

Comportements transactionnels & Fraude bancaires

03 Février 2026

Introduction & problématiques

- Cabinet de conseil spécialisé en data analytics et gestion du risque
- accompagnant les entreprises dans la détection de fraude, l'optimisation des processus et la prise de décision basée sur les données.



Comment analyser et visualiser efficacement les comportements transactionnels afin d'identifier les tendances, les anomalies et les zones de risque liées aux transactions bancaires, dans le but d'améliorer la prise de décision et la stratégie de lutte contre la fraude ?



Comment une entreprise peut-elle prédire les transactions frauduleuses par carte bancaire, tout en minimisant les faux positifs ?



Données

Données fictives :

<https://www.kaggle.com/code/aadyasingh55/transaction-fraud-detection/input>

- Période : Mois d'octobre 2024
- Volume : ~ 7,5 millions de lignes
- Variables : 24 colonnes
- Type : temporelles, commerciales, géographiques et financières

Plan

1. **powerBI : Tableau de board**

- Nettoyage & préparation des données
- Relation entre les tables
- Colonnes calculées & mesure DAX
- Visualizations
- Bilan d'analyse des tableaux de bords

2. **Machine Learning :**

- préparation des données pour ML
- modélisation & Evaluation des modèles

3. **Synthèse**

- Difficultés rencontrés
- Conclusion
- Recommandations métiers

Nettoyage & préparation des données

- Remplacer des valeurs "unknown" de la colonne "city" par le nom de la ville capitale du pays
- modifier le taux de fraude plus réaliste selon les catégories des commerçants puis pays
- Découpage de dataset en table de fait et de dimension
- table de dimension "currency" : supprimer les doublons & ajouter la colonne de taux de conversion en euros ("Exchange rate to euros")

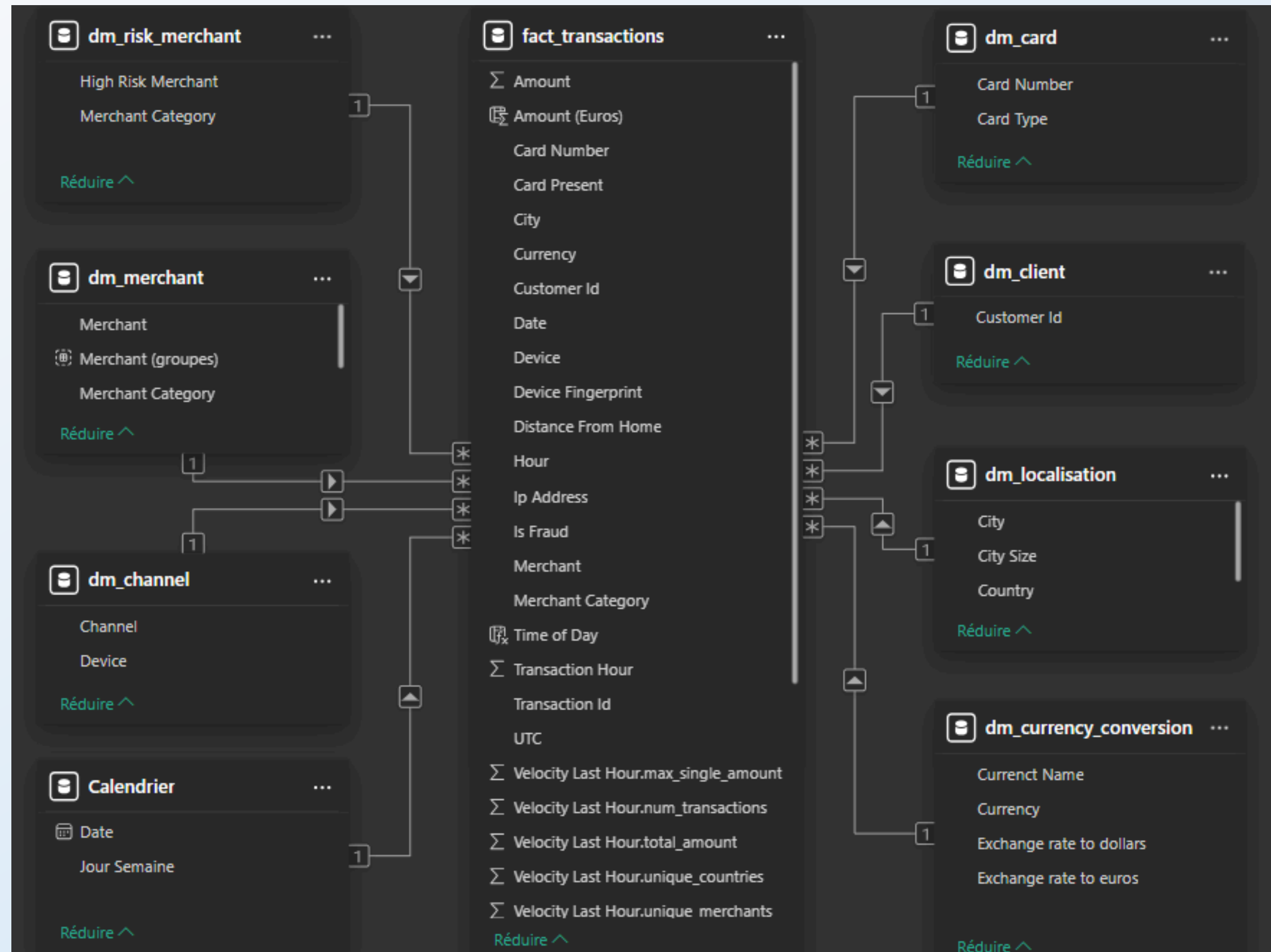


Nettoyage des données PowerQuery

- mettre 1ère ligne en l'en-tête & 1ère lettre en Masculine et le reste en minuscule
- pour les colonnes numériques, remplacer . par ,
- modifier le type des colonnes : date, décimale, entier ...
- supprimer les doublons sur les autres tables de dimensions
- supprimer les colonnes inutiles : les données du 30 Septembre
- des colonnes personnalisés : extraire date, heure,UTC de la colonne "Timestamp"



Relation entre les tables



Colonnes calculées

- Amount (Euros)
- Périod de la journée :
 - entre 6h et 12h : Matin
 - entre 12h et 18h : l'après -midi
 - entre 18h et 22h : soir
 - entre 22h et 6h : la nuit

Mesures DAX

- panier moyen des transactions
- panier moyen des fraudes
- panier médian des transactions
- panier médian des fraudes
- Taux de fraude (Nb Transactions)
- Taux de Fraude (Montant)

Nombre total de
transactions légitimes

5 795 990

Nombre total de
transactions
frauduleuses

376 888

Taux de Fraude
(Nb
transations)

6,11 %

Montant toal de
transactions legitimes
(€)

2 491,13M

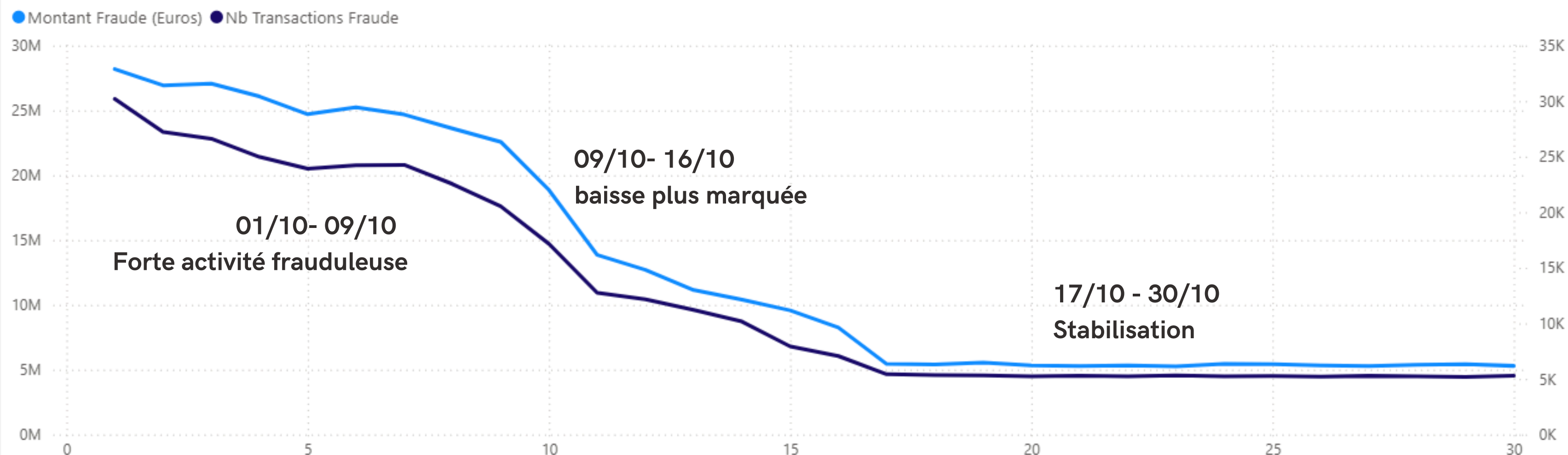
Montant total de
transactions
frauduleuses (€)

