

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 1 |



**C'est
moi
qui**

...



ÉTUDES DE LICENCE

FILS (CTI_FR) @ UPB



ÉTUDES DE MASTER

ACS (AI_ENG) @ UPB



INGÉNIEUR DE RECHERCHE

Intelligence artificielle



ASSISTANT D'ENSEIGNEMENT

Laboratoires IA @ UPB

Vision par Ordinateur (Computer Vision) @ UB



OBJECTIF HUMAIN

Mentorat et enseignement



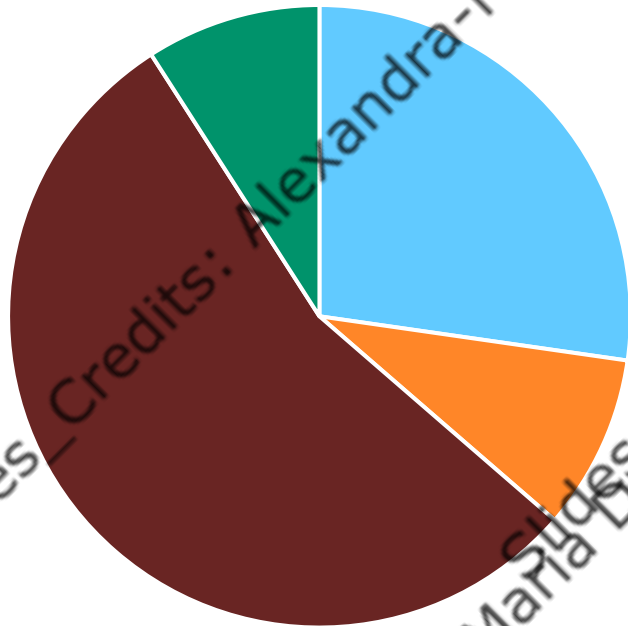
OBJECTIF TECHNIQUE

Résoudre le problème du changement climatique et continuer à apprendre

Dis-m'en plus sur les notes

QUANTITATIVE

Note Finale



■ Examen ■ Tests de cours ■ Les laboratoires ■ Bonus

QUALITATIVE

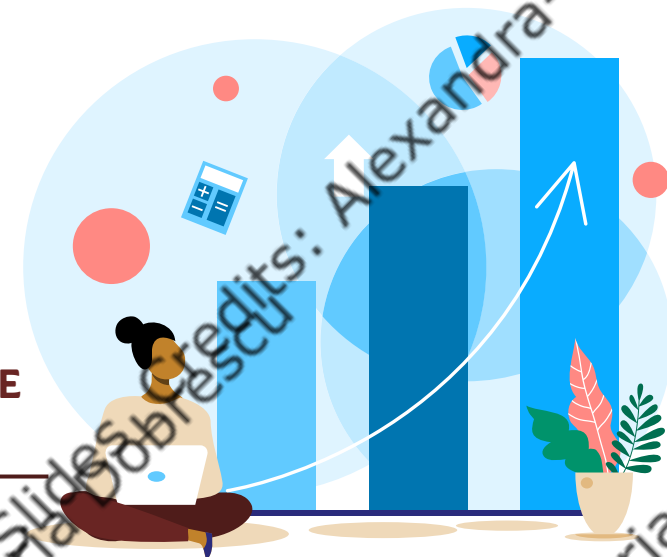
Nous

RESPECT

TRANSPARENCE

TRAVAIL

APPRENTISSAGE

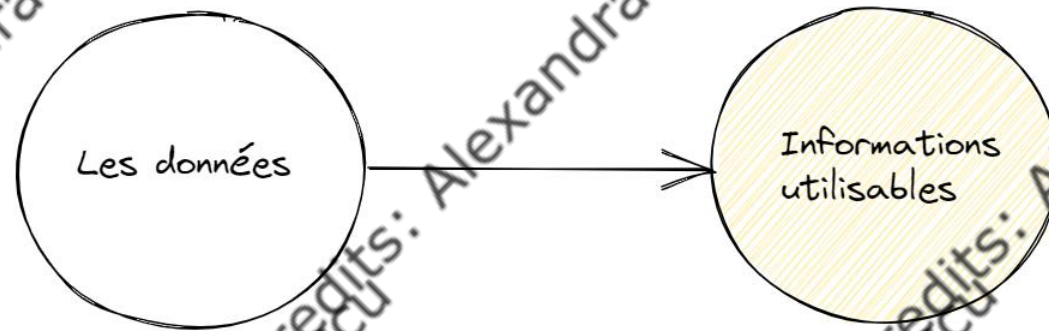


Introduction

- 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)

Définir l'analyse des données

L'ANALYSE DES DONNÉES EST LE PROCESSUS QUI CONSISTE À METTRE DE L'ORDRE ET DE LA STRUCTURE DANS LES DONNÉES COLLECTÉES.



L'analyse est effectuée à l'aide de méthodes systématiques:

- pour rechercher des tendances,
- des regroupements,
- d'autres relations entre différents types de données.

Définir la visualisation des données

LA VISUALISATION DES DONNÉES CONSISTE À PRÉSENTER LES DONNÉES SOUS FORME DE DIAGRAMME, DE GRAPHIQUE OU D'AUTRE FORMAT VISUEL, EN FONCTION DU PROBLÈME QUE L'ON SOUHAITE RÉSOUDRE.

POURQUOI ?

Pour faciliter l'analyse et l'interprétation.

COMMENT ?

D'une manière accessible et attrayante pour le destinataire.

OBJECTIF ?

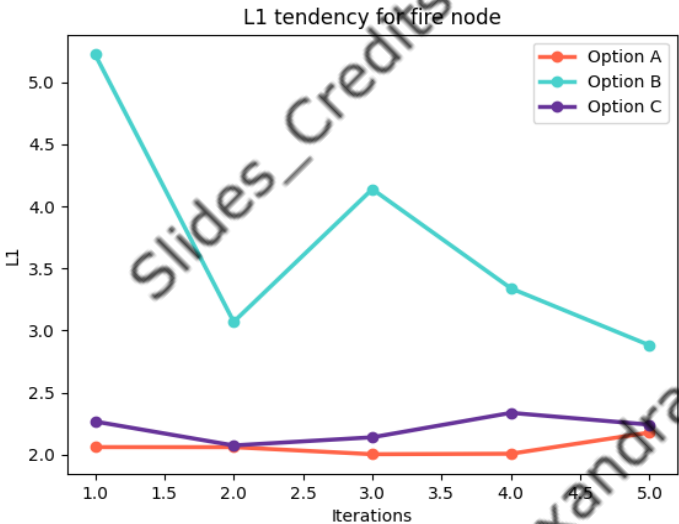
Communication.

En général, dans les problèmes de la vie réelle, il y a de grandes quantités de données et vous avez besoin de plusieurs outils de visualisation des données. Vous pouvez les combiner en fonction de votre domaine.

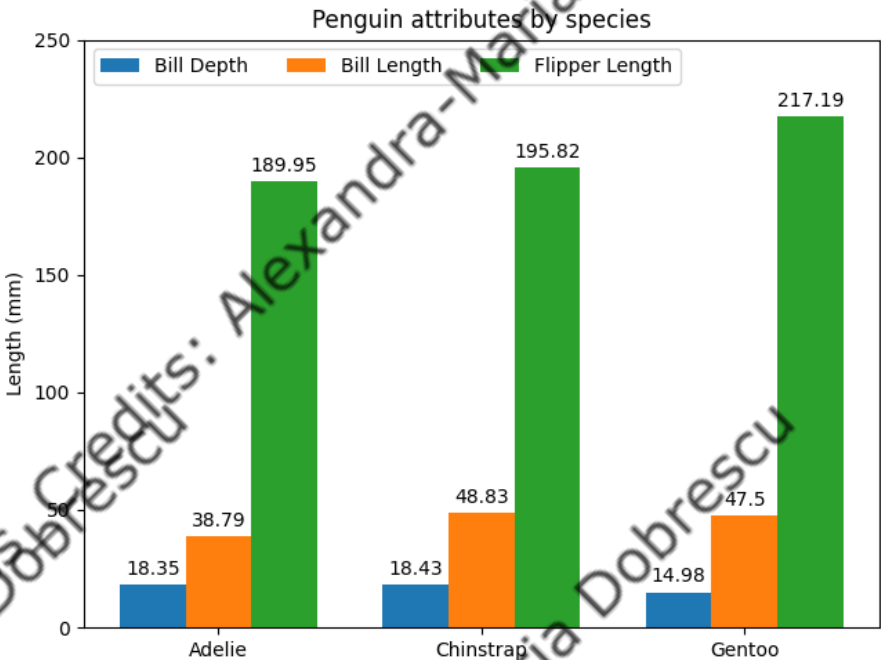
Définir la visualisation des données

TABLEAUX DE FRÉQUENCE, TABLEAUX CROISÉ

Fire Summary			
Input and Iteration (it)	L1	L2	Relative to Best
Single Link: it 1	3.191	0.002	0.627
Step 1 – A: it 3	2.0027	0.002	1
Step 1 – B: it 5	2.880	0.002	0.695
Step 1 – C: it 2	2.073	0.002	0.965
Step 2: it 4	2.691	0.002	0.743



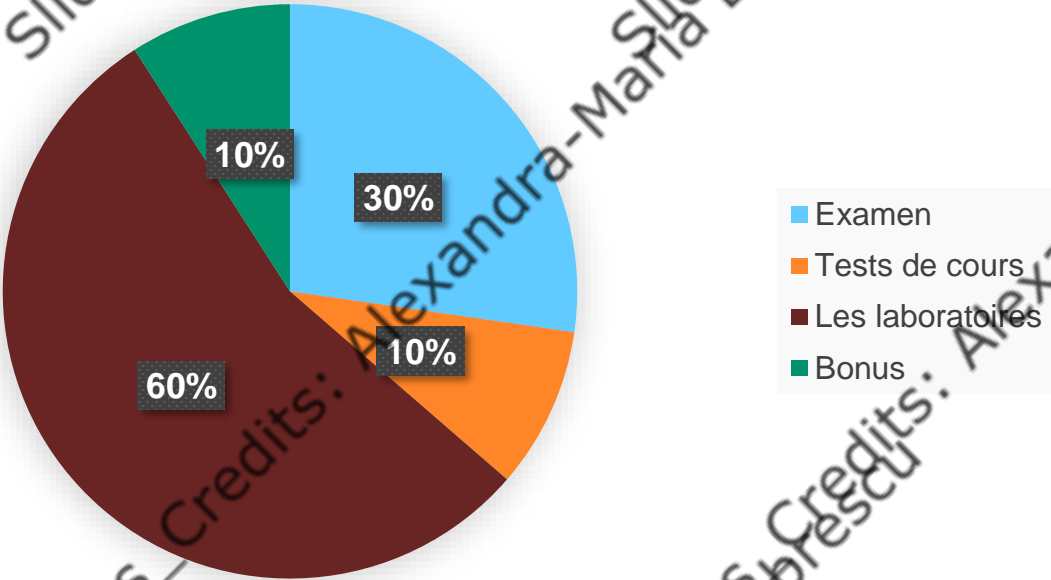
GRAPHIQUES
LINÉAIRES



DIAGRAMMES À BARRES [1]

