

# Analyse de Données Avancée

Airbnb à New York : Économie de partage ou para-hôtellerie ?

Marc-Antony TCHICOU

Kelvin WONG

Mouhamed CISSE

Alexis SCHNEIDER

Mohammed-Yahya DAMI

ENSIIE – Année 2025-2026

## Synthèse (Abstract)

Ce projet d'analyse de données vise à déconstruire la structure du marché Airbnb à New York pour l'année 2019, une période charnière de saturation de l'offre avant la pandémie. L'objectif est de déterminer dans quelle mesure la plateforme s'éloigne de sa promesse initiale d'économie de partage pour devenir un vecteur d'industrialisation locative. Allant du nettoyage des données à la modélisation spatiale (DBSCAN) et factorielle (ACP, CCA), nous mettons en évidence une dualité de marché. Les résultats révèlent une fracture géographique et économique nette : un centre hyper-professionnalisé opérant comme une "usine à touristes" et une périphérie résidentielle conservant une logique d'usage authentique. Ce rapport détaille la méthodologie, les algorithmes utilisés et l'interprétation économique des résultats.

## **Table des matières**

# 1 Introduction et Contexte de l'Étude

Depuis sa création en 2008, Airbnb a radicalement transformé l'industrie du tourisme en promettant une expérience de voyage centrée sur l'humain et l'immersion locale. Fondée sur le principe de l'*économie de partage*, la plateforme permettait initialement à des particuliers de louer une chambre ou leur logement principal pour compléter leurs revenus.

Cependant, dans des métropoles mondiales comme New York, ce modèle a évolué. La rentabilité élevée de la location courte durée a attiré des investisseurs et des opérateurs professionnels, soulevant une question fondamentale sur la nature actuelle de la plate-forme.

## 1.1 Problématique et Enjeux

Airbnb est-il toujours un outil de complément de revenu pour les résidents, ou est-il devenu une place de marché pour une industrie para-hôtelière dérégulée ? Cette étude cherche à quantifier cette transformation en répondant à la problématique suivante : **Comment l'activité des hôtes se structure-t-elle géographiquement et économiquement, et quels sont les facteurs déterminants de la performance sur ce marché hybride ?**

## 1.2 Jeu de Données et Démarche Méthodologique

L'analyse s'appuie sur le dataset *New York City Airbnb Open Data (2019)*, comportant environ 49 000 annonces. Pour traiter ce volume et extraire des tendances lourdes, nous avons adopté une démarche en quatre temps :

1. **Préparation et Statistiques Descriptives** : Nettoyage rigoureux et segmentation des acteurs.
2. **Analyse Factorielle (ACP CCA)** : Réduction de dimension pour identifier les stratégies sous-jacentes.
3. **Segmentation (Clustering)** : Classification non-supervisée pour établir des profils types.
4. **Analyse Spatiale** : Compréhension de la dynamique territoriale et de la ségrégation du marché.

# 2 Panorama du Marché : Préparation et Statistiques

La première phase du projet est fondamentale pour la robustesse des analyses ultérieures. Elle consiste à transformer des données brutes hétérogènes en indicateurs fiables.

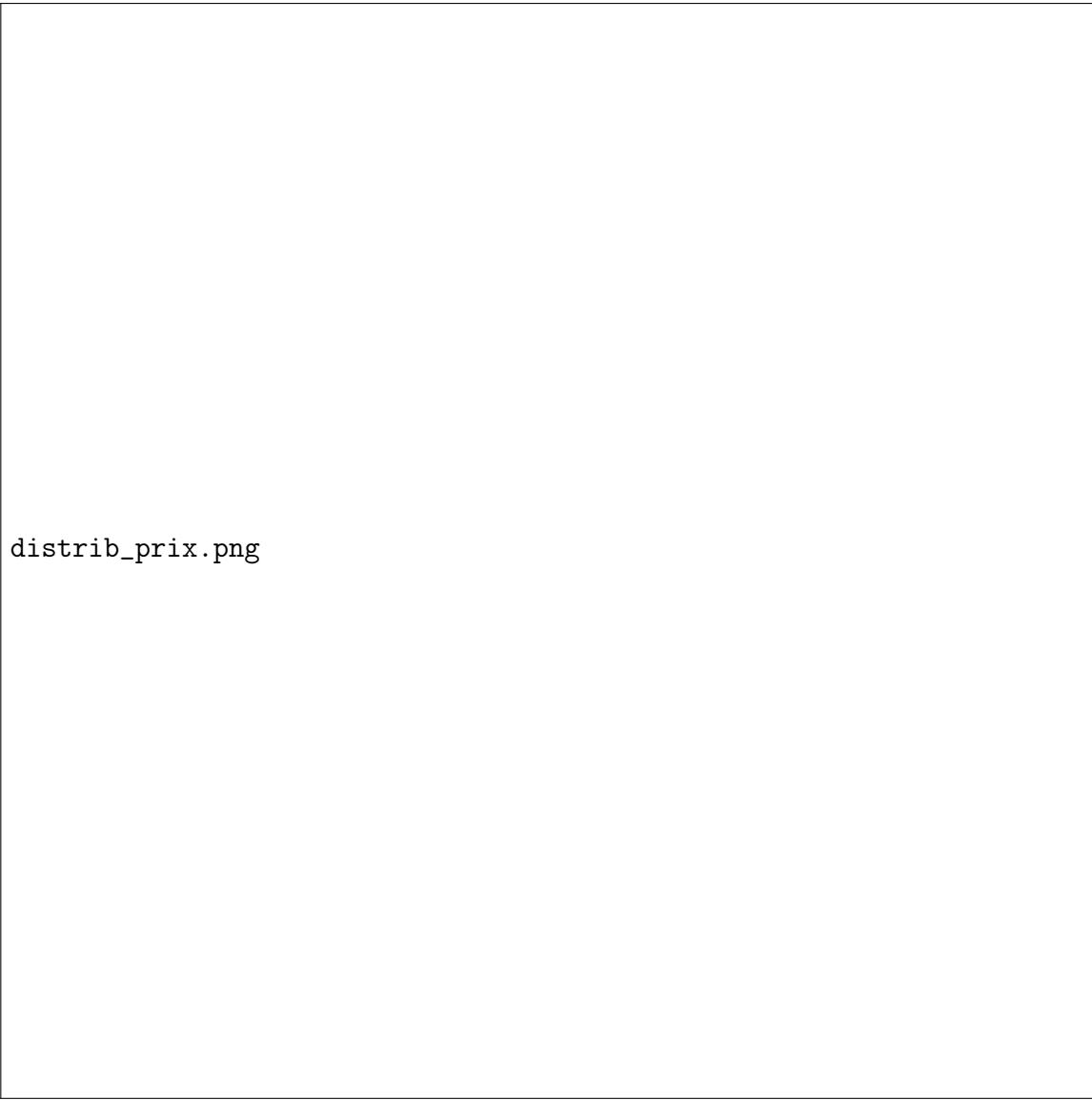
## 2.1 Protocole de Nettoyage et Feature Engineering

