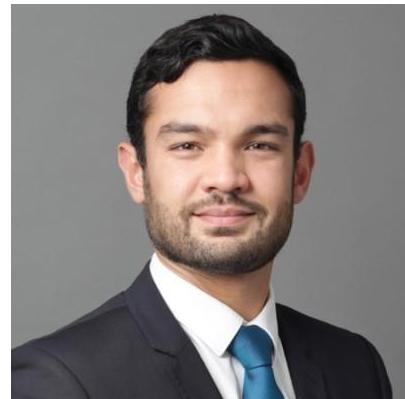


# IA en actuariat

## La formation DSA vous révèle comment passer de la théorie à la pratique



**Guillaume WELTERLIN**  
Actuaire Vie  
SCOR



**André GRONDIN**  
Directeur scientifique  
Institut du Risk Management



**Amandine GOMEZ**  
Manager conseil interne  
Generali

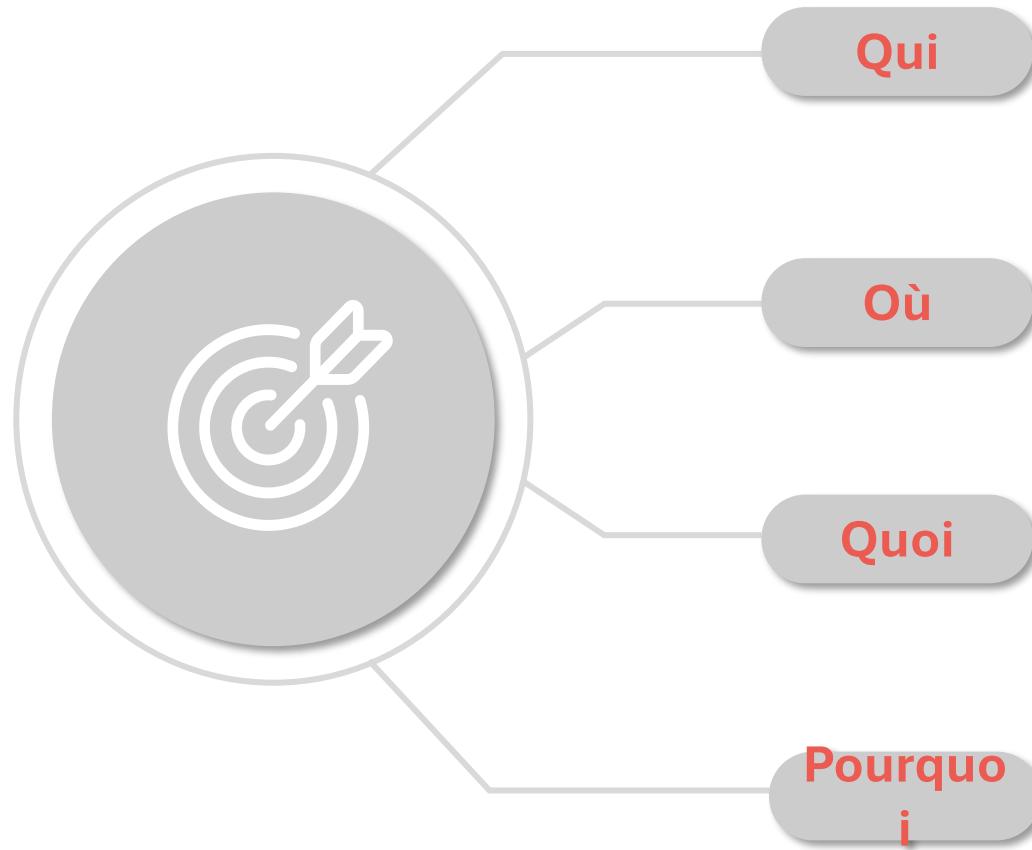


**Jean-Marc HECART**  
Fonction actuarielle  
Prepar Assurances

# Apport d'une démarche *Data Science* dans les travaux de fonction Actuarielle

Jean-Marc Hecart

# Contexte et cas d'étude



**Responsable de la fonction  
actuarielle**

**Compagnie d'assurance-  
vie**

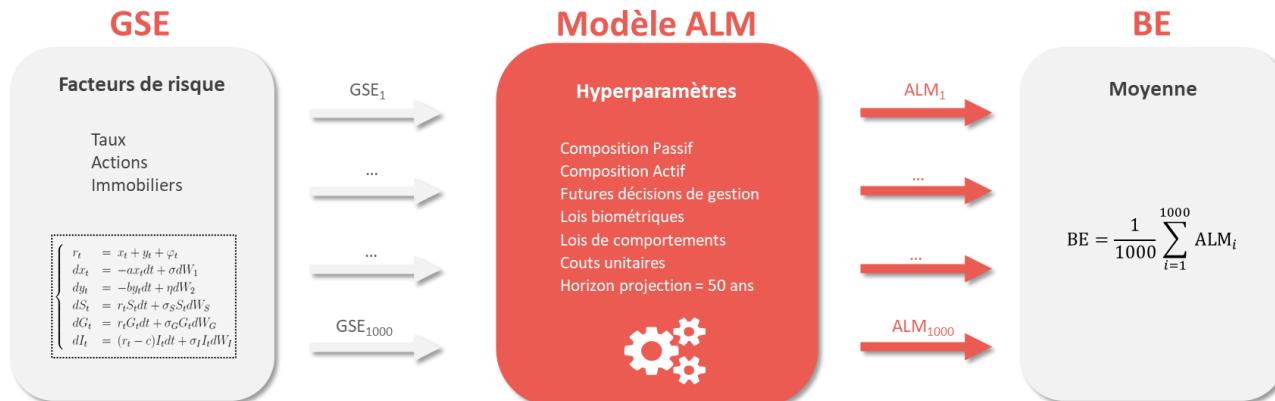
**Utiliser la Data Science**

*Data Engineering, Data Mining, Dataviz, Machine Learning*

**Compléter et approfondir les analyses**  
**Réduire le temps de réalisation des analyses**  
**Réduire la dépendance à Excel**  
**Réduire la dépendance aux équipes de production**