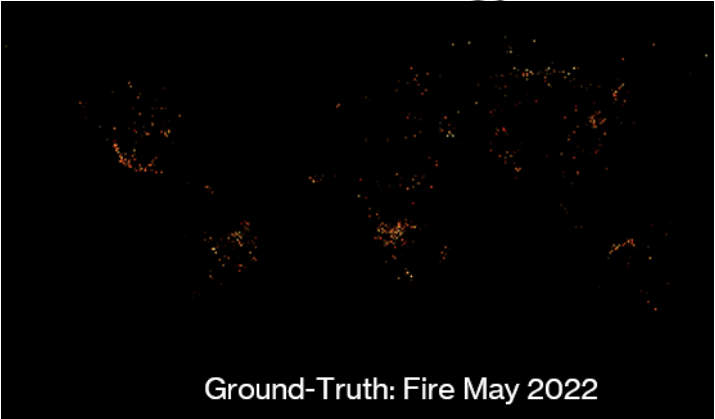
Définir la visualisation des données

Note Finale



DIAGRAMMES CIRCULAIRES

CARTES THERMIQUES [1], [2]

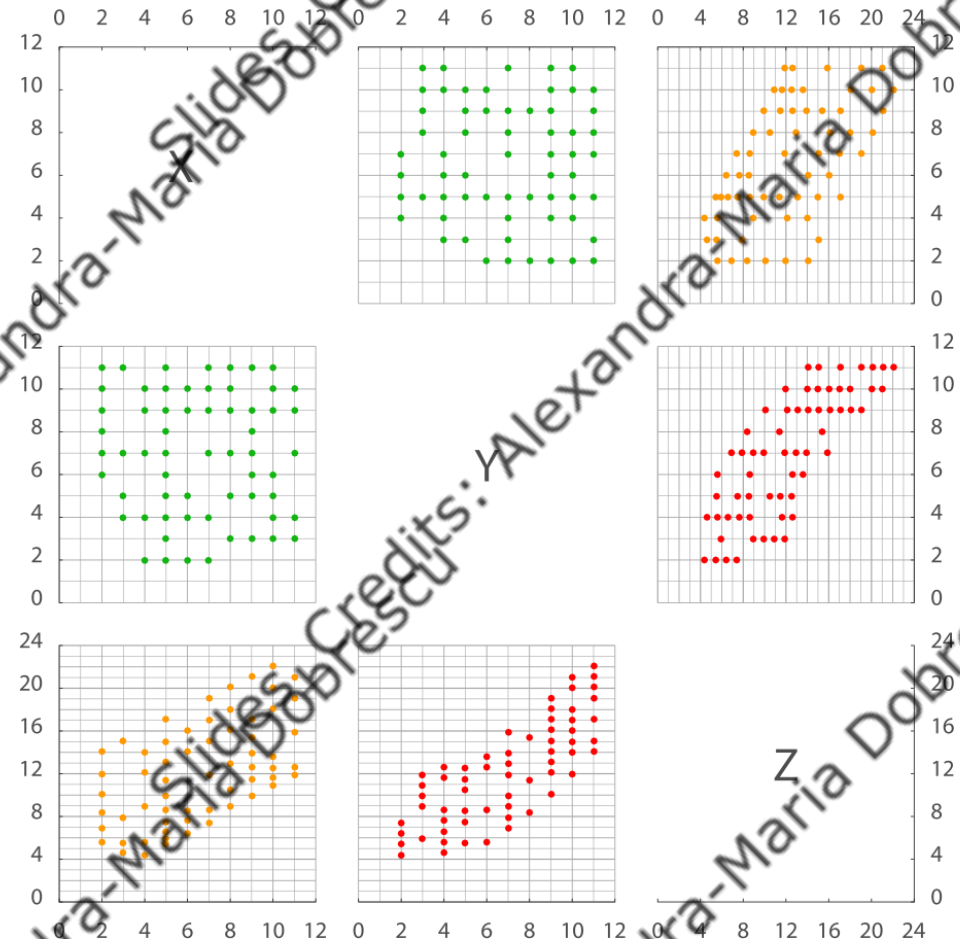
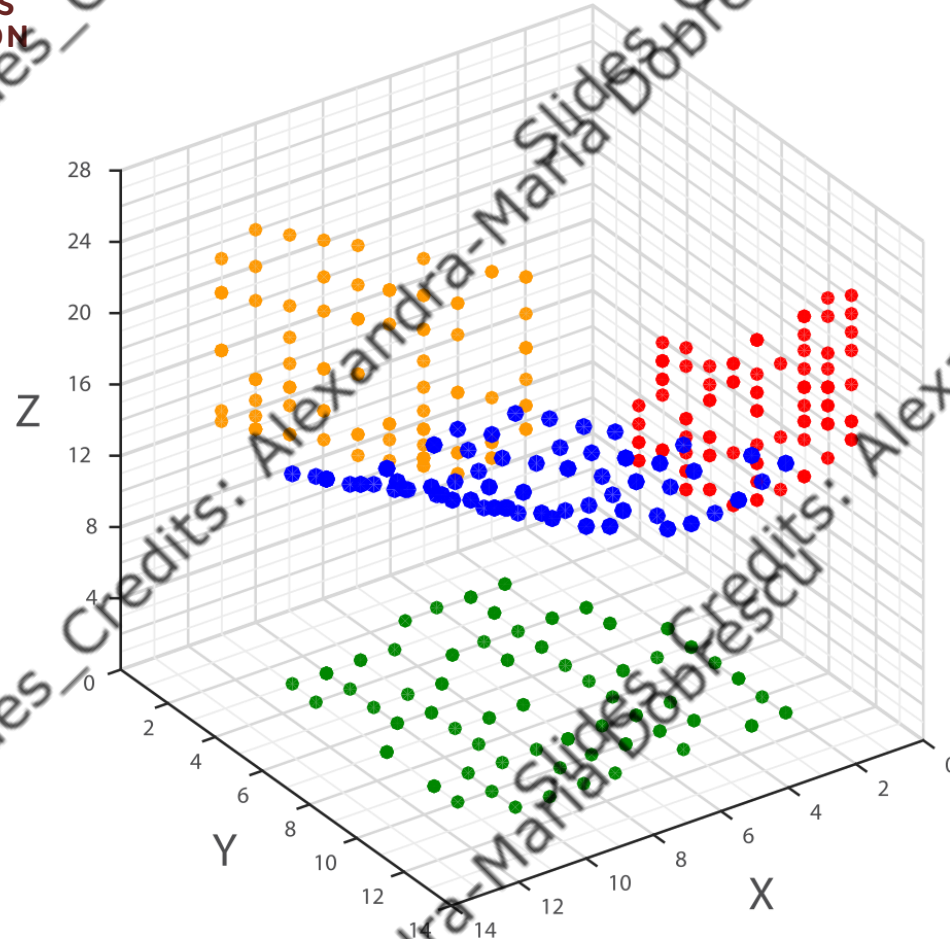


Harvest of local farmers (in tons/year)

	Home Inc.	Upland Bros.	Smith Gardening	AgriFun	Organiculture	BioGoods Ltd.	Cornylee Corp.
cucumber	0.8	2.4	2.5	3.9	0.0	4.0	0.0
tomato	2.4	0.0	4.0	1.0	2.7	0.0	0.0
lettuce	1.1	2.4	0.8	4.3	1.9	4.4	0.0
asparagus	0.6	0.0	0.3	0.0	3.1	0.0	0.0
potato	0.7	1.7	0.6	2.6	2.2	6.2	0.0
wheat	1.3	1.2	0.0	0.0	0.0	3.2	5.1
barley	0.1	2.0	0.0	1.4	0.0	1.9	6.3

Définir la visualisation des données

DIAGRAMMES
DE DISPERSION
[3]



Outils de visualisation des données

TABLEAUX DE FRÉQUENCE (FREQUENCY TABLES)

- À utiliser lorsque vous souhaitez afficher le nombre ou la fréquence de données catégorielles.
- Convient pour résumer des données comportant un petit nombre de catégories ou de valeurs.
- Utile pour créer une vue d'ensemble rapide de la distribution des données.

TABLEAUX CROISÉS (CONTINGENCY TABLES)

- À utiliser lorsque vous souhaitez explorer les relations entre deux ou plusieurs variables catégorielles.
- Ils sont utiles pour identifier des modèles et des associations dans les données, en particulier dans les enquêtes ou les expériences comportant plusieurs variables catégorielles.

Outils de visualisation des données

TABLEAUX DE FRÉQUENCE (FREQUENCY TABLES)

- À utiliser lorsque vous souhaitez afficher le nombre ou la fréquence de données catégorielles.
- Convient pour résumer des données comportant un petit nombre de catégories ou de valeurs.
- Utile pour créer une vue d'ensemble rapide de la distribution des données.

Exercice 1: **Catégorie A Fréquence**

Catégorie X 15

Catégorie Y 20

Catégorie Z 10

TABLEAUX CROISÉS (CONTINGENCY TABLES)

- À utiliser lorsque vous souhaitez explorer les relations entre deux ou plusieurs variables catégorielles.
- Ils sont utiles pour identifier des modèles et des associations dans les données, en particulier dans les enquêtes ou les expériences comportant plusieurs variables catégorielles.

Étant donné le tableau de fréquences de la variable A, construisez un tableau croisé.

Outils de visualisation des données

Exercice 2: Supposons que nous disposions d'un ensemble de données de notes d'étudiants, et que les notes soient classées comme suit : $\{+/-A, +/-B, +/-C, +/-D, F\}$.
Pour la colonne de fréquence, il y a un total de 86 étudiants dans cet ensemble de données.
Peut-on construire un tableau de fréquences avec des données discrètes ?

Outils de visualisation des données

DIAGRAMMES À BARRES (BAR CHARTS)

- À utiliser lorsque vous souhaitez afficher la fréquence, la distribution ou la comparaison de données catégorielles.
- Efficace pour montrer des données avec des catégories distinctes.
- Utiles pour comparer des données entre différentes catégories.