389,01M

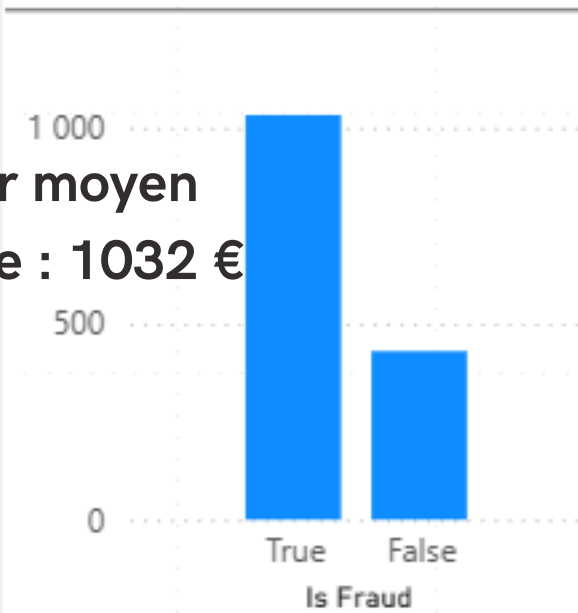
Taux de Fraude
(Montant)

13,51%

Nombre de transactions et somme du montant frauduleuses en fonction du jour

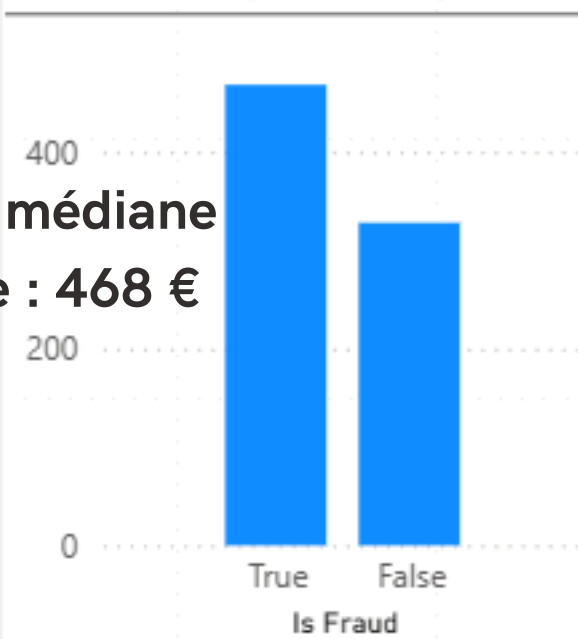


Moyenne de panier moyen (€)



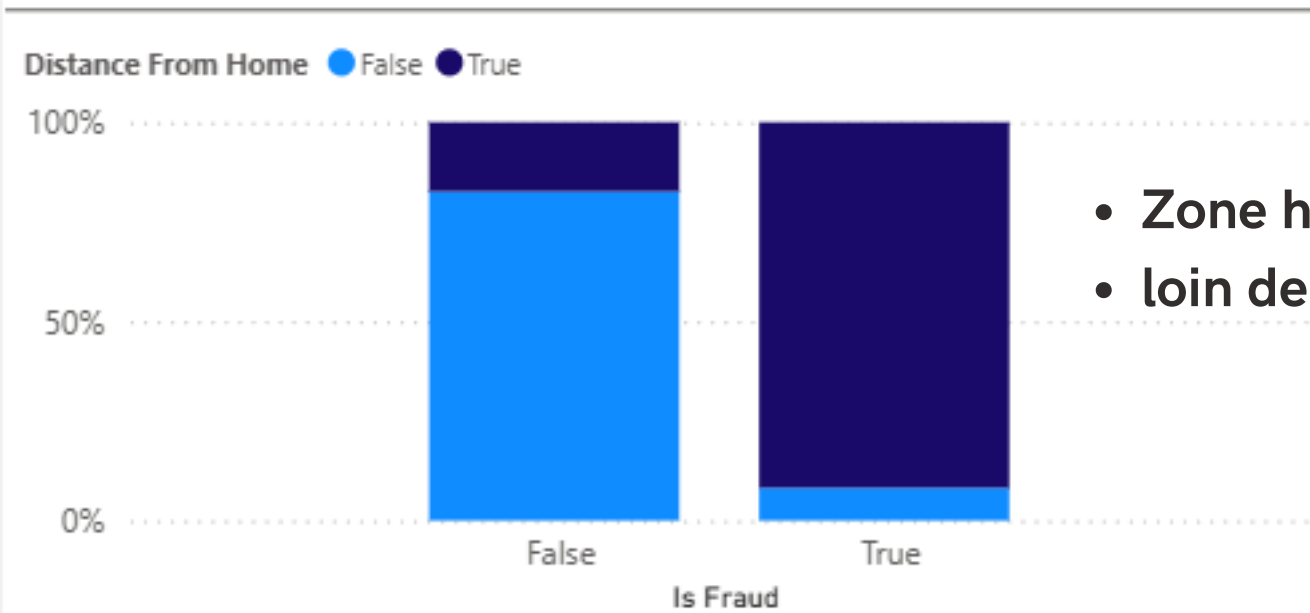
**Panier moyen
fraude : 1032 €**

Médiane de panier moyen (€)