## « Développer un modèle proxy du modèle ALM »

238 000 contrats d'épargnes et retraites  
7,5 Mrds € d'encours

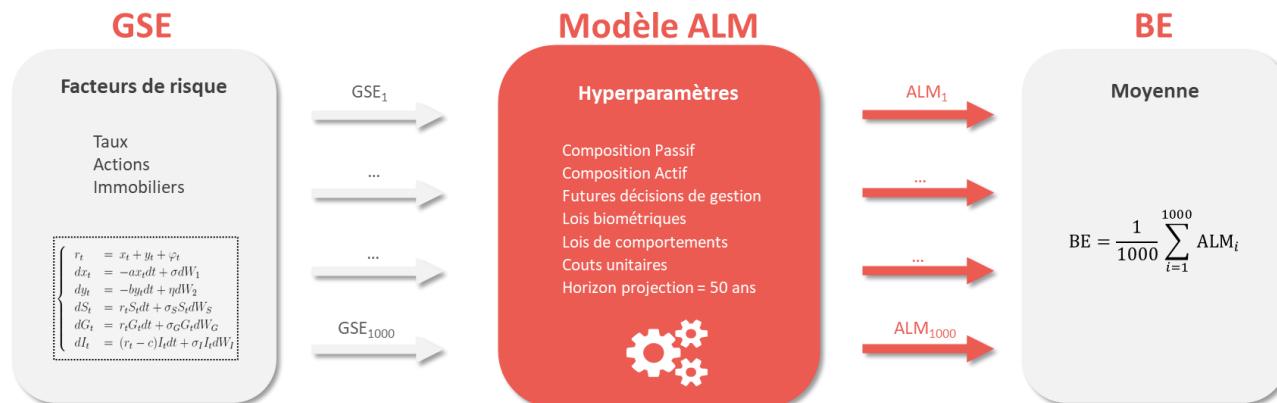


# Approche n°1 : *la naïve*

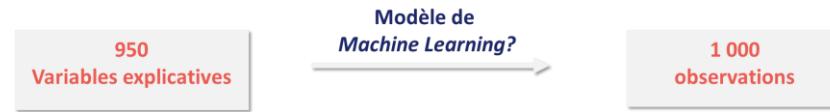
950  
Variables  
explicatives

Modèle de  
*Machine  
Learning?*

1 000  
observations

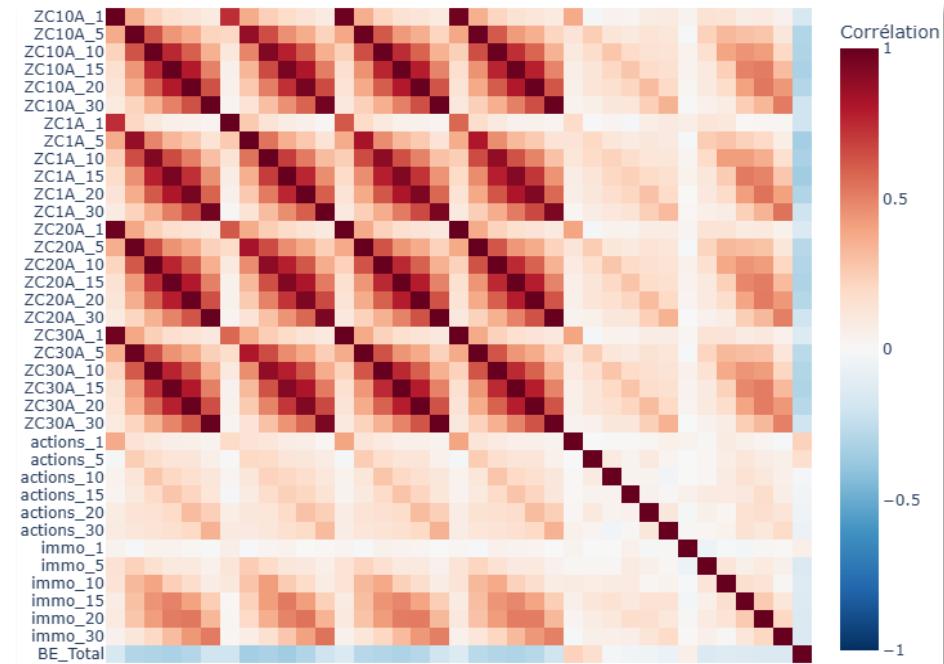


# Approche n°1 : la naïve



Réduction de  
la  
dimensionalit  
é

36  
Variables  
explicatives



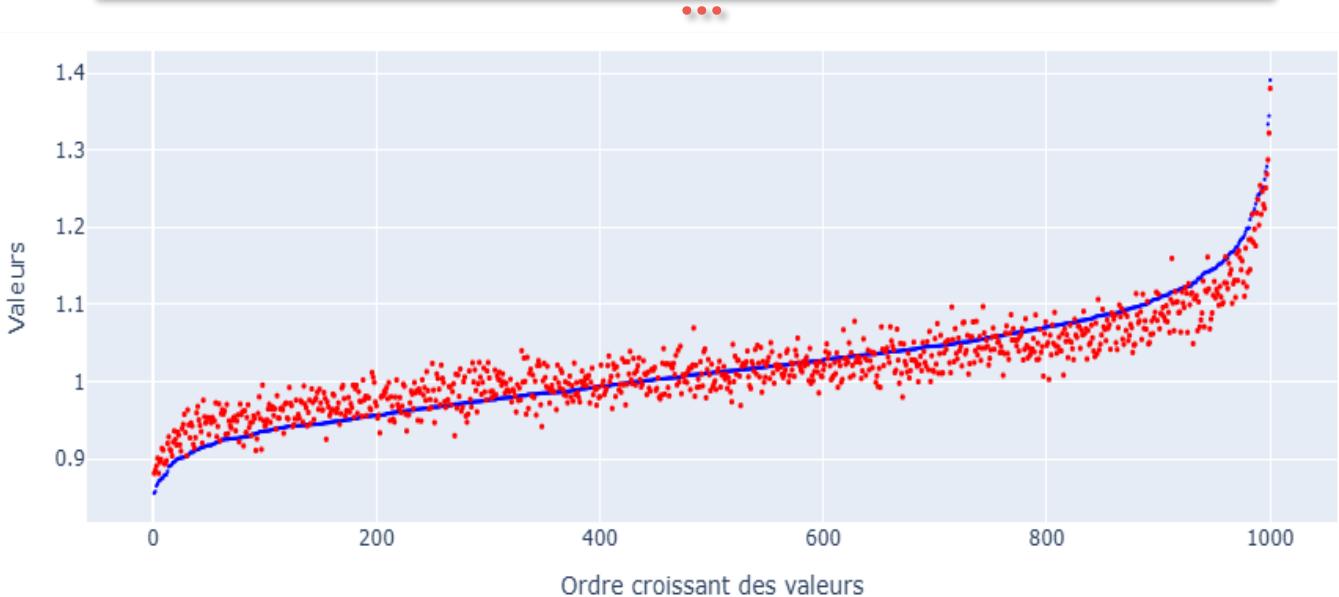
# Approche n°1 : la naïve

192 modèles testés par cross validation



# Approche n°1 : la naïve

Le meilleur modèle est un *Gradient Boosting*



... mais ne se généralise pas

	Scénarios				
	Central	2000 trajectoires	Baisse des taux	Hausse des rachats	Hausse des couts
BE prédit dans IC	✓	✓	✗	✓	✗
Score RMSE du BE	0,36	0,39	0,52	0,36	0,39
Score IC du BE	0,99	0,99	0	0,26	0

