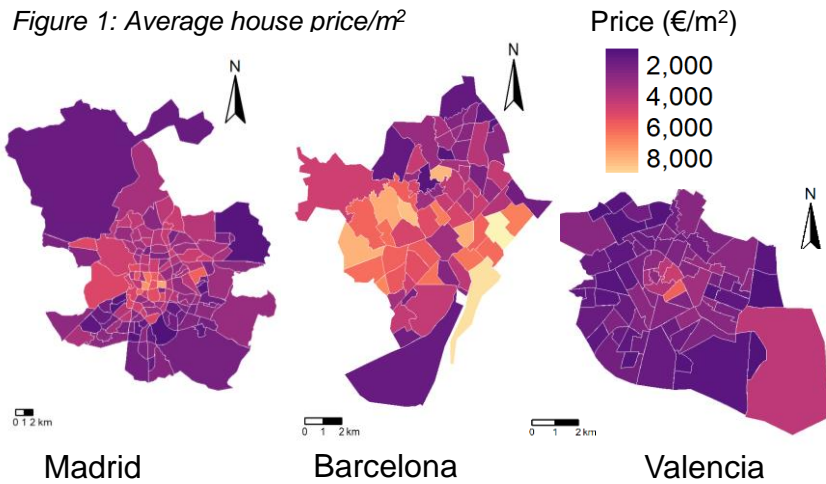


Spanish house prices: What factors are important?

Machine learning analysis of 2018 property advertisement data from Idealista by Gladys Kenyon

Figure 1: Average house price/m²



Background:

- Three largest Spanish cities
- Mean price: Madrid €396112, Barcelona €395770, Valencia €199678
- In Madrid and Valencia houses are most expensive in city centre, compared to West and East areas in Barcelona
- Patterns indicate there are local drivers of house prices

Regression Results:

- A Random Forest (RF) algorithm was used to build the model in R
- The model predictions were wrong by on average €1.15 per meter² (Table 1)
- RF explained greater house price variance than linear regression (Table 1)
- Six variables are important in all three models (Figure 2 highlighted)
- Locational features were more important than structural attributes
- Having a pool only important in Madrid

Figure 2: Ranking of variable importance

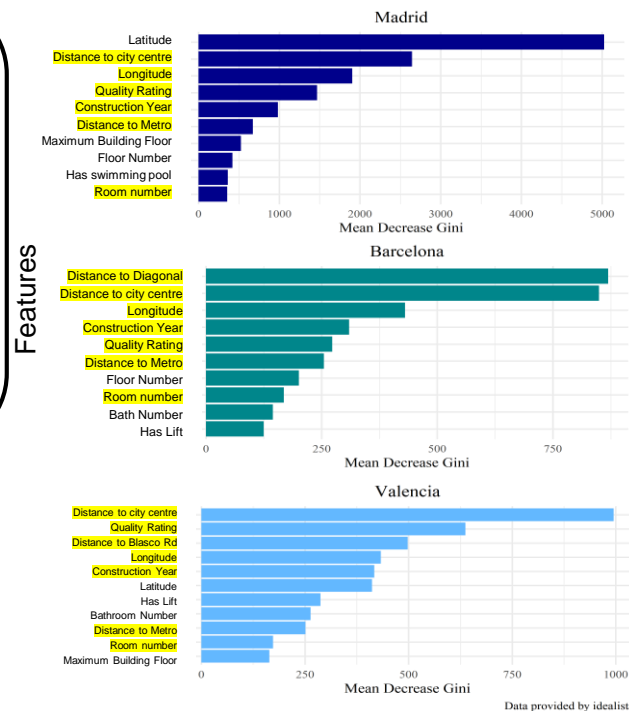
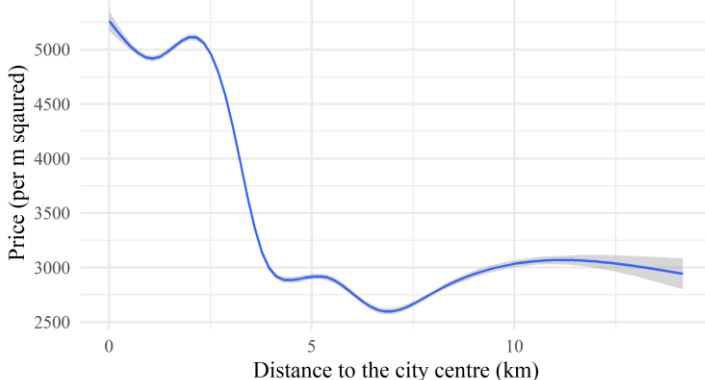


Table 1: Model performance measures

City	MAE	RMSE	POV	R ²
Madrid	1.15	1.22	82.88	56.64
Barcelona	1.13	1.19	67.99	45.37
Valencia	1.17	1.24	77.15	52.98

MAE = Mean Absolute Error, RMSE = Root Mean Square Error, POV and R = % of Variance

Figure 3: Non-linear relationship found in Madrid



Discussion and Conclusion:

- RF is effective for property appraisal as it exhibited low MAE and RMSE on test data
- Able to capture non-linear relationships in housing data (Figure 3)
- But RF can't predict prices outside of the outcome variable range
- Future analyses could include data on the built environment e.g. green space or socio-economic characteristics