



WE'LL TAKE YOU FURTHER

## Détection d'anomalies non supervisée en souscription

**Nabil RACHDI**

**Kezhan SHI**



# Sommaire

- I. Contexte et Objectifs**
- II. Approches stratégique et méthodologique**
- III. Cas d'application ALLIANZ**
- IV. Conclusion et perspectives**

# Sommaire

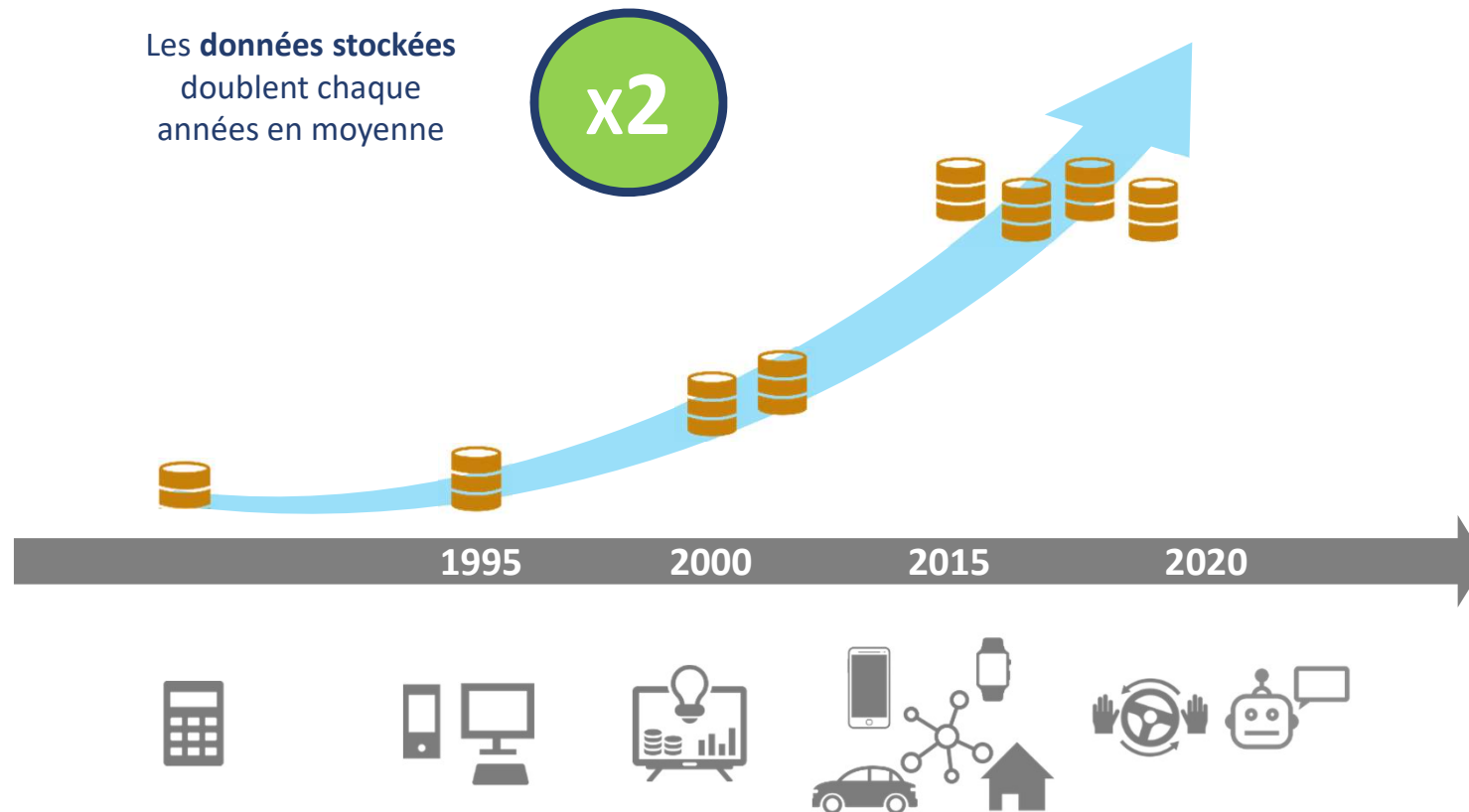
- I. Contexte et Objectifs**
- II. Approches stratégique et méthodologique
- III. Cas d'application ALLIANZ
- IV. Conclusion et perspectives



# ➤ Contexte et Objectifs

## Outils digitaux et vulnérabilité des données

- L'usage croissant d'outils digitaux rend **vulnérable** la masse de données à traiter



CONFIDENTIAL ©2018 ACTUARIS

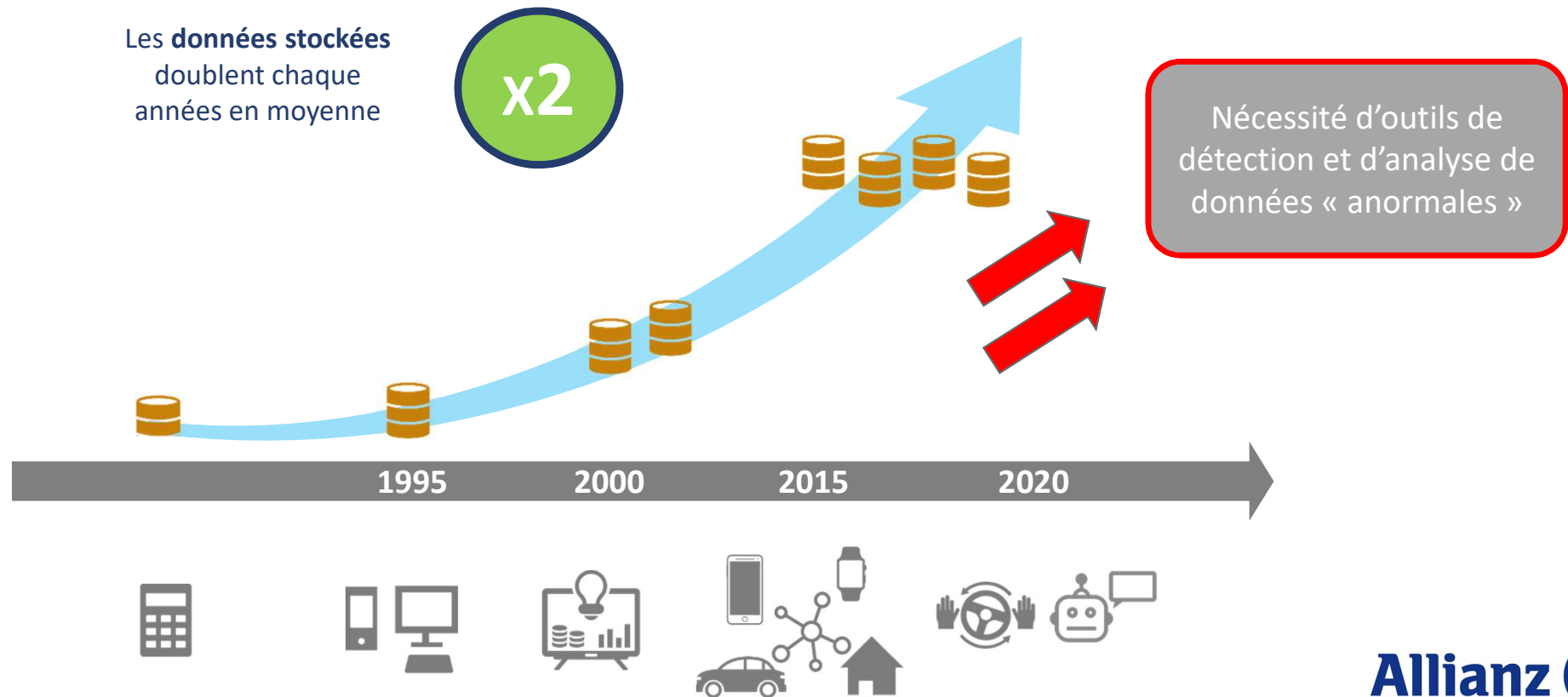
Allianz 

  
ADDACTIS Group  
Risk & insurance solutions

# ➤ Contexte et Objectifs

## Outils digitaux et vulnérabilité des données

- L'usage croissant d'outils digitaux rend **vulnérable** la masse de données à traiter



