

Etudiant : OUAHIB Yassine
Filière : Ingénierie des connaissances et sciences des données

Sommaire

I. Introduction	3
1. DÉFINITION DES BESOINS MÉTIERS	
2. DÉTERMINATION DES OBJECTIFS ET DE L'UTILITÉ DE LA CONNAISSANCE	
3. ÉVALUATION DE LA SITUATION	
II.Préparation des données	3
1. IMPORTATION DES DONNÉES	
2. EXPLORATION DES DONNÉES	
3. NETTOYAGE ET PRÉTRAITEMENT DES DONNÉES	
II. Sélection des données.....	5
IV. Modeling	6
1. RÉGRESSION LINÉAIRE	
a-Régression linéaire Simple	
b-Régression linéaire multiple	
2. RÉGRESSION RÉGULARISÉE	
a-Ridge	
b-Lasso	
c-Elastic Net	
3. RÉGRESSION NON LINÉAIRE	
a- Ajustement Exponentiel	
b- Ajustement Puissance	
c- Ajustement Logarithmique	
d- Ajustement Polynomiale	
V. Conclusion	0 15

I. Introduction

1. DÉFINITION DES BESOINS MÉTIERS

Dans le cadre du projet de prédiction des prix immobiliers en utilisant des modèles de régression, il est essentiel de comprendre les besoins métiers qui motivent cette démarche. L'objectif principal est de fournir une estimation précise des prix immobiliers afin d'aider les acteurs du marché immobilier, tels que les agents immobiliers, les acheteurs potentiels et les investisseurs, à prendre des décisions éclairées. En comprenant les besoins et les attentes de ces parties prenantes, nous pourrons développer un modèle de prédiction robuste qui répondra aux exigences spécifiques du marché immobilier.

2. DÉTERMINATION DES OBJECTIFS ET DE L'UTILITÉ DE LA CONNAISSANCE

L'objectif de ce projet est de développer un modèle de régression capable de prédire les prix immobiliers avec précision. Une fois le modèle construit, il pourra être utilisé pour estimer les prix des biens immobiliers sur la base de différentes variables telles que la taille, l'emplacement, les caractéristiques spécifiques de la propriété, etc. Cette connaissance prédictive sera d'une grande utilité pour les acteurs du marché immobilier, leur permettant de fixer des prix compétitifs, d'évaluer la valeur des biens immobiliers et d'optimiser leurs décisions d'achat ou de vente.

3. ÉVALUATION DE LA SITUATION :

Avant de commencer le développement du modèle de prédiction, il est important d'évaluer la situation actuelle du marché immobilier. Cela comprend l'analyse des données disponibles, la compréhension des tendances et des fluctuations du marché, ainsi que l'identification des variables qui ont le plus d'influence sur les prix immobiliers. Une évaluation minutieuse de la situation permettra de sélectionner les bonnes variables, de définir des hypothèses appropriées et de choisir les modèles de régression les plus pertinents pour obtenir des résultats précis et fiables.

II. Préparation des données

1- IMPORTATION DES DONNÉES

Nous avons commencé d'abord uploader notre dataset à partir de Kaggle composé de 81 colonnes et 1460 lignes. On y trouve comme exemple les variables suivant :

- **SalePrice:** le prix de vente de la propriété en dollars. C'est la variable cible que on essayez de prédire.
- **OverallQual:** Qualité générale des matériaux et de la finition
- **YearBuilt:** Date de construction originale
- **YearRemodAdd:** Date de rénovation
- **TotalBsmtSF:** Pieds carrés total de la surface du sous-sol
- **GarageCars:** Taille du garage en nombre de voitures qu'il peut contenir.

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities	...	PoolArea	PoolQC	Fence	MiscFeature	MiscVal	MoSold	YrSold	SaleType	SaleCondition	
0	1	60	RL	65.0	8450	Pave	NaN	Reg	Lvl	AllPub	...	0	NaN	NaN	NaN	0	2	2008	WD	Normal	
1	2	20	RL	80.0	9600	Pave	NaN	Reg	Lvl	AllPub	...	0	NaN	NaN	NaN	0	5	2007	WD	Normal	
2	3	60	RL	68.0	11250	Pave	NaN	IR1	Lvl	AllPub	...	0	NaN	NaN	NaN	0	9	2008	WD	Normal	
3	4	70	RL	60.0	9550	Pave	NaN	IR1	Lvl	AllPub	...	0	NaN	NaN	NaN	0	2	2006	WD	Abnorml	
4	5	60	RL	84.0	14260	Pave	NaN	IR1	Lvl	AllPub	...	0	NaN	NaN	NaN	0	12	2008	WD	Normal	
...	
1455	1456	60	RL	62.0	7917	Pave	NaN	Reg	Lvl	AllPub	...	0	NaN	NaN	NaN	0	8	2007	WD	Normal	
1456	1457	20	RL	85.0	13175	Pave	NaN	Reg	Lvl	AllPub	...	0	NaN	MnPrv	NaN	0	2	2010	WD	Normal	
1457	1458	70	RL	66.0	9042	Pave	NaN	Reg	Lvl	AllPub	...	0	NaN	GdPrv	Shed	2500	5	2010	WD	Normal	
1458	1459	20	RL	68.0	9717	Pave	NaN	Reg	Lvl	AllPub	...	0	NaN	NaN	NaN	0	4	2010	WD	Normal	
1459	1460	20	RL	75.0	9937	Pave	NaN	Reg	Lvl	AllPub	...	0	NaN	NaN	NaN	0	6	2008	WD	Normal	

1460 rows x 81 columns

Figure 1 : table de donnée « Sales Price » initialement importé

2- EXPLORATION DES DONNÉES

Après avoir importé notre dataset ainsi que les librairies nécessaires ,nous avons passé à l'exploration des données, étape cruciale dans le processus d'analyse de données qui consiste à examiner les données de manière approfondie afin de découvrir des tendances, des modèles et des relations entre les différentes variables.

