# Actuary and Data Scientist : Bring Complementary Areas of Experties in the Insurance Field

Léonard FONTAINE, GALEA & Associés
Joël ABOA, GALEA & Associés

# About the speakers

- **Léonard FONTAINE** – *Senior actuarial consultant and Partner at GALEA & Associés*
  *Léonard is a partner and a senior actuarial consultant. He carries out various technical insurance-related project on topics like pricing, reinsurance, data science and IFRS accounting standards. He also takes part in the work of the Accounting Commission in the Institute of Actuaries. He joined the firm in 2011.*
- **Joël ABOA** – *Consulting Actuary, GALEA & Associés*
  *Joël is a consulting actuary is specialized in healthcare, data science, claim reserving and reinsurance. He joined the firm in 2018.*

- GALEA & Associés is a French consulting firm specialized in actuarial sciences, that assists firms and insurance companies in their risk management and their social protection systems' monitoring (social security, medical expenses, pensions, Employee Savings Plans (ESP)). GALEA's field of work also covers data science and Mergers and Acquisitions (M&A).

# Table of contents

# 1. Actuary and Data Scientist : What's the Difference ?

## Actuary

**Applies business skills**
Insurance Underwriting
Reserving
Reinsurance
Asset and liability management
Risk management
Modeling
Data analysis
Expert judgement

**Analyses standards and regulations**
Prudential regulations and accounting standards

**Monitors economic and financial contexts**

**Understands marketing needs**

**Contributes to compliance**

**Comunicates with different stakeholders**
Directors, administrators, supervisors, regulators, accountants, sales representatives, etc.

## Data Scientist

**Applies data-oriented skills**

**Has IT expertise**
Developing and monitoring models
Market solutions understanding
Handling high-volume data

**Masters mathematics, and algorithms**
Selecting the most suitable models
Carrying out studies related to AI, Machine Learning, Neural network, etc.

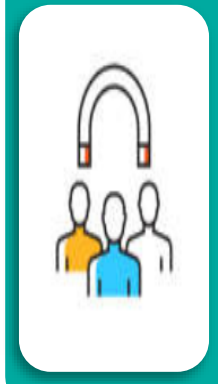
# 2. Scope of Application

## Behavioural aspects

- Multiple-Decrement Life Tables
- Surrender risk
- Prepayment laws on borrowers contrats (cf. French Bourquin law)
- change of guarantees/options
- Monitoring medical exepense & Income protection :development according to customer behaviour
- Pricing
- Fraud

## Profitability/adequacy of premiums

- Calculating customer lifetime value (on the entire or on a part of the portfolio)
- Determining of priority contracts termination
- Performing consumption analysis
- Conducting statistical studies of the claims



## Marketing

Conducting a study in order to :

- Segment clients
- Analyse their interaction
- Suggest a developpment strategy

## Risk Management

- Assessing premium rating by geographic area
- Scoring driver behaviour scoring
- Analyzing large claims
- Estimating bankruptcy scoring (for company)
- Ensuring the underwritings' quality
- Providing market risk expectation

## Preventive measures

Inlcuding :

- Absenteeism
- IoT : monitoring prevention
- Certain diseases control (strokes, major depressive disorder, etc.)

# 3. How to Carry Out a Data Project

## ▪ Stages

Two stages are essential for a Data Science project's implementation:

**Data**
(Processing data, exploratory data analysis, data visulization, etc.)

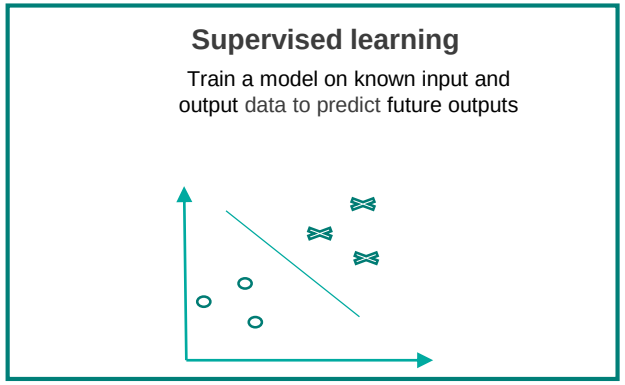
**Algorithms**
(Implementation, interpretation of results, etc.)

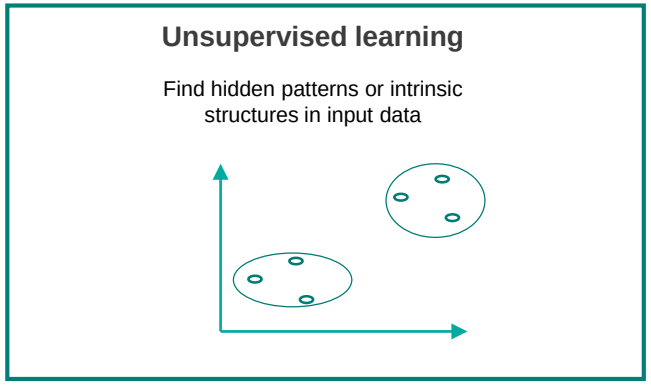
Data processing is **more time-consuming** (takes around 80% of the time). It consists of detecting outliers, adapting data to the study's aim, etc.

# 3. How to Carry Out a Data Project

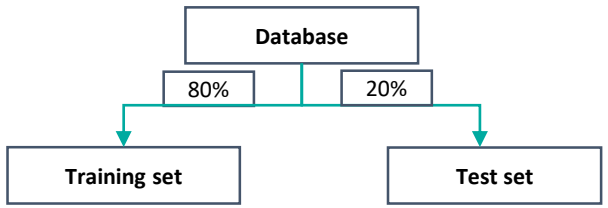
## Machine Learning algorithms

- There exist two Machine Learning techniques:

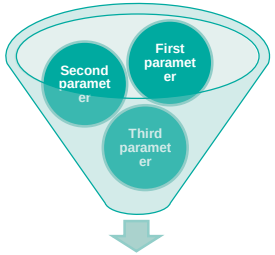



- Data splitting and model evaluation




1 *Splitting data into training and test set*



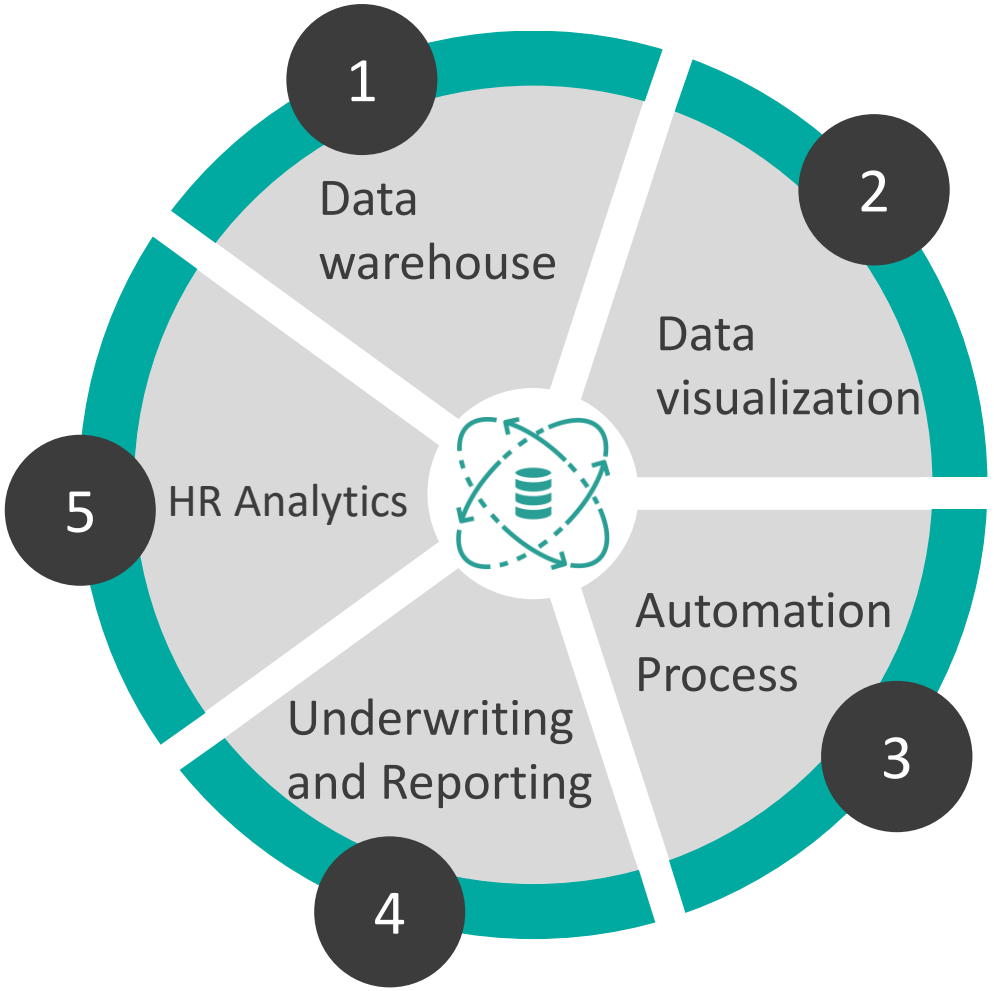

Final model

2 *Calibrating model on training set*

Risk function (e.g. : MSE)

