

Assignment 2 -2230035-Shubhi

by Shubhi Pushkarna

Submission date: 05-May-2023 10:10AM (UTC+0100)

Submission ID: 204863954

File name: 45643_Shubhi_Pushkarna_Assignment_2_-2230035-Shubhi_493224_852038523.pdf (1.56M)

Word count: 4491

Character count: 25164

DATA VISUALIZATION AND MINING

ASSIGNMENT 2 – (ACCA 7015)

Submitted by: Shubhi Pushkarna

Student ID- 2230035

Submitted to: Terry Walcott

Course: MSc Data Science and Analytics (2023)

TABLE OF CONTENTS

1. Introduction to Dataset	4
2. Problem Background.....	5
3. Dataset Description.....	6
4. Data pre-processing	9
5. Analysis of the dataset.....	17
6. Report and Evaluation	26
7. Reflection.....	29
8. The final say.....	30

TABLE OF FIGURES

Figure 1. Table 1-Attributes Description.....	7
Figure 2. Table 2-Attributes distribution.....	8
Figure 3. Screenshot of dataset in excel file	10
Figure 4. Screenshot of data visualization in Tableau.....	11
Figure 5. Screenshot of data interpreter being used for data cleaning.....	12
Figure 6. Screenshot of detecting missing/null values in the WIP column.....	13
Figure 7. Missingness gap graph.....	14
Figure 8. Comparison between actual and target productivity of the two departments.....	14
Figure 9. Average actual productivity in relation to quarters.....	16
Figure 10. Linear regression analysis on actual and targeted productivity.....	18
Figure 11. Linear Regression on incentive with actual productivity graph.....	19
Figure 12. Bar Graph between Incentive received per team.....	20
Figure 13. Scatterplot visualization between actual productivity and number of workers...	21
Figure 14. Screenshot of Dashboard having 4 different Graphs.....	23
Figure 15. Dashboard having graphs on Incentives, actual productivity and team.....	25

1. INTRODUCTION TO THE DATASET

In many developing nations, the garment sector contributes significantly to industrial output, employment, and commerce. As it generates a significant amount of foreign currency, it is essential to the development of a nation's economy. For instance, Bangladesh, which is still developing, is the second-largest exporter of clothing worldwide [1]. The Ready-Made Garments (RMG) industry contributed \$30.61 billion to Bangladesh's export profits in 2017, accounting for 81% of all export revenues and representing about 14.07% of the country's GDP, according to figures recently provided by the Export Promotion Bureau (EPB) [2]. The garment business is one of the oldest and most labor-intensive sectors in the world, regardless of economic considerations.

Worldwide, the textile, apparel, and footwear industries employ over 75 million people directly [3], therefore the competence and productivity of the workforce have a significant impact on the output of the clothing industry. Because of this, business owners in the industry need to have a strategic leadership vision in order to boost output and stay up with the sector's rapid changes [4]. According to Gunawan et al. [5], improving management skills in financial monitoring and activity evaluation at Indonesian Garment Medium Scale Manufacturers (IGMS) is essential to maintaining Indonesia's economy.

Numerous studies have been conducted and published in the past several decades in the area of the apparel business that deals with increasing output, enhancing quality, minimizing manufacturing costs, and simplifying the production process. There are, however, a few research in this field that are also interested in the use of data mining and machine learning techniques in

this industry. It makes intuitive sense that data mining and machine learning are methods used to uncover hidden knowledge that is useful [6]. A significant quantity of data is generated by the labour-intensive everyday operations and industrial processes in the clothing sector. Advanced computational approaches may be applied to this data to unearth hidden insights, improve production efficiency, and address a number of industrial and management issues.

2. PROBLEM BACKGROUND

A thorough garment production pipeline includes a number of sequential phases, including designing, sample confirmation, sourcing, and merchandising, lay planning, marker planning, spreading and cutting, sewing, washing, finishing, and packing. A good garment manufacturing always follows a line plan that specifies when production will begin, how many pieces are anticipated, and when the order has to be finished in addition to managing this pipeline. These sequential procedures must be carried out effectively and promptly if a production is to be completed in a set amount of time. To reach the production goals, the affiliated industrial engineers set a targeted productivity value against each working team in the manufacturing process [3]. However, it frequently happens that working teams' actual productivity falls short of their desired productivity due to a number of factors, including the number of team members, the volume of work in progress, over time, financial incentives, and so on. The production's ability to be finished on schedule is frequently hampered by this discrepancy between the working teams' targeted and actual productivity. A firm that makes clothes suffers a significant financial loss as a result of this issue. Therefore, to minimize the difference between the targeted and actual productivity, an industrial engineer should set the targeted productivity in a precise and data-driven manner when creating a line plan.

They can absolutely reduce financial loss, maximize profit, and improve production efficiency if they can forecast the real productivity of the working teams and establish the objective in accordance with the forecast.

This study uses TABLEAU (a data visualization tool) to tackle the issue using sophisticated computational data-driven technologies. I have suggested a regression model in this study forecast the productivity of garment workers based on historical data.

3. DATASET DESCRIPTION

The dataset used in this study was sourced from the Industrial Engineering (IE) division of a reputable company's Bangladeshi garment manufacturing facility. The gathered information includes the company's sewing and finishing department's production statistics for the three months from January to March 2015. There are 1197 occurrences in the collection, and 15 attributes. The dataset's data dictionary is shown in Table 1.

Table 2, provides a systematic differentiation of the attributes into categorical and numerical data types. This classification is essential in comprehending the nature of the data and in gaining insights into the underlying patterns and trends that may be present.

The combined nominal and numerical properties that make up this dataset have been used as dimensions and measures, respectively. Using Tableau, missing and null values have been removed throughout the data cleaning process.

Sr.No	Attribute	Description
1.	Date	Date in MM-DD-YYYY
2.	Department	Associated department with the instance
3.	Day	Day of the week
4.	Quarter	A portion of the month and the month is divided into four quarters.
5.	Team no	Associated team number with the instance
6.	No of workers	Number of workers in each team
7.	No.of style change	Shows 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.	Over time	Displays the amount of overtime by each team in minutes
12.	Incentive	Displays 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.

Figure 1. Table 1-Attributes Description

Sr.No	Attributes	Type	Category
1.	Date	Categorical	Dimensions
2.	Quarter	Categorical	Dimensions
3.	Department	Categorical	Dimensions
4.	Day	Categorical	Dimensions
5.	Team	Categorical	Dimensions
6.	Targeted Productivity	Numerical	Measures
7.	SMV	Numerical	Measures
8.	WIP	Numerical	Measures
9.	Overtime	Numerical	Measures
10.	Incentive	Numerical	Measures
11.	Idle time	Numerical	Measures
12.	Idle men	Numerical	Measures
13.	No of style change	Numerical	Measures
14.	No of workers	Numerical	Measures
15.	Actual productivity	Numerical	Measures

Figure 2. Table 2 -Attributes Distribution

TOOL USED FOR DATA VISUALIZATION

TABLEAU has been selected as the data visualisation tool for this dataset. As seen below, the dataset includes an Excel file. Excel makes it difficult to analyse and forecast output based on record-keeping. Tableau will offer rapid interactive visualisation based on the given dataset. Compared to Excel, Tableau's user interface is more straightforward and intuitive, making it simpler for non-technical people to interact with data. Without the requirement for intricate data modelling, tableau may combine data from several sources, including databases, spreadsheets, and cloud-based platforms. Excel is also capable of combining data from several sources, although it necessitates more human work. The fundamental properties that TABLEAU offers below seem to be the greatest fit for the dataset used in this report.

