678 Proposal

Yang Xiao

November 8, 2023

Personal Statement

I want to engage in research and work related to actuarial and data processing in the future, analyze problems from big data, and obtain meaningful conclusions and viewpoints. This project, centered on analyzing the Melbourne Housing Market, serves as a crucial step in this direction. It provides an opportunity to enhance my capabilities in data analysis and pattern recognition within the real estate sector, And this dataset has a relatively large amount of data, which is a challenge for me and can also increase my experience in analyzing big data, which is consistent with my career goals.

Research Question

The preliminary question I aim to answer is: How do various property characteristics and geographic factors influence the prices of houses across different suburbs in the Melbourne Housing Market?

Research Ideas and Methods

In order to study how the real estate characteristics and geographical factors in Melbourne affect housing prices, I will conduct exploratory data analysis. Firstly, the data will be preprocessed to remove invalid and unnecessary data. Then, the data will be processed according to our research objectives and the model we want to establish, and a multi-level model will be established. This model will consider changes at the real estate and suburban levels. Subsequently, model validation and improvement will be carried out to ensure accuracy.

Data Source

The dataset for this study is sourced from Kaggle (`https://www.kaggle.com/dansbecker/melbourne-housing-sn`
a well-known platform for data science datasets.

Proposed Timeline

- **Exploratory Data Analysis (EDA):** Completion by November 16, 2023.
- **Data Processing:** Completion November 23, 2023.
- **Modeling and Validation:** Completion by December 1, 2024.
- **Final Write-up:** Completion by December 7, 2024.

References

[1] Hox, Joop. *Multilevel modeling: When and why.* In Classification, data analysis, and data highways: proceedings of the 21st Annual Conference of the Gesellschaft für Klassifikation eV, University of Potsdam, March 12–14, 1997 (pp. 147-154). Springer, 1998.

[2] Dedrick, Robert F and Ferron, John M and Hess, Melinda R and others. *Multilevel modeling: A review of methodological issues and applications.* Review of educational research, 79(1):69–102, Sage Publications Sage CA: Los Angeles, CA, 2009.