



digital lifelong learning

---

# **CERTIFICAT EXÉCUTIF**

## **Analyse de données avec**

## **Excel, SPSS, IA/Python et PowerBI**

---

### **Module 1 : Introduction à**

### **l'analyse de données**

# Plan

- Définition et Importance de la Data Analytics
- Cycle de Vie de l'Analyse de Données
- Principaux Outils de Data Analytics
- Conclusion

# Définition de l'analyse de données

- L'analyse des données est un ensemble de techniques pour découvrir la structure, éventuellement compliquée, d'un tableau de nombres à plusieurs dimensions et la traduire en une structure plus simple qui la résume au mieux. Cette structure peut, le plus souvent, être représentée graphiquement. (J-P. Fénelon).

Jean-Paul Fénelon est un chercheur connu pour ses contributions à l'analyse des données, particulièrement dans les années 1980. Son ouvrage "Qu'est-ce que l'Analyse des Données?" publié en 1981 est une référence importante dans ce domaine.

# Définition de la Data Analytics

- L'analyse des données n'est pas une méthode unique mais un ensemble de méthodes et d'outils utilisés pour extraire des informations significatives à partir de données brutes.
- Ces techniques sont employées pour identifier les motifs, les relations et les tendances cachés dans les données.
- Les données peuvent être complexes, avec plusieurs dimensions et variables.
- Les données analysées sont souvent organisées sous forme de tableaux ou de matrices, avec des lignes et des colonnes représentant différentes dimensions et variables.

# Définition de la Data Analytics

- L'objectif de l'analyse est de simplifier cette complexité en une structure plus compréhensible.
- Cette simplification permet de résumer les données de manière à ce qu'elles soient plus faciles à interpréter.
- Souvent, la structure simplifiée est représentée visuellement, par exemple à l'aide de graphiques, de diagrammes ou de tableaux de bord.
- Les représentations graphiques permettent de visualiser les données de manière intuitive, facilitant ainsi la compréhension des informations clés.

# Définition de la Data Analytics

- La Data Analytics est un processus complexe qui se compose de plusieurs étapes clés, chacune étant essentielle pour transformer des données brutes en informations exploitables.
- La Data Analytics est le processus de **collecte, d'épuration, d'analyse** et de **visualisation** des données pour prendre des décisions éclairées.

# Importance de la Data Analytics

Elle est essentielle au quotidien pour :

**Améliorer l'efficacité opérationnelle** : consiste à optimiser les processus internes d'une organisation pour réduire les coûts, augmenter la productivité et améliorer la qualité des produits ou services.

- ✓ Automatisation des Tâches Répétitives
- ✓ Gestion de la Chaîne d'Approvisionnement
- ✓ Maintenance Prédictive

**Prendre des décisions stratégiques** : sont des décisions à long terme qui définissent la direction et les objectifs d'une organisation. La Data Analytics fournit des insights basés sur des données pour prendre ces décisions de manière informée.

- ✓ Analyse de Marché
- ✓ Prévisions Financières
- ✓ Gestion des Risques

# Importance de la Data Analytics

**Comprendre les tendances et les comportements** : consiste à analyser les données pour identifier les motifs, les préférences et les habitudes des clients ou des utilisateurs.

- ✓ Segmentation de la Clientèle
- ✓ Personnalisation des Offres
- ✓ Analyse des Sentiments

**Optimiser les processus et les ressources** : consiste à utiliser les données pour améliorer l'utilisation des ressources disponibles et à rendre les processus plus efficaces.

- ✓ Gestion des Ressources Humaines
- ✓ Gestion des Stocks
- ✓ Planification de la Production

# Cas d'utilisation au quotidien

Exemples d'utilisation de la Data Analytics:

## **Analyse des ventes pour optimiser les stocks**

- Prévisions de la Demande
- Optimisation des Niveaux de Stock
- Réapprovisionnement Automatisé

## **Suivi des performances des campagnes marketing**

- Analyse des Taux de Conversion
- Segmentation des Clients
- Optimisation des Canaux

## **Analyse des comportements des clients pour personnaliser les offres**

- Recommandations de Produits
- Personnalisation des Contenus
- Programmes de Fidélité

# Méthodes d'Analyse de Données Selon Type de Mesure

## Selon type de mesure

Mesure  
nominale

Mesure  
ordinale

Mesure  
métrique

# Méthodes d'Analyse de Données Selon Type de Mesure

## Mesure Nominale :

Variables catégorielles sans ordre intrinsèque. On utilise des chiffres sans aucune relation d'ordre, ni de distance, ni d'origine:

Femme (1)

Homme (2)

Exemple :

- ✓ Sexe: Femme , Homme
- ✓ Situation matrimoniale: marié, célibataire...

