Analyse Hédonique des Prix Immobiliers

Cette analyse a été réalisée à partir d'un échantillon de transactions immobilières provenant de la base de données Kaggle. Un premier modèle a été estimé en utilisant 1000 observations et les variables disponibles dans la base de données. Toutefois, certaines variables ont dû être omises, car l'outil Analyse de données / Régression d'Excel n'accepte qu'un maximum de 16 variables explicatives, alors qu'après le recodage des variables explicatives, le modèle en comptait plus de 20. Pour avoir le meilleur modèle, seules les variables jugées les plus pertinentes ont été conservées.

Malgré cette sélection, le R^2 obtenu était très faible (< 1 %), avec un seuil de significatif supérieur à 10 %. On ne peut donc pas conclure que ce modèle soit statistiquement significatif. Pour approfondir l'analyse, 3000 observations ont ensuite été utilisées dans un gabarit de régression linéaire sur Excel permettant l'intégration de 20 variables explicatives. Pour améliorer, le R^2 transformation logarithmique de variables comme Price, Lot_Size et Area_(Sqft) est recommandée dans l'analyse immobilière pour normaliser des distributions des extrêmes dans les prix. Ainsi, cette technique améliore la validité des modèles et la précision des prédictions (voir van Veen, 2023; Changyong et al., 2014; Dev.to, 2019)

Même avec ces ajustements, le modèle n'explique qu'environ 1,1 % de la variation du prix. Il est également important de souligner que la variable Zipcode a été omise : son recodage aurait généré un nombre trop élevé de modalités. Pourtant, cette variable aurait pu être déterminante, puisqu'elle reflète directement la localisation précise d'une propriété, un facteur essentiel dans la variation des prix immobiliers. Ainsi pour la localisation seulement les variables state et city ont été prise en compte.

Donc, pour confirmer mes résultats j'ai transféré, le fichier à ChatGPT, qui a modélisé la régression linéaire de façon plus complète, il est aussi arrivé à un R^2 similaire sachant qu'il a pris en compte la variable zip code dans le cas d'une régression linéaire.

Les résultats très faibles de R^2 nous indiquent donc que même avec plusieurs variables mises en place, il est difficile d'expliquer la fluctuation des prix avec un simple modèle prédictif de

régression linéaire. Ces résultats pourraient être expliqués par des variables manquantes indispensables comme certains facteurs qualitatifs (qualité de construction, état des rénovations, vue, bruit, réputation du quartier). Aussi même si deux biens paraissent similaires dans la base, leur valeur réelle peut diverger à cause de la spécificité locale : micro-emplacement dans la rue, proximité immédiate d'une école ou d'un parc, niveau de bruit, qualité des voisins. C'est caractéristique fines ne sont pas prises en comptes malgré leur importance. Aussi, il y a un manque d'information par exemple deux maisons construites la même année peuvent être codées comme identiques dans la base, mais si l'une a été entièrement rénovée et l'autre non, leur valeur diffère fortement, ce que le modèle ne capte pas et qui crée un écart important entre le prix réel et le prix estimé.

Ainsi, pour répondre à la question : « Peut-on vraiment évaluer à distance avec une étude de bureau ? ». Elle permet d'obtenir une estimation statistique utile grâce aux comparables et à la régression, mais elle reste insuffisante pour une évaluation précise. En effet, certains éléments qualitatifs fins comme les rénovations, l'état d'entretien ou la spécificité locale ne sont pas captés par le modèle et génèrent des écarts entre prix réels et prix estimés. Ainsi, une analyse de son bureau ou théorique donne une tendance générale, mais l'expertise terrain demeure indispensable pour refléter fidèlement la valeur marchande d'un bien.

Les coefficients révèlent certains résultats attendus : le type Condo a un effet négatif sur le prix par rapport aux maisons unifamiliales (XpertSource, 2025; Centris, 2025). Le statut Sold est associé à un prix plus élevé, ce qui correspond à la tendance du marché, où les propriétés vendues sont généralement transigées à des prix supérieurs (Centris, 2025). Toutefois, certains signes surprennent : Lot Size apparaît négatif, ce qui suggère que les grands terrains sont situés en périphérie avec un prix au pied carré inférieur (XpertSource, 2025). De même, le nombre de chambres ressort négatif, possiblement en raison du fait que certaines grandes maisons sont anciennes et se trouvent dans des zones moins chères (SCHL, 2025). Ces incohérences illustrent les limites du modèle hédonique qui ne tient pas compte de la qualité du bien, des rénovations ni de la localisation précise qui sont déterminant en immobilier (XpertSource, 2025; Centris, 2025; SCHL, 2025). (Annexe 1)