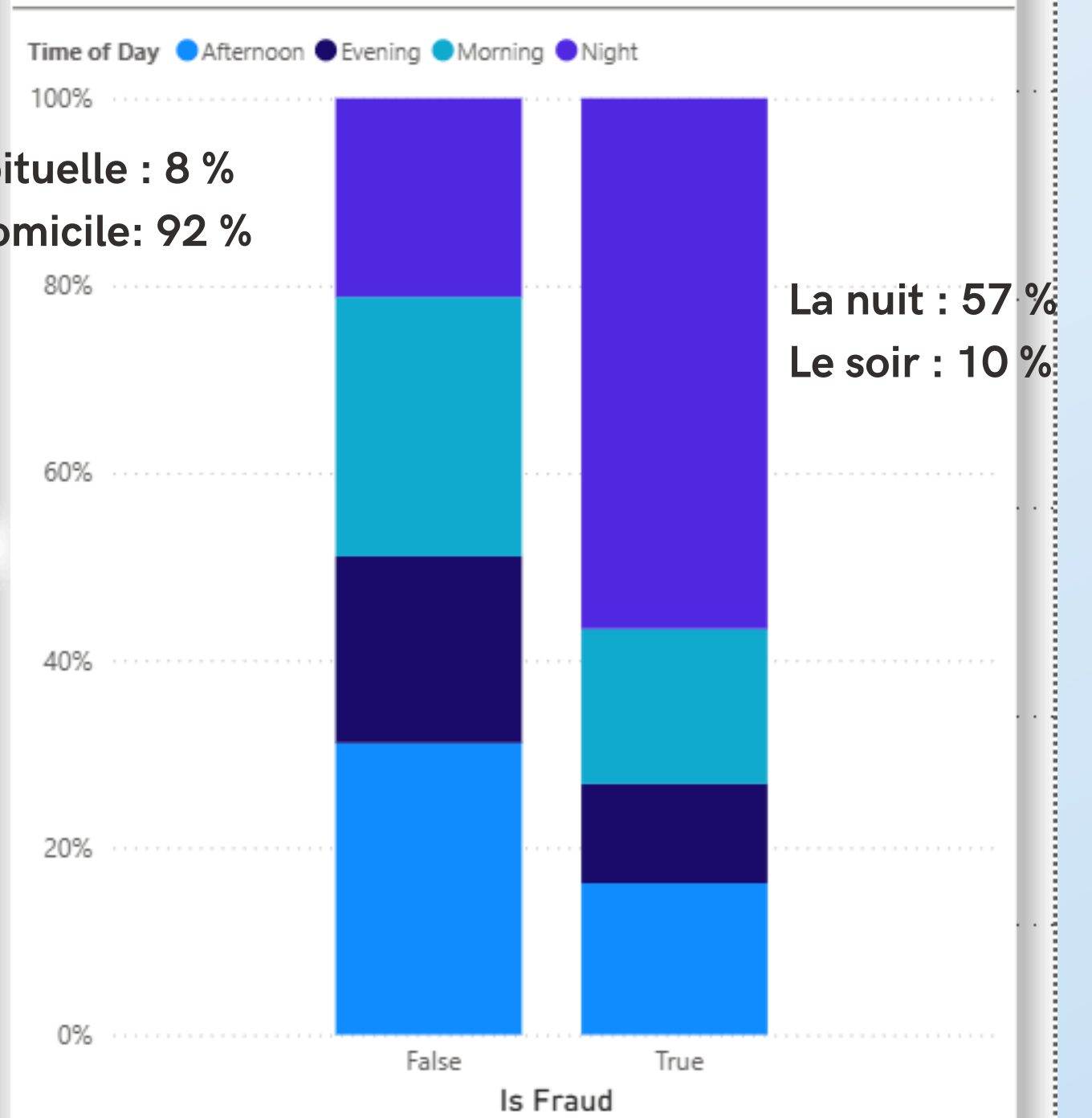
**Panier médiane
Fraude : 468 €**

Répartition des transactions selon la distance géographique



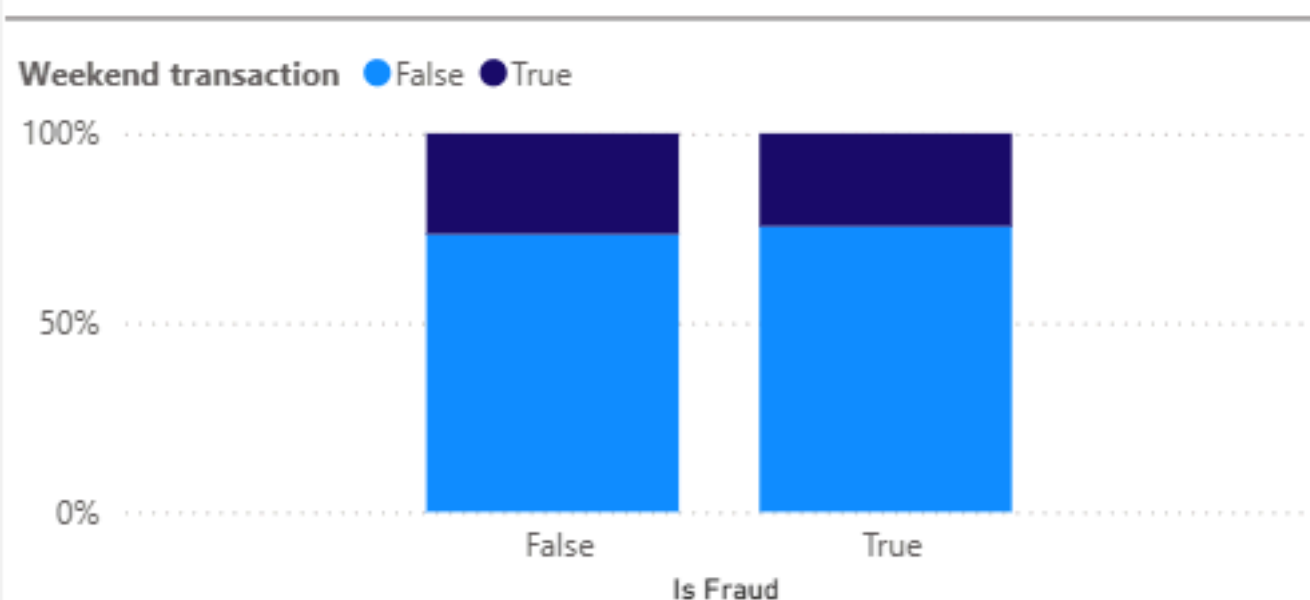
- Zone habituelle : 8 %
- loin de domicile: 92 %

Nombre de transactions en fonction de la période de la journée



**La nuit : 57 %
Le soir : 10 %**

Nombre de transactions total et frauduleuses - Nombre de transactions – Jours de semaine vs Week-end



Nombre de pays

12

Nombre total des villes

21

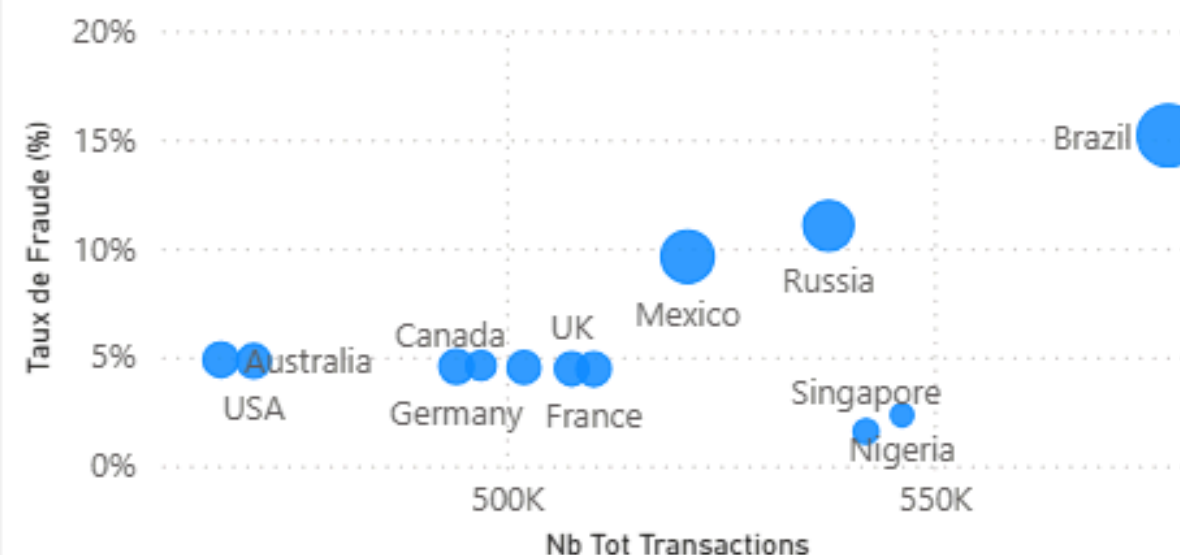
Nombre total de transactions
légitimes

5 795 990

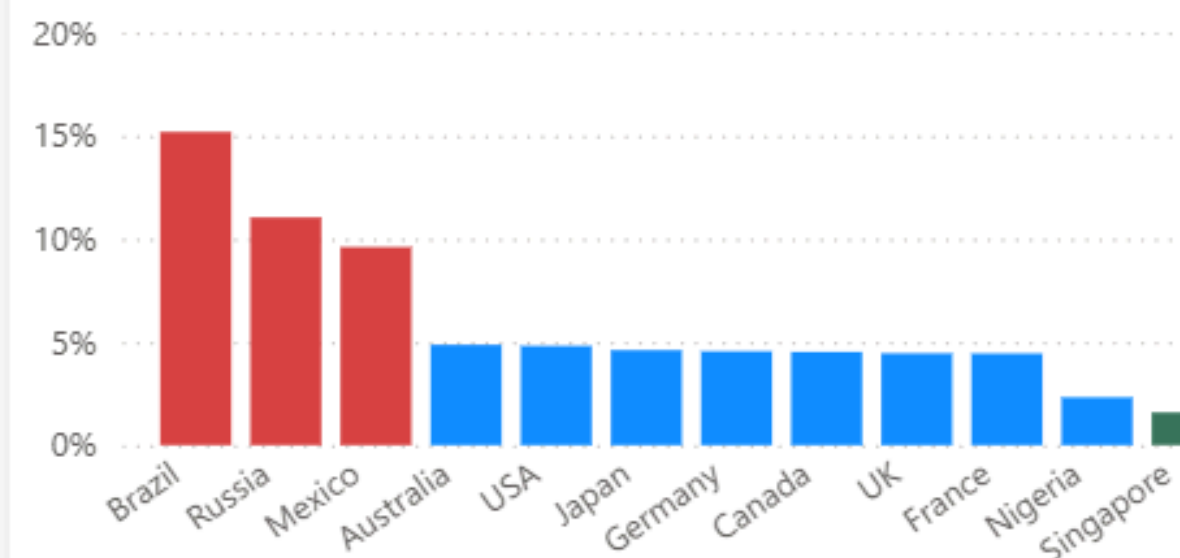
Nombre total de transactions
frauduleuses

