

Business Intelligence (BI) Informatique décisionnelle

Business Intelligence ou Informatique Décisionnelle

La Business Intelligence désigne un ensemble :

- d'**outils informatiques**
- de **ressources** : humaines, temps, budgets
- de **méthodes** alliant outils et ressources humaines

L'objectif est de **piloter** une entreprise et aider à la **prise de décision** : tableaux de bord, rapports analytiques et prospectifs



Business Intelligence ou Informatique Décisionnelle

La BI est basée sur :

- la **collecte** des données
- la **modélisation** des données
- la restitution de données **déstructurées** et **hétérogènes** générées par l'entreprise : archives papier, bases de données classiques, documents de bureautique (textes et feuilles de calcul), données collectées via des services en ligne, via les réseaux sociaux, etc.

Ces données sont traitées par des **outils d'extraction**, de transfert et de consolidation (en anglais *Extract Transform Load*, **ETL**) mis en place pour **normaliser** toutes les sources d'informations et établir une **cohérence** entre elles

Business Intelligence OU Informatique Décisionnelle

On **prépare les données** pour les rendre **présentables**

L'utilisateur a donc recours à des outils de **reporting** permettant :

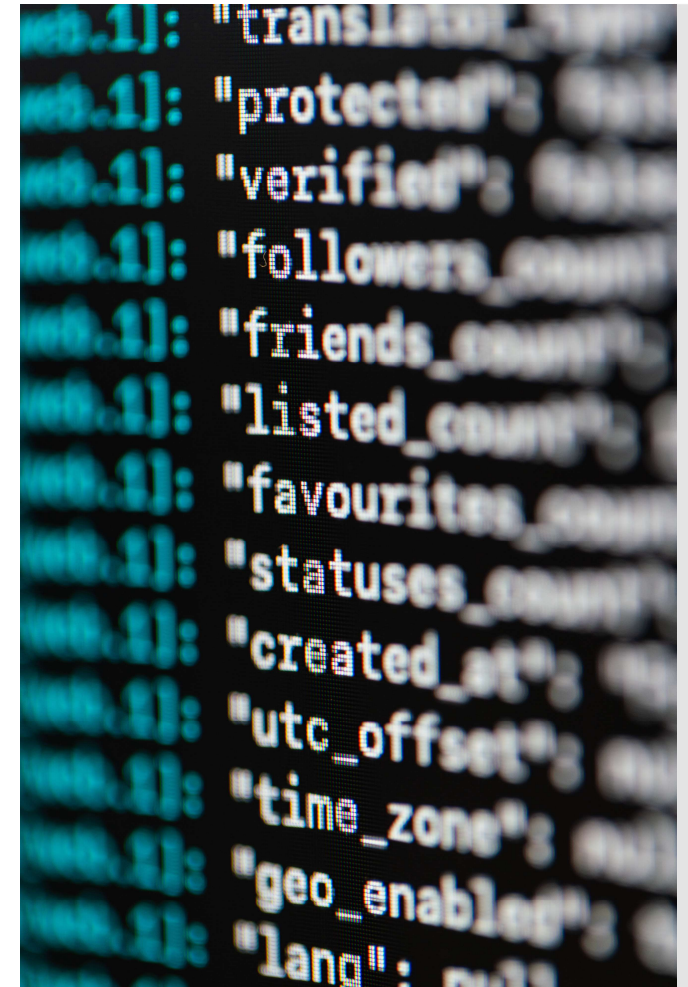
- de **sélectionner** les données selon divers **critères** : période, type de clientèle, régions, produits ...
- de **trier** les données en appliquant des **filtres**
- de **présenter** les données sous divers modes de visualisation : rapports, tableaux de bord
- de rendre les données plus facilement **exploitables** par les décideurs et acteurs de l'entreprise pour leur permettre de **prendre des décisions** plus avisées

Big Data

Le modèle des 5v

Volume

- les bases de données enregistrent des masses de données qu'il faut être en capacité de **collecter** et de **stocker**
- ces volumes sont **exceptionnels**



Big Data

Le modèle des 5v

Variété

- on collecte des données de formats **hétérogènes** qu'il faut pouvoir **corréler**
- exemples : un message texte, une image, une vidéo, un like, un tweet ... sont des données de différentes formes **difficiles à relier entre elles**
- plus la **variété** est **importante**, plus le **traitement** des données est **complexe**



Big Data

Le modèle des 5v

Vitesse

Ce critère caractérise à la fois :

- la capacité du système à **collecter rapidement** beaucoup de données et on parle de **collecte en temps réel**
- la capacité du système à **traiter** les données collectées le plus vite possible et on parle de **traitement en temps réel**

