

Coursework One - Question 1 - House Pricing Prediction - Students

November 14, 2021

1 Coursework One - Q1 - House Price Prediction - Student

Please List Members of Team and Student IDs below:

Member 1:

Member 2:

Member 3:

Which member of the team is responsible for this notebook:

1.1 Instructions

This is a dataset containing house prices and associated features.

1. Load the file housing/train.csv. This is the only data you will use.
2. Visualise and analyse the data.
3. Your aim is to predict the SalesPrice column so study its distribution.
4. Examine which features are most correlated with the sales price.
5. Look for missing data and other issues.
6. Visualise the data and look for outliers to remove.
7. Drop any columns that do not seem helpful for predicting.
8. Split the data into a 33% test and 66% training set.
9. Use a linear regression model and compare its training and test performance.
10. Use Scipy or Statamodels to get the p-values and to see which features are important.
11. Repeat the fitting but use a Lasso and Ridge Regression. Discuss your results.