CARTES THERMIQUES (HEAT MAPS)

- À utiliser lorsque vous souhaitez visualiser les relations entre deux variables catégorielles ou l'intensité d'une relation entre deux variables continues.
- Elles conviennent à l'affichage de vastes ensembles de données dont les valeurs sont représentées par des cellules codées en couleur.
- Efficace pour identifier des modèles, des grappes ou des tendances dans des données complexes.

GRAPHIQUES LINÉAIRES (LINE GRAPHS)

- À utiliser lorsque vous souhaitez afficher des tendances ou des changements dans les données au fil du temps ou dans des catégories ordonnées.
- Ils conviennent à la représentation de données continues présentant une séquence ou une progression claire.
- Idéal pour visualiser des séries de données temporelles.

GRAPHIQUES CIRCULAIRES (PIE CHARTS)

- À utiliser lorsque vous souhaitez afficher les parties d'un tout, où chaque catégorie représente une partie du total.
- Ils sont efficaces lorsque le nombre de catégories est faible, généralement moins de six ou sept, afin d'éviter les surcharges et les erreurs d'interprétation.

GRAPHIQUES DE DISPERSION (SCATTER PLOTS)

- À utiliser lorsque vous souhaitez visualiser la relation entre deux variables continues.
- Idéal pour identifier les corrélations, les valeurs aberrantes et les groupes de données.
- Efficace pour évaluer la force et la direction d'une relation entre des variables.

L'interprétation des données

L'INTERPRÉTATION DES DONNÉES EST LE PROCESSUS QUI CONSISTE À DONNER UN SENS AUX DONNÉES.

L'interprétation nécessite de tirer des conclusions sur :

- la généralisation,
- la corrélation,
- la causalité.

Vise à répondre à des questions d'apprentissage clés sur le projet/problème à résoudre.

L'interprétation des données

L'INTERPRÉTATION DES DONNÉES EST LE PROCESSUS QUI CONSISTE À DONNER UN SENS AUX DONNÉES.

L'interprétation nécessite de tirer des conclusions sur :

- **la généralisation,**
- la corrélation,
- la causalité.

Vise à répondre à des questions d'apprentissage clés sur le projet/problème à résoudre.

Une enquête menée auprès de 100 clients nous permet de conclure que la plupart des clients préfèrent le produit A au produit B, car 80 % d'entre eux ont choisi le produit A.

L'interprétation des données

L'INTERPRÉTATION DES DONNÉES EST LE PROCESSUS QUI CONSISTE À DONNER UN SENS AUX DONNÉES.

L'interprétation nécessite de tirer des conclusions sur :

- **la généralisation,**
- la corrélation,
- la causalité.

Vise à répondre à des questions d'apprentissage clés sur le projet/problème à résoudre.

Obs: On a tiré des conclusions plus larges à partir de données ou d'observations spécifiques.

Une enquête menée auprès de 100 clients nous permet de conclure que la plupart des clients préfèrent le produit A au produit B, car 80 % d'entre eux ont choisi le produit A.

L'interprétation des données

L'INTERPRÉTATION DES DONNÉES EST LE PROCESSUS QUI CONSISTE À DONNER UN SENS AUX DONNÉES.

L'interprétation nécessite de tirer des conclusions sur :

- la généralisation,
- **la corrélation**,
- la causalité.

Vise à répondre à des questions d'apprentissage clés sur le projet/problème à résoudre.

Il existe une forte corrélation positive entre les heures d'étude et les résultats aux examens, ce qui signifie que plus le temps d'étude augmente, plus les résultats aux examens ont tendance à augmenter.

L'interprétation des données

L'INTERPRÉTATION DES DONNÉES EST LE PROCESSUS QUI CONSISTE À DONNER UN SENS AUX DONNÉES.

L'interprétation nécessite de tirer des conclusions sur :

- la généralisation,
- **la corrélation**,
- la causalité.

Vise à répondre à des questions d'apprentissage clés sur le projet/problème à résoudre.

Obs: On a utilisé la relation statistique entre deux ou plusieurs variables, où les changements d'une variable sont associés aux changements d'une autre.

Il existe une forte corrélation positive entre les heures d'étude et les résultats aux examens, ce qui signifie que plus le temps d'étude augmente, plus les résultats aux examens ont tendance à augmenter.

L'interprétation des données

L'INTERPRÉTATION DES DONNÉES EST LE PROCESSUS QUI CONSISTE À DONNER UN SENS AUX DONNÉES.

L'interprétation nécessite de tirer des conclusions sur :

- la généralisation,
- la corrélation,
- **la causalité.**

Vise à répondre à des questions d'apprentissage clés sur le projet/problème à résoudre.

Il existe un lien de causalité entre le tabagisme et le cancer du poumon, des études scientifiques ayant montré que le tabagisme augmente le risque de développer un cancer du poumon.

L'interprétation des données

L'INTERPRÉTATION DES DONNÉES EST LE PROCESSUS QUI CONSISTE À DONNER UN SENS AUX DONNÉES.

L'interprétation nécessite de tirer des conclusions sur :

- la généralisation,
- la corrélation,
- **la causalité.**

Vise à répondre à des questions d'apprentissage clés sur le projet/problemème à résoudre.

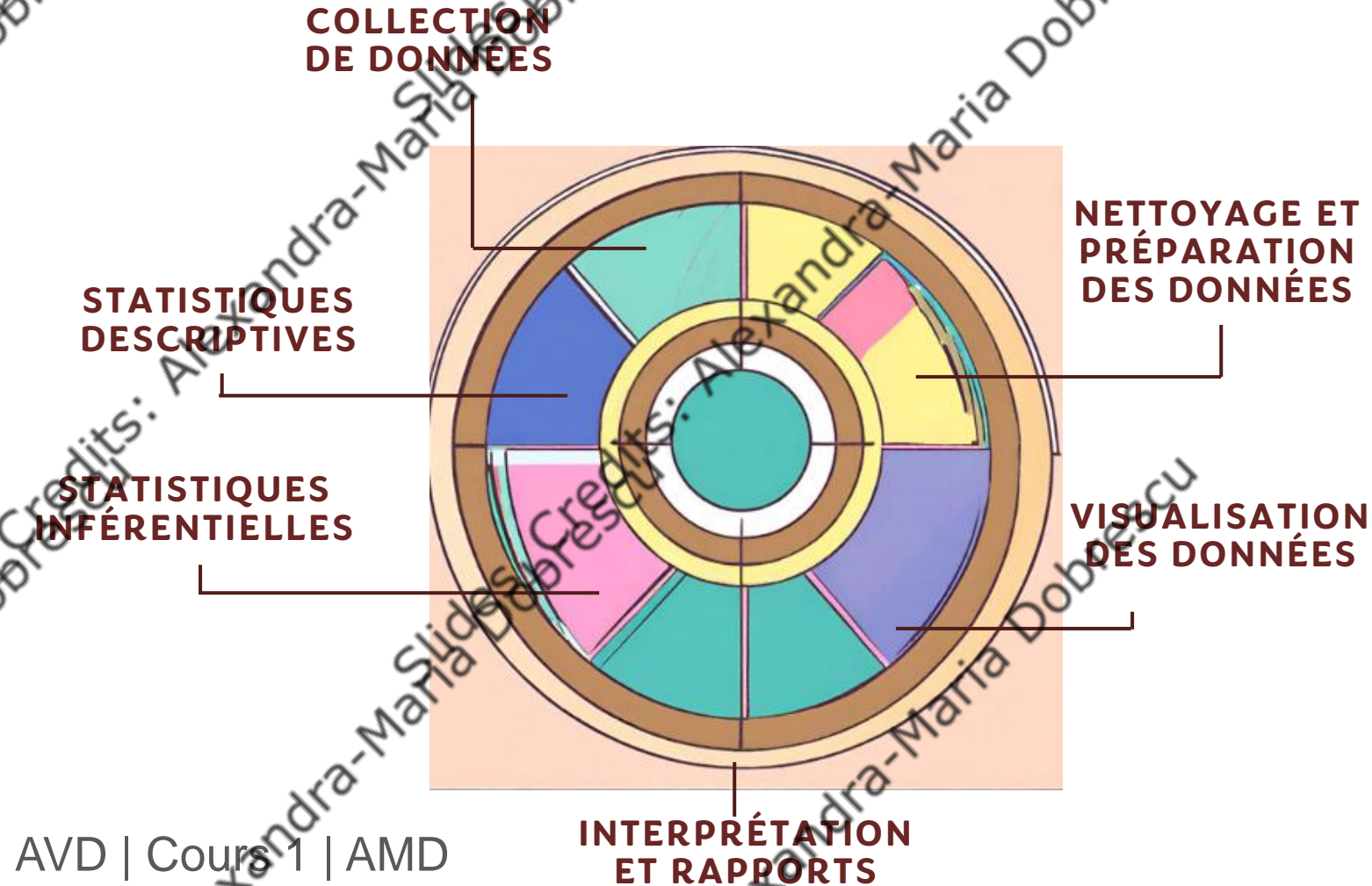
Obs: On a utilisé la relation de cause à effet, où un événement (la cause) entraîne un autre événement (l'effet).

Il existe un lien de causalité entre le tabagisme et le cancer du poumon, des études scientifiques ayant montré que le tabagisme augmente le risque de développer un cancer du poumon.

Aprofondir: Types d'analyse des données

QUANTITATIF

- Un processus systématique de collection, de traitement, d'interprétation et de conclusion de données numériques.
- Elle implique l'utilisation de techniques mathématiques et statistiques pour analyser et donner un sens à des informations quantitatives ou numériques.
- Ce type d'analyse est couramment utilisé dans divers domaines, notamment la science, les affaires, les sciences sociales ou l'économie.



