

i162 - Analyser et modéliser des données

Loré Godel

06/09/23

Analyser et modéliser des données



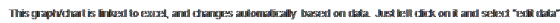
Définition :

- **Analyser des données** : Examiner/Interpréter les données accompagné de graphiques pour élaborer des réponses à des questions.
- **Modéliser des données** : Etablir la structure, les associations, les relations et les contraintes des données disponibles.

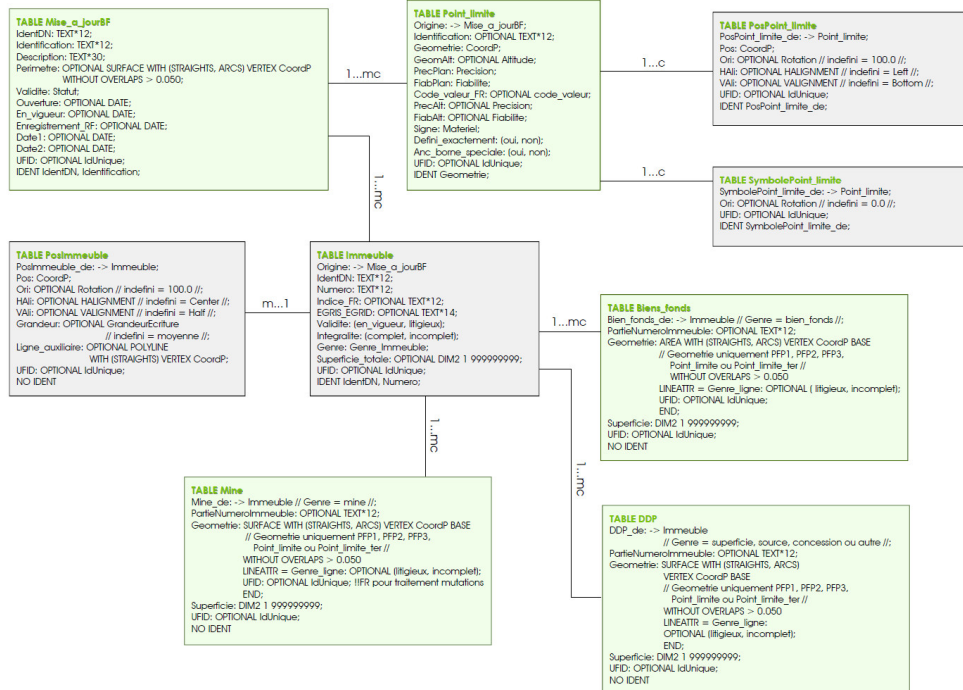
Exemples :

- **Analyse** de base de données ex.:

This slide shows the graphs of various metrics which he time, number of persons injured or killed by vehicle type



- **Modélisation** base de données ex.:



sources images :

1 (<https://www.imo-pe.fr/data.html>)

2 (<https://www.slideteam.net/blog/modeles-de-rapport-danalyse-ppt?lang=French>)

3 (<https://www.fr.ch/territoire-amenagement-et-constructions/cartes-plans-cadastre-et-geomatique/modeles-de>

donnees)

sources text :

modélisation (<https://www.talend.com/fr/resources/guide-modelisation->

donnees/#::~text=La%20mod%C3%A9lisation%20des%20donn%C3%A9es%20)

Différence entre données et informations

Les données ne sont PAS des informations !



Définition :

- **Données** : contiennent des chiffres, des énoncés et des caractères sous forme brute.
- **Information** : une nouvelle, un renseignement, une documentation sur quelque chose ou sur quelqu'un, portés à la connaissance de quelqu'un.

Exemples :

- **Données** :
 - Prénom
 - Nom
 - Age
 - Nationalité
 - Médicales
 - Scientifiques
 - Réseaux sociaux (likes, commentaires, partages)
- **Information** : Données traitées, organisées, structurées ou présentées dans un contexte donné afin de les rendre utiles -> "Il fait chaud aujourd'hui!"

source image :

1 (<https://waytolearnx.com/2018/08/difference-entre-donnee-et-information.html>)

sources text :

Données ([https://fr.wikipedia.org/wiki/Donn%C3%A9e_\(informatique\)](https://fr.wikipedia.org/wiki/Donn%C3%A9e_(informatique)))

Stocks de données



CD,HDD,SSD,USB... !



Ex. : Dropbox, Google Drive, et Amazon S3



Data center

Exemples :

- Stockage de médias (musique, films et photos numériques)
- Stockage de fichiers personnels (enregistrer un document word sur son pc pour qu'il reste le même jusqu'à la prochaine modification)
- Données de télémétrie (Les entreprises technologiques collectent des données de télémétrie à partir de leurs produits et services, telles que l'utilisation des applications, les performances des appareils, etc., pour améliorer leurs produits.)



