

Le marché de l'immobilier dans le comté de King, Washington



Lancelot Ravier

Onur Ozdemir

Othmane Laaraki

Quels sont les **facteurs** les plus significatifs **influençant** les **prix des logements** dans le comté de King,

et **comment** peuvent-ils être **quantifiés** en termes **d'impact** **sur le prix de vente** des biens ?

```

# traitement des valeurs extrêmes pour bedroom et bathrooms
df = df.drop(df[(df['bedrooms'] == 33)].index)
df = df.drop(df[(df['bathrooms'] == 0)].index)

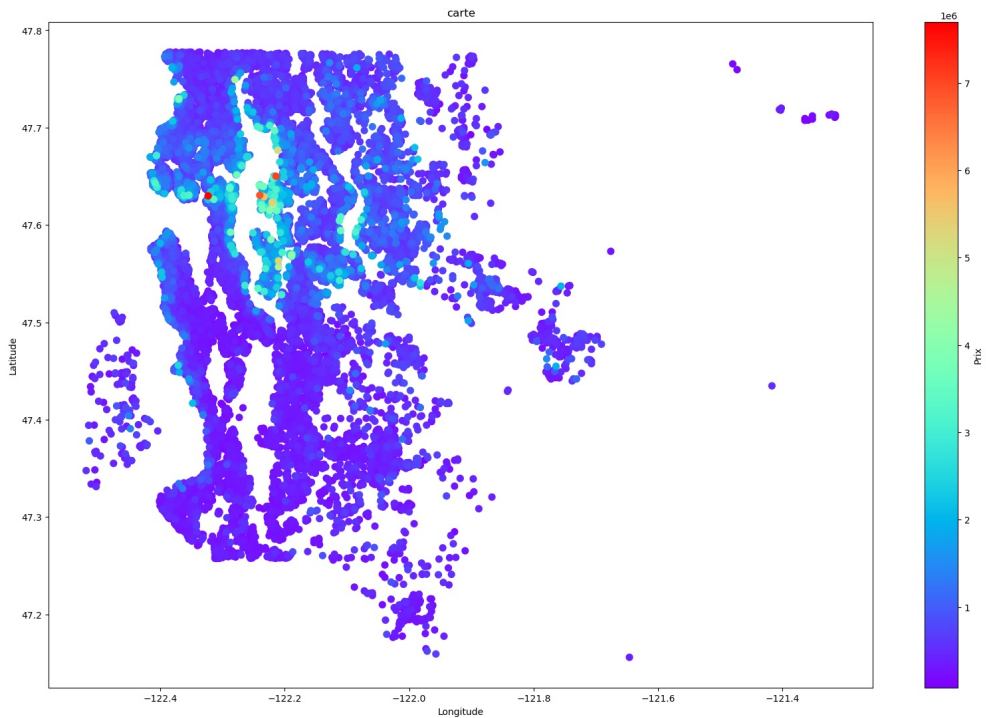
# creation d'une variable binanre pour yr_renovated
df['renovated'] = np.where(df['yr_renovated'] == 0, 0, 1)

# creation d'une variable age pour yr_built
df['age'] = 2015 - df['yr_built']

```

Biens vendus en \$, comté de King, de Mai 2014 à Mai 2015 - N = 21602

	Moyenne	Médiane	E.T.	Min	Max
price	540100	450000	367100	78000	7700000
sqft_living	2080	1910	918,2	370,0	13540
waterfront	0,007546	0,0000	0,08654	0,0000	1,000
view	0,2342	0,0000	0,7663	0,0000	4,000
condition	3,410	3,000	0,6505	1,000	5,000
lat	47,56	47,57	0,1386	47,16	47,78
long	-122,2	-122,2	0,1407	-122,5	-121,3
age	43,99	40,00	29,37	0,0000	115,0
l_price	13,05	13,02	0,5264	11,26	15,86



