

An isometric illustration in shades of blue and orange. It features various data visualization elements: a line graph with a rising trend, a bar chart with three bars of increasing height, and several donut charts. Small human figures are interacting with these elements: one stands next to the line graph, another stands next to the bar chart, and a third is seated at a desk with a laptop displaying a bar chart. The background is a dark blue with glowing lines and geometric shapes, suggesting a digital or data-driven environment.

# VALORISER LA DONNÉE EN ENTREPRISE

SÉBASTIEN QUINAULT / S1 2020/ M1 SARADS

# Qui suis-je ?

**Sébastien QUINAULT**

**Data scientist – Direction Data – Covéa**

**Mon parcours :**

**DUT STID -> Maitrise GIS**

**Développeur BI**

**Chargé études statistiques**

**Data analyst**

**Data scientist**

**<https://www.linkedin.com/in/sebastien-quinault>**

# Le contexte data



# L'impact du big data



## Volume

Terabytes - Petabytes

## Variety

Text – Files – Images - Video ...

## Velocity

Real-time – Batch - Stream



# Big Data

# Le « come-back » de l'IA

**L'IA est un concept : la machine est capable d'apprendre**

**wikipedia : Ensemble des théories et des techniques mises en oeuvre en vue de réaliser des machines capables de simuler l'intelligence**

# L'émergence de la data science

Ensemble de techniques permettant d'extraire de la valeur de données





# Le contexte réglementaire





**En 2020, l'entreprise « est » data  
centric  
La valeur se crée à partir de la  
donnée**



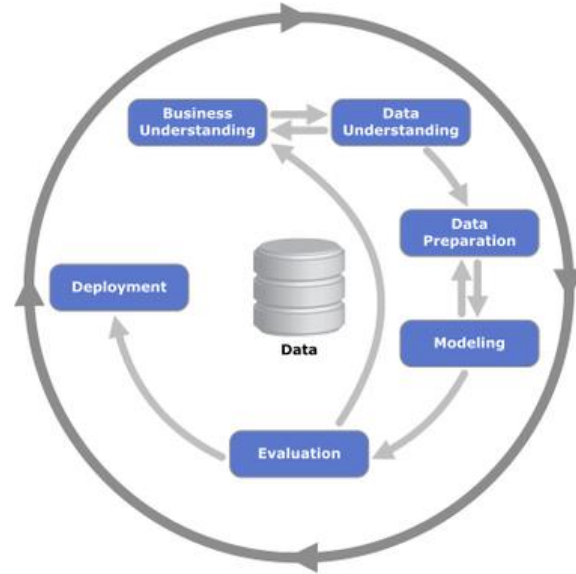
**Comment produire cette valeur ?  
Grace a vous spécialistes de la  
donnée**

