

Sentiment Analysis, Social Media and Urban Economics: The Case of Singaporean HDB and Covid-19

Srinaath Anbu Durai¹, Wang Zhaoxia²

¹ Department of Information Systems and Operations Management, HEC Paris, Paris, France

² School of Computing and Information Systems, Singapore Management University, Singapore

Abstract: Twitter sentiment analysis has been employed as a prognostic tool for predicting prices and trends in both stock and housing markets. Early studies in this domain drew inspiration from behavioural economics, establishing a link between sentiments or emotions and economic decision-making. However, recent investigations in this field have shifted their focus from the data utilized to the algorithms employed. A comprehensive literature review, with an emphasis on the data aspect, reveals a scarcity of research considering the influence of sentiments arising from external factors on stock or housing markets, despite abundant evidence in behavioural economics suggesting that sentiments induced by external factors impact economic decisions. To bridge this gap, this study explores the impact of Twitter sentiment related to the Covid-19 pandemic on housing prices in Singapore. Employing SNSCRAPE for tweet collection, sentiment analysis is conducted using VADER. Granger Causality is applied to investigate the relationship between Covid-19 cases and sentiment, while neural networks serve as prediction models. The research compares the predictive capacity of Twitter sentiment regarding Covid-19 with traditional housing price predictors, such as structural and neighbourhood characteristics. Findings indicate that utilizing Twitter sentiment related to Covid-19 yields superior predictions compared to relying solely on traditional predictors, outperforming two specific traditional predictors. Consequently, this study underscores the significance of considering Twitter sentiment related to external factors as crucial in economic predictions, demonstrating practical applications of sentiment analysis on Twitter data in real-world economic scenarios.

Keywords: Covid-19, Housing prices, Sentiment analysis, Twitter, Singapore

1. Introduction

Housing ownership plays a crucial role in economic well-being, as emphasized by Seiler et al. (2020). In Singapore, where over 80% of the population resides in Housing Development Board (HDB) Flats (Joo and Wong, 2008), the pricing dynamics of these residences hold national significance. New HDB Flat prices are regulated by the government, while resale HDB Flat prices are subject to market forces (Phang and Wong, 1997).

The global economic downturn triggered by the emergence of the Covid-19 virus in December 2019 impacted major economies worldwide by September 2020, including Singapore (Ryan, 2020). By April 2020, the lives of Singapore residents underwent substantial changes due to the pandemic (Lim et al., 2021).

Sentiment analysis, a method employing natural language processing techniques to discern subjective information in text (Wang et al., 2020), utilizes tools like VADER (Hutto and Gilbert, 2014) to derive sentiment scores from social media text data, such as Twitter. Described by Wojcik and Hughes (2019) as a "modern public square," Twitter serves as a platform for diverse

discussions and viewpoints. Given the profound impact of Covid-19 on daily life and the economy, quantifying public sentiment on Twitter provides an avenue to explore its influence on economic decisions.

While existing literature has employed Twitter sentiment to predict or analyze price trends in various markets, including the housing market (Velthorst and Güven, 2019) and stock market (Bollen and Mao, 2011) and much less attention has been paid to the impact of Twitter sentiment related to external factors such as Covid-19 on housing prices. Thus, this study aims to investigate the impact of Covid-19 on resale prices of HDB Flats in Singapore.

Contrary to previous studies, which often focused on algorithmic approaches rather than the data collected, this research recognizes the critical role of data in shaping model outcomes (Redman, 2018). Consequently, an examination of the data collection methods becomes imperative.

This study utilizes tweets gathered during the Covid-19 period for sentiment analysis, aiming to compare the performance of Covid-19 sentiment against traditional factors in influencing HDB Flat resale prices in Singapore.

This paper makes several noteworthy contributions:

1. It conducts a comprehensive survey of existing literature that utilizes social media data for predicting prices or examining price trends. Notably, the paper classifies this literature based on the nature of the data employed.
2. The paper identifies a significant gap in the existing body of work, revealing a scarcity of research that investigates the impact of sentiments, particularly those related to external factors like Covid-19, on housing prices.
3. To address the identified gap, the study specifically probes into the influence of Covid-19 on housing prices in Singapore. Twitter sentiment is employed as a key variable, and its impact is compared with that of other traditional predictors.
4. The research findings indicate that Twitter sentiments associated with Covid-19 emerge as a crucial informational source for predicting housing prices, particularly in comparison to traditional predictors, amidst the pandemic.
5. Notably, this paper stands as the pioneering research endeavor focusing on the interplay between housing prices and Covid-19 in the context of Singapore. To the best of our knowledge, no prior studies have delved into this specific intersection.

2. Literature Review

Human decision-making processes exhibit a degree of irrationality, with emotions playing a pivotal role in economic decisions, as noted by Nofsinger (2005). Existing literature highlights the influence of sentiments driven by various external factors, such as sports outcomes (Edmans et al., 2007) and temperature during trading days (Cao and Wei, 2005), on stock market prices. However, these studies faced challenges in quantifying public sentiment related to specific phenomena, resorting to a "mood variable" as a proxy (Edmans et al., 2007), as exemplified by Hirshleifer and Shumway (2003) using cloud cover on trading days as such a variable.