Dans cette étape on va Visualiser quelques figures illustrant la densité de la distribution des différentes colonnes, une autre sous forme histogramme pour visualiser la variable « sales prices » et bien d'autres afin de découvrir les tendances et des relations entre les différentes variables.

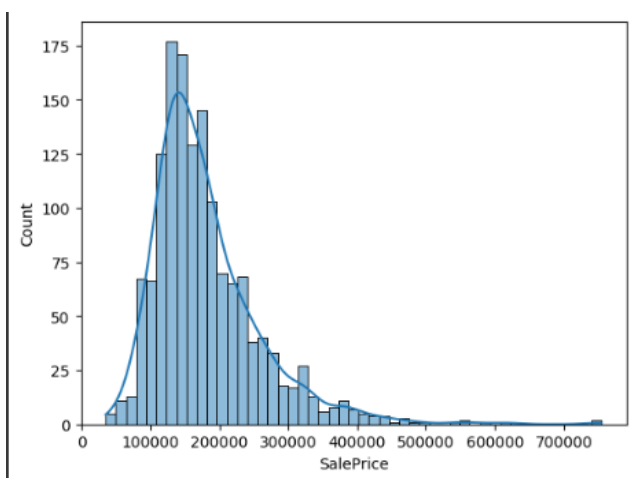


Figure 2: Histogramme des valeurs de "SalePrice"

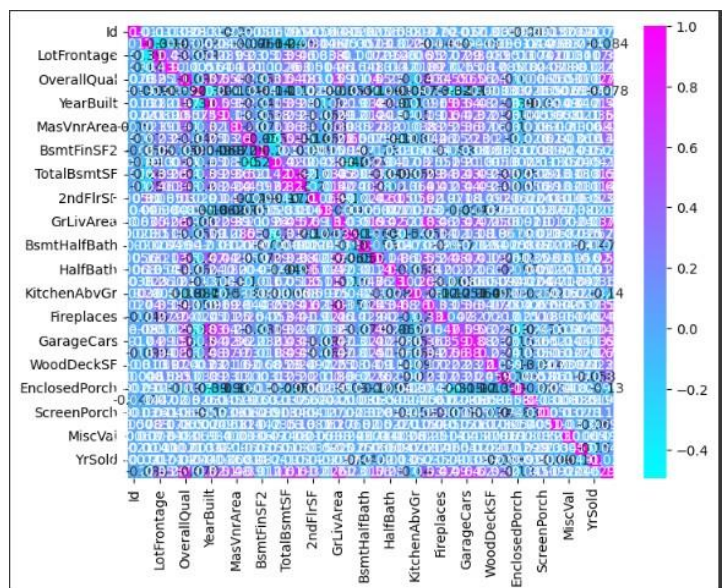


Figure 3 : Corrélation entre les variables

3- NETTOYAGE ET PRÉTRAITEMENT DES DONNÉES

En effet, la collection de données initiale présente des lacunes, des valeurs manquantes et des anomalies. De plus, il y a un grand nombre de variables qui ne sont pas normalisées, dont une grande partie est de type catégorie ou objet.

On vérifie d’abord les valeurs duplique, après les valeurs nulles puis on supprime les colonnes dont on n’aura pas besoin tels celle de l’ID et les colonnes de type objet ou catégorie. On voit bien dans la figure 4 qu’il existe encore des valeurs nulles représentées par des tirés blancs et qui ont été remplacées par leur moyenne. Il n'existe ainsi plus de valeur manquante.

Nous avons passé après pour traiter les valeurs aberrantes qui vont être remplacer par la valeur médiane et enfin nous avons vérifié les paires fortement corrélés pour éliminer les variables redondantes.

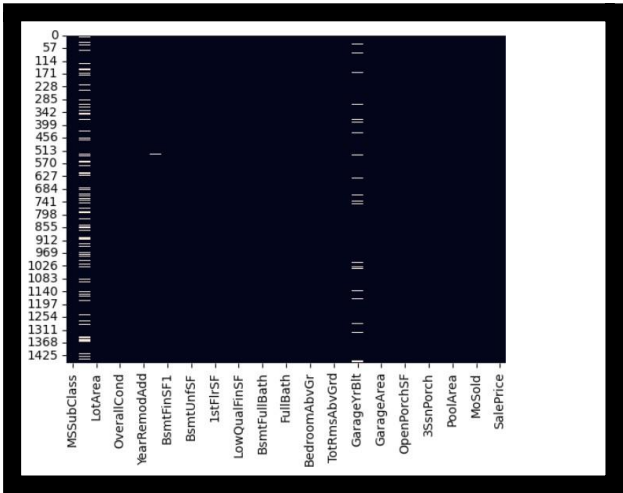


Figure 4: visualisation des valeurs manquantes

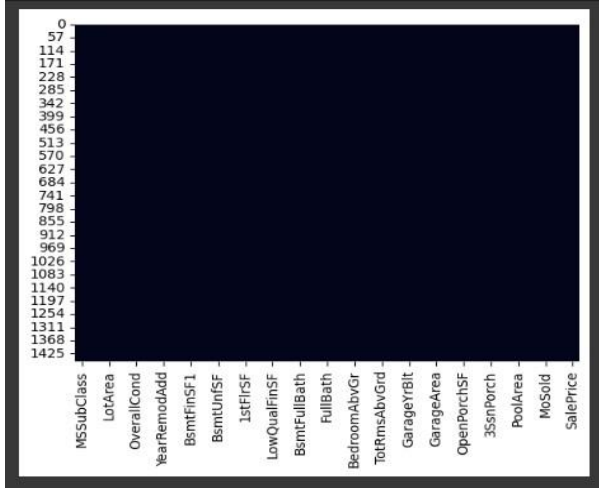


Figure 5: visualisation des valeurs manquantes après traitement

III. Sélection des données

La deuxième étape a été la sélection des données. Nous avons choisi les variables qui étaient les plus susceptibles d’avoir un impact sur les ventes de notre entreprise.