- ✓ Responsive dashboards for comparisons
- ✓ Clear understanding of graphs
- ✓ Provides quick insights to determine results better

4. DATA PRE-PROCESSING AND CLEANING

Pre-processing the dataset correctly is the most important step in utilizing machine learning to solve a real-world problem. It is one of the most crucial components in getting a promising outcome from the model. Data might be loud, distorted, lacking, uneven, and incomplete in the actual world. To ensure that a machine learning model can accurately understand the underlying patterns in the data, the data must be prepared for the model before being fitted into it. According on the format and structure of the data, the data pre-processing step of a machine learning pipeline may comprise a number of activities including data cleaning, feature encoding, instance

selection, normalization, transformation, feature extraction, and selection. Using Tableau, missing and null values have been removed throughout the data cleaning process.

Preprocessing data increases accuracy and dependability. Pre-processing data may increase the correctness and quality of a dataset, making it more dependable by removing missing or inconsistent data values brought on by human or machine mistake.

DATA PREPROCESSING IN TABLEAU

In Tableau, data pre-processing entails a number of steps, such as cleaning, transforming, and getting the data ready for analysis. Below are a few standard procedures followed for pre-processing data in Tableau.

STEP 1. Import the data into Tableau by connecting to the excel file. Below is the comparison of how data is presented in an excel file and Tableau.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	date	quarter	depart	day	team	targeted_smv	wip	over_time	incentive	idle_time	idle_men	no_of_sty	no_of_wo	actual_productivity							
2	1/1/2015	Quarter1	sweing	Thursday	8	0.8	26.16	1108	7080	98	0	0	0	59	0.940725						
3	1/1/2015	Quarter1	finishing	Thursday	1	0.75	3.94		960	0	0	0	0	8	0.8865						
4	1/1/2015	Quarter1	sweing	Thursday	11	0.8	11.41	968	3660	50	0	0	0	30.5	0.80057						
5	1/1/2015	Quarter1	sweing	Thursday	12	0.8	11.41	968	3660	50	0	0	0	30.5	0.80057						
6	1/1/2015	Quarter1	sweing	Thursday	6	0.8	25.9	1170	1920	50	0	0	0	56	0.800382						
7	1/1/2015	Quarter1	sweing	Thursday	7	0.8	25.9	984	6720	38	0	0	0	56	0.800125						
8	1/1/2015	Quarter1	finishing	Thursday	2	0.75	3.94		960	0	0	0	0	8	0.755167						
9	1/1/2015	Quarter1	sweing	Thursday	3	0.75	28.08	795	6900	45	0	0	0	57.5	0.753683						
10	1/1/2015	Quarter1	sweing	Thursday	2	0.75	19.87	733	6000	34	0	0	0	55	0.753098						
11	1/1/2015	Quarter1	sweing	Thursday	1	0.75	28.08	681	6900	45	0	0	0	57.5	0.750428						
12	1/1/2015	Quarter1	sweing	Thursday	9	0.7	28.08	872	6900	44	0	0	0	57.5	0.721127						
13	1/1/2015	Quarter1	sweing	Thursday	10	0.75	19.31	578	6480	45	0	0	0	54	0.712205						
14	1/1/2015	Quarter1	sweing	Thursday	5	0.8	11.41	668	3660	50	0	0	0	30.5	0.707046						
15	1/1/2015	Quarter1	finishing	Thursday	10	0.65	3.94		960	0	0	0	0	8	0.705917						
16	1/1/2015	Quarter1	finishing	Thursday	8	0.75	2.9		960	0	0	0	0	8	0.676667						
17	1/1/2015	Quarter1	finishing	Thursday	4	0.75	3.94		2160	0	0	0	0	18	0.593056						
18	1/1/2015	Quarter1	finishing	Thursday	7	0.8	2.9		960	0	0	0	0	8	0.540729						
19	1/1/2015	Quarter1	sweing	Thursday	4	0.65	23.69	861	7200	0	0	0	0	60	0.52118						
20	1/1/2015	Quarter1	finishing	Thursday	11	0.7	4.15		1440	0	0	0	0	12	0.436326						
21	1/3/2015	Quarter1	finishing	Saturday	4	0.8	4.15		6600	0	0	0	0	20	0.988025						
22	1/3/2015	Quarter1	finishing	Saturday	11	0.75	2.9		5640	0	0	0	0	17	0.98788						
23	1/12/2015	Quarter1	finishing	Saturday	9	0.8	4.15		960	0	0	0	0	8	0.956373						

Figure 3. Screenshot of dataset in excel file

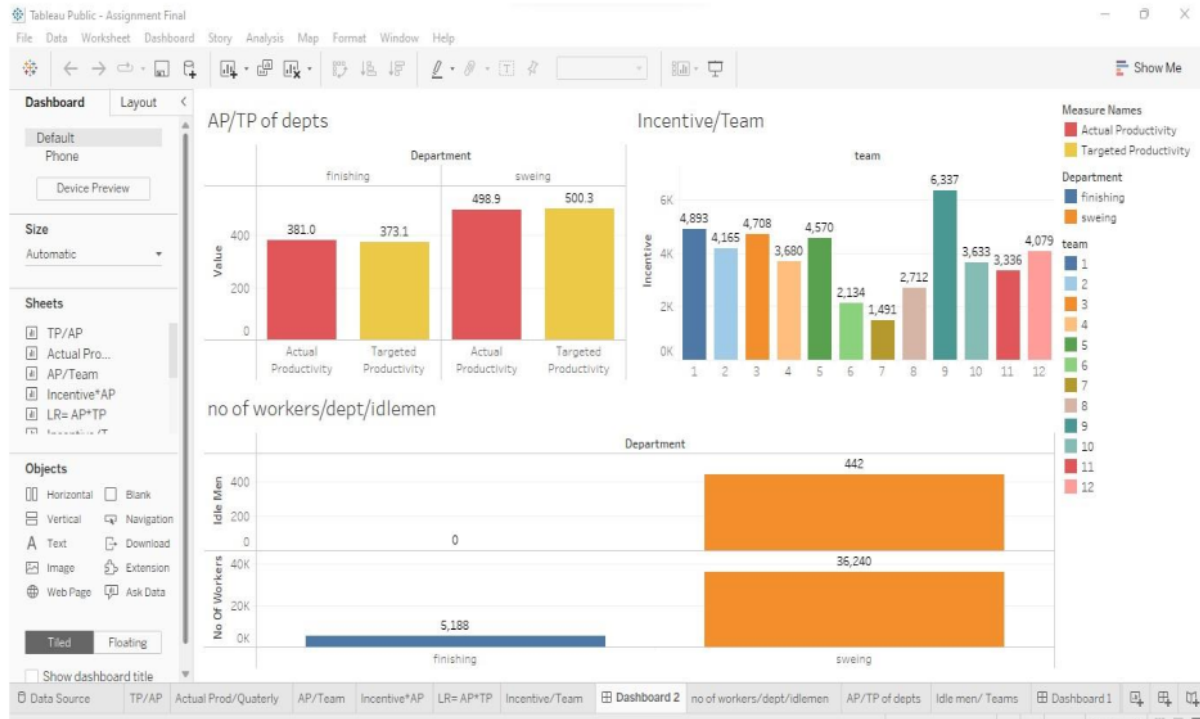


Figure 4. Screenshot of data visualization in Tableau

STEP 2. Cleaning up data by eliminating any duplicates, fixing any mistakes, and adding any values that are missing. To assist with this stage, Tableau's data cleansing tools, such as the Data Interpreter has been used.

