Communauté Economique et Monétaire De l'Afrique Centrale
(CEMAC)

Institut Sous-régional de Statistique et d'Economie Appliquée

(ISSEA)

Organisation Internationale
BP: 294 Yaoundé (Cameroun)
Tel: (+237) 622 220 134 Fax: (+237)622 229 521
Email : isseacemac@yahoo.fr

ANALYSE DES TRANSACTIONS FINANCIERES A PARTIR DES DONNEES DE LA PLATEFORME PAYSIM

Réalisé par

AMBASSA Samuel Lumière
ASSA Allo
AZONFACK Myriam
DOMEVENOU Komla Wisdom
KOULOU Crépin Anaklassé
MANGA FOUDA Léonard
OGNIMBA Sadri
SEYDOU Ferdinand

&

TEYANBAYE BERTORNGAÏ Christian

Sous la supervision de :

M. Serge NDOUMIN

Février 2025

TABLE DES MATIERES

INTRODUCTION

Le domaine des services financiers, et plus particulièrement celui des transactions d'argent mobile, est confronté à une pénurie notable de jeux de données publics. Cette rareté s'explique principalement par la nature hautement confidentielle des transactions financières, rendant difficile la mise à disposition de telles données pour la recherche.

Cette absence de données accessibles freine considérablement les avancées dans des domaines cruciaux tels que la détection de la fraude. Par exemple, une étude souligne le manque de jeux de données publics disponibles pour les transactions d'argent mobile, ce qui entrave le développement de modèles efficaces de détection de fraude.

Les statistiques récentes mettent en évidence l'ampleur du problème. En France, en 2023, la fraude aux moyens de paiement scripturaux a atteint un montant de 1,195 milliard d'euros, répartis sur 7,1 millions d'opérations frauduleuses . Plus spécifiquement, la fraude au virement a triplé depuis 2017, représentant 313 million d'euros en 2022. Dans le contexte des services d'argent mobile, la situation est tout aussi préoccupante. Par exemple, au Ghana, les femmes sont 89 % plus susceptibles que les hommes d'être victimes de fraude dans les services bancaires mobiles. Cette vulnérabilité accrue souligne l'importance de disposer de données pour développer des stratégies de protection adaptées.

C’est dans ce contexte que se place cette étude, menée à partir des données simulées de la plateforme PaySim, dans laquelle il est question d’appliquer les méthodes de traitement parallèle, d’utiliser le framework PySpark afin d’en dégager une analyse des transactions financières en dégageant 10 indicateurs clés avec

Objectifs

De manière générale, il s’agira dans cette étude d’analyser les transactions financières. Spécifiquement nous allons dégager quelques indicateurs pertinents à partir de la structure de la base de données et ensuite ressortir les tendances générales de chacun d’eux.

Méthodologie

Afin d'atteindre nos objectifs, nous utiliserons les techniques de statistiques descriptives pour décrire, analyser et commenter la tendance des indicateurs. Cependant en amont, pour les étapes de chargement des données, transformations, calcul des indicateurs, et visualisation nous ferons usage de Python et de PySpark (pour le traitement parallèle des données).

I. DONNEES DE L'ETUDES ET CHOIX DES INDICATEURS

I.1. DONNEES DE L'ETUDE

Les données utilisées dans cette étude sont issues d'une plateforme de simulations PaySim qui simule des transactions d'argent mobile en se basant sur un échantillon de transactions réelles extraites d'un mois de logs financiers provenant d'un service d'argent mobile mis en œuvre dans un pays africain. Les logs originaux ont été fournis par une entreprise multinationale, fournisseur du service financier mobile qui est actuellement opérationnel dans plus de 14 pays à travers le monde.

La base de données est constituée de 11 variables et de 6 362 620 lignes de transactions. Les vérifications faites sur cette base montre qu'elle ne comporte aucune valeur manquante.

