

Bootcamp AI

SEMAINE 1: LES BASES DE L'IA

Réalisé par :

Med Znouda

Ilyes Jennen

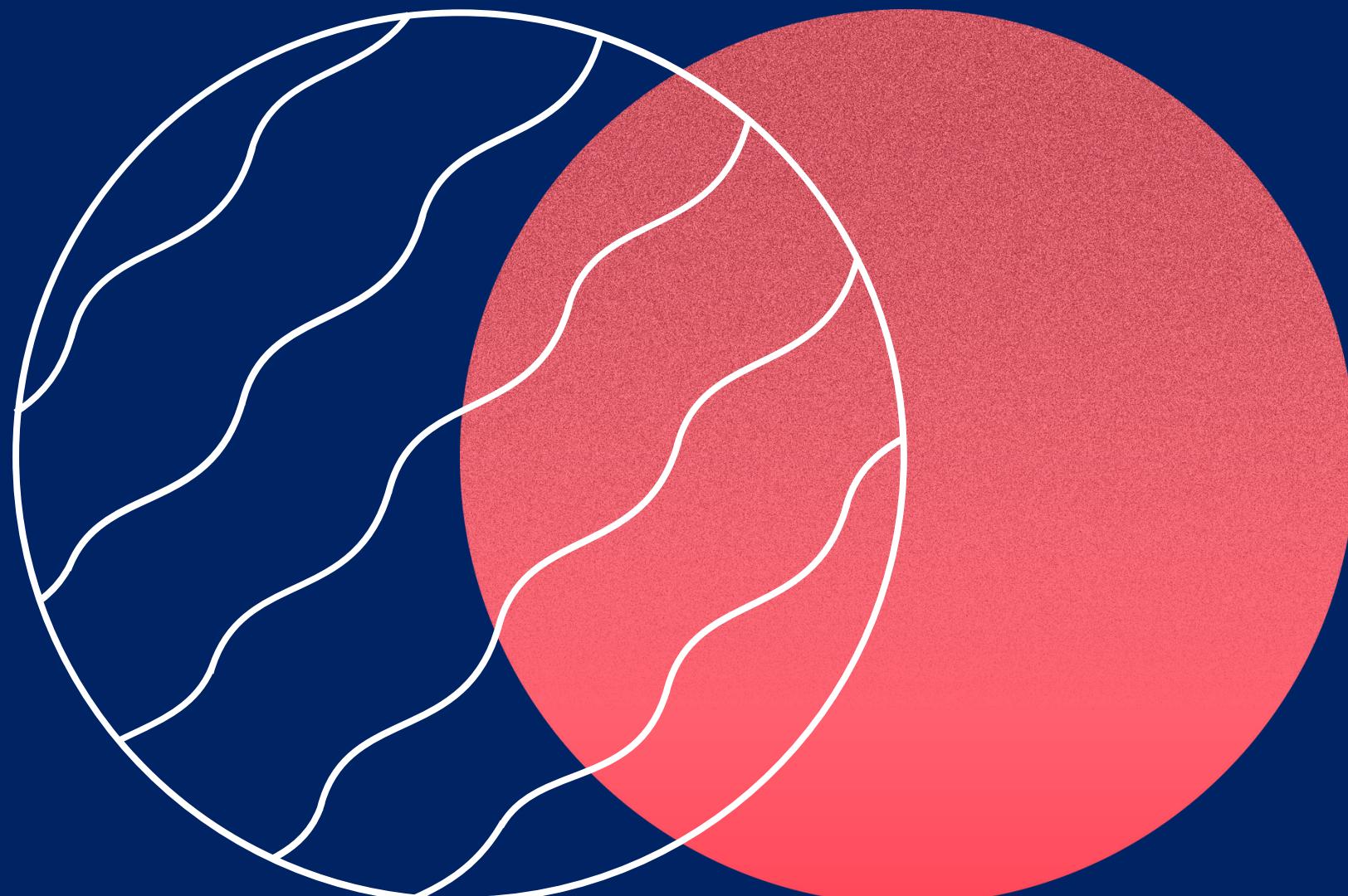
Brahim Bouguicha

Talan[★]
The future is a playfield

- I- LES MÉTHODOLOGIES D'UN PROJET IA
- II- LES RÔLES DANS L'IA
- III- RÉGLEMENTATION EN IA
- IV- COLLECTION DES DONNÉES
- V- ANALYSE ET EXPLORATION DES DONNÉES
- VI- PRÉTRAITEMENT DES DONNÉES
- VII- IA ET LES DÉFIS ÉTHIQUES

⋮ ⋮ ⋮ ⋮

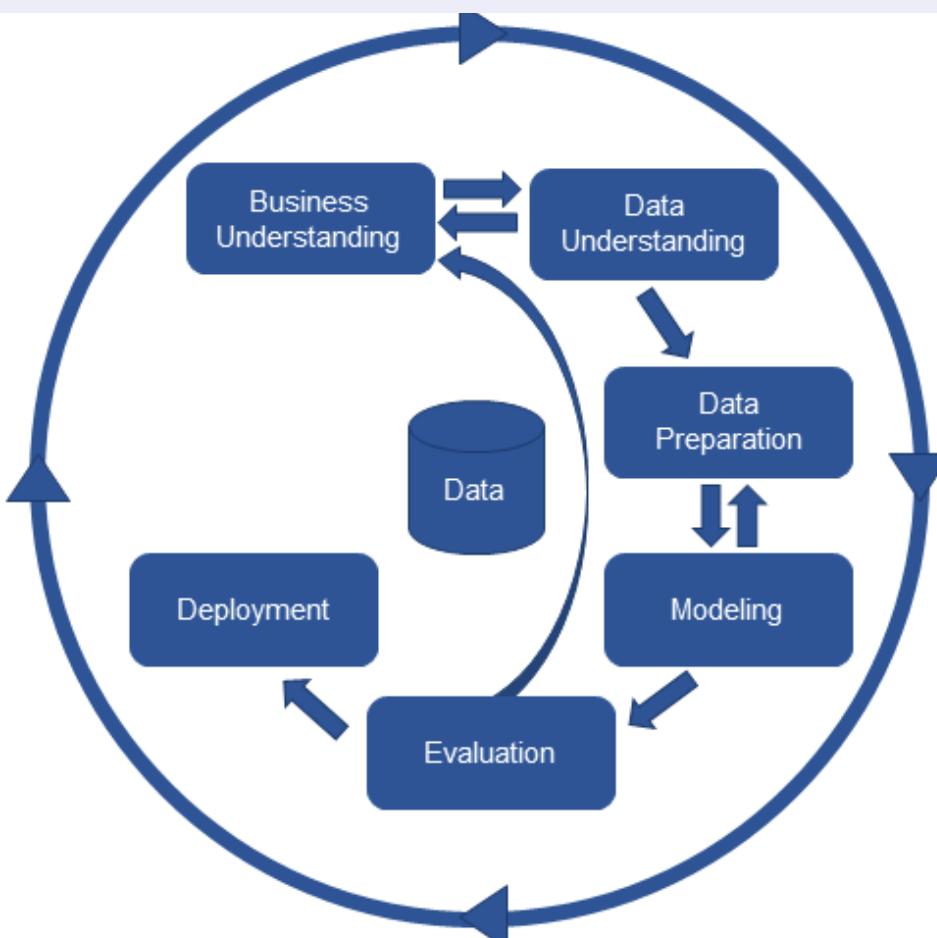




LES MÉTHODOLOGIES D'UN PROJET IA

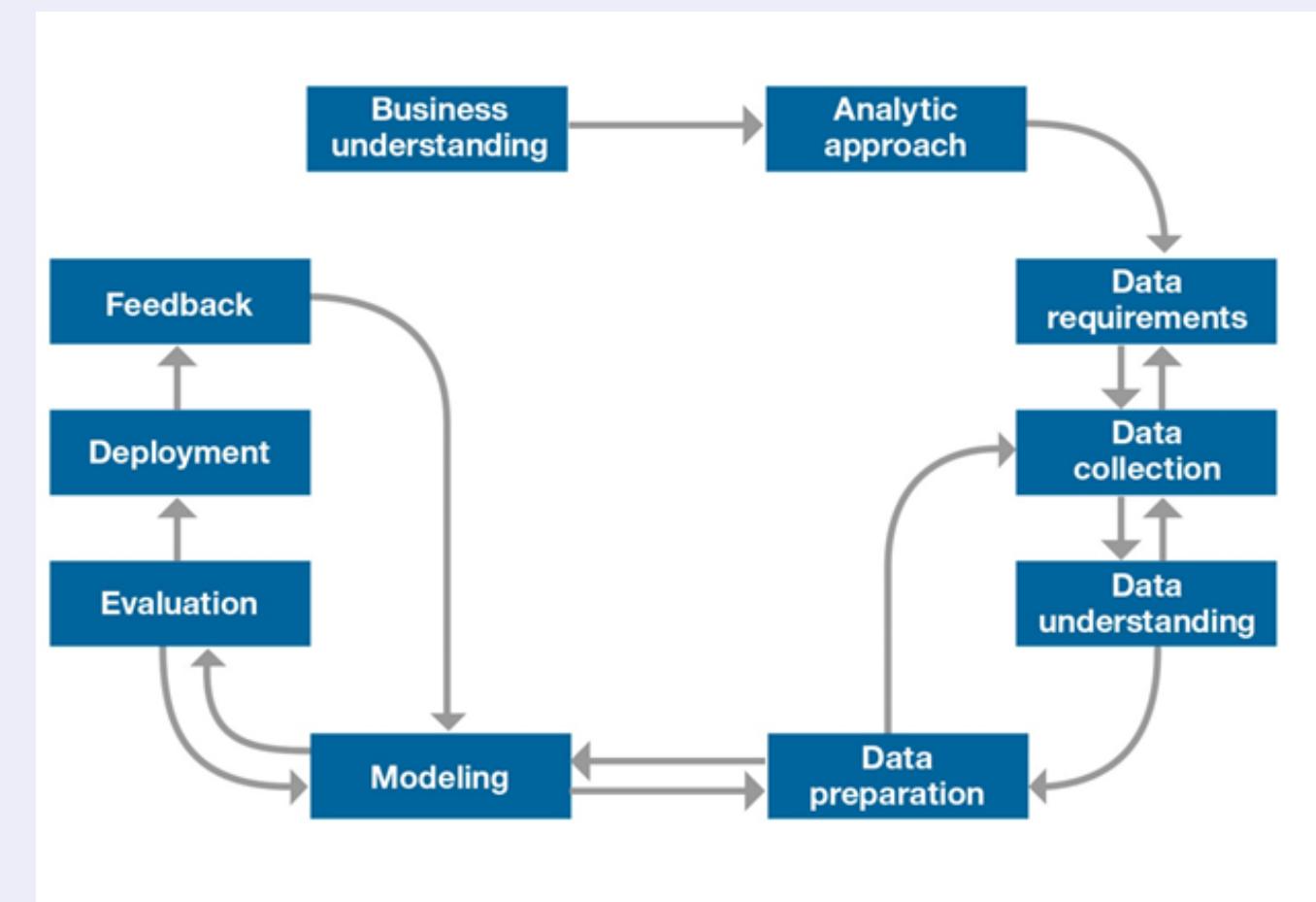
I-Les méthodologies d'un projet IA (1/2)

