

Garment Employee Productivity

Analysis and Visualisation

| Baldip Singh |

| Data Analytics and Intelligence |

| Quarter 2 2021 |

Contents

[Scenario 1](#_Toc86142393)

[Analyst: Baldip Singh 2](#_Toc86142394)

[The Data 2](#_Toc86142395)

[Data Preparation 3](#_Toc86142396)

[Data Types​ 3](#_Toc86142397)

[Incompleteness 4](#_Toc86142398)

[​Calculated Field 5](#_Toc86142399)

[The Analysis 5](#_Toc86142400)

[Model Assumptions 11](#_Toc86142401)

[Dashboard 12](#_Toc86142402)

[Sewing department 12](#_Toc86142403)

[Finishing department 13](#_Toc86142404)

[Conclusion 14](#_Toc86142405)

# Scenario

B & Y Apparel Corporation, is a medium sized garment company, primarily dealing with formal garments. B & Y Apparel Corporation have approached our duo of business analyst with questions which will be individually discuss.​

B & Y Apparel Corporation want to reconfigure the employee management system to maximise worker productivity.

Baldip is assigned to investigate the following:​

* Which attribute/s should be addressed to improve employee productivity?
* Propose a regression model based on one of the available variables which best predicts actual productivity. ​

B & Y Apparel Corporation has also provided a dataset,” garments\_worker\_productivity.csv”.

Primary tools being utilized by the analysts are Alteryx and MS Tableau to conduct this analysis.

# Analyst: Baldip Singh

## The Data

|  |  |  |  |
| --- | --- | --- | --- |
| **Time​** | **Job Details​** | **Team Details** **​** | **Performance​** |
| Date​ | No. Of style changes​ | Team number ​ | Targeted productivity​ |
| Quarter​ | Standard minute value (SMV)​ | Department ​ | Actual productivity​ |
| Day​ | Incentive​ | Work in progress ​ | *Net productivity*​ |
| ​ | Idle time​ | No of workers ​ | ​ |
| ​ | Idle men​ | Overtime​ | ​ |

The data supplied consists of 15 different attributes for each record. Each row represents a job that has taken place related to the production of garments. These 15 attributes can be categorized as shown in the above table. Time, Job Details, Team Details and Performance. From the scenario we can note that the Performance type attribute “Actual Productivity” will be the focus of this analysis. This is the productivity score for the team.

It is to note the “Net productivity” attribute was not supplied but calculated, discussed in the next section of the report.

The data supplied consists of 15 different attributes, information about the variables supplied is as follows:

Text

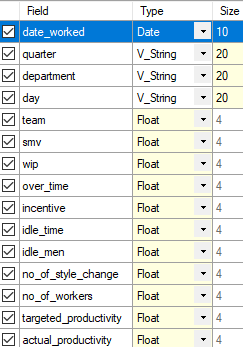
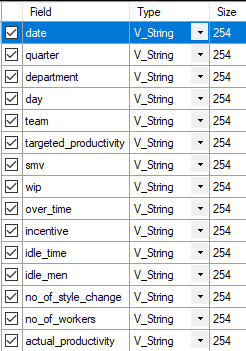
Description automatically generated

## Data Preparation

### Data Types​

Initially all data types were V\_String Size 254. I went ahead and changed the data types of the date attribute to Alteryx’s Date data type. Once we change the date data type from V\_String 254 to Date, we open the way for more avenues of analysis. If at later stages, the data set required to be filtered by date this would not be as easy if using the data type V\_String. It is also to note the Date data type is ten characters and 7 bytes in size vs the initial being 254 bytes which was above requirements, changing this down 20 where required and changing data types to float and date respectively has decreased the overall footprint of the data. However, the primary benefit of doing this is the fact the various Alteryx functions require the data to be of a specific type – therefore getting this step correct is critical as this is the foundation for our analysis.

Figure - Before and After Data Type changes.



Where necessary data types and sizes have been altered shown above.

### Incompleteness

When discussing completeness, it is important to identify fields where values are not present i.e., *Null* values. However, in some cases, due to the way the data has been collected, data may seem incomplete even with field entries, can be in the form of 0’s.

