



# Real-time Estimations and Online Learning for Industrial Assets at Total

Presented by **SPINNATO Juliette**



# AGENDA

- ▶ TOTAL: our Company and our data-driven philosophy
- ▶ From a Business use case to a data science “Proof Of Concept”
- ▶ Online learning for our industrial assets
- ▶ Conclusion and Next steps



# TOTAL: our Company and our data-driven philosophy

# TOTAL: Committed to Better Energy

Total is the world's **4<sup>th</sup>**-ranked oil  
and gas company<sup>1</sup>

and a global **leader** in solar energy  
with SunPower.

WITH OPERATIONS IN MORE THAN

**130 COUNTRIES,**

we have over **96,000 employees**  
who are fully committed to better energy.

<sup>1</sup> Based on market capitalization in U.S. dollars at December 31, 2015

# TOTAL: Committed to Better Energy

## TOTAL'S ACTIVITIES

### EXPLORE AND PRODUCE

- 1 OIL AND GAS
- 2 SOLAR
- 3 BIOMASS

### TRANSFORM AND DEVELOP

- 4 SPECIALTY CHEMICALS
- 5 POLYMERS
- 6 REFINING - PETROCHEMICALS

### SHIP AND MARKET

- 7 TRADING - SHIPPING
- 8 MARKETING AND SERVICES



# TOTAL: Our Data-Driven philosophy

## A DATA-DRIVEN ACTIVITY WITHIN TOTAL

- ▶ Since 2015, Total promotes internally **a strong data culture**
- ▶ To move **towards a data-driven and digitalized industry**
- ▶ Creation of **data analytics centers** within our company
  - Realization of several Proof of Concepts (PoC) to demonstrate **data science added-value for our Businesses**
  - Industrialization of the best PoCs into long-term solutions



# From a Business use case to a data science POC

# USE CASE



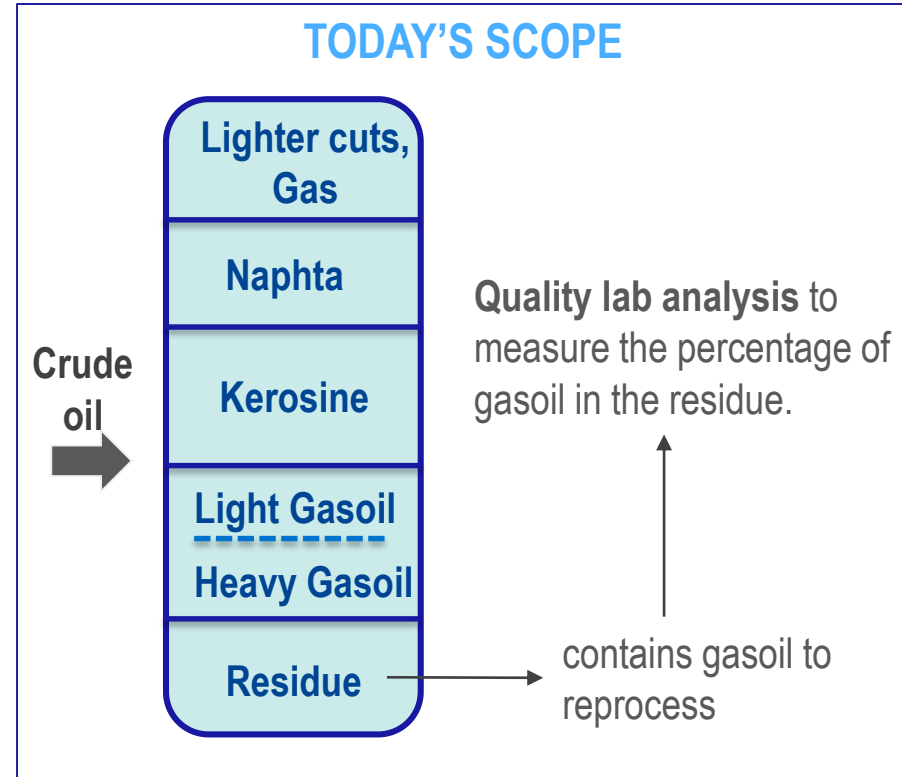
## BUSINESS OBJECTIVE

Increase lab & units efficiency using advanced analytics

- ▶ To reduce the amount of samples processed that bring little or no added value
- ▶ To help the process experts to improve units outputs



**A proof of concept:** estimating crude distillation quality for Antwerp refinery





# A DATA SCIENCE PROOF OF CONCEPT

## SUCCESSFUL METHODOLOGY

- ▶ **A 5 months project** following an AGILE method
- ▶ **A multi-disciplinary team** including Business experts, data scientist, and data owner

## DATA SOURCE

**80 relevant PI Tags™ identified and extracted** from the PI System®

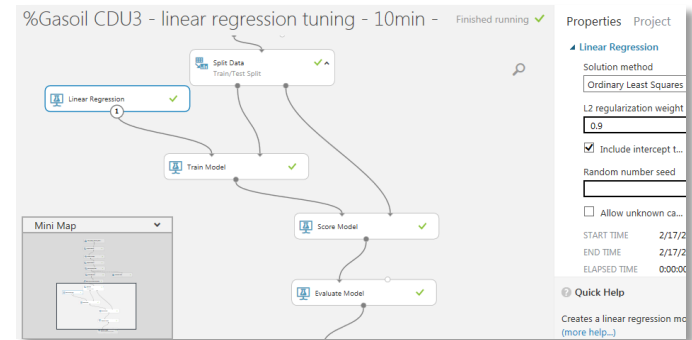
A one shot study: **data extraction using PI DataLink®**

