Problem Statement

Once there lived an atrocious King with the finest sword a man could bear at that time.

Alzar, the record keeper, lost papers that had prices for houses in the kingdom.

As he trembled with mortal fear, he went to Elric the sorcerer seeking for help.

"King is very specific and rather precise with numbers!" exclaimed Elric seeing the records.

Fortunately some records were still present, but they were too scattered!

King has commanded Alzar to present to him the complete record with price(in golden grains) of each house against its unique ID.

Now Elric invites you through time travel to help poor Alzar lest he should lose his life to sword.

Alzar will present to you the information that he has.

1) Each paper is specific to one builder family with details of houses that they built.

2) Alzar has sorted for you the house details with builder family name and 'Not Known' where builder's information was lost. "But certainly there are only ten builder families" he remarks.

"Careful! Black Magic has scraped off some more data from the records" says Elric as you begin to think upon…

FORMALLY:

1) .txt files (with builder names including 'Not Known' i.e., total 11 files) contain information about 20k houses describing their different features.
   You need to extract feature information from these files for all the 20k houses.
2) Above mentioned files do NOT contain price information.
3) house_prices.csv file contains information about 16.5k house prices.
4) missing.csv file contains 3.5k house IDs without price information.

You need to use house_prices.csv along with the feature information extracted from .txt files for training and testing your model.

Finally, predict the prices for houses mentioned in missing.csv and generate a file named predicted.csv with price information mentioned against their IDs.
(file similar to house_prices.csv with the same column names and sequence i.e., House ID followed by Golden Grains)

Problem set by Shubham Pandey