I.2. IDENTIFICATION DES INDICATEURS PERTINENTS

Indicateur	Définition	Utilité
1. Nombre total de transactions	Nombre absolu de toutes les transactions générées par PaySim sur une période donnée.	Mesure l'activité globale simulée. Permet d'analyser des scénarios d'activité normale ou suspecte.
2. Répartition des types de transactions	Classification des transactions générées par PaySim en	Identifie les canaux les plus utilisés et détecte des anomalies

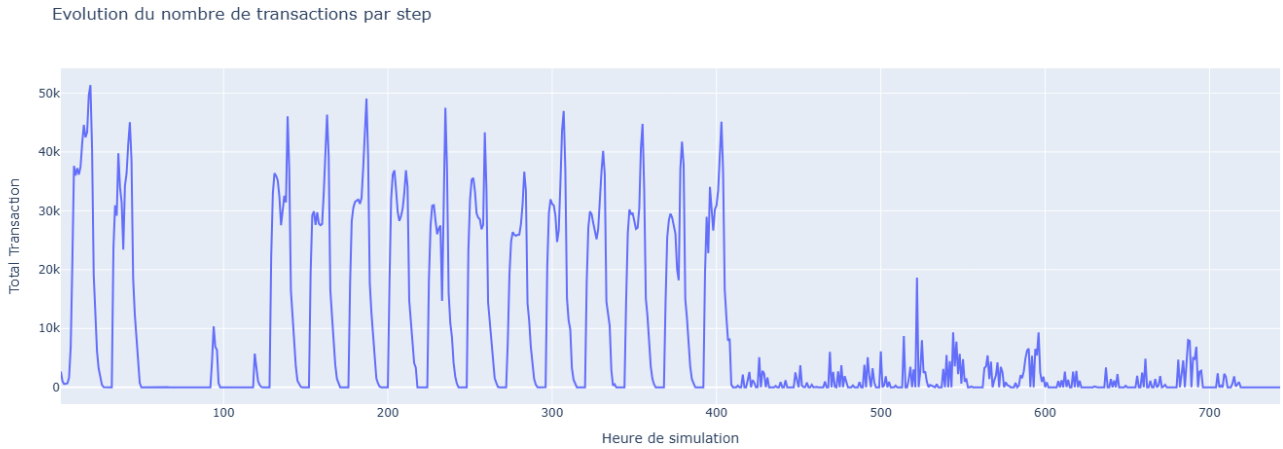
	catégories (ex. virements, paiements, etc.).	spécifiques à certains types de transactions.
3. Taux de fraude	Pourcentage de transactions frauduleuses générées par PaySim par rapport au total des transactions.	Évalue l'efficacité des systèmes de détection de fraude dans un environnement contrôlé.
4. Nombre de transactions suspectes	Transactions générées par PaySim et signalées comme potentiellement frauduleuses.	Permet de tester et d'ajuster les seuils d'alertes en simulant des transactions suspectes réalistes.
5. Montant total des transactions frauduleuses	Somme des valeurs financières des transactions frauduleuses générées par PaySim.	Quantifie l'impact économique simulé de la fraude, aidant à évaluer les stratégies de mitigation.
6. Fréquence des transactions par client	Nombre moyen de transactions effectuées par chaque client dans les données générées par PaySim.	Identifie des activités anormales (ex. clients avec une fréquence de transactions inhabituellement élevée).
7. Fréquence des transactions par destinataire	Nombre moyen de transactions reçues par chaque destinataire dans les données générées par PaySim.	Détecte des destinataires suspects (ex. comptes recevant un volume anormal de transactions).
8. Évolution des soldes des clients	Changement des soldes des comptes clients au fil du temps dans les données générées par PaySim.	Identifie des comportements suspects (ex. fluctuations brutales ou mouvements de fonds inhabituels).
9. Transactions avec solde initial nul	Transactions générées par PaySim où le solde du compte émetteur est nul avant la transaction.	Détecte des anomalies (ex. utilisation de comptes vides pour des activités frauduleuses).
10. Transactions avec solde final nul	Transactions générées par PaySim où le solde du compte émetteur est nul après la transaction.	Identifie des comportements suspects (ex. vidage de comptes ou transferts intégral des fonds).

II. ANALYSE DES TRANSACTIONS

Dans la plateforme de PaySim au total 6 868 768 de transactions ont été enregistrées. Ces transactions sont récoltées heure après heure sur un total de 31 jours. Une vue sur les transactions heure après heure montre que de manière globale sur les 400 premières heures, on a une période de forte activité, avec des pics réguliers de transactions dépassant les 40 000 voire 50 000. Après 400 heures l'intensité et la fréquence des pics diminuent drastiquement. On note une transition vers une activité plus irrégulière avec des montants transactionnels beaucoup plus faible.

