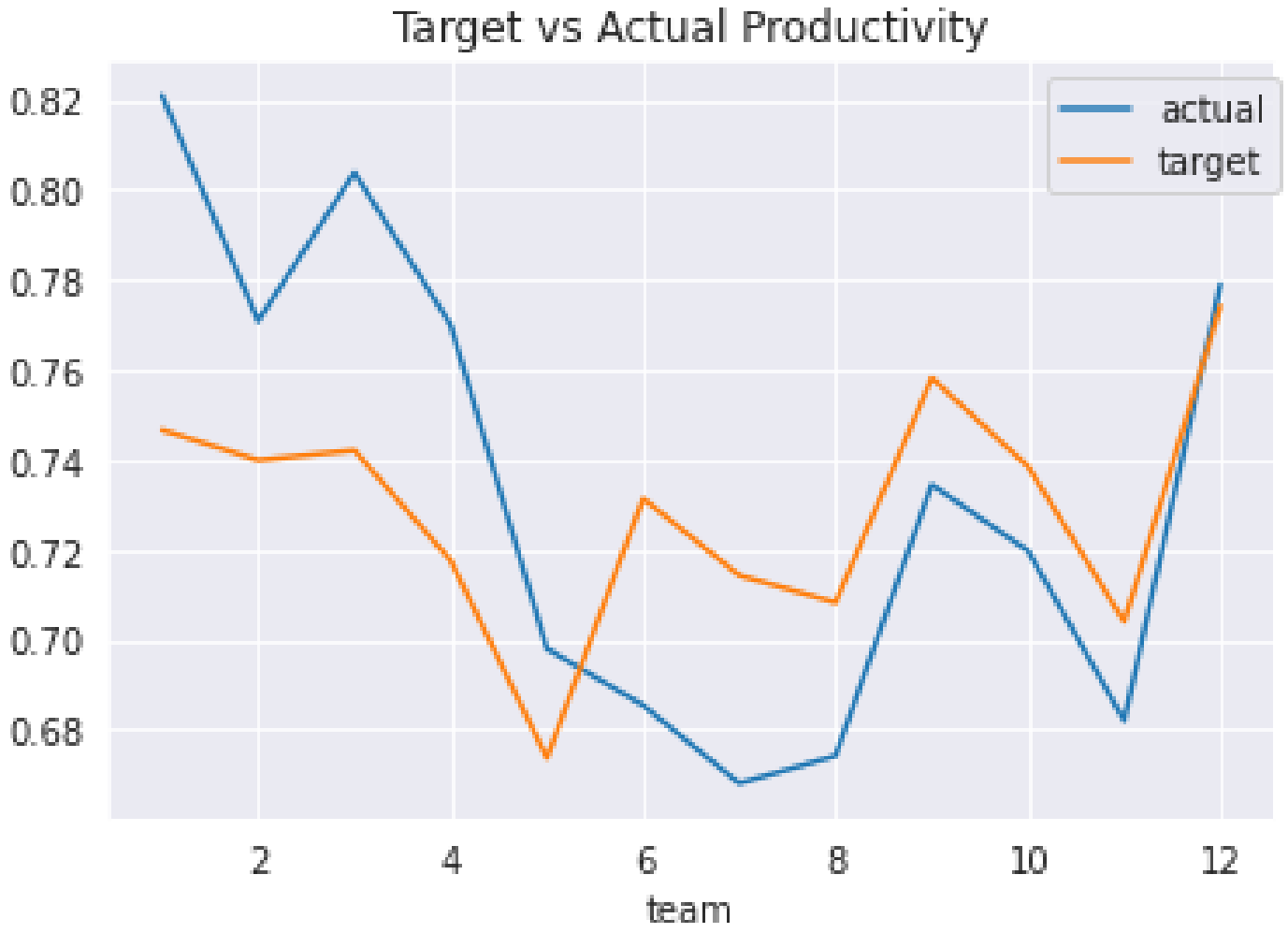


incentive