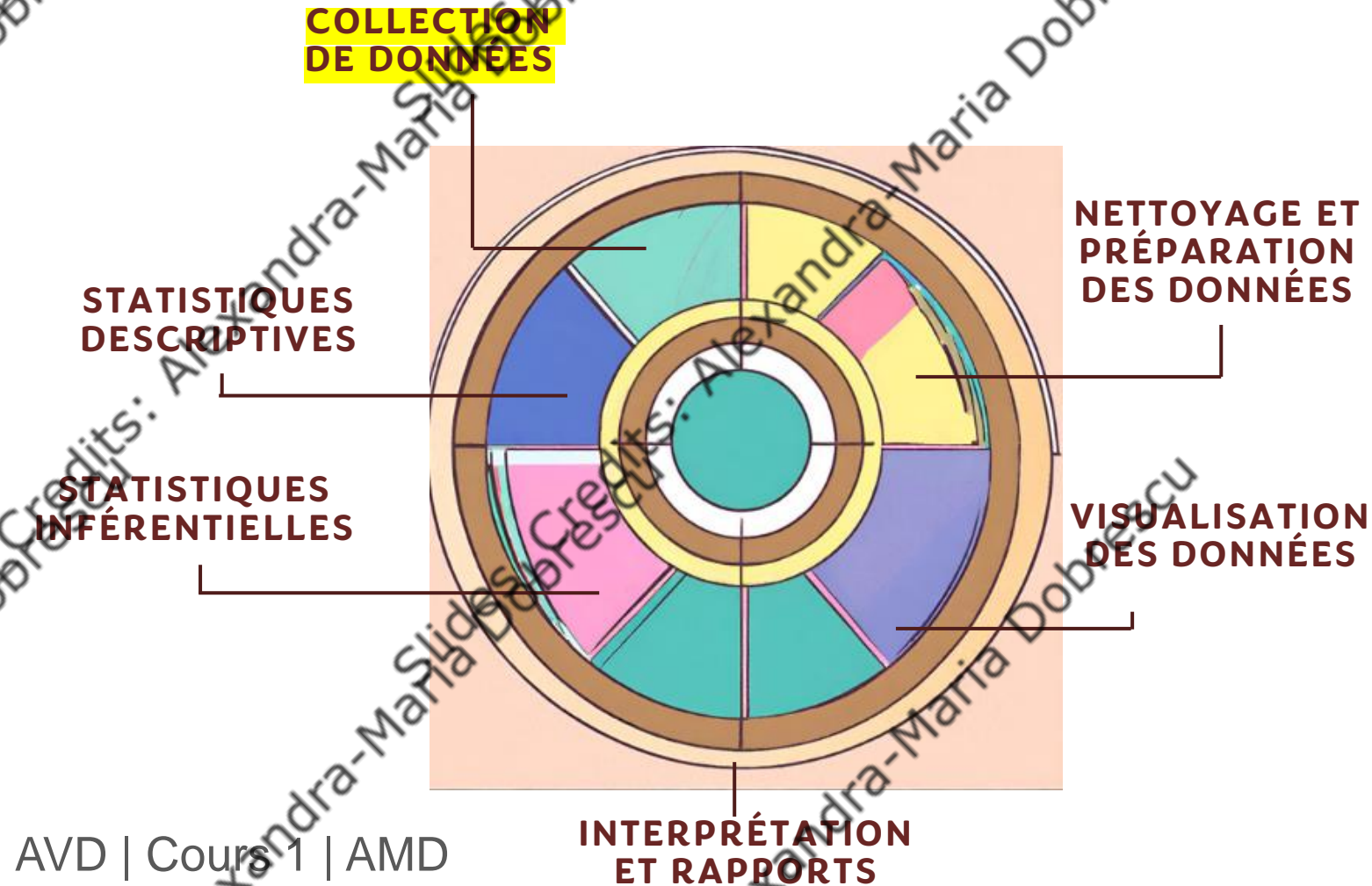
Aprofondir: Types d'analyse des données

QUANTITATIF

Comment ? Enquêtes, expérimentations, observations ou autres méthodes de collecte de données.

Sources ?

- transactions financières,
- partenariats commerciaux,
- écosystèmes IoT,
- applications mobiles,
- multimédia,
- médias sociaux,
- systèmes opérationnels.



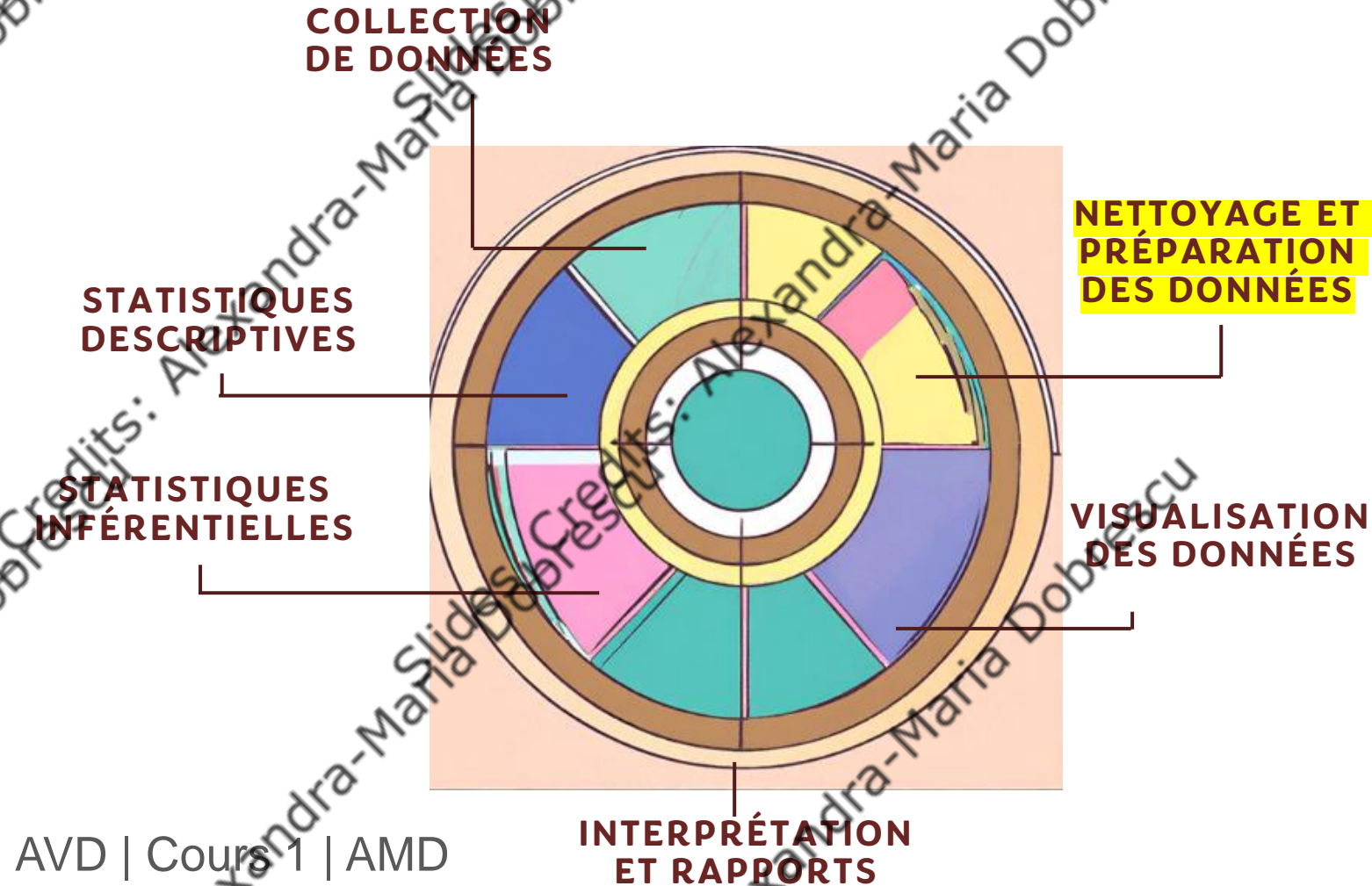
Aprofondir: Types d'analyse des données

QUANTITATIF

Objectif ? Nettoyer et organiser les données pour en assurer l'exactitude et la cohérence.

L'aspect caché, mais la vérité:

Gérer les valeurs manquantes, les valeurs du type «outliers» et les problèmes de formatage... et découvrir d'autres problèmes inconnus.



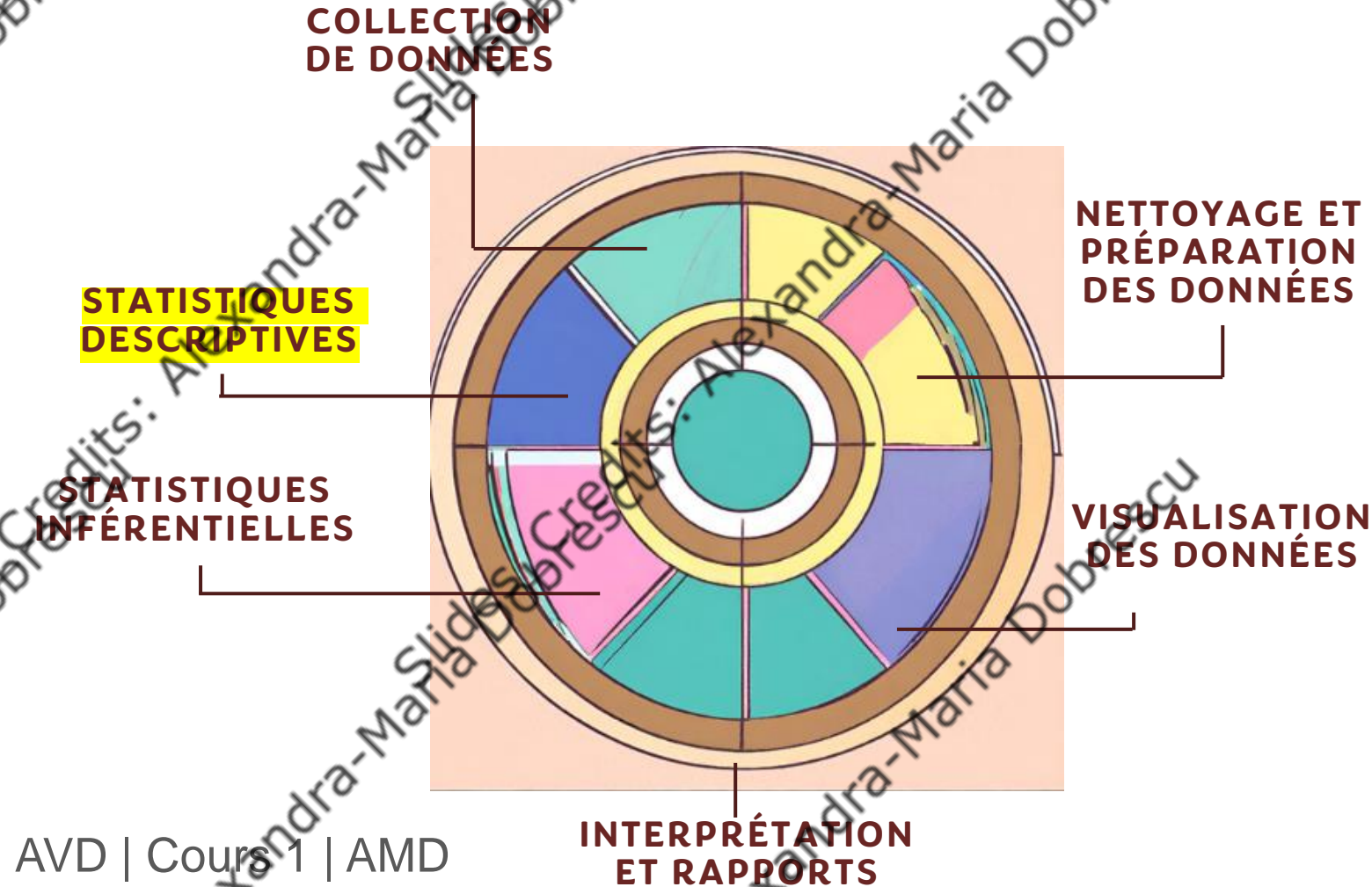
Aprofondir: Types d'analyse des données

QUANTITATIF

Objectif ? Décrire la tendance centrale et la variabilité des données.