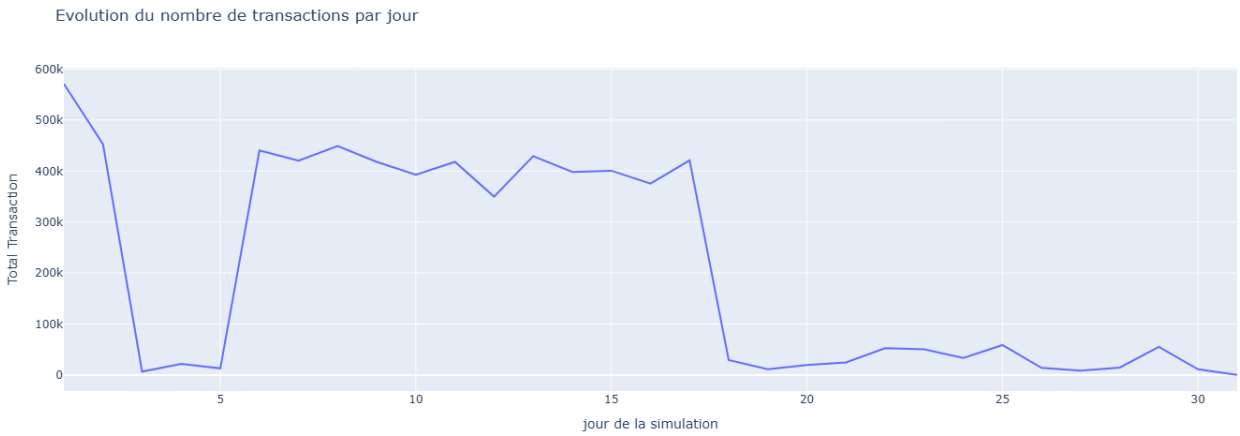
II.1. Vue globale sur les transactions de PaySim

Figure 1:Nombre de transactions par heure



Intuitivement, la répartition des transactions par jour suit la même tendance que celles par heure : une première période caractérisée par un fort nombre de transactions et une seconde de faible fréquence de transactions. Ces résultats paraissent tout à fait normaux puisque les débuts de mois sont souvent caractérisés par les retraits des salaires, les virements de salaires, payement des cotisations sociales, les payements de dettes, etc

Figure 2: Nombre de transactions par jour



II.2. Types de transactions

Une analyse plus approfondie suivant les types de transactions montre qu'en grande majorité elles portent sur les CASH OUT et des PAYMENT. Sur les plus de 6 millions de transactions, plus de 4 millions portent sur le CASHOUT et les PAYMENT ainsi en un mois, que ce soit de manière globale ou journalier(Figure 4), plus de transactions de retrait sont faites que les transactions de dépôt.