Pour cela, Nous avons utilisé deux techniques :

- La sélection descendante par étapes (Backward Stepwise Selection) :

une méthode de sélection de variables dans un modèle statistique qui commence avec toutes les variables et élimine une par une celles qui ont le moins d'impact sur le modèle jusqu'à ce qu'un modèle final soit atteint. Cette méthode est utile pour réduire la complexité du modèle et éviter le sur ajustement (overfitting).

- La corrélation :

Afin de trouver les variables qui ont une forte relation avec la variable cible, pour enfin construire un modèle prédictif plus performant.

```

Paires de variables corrélées :
('SalePrice', 'YearBuilt')
('SalePrice', 'YearRemodAdd')
('SalePrice', 'TotalBsmtSF')
('SalePrice', 'GrLivArea')
('SalePrice', 'FullBath')
('SalePrice', 'TotRmsAbvGrd')
('SalePrice', 'Fireplaces')
('SalePrice', 'GarageYrBlt')
('SalePrice', 'GarageArea')
('SalePrice', 'SalePrice')

```

Et voici la version finale des données à conserver :

	YearBuilt	YearRemodAdd	TotalBsmtSF	GrLivArea	FullBath	TotRmsAbvGrd	GarageArea	SalePrice
0	2003.0	2003	856.0	1710.0	2	8.0	548.0	208500.0
1	1976.0	1976	1262.0	1262.0	2	6.0	460.0	181500.0
2	2001.0	2002	920.0	1786.0	2	6.0	608.0	223500.0
3	1915.0	1970	756.0	1717.0	1	7.0	642.0	140000.0
4	2000.0	2000	1145.0	2198.0	2	9.0	836.0	250000.0
...
1455	1999.0	2000	953.0	1647.0	2	7.0	460.0	175000.0
1456	1978.0	1988	1542.0	2073.0	2	7.0	500.0	210000.0
1457	1941.0	2006	1152.0	2340.0	2	9.0	252.0	266500.0
1458	1950.0	1996	1078.0	1078.0	1	5.0	240.0	142125.0
1459	1965.0	1965	1256.0	1256.0	1	6.0	276.0	147500.0

1399 rows x 8 columns

Figure 6: table de donnée « Sales Price » final

IV. Modeling

Ensuite, nous sommes passés à l'étape de la modélisation. Nous avons 3 types de régression :

- Régression linéaire
- Régression régularisée
- Régression non linéaire

1-RÉGRESSION LINÉAIRE

a-Régression linéaire Simple

Dans la régression linéaire simple on a choisi comme variable explicative **TotalBsmtSF** (Total square feet of basement area) puis on compare le modèle obtenu avec celui dont la variable explicative est **GrLivArea**.

Resultat:

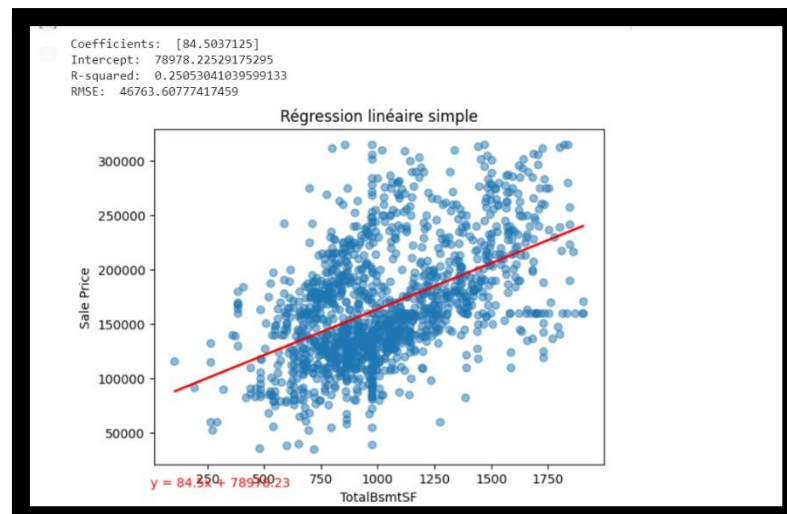


Figure 7: visualisation graphique de la régression linéaire simple

Les résultats de la régression linéaire affichent les coefficients de régression, l'ordonnée à l'origine, le coefficient de détermination (R-squared) et la racine carrée de l'erreur quadratique moyenne (RMSE).

Le coefficient de régression est de 84.5, ce qui signifie que pour chaque unité d'augmentation de la variable explicative "Total basement square feet", la variable cible "SalePrice" augmente en moyenne de 84.5 unités. L'ordonnée à l'origine est 78978.22., ce qui signifie que si "Total basement square feet" est nul, le prix de vente moyen est de - 78978.22 unités.