We shall delete the 'date' attribute as it is unlikely to be useful for our study.

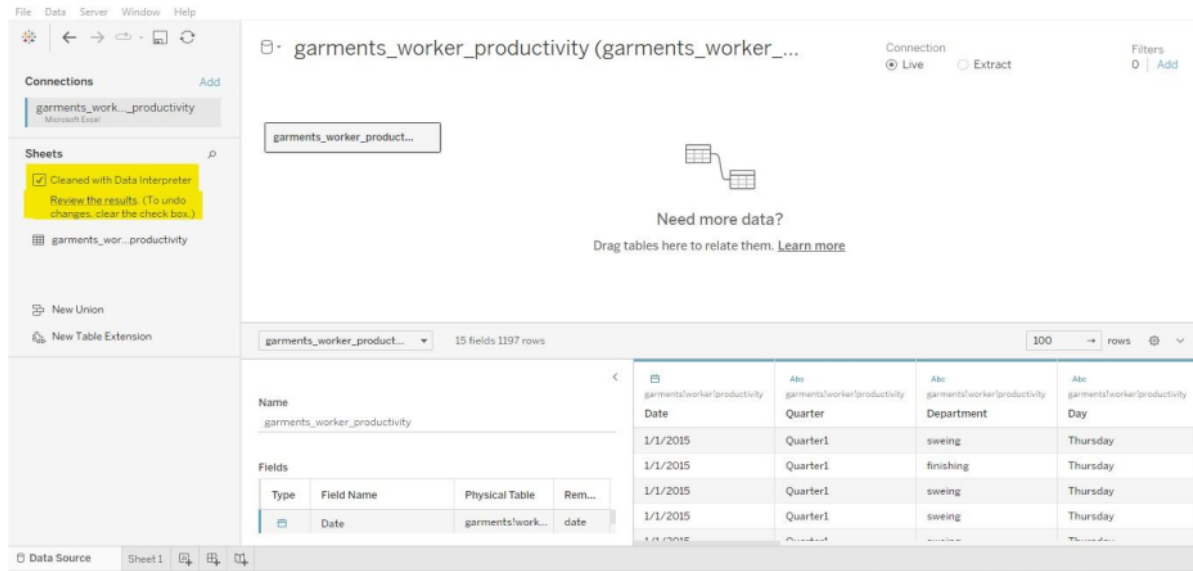


Figure 5. Screenshot of data interpreter being used for data cleaning

STEP 3. Finding the missing values and outliers in the columns is step three. It takes a mix of data exploration, data analysis, and data visualization approaches to extract outliers in Tableau. Hidden patterns and correlations in the remaining data can be revealed by locating and examining outliers.

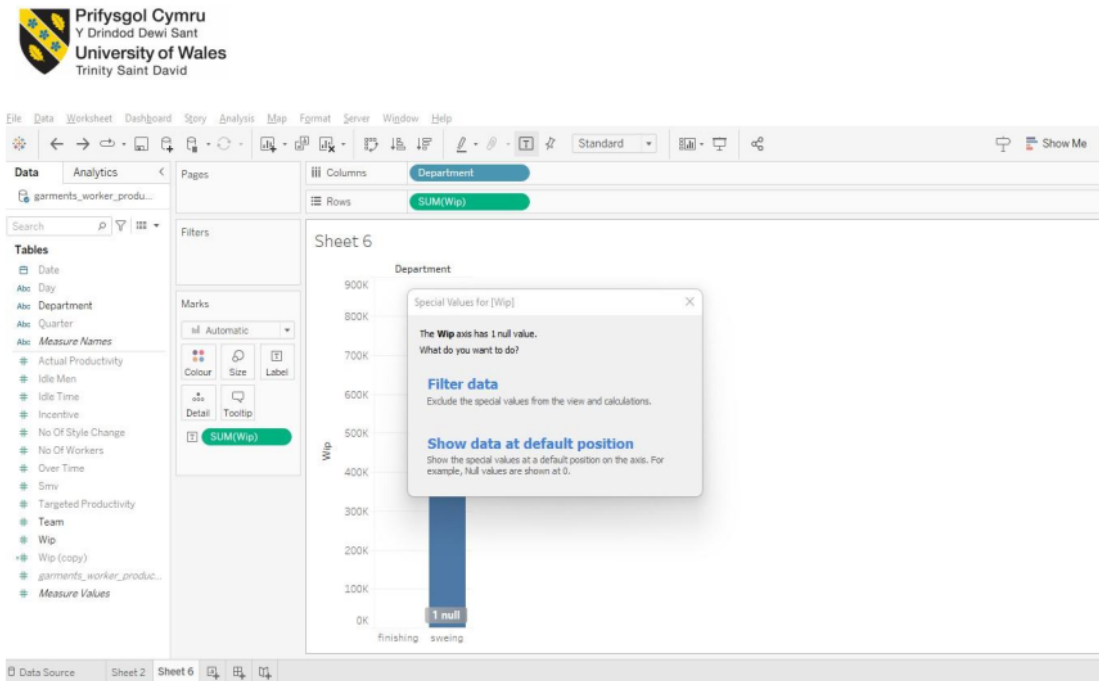
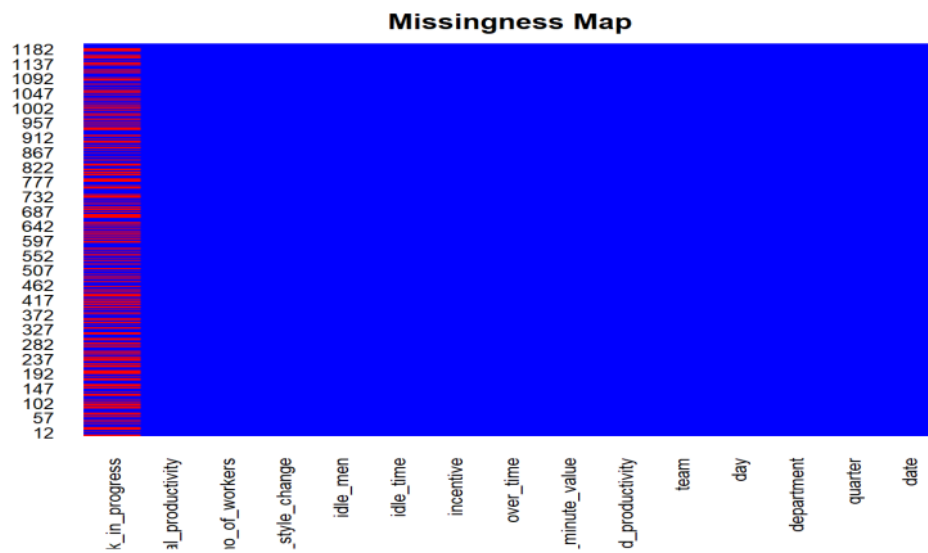


Figure 6. Screenshot of detecting missing/null values in the WIP column

Work in progress has a lot of NaN values. Therefore, we will enter a value of 0 in that cell to indicate that there are no items being worked on at the moment. There is no need to eliminate outliers because the dataset is tiny.