Figure 3: répartition des transactions par types

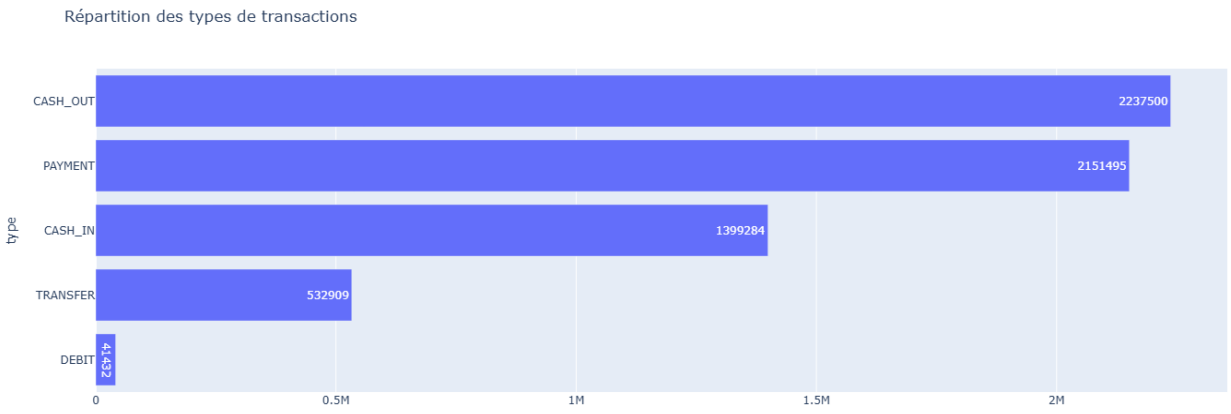
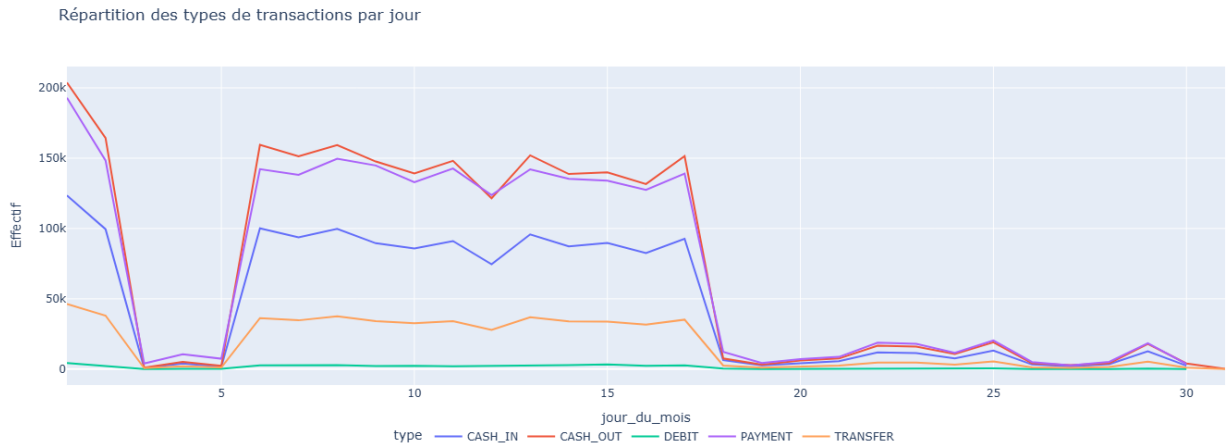


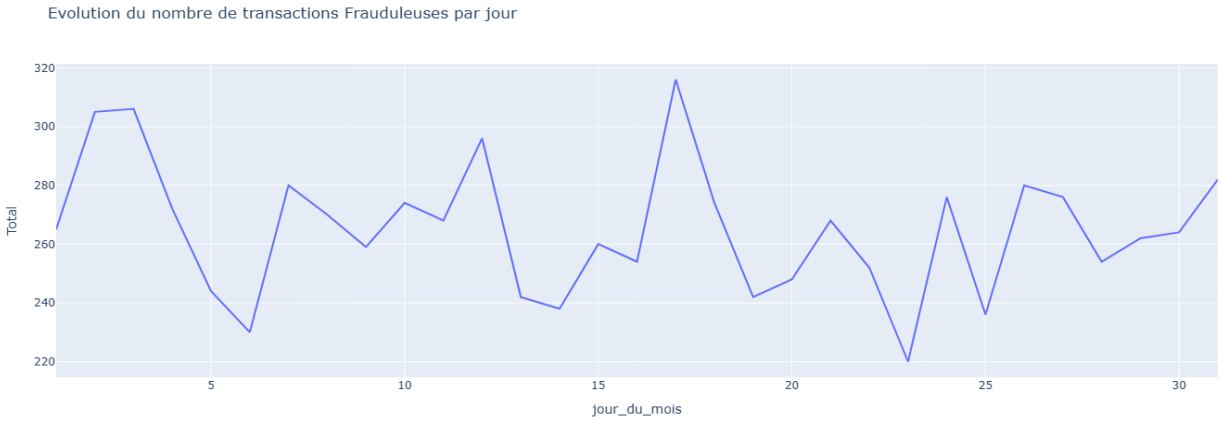
Figure 4: types de transactions par jour



II.3. Les transactions frauduleuses sur la plateforme PaySim

Les données disponibles nous permettent, entre autres, d'identifier les transactions dites frauduleuses et même suspecter d'être frauduleuses. Pour ce qui est des transactions frauduleuses, au total 8213 ont été recensé, soit 0,13% du total des transactions ce qui montre une très faible mais non négligeable activité des fraudeurs. Reparties suivant les jours du moi on remarque des activités irrégulières des fraudeurs avec des jours où le nombre de fraudes est particulièrement élevé, notamment vers le début du mois et autour du jour 17.