# Approche n°2 : *la complète*

Variables explicatives

Scénarios économiques

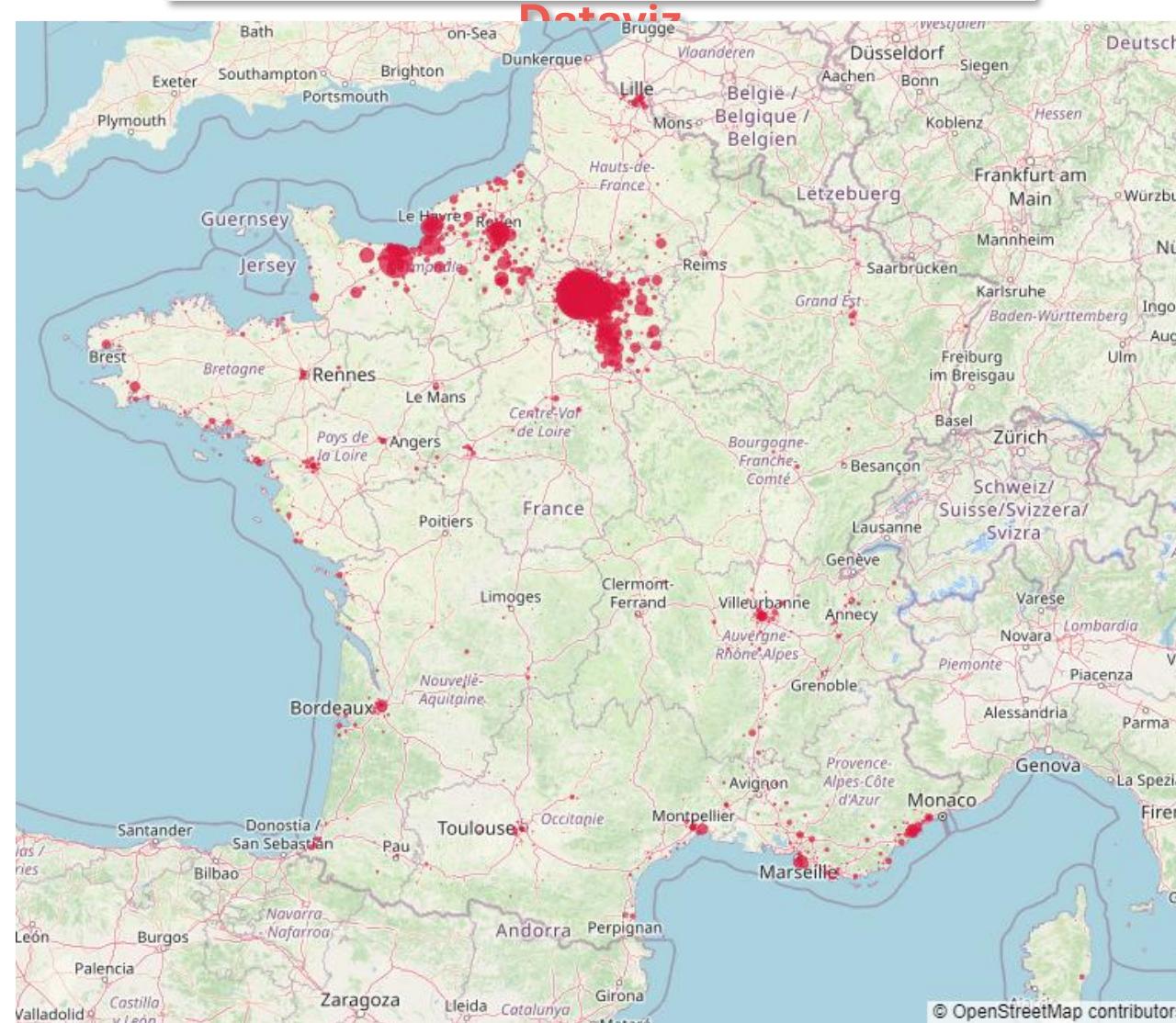
+

Données du Passif



# Approche n°2 : la complète

Data Engineering, Data Mining,



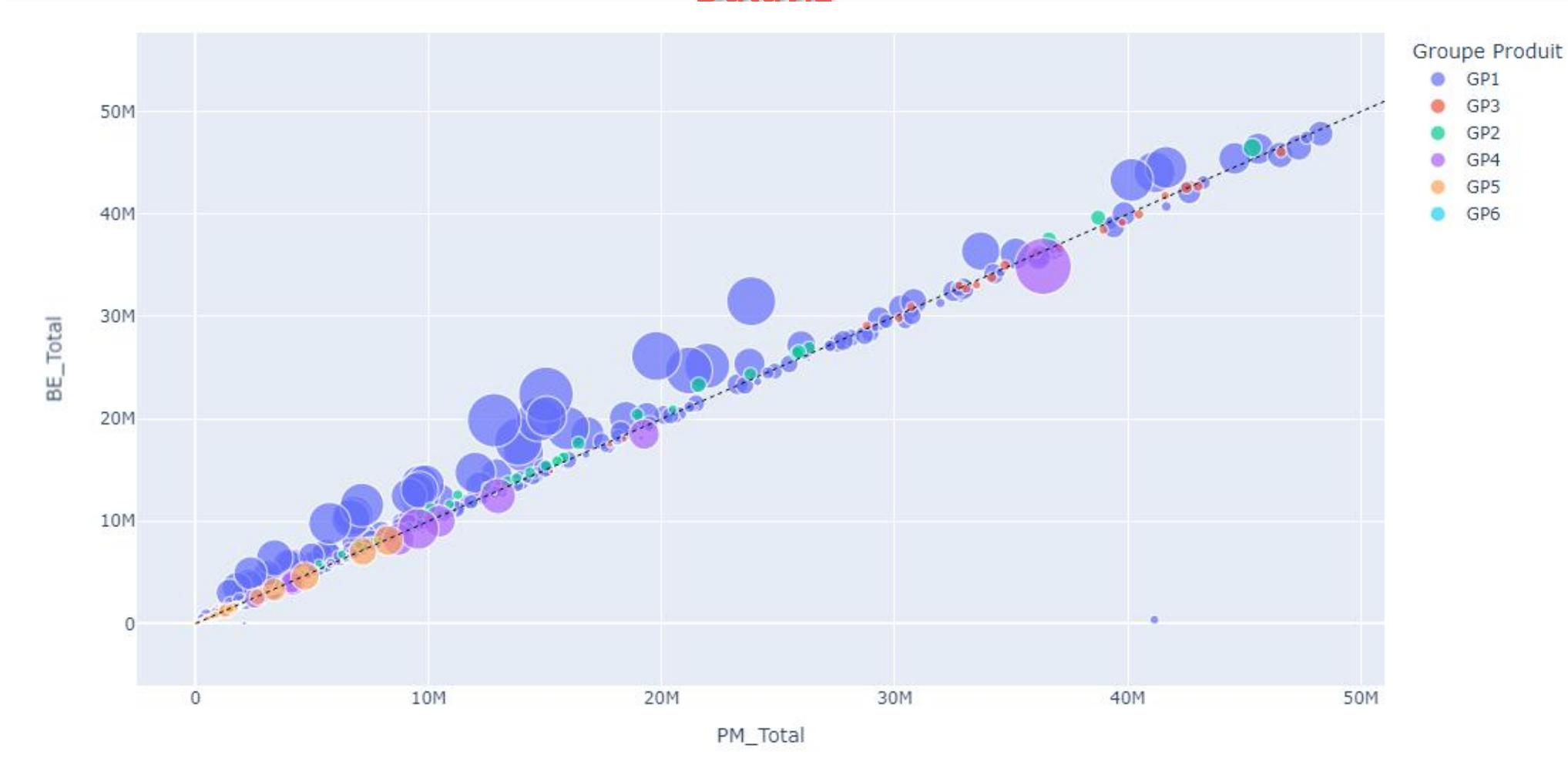
# Approche n°2 : la complète

Data Engineering, Data Mining,  
Dataviz



# Approche n°2 : la complète

Data Engineering, Data Mining,  
Dataviz



# Approche n°2 : la complète

## Machine Learning

**48 Variables explicatives**

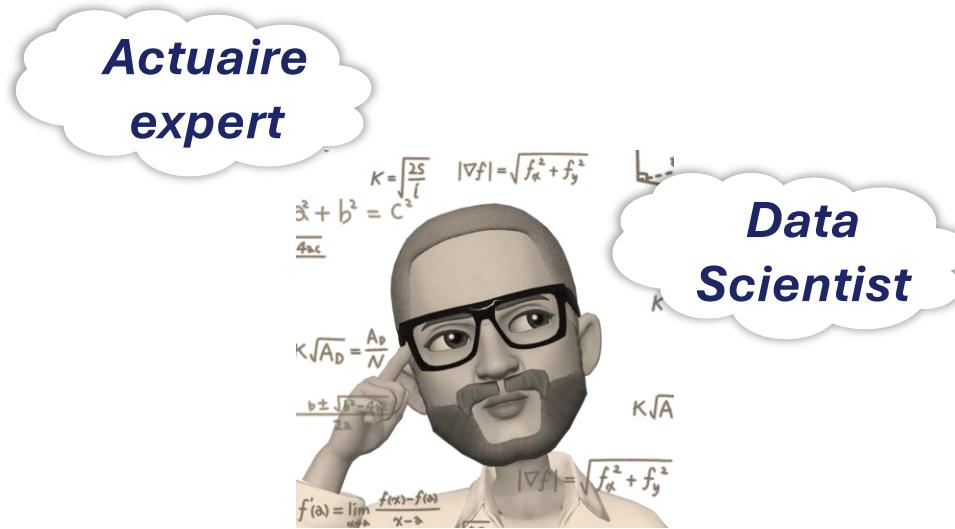
*Scénarios économiques + Model Points*

