Productivity Prediction of Garment Employees

PREPARED BY XYZ GROUP

MODULE: FORECASTING AND PREDICTIVE ANALYTICS - BM9719

16:00 – 18:00

Northumbria University NEWCASTLE

Table of Content

Introduction

- Productivity is a key factor in industrial globalization since it affects how effectively and competitively businesses operate on a global scale. World Bank research claims that achieving sustainable economic development and lowering poverty depend on productivity increases.
- The Productivity Prediction of Garment Employees dataset is available on the UC Irvine Machine Learning Repository (https://archive.ics.uci.edu/ml/datasets/Productivity+Prediction+of+Garment+Employees). The dataset was collected from a garment manufacturing company and contains 1197 instances with 15 attributes, including the target variable "productivity".
- The garment sector plays a vital role in international trade and significantly affects economic expansion and development. In order to pinpoint the elements influencing employee performance and enhance production procedures, there has been a rising interest in productivity data analysis of the apparel sector in recent years (Imran et al., 2021).
- The dataset's goal is to forecast employee productivity based on a variety of variables, including experience, education, and work rate.
- Tracking, analysing, and forecasting the productivity performance of the working teams in their factories is highly desired by the decision-makers in the apparel business.

Dataset Variables

Dependent Variable

Actual_productivity : The actual % of productivity that was delivered by the workers. It ranges from 0-1.

Independent Variables

- Date : Date in MM-DD-YYY.
- Day : Day of the Week.
- Quarter : A portion of the month. A month was divided into four quarters.
- Department : Associated department with the instance.
- Team : Associated team number with the instance.
- No_of_workers : Number of workers in each team.
- No_of_style_change : Number of changes in the style of a particular product.
- Targeted_productivity : Targeted productivity set by the Authority for each team for each day.
- SMV : Standard Minute Value, it is the allocated time for a task.
- WIP : Work in progress. Includes the number of unfinished items for products.
- Over_time : Represents the amount of overtime by each team in minutes.
- Incentive : Represents the amount of financial incentive (in BDT) that enables or motivates a particular course of action.
- Idle_time : The amount of time when the production was interrupted due to several reasons.
- Idle_men : The number of workers who were idle due to production interruption.

Objectives

- To identify factors that impact employee productivity such as worker experience, machine speed, incentive and working hours.
- To develop predictive models for productivity where the models can be used to forecast productivity and make adjustments to staffing levels or production schedules as needed.
- To evaluate the effectiveness of interventions aimed at improving productivity
- To use the dataset to determine the productivity levels against industry standards and identify areas where they may need to improve.
- To analyze the impact of employee satisfaction on productivity

DATA EXPLORATION

Data View

- Loaded the necessary packages for exploration.
- Imported the dataset
- We have 15 variables and 1,197 rows in the dataset

```r
library(tidyverse)
library(ggplot2)
library(gridExtra)
library(class)

# Imported Data set

garment <- read.csv("garments_worker_productivity.csv")
```

Filter

	date	quarter	department	day	team	targeted_productivity	smv	wip	over_time	incentive	idle_tim
1	2015-01-01	Quarter1	sweing	Thursday	8	0.80	26.16	1108	7080	98	
2	2015-01-01	Quarter1	finishing	Thursday	1	0.75	3.94	NA	960	0	
3	2015-01-01	Quarter1	sweing	Thursday	11	0.80	11.41	968	3660	50	
4	2015-01-01	Quarter1	sweing	Thursday	12	0.80	11.41	968	3660	50	
5	2015-01-01	Quarter1	sweing	Thursday	6	0.80	25.90	1170	1920	50	
6	2015-01-01	Quarter1	sweing	Thursday	7	0.80	25.90	984	6720	38	
7	2015-01-01	Quarter1	finishing	Thursday	2	0.75	3.94	NA	960	0	
8	2015-01-01	Quarter1	sweing	Thursday	3	0.75	28.08	795	6900	45	
9	2015-01-01	Quarter1	sweing	Thursday	2	0.75	19.87	733	6000	34	
10	2015-01-01	Quarter1	sweing	Thursday	1	0.75	28.08	681	6900	45	
11	2015-01-01	Quarter1	sweing	Thursday	9	0.70	28.08	872	6900	44	
12	2015-01-01	Quarter1	sweing	Thursday	10	0.75	19.31	578	6480	45	

Showing 1 to 12 of 1,197 entries, 15 total columns

Cleaned Dataset

- We removed the rows that are not complete. That was we have a clean dataset to analyze.

```
# Removed not completed rows

garment <- na.omit(garment)
```

Filter

	date	quarter	department	day	team	targeted_productivity	smv	wip	over_time	incentive
1	2015-01-01	Quarter1	sweing	Thursday	8	0.80	26.16	1108	7080	98
3	2015-01-01	Quarter1	sweing	Thursday	11	0.80	11.41	968	3660	50
4	2015-01-01	Quarter1	sweing	Thursday	12	0.80	11.41	968	3660	50
5	2015-01-01	Quarter1	sweing	Thursday	6	0.80	25.90	1170	1920	50
6	2015-01-01	Quarter1	sweing	Thursday	7	0.80	25.90	984	6720	38
8	2015-01-01	Quarter1	sweing	Thursday	3	0.75	28.08	795	6900	45
9	2015-01-01	Quarter1	sweing	Thursday	2	0.75	19.87	733	6000	34
10	2015-01-01	Quarter1	sweing	Thursday	1	0.75	28.08	681	6900	45
11	2015-01-01	Quarter1	sweing	Thursday	9	0.70	28.08	872	6900	44
12	2015-01-01	Quarter1	sweing	Thursday	10	0.75	19.31	578	6480	45
13	2015-01-01	Quarter1	sweing	Thursday	5	0.80	11.41	668	3660	50
18	2015-01-01	Quarter1	sweing	Thursday	4	0.65	23.69	861	7200	0

Showing 1 to 12 of 691 entries, 15 total columns

Manipulation Of Date Variable

- In this section, we manipulated the date variable in other to have variables such as day, month, year and day of the week.
- This was done using as.date functon. We have 18 variables in total after doing this.