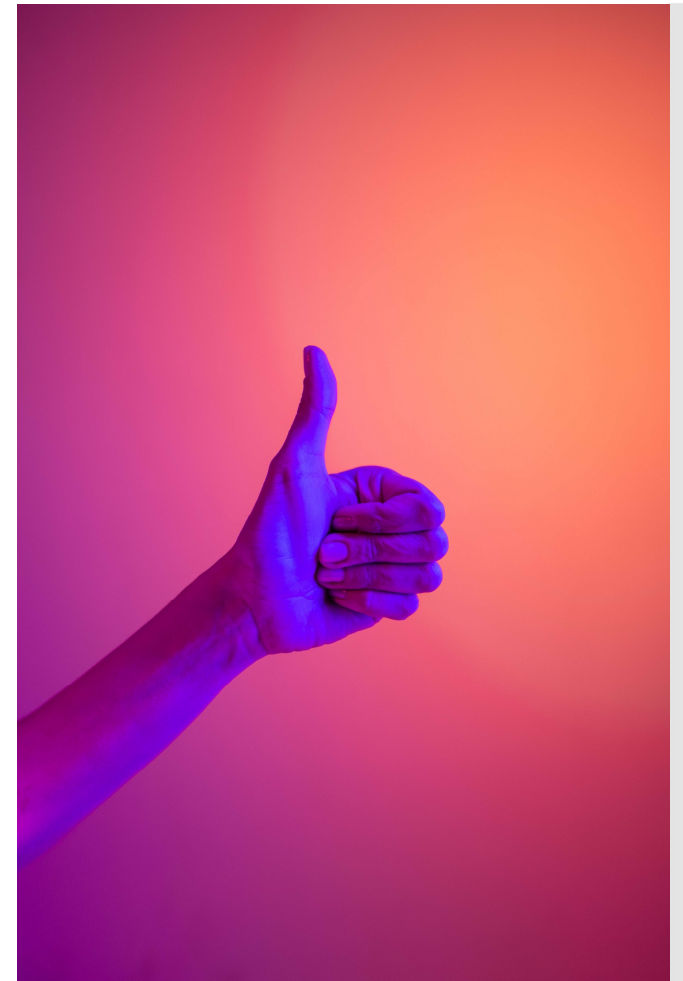


Big Data

Le modèle des 5v

Véracité

- pour assurer la **fiabilité** des données, la **qualification de la donnée** devient un critère essentiel car peu d'entreprises sont certaines de **l'exactitude** des données qu'elles collectent
- la question est : les données sont-elles **réelles** ? **pertinentes** ?

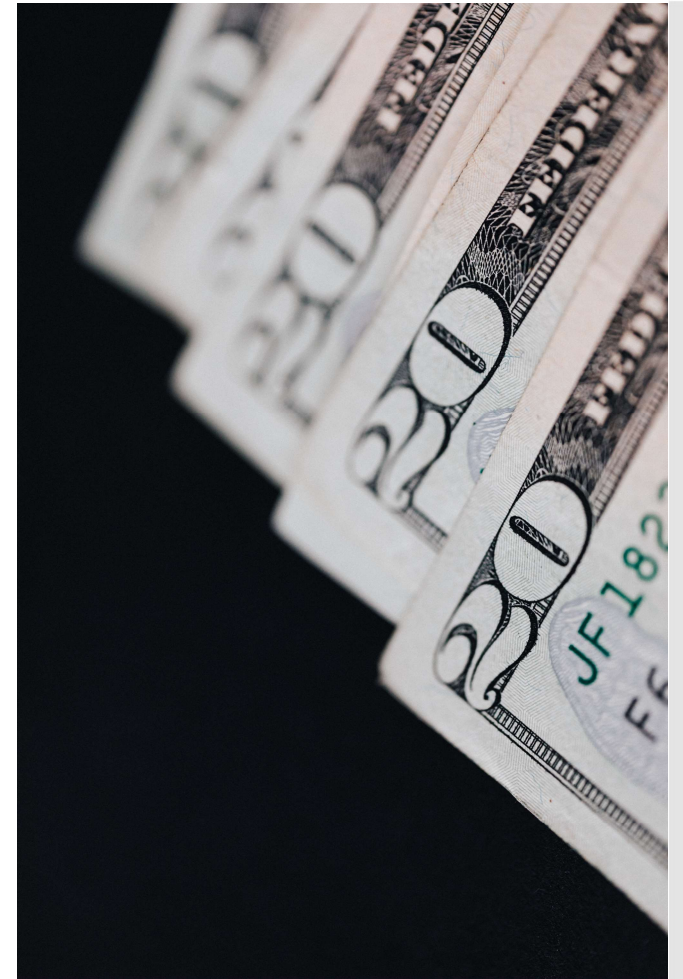


Big Data

Le modèle des 5v

Valeur

- la finalité de l'analyse des données est de les **valoriser** au travers de **croisements inédits** et les **faire parler** pour **obtenir de nouveaux indicateurs**
- la question est : quelle est la **valeur ajoutée** d'une donnée ?



Data Science

Définition

- la **Data Science**, littéralement **science des données**, est une **discipline** qui étudie les **informations** et leurs **sources**, ainsi que les **méthodes** permettant de les **transformer** en ressources utiles pour la création de **stratégies d'entreprise**
- **l'extraction** d'importants volumes de données permet aux organisations d'identifier des **modèles** qui les aident à maîtriser leurs coûts, à améliorer leur efficacité, à identifier de nouveaux débouchés commerciaux et à améliorer leur avantage concurrentiel
- ce processus, mené par des Data Scientists, est qualifié **d'exploration des données**, c'est le **Data Mining**

Data
Science

Data
Scientist

- le Data Scientist crée des **modèles de visualisation** des données
- il présente des **informations décisionnelles exploitables**, notamment aux **décideurs**, et en **explique clairement la signification**
- il facilite et améliore la **prise de décisions** dans les entreprises
- il exploite des données issues de **sources multiples** et doit donc disposer de **nombreuses compétences**
 - **sources multiples** : informations numériques venant d'ordinateurs, de smartphones, de l'Internet des objets (IoT), des réseaux sociaux, d'enquêtes, d'achats, de recherches et comportements sur Internet ...
 - **compétences** : la Data Science fait appel aux **mathématiques**, aux **statistiques**, à **l'informatique** et au **Data Mining** notamment

Data Science

Développement

- la Data Science se développe assez **lentement** dans l'entreprise car l'identification et l'analyse d'un volume colossal de données non structurées est **complexe, onéreuse** et **chronophage**
- cette discipline **manque de compétences** : la **formation** de spécialistes est **récente** et encore **incertaine** quant aux résultats et compétences réellement acquises
- l'avenir réside donc :
 - dans la **formation qualitative** de nouveaux data scientists
 - dans **l'amélioration des outils**, notamment **l'apprentissage statistique** qui relève de **l'intelligence artificielle** afin d'automatiser le traitement des données

Data Science

Et demain ?