CRISP-DM



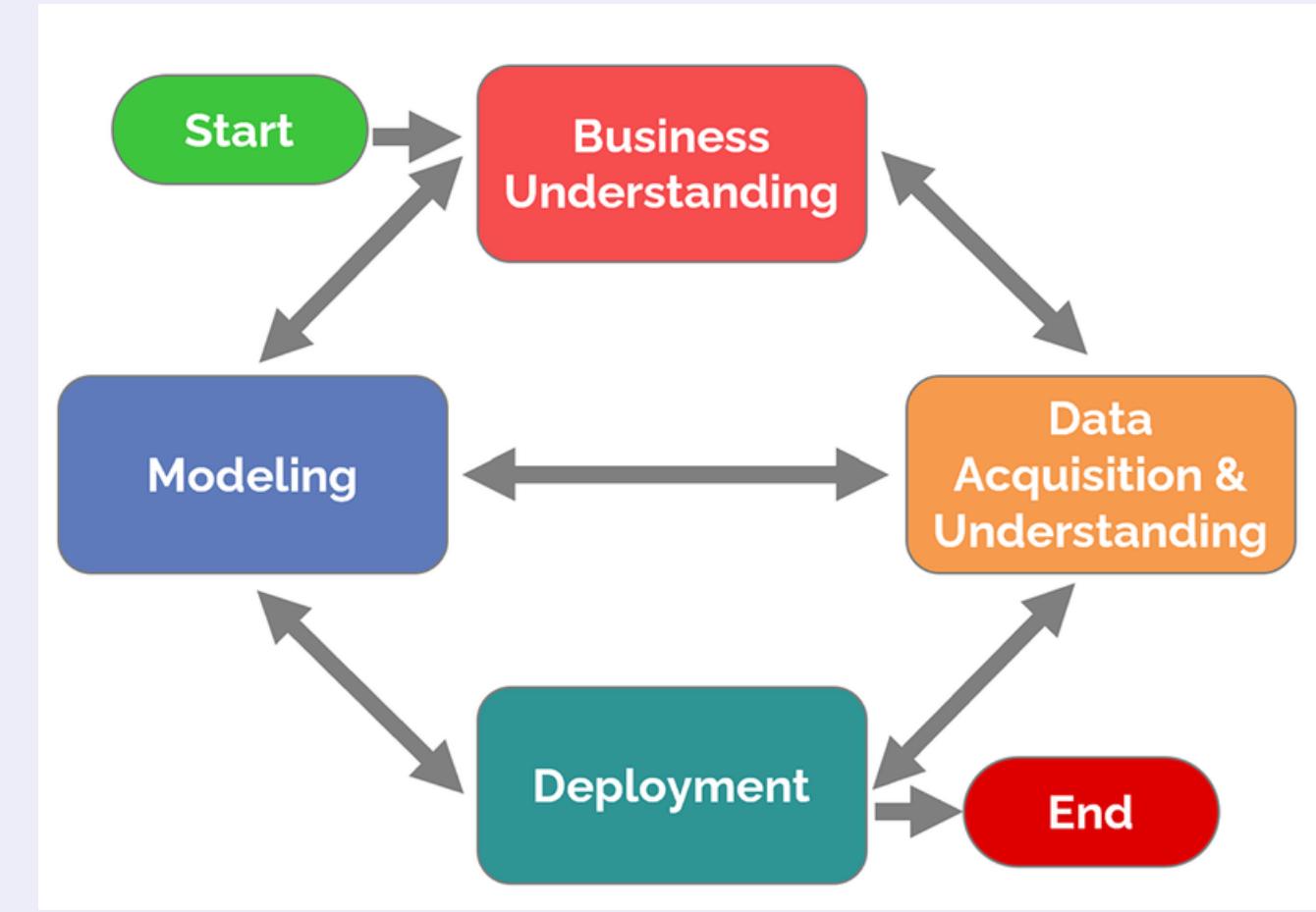
IBM Master plan

(The IBM Foundational
Methodology for Data Science)



TDSP

(The Team Data Science Process)



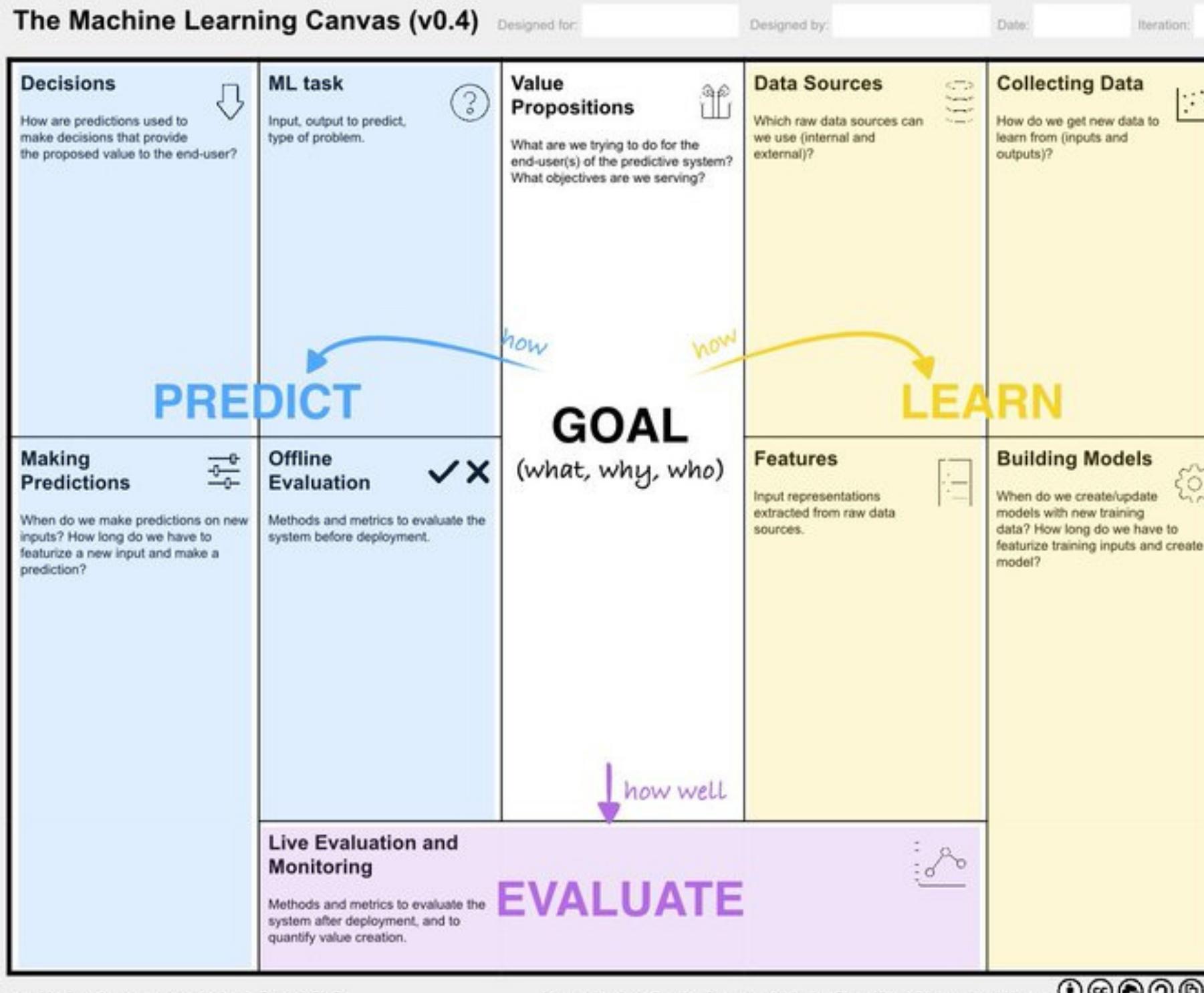
https://fr.wikipedia.org/wiki/Cross_Industry_Standard_Process_for_Data_Mining

<https://medium.com/@yassine.hammar1/data-science-methodology-or-the-master-plan-to-succeed-your-ds-project-d21397cf52b2>

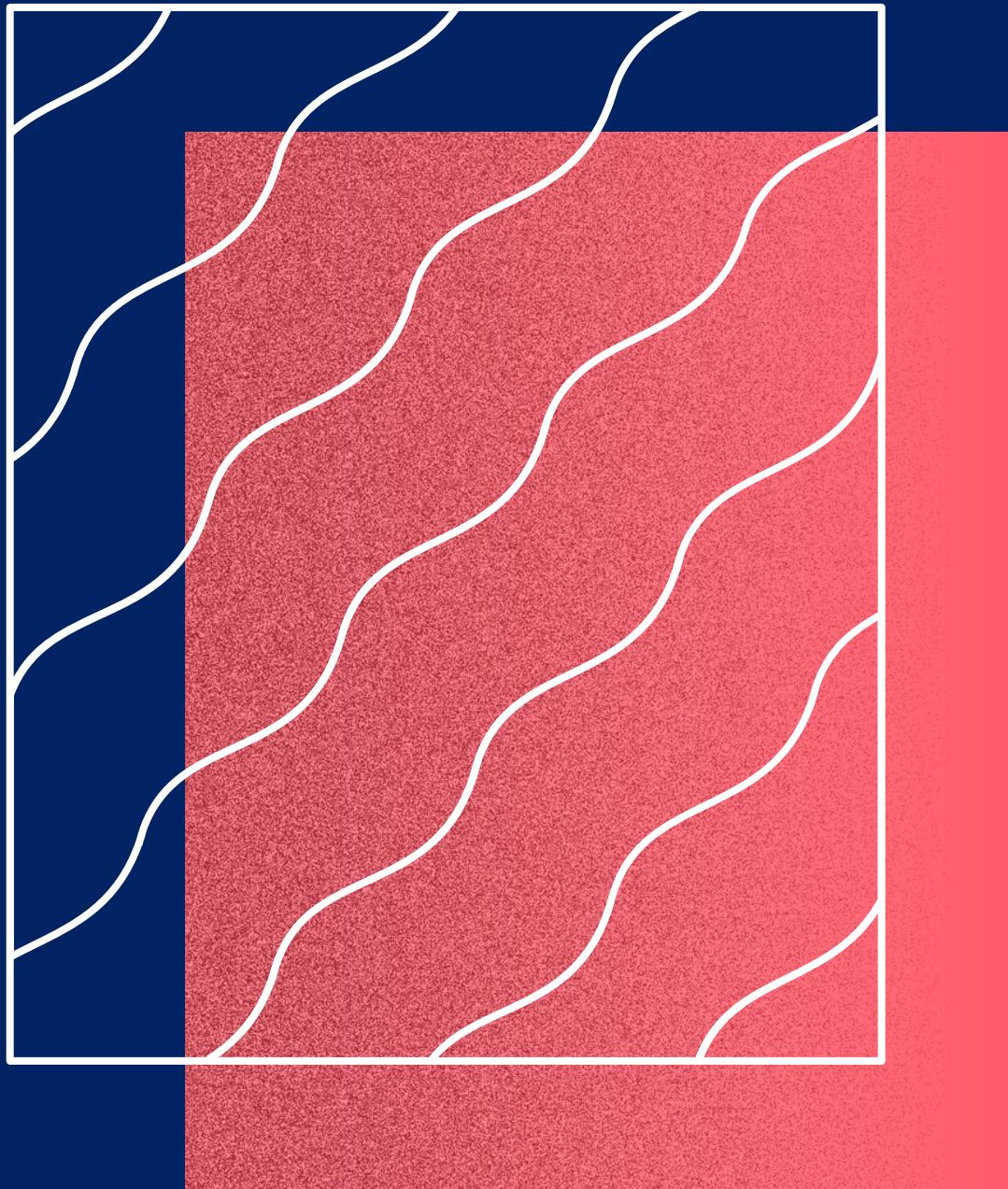
<https://www.datascience-prm.com/tdsp/>

I- Les méthodologies d'un projet IA (2/2)

Comment gérer un projet en data science?



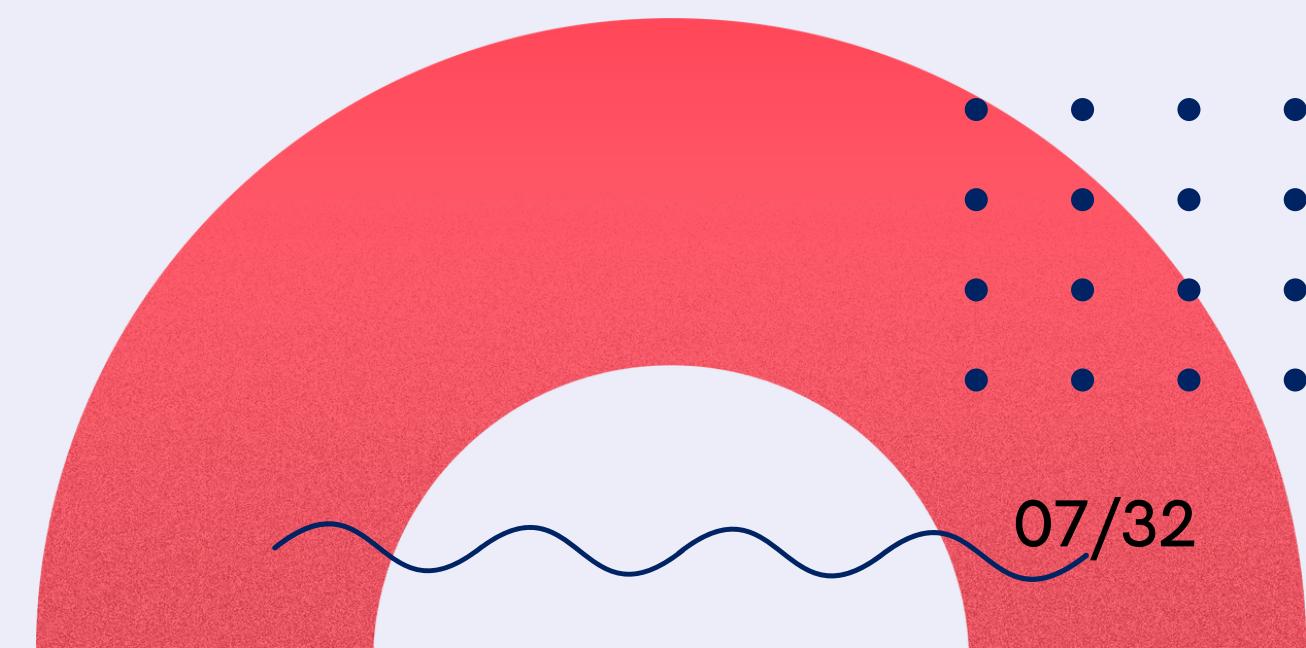
- the Machine Learning Canvas is providing our clients real business value by supplying the first critical entry point for their implementation of predictive applications