3 *Validating model on test set*

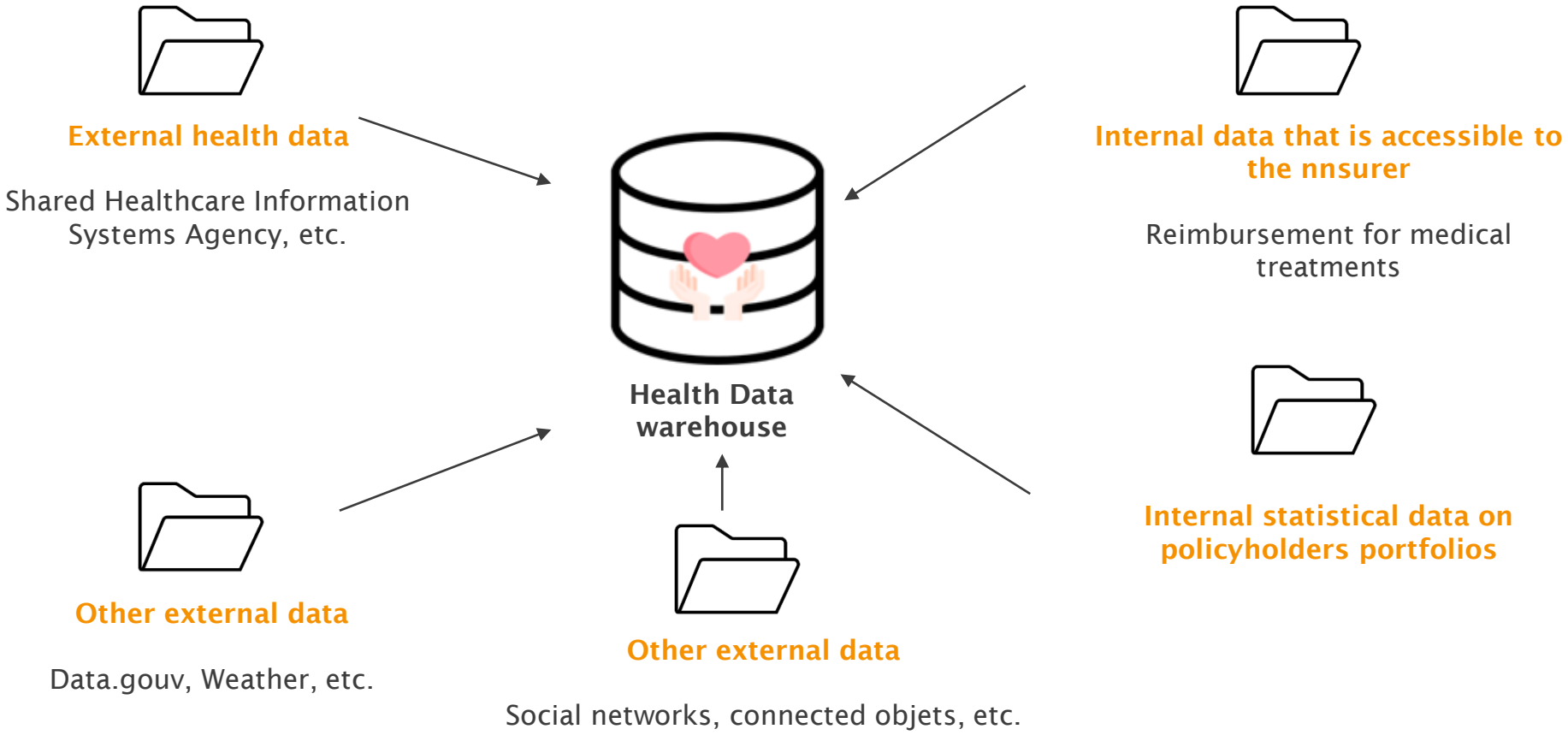
# 4. Practical examples



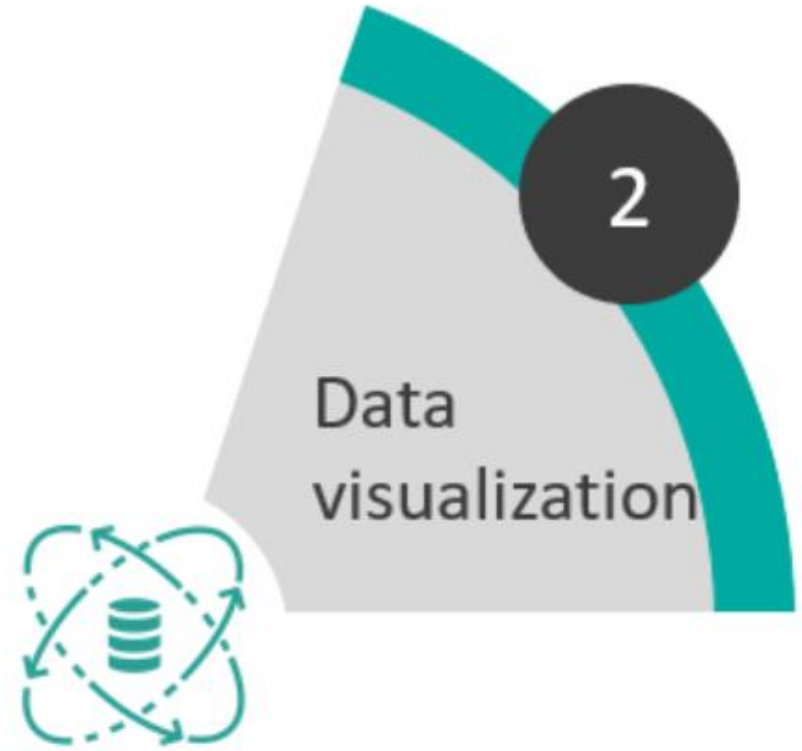
# 4. Practical examples


- **Internal data/External data – Securing – Data Quality and transparency – Patrimonial approach**

# Internal data/External data – Securing – Data Quality and transparency – Patrimonial approach

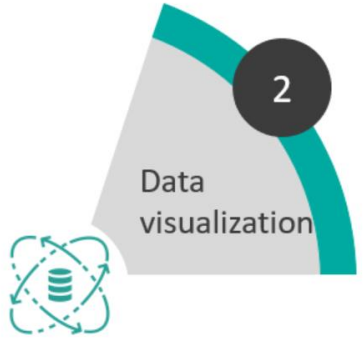



# 4. Practical examples




- **Data visualization as a powerful way of communication**

# Data visualization as a powerful way of communication




## What is data visualization ?

Data visualization is the graphic representation of data.

## Why do we need it ?

We need it to **translate datasets** into **visual graphics** in order to allow an easy understanding of complex relationships within the data.

## Common data visualization tools

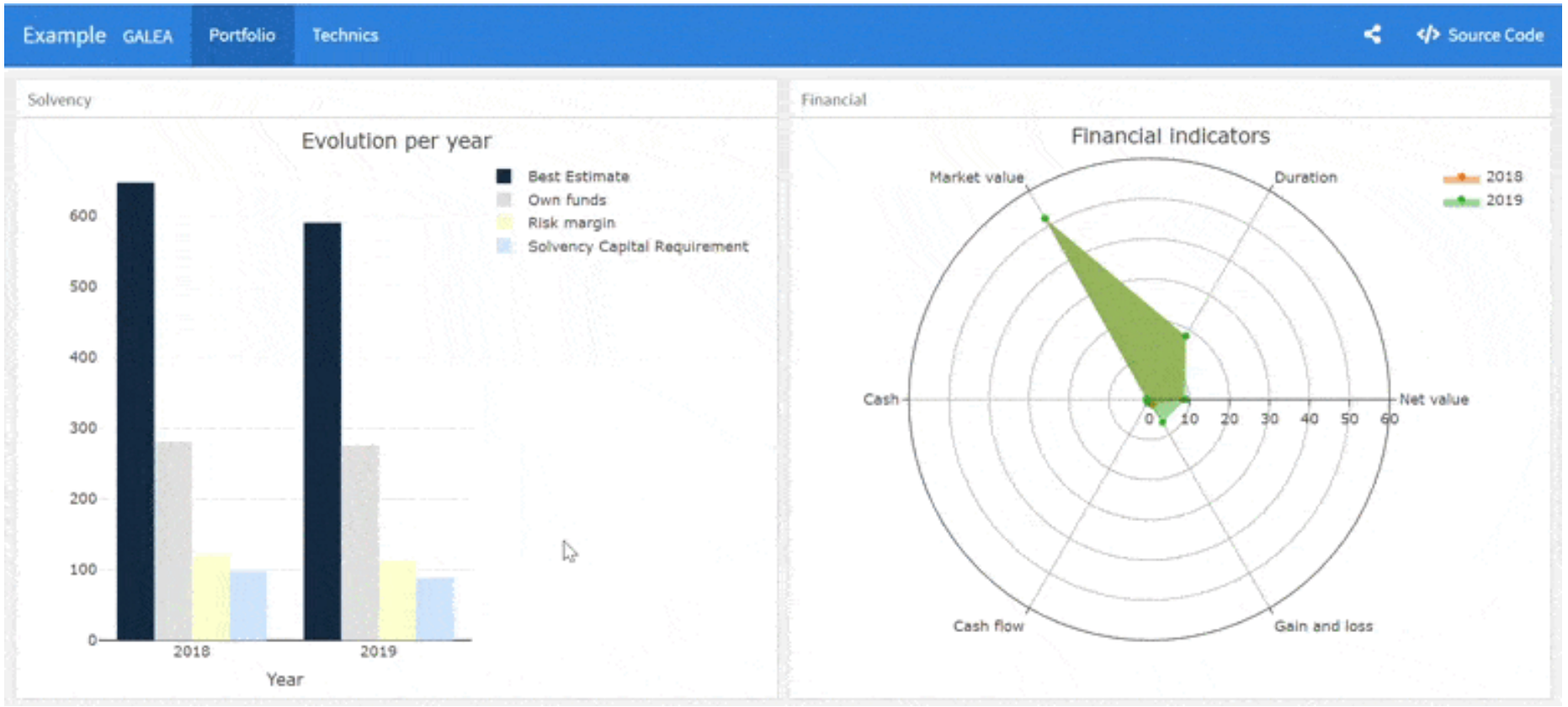
Actuaries and data scientists use R language, Plotly or Python to create interactive graphics and build dashboards.

