

Thomas Jarossay
Yassine Msilini
Lina Taguengayte
Mikail Yilmaz

Rapport d'étape

Présentation du dataset

Nous avons choisi un dataset qui contient des caractéristiques des maisons vendues dans le comté de King dans l'État de Washington. On a par exemple comme caractéristiques la date de vente, le prix, le nombre de chambres, de salle de bain etc... Ce sont les données des maisons vendues entre mai 2014 et mai 2015.

Ce dataset est disponible avec le lien suivant :

<https://www.kaggle.com/harlfoxem/housesalesprediction>

Données

Ce dataset comporte le prix de vente des maisons, le nombre de chambres, de salle de bains, le nombre de mètres carrés...

Ce data set comporte 21 labels et 21613 lignes.

	id	date	price	bedrooms	bathrooms	sqft_living	sqft_lot	floors	waterfront	view	...	grade	sqft_above	s
0	7129300520	20141013T000000	221900.0	3	1.00	1180	5650	1.0	0	0	...	7	1180	
1	6414100192	20141209T000000	538000.0	3	2.25	2570	7242	2.0	0	0	...	7	2170	
2	5631500400	20150225T000000	180000.0	2	1.00	770	10000	1.0	0	0	...	6	770	
3	2487200875	20141209T000000	604000.0	4	3.00	1960	5000	1.0	0	0	...	7	1050	
4	1954400510	20150218T000000	510000.0	3	2.00	1680	8080	1.0	0	0	...	8	1680	
...	
21608	263000018	20140521T000000	360000.0	3	2.50	1530	1131	3.0	0	0	...	8	1530	
21609	6600060120	20150223T000000	400000.0	4	2.50	2310	5813	2.0	0	0	...	8	2310	
21610	1523300141	20140623T000000	402101.0	2	0.75	1020	1350	2.0	0	0	...	7	1020	
21611	291310100	20150116T000000	400000.0	3	2.50	1600	2388	2.0	0	0	...	8	1600	
21612	1523300157	20141015T000000	325000.0	2	0.75	1020	1076	2.0	0	0	...	7	1020	

21613 rows × 21 columns

labels	Attributs
id	identifiant unique de la vente
date	date de la vente au format YearMoDaT000000 ex : 20150225T000000
price	prix de vente de la maison

bedrooms	nombre de chambre
bathrooms	nombre de salle de bain
sqft_living	taille de l'espace habitable en m²
sqft_lot	taille du terrain en m²
floors	nombre d'étage
waterfront	1 si la maison est sur un front de mer 0 sinon
view	un indice de 0 à 4 sur comment est la vue
condition	<p>indice de 0 à 5 sur l'état de la maison</p> <p>1 = Poor- Worn out. Repair and overhaul needed on painted surfaces, roofing, plumbing, heating and numerous functional inadequacies. Excessive deferred maintenance and abuse, limited value-in-use, approaching abandonment or major reconstruction; reuse or change in occupancy is imminent. Effective age is near the end of the scale regardless of the actual chronological age.</p> <p>2 = Fair- Badly worn. Much repair needed. Many items need refinishing or overhauling, deferred maintenance obvious, inadequate building utility and systems all shortening the life expectancy and increasing the effective age.</p> <p>3 = Average- Some evidence of deferred maintenance and normal obsolescence with age in that a few minor repairs are needed, along with some refinishing. All major components still functional and contributing toward an extended life expectancy. Effective age and utility is standard for like properties of its class and usage.</p> <p>4 = Good- No obvious maintenance required but neither is everything new. Appearance and utility are above the standard and the overall effective age will be lower than the typical property.</p> <p>5= Very Good- All items well maintained, many having been overhauled and repaired as they have shown signs of wear, increasing the life expectancy and lowering the effective age with little deterioration or obsolescence evident with a high degree of utility.</p> <p>source: http://info.kingcounty.gov/assessor/esales/Glossary.aspx?type=r</p>
grade	<p>indice de 1 à 13 sur la qualité de la maison</p> <p>1-3 Falls short of minimum building standards. Normally cabin or inferior structure.</p> <p>4 Generally older, low quality construction. Does not meet code.</p> <p>5 Low construction costs and workmanship. Small, simple design.</p> <p>6 Lowest grade currently meeting building code. Low quality materials and simple designs.</p> <p>7 Average grade of construction and design. Commonly seen in plats and older sub-divisions.</p> <p>8 Just above average in construction and design. Usually better materials in both the exterior and interior finish work.</p> <p>9 Better architectural design with extra interior and exterior design and quality.</p>

	<p>10 Homes of this quality generally have high quality features. Finish work is better and more design quality is seen in the floor plans. Generally have a larger square footage.</p> <p>11 Custom design and higher quality finish work with added amenities of solid woods, bathroom fixtures and more luxurious options.</p> <p>12 Custom design and excellent builders. All materials are of the highest quality and all conveniences are present.</p> <p>13 Generally custom designed and built. Mansion level. Large amount of highest quality cabinet work, wood trim, marble, entry ways etc.</p> <p>source: http://info.kingcounty.gov/assessor/esales/Glossary.aspx?type=1</p>
sqft_above	espace habitable en m ² au dessus du sol
sqft_basement	espace habitable en m ² en dessous du sol
yr_built	année de construction
yr_renovated	année de rénovation 0 si la maison n'a pas été rénové
zipcode	code postale
lat	latitude
long	longitude
sqft_living15	moyenne de l'espace habitable des 15 maisons les plus proches, en m ²
sqft_lot15	moyenne de la tailles des terrains des 15 maisons les plus proches, en m ²

Tâche

Notre objectif va être de prédire le prix d'une maison grâce à nos données. La tâche de notre modèle est donc de faire une régression linéaire.

Notre métrique de performance va être la différence entre notre prix de vente prédit et le prix de vente réel.