



Les Plus Beaux Logis de Paris

Les Plus Beaux Logis de Paris (Partie 1)

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

Milaine GUIAGAING – Consultante BI

Paris, le 11 mars 2024

Plan

Partie 1 : Analyse de l'historique du marché de l'immobilier - Prédiction des valorisations du portefeuille client

Etude de l'historique immobilier de Paris

Méthodologie suivie

Interprétation des résultats

Partie 2 : Clustering des biens issus de l'échantillon

Méthodologie suivie

Interprétation des résultats

Conclusion

Etude de l'historique immobilier de Paris

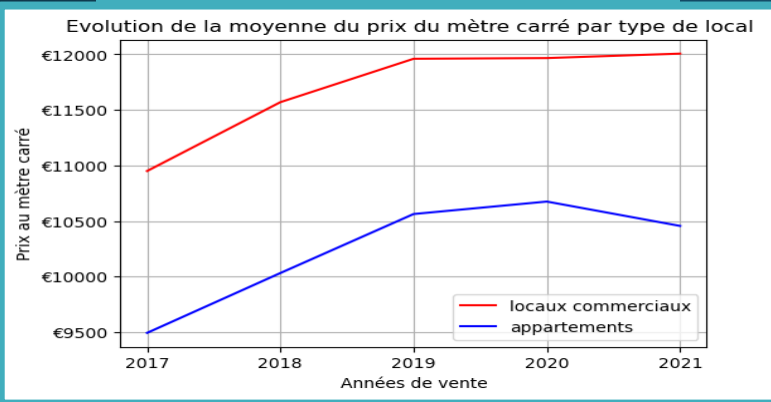
Evolution du marché par type local

	moyenne_prix_metre_carre_appart	moyenne_prix_metre_carre_locaux
date_mutation		
2017	9492.859239	10949.909217
2018	10031.403469	11569.499180
2019	10562.712563	11960.132050
2020	10674.872719	11966.470204
2021	10455.600179	12006.488691

Taux de variation de la moyenne de prix du m²

	Taux de variation_appartements	Taux de variation_locaux
date_mutation		
2017	NaN	NaN
2018	5.673151	5.658403
2019	5.296458	3.376403
2020	1.061850	0.052994
2021	-2.054100	0.334422

Evolution du marché par type local



Types de biens immobiliers:

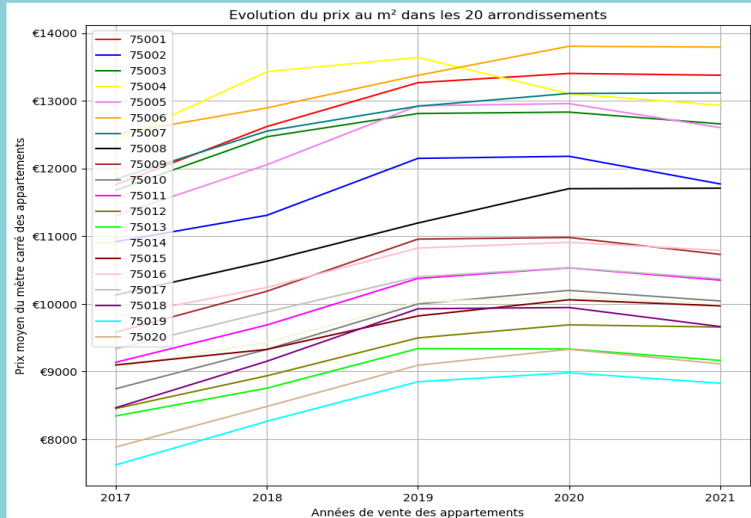
- Appartements
- Local industriel, commercial ou assimilé

Observations:

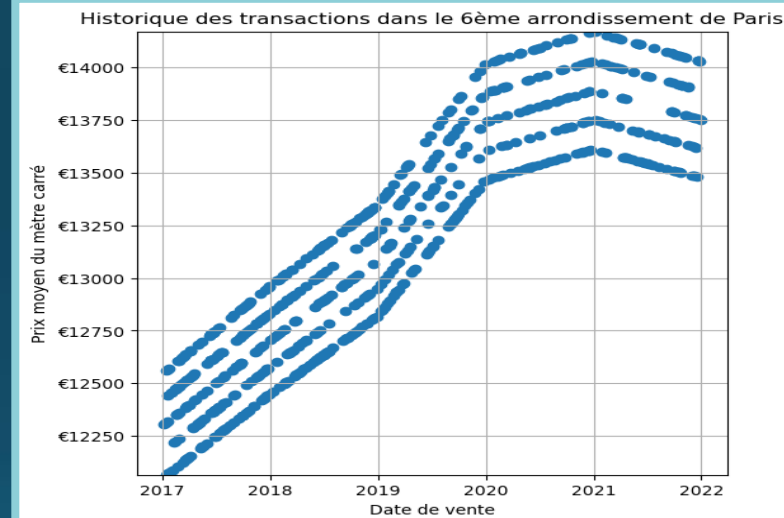
- Le prix du mètre carré est supérieur pour les locaux
- Augmentation de la moyenne du prix du mètre carré entre 2017 et 2020.
- Baisse du prix moyen du mètre carré en 2021 pour les appartements

Etude de l'historique immobilier de Paris

Evolution du prix du m² dans les 20 arrondissements



Historique des transactions dans le 6^e arrondissement



Observations:

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

Méthodologie suivie

Préparation du dataset

- Explorer les données afin de détecter les types et la complétude
- Identifier les types de biens immobiliers
- Déterminer le nombre de transactions global et par type de biens
- Calculer le prix du mètre carré

Etude de la relation entre deux variables

- Vérifier la corrélation entre le prix au m² et la date (Pearson: statistics = 0.90 et pvalue= 7.108e-263)
- Vérifier la corrélation entre la valeur foncière et la surface (Pearson: statistic=0.99, pvalue=0.0)

Mise en place d'un modèle de prédiction

- Transformer les variables catégorielles en variables binaires en utilisant le « one-hot-encoding »
- Utiliser train_test_split pour prélever 33% des données et entraîner notre algorithme avec le reste
- Créer et entraîner un modèle de régression linéaire

Méthodologie suivie

Calculer les métriques d'évaluation du modèle

- Le pourcentage d'erreur (MAPE=8.93%)
- La robustesse du modèle ($R^2=0.97$)
- La précision du modèle (RMSE=71404.03)

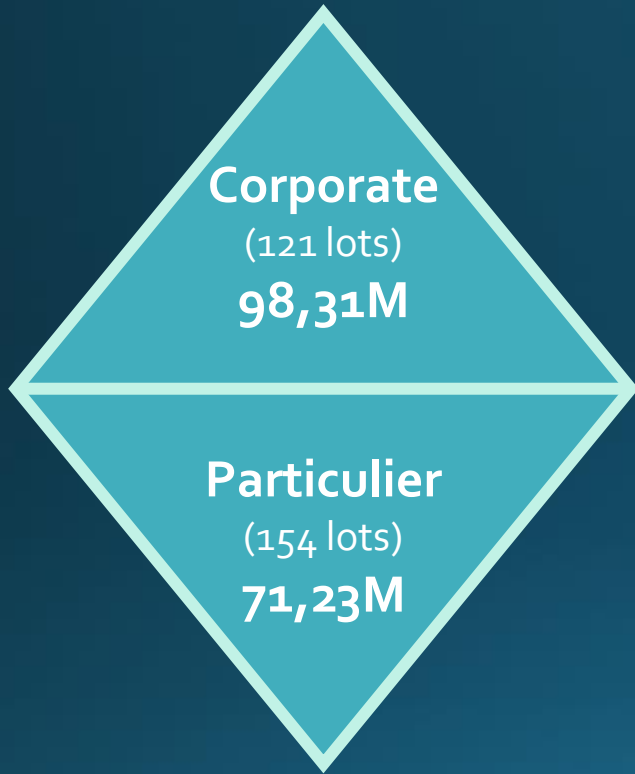
Appliquer le modèle au portefeuille client