**En utilisant les bonnes méthodes  
et les bons outils**

# Une méthode

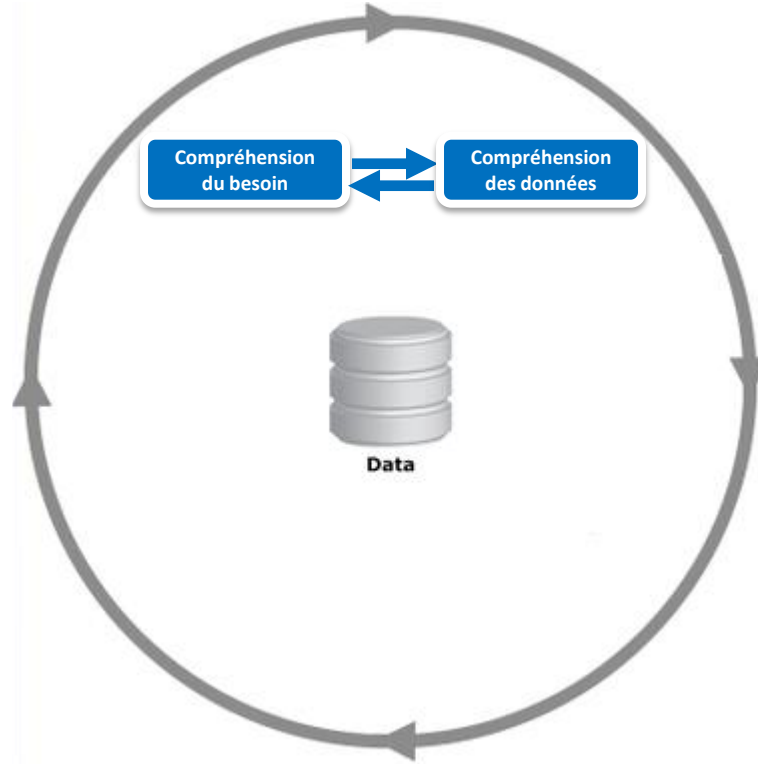