Cependant, cela n'a pas beaucoup de sens pratique dans ce contexte. Le coefficient de détermination (R-squared) est de 0.25, ce qui indique que seulement 25% de la variance dans les données est expliquée par la variable explicative "Total basement square feet". La racine carrée de l'erreur quadratique moyenne (RMSE) est de 46763.6, ce qui est la moyenne des erreurs de prédiction de la variable cible "SalePrice". Cela signifie que les valeurs prédites de "SalePrice" ont une erreur moyenne de 46763.6 unités par rapport aux valeurs réelles.

-On compare le modèle obtenu avec celui dont la variable explicative est « **GrLivArea** ».

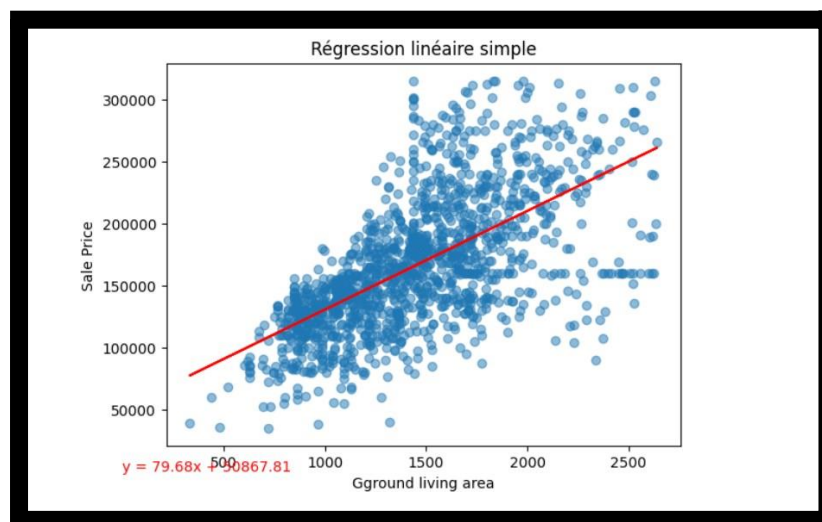


Figure 8: visualisation graphique de la régression linéaire simple la variable explicative est « GrLivArea »

Les résultats pour « **GrLivArea** » montrent un coefficient de régression plus faible mais un R-squared plus important, indiquant que "GrLivArea" est plus fortement corrélée avec "SalePrice" que "Total basement square feet". De plus, la RMSE est plus faible, ce qui suggère que les prédictions de "SalePrice" basées sur "GrLivArea" ont une erreur de prédiction moyenne moins importante que celles basées sur "Total basement square feet".

D'ailleurs la comparaison des modèles en utilisant le **critère d'information d'Akaike (AIC)** montre que le deuxième est le plus performant.

```
Modèle 1 MSE : 2186835012.056842
Modèle 2 MSE : 1769405321.828894
Modèle 1 AIC : 24072.901950615607
Modèle 2 AIC : 23833.884563446267
```

b-Régression linéaire multiple

En ce qui concerne la régression linéaire multiple, elle a été effectuée en premier lieu avec les 7 variables explicatives de l'ensemble de données.

```

Coefficients: [ 421.16175733  476.65364334  30.87801482  45.18428196  237.80254044
2405.7189249    48.45168211]
Intercept: -1744139.5799988285
R-squared (train): 0.6718002349413061
RMSE (train): 31218.76154244277
R-squared (test): 0.6676401138632928
RMSE (test): 29933.759746299987
Équation de la régression linéaire multiple : SalePrice = -1744139.58 + 421.16*YearBuilt + 476.65*YearRemodAdd + 30.88*TotalBsmtSF

```

Les résultats de la régression linéaire multiple indiquent les coefficients estimés pour chaque variable explicative ainsi que l'intercept. Voici une interprétation de ces résultats :

R-squared : indique que le modèle explique environ 64.66% de la variance dans les données d'entraînement et environ 63.99% de la variance dans les données de test. Cela suggère que le modèle est capable d'expliquer une part significative de la variation des prix de vente.

RMSE : indique un RMSE de 32396.14 pour les données d'entraînement et un RMSE de 31155.98 pour les données de test. Ces valeurs représentent l'écart moyen entre les prédictions du modèle et les prix de vente réels.

les résultats de la régression linéaire multiple montrent que les variables explicatives (TotRmsAbvGrd, YearBuilt, YearRemodAdd, GrLivArea, FullBath, GarageArea) ont une influence significative sur le prix de vente des biens immobiliers. Le modèle présente également une performance raisonnable avec un coefficient de détermination élevé et des erreurs relativement faibles.

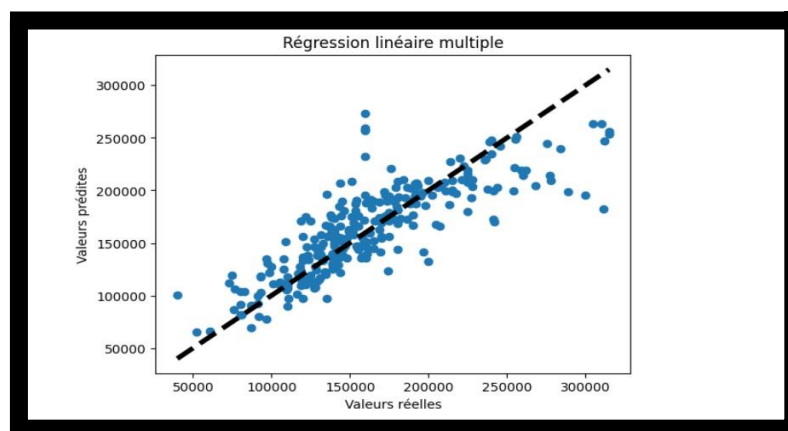


Figure 9: visualisation graphique de la régression linéaire multiple

Ensuite on vérifie si on peut optimiser plus ce modèle en utilisant la RFE : **Recursive Feature Elimination**