376 888

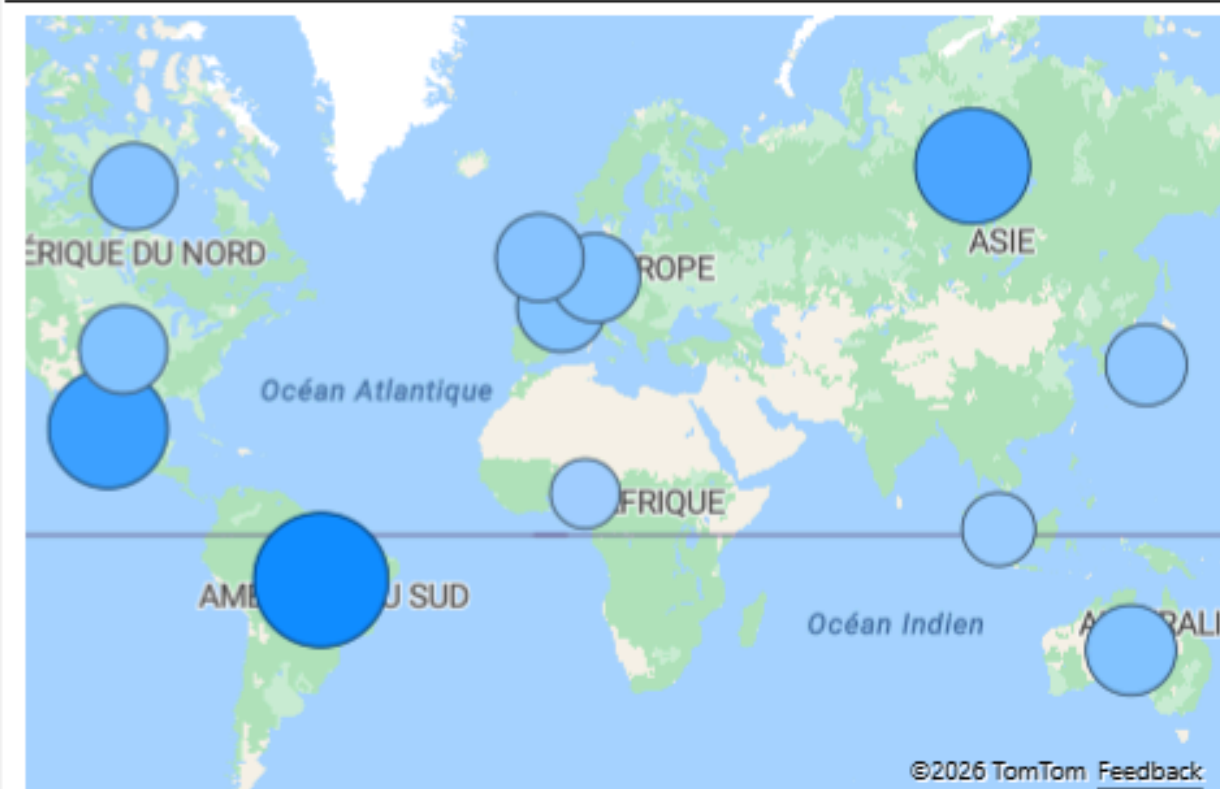
Volume de transactions, taux de fraude et montants frauduleux par pays



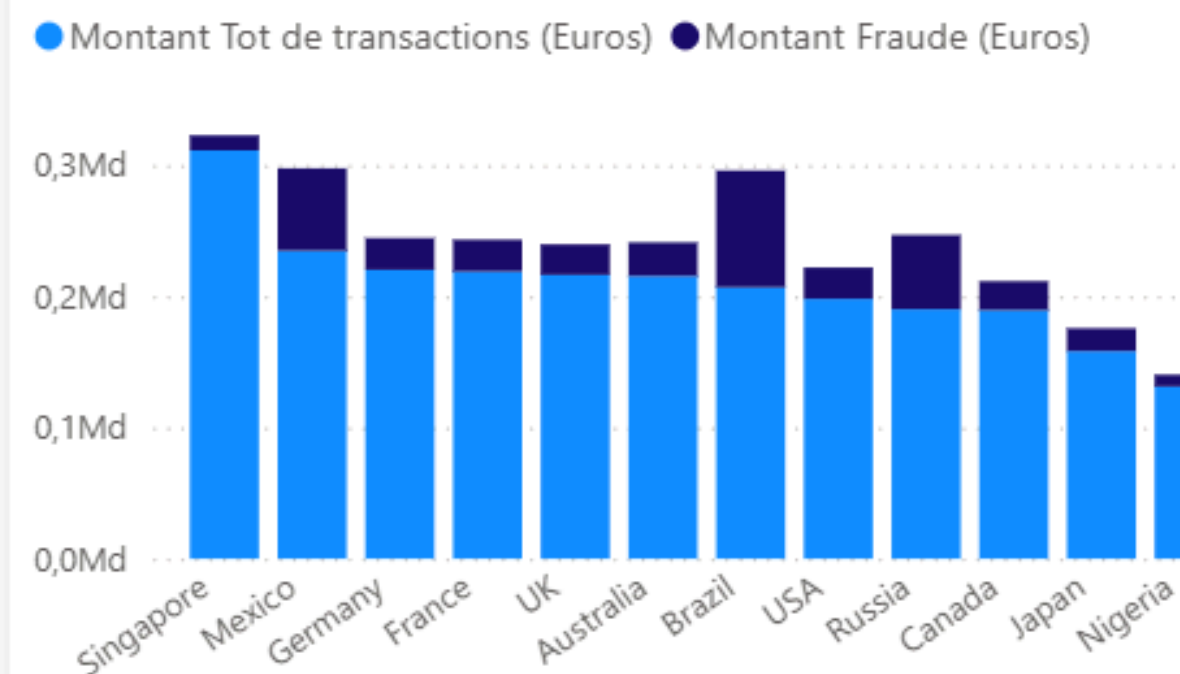
Taux de fraude par pays



Montant total de fraude par pays



Montant total de transactions légitimes et Frauduleuses (Euros) en fonction du pays





Localisation géographique: USA

Villes

Tout

USA

Médiane de panier moyen
frauduleux (€)

357,89

Moyenne de panier moyen
frauduleux (€)