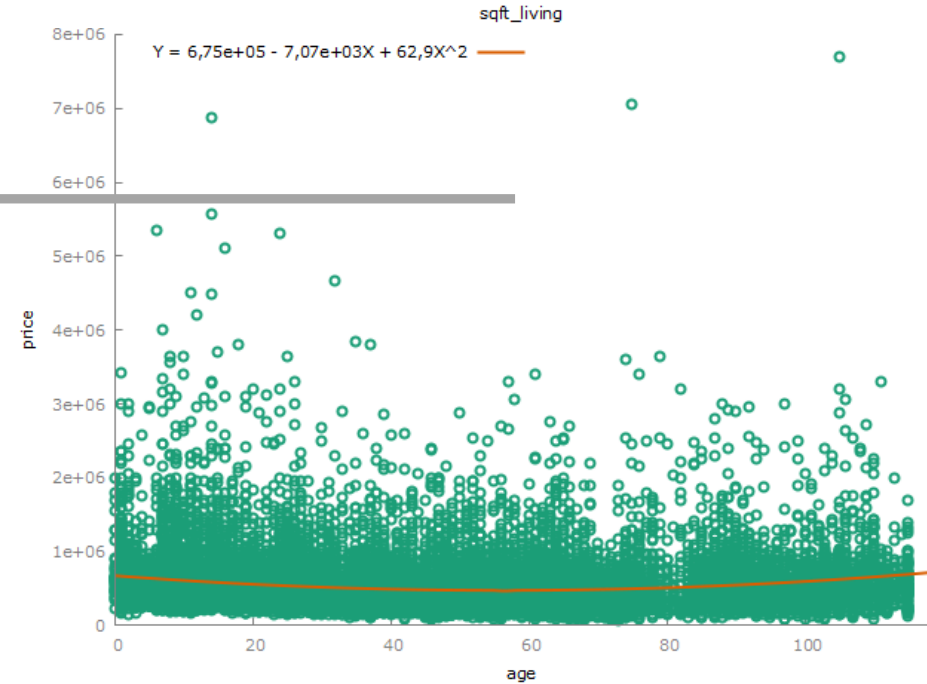
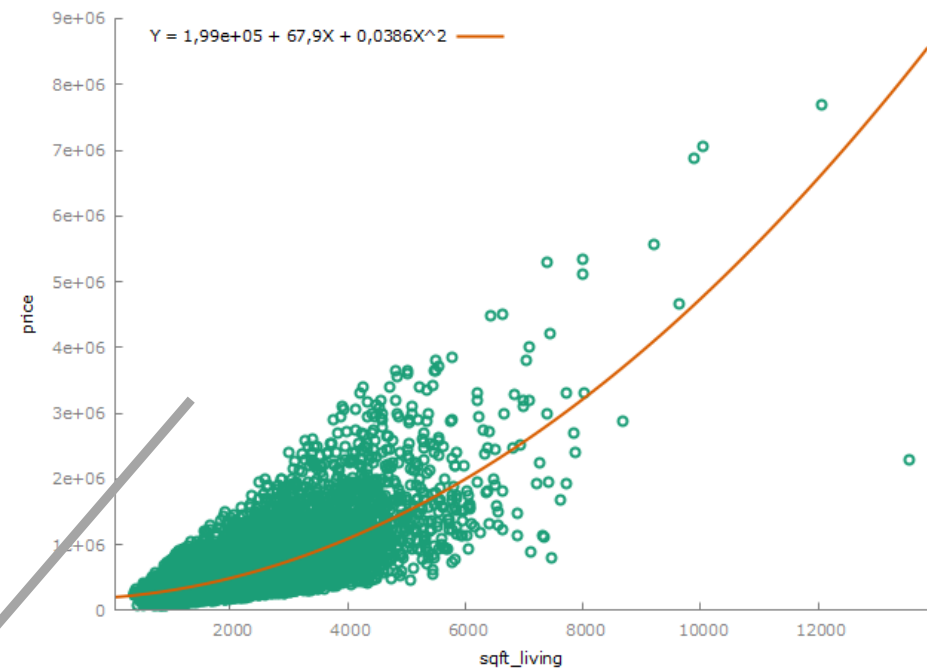
Variable dépendante

`l_price`

Régresseurs

`const`
`sqft_living`
`waterfront`
`view`
`condition`
`lat`
`long`
`age`
`sq_age`
`sq_sqft_living`

price vs sqft_living (avec ajustement quadratique)