[7]

Figure 7. Missing values gap graph

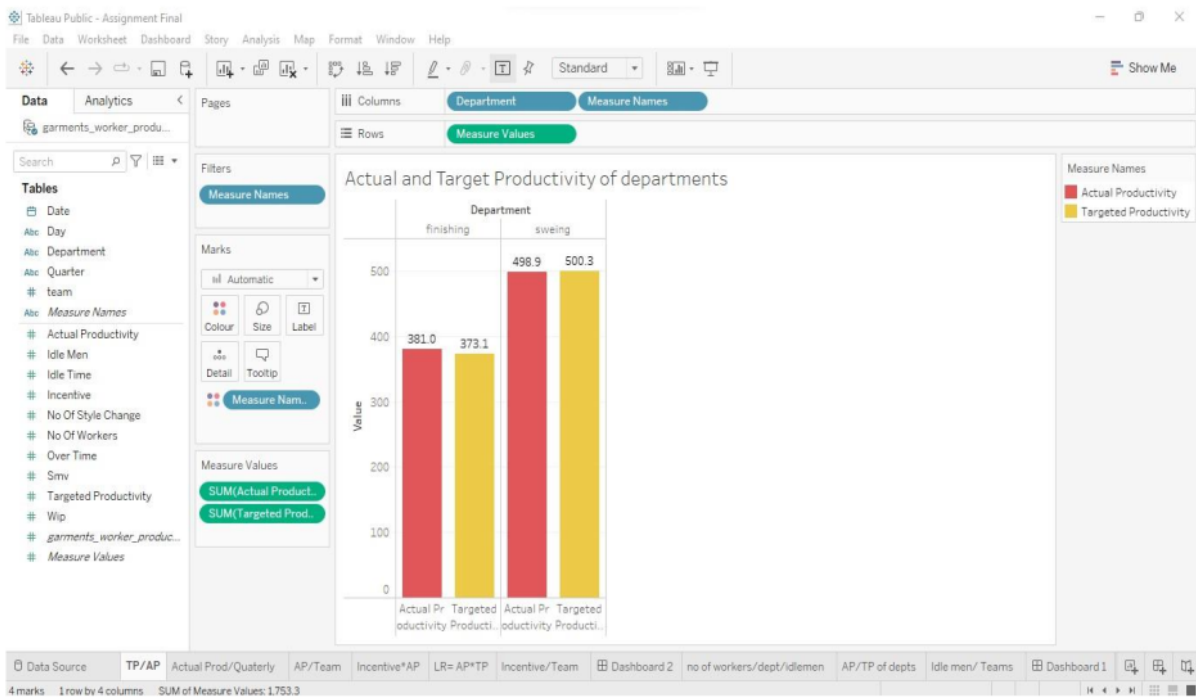


Figure 8. Comparison between actual and target productivity of the two departments

Targeted productivity is frequently set to 0.8, the maximum possible value. The variation in real productivity, however, appears to be rising as the intended productivity rises significantly. This shows that setting extremely lofty goals may not be wise.

In comparison to the finishing department, the sewing department's actual productivity appears to be more closely matched to its desired productivity.

The two departments' real production appears to be almost similar on average, but the sewing department has a more constant average actual productivity and a greater number of unproductive outliers.

The preceding graph showed that "higher the target productivity amount, higher the real productivity."

The goal productivity for the sewing department was about 500, while the overall productivity was just 499.

The finishing department's objective productivity is 373, but it really achieves a figure of 381.

Actual productivity is significantly influenced by target productivity.

Investigate how actual productivity is correlated with the quarters

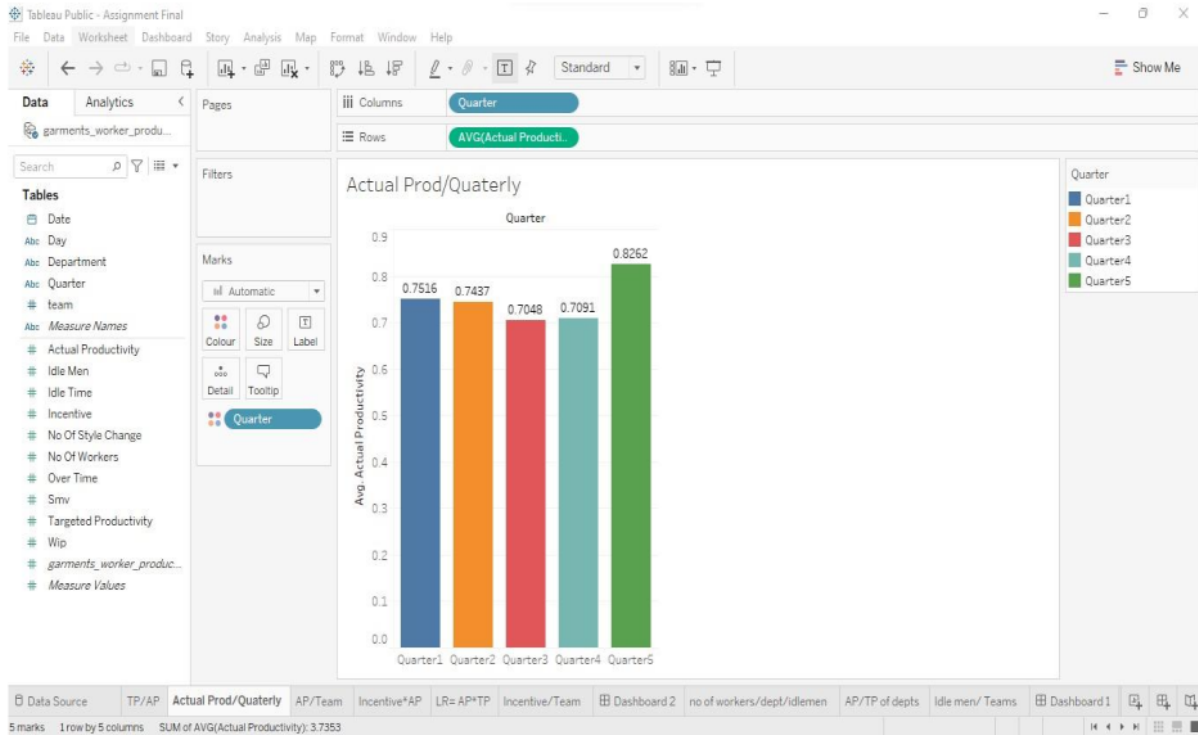


Figure 9. Average actual productivity with relation to quarters

In comparison to the other quarters, Quarter 5 appears to have higher actual productivity. This also explores that the fifth quarter is, on average, substantially more productive than the previous four. Additionally, it appears that actual productivity dips the most in Quarters 3 and 4.