Il fonctionne en éliminant de manière récursive les caractéristiques du modèle et en réajustant le modèle jusqu'à ce qu'un nombre souhaité de caractéristiques soit atteint. La nouvelle équation devient ainsi comme suit et les nouveaux résultats sont les suivants.

```

[ ] Coefficients: [ 495.41587081  456.27559353  47.48748825  2048.38484758  2309.17238742
61.37643471]
Intercept: -1829424.006314009
R-squared (train): 0.6465780428839434
RMSE (train): 32396.142552206624
R-squared (test): 0.639944855146595
RMSE (test): 31155.9844448567
Équation de la régression linéaire multiple : SalePrice = -1829424.01 + 495.42*YearBuilt + 456.28*YearRemodAdd + 47.49*GrLivArea + 2048.38*TotalBsmtSF

```

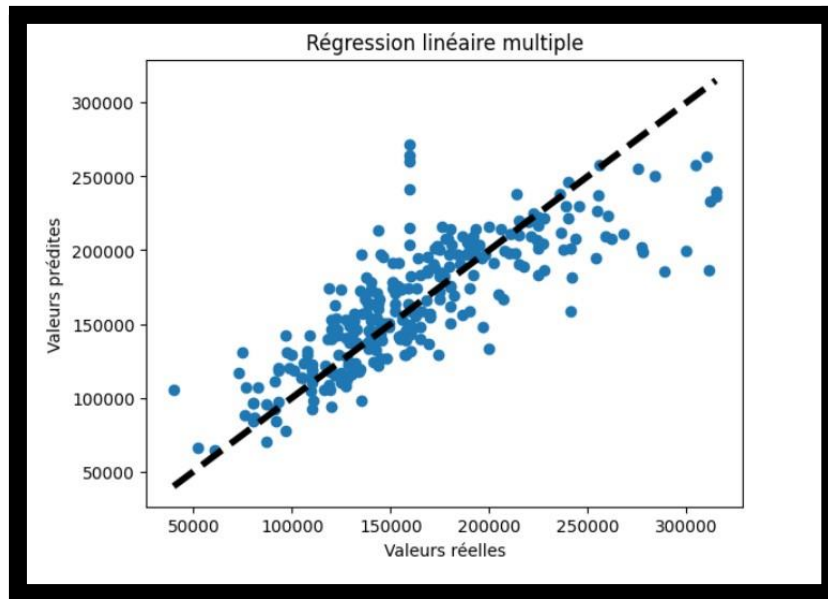



Figure 10: visualisation graphique de la régression linéaire simple en utilisant la RFE

Après la comparaison des deux modèles, on conclut que le premier modèle dont les variables explicatives sont 7 est le plus performant.

```
Modèle 1 MSE : 896029972.5492102
Modèle 2 MSE : 970695366.7281544
Modèle 1 AIC : 23074.489068089908
Modèle 2 AIC : 23162.052511766724
```

On retient ainsi le 2ème modèle de la RLS et le 1er de la RLM et c'est ce dernier qui donne les meilleurs résultats parmi les quatre modèles.

2-RÉGRESSION RÉGULARISÉE

a- Ridge

Tout d'abord, nous allons utiliser Ridge qui est une régression régularisée qui pénalise les coefficients des variables les plus importantes pour éviter le surajustement.

```
The model performance for training set is:
-RMSE is 29871.962883070522
-MSE is 892334166.487543
-R2 score is 0.6690109805970743
```

Un R2 score indique que le modèle explique environ 67% de la variance totale des valeurs de vente. ces résultats indiquent que le modèle de régression Ridge a une certaine capacité de prédiction des valeurs de vente, mais il reste une certaine marge d'erreur. Le R2 score suggère que les variables explicatives dans le modèle expliquent une proportion modérée de la variance dans les valeurs de vente.

b-Lasso:

Ensuite, nous allons utiliser Lasso qui pénalise les coefficients des variables les moins importantes pour éliminer les variables inutiles.

```
The model performance for training set is:
-RMSE is 29878.606342490377
-MSE is 892731116.9695061
-R2 score is 0.6688637417534764
```

Les résultats indiquent que le modèle de régression Lasso a une performance raisonnable sur l'ensemble d'apprentissage. De plus, le R^2 de 0.6689 indique que le modèle explique une part importante de la variance des prix. On remarque que le RMSE est assez proche de celui obtenu avec le modèle Ridge. Donc, on pourrait considérer que les modèles présentent des performances similaires.

c-Elastic Net:

Enfin on utilise Elastic Net, qui est un modèle de régression qui combine à la fois les régularisations de Ridge et de Lasso. Il permet de bénéficier des avantages des deux approches en contrôlant simultanément la magnitude des coefficients et en effectuant une sélection automatique des variables.

```
The model performance for training set is:
-RMSE is 29853.83328503186
-MSE is 891251361.8104762
-R2 score is 0.6694126198839324
```

Les valeurs de RMSE et de MSE sont relativement faibles, ce qui suggère que le modèle peut prédire les prix des logements avec une précision raisonnable. De plus, le R^2 de 0,6694 indique que le modèle explique une partie significative de la variance des prix.

- Dans ce cas, nous avons obtenu les résultats suivants :

Le modèle Ridge a la plus faible RMSE, ce qui signifie qu'il a la meilleure performance parmi les trois modèles testés. Le modèle Lasso a également une performance similaire, tandis que le modèle ElasticNet est légèrement moins performant que les deux autres.