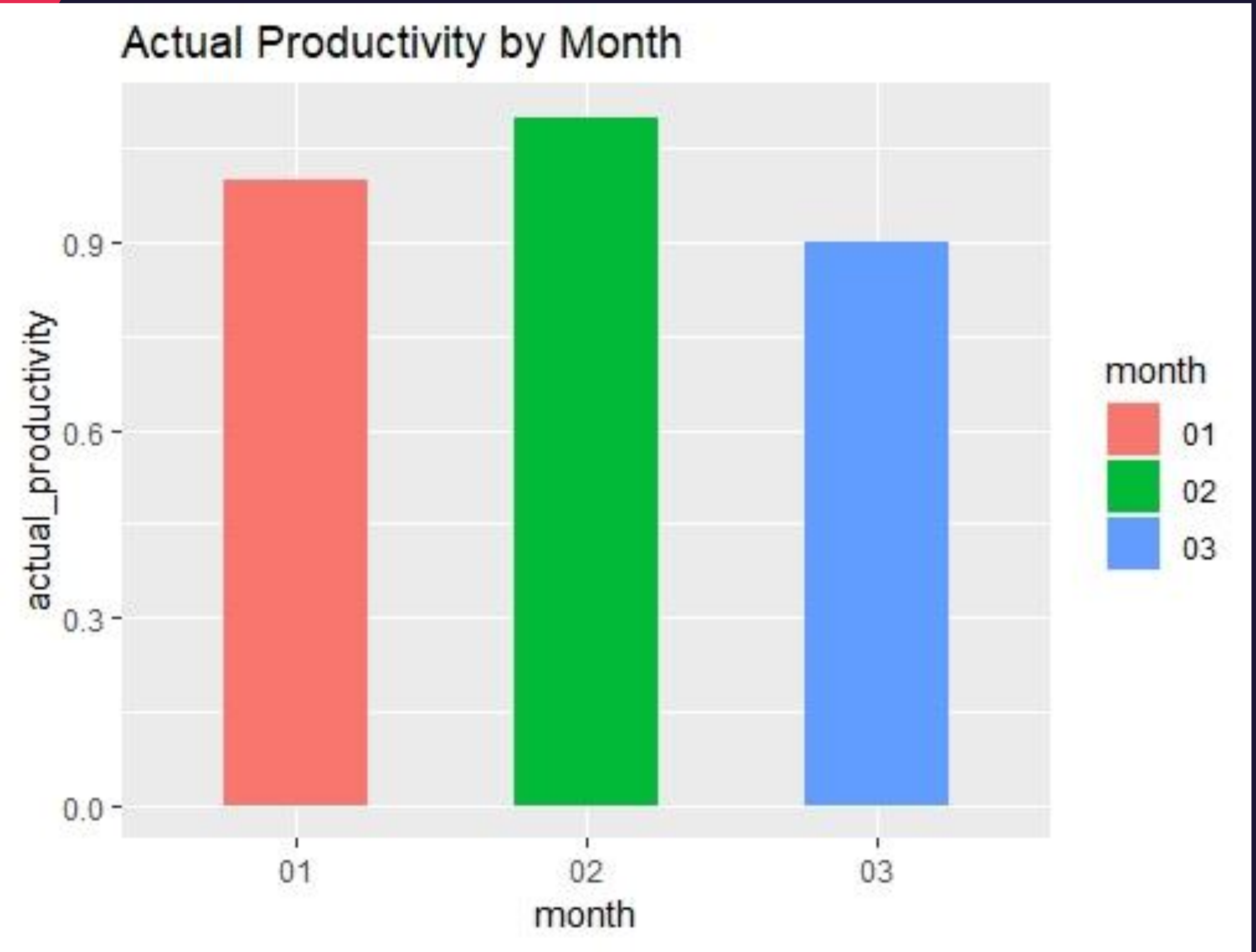
incentive	idle_time	idle_men	no_of_style_change	no_of_workers	actual_productivity	month	year	day_of_week
98	0	0	0	59.0	0.9407254	01	2015	Thursday
50	0	0	0	30.5	0.8005705	01	2015	Thursday
50	0	0	0	30.5	0.8005705	01	2015	Thursday
50	0	0	0	56.0	0.8003819	01	2015	Thursday
38	0	0	0	56.0	0.8001250	01	2015	Thursday
45	0	0	0	57.5	0.7536835	01	2015	Thursday
34	0	0	0	55.0	0.7530975	01	2015	Thursday
45	0	0	0	57.5	0.7504278	01	2015	Thursday
		0	0	57.5	0.7211270	01	2015	Thursday
		0	0	54.0	0.7122052	01	2015	Thursday
		0	0	30.5	0.7070459	01	2015	Thursday
		0	0	60.0	0.5211800	01	2015	Thursday

Showing 1 to 12 of 691 entries, 18 total columns

```r
# Changed  the datatype of date variable to have a better visualization of the dataset

garment$date <- as.Date(garment$date)
garment$month <- format(as.Date(garment$date), "%m")
garment$day <- format(as.Date(garment$date), "%d")
garment$year <- format(as.Date(garment$date), "%Y")
garment$day_of_week <- format(as.Date(garment$date), "%A")
```

