

ARCHITECTURE HADOOP

BIG DATA

Astrid Aurelien NKUMBE ENONGENE MSc 2 Info IA
Anne-Josée LOUIS MSc 2 Info IA

Sommaire :



1

Introduction

2

Ingestion et Normalisation

3

Enrichissement des données

4

Dashboard

5

Conclusion

Introduction

Dans le cadre de cette étude de cas, nous avons choisi un site de vente en ligne comme domaine d'application. Notre objectif est de mettre en place une plateforme de données globale pour cette entreprise de distribution de produits.

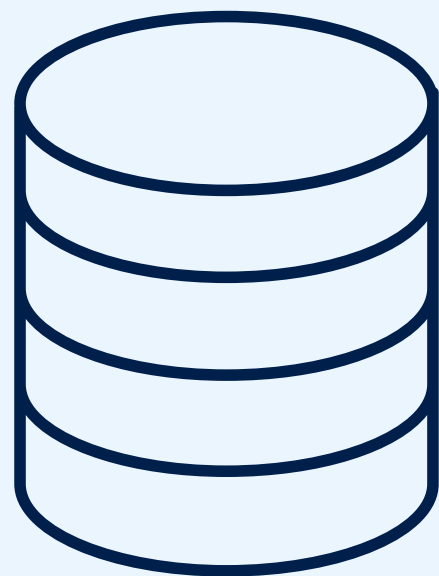


www.reallygreatsite.com





Sources de Données



E-commerce

Clients, commandes et livraisons



ERP (Interne)

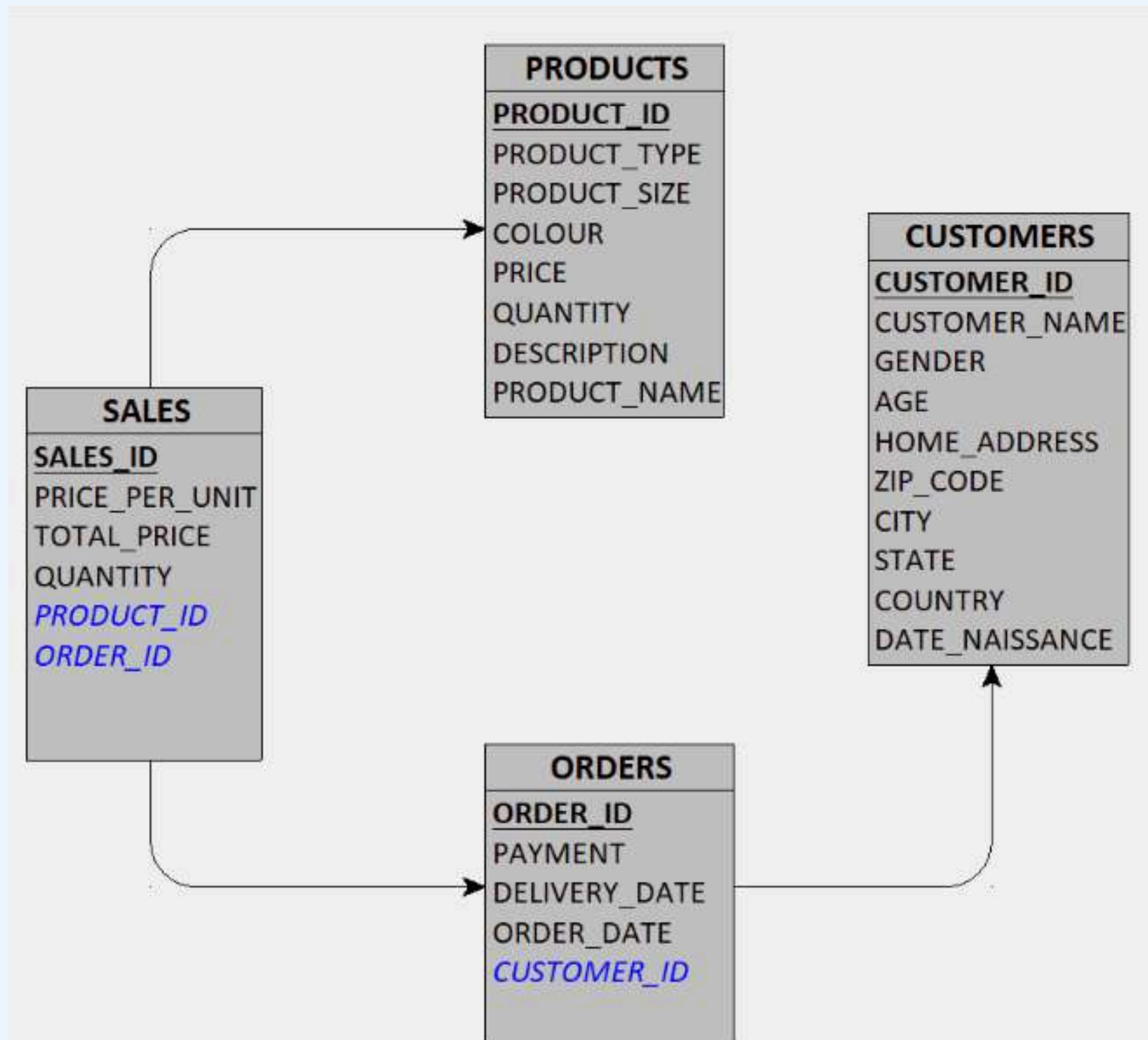
Commandes et les livraisons et feedback des clients