Comment ?

- Calcul de statistiques sommaires, telles que la moyenne, la médiane, le mode, l'écart-type et les centiles.
- Toute autre mesure pertinente pour le problème est plus que bienvenue.



Aprofondir: Types d'analyse des données

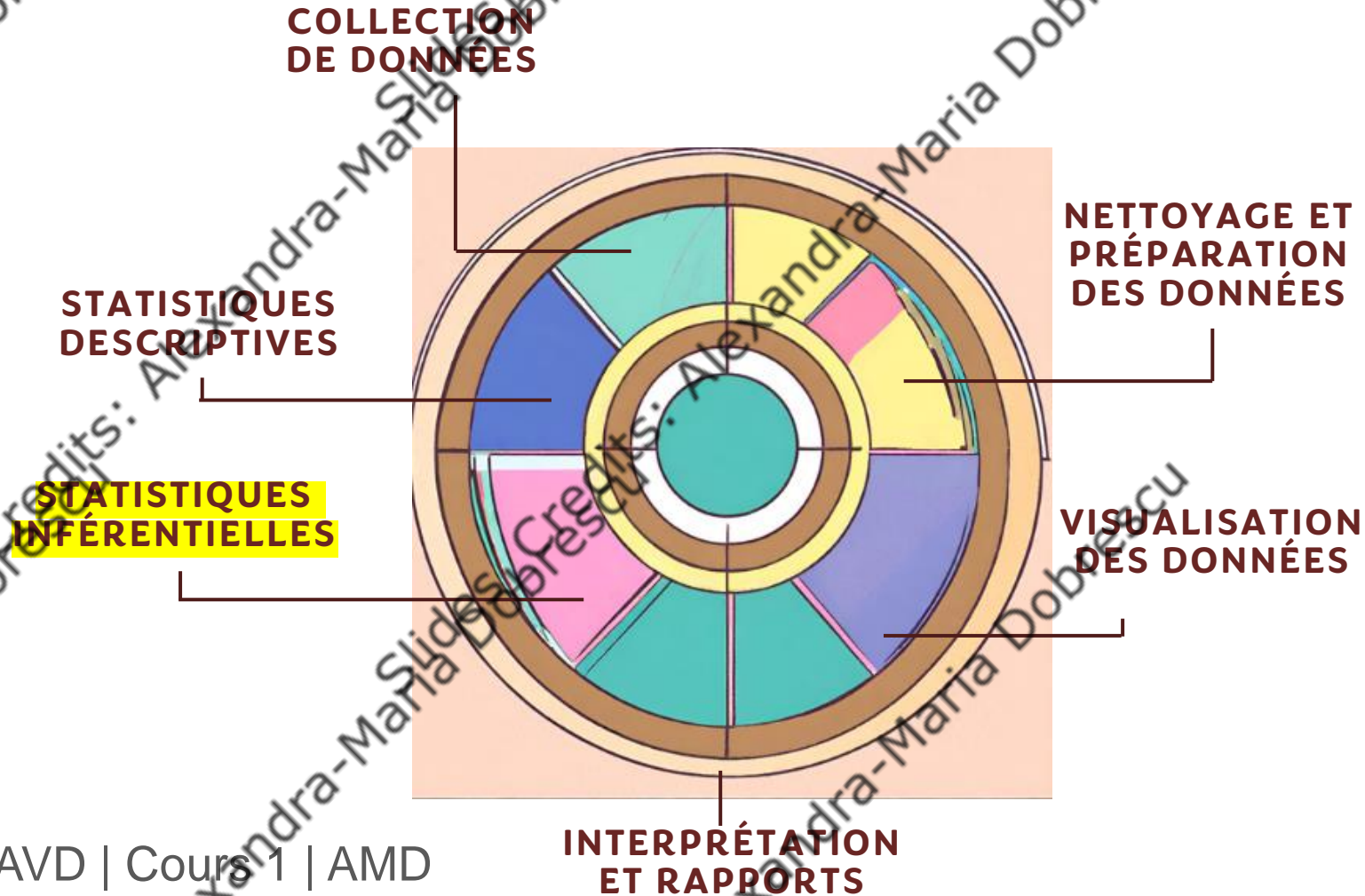
QUANTITATIF

Objectif ? Faire des déductions et tester des hypothèses sur la population dont les données sont tirées.

Comment ?

Les techniques statistiques telles que:

- les tests d'hypothèse,
- l'analyse de régression / classification,
- l'analyse de la variance (σ^2),
- l'analyse de corrélation.



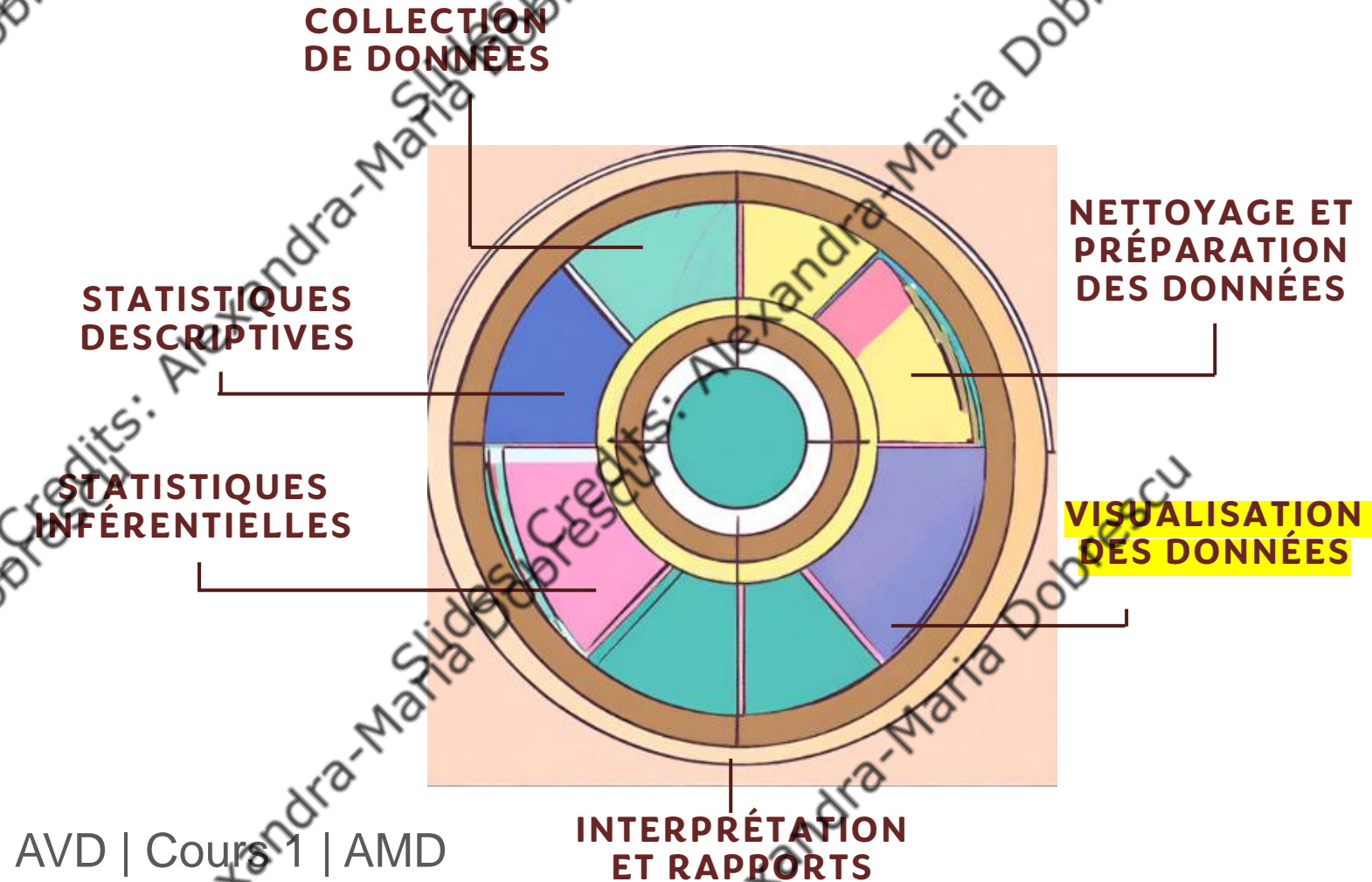
Aprofondir: Types d'analyse des données

QUANTITATIF

Objectif ? La compréhension et la communication des résultats.

Comment ?

Créer des représentations visuelles des données à l'aide de tableaux, de graphiques, d'histogrammes, de diagrammes de dispersion etc.