5. ANALYSIS OF THE DATASET

This dataset can be examined using regression or classification analytical tools by predicting the productivity range (0–1) or categorising the productivity range (0–1).

REGRESSION

One or more independent variables and a dependent variable can have a linear relationship, and this connection can be modelled statistically using the data mining technique named linear regression. Linear regression seeks to reduce the discrepancy between the anticipated and actual values to identify the line that best fits the connection between the variables.

Because of its simplicity and readability, linear regression is a widely used approach in data analytics and data science. The value of a continuous dependent variable can be predicted using one or more continuous or categorical independent variables. For instance, In this dataset, the analysis of actual and intended productivity and some other characteristics will help to forecast the productivity of the garment industry employees or teams.

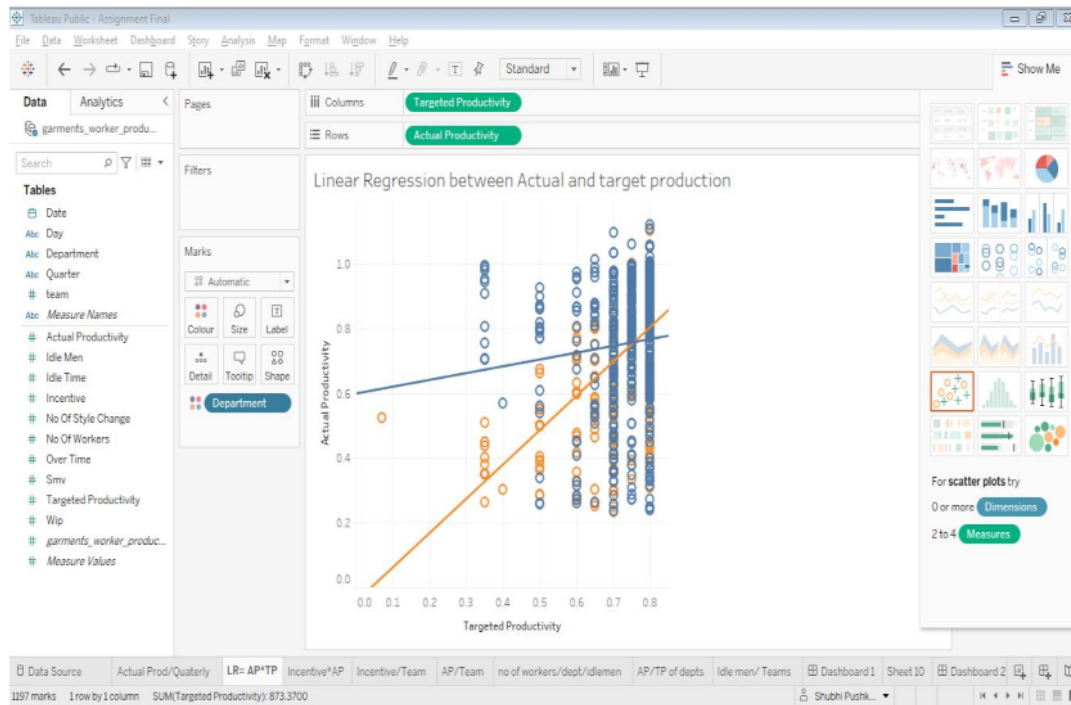


Figure 10. Linear regression analysis on actual and targeted productivity

These 2 variables exhibit a little heteroskedastic pattern.

Targeted productivity is frequently set to 0.8, the maximum possible value. The variation in real productivity, however, appears to be rising as the intended productivity rises significantly. This shows that setting extremely lofty goals may not be wise.

In comparison to the finishing department, the sewing department's actual productivity appears to be more closely matched to its desired productivity.

Additionally, we can see that the number for targeted productivity is typically a well-rounded value in contrast to the widely fluctuating value of actual productivity.

We can see that the sewing department's actual productivity appears to be more closely tied to its intended productivity than the finishing department, based on linear regression between the target productivity and actual productivity of the two separate departments, Sewing  Finishing 

Examine the connection between incentive and real production

According to intuition, the incentive should have the highest correlation with real production. Let's investigate more to learn more about that. I'll start with a scatterplot that illustrates the connection between incentives and actual productivity.

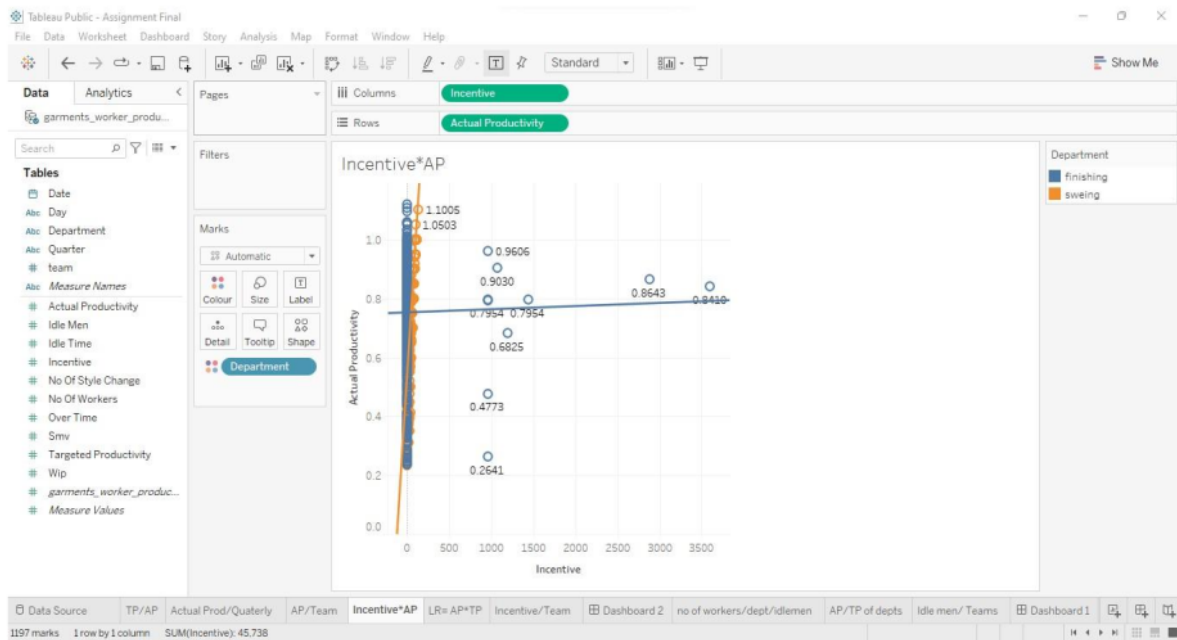


Figure 11. Linear Regression on incentive with actual productivity graph

We can see that actual productivity and incentive do, in fact, now have a rather significant positive association. When it comes to the relationship between incentives and real production, finishing and sweing are both strongly connected; the greater the actual output, the greater the incentive. The graph illustrates the relationship between the department's real production and incentive, demonstrating how closely related these two characteristics are.