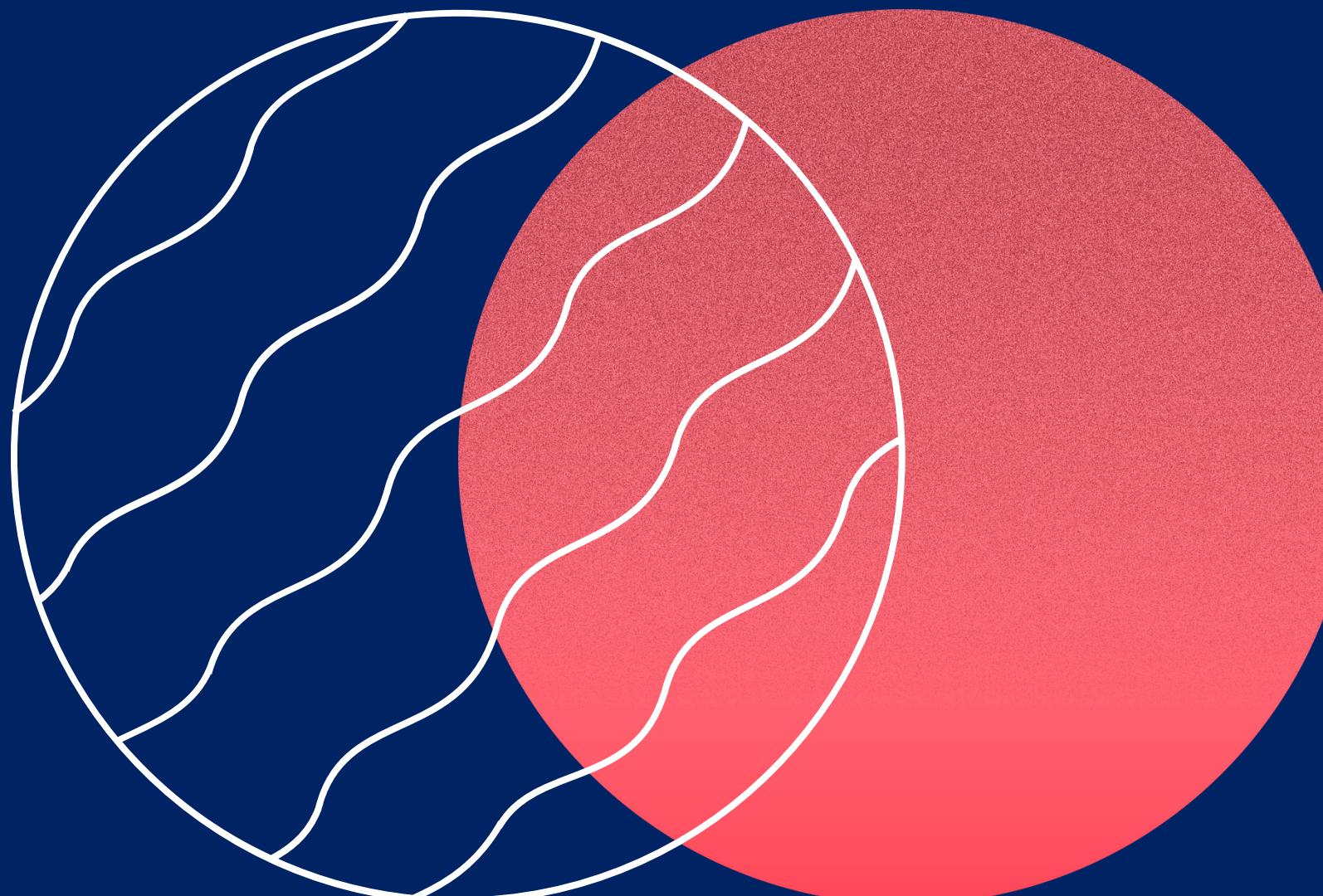


LES RÔLES DANS L'IA

II-DIFFÉRENTS RÔLES DANS L'IA

- AI Engineer/Consultant
- Data Scientist
- Data Engineer
- Data Analyst
- Data Architect
- ML/DL Engineer
- MLOps Engineer
- NLP Engineer





RÉGLEMENTATION EN IA

III-Réglementation en IA (1/5)

- Le RGPD limite dans certains cas le traitement des données personnelles dans un contexte d'IA. Mais cela pourrait éventuellement aider à créer la confiance nécessaire à l'acceptation de l'IA par les consommateurs et les gouvernements.



III-Réglementation en IA (2/5)

Le Projet de Règlement Européen

La confiance

La sécurité

Les droits fondamentaux

- Qu'est-ce qu'elle est interdit ?
- Où ?
- Qu'elles sont les risques pour l'entreprise ?
- Quand ?



En matière d'intelligence artificielle, la confiance n'est pas un luxe mais une nécessité absolue. En adoptant ces règles qui feront date, l'UE prend l'initiative d'élaborer de nouvelles normes mondiales qui garantiront que l'IA soit digne de confiance. En établissant les normes, nous pouvons ouvrir la voie à une technologie éthique dans le monde entier, tout en préservant la compétitivité de l'UE. À l'épreuve du temps et propices à l'innovation, nos règles s'appliqueront lorsque c'est strictement nécessaire: quand la sécurité et les droits

Margrethe Vestager

vice-présidente exécutive pour une Europe adaptée à l'ère du numérique

“

L'IA doit être au service des personnes et, par conséquent l'IA doit toujours respecter les droits des personnes. Nous voulons encourager nos entreprises, nos entrepreneurs, nos innovateurs à développer l'IA.

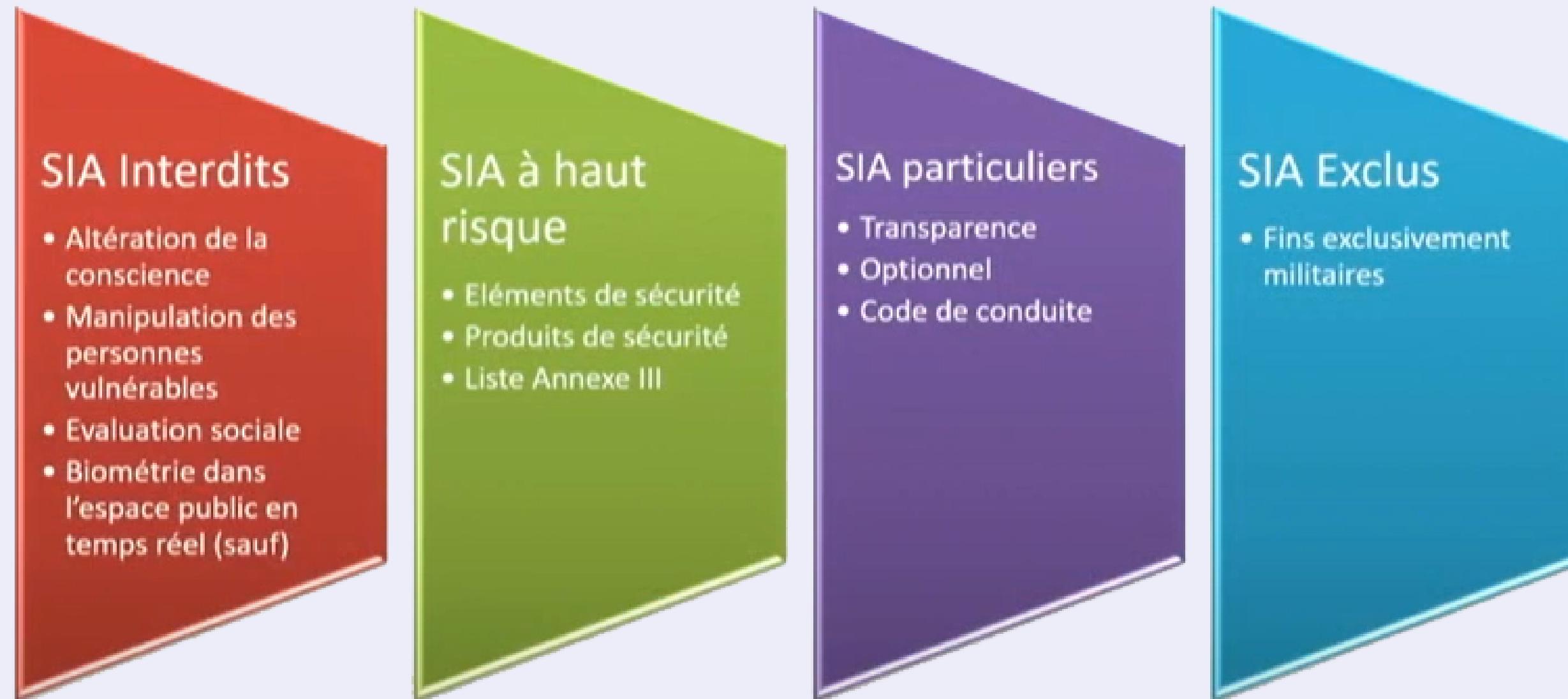
Ursula von der Leyen

Présidente de la Commission européenne



III-Réglementation en IA (3/5)

Une régulation horizontale, par typologie de SIA, et par les risques



classification des systèmes d'IA proposée par la Commission

III-Réglementation en IA (4/5)

Champ d'application géographique

- Utilisateurs de systèmes d'IA établis dans l'UE ;
- Fournisseurs établis dans l'UE ou hors UE mettant sur le marché européen des systèmes d'IA ;
- Utilisateurs et Fournisseurs établis hors UE, si les résultats produits par lesdits systèmes sont utilisés dans l'UE.

Perspective : calendrier d'application

- La Commission souhaite que la nouvelle législation devienne une loi **vers la fin de 2022** et les entreprises devront se conformer aux nouvelles exigences un an et demi à deux ans à compter de la date de son adoption définitive dans l'Union.

III-Réglementation en IA (5/5)