472

Nombre de villes

10

Taille des villes

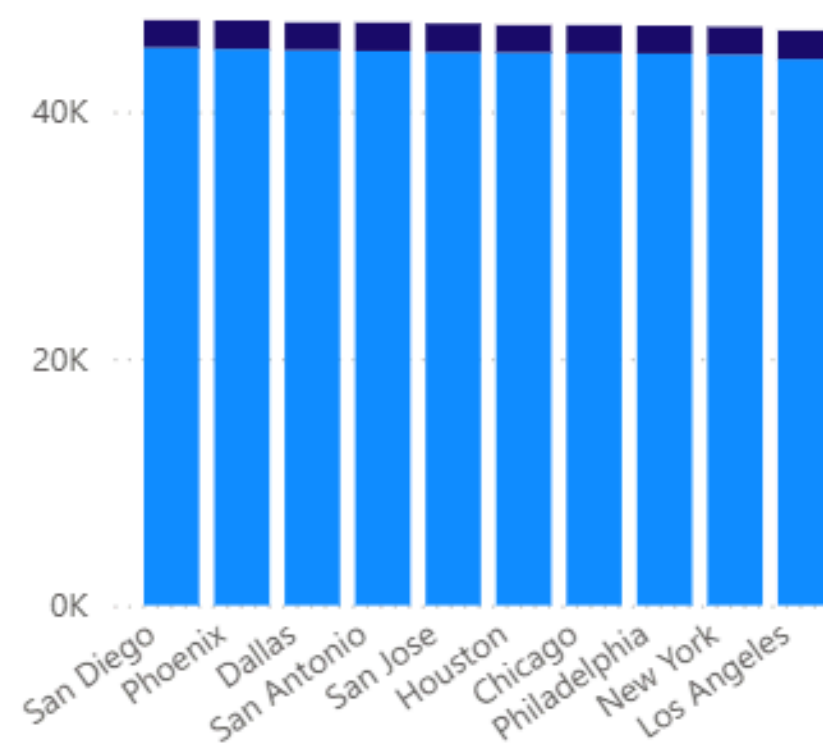
City Size Nombre de City

medium 6

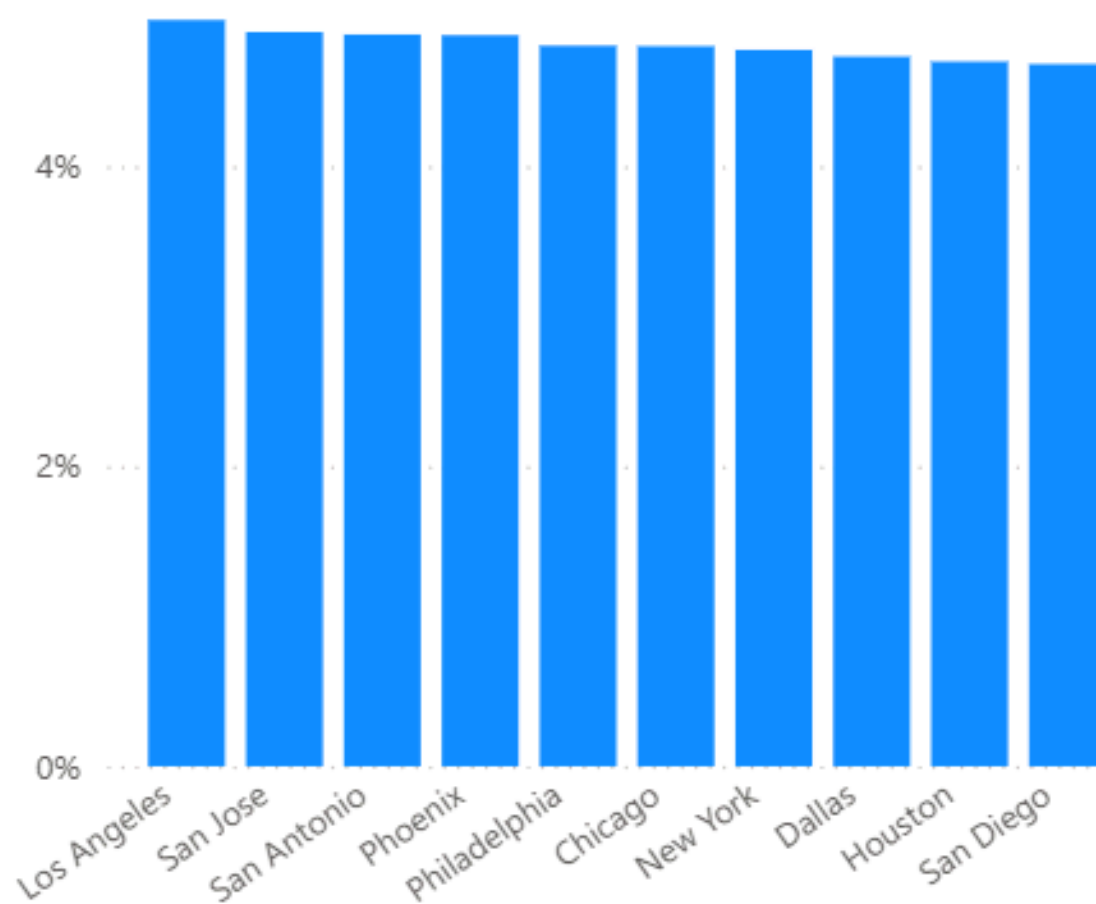
large 4

Nombre de transactions en fonction des
villes

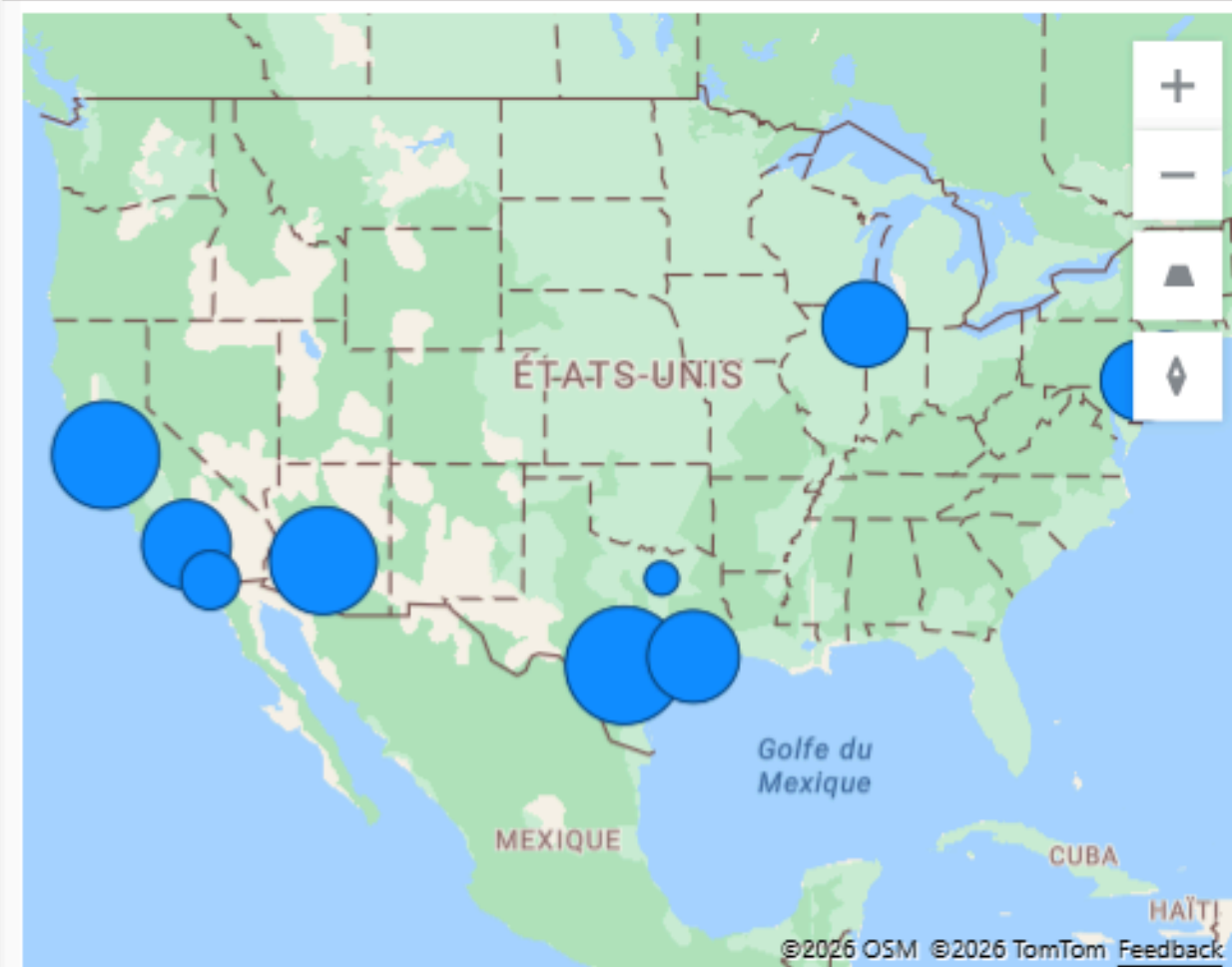
● Nb Tot Transactions ● Nb Transactions Fraude



