

The background of the slide is decorated with numerous paint splashes in various colors, including yellow, orange, green, and blue, creating a vibrant and artistic effect.

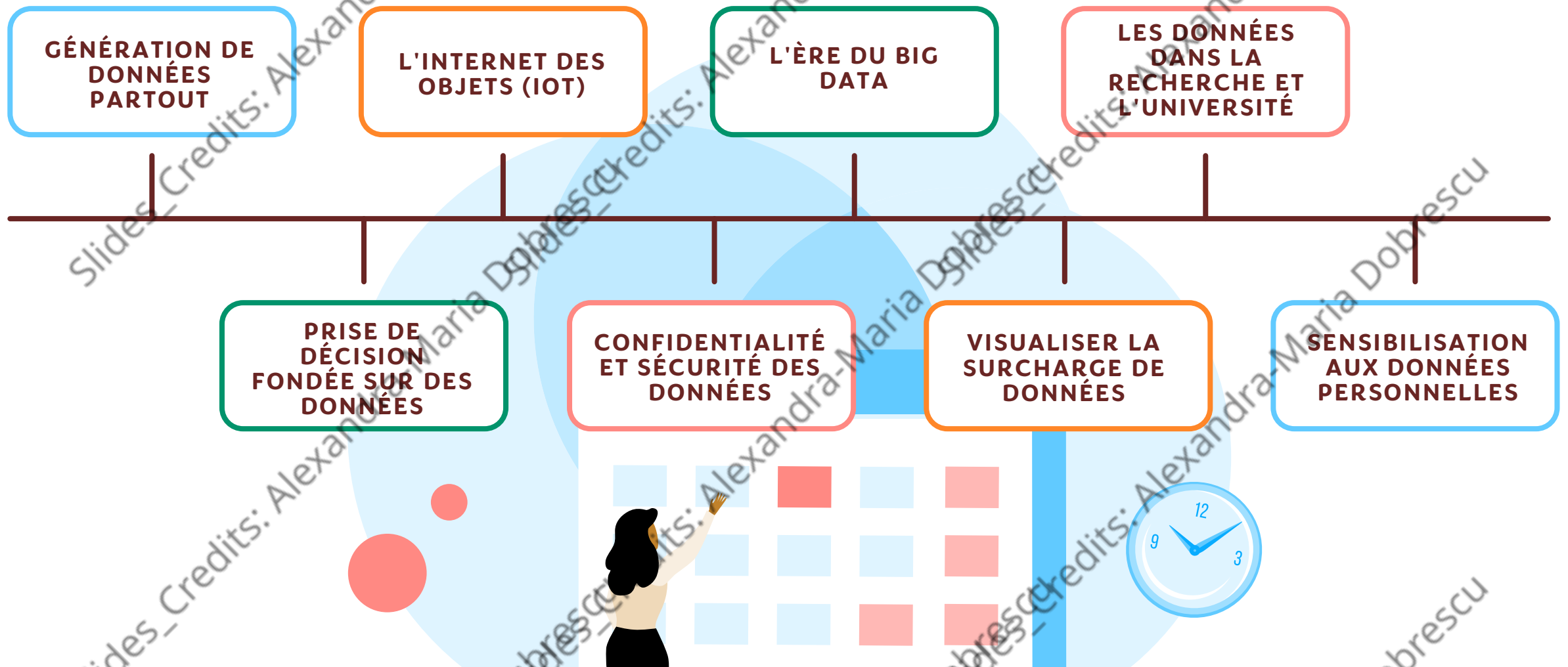
ANALYSE ET VISUALISATION DES DONNÉES

Alexandra-Maria DOBRESCU

| Cours 3 |

Introduction

- DÉFINIR L'ANALYSE DES DONNÉES
- DÉFINIR LA VISUALISATION DES DONNÉES
- APROFONDIR
- COMPARAISON
- POURQUOI L'ANALYSE DES DONNÉES EST IMPORTANTE DANS DIFFÉRENTS DOMAINES
- TYPES DE DONNÉES (STRUCTURÉES, NON STRUCTURÉES, SEMI-STRUCTURÉES)



Omniprésence des données

Les données dans la recherche et l'université

DANS LA RECHERCHE UNIVERSITAIRE, LES DONNÉES SONT LA BASE SUR LAQUELLE DE NOUVELLES CONNAISSANCES SONT CONSTRUITES. C'EST UNE RESSOURCE INDISPENSABLE POUR COMPRENDRE, EXPLORER ET DÉCOUVRIR DANS DIFFÉRENTS DOMAINES.

Méthodes de collection de données : Il peut s'agir d'enquêtes, d'expériences, d'observations, d'entretiens ou même de données obtenues à partir de sources et de bases de données existantes.

ImageNet: A Large-Scale Hierarchical Image Database

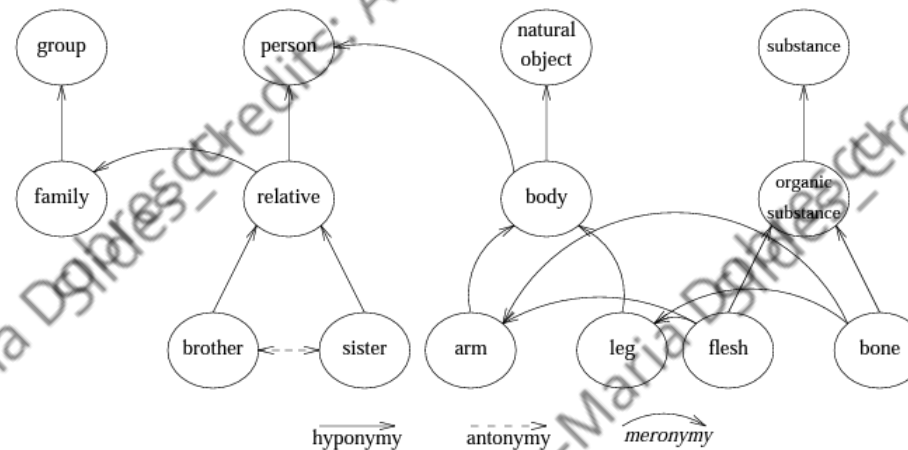
Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li and Li Fei-Fei
Dept. of Computer Science, Princeton University, USA
{jiadeng, wdong, rsocher, jial, li, feifeili}@cs.princeton.edu

- Fei-Fei Li, chercheur en intelligence artificielle, a commencé à travailler sur l'idée d'ImageNet en 2006.
- À une époque où la plupart des recherches sur l'IA se concentraient sur les modèles et les algorithmes, M. Li souhaitait élargir et améliorer les données disponibles pour entraîner les algorithmes d'IA.

Les données dans la recherche et l'université

- M. Li a pensé que la clé d'une meilleure performance des modèles pourrait être des ensembles de données plus importants reflétant la diversité du monde réel.
- Au cours de ses recherches sur les ensembles de données, Li a entendu parler du professeur Christiane Felbaum, co-développeur d'un ensemble de données appelé WordNet qui se composait de nombreux termes anglais organisés dans une structure ontologique [8].

