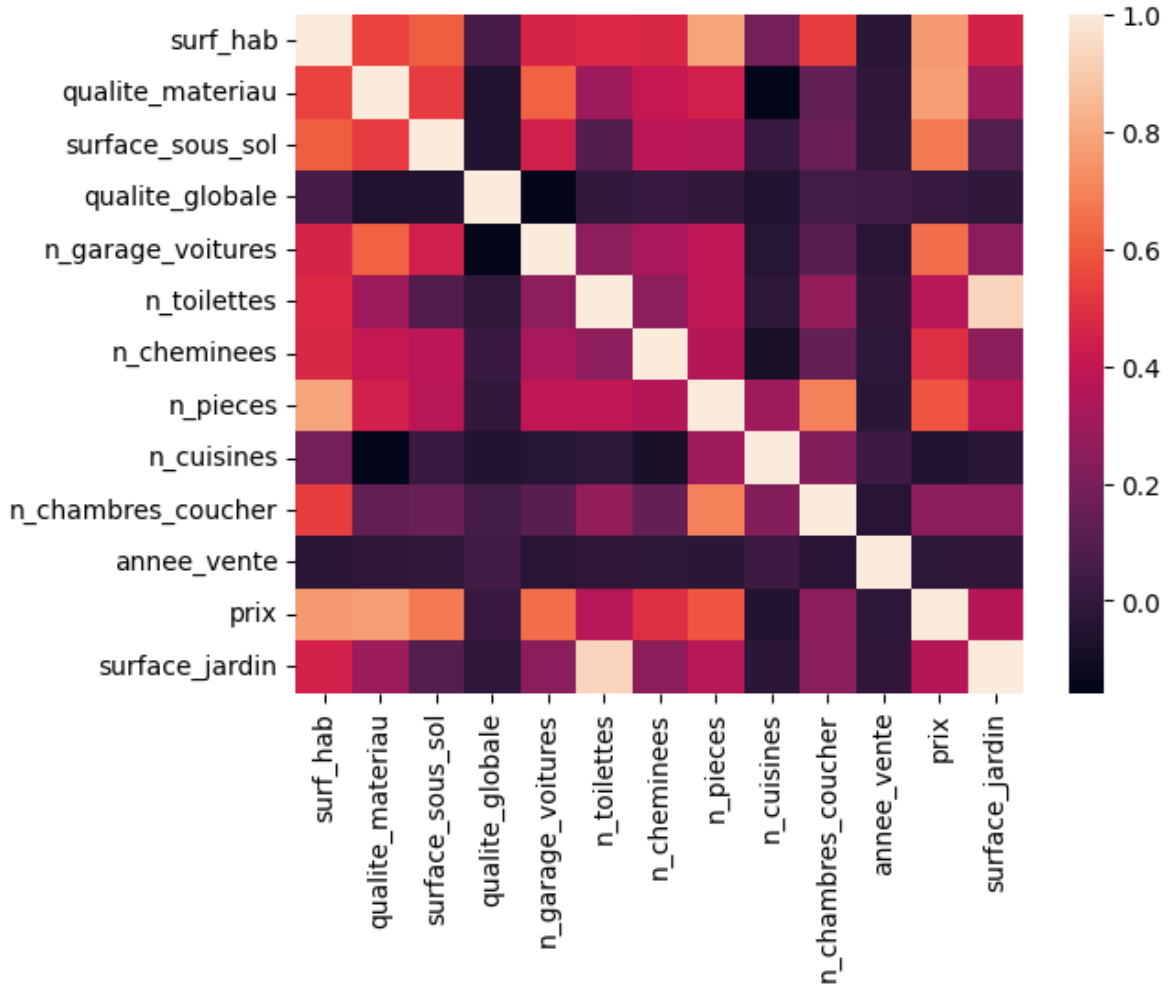
Project 1 - House Price Prediction

Presentation

1. Problem Presentation

- Goal: predict **house prices** from features (quality, number of rooms, surface, etc.).
- Pipeline includes data cleaning, encoding, and train/validation/test splitting.



