Predictive Maintenance Data Challenge

Team 2
Cornelia, Mariana, Marie, Konstantinos, Matt

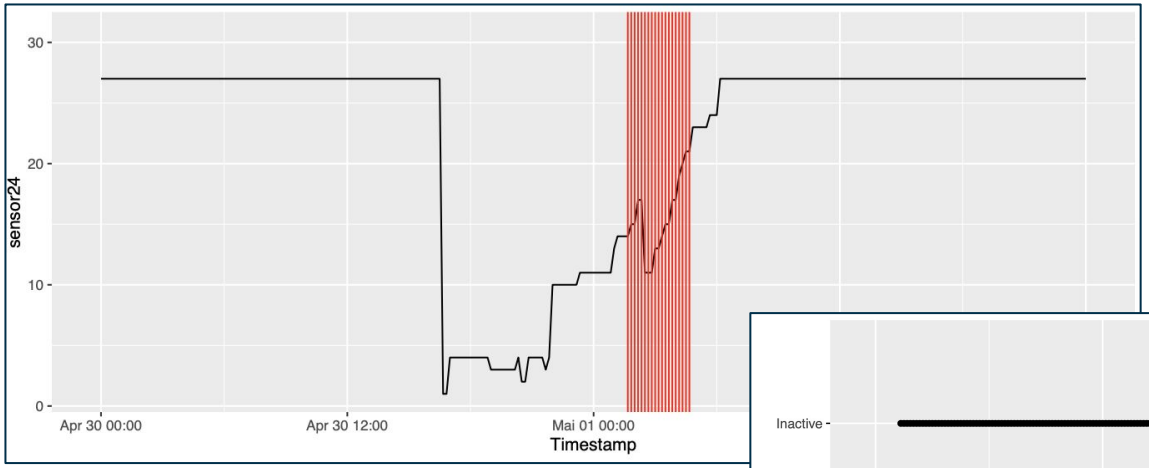
Introduction

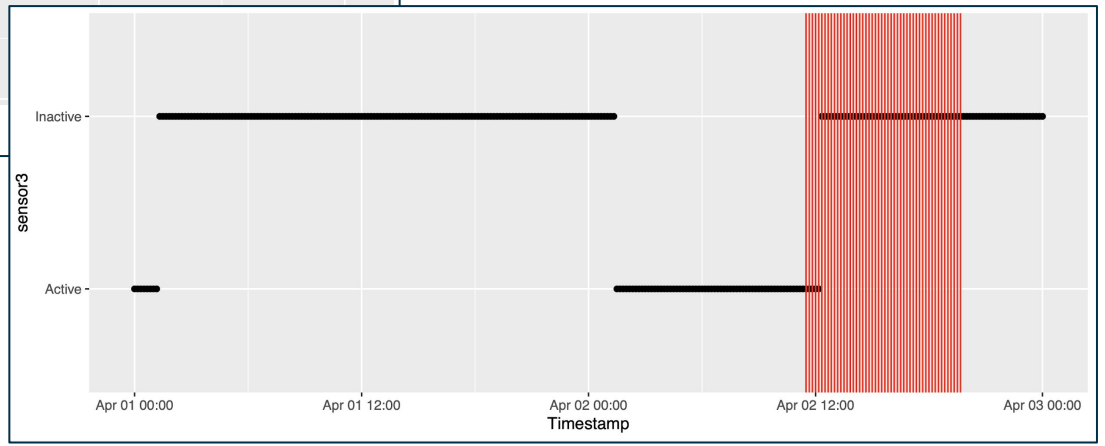
- Business-critical machines of Swiss pharmaceutical company are maintained on a fixed schedule
- Need for predictive maintenance to avoid unnecessary losses
- Dataset contains data from 48 sensors of one machine and actual 18 failures data (0: working, 1: failure), captured at 10 min intervals
- Goal - **Predict failure** of the machine to minimize technical interventions