Figure 2. Network representation of three semantic relations among an illustrative variety of lexical concepts

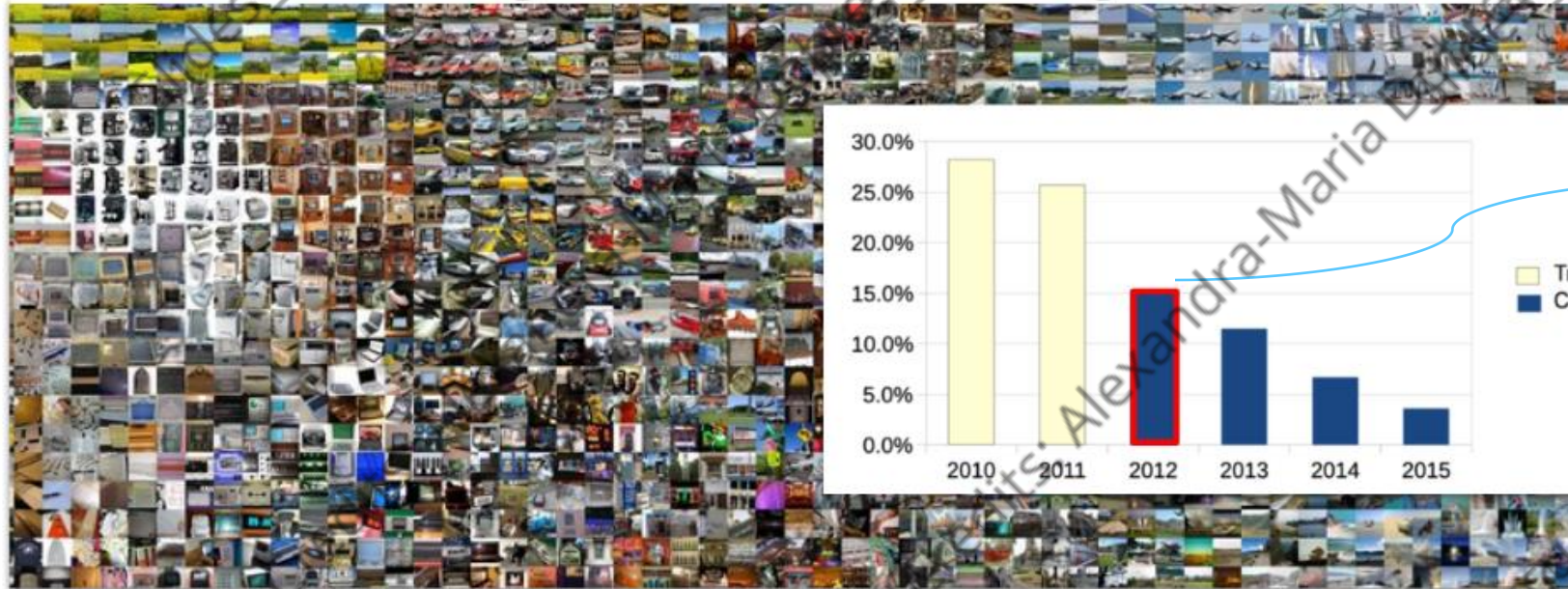


Les données dans la recherche et l'université

- L'idée derrière ImageNet est qu'une grande ontologie d'images - basée sur WordNet - pourrait être la clé du développement d'une recherche et d'une compréhension d'images avancées, basées sur le contenu.
- Deux ans plus tard, la première version d'ImageNet a été publiée avec 12 millions d'images structurées et étiquetées conformément à l'ontologie WordNet. Si une personne avait annoté une image/minute et n'avait rien fait d'autre pendant ces deux années (y compris dormir ou manger), cela aurait pris 22 ans et 10 mois.
- Pour y parvenir en moins de deux ans, M. Li s'est tourné vers Amazon Mechanical Turk, une plateforme de crowdsourcing où chacun peut engager des personnes du monde entier pour effectuer des tâches à moindre coût.
- L'équipe ImageNet a demandé aux "Turkers" de décider si une image représente ou non un mot donné (de l'ontologie WordNet). Plusieurs mesures ont été mises en œuvre pour garantir la précision de l'annotation, notamment l'attribution de plusieurs notes à chaque paire image-mot.

Les données dans la recherche et l'université

ImageNet Large Scale Visual Recognition Challenge



AlexNet

- Le 30 septembre 2012, un réseau neuronal convolutionnel (CNN) appelé AlexNet a obtenu une erreur de 15,3 % dans le top 5 du défi ImageNet 2012, soit plus de 10,8 points de pourcentage de moins que le deuxième.
- Selon The Economist, «soudain, les gens ont commencé à s'intéresser à la question, non seulement au sein de la communauté de l'IA, mais aussi dans l'ensemble du secteur technologique».

Les données dans la recherche et l'université



Prise de décision fondée sur les données

- Les décisions dans divers domaines sont de plus en plus fondées sur des données.
- L'analyse des données fournit des informations précieuses pour faire des choix éclairés.

Exemple : Les entreprises utilisent les données relatives aux ventes pour déterminer les stratégies de marketing et le développement des produits.

Exercice : Analyse des ventes en magasin.

Scénario : Vous êtes gérant d'un magasin et vous souhaitez analyser les données de vente mensuelles de votre magasin afin de prendre des décisions éclairées.

Prise de décision fondée sur les données

Exercice : Analyse des ventes en magasin.

Scénario : Vous êtes gérant d'un magasin et vous souhaitez analyser les données de vente mensuelles de votre magasin afin de prendre des décisions éclairées.

1) Collection de données

Un exemple simplifié de données sur les ventes mensuelles d'un magasin au cours d'une année:

D1 = {Janvier: 15k£, Février: 16.5k£, Mars: 18.2k£, Avril: 14.8k£, Mai: 17.3k£, Juin: 19.6k£, Juillet: 20.5k£, Août: 21.2k£, Septembre: 18.9k£, Octobre: 16.8k£, Novembre: 15.7k£, Décembre: 22k£}

2) Analyse des données

- a. **Calculez les indicateurs suivants:** Ventes annuelles totales (somme de toutes les ventes mensuelles), ventes mensuelles moyennes, mois avec les ventes les plus élevées, mois avec les ventes les plus faibles.

Prise de décision fondée sur les données

2) **Analyse des données**

b. Calculez le taux de croissance (Growth-Rate) des ventes d'une année sur l'autre.

$$growth_rate = \frac{(ventes_{annee_{encours}} - ventes_{annee_{precedente}})}{ventes_{annee_{precedente}}} \times 100$$

D2 = {Janvier: 16.2k£, Février: 17.4k£, Mars: 19£, Avril: 15.8k£, Mai: 18.6k£, Juin: 20.1k£, Juillet: 21.3k£, Août: 22k£, Septembre: 19.5k£, Octobre: 18.1k£, Novembre: 16.9k£, Décembre: 23.5k£}

3) **Visualisation des données**

Prise de décision fondée sur les données

4) Prise de décision

Discutez des décisions commerciales potentielles basées sur l'analyse. Par exemple, envisager des promotions pendant le mois où les ventes sont les plus faibles ou bien fixer des objectifs de vente pour l'année suivante.

- Quel a été le chiffre d'affaires total de l'année écoulée ?
- Quelle est la moyenne des ventes mensuelles ?
- Quel est le mois où les ventes ont été les plus élevées et quel a été le chiffre d'affaires de ce mois ?
- Quel est le mois où les ventes ont été les plus faibles et quel a été le chiffre d'affaires de ce mois ?
- Quel est le taux de croissance des ventes d'une année sur l'autre ?

Confidentialité et sécurité des données

- On doit reconnaître l'importance de la confidentialité et de la sécurité des données dans notre monde riche en données.
- Quelles sont les défis liés à la protection des informations sensibles?

Example : Les violations de données soulignent la nécessité d'adopter des mesures de cybersécurité solides.