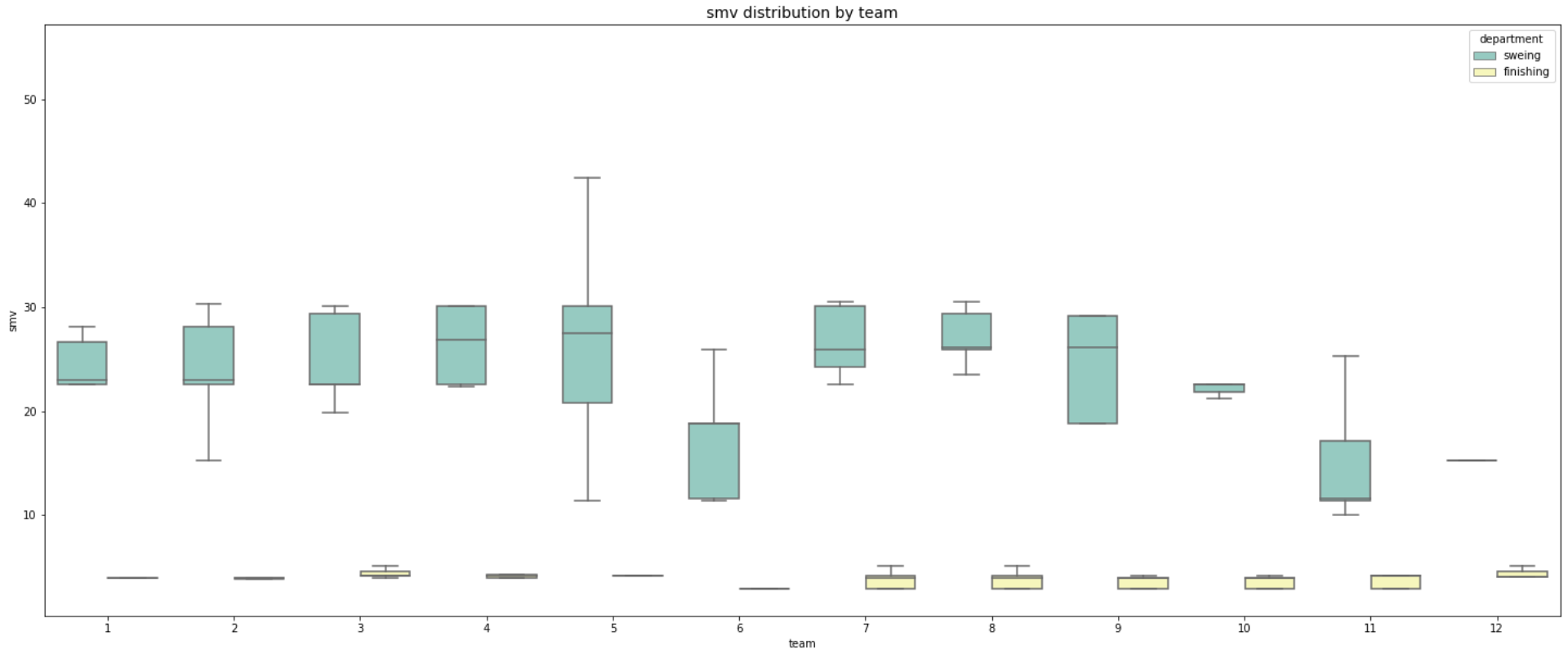
idle_men

Time series chart :