Test de White pour l'hétéroscédasticité -

Hypothèse nulle: homoscedasticité

Statistique de test: LM = 2025,4

avec p. critique = P(Khi-deux(52) > 2025,4) = 0

Tests de Student joints à 1% :

$$t_{sqft_living+2*sq_sqft_living} = 59.5894 > 2.5758$$

$$t_{age+2*sq_age} = -28.4517 > 2.5758$$

Modèle : MCO, utilisant les observations 1-21602

Variable dépendante: l_price

Écarts-types robustes (hétéroscédasticité), variante HC0

	coefficient	éc. type	t de Student	p. critique		I.C. 99%	
						t(21592, 0.001) = 2.576	
						[basse	haute]
const	-68,4642	2,06164	-33,21	6,84e-236	***	-73,7751	-63,1532
sqft_living	0,000487585	8,18443e-06	59,57	0,0000	***	0,000466501	0,000508668
waterfront	0,379973	0,0267956	14,18	1,94e-045	***	0,310946	0,449001
view	0,0973895	0,00324159	30,04	2,60e-194	***	0,0890390	0,105740
condition	0,0690004	0,00356351	19,36	7,98e-083	***	0,0598206	0,0781802
lat	1,54105	0,0133929	115,1	0,0000	***	1,50655	1,57555
long	-0,0586501	0,0164450	-3,566	0,0004	***	-0,101013	-0,0162869
age	-0,00707098	0,000247589	-28,56	4,02e-176	***	-0,00770878	-0,00643317
sq_age	7,37048e-05	2,27903e-06	32,34	4,13e-224	***	6,78339e-05	7,95758e-05
sq_sqft_living	-2,29261e-08	1,47763e-09	-15,52	5,34e-054	***	-2,67326e-08	-1,91197e-08
Moyenne var. dép.	13,04798	Éc. type var. dép.	0,526428				
Somme carrés résidus	1730,758	Éc. type régression	0,283121				
R2	0,710876	R2 ajusté	0,710755				
F(9, 21592)	5264,636	P. critique (F)	0,000000				

Test pour la normalité des résidus -

Hypothèse nulle: l'erreur est distribuée selon une loi normale

Statistique de test: Khi-deux(2) = 457,473

avec p. critique = 4,58066e-100

Test de Chow pour rupture structurelle à l'observation 10801 -

Hypothèse nulle: pas de rupture structurel

Statistique du test asymptotique: Khi-deux(10) = 19,256

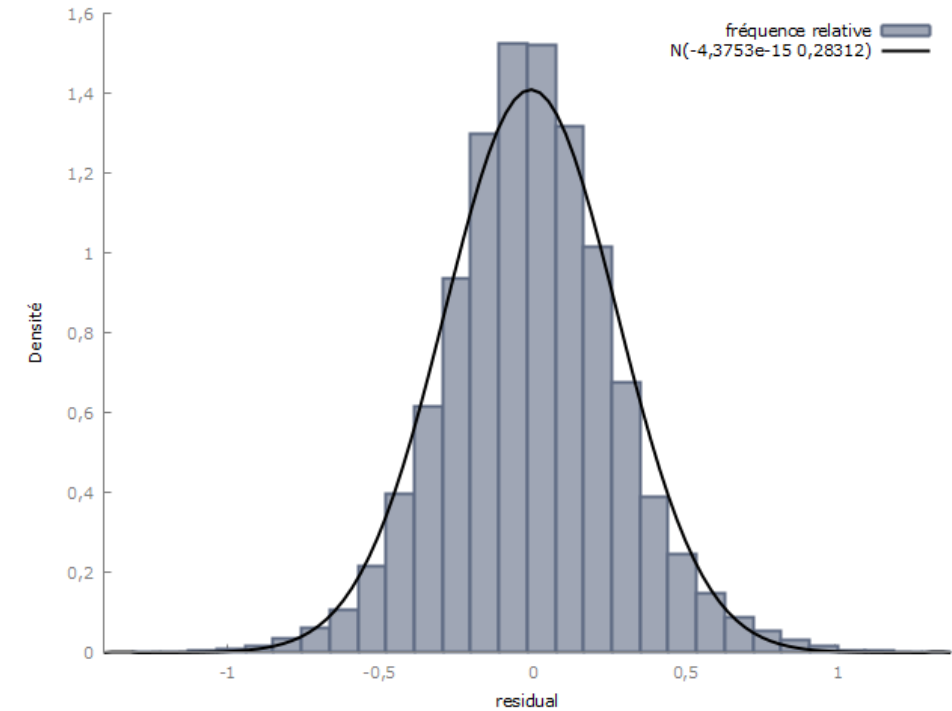
avec p. critique = 0,0371296

Test de Chow pour différence structurelle par rapport à renovated -

Hypothèse nulle: pas de différence structurelle

Statistique du test asymptotique: Khi-deux(10) = 148,464

avec p. critique = 7,71497e-27



Conclusion : Effets non linéaires / Overfitting

Ajout de nouvelles variables d'identification :

- Quartiers hypercentre
- Biens de luxe
- Rénovation

Autres types de régressions non linéaires :

- Ridge, Lasso