Confidentialité et sécurité des données

- On doit reconnaître l'importance de la confidentialité et de la sécurité des données dans notre monde riche en données.
- Quelles sont les défis liés à la protection des informations sensibles?
 - Menaces pour la cybersécurité,
 - Règlement sur la protection des données,
 - Volume et complexité des données,
 - Menaces d'initiés,
 - Sécurité de l'informatique en nuage,
 - Sécurité des appareils mobiles,
 - Risque lié aux tiers,
 - Chiffrement et accès aux données,
 - Technologies émergentes (les appareils IoT peuvent être vulnérables et l'IA peut nécessiter de vastes ensembles de données, ce qui soulève des problèmes de confidentialité)

Visualiser la surcharge de données

LES VISUALISATIONS EFFICACES QUI TRANSFORMENT LE VASTES ENSEMBLES DE DONNÉES EN INFORMATIONS CLAIRES.

Quelles sont les rôles clés de la visualisation des données?



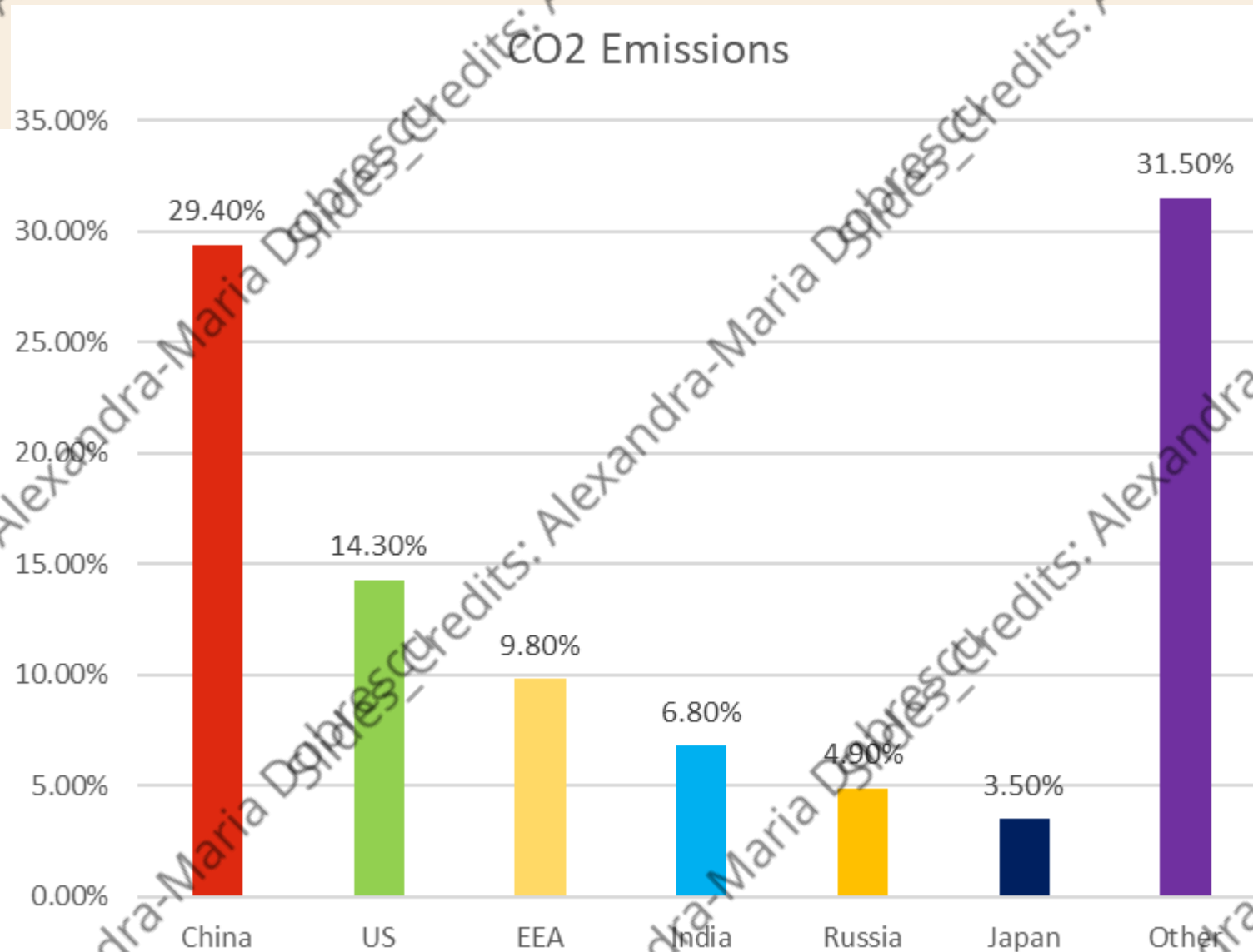
Visualiser la surcharge de données

LES VISUALISATIONS EFFICACES QUI TRANSFORMENT LE VASTES ENSEMBLES DE DONNÉES EN INFORMATIONS CLAIRES.

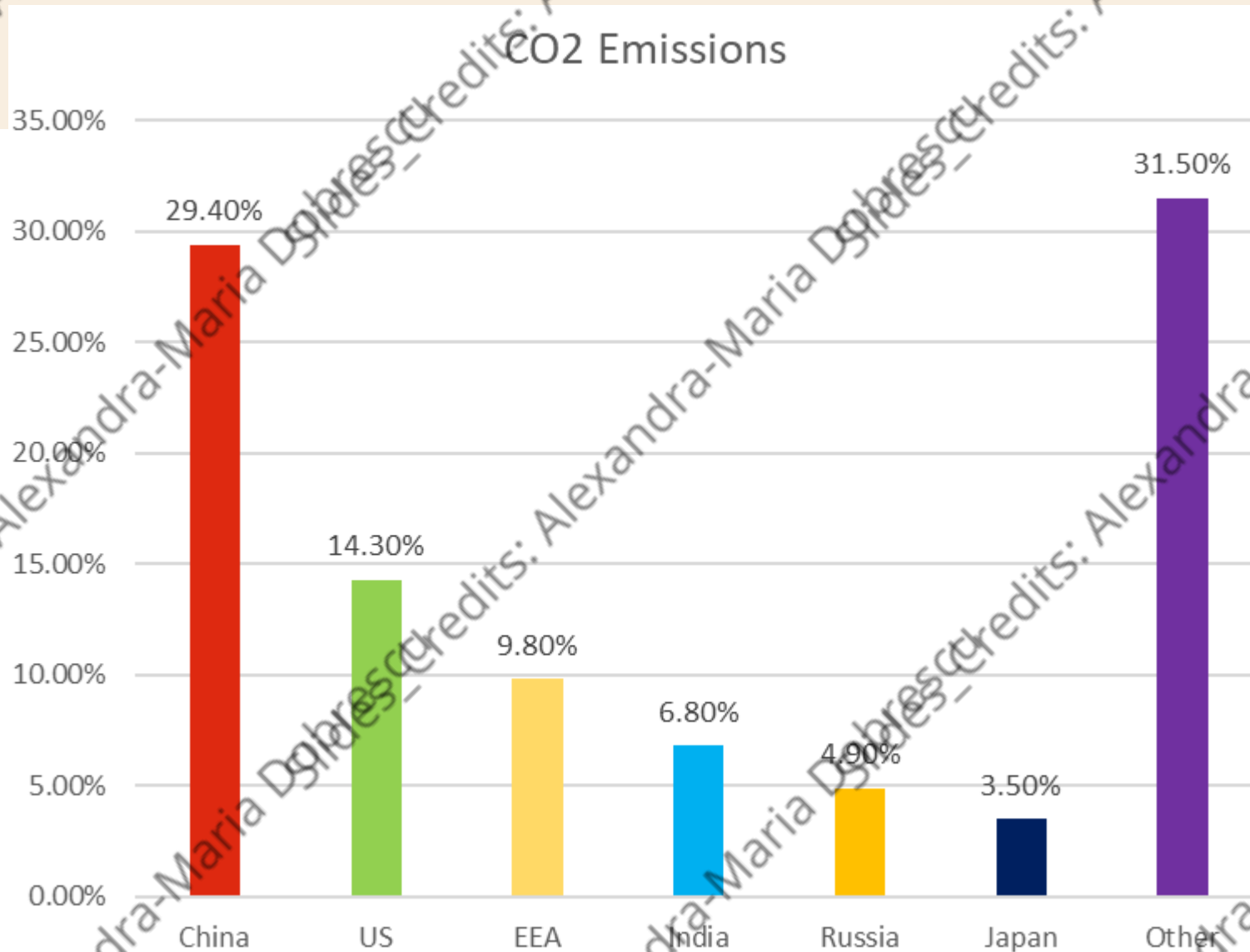
Quelles sont les rôles clés de la visualisation des données?