Dashboards are intended for managers as well as technical staff. Its primary advantage is to quickly update reportings and protect data.

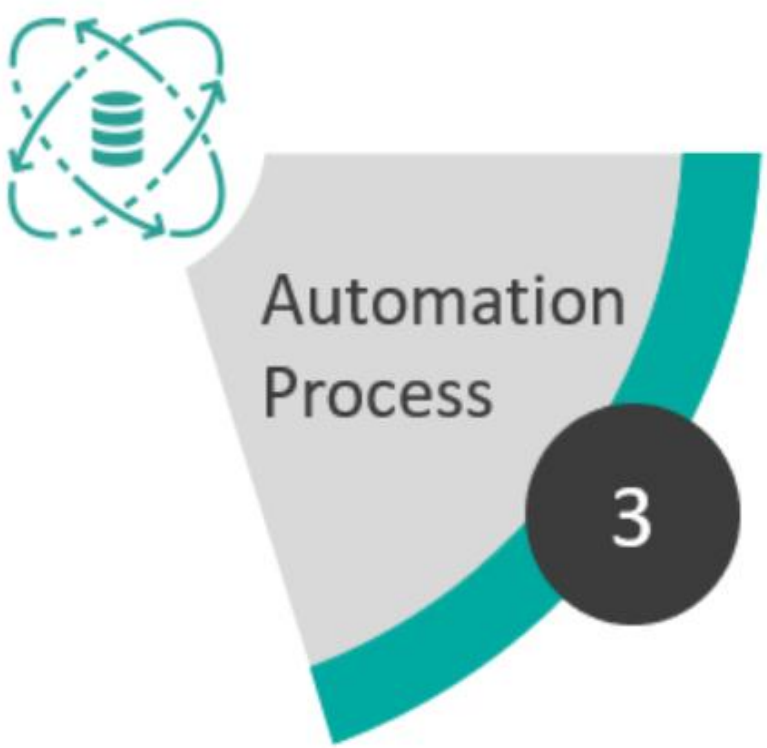
# Data visualization as a powerful way of communication



- **Dashboard with R : an example**

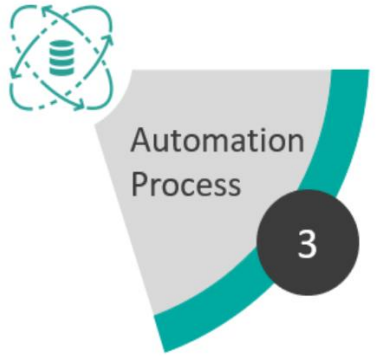


# 4. Practical examples




**Automation - Robotization - Artificial intelligence**

# Automation Process




- **Context**
  - Seeking for gain in productivity and flexibility in a tightly regulated environment at a national and international level.
  - Topical issues driven by regulatory inflation and competitive pressure on the market (see Appendix, slide 35)
  - **Paradox** : Produce and submit more complex accounting records with an improved analysis in a shorter deadline.
  - In order to respond to these issues, Automation Process seem to be necessary for Insurers.

# Automation Process

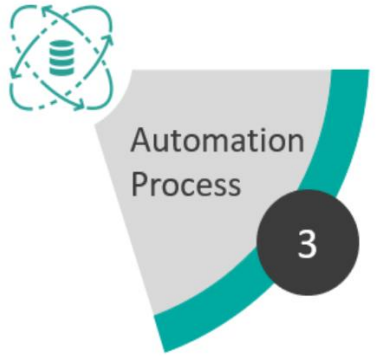


- **Distinguish automation from robotization and artificial intelligence**


| Automation | Robotization (RPA) | Artificial Intelligence |
|---|---|---|
| Program performing simple tasks. Can be manually activated. Very useful in Pack Office. | Software executing specific IT tasks, usually performed by a person on one's own. | IT and algorithmic techniques (Machine Learning, Deep Learning, etc.) to replicate cognitive recognition capabilities, interpretation… |

The aim of automation and robotization is not to entirely replace humans by machines, but to substitute them when tasks are repetitive, tedious, time-consuming, with low added value. Thus, one's saved time is spent on analyzing and creating value.

# Automation Process




## Automation objectives in Insurance

1. Increase a company's **productivity** and **efficiency**
2. **Gain in time** when it comes to executing low added value tasks
   → **Increase** time spent on analyzing
3. Increase **personal satisfaction** :
   - Reduction of manual, repetitive and tedious tasks
   - Increasing time spent on tasks with high added value
4. Improve planning **accuracy**
   - Minimization of operational risk: typos, accidental modifications or deletions…
   - Documentation : **Traceability**
5. Cost saving, FTE reduction : **High ROI**

## First applications in the insurance field

❑ **Fukoku Mutual Insurance**

すてきな未来応援します
フコク生命

Japanese Life Insurer implemented IBM's Waston Explorer system in order to analyze and process 600 cases per day (processing medical data, reading certificates, determining payments, billing expenses)

❑ **Natixis Insurance**

NATIXIS
ASSURANCES

Natixis implemented a robot to deal with contracts termination and email management. What used require 6 days of work needs only one from now on.

❑ **Aviva**

AVIVA

Robotization of tasks related to call centers and premium calculation.

Furthermore, several insurers have carried out first applications at the back office and management control : for example, monitoring dashboards.

# 4. Practical examples


- **Health Insurance Reporting : How to improve it with Data Science ?**
- **Auto pricing and the contribution of Telematics Data**

# Health Insurance Reporting : How to improve it with Data Science ?


## Issue

- Background information

Today, health reporting presents a static vision of the portfolio.

**Move on to a dynamic vision** : follow-up indicators, prevention policy, ...

- Lines of improvement
  - ❑ Data :
    - Making better use of available data
    - Integrating external data
  - ❑ Models :
    - Using Data Science