Externe

Informations sur la météo et le trafic

Source de données 1



Source de données 2

erp_orders
<u>order_id</u>
order_date
expected_delivery_date
actual_delivery_date
delivery_status
delivery_cost

customer_feedback
<u>feedback_id</u>
satisfaction_rating
order_id
comment

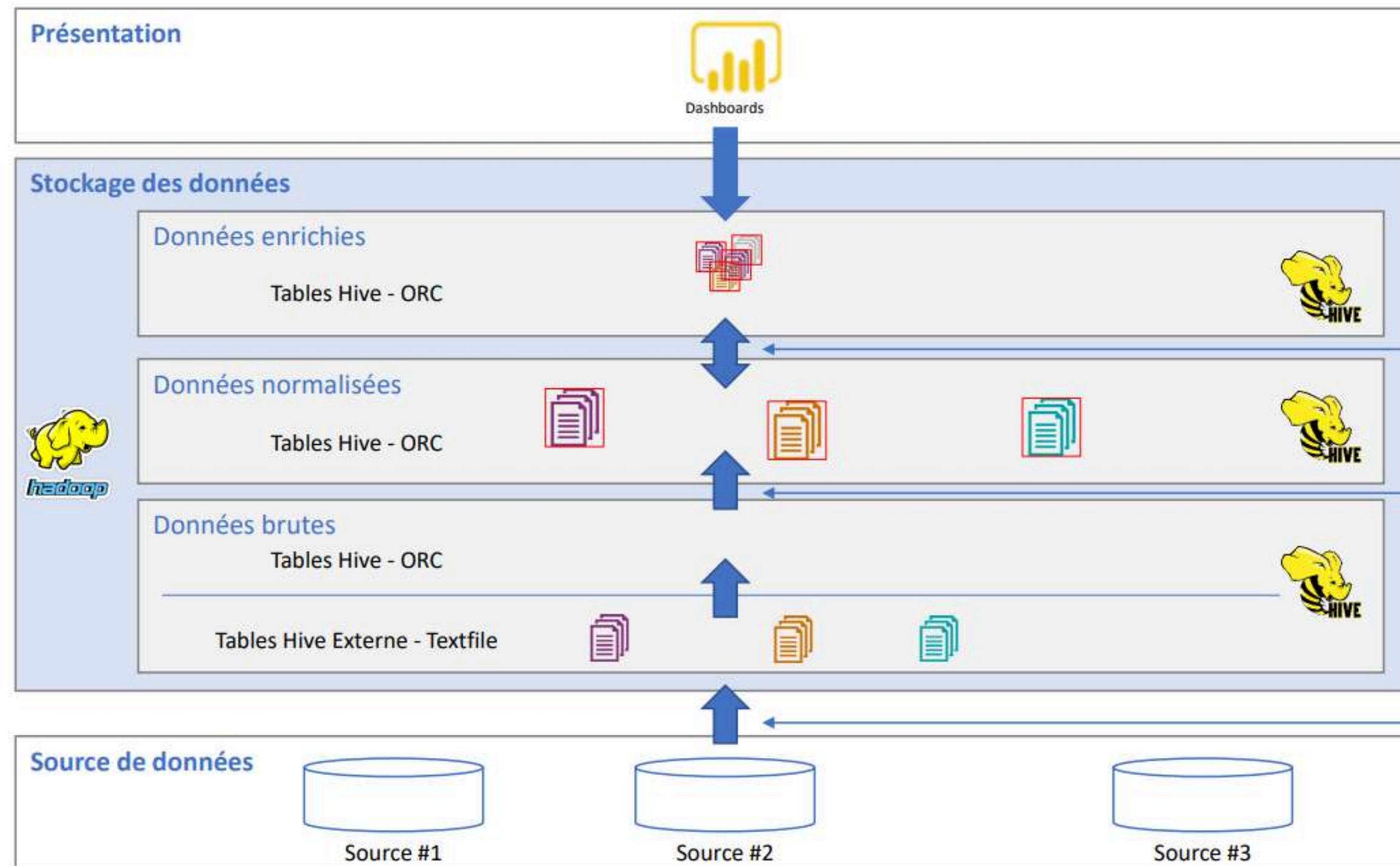


Source de données 3

weather_data
<u>data_id</u>
date_data
weather_condition
traffic_intensity



Organisation du stockage des données sur HDFS



L'utilisation de la