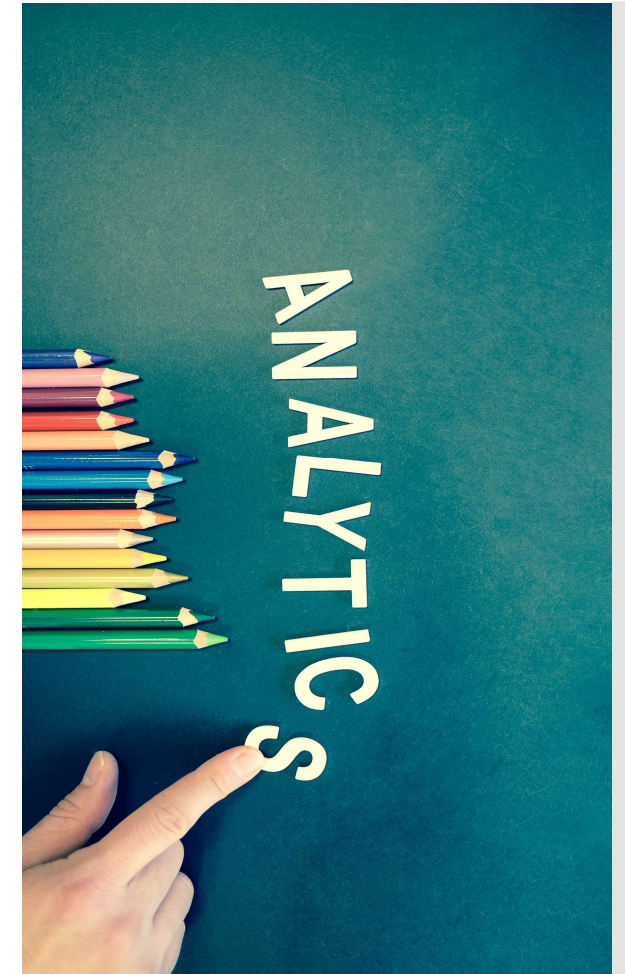
- ces algorithmes évolués font leur **propre apprentissage** et peuvent traiter de très grandes quantités de données en peu de temps, **comparativement à ce que l'homme peut faire**
- les applications d'apprentissage automatique utilisées dans le champ de la Data Science intègrent par exemple la **reconnaissance d'images** et la **reconnaissance vocale**
- exemple : des algorithmes d'apprentissage automatique sont embarqués dans les véhicules autonomes afin de **parfaire leurs comportements**

Data Mining
Prévoir
et anticiper l'avenir

Data Mining

Définition

- le Data Mining (exploration de données) est un processus informatique qui analyse de grandes quantités d'informations
- ce processus est la source des (Big) Data Analytics, c'est-à-dire :
 - des analyses prédictives
 - de l'exploitation des données collectées
- autres termes utilisés :
 - forage de données
 - exploration de données
 - fouilles de données



Data
Mining

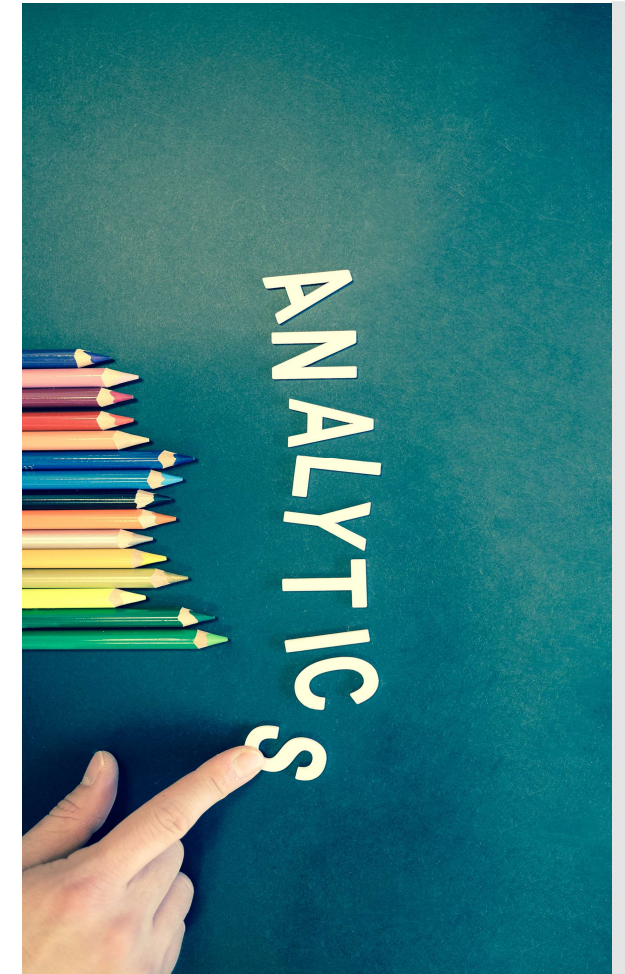
Définition

Objectifs ?