## MACHINE LEARNING MODELING

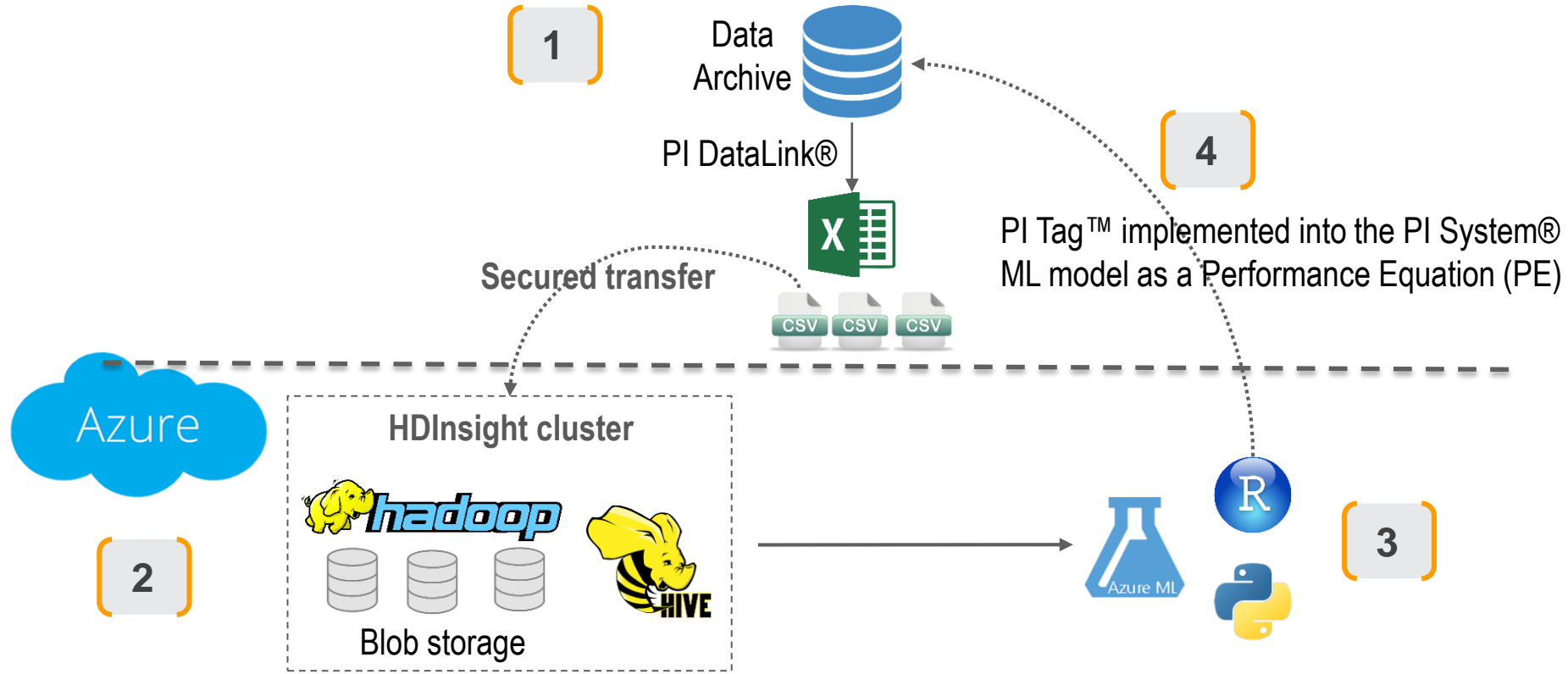
- ▶ **White box machine learning** algorithm
- ▶ **Highly correlated variables** (flow, pressures, temperatures, ...)

## Linear Ridge regression

Azure ML Studio:



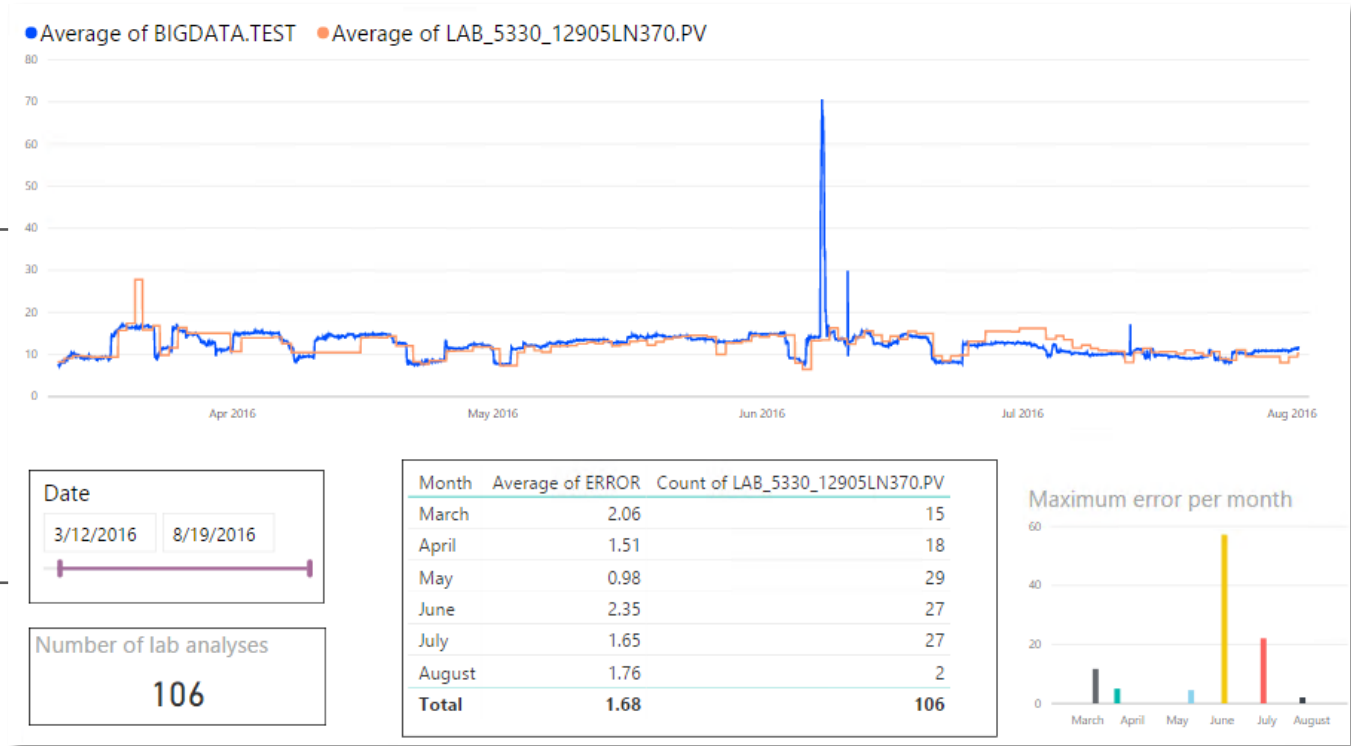
# ARCHITECTURE & WORKFLOW



# RESULTS

## REAL-TIME ESTIMATION AND MODEL FOLLOW-UP

Model estimation  
versus  
lab analysis



KPIs to follow the  
performances

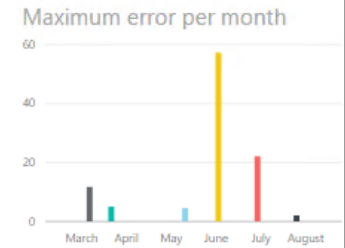
Date

3/12/2016 8/19/2016

Number of lab analyses

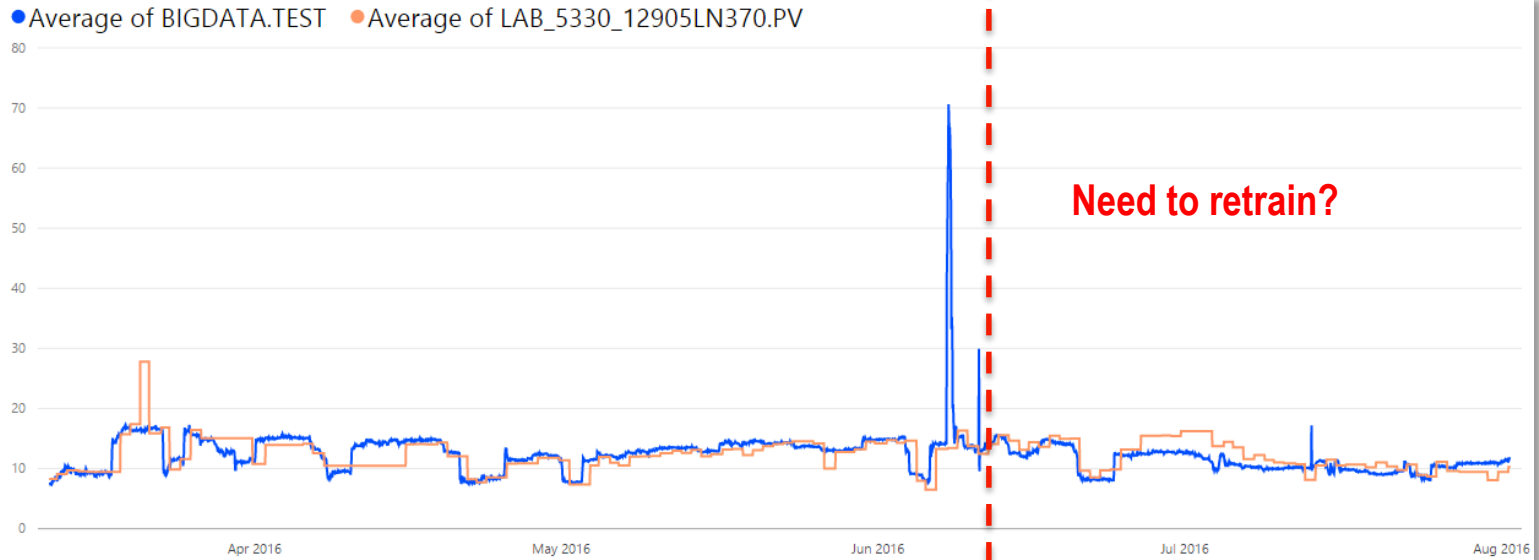
106

Month	Average of ERROR	Count of LAB_5330_12905LN370.PV
March	2.06	15
April	1.51	18
May	0.98	29
June	2.35	27
July	1.65	27
August	1.76	2
<b>Total</b>	<b>1.68</b>	<b>106</b>



# RESULTS

## MODEL SUSTAINABILITY



- ▶ When do we decide that the model is no longer up-to-date?
- ▶ When do we decide to retrain the model?