Solving the Problem

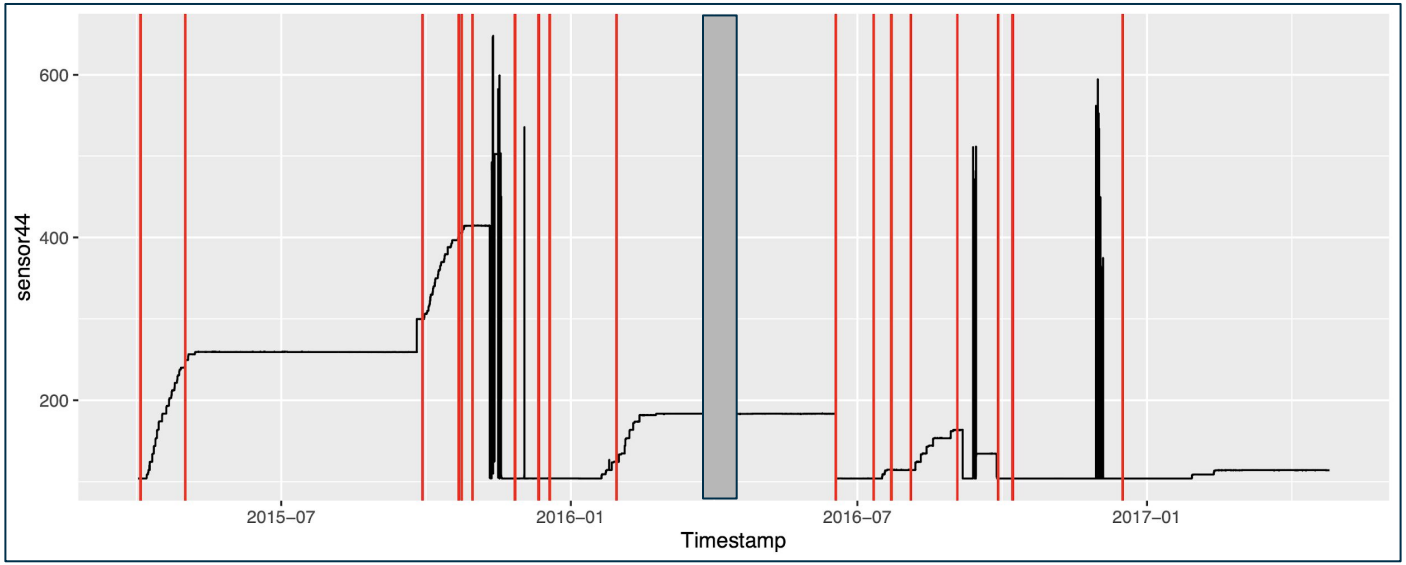
- **Clean data** - remove “bad input” data from numerical and categorical variables
- **Reduce dimensionality**
  - Remove insignificant variables with wilcox and chi-squared tests
  - Remove highly correlated variables after checking for correlation
- **Compute central tendencies** - come up with thresholds to identify faults
- Identify sensors involved in faults - using thresholds to create predictive model

