

Nom de la SAE	Intégration de données dans un datawarehouse		semestre / Période	Semestre 3
volume horaire consacré par l'étudiant	avec enseignant	3h	en autonomie	12h
coéquipiers :	Samuel Darmalingon		Enzo Leriche	

Sujet spécifique	Construction d'indicateurs .
Objectifs	<ul style="list-style-type: none"> • Faire comprendre à l'étudiant l'apport de l'entrepôt de données dans un environnement décisionnel en mettant en place et en construisant des indicateurs pertinents • L'amener à réfléchir à la structuration des données nécessaire • Introduire l'alimentation d'un entrepôt de données par un ETL

<p>Livrables</p>	<div><div>Olist a Brazilian e-commerce</div><div>Enzo LERICHE, Samuel DARMALINGON, Ayoub ERRAHMANI</div><div><ul style="list-style-type: none">1 Importation des différentes tables<ul style="list-style-type: none">1.1 Principal Dataset1.2 Other Datasets2 Analyses Statistiques<ul style="list-style-type: none">2.1 Statistiques descriptives2.2 Représentation graphique2.3 Valeurs manquantes<ul style="list-style-type: none">2.3.1 Nombre de valeurs manquantes2.3.2 Représentation du nombre de valeurs manquantes3 Jointure4 Indicateurs</div></div>
------------------	--

1 Importation des différentes tables

1.1 Principal Dataset

Orders dataset

- Order ID
- Customer ID
- Order Status
- Order Purchase Timestamp
- Order Approved at
- Order Delivered Carrier date
- Order Delivered Customer date
- Order Estimated delivered date

Orders dataset

order_id	customer_id	order_status	order_purchase_timestamp	order_approved_at	order_deliver
e48f51c8de54678b7cc49136f2d6af7	9ef432eb6251297304e76186b10a928d	delivered	2017-10-02 10:56:33	2017-10-02 11:07:15	2017-10-04 19:5
53c0b2fc8bc7de0b6741e2150273451	b0830fb4747a6c6d20dea0b8c802d7ef	delivered	2018-07-24 20:41:37	2018-07-26 03:24:27	2018-07-26 14:5
47770eb9100c2doc44946d9cf07ec65d	41ce2a54c0b03bf3443c3d931a367089	delivered	2018-08-08 08:38:49	2018-08-08 08:55:23	2018-08-08 13:3
949d5b44dbf5de918fe9c1697b45f8a	f88197465ea7920adcdbec7375364d82	delivered	2017-11-18 19:28:06	2017-11-18 19:45:59	2017-11-22 13:3
ad21c59c0840e6cb83agceb5573f8159	8ab97904e6daea8866dbdc4fb7aad2c	delivered	2018-02-13 21:18:39	2018-02-13 22:20:29	2018-02-14 19:4
a4591c265e18cbidcee52889e2d8acc3	503740e9ca751ccdda7ba28e9ab8f608	delivered	2017-07-09 21:57:05	2017-07-09 22:10:13	2017-07-11 14:5

2.2 Représentation graphique

2.3 Valeurs manquantes

Ensuite nous avons regarder si il y avaient des valeurs maquantes dans les différentes tables.

2.3.1 Nombre de valeurs manquantes

Code

##	Table	nb_na
## 1	Orders	0
## 2	Customers	0
## 3	Geolocation	0
## 4	Items	0
## 5	Payments	0
## 6	Pdttrans	0
## 7	Products	1838
## 8	Reviews	0
## 9	Sellers	0

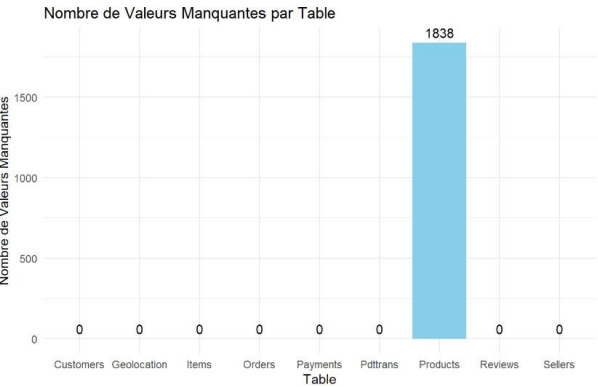
Ainsi, nous constatons qu'il y a aucunes valeurs maquantes dans les différentes tables. A part, pour la table de "Product" qui contient 1838 valeurs maquantes.

