New Taipei City Property Values

An Analysis of Property Value in New Taipei done for a real estate firm

The Data

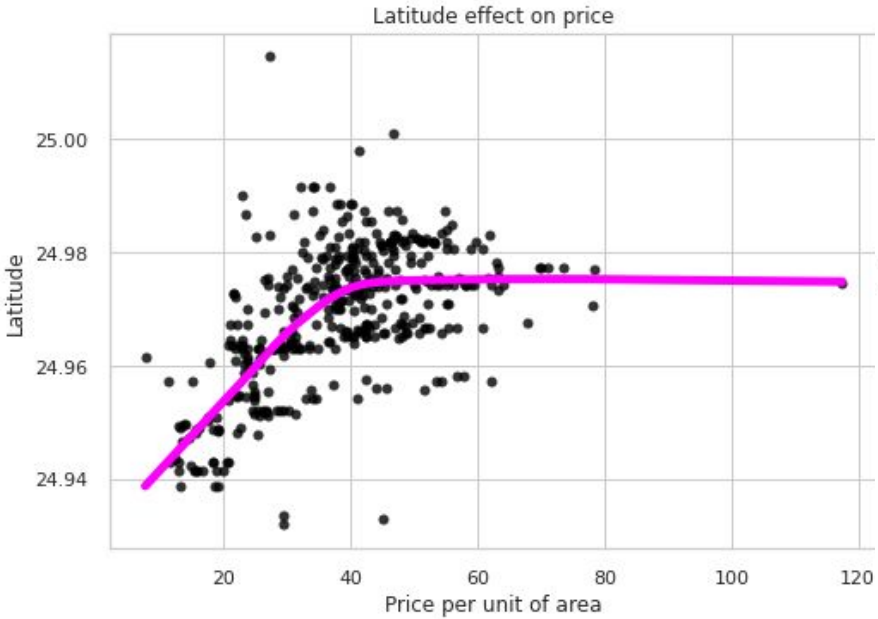
The Data is the market historical real estate valuation collected from Sindian Dist., New Taipei City, Taiwan.

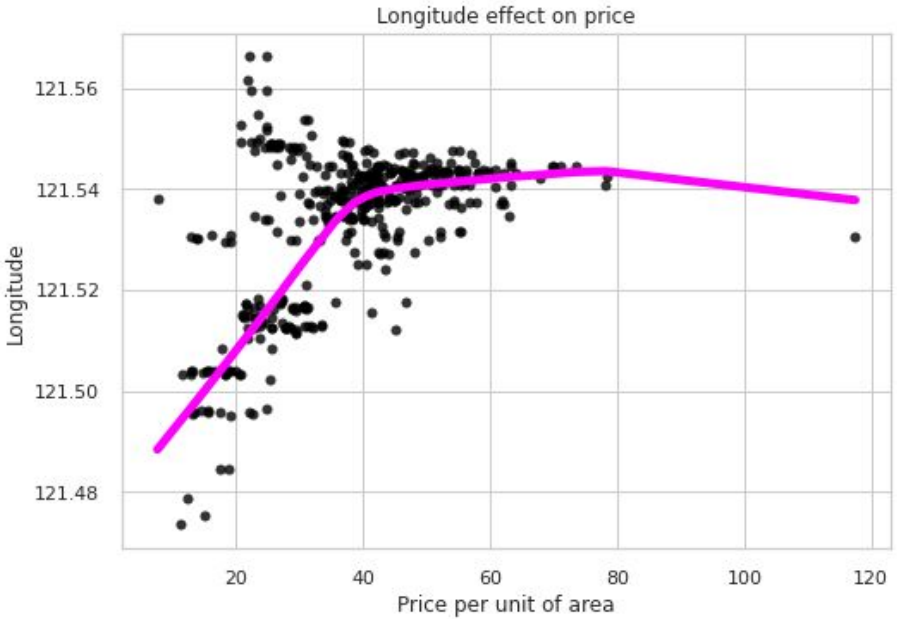
Features: the features in our data set are transaction date by year and month, house age in years, distance to the nearest Mass Rapid Transit station, number of convenience stores nearby, latitude, longitude, and the target variable is price per ping (a ping is $3.3m^2$).

Problem

A real estate investment firm located out of new Taipei City is struggling to predict what properties are undervalued and overvalued. They have provided this dataset hoping to get a model that can predict the value of any given property.

