In this study, we experiment on using a Bayesian hierarchical approach to identify submarkets of the Denver real estate market. The goal is to explore whether these submarket classifications, as a combination of geographical, physical and socio-environmental factors, will add value to the prediction of demand for residential housing in the Denver area.

We used an existing dataset from the Fall 2020 IACS Capstone Team. More specifically, our data is a combination of MLS listing data, REX data, and various sources of census-tract data. Overall, we evaluate the quality of the dataset by looking at the data distribution to remove outliers and using preprocessing scripts to remove suspicious observation points. Most of the feature selection are carried out automatically to avoid biases. The quality of the data does come into question when we detect that some features have conflicting values with one another. Nevertheless, without other data sources to compare with or any verification service, it is difficult to identify and correct these values.

Our Bayesian models produce slightly better results for the testing set compared to our baseline models. The aggregated metrics also show that the models can make reasonable predictions on the total numbers of houses that will be sold during a certain time period, which is beneficial for reducing information asymmetries amongst real estate investors and property agents. The generality of the results, however, require more testing. Nonetheless, our project produces some insights on a new strategy employed in the real estate market to predict metrics that are rarely focused and hard to quantify such as demand.