Le dataset initial comportait du bruit statistique qu'il a fallu traiter pour éviter des biais dans les modèles de Machine Learning.

- **Traitement des valeurs aberrantes** : Une suppression des prix nuls (0\$) a été effectuée. De plus, la distribution des prix étant fortement asymétrique (étalée vers la droite avec des biens de luxe extrêmes), une transformation logarithmique (`log_price`) a été appliquée pour normaliser la variable.
- **Segmentation Métier des Hôtes** : Une simple distinction binaire étant insuffisante, nous avons créé la variable `host_type` basée sur le volume d'annonces (`calculated_host_listings_count`) et la disponibilité (`availability_365`). Trois segments ont été définis :
  - *Occasionnel* : 1 annonce.
  - *Multi-propriétaire* : 2 à 5 annonces.
  - *Professionnel* : Plus de 5 annonces.

## 2.2 Analyse Descriptive : La Fracture du Marché

L'analyse univariée et bivariée met en évidence une polarisation géographique et économique.



distrib\_prix.png

FIGURE 1 – La fracture des prix : Manhattan affiche une médiane élevée et une forte volatilité (Luxe), contrairement à la périphérie.

Comme illustré par la Figure 1, Manhattan affiche des médianes nettement supérieures et une volatilité extrême comparée aux arrondissements périphériques. Par ailleurs, un **paradoxe de l'offre** apparaît (voir Figure 2) : bien que les hôtes "Occasionnels" soient majoritaires en nombre (66%), ils ne pèsent que très peu dans la disponibilité temporelle réelle (78 jours loués/an). Les "Professionnels" assurent une disponibilité quasi-permanente (255 jours/an).