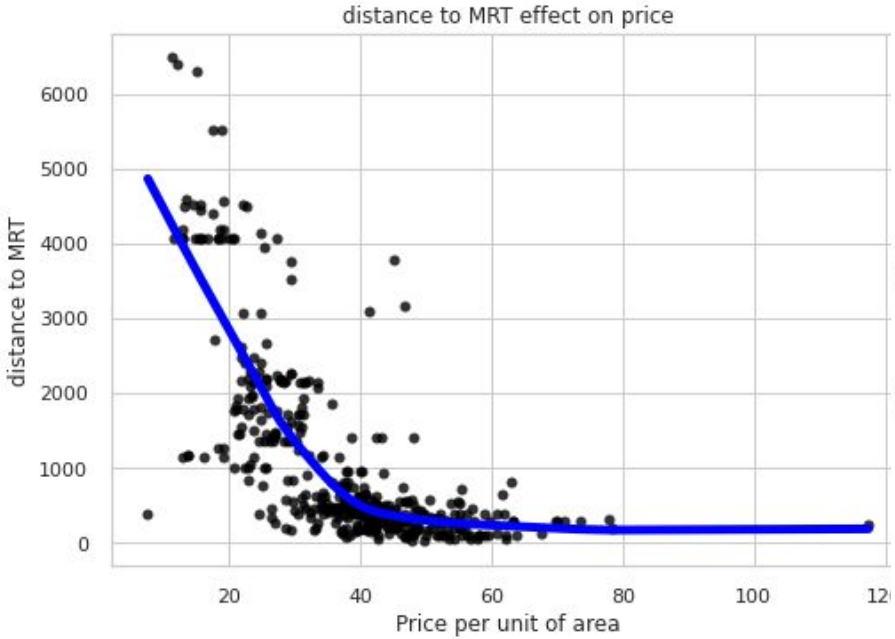
Visuals:

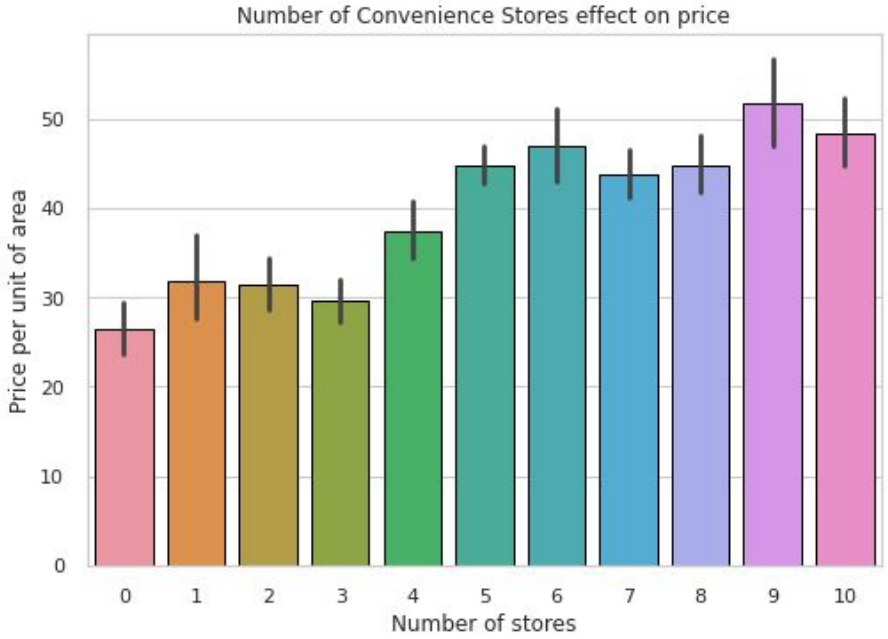




The further North and East, the higher the value up to a point. The value gains level off at that point.

Visuals cont.





Proximity to convenience stores and distance from mass transit stations are also predictors of value.

Strengths

- The model is a bagged forest, it's strengths can be best described by two metrics:
  - **Predictive power:** we have several strong explanatory variables, as shown in the visuals. Location is possibly the biggest one, particularly proximity to coast, which is to the north and east. Being near convenience stores, which are far more flexible businesses in Taiwan, is also a strong explanatory variable, and distance from mass rapid transit are all variables which lend strength to our model
  - **Low Variance**: our model's predictions are on average from the actual figure. **Our mean value per ping is 38.5, and our RMSE is 6.2, meaning that our model's average deviation from the real value is only around 15%.** Essentially, this means the average prediction will be able to predict a home's value within a 15% confidence interval.

Weaknesses

- The most important thing to remember about this model is that **it is limited**. The data is from:
  - One city in Taiwan
  - One year long period

  So as long as the model is applied to this city in Taiwan, we just need to see if the real estate market is comparable to that year.

- This data does not show us home size, just price per ping. Size probably would be a very useful variable. I am willing to bet a useful variable like that would really help us shore up our R^2 value, probably reduce variance even further. If that data can be collected, I would be willing to bet we could get an extremely strong model.
- Because of the 15% variance, If we can't afford the outcome price being 15% lower than predicted, a risk averse investing approach would be avoiding said property.

Implications

There are some very clear patterns that do emerge when looking at this data.

1. The North and East areas of the city are preferable to the south and east, but only up to a point. I would assume this is because it is closer to the coast
2. Similarly, being further from mass rapid transit is an indicator of value, but at a certain point we get no more returns.
3. Finally, proximity to convenience stores indicates value.
4. One more takeaway is that these data points are not the only important data points out there. Our variance is high and R^2 is low enough that we are probably missing a couple key variables.
5. Because these results are based on price per unit of area, this is probably most useful for buying land to put a new building on, or buying a building to demolish and build real estate on.