

# **Hackathon 2025 : Comment démasquer les fraudeurs ?**

---

*Par Nexialog Consulting*

*Par Nexialog Consulting – Sujet en Apprentissage Non Supervisé*

## **Contexte**

NexiaBank est un leader européen du crédit à la consommation, avec plus de 250 collaborateurs et une présence dans 35 pays. Chaque jour, des milliers de paniers d'achat sont financés via ses filiales. En finançant des millions de transactions issues de divers canaux (magasins, concessions automobiles, en ligne), l'entreprise est exposée à des risques, dont la fraude.

Dans une optique de renforcement proactif de la détection de comportements suspects, l'analyse non supervisée permet d'identifier des paniers clients atypiques, sans recourir à un historique de fraudes connues. Cette approche est pertinente pour détecter des schémas émergents mal étiquetés, ou encore inconnus.

L'équipe RiskData, au sein de la direction des risques, souhaite expérimenter ces approches exploratoires pour compléter les modèles de détection de fraude existants.

## **Objectif du challenge**

L'objectif de ce challenge est de **concevoir une solution permettant d'identifier automatiquement les paniers clients les plus atypiques**, à partir des données paniers fournis par l'un des partenaires de cette banque. Ces paniers considérés comme « anormaux » ou « hors norme » pourraient révéler des situations de fraude potentielle ou d'utilisation abusive, nécessitant une attention particulière.

Les participants devront développer une pipeline d'analyse non supervisé capable de :

- Calculer un score d'anomalie, reflétant le degré d'atypie du panier (un score élevé signifiant une anomalie plus marquée).
- Générer un fichier de résultats structuré et exploitable par les équipes métier pour faciliter l'investigation et la prise de décision.

Au-delà de l'aspect technique, l'enjeu clé du challenge est de **proposer une méthode innovante, convaincante et opérationnelle, capable d'améliorer significativement la**

**détection de la fraude.** Il s'agira également de valoriser la pertinence métier des résultats auprès des directions impliquées.

## Descriptions des données

### Base de données

La base contient une liste d'achats effectués chez un des partenaires financés. Les informations décrivent exclusivement le contenu du panier.

Pour chaque observation de la base, il y a 146 colonnes dont 144 peuvent être regroupées en 6 catégories :

- ITEM,
- CASH\_PRICE,
- MAKE,
- MODEL,
- GOODS\_CODE,
- NBR\_OF\_PROD\_PURCHAS.

Le panier se décompose au maximum en 24 items. Par exemple, si un panier contient 3 items alors toutes les informations relatives à ces 3 items seront renseignées dans les COLONNES ITEM1, ITEM2, ITEM3, CASH\_PRICE1, CASH\_PRICE\_2, CASH\_PRICE3, MAKE1, MAKE2, MAKE3, MODEL1, MODEL2, MODEL3, GOODS\_CODE1, GOODS\_CODE2, GOODS\_CODE3, NBR\_OF\_PROD\_PURCHAS1, NBR\_OF\_PROD\_PURCHAS2 ET NBR\_OF\_PROD\_PURCHAS3. Les variables restantes (celles avec un indice > 3) seront vides.

Un item correspond à un produit ou un regroupement de produits équivalents. Par exemple, si un panier contient 3 iPhone 14, alors ces 3 produits seront regroupés dans un seul item. En revanche, si le client achète 3 produits iPhone différents, alors nous considérerons ici 3 items.

La variable NB\_OF\_ITEMS correspond au nombre d'items dans le panier, tandis que la somme des variables NBR\_OF\_PROD\_PURCHAS correspond au nombre de produits.

### Description des variables en entrée (X)

Variable	Description	Exemple
ID (NUM)	Identifiant unique	1
ITEM1 A ITEM24 (CHAR)	Catégorie du bien de	Computer

	l'item 1 à 24	
CASH_PRICE1 A CASH_PRICE24 (NUM)	Prix de l'item 1 à 24	850
MAKE1 A MAKE24 (CHAR)	Fabriquant de l'item 1 à 24	Apple
MODEL1 A MODEL24 (CHAR)	Description du modèle de l'item 1 à 24	Apple Iphone XX
GOODS_CODE1 A GOODS_CODE24 (CHAR)	Code de l'item 1 à 24	2378284364
NBR_OF_PROD_PURCHAS1 À NBR_OF_PROD_PURCHAS24 (NUM)	Nombre de produits dans l'item 1 à 24	2
NB_OF_ITEMS (NUM)	Nombre total d'items	7

### Taille de la base

**Taille :** 115 988 observations, 146 colonnes.

### Echantillons

- **Echantillon d'entraînement :** Taille : 92 790 observations, 146 colonnes.
- **Echantillon de test :** Taille : 23 198 observations, 146 colonnes.

### Consignes générales

Les participants sont invités à concevoir une solution complète d'analyse non supervisée pour identifier des comportements d'achat atypiques. Il est important montrer les facteurs qui servent à suspecter la fraude pour les paniers identifiés. Ils pourront mobiliser plusieurs familles de méthodes.

**Bon courage et amusez-vous bien !!**