Emily Phillips
Dr. Catherine Williams
DSC 680 – Applied Data Science
4/18/2022

The Booming House Market – What Makes an Area Hot for Buying?

The housing market is in a constant flux of change and with it being at a positive progression recently, I want to investigate what makes metro areas in the United States "hotter" than others. Realter.com assigns a Market Hotness Index for metro geographical areas based on various factors; using this target variable, my project will be focused on identifying what signifies the hotness of housing markets and predicting it.

The research questions I am going to explore are:

1. Which regions of the United States have the hottest housing markets?
2. Is there a cutoff for the market hotness index where a significant difference between markets becomes apparent?
3. Does the allure of the certain cities (favorable place to live) play into the cities' assigned market hotness index?
4. Can we predict the market hotness index of cities on a monthly/yearly basis?

I am getting my data from two different sources: Realtor.com and the Simplemaps: World Cities Database. For Realtor.com, I am going to utilize two different datasets: one pertains to monthly data for metro market trends and statistics on active for-sale listings, and the other dataset contains information on the assigned Market Hotness Index for given metro areas which is based on days on market (supply index) and realtor.com views per property (demand index). The second source of data, Simplemaps: World Cities Database, will provide data on U.S. cities regarding their geographical location and population count. This data will be utilized to take metro area sizing into account and help determine if it has any impact on market hotness.

Emily Phillips
Dr. Catherine Williams
DSC 680 – Applied Data Science
4/18/2022

I will be employing various forms of exploratory data analysis (EDA) and modeling techniques for this project. The EDA portion will use data visualization methods such as bar charts, line charts to track timely trends, and geospatial charts to depict geographical changes. For modeling methods, the focus will be on predicting the market hotness index for a given city in a given month or year, so I will be utilizing the Market Hotness Index as my target variable. Since it is of a numeric type, I'll be using regression modeling techniques to model my data and yield predictions from it. These techniques could include neural networks, linear regression, time-series analysis, etc.

I think the mobility of people moving around America plays a huge role into where those people choose to settle and buy homes to build their lives. An ethical concern with looking into this type of data is that the location choice is often very subjective, and this subjectivity is very difficult to capture in a dataset unless it contains survey data.

I do not have very much experience with modeling time-series data, which could be a challenge for this project given that the Realtor.com datasets contain monthly data. However, it will be a great learning experience for me! Many numeric variables also exist in the datasets, and they will need to be standardized to ensure that no issues arise with modeling variables of different measurement types.

References

1. https://realestate.usnews.com/real-estate/articles/how-to-decide-where-to-live
2. https://realestate.usnews.com/places/methodology
3. https://www.bls.gov/data/#maps
4. https://www.realtor.com/research/data/
5. https://simplemaps.com/data/us-cities
6. https://fortune.com/2022/04/16/real-estate-for-sale-cheapest-most-expensive-housing-markets/