# Health Insurance Reporting : How to improve it with Data Science ?


## Study process

- Objective

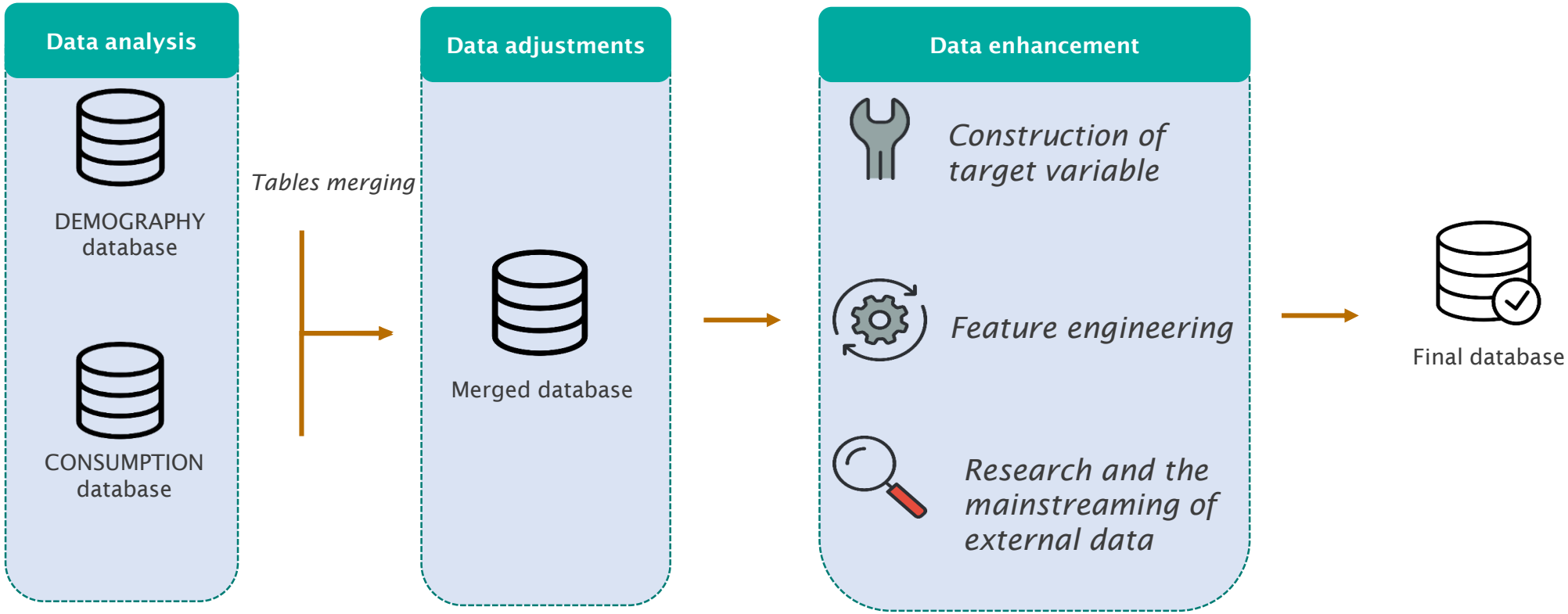
Use Data Science models for a better understanding of healthcare consumption

- Different focal areas should be analyzed :

❑ Healthcare consumption analysis through pricing
❑ « Large consumers »
❑ Health reportings

# Health Insurance Reporting : How to improve it with Data Science ?


## Data processing



# Health Insurance Reporting : How to improve it with Data Science ?


- **Modeling**

| Algorithm | Simplicity | Learning speed | Interpretability | Predictive power |
|---|---|---|---|---|
| GLM | ★★★ | ★★★★★ | ★★★★ | ★ |
| CART | ★★★★ | ★★★★★ | ★★★★★ | ★★ |
| Random Forest | ★★ | ★ | ★★ | ★★★ |
| Extreme Gradient Boosting | ★ | ★ | ★★ | ★★★★ |

# Health Insurance Reporting : How to improve it with Data Science ?


- **Results**

## Premium

| Model | Frequency | Average expenditures |
|---|---|---|
| | MSE on test set | MSE on test set |
| GLM | 22,6 | 82,9 |
| CART | 23,5 | 59,6 |
| Random Forest | 22,1 | 56,3 |
| XGBoost | 23,2 | 54,5 |

| | |
|---|---|
| | Most relevant model |
| | 2nd most relevant model |

## « Large consumers »

Performance : AUC ≈ 70 %

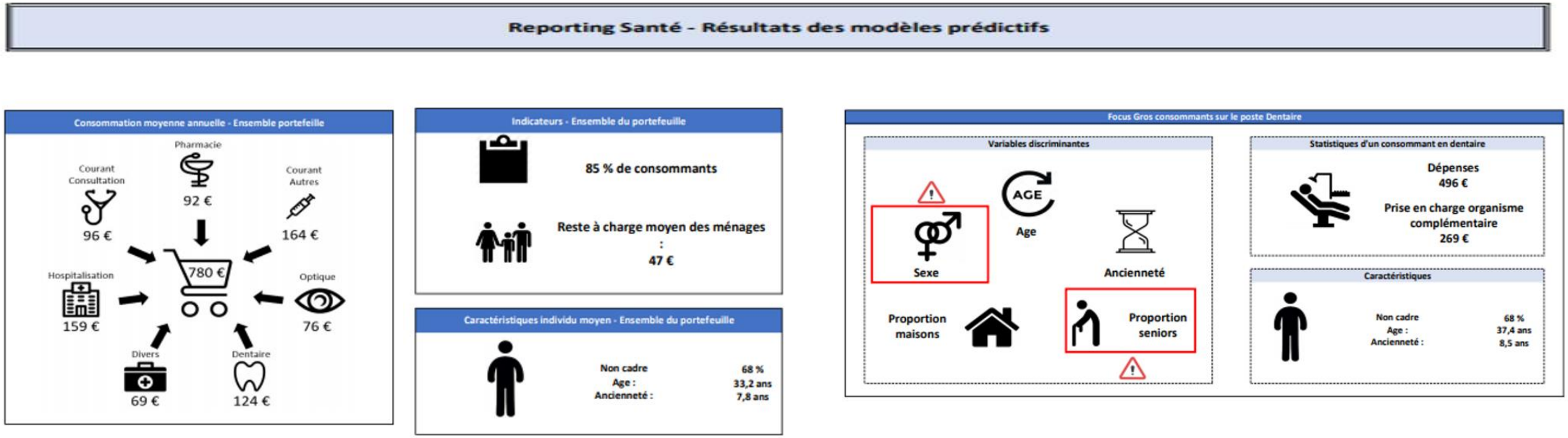
Dental care

| Frequency | Average expenditures | « Large consumers » |
|---|---|---|
| Age | CSP | Age |
| CSP | Dentist number | Healthcare equipment number |
| Dentist number | Number of healthcare equipment | Seniority |
| Healthcare equipment number | Seniority | Sex |
| Activity ratio | Households ratio | Senior ratio |

# Health Insurance Reporting : How to improve it with Data Science ?



- Results
  - Health reporting



# Auto pricing and the contribution of Telematics Data


- **Issues**

Use both GLM and Machine Leaning models to :

  - Model the number of claims
  - Predict the losses
  - Assess the relevance of these models

- **Process**
  - Processing data
  - Modeling
  - Integrating telematics data

# Auto pricing and the contribution of Telematics Data


## Processing data

### Data collection

- Database construction
- Exploratory data analysis
- *Feature engineering*
- Missing value treatment
- Methods for attenuating the effects of large claims

### Modeling

- Training set / Test test
- Models' implementation (**GLM, CART, *Random Forest*, XGBoost**)
- Variable importance
- Measuring models errors (**RMSE**)
- Results / analyses

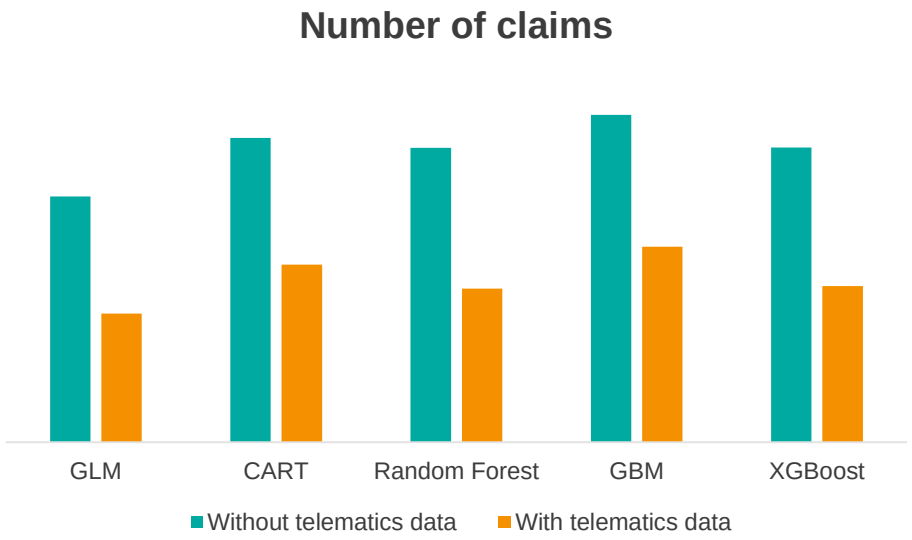
### Integrating telematics data

- Geocoding ZIP codes
- Processing telematics data
- Merging with previous database
- Modeling + impacts analysis

# Auto pricing and the contribution of Telematics Data


## Results

Models evaluation (RMSE) on test set :




# Auto pricing and the contribution of Telematics Data


## Conclusion

This study shows that Data Science models are comparable to traditional models (GLM) and allow us to have better results. Insurers can use them in their pricing policy.

## Discussion

Adding telematics data helps getting better results but is it really worth it? Processing telematics data can be tough and data science models implementation take a lot of time (e.g XGBoost). As an actuary or a data scientist, one needs to be pragmatic in making decisions.