# Méthodes d'Analyse de Données Selon Type de Mesure

## **Mesure Ordinale :**

Variables catégorielles avec un ordre. Les chiffre qui identifient la relation d'ordre entre les propriétés d'objet sans aucune relation de distance.

## **Exemple :**

- ✓ Classement des goûts des clients selon un critère
- ✓ classe d'âge (15- 25), ( 26- 35)....

# Méthodes d'Analyse de Données Selon Type de Mesure

## Mesure Métrique :

Variables quantitatives dont les valeurs ont une relation d'ordre et de distance.

### Exemple :

- ✓ Le nombre de points de vente d'une marque
- ✓ La valeur d'investissement dans une ville

# Cycle de Vie de l'Analyse de Données

Le cycle de vie de l'analyse de données comprend les étapes suivantes :

1. Collecte
2. Épuration
3. Analyse
4. Visualisation

# 1. Collecte des Données

La collecte de données est la première étape du cycle de vie de l'analyse de données. Elle peut être réalisée via:

**Sources internes (bases de données, systèmes ERP)** : sont les systèmes et bases de données propres à une organisation. Ces sources contiennent des données générées et stockées par l'entreprise elle-même.

**Sources externes (réseaux sociaux, études de marché)** : sont des données obtenues de l'extérieur de l'organisation. Ces données peuvent fournir des informations complémentaires et élargir la perspective analytique.

**Outils de collecte (formulaires en ligne, capteurs IoT)** : sont des instruments et des technologies utilisés pour rassembler des données de manière systématique et efficace.

## 2. Épuration des Données

L'épuration des données est cruciale pour assurer leur qualité. Elle comprend :

La suppression des doublons

**Problèmes Causés par les Doublons :**

- ✓ Biais dans les Analyses
- ✓ Augmentation des Coûts
- ✓ Confusion et Mauvaises Décisions

La correction des erreurs

Le traitement des valeurs manquantes

La normalisation des formats

## 2. Épuration des Données

L'épuration des données est cruciale pour assurer leur qualité. Elle comprend :

La suppression des doublons

La correction des erreurs

### **Types d'Erreurs**

- ✓ Fautes de Saisie
- ✓ Erreurs de Format
- ✓ Données Incorrectes

Le traitement des valeurs manquantes

La normalisation des formats

## 2. Épuration des Données

L'épuration des données est cruciale pour assurer leur qualité. Elle comprend :

La suppression des doublons

La correction des erreurs

Le traitement des valeurs manquantes

### **Problèmes Causés par les Valeurs Manquantes**

- ✓ Analyses Biaisées
- ✓ Réduction de la Taille de l'Échantillon
- ✓ Complications dans les Modèles

La normalisation des formats

## 2. Épuration des Données

L'épuration des données est cruciale pour assurer leur qualité. Elle comprend :

La suppression des doublons

La correction des erreurs

Le traitement des valeurs manquantes

La normalisation des formats

### **Problèmes Causés par les Formats Incohérents**

- ✓ Difficultés d'Intégration
- ✓ Erreurs d'Analyse
- ✓ Complexité Accrue

### 3. Analyse des Données

L'analyse des données permet de découvrir des insights et de prendre des décisions. Elle inclut :

**Les statistiques descriptives (moyenne, médiane, mode)** : résument et décrivent les caractéristiques principales d'un ensemble de données. Elles fournissent un aperçu rapide et compréhensible des données sans tirer de conclusions au-delà de ce que montrent les données actuelles.

- Environ 70% des entreprises utilisent des statistiques descriptives pour leurs rapports réguliers.
- Les mesures de tendance centrale et de dispersion sont utilisées dans 95% des analyses de performance des ventes.

**Les analyses prédictives**

**Les analyses exploratoires**

### 3. Analyse des Données

**Les analyses prédictives (modèles de régression) :** Les analyses prédictives utilisent des techniques statistiques et des algorithmes de machine learning pour prévoir des événements futurs en se basant sur des données historiques. Elles permettent aux entreprises de prendre des décisions proactives.

- Environ 37% des entreprises utilisent des analyses prédictives pour prendre des décisions stratégiques.

### 3. Analyse des Données

**Les analyses exploratoires (corrélations, clusters) :** Les analyses exploratoires visent à découvrir des motifs, des relations et des structures cachées dans les données sans hypothèse préalable. Elles sont souvent utilisées pour formuler de nouvelles questions de recherche et orienter les analyses futures.

- Les entreprises utilisant l'analyse des clusters pour la segmentation de la clientèle voient une augmentation de 20% à 25% de l'efficacité de leurs campagnes marketing.