# CHALLENGES

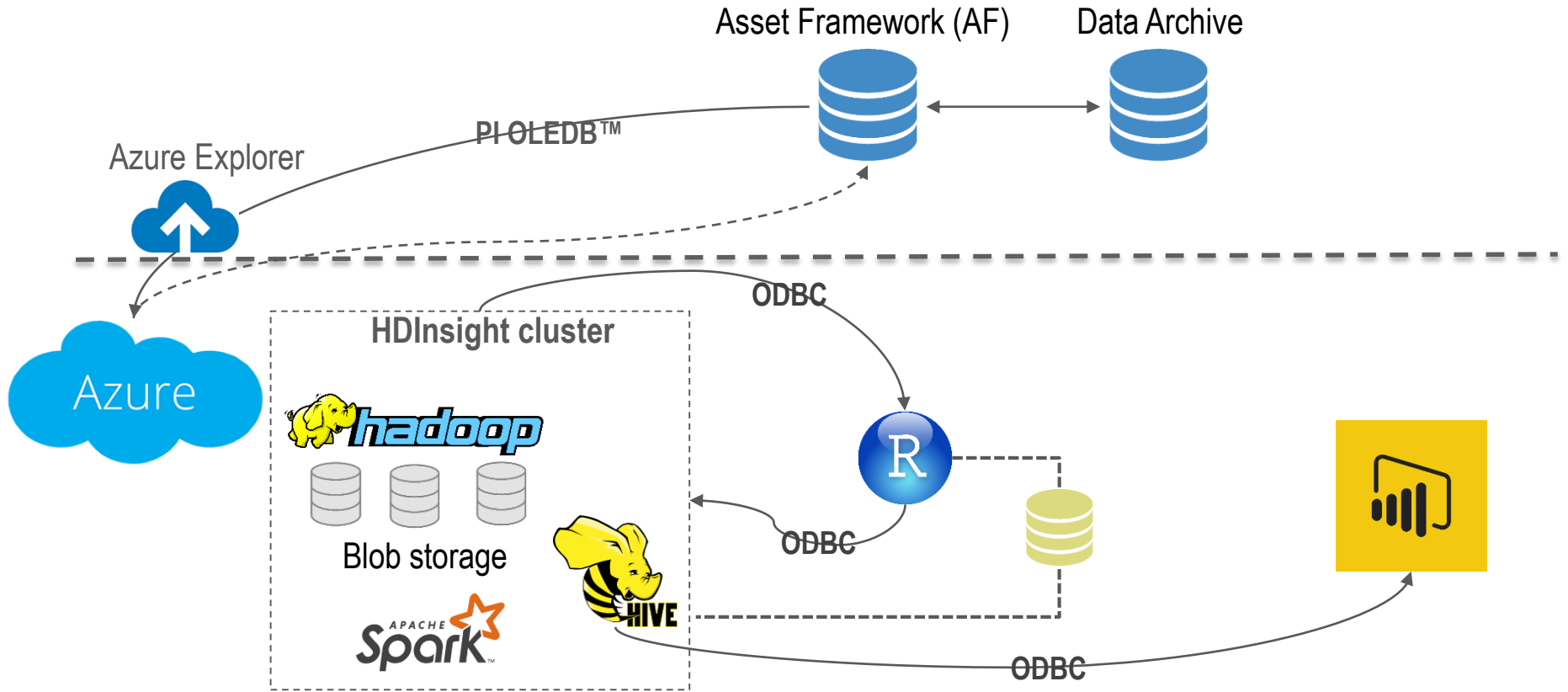
## Maintaining a ML model implemented in a PI System® : **KEY QUESTIONS**

- ▶ How do we **decide to retrain** the model: detect changes made on the industrial asset, launch retraining based on a statistical and/or business criteria
- ▶ What is **the best dataset** to take into account?
- ▶ How **to keep the historic** of our online learning?