# 4. Practical examples


- Absenteeism
- Gender equality in France
- Turn-over

# Absenteeism



- **Meaning**
  - The meaning will vary depending on the context
  - « Planned » absences
    → Maternity, holiday, …
  - « Unplanned » absences
    → Disease, accident, exceptional holiday

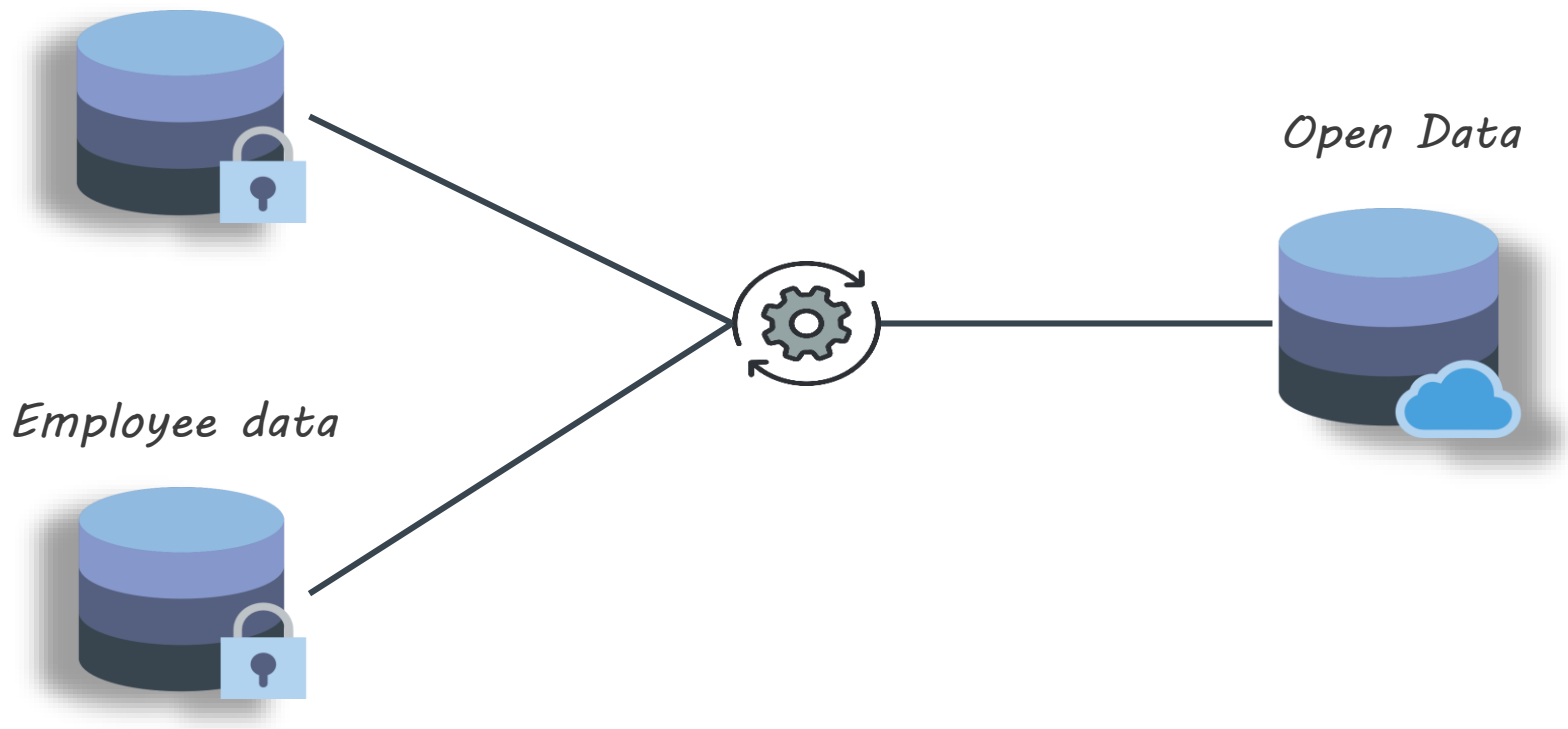
- **Representation**
  - Reasons/Causes
  - Duration
  - Renewal
  - Influence

→ In a prevention framework :
Case study of the **« Unplanned » absences** focused on **influence**

# Absenteeism

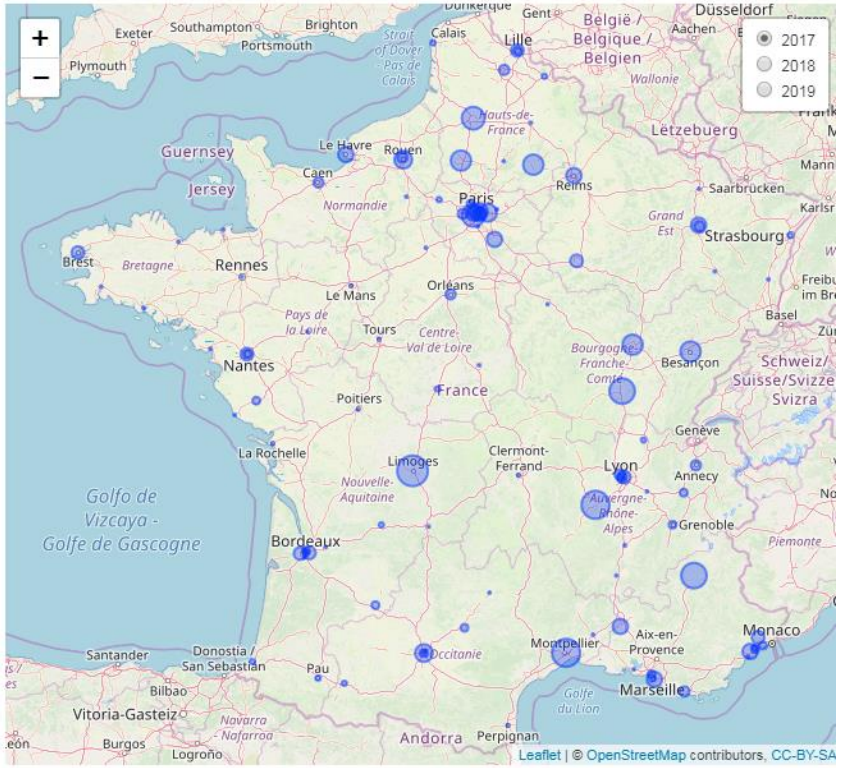
- **Processing data : Adding external data**

« Unplanned » absences data



# Absenteeism

- **Data mining**
  - Explanatory factors
    - Age
    - Sex
    - Seniority
    - Family status
    - Number of dependent children
    - Socio-professional categories
    - Contract
    - Administrative partitioning (Dilution of Precision, unit)
    - Epidemic
    - Duration
    - Period (seasonality)

