

Compte Rendu Lab 4

Initiation à MongoDB

Nom et Prénom : AMRANI Hamza

Filière : IDF

I. Manipulation avec Python et PyMongo

1. Installation et Connexion

installer PyMongo :

```
!pip install pymongo
```

Se connecter à MongoDB Atlas :

```
Connected successfully! Found 1000 transactions
```

2. Agrégation avec Python

```
Transaction Totals by Type:
```

```
-----  
Transaction Type: Deposit  
Total Amount: $252,042.62
```

```
-----  
Transaction Type: Transfer  
Total Amount: $291,776.55
```

```
-----  
Transaction Type: Withdrawal  
Total Amount: $227,346.12
```

3. Autres Opérations d'Agrégation

Montant moyen par type de transaction

Top 5 des comptes par montant envoyé

Analyse de fraude par type de transaction

```
Average Transaction Amount by Type:
```

```
Deposit: $797.60 (Count: 316)  
Transfer: $780.15 (Count: 374)  
Withdrawal: $733.37 (Count: 310)
```

```
Top 5 Accounts by Amount Sent:
```

```
1. Account ACC37810: $2,757.77 (2 transactions)  
2. Account ACC75741: $1,918.82 (2 transactions)  
3. Account ACC89865: $1,692.55 (2 transactions)  
4. Account ACC44804: $1,497.76 (1 transactions)  
5. Account ACC67128: $1,495.01 (1 transactions)
```

```
Fraud Analysis by Transaction Type:
```

```
Deposit: 167 fraudulent transactions, Total: $130,061.30  
Transfer: 170 fraudulent transactions, Total: $135,385.93  
Withdrawal: 144 fraudulent transactions, Total: $113,416.67
```

4. Code Complet et Détails

Pour voir l'implémentation complète de toutes les analyses, agrégations et visualisations réalisées avec Python et PyMongo, veuillez consulter le fichier Jupyter Notebook :

`MapReduce.ipynb`

Ce notebook contient :

- La connexion complète à MongoDB Atlas
- Toutes les requêtes d'agrégation détaillées
- Les analyses statistiques avancées
- Les visualisations de données avec Matplotlib/Seaborn
- Les opérations MapReduce
- Le code source complet avec commentaires

5. Conclusion

PyMongo permet de manipuler facilement les données MongoDB depuis Python. Les pipelines d'agrégation offrent des capacités d'analyse puissantes pour :

- Calculer des statistiques (sommes, moyennes, comptages)
- Filtrer et trier les données
- Grouper par différents critères
- Identifier les tendances et anomalies (fraudes)

L'intégration avec Python permet également de créer des visualisations et d'automatiser les analyses, comme démontré dans le notebook `MapReduce.ipynb`.

II. Dashboard avec MongoDB Charts

1. Activation de MongoDB Charts

a) Étapes d'activation

1. Sélectionner le cluster
2. Cliquer sur “Charts”
3. Créer un nouveau tableau de bord

b) Connexion de la Source de Données

1. Cliquer sur “Add Data Source”
2. Sélectionner la base de données BankDB
3. Sélectionner la collection `transactions`

2. Crédit des Visualisations

Le dashboard final contient 9 visualisations organisées en 3 sections : KPIs, Visualisations Principales, et Analyses Détaillées.