**1 544 000 observations**

*Valeur des engagements sur chaque Model Points*

**✗ Temps de cross validation > 7 heures**

# Approche n°3 : *la réflexion*



Data  
Scientist

- 1 Canton × Type de support
- 2 Facteurs d'actualisations
- 3 BE équivalent certain
- 4 Pseudo-duration

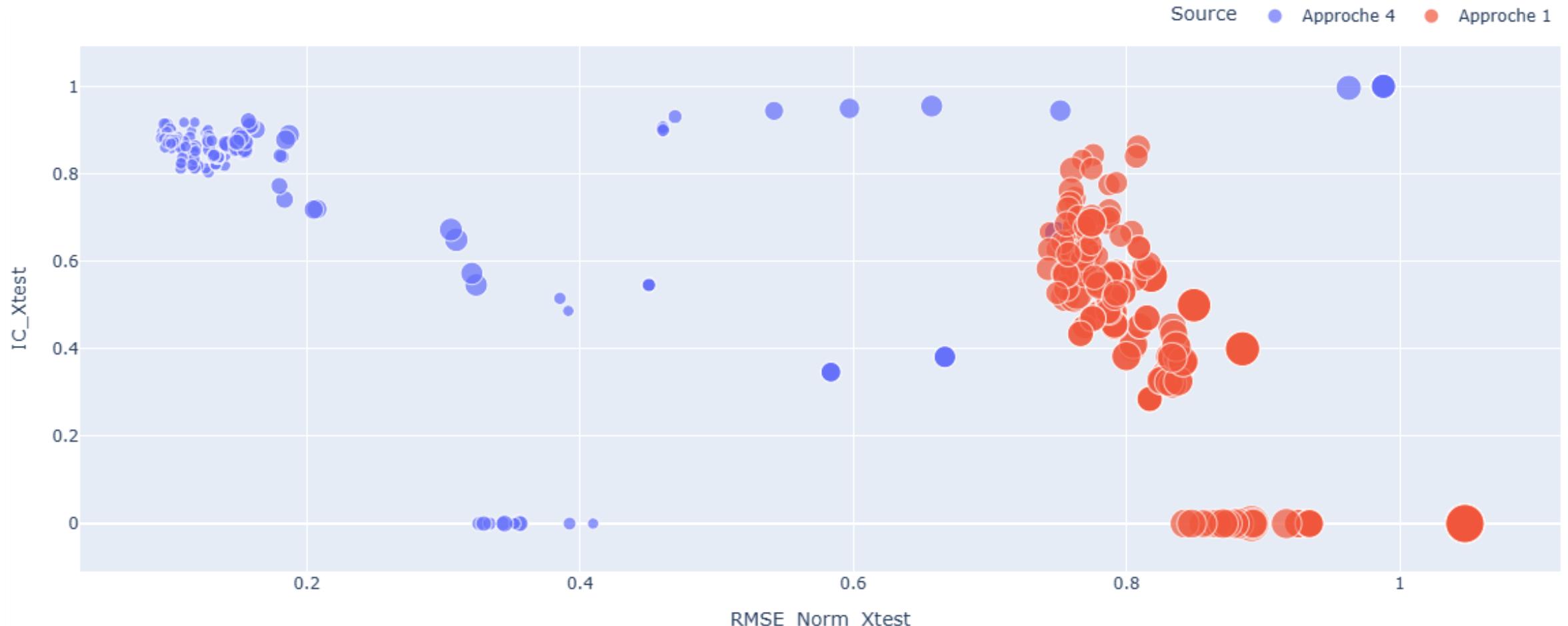
# Approche n°3 : la réflexion

Est-ce de bonnes idées ?



# Approche n°3 : la réflexion

Approche n°3 meilleure que l'approche n°1

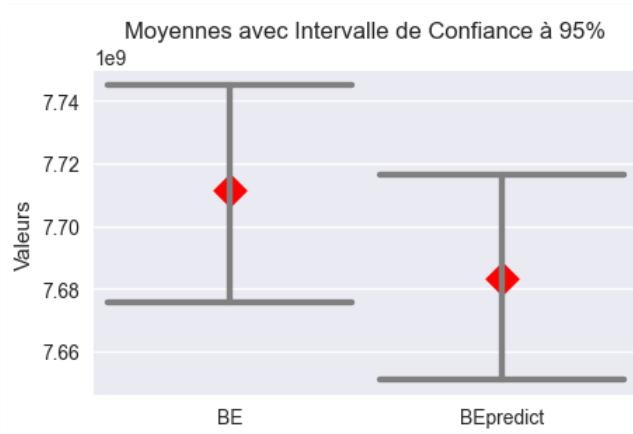


# Approche n°3 : la réflexion

Ce n'est pas encore ça !