- **Reconnaissance des trends**
- **Communication**
- **Prise de décision**
- **Comparaison**
- **Simplification et clarification**

Bonne ou Mauvaise visualisation des données? [10]



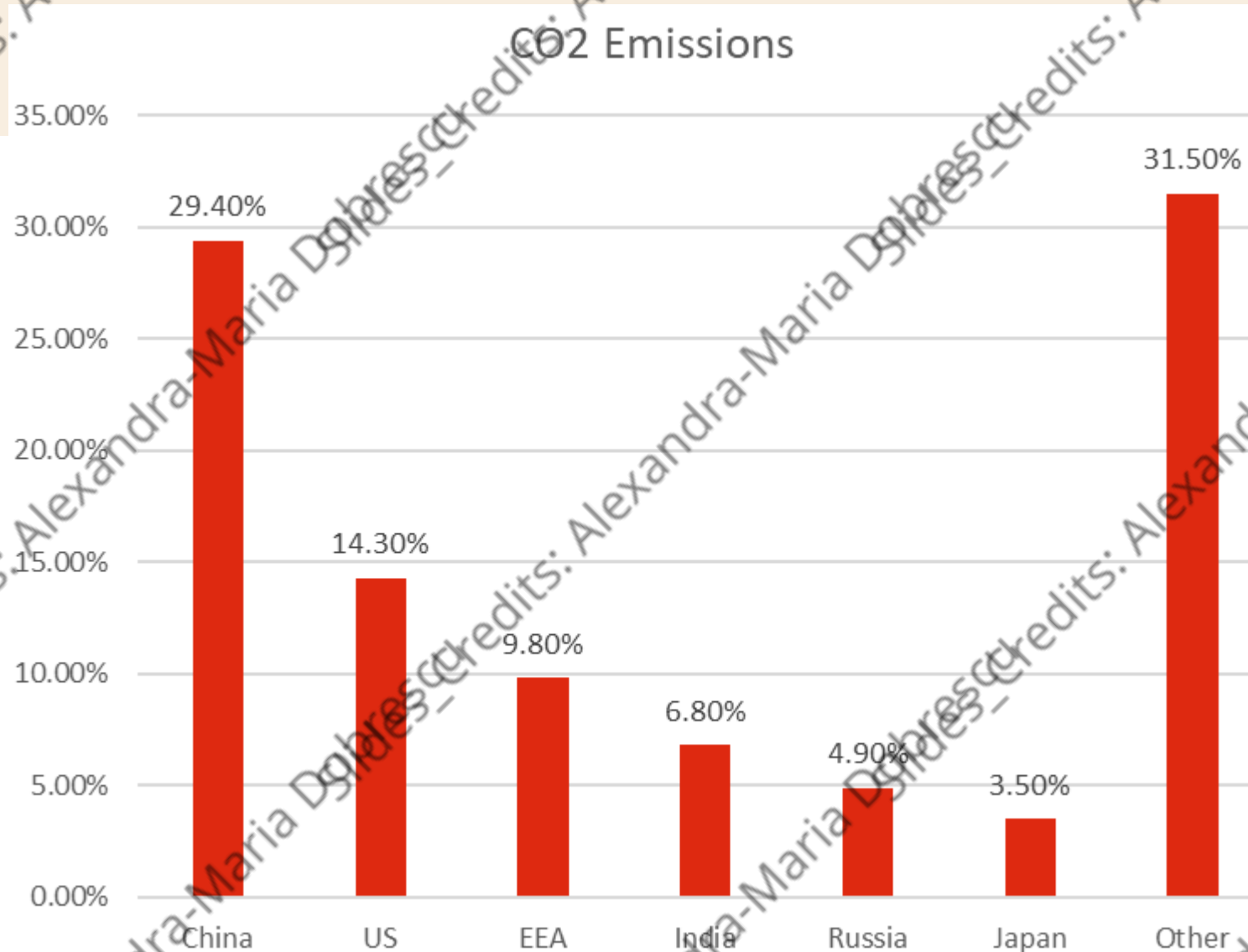
Bonne ou Mauvaise visualisation des données? [10]



