MA5851 Assigment 3, Document One

1 Background of the Problem

According to ABS (2021), housing continues to be Australia's largest asset class, worth approximately $7.7 trillion, and yet, the study of real estate sentiment in Australia has been profoundly under-researched to date (Nanda & Heinig, 2018). Conversely, current literature reviews have established a strong connection between investor sentiment and stock returns in general, which in turn would suggest that there is also some degree of correlation between real estate sentiment and the strength of the property market as well (Bisen, 2020; Bharathi & Geetha, 2017; Corredor et al., 2011). In view of these findings, it was clear that there existed an opportunity to exploit this underutilisation in real estate sentiment within Australia to gain a competitive edge in estimating where and when to invest in housing in Australia.

As such, this Project aims to autonomously estimate the relative strengths of the property market across all states and territories in Australia, as a function of time, by relying on trusted Australian news sources.?

2 The Proposed Solution

The first component in the data ingestion architecture involved identifying the news article addresses of interest from all pages contained within the trusted news source websites. The decision was made to include news articles from all available pages on the trusted news source websites in order to afford more power to the models that formed part of the solution.

These website addresses were collected using the second component of the solution, which consisted of a web crawler powered by the *Selenium* package. Once the web crawler had extracted all of the news article website addresses from the trusted news sources, the web crawler would then also extract the title, publication date, body content, and any other relevant data, for each of the website addresses collected previously.

The third component involved the natural language processing (NLP) pipeline. The first NLP task in this pipeline involved the implementation of Named Entity Recognition (NER) to identify all locations (LOCs) and geopolitical entities (GPEs) from the body text of each news article. This LOC and GPE data would then be rolled up to the state and territory level using the *Nominatim* geocoding software in order to achieve consistent data granularity for comparison purposes. The second NLP task involved testing a variety of traditional machine learning methods to perform sentiment analysis on sentences of the news article body text containing the identified LOCs and GPEs. Subsequent to the validation process, the best sentiment predictor would be selected for application.

The output from the NLP pipeline served as scoring weights to the frequency of published news articles per state or territory. For example, if all news article mentions of Queensland for a given week were negative and all mentions of New South Wales were positive, this would leave a -1 score for Queensland and a +1 score for New South Wales for that week. The relative scoring across all states and territories over time provided an estimate of the relative strength of the property market across all states and territories in Australia. These scores were then plotted as a function of time, however, an online dashboard of the results was not produced and instead kept as a future recommendation.

Publications and communications of project developments for this Project were executed using Github.

3 References

ABS (2021). Residential Property Price Indexes: Eight Capital Cities. Retrieved from: https://www.abs.gov.au/statistics/economy/price-indexes-and-inflation/residential-property-price-indexes-eight-capital-cities/latest-release#:~:text=last%20twelve%20months.-,Total%20value%20of%20the%20dwelling%20stock,rose%20by%2044%2C100%20to%2010%2C602%2C700.

Bisen, V. (2020). How Sentiment Analysis in Stock Market Used for Right Prediction? Retrieved from: https://medium.com/vsinghbisen/how-sentiment-analysis-in-stock -market-used-for-right-prediction-5c1bfe64c233

Bharathi, S. & Geetha, A. (2017). Sentiment Analysis for Effective Stock Market Prediction. Retrieved from: https://www.researchgate.net/publication/317214679_Sentiment_Analysis_for_Effective_Stock _Market_Prediction

Corredor, P., Ferrer, E. & Santamaria, R. (2011). Investor Sentiment Effect in Stock Markets: Stock Characteristics or Country-Specific Factors? Retrieved from: https://www.researchgate.net/publication/228318309_Investor_Sentiment_Effect_in Stock_Markets_Stock_Characteristics_or_Country-Specific_Factors

Nanda, A & Heinig, S. (2018). Measuring sentiment in real estate - a comparison study. Retrieved from: http://centaur.reading.ac.uk/72893/1/170930_Revised_Manuscript.pdf

[ ]: