A study about French counties and house prices

By Ashish SAUBA

Goal of the project

Where is a “good place” to buy a house in France and at what price?

Good place defined as :

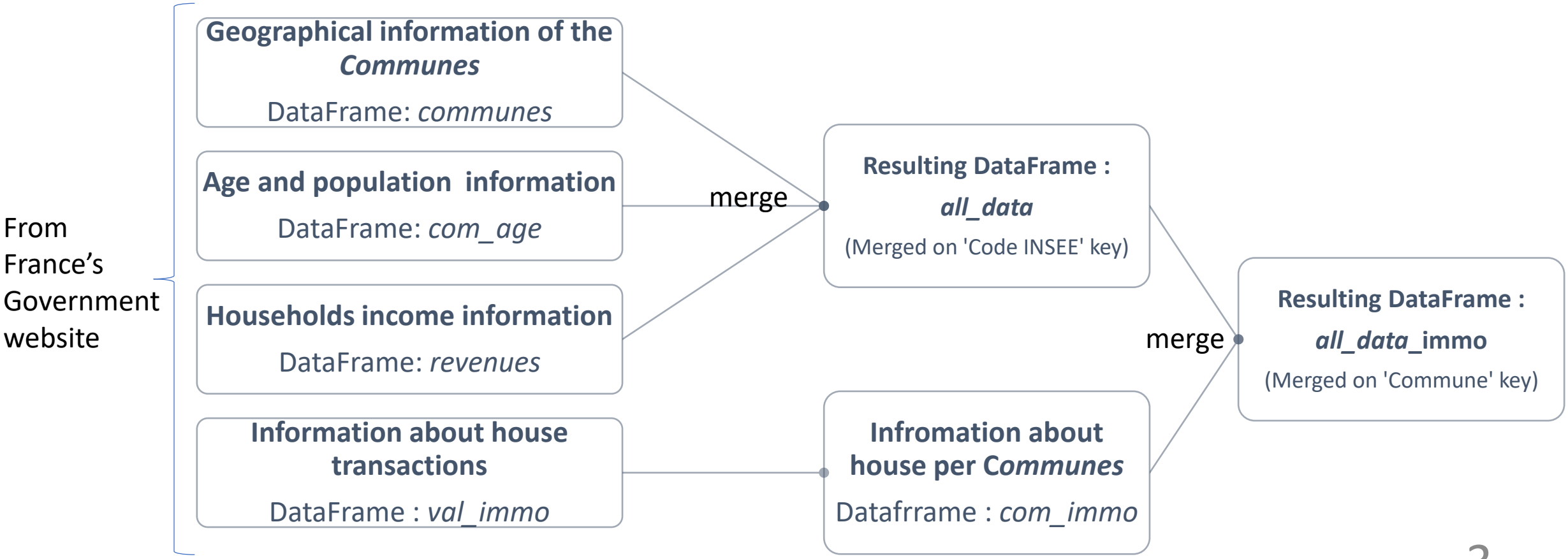
- With a dense enough population
- With a relatively young population
- Where people have a decent income
- With a fair house price
- With a good nightlife

The project audience are :

- People who are willing to relocate and buy a house
- Real estate agencies
- Real estate developers

Data acquisition and cleaning

A quick overview of the data pre-processing steps :



Exploratory data analysis

Information about house transactions

DataFrame : *val_immo*

→We filtered the transaction records and kept only transactions:

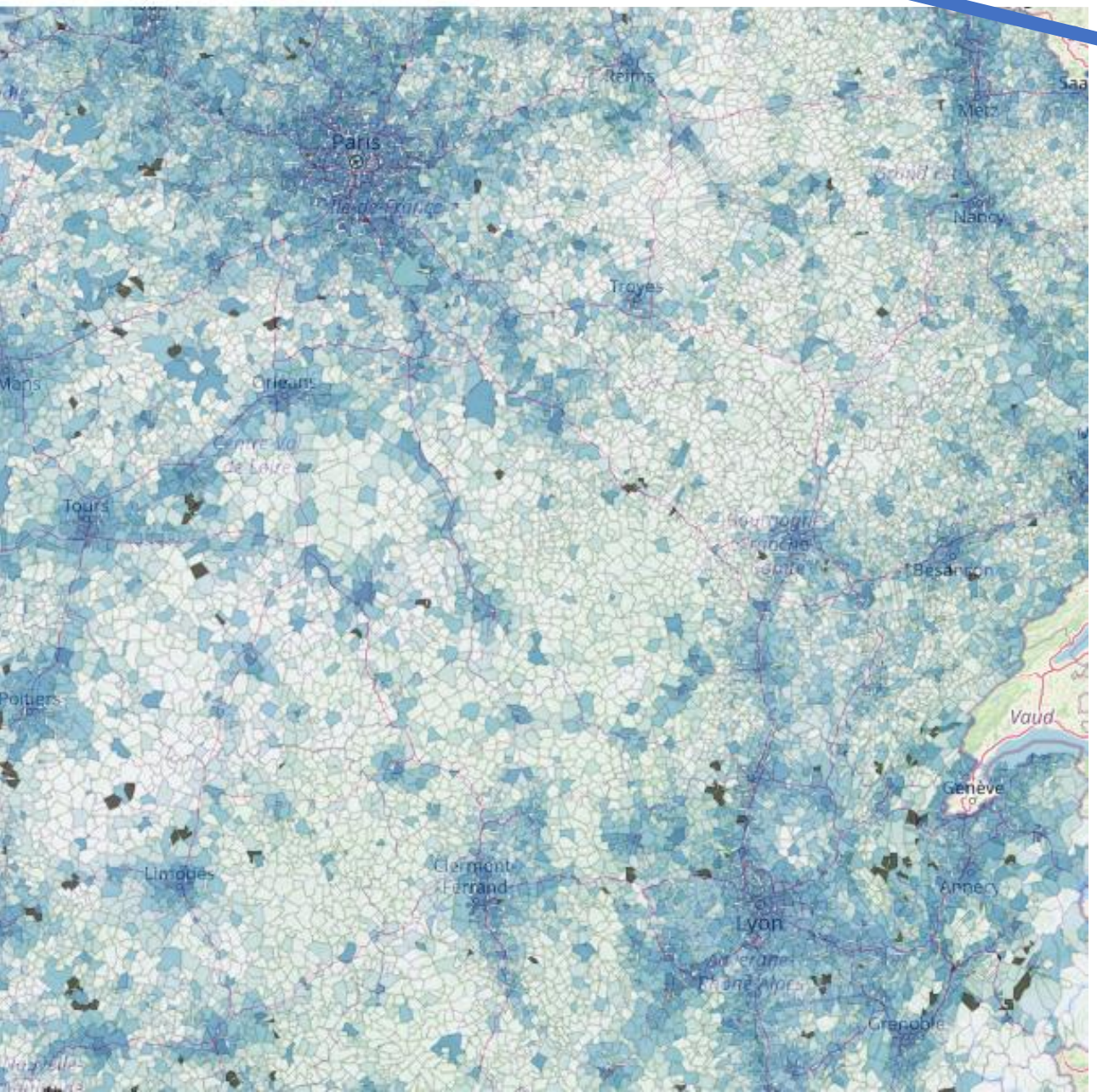
- concerning sales
- concerning houses
  - with no outbuildings
  - and with less or equal to 1000m² of land surface
  - and with less or equal to 200m² of living area
  - and with less or equal to 10 main rooms

Exploratory data analysis

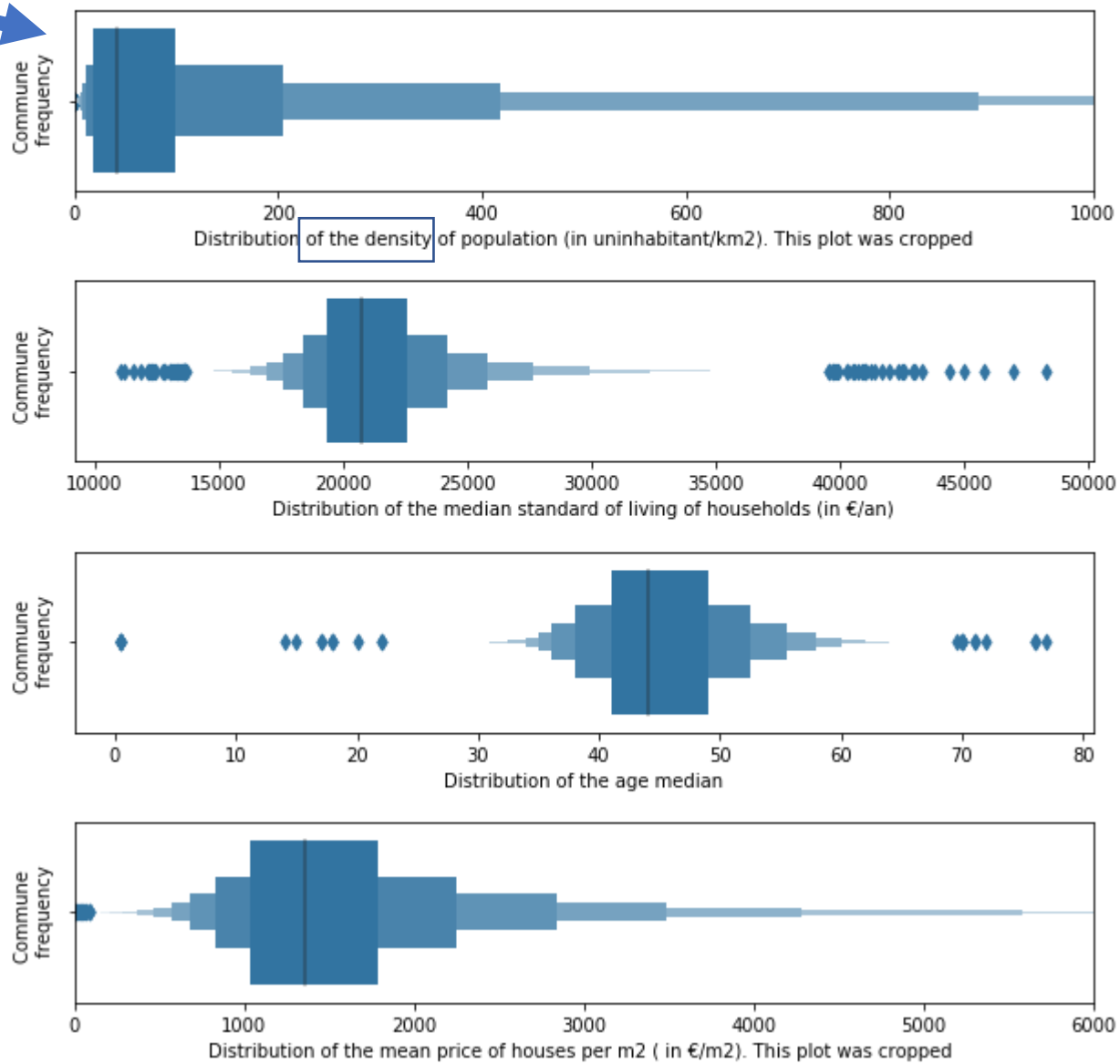
- Interactive maps of the counties' statistics
- Distribution plots of the counties