couleurs

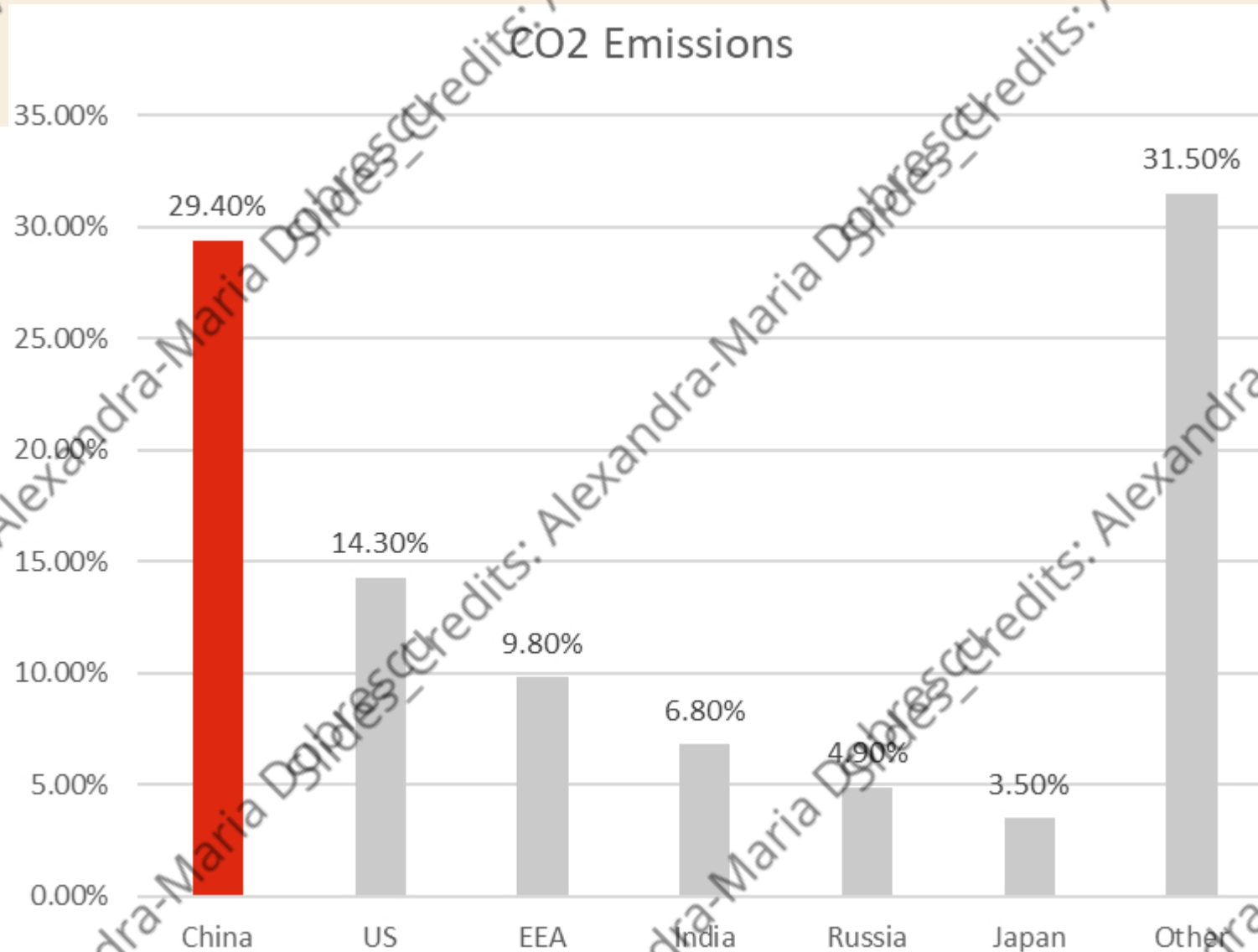
compréhensible

Bonne ou Mauvaise visualisation des données? [10]



Bonne ou Mauvaise visualisation des données? [10]

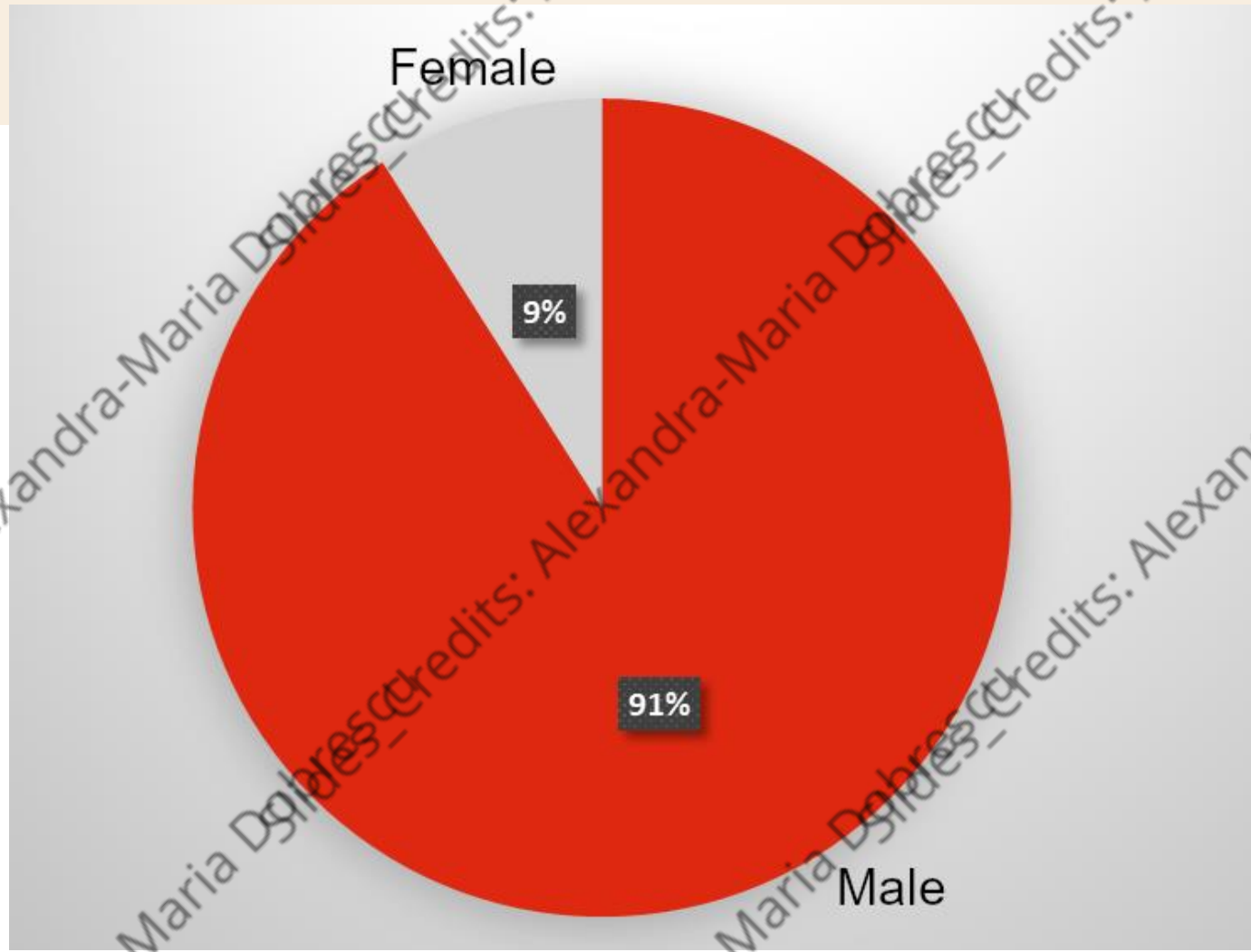
On doit mettre en évidence les informations souhaitées.



Pensez toujours aux catégories de couleurs que notre cerveau doit traiter.

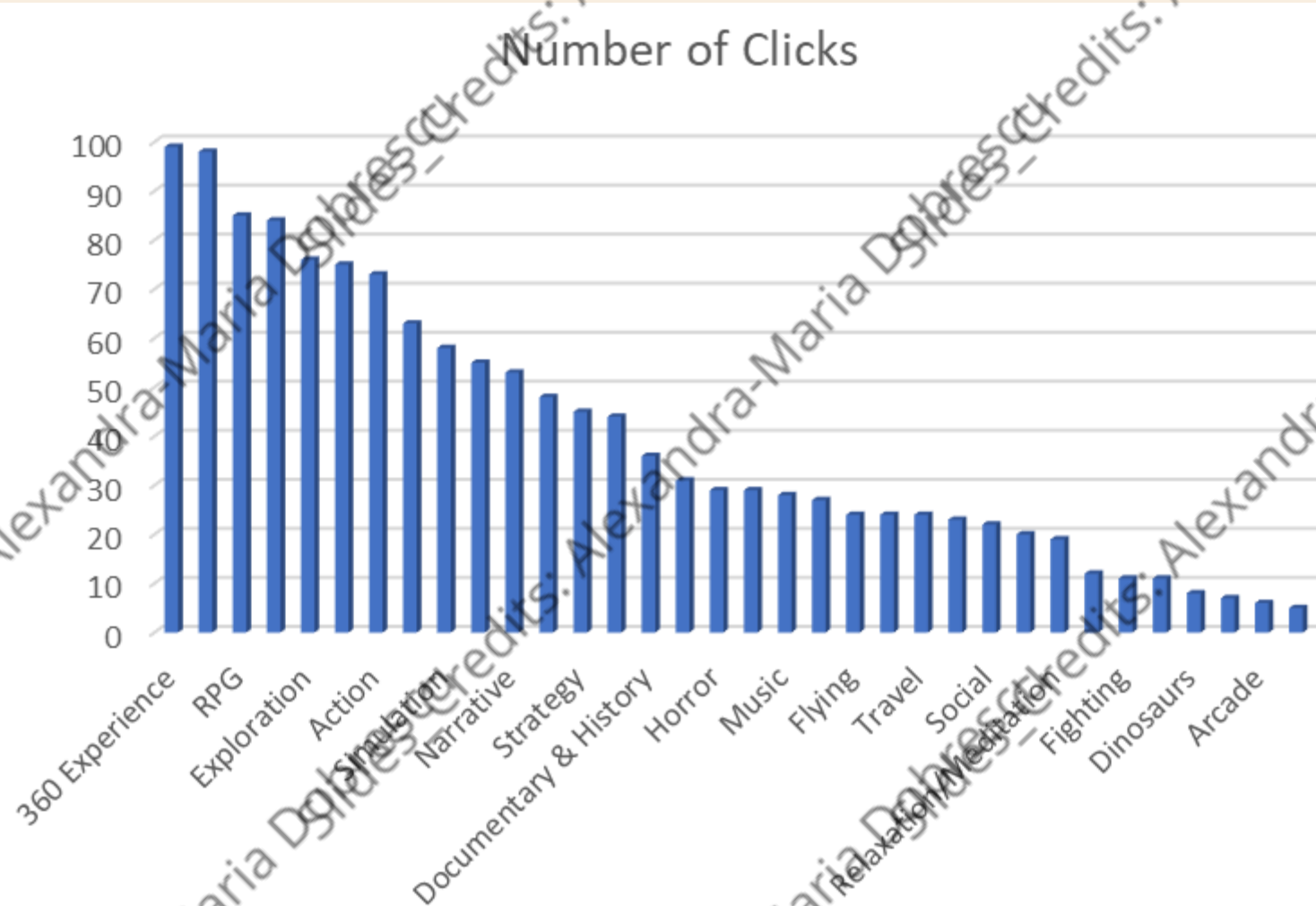
Bonne ou Mauvaise visualisation des données? [10]