2.3.2 Représentation du nombre de valeurs manquantes

2.3.2.1 Nombre de valeurs manquantes par table

2.3.2.2 Colonnes des valeurs manquantes dans la table Products

Code



Nous pouvons l'observer sous la forme d'un graphique en barres qui montre le nombre de valeurs manquantes dans chaque

3 Jointure

Avant de créer les différents indicateurs, il est crucial de consolider les tables en effectuant des jointures. Cette étape fondamentale permet de regrouper des données dispersées au sein de plusieurs tables, simplifiant ainsi l'intégration d'informations, l'établissement de relations logiques, la création de variables composites, et la contextualisation des données. Les jointures optimisent le processus d'analyse en assurant une connexion préalable entre les données, ce qui favorise la création d'indicateurs pertinents et permet une analyse approfondie des informations consolidées

- 3.1 Orders avec Items
- 3.2 Reviews avec Items et Products
- 3.3 Items avec Products
- 3.4 Payments avec Items
- 3.5 Orders avec Customers et Geolocation
- 3.6 Reviews avec Orders
- 3.7 Payments avec Items et Products
- 3.8 Items avec Products et Sellers

Cette jointure servira à créer l'indicateur "Nombre moyen d'articles par commande".

Code

Jointure Orders avec Items

order_id	customer_id	order_status	order_purchase_timestamp	order_approved
00010242fe8c5a6d1ba2dd792cb16214	3ce436f183e68e07877b285a838db11a	delivered	2017-09-13 08:59:02	2017-09-13 09:43
00018f7f2f0320c557190d7a144bdd3	f6dd3ce061db4e3987629fe6b26e5cee	delivered	2017-04-26 10:53:06	2017-04-26 11:05
000229ec398224ef6ca0657da4fc703e	6480ae5e4333f3693df5ad4372dab6d3	delivered	2018-01-14 14:33:31	2018-01-14 14:48
00024acbcdfoa6daa1e931b038114c75	d4eb9395c8c0431ee92fce09860c5a06	delivered	2018-08-08 10:00:35	2018-08-08 10:10
00042b26cf59d7ce69dfabb4e55b4fd9	58dbdb2d70206bf40e62cd34e84d795	delivered	2017-02-04 13:57:51	2017-02-04 14:10
00048cc3ae777c65dbb7d2a0634bce1ea	816cbea060fe5b680b30cfc07a506742	delivered	2017-05-15 21:42:34	2017-05-17 03:55

4.9 Calculer le nombre total de produits vendus par vendeur 4.10 Calculer la moyenne du temps de livraison

4.1.1 Explication de l'indicateur

Pertinence : Cet indicateur mesure la quantité moyenne d'articles inclus dans chaque commande. Il offre un aperçu de la taille typique des paniers d'achat, ce qui peut être crucial pour comprendre les habitudes d'achat des clients et orienter les décisions stratégiques en matière de gestion des stocks, de promotions, et de l'expérience client.

Méthodologie de calcul : Pour obtenir ce nombre moyen, nous divisons le total des articles par le nombre total de commandes.

Résultats attendus : Un chiffre plus élevé suggère généralement un panier d'achat plus important, tandis qu'un chiffre plus bas indique une tendance à des achats plus restreints. Ces informations peuvent influencer les stratégies de vente croisée, de promotions groupées, et la planification de l'inventaire.

4.1.2 Calcul de l'indicateur

Voici comment nous calculons l'indicateur, nous vous invitons à cliquer sur code.

