



Université de Caen
Normandie



IUT Grand Ouest Normandie
Pôle de Caen

Bachelor Universitaire de Technologie
Science des Données

Troisième année
Ressource : Base de données NoSQL

Rapport d'analyse des données Olist avec MongoDB



Auteur

Nomel DJAHOUA

Année universitaire 2024-2025

SOMMAIRE

	Page
1 Introduction	3
2 Analyses effectuées et visualisations	3
2.1 Prix moyen des articles par catégorie de produit	3
2.2 Scores moyens par vendeur	4
2.3 Délai moyen de livraison	4
2.4 Les états les plus prolifiques	5
2.5 Délais de livraison les plus courts par État	5
2.6 Catégorie de produit la mieux notée	6
3 Conclusion	6

1 - Introduction

Cet ensemble de données a été fourni par Olist, le plus grand site de e-commerce au Brésil. La plateforme connecte de petites entreprises à travers le pays en leur permettant de vendre leurs produits via Olist Store et d'utiliser leurs partenaires logistiques. Les clients reçoivent une enquête de satisfaction après chaque achat, permettant d'évaluer l'expérience d'achat. Dans ce projet, une migration vers un environnement NoSQL avec MongoDB a été réalisée. Des opérations d'analyse descriptive ont été effectuées pour répondre à des questions clés sur les performances des produits et des vendeurs, les délais de livraison, et les états les plus prolifiques.

Migration des données vers MongoDB

Les 7 fichiers CSV suivants ont été importés dans MongoDB sous forme de collections :

- **olist_customers_dataset.csv**
- **olist_order_items_dataset.csv**
- **olist_order_payments_dataset.csv**
- **olist_order_reviews_dataset.csv**
- **olist_orders_dataset.csv**
- **olist_products_dataset.csv**
- **olist_sellers_dataset.csv**
- **product_category_name_translation.csv**

Chaque collection a été indexée pour optimiser les requêtes. Des opérations de fusion (*merge*) ont permis de relier différentes sources d'informations pour les analyses suivantes.

2 - Analyses effectuées et visualisations

2.1 Prix moyen des articles par catégorie de produit

L'objectif était de déterminer les 10 catégories de produits avec le prix moyen le plus élevé. Les résultats montrent que les catégories comme "computers" et "small_appliances_home_oven_and_coffee" figurent parmi les plus chères.