The universal applicability of the mood variable method is constrained, as the impact of weather conditions, like more sunshine, may elicit positive sentiment in New York but have the opposite effect in Singapore. Leveraging sentiment analysis on Twitter data provides a viable approach to quantifying public sentiment toward events or phenomena and studying their consequential impact on the economic decisions of the public. This approach was pioneered by Bollen and Mao (2011), who utilized Twitter sentiment to predict stock prices and establishing a connection between behavioral economics and natural language processing by drawing upon the works of Hirshleifer and Shumway (2003) and Nofsinger (2005). Subsequently, a plethora of similar research endeavors has emerged in this domain.

Scholarly works aligning with this trajectory can be categorized into three groups, delineated by the criteria employed for the collection of tweets intended for sentiment analysis:

1. Generic collection
2. Topic-based collection
3. Object-based collection

Within the generic tweet collection approach, tweets are amassed without specific criteria, relying solely on the expression of emotions. For instance, Bollen and Mao (2011) employed this method by gathering tweets that "contain explicit statements of their author's mood states," encompassing expressions such as "i feel", "i am feeling", "i'm feeling", "i don't feel", "I'm", "Im", "I am", and "makes me". In contrast, in topic-based tweet collection, keywords relevant to the subject matter guide the tweet selection process. As illustrated by Ozturk and Ciftci (2014), this approach involves collecting tweets containing specified terms such as "USD/TRY", "#USD/TRY", "Dollar", and "#Dollar". Lastly, object-based tweet collection relies on keywords associated with the object of interest, typically companies. Ruan et al. (2018) is an example of this approach, where tweets containing terms like "\$AAPL", "\$FB", "\$GOOG", "\$NFLX", "\$AMZN", "\$GE", "\$MSFT", and "\$GILD" were gathered.

However, studies such as those conducted by Hirshleifer and Shumway (2003) and Edmans et al. (2007) provide evidence that sentiment influenced by external factors has an impact on stock market prices. Notably, the works do not specifically investigate the influence of sentiment related to an external factor on the stock market. In contrast, Ilyas et al. (2020) examined the impact of Twitter sentiment regarding Brexit on the FTSE 100 Index, and Kinyua et al. (2021) explored the effects of the sentiment expressed in President Trump's tweets on the S&P 100 and DJIA indices, serving as instances where such investigations were conducted. The literature suggests that emotions play a role in influencing decisions related to house purchases (Jørgensen, 2016). Although this aspect has not been explicitly explored, similar methods of tweet collection have been employed for forecasting prices and examining price trends in the housing market. Tan and Guan (2021), for instance, engaged in generic tweet collection by gathering tweets within the geographic confines of the United States to investigate their influence on housing prices nationwide and within the specific geographic location of Manhattan to understand their impact on housing prices in that area. Another example is presented by Velthorst and Güven (2019), who adopted a topic-based tweet collection approach, collecting tweets containing Dutch terms like "woningmarkt" and "huizenmarkt" (both signifying "housing market") and "huizenprijzen" (indicating "housing prices") to forecast market trends in the housing market of Netherlands. In a different context, Hannum et al. (2019) utilized an object-based tweet collection method, amassing tweets containing the names of at least one of the 39 districts of Istanbul to scrutinize their impact on housing prices within those specific districts.

Certain investigations leverage sentiment analysis on data from alternative social media platforms such as Weibo (Deng et al., 2018) and financial news sources (Hu et al., 2021) to anticipate stock market prices. Instances also exist where sentiment analysis of Weibo data (Li et al., 2022) and news articles (Hausler et al., 2018) has been employed for forecasting housing prices. To the extent of our knowledge, no research has employed sentiment analysis of social media data, particularly concerning an external factor like Covid-19, to predict prices or study trends in the housing market.

The global impact of the Covid-19 pandemic has reverberated across housing markets worldwide (Yiu, 2021). Utilizing Twitter data and sentiment analysis provides an avenue to quantify public sentiment towards Covid-19 (Ridhwan and Hargreaves, 2021). Consequently, it becomes feasible to investigate the repercussions of Twitter sentiment on resale HDB prices in Singapore. To the best of our knowledge, no study has explored the impact of Covid-19 on the housing market in Singapore.

3. Data Collection

The HDB resale data and the HDB resale price index data have been sourced from data.gov.sg and meticulously filtered for the time span of January 2018 to February 2022, encompassing the prominent phase of the Covid-19 pandemic in Singapore. The resale prices have been adjusted to Quarter I of 2018. Concatenating block and street names yielded the addresses, and the age of each property is computed based on the lease commencement date. MRT station details are obtained from mrtmapsingapore.com, bus interchange data is sourced from landtransportguru.net, primary school information is extracted from moe.gov.sg, and hawker centre details are gathered from sgclean.gov.sg, with corresponding addresses retrieved from streetdirectory.com. Utilizing the addresses and the OneMapAPI (One Map API, 2018), longitude and latitude coordinates are computed. Various proximity metrics, including Proximity to the Nearest MRT Station, Proximity to the Nearest Bus Interchange, Proximity to the Nearest Primary School, Proximity to the Central Business District (CBD), and Proximity to the Nearest Hawker Centre, are calculated for each Flat using the GeoPy library (Esmukov, 2021). The dataset comprises a total of 100,506 resale transactions.