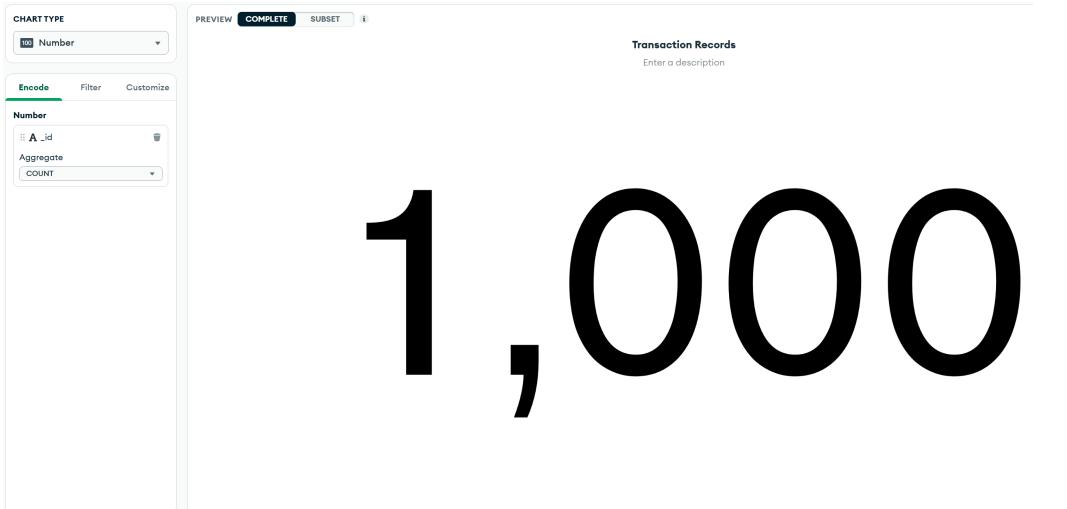
a) Section 1 : KPIs (Key Performance Indicators)

1. Transaction Records Agrégation utilisée : COUNT

Compte le nombre total de documents (transactions) dans la collection.

Pipeline MongoDB équivalent :

```
db.transactions.aggregate([
  { $count: "total" }
])
```

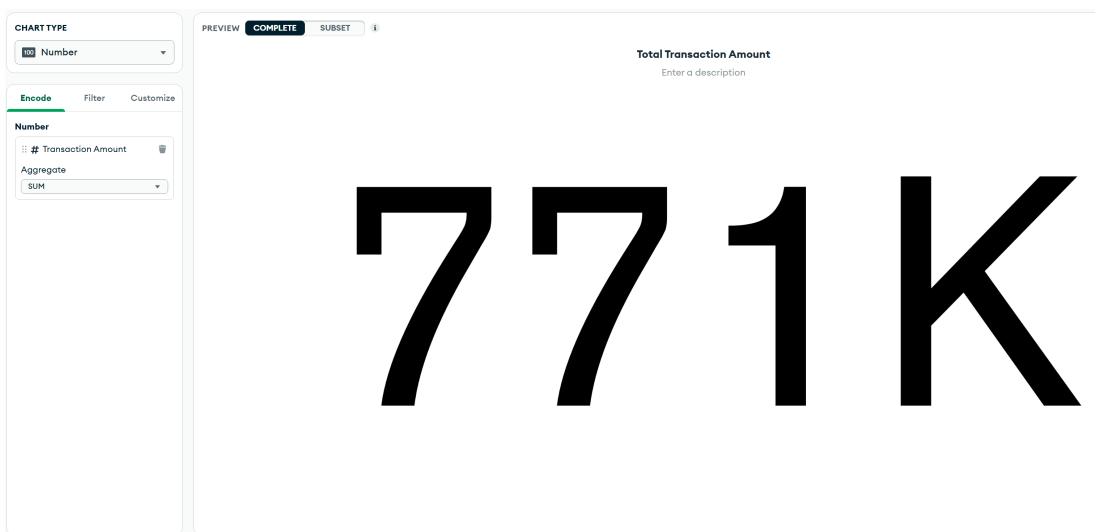


2. Total Transaction Amount Agrégation utilisée : SUM

Additionne tous les montants de transactions pour obtenir le total général.