## 4. Visualisation des Données

La visualisation des données aide à comprendre et communiquer les résultats de l'analyse. Les outils couramment utilisés sont :

- Graphiques
- Tableaux de bord interactifs
- Cartes géographiques

## 4. Visualisation des Données

- **Graphiques** : sont des représentations visuelles des données qui permettent de voir rapidement les tendances, les variations et les relations entre les variables. Ils sont essentiels pour présenter les résultats de manière claire et concise.
- **Tableaux de bord interactifs**
- **Cartes géographiques**

## 4. Visualisation des Données

- **Graphiques**

- **Tableaux de bord interactifs** : Les tableaux de bord interactifs sont des interfaces visuelles qui permettent aux utilisateurs de suivre, analyser et visualiser les données en temps réel. Ils combinent plusieurs graphiques, indicateurs et filtres, offrant une vue d'ensemble complète et interactive des données.

- **Cartes géographiques**

## 4. Visualisation des Données

- **Graphiques**
- **Tableaux de bord interactifs**
- **Cartes géographiques** : Les cartes géographiques sont des visualisations spatiales qui permettent de représenter des données en fonction de leur localisation géographique. Elles sont particulièrement utiles pour analyser des données à dimension géographique, comme les ventes par région ou les incidents par emplacement.

# Principaux Outils d'Analyse de Données

L'analyse de données est facilitée par divers outils qui permettent de manipuler, visualiser et interpréter les données. Voici une description détaillée de quatre outils couramment utilisés : Excel, Python, SPSS, et Microsoft Power BI.

# Principaux Outils d'Analyse de Données



**Microsoft Excel** : est un tableur largement utilisé pour l'analyse de données, offrant des fonctionnalités puissantes pour les calculs, les graphes et la visualisation des données.

## Caractéristiques Principales :

- Fonctions et Formules
- Tableaux Croisés Dynamiques
- Graphiques et Visualisations
- Macros et VBA
- Outils d'Analyse de Données

Environ 81 % des entreprises utilisent Excel pour des tâches analytiques et des rapports financiers.  
[\(AlloExcel\)](#) [\(atelier-mosesu\)](#).

# Principaux Outils d'Analyse de Données



**Python** : est un langage de programmation polyvalent et populaire pour l'analyse de données, offrant des bibliothèques puissantes pour le traitement des données, les statistiques, et le machine learning.

## Caractéristiques Principales :

- **Pandas**
- **NumPy**
- **Matplotlib et Seaborn**
- **Scikit-learn**

Environ 60% des data scientists utilisent Python pour leurs projets d'analyse de données.  
[\(Geekflare\)](#) [\(Developpez Python\)](#).

# Principaux Outils d'Analyse de Données



**SPSS (Statistical Package for the Social Sciences)** : est un logiciel d'analyse statistique utilisé principalement dans les sciences sociales et la recherche.

## Caractéristiques Principales :

- **Interface Utilisateur Graphique**
- **Analyses Statistiques Avancées**
- **Manipulation des Données**
- **Rapports et Visualisations**
- **Extensions et Scripts**

Environ 84% des chercheurs en sciences sociales utilisent SPSS pour l'analyse de leurs données. ([Techno-Science.net](http://Techno-Science.net)).

# Principaux Outils d'Analyse de Données



## Power BI

*Microsoft PowerBI* : est un outil de business intelligence qui permet de créer des tableaux de bord interactifs et des rapports visuels à partir de diverses sources de données.

### Caractéristiques Principales :

- Tableaux de Bord Interactifs
- Intégration de Données
- Visualisations Riches
- Fonctionnalités de Drag-and-Drop
- Collaborations et Partage

Les entreprises utilisant Power BI rapportent une amélioration de 25% de leur prise de décision basée sur les données. ([Power BI](#)) ([Hub Collab](#)).

# Principaux Outils d'Analyse de Données

Chacun de ces outils - Excel, Python, SPSS, et Microsoft Power BI - offre des capacités uniques pour l'analyse des données.

- ✓ Excel est idéal pour les tâches analytiques quotidiennes et la visualisation de base,
  - ✓ Python excelle dans les analyses avancées et le machine learning,
  - ✓ SPSS est spécialisé dans les analyses statistiques complexes,
  - ✓ et Microsoft Power BI est parfait pour la création de tableaux de bord interactifs et la visualisation en temps réel.
- ➔ Ensemble, ces outils permettent une analyse des données complète et efficace, adaptée à divers besoins professionnels et académiques.

# Conclusion

---

La Data Analytics est un outil puissant qui aide à transformer des données brutes en informations exploitables. Elle joue un rôle crucial dans la prise de décisions, l'amélioration des processus et la compréhension des tendances.