Further segmentation of the dataset is performed into two distinct periods: Pre-Covid, spanning from January 2018 to December 2019, representing the timeframe before the onset of the Covid-19 pandemic, and Covid, spanning from May 2020 to February 2022. The number of resale transactions in the Pre-Covid dataset and the Covid dataset is 43,743 and 50,830, respectively.

Tweets were systematically gathered utilizing the Python library SNSCRAPE (Snsrape, 2022) within the timeframe from 15th February 2020 to 15th February 2021, focusing on the geographical location of Singapore, utilizing the English language, and employing specific keywords including 'Covid-19', 'COVID', 'coronavirus', '2019nCoV', 'nCoV2019', 'SARS-COV-2', and 'circuit breaker' akin to the methodology outlined in Ridhwan and Hargreaves (2021). Duplicated tweets were eliminated based on Tweet ID and @ mentions, retweet symbols, hyperlinks, and special characters were removed.

Subsequently, sentiment scores were computed for each tweet through the utilization of VADER (Hutto and Gilbert, 2014). These sentiment scores were then leveraged to calculate daily average sentiment scores. Data pertaining to daily Covid-19 case numbers, spanning from February 15th, 2020, to February 15th, 2022, was acquired from moh.gov.sg.

4. Covid-19 Cases and Covid-19 Twitter Sentiment

The impact of Covid-19 in Singapore has been overwhelmingly negative, leading to widespread public panic (Ho et al., 2020). However, an analysis of tweets using the VADER sentiment analysis tool reveals an interesting pattern. Contrary to the prevailing negative sentiment, the number of positive tweets related to Covid-19, as illustrated in Fig. 1, surpasses the count of negative tweets, aligning with the findings of Ridhwan and Hargreaves (2021).

This apparent contradiction may be attributed to the inherent limitations of VADER, a lexicon-based sentiment analysis tool (Hutto and Gilbert, 2014). Notably, VADER faces constraints due to its lexicon, rendering it incapable of handling newly coined words or neologisms (Kannan et al., 2016). To address this limitation, we augmented VADER's lexicon by incorporating relevant keywords used for tweet collection, such as 'COVID-19', 'Covid-19', 'Covid19', 'covid', 'coronavirus', '2019nCoV', 'SARS-COV-2', 'circuit breaker', 'cases', 'circuitbreaker', 'stayhome', 'death', and 'deaths'. Words associated with negativity, including 'Covid positive', were classified as negative, while terms like 'vaccine' and 'vaccines' were deemed positive. To account for the neutral connotation of 'positive' in the era of Covid-19, because 'Covid positive' is negative, it

was categorized as neutral. This modified lexicon, referred to as the Adjusted VADER, was then employed to recalculate sentiment scores for each tweet.

Figure 2 displays the histogram of tweets analyzed using the Adjusted VADER, revealing that negative tweets now surpass positive ones. This outcome is consistent with the nature of Covid-19, demonstrating the effectiveness of the Adjusted VADER in capturing nuanced sentiments that may be overlooked by the original VADER tool.

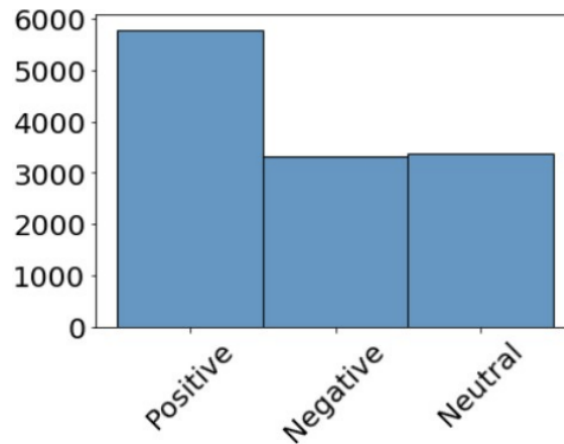


Figure 1: Histogram of Tweets - VADER

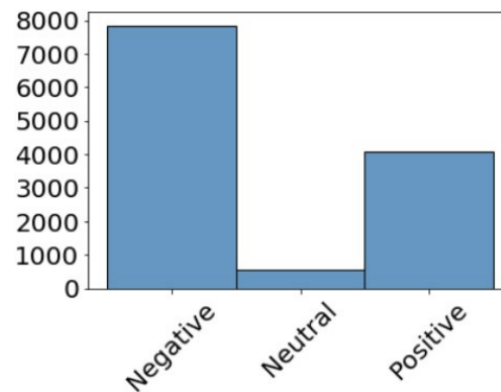


Figure 2: Histogram of Tweets - Adjusted VADER

Table 1: Granger Causality – Covid Cases on Adjusted VADER Sentiment

Lags	SSR based F Test	P - value	SSR based Chi2 Test	P - value
1	42.515	0.0000	42.690	0.0000
2	17.805	0.0000	35.855	0.0000
3	11.244	0.0000	34.060	0.0000
4	6.976	0.0000	28.253	0.0000
5	5.047	0.0001	25.625	0.0001
6	4.335	0.0003	26.487	0.0002

Following this, mean daily sentiment scores were computed. As delineated in both Table 1 and Table 2, at a significance level of 5%, the daily count of Covid-19 cases exhibits Granger causality on the daily mean Covid-19 sentiment score derived from the Adjusted VADER. Notably, the reciprocal relationship is not observed.