Cela suggère que la méthode de régularisation **Ridge** peut être la plus appropriée et pour la tâche de prédiction des prix de vente des maisons.

Nous pouvons également afficher les coefficients des variables sélectionnées par chaque modèle.

```
Ridge selected variables:
TotRmsAbvGrd      1159.437916
YearBuilt          417.068599
YearRemodAdd       476.035285
TotalBsmtSF        30.917528
GrLivArea          48.443839
FullBath           109.432918
GarageArea         48.457008
dtype: float64
Lasso selected variables:
TotRmsAbvGrd      1211.596848
YearBuilt          417.356038
YearRemodAdd       475.149438
TotalBsmtSF        30.940845
GrLivArea          48.386307
GarageArea         48.493735
dtype: float64
ElasticNet selected variables:
TotRmsAbvGrd      185.268011
YearBuilt          412.301057
YearRemodAdd       465.788450
TotalBsmtSF        31.039977
GrLivArea          51.049611
FullBath           20.236219
GarageArea         48.799260
dtype: float64
```

Le modèle Lasso a sélectionné les mêmes variables que Ridge, à l'exception de GarageCars pour lequel le coefficient est très faible.

Le modèle ElasticNet a également sélectionné les mêmes variables que Ridge et Lasso, avec des coefficients légèrement différents.

Ces résultats suggèrent que ces variables sont les plus importantes pour prédire les prix de vente des maisons.

- **Affichage du graphique de convergence ridge**

Le graphique montre comment les coefficients varient en fonction d'hyper paramètre alpha.

L'axe des x est échelonné logarithmiquement pour mieux visualiser la variation des coefficients à travers une large plage de valeurs alpha. lorsque le paramètre alpha augmente, les coefficients de régression se rapprochent de zéro, ce qui signifie que la régularisation est plus forte. Cela se traduit par une diminution de la complexité du modèle et une réduction du risque d'overfitting sur les données d'entraînement. Les coefficients qui convergent lentement vers zéro sont associés à des prédicteurs qui ont une plus grande influence sur la prédiction des saleprices.

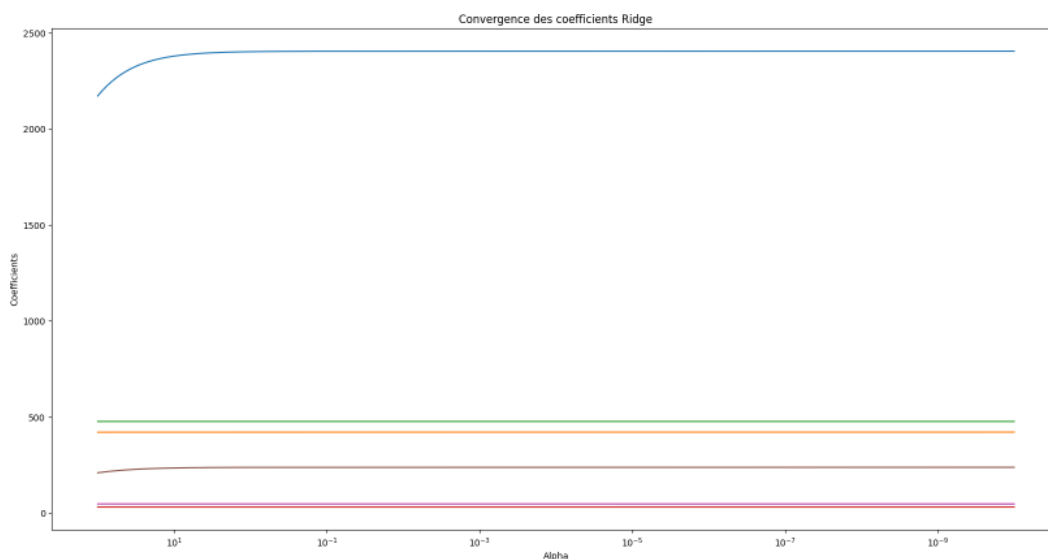


Figure 11: visualisation graphique de la convergence Ridge

- **Affichage du graphique de trajectoire lasso**

On entraîne un modèle Lasso avec une régularisation L1 pour différentes valeurs d'hyperparamètre alpha stocke les coefficients de régression pour chaque valeur d'alpha et on trace la trajectoire des coefficients Lasso en fonction des valeurs d'alpha.