Visualizing Sensors Fault Regions

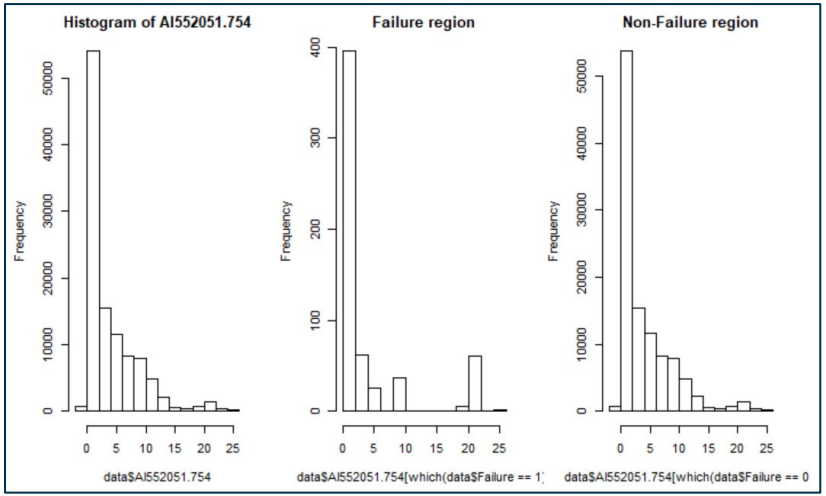




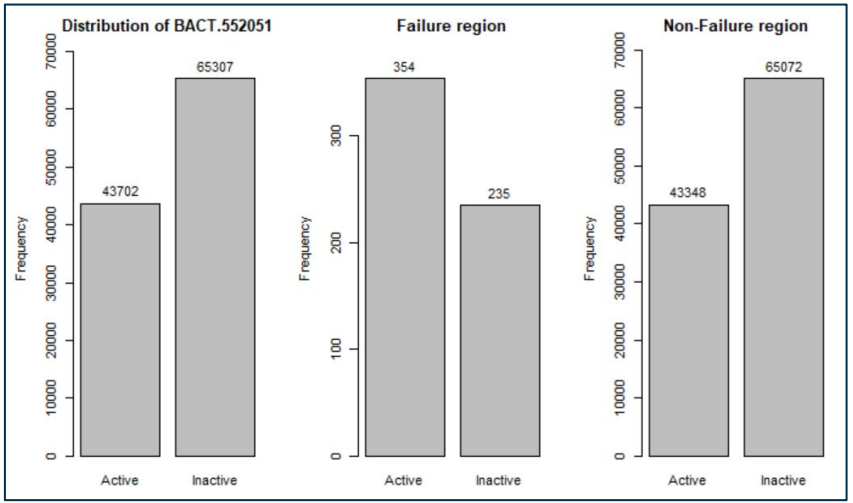
Analyzing Sensors Fault Regions



Distribution of Sensors Data



Numeric



Categorical

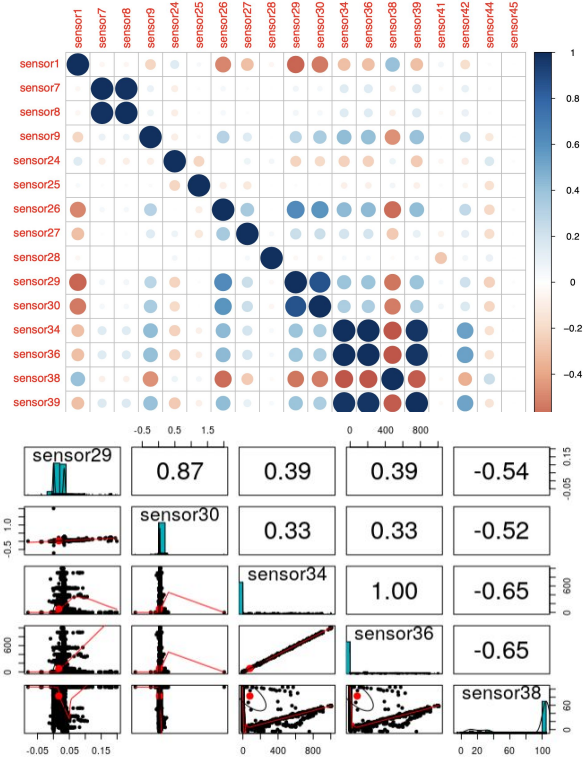
Performing Statistical Tests

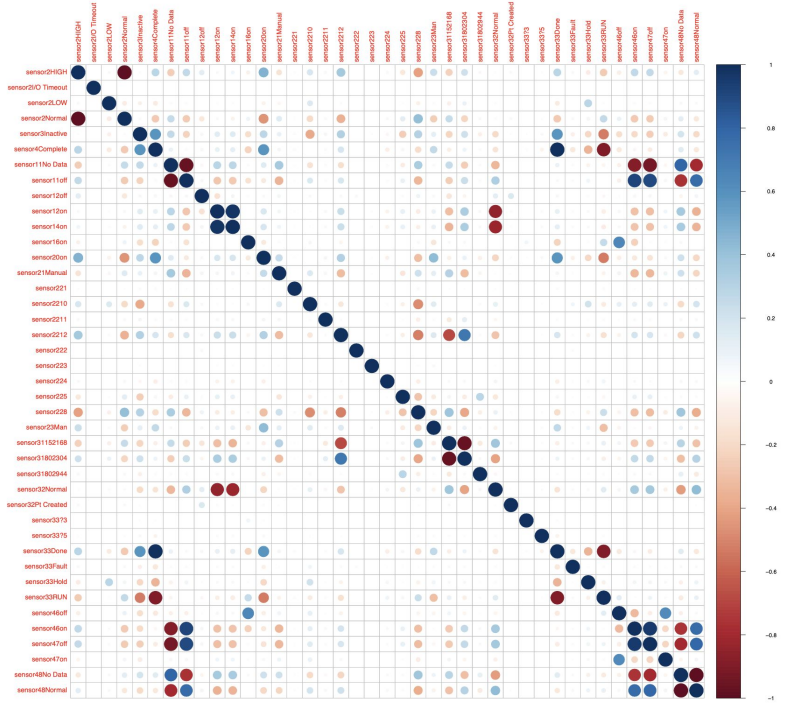
Type of Data: **Categorical** **Numerical**