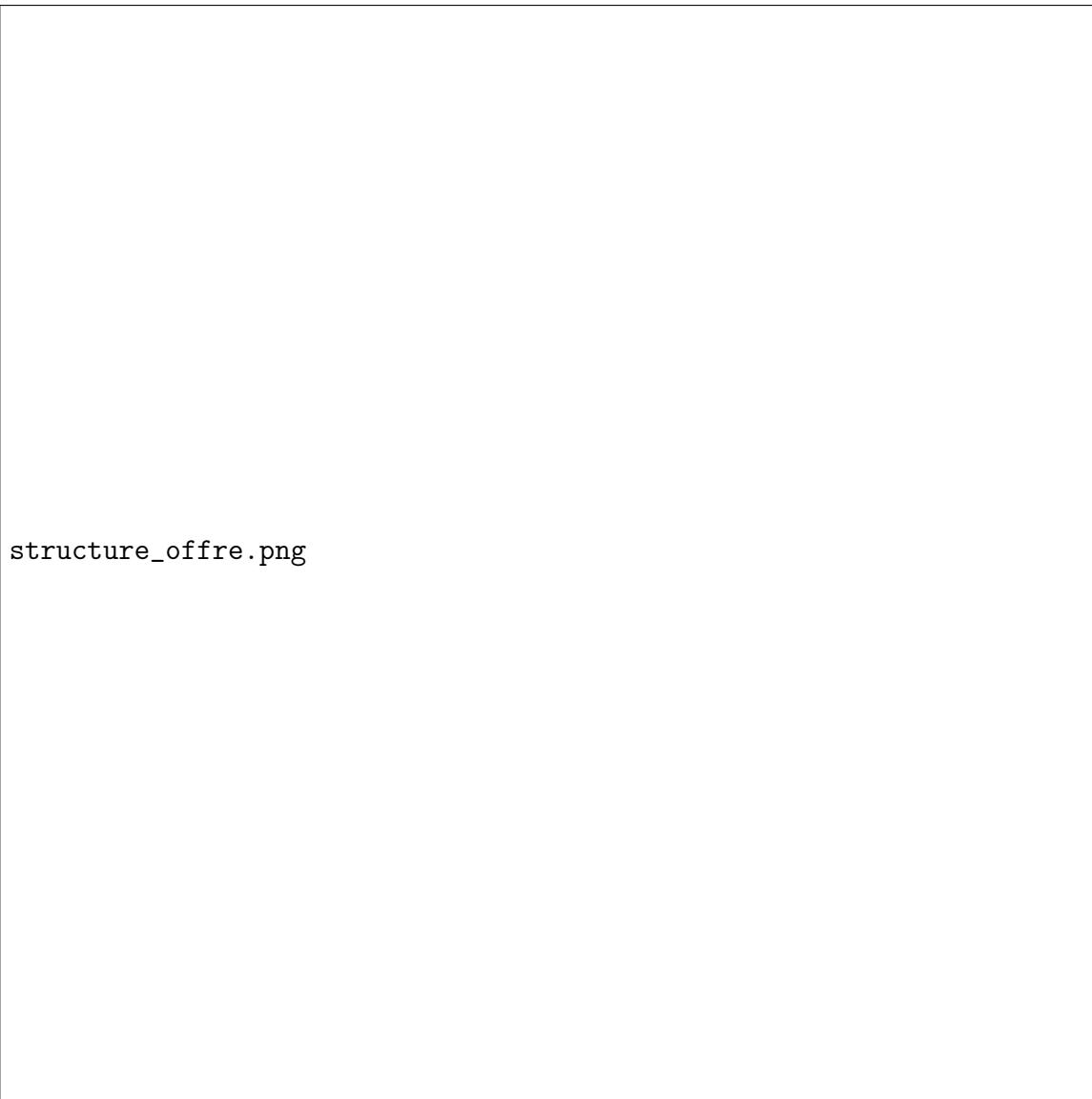


FIGURE 2 – Le paradoxe de l’offre : Les occasionnels sont nombreux mais louent peu souvent, contrairement aux Pros.

### 2.3 Validation Statistique

Le test de Kruskal-Wallis confirme que les différences de prix et de disponibilité entre les types d’hôtes sont statistiquement significatives ( $p\text{-value} \approx 0$ ). Cela valide l’hypothèse que les multi-propriétaires et les occasionnels n’opèrent pas sur le même marché économique.

## 3 Analyses Factorielles (ACP et CCA)

Afin de comprendre les structures latentes du marché sans a priori, nous avons déployé des méthodes d’analyse de données multivariées.

### 3.1 Analyse en Composantes Principales (ACP)

L'ACP a été appliquée sur les variables quantitatives décrivant l'activité (`price`, `minimum_nights`, `availability_365`, etc.). L'analyse révèle deux dimensions principales expliquant plus de 50% de l'inertie.

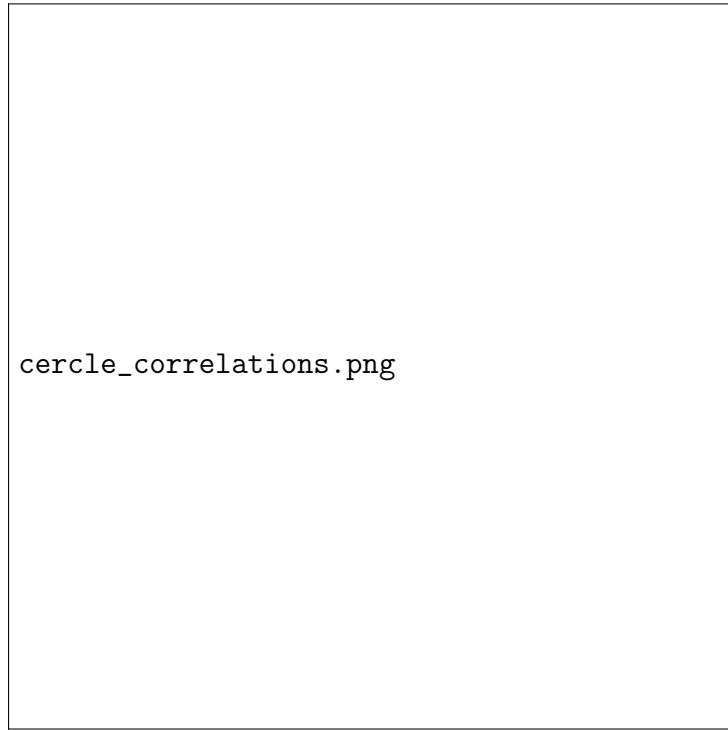


FIGURE 3 – Le cercle des corrélations met en évidence deux axes : l'intensité commerciale (horizontal) et la structure d'offre (vertical).

- **Axe 1 (28.2% variance) - L'Intensité Commerciale** : Oppose les logements à forte rotation (beaucoup d'avis) aux logements "dormants".
- **Axe 2 (23.3% variance) - La Barrière à l'Entrée** : Distingue l'activité professionnelle (disponible toute l'année, gestion de parc) de l'activité ponctuelle.

La projection des individus (Figure 4) montre une déconnexion nette : les professionnels se regroupent dans le quadrant "Haute Disponibilité", tandis que les occasionnels sont dispersés.

