

Capstone Project Proposal

1. What is the problem you want to solve?
2. Who is your client and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn't have otherwise?
3. What data are you going to use for this? How will you acquire this data?
4. In brief, outline your approach to solving this problem (knowing that this might change later).
5. What are your deliverables? Typically, this would include code, along with a paper and/or a slide deck.

- 1) I am going to be predicting house prices based on zip code, and other attributes within each zip code.
- 2) There are multiple clients that would benefit from this analysis. First, this would help anyone that would be potentially in the housing market in the upcoming months/years. There is a need to understand whether or not a house should be purchased when moving to a new city or as a possible income property. There are many other consumer related decisions that can be made as a result of this data. In addition, there is also commercial/enterprisal usage. This data could influence the decision for a company to have headquarters in a certain city or not, as the salary they would need to pay to their employees would need to be higher if they picked a high COL city to headquarter in. In addition, this could provide valuable data to realtors. If it is evident from data that a certain area seems to be on an upward trend of prices, it would be a good business decision for the realty company to focus efforts of marketing and perhaps hire or station more real estate agents in that area, leading to larger commissions for the agents and more income for the company.
- 3) Data based by zip code is available on realtor.com, Zillow, and other real estate websites. There is also valuable housing data at the Federal Housing Finance Agency as well as the US Department of Housing and Urban Development. This data is all free to download.
- 4) After acquiring the data, the next logical steps would include taking a look at the data, merging sets if necessary to make the analysis easier, cleaning the data, doing some visualization on the available data without any predictions, then calculating the predictions by doing TSA or other forms of data analysis, or a ML algorithm using the existing data as a training set, and use predictions from there.
- 5) Deliverables in this would be code, a slide deck, as well as a write up based on what I did including the visualizations I completed. If possible, this could also include a front end UI that would allow the client to input certain factors and return a predicted price.