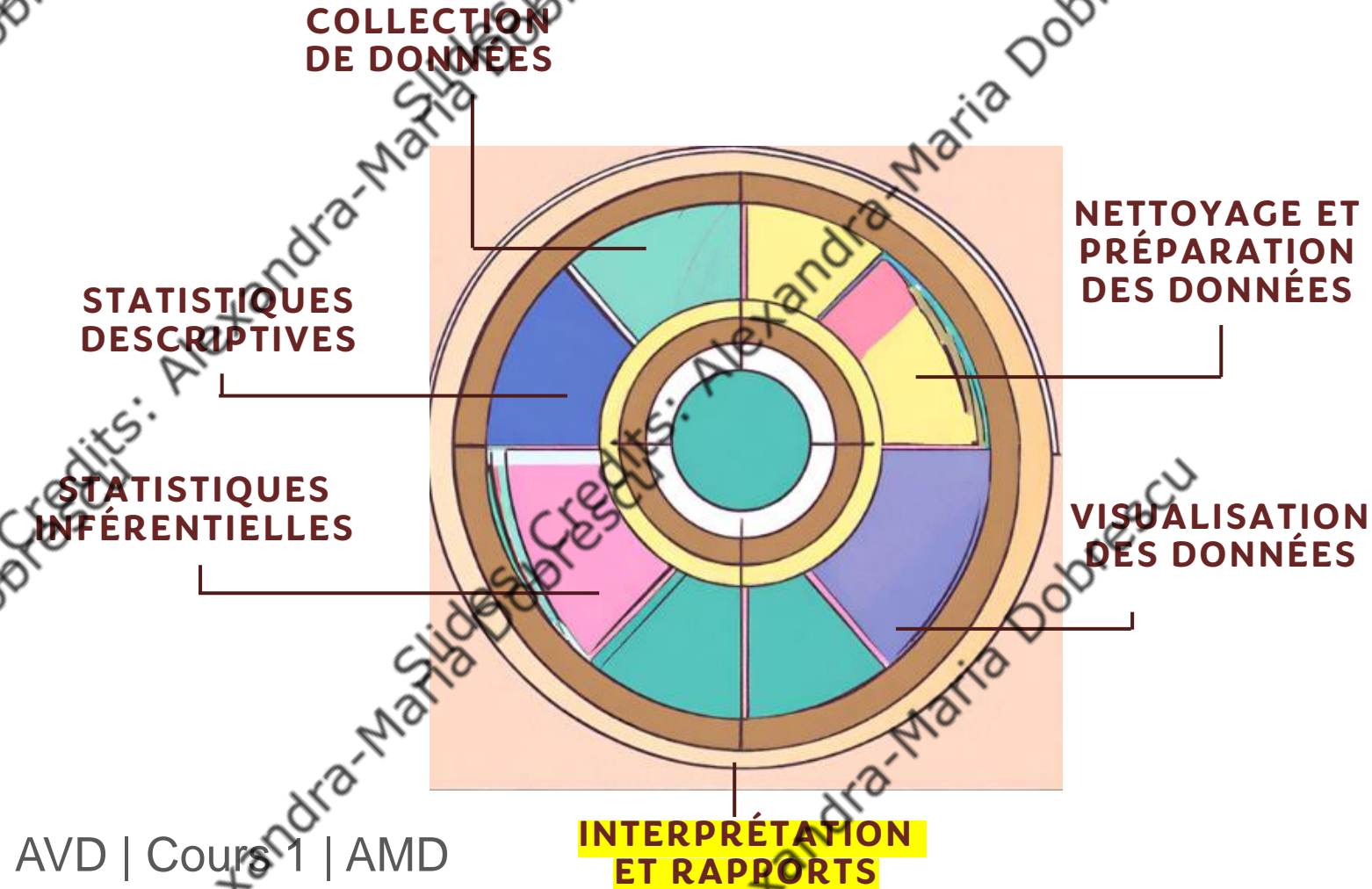
Aprofondir: Types d'analyse des données

QUANTITATIF

Objectif ? Interpréter les résultats de l'analyse et tirer des conclusions significatives sur la base de preuves statistiques.

Comment ? Les résultats sont souvent présentés dans des.

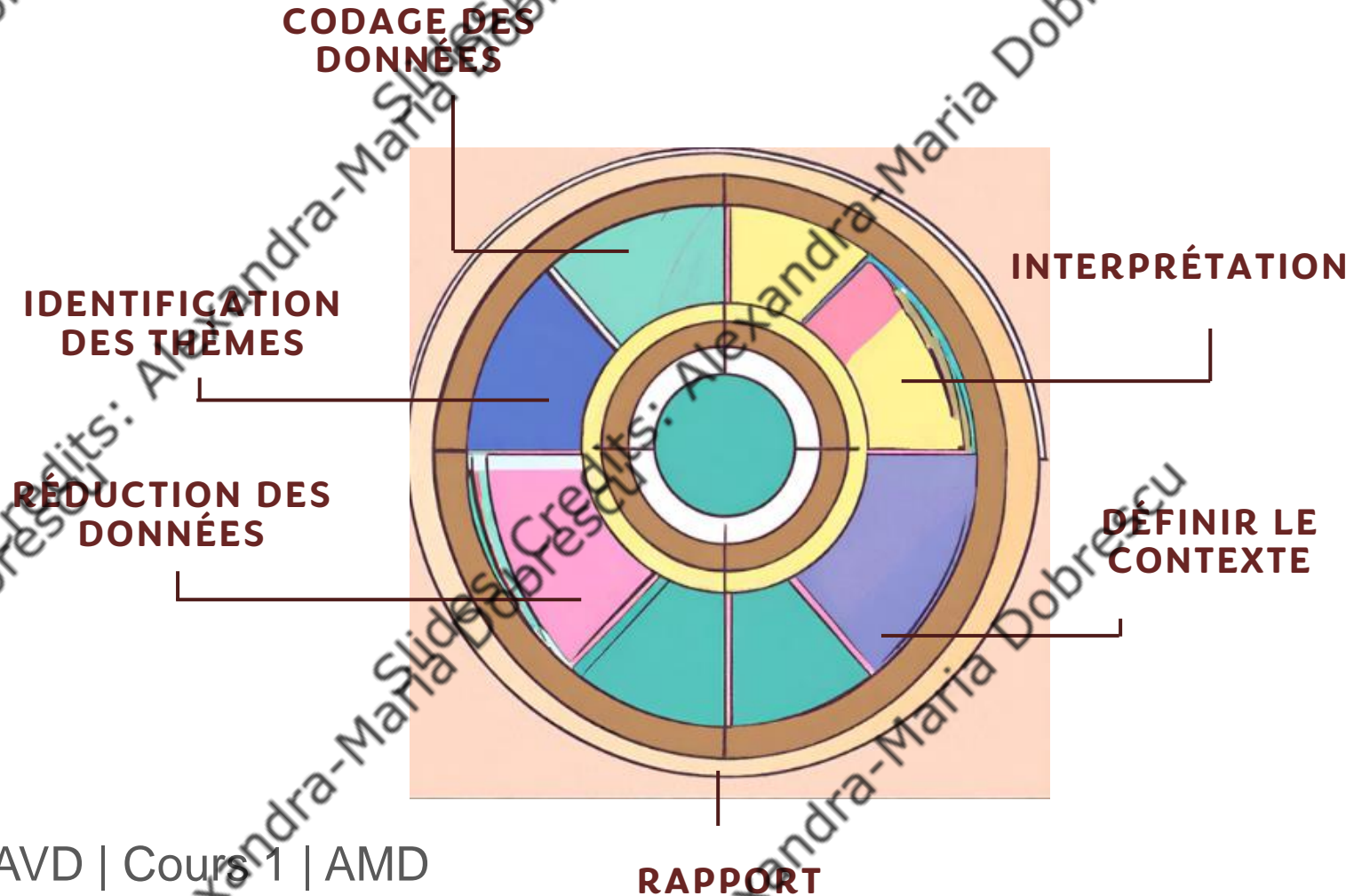
- documents de recherche,
- des rapports,
- des présentations,
- toute autre preuve matérielle ;).



Aprofondir: Types d'analyse des données

QUALITATIVE

- L'analyse de données qualitatives est un processus systématique et subjectif d'examen et d'interprétation de données non numériques, souvent sous la forme de textes, d'images, des fichiers audio ou de vidéos, afin d'identifier des thèmes, des modèles, des idées et des significations.
- Ce type d'analyse est généralement utilisé dans des domaines tels que les sciences sociales, les sciences humaines et la recherche qualitative afin d'explorer et de comprendre la richesse et la complexité des données qualitatives.

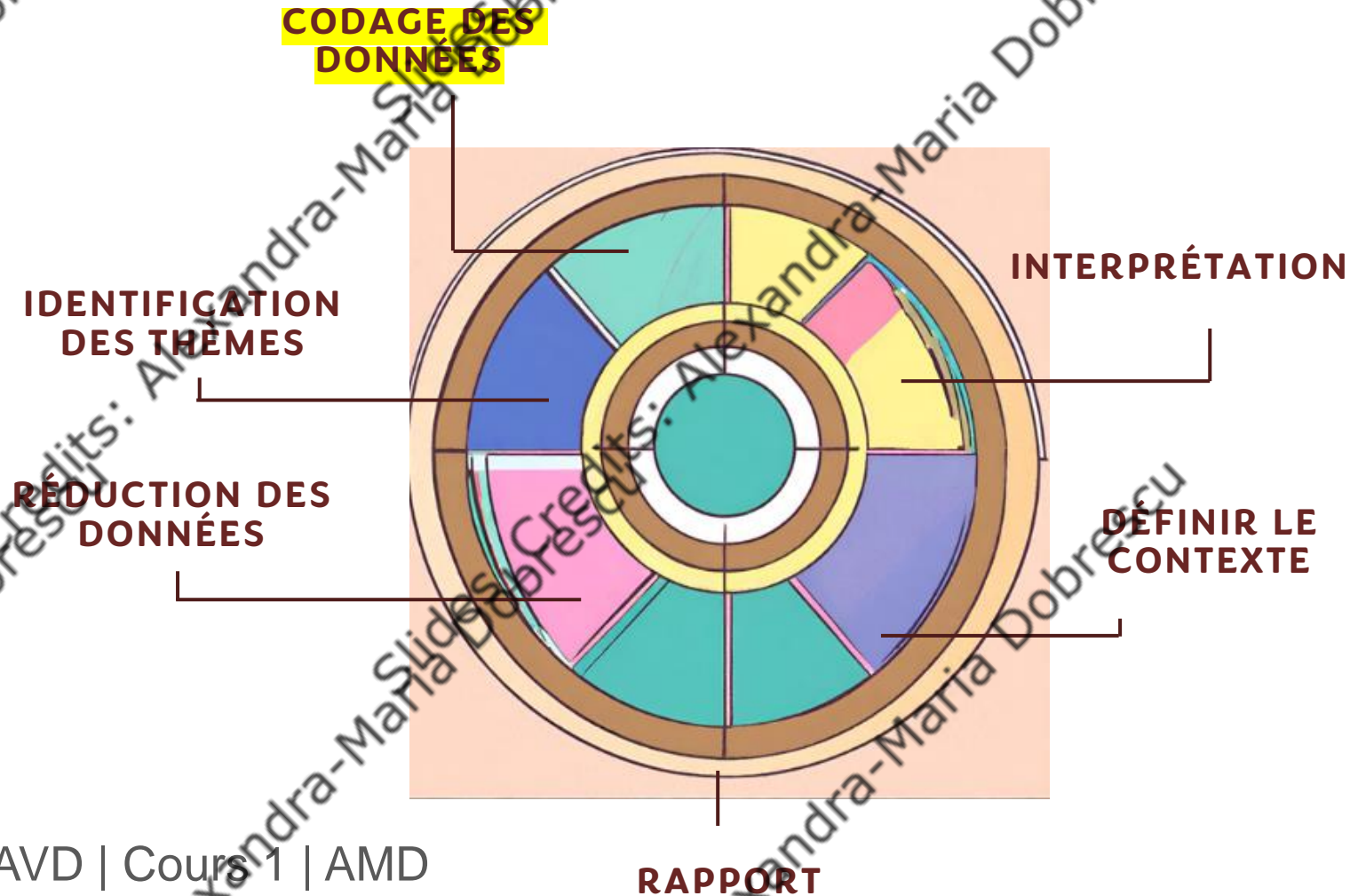


Aprofondir: Types d'analyse des données

QUALITATIVE

Comment ?