Table 2: Granger Causality - Adjusted VADER Sentiment on Covid Cases

Lags	SSR based F Test	P - value	SSR based Chi2 Test	P - value
1	0.061	0.8056	0.061	0.8052
2	0.198	0.8207	0.398	0.8195
3	0.339	0.7969	1.028	0.7945
4	0.529	0.7143	2.144	0.7094
5	0.899	0.4811	4.565	0.4712
6	0.792	0.5760	4.841	0.5643

5. Covid-19 Twitter Sentiment as a Predictor for Housing Prices

Utilizing the Hedonic Price Model, the pricing function for housing can be expressed as $P = f(S, N)$, where S represents the structural attributes of the dwelling and N denotes the neighborhood characteristics (Rosen, 1974). Following the Monocentric City Model, predominant economic activities are concentrated around a central point, namely the Central Business District (CBD). Consequently, residential property values exhibit a negative correlation with distance from the CBD (Alonso, 1964). In the context of HDB resale Flat prices in Singapore, structural features such as Floor Area, Flat Type, Flat Model, Flat Age, and Flat Storey, along with neighborhood characteristics like Proximity to the Nearest MRT Station, Proximity to the Nearest Bus Interchange, Proximity to the CBD, Proximity to the Nearest Hawker Centre, and Proximity to the Nearest Primary School, exert significant influences on pricing dynamics (Yuen, 2005; Belcher and Chisholm, 2018).

$$\text{Price}_{it} = \alpha_i + \beta_1 \text{type}_i + \beta_2 \text{storey}_i + \beta_3 \text{area}_i + \beta_4 \text{Flatmodel}_i + \beta_5 \text{age}_{it} + \beta_6 \text{interchange}_i + \beta_7 \text{MRT}_i + \beta_8 \text{CBD}_i + \beta_9 \text{school}_i + \beta_{10} \text{hawker}_i + \beta_{11} \text{dummyquarterly} + \beta_{12} \text{dummytown} + \epsilon_{it} \quad (1)$$

To validate this assertion, we employ Model (1) on the Pre-Covid dataset, and the outcomes are detailed in Table 3. At a significance level of 5%, it is observed that all the aforementioned characteristics exert a statistically significant influence on the resale prices of HDB Flats in Singapore ($R^2 = 0.848$). Moreover, each characteristic is individually utilized as a predictor for resale HDB Flat prices. The training and test sets, maintained at a 0.25: 0.75 test-train split ratio, are consistently derived from the Pre-Covid dataset for all predictors.

The prediction process involves the utilization of neural networks comprising four hidden layers, employing two distinct activation functions: Scaled Exponential Linear Unit (SeLU) and Rectified Linear Unit (ReLU). The construction of these neural networks is facilitated through the Keras library (Chollet, 2015), and each network undergoes training for 1000 epochs with early stoppage, the outcomes of which are depicted in Fig.3 and Fig.4. The findings indicate that, in terms of predictive accuracy for resale HDB Flat prices, structural characteristics outperform neighborhood characteristics.

Table 3: Results of Model 1

Characteristics	F - statistic	P - value
Flat Type	535.43	0
Flat Storey	528.46	0
Flat Model	1056.68	0
Floor Area	530.42	0
Age	13158.18	0
Proximity to Nearest Bus Interchange (Bus Interchange)	779.58	0
Proximity to Nearest MRT station (MRT)	246.32	0

Proximity to Central Business District (CBD)	142.33	0
Proximity to Nearest Primary School (Primary School)	17.22	0
Proximity to Nearest Hawker Centre (Hawker Centre)	18.70	0