# CRISP-DM

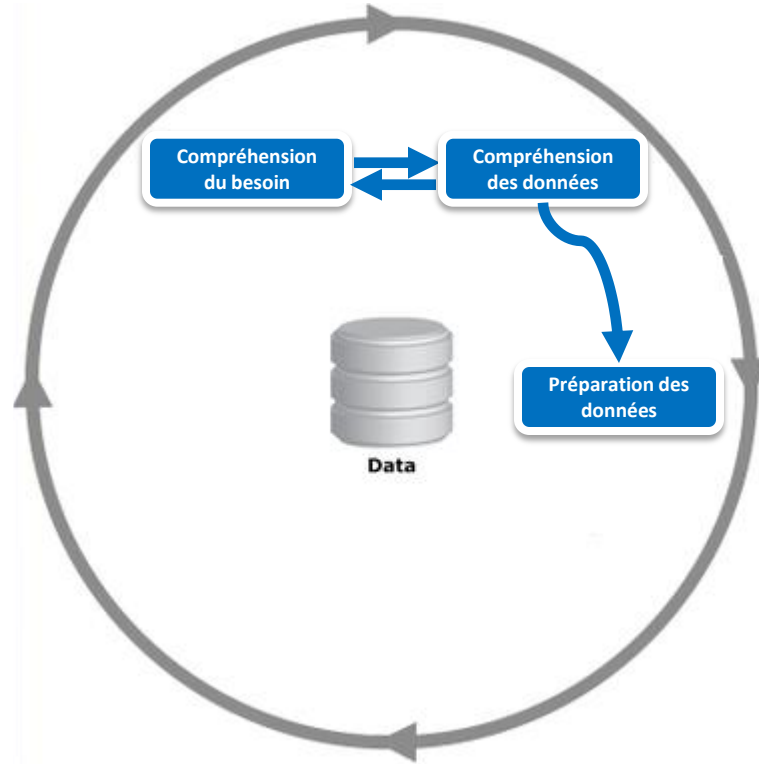




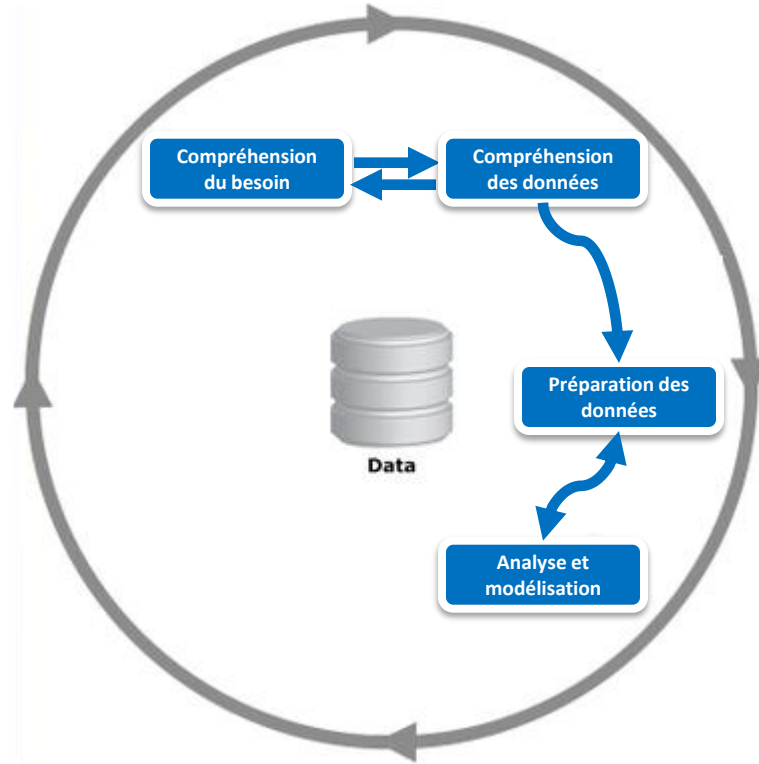
# Comprendre le besoin métier et les données disponibles