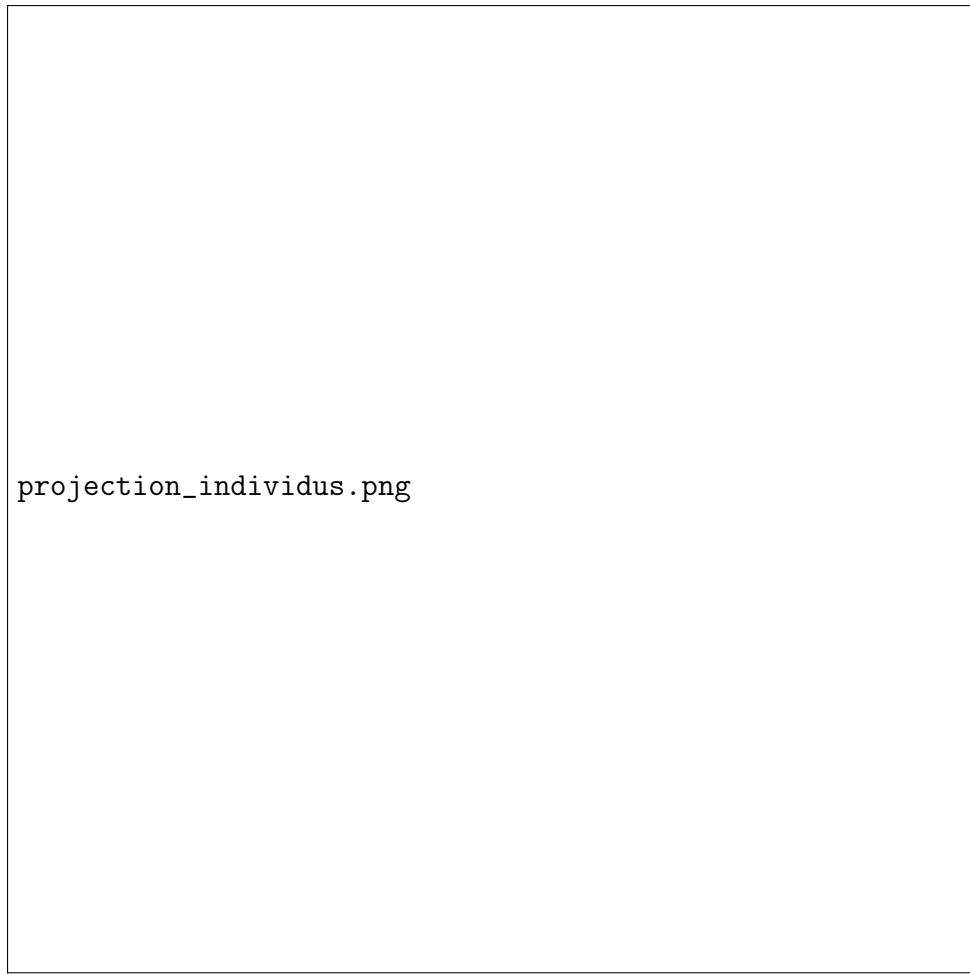
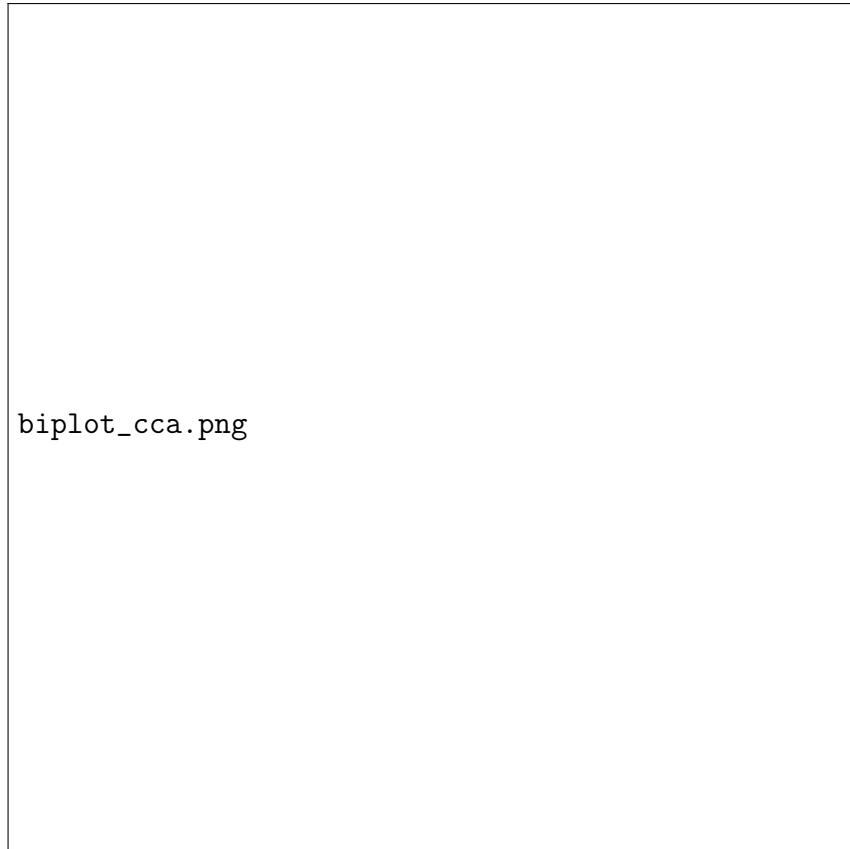


FIGURE 4 – Projection des individus : Les professionnels (en rouge/vert) se distinguent clairement des occasionnels (bleu).

### 3.2 Analyse Canonique des Corrélations (CCA)

La CCA nous a permis d'étudier le lien entre la *stratégie de l'hôte* (groupe X) et la *performance marché* (groupe Y). L'analyse révèle une corrélation forte sur le premier axe, interprétable comme l'axe des "**Opérateurs de Flux**". Il démontre que maximiser la disponibilité est le levier le plus puissant pour maximiser l'engagement client. Une seconde dimension lie le nombre d'annonces au prix, illustrant une logique de "Rente Immobilière".



biplot\_cca.png

FIGURE 5 – La CCA illustre deux stratégies gagnantes : l’usine à touristes (axe horizontal) et l’empire immobilier (axe vertical).

““latex ““latex

## 4 Segmentation comportementale et structure du marché Airbnb

L’analyse descriptive met en évidence des différences significatives entre catégories d’hôtes, mais elle ne permet pas d’identifier les logiques économiques profondes à l’œuvre. Afin de dépasser cette lecture administrative, nous avons appliqué un clustering K-Means sur les axes factoriels issus de l’Analyse en Composantes Principales (PCA). La méthode du coude a conduit à un nombre optimal de quatre clusters, révélant une segmentation comportementale claire du marché Airbnb new-yorkais.

### 4.1 Méthodologie de segmentation

Le clustering est réalisé sur les coordonnées des annonces projetées sur les deux premières composantes principales, qui résument l’essentiel de la variance comportementale. Ces axes opposent principalement :

- une logique d’usage intensif et commercial du logement,

— une logique d'usage résidentiel et occasionnel.  
 Cette approche permet de segmenter les annonces selon leur comportement réel, indépendamment du statut déclaré de l'hôte.

## 4.2 Choix du k optimal

La méthode du coude, fondée sur l'analyse de la décroissance de l'inertie intra-cluster en fonction du nombre de groupes, a conduit à un nombre optimal de quatre clusters, révélant une segmentation comportementale claire du marché Airbnb new-yorkais.

