



FRAUD ALERT



CONFIRM

Click here for more information

AIA - BLOC_03 Fraud Detection

Éric NGUYEN



Brief Marketing



Détection de Fraude

L'équipe Business a besoin de monitorer la fraude en temps réel et d'endiguer la perte économique.

L'équipe Data devra créer une application qui :

- détecte la fraude en temps réel
- notifie le Business dès qu'une fraude est détectée
- envoie un rapport récapitulatif des fraudes de la journée précédente.

Target :

Fraud / non Fraud

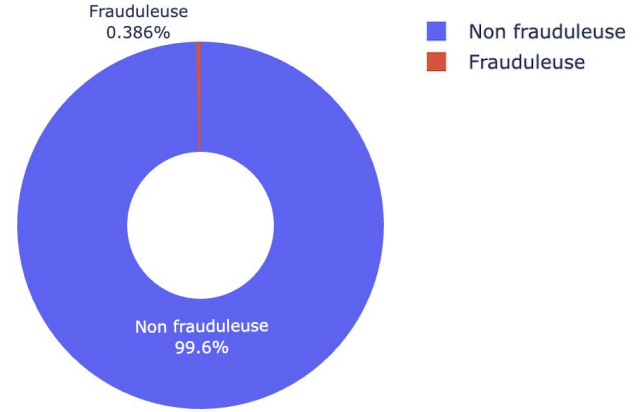
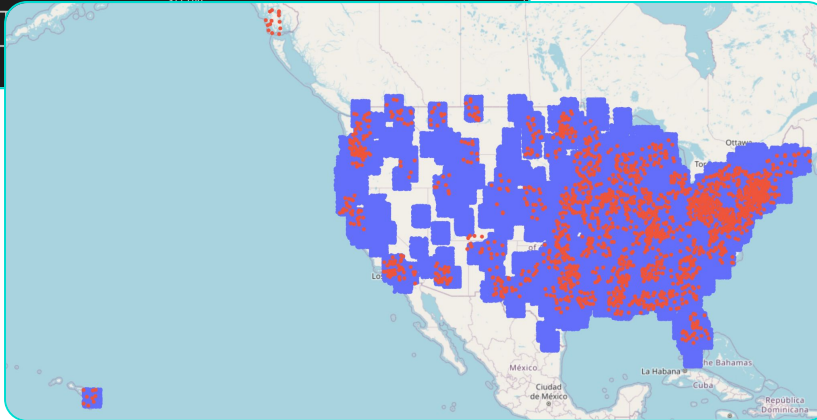


EDA Dataset Kaggle

Data Summary				Data Types			
Dataframe		Values		Column Type		Count	
Number of rows		555719		string		12	
Number of columns		23		int64		6	
				float64		5	

number										
column	NA	NA %	mean	sd	p0	p25	p50	p75	p100	hist
Unnamed: 0	0	0	277900	160400	0	138900	277900	416800	555700	
cc_num	0	0	4.178e+17	1.31e+18	604200000000	1800000000000000	3521000000000000	4635000000000000	4.992e+18	
amt	0	0	69.39	156.7	1	9.63	47.29	83.01	22770	
zip	0	0	48040	26800	1257	26290	48170	72010	99210	
lat	0	0	38.54	5.061	20.03	34.67	41.89	65.69		
long	0	0	-90.23	13.72	-165.7	-96.8	-87.48	-80.18	-67.95	
city_pop	0	0	80220	300400	23	741	2400	19600	2807000	
unix_time	0	0	13810000	5201000	13720000	13760000	13810000	13860000	13890000	
merch_lat	0	0	38.54	5.096	19.03	34.76	39.38	41.95	66.68	
merch_longitude	0	0	-90.23	13.73	-166.7	-96.91	-87.45	-80.26	-66.95	
is_fraud	0	0	0.00386	0.06201	0	0	0	0	1	