Relationship between incentives and teams

Teams are formed among employees in the garment industry to improve management and output. The "*team*" attribute in this dataset is transformed from a measure to a series of dimensions before being translated to numerical data. The actual productivity of each team is different, therefore every team receives a different amount of incentives.

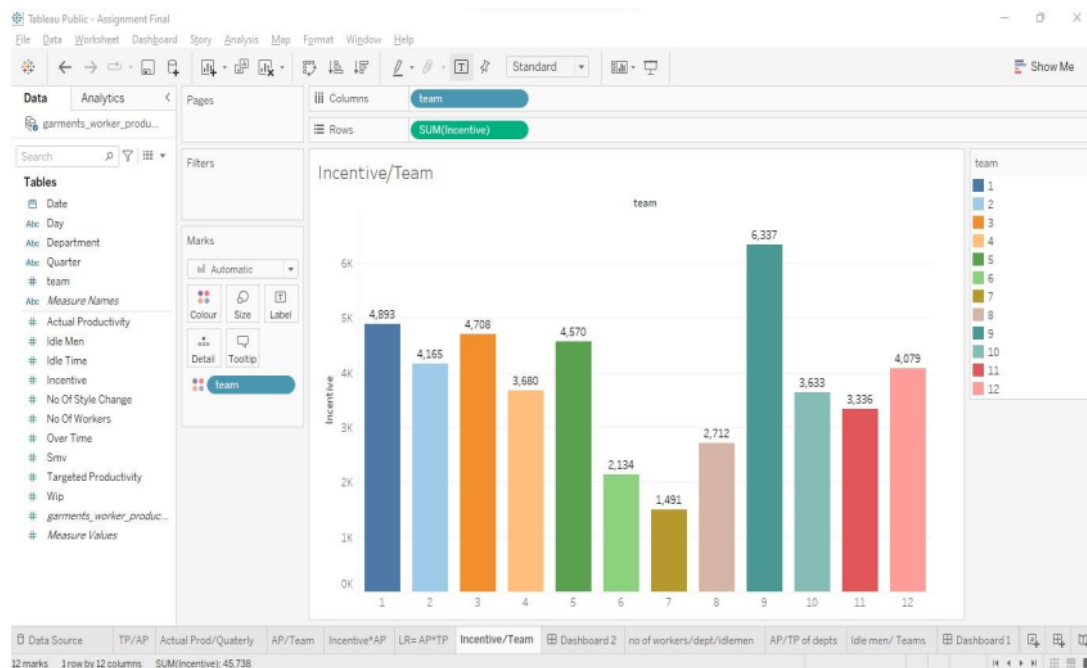


Figure 12. Bar Graph between Incentive received per team

We can see from the bar graph above that Team 9 earned the most incentive, with the majority of other teams receiving only half as much.

It is reasonable to assume that this will at least somewhat affect each team's production, with Team 9 perhaps being one of the most productive teams, if not the most prolific. Let's find out whether that's the case.

Relationship between actual productivity and number of workers

Let's check to see if actual productivity is impacted by the number of employees. A scatterplot between these two variables will be my first analysis for this section.

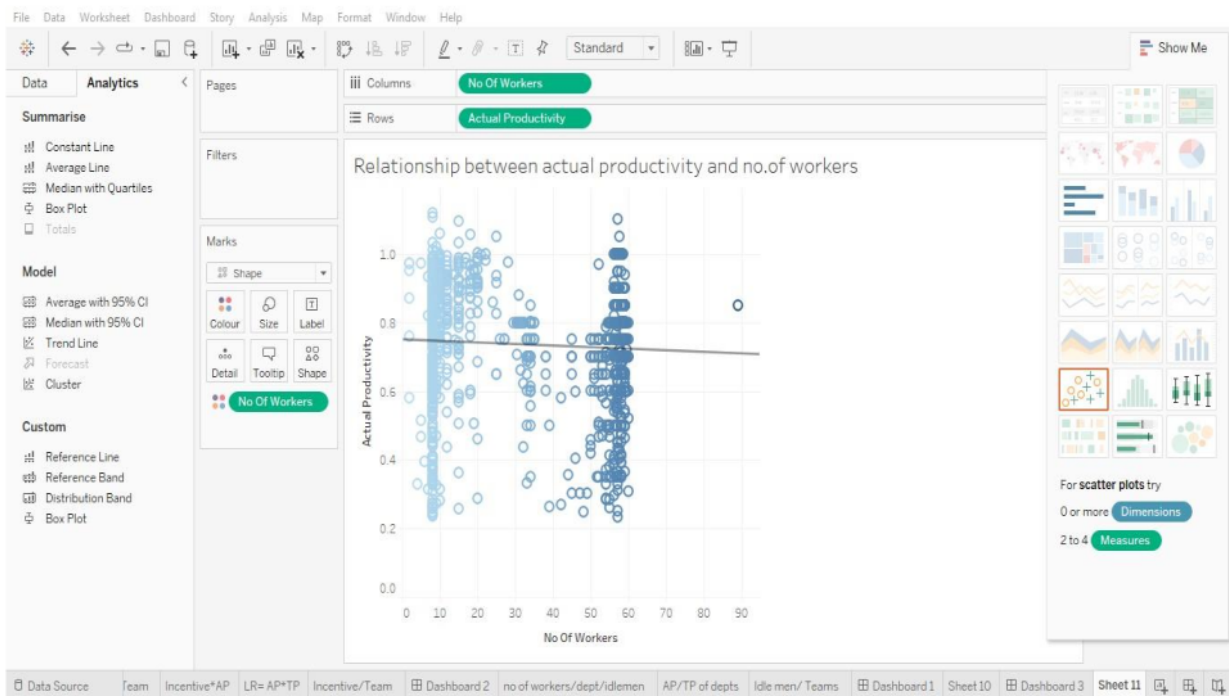


Figure 13. Scatterplot visualization between actual productivity and number of workers

We can see an intriguing trend where, on average, real productivity increases up to a certain number of people, then declines, then increases once again.

As per the Law of Diminishing Marginal Returns, this can be explained. The productivity rate initially rises with the addition of more people, but when the number of workers grows too large, the rate declines. If we keep hiring additional people, eventually the productivity rate may start to increase once again as some tasks become more efficient with more people present, and so forth [8].

ANALYSIS THROUGH DASHBOARD USING TABLEAU

A Tableau dashboard is an assortment of visuals, charts, and tables presented on one page to offer a thorough view of data. With the aid of dashboards, one can swiftly analyze big datasets and provide decision-makers with clear, visually attractive information.

Following are the advantages that were utilized while using the dashboard in Tableau:

Better Data Analysis: Dashboards give users a rapid, complete overview of data, allowing them to quickly analyse massive volumes of data and discover trends and patterns that may not be readily apparent when examining raw data.

Real-time monitoring: Dashboards that automatically update and offer real-time insights are particularly helpful for keeping an eye on important indicators like sales, revenue, and customer engagement and in this case, actual productivity, incentives and targeted productivity.