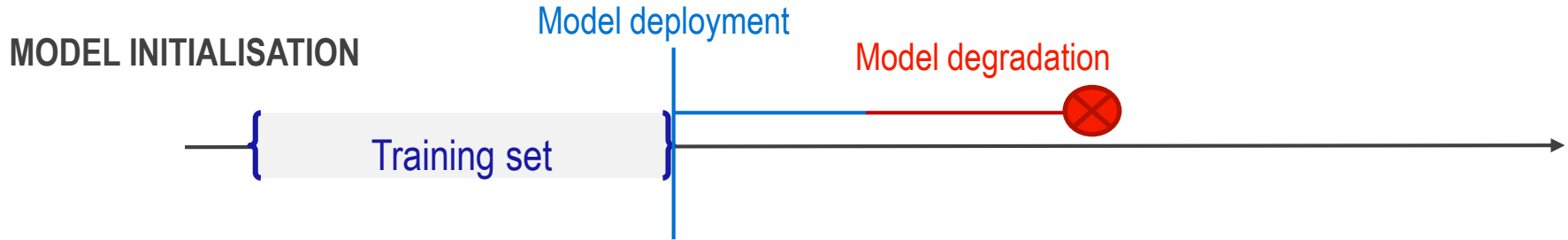


# Online learning for our industrial assets

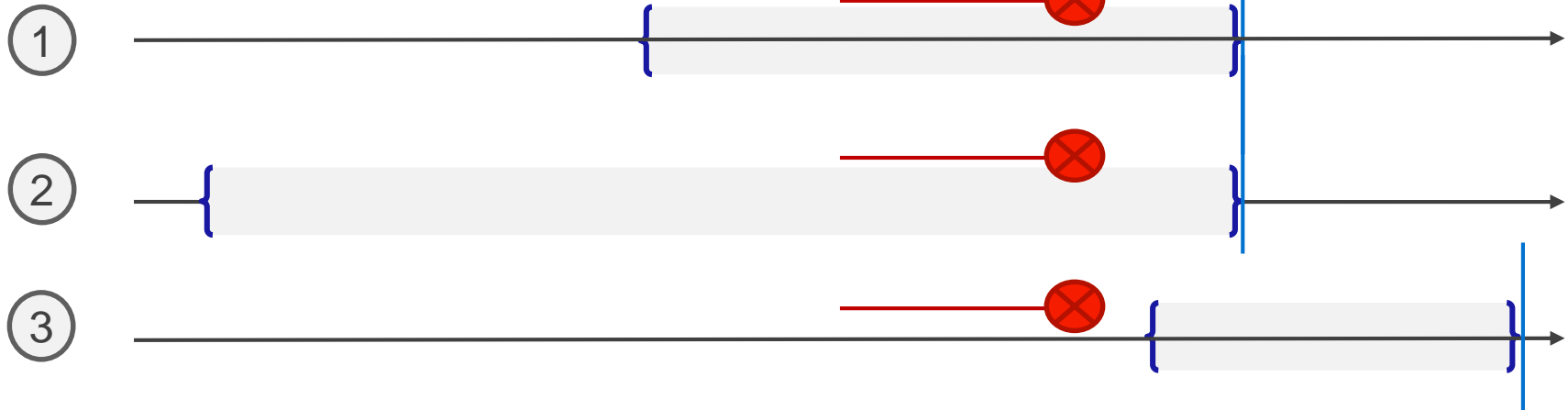
# TARGET ARCHITECTURE



# ONLINE LEARNING: Dataset(s) & model(s)



## ONLINE LEARNING – Scenario examples





# ONLINE LEARNING: Expert System

How to...

- ▶ ...automatically define the different datasets and
- ▶ ...evaluate models



## ***Start\_Control\_Algorithm ()***

Inter\_V: The time interval controls tags = 30 minutes

Inter\_M: Variable check interval = 10 minutes

Perf\_model = F (Mean error, accuracy, actual deviation)

Perf\_threshold : Performance threshold set by Business

### ***For each Inter\_M***

***If ( Perf\_model < Seuil\_Perf ) then***

*Create N datasets*

*Dataset = {old dataset, Delta-2, Delta-1, Delta 1, Delta2 }*

*Modeles\_test <- Create 2^N prediction models*

*Model <- Max (Modeles\_test ) # the best model of Modeles\_test*

***End if***

***End for***

***End\_Control\_Algorithm***

# ONLINE LEARNING: First results

## MODELING CHOICES

- ▶ Elastic-net algorithm: Variables selection and **optimized linear regression**



R library : “*glmnet: Lasso and Elastic-Net Regularized Generalized Linear Models*”

Ref: [T. Hastie et al. , “*The Elements of Statistical Learning*”, Springer Series in Statistics, 2<sup>nd</sup> Edition, Springer, 2009]

## MODELS EVALUATION

- ▶ **Performance evaluation:** Mean-absolute error (MAE)
- ▶ **Business evaluation:**

Model equation highlights the main contributions to estimate the target

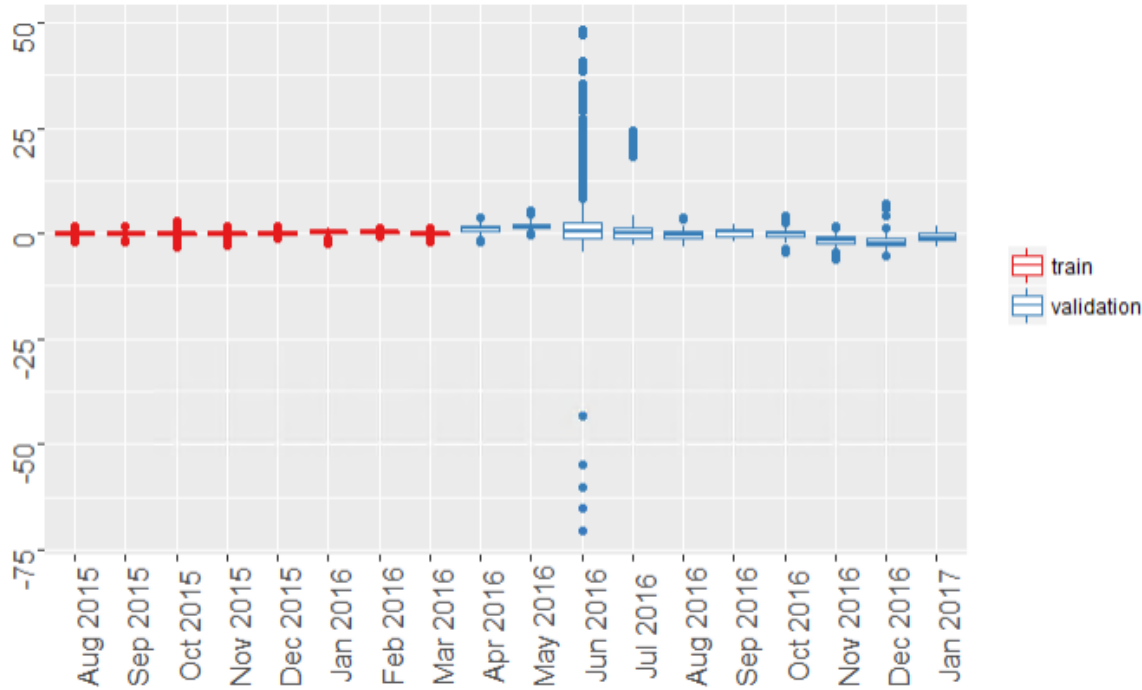
Descriptive statistics



# ONLINE LEARNING: First results

## R RESULTS - Example

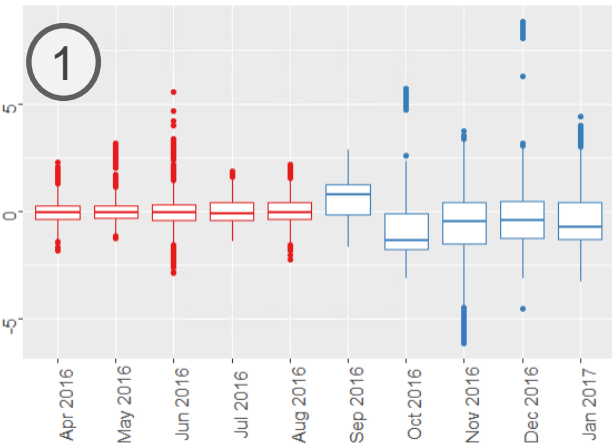
## MODEL INITIALISATION – Estimation Error