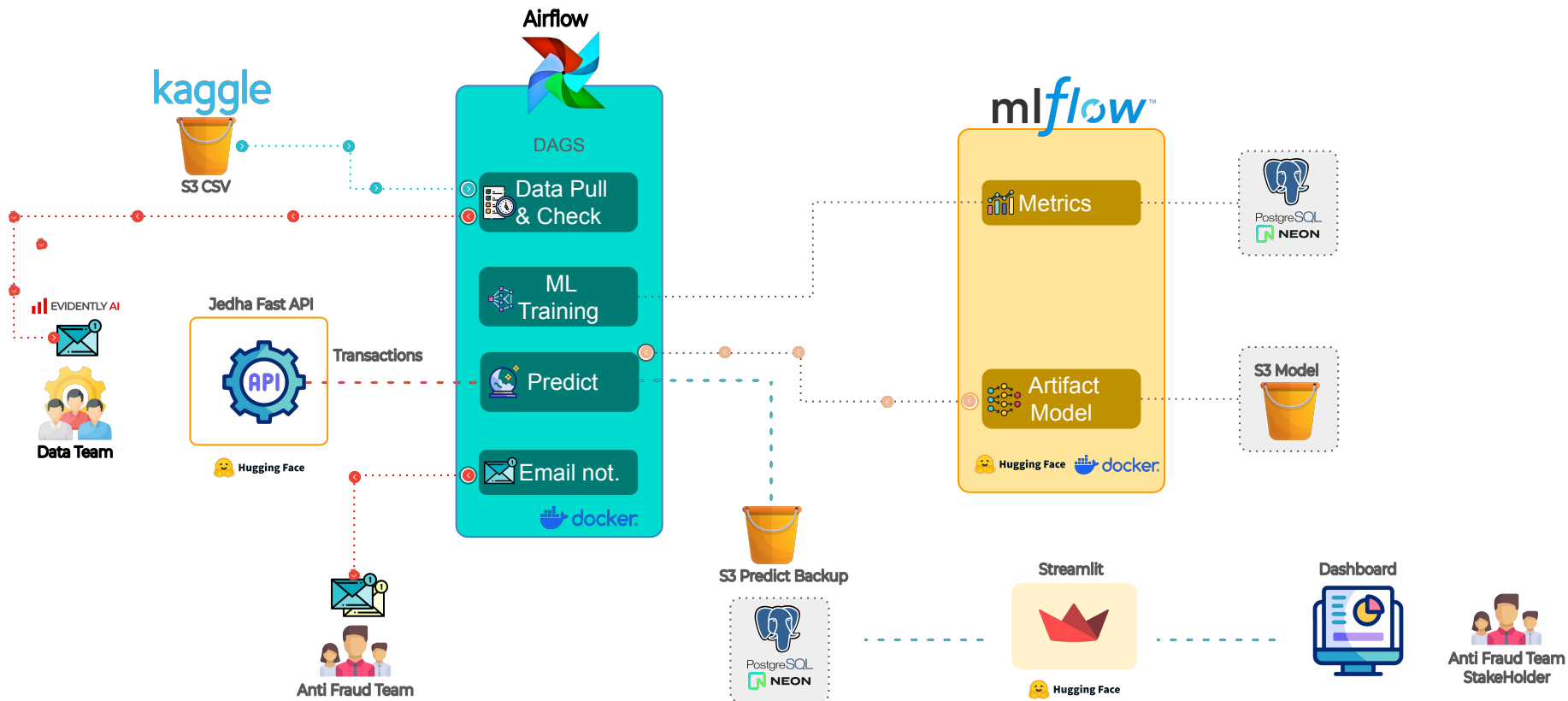
string		
column	NA	NA %
trans_date_trans_time	0	0



- Dataset déséquilibré
- Pas de données manquantes
- Des outliers
- Données géographiques à creuser



Architecture - CI/CD Fraud Detection





Malgré 2 alertes, pas de problème :

ce sont des colonnes sans importances qui ne sont pas nécessaires au machine learning.

