Predicting House Prices: How Data Science Can Help You Understand the Housing Market?

Tianjiao Jiang

Table of contents

Predicting House Prices: How Data Science Helps You Understand the Housing Market

Buying a house is a big decision to make in everyone's life. It takes a significant amount of money and you might even need to take a loan for which you need to work for several years to repay. Therefore it is important that you make sure the house you intended to buy is worthy! One of the first questions that comes up in your mind when you consider a house is: **How much is this house really worth?**

Well, imagine if there's a way to predict the price for it before you even see the listing. Sounds cool, right? Thanks to the tool in data science called **Ridge Regression**, we can actually make pretty good guesses about house prices using data we already know.

In this blog, we'll take a closer look at how Ridge Regression works and how data scientists use it to predict house prices. Whether you're just curious or starting to dive into the world of data science, we'll keep things simple and fun!

So, What Exactly Is Ridge Regression?

At its core, Ridge Regression is like a smart calculator. It looks at lots of houses, sees what they have in common, like size, number of bedrooms, or whether they have a fireplace, and then uses the information to guess the price of a new house.

Think of it like this: imagine you're trying to guess the price of a toy. You know that bigger toys, like big trucks, tend to worth more money. On the other hand, small toys with less functions are usually cheaper. Ridge Regression works in a similar way for houses. It looks at their features and uses that to predict how much the house would cost.

How Does It Actually Work?

Let's say you've got a collection of toys, and you want to figure out how much each one is worth. You might start to notice patterns—like larger toys being worth more. Ridge Regression works the same way, but with houses.

It takes a bunch of features, like: - **How big the house is** (measured in square meters), - **If there's a garage** or not, and - **If there's a fireplace** or some other special feature.

Then, it looks for patterns in those features to guess how much a house might cost.

What makes Ridge Regression unique is that it's **really good at handling a lot of features at once**, without getting "too caught up" in one specific detail. For example, a large house might cost more, but if it also has a huge backyard or a super nice kitchen, it could be even more expensive. Ridge Regression helps balance all of these factors and come up with a pretty solid guess.

Using our model, the predicted result is as below. As you can see, with basic information for the house listed in the first 5 columns, we are able to predict the housing price. How interesting is that!

	meters	garage	firepl	bsmt	bdevl	Predicted_Values
0	174.23	Y	Y	Y	N	537035
1	132.76	Y	N	Y	Y	457895
2	90.82	Y	N	N	Y	360993
3	68.54	N	N	N	N	212530
4	221.3	Y	Y	Y	Y	700375
5	145.03	N	N	N	Y	445199

The Data: What We Looked At

To test this model, we used a dataset from **Strathcona County**, which contains information about different houses, like:

- **Meters**: The size of the house in square meters.
- **Garage**: Whether or not the house has a garage.
- **Fireplace**: Whether it has a fireplace or not.
- **Building Evaluation**: Whether the house has been recently evaluated.

We looked at how these features affected house prices. According to the figure below, we noticed that larger houses tend to cost more, and houses with a garage or a fireplace are also tend to be more expensive.
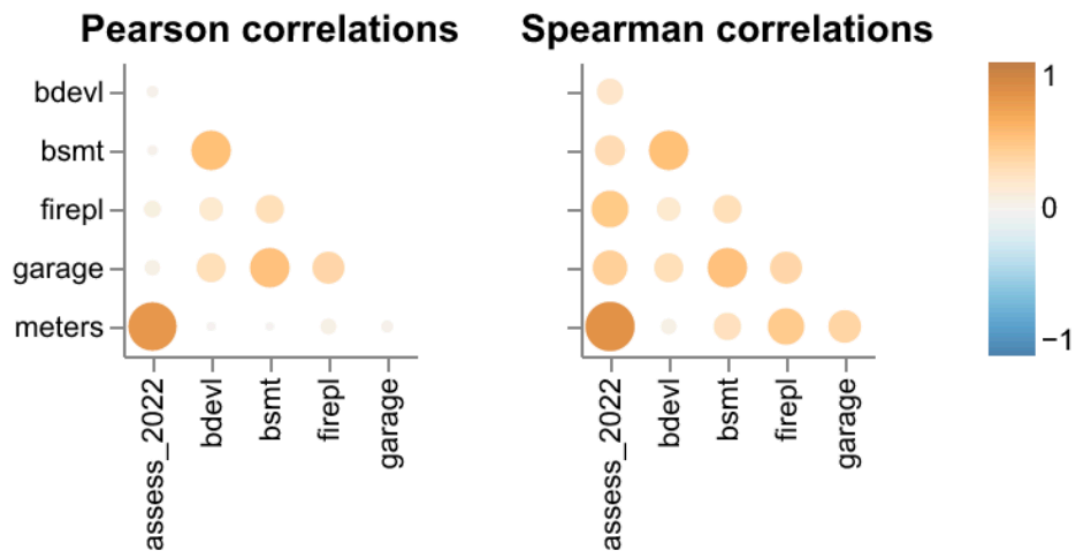
But it's not just about size or amenities; Ridge Regression helps us pull all of this information together to make more accurate predictions about a house's price. Table 1 below shows the first few rows of the dataset.

Table 1: Preview of the 2023 Property Tax Assessment dataset after selecting relevant columns.

	meters	garage	firepl	bsmt	bdevl	assess_2022
0	150.59	Y	Y	Y	N	382460
1	123.56	N	Y	N	N	280370
2	104.98	N	N	N	N	402000
3	66.611	N	N	N	N	3690
4	123.83	Y	Y	Y	Y	295910
5	205.4	N	Y	Y	Y	419000
6	120.68	N	Y	Y	N	289380
7	83.148	N	N	Y	Y	326170
8	120.77	N	N	Y	Y	419000
9	115.94	Y	N	N	N	290000

Why Does This Matter?

You might be wondering: **Why is all of this important?** Well, understanding how to predict house prices with data has a huge impact on the real world:

- **Buyers** can make better, more informed decisions about what they're willing to pay for a house.
- **Sellers** can set a fair price for their homes, making it more likely to sell.
- **Cities and governments** can use this kind of data to make sure homes are affordable and plan for future growth.

And it's not just about houses. The techniques that help us predict housing prices can also be applied to other things like **stock prices**, **project costs**, and even things like **weather forecasts**!

What's Next for Data Science in the Housing Market?

As data science continues to evolve, predictive models like Ridge Regression are getting even smarter. In the future, we'll likely include more data—like the number of bedrooms, the year the house was built, or even how close the house is to schools or parks. The more data we have, the better our predictions become.

The figure below shows how accurate our model is. We can see that the even though model is not perfect, but it's better than nothing! If you do not have any clue about what the price the house would be at, Using Ridge Regression can be a good starting point.

Cross-validation results showing the mean and standard deviation for train and validation scores across 5 folds.

	Unnamed: 0	mean	std
0	fit_time	0.019	0.005
1	score_time	0.004	0.001
2	test_score	0.564	0.235
3	train_score	0.575	0.078

So, whether you're just starting to explore data science or are already diving deep into the field, this is just the beginning. The world of predictions is expanding, and there's so much more to discover.

Key Takeaways

- Predicting house prices is like guessing the value of a toy collection—looking at different features helps you figure out the price.
- **Ridge Regression** is a tool that helps us make predictions by looking at key features of houses and adjusting for how they might relate to each other.
- Predicting house prices isn't just fun. This data science technique can be applied to a whole range of areas, like economics, finance, health care, and even entertainment!