- **comprendre** les données acquises
- **extraire** les informations les plus utiles
- les **transformer en connaissances**

Comment ?

- en établissant des **relations** entre les données
- en **repérant des patterns (modèles)**



Une première illustration

Des produits pour bébés ...

- une chaîne de grands magasins remarque que les achats de **produits pour bébés**, tard le soir, sont **corrélés** à l'augmentation des achats de **bières**
- coïncidence ? non ... **l'analyse approfondie des profils des acheteurs** révèle que les clients sont principalement des jeunes pères âgés de 25 à 35 ans
- l'analyse montre également que ces **ventes corrélées** sont plus importantes les **jeudis** et **samedis**



... et
des bières !

- l'enseigne commercialise alors ses produits pour bébés et les bières dans des **linéaires très proches**
- les **ventes combinées** des deux gammes de produits **montent en flèche**
- l'enseigne **ne solde jamais** ces deux produits les **jeudis** et les **samedis**
- **l'exploration des relations** entre des données et des événements à priori sans rapport améliore la **prise de décision commerciale**



Data Mining en 5 phases

- 1 - collecte, extraction, transformation, chargement des données dans un dépôt (voir Data Warehouse)
- 2 - stockage organisé ou non dans une base de données, relationnelle ou non
- 3 - accès aux données pour les :
 - informaticiens (concepteurs / développeurs)
 - data analysts qui expriment les besoins et exploitent les informations préparées par les informaticiens
- 4 - analyse des données et recherche des patterns (modèles) :
 - logiciels
 - data analysts
- 5 - présentation et partage des données sous un format exploitable et compréhensible : rapports, graphiques, tableaux ...

Analyse grâce à 4 types de relations

1 – classes

- l'analyse de certaines données stockées permet de localiser d'autres données en **groupes prédéterminés**
- exemple de minage par classes de données : une entreprise de livraison de repas à domicile mine les commandes passées par ses clients pour déterminer :
 - les dates et créneaux horaires des commandes
 - le contenu des commandes
- cela leur permet **d'optimiser** :
 - leurs stocks de produits
 - la disponibilité de leurs livreurs
 - leurs tournées de livraison ...

Analyse grâce à 4 types de relations

2 – clusters

- les données sont découpées en clusters par rapport à des relations logiques (cohérentes)
- exemple de minage par cluster : le Web
 - il contient des milliards de pages
 - les résultats fournis par un moteur de recherche peuvent donner des millions de réponses
 - le clustering est utilisé pour **regrouper** ces résultats en quelques **clusters**, chacun reprenant un **élément spécifique** de la requête
 - par exemple, une requête sur le cinéma donne des pages Web regroupées dans des clusters de catégories : films, critiques, bandes-annonces, actrices et acteurs, cinémas ...
 - ceci **facilite l'exploitation ultérieure** des données

Analyse grâce à 4 types de relations

3 - associations (corrélations)

- les données sont minées pour identifier des **associations** (ou corrélations)
- exemple de minage associatif : les couches pour bébés et les bières achetées par les papas âgés de 25 à 35 ans spécialement les jeudis et samedis

4 - patterns (modèles) séquentiels

- les données sont minées pour prévoir **des modèles comportementaux** et des **tendances**
- on parle aussi **d'anticipation**
- exemple : un vendeur d'équipements extérieurs peut prévoir les ventes de sacs à dos en se basant sur l'analyse des ventes des sacs de couchage et de chaussures de randonnée

Data Mining

Infrastructures et coûts liés

Le **prix des infrastructures** est très **variable** : de quelques **milliers** à plusieurs **millions** de dollars selon la **nature des applications** et le **volume des données** stockées

Les critères :

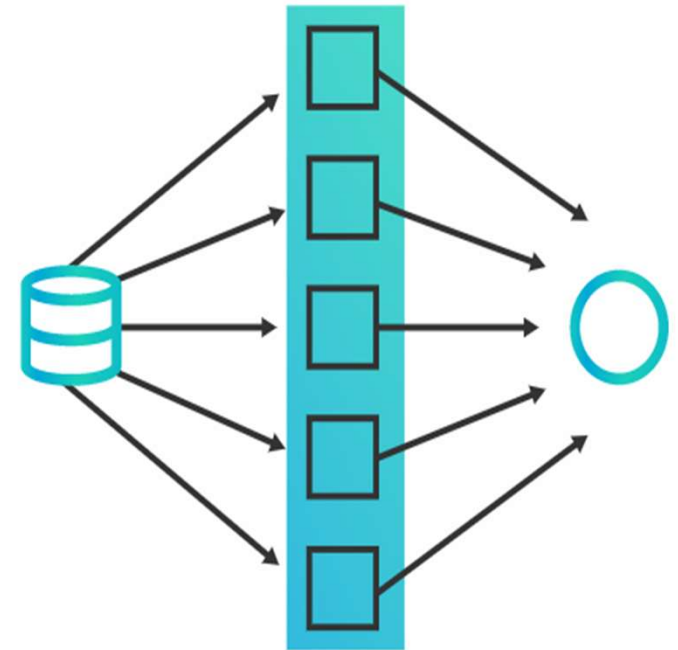
- le **volume** des données à stocker et à traiter
- la **complexité des requêtes** à développer
- la **complexité des requêtes** à exécuter : plus une requête est complexe plus elle demande de **puissance machine**
- l'usage de produits **Open Source** ou **sous licences**