Un régime de sanctions substantielles pour non-conformité

30 millions ou 6%
CA

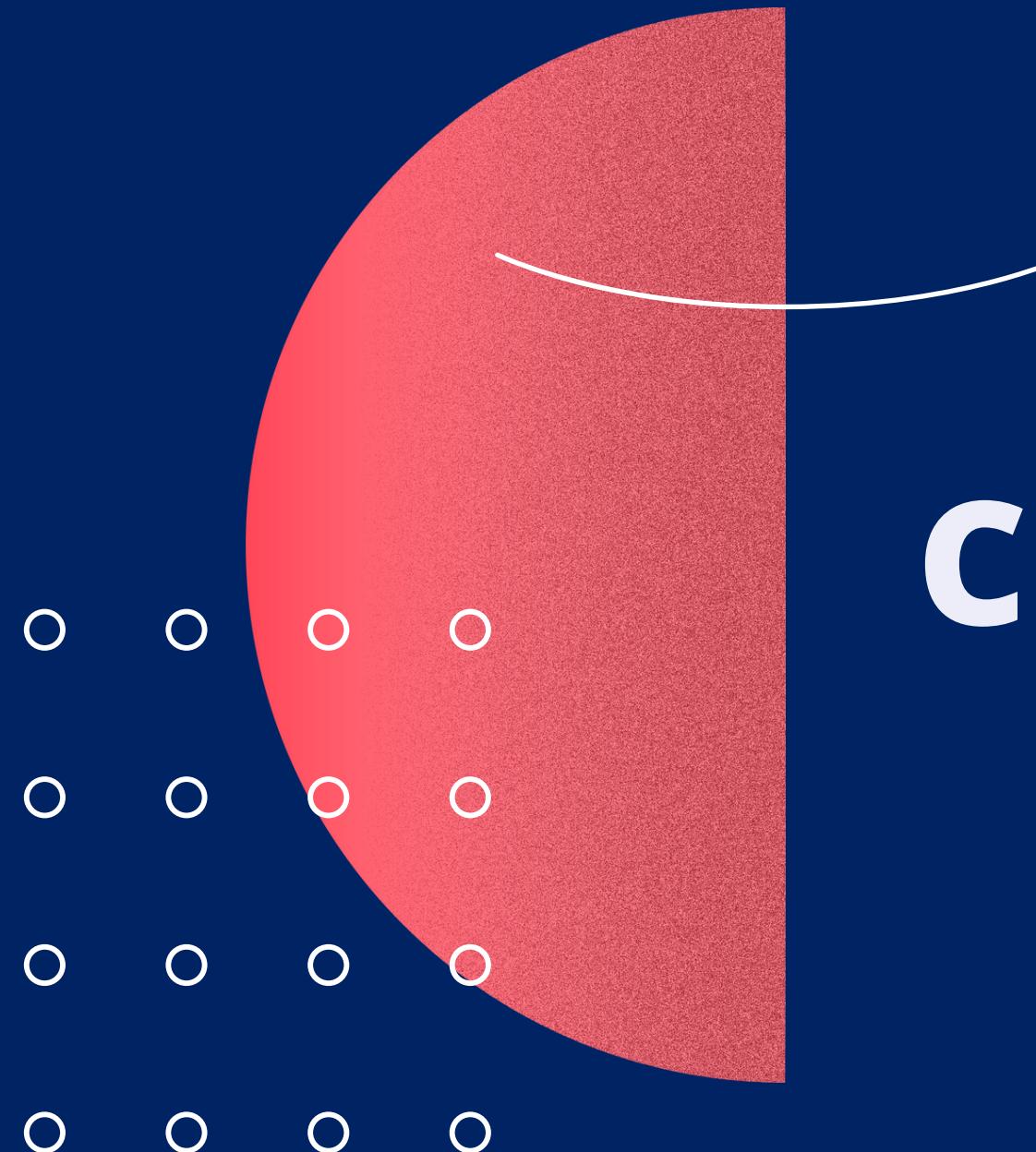
- SIA interdits (art. 5)
- SIA et données (art.10)

20 millions ou 4%
CA

- Non-conformité des SIA réglementés

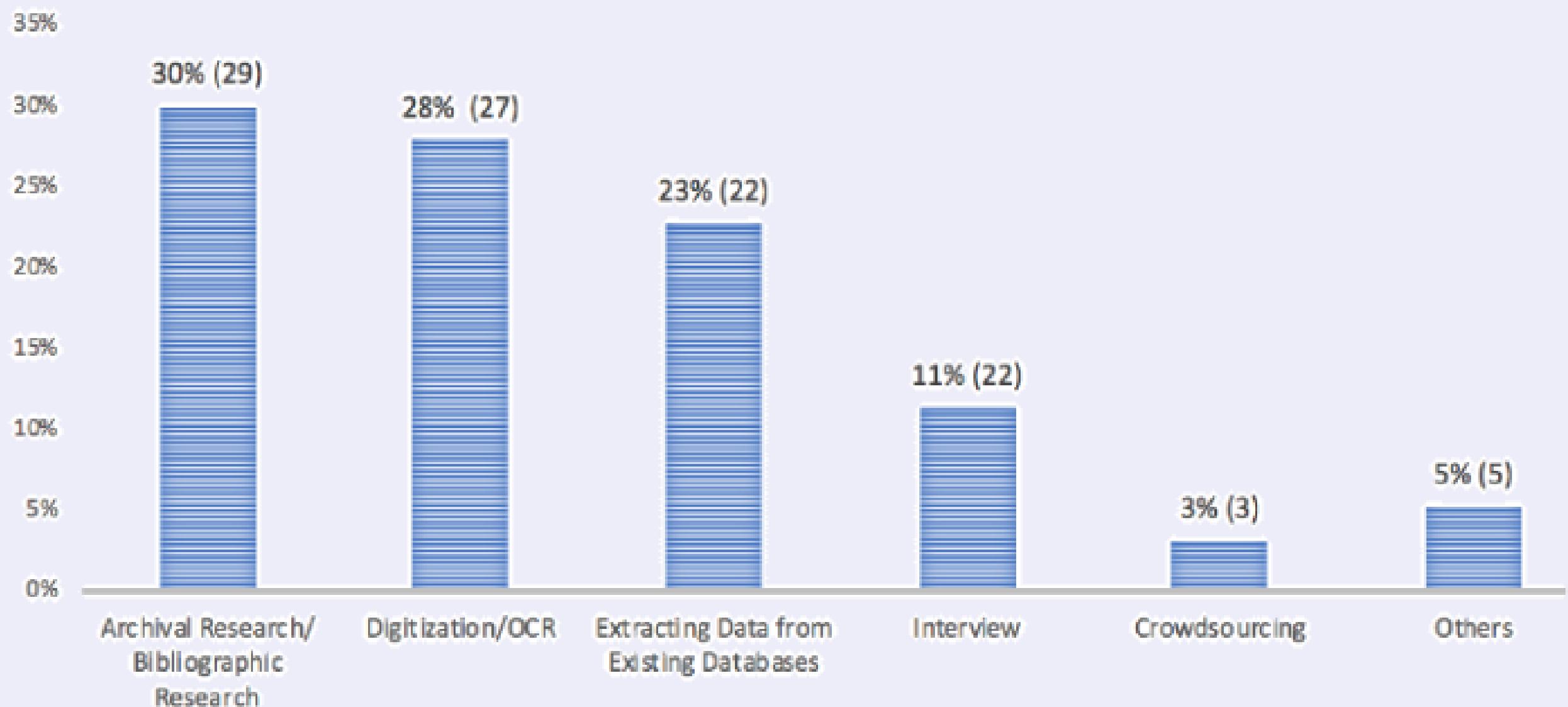
10 millions ou 2%
CA

- Informations inexactes ou transparences : Organismes notifiés
- Informations inexactes ou transparences: autorités nationales compétentes

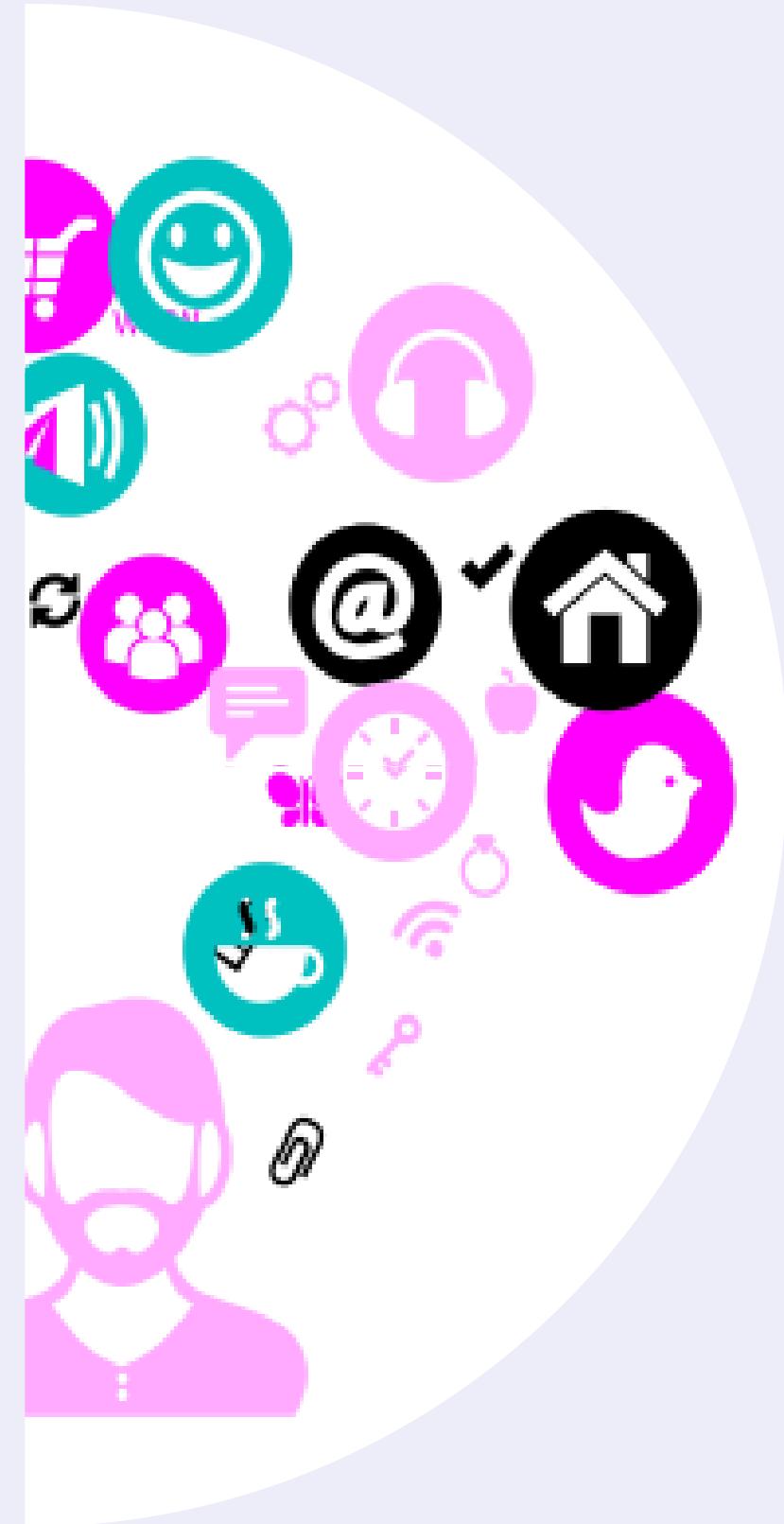


COLLECTION DES DONNÉES

IV-COLLECTION DES DONNÉES (1/3)



Répartition des sources de collecte des données



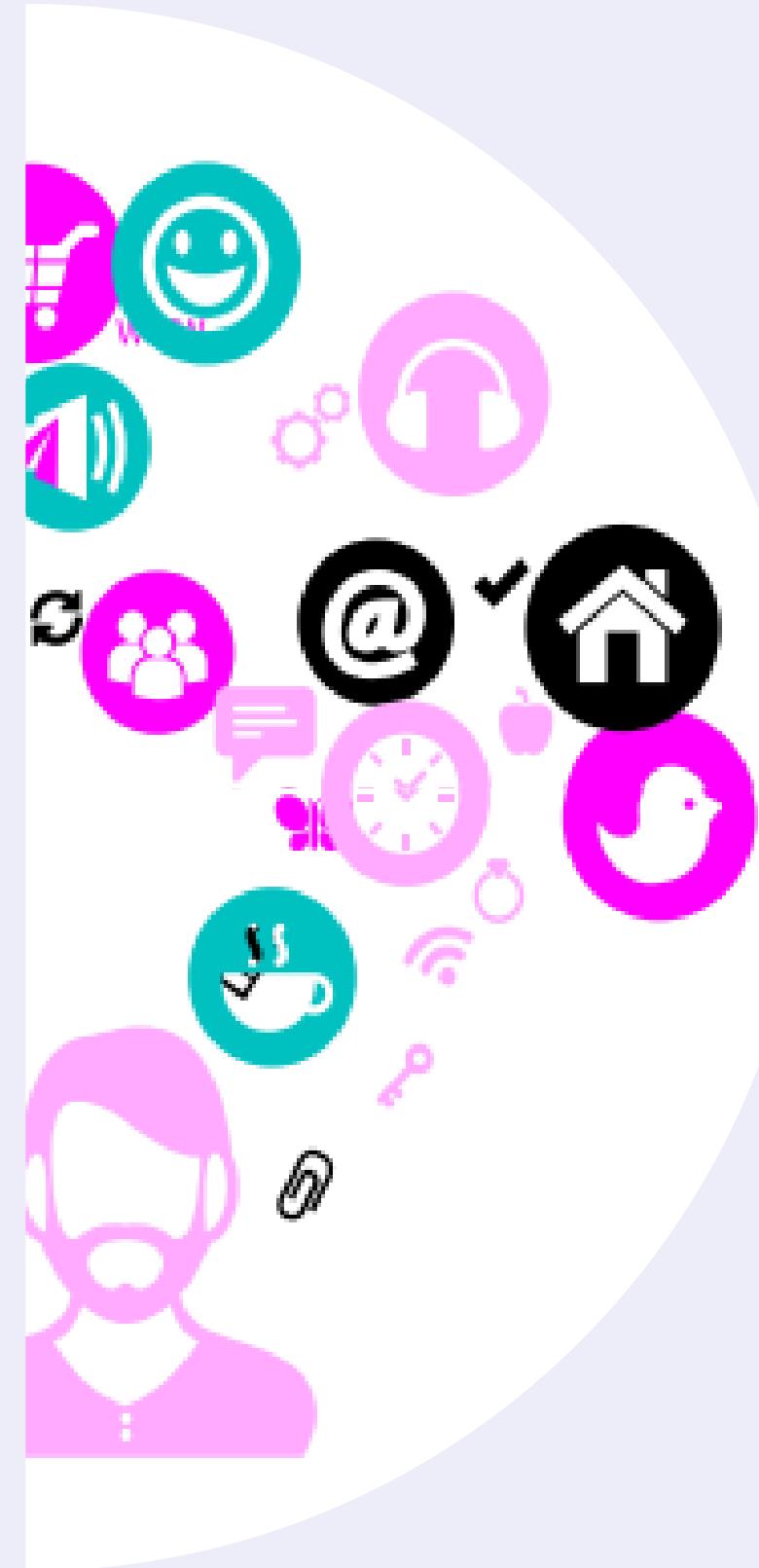
IV-COLLECTION DES DONNÉES (2/3)