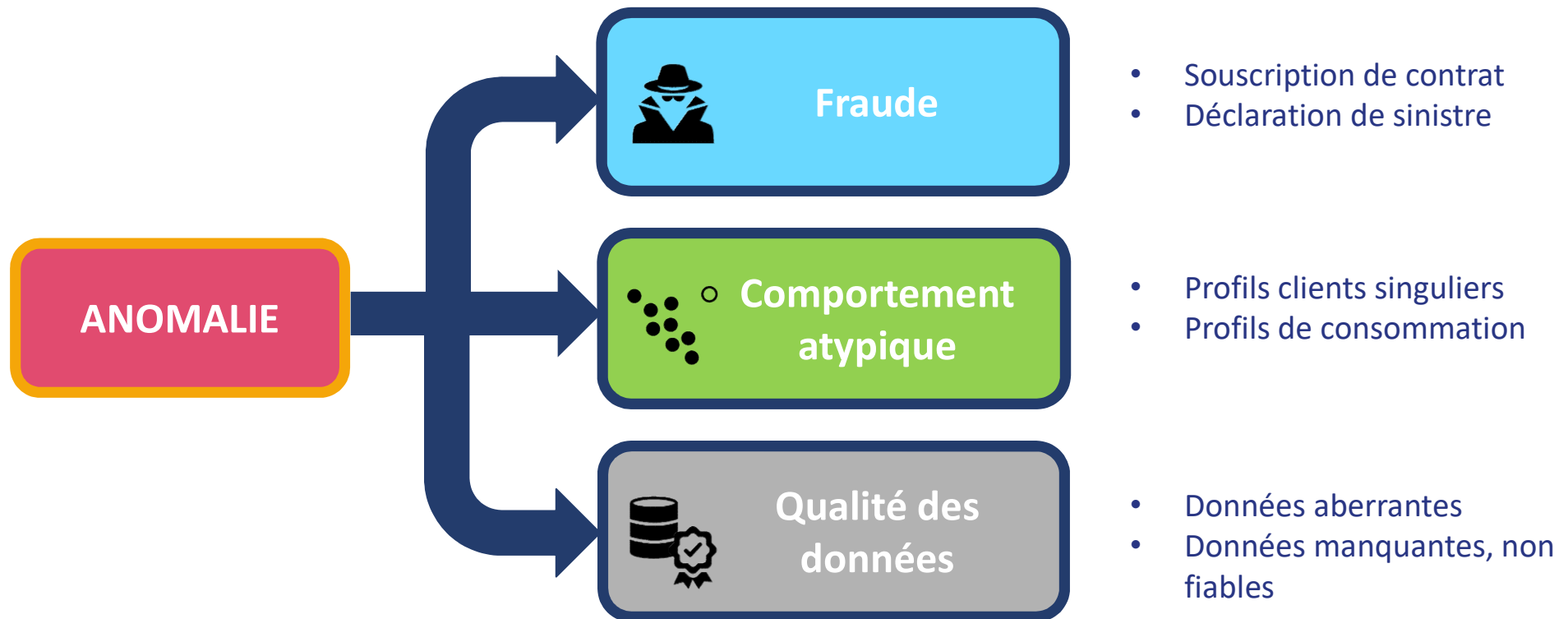
Allianz 

  
ADDACTIS Group  
Risk & insurance solutions

# ➤ Contexte et Objectifs

## La détection d'anomalies

- Cette vulnérabilité donne lieu à des **anomalies** pouvant correspondre à différentes causes



CONFIDENTIAL©2018 ACTUARIS

# ► Contexte et Objectifs

## Des enjeux économiques et réglementaires importants

- La détection d'anomalies représente des enjeux économiques et réglementaires importants



Coût estimé (% primes IARD) du à la **fraude** à l'assurance dommage en France, soit **2.5Md€** (rapport Alfa de 2015)



Impact que peut avoir des **erreurs** ou **imprécision des données** sur le calcul d'un SCR



**Sanction financière** pouvant atteindre **4% du CA mondial** annuel d'une entreprise en cas de **manquement à la RGPD**

# ► Contexte et Objectifs

## Quelques applications de la détection d'anomalies

### ■ Fraude à la transaction financière (cartes de crédit)



#### Indicateurs à dispositions

- Temps de transaction
- Montant de transaction
- Lieux
- Date
- Etc.

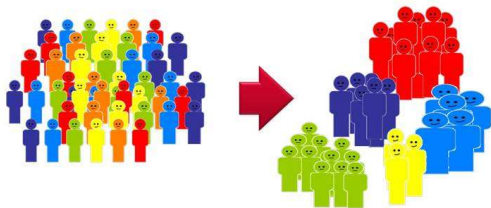
### ■ Qualité des données



#### Analyse à 4 niveaux

- Exactitude
- Conformité
- Complétude
- Consistance

### ■ Segmentation, comportements atypiques



Segmentation de données  
(individus, contrats, etc.)

en fonction de **caractéristiques** (consommation, sinistralité, etc.)



# ➤ Contexte et Objectifs

## Objectifs

Les principaux objectifs de cet atelier sont de

- Présenter l'approche d'ACTUARIS pour la détection d'anomalies par la Data Science
- Présenter une méthodologie pour définir la « normalité » et identifier les facteurs impactant qui conduisent à l'« anormalité »
- Démontrer la méthodologie sur un cas d'étude ALLIANZ

# Sommaire

- I. Contexte et Objectifs
- II. Approches stratégique et méthodologique**
- III. Cas d'application ALLIANZ
- IV. Conclusion et perspectives

# ➤ Approche stratégique et méthodologique

## Approche stratégique



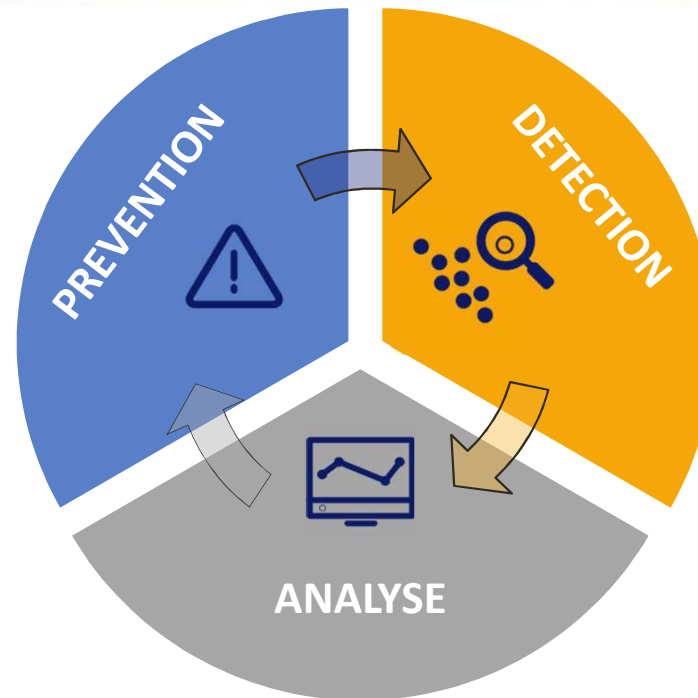
Des contrôles **aléatoires** des échantillons aux contrôles **exhaustifs** sur des points spécifiques

### Plusieurs approches:

- Approche « manuelle »: contrôles manuels aléatoires
- Approche « Big Data »
  - Détection des anomalies ou des points des non-conformités
  - Demande de correction avec identification des anomalies
  - Récupération financières avec la remise en conformité
- Approche automatisation des tâches spécifiques
  - Analyse des tâches spécifiques des gestionnaires
  - Approche méthodologique (déterministe, supervisé, et non supervisé)
  - Envoi massif et correction par les intermédiaires (avec les assurés)

# ➤ Approche stratégique et méthodologique

## Approche stratégique



- Changement de paradigme
  - Prévention : éviter que les anomalies puissent avoir lieu
  - Métier de contrôleurs vers métier de formateur
  - Sanction ponctuelle vers contrôle exhaustif (effet dissuasif)



# ➤ Approche stratégique et méthodologique

## Approche méthodologique

### 3 principales approches:

- **Détections déterministes**

- Application des règles métiers
- Application avec règles de calcul
- Comparaison avec des données externes

➔ Besoin des approches statistiques ou algorithmiques quand le nombre de variables est important

- **Apprentissage non supervisé**

- Réduction de dimensions
- Segmentation des observations

➔ Meilleure vision d'ensemble et détection des anomalies

- **Apprentissage supervisé**

- Utilisation des bases libellées pour construire des règles automatisées

➔ Combinaison de plusieurs variables pour une meilleure prédiction

**En pratique, ces trois approches peuvent être combinées**

# ➤ Approche stratégique et méthodologique

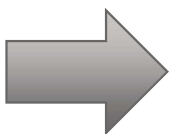
## Approche méthodologique

- 2 cadres d'analyse:

« SUPERVISE »



Les **données** à disposition sont **labellisées** de présence ou non d'anomalie



### MODELES SUPERVISES

- Qualité de labellisation de l'anomalie (degré de certitude)
- Prise en compte de la faible occurrence des anomalies (souvent moins d'1% )
- Détection d'anomalie par Machine Learning supervisé
- Identification des indicateurs ayant conduit à l'anomalie

# ➤ Approche stratégique et méthodologique

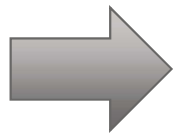
## Approche méthodologique

- 2 cadres d'analyse:

« NON SUPERVISE »



Les **données** à disposition  
sont **ne sont pas** labellisées.