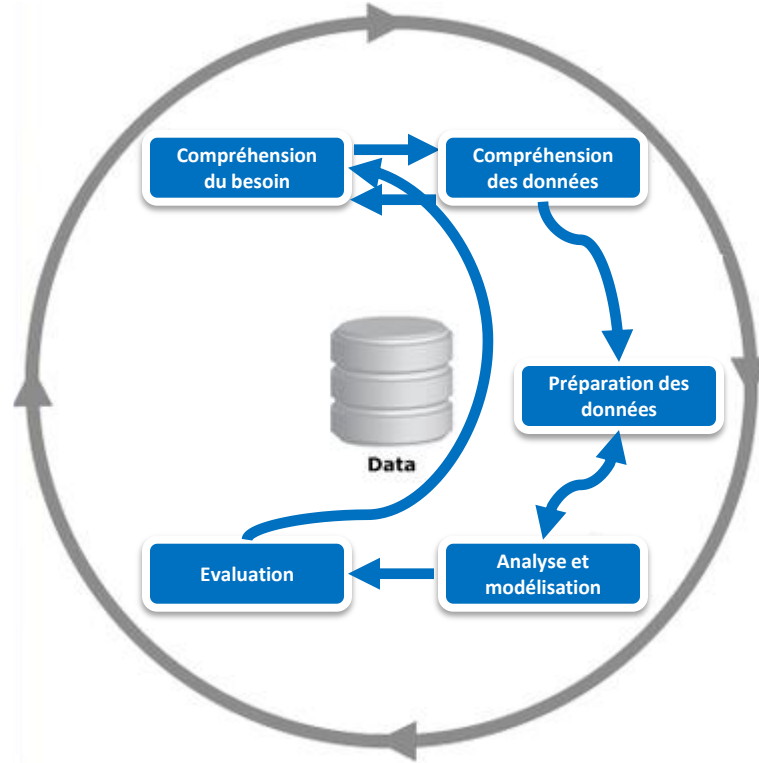
# Préparer les données



# Analyser et modéliser

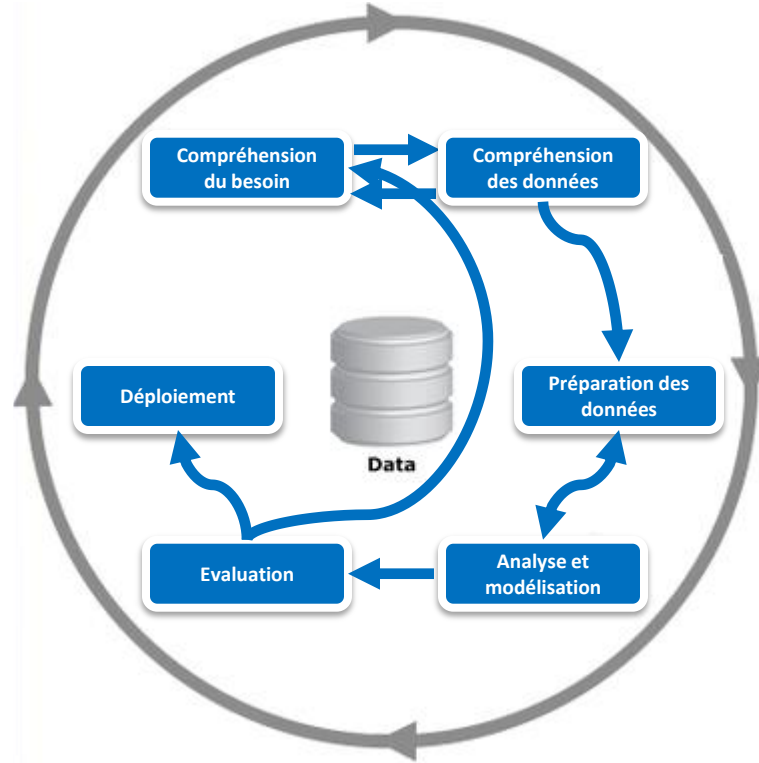


# Evaluer vos résultats





# Délivrer un produit



# Des outils



# Mais lesquels ?

# Accéder aux données – les classiques

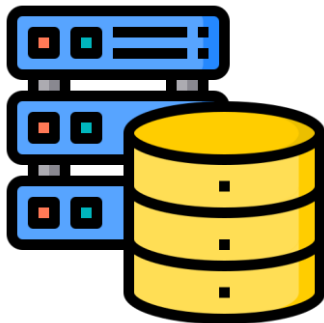


PostgreSQL





# Accéder aux données – big data



# Accéder aux données

**La valeur sure !**



# Le traitement de la donnée

En programmant



Avec des outils ‘user-friendly’



TRIFACTA

alteryx

# L'exploration de données

En programmant



Avec des outils 'user-friendly'





# Analyser et modéliser

En programmant



Avec des outils ‘user-friendly’