Taux de fraude par villes



Montant fraude en fonction des villes



Analyse catégorie de commerçants

Nombre de catégorie

8

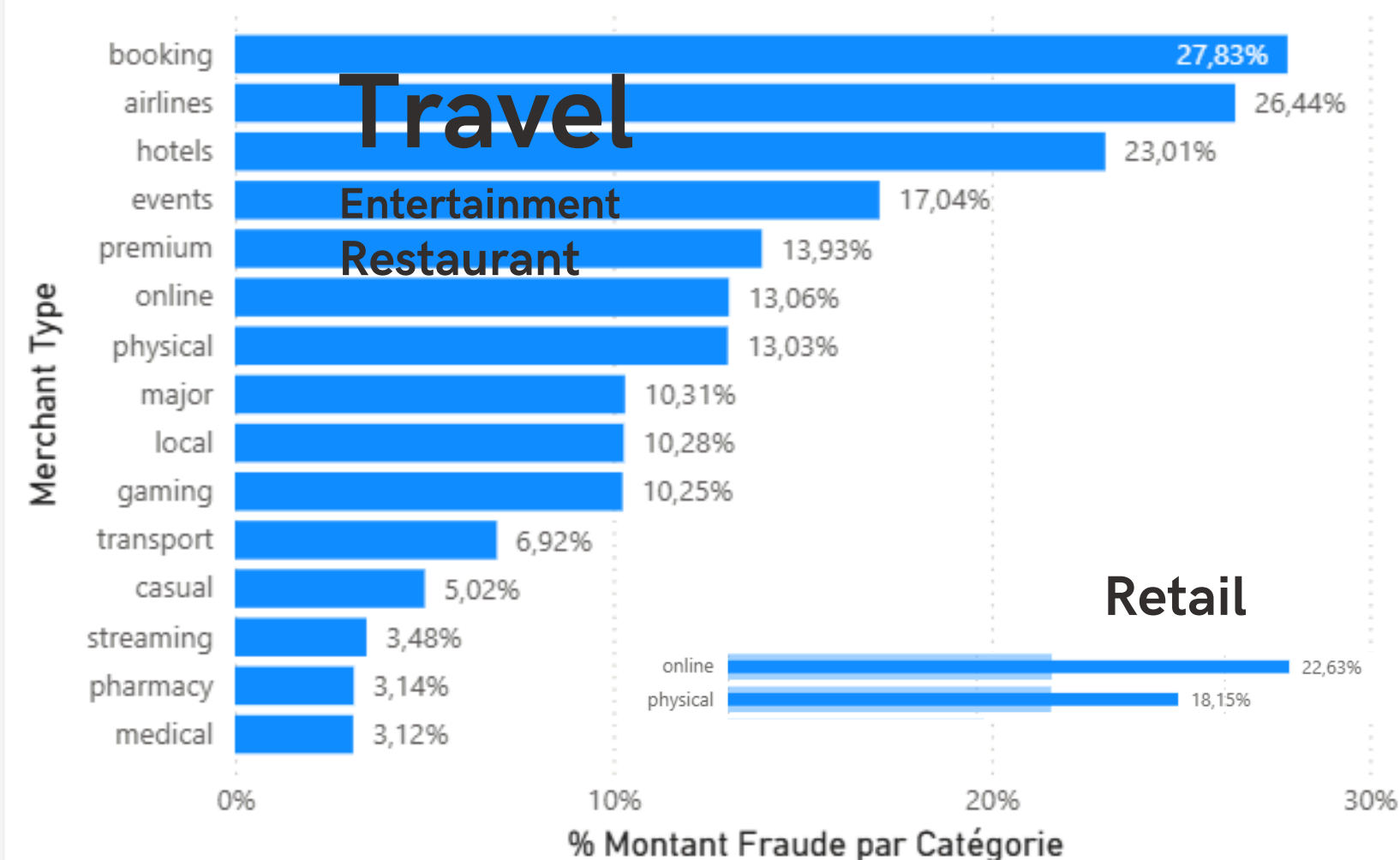
Nombre de type
commerçants

17

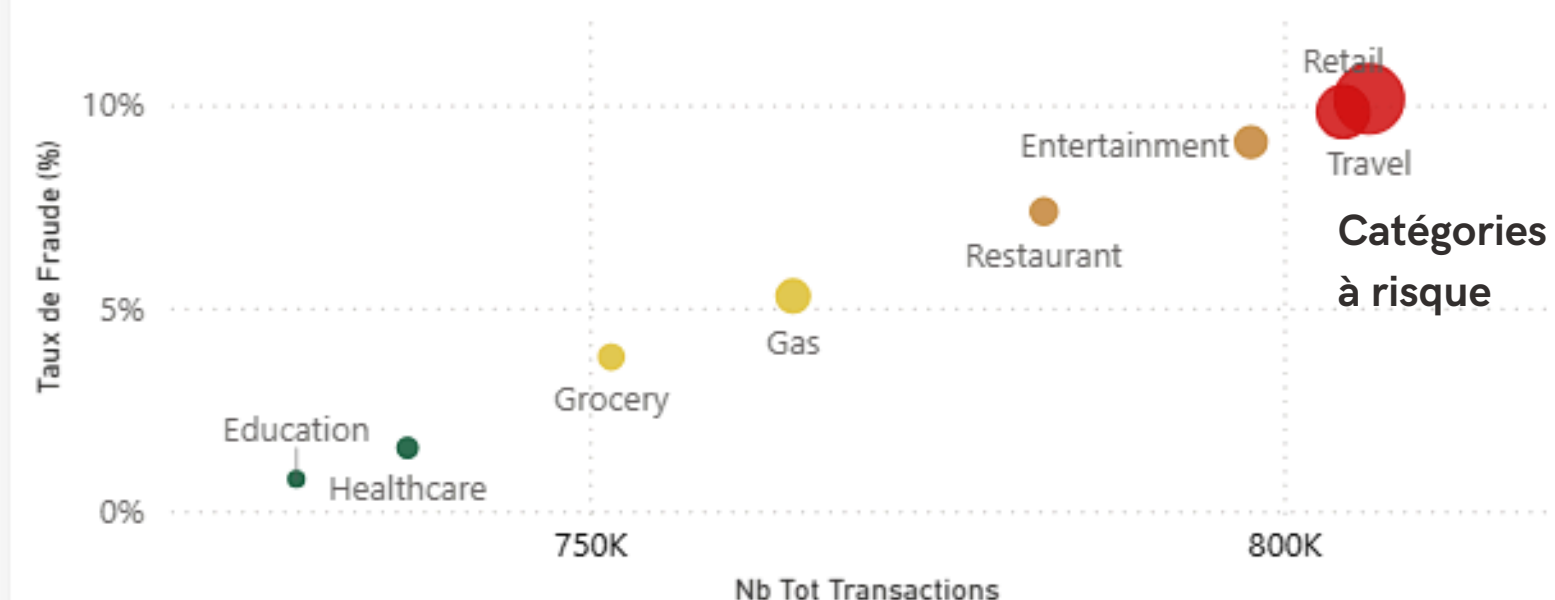
Nombre de
commerçants

105

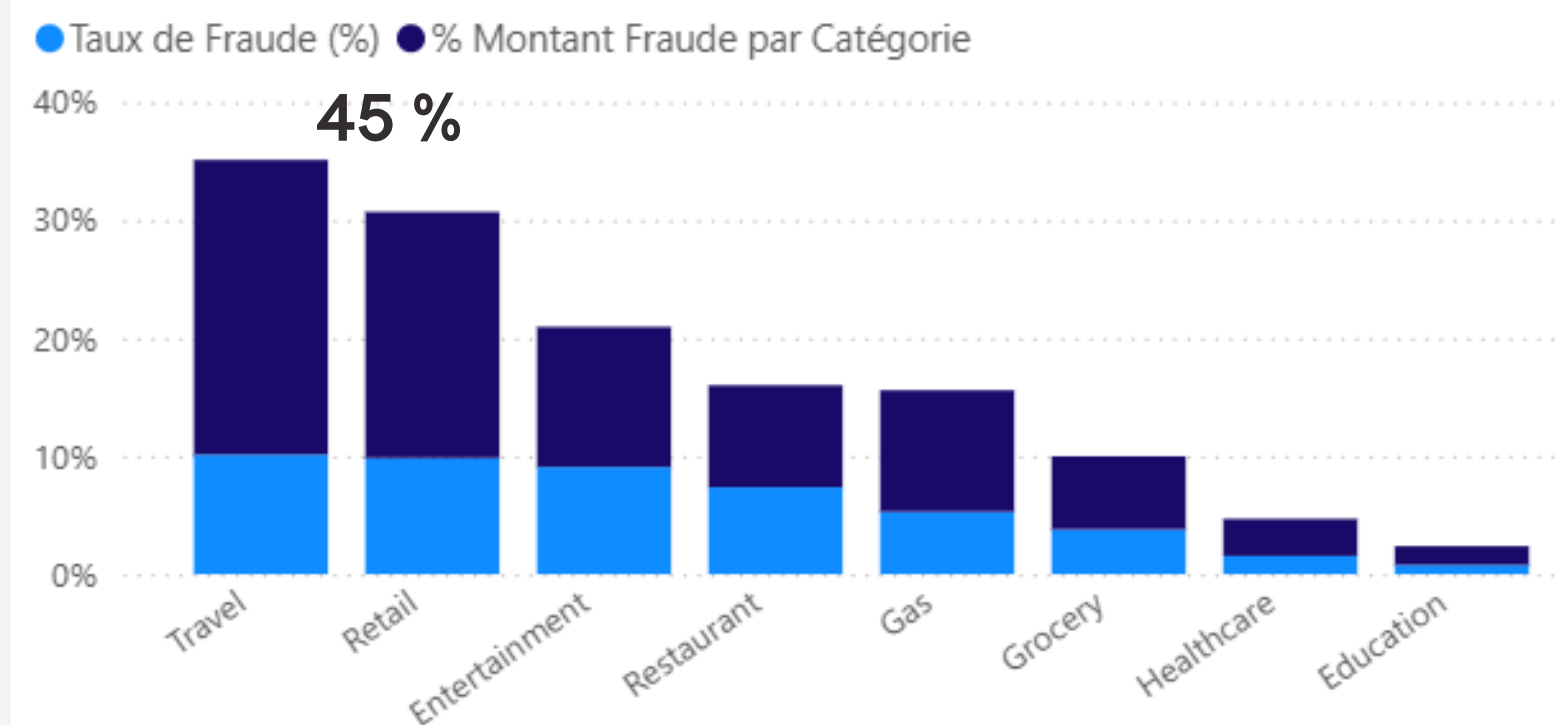
Taux de Fraude par type de commerçants



Analyse des transactions et de la fraude par catégories



Taux de fraude par catégorie de commerçants





VAK Consulting

Analyse commerçants

Retail

Nombre de commerçants

15

Médiane de panier moyen
frauduleux (€)

425,85

Moyenne de panier moyen
frauduleux (€)