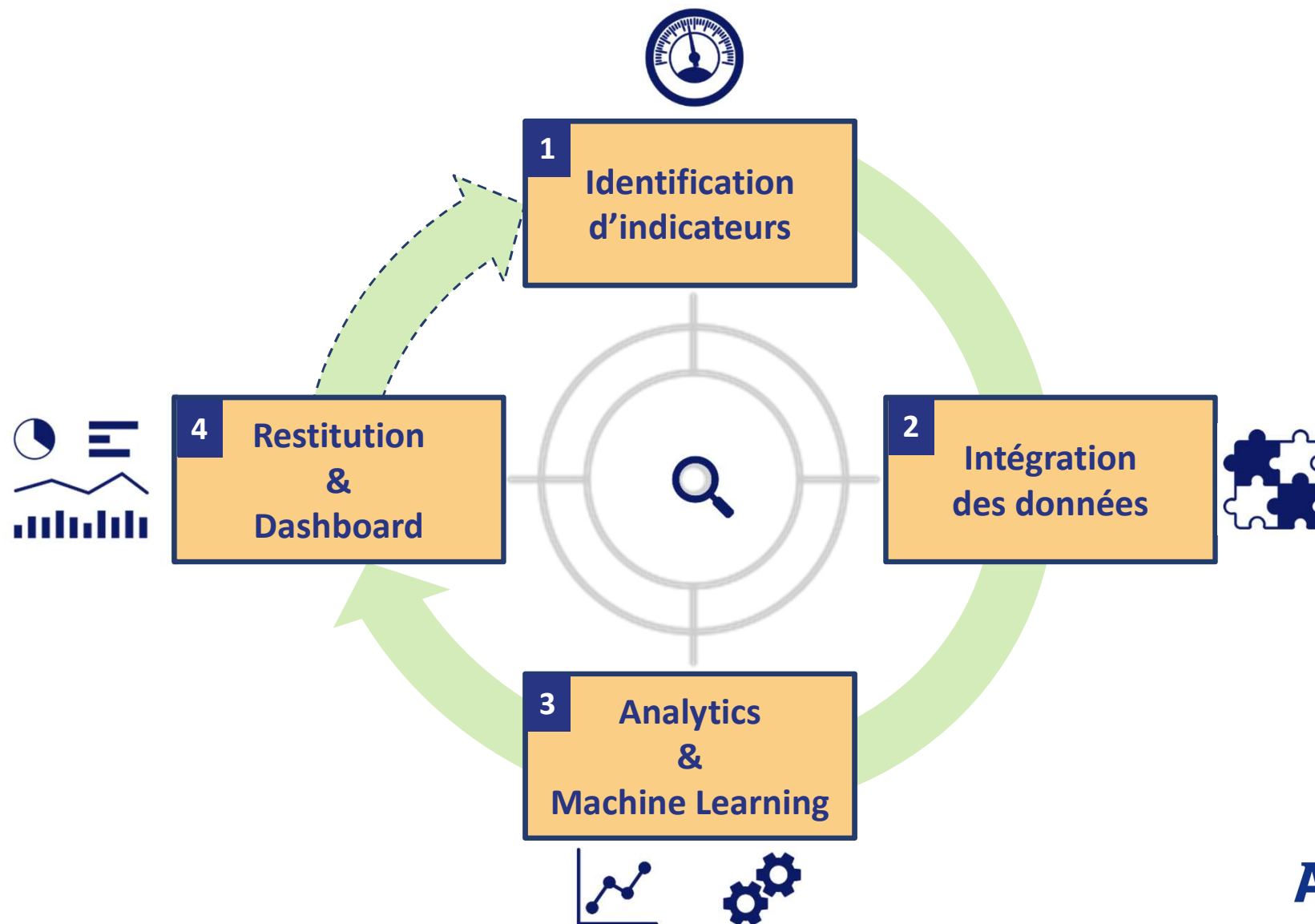


### MODELES NON SUPERVISES

- Définition de l'anomalie
- Collecte des indicateurs (données sinistres, contrats, transactions, etc. )
- Détection d'anomalies par Machine Learning non-supervisé
- Analyse des données normales vs. anormales