Pipeline MongoDB équivalent :

```
db.transactions.aggregate([
  {$group: { _id: null,
    totalAmount: { $sum: "$Transaction Amount" } } } ])
```

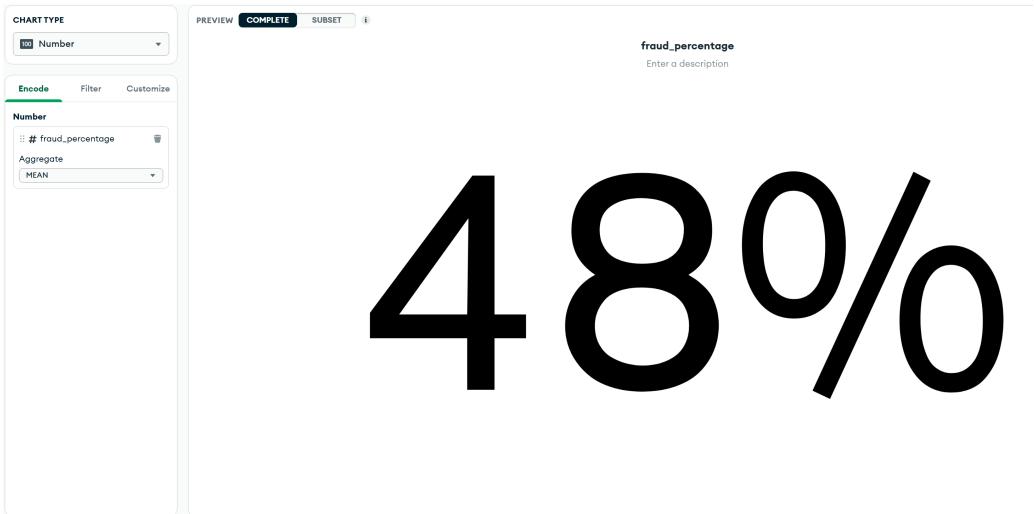


3. fraud_percentage Agrégation utilisée : MEAN (AVERAGE)

Le champ calculé transforme chaque transaction : si “Fraud Flag” = “True” → 100, sinon → 0. La moyenne de ces valeurs donne le pourcentage de fraudes.

Champ calculé :

```
{$cond: {$eq: ["$Fraud Flag", "True"]}, 100, 0]}
```

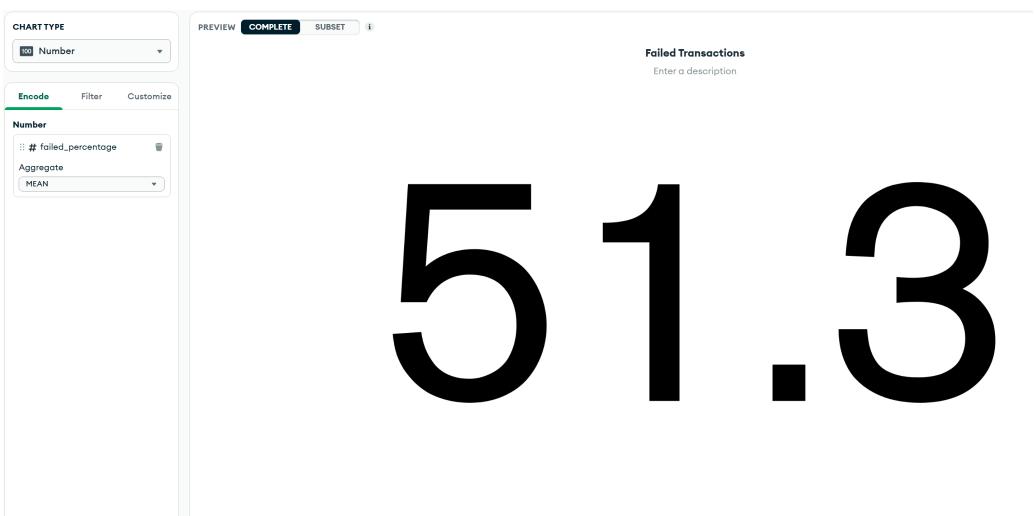


4. Failed Transactions Agrégation utilisée : MEAN (AVERAGE)

Calcule le pourcentage de transactions échouées en utilisant la même logique que pour le pourcentage de fraudes.