- précis:
Stockage de données avec une grande attention aux détails et à la précision des informations.
- exhaustif:
Inclut un large éventail d'informations et de détails sur un sujet particulier.
- redondant:
Stockage de données où les mêmes informations sont répétées à plusieurs endroits
- incohérent:
Stockage de données où les informations ne sont pas cohérentes ou ne suivent pas un modèle de données défini

Exemple:

- précis : Carte de crédit.
- exhaustif :

| ID Produit | Nom du Produit | Catégorie | Marque | Poids (g) | Pri (€) |
|---------------|----------------|----------------------|------------------|--------------|------------|
| 1 | Pommes | Fruits | Ferme X | 150 | 1. |
| 2 | Lait | Produits laitiers | Marque A | 1000 | 0. |
| 3 | Pain de blé | Boulangerie | Boulangerie Y | 400 | 2. |
| 4 | Shampooing | Soins personnels | Marque B | 250 | 4. |

chaque produit est décrit avec une grande quantité d’informations détaillées.

- redondant :

| ID Client | Nom | Adresse | Ville |
|-----------|-------|-------------------|-----------|
| 1 | Alice | 123 Rue de Paris | Paris |
| 2 | Bob | 456 Avenue du Sud | Lyon |
| 3 | Carol | 789 Rue du Nord | Marseille |
| 4 | Dave | 123 Rue de Paris | Paris |

“123 Rue de Paris” se répète

- incohérent :

| Nom | Âge | Sexe |
|-------|---------|-------|
| Alice | 25 | Femme |
| Bob | Aucun | Homme |
| Carol | 32 | Femme |
| Dave | Inconnu | Homme |

"Aucun" et "Inconnu" ne sont pas des chiffres et vont poser problème lors de l'analyse et du traitement des données.

sources images :

1 (<https://quizizz.com/admin/quiz/5e3b6f663bd6ad001bd2679a/storage-media-questions-from-students-cie-ict-igcse-unit-3>)

2 (<https://medium.com/meet-lima/whats-the-cloud-what-s-a-private-cloud-7fcdd8b535d1>)

3 (<https://home.cern/fr/news/news/computing/cern-data-centre-passes-100-petabytes>)

sources texte:

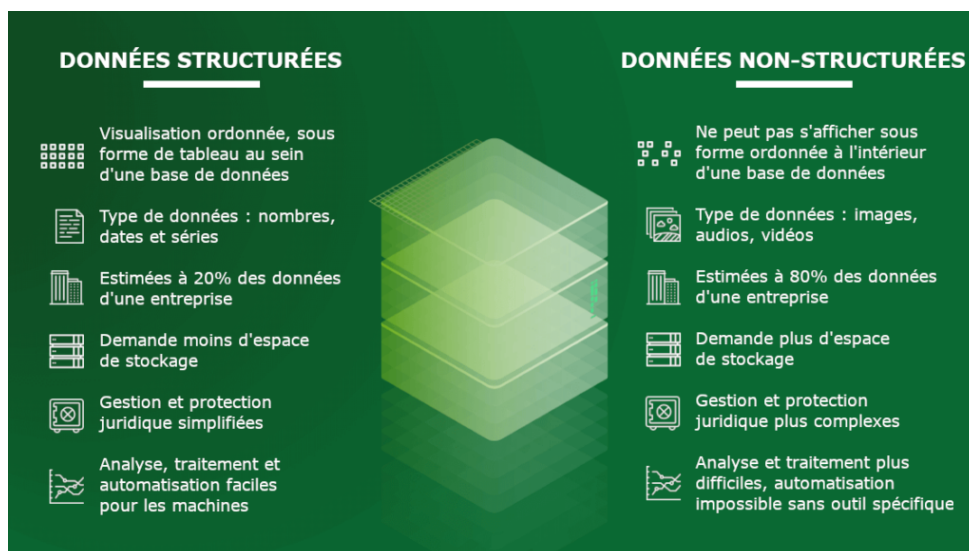
1

(<https://fr.differbetween.com/article/what-is-the-difference-between-data-redundancy-and-data-inconsistencyd>)

prompt ChatGPT:

"stock de données exhaustif exemple" "peux-tu me donner un exemple de stock de données redondant ?" "peux-tu me donner un exemple de stock de données incohérent ?"

Stocks de données structurés ou non structurés



Définition :

- **Données structurées:** Organisées de manière très formelle et tabulaire, elles sont généralement stockées dans des bases de données relationnelles ou des feuilles de calcul.
- **Données non structurées:** Sans format fixe ou de structure préalablement définie, elles peuvent prendre la forme de texte brut, de médias tels que des images ou des vidéos.

Exemple :

- **structuré:**

- carte d'identité
- factures
- tableaux de bord financiers
- bases de données de ressources humaines

- **non structuré:**

- signature sur la carte d'identité
- Articles de blog
- vidéos YouTube
- images Instagram
- enregistrements audio de conversations
- courriels

source image :

1 (<https://www.group-edt.fr/les-differents-types-de-donnees/>)

Qu'est-ce qu'un format pour une donnée en informatique ?

