
Les Plus Beaux Logis de Paris

Partie 1



Les Plus Beaux Logis de Paris

Titre du projet : Analyser l'évolution des prix de l'immobilier avec Python

Prénom : Chehrazad

Nom : El Botout

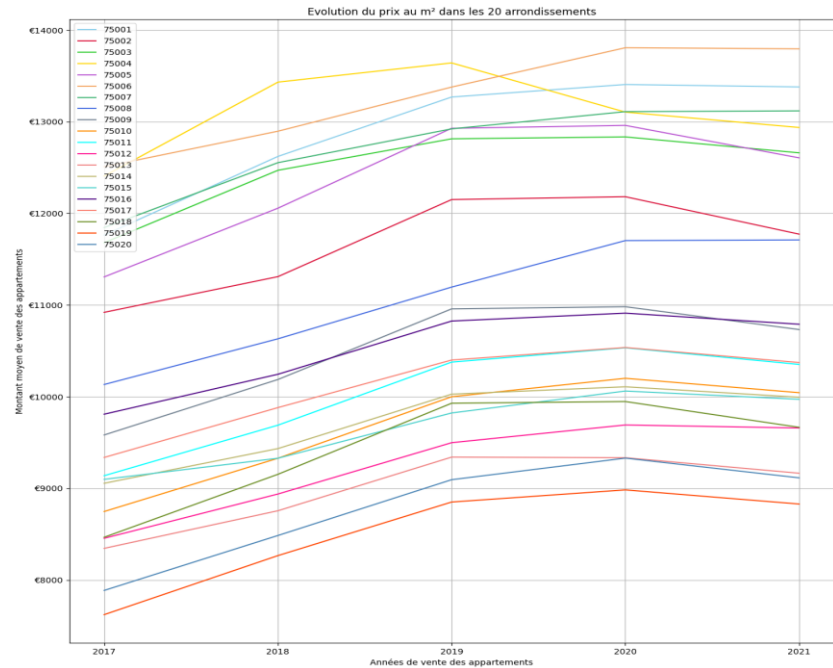
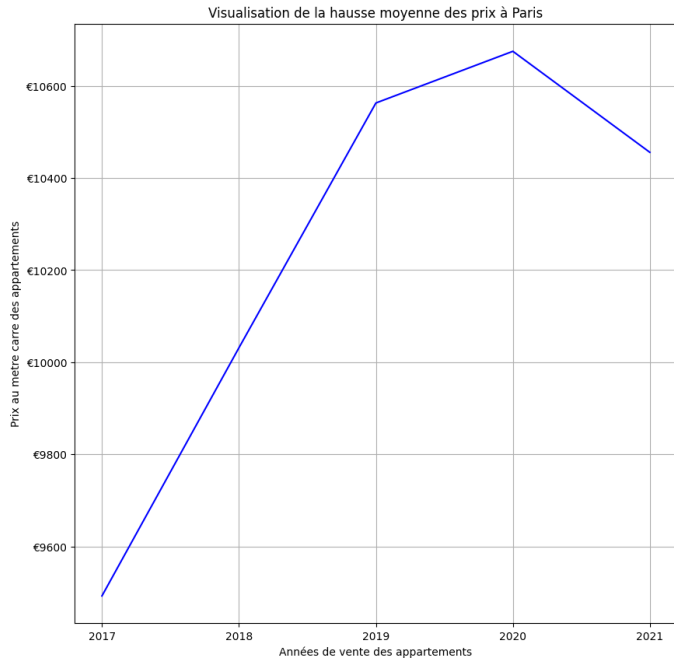
Date : 01/01/2024