Champ calculé :

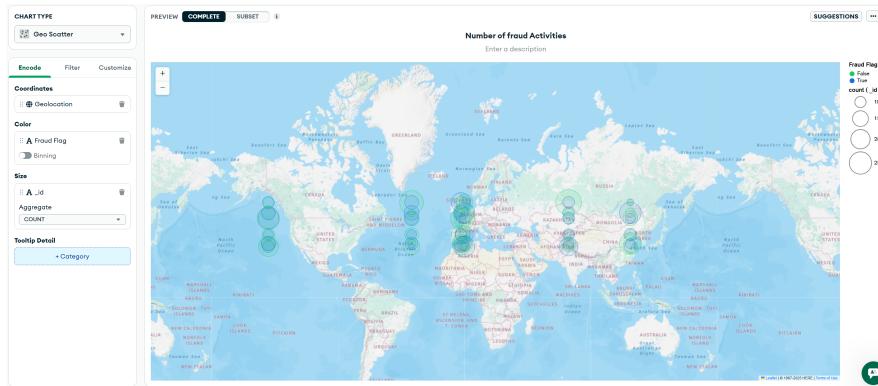
```
{$cond: {$eq: ["$Transaction Status", "Failed"]}, 100, 0]}
```



b) Section 2 : Visualisations Principales

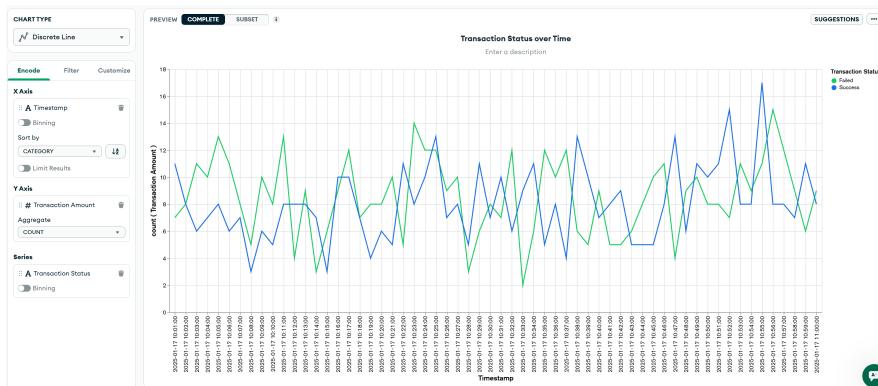
5. Number of Fraud Activities Agrégation utilisée : COUNT

Affiche un point pour chaque localisation géographique. La taille du cercle représente le nombre de transactions (COUNT). La couleur indique si c'est une fraude (True = bleu/vert) ou non (False = gris).



6. Transaction Status over Time Agrégation utilisée : COUNT

Compte le nombre de transactions par intervalle de temps, groupé par "Transaction Status" pour créer deux lignes distinctes (Failed = vert, Success = bleu).



7. Number of Fraud By Device Agrégation utilisée : COUNT

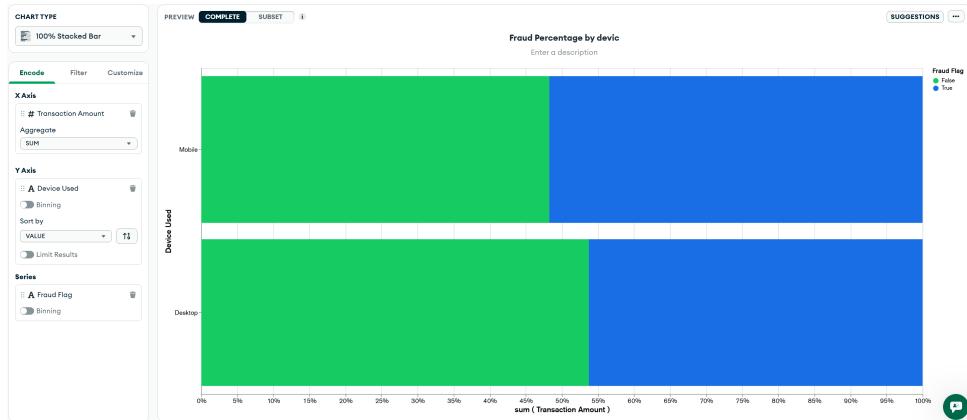
Compte le nombre de transactions par type d'appareil (Device Used) et affiche la répartition en pourcentage.