Annexe 1

RAPPORT DÉTAILLÉ

Ajustement	
racine de R-carré	0.108076842
R-carré (R^2)	0.011680604
R-carré ajusté	0.005045361
AIC	-2596.675995
BIC	-2470.542276

seuil expérimental < 0.01 cellule vert foncé
0.01 ≤ seuil expérimental ≤ 0.05 cellule saumon

ANALYSE DE VARIANCE (ANOVA)

	Degré de liberté	*Somme des carrés*	*Moyenne des carrés*	*F*	*Seuil expérimental*
Régression	20	14.71301767	0.735650883	1.760388343	0.019542198
Résidus	2979	1244.898031	0.417891249		
Total	2999	1259.611048			

niveau de confiance = 95%

	Coefficients	*Erreur-type*	*Statistique t*	*Seuil exp.*	*Limite inférieure*	*Limite supérieure*
Constante	14.1490407	1.135110609	12.46490042	8.43203E-35	11.92336051	16.3747209
San Diego	0.044205559	0.037565182	1.176769448	0.239381558	-0.029450772	0.11786189
Fresno	0.074822262	0.037571073	1.991485896	0.046518664	0.001154381	0.148490143
Sacramento	0.046287273	0.037214052	1.243811695	0.213666739	-0.026680575	0.11925512
Los Angeles	0.036295851	0.037054579	0.979524047	0.327400652	-0.036359308	0.10895101
CA	0.018569678	0.036551964	0.508035037	0.611466386	-0.053099974	0.090239331
IL	0.040982442	0.037511592	1.092527386	0.274689703	-0.032568812	0.114533696
TX	-0.004227469	0.038054633	-0.11108947	0.911552883	-0.078843495	0.070388557
FL	0.057175178	0.037908087	1.50825805	0.131594536	-0.017153507	0.131503863
Bedrooms	-0.013177752	0.006941411	-1.898425512	0.057736607	-0.026788197	0.000432694
Bathrooms	0.004459065	0.010540427	0.423044082	0.672293635	-0.016208189	0.025126319
Ln(Area)	-0.03995364	0.02364807	-1.689509515	0.091226468	-0.086321845	0.006414566
Ln(Lotsize)	-0.053491853	0.019887067	-2.68978095	0.007189767	-0.09248563	-0.014498076
Year Built	3.46102E-06	0.000556554	0.006218665	0.995038671	-0.001087808	0.00109473
Days on Market	0.000198371	0.000339775	0.583831672	0.559377769	-0.000467846	0.000864588
Townhouse	-0.015797326	0.037470195	-0.421597108	0.673349536	-0.089267409	0.057672757
Apartment	-0.019590019	0.037328483	-0.524800826	0.599760708	-0.09278224	0.053602202
Multi-Family	0.001468571	0.038231462	0.038412634	0.969361262	-0.073494174	0.076431316
Condo	-0.096118428	0.037389806	-2.570712147	0.01019711	-0.169430887	-0.022805969
For sale	0.01331712	0.029326513	0.454098321	0.649791143	-0.044185151	0.070819392
Sold	0.056814576	0.02868829	1.980409976	0.047749337	0.000563706	0.113065446

Bibliographie

- Centris. (2025). Immobilier à Montréal : quel est le prix des maisons en 2025? https://www.centris.ca/fr/blogue/immobilier/immobilier-a-montreal-quel-est-le-prix-des-maisons-en-2025

- SCHL. (2025). Perspectives du marché de l'habitation de 2025. https://www.cmhc-schl.gc.ca/professionnels/marche-du-logement-donnees-et-recherche/marches-de-lhabitation/rapports-sur-le-marche-de-lhabitation/perspectives-du-marche-de-lhabitation

- XpertSource. (2025). Condo vs maison : 6 facteurs à considérer avant d'acheter. https://xpertsource.com/blogue/acheter-bien-immobilier/condo-vs-maison-facteurs-a-considerer-avant-dacheter
- van Veen, J. (2023). A transformation for the pricing of real estate data. Erasmus University. https://thesis.eur.nl/pub/67189/Final_Thesis_Jordi_van_Veen_Erasmus.pdf

- Changyong, F., et al. (2014). Log-transformation and its implications for data analysis. PMC. https://pmc.ncbi.nlm.nih.gov/articles/PMC4120293/

- Rokaandy. (2019). Logarithmic Transformation in Linear Regression Models: Why & When. Dev.to. https://dev.to/rokaandy/logarithmic-transformation-in-linear-regression-models-why-when-3a7c