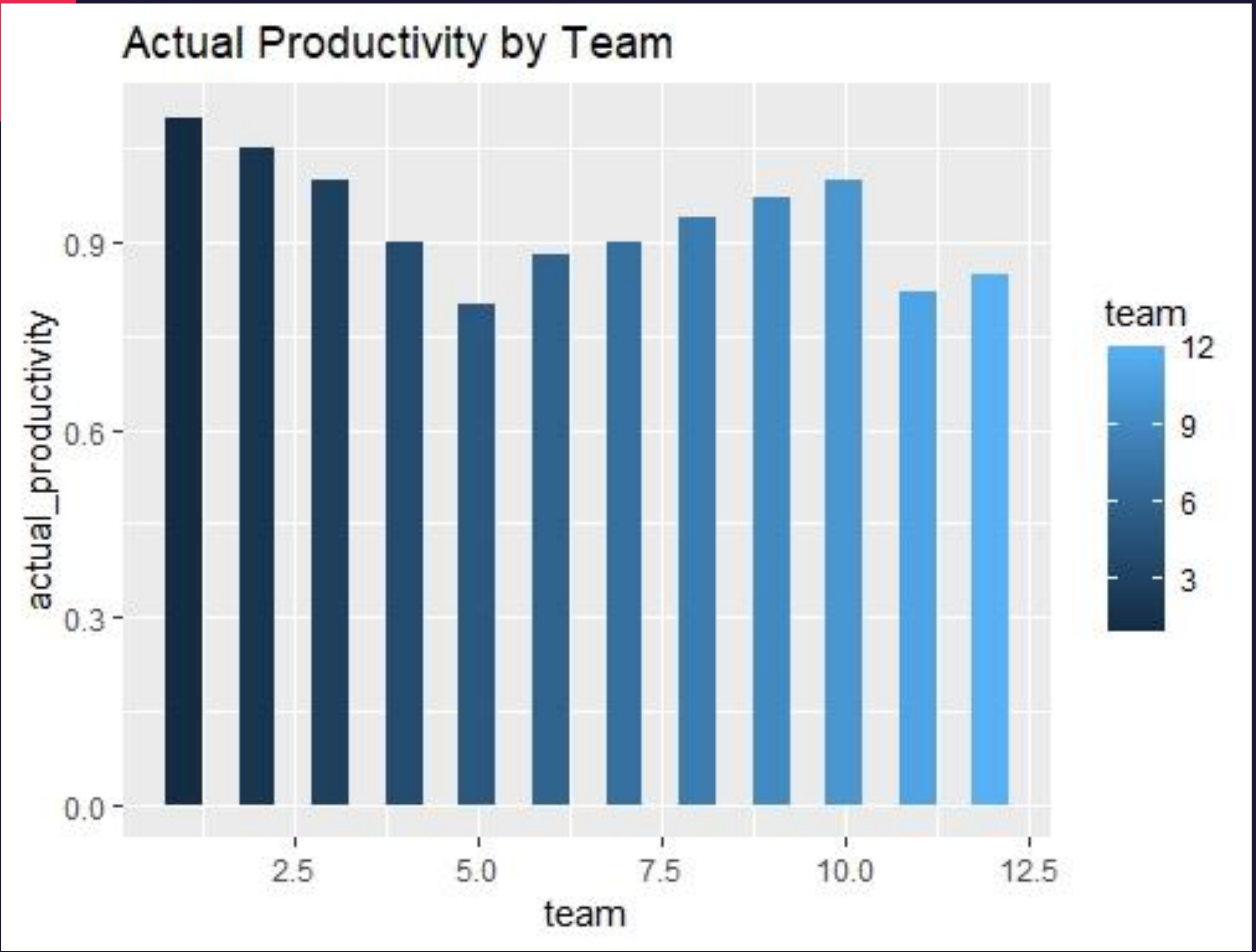
DATA ANALYSIS VISUALIZATION

Actual Productivity By Month



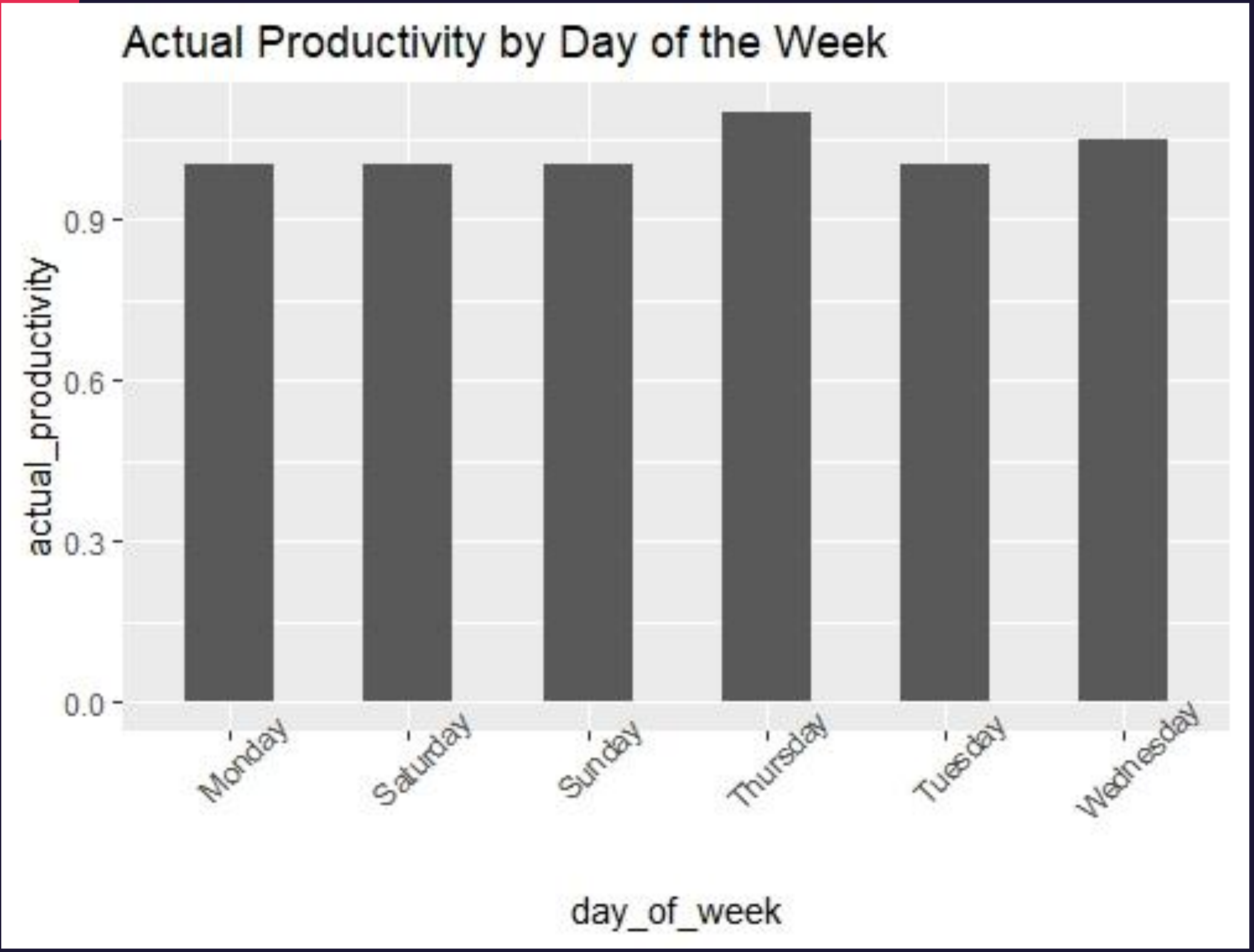

```r
garment %>%
  group_by(month) %>%
  ggplot(aes(x = month, y = actual_productivity, fill = month))+
  geom_col(width=0.5, position = position_dodge(width=0.5))+
  labs(title = "Actual Productivity by Month")
```

Actual Productivity By Team




```r
garment %>%
  group_by(team) %>%
  ggplot(aes(x = team, y = actual_productivity, fill = team))+
  geom_col(width=0.5, position = position_dodge(width=0.5))+
  labs(title = "Actual Productivity by Team")
```

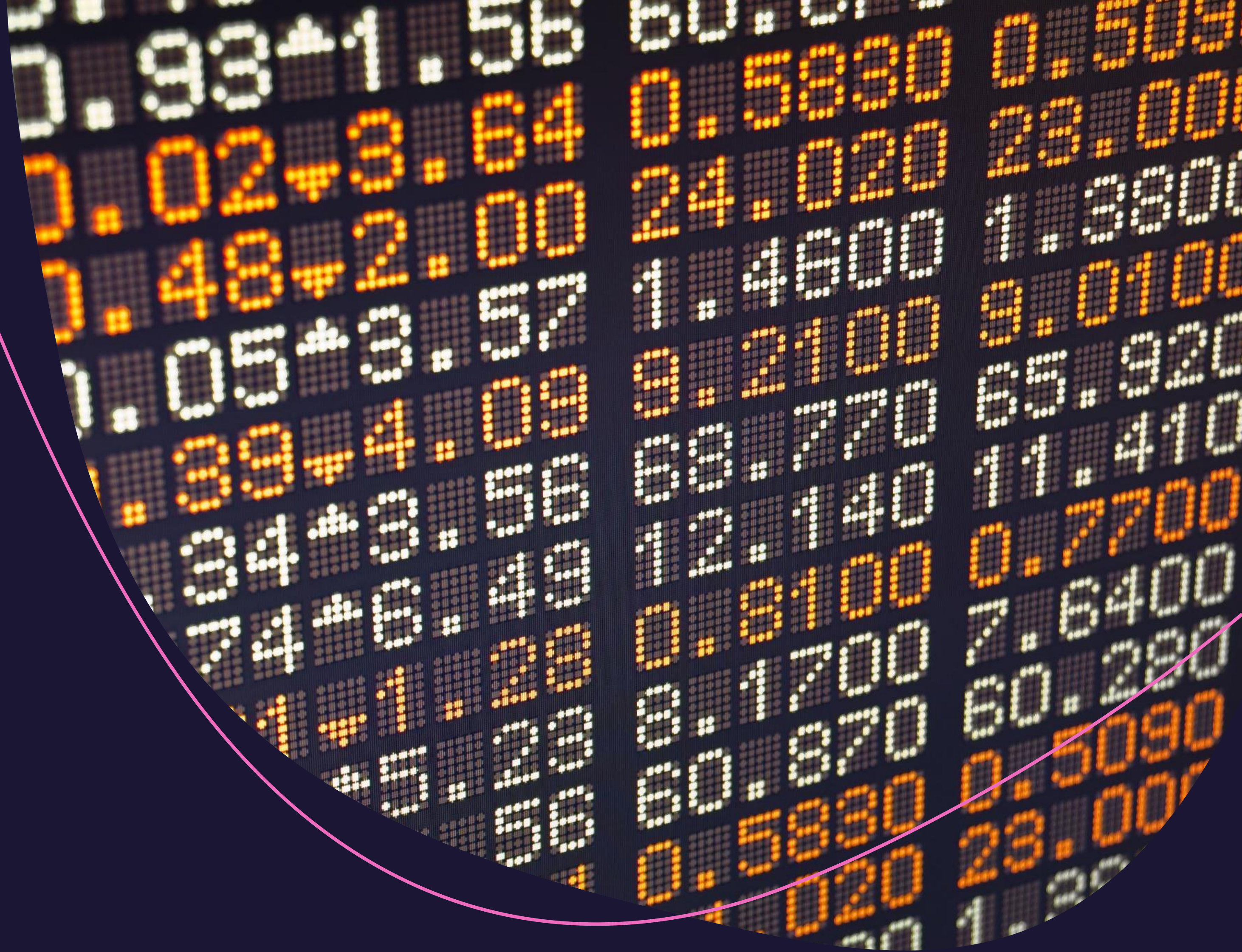
Actual Productivity By Day Of The Week



- In this data set, there is no “Friday”, so we believe this must have been their day of no work as they work on Sundays.

```
garment %>%
  group_by(day_of_week) %>%
  ggplot(aes(x = day_of_week, y = actual_productivity))+
  geom_col(width=0.5, position = position_dodge(width=0.5))+
  theme(axis.text.x = element_text(angle = 45))+
  labs(title = "Actual Productivity by Day of the Week")
```

DECISION TREE CLASSIFICATION



Removed Unnecessary Variables For Classification

```r
################ Decision Tree Classification ###############

# Removed unnecessary columns for Classification (Date, no_of_style_change and Year was removed)

garment <- garment[, c(2, 3, 4, 5, 7, 8, 9, 10, 11, 14, 15, 16, 18)]
```

- In this section, we removed the variable that we will not be using for classification.
- Decision tree requires numeric variables and "Date", is date data type. We have just a year dataset and also, no_of_style_change as no real value and can't seem to impact our dependent variable.
- We now have 13 columns to work on for classification which actual_productivity will be our dependent variable.