#### Null Values

Upon analyising the null values through the summary tool within the Alteryx workflow, the below was shown. Of all the attributes only,” wip – work in progress*”* possessed Null values.



Of the 1198 records, 506 entries went without entries. This was deemed too excessive as the data set is already quite small. Before eliminating this attribute, an important relationship was discovered by chance. Only jobs that got associated with department Sewing had entries for “wip – work in progress”. Where the job was associated with department “Finishing” there were null entries. Without inquiring further, it seems as though this is an issue in the way the data is being collected in that the Finishing departments jobs must be complete by nature the wip must always be nil. In that case this further solidifies the need to drop this attribute as it is only relevant to one department. This can further be investigated by modelling actual productivity by department where Sewing department will have “wip” to pick up for any significance.

#### Zero Entries

When analysing the attributes number of style changes, idle men, and idle time it was to be noted the data was skewed more towards zero. Majority of the row entries were zero which immediately caught my eye and had to assessed.

Diving deeper into the number of style changes revealed only 147 rows with entries other than 0. Only other values that appeared were 1 or 2. This meant majority of the time the jobs did not undergo style changes and when they did it would only be either 1 or 2. 113 times the number of styles changes was 1 vs 34 times 2.

Idle time and Idle men, when checking for patterns it was revealed that the fields idle time and idle men went hand in hand. Only when the job was interrupted could we have an Idle time reading and further only when we have idle time can we have idle men. This was proven in the data in the 18 rows.

What this showed was that only 18/1198 jobs were interrupted. 147/1198 job required a style change. Although the range of values for these attributes where not great decision was made to keep them as these rows could still provide some signal for this model. If the model deemed, they were not significant they will then be dropped.

### ​Calculated Field

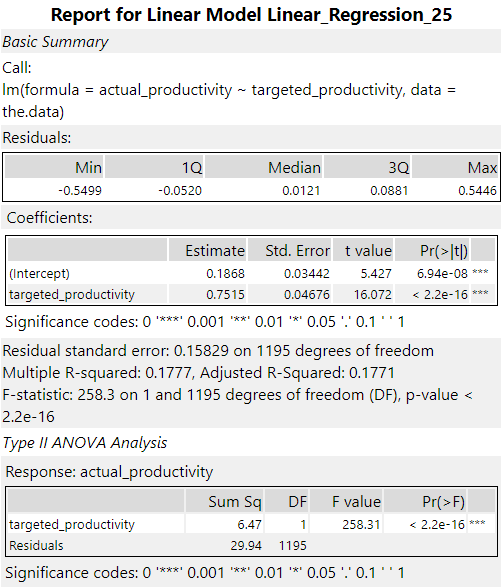
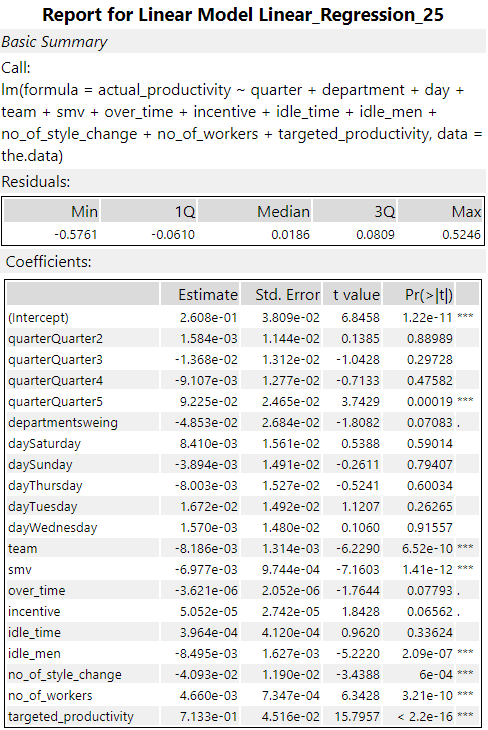
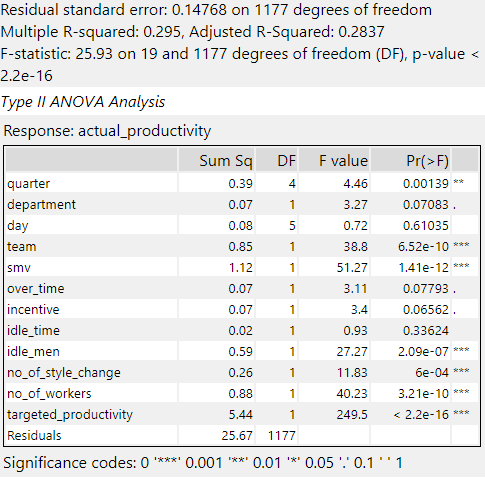
After Initial observations of the data, it was quite clear the attributes target productivity and actual productivity were not enough alone to truly show the relationship and performance of the team. Actual productivity can be deceptive when dealing with higher values as a team could still be under performing with high productivity. It was then decided net productivity attribute would be added and be a calculated field of the Actual subtracting the Target productivity. This was done so that when a team was under achieving a target the value would be a negative value in line with natural interpretation *negative* being *“bad”*. This gave a more powerful indicator of how the team was performing.