Catégorisation et étiquetage de segments de données qualitatives (par exemple, des extraits de texte ou des réponses à des entretiens) sur la base de thèmes, de concepts ou d'idées communs.



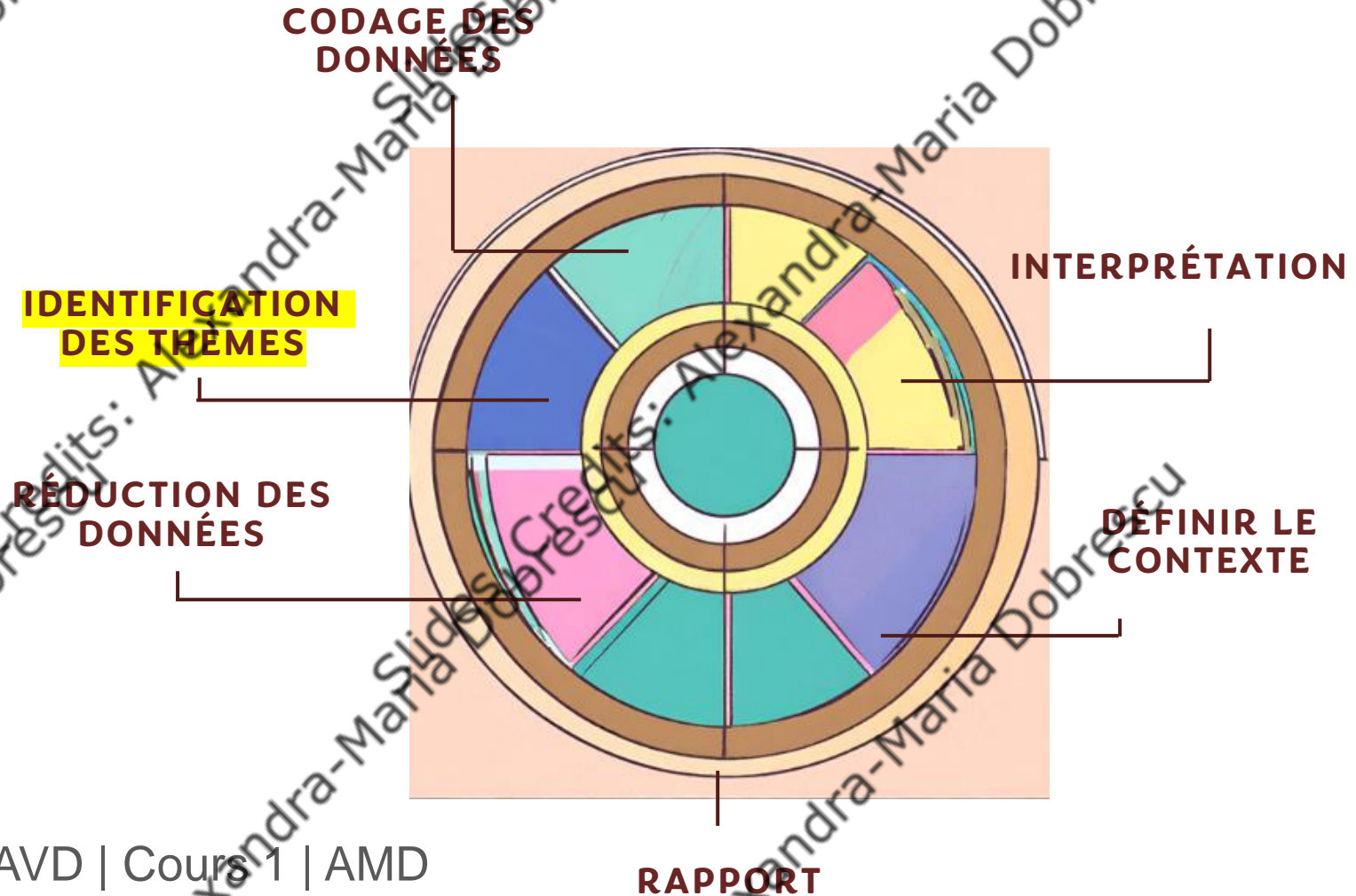
Aprofondir: Types d'analyse des données

QUALITATIVE

Objectif ? Mieux comprendre les questions ou les objectifs de la recherche.

Comment ?

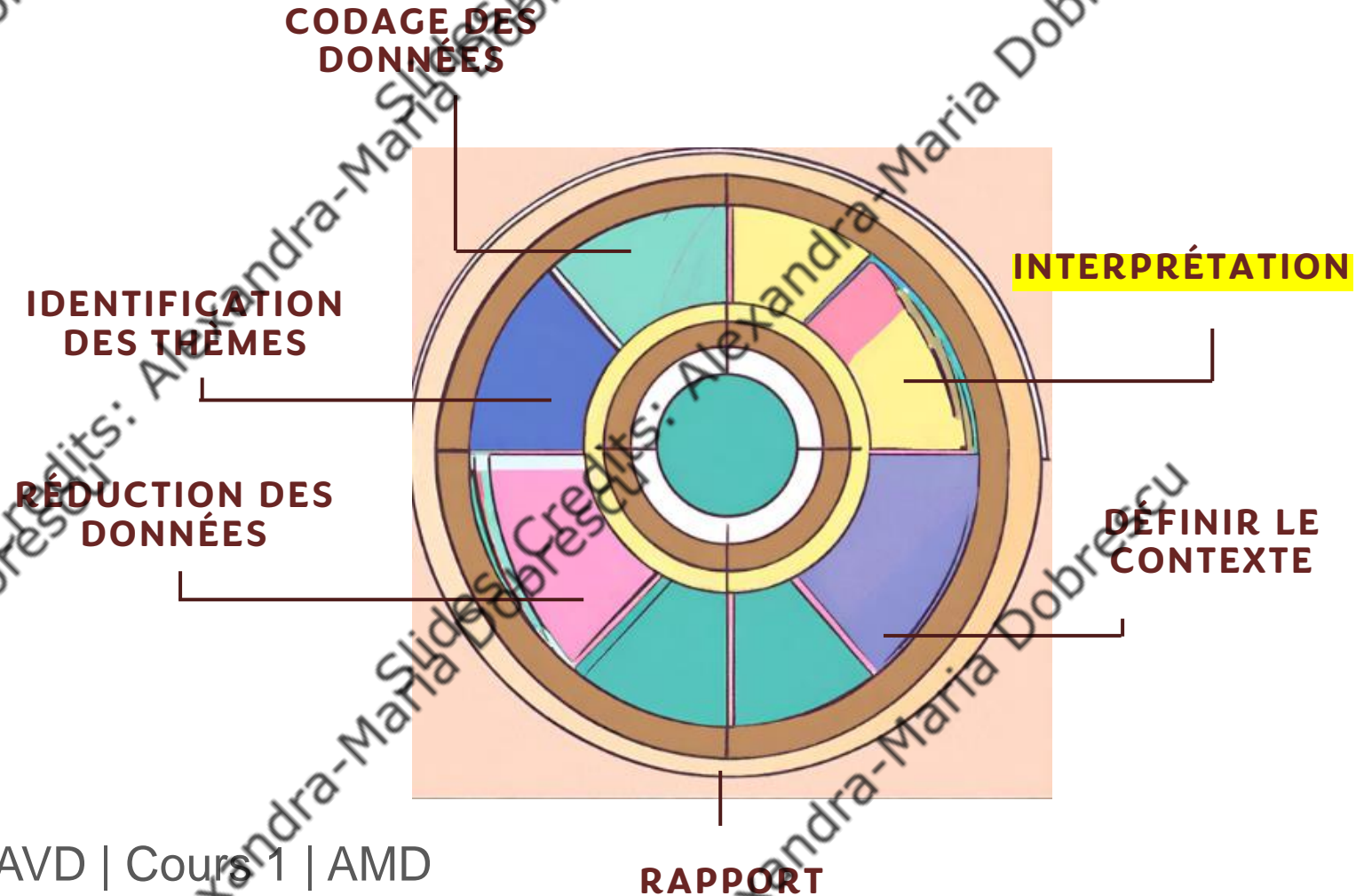
Identification des thèmes récurrents ou tendances dans les données qui permettent de mieux comprendre les questions ou les objectifs de la recherche.



Aprofondir: Types d'analyse des données

QUALITATIVE

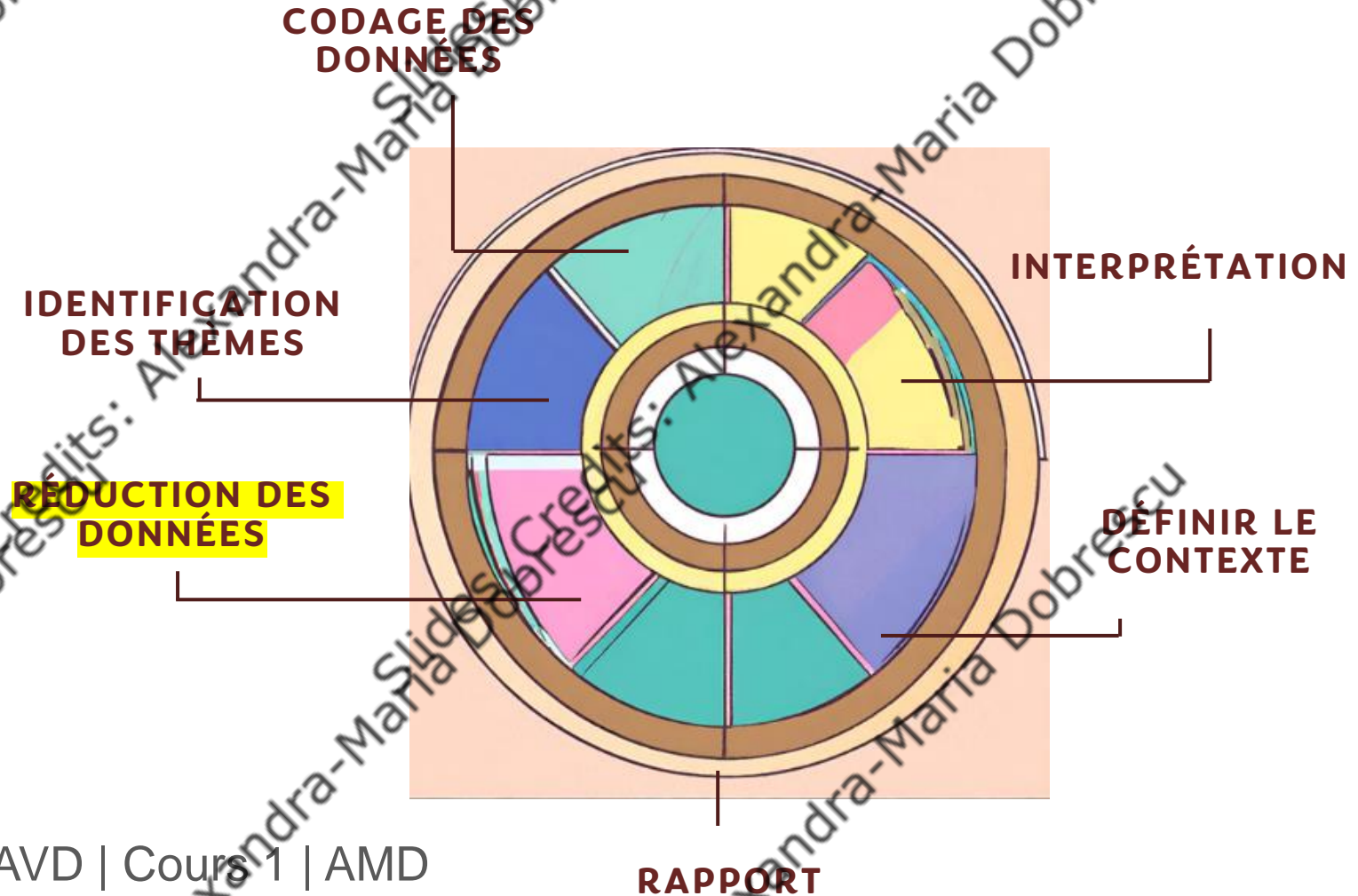
Objectif ? Mieux comprendre les phénomènes étudiés et de développer des idées ou des théories.