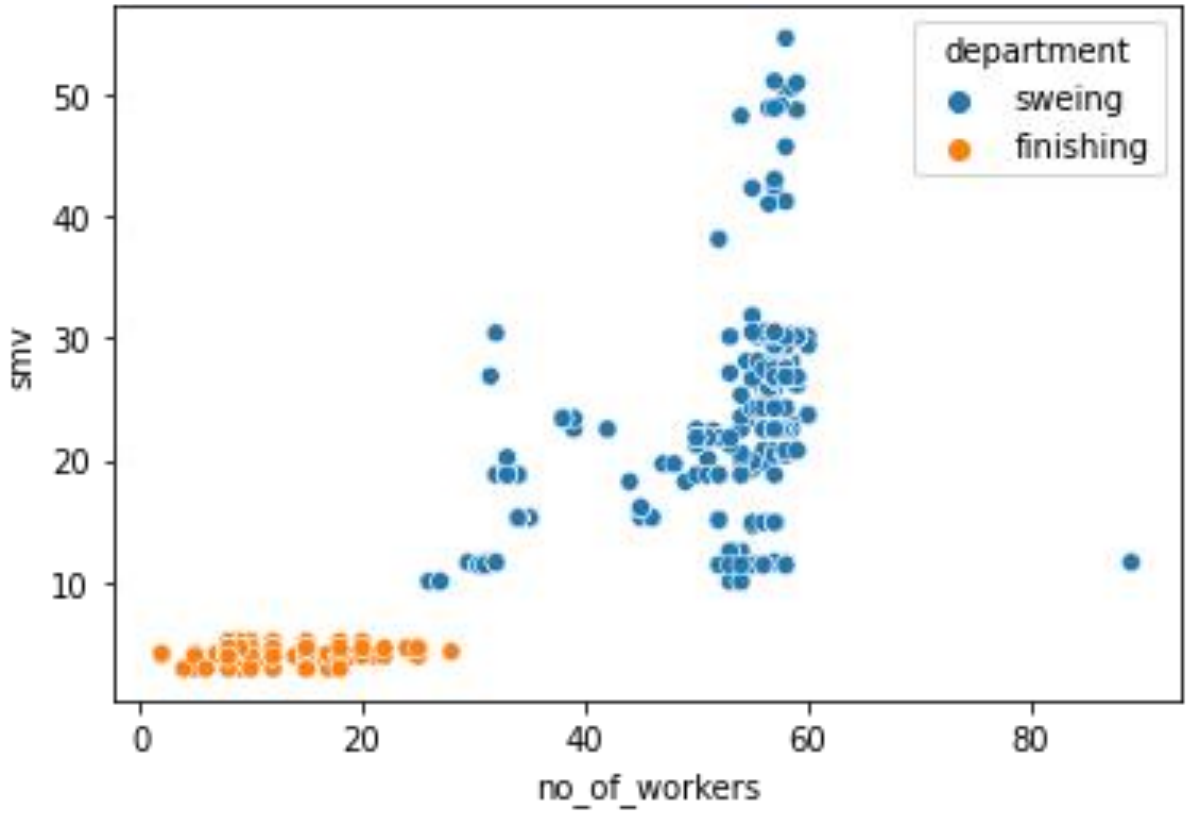
Actual prodactivity

Bar chart for understanding Numerical feature:



When we look into the smv boxplot on team basis with department separation, it can be clearly seen that while there are fluctuations between teams in the sewing department, the finishing department has almost evenly distributed smv values for each team.

Bar chart for understanding Numerical feature:



For the finishing department SMV does not change with no_of_workers

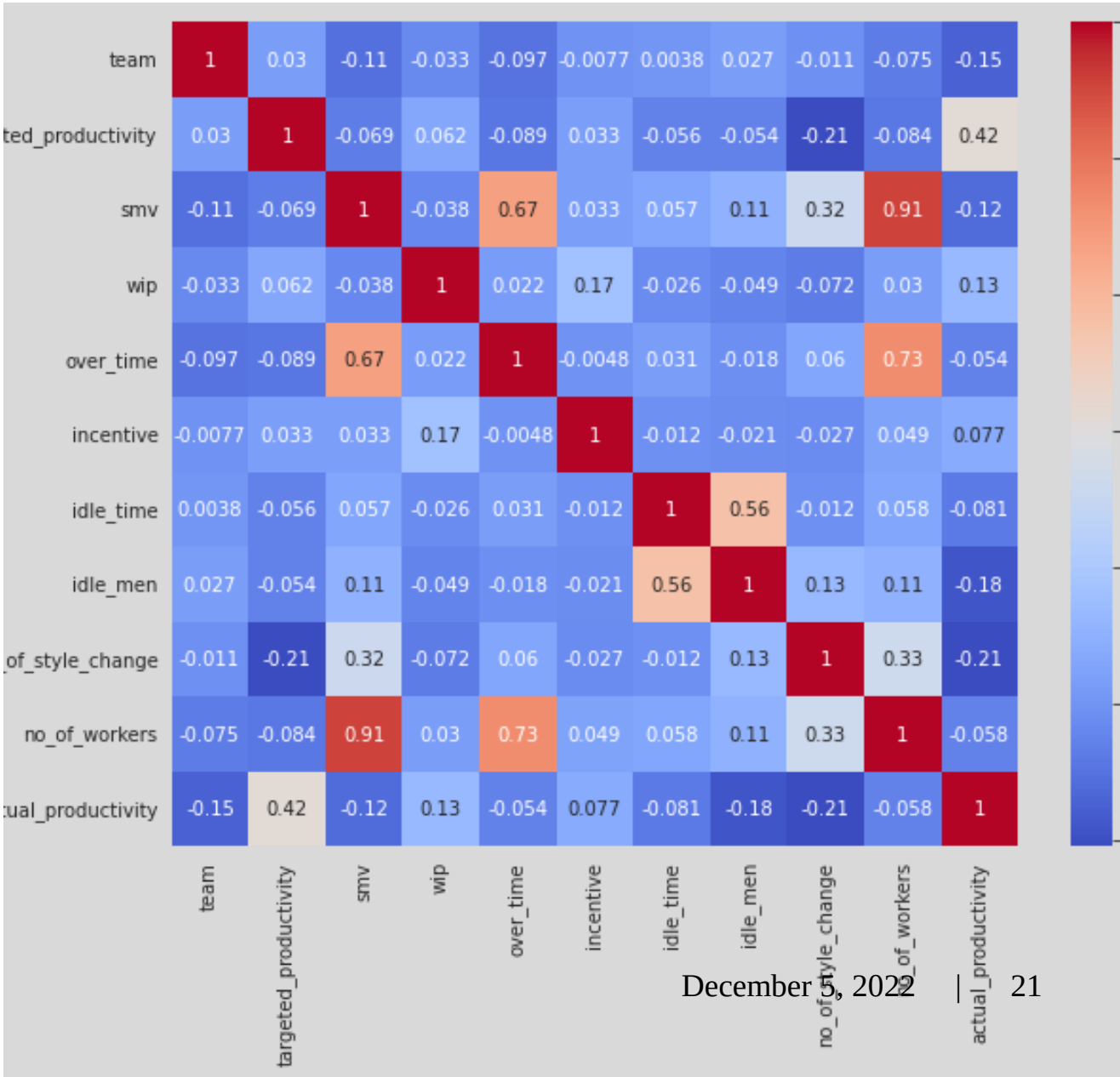
Correlation

Highest Positive Correlations:

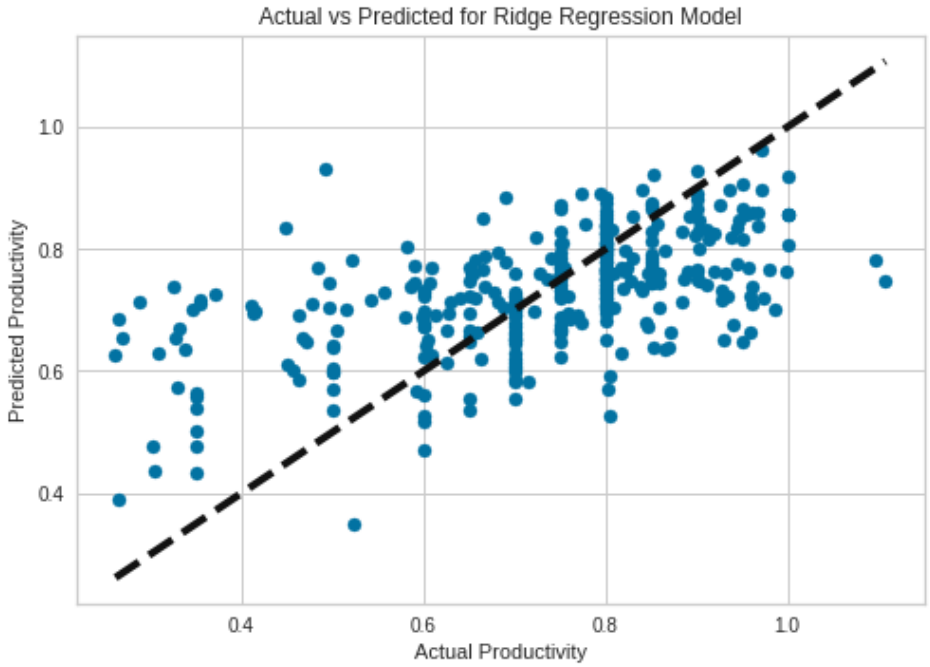
- No_of workers and smv (0.91)
- No_of workers and over_time (0.73)
- Over_time and smv (0.67)
- Idle_men and Idle_time (0.56)