Non-parametric

Tests used: **Chi-squared** **Wilcox**

Visualizing Sensors Correlation

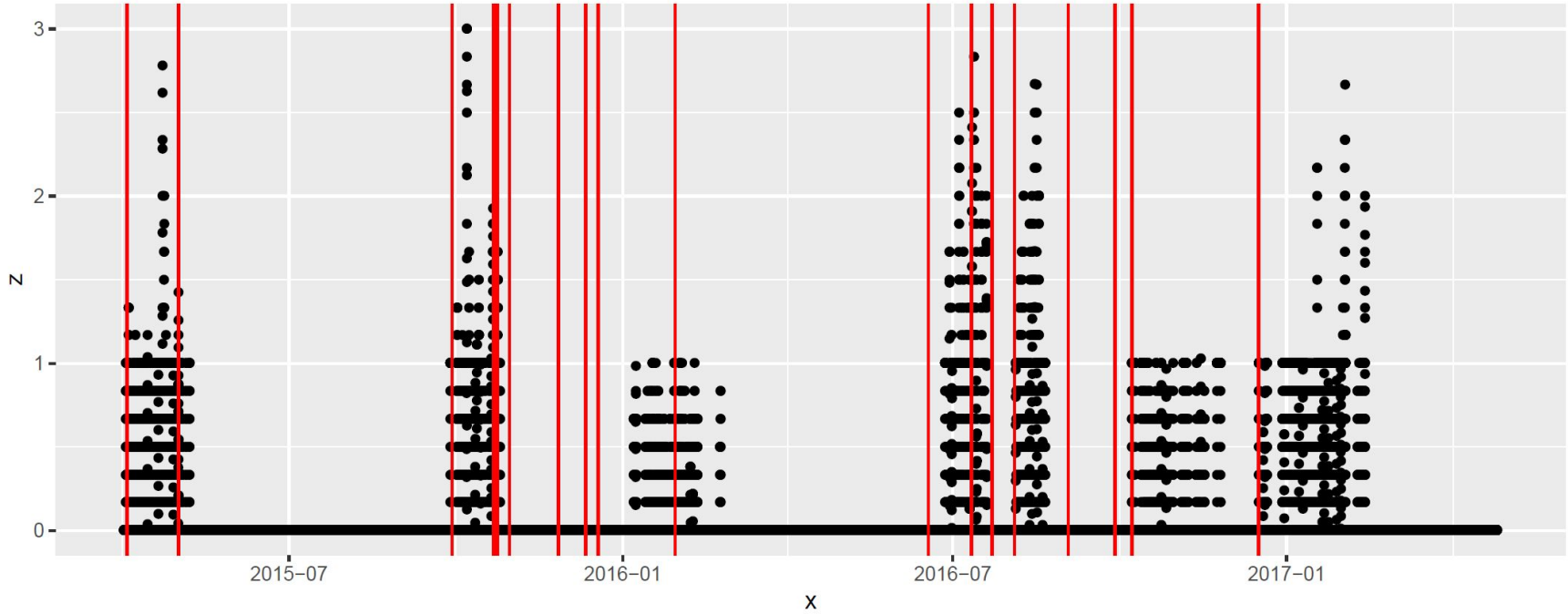




Central Tendencies - Identifying Faults

NAME	NORMAL	PRE_FAULT	FAULT
[1] "AI552051.754"	"33284.5111111111"	"39073.2444444444"	"3.7487716624881"
[1] "FI552051_718"	"4.26944444444444"	"7.77222222222222"	"0.0304368058894558"
[1] "FIC552051_718"	"65.1694444444445"	"75.6916666666667"	"28.8653327782143"
[1] "FQI552051.718"	"413.011111111111"	"752.322222222222"	"25.6038650358469"
[1] "PHASE.552051"	"16.5083333333333"	"15.35"	"19.8537414965986"
[1] "PHTIME.552051"	"24436.7611111111"	"24714.5861111111"	"330.490371014189"
[1] "PI552051.641"	"27078.6833333333"	"29227.2583333333"	"0.0162700541880952"
[1] "PI552051.642"	"14706.725"	"12411.7805555556"	"0.0532346108639456"
[1] "PI552051.645"	"7603.99444444444"	"6694.50555555556"	"0.00906095916496599"
[1] "PI552051.646"	"32318.2638888889"	"32063.7833333333"	"0.0213336453348639"
[1] "PI552051.651"	"11318.1333333333"	"11857.6888888889"	"0.0331917325187075"
[1] "SIC552051.801"	"24509.075"	"25042.5277777778"	"85.8170989730119"
[1] "WI552051.675"	"37367.2583333333"	"33723.3333333333"	"13683333333333896"

Central Tendencies - Identifying Faults



Sensor: CYCLE.552051

Sensors Involved in Faults

```r
fit <- glm(Failpred ~ sensor1 + sensor5 + sensor9 + sensor24 + sensor25 + sensor26 + sensor29 + sensor38 + sensor41 + sensor44, family =