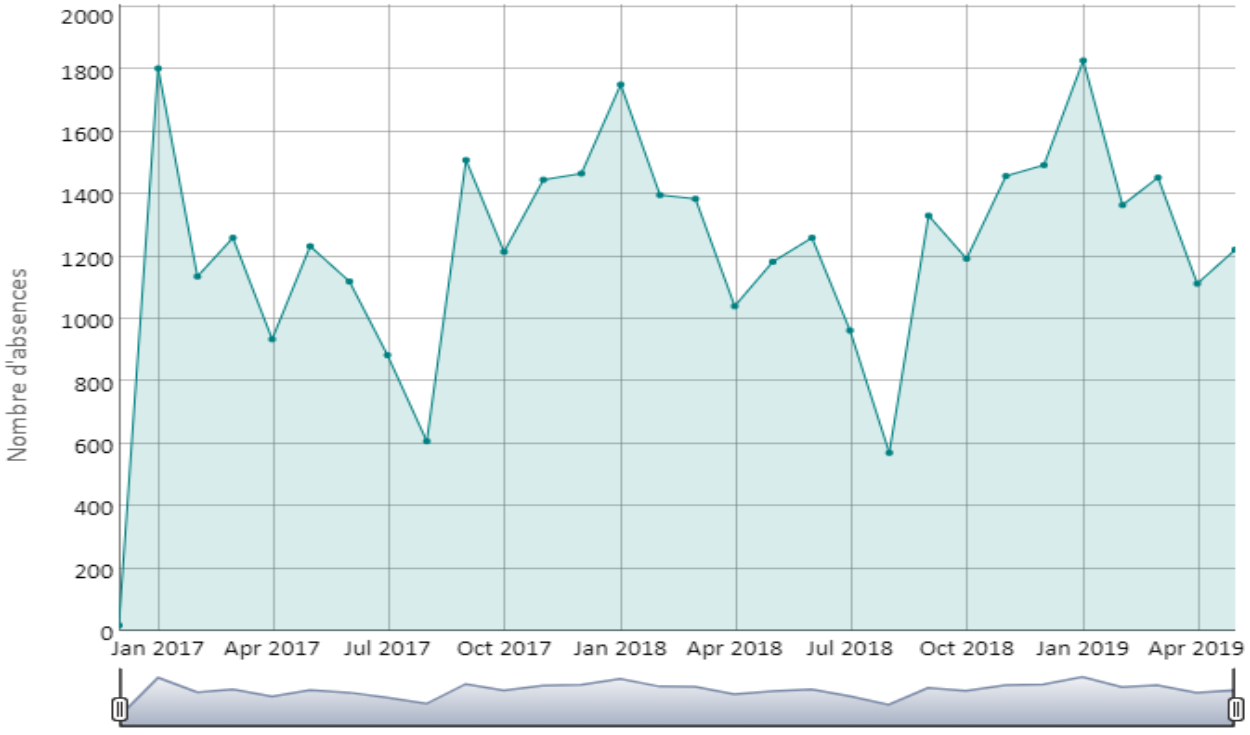



*Absenteeism rate according to structures*

# Absenteeism

- **Data mining**

**High temporal component**

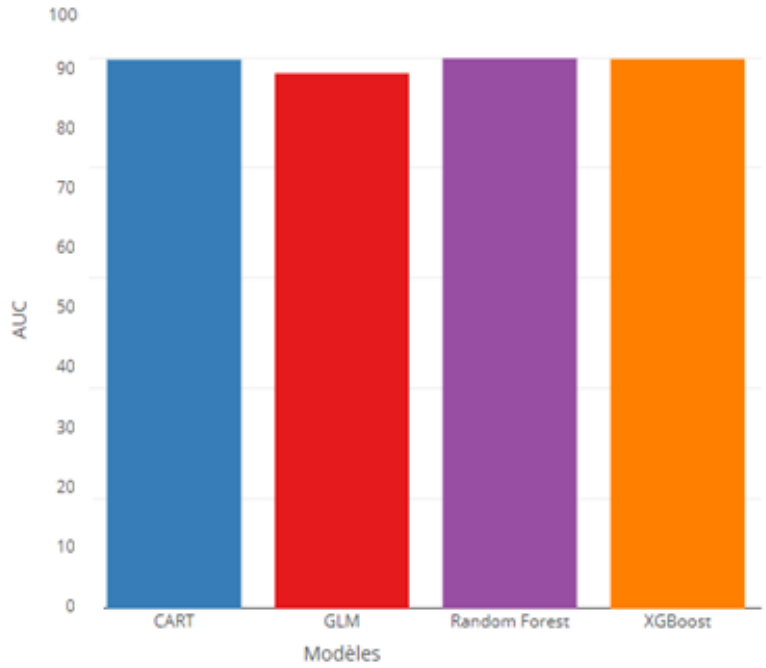



*Period of absence*

# Absenteeism

- **Modeling**
  - ❑ **Four models:**
    - Generalized linear model (GLM)
    - Decision tree
    - Random forest
    - eXtreme Gradient Boosting (XGBoost)
  - ❑ **Performances:**
    - AUC criterion

*Models's AUC*




- **Model choice**

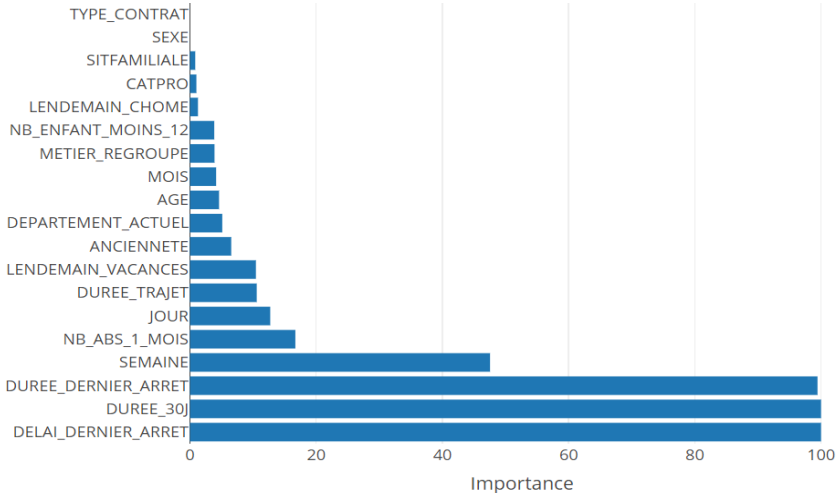
Decision tree is chosen: this model's interpretability is preferred in comparison to other more efficient models.

# Absenteeism

## Key factors :

- Absence history
- Seasonality
- Travel time
- Context

*Discriminant variables – Decision tree*




## Conclusion

We measure and diagnose the main parameters related to absence. However, in many situations, this is not enough to take action.

One needs to develop an entire project approach and involve all stakeholders : executive management, human resource department, staff representatives and employees.

# Gender equality in France



## Observation

On average a woman earns about **25%** less than a man over a year.

For equal work, there's a **9%** gap.

## Legal obligations

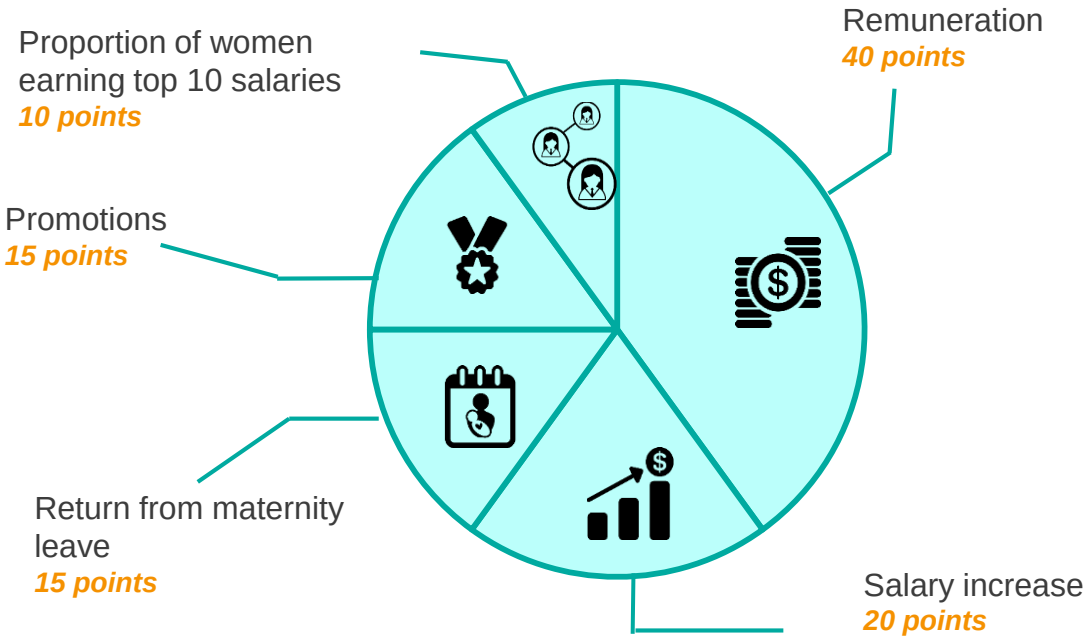
Achieve equal remuneration in 2022 in four steps :

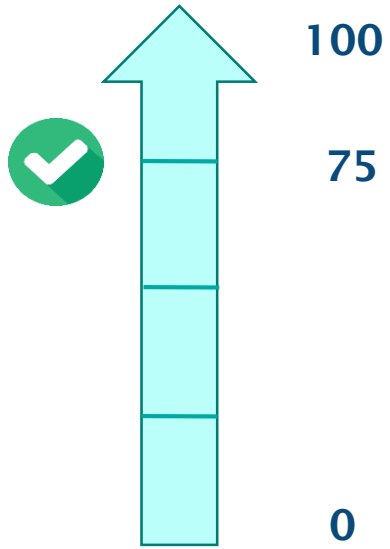
- Measure : Calculate gender Inequality Index (5 indicators)
- Transparency : Publish results on government website
- Adjustment : Set 3 years to close the gender-related wage gap
- Control : Put financial sanctions

# Gender equality in France

- **Focus on Gender Inequality Index calculation (5 indicators)**

**Index** = + + + +

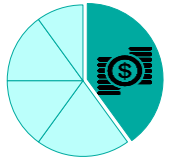



**An actuarial point of view :**
The Index should be challenged because the defined criteria have limits (e.g. : Remuneration is sensible to grade, seniority etc.) Using predictive modeling can help to handle it.