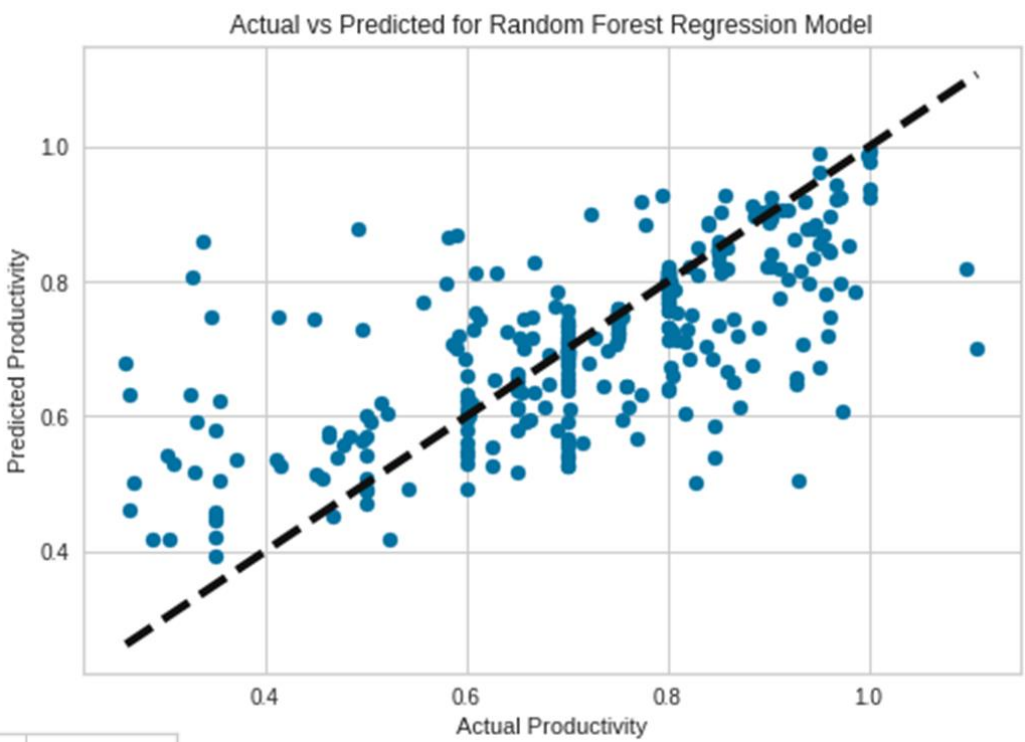
Positive Correlations:

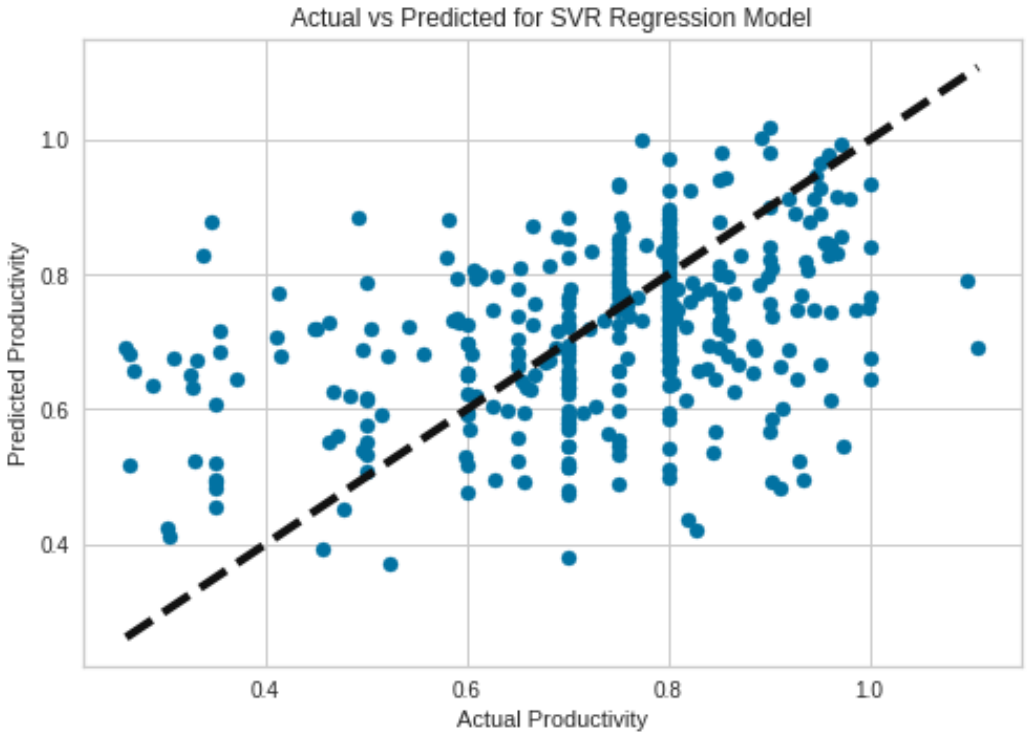
- No_of workers and no_of_style_change(0.33)
- No_of_style_changehas and smv (0.32)