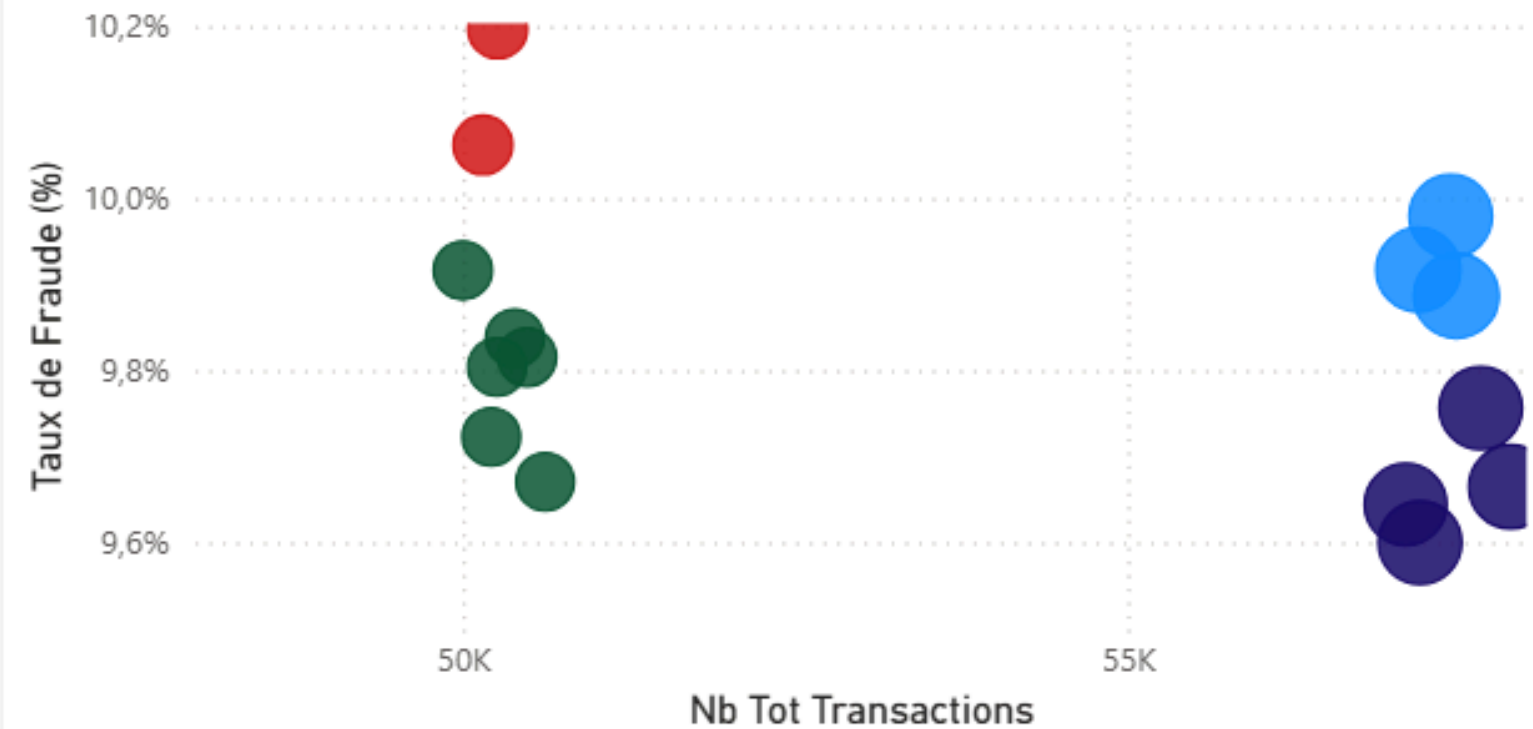
574

Top 5 commerçants à risque

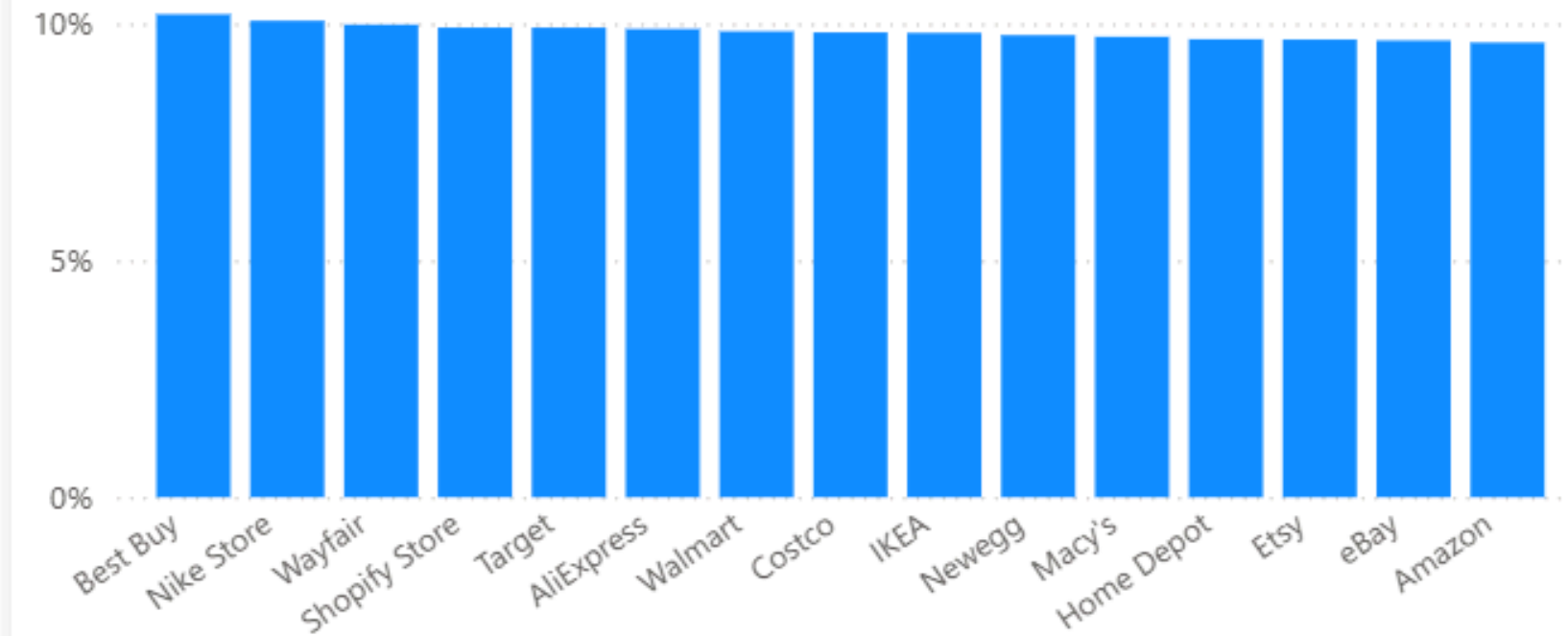
Commerçant	Montant de fraude (€)
AliExpress	9 382 738
Shopify Store	9 214 973
Etsy	9 041 866
Amazon	8 987 001
Newegg	8 914 435

Analyse des transactions et de la fraude par commerçants

Merchant (groupes) ● AliExpress & Shopify ... ● Autre ● Best Buy & Nik... ● Costco & Ho...



Taux de fraude par merçants



Bilan global

- Fraude plus élevée en **début de mois**, puis **diminution progressive** et **stabilisation** en fin de mois.
- Les comportements frauduleux sont stables, prévisibles et cohérents : montants **modérés à élevés**, **localisation inhabituelle**, et **forte concentration nocturne**.
- Pays les plus à risque : **Brésil, Russie, Mexique**.
- Pays le moins à risque : **Singapour**.
- Catégories les plus à risque : **Travel** (Voyage) et **Retail**.
- Catégories les moins à risque : **Education et Healthcare**.