## 4.3 Visualisation des clusters sur les axes factoriels

La Figure ?? présente la projection des annonces sur les deux premières composantes principales de l'ACP, colorées selon leur appartenance aux clusters issus du K-Means.

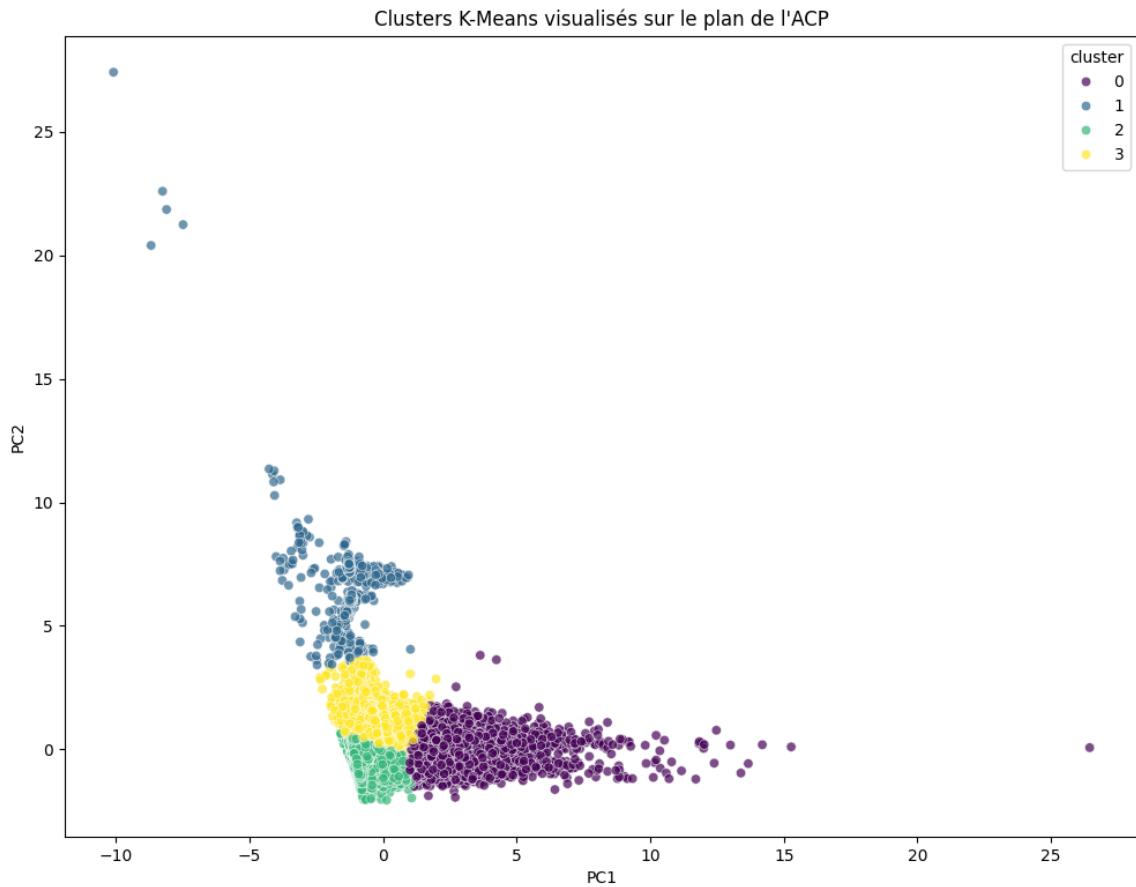


FIGURE 6 – Projection des clusters K-Means sur les deux premiers axes de l'ACP

Cette visualisation met en évidence une structuration nette de l'espace factoriel. Le Cluster 1 se distingue clairement par une position associée à une forte disponibilité et à des niveaux de prix élevés, caractéristiques d'un usage commercial intensif du logement. À l'inverse, le Cluster 2 occupe une zone opposée, traduisant un usage résidentiel et occasionnel. Les

Clusters 0 et 3 se situent dans des régions intermédiaires, confirmant leur rôle de profils hybrides entre partage et exploitation semi-professionnelle.

La bonne séparation visuelle des groupes valide la pertinence du choix des axes factoriels et du partitionnement en quatre clusters.

## 4.4 Caractérisation quantitative des clusters

Le Tableau ?? présente les moyennes des principales variables numériques par cluster.

TABLE 1 – Moyenne des variables numériques par cluster

Cluster	Prix moyen (\$)	Nuits min.	Disponibilité (jours/an)	Nombre d'avis
0	117	2	167	95
1	257	64	285	2.8
2	121	4	25	8
3	181	12	270	13

Ces statistiques révèlent des profils nettement différenciés, que l'on peut interpréter économiquement.

## 4.5 Interprétation économique des clusters

### Cluster 1 : Business para-hôtelier

Le Cluster 1 correspond à la forme la plus industrialisée du marché Airbnb. Il affiche le prix moyen le plus élevé (257\$) et une disponibilité quasi permanente (285 jours par an). La durée minimale de séjour élevée (64 nuits) indique un positionnement orienté vers une clientèle de moyen terme.

Ce cluster est composé à 86.7% d'opérateurs professionnels. Les logements ne sont pas habités par les hôtes et sont exploités comme des actifs locatifs dédiés. Il s'agit d'une activité économique structurée, assimilable à de la para-hôtellerie.

### Cluster 2 : Économie du partage authentique

À l'opposé, le Cluster 2 incarne le modèle originel d'Airbnb. Il est majoritairement composé d'hôtes occasionnels (78.7%), avec des prix modérés (121\$) et une disponibilité très faible (25 jours par an).

Le logement est occupé par l'hôte la majeure partie de l'année. Airbnb constitue ici un complément de revenu ponctuel et non une activité commerciale principale.

### Cluster 0 : Partage intensif

Le Cluster 0 représente un profil intermédiaire. Les prix sont bas (117\$), les séjours courts et le nombre d'avis élevé, indiquant une forte rotation des voyageurs.

