setwd("/home/data/Forum\_Jigsaw\_Challenge/JigsawChallenge\_and\_contest\_Jan\_2018")

getwd()

#read the data

Maintenance <- read.csv("P.Maintenance\_Project\_dataset.csv",header=T,sep=",")

head(Maintenance)

#dimensions of dataset

dim(Maintenance)

colnames(Maintenance)<-c("serial\_no","lifetime","broken","pressureInd\_1","pressureInd\_2","pressureInd\_3","team","provider")

#structure of the dataset

str(Maintenance)

head(Maintenance)

tail(Maintenance)

#data exploration analysis

#ccheck if there are any missing values in the dataset

colSums(is.na(Maintenance))

#analysis of data

table(Maintenance$team)

table(Maintenance$provider)

table(Maintenance$broken)

table(Maintenance$lifetime)

library("ggplot2", lib.loc="/usr/local/lib/R/site-library")

library("RColorBrewer", lib.loc="/usr/local/lib/R/site-library")

par(mfrow=c(2,2))

#analysis of plot through visualisations

hist(Maintenance$pressureInd\_1)

hist(Maintenance$pressureInd\_2)

hist(Maintenance$pressureInd\_3)

dev.off()

qplot(Maintenance$team)

qplot(Maintenance$provider)

par(mfrow=c(2,2))

# check if there are any outliers in the dataset

boxplot(Maintenance$pressureInd\_1, ylab="pressureInd1",col= heat.colors(3))

boxplot(Maintenance$pressureInd\_2, ylab="pressureInd2", col= "blue")

boxplot(Maintenance$pressureInd\_3, ylab="pressureInd3", col = "green")

# detailed statistical analysis of the data

summary(Maintenance)

str(Maintenance)

# check if the linear regression model is fit for the data

linearegmodel = lm(lifetime~ .-broken,data = Maintenance)

summary(linearegmodel)# doesnt give good insights and hence try any other model for predictive maintenance

#using survival analysis for predictive maintenance

install.packages("survival")

library(survival)

dependentvars = Surv(Maintenance$lifetime,Maintenance$broken)

# using survreg function of survival analysis

survreg = survreg(dependentvars~ pressureInd\_1+pressureInd\_2+pressureInd\_3+team+provider,dist = 'gaussian',data = Maintenance)

# statistical analysis after applying survreg function

summary(survreg)

#predict function of the survival analysis model

Ebreak<-predict(survreg, newdata = Maintenance, type = 'quantile',p=.5)

# assign column names to the predicted dataframe

Forecast=data.frame(Ebreak)

Forecast$lifetime=Maintenance$lifetime

Forecast$broken=Maintenance$broken

Forecast$provider=Maintenance$provider

Forecast$team=Maintenance$team

# calculation of the remaining lifetime of the machines

Forecast$remainingLT=Forecast$Ebreak-Forecast$lifetime

View(Forecast)

# order the dataset in incresion order of remaining Lifetime

Forecast=Forecast[order(Forecast$remainingLT),]

# taking only those machines in the dataset which are working and predict the future of breakdown of those machines.

Priority=Forecast[Forecast$broken==0,]

View(Priority)

#visualization analysis

par(mfrow=c(1,1))

boxplot(Priority$lifetime,Priority$remainingLT,col=heat.colors(3))

Maintenance\_broken <- maintenance %>% filter(broken == 1)

Maintenance\_working <- maintenance %>% filter(broken == 0)

boxplot(lifetime~broken,data=Maintenance\_broken, main="Borken machines", xlab="", ylab="", col="#357EC7")

boxplot(lifetime~broken,data=Maintenance\_working, main="active machines", xlab="", ylab="", col="#357EC7")

boxplot(lifetime~team,data=maintenance\_broken, main="Per team", xlab="", ylab="",col="#357EC7")

boxplot(lifetime~provider,data=maintenance\_broken, main="Per provider", xlab="", ylab="",col="#357EC7")

# structure of the priority or active machines dataset

str(Priority)

# classify the remainingLt of machines into urgent, good or medium class.

Priority$class <- cut(Priority$remainingLT, c(-10,1,4,1000))

levels(Priority$class) <- c('Urgent', 'Medium', 'good')

#summary of the priority dataset model

summary(Priority)

str(Priority)

# visualisation of the insights of the priority model

qplot(Priority$class,col="heat.colors(3)",fill = "")

qplot(Priority$provider,col="heat.colors(3)",fill = "")

qplot(Priority$team,col="heat.colors(3)",fill = "")

ggplot(Priority, aes(x=provider, fill = "", color = "red"))+ geom\_bar(aes(fill=class)) + labs(title= "Preventive Maintenance Provider Prediction")

ggplot(Priority, aes(x=team, fill = "", color = "orange"))+ geom\_bar(aes(fill=class)) + labs(title= "Preventive Maintenance Team Prediction")

View(Priority)

## Analytical solution

#The solution will predict \*\*which machine will break next\*\* and propose some reasons about

#\*\*why these machines have different lifetimes.\*\* A survival analysis is a good choice as we can visualise the machine's lifetime very easily.

## Assessing Feasibility

#The business has data on each machine for the last pasts years and they are able to provide some more

#information such as the team that used it and its provider. These information seems sufficient to start an analysis.# Data exploration

## Data quality report

#As we summarise the data, we can see that the business have used 90000 machines.

#Machine have an average lifetime of 55 weeks, with some brand new machines and others that are running since almost two years.

#In our dataset almost \*\*40 % of the machines have being broken in the past two years.\*\*

# Survival analysis {.tabset}

#As we have seen above, all variables tell us something about when a machine will break.

#We decide to use them all into our survival analysis model.

#We create a model using the \*\*gaussian method\*\* and use directly the maintenance dataset presented above.

## Predictions

#Using our survival analysis, we can now predict which machine will break next and therefore prioritise the maintenance on these machines to avoid the supply chain to stop.

#We have given the classification as urgent, medium and good to machines so that the client can look and maintain the machines accordingly.

### Management

#The \*\*second action is to show these figures to team C and provider 3 and monitor the improvement.\*\*

#Now the management could focus on the enhancement of the machines' lifetimes and see what are the best methods.

#Of course we should interview the workers as we could have new insights on how we could push the lifetime of each machine further.