	quarter	department	day	team	smv	wip	over_time	incentive	idle_time	no_of_workers	actual_productivi
1	Quarter1	sweing	01	8	26.16	1108	7080	98	0	59.0	(
3	Quarter1	sweing	01	11	11.41	968	3660	50	0	30.5	(
4	Quarter1	sweing	01	12	11.41	968	3660	50	0	30.5	(
5	Quarter1	sweing	01	6	25.90	1170	1920	50	0	56.0	(
6	Quarter1	sweing	01	7	25.90	984	6720	38	0	56.0	(
8	Quarter1	sweing	01	3	28.08	795	6900	45	0	57.5	(
9	Quarter1	sweing	01	2	19.87	733	6000	34	0	55.0	(
10	Quarter1	sweing	01	1	28.08	681	6900	45	0	57.5	(
11	Quarter1	sweing	01	9	28.08	872	6900	44	0	57.5	(
12	Quarter1	sweing	01	10	19.31	578	6480	45	0	54.0	(
13	Quarter1	sweing	01	5	11.41	668	3660	50	0	30.5	(
18	Quarter1	sweing	01	4	23.69	861	7200	0	0	60.0	(

Showing 1 to 12 of 691 entries, 13 total columns

Changed Character Variables To Numeric

- In this section, we changed the variables in character data type to numeric data type.
- As we need numeric variables for classification

```
garment$quarter <- as.factor(garment$quarter)
garment$quarter <- as.numeric(garment$quarter)

garment$department <- as.factor(garment$department)
garment$department <- as.numeric(garment$department)

garment$day <- as.factor(garment$day)
garment$day <- as.numeric(garment$day)

garment$month <- as.factor(garment$month)
garment$month <- as.numeric(garment$month)

garment$day_of_week <- as.factor(garment$day_of_week)
garment$day_of_week <- as.numeric(garment$day_of_week)
```

	quarter	department	day	team	smv	wip	over_time	incentive	idle_time	no_of_workers	actual_producti
1	1	1	1	8	26.16	1108	7080	98	0	59.0	(
3	1	1	1	11	11.41	968	3660	50	0	30.5	(
4	1	1	1	12	11.41	968	3660	50	0	30.5	(
5	1	1	1	6	25.90	1170	1920	50	0	56.0	(
6	1	1	1	7	25.90	984	6720	38	0	56.0	(
8	1	1	1	3	28.08	795	6900	45	0	57.5	(
9	1	1	1	2	19.87	733	6000	34	0	55.0	(
10	1	1	1	1	28.08	681	6900	45	0	57.5	(
11	1	1	1	9	28.08	872	6900	44	0	57.5	(
12	1	1	1	10	19.31	578	6480	45	0	54.0	(
13	1	1	1	5	11.41	668	3660	50	0	30.5	(
18	1	1	1	4	23.69	861	7200	0	0	60.0	(

Showing 1 to 12 of 691 entries, 13 total columns

Converted Actual Productivity Variable Into A Categorical Variable

- In this section, we converted our dependent variable into character and factor “Low”, “Medium”, “High
- This is done in other to classify our productivity level into three.

  Where productivity from:
  0 - 0.3 = Low
  0.3 – 0.6 = Medium
  0.6 – 1 = High

```r
# Converted Actual productivity variable into a categorical variable
garment$actual_productivity <- ifelse(garment$actual_productivity <= 0.3, "Low",
                                ifelse(garment$actual_productivity <= 0.6, "Medium", "High"))

garment$actual_productivity <- as.factor(garment$actual_productivity)
```

lay	team	smv	wip	over_time	incentive	dle_time	no_of_workers	actual_productivity	month	day_of_week
1	8	26.16	1108	7080	98	0	59.0	High	1	4
1	11	11.41	968	3660	50	0	30.5	High	1	4
1	12	11.41	968	3660	50	0	30.5	High	1	4
1	6	25.90	1170	1920	50	0	56.0	High	1	4
1	7	25.90	984	6720	38	0	56.0	High	1	4
1	3	28.08	795	6900	45	0	57.5	High	1	4
1	2	19.87	733	6000	34	0	55.0	High	1	4
1	1	28.08	681	6900	45	0	57.5	High	1	4
1	9	28.08	872	6900	44	0	57.5	High	1	4
1	10	19.31	578	6480	45	0	54.0	High	1	4
1	5	11.41	668	3660	50	0	30.5	High	1	4
1	4	23.69	861	7200	0	0	60.0	Medium	1	4

Showing 1 to 12 of 691 entries, 13 total columns

Splited The Data Into Training And Test Sets

- In this section, we randomize our data set and then split the data set into training and test data. We used 80% values for training and then 20% values for testing.

```
# Splited the data into training and testing sets

set.seed(555)

ind <- sample(2,
              nrow(garment),
              replace = TRUE,
              prob = c(0.8, 0.2))

train <- garment[ind==1,
test <- garment[ind==2,]
```

Data	
garment	691 obs. of 13 variables
test	137 obs. of 13 variables
testnn	1197 obs. of 15 variables
train	554 obs. of 13 variables
Values	
ind	int [1:691] 1 2 1 1 1 1 2 1 1 2 ...

Decision Tree Model

- In this section, a decision tree model was used.
- Party package was used to partition.

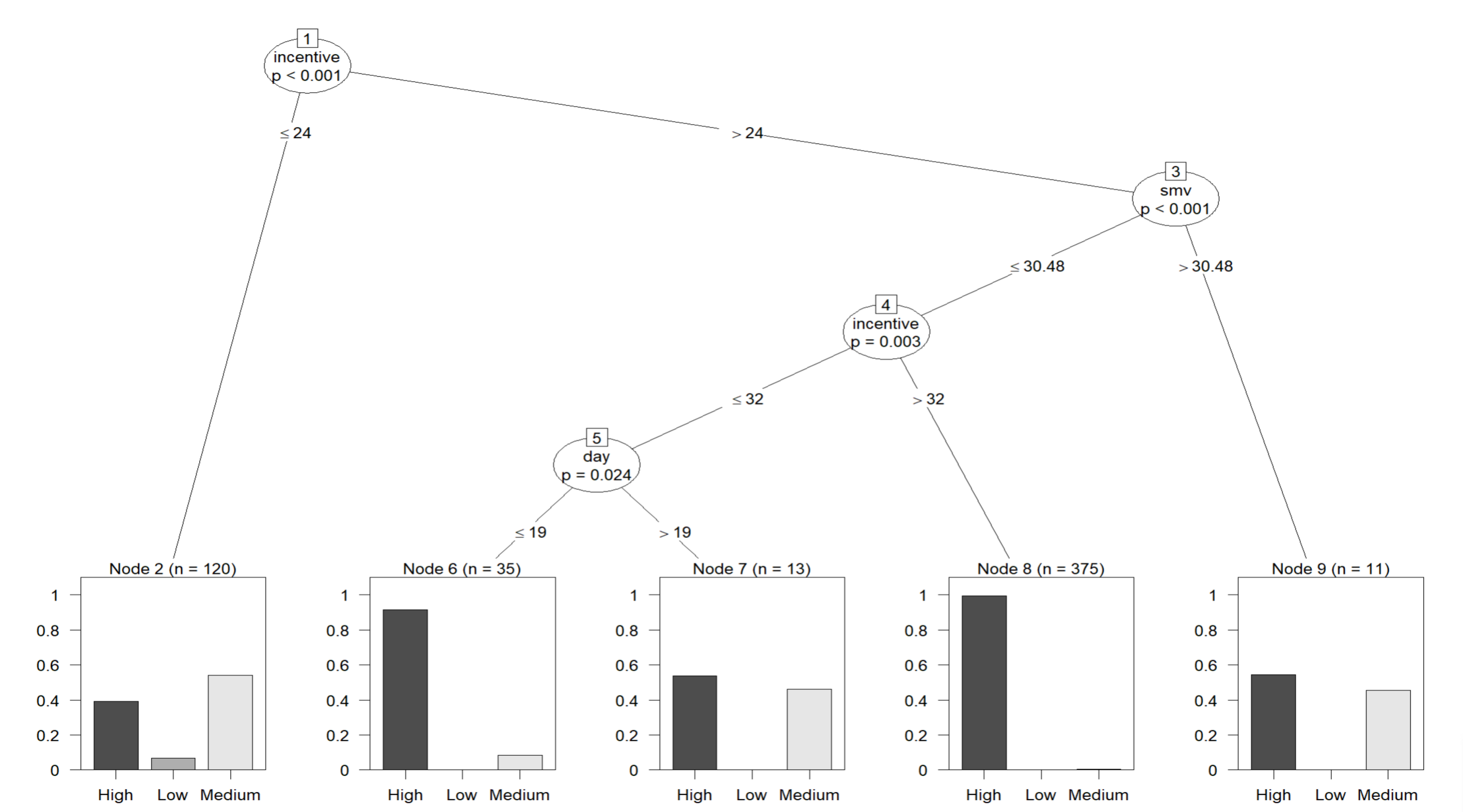
```r
# Decision Tree Model

library(party) #for partition

tree <- ctree(actual_productivity~., train)
```