On doit mettre en évidence les informations souhaitées.

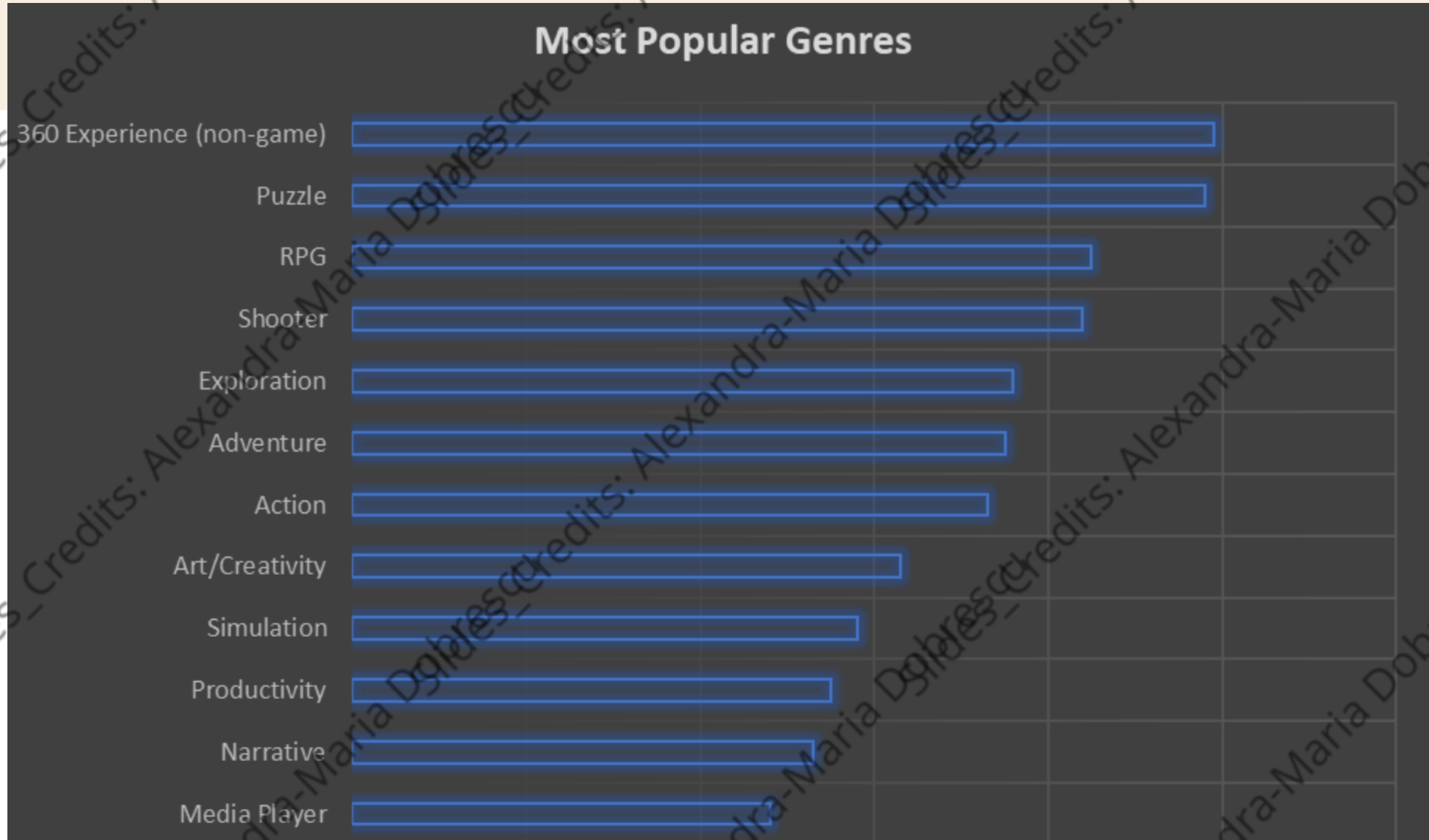


Pensez toujours aux catégories de couleurs que notre cerveau doit traiter.

Bonne ou Mauvaise visualisation des données? [10]



Bonne ou Mauvaise visualisation des données? [10]



Bonne ou Mauvaise visualisation des données? [10]

	A	B	C	D	E	F	G
1	age	first_name	last_name	gender	work_email	work_phone	mobile_phone
2	60	Tull	Mattersley	Male	tull.mattersley@myoffice.com	657-751-8755	611-983-0058
3	60	Donalt	Moreno	Male	donalt.moreno@myoffice.com	645-964-9614	536-535-9404
4	60	Elaina	Teodorski	Female	elaina.teodorski@myoffice.co	330-435-4970	810-309-8959
5	60	Filmer	Justham	Male	filmer.justham@myoffice.com	243-635-8104	330-856-6074
6	60	Steffi	Scullard	Female	steffi.scullard@myoffice.com	542-990-9820	283-515-9623
7	60	Cami	Drohane	Female	cami.drohane@myoffice.com	594-872-4666	414-820-3045
8	59	Harper	Osgar	Male	harper.osgar@myoffice.com	506-536-8971	150-514-2093
9	59	Rebeca	Fawdrey	Female	rebeca.fawdrey@myoffice.cor	568-439-3661	463-609-2353
10	59	Kelsy	Baggalley	Female	kelsy.baggalley@myoffice.con	871-588-5503	749-886-7156
11	59	Fernanda	Dagon	Female	fernanda.dagon@myoffice.cor	338-291-0721	561-434-7051
12	59	Nicolis	Orrum	Male	nicolis.orrum@myoffice.com	770-777-8169	756-857-0727
13	59	Killian	Jannings	Male	killian.jannings@myoffice.con	433-869-3375	822-864-0872
14	59	Reynold	Harcourt	Male	reynold.harcourt@myoffice.cc	300-763-5401	108-677-0881
15	59	Sheba	Tunkin	Female	sheba.tunkin@myoffice.com	980-361-9731	950-568-9340
16	59	Cora	Goude	Female	cora.goude@myoffice.com	635-960-6316	703-940-1949
17	59	Antonin	Tarren	Male	antonin.tarren@myoffice.com	638-708-9787	919-101-0691
18	59	Wainwright	Tingle	Male	wainwright.tingle@myoffice.c	580-463-0646	518-857-9368
19	59	Kevon	Cottee	Male	kevon.cottee@myoffice.com	519-361-6934	505-338-4279
20	59	Kalie	Riden	Female	kalie.riden@myoffice.com	841-712-1156	857-341-9879
21	59	Towny	Sharper	Male	towny.sharper@myoffice.com	252-465-0152	169-642-4792
22	59	Harcourt	Batts	Male	harcourt.batts@myoffice.com	698-217-1895	634-884-6545
23	59	Nathanil	Wigfall	Male	nathanil.wigfall@myoffice.cor	687-646-0510	646-982-9324
24	59	Melamie	Tuckey	Female	melamie.tuckey@myoffice.coi	117-191-2575	996-276-5495
25	59	Carlen	Peasnone	Female	carlen.peasnone@myoffice.co	897-338-7257	555-623-8529
26	59	Janean	Maus	Female	janean.maus@myoffice.com	618-769-1825	218-782-5308
27	59	Drusie	Hordle	Female	drusie.hordle@myoffice.com	600-521-9754	590-116-2562
28	59	Jackie	Gutherson	Female	jackie.gutherson@myoffice.co	559-325-8387	104-406-4731
29	59	Walther	Darrington	Male	walther.darrington@myoffice.	797-699-7001	654-578-1765
30	59	Hedi	Swiers	Female	hedi.swiers@myoffice.com	951-138-9266	431-146-1840
31	59	Natalya	Perch	Female	natalya.perch@myoffice.com	821-338-3983	335-399-7032
32	59	Ciro	Mundall	Male	ciro.mundall@myoffice.com	674-503-8685	132-811-8689
33	59	Iris	Hebron	Female	iris.hebron@myoffice.com	254-686-3126	463-858-3510