binomial (link = "logit"), data=dfusefuldata2)
summary(fit)
```

```
Call:
glm(formula = Failpred ~ sensor1 + sensor5 + sensor9 + sensor24 +
    sensor25 + sensor26 + sensor29 + sensor38 + sensor41 + sensor44,
    family = binomial(link = "logit"), data = dfusefuldata2)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.1788  -0.0871  -0.0591  -0.0491   3.7636

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -7.976e+00  3.323e-01 -23.999  < 2e-16 ***
sensor1      9.375e-02  1.163e-02   8.059 7.70e-16 ***
sensor5      6.318e-01  1.714e-01   3.687 0.000227 ***
sensor9      3.744e-03  8.320e-04   4.499 6.81e-06 ***
sensor24    -4.875e-02  5.828e-03  -8.364  < 2e-16 ***
sensor25     4.120e-04  5.975e-05   6.896 5.36e-12 ***
sensor26     2.245e+01  3.825e+00   5.870 4.37e-09 ***
sensor29     2.563e+01  2.136e+00  12.000  < 2e-16 ***
sensor38     6.165e-03  2.206e-03   2.794 0.005201 **
sensor41     3.159e-02  9.307e-03   3.394 0.000688 ***
sensor44     3.626e-03  5.530e-04   6.557 5.50e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 4625.5  on 108973  degrees of freedom
Residual deviance: 4321.8  on 108963  degrees of freedom
  (35 observations deleted due to missingness)
AIC: 4343.8

Number of Fisher Scoring iterations: 9
```

Bonus

Logistic Regression Model