# ➤ Approche stratégique et méthodologique

## Approche méthodologique

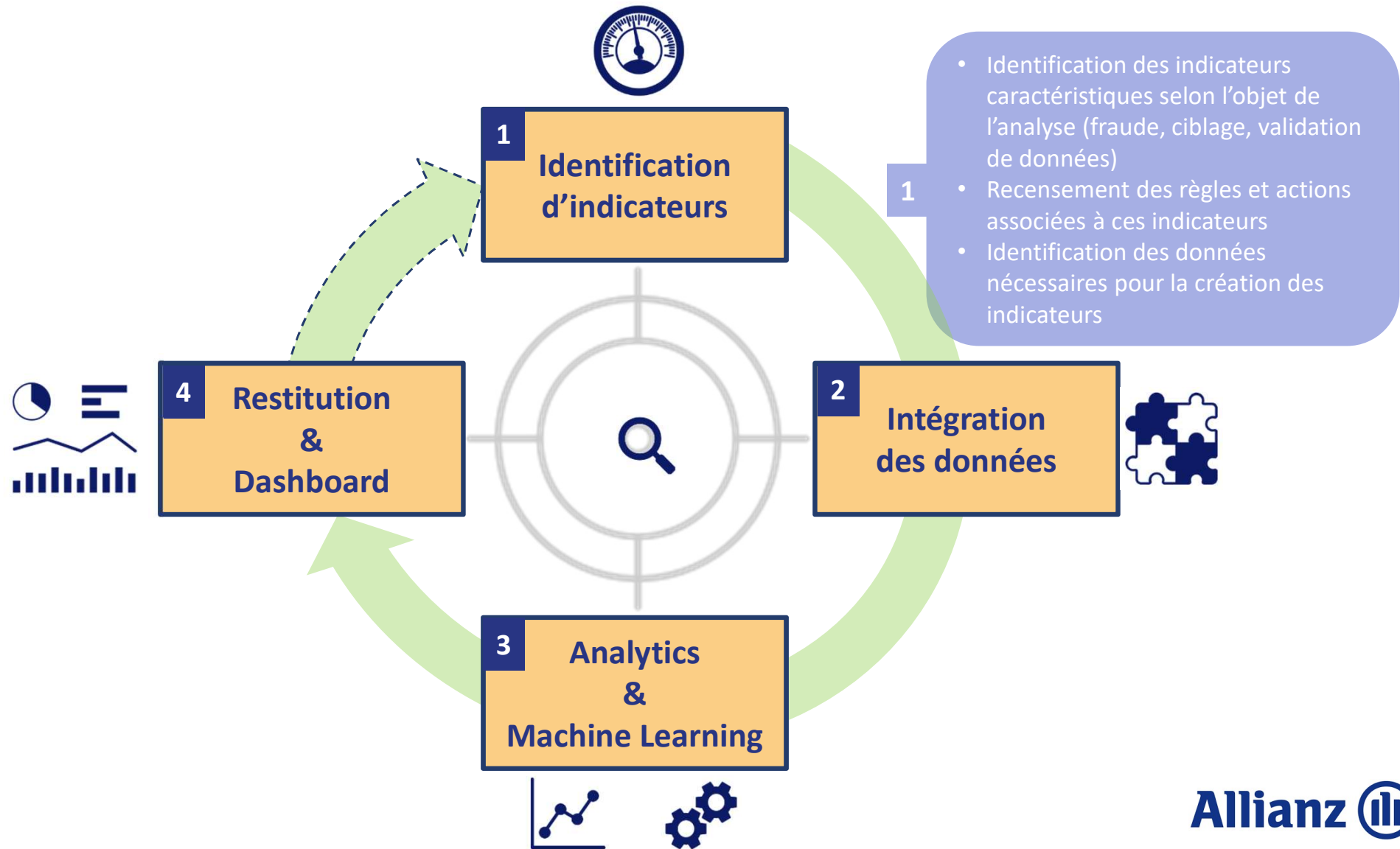


CONFIDENTIAL@2018 ACTUARIS



# ➤ Approche stratégique et méthodologique

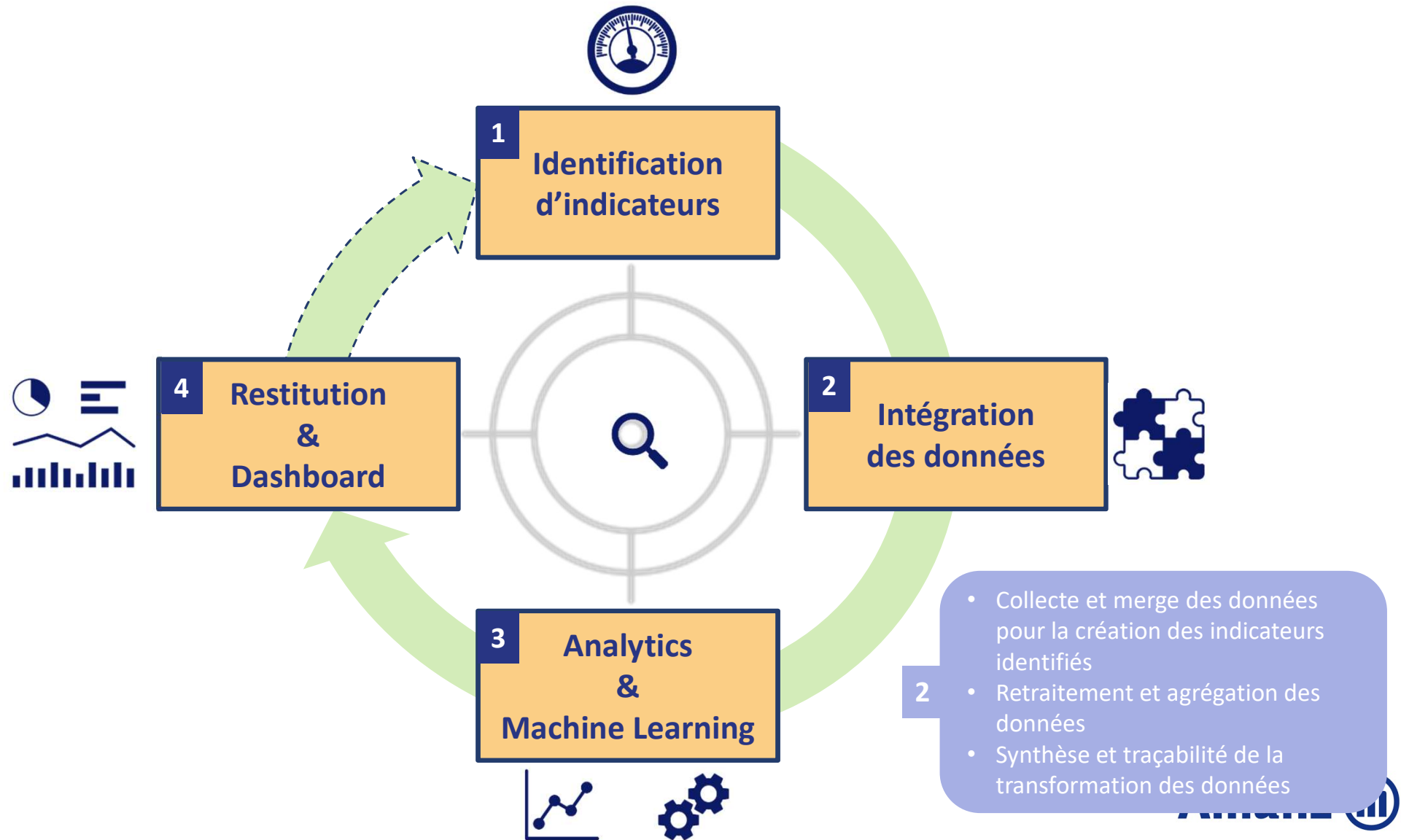
## Approche méthodologique



CONFIDENTIAL@2018 ACTUARIS

# ➤ Approche stratégique et méthodologique

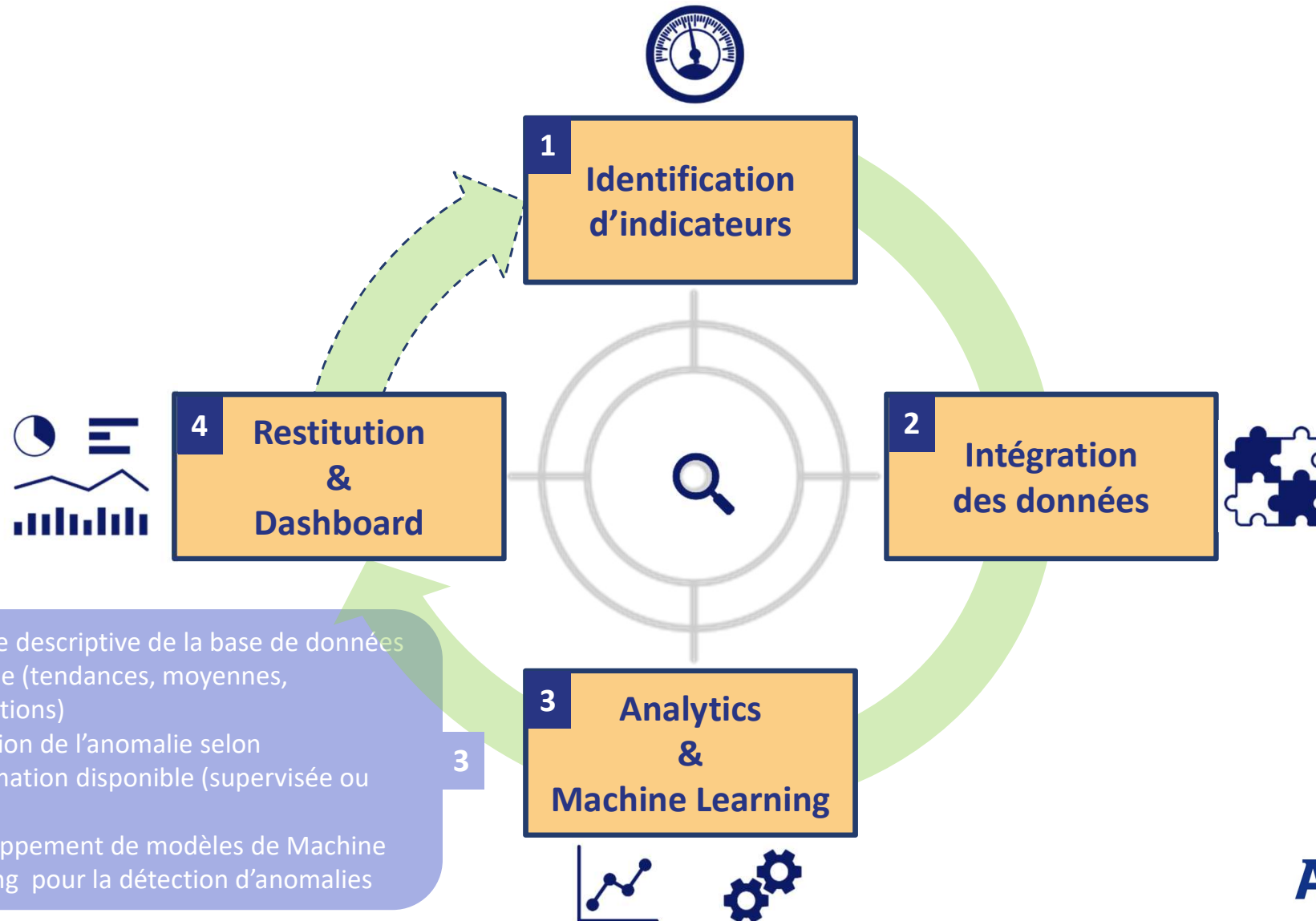
## Approche méthodologique



CONFIDENTIAL@2018 ACTUARIS

# ➤ Approche stratégique et méthodologique

## Approche méthodologique

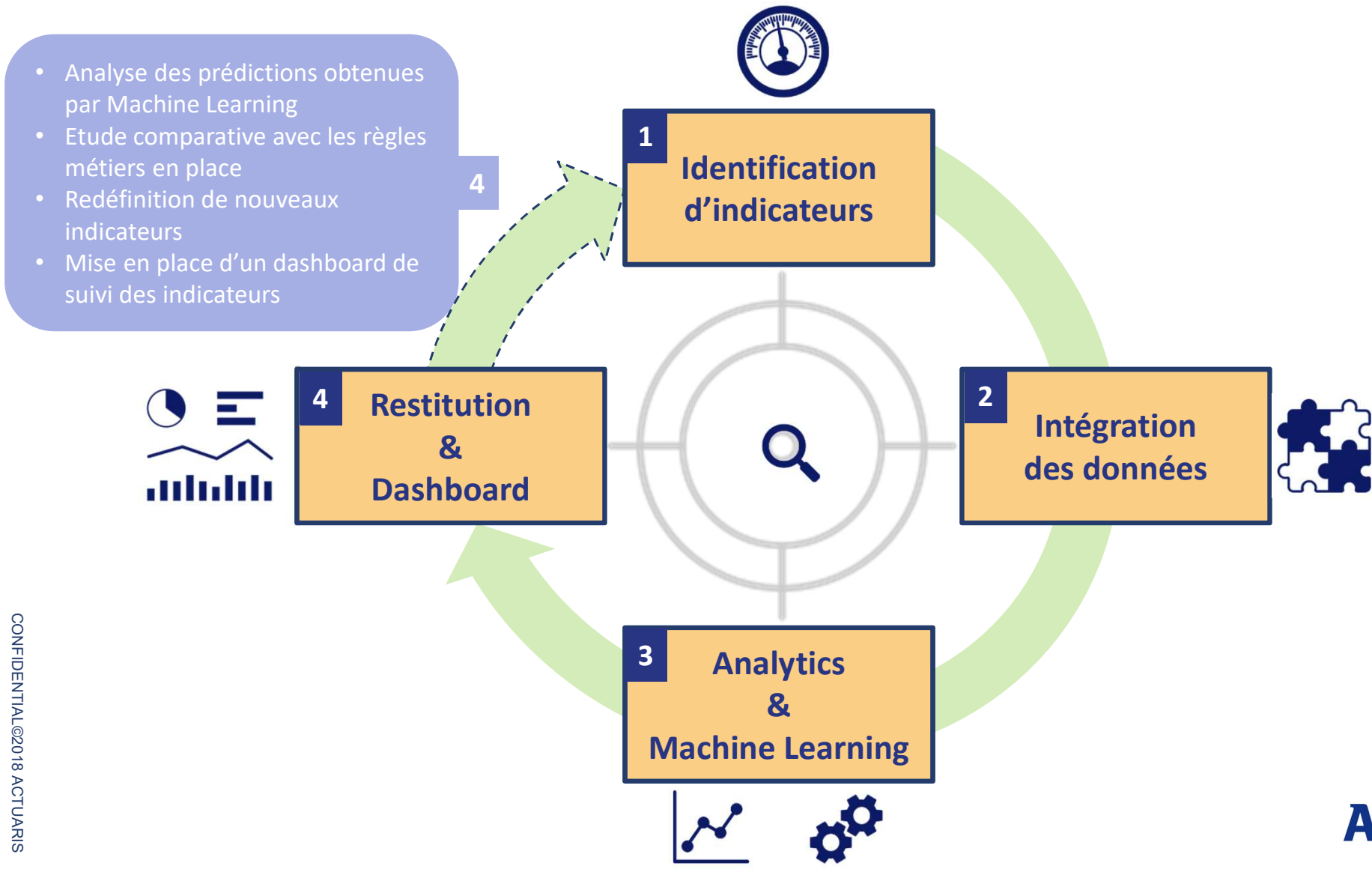


Allianz

**ACTUARIS**  
ADDACTIS Group  
Risk & insurance solutions

# ➤ Approche stratégique et méthodologique

## Approche méthodologique



CONFIDENTIAL@2018 ACTUARIS





# Sommaire

- I. Contexte et Objectifs
- II. Approches stratégique et méthodologique
- III. Cas d'application ALLIANZ**
- IV. Conclusion et perspectives

# ➤ Cas d'application ALLIANZ

## Présentation

### ■ Contexte

- Construction d'un outil de pilotage, de contrôles et de surveillance de portefeuille
- Des informations de différentes granularités
  - Vision très macro
  - Vision des indicateurs
  - Détection des contrats présentant des anomalies

### ■ Problématique

- Construction des indicateurs
  - Approche métiers
  - Approche statistique
- Pertinence des indicateurs

# ➤ Cas d'application ALLIANZ

## Résultats

- Données à disposition

1  
Identification  
d'indicateurs

2  