→ Gave interesting insight about the French territory

Representation of the feature : density



Distribution of the Communes acrosss the studied features



Exploratory data analysis

A quick filtering provides us easy access to “good places” depending on the value of the features we want :

Min density 2000.00

Min revenu… 18000.00

Max age m… 41.00

Max price 2300.00

32 result found

Parameters : filter_result(density= 2000.0 , revenue= 18000.0 , age= 41.0 , price= 2300.0)

	Code INSEE	Code Postal	Commune	Code Département	Code Région	Superficie_km2	lat	lng	age_r
30844	62041	62000	ARRAS	62	31.0	11.93	50.289896	2.765873	38.30
23974	29019	29200	BREST	29	53.0	49.12	48.400500	-4.502791	38.45

Exploratory data analysis

We also used Foursquare's database to inquire about the amount of bars in those places. This gives us a good insight about nightlife.

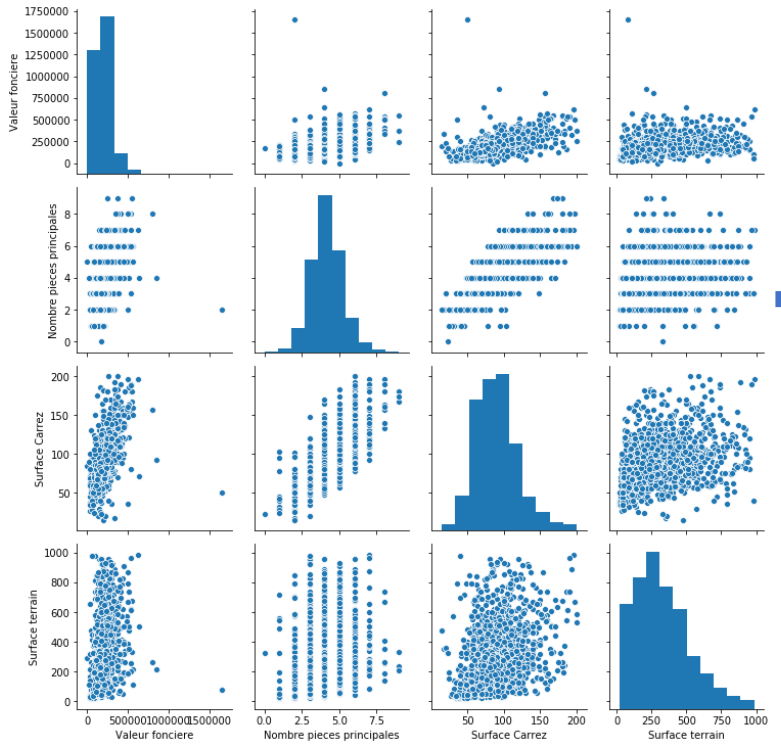
Commune	population	N	nb_venues
CLERMONT-FERRAND	142686.0	1382.0	50
BREST	139342.0	2938.0	48
LE MANS	142991.0	5398.0	35
CROIX	21271.0	1093.0	32
SAINT-MAURICE	14312.0	55.0	28
LOOS	22076.0	817.0	28
SAINT-ANDRE-LES-VERGERS	12116.0	436.0	26
LE HAVRE	170352.0	4106.0	23
POITIERS	87961.0	2087.0	22
WATTRELOS	41341.0	1960.0	20

Therefor, by filtering places :

- With a dense enough population
- With a relatively young population
- Where people have a decent income
- With a fair house price
- With a good nightlife

Clermont-Ferrand is one of the city that stood out!