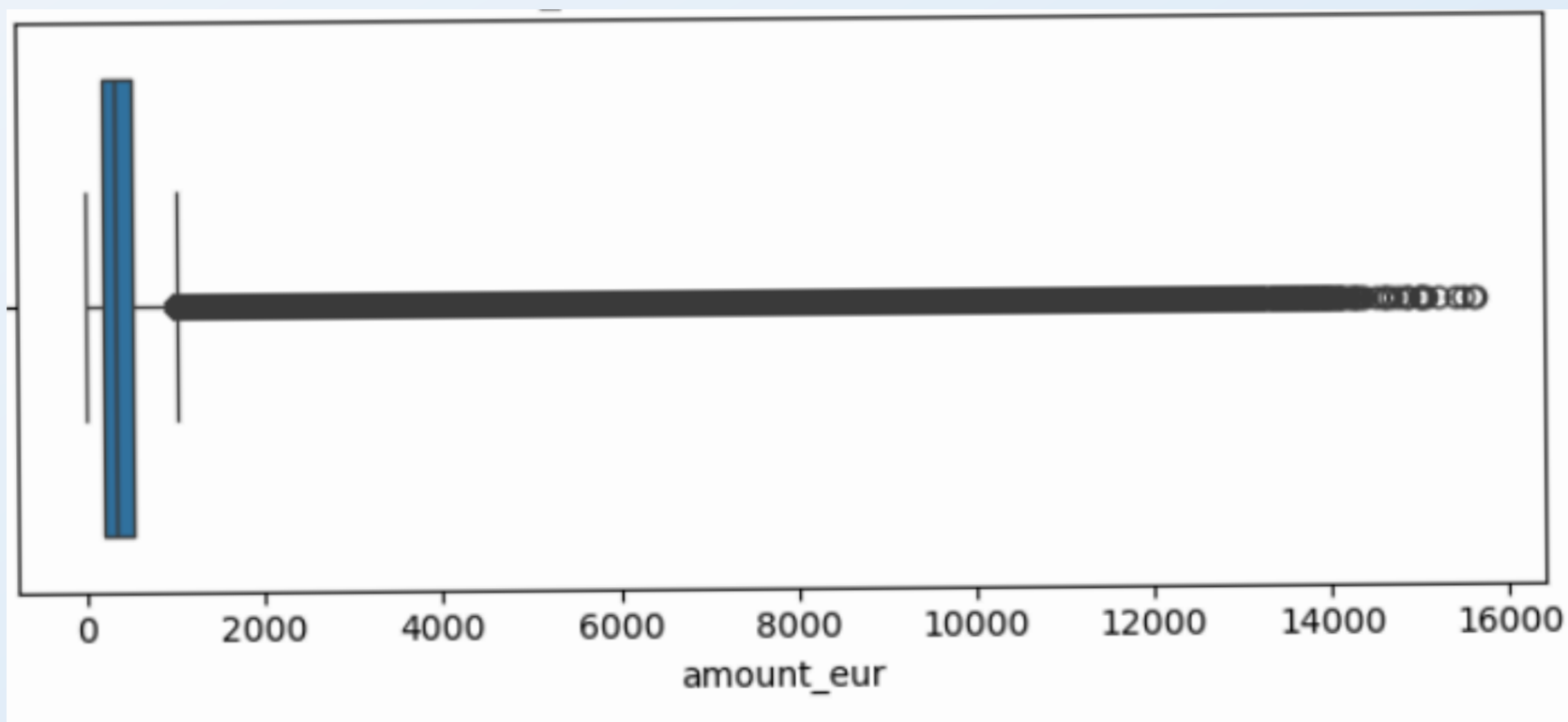
Machine Learning

Comment une entreprise peut-elle prédire les transactions frauduleuses par carte bancaire, tout en minimisant les faux positifs ?

Données & préparation (ML)

- Diviser la colonne timestamp en time et date
- Vérifier les données nuls et les supprimer
- Créer la colonne amount_eur
- Colonne période de la journée
- Créer des dummies pour les colonnes periode de la journée, pays, catégories commerçants
- Type booléan → type int
- Descriptives statistiques pour la colonne Amount_eur

```
count    6395838.0
mean      468.0
std       565.0
min        0.0
25%      186.0
50%      330.0
75%      522.0
max     15622.0
```



period_of_day_evening	period_of_day_morning	period_of_day_night
False	False	False
True	False	False
False	False	False
False	False	False
True	False	False

Modélisations & Evaluation des modèles

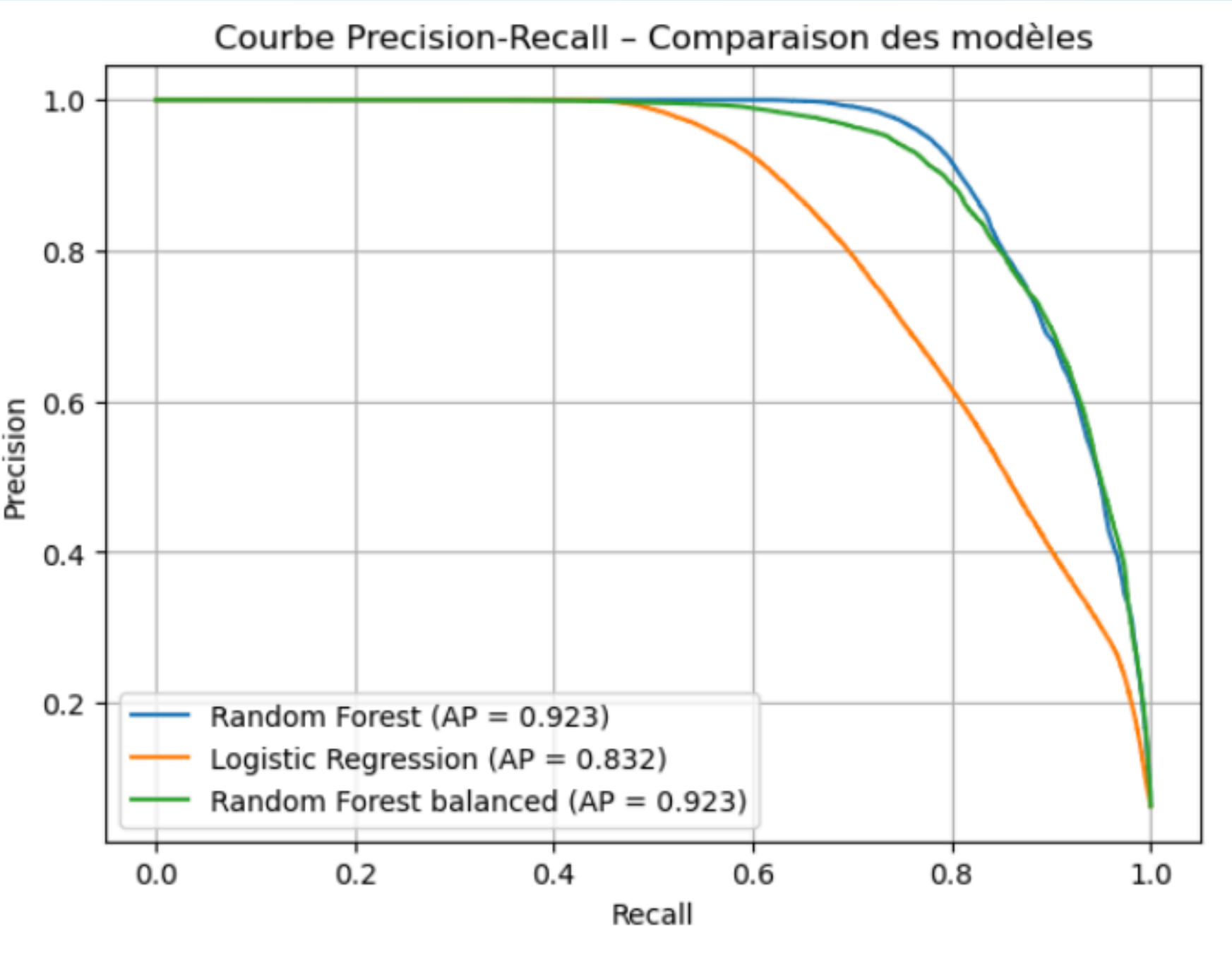
	Régression logistique ★★	Random forest ★★★★	Random forest balanced ★★★
Matrice de confusion	[[1191446 6325] [30023 51374]]	[[1196492 1279] [21491 59906]]	[[1130555 67216] [4642 76755]]
Précision globale	97.16%	98.22%	94.38%
Précision	89.04%	97.91%	53.31%
Rappel (Recall)	63.12%	73.60%	94.30%
F1-Score	73.87%	84.03%	68.12%

FP * 50

FN / 4,6

TN / 1,1

TP * 1,3



Difficultés rencontrées

- Nombre de ligne élevés → travail pas chucks (code RAM-Friendly)
- Données fictives trop parfait → équilibrer les données pour les rendre plus réaliste (catégorie des commerçants, pays)
- Pas assez de données numériques (ML) → des dammies
- Jeu de données fortement déséquilibré → modèles "balanced"

Conclusion globale

1. Métier

La fraude est **minoritaire à modéré en volume** mais **majeure en impact financier**.

2. Data

Les comportements frauduleux présentent des patterns clairs et exploitables : **montants élevés, localisation inhabituelle, forte activité nocturne, catégories spécifiques**.

3. ML

Les modèles de machine learning permettent une **détection performante**, avec un bon **équilibre** entre **détection de la fraude** et **réduction des faux positifs**.

4. Valeur business

Une telle solution permettrait de réduire significativement les pertes financières.

Recommandations métiers – Stratégie antifraude

1. Renforcement temporel

- En début de mois, la fraude est plus élevée.
→ mécanismes de contrôle renforcés : **Modèle Random Forest balanced**

2. Seuils basés sur le montant

- des seuils de contrôle à partir de montants proches du panier médian frauduleux.

3. Contrôle géographique

- Double contrôle , plafond réduits, validation manuelle

4. Renforcement nocturne

- seuils plus strict, blocage automatique des montants élevés en fonction de la probabilité calculée par les modèles.

5. Ciblage par catégorie

- Travel & Retail (Online)



Merci !