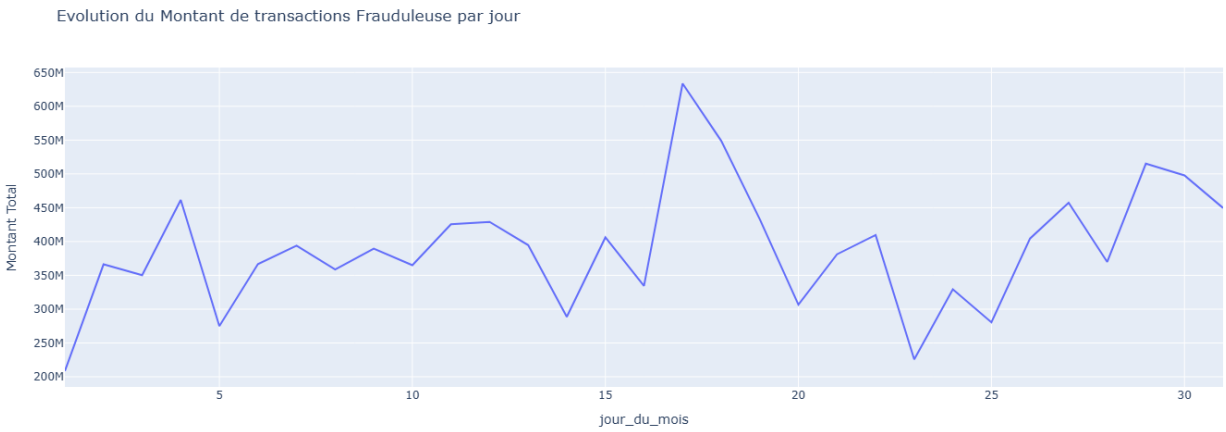
Figure 5: nombre de transactions frauduleuse par jour



Le montant total des transactions frauduleuses fluctue fortement au fil des jours, ce qui montre une fois de plus, une dynamique irrégulière des fraudes. On observe un pic important autour du jour 17, où le montant dépasse **600 millions.** l y a plusieurs creux, notamment autour des jours **6, 15, 20 et 24**, où le montant des fraudes chute sous les **300 millions.** Toutefois aucune

transaction frauduleuse n'est en deçà de 200 millions. Il faut également noter que contrairement au simple nombre de transactions frauduleuses, ici les montants varient de façon plus chaotique. Cela suggère que certaines journées avec **peu de transactions frauduleuses** peuvent néanmoins afficher **des montants très élevés**, ce qui pourrait indiquer la présence de fraudes à grande échelle ces jours-là

Figure 6: montant des transactions frauduleuses par jour



En ce qui concerne les transactions suspectées, elles sont au nombre de 16 ce qui représente une part très faible par rapport au nombre de transactions totale. Cependant sur le graphique . ; on peut remarquer que pareillement aux transactions identifiées de fraudes, le nombre de transactions suspectées est élevé au début de la deuxième moitié du mois (jour 16,17, 18) et vers la fin du mois.

II.4. Les transactions suspectées de fraude

Figure 7: transacttions suspectées

En s'intéressant aux fréquences des transactions, que ce soit du côté des clients ou des destinataires, il ressort qu'aucun client n'a fait plus de **3 transactions** dans le mois, donc les clients les plus actifs sont ceux qui ont eu à faire trois transactions. Du coté des destinataires, celui ayant reçu plus de transactions a à son actif 113 transactions, ce qui s'éleve à un montant total d'environ 7,8 millions, suivi de 5 autres destinataires ayant reçu plus de 100 transactions dans le mois. Cette analyse nous a permis de dresser **le top 10 des clients** les plus actifs ainsi que les destinataires les plus réceptifs(voir annexe).

II.5. Solde initial nul et solde final nul

Un regard porté sur le solde initial et final des transactions montre que plus du tiers (2 102 449) du total des transactions enregistrées dans le mois sont initiées avec des soldes nuls et le même nombre finissent leur transaction avec un solde nul également. Cette remarque traduit deux plausibles situations des transactions. En effet, si ces transactions concernent les mêmes clients cela pourrait traduire le fait que plus du tiers des transactions proviennent des comptes intermédiaires utilisés pour traiter des transactions sans conserver de solde en fin d'opération, comme les clients utilisant chaque crédit ou virement pour immédiatement rembourser une transaction passée(découvert, dette).