c) Section 3 : Analyses Détaillées

8. Fraud Percentage by device Agrégation utilisée : COUNT avec transformation en pourcentage

Compte les transactions par appareil et par statut de fraude, puis normalise à 100% pour montrer les proportions.



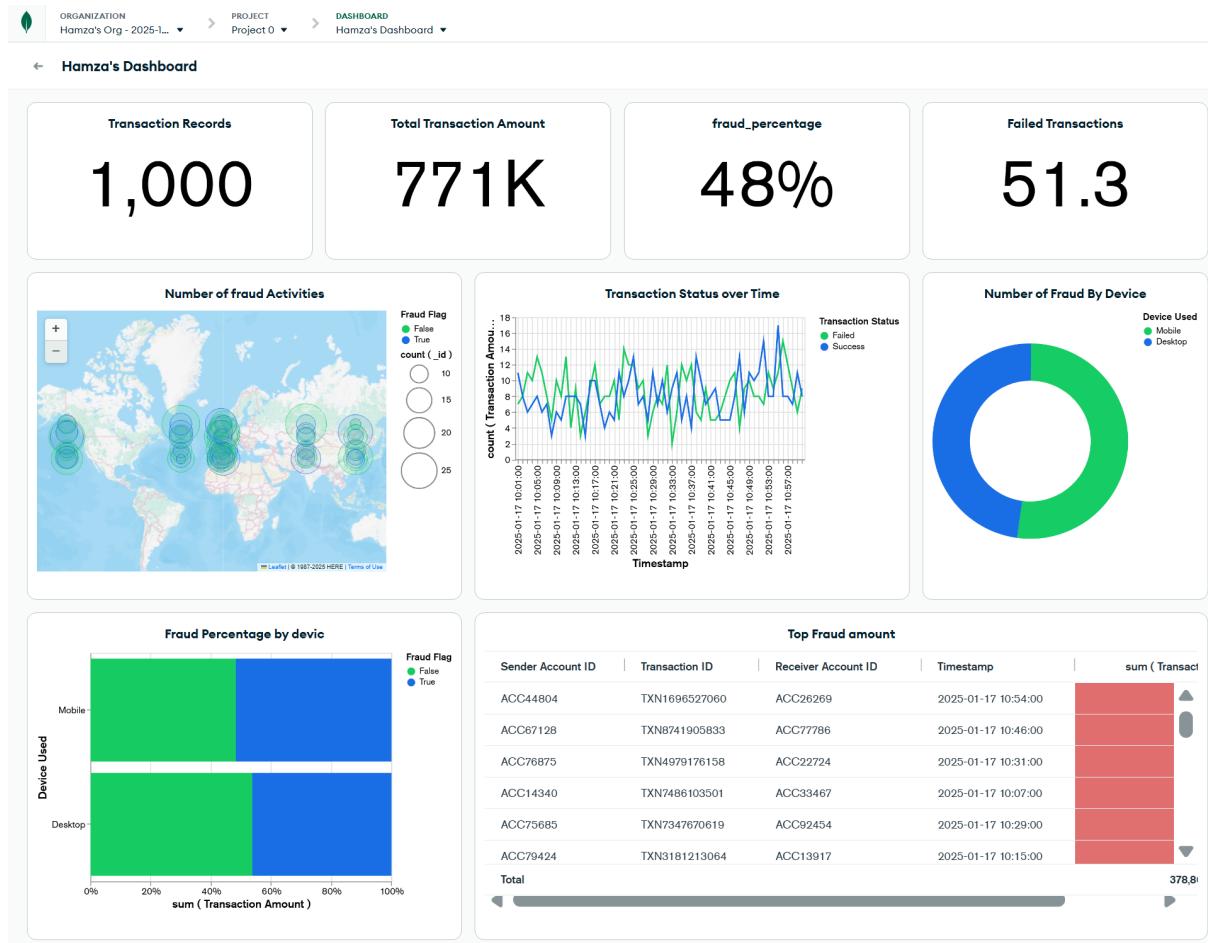
9. Top Fraud amount Agrégation utilisée : SUM