Meilleur modèle  
**XGBoost**

	Scénarios				
	Central	2000 trajectoires	Baisse des taux	Hausse des rachats	Hausse des couts
BE prédit dans IC	✓	✓	✓	✗	✓
Score RMSE du BE	0,03	0,12	0,14	0,18	0,02
Score IC du BE	0,97	0,75	0,29	0,00	0,95



# Approche n°4 : *la segmentation*

8  
Cantons

1 000  
observations

16  
Variables  
explicatives



Scénarios	BE prédit dans IC							
	Ptf1_EURO	Ptf1_UC	Ptf2_EURO	Ptf2_UC	Ptf3_EURO	Ptf3_UC	Ptf4_EURO	Ptf4_UC
Central	✓	✓	✓	✓	✓	✓	✓	✓
2000 trajectoires	✓	✓	✓	✓	✓	✓	✓	✓
Baisse des taux	✓	✓	✓	✓	✓	✓	✓	✓
Hausse des rachats	✓	✓	✓	✓	✓	✓	✓	✓
Hausse des couts	✓	✓	✓	✓	✓	✓	✓	✓

# Approche n°4 : la segmentation

Le BE de l'Actif général est une régression linéaire !



# Conclusions

## Résultats



- Cas d'étude entièrement résolu
- Mise en place nouvelles analyses
- Gain de temps
- Meilleur connaissance du BE

## Difficultés



- Trouver et analyser les données
- Volume de code

## Perspectives



- Challenger l'approche en TVOG élevée
- BE simplifié pour révision S2 ou ORSA

# La transformation automatique et le machine learning pour la modélisation du risque d'incidence en arrêt de travail

Guillaume Welterlin

# Introduction: contexte et objectifs

## Contexte

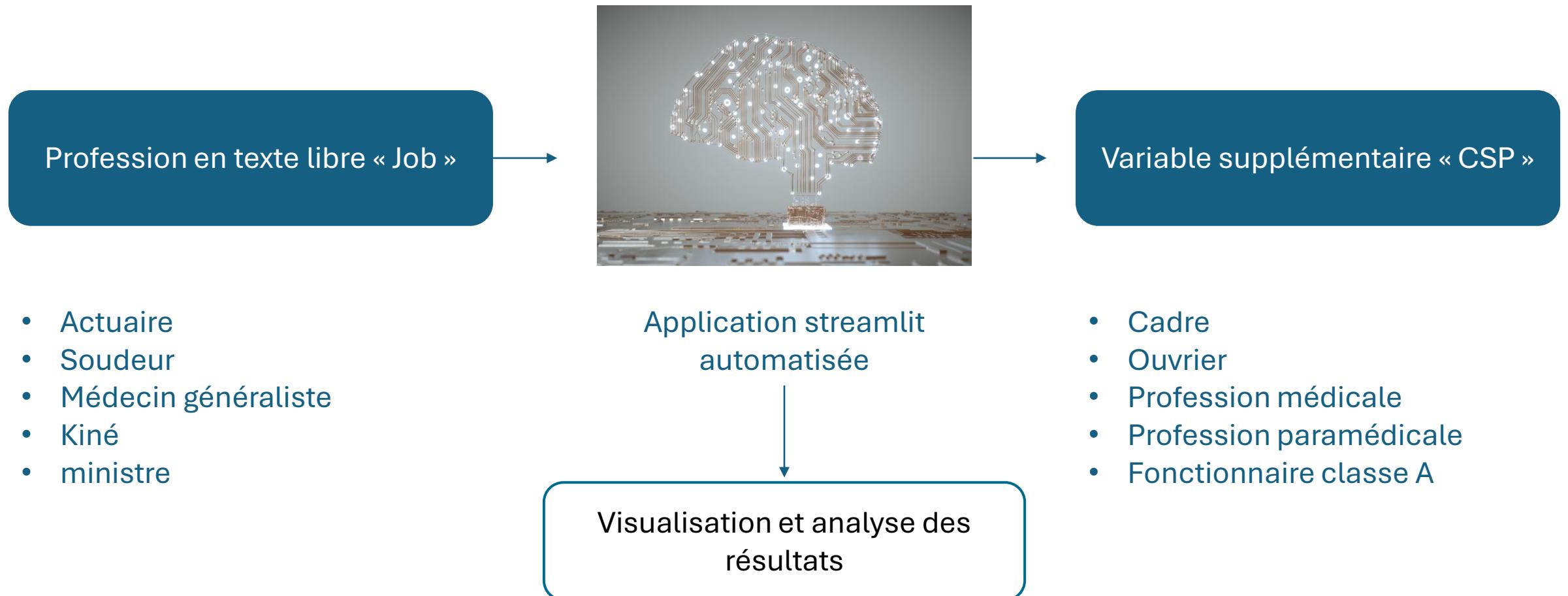
- Projet interne SCOR – Revue des tables d'expérience sur l'assurance emprunteur (Mortalité, Incidence en arrêt de travail (AT), Maintien en arrêt de travail (AT))
- Base utilisée :
  - Base de donnée assureur réassuré par SCOR
  - 600 000 assurés
  - 6 000 sinistres AT observés
- Données importantes absentes : la CSP (**catégorie socioprofessionnelle**)

## Objectifs

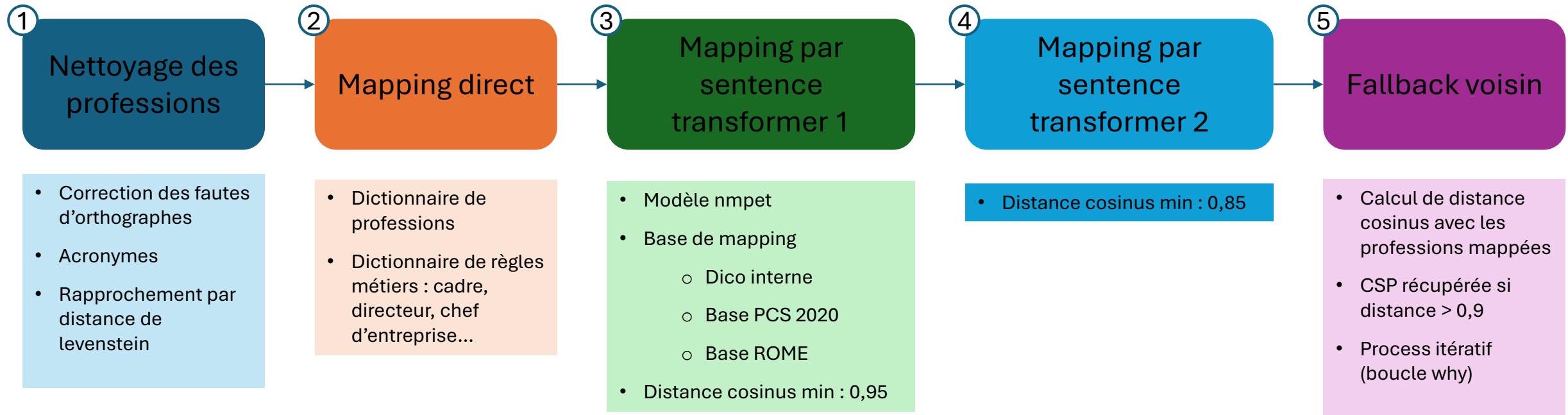
- Développement d'un algorithme qui assimile une CSP à une profession
- Création de lois d'incidence en AT (arrêt de travail)
- Challenger la méthode traditionnelle avec des algorithmes de Machine Learning