Intégration  
des données

### Indicateurs

	ind2	ind3	ind4	ind5	ind6	ind7	ind8	ind9	ind10	ind11	...	ind103	ind104	ind105	ind106	ind107	ind108	ind109	ind110	ind111	ind112	
A0932	2	377	257	120	278	208	70	1330997.7	713177.8	308125.30	...	0	0	0	0	0	0	0	0	0	0	0
A1529	0	170	102	68	141	70	71	499087.0	254397.1	168889.92	...	0	0	0	0	0	0	0	0	0	0	0
A0623	2	177	134	43	138	93	45	751660.1	424913.7	212714.14	...	5	2	2	0	0	3	0	1	1	0	0
A1433	1	313	205	108	203	138	65	1062142.1	568537.7	268661.69	...	0	0	0	0	0	0	0	0	0	0	0
A1610	1	244	134	110	155	104	51	685933.3	362942.5	255517.57	...	0	0	0	0	0	0	0	0	0	0	0
A1881	2	186	120	66	153	109	44	650973.2	346801.5	180227.30	...	19	10	5	0	0	0	0	0	0	0	0
A1608	3	268	151	117	139	85	54	1404774.0	455130.4	331527.78	...	0	0	0	0	0	0	0	0	0	0	0
A1401	0	84	62	22	48	38	10	334280.6	172451.4	84303.98	...	26	9	10	1	0	10	0	5	3	0	0
A1999	1	201	158	43	161	123	38	877643.1	422781.7	154770.76	...	97	25	34	2	1	36	6	7	8	1	0
A0574	2	509	358	151	306	232	74	1106335.1	717897.0	267628.04	...	0	0	0	0	0	0	0	0	0	0	0
...																						

CONFIDENTIAL@2018 ACTUARIS

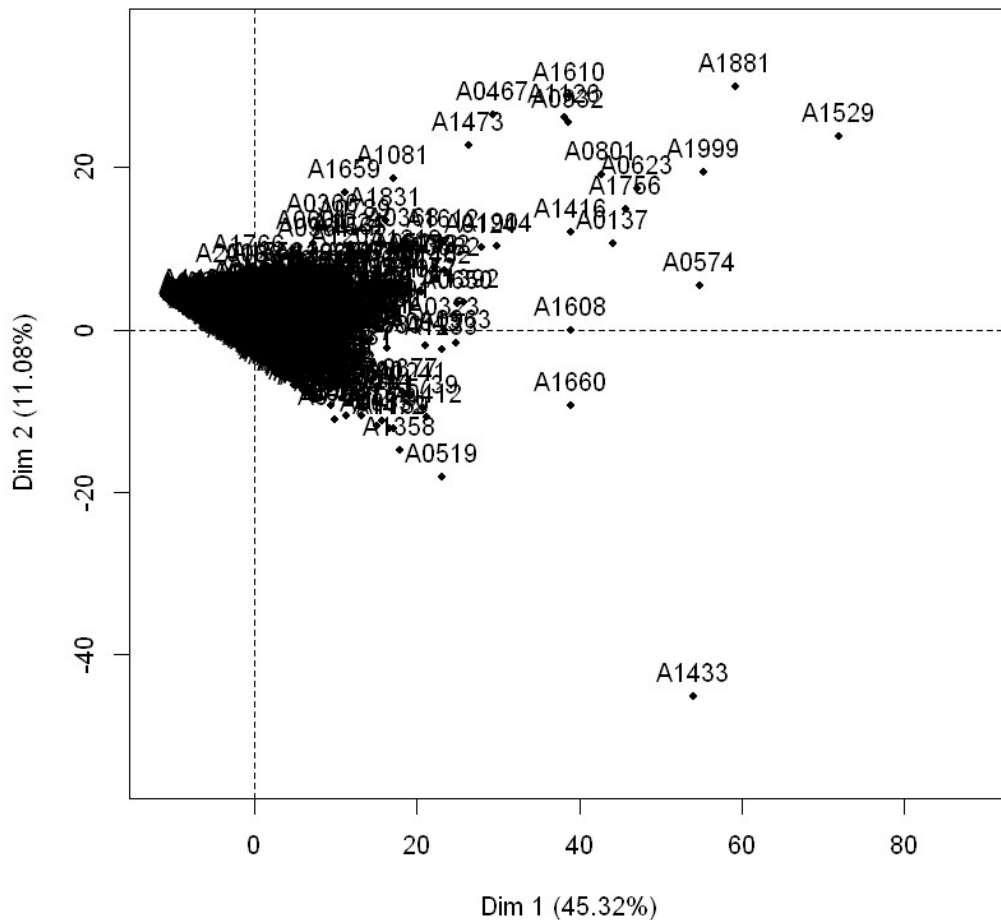
Allianz 

 ACTUARIS  
ADDACTIS Group  
Risk & insurance solutions

# ➤ Cas d'application ALLIANZ

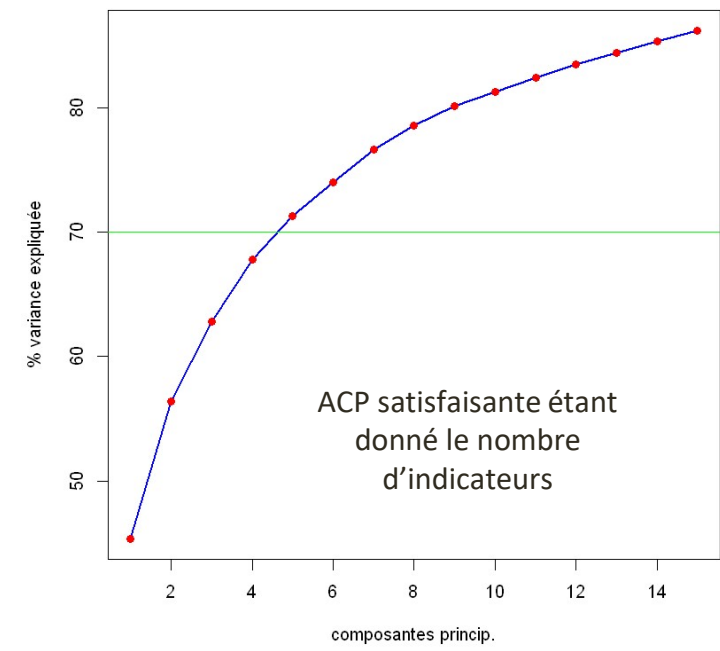
## Résultats

### ■ Réduction de dimension (Analyse en Composantes Principales)



### 3 Analytics & Machine Learning

Variance cumulée vs. Composantes princ.