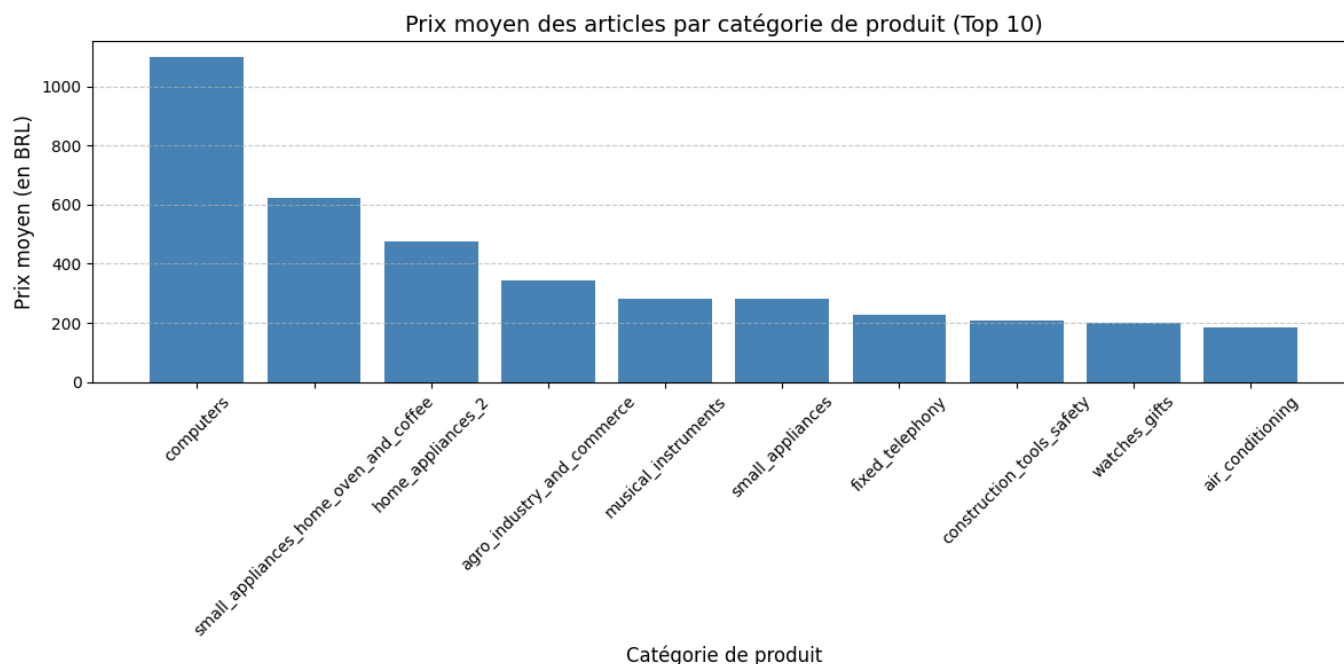


Fig. 1 - Ce graphique montre les catégories de produits avec le prix moyen le plus élevé. La catégorie "computers" se distingue avec un prix moyen supérieur à 1000 BRL.

2.2 Scores moyens par vendeur

Une analyse a été effectuée pour identifier les 10 vendeurs ayant les meilleurs scores moyens de satisfaction donnés par les clients. Les résultats incluent également le nombre de commandes associées.

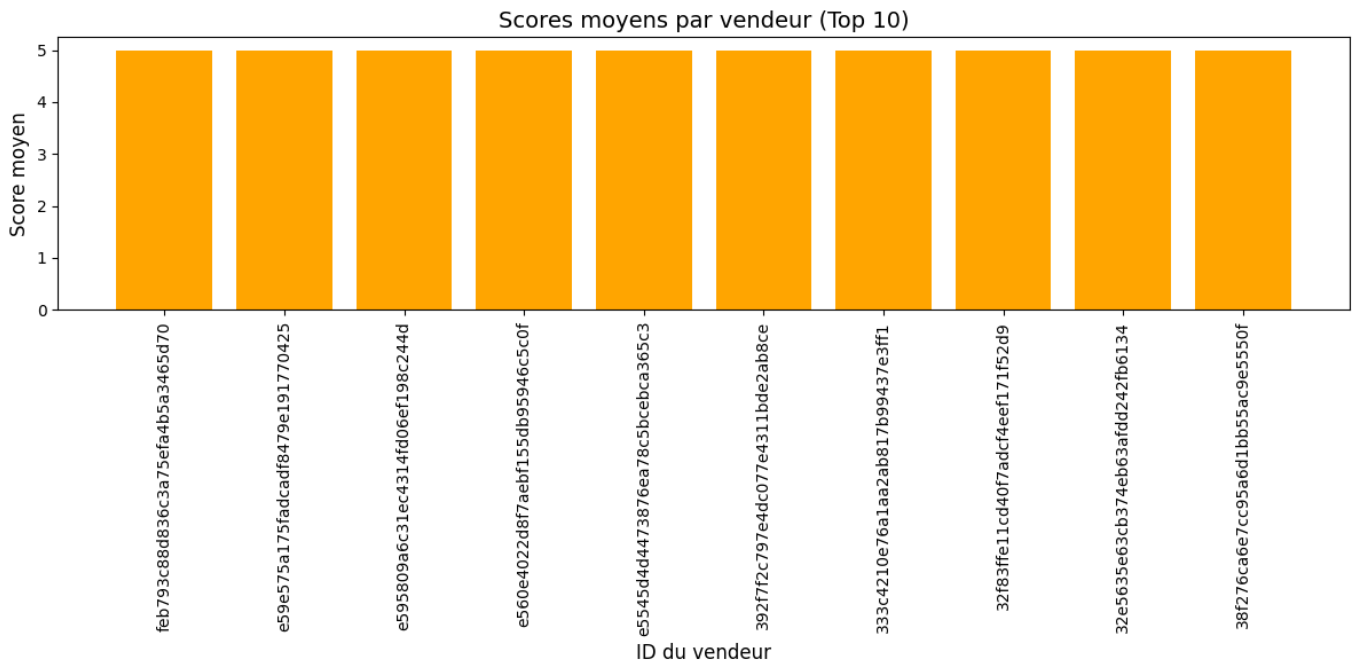
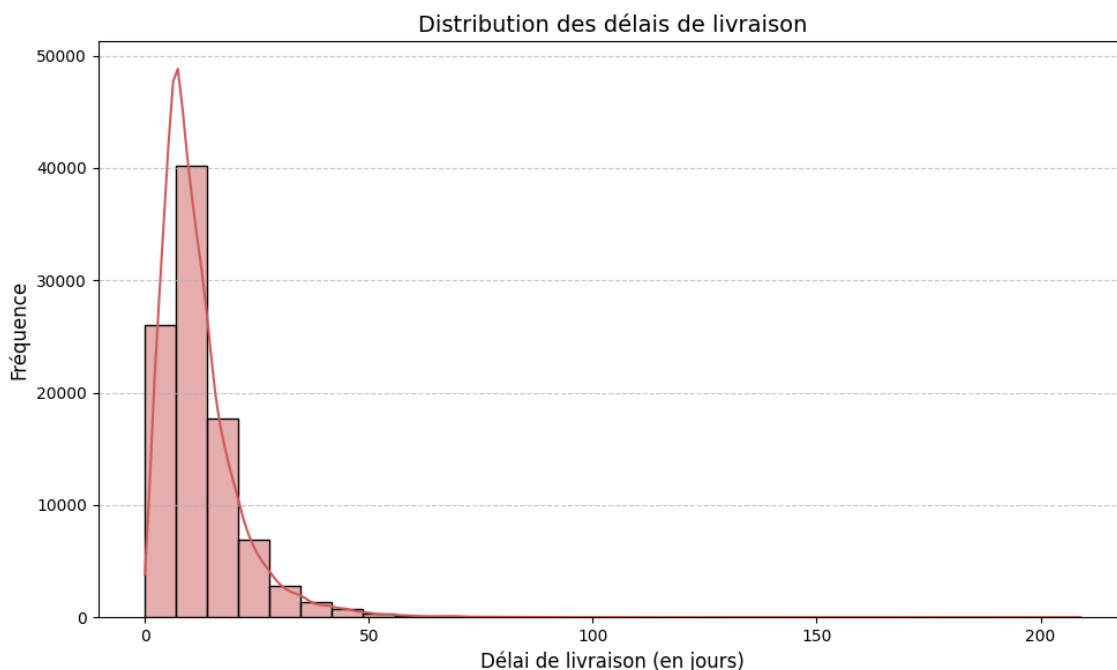