# 1. Traitement automatique des professions

# Principe de l'application streamlit



# Un pipeline automatisé de transformation des profession CSP



# Résultats de l'application sur la base de données étudiée

## Input : base de données initiale nettoyée

600 000 assurés

26 000 professions distinctes



## Résultat

Mapping direct

Sentence transformers

Total

% Nombre d'assurés

90 %

8,5 %

98,5 %

% Nombre de professions

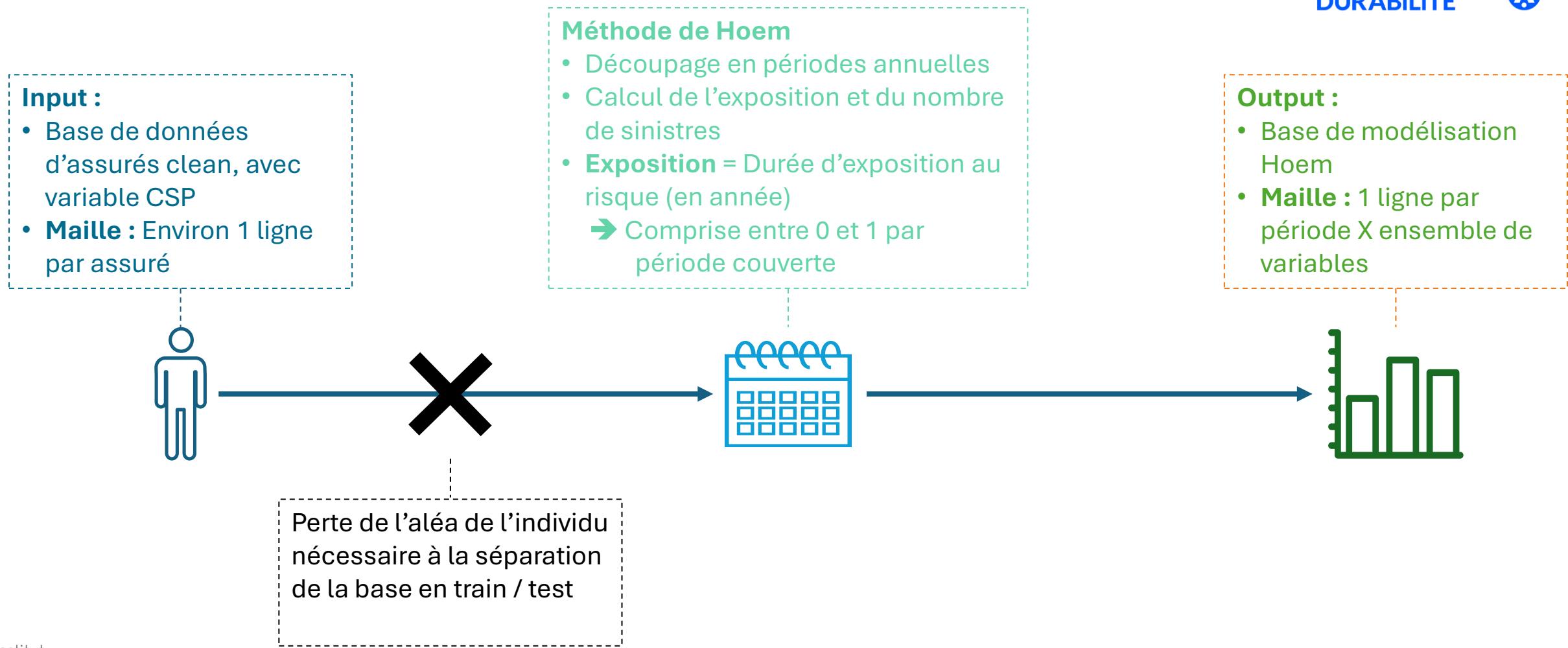
65 %

30 %

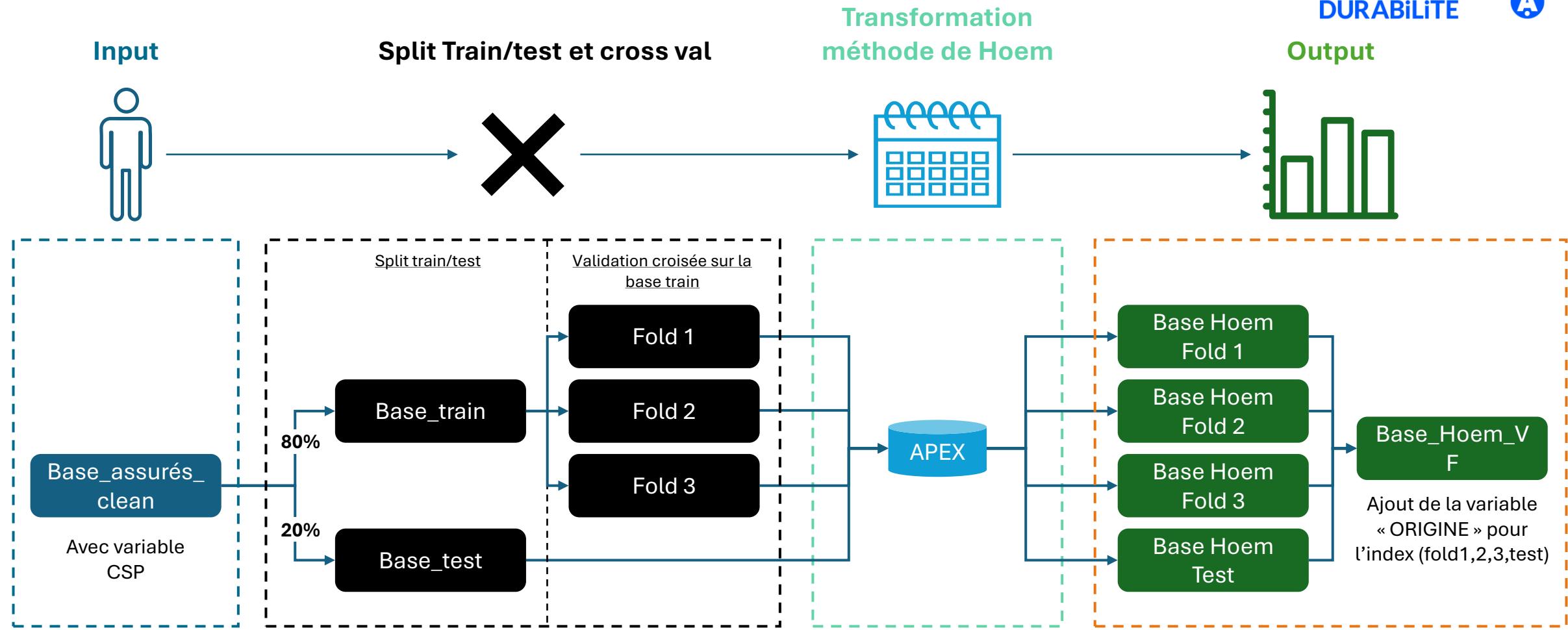
95 %

## **2. Construction de la base de modélisation**

# Transformation de la base individuelle en base Hoem pour la modélisation temporelle



# Séparation des bases tests et train avant transformation de Hoem



# Transformation de la base Hoem pour la modélisation

1. Sélection des variables impactantes pour le risque

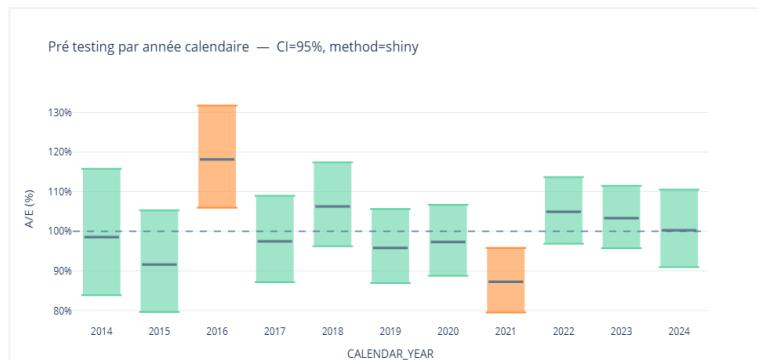
2. Identification de l'absence de leakage par année calendaire

3. Réduction du nombre de modalités

## Variables sélectionnées :

Age atteint, Sexe, CSP,  
Ancienneté dans la police,  
Option MNO, Statut fumeur,  
Région géographique,

## Pas de tendance par année calendaire



Nombre de sinistres : 5,2 K

Exposition : 1,3M

Nombre initiale de lignes : 4 M

Expo moyenne : 0,3

Variables	Modalités avant	Modalités après
Ancienneté	13	3
CSP	26	3
Région	14	3

Nombre de sinistres : 5,2 K

Exposition : 1,3M

Nombre initiale de lignes : 43 000

Expo moyenne : 30

### 3. Calibration des modèles

# Modélisation – GAM et XGB



Variable cible Y : Nombre de sinistres, avec offset log(exposition)

Critères	GAM – Poisson	XGB - Poisson