➔ L'ACP révèle des disparités entre les données

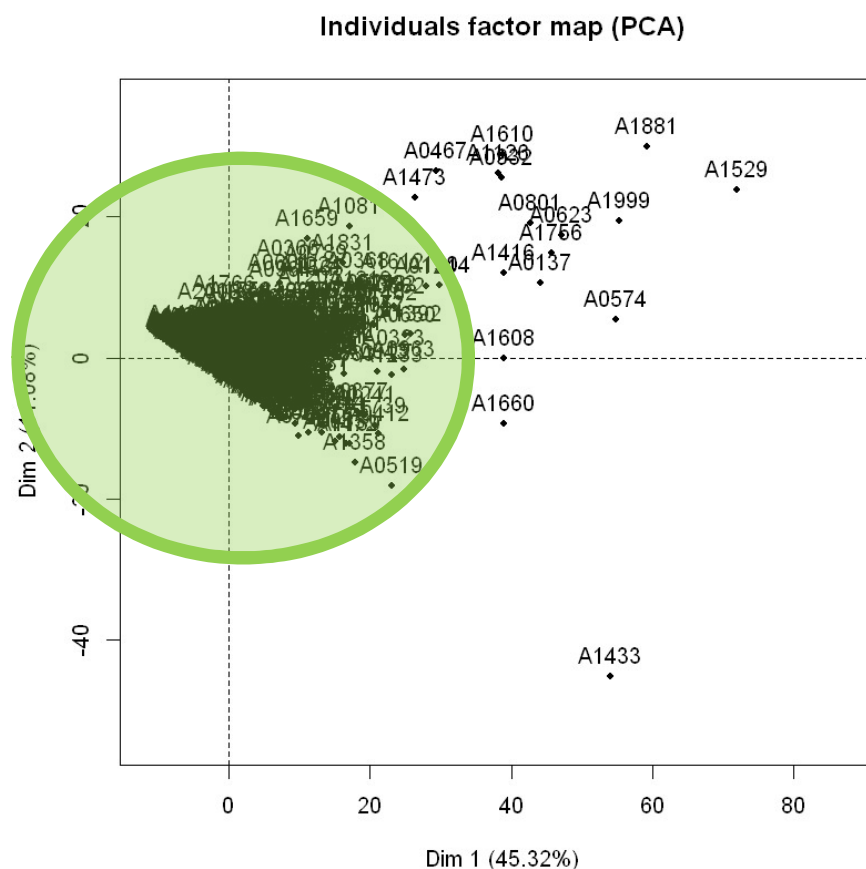
CONFIDENTIAL@2018 ACTUARIS

# ➤ Cas d'application ALLIANZ

## Résultats

- Identification des données « normales »

Scoring et discrimination des données normales/anormales par la **norme euclidienne** de ces individus dans l'espace transformé par l'ACP



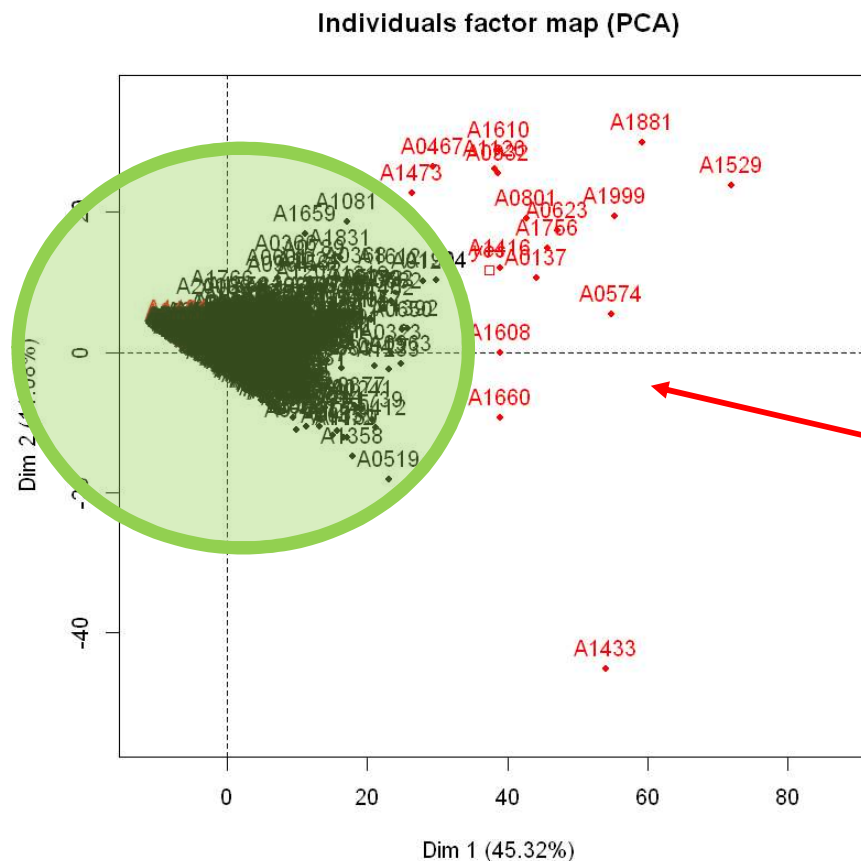


# ➤ Cas d'application ALLIANZ

## Résultats

- Identification des données « normales »

Scoring et discrimination des données normales/anormales par la **norme euclidienne** de ces individus dans l'espace transformé par l'ACP



Scoring des données sujettes à anomalies

# ➤ Cas d'application ALLIANZ

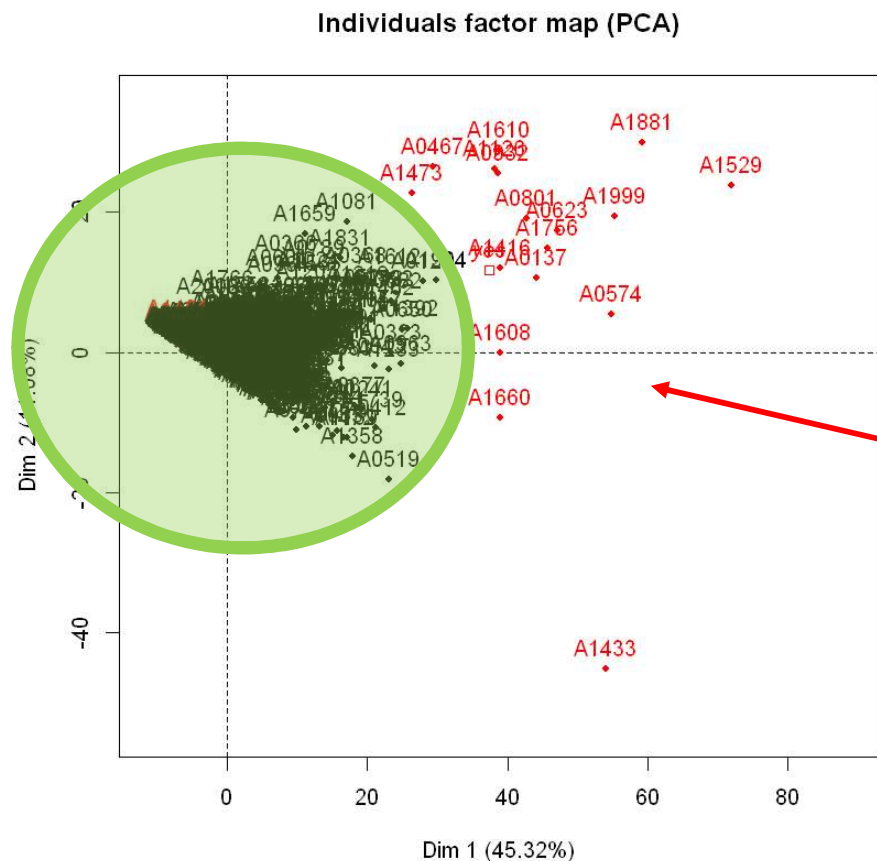
## Résultats

- Identification des données « normales »

Scoring et discrimination des données normales/anormales par la **norme euclidienne** de ces individus dans l'espace transformé par l'ACP