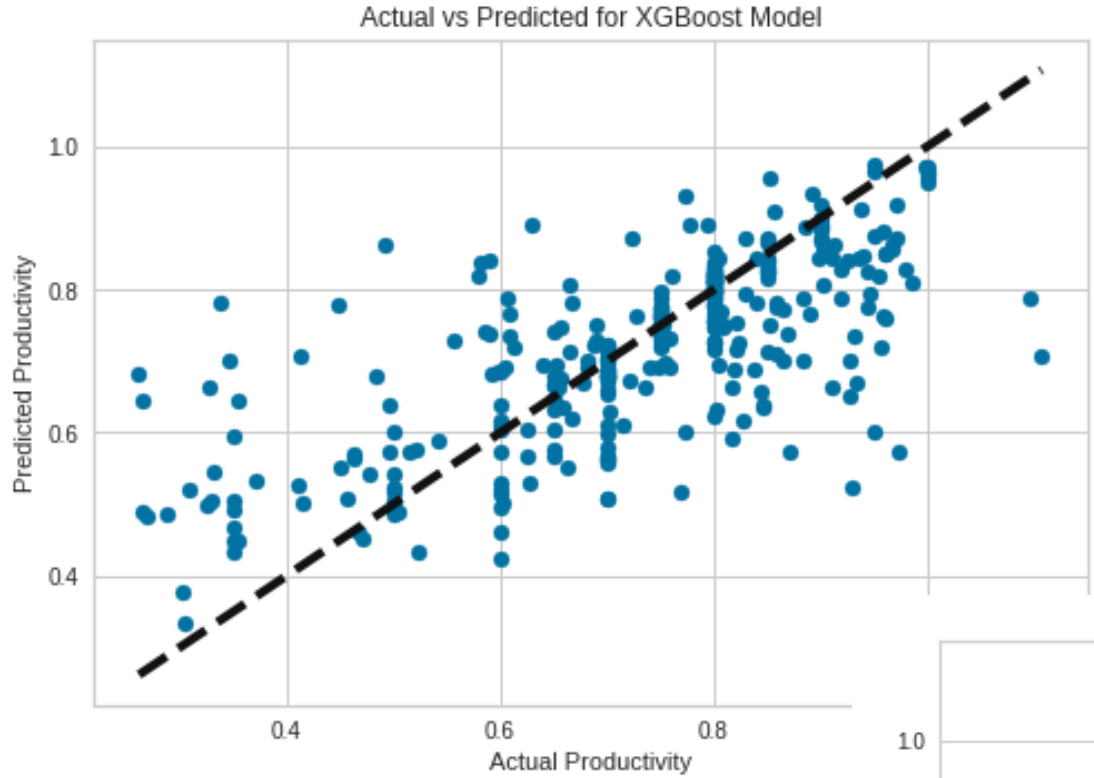


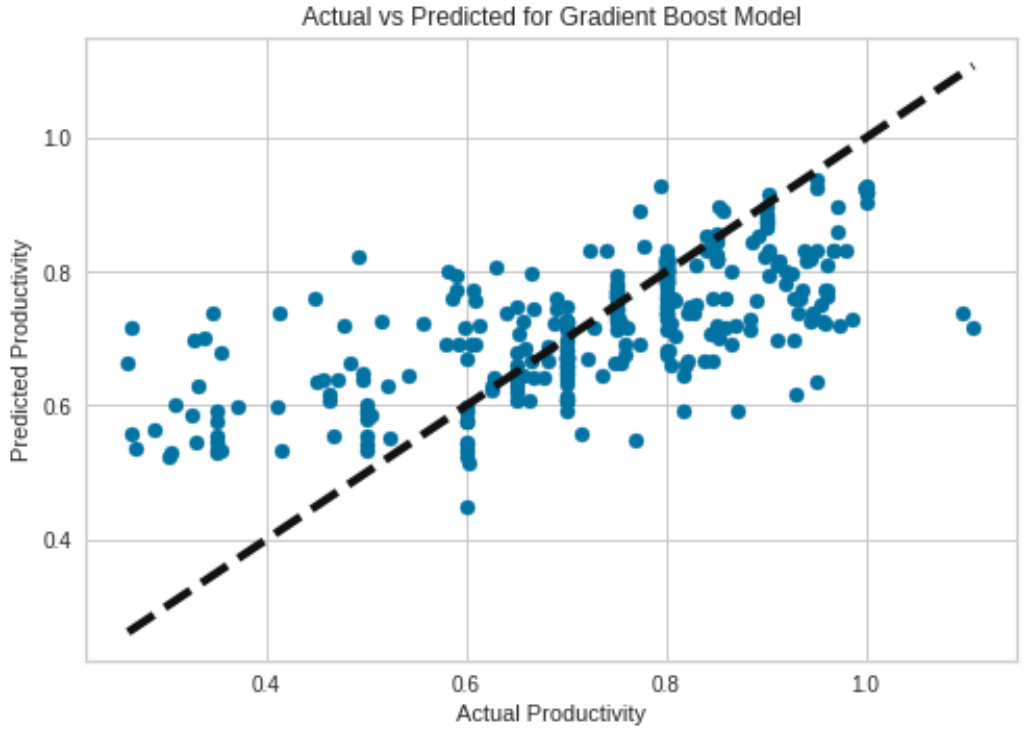

Actual vs Predicted for Decision Tree Regression Model
Predicted Productivity
Actual Productivity
0.2
0.4
0.6
0.8
1.0
0.4
0.6
0.8
1.0


Actual vs Predicted for Ridge Regression Model
Predicted Productivity
Actual Productivity
0.4
0.6
0.8
1.0
0.4
0.6
0.8
1.0


Actual vs Predicted for Random Forest Regression Model
Predicted Productivity
Actual Productivity
1.0
0.8
0.6
0.4
0.4
0.6
0.8
1.0


Actual vs Predicted for SVR Regression Model
Predicted Productivity
Actual Productivity
1.0
0.8
0.6
0.4
0.4
0.6
0.8
1.0


Actual vs Predicted for XGBoost Model
Predicted Productivity
Actual Productivity
1.0
0.8
0.6
0.4
0.4
0.6
0.8


Actual vs Predicted for Gradient Boost Model
Predicted Productivity
Actual Productivity
1.0
0.8
0.6
0.4
0.4
0.6
0.8
1.0

Result

- All the methods had good metrics for classification and prediction.
- Random forest, Xboost and Gradient Boost had the highest accuracy.
- targerted_productivity can have a strong positive impact on actual_productivity.
- Quarter is another factor influencing actual productivity, with Quarters 5, 1, and 2 being the most productive, respectively.
- Team can also affect actual productivity, with teams 1, 12, and 3 being the most productive, respectively.