Type de modèle	Statistique – GLM étendue avec splines	Machine Learning – Arbres ensemblistes
Offset Exposition	Pris en compte log(expo) <span style="color: green;">+</span>	Pris en compte – objectif : count et base_margin <span style="color: green;">+</span>
Effets non linéaires	Oui mais via splines à paramétrier <span style="color: red;">-</span>	Capturés automatiquement <span style="color: green;">+</span>
Interactions entre variables	Non – à spécifier <span style="color: red;">-</span>	Détectées implicitement <span style="color: green;">+</span>
Interprétabilité	Forte – Coefficient et courbe <span style="color: green;">+</span>	Limitée – Analyse SHAP nécessaire <span style="color: red;">-</span>

- Hyperparamètres avec Optuna sous python
- Métrique : RMSE pondérée par l'exposition

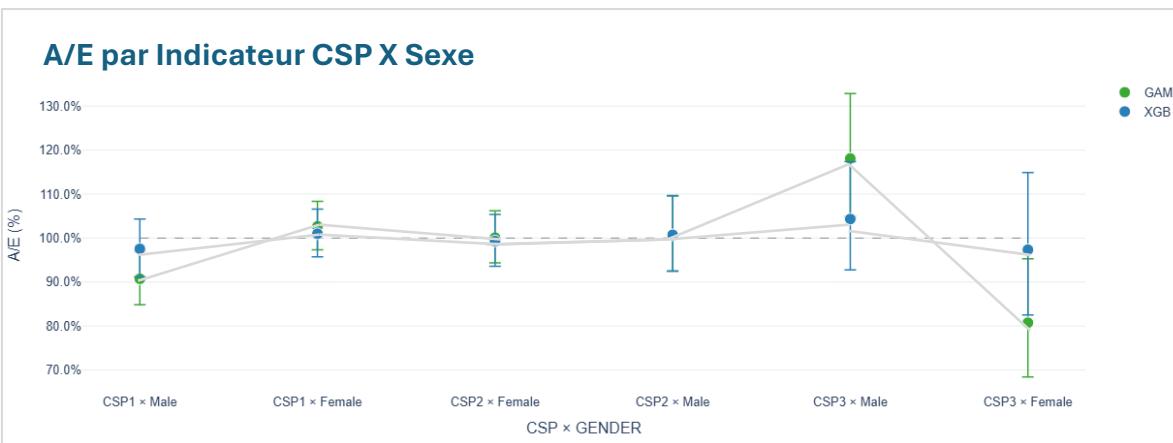
# Comparaison des résultats – Base train univariée

Modèle	RMSE pondérée expo	A/E
GAM	0,022364	0,9911
XGB	0,022355	0,9991



→ XGB plus performant mais bon résultat pour les deux modèles. A/E entre 98 et 102%

# Comparaison des résultats – Base train bivariés



**XGB significativement meilleur en bivarié**

- S/P toujours proche de 100%

**GAM est beaucoup moins précis en bivarié**

- A/E entre 80% et 110%

## 4. Interprétabilité des modèles et création d'un modèle XGB as if

# Interprétabilité Univarié – Variable qualitative

Variables	Coeff GAM	Coeff XGB	XGB/GAM -1
CSP 2 vs 1	2,10	2,26	8%
CSP3 vs 1	3,59	3,88	8%
Durée 0-1 ans	0,75	0,8	7%
Durée 2-3 ans	0,97	0,98	1%
Capital > 400k	0,94	0,95	1%
Région Ile de France	0,43	0,44	2%
Région Bretagne/PACA	1,27	1,24	-2%
Sans option MNO	0,6	0,62	3%
Fumeur	1,43	1,36	-5%