Exploratory data analysis : In Clermont-Ferrand



A quick analysis with a correlogram shows that the price of the houses is correlated to :

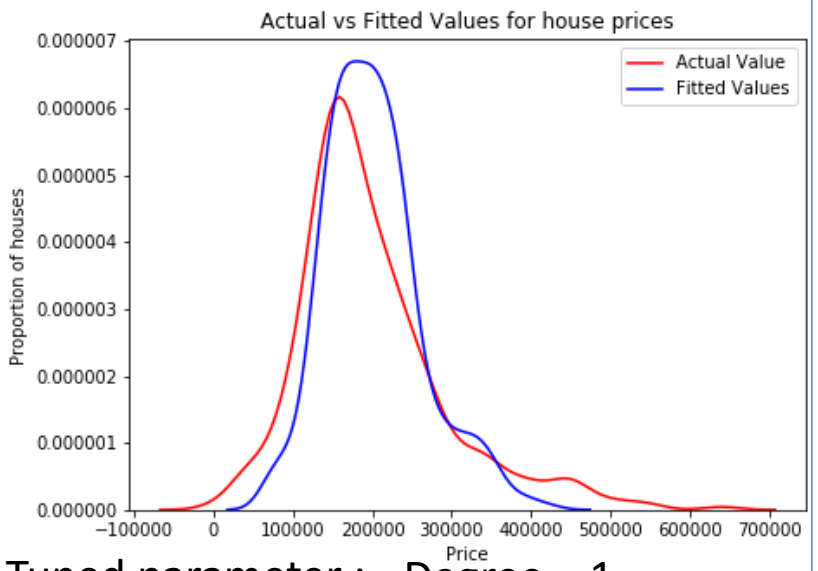
- the number of rooms
- the land area
- the living area

Modelling house prices with regression models

- We tried modelling house prices with different models :

Linear regression model

Polynomial regression model



Tuned parameter : Degree = 1
(so basically
a linear regression model)

Ridge regression model

Tuned parameter : Alpha = 10000

Choosing a model

Here are the scores of the different models :

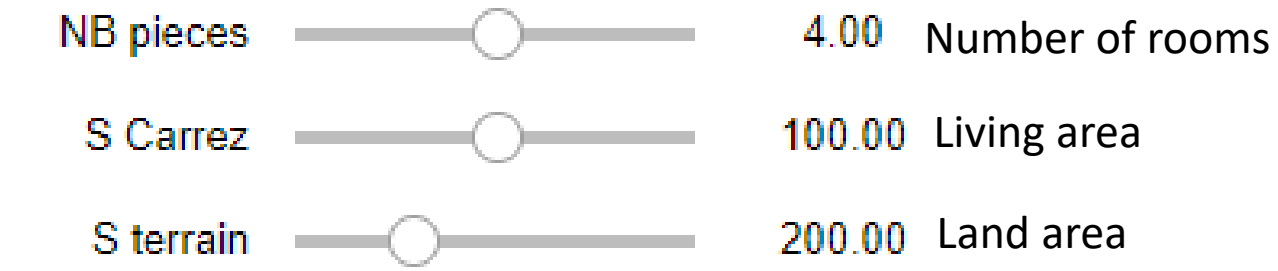
	R2 score	MSE score
Linear regression	0.404699	6.043176e+09
Polynomial regression	0.404699	6.043176e+09
Ridge regression	0.406139	6.028566e+09

→ We can conclude that the difference in performance of different models is not very significant.
→ For the sake of simplicity a **linear regression** model was chosen to estimate house prices.

Note :The polynomial regression model performs better with a degree of 1 which corresponds to a linear model, hence their exact same score

Creating a price estimator with linear regression

A house price estimator was then built buy using the multiple linear regression :



We can adjust the different parameters as needed

Conclusion

Initial question :

Where is a “good place” to buy a house in France and at what price?

This project showed that with freely available data, and basic data science skills we could point out a list of “good places” to buy a house and fairly estimate the price of a specific type house.

However, the R^2 score of the price estimator is modestly low. This score can be easily explained by the fact that the features used to train the model are not enough to explain the price of a property. For example, the age or state of the property, the safety of the borough, the proximity to transit facilities or city centres, and other factors can strongly influence the price of a property. There is room for improvement.