Name	Type	Va
tree	S4 (party::BinaryTree)	S4 [illegible]t of class BinaryTree
data	S4 (modeltools::ModelEnvFormu	S4 [illegible]ct of class ModelEnvFormula
responses	S4 (party::ResponseFrame)	S4 object of class ResponseFrame
cond_distr_response	function	function(newdata = , mincriterion = 0, ...) { ... }
predict_response	function	function(newdata = , mincriterion = 0, type = c("response", "node", "prob"), ...
prediction_weights	function	function(newdata = , mincriterion = 0, ...) { ... }
get_where	function	function(newdata = , mincriterion = 0, ...) { ... }
update	function	function(weights =) { ... }
tree	list [10] (S3: SplittingNode)	List of length 10
where	integer [554]	8 8 8 8 8 8 ...
weights	double [554]	1 1 1 1 1 1 ...

Print Tree Output

```
> print(tree)

	 Conditional inference tree with 5 terminal nodes

Response:  actual_productivity
Inputs:  quarter, department, day, team, smv, wip, over_time, incentive, idle_time, no_of_workers, month, day_of_week
Number of observations:  554

1) incentive <= 24; criterion = 1, statistic = 156.79
  2)*  weights = 120
1) incentive > 24
  3) smv <= 30.48; criterion = 1, statistic = 25.39
    4) incentive <= 32; criterion = 0.997, statistic = 13.367
      5) day <= 19; criterion = 0.976, statistic = 9.539
        6)*  weights = 35
      5) day > 19
        7)*  weights = 13
    4) incentive > 32
      8)*  weights = 375
  3) smv > 30.48
    9)*  weights = 11
```

Decision Tree Plot



Explanation On Decision Tree Plot

- In this section, we can see where incentive is equal to or less than 24, there is 120 values classified and the productivity level is more of medium and then high and the low.
- Where incentive is higher than 24 and smv is greater than 30.48 the it falls under Node 9 and it has 11 variables where we have high and medium productivity.
- Where incentive is higher than 24 and smv is less than 30.48 falls under Node 8 and the productivity is very high and we have 375 values in this classification.
- Where incentive is less than 32 and day variable is greater than 19, it falls under Node 7 and have 13 variables where the productivity level is high and medium.
- Where incentive is less than 32 and day variable is less than or equal to 19, we have 35 variables and we have more high productivity than medium.

Prediction

```
> # Prediction
> predict(tree, train)
  [1] High   High   High   High   High   High   High   Medium High   High   High   High   High   High   High   High  
 [17] Medium High   High   High   High   High   High   High   High   High   High   High   High   High   High   High  
 [33] High   High   High   High   High   High   High   Medium High   High   High   High   High   High   High   High  
 [49] High   High   High   High   High   High   High   High   High   High   High   High   High   High   High   High  
 [65] High   High   High   High   High   High   High   High   High   High   High   High   High   High   High   High  
 [81] High   Medium High   Medium High   High   High   High   High   High   High   Medium High   Medium High   High  
 [97] High   High   High   High   High   Medium Medium Medium High   High   High   High   Medium Medium Medium Medium
[113] High   High   High   High   High   High   High   High   Medium Medium High   High   High   High   Medium High  
[129] High   High   Medium High   High   High   High   High   High   High   High   High   Medium High   High   High  
[145] High   High   High   High   High   Medium High   High   High   High   High   High   Medium High   High   High  
[161] High   High   High   High   High   High   Medium High   High   High   High   High   High   High   Medium Medium
[177] High   High   High   High   High   High   High   High   High   High   Medium Medium High   High   High   High  
[193] High   High   High   High   High   Medium High   High   High   High   High   High   High   High   Medium High  
[209] High   High   High   Medium High   High   Medium High   High   High   High   High   High   High   Medium Medium
[225] Medium High   High   High   High   High   High   High   High   Medium High   High   High   High   High   High  
[241] High   High   Medium High   High   High   High   High   High   High   High   High   Medium Medium High   High  
[257] High   High   High   High   High   High   Medium Medium Medium High   High   High   High   High   High   High  
[273] Medium Medium Medium High   High   High   High   High   High   High   Medium High   Medium Medium High   High  
[289] High   High   High   High   Medium Medium Medium High   High   High   High   High   High   High   Medium High  
[305] High   High   High   High   High   High   High   High   Medium Medium High   High   High   High   High   High  
[321] Medium High   High   High   High   High   High   Medium High   High   High   High   High   High   High   High  
[337] High   High   High   Medium Medium Medium High   High   High   Medium Medium Medium Medium Medium Medium Medium
[353] Medium High   Medium High   Medium Medium High   High   Medium Medium High   Medium Medium High   Medium Medium
[369] High   High   Medium Medium High   High   High   Medium Medium High   High   Medium High   High   High   Medium
[385] Medium Medium Medium Medium High   High   Medium High   High   Medium Medium Medium High   High   High   High  
```

Misclassification Error For Train Data

```
# Misclassification error - train data
pred1 <- predict(tree, train)
tab1 <- table(Predicted = pred1, Actual = train$actual_productivity)

# False classication is about 12% (0.1281588) AND Accuracy is 88% which is a good accuracy.

1- sum(diag(tab1))/sum(tab1)
```

```
> pred1 <- predict(tree, train)
> tab1 <- table(Predicted = pred1, Actual = train$actual_productivity)
> 1- sum(diag(tab1))/sum(tab1)
[1] 0.1281588
```

Misclassification Error - Test Data

```
# Misclassification error - test data
pred2 <- predict(tree, test)
tab2 <- table(Predicted = pred2, Actual = test$actual_productivity)

# False classication is about 16% (0.1678832) AND Accuracy is 84% which is a good accuracy.

1- sum(diag(tab2))/sum(tab2)
```

```
> pred2 <- predict(tree, test)
> tab2 <- table(Predicted = pred2, Actual = test$actual_productivity)
> 1- sum(diag(tab2))/sum(tab2)
[1] 0.1678832
```