## The Analysis

Diagram

Description automatically generated

As B & Y Apparel required a linear regression model each individual attribute was removed using the linear regression function on Alteryx. Removing each attribute one after another it was noted that target productivity had the lowest pvalue – most significant.



All variables regression model. Initial.

Single attribute model. Final.

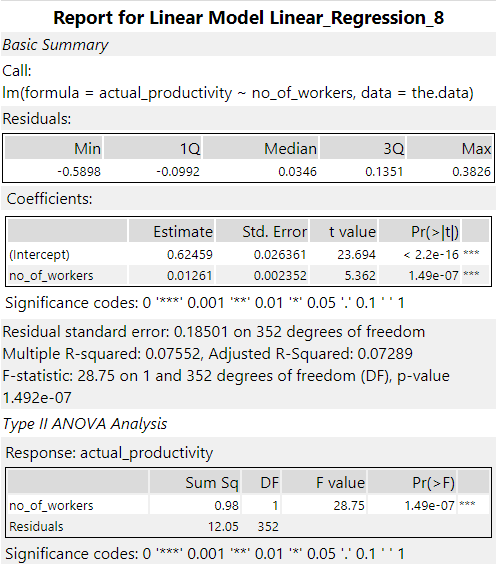
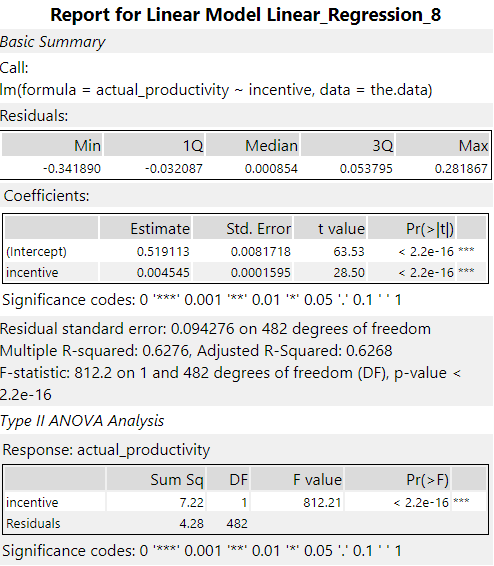
As the requirements were to build a model on one attribute – target productivity was selected as this produced the best linear regression model in terms of having an extremely low p-value and the highest r-squared out of the attributes.

Upon analyzing the average number of workers within each team it was noted that teams were made up of 2-70 workers. Plotting the number of workers per instance of the team number showed two distinct grouping towards either end of the scale. For example, for team 4:

Chart

Description automatically generated

Connecting the dots, the data was of two different departments, looking at the average number of people within each team based on department revealed clear distinctions. The sewing department on average had almost 50 more workers on the team. It was Then noted that perhaps it would be best to create linear regression models for different departments i.e., splitting the data up on department.



