

02. Traitement de données – Découverte



Description de l'activité

Dans cette activité, on découvre la thématique à travers une vidéo et un QCM. Ensuite on effectue de façon simple et graphique des tris sur un tableau de données.

Objectifs pédagogiques ou compétences

Objectifs généraux	Objectifs intermédiaires	Compétences
Notions de cours	- Connaître la notion de données et les mots-clés en lien. - Comprendre l'importance des données et de leur traitement automatique	- Traiter des données de façon simple (filtrer, trier)

Matériel et outils

- 1 poste par élève

Tags

#données structurées ; #données ; #tableur ; #filtrer ; #trier ; #descripteurs; #stocker

Déroulé de l'activité

Introduction : (~30 minutes)

- **Présenter les objectifs de la séance (contenu théorique et productions attendues) (2-3 minutes)**
- **Introduire la thématique des données : (~30 minutes)**

Pour lancer la thématique des données, on propose aux élèves une première activité de classement, à faire en petits groupes de 4 ou 5 élèves. L'objectif est de mieux leur faire comprendre le besoin de classer les informations pour automatiser le travail.

- Dans un premier temps, les élèves définissent 5 critères, et remplissent un tableau sans noter le nom de la personne.
- Ensuite, chaque groupe échange son tableau avec un autre, chacun devant retrouver qui correspond aux descriptions.
- On vérifie si les propositions sont correctes (on peut à ce moment réunir les 2 groupes et les laisser discuter plutôt que de tout vérifier ensemble), puis on discute de l'efficacité des critères choisis.
- L'enseignant.e peut demander aux élèves si les critères proposés permettent de cibler efficacement une personne dans un plus grand groupe (à l'échelle du lycée, de la ville, de la planète, ...)

Ensuite, les élèves retournent avec les mêmes groupes et affinent les critères pour permettre d'identifier la personne correspondant aux mieux à chaque cas spécial. Cette partie a pour objectif de rappeler que les types de données à cibler et traiter sont nombreuses, en fonction de multiples besoins. Les élèves peuvent présenter lors de la correction la personne du groupe la plus à même d'être concernée.

Enfin, on conclut sur un petit point sur le traitement automatisé des données. Les élèves répondent avec leur groupe ou en binôme, sans effectuer de recherches, le but étant de faire le point sur leurs connaissances.

Remarque : Avec moins de temps, on peut demander aux élèves s'ils savent ce que sont les données, comment on en crée, qui les récupère et pourquoi (usages avantageant l'utilisateur, mais aussi usages qui peuvent avoir un impact négatif).

En conclusion, vous trouverez des pistes de réponses.

Déroulé – Exercices multiples (~1h30)

- **Présentation du site (<5 minutes)**

L'enseignant.e présente le site <https://parcours.algorea.org/contents/4707-4702-1067253748629066205-653650670442840123-1347056385813169871/> et explique rapidement le fonctionnement du site :

- Lecture du texte explicatif lié à chaque exercice
- Espace de manipulation (ici, des listes déroulantes à fonction de filtres)

- Espace de renseignement des réponses ou de simple validation
- Le système à plusieurs étoiles.

En cas d'erreur, le feedback les guide pour se corriger. Le feedback peut également apporter des informations complémentaires. L'enseignant.e circule et répond aux questions.

Remarque : on peut préférer faire les niveaux à 2 étoiles de chaque exercice dans un premier temps puis revenir pour les niveaux avancés, ou bien faire les différents niveaux à la suite avant de passer à l'exercice suivant. L'enseignant.e peut noter les élèves sur leurs performances sur les niveaux les plus avancés.

- **Vidéo d'introduction (20 minutes)**

Les élèves visionnent la vidéo puis répondent au QCM.

- **Je choisis - (~15 minutes)**

Cette partie consiste à utiliser une ou des listes déroulantes permettant de trier une liste de données afin d'obtenir la liste cible.

- **Quiz (~10 minutes)**

Les élèves réalisent un quizz de trois questions en glissant-déposant des mots dans un texte. Cela permettant de formaliser les connaissances précédentes et introduire le vocabulaire.

- **Non, pas ceux-là ! – Attention au prix ! – Dans un certain ordre ! (~30 minutes)**

Les élèves font les 3 activités.