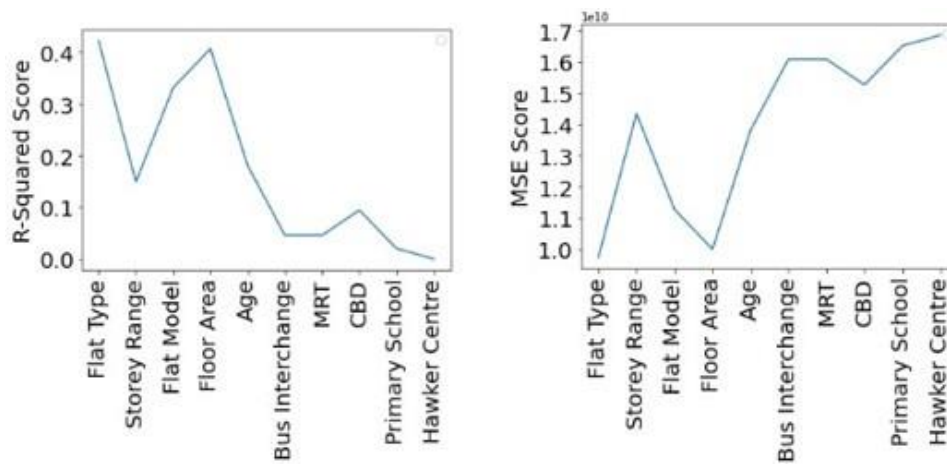


Figure 3: Prediction of Pre Covid HDB Prices using SeLU Activation

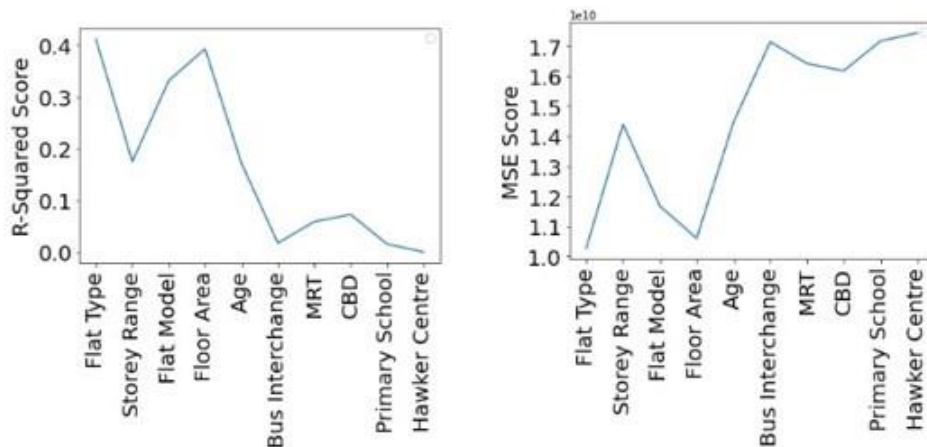


Figure 4: Prediction of Pre Covid HDB Prices using ReLU Activation

With the established statistical significance of conventional housing attributes, encompassing both structural and neighborhood characteristics, in influencing housing prices, the next step involves contrasting their predictive efficacy for HDB resale prices with that of Twitter sentiment related to Covid-19.

The average duration for housing purchase decisions in Israel was approximately 8.2 weeks in 2010, with a local population of 7.6236 million (Genesove and Han, 2012; Population, total-Israel, 2021). In comparison, Singapore had a population of 3.9942 million in 2018 (Population trends 2018, 2018). Considering that the number of houses in a country is intricately tied to its population (Mulder, 2006), it follows that the timeframe for housing purchase decisions is contingent upon population size. Consequently, the estimated average duration for housing purchase decisions in Singapore is around 4.3 weeks, or roughly one month. Hence, the monthly average Covid-19 sentiment score, computed using the Adjusted VADER, is selected as a predictor.

Prior to forecasting, Model (2) is applied to the Covid dataset, and the outcomes are presented in Table 4. At a 5% significance, the monthly average Covid-19 sentiment score, computed through the Adjusted VADER, exhibits statistical significance, contrasting with two conventional neighborhood characteristics—Proximity to Nearest Primary School and Proximity to Nearest Hawker Centre—which do not contribute significantly ($R^2 = 0.902$).

$$\text{Price}_{it} = \alpha_i + \beta_1 \text{type}_i + \beta_2 \text{storey}_i + \beta_3 \text{area}_i + \beta_4 \text{Flatmodel}_i + \beta_5 \text{age}_{it} + \beta_6 \text{interchange}_i + \beta_7 \text{MRT}_i + \beta_8 \text{CBD}_i + \beta_9 \text{school}_i + \beta_{10} \text{hawker}_i + \beta_{11} \text{sentiment}_t + \beta_{12} \text{dummyquarterly} + \beta_{13} \text{dummytown} + \varepsilon_{it} \quad (2)$$

Table 4: Results of Model 2

Characteristics	F - statistic	P - value
Flat Type	67.07	0
Flat Storey	582.52	0
Flat Model	461.49	0
Floor Area	4806.18	0
Age	24522.50	0
Proximity to Nearest Bus Interchange (Bus Interchange)	712.04	0
Proximity to Nearest MRT station (MRT)	237.22	0
Proximity to Central Business District (CBD)	150.86	0
Proximity to Nearest Primary School (Primary School)	1.44	0.231
Proximity to Nearest Hawker Centre (Hawker Centre)	0.03	0.870
Monthly Average Covid-19 Sentiment Score (Sentiment)	57.28	0

