

CSE 450 - Machine Learning

Module 03 — Housing Estimates, Class Discussion Questions

[Module 1](#)[Module 2](#)[Module 3](#)[Module 4](#)[Module 5](#)[Module 6](#)[Module 7](#)

Photo by Campaign Creators on Unsplash

1. Questions

You're about to go into a strategy meeting with the CEO, Vice President of Human Resources, and Vice President of Finance. They want to make sure you have the data required to answer the questions they're most interested in.

Be prepared to answer the following questions:

1.1 - PROBLEM TYPE



DEVON, THE CEO SAYS:

“ I just sat through four hours of machine learning training with the board of directors this past week, so I'm curious to get your take on this.

Looking at the data and our business model, what kind of machine learning problem do you think we're looking at here?

1.2 - MODEL CONFIDENCE



CECIL, THE VP OF CUSTOMER RELATIONS ASKS:

“ My biggest concern right now is making sure that whatever method we come up with to predict housing prices, we can also attach some kind of empirical confidence metric.

Based on your initial analysis of the data, your team feels you can best show confidence in your model by using:

- A. The sum of squares error (SSE).
- B. The mean squared error (MSE).
- C. The root mean squared error (RMSE).

D. The R^2 value.



1.3 - INSURANCE QUESTION



WILLIAM, THE VP OF FINANCE ASKS:

“ Our insurance customers are particularly interested in making sure that homes in... unsavory neighborhoods, aren't estimated high.

Is there a way we can easily identify properties in low income areas and have the model lower those estimates to protect our insurance customers' interests?

Based on your initial analysis of the data, your team feels:

- A. We can lower the predicted price for specific neighborhoods before training the model.
- B. We can add average income or other demographic information for the area as features.
- C. For specific zip codes we could add a step to our pipeline that reduces the predicted price by a specific percentage prior to outputting the final price result to ensure the properties aren't being overvalued.
- D. Taking this kind of action would be a violation of federal laws and/or ethics.

1.4 - DATA ANALYSIS



JOHNNY, THE DATA SCIENCE INTERN ASKS:

“ The head of data science says we should use gradient boosted trees for this analysis.



I've noticed that a lot of the features use pretty different ranges.

For example, how should we handle square footage?

-
1. [CEO photo by Oz Seyrek on Unsplash](#) ↵
 2. [VP of Customer Support photo by Christina @ wocintechchat.com](#) ↵
 3. [VP of Finance photo by steffen Wienberg on Unsplash](#) ↵
 4. [Data Science Intern photo by Fábio Lucas on Unsplash](#) ↵