Data Mining

Infrastructures et coûts liés

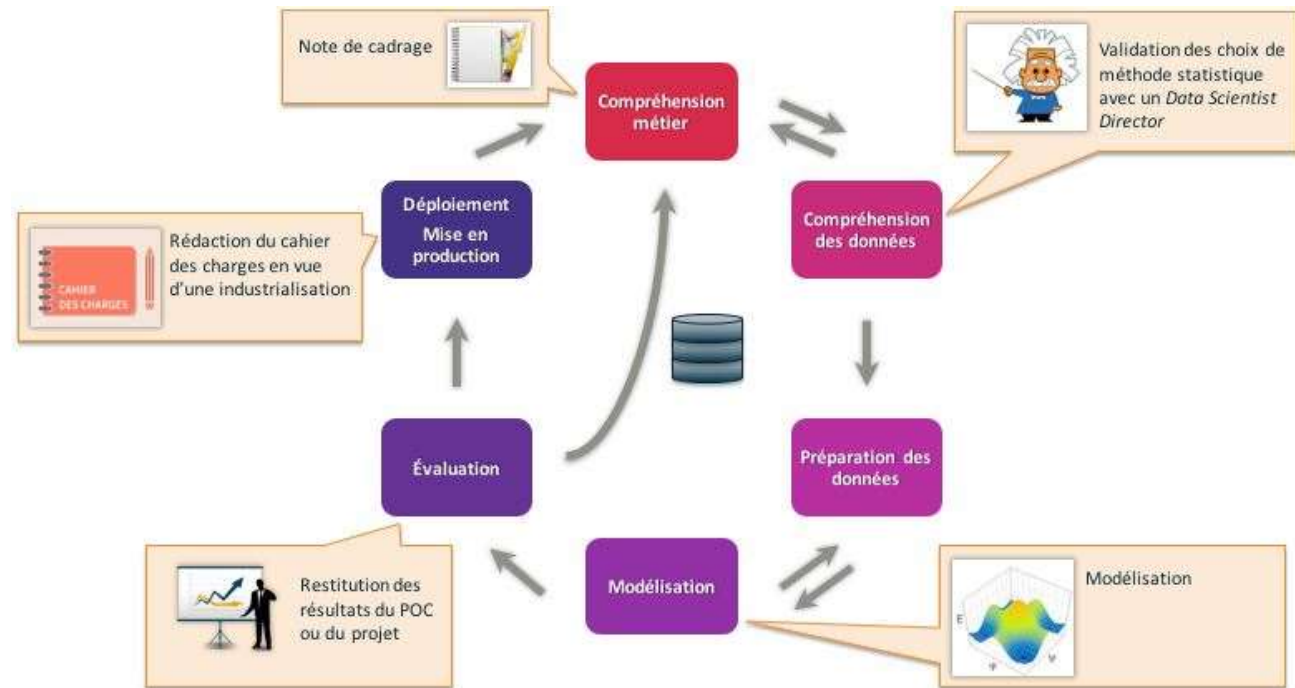
- le **type de base de données** :
 - **base relationnelle** : efficacité communément admise jusqu'à **50 gigabytes** de données
 - **base non relationnelle** : (NoSQL par exemple) pour traiter des volumes supérieurs. Ce type de base est couplé à des **architectures matérielles plus évoluées** (MPP : Massively Parallel Processing)



MPP : Massively Parallel Processing

Data Mining

La méthode CRISP



Cross-Industry Standard Process (for Data Mining)

<https://www.ibm.com/docs/fr/spss-modeler/saas?topic=dm-crisp-help-overview>

Data Mining

La méthode CRISP

Les 6 étapes

Méthode en 6 étapes créée par IBM au cœur des **années 60**

1 - compréhension du métier → connaître :

- les éléments métiers
- les problématiques à résoudre et les objectifs visés

2 - compréhension des données → déterminer :

- quelles sont les **données à analyser**
- la **qualité** des données existantes (intégrité, véracité)
- la **quantité** de données nécessaires à une analyse pertinente
- les **liens** à identifier entre les données avec l'aide de l'entreprise car la **connaissance « métier »** est indispensable

Data Mining

La méthode CRISP

Les 6 étapes

- 3 - construction de la base de données (Data Hub)
 - **préparation** des données (harmonisation des formats)
 - **classement** des données selon critères fournis par l'entreprise
 - **nettoyage** des données
 - rendre les données **compatibles** avec les algorithmes qui seront utilisés
 - création d'une **base de données structurée ou non**

Data Mining

La méthode CRISP

Les 6 étapes

4 – **modélisation** → générer de la **connaissance**

- choix, paramétrage et tests de différents algorithmes pour créer un modèle
- 1^{er} objectif : être **explicatif**, justifier le passé et le présent en fonction des données que l'on a collectées
- 2^{ème} objectif : être **prédictif**, expliquer l'avenir
- 3^{ème} objectif : être **prescriptif**, anticiper les futures situations en proposant déjà des **optimisations possibles**

Data Mining

La méthode CRISP

Les 6 étapes

5 - évaluation

- vérifier si le modèle et les connaissances acquises répondent aux objectifs de l'entreprise
- corriger / améliorer le modèle

6 - déploiement

- mise en production du modèle
- diffuser de manière intelligible la connaissance obtenue
- exemples : communication web, communication interne (rapports, tableaux ...)