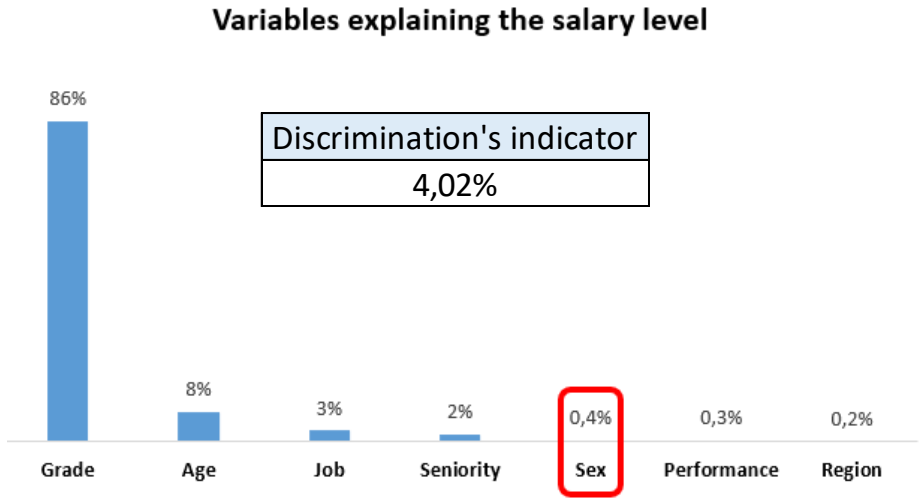
# Gender equality in France

## ▪ Modeling Remuneration's indicator

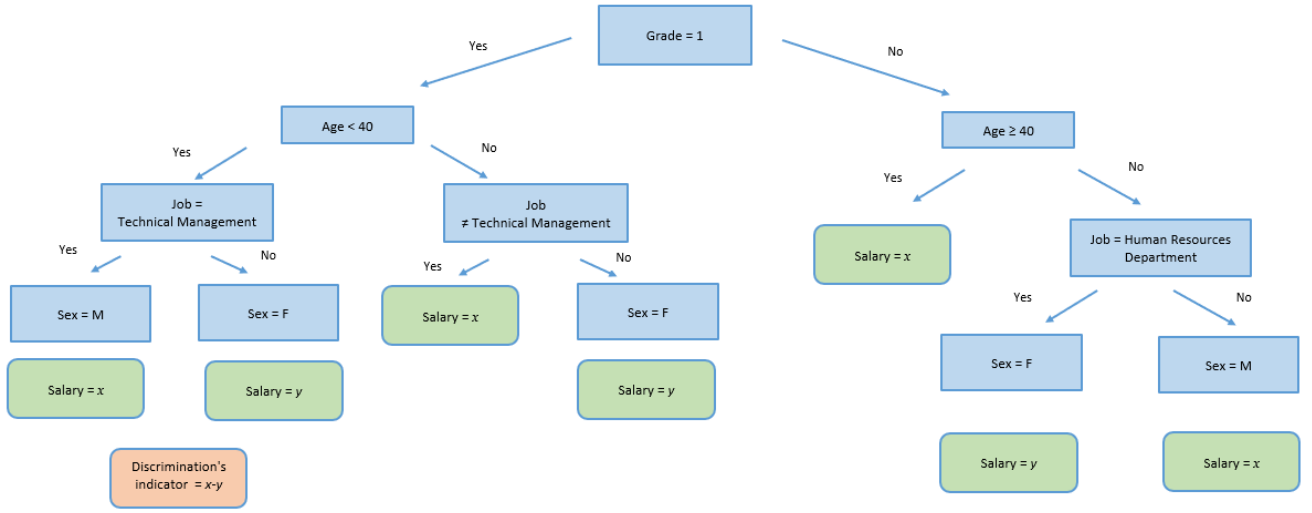


*40 points*

GLM model : Calculating an indicator related to the pay gap between women and men




CART model : Identifying discriminated women



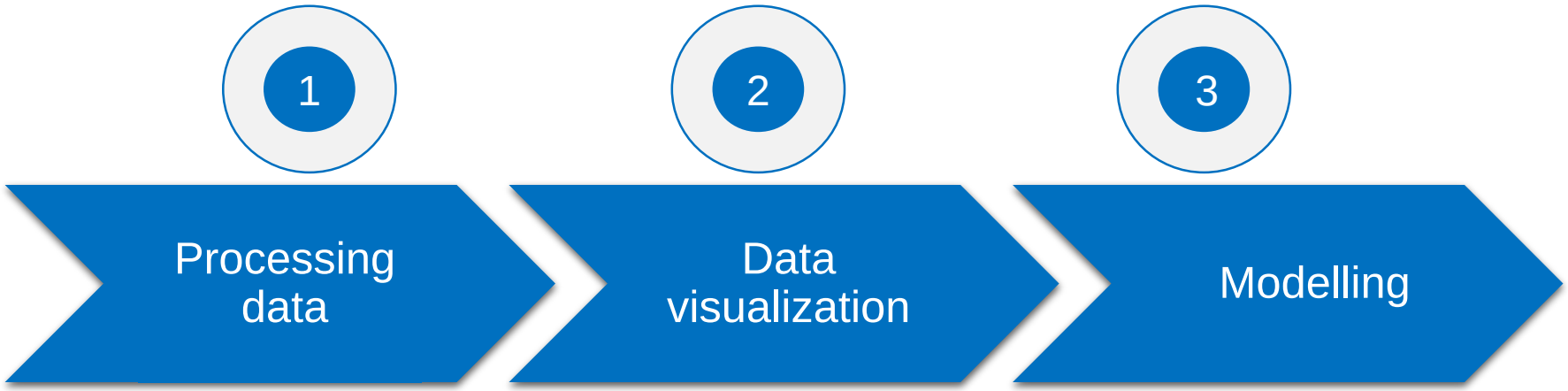
# Turn-over



- **Objective**

The objective is to identify any dissatisfaction that could lead to a early departure of an employee in order to readjust human resources management.