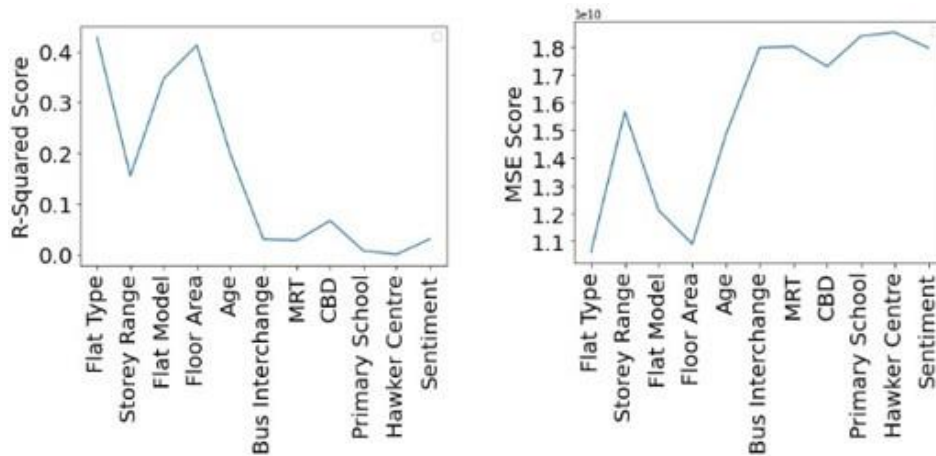


Figure 5: Prediction of Pre Covid HDB Prices using SeLU Activation

The predictive capability of the monthly average Covid-19 sentiment score in predicting resale HDB Flat prices is evaluated in comparison with the structural and neighborhood characteristics outlined earlier. This assessment is conducted on the Covid dataset employing a methodology consistent with that applied to the Pre-Covid dataset, maintaining the same test-train split ratio (0.25–0.75), and the outcomes are illustrated in Fig.5 and Fig.6.

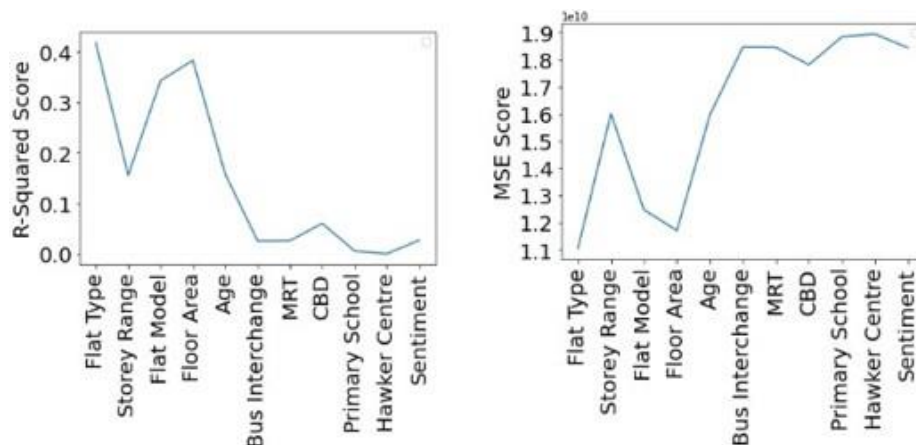


Figure 6: Prediction of Pre Covid HDB Prices using ReLU Activation

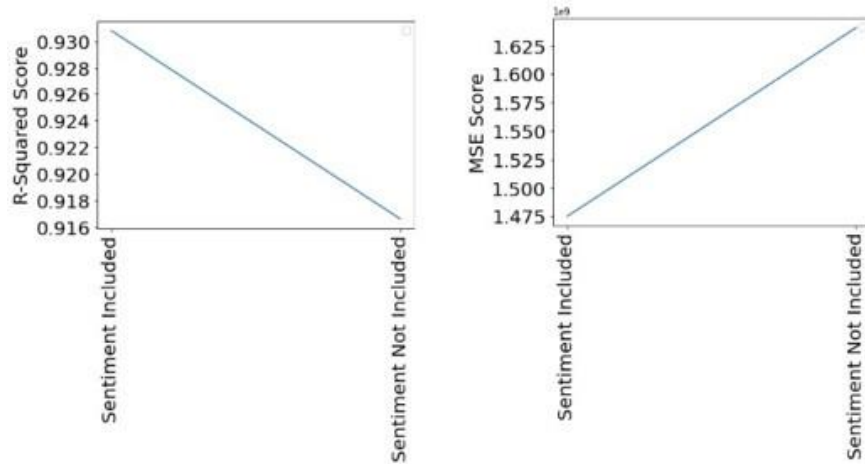


Figure 7: Prediction of Covid HDB Prices using SeLU Activation with and without Sentiment

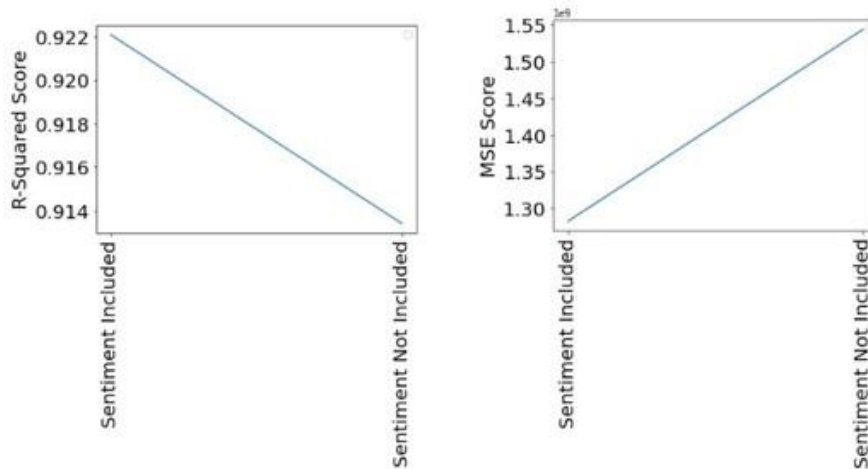


Figure 8: Prediction of Covid HDB Prices using ReLU Activation with and without Sentiment

Furthermore, employing the identical training and test sets, the R-squared value of a composite predictor, incorporating both the traditional structural and neighborhood characteristics along with the monthly average Covid-19 sentiment score, is juxtaposed with a composite predictor utilizing solely the traditional structural and neighborhood characteristics. The findings of this comparative analysis are presented in Fig.7 and Fig.8.