plateforme Cloudera
Quick start est
recommandée

Transformation & enrichissement



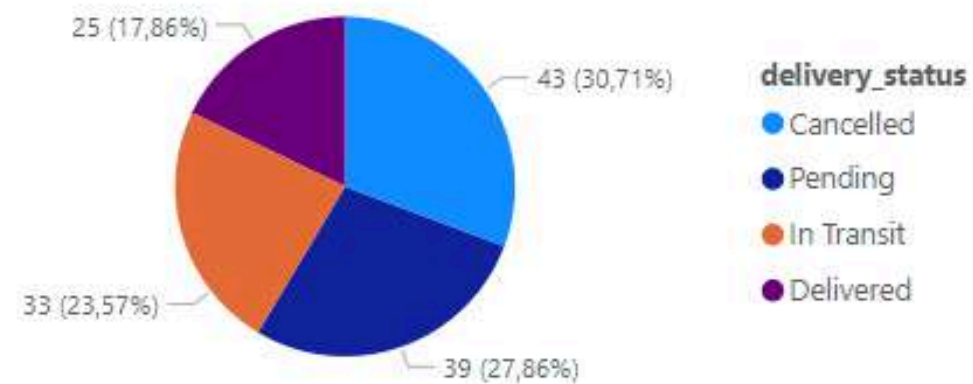
Ingestion



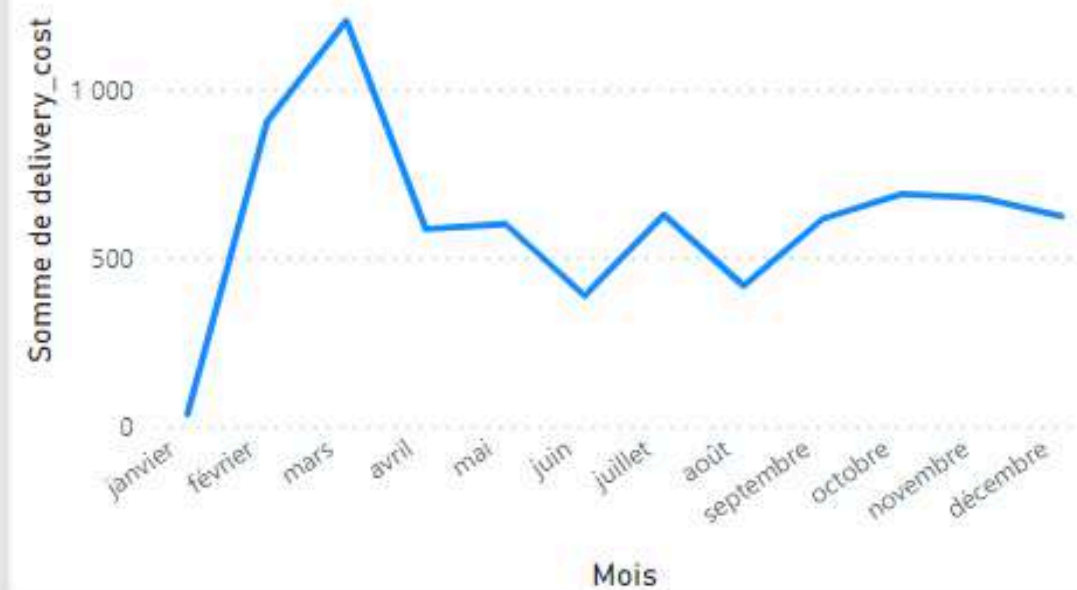
Dashboard

Globale

Nombre de order_date par delivery_status



Somme de delivery_cost par Mois



7 364

Somme de delivery_cost

100

Nombre de order_id

delivery_status	Nombre de order_id	Somme de delivery_cost
Cancelled	31	2 397,92
Delivered	18	1 211,90
In Transit	26	1 785,07
Pending	25	1 969,41
Total	100	7 364,30