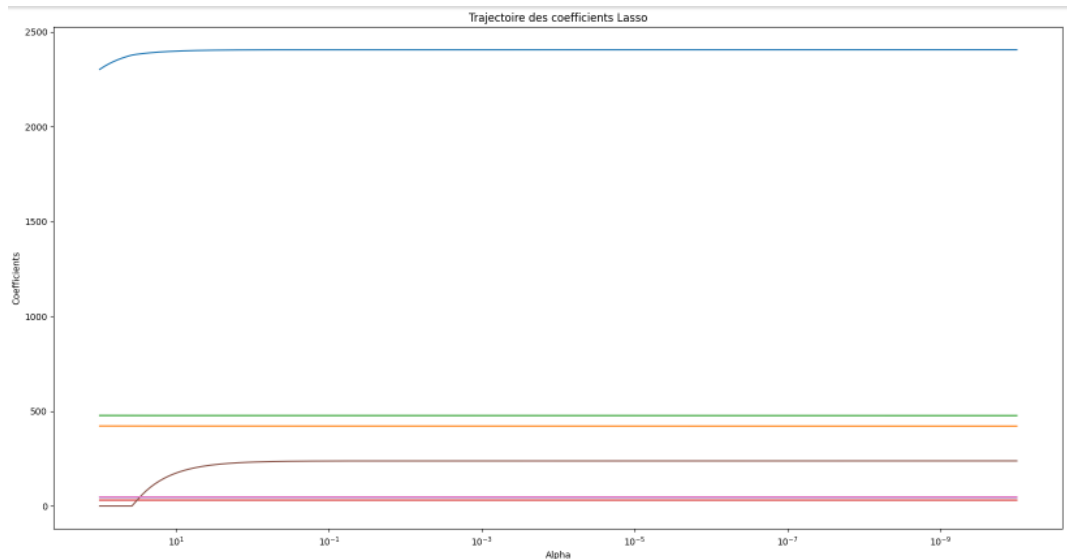


Figure 12: visualisation graphique de la convergence Lasso

- **Affichage du graphique de trajectoire Elastic Net**

La courbe bleue représente les coefficients de régression régularisés L1 (Lasso) et la courbe verte représente les coefficients de régression régularisés L2 (Ridge).

La trajectoire des coefficients L1 (Lasso) diminue plus rapidement que la trajectoire des coefficients L2 (Ridge) lorsque les valeurs d'alpha augmentent.

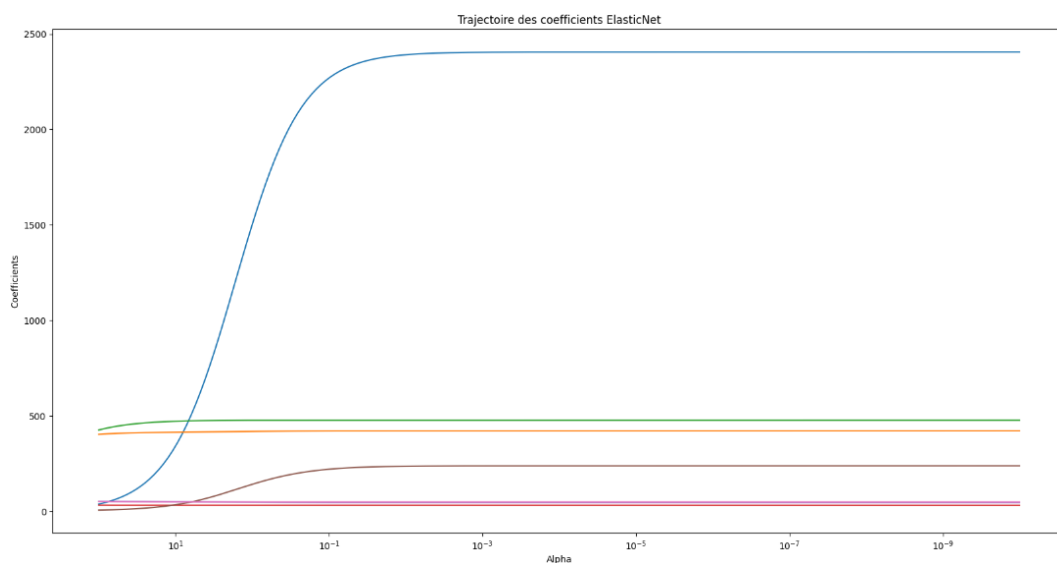


Figure 13: visualisation graphique de la convergence Elastic net

2- RÉGRESSION NON LINÉAIRE

a- Ajustement Exponentiel:

```
les coefficients : [0.00049243]
11.255969473415313
The model performance for training set is:
-RMSE is 0.24730304501789427
-MSE is 0.061158796075122646
-R2 score is 0.42132479237873155
```

Dans le cadre d'une analyse de régression non linéaire avec ajustement exponentiel, les résultats révèlent que ce modèle présente une performance passable .Ou, le coefficient indique que chaque augmentation unitaire dans la variable indépendante est associée à une augmentation exponentielle de 0.00049243 dans la variable cible. Ensuite, lorsque toutes les autres variables sont absentes, la variable cible serait proche de 11.255969473415313. Enfin, R^2 de 0.42132479237873155 suggère que le modèle explique approximativement 42.13% de la variance de la variable cible.

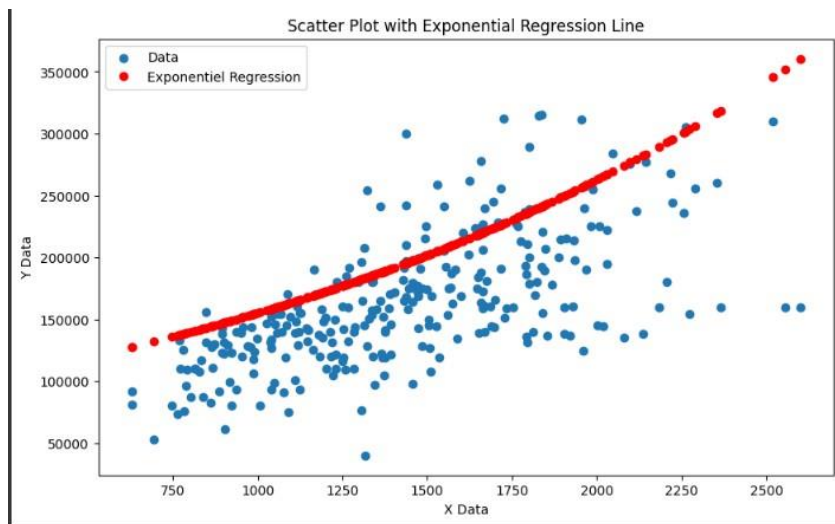


Figure 14: visualisation de Nuage de points avec courbe de régression exponentielle

b- Ajustement Puissance:

```
les coefficients : [0.73399844]
6.660101120420158
The model performance for training set is:
-RMSE is 0.2410283018294537
-MSE is 0.05809464228279023
-R2 score is 0.45031734857266736
```

Dans le cas d'un ajustement de puissance dans la régression, les résultats montrent que ce modèle présente une performance acceptable .Dont le coefficient suggère que chaque augmentation unitaire dans la variable indépendante est associée à une augmentation de puissance de 0.73399844 dans la variable cible.