The findings indicate the following:

1. The monthly average Covid-19 sentiment score surpasses the neighborhood characteristics (Proximity to the Nearest Hawker Centre and Proximity to the Nearest Primary School) as a more effective predictor, as evidenced by a higher R-Squared score and a lower Mean Squared Error.
2. The combined predictor incorporating the monthly average Covid-19 sentiment score outperforms the counterpart that excludes it, exhibiting superior results in terms of both a higher R-Squared score and a lower Mean Squared Error.

The findings suggest that incorporating Covid-19 sentiment enhances predictive accuracy for housing prices in Singapore during the Covid-19 pandemic compared to relying solely on traditional predictors. This observation is noteworthy as existing literature underscores the influence of neighborhood and structural characteristics on housing prices (Belcher and Chisholm, 2018; Yuen, 2005), yet there is no precedent in the literature for an external factor such as Covid-19 exerting an influence on housing prices.

6. Limitations and Future Work

A review of existing literature utilizing sentiment analysis of social media data for the prediction of prices, or the examination of price trends reveals an absence of prior investigations employing sentiment related to external factors, such as Covid-19, for the prediction of housing prices. This study addresses this gap by proposing an analysis of resale HDB Flat prices in the context of Twitter sentiment during the Covid-19 pandemic. The research aims to scrutinize the impact of Covid-19 on HDB resale prices in Singapore, leveraging sentiment analysis of Twitter data, and contrasts this with the influence of traditional predictors of housing prices, namely structural and neighborhood characteristics. The findings indicate that Twitter sentiment regarding Covid-19 emerges as a noteworthy predictor of HDB resale prices in comparison to conventional predictors. This research draws upon insights from behavioral economics, natural language processing, and urban economics, thereby opening avenues for potential interdisciplinary investigations in the future.

While this study furnishes valuable insights and practical implications regarding the utilization of Twitter sentiment analysis for predicting resale HDB Flat prices during the Covid-19 pandemic, it is constrained by the fact that it exclusively focuses on HDB resale prices in Singapore, with no inclusion of private property prices.

Subsequent research endeavors may encompass the prediction of prices for private properties and incorporate additional neighborhood characteristics, such as the Proximity to the Nearest Expressway. Furthermore, future investigations could delve into elucidating the economic implications of Twitter sentiment related to Covid-19. Additionally, there is scope for examining how the influence of this sentiment on housing prices undergoes variations across distinct phases of the pandemic.

References

- Alonso, W. (1964), Location and Land Use: Toward a General, East-West Center Press.
- Belcher, R. N. & Chisholm, R. A. (2018), 'Tropical vegetation and residential property value: A hedonic pricing analysis in Singapore', *Ecological economics* 149, 149–159.
- Bollen, J. & Mao, H. (2011), 'Twitter mood as a stock market predictor', *Computer* 44(10), 91–94. 8 [CrossRef](#)
- Cao, M. & Wei, J. (2005), 'Stock market returns: A note on temperature anomaly', *Journal of Banking & Finance* 29(6), 1559–1573.
- Chollet, F. (2015), 'Keras'. URL: <https://github.com/fchollet/keras>, 2015
- Deng, Y., Xie, Q. & Wang, Y. (2018), Research on investor sentiment and stock market prediction based on Weibo text, in '2018 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC)', IEEE, pp. 1–5. [CrossRef](#)
- Edmans, A., Garcia, D. & Norli, Ø. (2007), 'Sports sentiment and stock returns', *The Journal of finance* 62(4), 1967–1998. [CrossRef](#)
- Esmukov, K. (2021), 'Geopy'. URL: <https://pypi.org/project/geopy/>.
- Genesove, D. & Han, L. (2012), 'Search and matching in the housing market', *Journal of Urban Economics* 72(1), 31–45.
- Hannum, C., Arslanli, K. Y. & Kalay, A. F. (2019), 'Spatial analysis of twitter sentiment and district-level housing prices', *Journal of European Real Estate Research* 12(2), 173–189.
- Hausler, J., Ruscheinsky, J. & Lang, M. (2018), 'News-based sentiment analysis in real estate: a machine learning approach', *Journal of Property Research* 35(4), 344–371.
- Hirshleifer, D. & Shumway, T. (2003), 'Good day sunshine: Stock returns and the weather', *The journal of Finance* 58(3), 1009–1032.
- Ho, C. S., Chee, C. Y., Ho, R. C. et al. (2020), 'Mental health strategies to combat the psychological impact of covid-19 beyond paranoia and panic', *Ann Academy Med Singapore* 49(1), 1–3. [CrossRef](#)