Better Data Analysis: Dashboards gives a rapid, complete overview of data, allowing us to quickly huge data sets and identify patterns and trends that may not be immediately obvious from looking at raw data.

Interactive and customizable: Dashboards provides interactive features like filters, drill down functions and many in order to explore the data more accurately and easily.

Tableau dashboards, in general, offer a potent approach to derive insights from data and convey these insights to others.

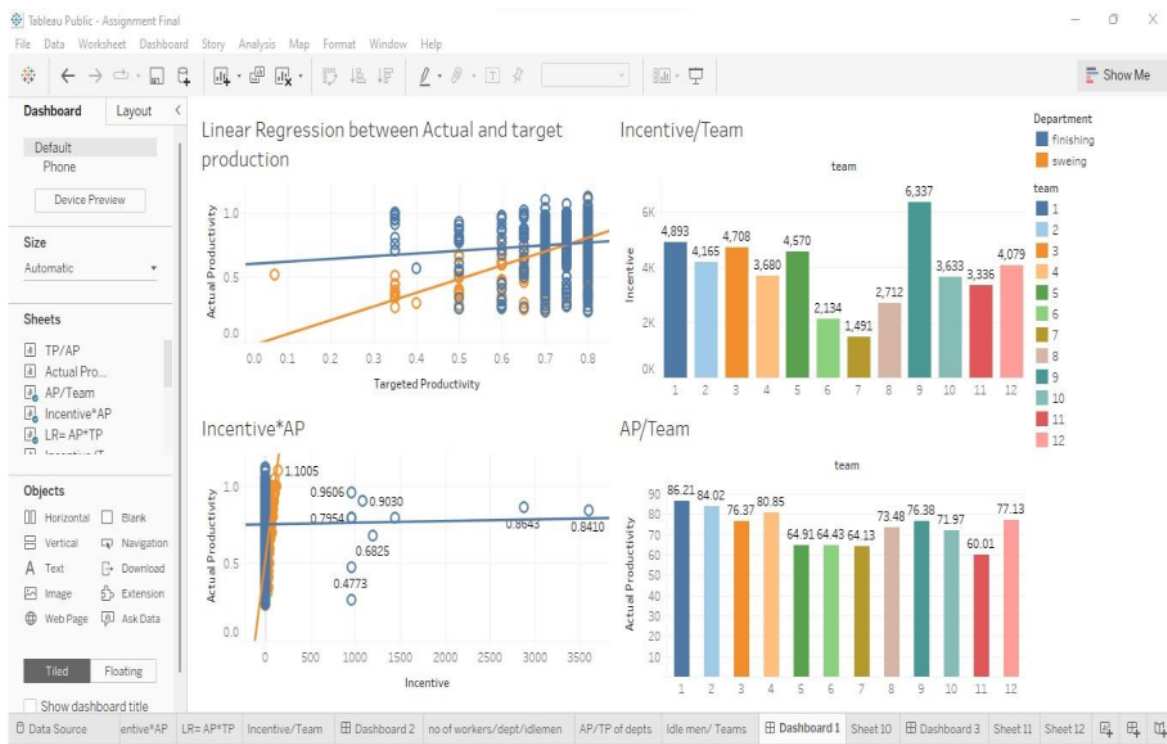


Figure 14. Screenshot of Dashboard having 4 different Graphs

The screenshot represents four different graphs on a single sheet known as *Dashboards*.

On the second graph, the relationship between actual productivity and target production, and incentive is shown through linear regression. When comparing the analyses from the dashboard, it is evident that the Sewing Department's actual productivity is more closely related to its intended production. As a result, it is clear from another graph showing real productivity and reward that the sewing department received more incentives than the finishing section.

Furthermore, Comparing the other two graphs, it can be seen that

1. Team 9 earned the largest reward, with the majority of other teams receiving only around half as much.
2. Team 9 is neither the most nor, on average, in the top 5 most productive teams. In that order, Teams 1, 12, and 3 are the most effective. The average reward given to each of them was just about 3500 BDT.
3. The teams who received the least amount of incentive, which are Teams 6, 7, and 8, on the other hand, had the lowest average levels of real production.

It appears that the additional incentive no longer affects real productivity at a certain point.

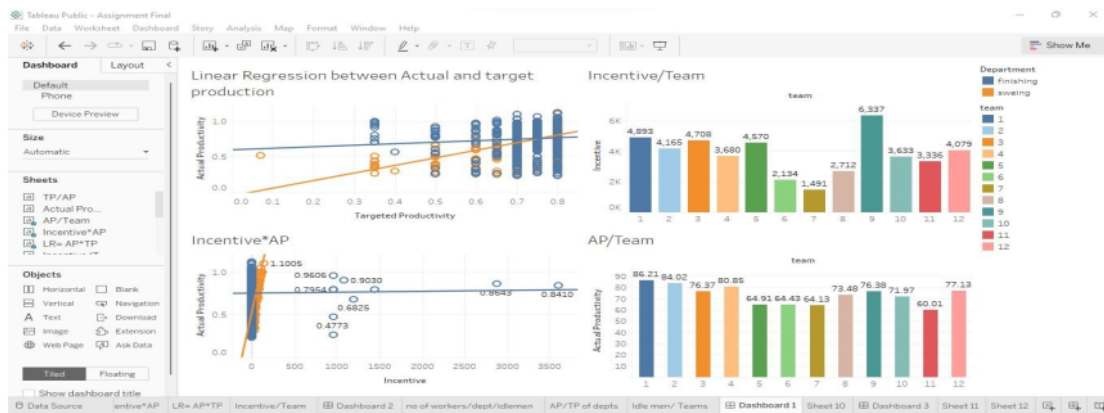


Figure 15. Dashboard having graphs on Incentives, Actual productivity and Team

6. REPORT AND EVALUATION

After performing extensive EDA (Exploratory data analysis) and creating a number of models, some of which were overfit, I feel I have found a number of crucial elements that have a significant impact on real productivity. Based on EDA, it seems that targeted productivity may have a kind of self-fulfilling prophecy effect, strongly influencing actual productivity.

Another element that may have an impact on real output is the quarter, with Quarters 5, 1, and 2 being the most fruitful, correspondingly.

The real production can also be influenced by the team, with teams 1, 12, and 3 being the most productive, correspondingly.

Although its impact is more complex and not always linear, the number of workers can also have an impact on actual productivity, as shown by EDA and model development.

Incentives are still another crucial element that can be a useful predictor of actual productivity, but only after some outliers and zero values have been eliminated. This makes sense because incentives that are zero should not have an impact on actual productivity.

Even though I don't think SMV (Standard minute value) is important, I nevertheless took it into account because it has a strong correlation with the number of employees. An interaction phrase can adequately describe this relationship.

Evaluation

The dataset was analysed with TABLEAU using linear regression and other approaches, and some of the findings are shown here.

