

Les Plus Beaux Logis de Paris (Partie 1)



Les Plus Beaux Logis de Paris

Analyse de l'historique du marché de l'immobilier
Prédiction des valorisations du portefeuille client

Amina Feredj - Consultant BI

Paris, le 29 février 2024

Partie 1

Analyse de l'historique du
marché de l'immobilier

Prédiction des valorisations du
portefeuille client

Importer les librairies

Etude de l'historique immobilier parisien

Méthodologie suivie

Interprétation des résultats

Partie 2

Classification non supervisée
des biens issus de l'échantillon

Méthodologie suivie

Interprétation des résultats

Conclusion

Importer les librairies

Importer les librairies via la commande « import » suivi du nom de la librairie.

On peut ajouter « as » pour définir un alias

Pandas:

Manipuler et analyser des données

Numpy:

Faire des opérations mathématiques sur des tableaux

Plotly express:

Crée des visualisations de données interactives

Seaborn :

Résumer et visualiser des données

Train test split de la librairie sk learn:

Pour faire des prédictions

Datetime:

Manipuler les dates et les heures

Etude de l'historique immobilier parisien

Observations:

- Augmentation de la moyenne du prix du mètre carré entre 2017 et 2020.
- Baisse du prix moyen du mètre carré en 2021
- Le calcul de pourcentage de variation fait ressortir une baisse depuis 2020

Evolution du marché des appartements

moyenne_prix_metre_carre_appartements	
date_mutation	
2017	9492.859195
2018	10031.403432
2019	10562.712581
2020	10674.872650
2021	10455.600126

Observations:

- Augmentation de la moyenne du prix du mètre carré entre 2017 et 2019.
- Stabilisation du prix moyen du mètre carré depuis 2020
- Le calcul de pourcentage de variation fait ressortir une hausse depuis 2017

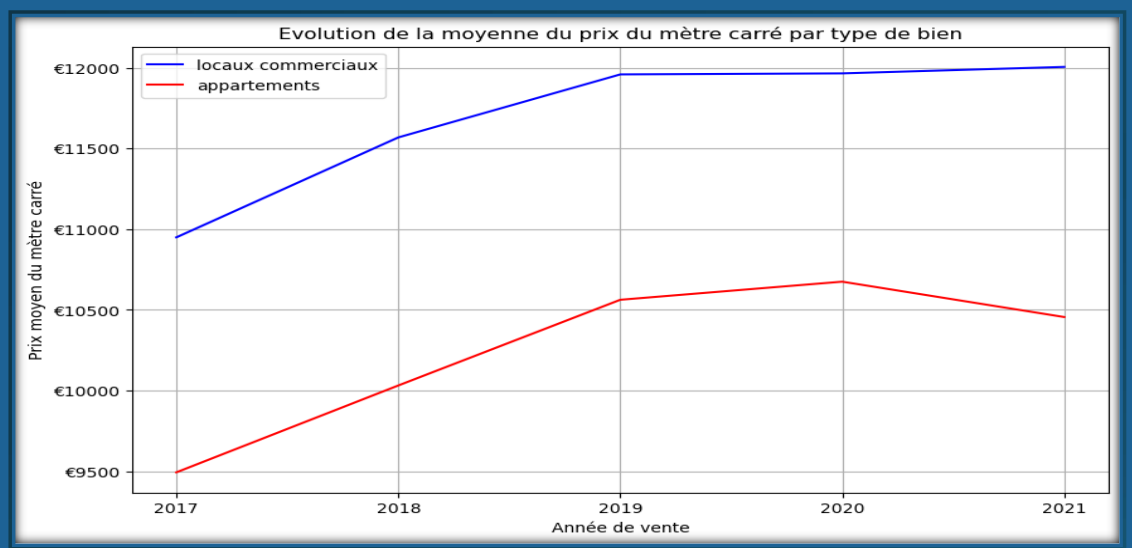
Evolution du marché des locaux

moyenne_prix_metre_carre_locaux	
date_mutation	
2017	10949.909217
2018	11569.499180
2019	11960.132050
2020	11966.470204
2021	12006.488691

Etude de l'historique immobilier parisien

Observations:

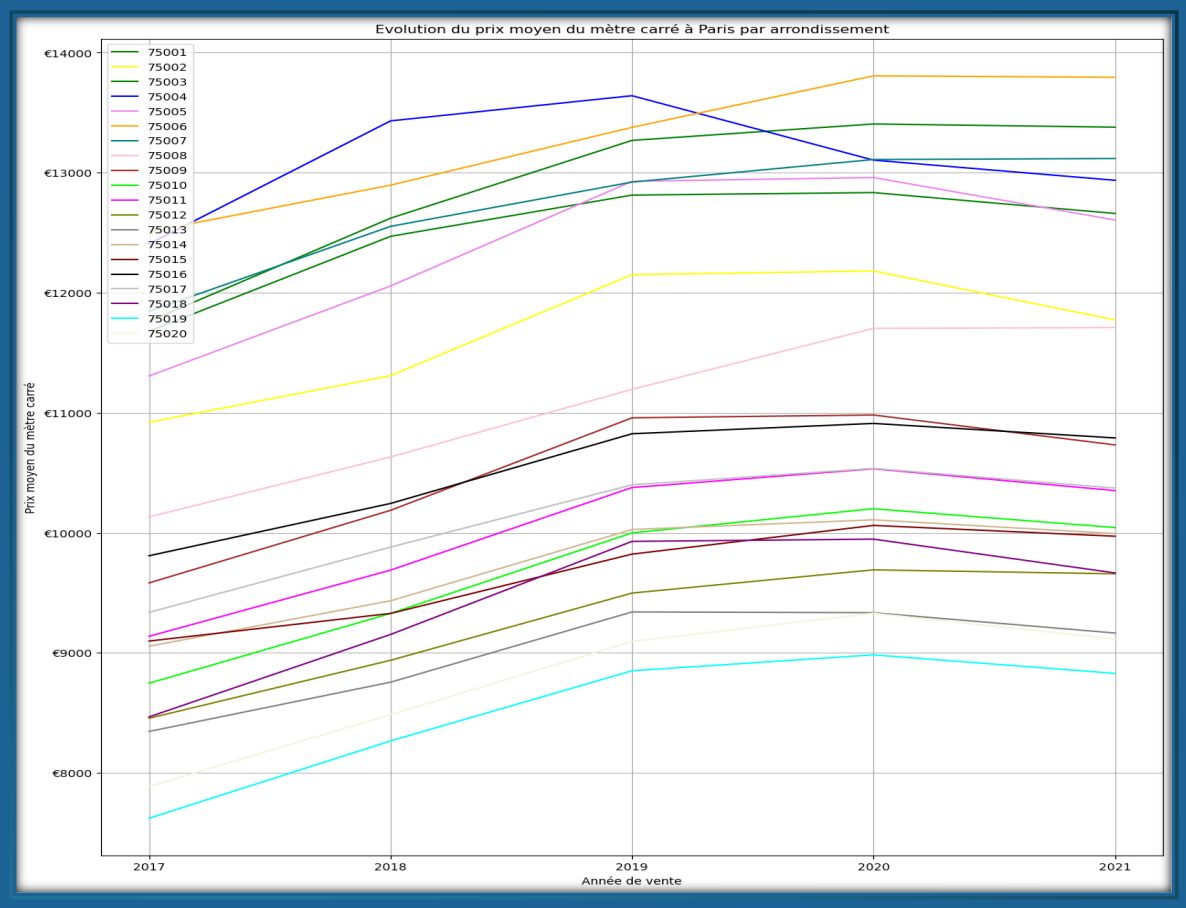
Le prix moyen du mètre carré est supérieur pour les locaux



	pourcentage_variation_locaux	pourcentage_variation_appartements
date_mutation		
2017	NaN	NaN
2018	5.658403	5.673151
2019	3.376403	5.296459
2020	0.052994	1.061849
2021	0.334422	-2.054100

Observations:

Le prix moyen du mètre carré est le plus élevé dans le 75006 et le plus bas dans le 75019.



Préparation du dataset

- Calculer et créer une colonne prix du mètre carré
- Identifier la nature des données et faire le nettoyage
- Identifier les types de biens immobiliers
- Calculer le nombre de transactions: 26196

Analyse des corrélations

- Calculer le prix moyen du mètre carré par année
- Comparer les prix moyens du mètre carré entre arrondissements pour 2021
- Analyser la corrélation entre le prix du mètre carré et la date (Pearson: statistics = 0.90 et pvalue= 7.108e-263)
- Analyser la corrélation entre le prix du mètre carré et la valeur foncière (Pearson: statistic=0.99, pvalue=0.0)

Mise en place d'un modèle de régression linéaire

- Transformer les données catégorielles en données binaires avec one hot encoding pour limiter les biais
- Utiliser train_test_split pour prélever 33% des données et entraîner notre algorithme avec le reste
- Définir les tableaux de données (X) surface et (Y) valeur foncière
- Créer et entraîner un modèle de régression linéaire

Vérifier les résultats et la robustesse du modèle

- Calculer l'intercept et le coefficient
- Comparer les résultats de l'algorithme avec les 33% mis de côté
- Vérifier la robustesse du modèle ($R^2 = 0,97$)
- Calculer le pourcentage d'erreur (8,93%)