Bonne ou Mauvaise visualisation des données? [10]

The image shows a data visualization interface with a sidebar on the left and a main content area on the right. The sidebar contains a search bar, a 'Tags' section, and a 'Views' section with a list of filters. The main content area displays three employee records, each with a detailed view of their information.

Filter	Count
First Name	3,474 +
Last Name	4,307 +
Gender	2 +
Attrition	2 +
Business Travel	3 +
Department	3 +
Education Field	6 +
Job Role	9 +
Marital Status	3 +
Employment S...	2 +
Over 18	1 +
Overtime	1 +
Pay Type	2 +
Employee Eth...	7 +
Asset Category	4 +
Asset Descript...	10 +
Frequency	3 +

Employee	Work Email	Home Email	Monthly Income	Reporting To	Termination Reason	Location	Role	Ethnicity
Mattersley	tull.mattersley@myoffice.com	tull.mattersley@abc.com	5,405 / 5,000 - 6,000	Tull Mattersley	Misconduct	London, UK	Sales Executive	Asian
Moreno	donalt.moreno@myoffice.com	donalt.moreno@abc.com	5,405 / 5,000 - 6,000	Donalt Moreno		London, UK	Sales Executive	Hispanic or Latino
Teodorski	elaina.teodorski@myoffice.com	elaina.teodorski@abc.com	5,405 / 5,000 - 6,000	Elaina Teodorski		Lindon, Utah	Sales Executive	White

Sensibilisation aux données personnelles

IDEA 1: <<LA SENSIBILISATION AUX DONNÉES PERSONNELLES>>, CONNUE COMME *PERSONAL DATA AWARENES* EN ANGLAIS, EST LA COMPRÉHENSION CONSCIENTE:

- DE LA VALEUR,
- DE LA CONFIDENTIALITÉ
- ET DE LA SÉCURITÉ DES INFORMATIONS PERSONNELLES À L'ÈRE NUMÉRIQUE.

IDEA 2: IL PERMET AUX INDIVIDUS DE FAIRE DES CHOIX INFORMÉS ET DE PROTÉGER LEURS DONNÉES DANS UN MONDE DE PLUS EN PLUS INTERCONNECTÉ.

Les données personnelles dans le contexte de l'IdO

CAS NÉGATIF : ACCÈS NON AUTORISÉ AUX DONNÉES DES APPAREILS IOT

«En 2020, un grand constructeur automobile a été victime d'une faille de sécurité dans ses systèmes de voitures connectées. Des pirates informatiques ont pu accéder sans autorisation aux données relatives à la localisation GPS, à la vitesse et à l'historique de conduite de la voiture. Cette faille a suscité des inquiétudes quant à l'utilisation potentiellement abusive des données de conduite personnelles, y compris le suivi des conducteurs sans leur consentement.»

Les données personnelles dans le contexte de l'IdO

CAS POSITIF : SURVEILLANCE DE LA SANTÉ ET DÉTECTION PRÉCOCE

« Dans un exemple positif, les dispositifs IdO ont été utilisés pour le suivi de la santé et la détection précoce de maladies. L'un de ces exemples concerne un dispositif de suivi de la condition physique, doté de capacités de surveillance de la fréquence cardiaque. L'appareil recueillait en permanence des données sur la fréquence cardiaque des utilisateurs. En analysant ces données, les chercheurs ont découvert des schémas susceptibles d'indiquer des signes précoces de troubles cardiaques. Les utilisateurs ont alors été avertis qu'ils devaient consulter un professionnel de la santé, ce qui a permis de sauver des vies. Ce cas montre comment les dispositifs IdO, lorsqu'ils sont utilisés de manière responsable, peuvent avoir un impact positif significatif sur la santé et le bien-être des personnes. »

Types de données

DONNÉES STRUCTURÉES : DES DONNÉES HAUTEMENT ORGANISÉES ET FORMATÉES QUI SUIVENT UNE STRUCTURE SPÉCIFIQUE ET PRÉDÉFINIE. ELLES SONT GÉNÉRALEMENT STOCKÉES DANS DES BASES DE DONNÉES RELATIONNELLES ET SE COMPOSENT DE LIGNES ET DE COLONNES, CE QUI PERMET DE LES INTERROGER ET DE LES ANALYSER FACILEMENT.

Exemples : Les données contenues dans les feuilles de calcul, les informations sur les clients dans les systèmes de gestion de la relation client et les dossiers financiers dans les logiciels de comptabilité sont des exemples courants de données structurées.

Caractéristiques : Les données structurées sont faciles à traiter, à interroger et à analyser en raison de leur format uniforme. Elles sont idéales pour des tâches telles que l'établissement de rapports, l'analyse de données et la business intelligence.



Données structurées

AVANTAGES

Traitement efficace: Les données structurées sont conçues pour être traitées rapidement et facilement par les systèmes informatiques, ce qui les rend très efficaces.

Applicabilité universelle: Les données structurées sont utiles à divers secteurs en raison de leur facilité de traitement.

Possibilité de recherche: Les données structurées peuvent être facilement recherchées à l'aide d'outils d'automatisation traditionnels basés sur des règles.

Compatibilité: Les anciens logiciels d'automatisation tels que l'automatisation des processus robotiques (RPA) et les anciennes technologies telles que la reconnaissance optique des caractères (OCR) peuvent traiter efficacement les données structurées.

Technologie moins avancée: Les documents structurés requièrent une technologie moins avancée, ce qui est avantageux pour les organisations ayant des processus documentaires et des systèmes de données patrimoniaux plus anciens.

Données structurées

LIMITES

L'écriture manuscrite et les éléments non structurés:

Les données structurées ne sont pas adaptées aux documents comportant des éléments manuscrits ou non structurés, ce qui peut entraîner l'absence d'informations essentielles.