REGRESSION FORCASTING

SPSS Clean Data View

*Untitled5 [DataSet5] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help

1 : date 01/01/2015 Visible: 15 of 15 Variables

	date	quarter	department	day	team	targeted_productivity	smv	wip	over_time	incentive	idle_time	idle_men	no_of_style_change	no_of_workers	actual_productivity
1	01/01/2015	Quarter1	sweing	Thursday	8	.80	26.16	1108	7080	98	.0	0	0	59.0	.94
2	01/01/2015	Quarter1	sweing	Thursday	11	.80	11.41	968	3660	50	.0	0	0	30.5	.80
3	01/01/2015	Quarter1	sweing	Thursday	12	.80	11.41	968	3660	50	.0	0	0	30.5	.80
4	01/01/2015	Quarter1	sweing	Thursday	6	.80	25.90	1170	1920	50	.0	0	0	56.0	.80
5	01/01/2015	Quarter1	sweing	Thursday	7	.80	25.90	984	6720	38	.0	0	0	56.0	.80
6	01/01/2015	Quarter1	sweing	Thursday	3	.75	28.08	795	6900	45	.0	0	0	57.5	.75
7	01/01/2015	Quarter1	sweing	Thursday	2	.75	19.87	733	6000	34	.0	0	0	55.0	.75
8	01/01/2015	Quarter1	sweing	Thursday	1	.75	28.08	681	6900	45	.0	0	0	57.5	.75
9	01/01/2015	Quarter1	sweing	Thursday	9	.70	28.08	872	6900	44	.0	0	0	57.5	.72
10	01/01/2015	Quarter1	sweing	Thursday	10	.75	19.31	578	6480	45	.0	0	0	54.0	.71
11	01/01/2015	Quarter1	sweing	Thursday	5	.80	11.41	668	3660	50	.0	0	0	30.5	.71
12	01/01/2015	Quarter1	sweing	Thursday	4	.65	23.69	861	7200	0	.0	0	0	60.0	.52
13	01/03/2015	Quarter1	sweing	Saturday	1	.80	28.08	772	6300	50	.0	0	0	56.5	.80
14	01/03/2015	Quarter1	sweing	Saturday	3	.80	28.08	913	6540	50	.0	0	0	54.5	.80
15	01/03/2015	Quarter1	sweing	Saturday	8	.80	26.16	1261	7080	50	.0	0	0	59.0	.80
16	01/03/2015	Quarter1	sweing	Saturday	12	.80	26.16	844	7080	63	.0	0	0	59.0	.80
17	01/03/2015	Quarter1	sweing	Saturday	11	.80	11.61	1005	7080	50	.0	0	0	29.5	.80
18	01/03/2015	Quarter1	sweing	Saturday	5	.80	11.61	659	7080	50	.0	0	0	31.5	.80
19	01/03/2015	Quarter1	sweing	Saturday	6	.80	25.90	1152	6720	50	.0	0	0	56.0	.80
20	01/03/2015	Quarter1	sweing	Saturday	7	.80	25.90	1138	6720	38	.0	0	0	56.0	.80
21	01/03/2015	Quarter1	sweing	Saturday	10	.75	19.31	610	6480	56	.0	0	0	54.0	.79
22	01/03/2015	Quarter1	sweing	Saturday	2	.75	19.87	944	6600	45	.0	0	0	55.0	.75
23	01/03/2015	Quarter1	sweing	Saturday	4	.70	23.69	544	13800	0	.0	0	0	60.0	.70
24	01/03/2015	Quarter1	sweing	Saturday	9	.70	28.08	1072	6900	40	.0	0	0	57.5	.70
25	01/04/2015	Quarter1	sweing	Sunday	6	.80	11.61	539	6975	50	.0	0	0	31.0	.88
26	01/04/2015	Quarter1	sweing	Sunday	9	.80	26.16	1278	7080	60	.0	0	0	59.0	.85
27	01/04/2015	Quarter1	sweing	Sunday	7	.80	25.90	1227	7020	60	.0	0	0	56.5	.85
28	01/04/2015	Quarter1	sweing	Sunday	8	.80	25.90	1039	6780	45	.0	0	0	56.5	.85
29	01/04/2015	Quarter1	sweing	Sunday	4	.80	28.08	878	4260	50	.0	0	0	55.5	.80
30	01/04/2015	Quarter1	sweing	Sunday	1	.80	26.16	1033	7080	63	.0	0	0	59.0	.80
31	01/04/2015	Quarter1	sweing	Sunday	2	.80	28.08	782	6660	50	.0	0	0	55.5	.80
32	01/04/2015	Quarter1	sweing	Sunday	12	.80	11.61	1216	6975	50	.0	0	0	31.0	.80

Data View Variable View

IBM SPSS Statistics Processor is ready Unicode:ON Classic

Step Wise Linear Regression Output

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.804[a]	.647	.646	.09207
2	.808[b]	.653	.652	.09125

a. Predictors: (Constant), incentive

b. Predictors: (Constant), incentive, smv

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	10.690	1	10.690	1260.936	<.001[b]
	Residual	5.841	689	.008		
	Total	16.531	690			
2	Regression	10.803	2	5.401	648.708	<.001[c]
	Residual	5.728	688	.008		
	Total	16.531	690			

a. Dependent Variable: actual_productivity

b. Predictors: (Constant), incentive

c. Predictors: (Constant), incentive, smv

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	.521	.007		78.429	.000
	incentive	.005	.000	.804	35.510	<.001
2	(Constant)	.566	.014		40.989	<.001
	incentive	.004	.000	.797	35.345	<.001
	smv	-.002	.001	-.083	-3.679	<.001

a. Dependent Variable: actual_productivity

- In the next slide, we will be forecasting using the variable smv and incentive to predict our independent variable, productivity.

Forcasting With Selected Model

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	.521	.007		78.429	.000
	incentive	.005	.000	.804	35.510	<.001
2	(Constant)	.566	.014		40.989	<.001
	incentive	.004	.000	.797	35.345	<.001
	smv	-.002	.001	-.083	-3.679	<.001

a. Dependent Variable: actual_productivity

- Formula to forecast with STEPWISE:

  Y = B0 + B1Incentive + B2smv, where B0 is constant.

- So, to forecast we are using extrapolations (80) as incentive and 30.3 for the value of smv in other to forecast percentage of an actual productivity.

  So, Y = .566 + (.004 * 80) + ((-.002) * (30.3) = 0.8254

- So, our forecast Y (Actual Productivity) = 0.8254

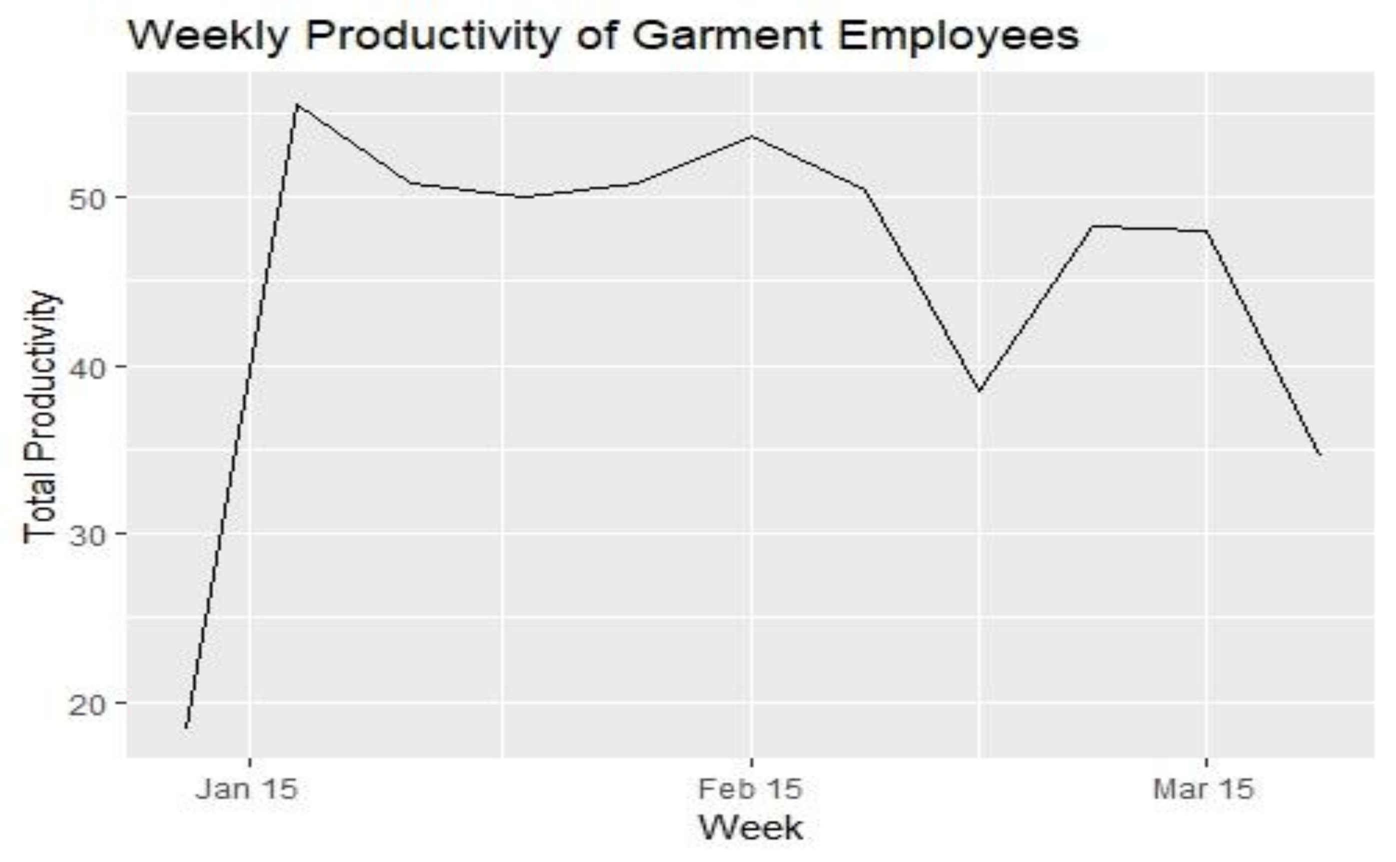
TIME SERIES FORCASTING

Data Manipulation For Time Series

```r
###### Time series ##########

garmentTS <- read.csv("garments_worker_productivity.csv")

garmentTS<- na.omit(garmentTS)

# Convert date column to Date format

garmentTS$date <- ymd(garmentSE$date)

# Aggregate data by week
garments_weekly <- garmentTS %>%
  group_by(week = floor_date(date, "week")) %>%
  summarise(total_productivity = sum(actual_productivity))

# Check the data
view(garments_weekly)

install.packages("forecast", repos="http://cran.us.r-project.org")
library(forecast)
```