- no_of_workers can also play a role in influencing actual_productivity
- Incentive is also another important factor that can be a meaningful predictor of actual_productivity.
- understand that date and time features have the least impact on productivity.

Ryerson University

	models	mae	mse	rmse	mape	R2
0	Linear Regression	0.104630	0.020869	0.144462	17.649078	0.245628
1	Lasso Regression	0.127489	0.027683	0.166383	22.537563	-0.000680
2	Ridge Regression	0.102079	0.018746	0.136915	17.483005	0.322390
3	Decision Tree Regression	0.097947	0.030112	0.173529	16.671211	-0.088481
4	Random Forest Regression	0.076509	0.014719	0.121320	13.286285	0.467959
5	Support Vector Regression	0.125738	0.026929	0.164101	20.415145	0.026575
6	KNeighbour Regression def	0.103148	0.022045	0.148477	18.535759	0.203116
7	KNeighbour Regression 3	0.100779	0.022295	0.149316	17.986969	0.194076
8	XGBoost	0.079848	0.014027	0.118434	13.393031	0.492969
9	Gradient Boost	0.086104	0.015304	0.123711	15.096265	0.446784

This research can be continued and advanced with a focus on imbalance and change to balance data. It would be applied with another Machin learning Algorithm, Natural neural Network and AI Technic.

Ryerson University

Thank you for your Time