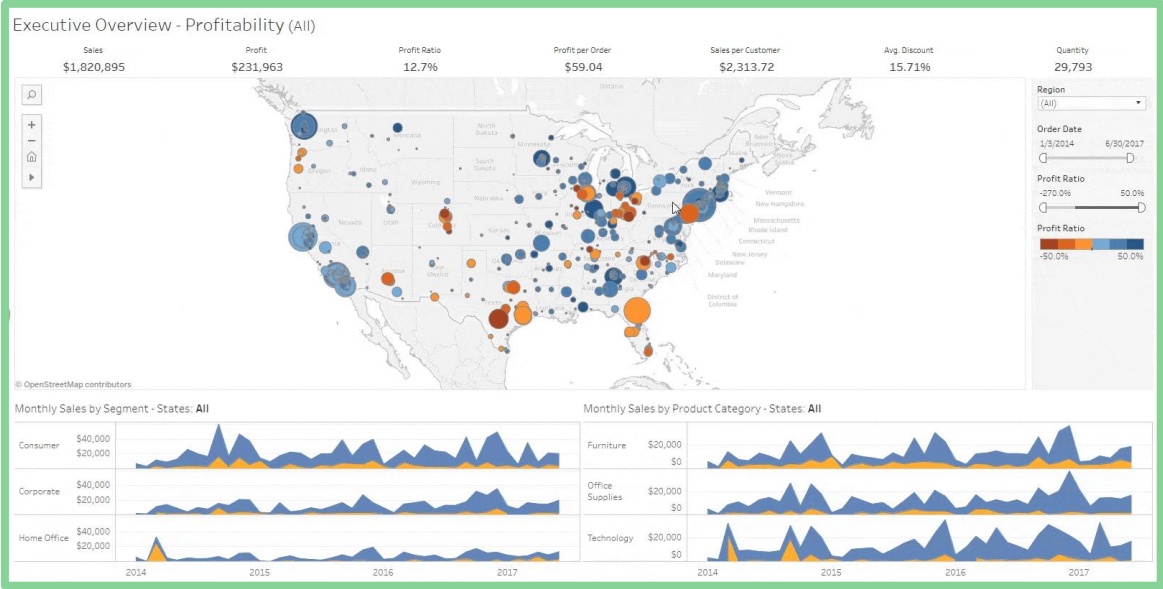


# Analyser et modéliser

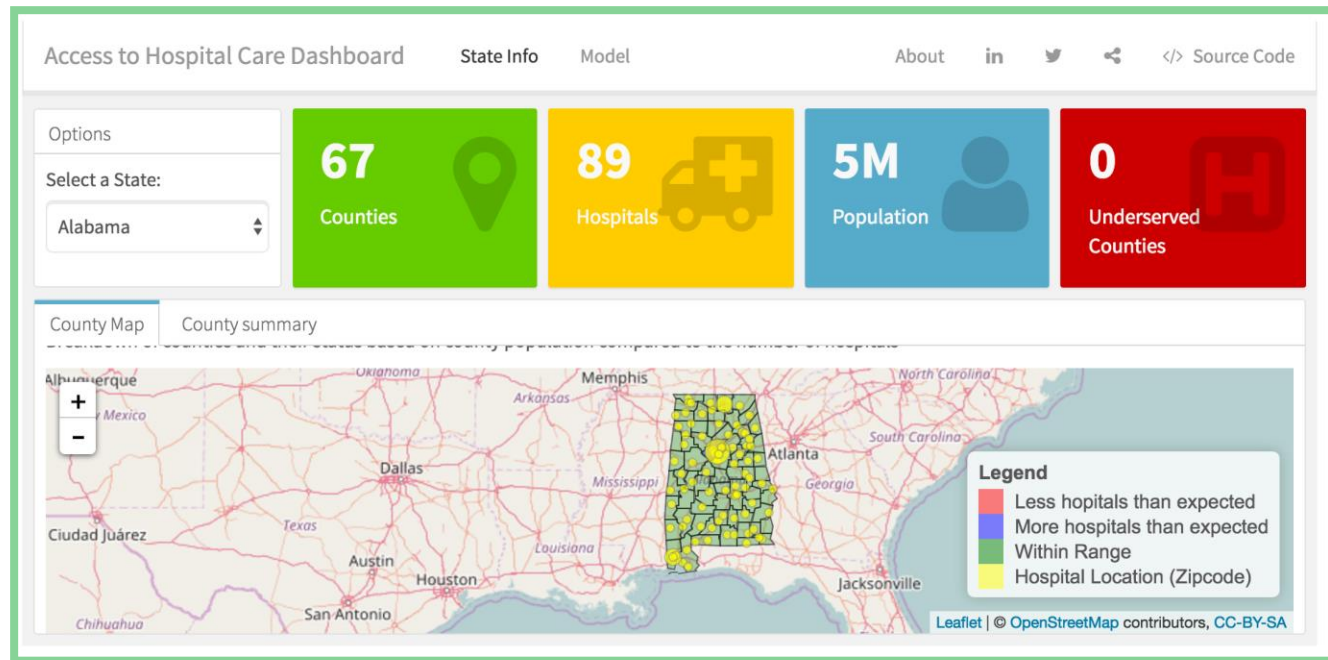
## Les frameworks de programmation



# Préserver les résultats



# Préserver les résultats



# Préserver les résultats



Streamlit



# Déployer vos résultats aux clients

**C'est la partie délicate**

**Peu de solutions « miracles »**

**La meilleure : impliquer l'IT dès le départ**



# Pour utiliser R et Python

## notebooks



## IDE



# #projet





# #opendata



[data.gouv.fr](https://data.gouv.fr)