Data Mining

Quelques logiciels

Quelques outils Open Source :

- Knime <https://www.knime.com>
- Orange Data Mining <https://orangedatamining.com>
- RapidMiner <https://rapidminer.com>
- Weka (Université de Waikato - NZ)
<https://www.cs.waikato.ac.nz/ml/weka/index.html>

Quelques produits à licences :

- SPSS (IBM) <https://www.ibm.com/fr-fr/spss>
- SAS https://www.sas.com/fr_fr/home.html
- Azure Analysis Services (Microsoft)
<https://azure.microsoft.com/fr-fr/products/analysis-services/>

Data Mining

Inconvénients pour les entreprises

- les **coûts** sont difficilement maitrisables à ce jour :
 - **technologies de pointe** en constante évolution
 - **compétences humaines rares**
- les entreprises sont tributaires :
 - du coût des **infrastructures**
 - du coût des **développements** et améliorations des **logiciels** car ce sont des technologies évoluées, développées et entretenues par des **spécialistes**
 - du coût de **recrutement** des spécialistes du domaine
 - des **rémunérations** des spécialistes : informaticiens, data analysts
- elles n'ont **aucune garantie** du **retour sur investissement**

Data Mining

Inconvénients pour les clients et utilisateurs que nous sommes

- la **collecte massive sans consentement** : réseaux sociaux, sondages ...
- l'utilisateur / client reste le premier **responsable de ses propres informations**
- la **sécurité des données**
- le **profilage** et la nécessité de **l'encadrer** (voir RGPD)



Data Mining

Formations

- Universités
- Ecoles informatiques avec spécialisation Big Data / Data Mining / Machine Learning
- Ecoles non spécialisées proposant des modules dédiés
- MOOC
 - <https://www.fun-mooc.fr/fr/cours/sinitier-la-data-science-et-ses-enjeux/>
 - <https://www.my-mooc.com/fr/>
- Attention: la **qualité** de tous ces enseignements reste encore à **démontrer** sur le **long terme** (domaine trop récent)



Data Mining

Un métier d'avenir ?

- la **demande en Data Scientists** au niveau mondial augmente chaque année
- ce secteur n'est pas encore saturé : les grands acteurs du numérique prévoient que la demande pour ces spécialistes **augmentera de 15 à 20% par an au moins jusqu'en 2025**
- les prévisions ne peuvent pas aller au-delà, car nous sommes dépendants :
 - de **l'évolution des technologies** actuelles
 - de notre **capacité à former** ces spécialistes
 - du nombre de **nouvelles entreprises** qui feront appel à l'analyse des données
 - de l'augmentation du **volume mondial des données**

Data
Mining

Illustrations

La médecine

La possession des données
des patients telles que :

- les **dossiers médicaux**
- les **examens** physiques
- la **fréquence** des consultations
- la **nature** des interventions

permet de **prescrire des**
traitements plus efficaces



Data
Mining

Illustrations

La médecine

Cela permet une **gestion plus efficace et rentable** des ressources de santé :

- en **identifiant** les risques
- en **prédisant** les maladies dans certains segments de la population, régions ou pays
- en **prévoyant une durée d'hospitalisation**

Cela peut **renforcer les liens avec les patients** avec une meilleure connaissance de leurs besoins



Data
Mining

Illustrations

Les médias

Les médias utilisent l'exploration de données en temps réel pour **mesurer leurs audiences**, télévision et radio

Ces systèmes collectent et analysent des informations parfois anonymes provenant des **émissions** ainsi que des **rediffusions** (replay, podcasts ...)



Data
Mining

Illustrations

Les médias

Ils peuvent ainsi :

- connaître les centres d'intérêts des téléspectateurs et auditeurs
- faire des recommandations personnalisées

Les médias fournissent également ces éléments à leurs annonceurs

Ces annonceurs vont les utiliser pour cibler plus précisément leurs clients potentiels



Banques :
du
mécontentement
...

Pour répondre au **mécontentement** de leurs clients quant à leur qualité de services, les banques ont fait appel au Data Mining

Les données analysées concernent les **sites web** et les **applications mobiles** des banques :

- leur **usage**
- leur **fréquentation**
- les **temps passés** sur ces sites et applications
- les **commentaires** des clients

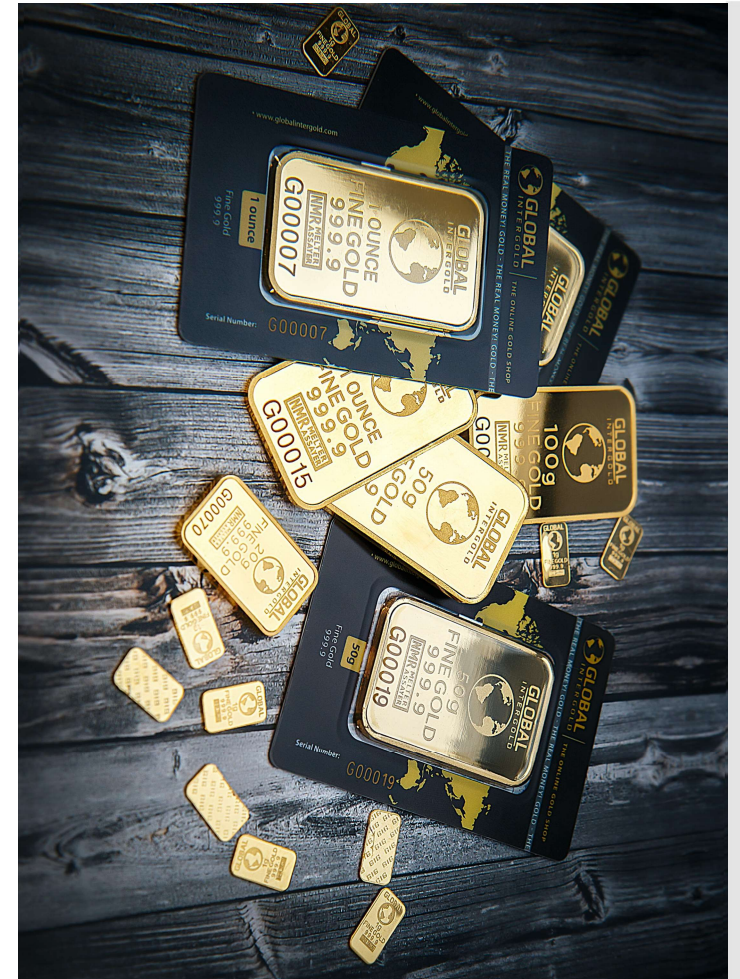


... à la fidélisation

Elles constatent :