Aprofondir: Types d'analyse des données

QUALITATIVE

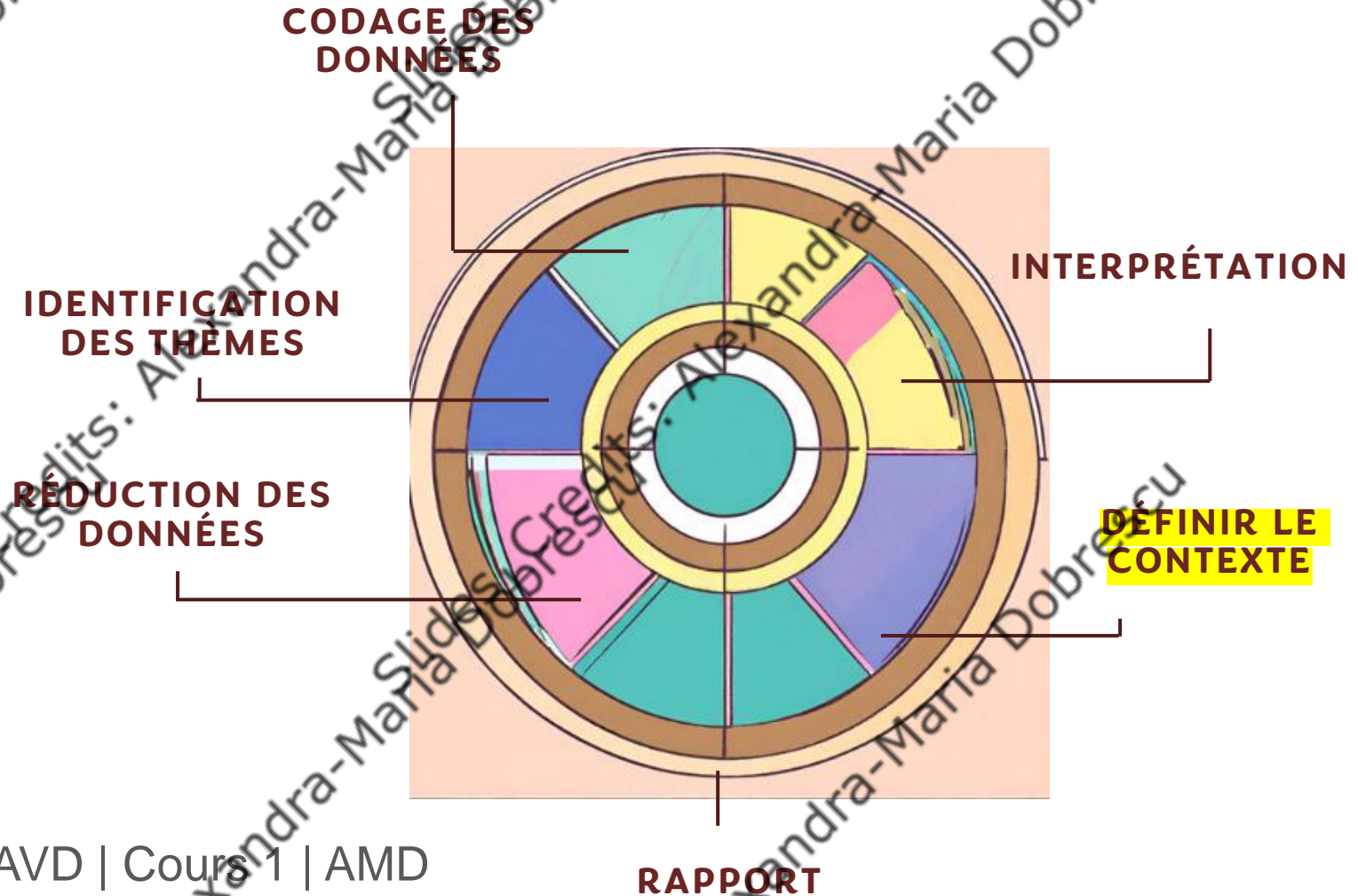
Objectif ? Condenser et résumer de grands volumes de données qualitatives afin d'en extraire des résultats significatifs et gérables.



Aprofondir: Types d'analyse des données

QUALITATIVE

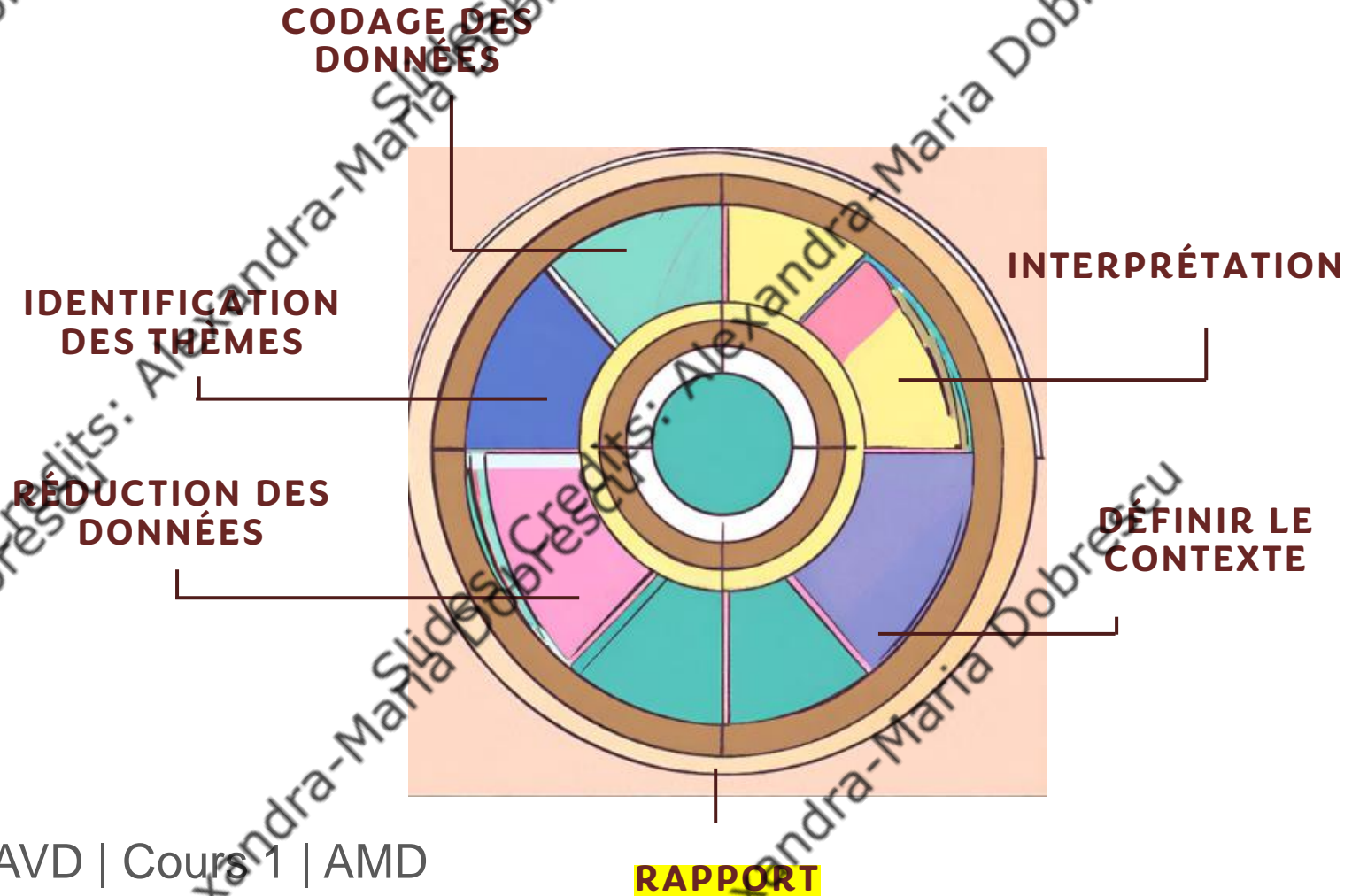
Objectif ? Prise en compte du contexte général et des circonstances spécifiques entourant les données afin de fournir une compréhension nuancée du sujet.



Aprofondir: Types d'analyse des données

QUALITATIVE

Objectif ? Présenter les résultats de l'analyse des données qualitatives de manière cohérente et organisée, souvent au moyen de récits, de citations et de résumés thématiques.



Bibliographie

- [1] Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. Computing in science & engineering, 9(03), 90-95.
- [2] NASA Earth Observations. Our mission. URL: <https://neo.sci.gsfc.nasa.gov/about/>. (accessed: 20.08.2022).
- [3] https://en.wikipedia.org/wiki/Scatter_plot#/media/File:Matriz_de_gr%C3%A1ficos_de_dispers%C3%A3o.svg
- [4] Knafllic, C. N. (2015). Storytelling with data: A data visualization guide for business professionals. John Wiley & Sons.