- **Quiz (~15 minutes)**

On finit avec un dernier quiz.

Le dernier exercice n'est pas inclus mais peut l'être si la classe est à l'aise avec l'activité.

Conclusion (15 minutes)

- **Bilan de la séance : (5 minutes)**

Pour clôturer la séance, on peut revenir sur les principales difficultés rencontrées pendant l'activité. On revoit également les principaux mots-clés : Objet, Descripteur, Type, Collection, Valeur, ...

Éventuellement, il est possible de finir sur un court échange autour :

- **L'importance de l'automatisation des données (5 minutes)**

Si besoin, on peut expliquer davantage en quoi consiste le traitement automatique des données, discuter des avantages, éventuellement des inconvénients ou réserves qu'on peut émettre.

Proposition :

Le traitement automatique des données (TAD), ou traitement automatisé de l'information, fait référence au processus par lequel des machines (ordinateurs et logiciels spécialisés), sont utilisées pour collecter, stocker, manipuler, analyser et interpréter des données numériques. Ce processus est devenu essentiel dans de nombreux domaines : affaires, recherche scientifique, médecine, industrie, entreprises, sport, loisirs, ...

Avantages :

- **Vitesse et efficacité :** Les machines peuvent traiter d'énormes quantités de données beaucoup plus rapidement que les êtres humains, ce qui permet d'obtenir des résultats en temps réel ou dans des délais très courts.
- **Précision :** Les machines sont moins sujettes aux erreurs humaines. Une fois correctement programmées, elles effectuent des calculs et des analyses de manière cohérente et précise.
- **Stockage et récupération faciles :** Les données numériques peuvent être stockées de manière organisée et facilement récupérées lorsque nécessaire, ce qui permet de gagner du temps et d'éviter la perte de données.
- **Analyse avancée :** Les logiciels de traitement de données peuvent effectuer des analyses complexes qui servent de base aux Intelligences Artificielles, telles que l'apprentissage automatique, la modélisation statistique et l'exploration de données, pour extraire des informations précieuses et des tendances à partir des données.
- **Automatisation des tâches répétitives :** Le TAD peut automatiser des tâches routinières, et ainsi, nous évite d'avoir à le faire nous-même (dans l'activité, les élèves auraient dû faire des allers-retours dans les 2 fichiers pour les 37 993 communes).
- **Accessibilité :** Les données peuvent être accessibles à partir de n'importe quel endroit avec une connexion Internet, ce qui facilite le travail à distance et la collaboration entre équipes dispersées géographiquement.

Inconvénients et réserves potentiels :

- **Dépendance technologique :** La dépendance à l'égard de la technologie peut entraîner des problèmes en cas de défaillance matérielle ou de pannes informatiques, et bloquer de plus en plus de secteurs. Avec l'émergence de machines autonomes et des objets connectés (transports en commun, maisons avec volets, portes et autres objets connectés par exemple), mais aussi l'usage des données dans des métiers tels que ceux de la santé et l'administration, il ne s'agit pas seulement de quelques entreprises bloquées pour 2 ou 3 heures, mais d'un réel impact sur nos vies.
- **Sécurité :** Les données numériques sont vulnérables aux menaces de sécurité (piratage informatique, logiciels malveillants, vol de données, ...).
- **Protection de la vie privée :** La collecte et le traitement automatisés des données soulèvent des questions concernant la vie privée et la confidentialité. Les organisations doivent être conformes aux réglementations

de protection des données, telles que le RGPD en Europe, pour éviter des problèmes juridiques. La question du réel respect des données se pose également ([article sur Amazon France](#)).

- **Biais algorithmique** : Les systèmes automatisés peuvent être biaisés si les données utilisées pour les former sont elles-mêmes biaisées. Cela peut entraîner des discriminations injustes, en particulier dans les domaines tels que les prêts bancaires, la justice pénale et l'embauche.
- **Perte de compétences manuelles** : L'automatisation excessive peut entraîner une perte de compétences manuelles, car les employés peuvent devenir dépendants des systèmes automatisés et perdre la capacité de résoudre manuellement certains problèmes.
- **Déshumanisation** : Dans certaines industries, l'automatisation peut entraîner la suppression de postes de travail, ce qui peut avoir des implications sociales, économiques et psychologiques.