Fig. 2 - Ce graphique présente les vendeurs ayant les scores moyens de satisfaction les plus élevés. Tous les vendeurs affichent un score supérieur à 4,5, indiquant une grande satisfaction des clients.

2.3 Délai moyen de livraison

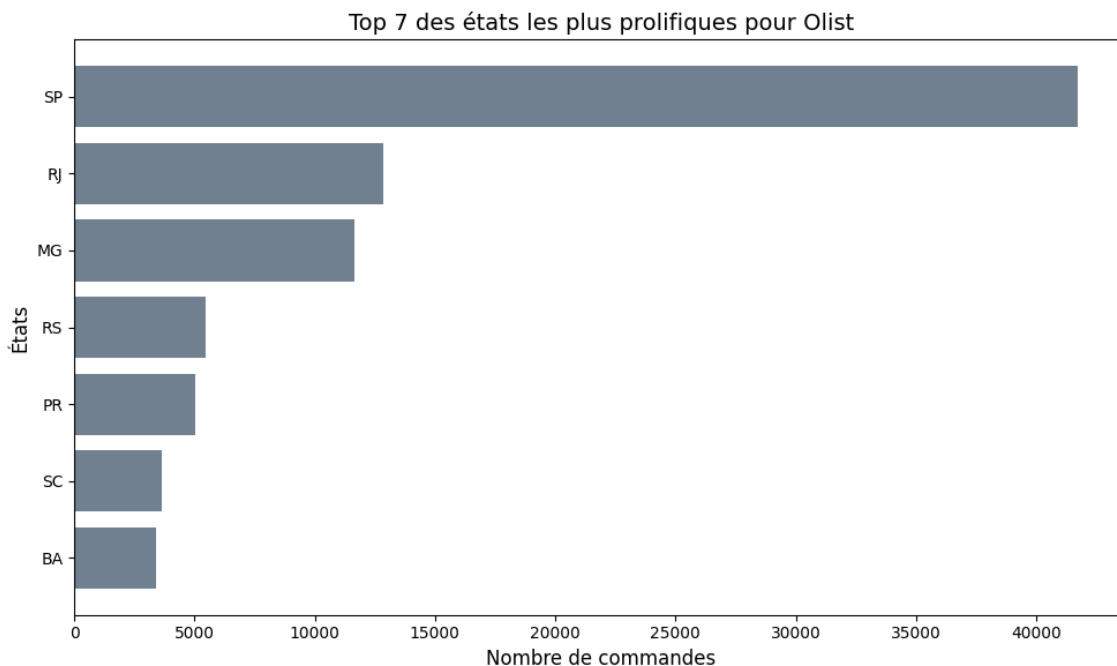
Une nouvelle variable calculant le délai de livraison en jours a été créée. Le délai moyen de livraison chez Olist est estimé à **12 jours**. Une distribution des délais de livraison a été visualisée pour analyser leur répartition.