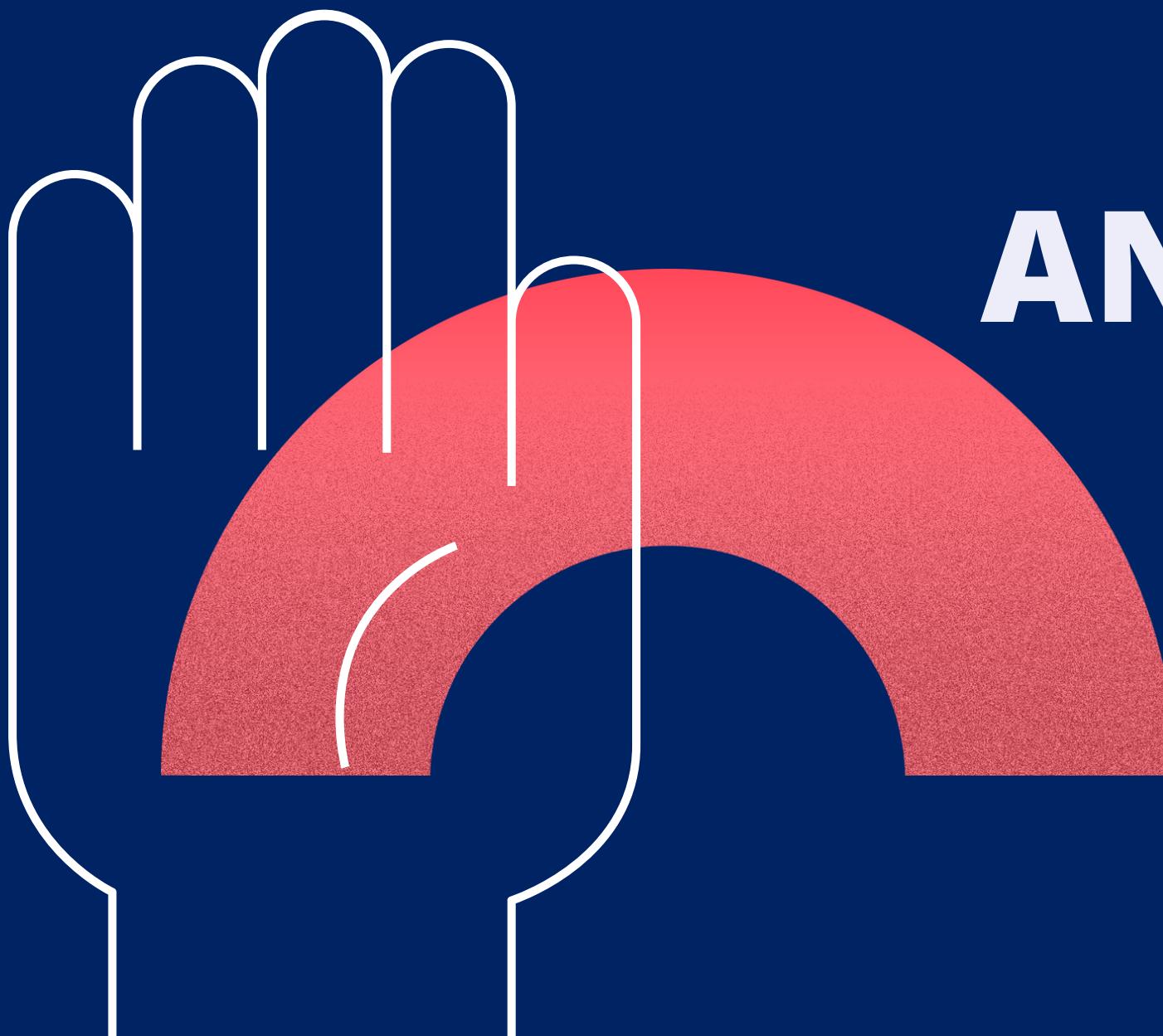
Data labeling

- ***Manuellement***
- ***Automatiquement***
 - Programmatic labeling.
 - Minimally supervised labeling.
 - Unsupervised labeling.

IV-COLLECTION DES DONNÉES (3/3)

Web scraping





ANALYSE ET EXPLORATION DES DONNÉES

V-ANALYSE ET EXPLORATION DES DONNÉES(1/3)

- C'est quoi ?

...
...
...
...

La quantité de données est-elle adéquate?

Les données sont-elles sous la bonne forme (type de données)?

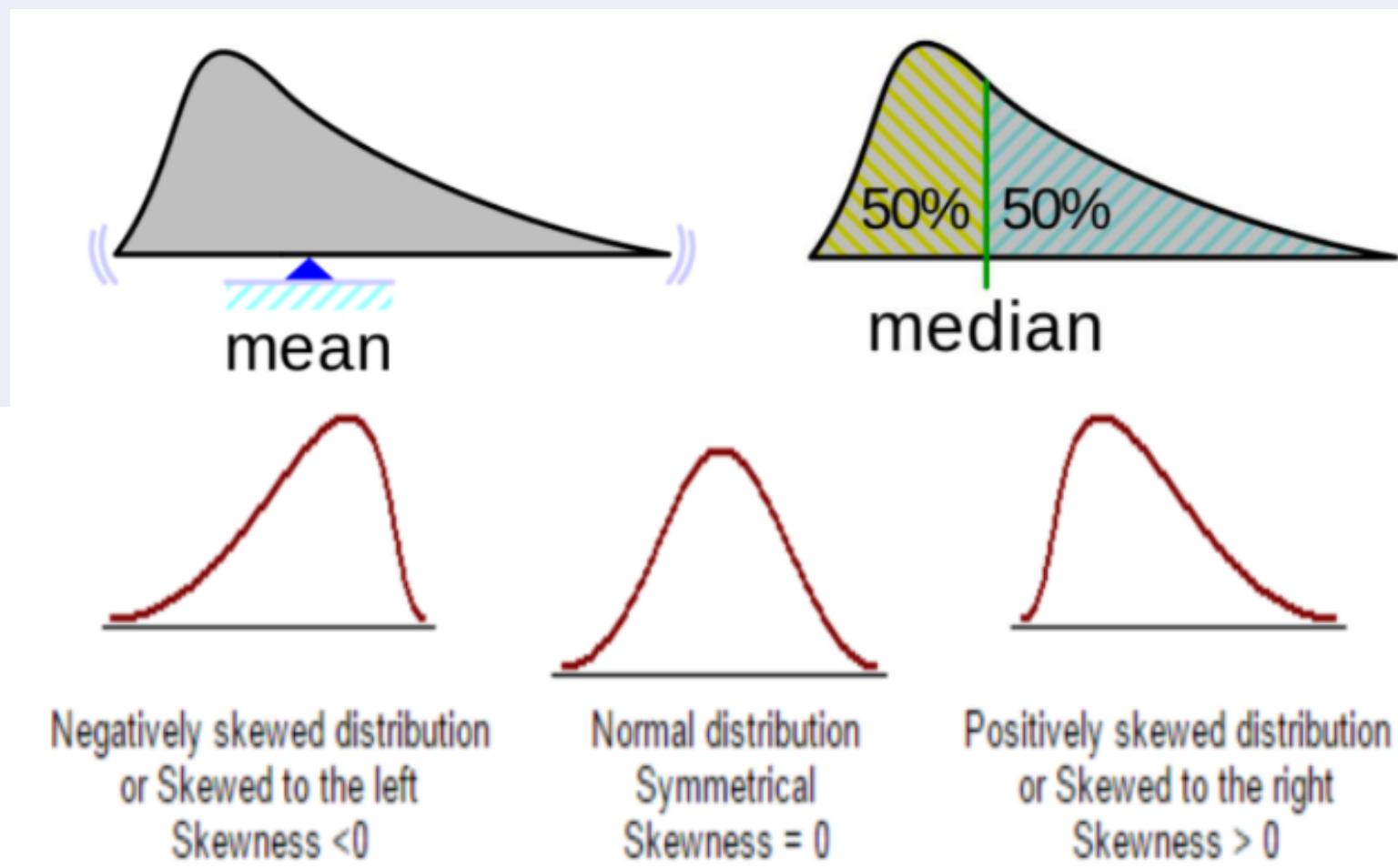
Quel est le niveau du bruit dans les données (data noisy) ?

V-ANALYSE ET EXPLORATION DES DONNÉES (2/3)

- Outils

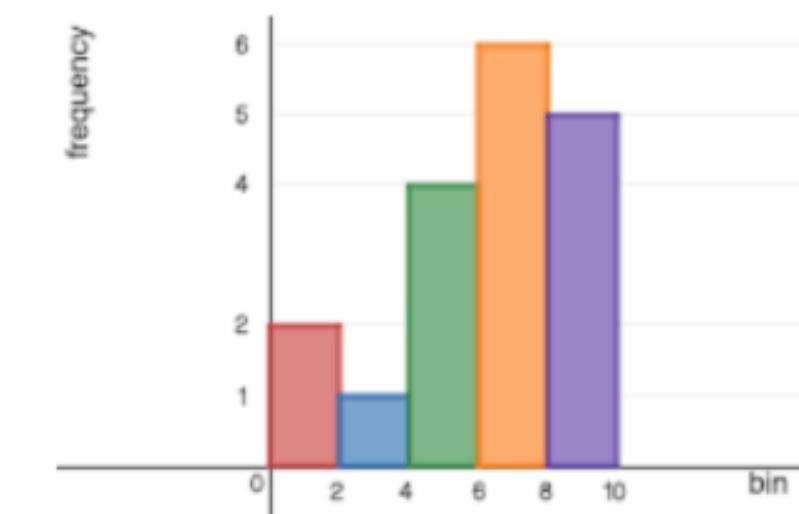
Mesures Statistiques

- Mean
- Median
- Mode
- Range
- Quantiles
- Variance



Visualisation (Distribution des valeurs)

- Histogram (numeric variable)
- Boxplot (numeric variable)
- Barplot, Piechart (categorical variable)
- Scatterplot(numeric variable)