Colonnes concernées : unnamed / unix_time



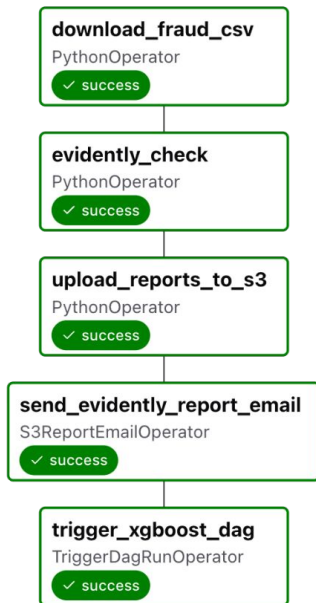
AIRFLOW



DAGs

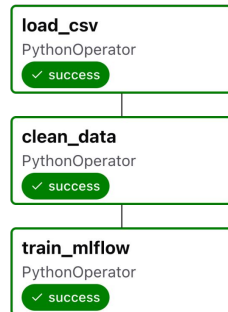
1

DATA QUALITY



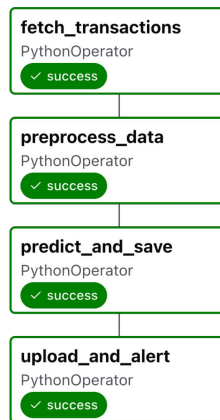
2

TRAIN



3

PREDICTION REAL TIME

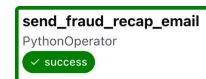


Toutes les 5 min



4

RECAP



9h du matin





Email Notifications



Gmail interface showing a list of email notifications. The list includes messages from 'Airflow Fraud Detection' and 'Airflow || Evidently: Data Quality Fraud'. The messages are categorized by status: 'Alerte Fraude détectée !', 'Rien à Signaler', and 'Transaction Normale'.

Rechercher dans les messages

Nouveau message

Boîte de réception 148

Messagerie suivie

En attente

Messagerie envoyée

Brouillons

Achats

Plus

Libellés

1-50 sur 216

Fraud Detection - Airflow Fraud Detection : Alerte Fraude détectée ! ALERTE FRAUDE DÉTECTÉE 1 transaction(s) suspecte(s) détectée(s). ID Transaction Marchand Catégorie Montant Ville ...

Fraud Detection - Airflow Fraud Detection : Rien à Signaler Transaction Normale Aucune fraude détectée sur 1 transaction(s) analysée(s). ID Transaction Marchand Catégorie Montant Ville S...

Fraud Detection - Airflow Fraud Detection : Rien à Signaler Transaction Normale Aucune fraude détectée sur 1 transaction(s) analysée(s). ID Transaction Marchand Catégorie Montant Ville S...

Fraud Detection - Airflow Fraud Detection : Rien à Signaler Transaction Normale Aucune fraude détectée sur 1 transaction(s) analysée(s). ID Transaction Marchand Catégorie Montant Ville S...

Fraud Detection - Airflow Fraud Detection : Rien à Signaler Transaction Normale Aucune fraude détectée sur 1 transaction(s) analysée(s). ID Transaction Marchand Catégorie Montant Ville S...

Fraud Detection - Airflow Fraud Detection : Rien à Signaler Transaction Normale Aucune fraude détectée sur 1 transaction(s) analysée(s). ID Transaction Marchand Catégorie Montant Ville S...

Fraud Detection - Airflow Fraud Detection : Rien à Signaler Transaction Normale Aucune fraude détectée sur 1 transaction(s) analysée(s). ID Transaction Marchand Catégorie Montant Ville S...

Fraud Detection - Airflow || Evidently: Data Quality Fraud - Airflow: Evidently Data Quality Check Statut Global: Des problèmes ont été déte...

Fraud Detection - Airflow Fraud Detection : Alerte Fraude détectée ! Boîte de réception x

16:18 (il y a 2 minutes)

Fraud Detection À moi

Ce message semble être en anglais Traduire en français

ALERTE FRAUDE DÉTECTÉE

1 transaction(s) suspecte(s) détectée(s).

ID Transaction	Marchand	Catégorie	Montant	Ville	Statut
2a5047ad342d8ca9d80e9c4cc0e08d	Bolton Ltd	shopping_pos	774.10	Thral	FRAUDE

Voir le fichier complet : Lien S3

Dashboard de suivi : Voir Dashboard

Boîte de réception 148 - Airflow || Evidently: Data Quality Fraud Boîte de réception x

16:18 (il y a 2 minutes)

Airflow À moi

Airflow: Evidently Data Quality Check

Statut Global: Des problèmes ont été détectés

Date: 30/10/2023 à 15:11:45

Résumé

Total lignes
555,719
277,859

- Airflow Fraud Detection : Rien à Signaler Boîte de réception x

16:12 (il y a 3 minutes)

Fraud Detection À moi

Transaction Normale

Aucune fraude détectée sur 1 transaction(s) analysée(s).