- l'importance des **services mobiles**
- la nécessité de les rendre plus **accessibles**
- la nécessité de les **personnaliser**
- la nécessité de les **sécuriser**
- et la très grande part qu'ils jouent dans **l'appréciation de la qualité des services** et de la banque

Afin de rétablir la confiance et construire une **relation client durable**, ces **services** sont alors **améliorés** et deviennent même des **arguments commerciaux**



Aéronautique

Maintenance prédictive et sécurité

- le Big Data et le Data Mining contribuent à notre **sécurité** pendant un vol
- grâce à des **algorithmes** capables de **prévoir** la date à laquelle les pièces seront déficientes, prenant en compte des facteurs tels que le temps, les contraintes exercées ou la fréquence d'utilisation, une **maintenance préventive** est mise en place
- de potentiels **accidents** sont **évités**
- l'industrie aéronautique est gagnante en investissant dans des **processus de maintenance efficaces**



Compagnies aériennes

Le Dynamic Pricing

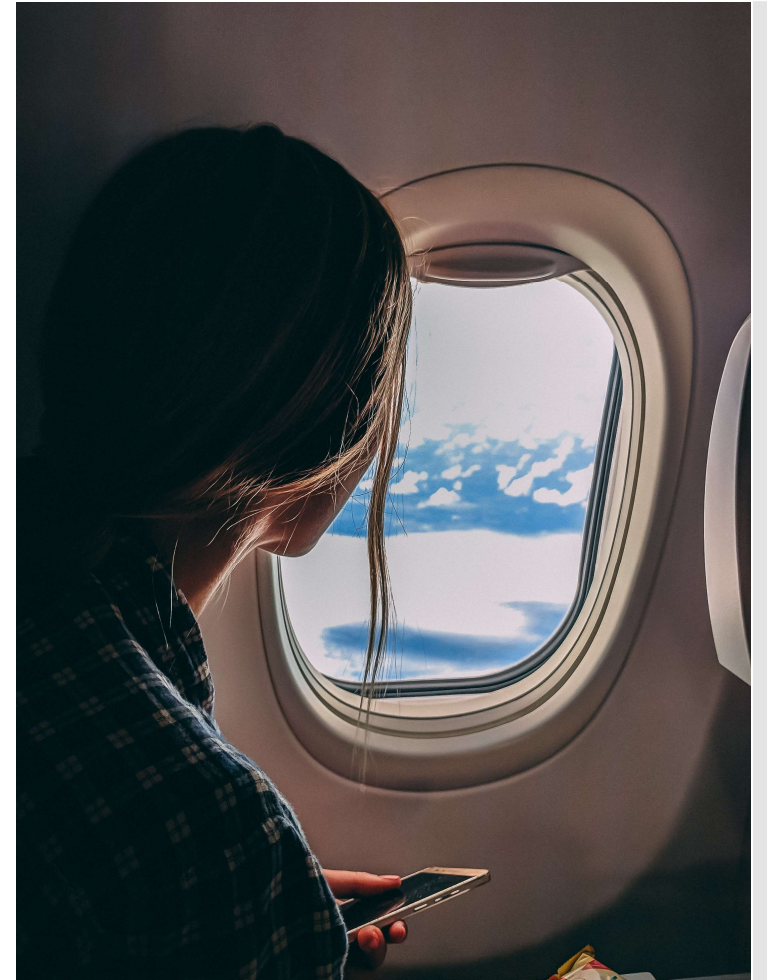
- les compagnies aériennes ajustent les **prix des billets** en fonction des **tendances**, des **profils**, des **destinations** et des **habitudes** de voyages des clients
- le principe est le suivant : la compagnie **identifie** la personne qui cherche un vol, puis **exploite ses données** grâce à son **compte client** et à son **historique** pour faire varier le prix du billet



Compagnies aériennes

Le Dynamic Pricing

- le **taux de conversion** (visite d'un site ou usage d'une application mobile qui se **transforme en commande ferme**) augmente en proposant :
 - des **tarifs plus abordables** aux **nouveaux clients**
 - des **billets plus chers** aux **voyageurs business** qui sont moins susceptibles d'être rebutés par des prix plus élevés
- **2 places similaires** dans le même avion ne sont donc **pas vendues au même prix**



Crème solaire

Des clients inattendus

- une entreprise internationale de cosmétiques demande une **étude du profil de ses clients**
- l'agence qui en est chargée est **très expérimentée** dans ce domaine car elle exploite les traces laissées par les internautes depuis le tout début d'internet
- et **pourtant** ... l'agence commente son étude comme ceci :



Crème solaire

Des clients
inattendus :

les golfeurs !