Ce cluster correspond à des hôtes actifs, utilisant intensivement leur logement personnel sans atteindre un niveau de professionnalisation complet.

### Cluster 3 : Profils hybrides semi-professionnels

Le Cluster 3 se caractérise par une forte disponibilité (270 jours par an) et des durées minimales de séjour intermédiaires (12 nuits). Il regroupe un mélange d'hôtes occasionnels, multi-annonces et professionnels.

Ce profil cible une clientèle spécifique, comme les expatriés ou voyageurs d'affaires, avec une logique proche de la location meublée de moyen terme.

## 4.6 Répartition des profils d'hôtes par cluster

La Figure ?? illustre la répartition des types d'hôtes dans chaque cluster.

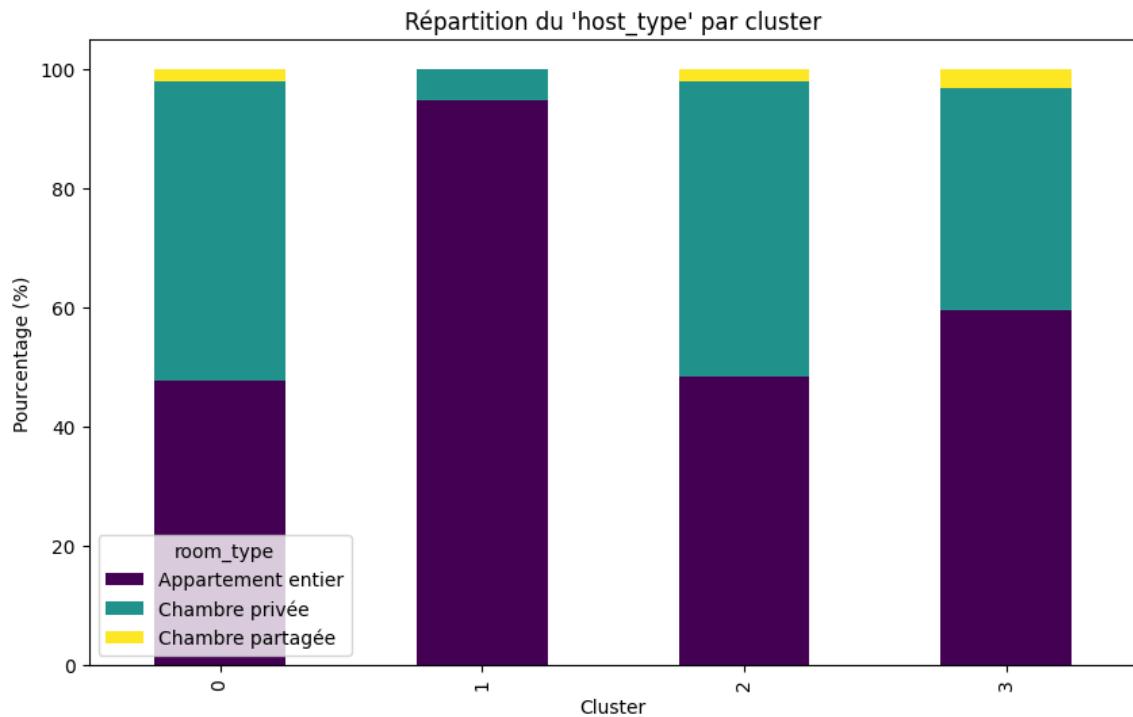


FIGURE 7 – Répartition des profils d'hôtes en fonction des clusters

Cette figure confirme que :

- le Cluster 1 est quasi exclusivement professionnel,
- le Cluster 2 est dominé par les hôtes occasionnels,
- les Clusters 0 et 3 constituent des zones de transition entre partage et business.

## 4.7 Spécialisation immobilière par cluster

La Figure ?? montre la relation entre clusters et types de logements.