ID Transaction	Marchand	Catégorie	Montant	Ville	Statut
8114c8e33fe9355abc34778fe9a45f06	Mueller, Gerhold and Mueller	health_fitness	23.36	Issaquah	LÉGITIME

Voir le fichier complet : Lien S3

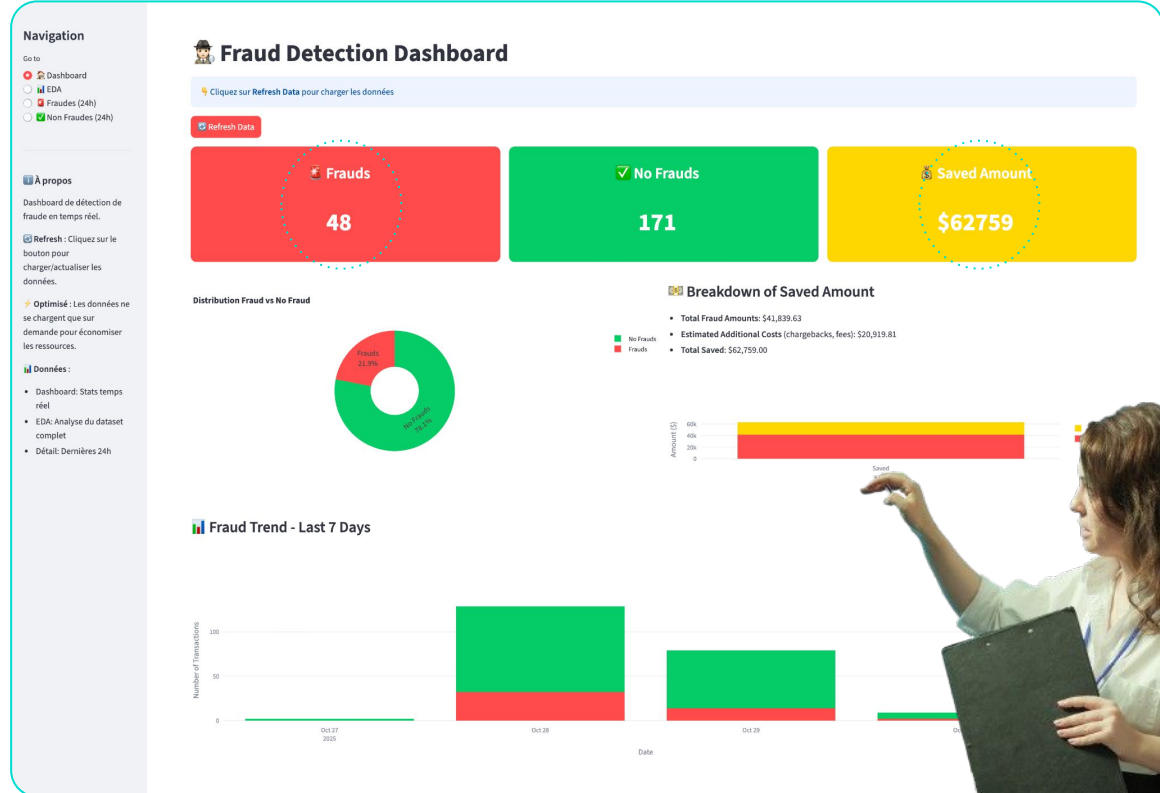
Dashboard de suivi : Voir Dashboard



Dashboard



Demo Dashboard



StakeHolder



Démo Vidéo

Evidence Test Suite - Results Overview

Summary Overview

Test Suite	Passed	Failed	Skipped	Not Run
Test Suite	12	0	0	0

Details of Tests

Test Name	Status	Message
Test 1	Passed	Test 1 passed successfully.
Test 2	Passed	Test 2 passed successfully.
Test 3	Passed	Test 3 passed successfully.
Test 4	Passed	Test 4 passed successfully.
Test 5	Passed	Test 5 passed successfully.
Test 6	Passed	Test 6 passed successfully.
Test 7	Passed	Test 7 passed successfully.
Test 8	Passed	Test 8 passed successfully.
Test 9	Passed	Test 9 passed successfully.
Test 10	Passed	Test 10 passed successfully.
Test 11	Passed	Test 11 passed successfully.
Test 12	Passed	Test 12 passed successfully.
Test 13	Passed	Test 13 passed successfully.
Test 14	Passed	Test 14 passed successfully.
Test 15	Passed	Test 15 passed successfully.
Test 16	Passed	Test 16 passed successfully.

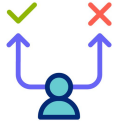




What's next?



Déployer le Modèle (MLFLOW) microservice à la passerelle de paiement (blocker avant la Fraude effective)



Feedback Faux positif / Faux Négatif (bouton sur le Dashboard) - réentraînement du modèle en continue.



Pipeline Jenkins CI/CD (test unit / test integration)



Jedha

Any questions ?