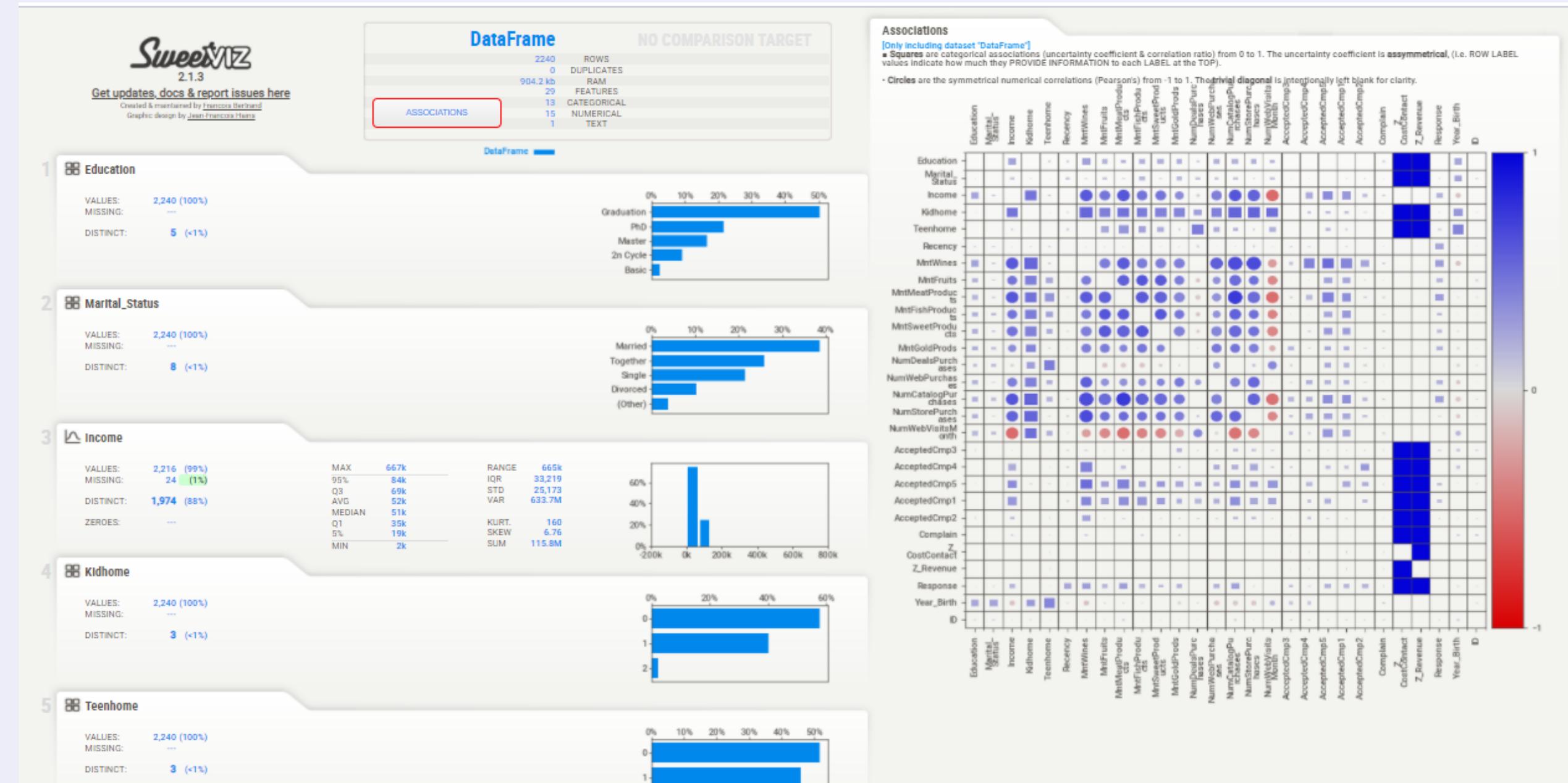
une représentation graphique permettant de représenter la répartition d'une variable continue

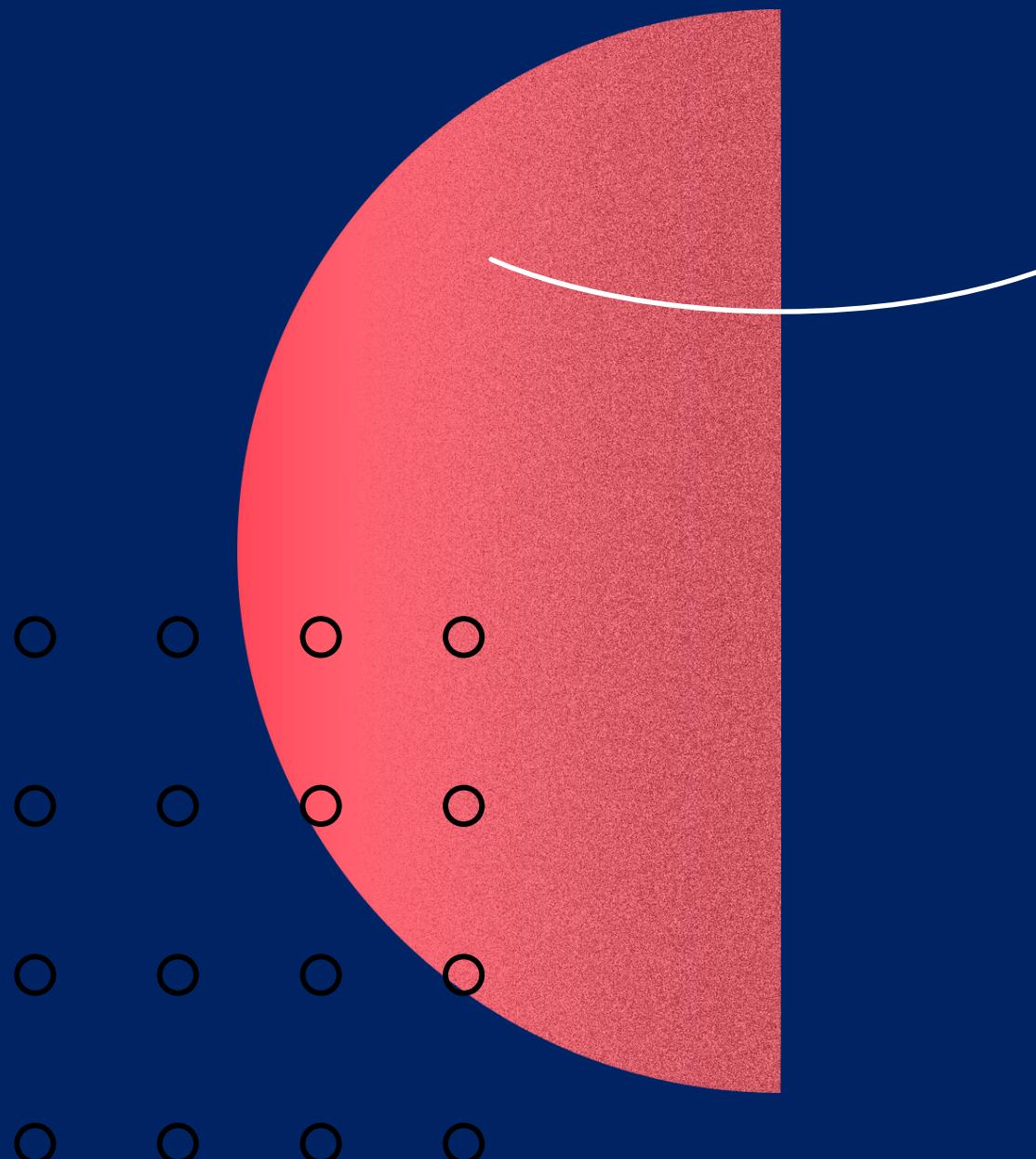
Une valeur aberrante est une observation qui se situe à une distance anormale des autres valeurs dans un échantillon aléatoire d'une population.

V-ANALYSE ET EXPLORATION DES DONNÉES (3/3)

• Sweetviz

Sweetviz est une bibliothèque Python open-source qui génère de magnifiques visualisations à haute densité pour lancer l'AED (analyse exploratoire des données) avec seulement deux lignes de code.





PRÉTRAITEMENT DES DONNÉES

VI-PRÉTRAITEMENT DES DONNÉES (1/6)

Les Techniques de traitement des datasets déséquilibrées

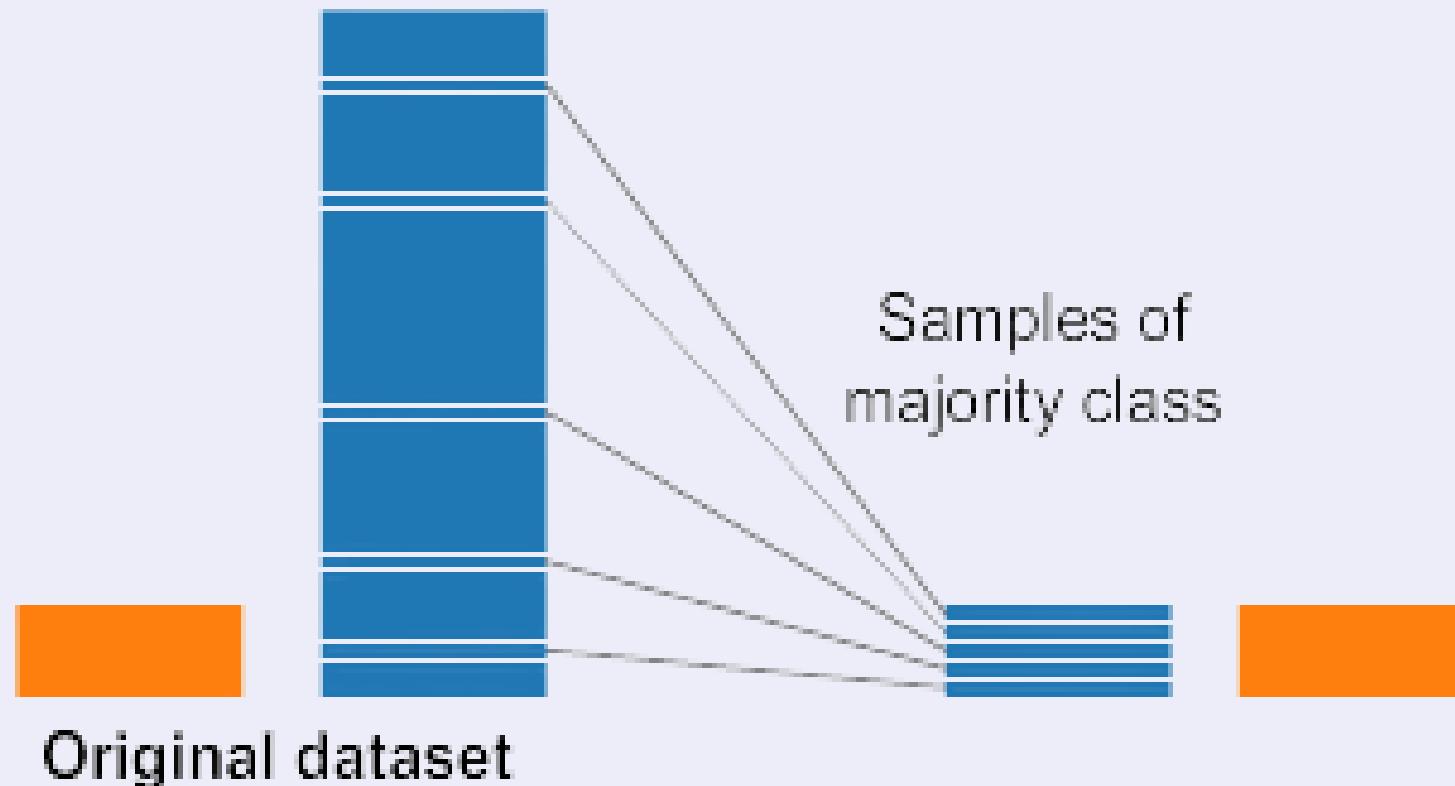
1-UNDERSAMPLING

- Random undersampling.
- Règle du plus proche voisin condensé (CNN).

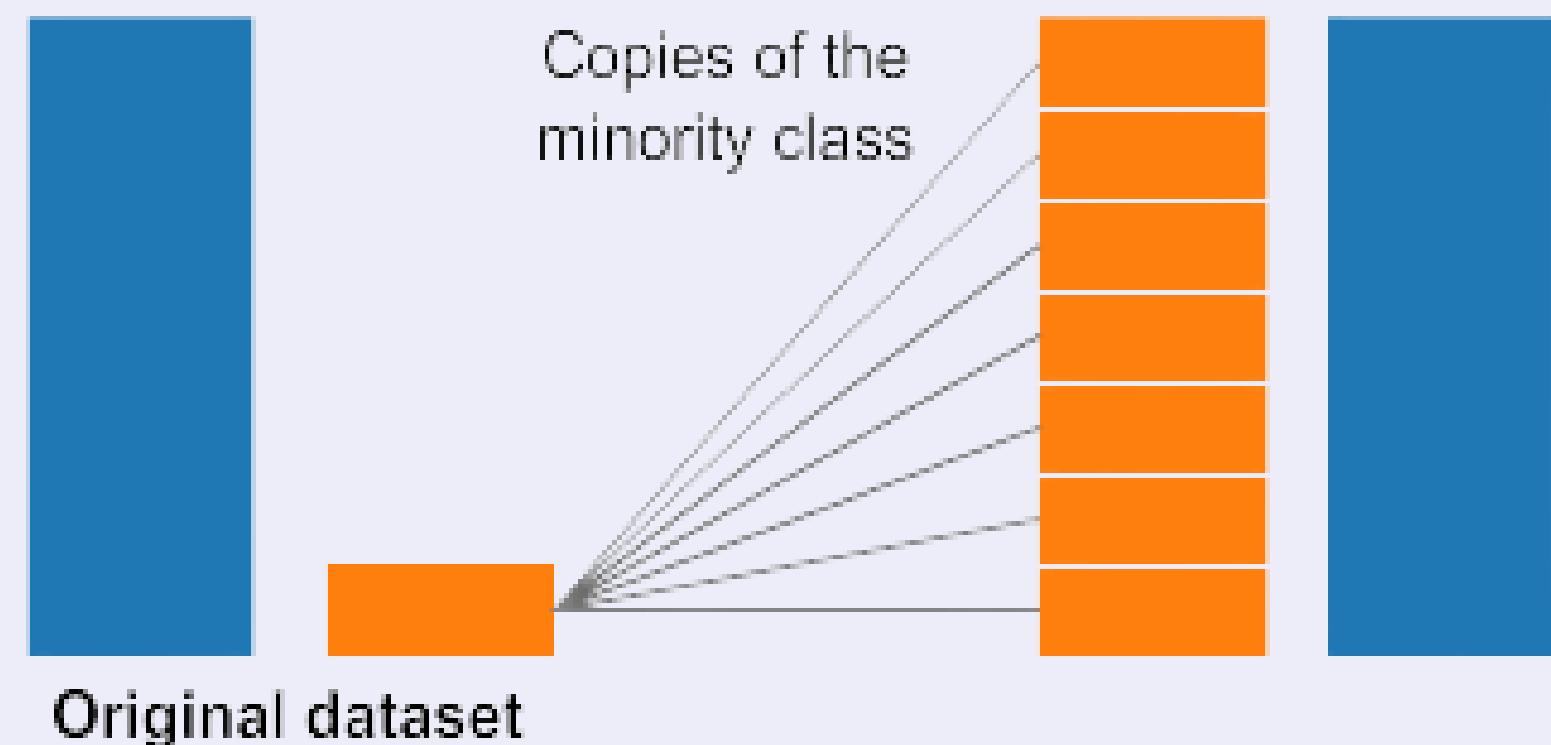
2-OVERSAMPLING

- Random oversampling.
- SMOTE.