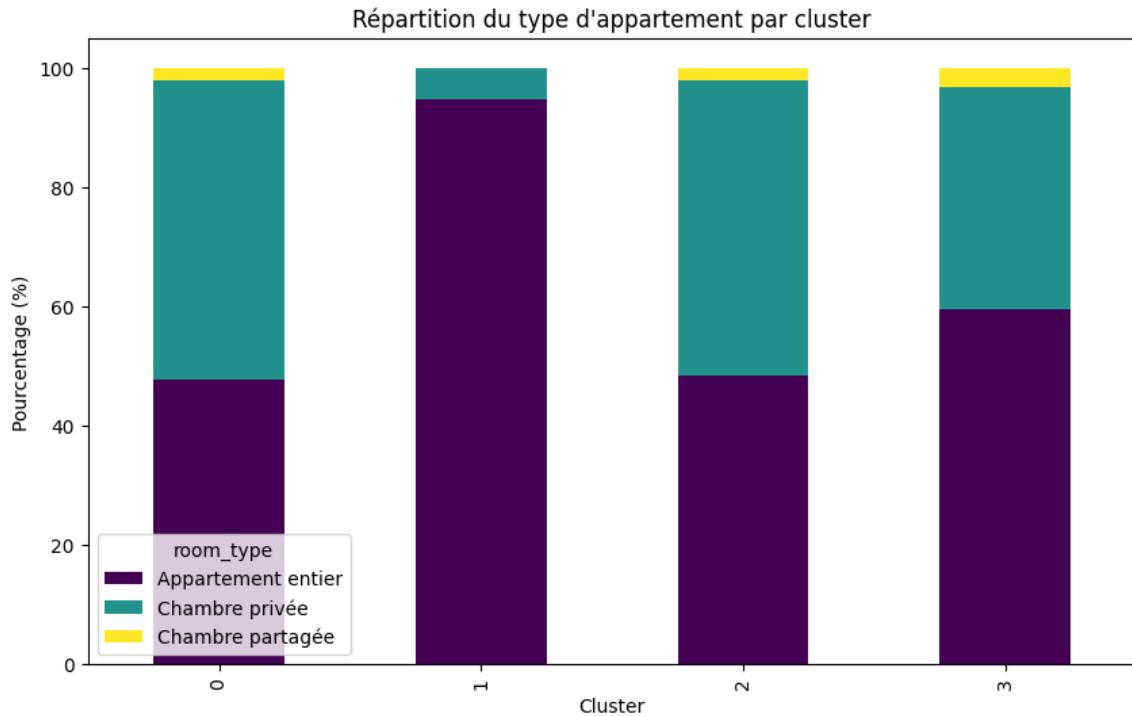


FIGURE 8 – Répartition des types de logements par cluster

Le Cluster 1 est composé à près de 80% de logements entiers. Il ne s'agit plus de partage d'un espace de vie, mais d'une privatisation du parc immobilier résidentiel. À l'inverse, le Cluster 2 propose majoritairement des chambres privées, maintenant un lien direct entre hôte et voyageur.

## 4.8 Segmentation et appropriation géographique

La Figure ?? présente la répartition des clusters selon les arrondissements.

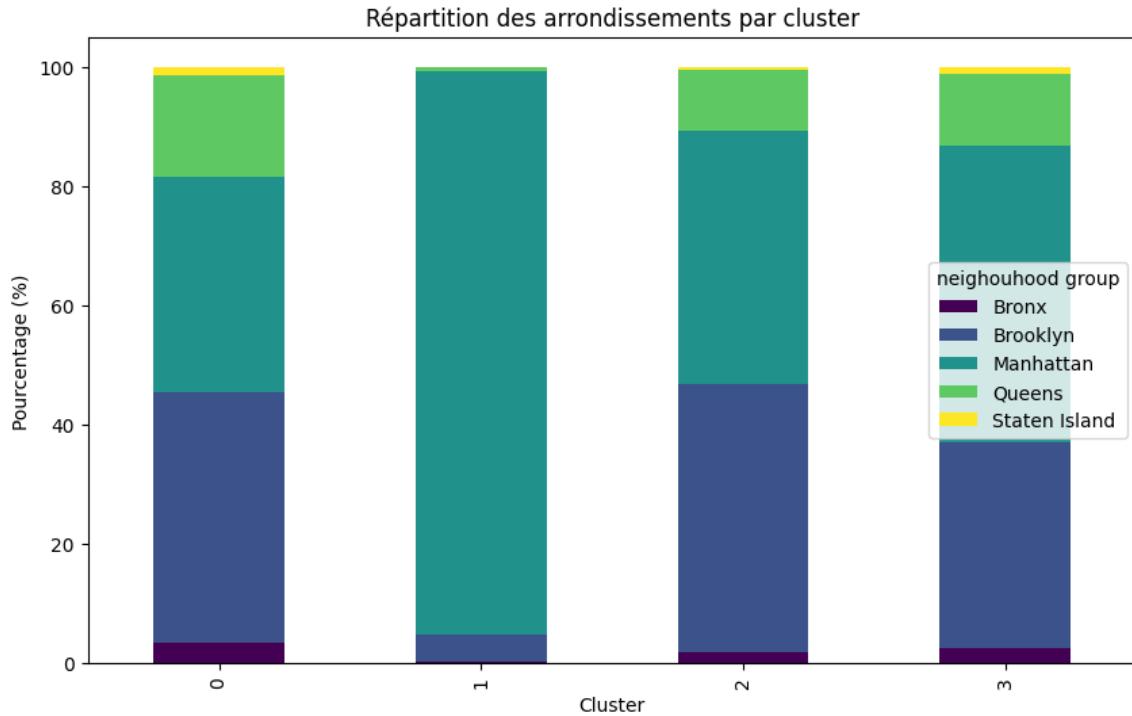


FIGURE 9 – Répartition géographique des clusters par arrondissement

On observe une ségrégation spatiale marquée :

- le Cluster 1 se concentre majoritairement à Manhattan, dans les zones à forte tension immobilière,
- les clusters orientés partage sont davantage présents dans Brooklyn, Queens et le Bronx,
- le marché Airbnb reproduit et accentue la géographie sociale et immobilière de New York.

## 4.9 Synthèse

La segmentation comportementale met en évidence une dualité structurelle du marché Airbnb new-yorkais. D'un côté, un pôle para-hôtelier professionnel, fortement concentré spatialement et fondé sur la privatisation du logement. De l'autre, un pôle de partage résidentiel, marginalisé géographiquement mais fidèle à la philosophie initiale de la plateforme. Les clusters intermédiaires jouent un rôle de transition et illustrent la transformation progressive du marché. “

## 5 Dynamique Spatiale et Dualité du Marché

La dimension géographique est la clé de voûte de la compréhension du marché Airbnb new-yorkais. L'analyse spatiale révèle que la plateforme ne lisse pas les inégalités territoriales, mais tend au contraire à les exacerber en créant des pôles d'hyper-activité.

## 5.1 Le "Super-Cluster" Touristique : Au-delà des Frontières Administratives

L'application de l'algorithme de densité DBSCAN a permis de dépasser la lecture administrative classique par arrondissements (*Boroughs*). La carte de densité ci-dessous met en lumière un phénomène majeur : la formation d'un continuum économique unifié.



FIGURE 10 – Concentration massive de l'activité et des prix élevés dans le centre névralgique, ignorant la frontière de l'East River.

Nous observons un "Super-Cluster" qui englobe le sud de Manhattan (Financial District, Soho) et traverse l'East River pour inclure Brooklyn Heights et Williamsburg. Dans cette zone, la densité d'annonces est maximale et les prix moyens dépassent systématiquement les 200\$. L'algorithme démontre que pour le touriste (et donc pour le marché), la frontière administrative entre Manhattan et Brooklyn n'existe plus : seule compte la centralité et l'accessibilité aux lieux d'intérêt.