Weekly Productivity Of Garment Employees

```
# visualize the weekly time series
ggplot(garments_weekly, aes(x = week, y = total_productivity)) +
  geom_line() +
  scale_x_date(date_breaks = "1 month", date_labels = "%b %y") +
  labs(x = "Week", y = "Total Productivity", title = "Weekly Productivity of Garment Employees")
```



	week	total_productivity
1	2014-12-28	18.40110
2	2015-01-04	55.56048
3	2015-01-11	50.87640
4	2015-01-18	50.03275
5	2015-01-25	50.78599
6	2015-02-01	53.60097
7	2015-02-08	50.37758
8	2015-02-15	38.44246
9	2015-02-22	48.25157
10	2015-03-01	48.03606
11	2015-03-08	34.54564

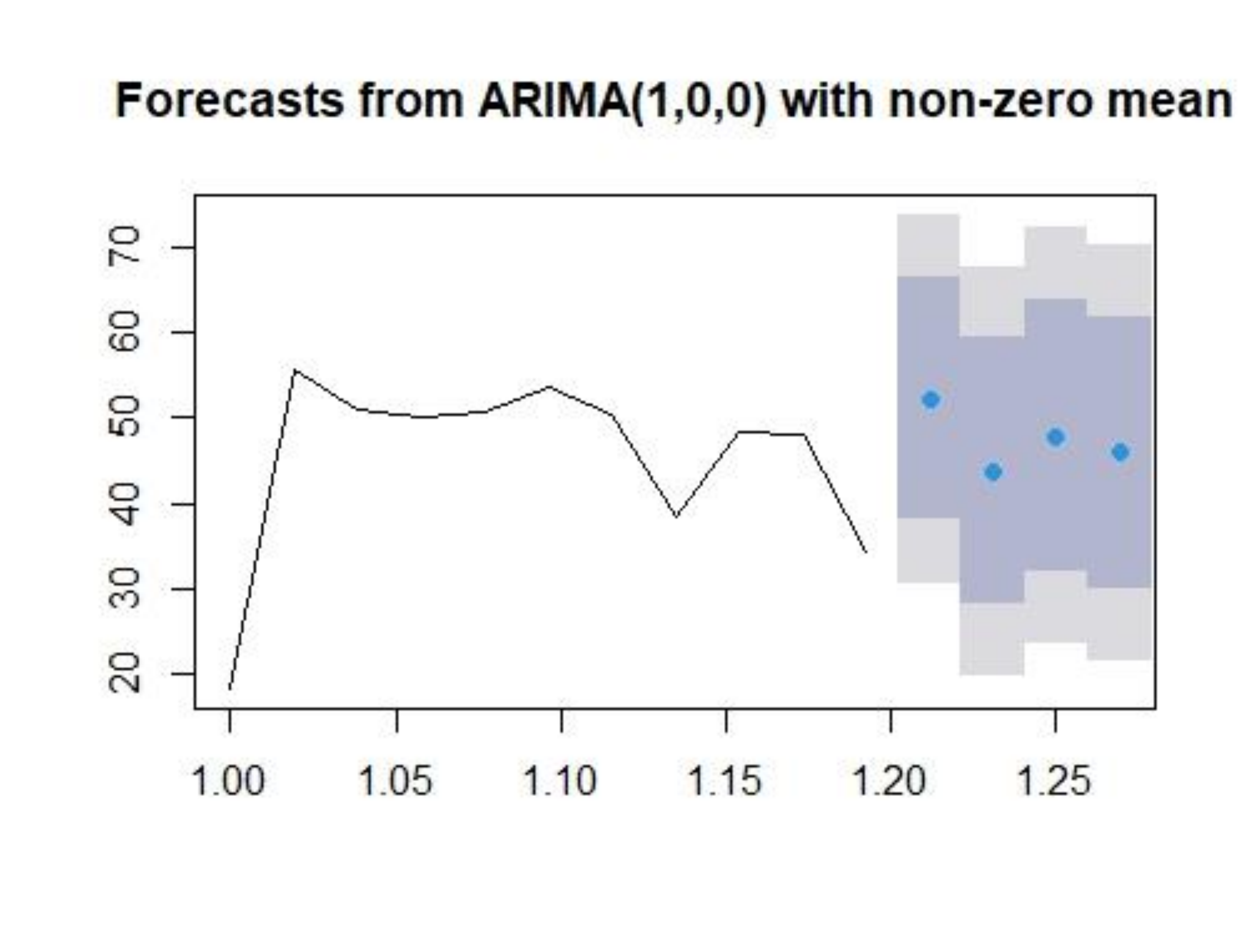
Time Series Model And Weekly Forecast

```r
# Fit a time series model
garments_ts <- ts(garments_weekly$total_productivity, frequency = 52)
fit <- auto.arima(garments_ts)

# Make predictions for the next 4 weeks
forecast <- forecast(fit, h = 4)
print(forecast)
```

```
> print(forecast)
         Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
1.211538       52.25894 38.20429 66.31359 30.76420 73.75367
1.230769       43.80090 28.22620 59.37560 19.98145 67.62035
1.250000       47.83958 31.93863 63.74053 23.52118 72.15798
1.269231       45.91113 29.93673 61.88553 21.48039 70.34186
```

Weekly Time Series Forecast Plot

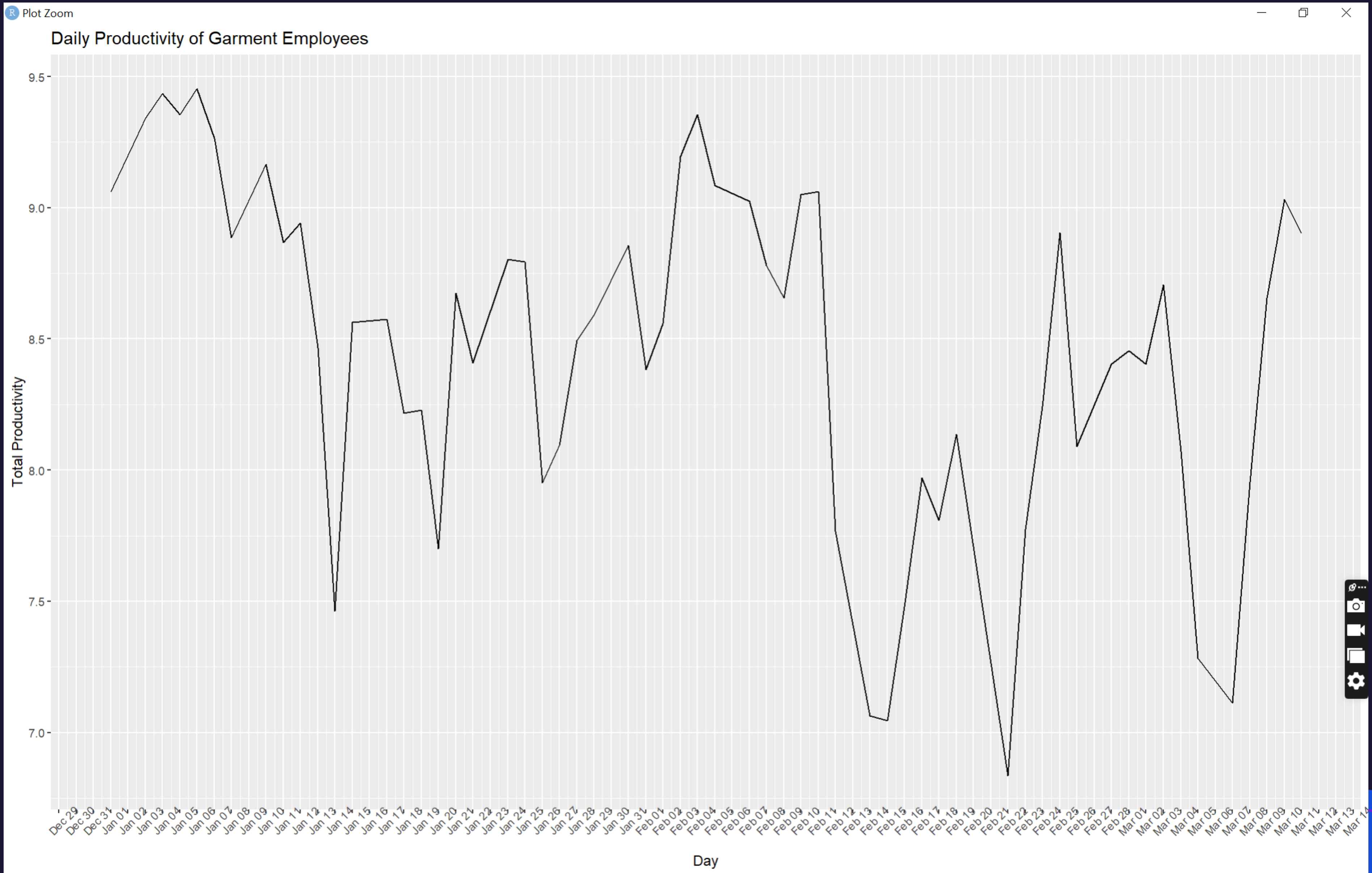
- We can see in the visual the forecast for the next 4 weeks.