Variations importantes  
=> Captées par des interactions avec XGB

$$\text{Coeff GAM} = \exp(\beta_m)$$

$$\text{Coeff XGB} = \exp(\text{Shap}_m - \text{shap}_{ref})$$

# Des coefficients d'interaction qui améliore la précision du modèle XGB

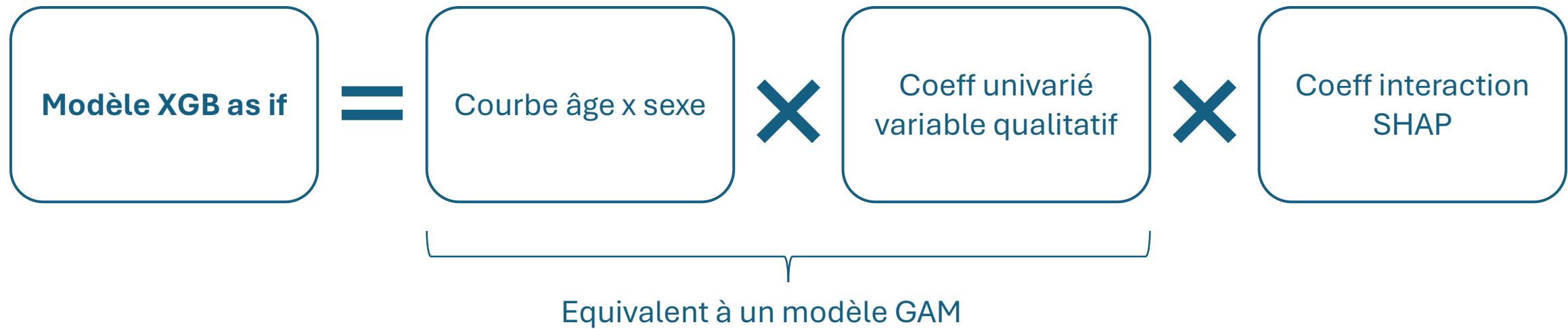
CSP / Sexe	Hommes	Femmes
CSP 1	0,99	1,01
CSP 2	1,0	0,99
CSP 3	1,08	0,93

CSP / MNO	0	1
CSP 1	0,98	1,01
CSP 2	1,03	0,99
CSP 3	1,04	0,98

MNO / Sexe	0	1
CSP 1	0,98	0,97
CSP 2	1,04	1,02

- Ecart -7% à +8% selon CSP X Sexe X MNO
- Calcul grâce aux interactions SHAP

# Création d'un modèle XGB as if exploitable sous Excel



Coefficients exportables sous Excel comme pour le GAM

# Impact du choix de modèle sur une étude de rentabilité

## Statistiques du portefeuille



## Résultats

$$\frac{PV \text{ claim XGB}}{PV \text{ claims GAM}} - 1 =$$

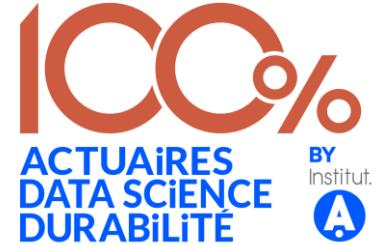
CSP\Sexe	CSP 1	CSP 2	CSP 3	Total
<b>Hommes</b>	-10%	-2%	4%	-9%
<b>Femmes</b>	0%	+3%	-5%	0%
<b>Total</b>	-5%	1%	-1%	-4%

Ecart important sur les hommes, principalement CSP 1 → impact modèle significatif sur la tarification

# CIA : Quand l'actuariat entre en mission... d'innovation !

**Amandine Gomez**

# Transition technologique : Bénéficier de l'IA Générative pour innover



## Actuariat : pourquoi est-ce difficile d'innover?



Beaucoup de travaux récurrents qui laissent peu de temps à l'innovation

*Obligations réglementaires, comptables et concurrentielles avec des cycles répétitifs, des contrôles stricts et peu de marge pour l'innovation.*



Manque d'appétence pour les outils data/IA avancés

*Une option, pas une priorité (en général 😊)*



Des experts, mais un peu chacun dans sa bulle

*Cloisonnement entre branches, peu de transversalité*



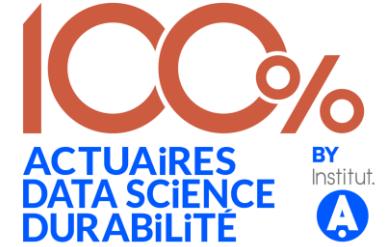
Un écosystème en mouvement permanent

Des résultats concrets encore à prouver



Beaucoup de hype

# Embarquer les équipes : Cellule innovation actuarielle



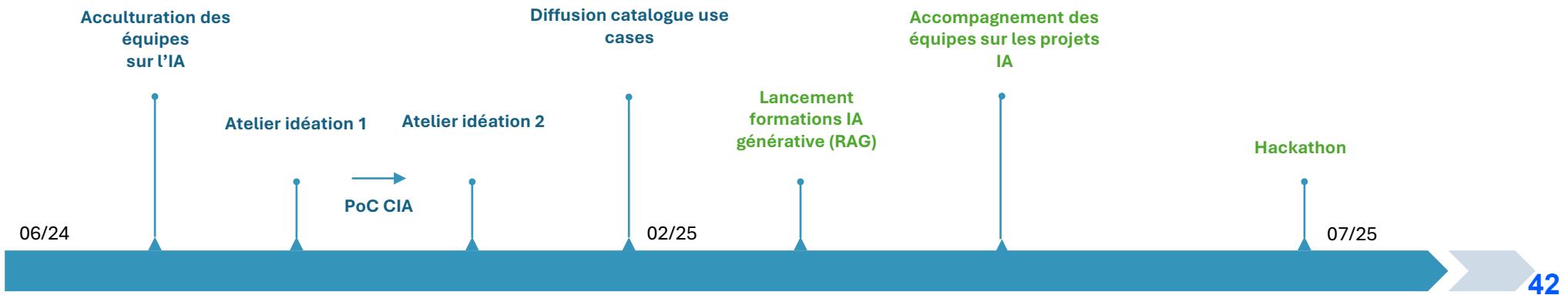
## Création de la CIA

- Un **collectif** de 16 personnes
- Représentants de **chacune des directions techniques**

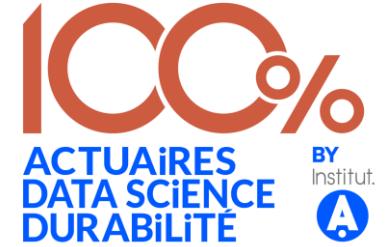


**OBJECTIFS :** Moteur d'innovation en actuariat

- Encourager la coopération entre les différentes équipes
- Soutenir le développement des compétences au sein des équipes
- Promouvoir la transformation à travers l'expérimentation
- Favoriser le rapprochement entre les métiers et l'IT



# Rex à 1 an: Success story de la CIA



## KPI

- **25 présentations** ( équipes, ateliers d'idéation, Codir, conférences internes et externes)
- **27 cas d'usage** identifiés dont 9 accompagnés et validés par l'IT
- **90 personnes** formées à la RAG soit 73% des effectifs éligibles
- Organisation d'un **hackathon** impliquant 8 équipes, 40 stagiaires et 8 encadrants de la CIA
- **Extension** aux directions Risques et Finance

### TRANSVERSALITÉ

Richesse des échanges dû à la transversalité du collectif et ouverture à plusieurs directions

### ENGAGEMENT

Réalisations des travaux et engagement des collaborateurs supérieurs à l'attendu

### DYNAMISME

Bonne dynamique dans le collectif et soutien/sponsor de la direction

# Formation Data Science pour l'assurance

Pour nous contacter ou vous inscrire : [formation@institut-du-risk-management.fr](mailto:formation@institut-du-risk-management.fr)

Dites-le-moi et je l'oublierai

Montrez-moi et peut-être que je m'en souviendrai

**Nous vous guiderons, nous vous impliquerons,  
Vous serez certifié et vous déployerez dans votre entreprise**