Project 1
Milestone 1 – Proposal
Laura Hoffmann
DSC680

Topic
Analyze the housing markets in Australia to gain a better understanding of pricing, and possibly explore building a model for price prediction.

Business Problem
The housing market has become a turbulent market to navigate with a lot of unpredictable factors about our future, especially after the spread of a global pandemic. My sister moved to Australia in 2018 and is now engaged to an Australian and they are perusing the housing market in such a questionable and uncertain current world climate. I would like to find a dataset for the housing market in Australia that would aid in the analyzing and exploration of the market so I could better help my sister feel more comfortable in her house purchasing endeavors.

Datasets
https://www.kaggle.com/anthonypino/melbourne-housing-market?select=Melbourne_housing_FULL.csv (Melbourne - Full)

https://www.kaggle.com/anthonypino/melbourne-housing-market?select=MELBOURNE_HOUSE_PRICES_LESS.csv (Melbourne - Some columns removed)

https://www.kaggle.com/datasets/mihirhalai/sydney-house-prices?select=SydneyHousePrices.csv (Sydney)

Methods
Cleaning the data: Deciding whether to remove rows that have null values for analysis, or if certain variables have a heavy effect on price, decide if I need to fill those values with mean or median or another value.

Visualizing the data: use the Plotly package for interactivity, determine if there are outliers by visualizing distributions.

Modeling the data: run a few different models to determine which might perform better given the data and then hyper-tune whichever originally performs the best.

Ethical Considerations
I can't think of a lot of ethical considerations when exploring a housing market because it's not necessarily a business-related endeavor. I do however need to recognize that there might be a difference in the current market because the data is a few years old.

Challenges

One challenge might be in understanding how Covid might've affected the housing market. Being able to properly deal with missing data or outlying data might be another challenge that could be discovered in the exploratory process.

References

Exploratory Analysis:

https://www.kaggle.com/code/alexgeiger/insightful-vast-usa-statistics-eda-efa/notebook

https://www.kaggle.com/code/kabure/financial-hedging-eda-plotly-charts/notebook

https://www.kaggle.com/code/janiobachmann/melbourne-comprehensive-housing-market-analysis/notebook

https://www.kaggle.com/code/furkankarakuz/sydney-house-prices-multiple-linear-regression

Handling missing values:

https://www.kaggle.com/code/dansbecker/handling-missing-values/notebook

Handling outliers:

https://www.kaggle.com/code/nareshbhat/outlier-the-silent-killer/notebook

Model Building:

https://www.kaggle.com/code/emanueleamcappella/random-forest-hyperparameters-tuning/notebook