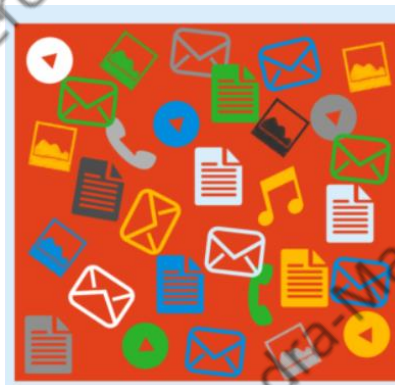
Sources de données omnicanales: Dans le cas de sources de données modernes sans structures bien définies, telles que les médias sociaux et le courrier électronique, les techniques d'extraction traditionnelles basées sur des règles peuvent être limitées.

Types de données

DONNÉES NON STRUCTURÉES : DES DONNÉES QUI N'ONT PAS DE STRUCTURE PRÉDÉFINIE ET QUI NE SONT PAS ORGANISÉES DE MANIÈRE À ÊTRE FACILEMENT TRAITÉES PAR DES MACHINES. ELLES COMPRENNENT SOUVENT DU TEXTE, DES IMAGES, DU CONTENU AUDIO ET VIDÉO..

Exemples : Le texte dans les courriels, les messages sur les médias sociaux, les images, les fichiers audio et les clips vidéo sont des exemples courants de données non structurées.

Caractéristiques : Les données non structurées sont plus difficiles à traiter car elles n'ont pas de format cohérent. Le traitement du langage naturel (NLP) et l'apprentissage automatique sont souvent utilisés pour extraire des informations des données non structurées.



Données non structurées

AVANTAGES

Valeur commerciale débloquée : Les données non structurées, qui représentent une part importante des données d'entreprise, recèlent une valeur commerciale inexploitée pour l'analyse avancée, 80 à 90 % des nouvelles données étant non structurées. Les entreprises qui savent en tirer parti acquièrent un avantage concurrentiel.

Analyse des intentions et des sentiments : Les données non structurées permettent de mettre en œuvre des techniques d'analyse avancées telles que l'analyse des sentiments et des intentions, ce qui aide les entreprises à comprendre les intentions et les émotions des clients, à automatiser le traitement des courriels et à améliorer la satisfaction de la clientèle.

Avantage concurrentiel: L'analyse avancée des données non structurées révèle des informations précieuses, des modèles et des tendances, ce qui permet aux entreprises de prendre des décisions fondées sur des données et d'acquérir un avantage concurrentiel.

Données non structurées

LIMITES

Extraction de données: Les données non structurées sont dépourvues de schémas prévisibles, ce qui complique l'application des méthodes traditionnelles d'extraction de données.

Reconnaissance de l'écriture manuscrite: La technologie traditionnelle de reconnaissance optique des caractères (OCR) a du mal à traiter les éléments manuscrits non structurés, ce qui nécessite un examen manuel pour une extraction précise des données.

Précision et qualité des données: La précision de l'OCR diminue lorsque les scans sont de mauvaise qualité ou que le texte présente des taches, ce qui nuit à la vitesse et à la précision du traitement des données.

Confidentialité et sécurité des données: Les données non structurées peuvent contenir des informations sensibles, ce qui nécessite de les protéger contre les accès non autorisés et de se conformer aux réglementations en matière de protection des données.

Données non structurées

LIMITES

Problèmes de stockage évolutif: Le stockage, le traitement et l'extraction de données non structurées volumineuses et diverses posent des problèmes, ce qui nécessite d'investir dans des solutions de stockage évolutives et des outils de traitement de données avancés.

Défis d'intégration: L'intégration des données non structurées dans les systèmes existants peut nécessiter des techniques de mappage et de consolidation des données afin de garantir la compatibilité et la fluidité des flux de données.



Types de données



DONNÉES SEMI-STRUCTURÉES : SE SITUENT ENTRE LES DONNÉES STRUCTURÉES ET NON STRUCTURÉES. ELLES PRÉSENTENT UNE STRUCTURE PARTIELLE, SOUVENT SOUS LA FORME DE BALISES (TAGS) OU DE MÉTADONNÉES QUI FOURNISSENT UNE CERTAINE ORGANISATION. BIEN QU'ELLES NE SOIENT PAS AUSSI BIEN ORGANISÉES QUE LES DONNÉES STRUCTURÉES, ELLES SONT PLUS STRUCTURÉES QUE LES DONNÉES NON STRUCTURÉES.

Exemples : Les documents au format XML ou JSON, les pages web HTML et les bases de données NoSQL qui utilisent des schémas flexibles contiennent des données semi-structurées.

Caractéristiques : Les données semi-structurées offrent un certain degré de flexibilité, ce qui les rend adaptées aux scénarios dans lesquels la structure des données peut varier. Elles sont couramment utilisées dans les applications web, la gestion de contenu et l'échange de données. Elles peuvent également fournir des informations supplémentaires pour les données non structurées.

Données semi-structurées

AVANTAGES

Analyse des données: Les données semi-structurées contiennent souvent des informations contextuelles supplémentaires, telles que des métadonnées ou des étiquettes, ce qui améliore la précision et la pertinence de l'analyse des données.

Flexibilité: Les données semi-structurées sont plus souples pour le stockage et la gestion des données, car elles n'adhèrent pas à des modèles rigides, ce qui permet d'intégrer plus facilement de nouveaux types de données dans les systèmes existants.

Évolutivité : Les données semi-structurées sont hautement évolutives et s'adaptent aux systèmes informatiques distribués et aux options de stockage telles que les bases de données sur site, les lacs de données et le stockage en nuage, ce qui permet de traiter efficacement de gros volumes de données.

Intégration: Les données semi-structurées s'intègrent de manière transparente à d'autres types de données, y compris les données non structurées, ce qui permet de combiner et de comparer rapidement des données provenant de diverses sources.

Données semi-structurées

LIMITES

Extraction et intégration des données:

L'extraction et l'intégration des données semi-structurées peuvent être plus complexes et nécessitent souvent des outils spécialisés tels que le traitement intelligent des documents (IDP) avec des processus d'extraction de données OCR contextuels personnalisés en raison de la diversité des modèles et des éléments de données.

Qualité et cohérence des données: Garantir la qualité des données dans des environnements de données semi-structurées est difficile, car l'absence de modèles et de schémas de données stricts peut entraîner des incohérences, des duplications et des divergences. La résolution de ces problèmes nécessite une attention humaine ou une technologie d'IA avancée pour le nettoyage et la normalisation des données.

Bibliographie

- [1] https://www.freepik.com/premium-vector/round-social-media-icons-network-platforms-logos_12628744.htm
- [2] <https://www.taggdigital.com/blog/what-are-smartwatch-sensors-and-how-do-they-function>
- [3] <https://www.goodhousekeeping.com/uk/product-reviews/tech/g685723/best-smart-heating-thermostats/#product-e8b0b9d6-3938-4403-8422-24f8da27ff43>
- [4] Kramer, O., & Kramer, O. (2016). Scikit-learn. Machine learning for evolution strategies, 45-53.
- [5] <https://www.geeksforgeeks.org/difference-between-cassandra-and-mongodb/>
- [6] <https://www.image-net.org/about.php>
- [7] G. Miller, R. Beckwith, C. Felbaum, D. Gross, K. Miller, Introduction to WordNet: An On-line Lexical Database (1993)
- [8] Knafllic, C. N. (2015). Storytelling with data: A data visualization guide for business professionals. John Wiley & Sons.
- [9] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25.
- [10] <https://www.polygonsearch.com/blog/10-good-and-bad-examples-of-data-visualization>