Code

4.1.3 Représentation graphique

Voici la représentation graphique de l'indicateur

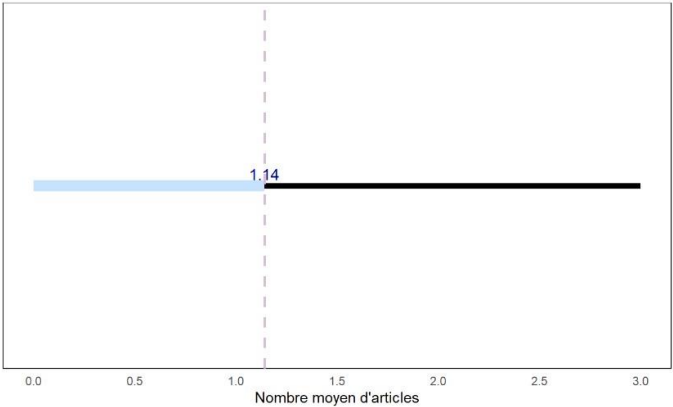
Code

Voici la représentation graphique de l'indicateur

C

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

Nombre moyen d'articles par commande



En moyenne, il y a 1,14 articles par commande.

Bilan de la SAE

(reproduire le tableau autant de fois que de compétences mobilisées dans la SAÉ)

Compétence	Traiter des données à des fins décisionnelles
Apprentissages critiques sollicités	Comprendre l'organisation des données de l'entreprise
	Réaliser le rôle central et spécifique de l'entrepôt de données dans la chaîne décisionnelle
	Comprendre la nécessité de tester, corriger et documenter un programme
Composantes essentielles à respecter	En s'inscrivant dans une démarche de documentation des réalisations adaptée au public visé
	En écrivant un programme correctement structuré et documenté, respectant les bonnes pratiques
	En traduisant correctement les demandes métier en programmes, avec le respect du cahier des charges s'il existe

Ma démarche

Savoirs / connaissances	Savoir-faire	Savoir-être
Construction d'indicateurs pertinents.	<ul style="list-style-type: none">• Traiter• Valoriser	<ul style="list-style-type: none">• Collaboration en équipe

Evaluation du résultat

- Ce que je trouve bien réalisé, pourquoi ?

Ce que je trouve bien réaliser pour moi sont les graphiques concernant les différents indicateurs construits grâce à la base de données de plus un dictionnaire de données a été sollicité et réalisé pour mener à bien ce projet .

- Ce que je n'ai pas bien compris ; ce qui serait à améliorer pour une prochaine fois :
Pourquoi ? comment ?

Il fallait mieux organiser notre rendu car on se rend compte qu'on a du mal a ce orienter vers les pages.

Eléments de preuve, ce que je peux montrer

(Choisir des éléments précis à mettre annexe)

- 1) extrait de code pour la construction d'un indicateur

Fiche bilan SAE

```
## Nombre de vendeurs et de clients

### Explication de l'indicateur
*Pertinence :* Cet indicateur mesure le nombre total de vendeurs distincts et de clients distincts le système. Il offre un aperçu du volume de l'écosystème de vente et de l'étendue de la clientèle.

*Méthodologie de calcul :* Pour obtenir le nombre de vendeurs, comptez le nombre d'identifiants uniques de vendeurs dans votre ensemble de données. Pour le nombre de clients, comptez le nombre d'identifiants uniques de clients.

### Calcul de l'indicateur
Voici comment nous calculons l'indicateur, nous vous invitons à cliquer sur code.


```
{r}
nombre_total_vendeurs <- length(unique(Sellers$seller_id))
nombre_total_clients <- length(unique(Customers$customer_id))
```



### Représentation graphique
Voici la représentation graphique de l'indicateur


```
{r}
ggplot() +
 geom_text(aes(x = 0.5, y = 0.5, label = paste("Nombre total de vendeurs :",
round(nombre_total_vendeurs))),
```


```

2) exemple de graphique réalisé sur un indicateur

Nombre moyen d'articles par commande