	Training	Validation
<b>MAE</b>	0.36	2.08

# ONLINE LEARNING: First results

## R RESULTS - Example

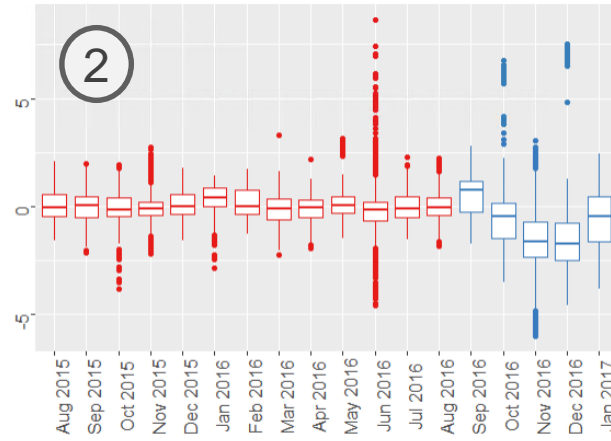


Training Validation

MAE

1.32

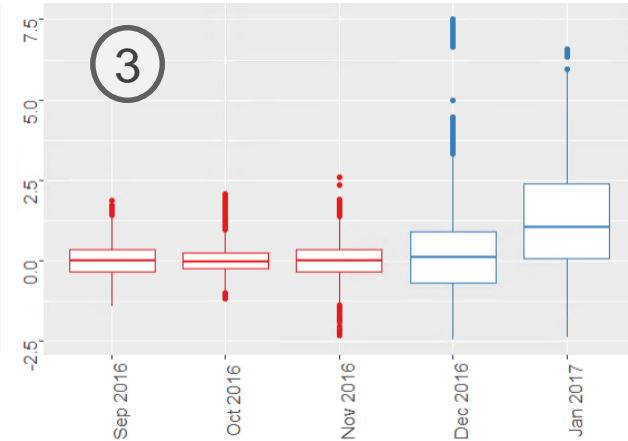
0.93



Training Validation

0.53

1.38



Training Validation

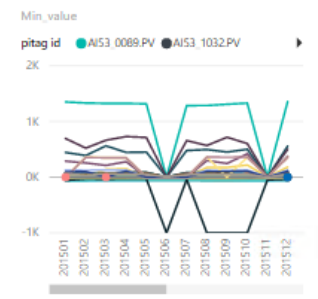
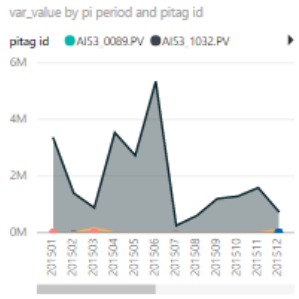
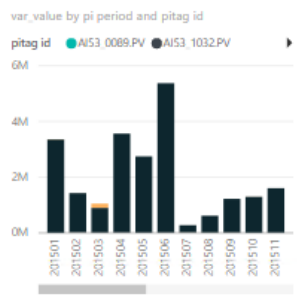
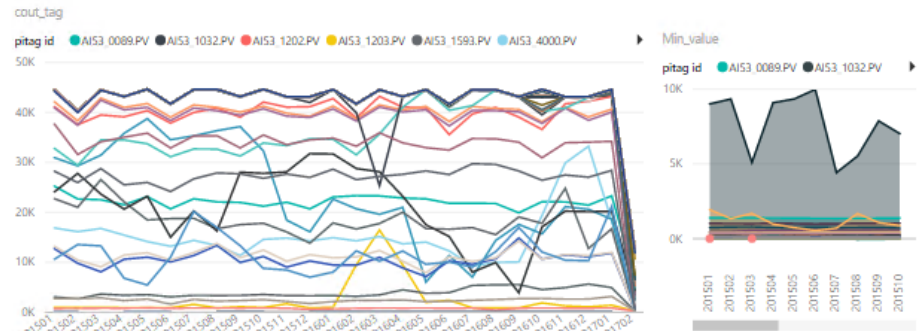
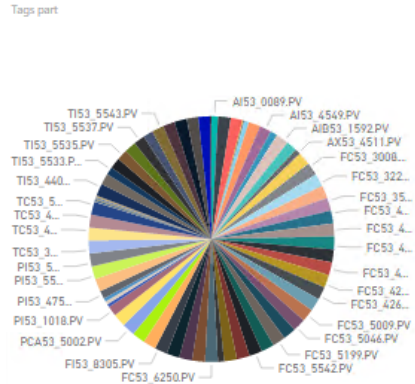
2.52

1.52

# ONLINE LEARNING: Models follow-up

## Power BI Dashboard

PI Tags™  
descriptive statistics  
to detect important  
changes among  
input variables