Appliquer le modèle au portefeuille client

- Importer le fichier et identifier la nature des données
- Importer la librairie datetime
- Transformer les variables catégorielles en variables binaires avec one hot encoding
- Transformer les variables en (str) pour pouvoir appliquer l'algorithme

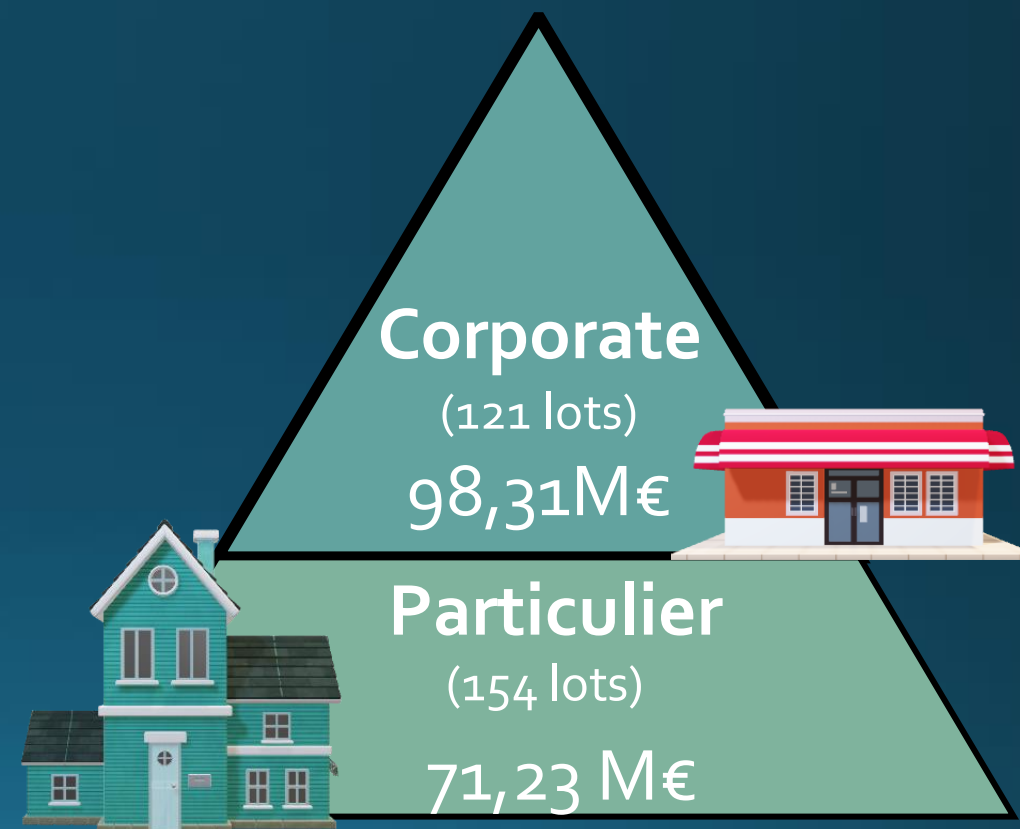
Analyser et présenter les résultats

- Calculer le montant des valorisations du portefeuille particulier
- Calculer le montant des valorisations du portefeuille corporate
- Calculer le nombre de lots
- Calculer le pourcentage de valorisation du parc immobilier par année
- Présenter les résultats

Interprétation des résultats

Comparaison entre les deux segments

- La valorisation du segment corporate est supérieur à celle du segment particulier malgré le nombre de lots.
- La valorisation des dernières années penche en faveur du segment Corporate.
- Le modèle utilisé a enregistré 8,9% de marge d'erreur
- Le résultat ne prend pas considération d'autres paramètres(ex: quartier, situation géopolitique)



Les Plus Beaux Logis de Paris (Partie 2)



Les Plus Beaux Logis de Paris

Classification non supervisée des biens issus de
l'échantillon

Amina Feredj - Consultant BI

Paris, le 29 février 2024

Méthodologie suivie

Préparation du dataset

- Identifier la nature des données
- Supprimer les colonnes inutiles (code_postal, nom_commune)
- Calculer et créer une colonne prix du mètre carré
- Afficher un graphique en nuage de points pour voir si des groupes homogènes se forment
- Convertir les données catégorielles en données quantitatives

Classification non supervisée

- Utiliser l'algorithme Kmeans pour créer 2 clusters (Appartement et local)
- Tester différentes valeurs pour le nombre d'itérations pour explorer les possibilités de recherches et éviter de tomber dans un optimum local et trouver l'équilibre.
- Tester l'algorithme pour classer les biens en attribuant (1 et 0)