La distribution montre que la majorité des commandes sont livrées en moins de 20 jours, avec une légère présence de délais plus longs dépassant 50 jours.

2.4 Les états les plus prolifiques

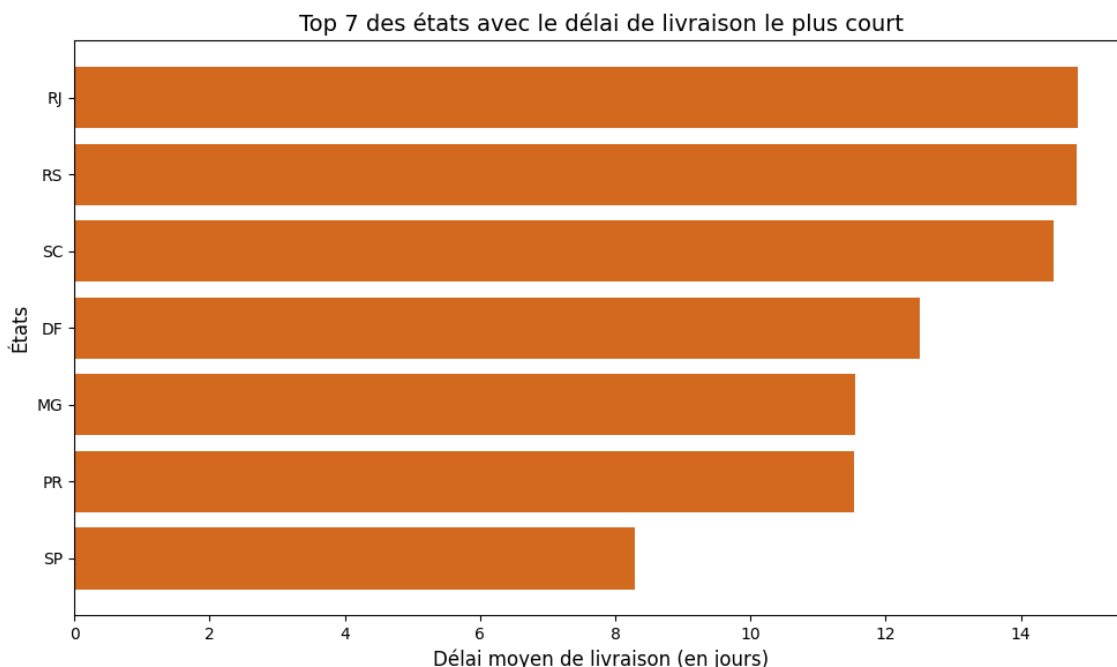
Les 7 états les plus prolifiques en termes de nombre de commandes ont été identifiés. L'État de São Paulo (extitSP) arrive largement en tête.



Ce graphique classe les états selon le nombre de commandes. São Paulo domine avec plus de 40 000 commandes, suivi par Rio de Janeiro et Minas Gerais.

2.5 Délais de livraison les plus courts par État

Les délais moyens de livraison ont été calculés pour chaque État. Les 7 États avec les délais les plus courts incluent Rio de Janeiro (extitRJ) et Rio Grande do Sul (extitRS).



Les États avec les délais moyens de livraison les plus courts varient entre 7 et 10 jours. Rio de Janeiro et Rio Grande do Sul figurent en tête de ce classement

2.6 Catégorie de produit la mieux notée

Les catégories de produits ont été évaluées selon leur score moyen. La catégorie *cds_dvds_musicals* a obtenu le meilleur score moyen.



Ce graphique présente les catégories de produits les mieux notées. "cds_dvds_musicals" arrive en tête avec un score moyen proche de 5, indiquant une très haute satisfaction des clients.

3 - Conclusion

Ce projet a permis de migrer efficacement les données vers MongoDB et de répondre aux questions clés concernant les performances des produits, la satisfaction des clients, et les délais de livraison. Les analyses et visualisations obtenues offrent des informations précieuses pour améliorer l'expérience client chez Olist. Les résultats sont sauvegardés dans MongoDB pour faciliter les analyses futures et améliorer la stratégie commerciale d'Olist.