The above left displays the Sewing departments final linear regression model. Showing an RSquared value of .626 it plays in stark comparison to the final linear regression model for the Finishing department with a RSquared value of .073. The model for the Sewing department also shows a much smaller p-value.

To create the test and training data sets alongside the model file to apply the follow Alteryx workflow was implemented – split chosen was 70:30 through the create sample’s function, with 70 being the training sample.

A picture containing timeline

Description automatically generated

To arrive at the .csv files shown above the original workflow had to be altered to the one shown below.

Diagram

Description automatically generated

In comparison to the original model containing both departments’ data, the Sewing department model has been better at predicting the actual productivity within the sewing department.

Diagram

Description automatically generated

Using the score tool, the model was applied to the Sewing Test data set creating a score field which showed the predicted value. On average within the test data set the Sewing model managed to predict the actual productivity with an average error of -.00235. This means the model consistently predicted just below the actual value, which is to be noted when deploying the final model.

In comparison to the model associated to the finishing department, it showed a much lower r squared value compared to the model created from using both department data sets. Therefore, it is best to continue ahead with the Sewing department and re visiting the Finishing department at a later stage.

## Model Assumptions

Chart, diagram

Description automatically generated

**Residual vs Fitted:** ​  
No real non-linear relationships visible.​  
​**Scale-Location: ​**  
Residuals are generally spread evenly; line is horizontal enough.​  
​**Residuals vs Leverage: ​**  
Observations are independent of each other, although some outliers are present the model is good enough.​  
​**Normal Quantile-Quantile plot: ​**  
Values follow the dashed line for the most part​

​Above plots for the sewing department data set.

## Dashboard

### Sewing department

Chart

Description automatically generated

As determined by the model Incentive has deemed to be a powerful influencer on the actual productivity and sub sequent the net productivity. After an incentive of 55 we are more likely to achieve over our targets (top left). It is also to note within the department Wednesday, Thursday and Sunday are days the teams are more likely to achieve less than targeted. Final note of interest is that on average teams 1 to 4 are achieving targets whereas teams 5 to 12 are under achieving. Looking back at incentives teams that are achieving targets are shown to have a higher incentive, teams that achieved targets had an average incentive of 56.2 vs 38.6 of those that did not meet targets.

### Finishing department

Chart

Description automatically generated

With a quick overview of the Finishing department, we can see the strength of the relationship between the number of workers and net productivity (top left). As the number of workers on the team is increasing the higher the likelihood of achieving target productivity. On average teams had 13 workers (bottom left), alongside being productive every day (right) overall this department achieves targets and therefore reworking the Sewing department is a much better first step.

## Conclusion

According to the data supplied by B & Y Apparel Corporation, it is in the best interest of the company to begin reconfiguring the Sewing department first as there are variables that show high likelihood of reaping immediate benefits. As things stand on average the Sewing department is falling short of the targeted productivity whereas the Finishing department is achieving its set targets. Therefore, the sewing department should be made priority in increasing actual team productivity. Experimenting with the incentives on offer for employees would be a good start. Where possible it is in the company’s best interest to keep incentives above 60 to ensure targets are met as this is the number where the targets are more likely to be met.

In comparison to the Finishing department the signal is not as strong as the one within the Sewing department therefore it is my suggestion that more time be invested into the working of the Finishing department, so a better understanding of the necessary data required is obtained. The number of workers per team is looking promising as things stand, teams with over 10 works on average are achieving productivity targets.

Initially it was requested one model to be made however discovering patterns in the data suggests models be built based on each department. Sewing department end up having 692 records vs Finishing 516. As the model for Finishing department does not seem fit for use only the final model for the Sewing department is shown below. It is to note this model predicted just below the actual prediction on average by 0.002, this will need to be considered during the interpretation of this model.

0.004545 \* incentive + 0.519113 = actual productivity

R Squared = 0.6268

P-Value = 2.2e-16

Given time to create THE best model for any given data set, multiple different modeling techniques should be utilized and compared against each other. For this reason, I believe to further refine these results other models should be experimented with.

To get a stronger model it would be best advised to incorporate more than one variable to use as predictors for the target variable actual productivity. My fellow analyst will discuss this further.