Premier Mois par Year

2022

2023

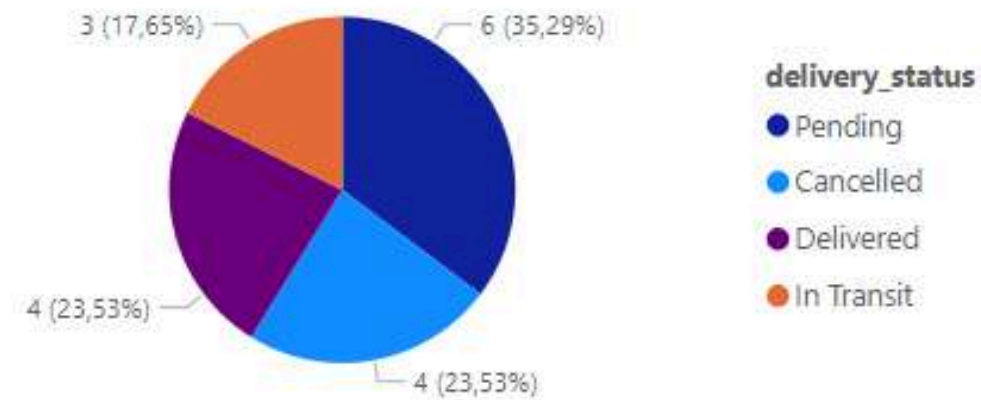
2024



Dashboard

Année 2022

Nombre de order_date par delivery_status



Somme de delivery_cost par Mois



887

Somme de delivery_cost

13

Nombre de order_id

delivery_status	Nombre de order_id	Somme de delivery_cost
Cancelled	4	127,43
Delivered	3	293,89
In Transit	3	123,95
Pending	3	341,35
Total	13	886,62