Par ailleurs en répartissant de manière journalière, les transactions à solde initial et final suivent la même tendance à quelques centimes près : une première période située dans la première moitié du mois caractérisé par un nombre de transactions à solde initiale (ou final) nul élevé et

la seconde moitié du mois caractérisé par de faible nombre de transactions à solde finale nul (ou initiale

Figure 8: transactions avec solde initial nul par jour

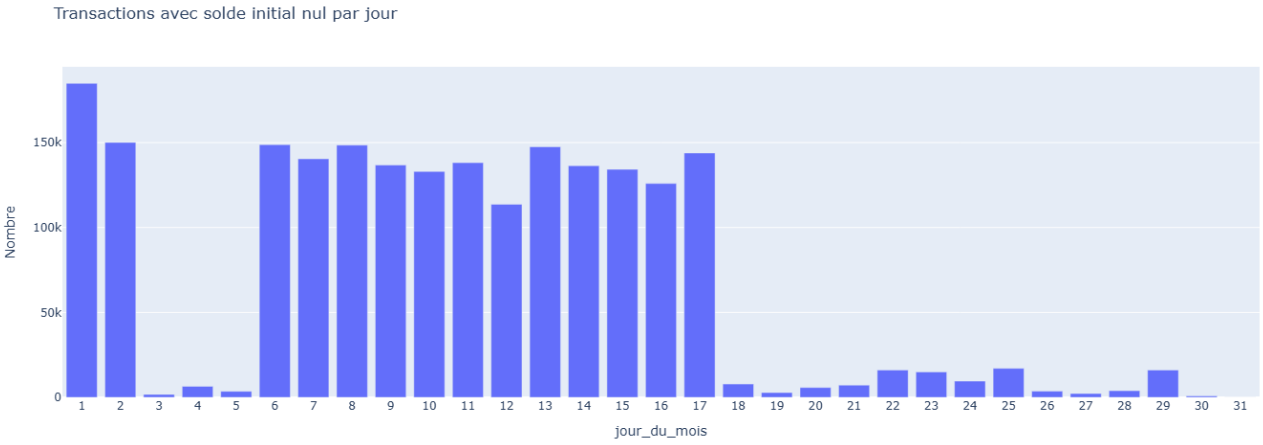
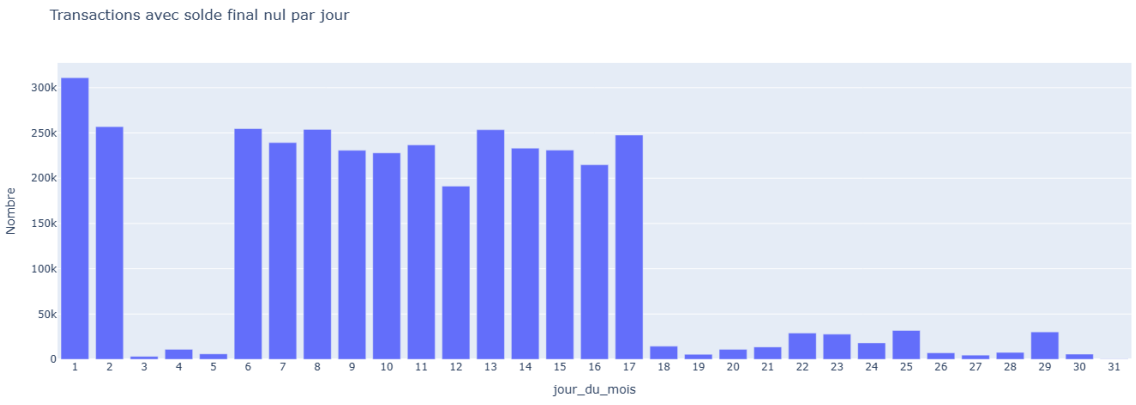


Figure 9:transactions avec solde final nul par jour



I. DASHBOARD ET LIENS IMPORTANTS

A partir de ces travaux de calculs d'indicateurs nous avons réalisé un dashboard que nous avons déployer sur un SAS (Dash) accessible à l'adresse https://paysim-transaction.onrender.com/app_pages/page1. Développé avec l'outil Dash de Python, ce Dashboard combine interactivité, performance et simplicité, il présente les indicateurs calculés dans le cadre de cette étude. On y trouve aussi les membres du groupe qui ont constitué l'équipe de conception du projet avec les rôles correspondants.