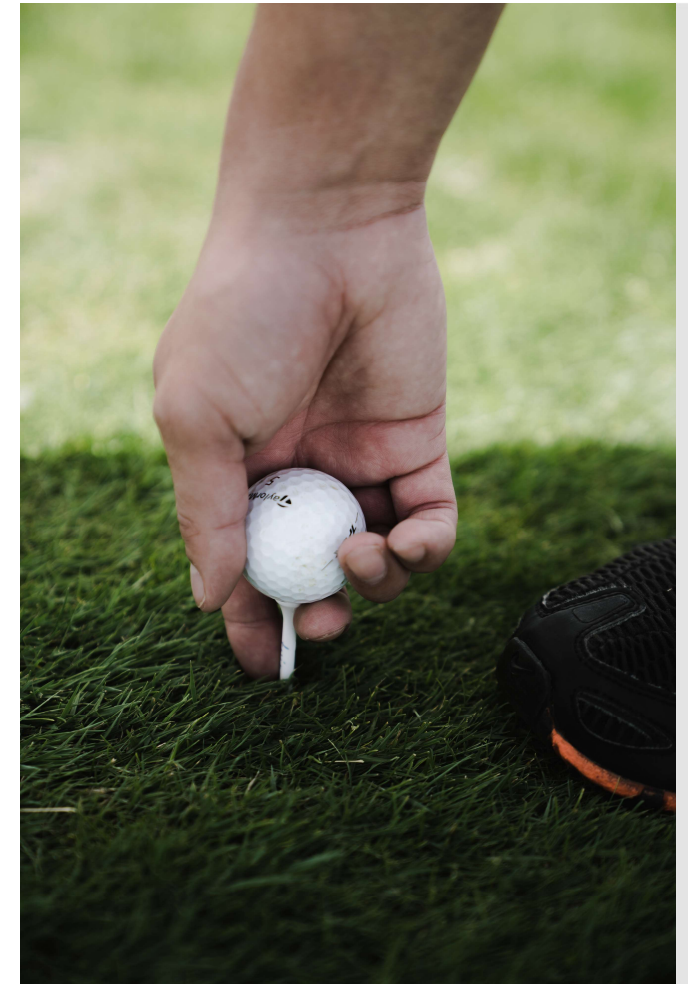
"On s'attendait à trouver des femmes dont l'âge variait en fonction des lignes de produits.

En fait on trouve des hommes en masse sur une catégorie de produits : les crèmes solaires.

En croisant les données on se rend compte que ces messieurs aiment aussi le golf.

Ils se rendent sur les greens quand il fait beau, donc ils ont besoin de crème solaire."

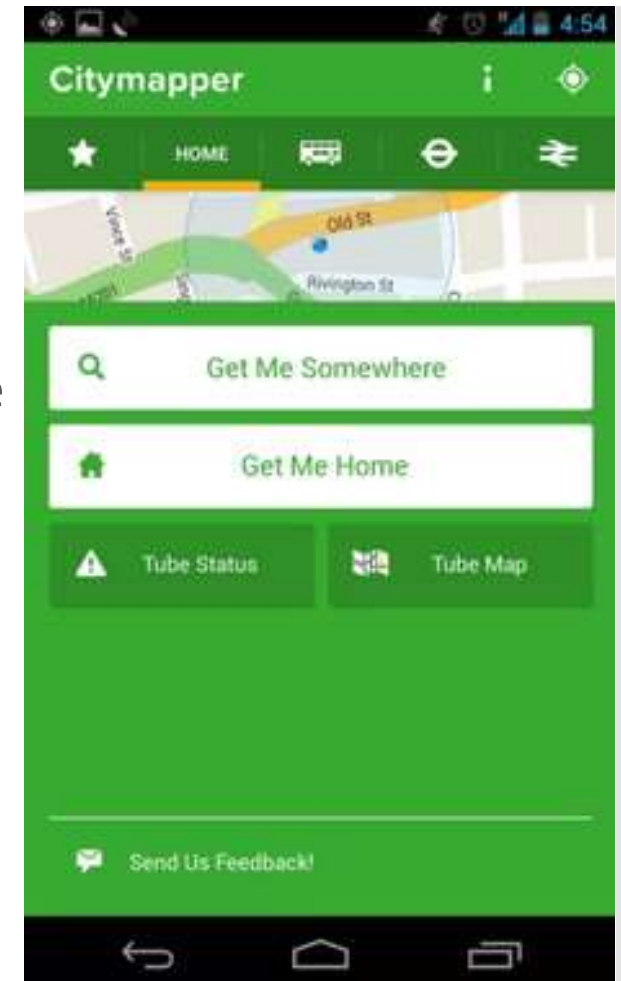
- l'entreprise **revoit son marketing** : plutôt que de mettre seulement une mère et ses enfants dans ses publicités pour les crèmes ...
- ... elle ajoute le papa !



Transports

Analyse et développement du réseau

- Citymapper est une application mobile qui **agrège** les informations utiles pour se déplacer en bus, métro, vélo, taxi, VTC ou à pied
- depuis sa création en 2013, elle **compile les données** des itinéraires de ses usagers
- à sa création elle **identifie des lacunes** notamment dans le réseau de transports de Londres
- elle identifie les **zones géographiques** où il **manque** des liaisons, les heures de la journée où l'**offre** de transport est **insuffisante**



Transports

Analyse et développement du réseau

- en 2017, elle lance ses propres bus qui ont un grand succès
- la régie des transports de la capitale anglaise réagit en développant son offre aux mêmes endroits et mêmes horaires
- Citymapper étudie alors les mêmes opportunités dans les plus grandes cités du monde