---

2. Tested Models and Complexities

- Model 1: **Linear Regression** (`sklearn.linear_model.LinearRegression`) – low complexity.
- Model 2: **Polynomial Regression (degree 2)** (`sklearn.preprocessing.PolynomialFeatures` + `LinearRegression`) – higher complexity.
- Model 3: **Ridge Regression** (`sklearn.linear_model.Ridge`) – adds regularization to reduce overfitting.

---

3. Polynomial Degree Choice and Learning Curve

- Tested polynomial regression of degree 2.
- Captures non-linearities but may lead to overfitting.

---

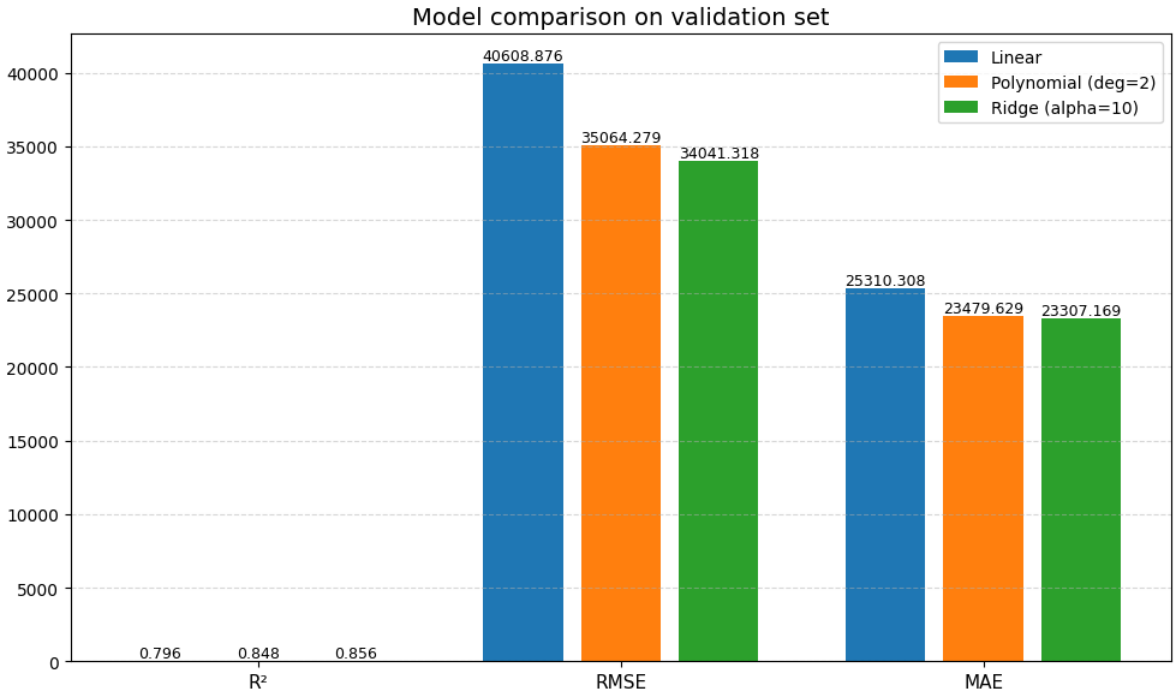
4. Regularization

- Added Ridge regularization to reduce variance and stabilize predictions.
- Shrinks coefficient magnitudes.
- Improves generalization compared to plain polynomial regression.

---

5. Final Model and Cost Function Comparison

- Evaluation metrics: R^2, RMSE, MAE.
- Best compromise obtained with Ridge Regression.
- Comparison across at least 3 (model / cost function) combinations.


Model comparison on validation set
Linear
Polynomial (deg=2)
Ridge (alpha=10)
40608.876
35064.279
34041.318
25310.308
23479.629
23307.169
0.796
0.848
0.856
0
5000
10000
15000
20000
25000
30000
35000
40000
R²
RMSE
MAE