I. Analyse du marché de l'immobilier

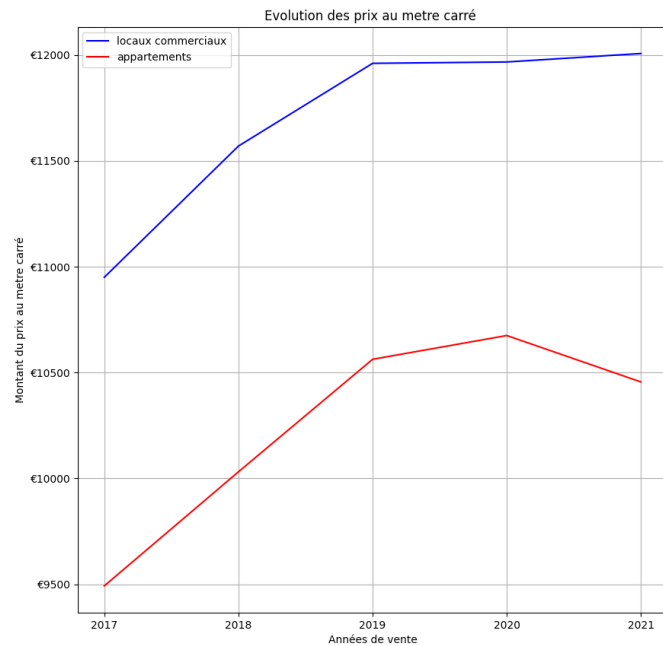
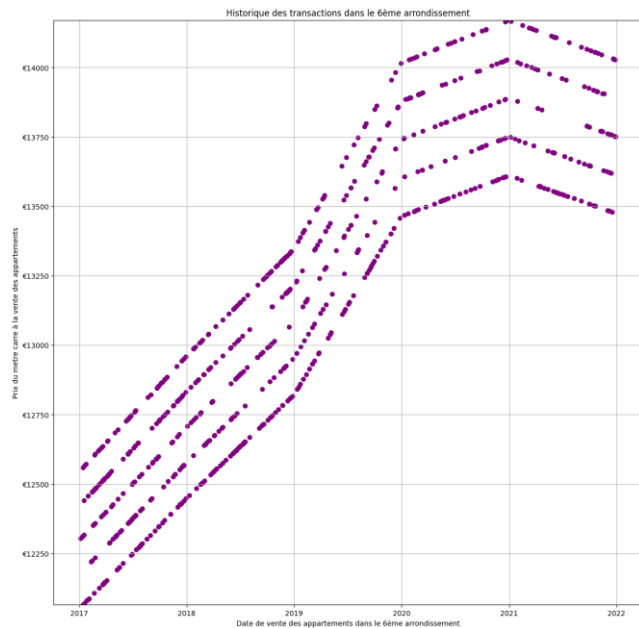


- Notre jeu de données contient les biens immobiliers à Paris entre **2017 et 2021**
- Il y a deux catégories de biens : les locaux commerciaux pour lesquels il y a 1843 transactions et les appartements pour lesquels il y a 24353 transactions
- Vérification du type de données et étude des relations entre les variables en créant la colonne « prix_m2 »
- Création d'un Dataframe « df_ventes_appartements » et d'une visualisation : **On constate une hausse exponentielle des prix au m2 entre 2017 et 2020, puis une légère baisse entre 2020 et 2021**
- Création d'un Dataframe pour avoir l'évolution des prix par arrondissement nommé : « df_appartements_par_arrondissements », puis d'une visualisation de l'évolution du prix au m2 dans les 20 arrondissements : **On constate à peu près la même tendance sur tous les arrondissements.**
- **Nous constatons que l'évolution des prix dans le 1^{er} arrondissement suit la tendance globale, quant au 6^{ème} arrondissement, on observe que les prix sont globalement plus élevés que les autres et qu'ils se stabilisent à partir de 2020.**