Liens importants :

- Dépôt GitHub principal : https://github.com/Teyan-Chris/Project-BigData-final
- Dépôt GitHub du Dashboard : https://github.com/kouloucrepin/paysim_transaction/
- Source des données : https://www.kaggle.com/datasets/ealaxi/paysim1
- Application Dash : https://paysim-transaction.onrender.com/app_pages/page1

CONCLUSION

Ce projet d'analyse des transactions financières, réalisé à partir des données synthétiques générées par la plateforme PaySim, a permis d'explorer en profondeur les activités transactionnelles et les risques de fraude dans un environnement simulé. En combinant des techniques de statistiques descriptives et des outils modernes comme Python et PySpark, nous avons identifié et analysé 10 indicateurs clés.

Les résultats ont mis en évidence des tendances significatives, telles que :

- Une activité transactionnelle intense en début de mois, corrélée aux cycles économiques courants (salaires, paiements de dettes).
- Des pics de fraudes autour du 17e jour, avec des montants dépassant les 600 millions, suggérant des opérations frauduleuses à grande échelle lors de périodes spécifiques.

- La présence de transactions à solde initial ou final nul dans plus d'un tiers des cas, indiquant potentiellement l'utilisation de comptes intermédiaires pour des opérations suspectes.
- Un taux de fraude global de 0,13%, bien que faible, soulignant la nécessité de systèmes de détection robustes pour identifier des anomalies même rares.

En guise de recommandations, il serait pertinent :

- De renforcer les contrôles durant les périodes à haut risque identifiées (début de mois, jour 17).
- D'approfondir l'analyse des comptes à solde nul pour détecter d'éventuels schémas frauduleux.
- D'intégrer des algorithmes de machine learning pour affiner la détection des transactions suspectes.

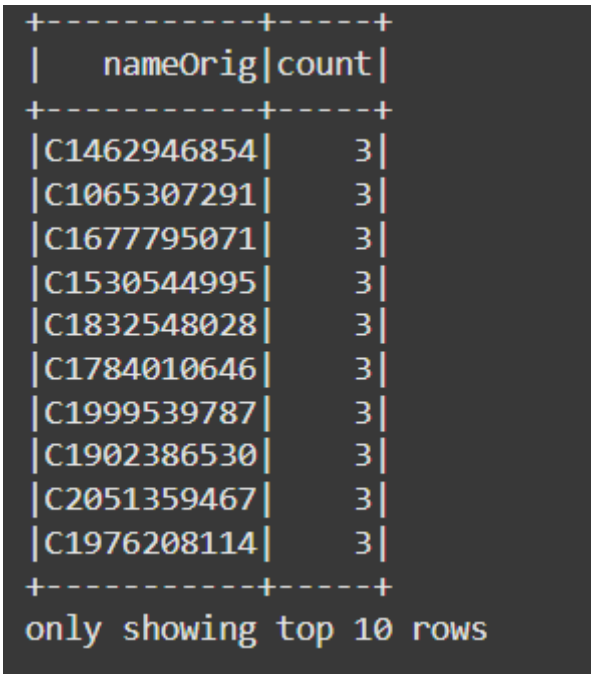
En perspective, ce travail ouvre la voie à des recherches futures, notamment l'intégration de données réelles pour valider les modèles. Ce projet démontre ainsi l'utilité des données synthétiques pour étayer des stratégies de transactions financières, tout en respectant les contraintes de confidentialité des données financières.

ANNEXES

Annexe 1: TOP 10 des destinataires les plus recepteurs

nameDest	Effectif	Amount
C1789550256	99	1.778853e+08
C665576141	105	8.874938e+07
C1286084959	113	7.742894e+07
C2083562754	102	5.307394e+07
C1590550415	101	4.320610e+07
C985934102	109	4.242289e+07
C1023714065	97	4.240292e+07
C248609774	101	4.068016e+07
C451111351	99	3.745303e+07
C1360767589	98	3.512129e+07

Annexe 2: TOP 10 des clients avec le plus de transactions



```
+-----------+-----+
|   nameOrig|count|
+-----------+-----+
|C1462946854|    3|
|C1065307291|    3|
|C1677795071|    3|
|C1530544995|    3|
|C1832548028|    3|
|C1784010646|    3|
|C1999539787|    3|
|C1902386530|    3|
|C2051359467|    3|
|C1976208114|    3|
+-----------+-----+
only showing top 10 rows
```