Undersampling

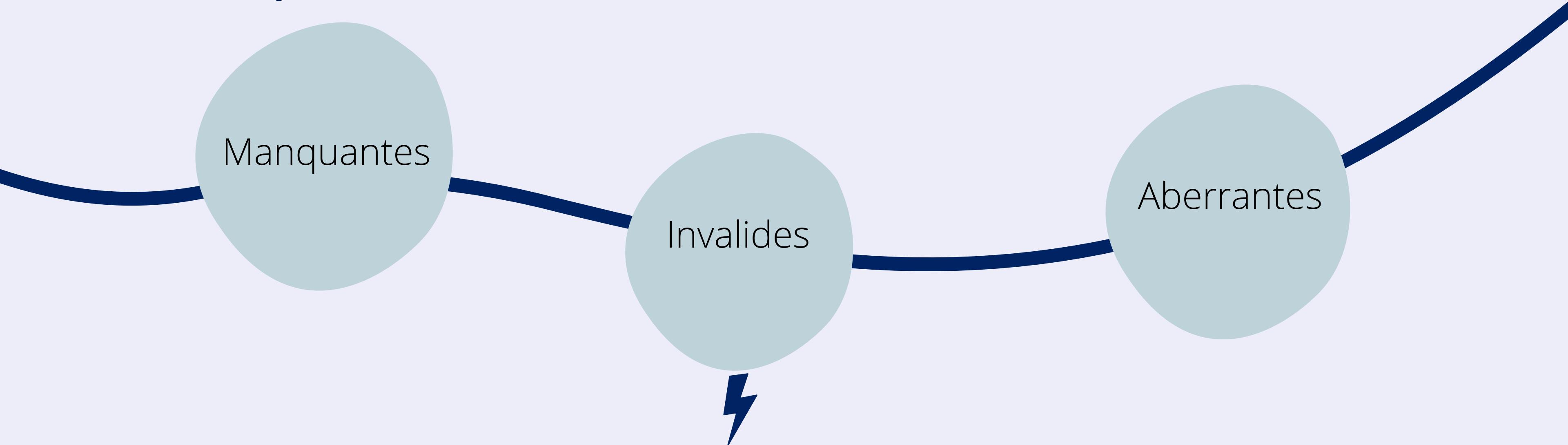


Oversampling



VI-PRÉTRAITEMENT DES DONNÉES (2/6)

Les Techniques de traitement des mauvaises valeurs

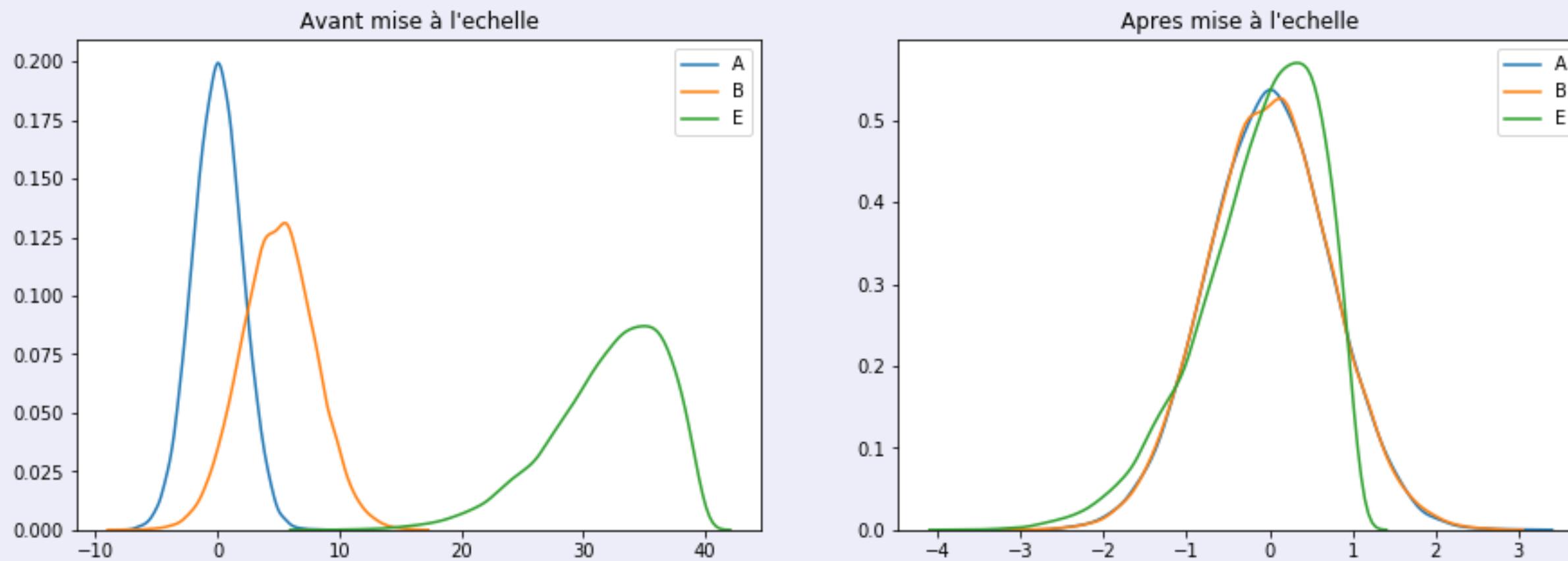


- Remplacer la valeur (mean, median, mode).
- Supprimer l'observation.
- Garder la valeur(détection d'anomalie).

VI-PRÉTRAITEMENT DES DONNÉES (3/6)

Feature engineering

- Scaling or Scale Normalization (Mise à l'échelle).
- Log Normalization.
- One-hot encoding.
- Label encoding.
- Grouping sparse categories.



id	X = a	X = b	X = c
1	1	0	0
2	0	0	1
3	1	0	0
4	0	1	0
5	1	0	0
6	0	0	1
7	0	0	1
8	0	1	0

id	X
1	a
2	c
3	a
4	b
5	a
6	c
7	c
8	b

id	X = a	X = b
1	1	0
2	0	0
3	1	0
4	0	1
5	1	0
6	0	0
7	0	0
8	0	1

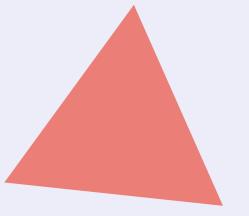
One-Hot Encoding →

Dummy Encoding →

Feature selection

- **Manuellement:** Les attributs choisis directement par le data scientist.
 - Supprimer les attributs redondants et dupliqués uniquement en fonction de leurs descriptions et d'une bonne compréhension du domaine du problème.
- **Automatiquement:** Les attributs sont choisis par un algorithme.
 - Coefficient de corrélation.
 - Chi-square.

VI-Prétraitement des données (5/6)



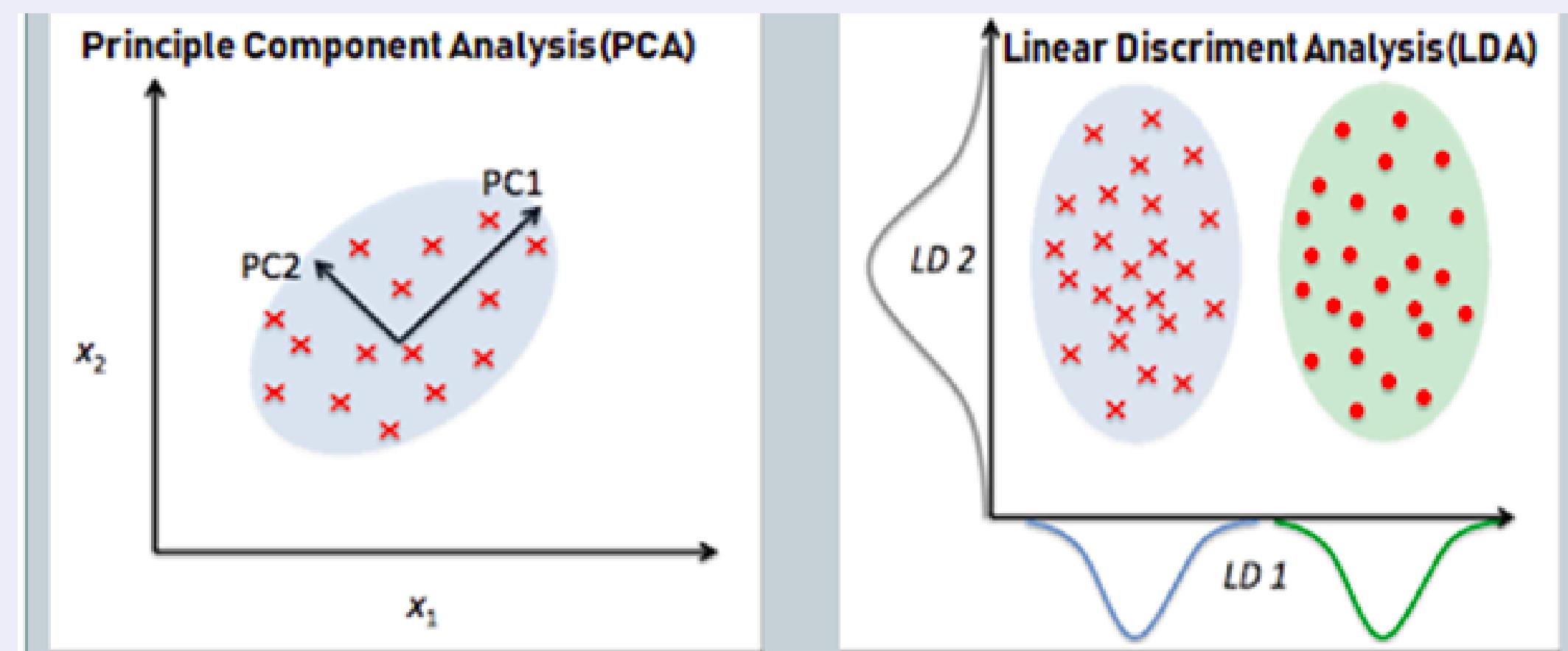
Dimensionality reduction

- PCA(Principal Components Analysis)

Il considère chaque attribut comme un axe dans un espace à n dimensions. Il effectue une transformation linéaire de cet espace en un espace p-dimensionnel plus petit .