- **Les métiers en lien (10 minutes)**

On peut également évoquer les principaux métiers en lien avec l'analyse et le traitement des données. En fonction du temps, on peut demander aux élèves de faire des recherches, et éventuellement construire leur propre « top 3 » des métiers qui les intéressent dans ce domaine. Voici quelques exemples :

- **Analyste de données** : Analyse les données pour identifier des tendances, des modèles et des informations utiles pour la prise de décision.
- **Data Scientist** : Utilise des techniques avancées d'analyse de données, de machine Learning et de statistiques pour extraire des prédictions à partir des données.
- **Ingénieur en traitement des données** : Conçoit et développe des systèmes de gestion des données, des bases de données et des pipelines de traitement des données.
- **Analyste de business intelligence** : Collecte, organise et analyse des données pour aider les entreprises à prendre des décisions stratégiques et à surveiller leurs performances.
- **Ingénieur en apprentissage automatique** : Développe des modèles d'apprentissage automatique pour automatiser des tâches et créer des systèmes intelligents.
- **Analyste de données marketing** : Analyse les données marketing pour évaluer l'efficacité des campagnes, identifier les segments de clientèle et améliorer les stratégies marketing.
- **Analyste financier** : Utilise les données financières pour évaluer les performances de l'entreprise, prévoir les tendances économiques et recommander des stratégies d'investissement.
- **Data Engineer** : Gère l'infrastructure de données, construit et optimise les pipelines de données et assure la qualité des données pour les analyses ultérieures.
- **Architecte de données** : Conçoit l'architecture des systèmes de données, définit les flux de données et garantit la cohérence et l'intégrité des données.
- **Analyste en cybersécurité** : Analyse les données de sécurité pour détecter et prévenir les menaces, les attaques et les vulnérabilités dans les systèmes informatiques.
- **Data Analyst en santé** : Analyse les données médicales pour identifier des modèles de santé, améliorer les soins aux patients et prendre des décisions éclairées en matière de santé.
- **Analyste en sciences sociales** : Utilise les données pour étudier les comportements humains, les tendances sociales et les modèles sociétaux.

EVALUATION :
Les niveaux 2
et/ou 3 peuvent
être notés.

Introduction aux données structurées

Fiche activité - Correction

Introduction – Créer des données et les exploiter

● Données générales

Avec votre groupe, vous allez choisir **5 critères permettant**, à n'importe quelle autre personne de la classe (en dehors de votre groupe) de deviner qui est la personne correspondant aux données.

Exemples de critères :

- Date de naissance
- Ville de résidence
- Taille
- Couleurs de cheveux
- Sexe
- Vêtements et/ou bijoux
- Centres d'intérêts
- ...

● Données spécifiques

Maintenant, allons un peu plus loin ... Pour chaque cas, donnez une liste de plusieurs critères permettant de cibler au mieux leurs candidats idéaux, et ce à l'échelle de votre ville. Dans votre groupe, qui serait concerné ?

- Le club de football local cherche des personnes sportives pour relancer le club amateur des seniors. Ce club ne se retrouve que les jeudis de 18h à 20h, et organise ses matchs les samedis matin.

Pistes de critères : âge, proximité de la ville, sports pratiqués et fréquence, disponibilité sur les horaires mentionnés, ...

- Le dernier samedi soir du mois prochain, un quiz musical sur le thème du rap/hip-hop des années 2000-2010 a lieu. Les questions sont basées sur la reconnaissance des titres et artistes, mais aussi sur la reconnaissance des paroles.

Pistes de critères : styles de musiques écoutés et leur année de création, connaissance des titres et artistes, connaissance des paroles, gestion du stress, disponibilité le jour du quiz, ...

- Une entreprise d'e-commerce proposant des chaussures et accessoires en cuir végétal à destination des enfants (0-10 ans) souhaite mieux cibler ses clients pour augmenter ses ventes.

Pistes de critères : convictions (via leurs recherches web ou leurs profils sur les RS par exemple), âge, avoir des enfants ou petits-enfants, revenus, lieu d'habitation, ...