- Faire une analyse exploratoire et transformer les variables catégorielles utilisées dans le modèle en variables binaires
- Appliquer l'algorithme de prédiction

Analyser les résultats

- Calculer le montant des valorisations du portefeuille particulier
- Calculer le montant des valorisations du portefeuille corporate
- Calculer le pourcentage de valorisation du parc immobilier par année

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: situation géopolitique, épidémie)



Les Plus Beaux Logis de Paris

Les Plus Beaux Logis de Paris (Partie 2)

Clustering des biens issus de l'échantillon

Milaine GUIAGANG – Consultante BI

Paris, le 11 mars 2024

Méthodologie suivie

Préparation du dataset

- Explorer les données afin de détecter les types et la complétude
- Supprimer les colonnes inutiles
- Détecter des groupes homogènes à l'aide d'un nuage de points
- Calculer le prix du mètre carré

Clustering

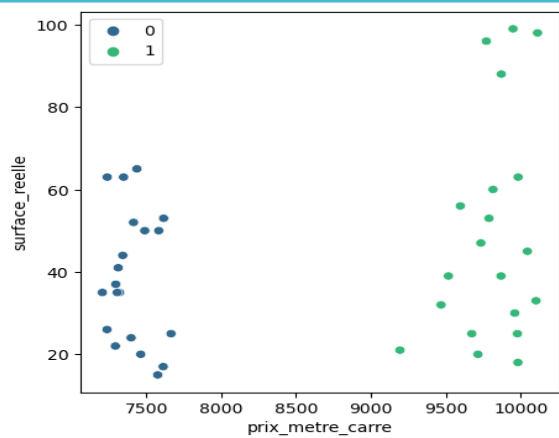
- Utiliser l'algorithme Kmeans pour créer 2 clusters (Appartement et Local commercial)
- Tester différentes valeurs pour le nombre d'itérations pour explorer les possibilités de recherches et trouver l'équilibre.
- Tester l'algorithme pour classer les biens en attribuant (1 et 0)

Vérification des résultats

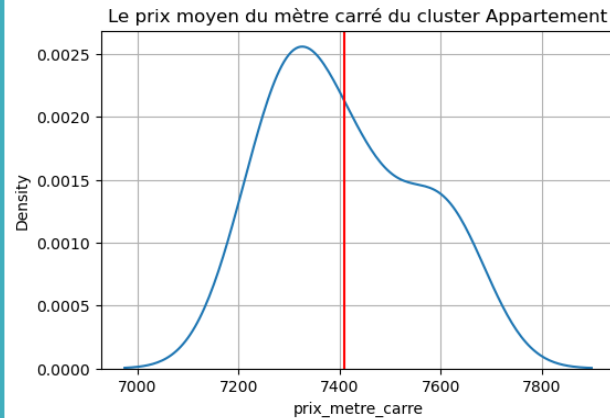
- Hypothèse: pour deux superficies identiques dans le 19ème arrondissement celle qui a le prix du mètre carré le plus élevé correspond à un local
- Remplacer les 1 et 0 par les biens correspondants
- Compter les proportions (20 appartements et 20 locaux)

Interprétation des résultats

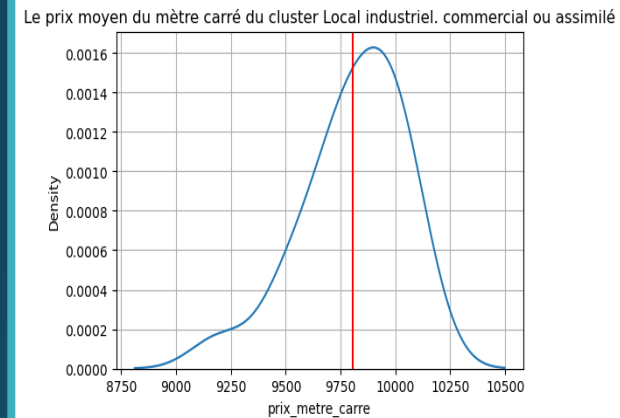
Représentation des Clusters



Centroïde du cluster appartement



Centroïde du cluster local



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.

Conclusion

- 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 Kmeans 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.