Time Series Aggregate And Visualisation For Day

```
# Aggregate data by day
garments_day <- garmentSE %>%
  group_by(date = floor_date(date, "day")) %>%
  summarise(total_productivity = sum(actual_productivity))

# visualize the daily time series

ggplot(garments_day, aes(x = date, y = total_productivity)) +
  geom_line() +
  scale_x_date(date_breaks = "day", date_labels = "%b %d") +
  theme(axis.text.x = element_text(angle = 45))+
  labs(x = "Day", y = "Total Productivity", title = "Daily Productivity of Garment Employees")
```

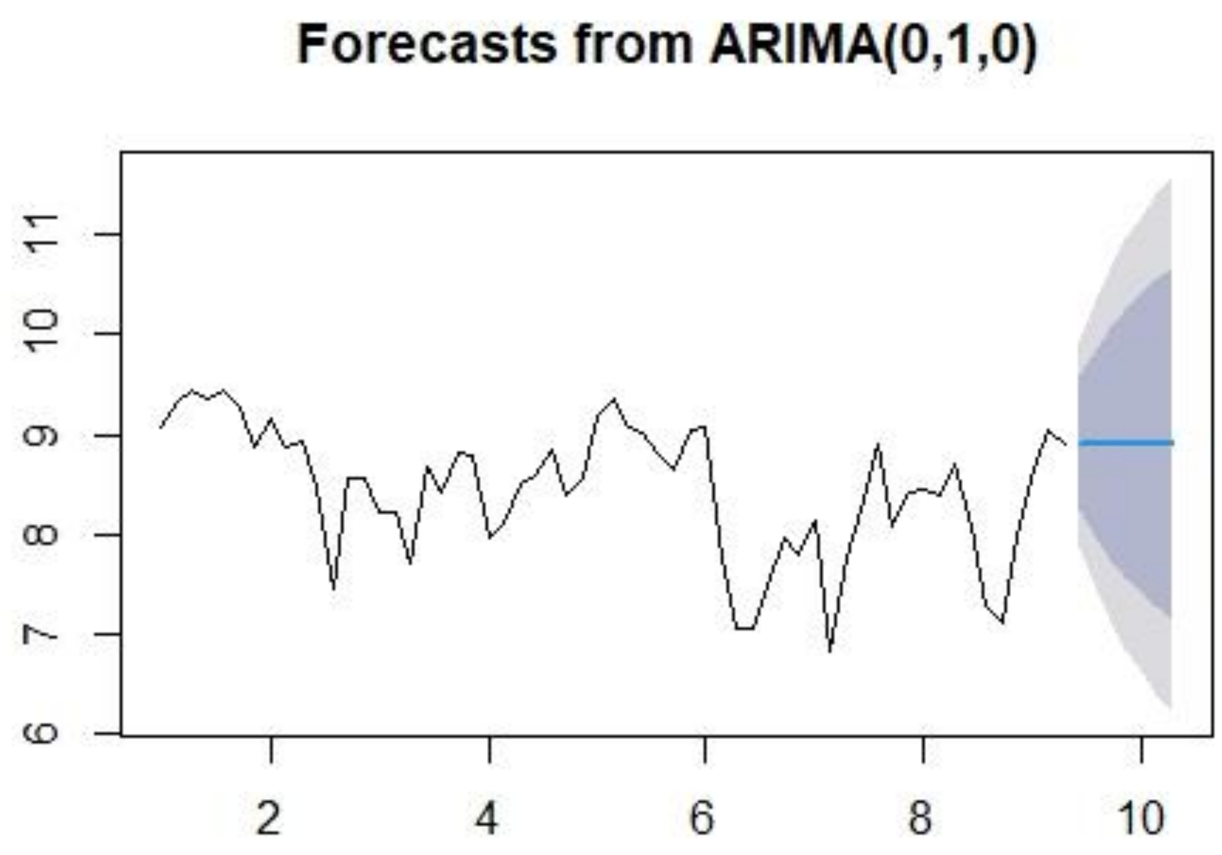
Daily Productivity Of Garment Employees



Daily Time Series Forecast Plot

```r
# Fit a time series model
garments_day_ts <- ts(garments_day$total_productivity, frequency = 7)

fitday <- auto.arima(garments_day_ts)

forecastday <- forecast(fitday, h = 7)
print(forecastday)

plot(forecastday)
```

- We can see in the image the forecast of productivity for the next 7 day.



Recommendation and Conclusion

- We can see from the decision tree that where the higher incentive is given to the employees, there's high productivity. So, incentives should be introduced to employees for motivation at work.
- Also, the first 19 days of the month show high productivity days from the staff. So, these days should be the main target for employers to encourage productivity from employees.
- With the regression analysis, we could see that there's high significance between incentive, smv and productivity.
- For the forecast with regression, we used an extrapolation where incentive given to the employees was 80, the turn out for the actual productivity we got was high, which justifies our recommendation that incentive should be exploited by the employers.

Professional, Ethical, And Legal Issues That Can Affect The Productivity Of Employees

- Harassment and Discrimination
- Poor Working Conditions
- Lack of Training and Development
- Conflict with Co-Workers
- Burnout
- Employees may feel confused and uncomfortable about their employment as a result of ethical quandaries such conflicts of interest or dubious corporate practises.
- Legal problems, such as labour code infractions or contract breaches, might result in litigation and harm the company's reputation.

References

- Imran, A. A., Rahim, M. S., & Ahmed, T. (2021). Mining the productivity data of the garment industry. *International Journal of Business Intelligence and Data Mining*, *19*(3), 319. https://doi.org/10.1504/ijbidm.2021.118183
- Balla, I., Rahayu, S., & Purnama, J. J. (2021, March 15). GARMENT EMPLOYEE PRODUCTIVITY PREDICTION USING RANDOM FOREST. *Jurnal Techno Nusa Mandiri*, *18*(1), 49–54. https://doi.org/10.33480/techno.v18i1.2210.
- *Global Productivity: Trends, Drivers, and Policies*. (n.d.). World Bank. Retrieved from https://www.worldbank.org/en/research/publication/global-productivity
- *UCI Machine Learning Repository: Productivity Prediction of Garment Employees Data Set*. (n.d.). UCI Machine Learning Repository: Productivity Prediction of Garment Employees Data Set. https://archive.ics.uci.edu/ml/datasets/Productivity+Prediction+of+Garment+Employees