D'autres métriques et méthodes d'identification existent

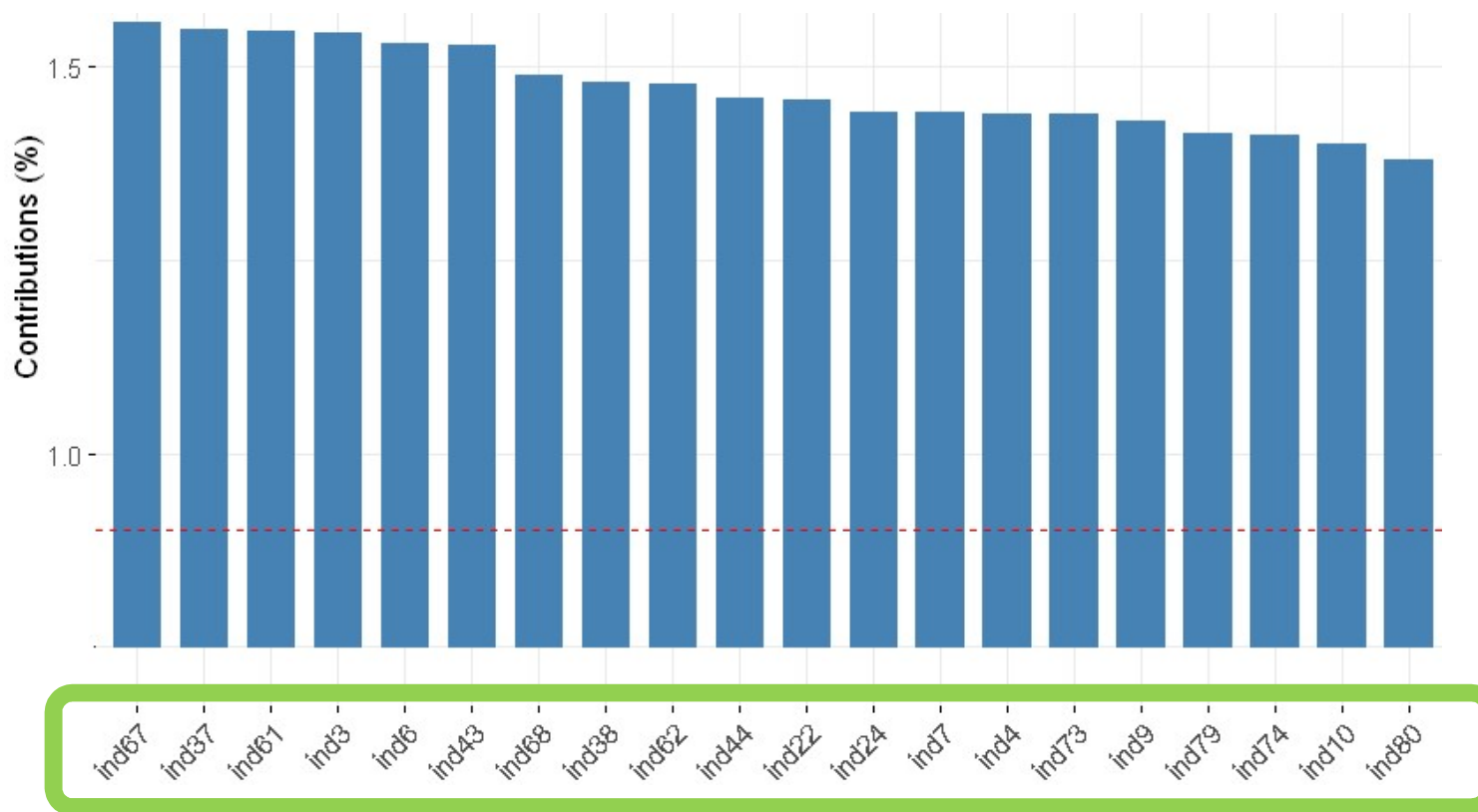


Scoring des données sujettes à anomalies

# ➤ Cas d'application ALLIANZ

## Résultats

### ■ Identifications des indicateurs importants



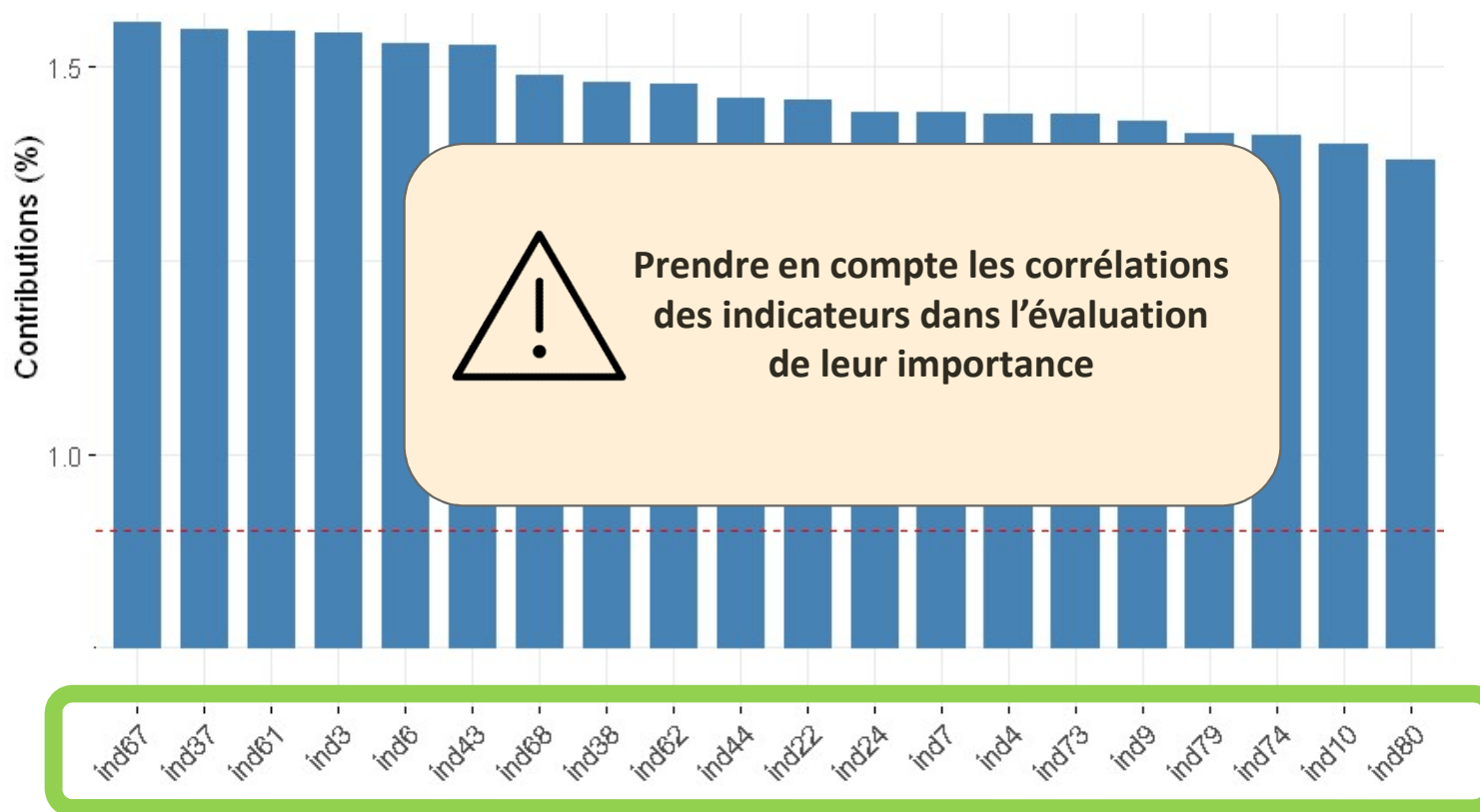
**Top 20** des indicateurs les plus importants sur les 2 premiers axes de l'ACP