# #formation



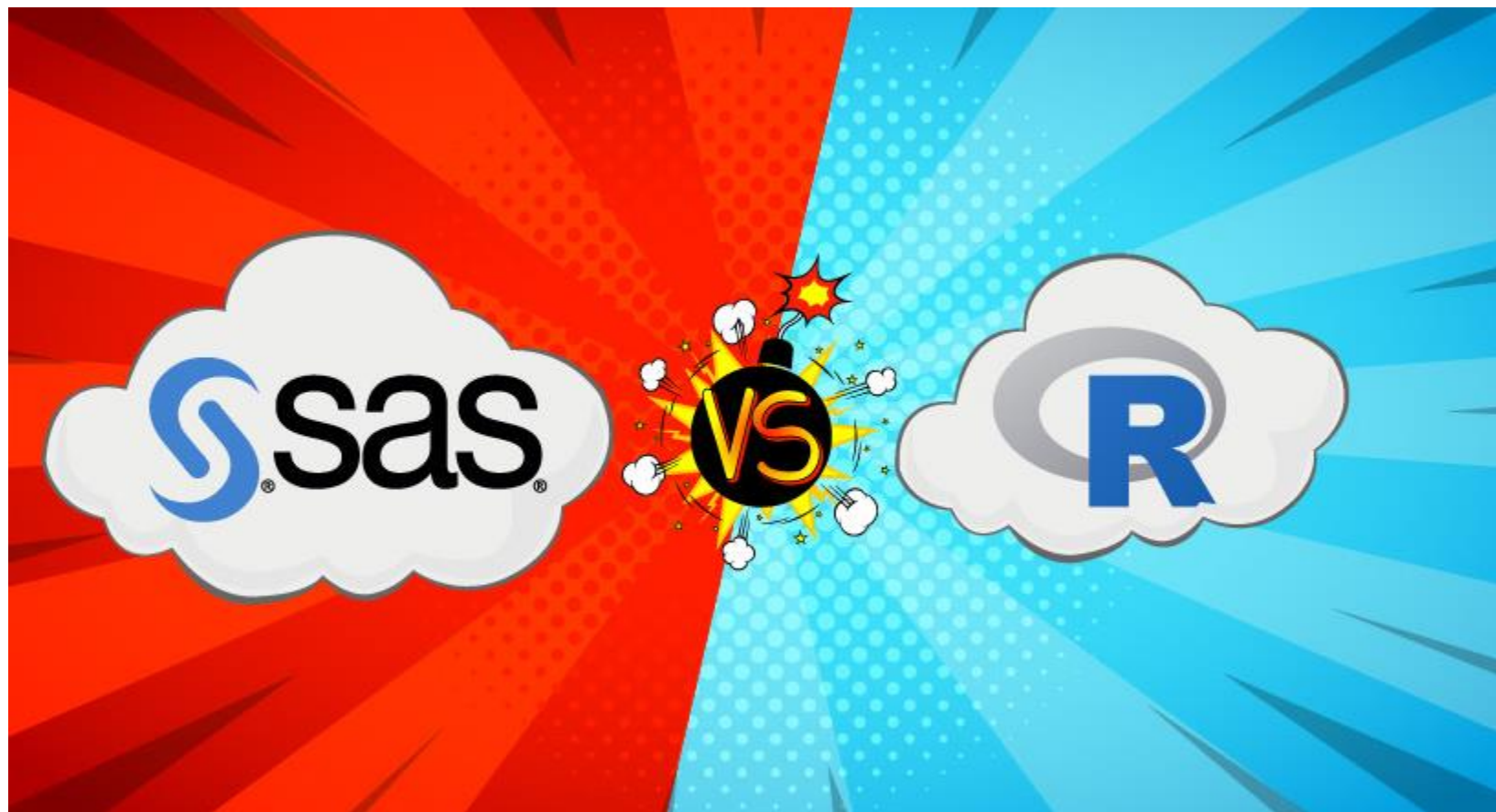
# #next step / 2025



Google Cloud Platform







# Quelques différences

**Modèle économique : licensing vs open source**

**Courbe d'apprentissage + IDE**

**Disponibilité des nouveautés**

**Restitution des résultats**

**Support éditeur / communauté / documentation**

**Patrimoine applicatif existant**



VS



# Quelques différences

**Deux langages open source avec 2 objectifs différents**

**Python + maintenable, + facile à apprendre**

**Présentation des résultats**

**API de machine learning**

**Les packages R**

**Comment choisir ?**