- Hu, Z., Wang, Z., Ho, S.-B. & Tan, A.-H. (2021), Stock market trend forecasting based on multiple textual features: A deep learning method, in '2021 IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI)', IEEE, pp. 1002–1007.
- Hutto, C. & Gilbert, E. (2014), Vader: A parsimonious rule-based model for sentiment analysis of social media text, in 'Proceedings of the international AAAI conference on web and social media', Vol. 8, pp. 216–225.
- Ilyas, S. H. W., Soomro, Z. T., Anwar, A., Shahzad, H. & Yaqub, U. (2020), Analyzing brexit's impact using sentiment analysis and topic modeling on twitter discussion, in 'The 21st Annual International Conference on Digital Government Research', pp. 1–6.
- Joo, T. T. K. & Wong, T.-C. (2008), 'Public housing in Singapore: A sustainable housing form and development', *Spatial Planning for a Sustainable Singapore* p. 135. [CrossRef](#)
- Jørgensen, C. J. (2016), 'The space of the family: emotions, economy and materiality in homeownership', *Housing, Theory and Society* 33(1), 98–113.
- Kannan, S., Karuppusamy, S., Nedunchezian, A., Venkateshan, P., Wang, P., Bojja, N. & Kejariwa, A. (2016), 'Big data analytics for social media', *Big data: principles and paradigms*, Cambridge-MA, Morgan Kaufmann-Elsevier pp. 63–94. [CrossRef](#)
- Kinyua, J. D., Mutigwe, C., Cushing, D. J. & Poggi, M. (2021), 'An analysis of the impact of president trump's tweets on the DJIA and S&P 500 using machine learning and sentiment analysis', *Journal of behavioral and experimental finance* 29, 100447.
- Li, J., Wang, Y. & Liu, C. (2022), 'Spatial effect of market sentiment on housing price: evidence from social media data in China', *International Journal of Strategic Property Management* 26(1), 72–85.
- Lim, V. W., Lim, R. L., Tan, Y. R., Soh, A. S., Tan, M. X., Othman, N. B., Dickens, S. B., Thein, T.-L., Lwin, M. O., Ong, R. T.-H. et al. (2021), 'Government trust, perceptions of covid-19 and behavior change: cohort surveys, Singapore', *Bulletin of the World Health Organization* 99(2), 92. [CrossRef](#)
- Mulder, C. H. (2006), 'Population and housing: a two-sided relationship', *Demographic research* 15, 401–412.
- Nofsinger, J. R. (2005), 'Social mood and financial economics', *The Journal of Behavioral Finance* 6(3), 144–160. 9 One Map API (2018). URL: <https://www.onemap.gov.sg/home/>.
- Ozturk, S. S. & Ciftci, K. (2014), 'A sentiment analysis of twitter content as a predictor of exchange rate movements', *Review of Economic Analysis* 6(2), 132–140. [CrossRef](#)
- Phang, S.-Y. & Wong, W.-K. (1997), 'Government policies and private housing prices in Singapore', *Urban studies* 34(11), 1819–1829. [CrossRef](#)
- Population, Total-Israel (2021). URL: <https://data.worldbank.org/indicator/IL>.
- Population Trends 2018 (2018). URL: <https://www.singstat.gov.sg/-/media/files/publications/population/population2018.pdf>.
- Redman, T. C. (2018), 'If your data is bad, your machine learning tools are useless', *Harvard Business Review* 2.
- Ridhwan, K. M. & Hargreaves, C. A. (2021), 'Leveraging twitter data to understand public sentiment for the covid-19 outbreak in Singapore', *International Journal of Information Management Data Insights* 1(2), 100021.
- Rosen, S. (1974), 'Hedonic prices and implicit markets: product differentiation in pure competition', *Journal of political economy* 82(1), 34–55.
- Ruan, Y., Duresi, A. & Alfantoukh, L. (2018), 'Using twitter trust network for stock market analysis', *Knowledge Based Systems* 145, 207–218.
- Ryan, J. M. (2020), Covid-19: Global pandemic, societal responses, ideological solutions, in 'COVID-19', Routledge, pp. 1–8. [CrossRef](#)
- Seiler, M. J., Zhang, Y. & Zhao, L. (2020), 'The effect of real estate ownership on subjective well-being', *Journal of Real Estate Portfolio Management* 25(2), 128–137.
- Snsrape (2022). URL: <https://github.com/JustAnotherArchivist/snsrape>

- Tan, M. J. & Guan, C. (2021), 'Are people happier in locations of high property value? spatial temporal analytics of activity frequency, public sentiment and housing price using twitter data', *Applied Geography* 132, 102474. [CrossRef](#)
- Velthorst, M. & Güven, Ç. (2019), Predicting housing market trends using twitter data, in '2019 6th Swiss Conference on Data Science (SDS)', IEEE, pp. 113–118.
- Wang, Z., Ho, S.-B. & Cambria, E. (2020), 'Multi-level fine-scaled sentiment sensing with ambivalence handling', *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 28(04), 683–697. [CrossRef](#)
- Wojcik, S. & Hughes, A. (2019), 'Sizing up twitter users', PEW research center 24. Yiu, C. Y. (2021), 'Why house prices increase in the covid-19 recession: A five-country empirical study on the real interest rate hypothesis', *Urban Science* 5(4), 77.
- Yuen, B. (2005), 'Romancing the high-rise in Singapore', *Cities* 22(1), 3–13. [CrossRef](#)