L'extension d'un fichier n'est PAS le format d'un fichier

Différence : l'extension de fichier indique le type de fichier, tandis que le format de fichier spécifie la manière dont les données sont organisées et encodées à l'intérieur du fichier. **



Définition : Le format de données = format de fichier, est une spécification qui définit la structure, l'encodage et l'organisation des données à l'intérieur du fichier. Il est essentiel pour l'interopérabilité entre les logiciels et les systèmes, ainsi que pour garantir que les données sont lues et interprétées correctement.

Exemple :

- JPEG pour les images
- MP3 pour l'audio
- JSON pour l'échange de données
- CSV pour les données tabulaires

Il existe beaucoup de formats...

source image :

1 (<http://editorialisation.org/ediwiki/index.php/Accueil>)

Code ou codage d'une donnée.

| Aspect | Code | Codage |
|----------|------------------------------------|---|
| Nature | Instructions | Transformation des informations |
| Language | Langage de programmation | Format ou représentation |
| But | Indiquer à l'ordinateur quoi faire | Rendre les données compréhensibles par l'ordinateur |

Définition:

- **code:** mots-clés/ symboles/chiffres et structures logiques qui permettent à toute sorte de dispositifs électroniques de comprendre les tâches à accomplir et de les exécuter.
 - Dispositifs électroniques capables de comprendre/d'exécuter du code:
 - Ordinateurs personnels (PC et Mac)
 - Serveurs
 - Superordinateurs
 - Automates et dispositifs embarqués
 - Calculatrices
 - IoT (Internet des objets)
- **codage:** Permet de représenter l'information de manière à ce qu'elle puisse être traitée, stockée, transmise et interprétée par des systèmes

informatiques.

Les données sont généralement stockées sous forme de séquences de bits (0 et 1) dans les systèmes informatiques.

Exemple:

- **Code:**

- Python :

```
# Ceci est un commentaire en Python
nom = "Alice"
print("Bonjour, " + nom + "!")
```

affiche "Bonjour, Alice!" à l'écran.

- Fortran (Formula Translation) :

```
program bonjour
  print *, "Bonjour, Monde!"
end program bonjour
```

- COBOL (Common Business-Oriented Language):

```
IDENTIFICATION DIVISION.
PROGRAM-ID. BonjourMonde.

PROCEDURE DIVISION.
  DISPLAY 'Bonjour, Monde!'.
```

- HTML :

```
<!DOCTYPE html>
<html>
<head>
  <title>Exemple de page web</title>
</head>
<body>
  <h1>Bienvenue sur ma page web</h1>
  <p>Ceci est un paragraphe de texte.</p>
</body>
</html>
```

- **Codage:**

- L'encodage des caractères textuels en binaire (comme UTF-8 pour les caractères Unicode)
 - la compression d'images JPEG
 - la conversion d'audio en MP3

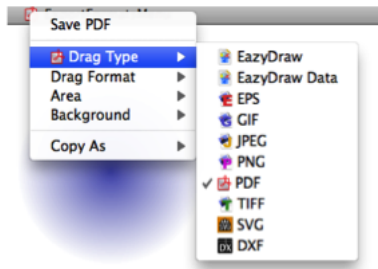
| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 |
| 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 |
| 1 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 |
| 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 |
| 1 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 |
| 1 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 1 |
| 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 |
| 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

Codage d'une image.

Source image :

5 (https://www.loutrel.fr/wikisn/doku.php?id=les_exposes:image_format_d_une_image_bmp)

Différents formats de données



Chaque format de données est adapté à des besoins particuliers en termes de stockage, de transmission et de traitement. Voici quelques-uns des formats de données les plus courants :

| Catégorie | Format | Description |
|-------------------------------|-----------|---|
| Documents | PDF | Portable Document Format, utilisé pour les documents avec mise en page fixe. |
| | DOCX | Format de fichier de document Microsoft Word. |
| | ODT | Open Document Text, format ouvert de traitement de texte. |
| Base de données | SQL | Structured Query Language, utilisé pour interagir avec les bases de données relationnelles. |
| | NoSQL | Divers formats de base de données non relationnelles, tels que MongoDB, Cassandra, etc. |
| | Shapefile | Format de données géographiques vectorielles largement utilisé. |
| Code source | C | Langage de programmation C. |
| | Java | Langage de programmation Java. |
| | Python | Langage de programmation Python. |
| | HTML/CSS | Langages de balisage et de style pour la création de pages web. |
| Compression de données | ZIP | Format de compression de fichiers. |

| Catégorie | Format | Description |
|-----------------------------|--------|---|
| | RAR | Format de compression de fichiers. |
| | 7z | Format de compression de fichiers très efficace. |
| Échange de données | SOAP | Simple Object Access Protocol, utilisé pour l'échange de messages entre applications. |
| | REST | Representational State Transfer, un style d'architecture pour les services web. |
| Fichiers exécutables | EXE | Fichiers exécutables sur les systèmes Windows. |
| | APP | Fichiers exécutables sur les systèmes macOS. |

Source image :

1 (<https://www.eazdraw.com/fileFormats.htm>)