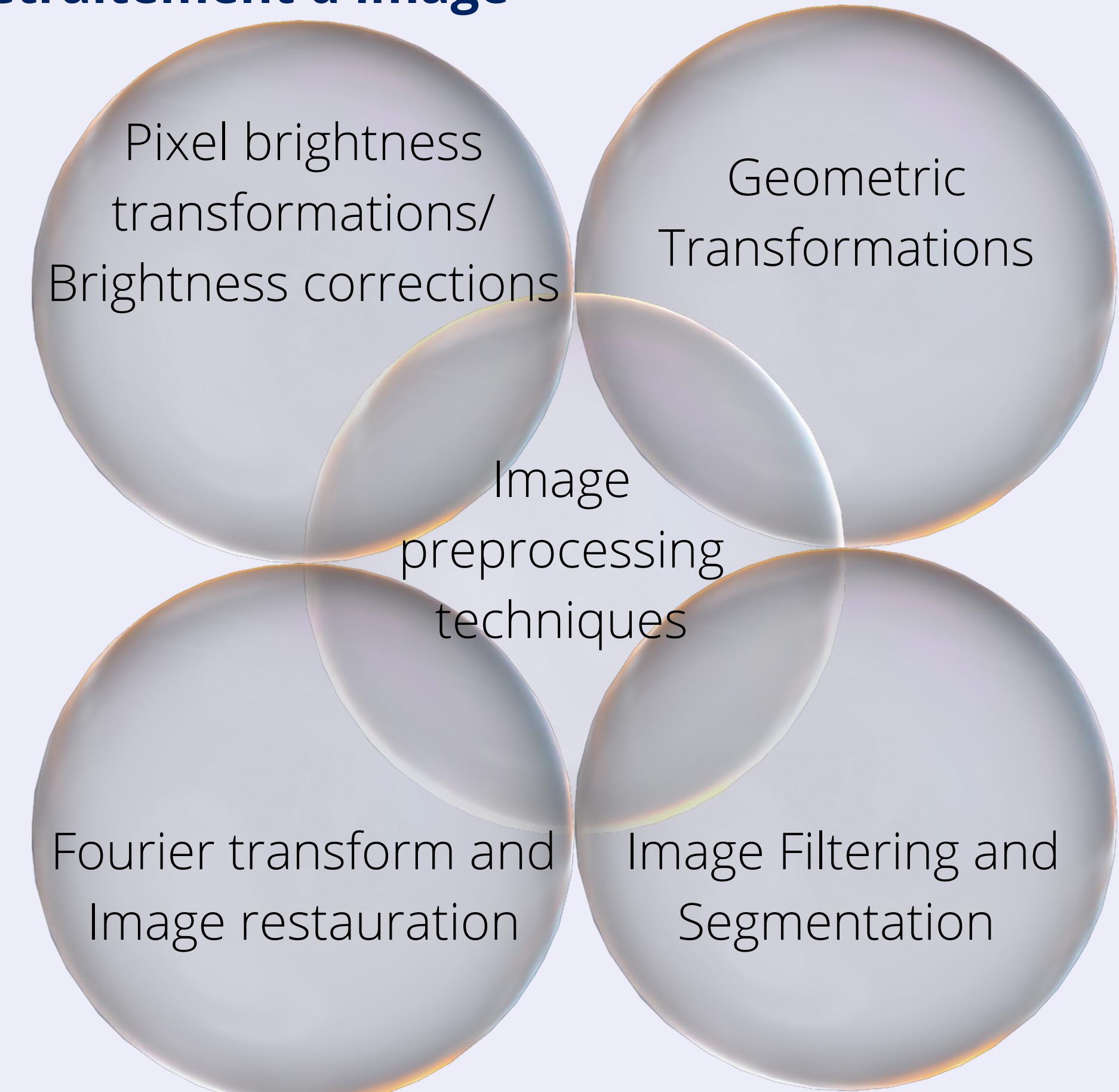
- LDA(Linear Discriminant Analysis)

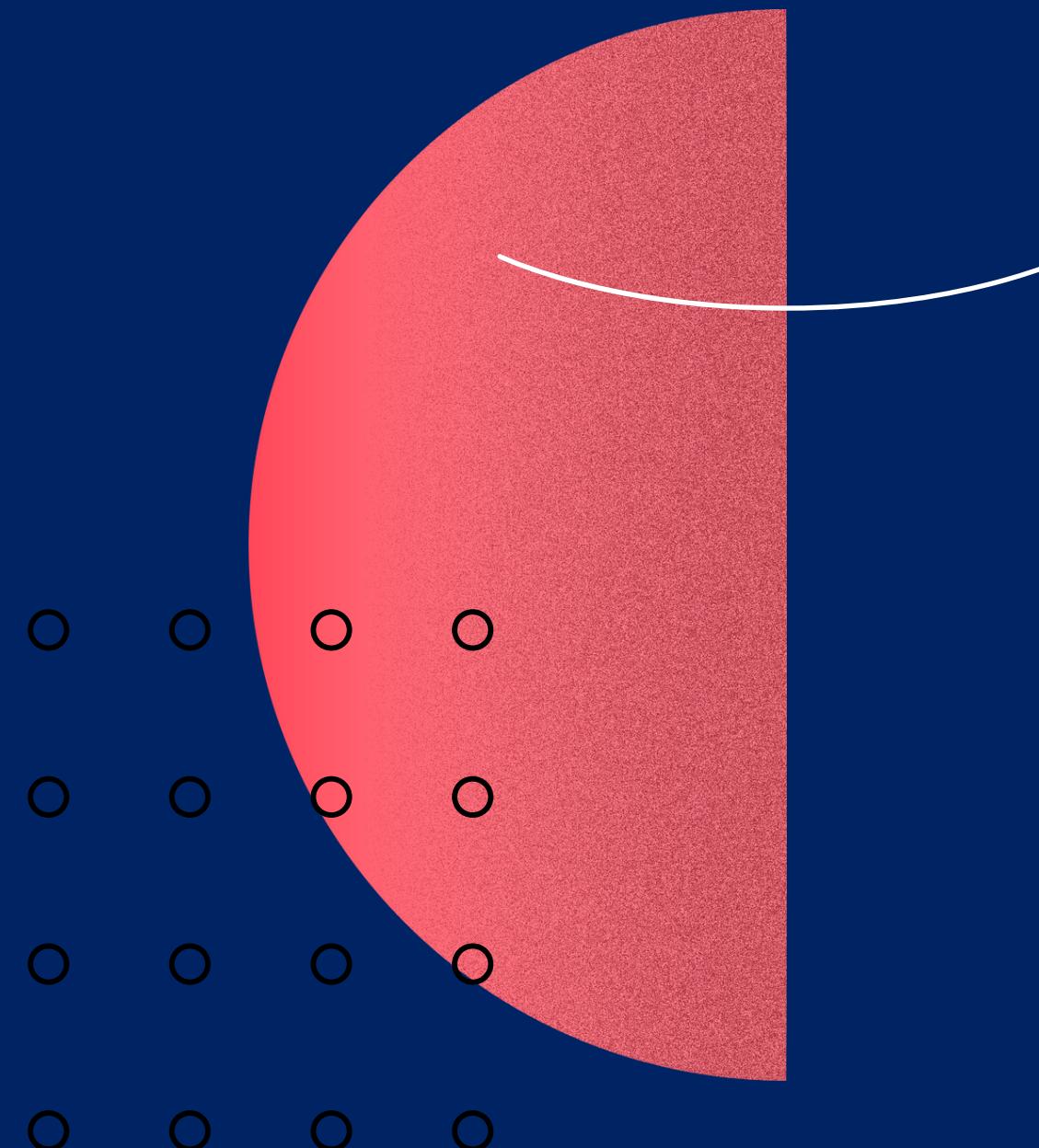
Il fournit une projection d'un ensemble de données d'apprentissage qui sépare mieux les exemples par leurs classes attribuées.



VI-Prétraitement des données (6/6)

Techniques de prétraitement d'image

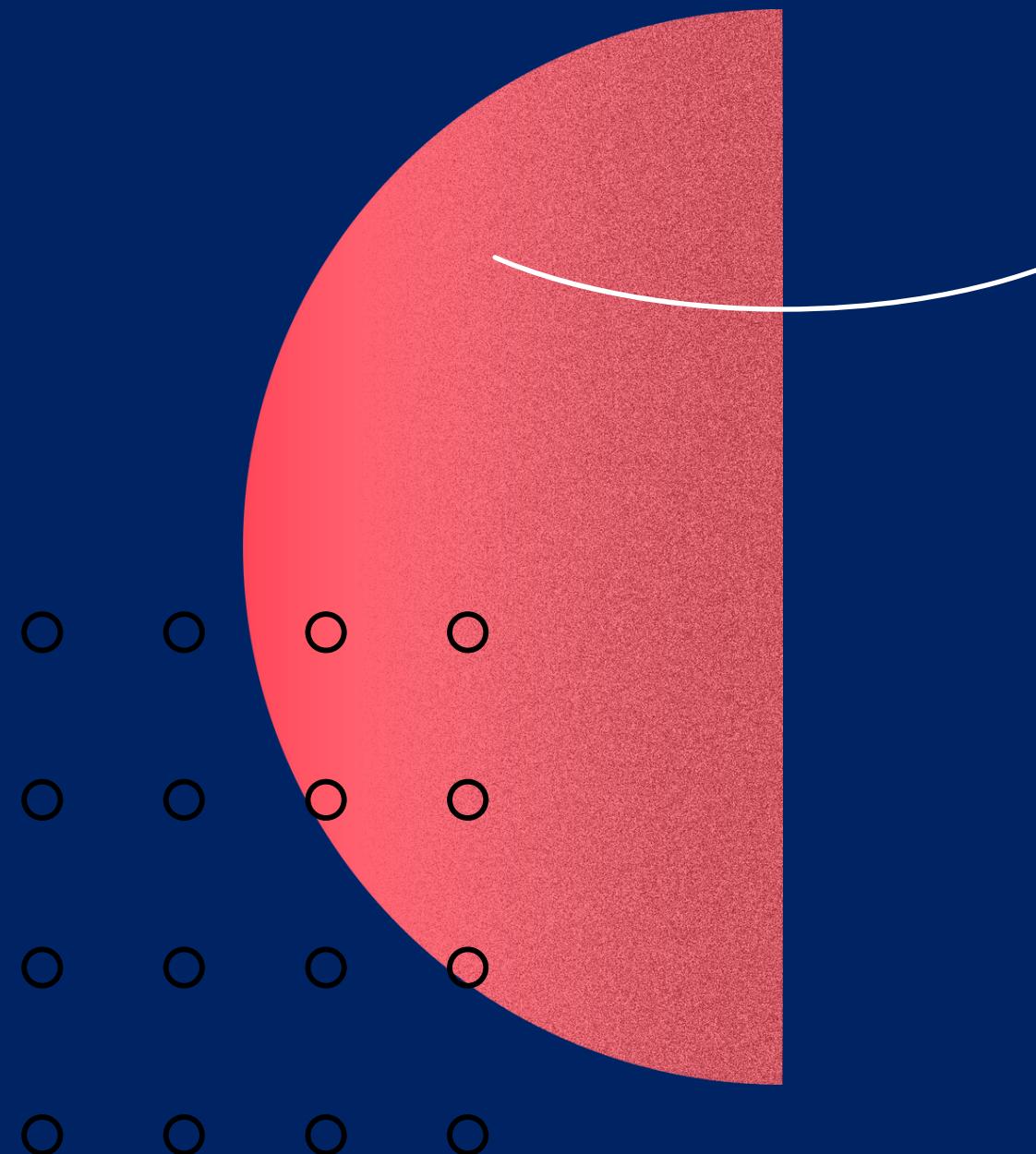




IA ET LES DÉFIS ÉTHIQUES

VII-IA ET LES DÉFIS ÉTHIQUES

- Les décisions vitales des voitures autonome: Sauver la vie d'un piéton plutôt que celle du conducteur, d'un adulte plutôt que d'un enfant .
- La reconnaissance faciale :Lorsqu'il s'agit de reconnaître le genre d'un homme blanc, des logiciels affichent un taux de réussite de 99 %. La tâche se complique lorsque la peau d'une personne est plus foncée, ou s'il s'agit d'une femme .
- Suite d'une vidéo montrant des hommes noirs, un algorithme de Facebook a proposé de continuer à voir des vidéos « sur les primates ».



DEMO

RESOURCES

Les nouvelles règles encadrant l'intelligence artificielle:

https://ec.europa.eu/commission/presscorner/detail/fr/ip_21_1682

<https://www.alcees.com/analyse-sia-versquelle-regulation-notre-guide-complet-pour-tout-comprendre5/>

<https://www.alain-bensoussan.com/avocats/nouveau-reglement-sur-lia-pour-une-ia-digne-de-confiance/2021/10/25/> (Alain Bensoussan a présenté « le nouveau règlement sur l'IA,» lors du salon Big Data & AI Paris 2021)

The Machine Learning Canvas:

<https://www.ownml.co/machine-learning-canvas>

SweetViz Library – EDA in Seconds:

<https://pypi.org/project/sweetviz/>

The IBM Foundational Methodology for Data Science (IBM Master plan):

<https://www.ibm.com/downloads/cas/WKK9DX51>



Merci de votre
attention



Talan[★]
The future is a playfield