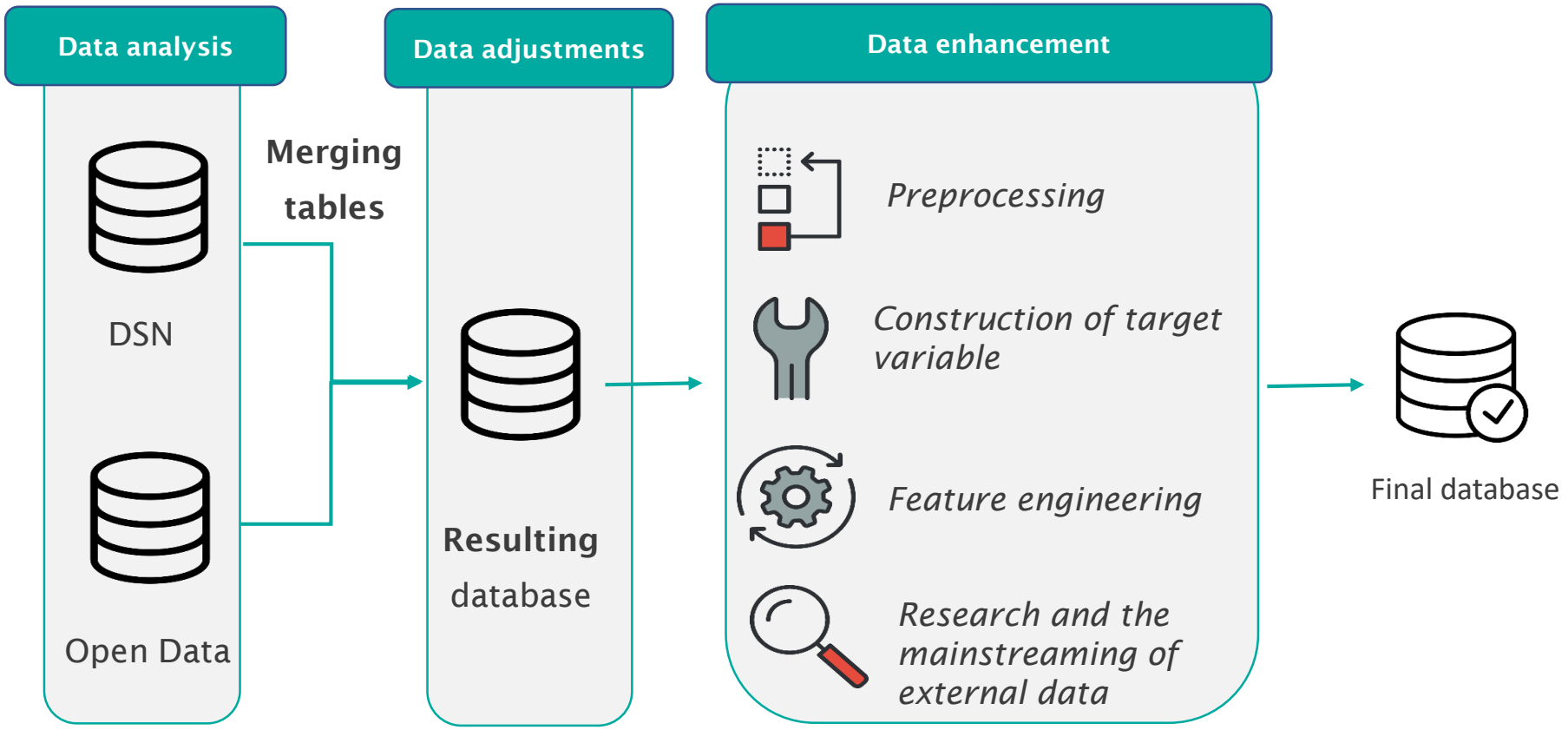
- **Process**



# Turn-over



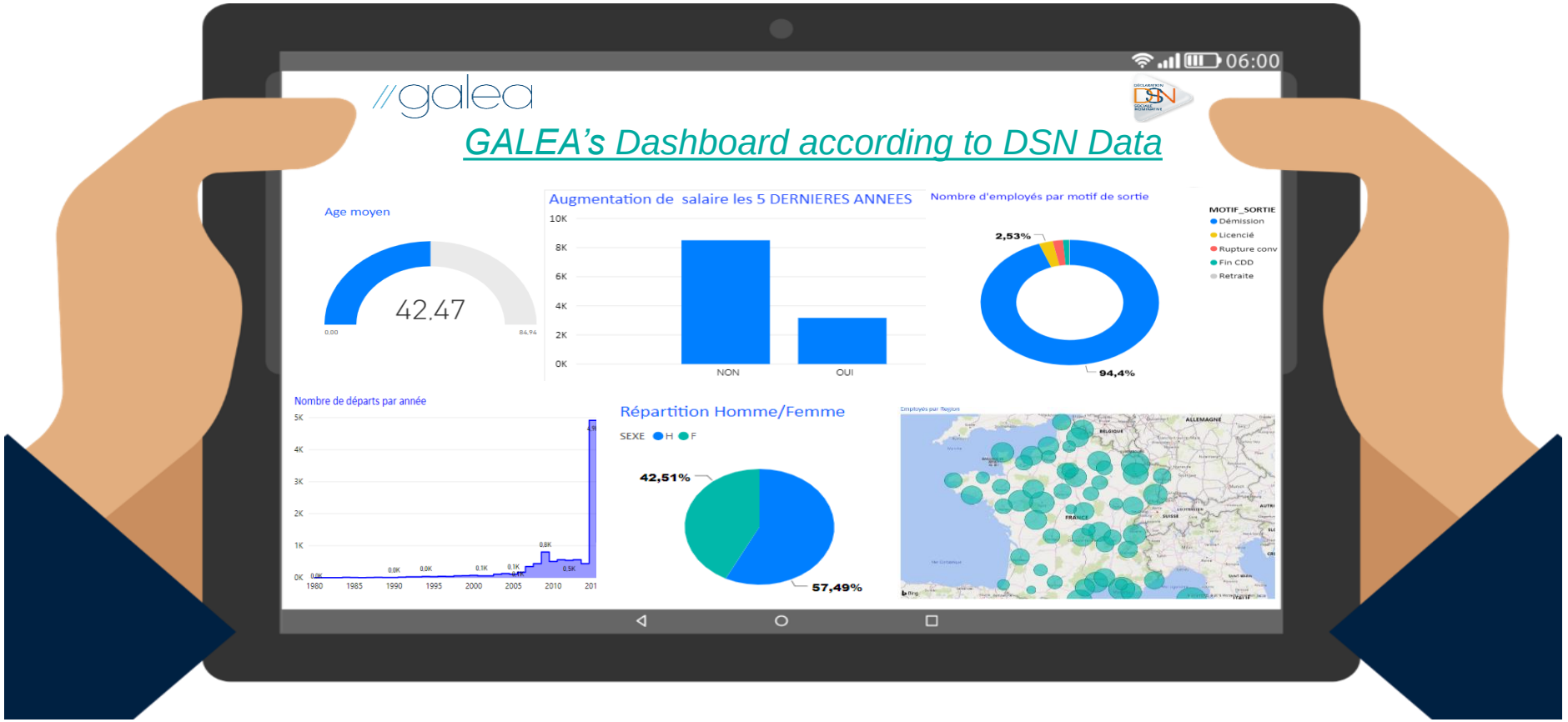
Processing data | Data visualisation | Modelling



# Turn-over



Processing data | Data visualisation | Modelling

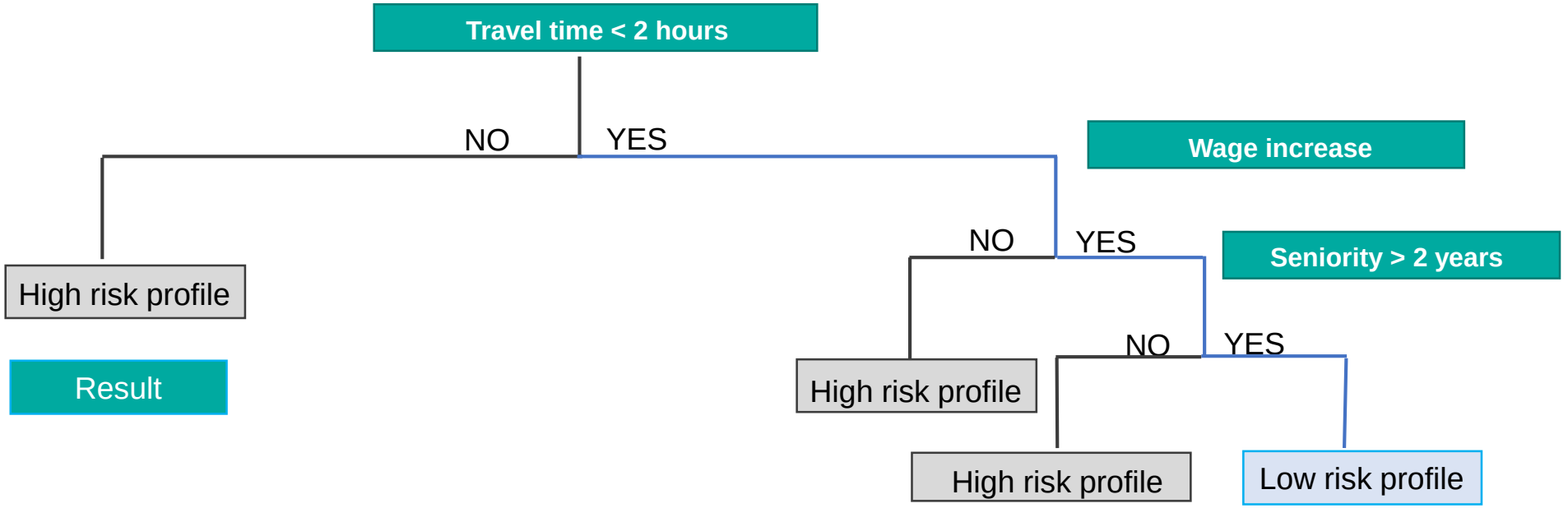


# Turn-over


Processing data → Data visualisation → Modelling

***This employee is leaving the company ?***

- Age : 37 years
- Sex : M
- Travel time : 1 hour
- Seniority : 3 years
- Chance of career progression : yes



# Appendix

Les Echos, *IFRS17 ou l'avènement des robots?*
26 mars 2018

*« […] IFRS17 introduit tout à la fois de nouvelles modalités de comptabilisation du chiffre d'affaires, une valorisation des passifs d'assurance reposant sur un modèle prospectif et une refonte complète des états financiers. En clair, cela signifie la manipulation d'un volume considérable de données établies à une maille fine afin de produire les indicateurs comptables et analyser les marges.*

*[…] Les directeurs financiers et de l'actuariat sont ainsi confrontés à un défi s'apparentant à résoudre la quadrature du cercle : comment produire des données comptables plus complexes, mieux analysées, dans des délais toujours plus contraints et à iso-ressources (au mieux) ?*

*[…] La robotisation ciblée des processus multinormes doit permettre d'éviter cet écueil : les automates prennent en charge les travaux à faible valeur ajoutée - par ailleurs rébarbatifs et générateurs d'irritants - au profit des travaux d'analyse et du renforcement de l'efficience opérationnelle. Le caractère non intrusif des technologies sous-tendant la robotisation garantit une mise en oeuvre rapide et maîtrisée. […] »*

Les Echos, *Robotisation dans l'industrie financière, comment en faire une opportunité… pour l'homme!,* 12 janvier 2018

*« Avec de lourdes contraintes de mises en conformité et des taux d'intérêt bas, les acteurs du secteur financier tentent de retrouver leurs marges en identifiant de nouveaux leviers de productivité. Cela passe par l'informatique et notamment par l'automatisation des tâches simples, répétitives et chronophages. La logique est claire : s'appuyer sur un logiciel capable de réduire par trois ou par quatre le temps de traitement d'un processus. De fait, on évite de demander à un salarié de collecter laborieusement des masses d'informations issues de trente voire quarante systèmes informatiques.*
*En prenant en charge les tâches basiques et répétitives, le robot vient aider le salarié et le soulager d'une « douleur ». En effet, le salarié n'a plus besoin d'effectuer un travail de fourmi et peut se concentrer sur la création de valeur. Il va également pouvoir développer ses compétences.*
*En déployant des technologies d'automatisation, les entreprises vont augmenter la productivité des salariés et assurer une meilleure qualité de vie au travail. Libéré des tâches sans valeur ajoutée, le « collaborateur augmenté » travaille davantage la relation avec le client, la qualité et la profondeur de ses conseils et donc, in fine, apporte plus de satisfaction, à son client mais à lui-même également ! »*

# Thank you very much for your attention!


## Contact details

**Léonard FONTAINE**

**Joël ABOA**

GALEA & Associés

Email : lfontaine@galea-associes.eu; jaboa@galea-associes.eu

Phone : +33 01 43 22 11 11

http://www.galea-associes.eu/