1. Most attribute relationships are weak or non-existent. Very few relationships, both moderate and strong, are positive.
2. For all variables, there were no differences between the days of the week.
3. In comparison to the other quarters, Quarter 5 appears to have higher actual_productivity.
4. Actual productivity appears to have peaked during the third and fourth quarters. In idle men attribute, good number of outliers has been detected.
5. The two departments' real production appears to be essentially the same on the whole. However, more unproductive outliers and a more stable average actual productivity can be seen in the sewing department than in the finishing one.
6. No of workers and SMV show a significant positive association. This makes obvious since we would anticipate that the more personnel assigned to a task, the longer it will take.
7. Team 9 earned the largest reward, with the majority of other teams receiving only around half as much.
8. Team 9 is neither the most nor, on average, in the top 5 most productive teams. In that order, Teams 1, 12, and 3 are the most effective. The average reward given to each of them was just about 3500 BDT.

9. The teams who received the least amount of incentive, which are Teams 6, 7, and 8, on the other hand, had the lowest average levels of real production. It appears that the additional incentive no longer affects real productivity at a certain point.
10. Actual productivity often increases up to a particular number of workers before declining and then increasing once again.
11. There is no obvious trend in the actual productivity figures, which appear to be rather constant throughout the amount of WIP. Although there is a tenuous correlation, the values of actual productivity appear to slightly decline as over-time increases. While there is a slight decrease in actual productivity as SMV rises, this relationship is tenuous

Future Work

These models can definitely be refined further, maybe more data will benefit them as well. Other more advanced machine learning models such as Random Forest, or Support Vector Machines can also be implemented with potentially much better results.

7. REFLECTION

When I think back on the process of writing this report, I believe I have improved my research, analysis, and communication abilities. I studied specifically how to collect and analyse data using a variety of techniques and technologies, including surveys, interviews, and data visualization tools like Tableau.

Effective time management was one of the difficulties I encountered while working on this project. I discovered that in order to do the research and analysis within the allotted time, I had to be highly disciplined and organized. But I also discovered that having specific objectives and due dates made it easier for me to stay on course and finish the project on schedule.

Effective data interpretation and communication was another problem I encountered. I came to see how crucial it was to analyse the data and deliver it in a clear, succinct manner. I have to use critical thinking to consider how the information may be best conveyed to various audiences.

Overall, I believe that this project has aided me in honing my research, analytical, and communication abilities, all of which I am confident will be useful to me in my future academic and professional endeavours. Time management, organisation, and effective communication are other abilities I learnt the value of and will keep honing in the future.

8. THE FINAL SAY (CONCLUSION)

To summarize, the dataset, which has 15 characteristics and 1197 occurrences, was downloaded from the original UCI repository. The combined nominal and numerical properties that make up this dataset have been used as dimensions and measures, respectively. Using Tableau, missing and null values have been removed through the data cleaning process.

Despite having few attributes, the chosen dataset provides a thorough understanding of the information. The dataset offers rich knowledge about a variety of parameters, simplifying the analysis and comparison of productivity and the ability to make trustworthy decisions in order to estimate performance output based on productivity. Using this data collection, the linear regression model was followed to best predict employee productivity. Furthermore, using this dataset in the future, it can also be analyzed by the best classification model other than the regression technique to determine whether actual productivity met the desired productivity.

In order to examine many graphs at once and assess the best and most accurate findings, tableau's *dashboard* feature was employed. The dataset is not extremely vast and provides clarity on the properties, allowing for the comparison of various graphs and the ability to analyze better findings. If properly analyzed, the dataset offers a sufficient number of comparisons between the characteristics, enabling reliable prediction.

To get raw data into a clean, organised, and structured format that can be quickly analysed using statistical techniques like linear regression, data pre-processing and cleaning approaches were used. "Date" attribute was deleted during this proper and "Team" attribute was converted into a dimension. The data set has 506 missing values.

Overall, the dataset included a reasonable number of pertinent qualities for analyses, however the data could have been more effectively examined and predictions could be produced with greater accuracy if attributes like "No of holidays," "Leaves taken by employee," and "Team capacity" had been included.

9. REFERENCE LIST

- [1] "Exporters Hardly Grab Orders Diverted from China." *The Daily Star*, 11 Aug. 2017,
www.thedailystar.net/business/exporters-hardly-grab-orders-diverted-china-1446907
- [2] M. S. Islam, M. A. Rakib and A. T. M. Adnan, "Ready-Made Garments Sector Bangladesh: Its Contribution and Challenges towards Development", *Stud*, vol. 5, no. 2, 2016
- [3] Al Imran, Abdullah , et al. *Deep Neural Network Approach for Predicting the Productivity of Garment Employees*. 2019.
- [4] Nimlaor, Chavalit, et al. "Thai Garment Industry Organisational Characteristics and Strategic Factors That Affect Corporate Performance." *International Journal of Management and Enterprise Development*, vol. 14, no. 2, 2015, p. 126,
<https://doi.org/10.1504/ijmed.2015.070096>. Accessed 1 June 2020
- [5] Gunawan, Agus, et al. "Increasing the Managerial Capabilities in Indonesian Garment Manufacturing." *International Journal of Economic Policy in Emerging Economies*, vol. 3, no. 4, 2010, p. 346, <https://doi.org/10.1504/ijepee.2010.037582>. Accessed 12 May 2022.
- [6] Hand, David J. "Principles of Data Mining." *Drug Safety*, vol. 30, no. 7, 2007, pp. 621–622,
<https://doi.org/10.2165/00002018-200730070-00010>.
- [7] rpubs.com. (n.d.). *RPubs - Garments Productivity Prediction (WQD7004 Group 7)*. [online] Available at: <https://rpubs.com/jialing2511/wqd7004occ2group7>
- [8] "Exploratory Data Analysis and Prediction: Productivity of Garment Industry Employees." *Vietpham.com*, 22 Dec. 2021, www.vietpham.com/productivity-garment-industry/.

Assignment 2 -2230035-Shubhi

GRADEMARK REPORT

FINAL GRADE

GENERAL COMMENTS

Instructor

76/100

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8

PAGE 9

PAGE 10

PAGE 11

PAGE 12

PAGE 13

PAGE 14

PAGE 15

PAGE 16

PAGE 17

PAGE 18

PAGE 19

PAGE 20

PAGE 21

PAGE 22

PAGE 23

PAGE 24

PAGE 25

PAGE 26

PAGE 27

PAGE 28

PAGE 29

PAGE 30

PAGE 31

PAGE 32

GRADING FORM: DATA MINING / VIS A2_2022/3

SHUBHI PUSHKARNA

76

KNOWLEDGE AND UNDERSTANDING

Dataset description (15 marks)



Firm understanding of the garment manufacturing industry is provided. Description of attributes is sufficiently clear.

13

COGNITIVE/ INTELLECTUAL SKILLS

Primary and Secondary data analysis, reflection (35 marks)



Descriptive statistics used to demonstrate both missing and null values excellently. Productivity across quarters showcases a need for introducing a particular algorithm.

31

PRACTICAL APPLICATION SKILLS

Primary and Secondary data processing (20 marks)



Practical understanding of established machine learning techniques (regression in particular) addressed fairly well in assignment.

16

TRANSFERABLE SKILLS

Data reporting and report writing (30 marks)



Suitably constructed assignment that has addressed most themes required for its submission. In terms of meeting the requirements for skills for professional practice this is done fairly well.

16