## 5.2 La Ségrégation Spatiale : Business vs Partage

L'analyse la plus marquante réside dans la visualisation de la "Dualité du Marché". En cartographiant séparément les multi-propriétaires (Business) et les hôtes occasionnels (Partage), une ségrégation géographique nette apparaît.

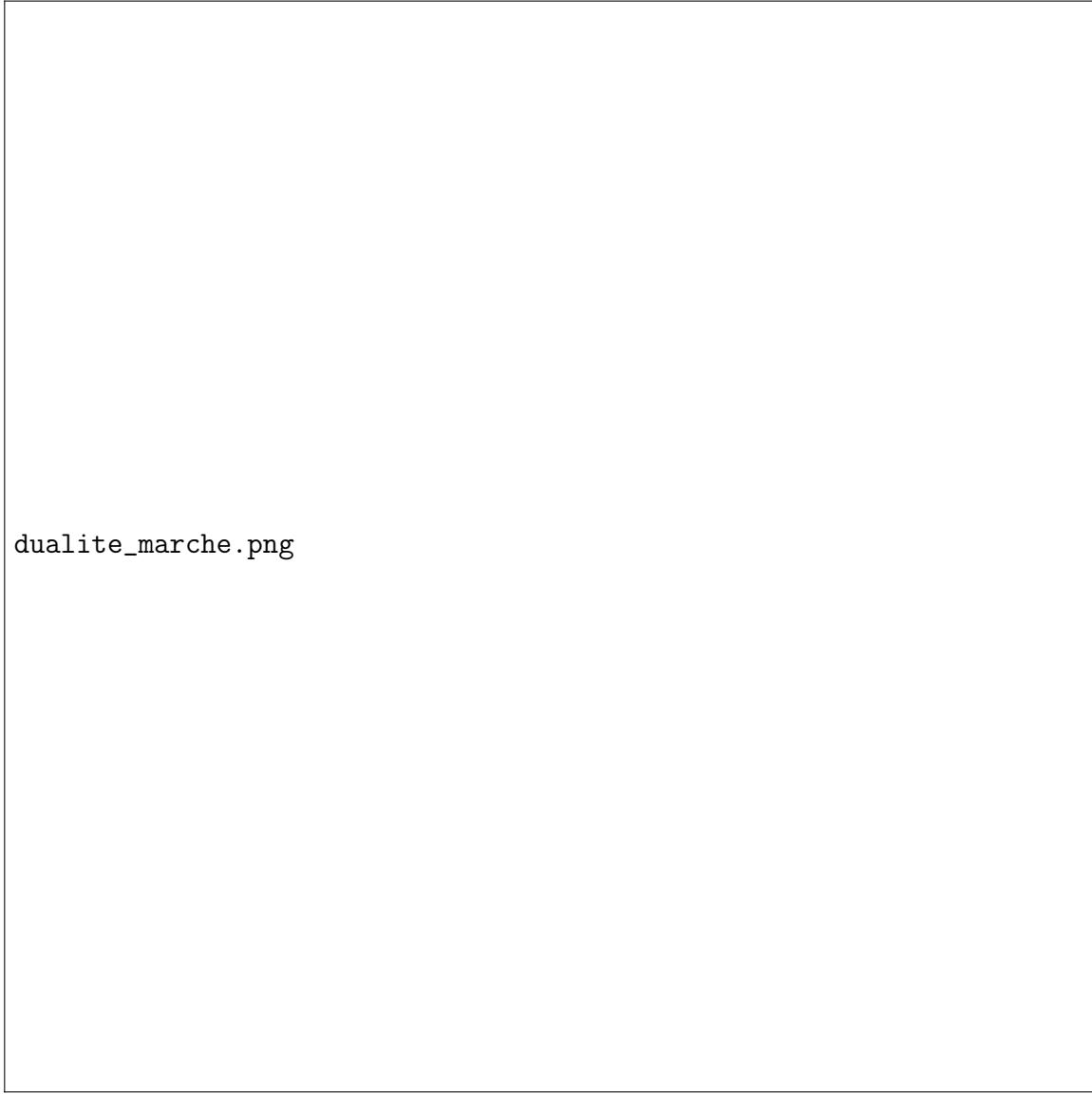


FIGURE 11 – Ségrégation spatiale : Les multi-propriétaires (Rouge) saturent le centre économique, tandis que l'économie de partage (Vert) domine la périphérie résidentielle.

- **Le Centre "Hôtelier" (Zone Rouge)** : Le cœur de Manhattan est dominé par les acteurs professionnels. Dans des quartiers comme Midtown ou le Theater District, la pression foncière est telle que seuls les opérateurs capables de générer des taux d'occupation élevés (type hôtelier) peuvent s'y maintenir. L'économie de partage y est marginalisée.
- **La Périphérie "Authentique" (Zone Verte)** : Dès que l'on s'éloigne de ce centre névralgique pour aller vers le Bronx, le Queens ou le sud de Brooklyn, la logique s'inverse. Les opérateurs professionnels disparaissent au profit des résidents

locaux. C'est dans ces zones "vertes" que survit l'esprit original d'Airbnb : une activité d'appoint, dispersée, moins chère et ancrée dans le tissu résidentiel local.

Cette analyse spatiale confirme que la professionnalisation d'Airbnb agit comme une force centrifuge, repoussant l'usage "amateur" vers les marges de la ville tout en transformant le centre en une zone d'exploitation industrielle.

## 6 Conclusion Générale

Au terme de cette analyse, la réponse à notre problématique est claire : le marché Airbnb new-yorkais est devenu **\*\*hybride et dual\*\***.

D'un côté, le centre fonctionne selon une logique industrielle (disponibilité permanente, gestion professionnelle, prix élevés). De l'autre, la périphérie maintient une logique de partage authentique. La plateforme a absorbé la logique hôtelière en son centre tout en créant un marché social complémentaire en périphérie. Cette étude démontre que loin d'uniformiser le territoire, Airbnb révèle et exploite les fractures économiques de la métropole.