Premier Mois par Year

2022

2023

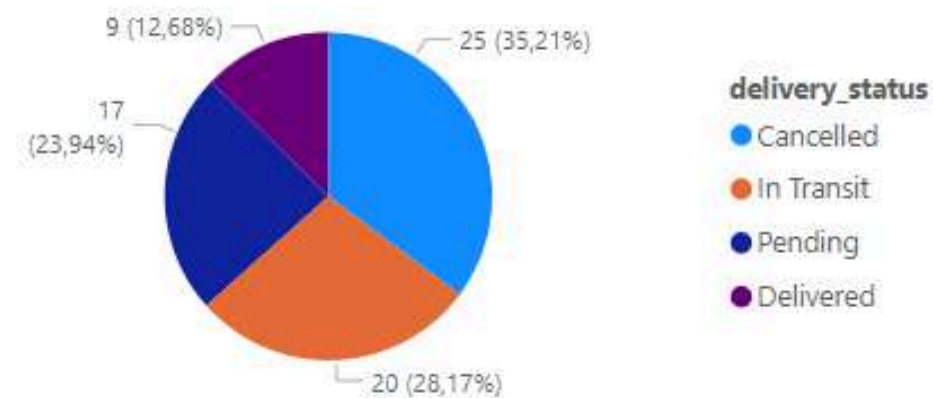
2024



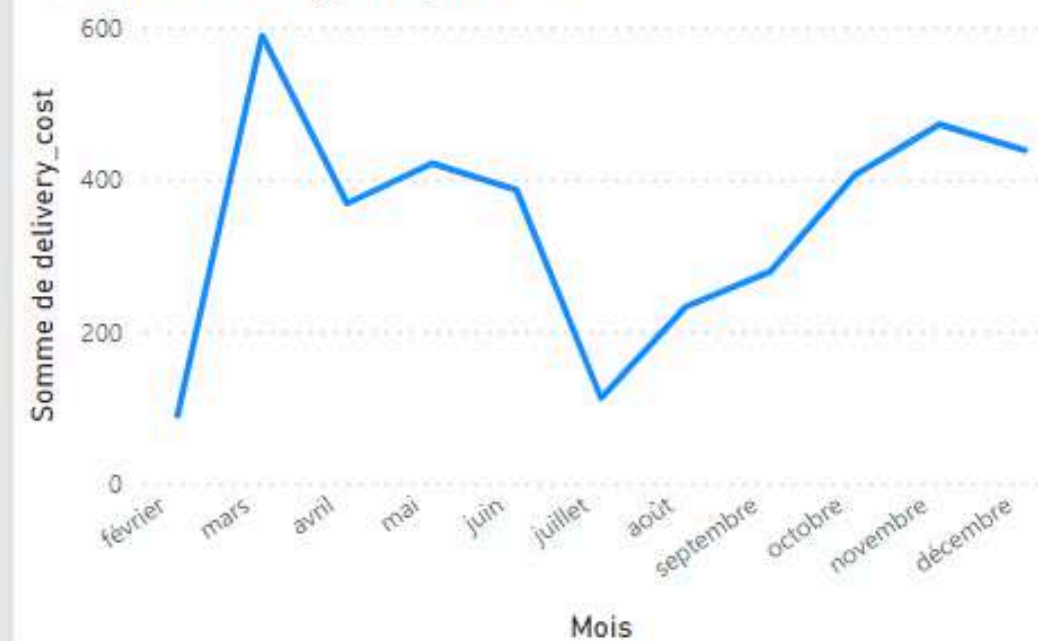
Dashboard

Année 2023

Nombre de order_date par delivery_status



Somme de delivery_cost par Mois



3 797

Somme de delivery_cost

54

Nombre de order_id

delivery_status	Nombre de order_id	Somme de delivery_cost
Cancelled	16	1 401,13
Delivered	7	406,95
In Transit	16	1 095,46
Pending	15	893,69
Total	54	3 797,23

Premier Mois par Year

2022

2023

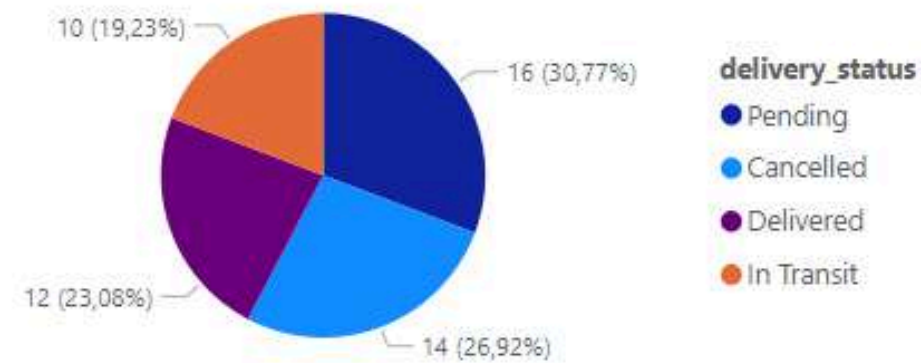
2024



Dashboard

Année 2024

Nombre de order_date par delivery_status



Somme de delivery_cost par Mois



2 680

Somme de delivery_cost

33

Nombre de order_id

delivery_status	Nombre de order_id	Somme de delivery_cost
Cancelled	11	869,36
Delivered	8	511,06
In Transit	7	565,66
Pending	7	734,37
Total	33	2 680,45

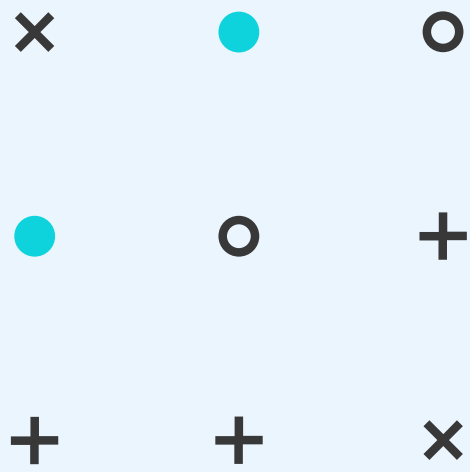
Premier Mois par Year

2022

2023

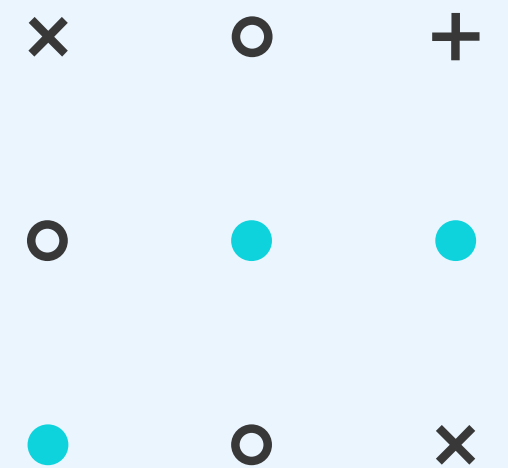
2024





Conclusion

En ingérant et en normalisant les données de plusieurs sources, en enrichissant ces données et en les visualisant à travers un tableau de bord, nous pouvons obtenir des insights précieux qui peuvent aider à la prise de décision et à l'amélioration des performances de l'entreprise.





MERCI

Pour votre attention !