- Suite à un départ, une équipe d'eSport française (genre : MOBA) souhaite recruter leur prochain membre.

Pistes de critères : rang dans le jeu, années de pratique, prix ou tournois gagnés, âge, fréquence de jeu, connaissance des tactiques, gestion du stress, disponibilité, capacité à communiquer, capacité à gérer la frustration, ...

- **Traitement automatique des données :**

- Qu'est-ce qu'une donnée ?

Une donnée est une représentation symbolique d'informations sous une forme brute ou non traitée. Elle peut prendre diverses formes, telles que des nombres, des textes, des images, des sons, etc. Les données peuvent être collectées, stockées, et utilisées pour diverses finalités, notamment pour l'analyse, la prise de décision, la recherche, et bien d'autres domaines.

- Qui peut récupérer des données ? Comment ?

Les données peuvent être récupérées par divers acteurs, notamment :

- **Les individus :** Ils peuvent collecter des données personnelles à travers des enquêtes, des formulaires en ligne, ou en utilisant des dispositifs de suivi.
- **Les organisations :** Les entreprises, les gouvernements, et les organismes à but non lucratif collectent souvent des données pour des études de marché, des rapports statistiques, ou pour des services publics.
- **Les machines :** Les dispositifs automatisés, tels que les capteurs, les caméras, et les ordinateurs, peuvent recueillir des données en continu.

La collecte de données peut se faire par le biais de sondages, de questionnaires, de capteurs, de dispositifs de suivi, d'applications mobiles, de formulaires en ligne, et de nombreuses autres méthodes.

- Quels sont les avantages du traitement automatique des données ?

- **Rapidité :** Les machines peuvent traiter d'énormes volumes de données beaucoup plus rapidement que les humains.
- **Précision :** Les erreurs humaines sont réduites, ce qui garantit des résultats plus précis.
- **Analyse avancée :** À partir de très grandes quantités de données, les algorithmes peuvent détecter des tendances et des modèles complexes dans les données.
- **Automatisation :** Les machines / programmes peuvent effectuer la collecte et l'analyse des données (plus rapidement et en fonction des cas, plus efficacement), et ainsi nous permettre d'économiser du temps.
- **Prise de décision éclairée :** Le traitement automatique des données peut fournir des informations essentielles pour la prise de décisions stratégiques dans les domaines tels que les affaires, la médecine, la recherche scientifique, etc.

Introduction aux données structurées

Fiche activité

Introduction – Créer des données et les exploiter

- **Données générales**

Avec votre groupe, vous allez choisir **5 critères permettant**, à n'importe quelle autre personne de la classe (en dehors de votre groupe) de deviner qui est la personne correspondant aux données.

	Critère 1	Critère 2	Critère 3	Critère 4	Critère 5
Elève 1					
Elève 2					
Elève 3					
Elève 4					
Elève 5					

- **Données spécifiques**

Maintenant, allons un peu plus loin ... Pour chaque cas, donnez une liste de plusieurs critères permettant de cibler au mieux leurs candidats idéaux, et ce à l'échelle de votre ville. Dans votre groupe, qui serait concerné ?

- Le club de football local cherche des personnes sportives pour relancer le club amateur des seniors. Ce club ne se retrouve que les jeudis de 18h à 20h, et organise ses matchs les samedis matin.

.....

.....

.....

.....

.....

.....

- Le dernier samedi soir du mois prochain, un quiz musical sur le thème du rap/hip-hop des années 2000-2010 a lieu. Les questions sont basées sur la reconnaissance des titres et artistes, mais aussi sur la reconnaissance des paroles.

.....

.....

.....

.....

.....

.....

.....

- Une entreprise d'e-commerce proposant des chaussures et accessoires en cuir végétal à destination des enfants (0-10 ans) souhaite mieux cibler ses clients pour augmenter ses ventes.

.....

.....

.....

.....

.....

.....

.....

- Suite à un départ, une équipe d'eSport française (genre : MOBA) souhaite recruter leur prochain membre.

.....

.....

.....

.....

.....

.....

.....

- **Traitement automatique des données :**

- Qu'est-ce qu'une donnée ?

.....

- Qui peut récupérer des données ? Comment ?

Quels sont les avantages du traitement automatique des données ?