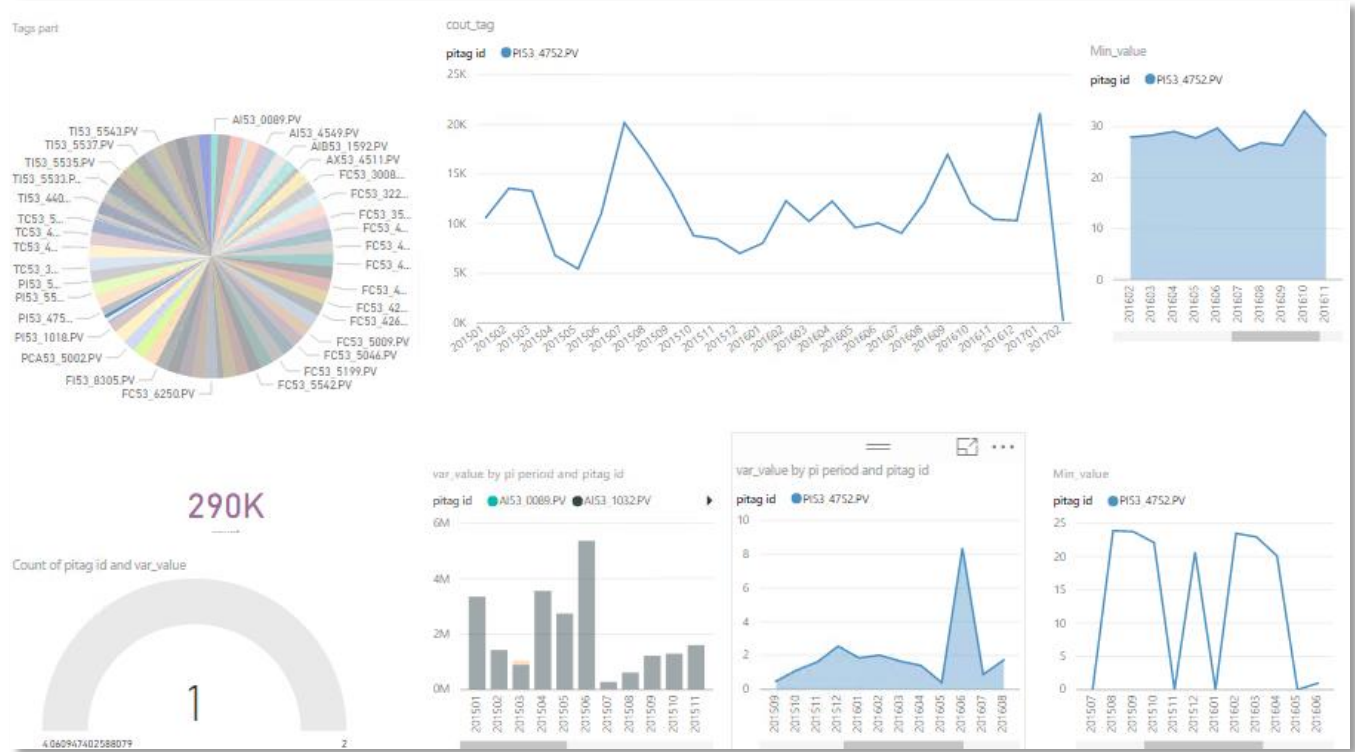


# ONLINE LEARNING: Models follow-up

## Power BI Dashboard

Dynamic dashboard

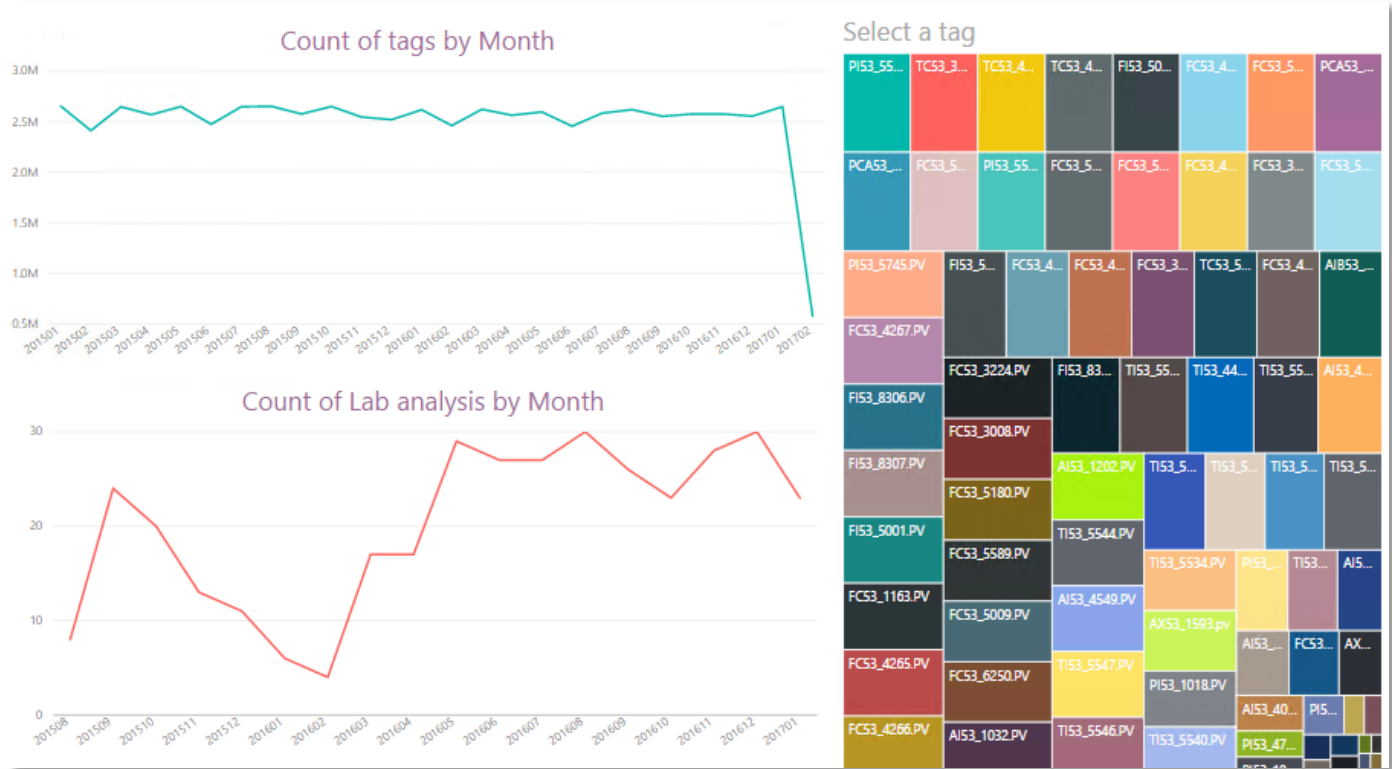
Zoom on a particular variable



# ONLINE LEARNING: Models follow-up

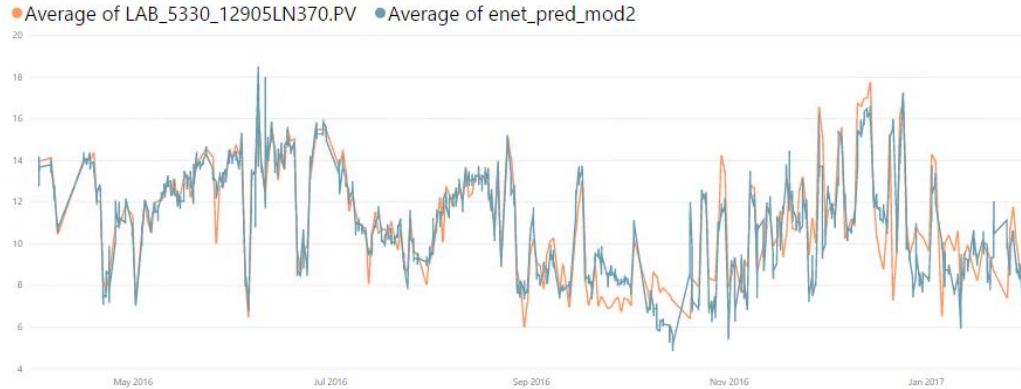
## Power BI Dashboard

Data quantity to monitor models retraining ←

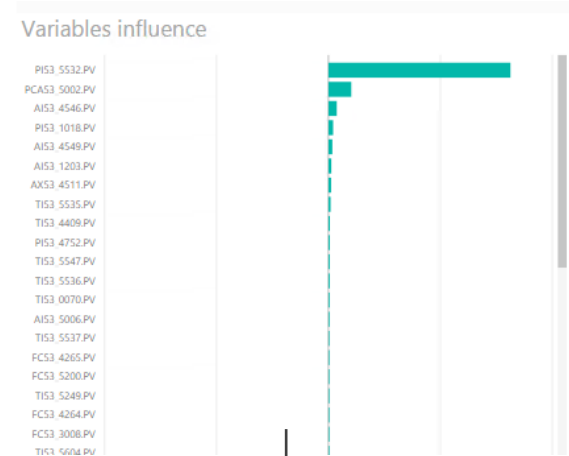


# ONLINE LEARNING: Models follow-up

## Power BI Dashboard



Model performances in time



Variables influence for each model





# Conclusion and Next steps

# CONCLUSIONS

## BUSINESS OBJECTIVE

Increase lab & units efficiency using advanced analytics

## DATA SCIENCE SOLUTION

Online learning to estimate the % of gasoil remaining in the residue of a distillation unit

- ▶ **An automated workflow** to extract data, compute several models and provide data scientists and Business experts with an updated dashboard

## ADDED VALUE

### Short term impact

- ▶ The proposed workflow is **easy to both install and generalize many ML models** for quality estimation
- ▶ **Provide a clear and updated view of unit's quality** to the Business at any time

### Long term impact

- ▶ Important **reduction of processed samples** that bring little or no added value

# NEXT STEPS

- ▶ Create more rules to launch **a model retraining automatically**
- ▶ Give Business experts **the possibility to manually launch a model retraining**
- ▶ Send online results back to the refinery's Asset Framework (AF)