Ainsi que, R^2 de 0.45031734857266736 suggère que le modèle explique environ 45.03% de la variance de la variable cible.

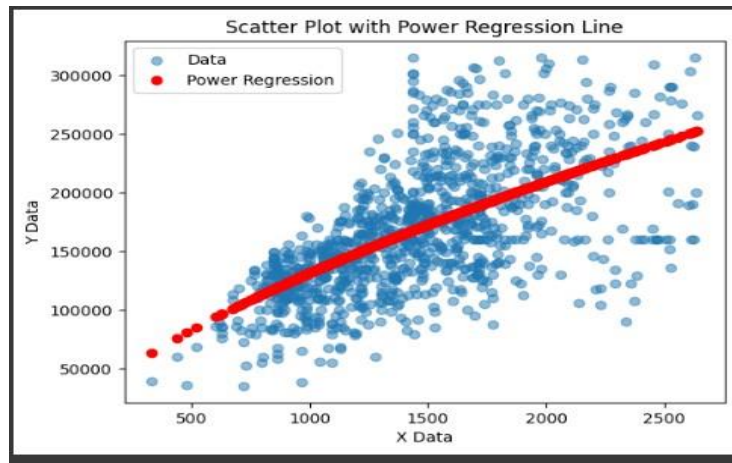


Figure 15: visualisation de Nuage de points avec courbe de régression puissance

c- Ajustement Logarithmique:

```
les coefficients : [114261.33202918]
-659645.5873914409
The model performance for training set is:
-RMSE is 39253.35286257167
-MSE is 1540825710.9535637
-R2 score is 0.42846927721392425
```

Dans le cas d'un ajustement logarithmique dans la régression, les résultats pour R^2 de 0.42846927721 suggère que le modèle explique environ 42.85% de la variance de la variable cible. cela indique que ce modèle présente une performance raisonnable

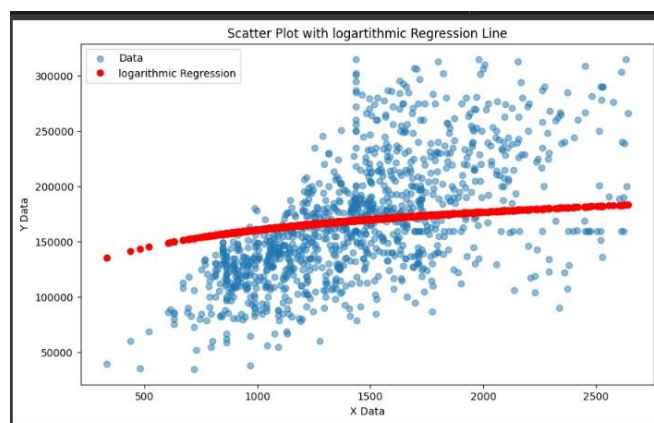


Figure 16: visualisation de Nuage de points avec courbe de régression logarithmique

d- Ajustement Polynomiale:

Ces mesures évaluent la qualité de l'ajustement du modèle aux données d'apprentissage. Les résultats suggèrent que le modèle de régression polynomiale avec un degré de 2 ou 3 explique environ 43.07 % de la variance de la variable cible.

```
The model performance for training set is:
-RMSE is 39177.91360528911
-MSE is 1534908914.4634979
-R2 score is 0.43066396474445134
```

Degré 2

```
The model performance for training set is:
-RMSE is 39177.91360528911
-MSE is 1534908914.4634979
-R2 score is 0.43066396474445134
```

Degré 3

En conclusion, à la lumière de notre analyse approfondie sur la prédiction des prix des maisons en utilisant différents modèles de régression, nous pouvons tirer plusieurs enseignements importants.

Tout d'abord, il est crucial de choisir le modèle de régression approprié en fonction des caractéristiques spécifiques de l'ensemble de données et des objectifs de prédiction. Dans notre étude, nous avons examiné divers modèles, y compris des modèles linéaires et non linéaires tels que le modèle exponentiel, le modèle de puissance, le modèle logarithmique et les modèles polynomiaux. Ces modèles offrent une flexibilité pour capturer les différentes relations entre les variables indépendantes et dépendantes.

Lors de l'évaluation des performances des modèles, nous avons utilisé des mesures clés telles que le RMSE, le MSE et le R2 score. Ces mesures nous ont permis d'évaluer à la fois la précision des prédictions et la capacité du modèle à expliquer la variabilité des données.

Notre analyse a révélé que le modèle de puissance non linéaire a obtenu les meilleures performances pour la prédiction des prix de maisons, avec un RMSE et un MSE inférieurs et un R2 score plus élevé par rapport aux autres modèles. Cela suggère que la relation entre les variables est mieux capturée par ce modèle non linéaire.

En résumé, notre étude démontre l'importance de la prédiction des prix des maisons dans le domaine de l'immobilier en utilisant des techniques de machine learning avancées. Les modèles de régression, tels que le modèle de puissance non linéaire, permettent de capturer les relations complexes entre les caractéristiques des maisons et leurs prix, fournissant ainsi des prédictions précises et utiles pour les décisions liées au marché immobilier. Cependant, il est essentiel de prendre en compte les facteurs externes et les spécificités du marché immobilier pour obtenir des résultats pertinents et fiables.