Filtre uniquement les fraudes (Fraud Flag = "True") avec montants > 1000, groupe par Sender, Transaction ID, Receiver et Timestamp, puis somme les montants. Le tableau est trié par montant décroissant avec un gradient de couleur (plus foncé = montant plus élevé).

Top Fraud amount					
Enter a description					
Sender Account ID	Transaction ID	Receiver Account ID	Timestamp	sum (Transaction Amount)	
ACC44804	TXN1696527060	ACC26269	2025-01-17 10:54:00	1,497.76	
ACC67128	TXN8741905833	ACC77786	2025-01-17 10:46:00	1,495.01	
ACC76875	TXN4979176158	ACC22724	2025-01-17 10:31:00	1,494.2	
ACC19340	TXN7486103001	ACC33467	2025-01-17 10:07:00	1,494.03	
ACC75685	TXN7347670619	ACC92454	2025-01-17 10:29:00	1,488.93	
ACC79424	TXN9181213064	ACC13917	2025-01-17 10:15:00	1,480.5	
ACC63042	TXN8679057633	ACC67654	2025-01-17 10:24:00	1,479.46	
ACC74504	TXN4101163005	ACC51180	2025-01-17 10:53:00	1,478.97	
ACC54544	TXN82368803729	ACC22881	2025-01-17 10:13:00	1,473.03	
ACC47841	TXN2545528308	ACC84633	2025-01-17 10:32:00	1,468.59	
ACC79263	TXN1696074130	ACC45900	2025-01-17 10:50:00	1,467.44	
ACC72542	TXN4172589703	ACC82377	2025-01-17 10:01:00	1,457.37	
ACC61170	TXN8594411611	ACC73104	2025-01-17 10:26:00	1,456.87	
ACC723612	TXN969738753	ACC47572	2025-01-17 10:23:00	1,452.12	
Total					378,863,90000000014

3. Dashboard Final

Le dashboard ci-dessous présente les 9 visualisations organisées de manière cohérente avec les KPIs en haut, les visualisations principales au milieu, et les analyses détaillées en bas.



4. Insights Clés

D'après le dashboard, les insights principaux sont :

- **Volume :** 1,000 transactions au total avec un montant total de 771K
- **Taux de fraude élevé :** 48% des transactions sont frauduleuses
- **Taux d'échec important :** 51.3% des transactions ont échoué
- **Distribution par appareil :** Répartition équilibrée entre Mobile et Desktop
- **Géographie :** Fraudes concentrées en Amérique du Nord, Europe et Asie
- **Top fraude :** La transaction la plus élevée atteint 1,497.76.

5. Conclusion

MongoDB Charts permet de créer des dashboards interactifs directement connectés à MongoDB Atlas, facilitant l'analyse de données en temps réel sans exportation. Les agrégations (COUNT, SUM, MEAN) sont essentielles pour transformer les données brutes en insights exploitables. Le dashboard créé démontre la capacité à identifier rapidement les problèmes (taux de fraude élevé de 48%) et à visualiser les distributions géographiques et temporelles des transactions.