# Real-time Estimations and Online Learning for Industrial Assets at Total



## COMPANY and GOAL

**TOTAL: committed to better energy**

The Business objective is to increase lab & units efficiency using data science methodology and tools

## CHALLENGE

To build a **data-driven model** based on PI Tags™

In order to anticipate possible model degradation over time, an **online retraining** has also been developed.

- A project for Antwerp Refinery (Belgium) to estimate the % of gasoil in the residue of a Crude Distillation Unit.

## SOLUTION

Definition of an online machine learning workflow using data science and PI systems tools

- Asset Framework (AF) and PI OLEDB™ were used to efficiently extract data
- Data science tools



## RESULTS

An **automated workflow** scalable, easy to use and that provides good quality estimation

- Model estimation providing an error of less than 1.5 point
- A clear and updated view of unit's quality to the Business at any time
- Lab samples reduction

# Contact Information

**Juliette Spinnato**

[juliette.spinnato@total.com](mailto:juliette.spinnato@total.com)

Data Scientist

Total



## Questions

Please wait for the **microphone** before asking your questions



State your **name & company**

## Please remember to...

Complete the Online Survey for this session

Download the Conference App for OSISOFT USERS CONFERENCE 2017



- View the latest agenda and create your own
- Meet and connect with other attendees



search OSISOFT in the app store

<http://bit.ly/uc2017-app>

감사합니다

谢谢

Danke

Merci

Gracias

**Thank You**

ありがとう

Спасибо

Obrigado