# ➤ Cas d'application ALLIANZ

## Résultats

### ■ Identifications des indicateurs importants



**Top 20** des indicateurs les plus importants sur les 2 premiers axes de l'ACP



# ➤ Cas d'application ALLIANZ

## Résultats

- Résumé





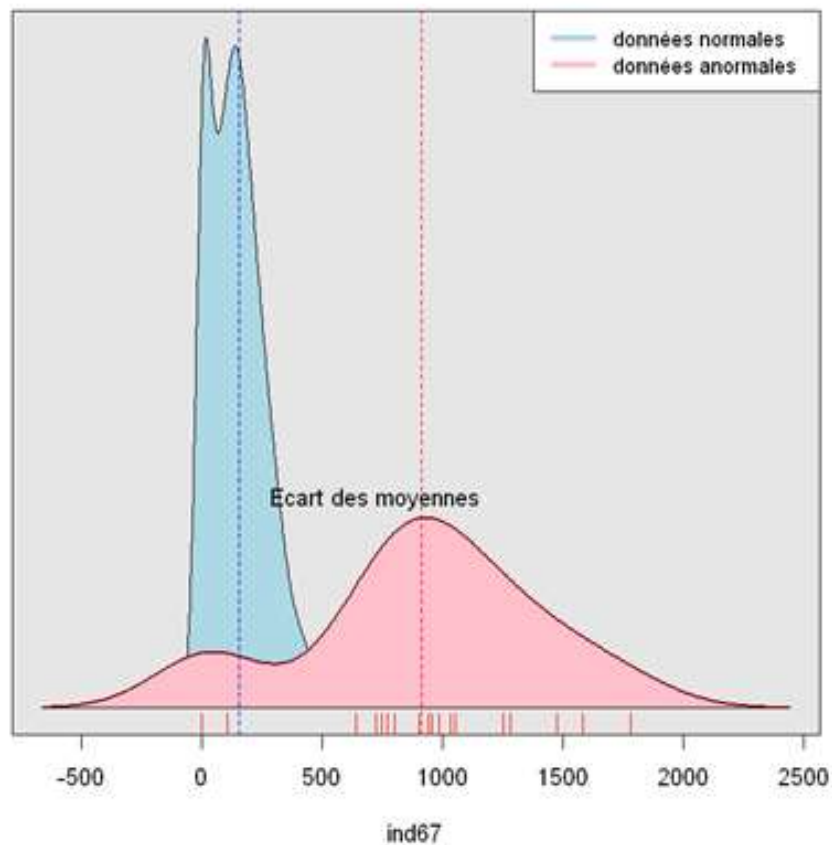
# ➤ Cas d'application ALLIANZ

## Résultats

### ■ Identifications des indicateurs importants

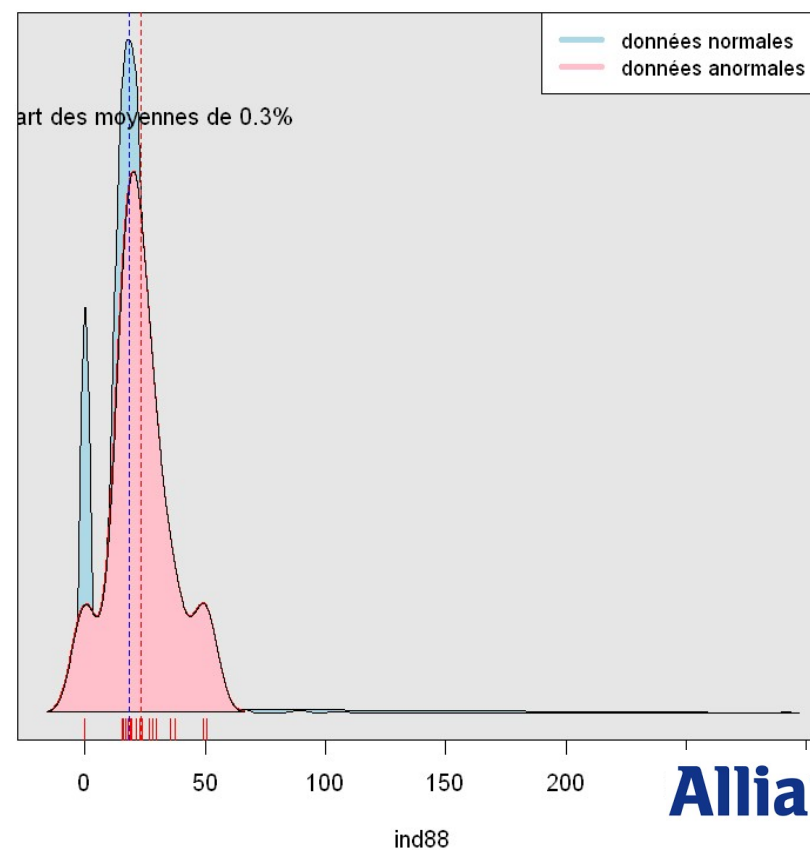
Indicateur très influent

Analyse de l'indicateur ind67



Indicateur très peu influent

Analyse de l'indicateur ind88



CONFIDENTIAL©2018 ACTUARIS

Allianz 

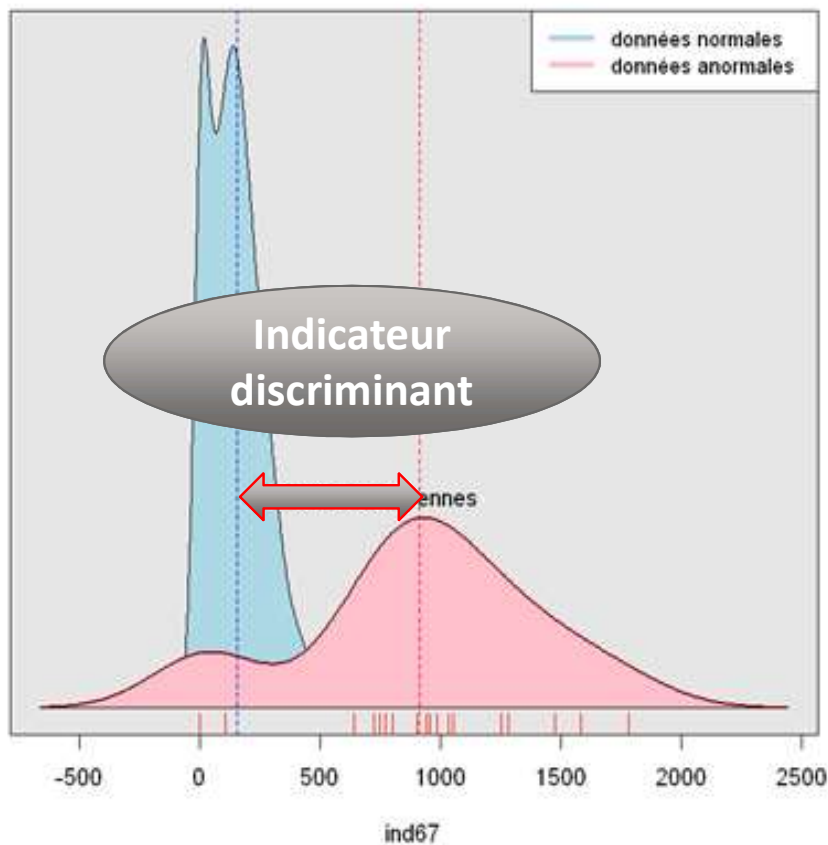
# ➤ Cas d'application ALLIANZ

## Résultats

### ■ Identifications des indicateurs importants

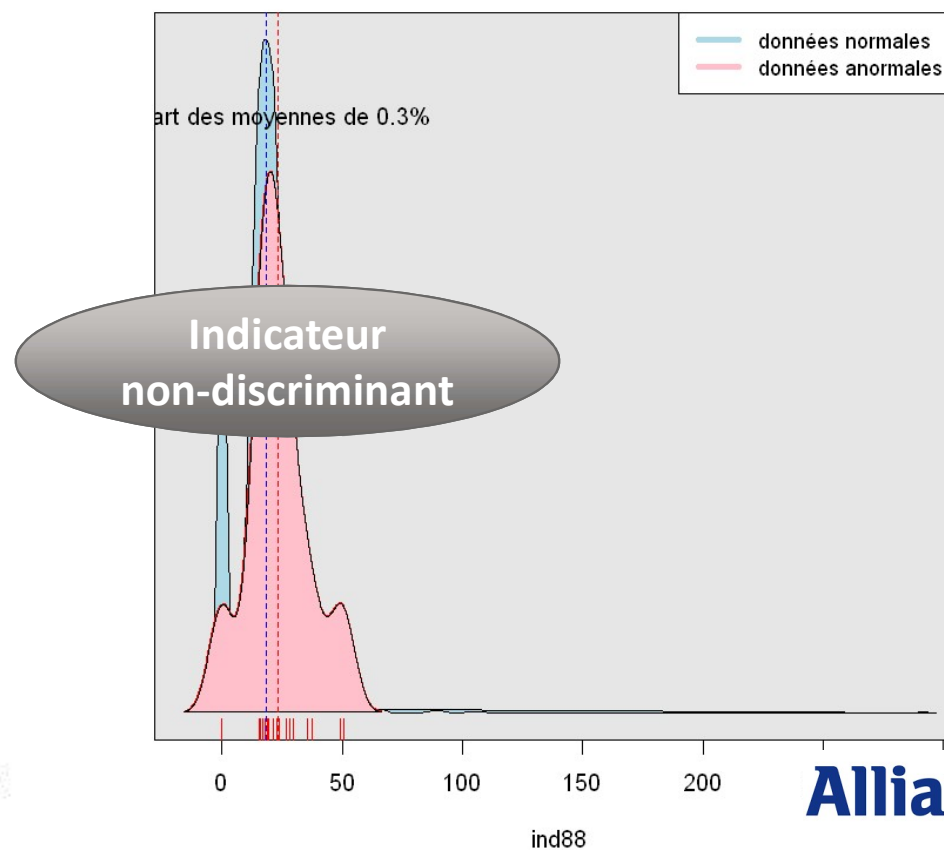
Indicateur très influent

Analyse de l'indicateur ind67



Indicateur très peu influent

Analyse de l'indicateur ind88



CONFIDENTIAL©2018 ACTUARIS

Allianz

# ➤ Cas d'application ALLIANZ

## Résultats

- La « visualisation » des anomalies est un aspect important en détection d'anomalies
  - Une vue synthétique
  - Echanges facilités avec les métiers
  - Analyse et rendu immédiat lors de changement de méthodes (ou seuil de détection, ...)

# ➤ Démonstration sur Dashboard



# Sommaire

- I. Contexte et Objectifs
- II. Approches stratégique et méthodologique
- III. Cas d'application ALLIANZ
- IV. Conclusion et perspectives**



# ➤ Conclusions

- La détection d'anomalies est un axe important d'analyse de données, avec des résultats et **impacts immédiats**
- La **nature de l'information** disponible (supervisée ou non) et son **volume** (nombre d'indicateurs) induit des développements plus ou moins complexes de méthodes de Machine Learning
- La **combinaison de méthodes** de réduction de dimension, de scoring et d'identification de variables d'importance a donné des **résultats très satisfaisants sur le portefeuille d'étude d'ALLIANZ**
- La **DataVisualisation** est un aspect important de la détection d'anomalies: facilite les interactions utilisateurs/méthodes et datascientist/métier

Merci pour votre attention !