Vérification des résultats

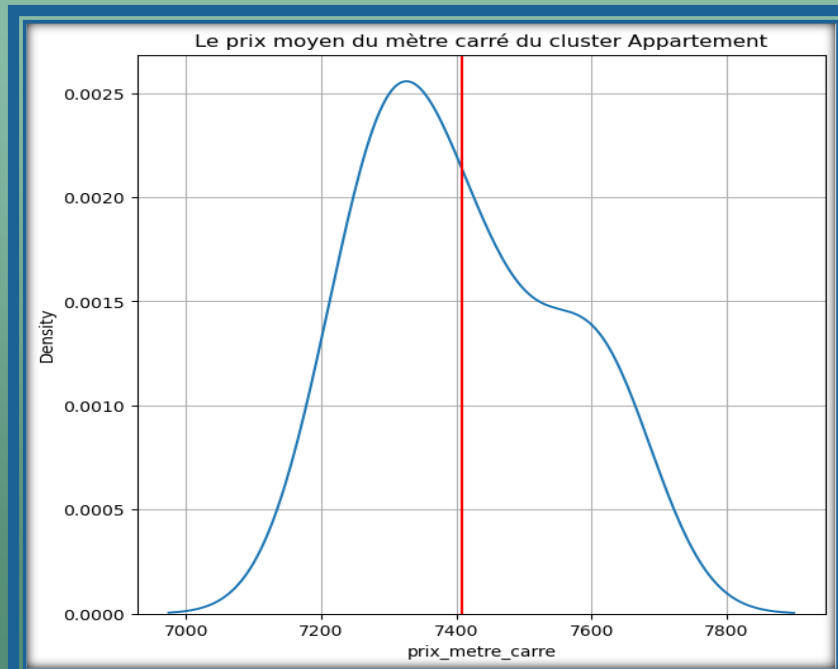
- Hypothèse: pour deux superficies identiques dans le 19^{ème} arrondissement celle qui a le prix du mètre carré le plus élevé correspond à un local
- Remplacer la valeur 1 par « local » et 0 par « appartement »
- Compter les proportions (20 appartements et 20 locaux)
- Vérifier avec les services commerciaux la pertinence du clustering

Interprétation des résultats

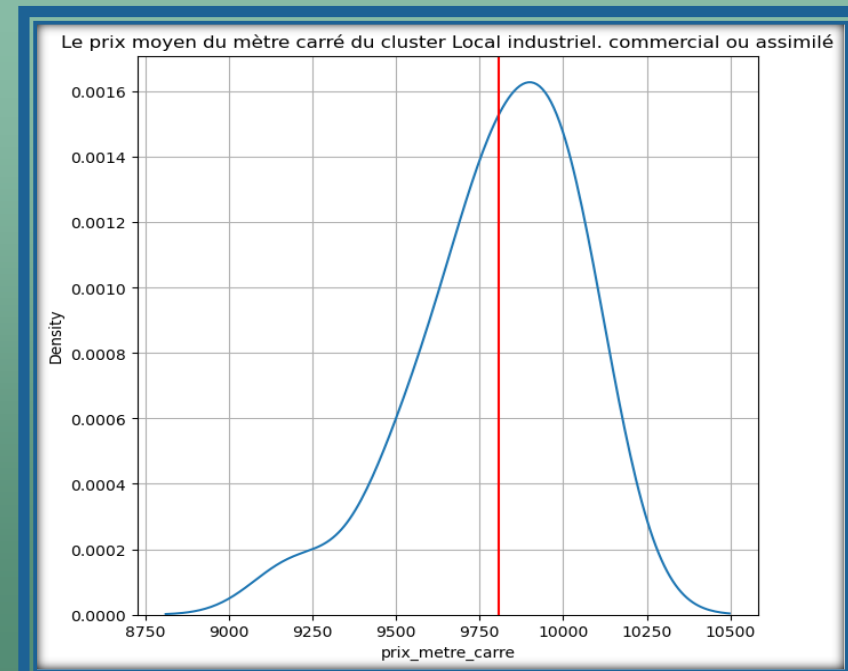
Observations:

- Le centroïde du cluster « Appartement » est de 7408,77€/m² et celui du cluster « Local industriel. commercial ou assimilé » est de 9806,92€/m².
- La surface moyenne du cluster « Appartement » est de 38,60m² et celle du cluster « Local industriel. commercial ou assimilé » est de 49,35m².

Centroïde du cluster appartement



Centroïde du cluster Local industriel



Conclusions

- Le clustering a identifié deux groupes homogènes et suffisamment distincts comprenant chacun 20 biens.
- Deux variables ont été prises en considération, le prix du mètre carré et la surface pour poser l'hypothèse
- Vérifier les résultats avec les équipes commerciales peut s'avérer utile
- La méthode Kmean peut s'avérer moins adapter pour des volumes plus importants
- Il serait plus opportun de prendre des paramètres supplémentaires pour identifier les clusters comme l'état du bien et le quartier pour plus d'efficacité.