Analyse du marché de l'immobilier



Analyse du marché de l'immobilier





II. Méthodologie suivie

- *Nous utilisons la corrélation de Spearman pour prouver la relation entre le prix au m2 et la date puis entre la valeur foncière et la surface. **Les corrélations sont confirmées.***
- *Création d'un Dataframe qui ne contient que les locaux commerciaux puis création de la colonne « prix_m2 ». **Nous constatons que les prix des locaux commerciaux sont plus élevés que les appartements.***
- *Transformation des variables catégorielles en nombre à l'aide du One hot encoder . Un tableau est donc créé avec une colonne pour chaque code postal.*
- *Importation du Train split avec le Sklearn model : x = Valeur réelle, y = valeur prédite*
- *Conversion des données en chaîne de caractères (string)*
- *Utilisation de la méthode de la régression linéaire pour calculer le pourcentage d'erreur.*

III. Résultat des prédictions



- *Résultat de ma prédiction sur la valorisation au 31 décembre 2022 du portefeuille des actifs immobiliers de l'entreprise :*
- *Importation du dataframe portefeuille actifs puis utilisation du One hot encoder pour transformer les variables.*
- *Ajout de la colonne 'prediction'.*
- ***La valorisation du segment particulier est de 71,24 millions d'euros.***
- ***La valorisation du segment corporate est de 98,32 millions d'euros.***
- *Après test du modèle, nous obtenons une marge **d'erreur de 6,84%** sur la valeur foncière.*
- *Les estimations sont basées sur les données disponibles au moment de l'entraînement du modèle, les conditions du marché peuvent évoluer, ce qui peut affecter la précision des estimations.*
- *Les prédictions du modèle peuvent ne pas capturer toutes les nuances du marché immobilier.*

Les Plus Beaux Logis de Paris

Partie 2



Les Plus Beaux Logis de Paris

Analysez l'évolution des prix de l'immobilier avec Python

Chehrazad

El Botout

01/01/2024

I. Méthodologie suivie



- *Importation du Dataframe « échantillons à classer »*
- *Même nettoyage de données que pour les autres fichiers → suppression des colonnes « valeur foncière » et « surface » pour créer la colonne « prix_m2 »*
- *Utilisation du Kmeans pour analyser les clustering et classer nos biens dans appartements ou locaux en fonction de leur prix au m2*
- *La courbe de coude avec les centroïdes dans l'algorithme K-means permet de choisir visuellement le nombre optimal de clusters en identifiant le point où la courbe commence à s'aplatir, et les centroïdes représentent les centres de chaque cluster.*
- *Création de la colonne « label_prediction » pour afficher un tableau où 1 représente les appartements et 0 les locaux commerciaux.*

II. Résultat de la classification

```
#On affiche les résultats
df_opportunité_2['prediction_label'] = df_opportunité_2['prediction_label'].replace({0: 'locaux', 1: 'appartements'})
df_opportunité_2_sorted = df_opportunité_2.sort_values(by='prix_m2', ascending=False)
print(df_opportunité_2_sorted)
```

	code_postal	prix_m2	prediction_label
32	75019	10113.195822	locaux
27	75019	10103.161577	locaux
1	75019	10045.572493	locaux
38	75019	9983.591800	locaux
5	75019	9982.016919	locaux
18	75019	9978.784822	locaux
26	75019	9961.146860	locaux
17	75019	9949.544452	locaux
0	75019	9871.444128	locaux
13	75019	9869.462342	locaux
22	75019	9815.972258	locaux
8	75019	9789.051438	locaux
39	75019	9771.486096	locaux
30	75019	9734.720495	locaux
36	75019	9715.531658	locaux
6	75019	9674.262777	locaux
28	75019	9597.631488	locaux
11	75019	9518.076103	locaux
3	75019	9469.142168	locaux
2	75019	9194.697790	locaux
31	75019	7666.071700	appartements
9	75019	7616.420959	appartements
29	75019	7613.514621	appartements
37	75019	7583.980837	appartements
33	75019	7577.044521	appartements
10	75019	7490.469041	appartements
4	75019	7463.610005	appartements

Ce code effectue la classification des données dans la catégorie correspondante, suite à la prédiction effectuée. Nous constatons que le prix au m2 moyen des locaux commerciaux est plus élevé que le prix des appartements. Cette conclusion suit la logique et la tendance de toute notre analyse.