

IMMOBILIER



Analyse data-centric

LOU

Une société immobilière souhaiterait récupérer et stocker des données Open Data des ventes immobilières afin de mener une étude lui permettant d'avoir une meilleure connaissance du marché en utilisant l'IA dans un futur proche. En amont de cela, elle souhaiterait donc pouvoir créer un flux permettant le stockage de données mais aussi des analyses de l'historique.

Nettoyage de données

J'ai pris le parti de nettoyer les données pour n'avoir qu'une ligne par Transaction. En effet, en analysant, je me suis rendu compte qu'il y avait des doublons et que les dépendances, caves, garages, etc. qui possédaient la même Valeur foncière. Dans un premier temps, l'objectif était de les linker ensemble. Finalement, j'ai préféré garder que les transactions Appartement, Maison et Local Commercial qui représentent les lots principaux des ventes.

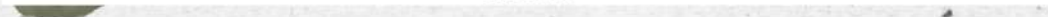
Dans un premier temps, pour obtenir ce résultat, j'ai contrôlé les colonnes et leur type ainsi que la proportion de NaN au sein des colonnes. Avec l'équipe, nous avons décidé de ne conserver que les colonnes avec moins de 60 % de NaN. Les colonnes Surface carrez n'étaient pas concernées et nous les avons gardées pour les analyses futures.

Par la suite, j'ai concaténé les colonnes N° voie, type de voie et Voie afin d'obtenir une colonne Adresse. J'ai fait la même chose avec les colonnes Code département et Code commune afin d'obtenir un Code postal propre. Au préalable, il aura fallu préparer ces dites colonnes. Enfin, une fois terminé, j'ai décidé de supprimer les doublons.

J'ai finalement trouvé un dataset des régions et départements que j'ai merge avec notre dataframe afin de pousser les analyses un peu plus loin. Au final, j'ai traité le dataframe obtenu pour n'avoir qu'une seule transaction passant de plus de 4 millions de lignes à 1 471 825 lignes. J'en ai profité pour remplacer tous nos NaN par des 0.

Enfin, j'ai renommé et réorganisé toutes les colonnes afin d'obtenir un dataframe propre et prêt à l'emploi. Je l'ai exporté en CSV pour que mon équipe puisse travailler dessus. À partir de là, nous avons pu préparer les dataframe correspondant à nos tables qui ont été établies dans notre MCD. Cela nous a permis d'importer sans souci nos tables dans notre base de données.







Ces trois graphiques traitent des moyennes en fonction du type de logement. Nous avons donc :

- Moyenne du prix
- Moyenne des m²
- Moyenne du Prix au m²

On constate alors que les locaux commerciaux possèdent à chaque fois la plus grande part du donut. Cela se justifie par le fait qu'ils soient plus grands et donc plus chers que les autres types.

Conclusion

Avec toutes les analyses effectuées, les maisons se démarquent clairement du type de logement. C'est celui qui se vend le mieux et ce dans la majorité des régions. On constate que les départements où les régions influent énormément sur les achats. Les maisons font partie des valeurs foncières les plus élevées.

Pour autant, ces derniers ne sont pas soumis à la saisonnalité puisque les ventes restent constantes sur les 4 trimestres 2022. On peut à la rigueur soulignée les faits que les ventes augmentent légèrement au printemps et à l'été, le soleil donne envie aux acheteurs d'investir. Cependant, il y a une grosse chute des ventes au mois d'Août et au mois de Novembre.

Si l'on s'attarde sur les appartements, on rapporte que ce sont les 2 pièces puis les 3 pièces qui vont se vendre globalement mieux.

Concernant les locaux commerciaux, eux, se démarquent vraiment lorsqu'on effectue les moyennes par prix, surfaces et prix au m², sinon on ne les voit pas tant que ça.