MGT 6203

Final Group Report

on

Understanding the impact of COVID 19 on housing prices in Singapore

Team #: 43

Timothy Low

Mohammad Noman H Chowdhury

Fan Qi

Yu Jie Tan

Ya-Hsin Tsai

Submitted on

21st November, 2022

Georgia Institute of Technology, USA

TABLE OF CONTENT

TABLE OF FIGURES

TEAM DETAILS

Team Members

1. Timothy Low
    - GT id: tlow8
    - BSc in Mathematics
    - Data science, analytics and engineering work in various industries; logistics, maritime.
2. Mohammad Noman H Chowdhury
    - GT id: mchowdhury67
    - BSc in Computer Science, MBA in Finance; PhD in Information Systems.
    - Professional experience includes business and systems analysis roles in Telecommunication and software development industry.
3. Fan Qi
    - GT id: fqi7
    - BE in Civil Engineering, MS in Civil Engineering
    - Experience in data analysis and visualization in engineering industry
4. Yu Jie Tan
    - GT id: ytan98
    - BSc in Science (Business Analytics)
    - Experience in data engineering, data science and software development in Telecommunications industry.
5. Ya-Hsin Tsai
    - GT id: ytsai86
    - Current role: financial data analyst,
    - BSc in quantitative finance
    - Previous works include text mining project on news and major index performances and a bit of econometric analysis on carbon tax

1. Background Information

Studies have identified several economic factors (e.g., Foreign direct investment (FDI), interest rate, unemployment, inflation) affecting houses prices in general (e.g., Latif et al., 2020). Government monetary policies have an impact on most of the factors. Fluctuations of housing prices are also found to be prone to events like pandemic, tensions in world politics, news events etc. For example, Wong (2007) investigated the impact of SARS on housing prices in Hong Kong and found calculated expected price falling under the rational asset-pricing model. Similarly, Fritsche (2021) identified a continued upward trend of housing prices in EU countries since the outbreak of Covid-19. Impacts are also investigated in other country-contexts, for example, in China (Yang, & Zhou, 2021), USA (Li, Zheng, 2021), Turkey (Subaşı, & Baycan, 2022), Vietnam (Ha, 2021) etc. Clearly, there is a clear trend of interest among researchers as well as potential buyers regarding variability of house prices in a post-pandemic world.

2. Problem Statement and Research Questions

In this project, we attempted to identify change points, as well as their underlying potential causes relevant to pandemic (e.g., infection rate, news reports, govt. announcements), in housing price trends in Singapore during the last 5 years. We further analyzed the factors dominantly driving house prices post-pandemic and will develop and assess predictive models to predict house prices.

2.1 Primary Research Question (RQ)

RQ: What is the impact of global events on housing prices in Singapore, and how can we predict the trend?

2.2 Supporting Research Questions (SQ)

1. What are the potential change-points in the housing price trend in Singapore during the last 5 years? (SQ1)
2. What could be the potential causes behind those change-points? (SQ2)
3. What factors were the most dominant in driving house prices due to global events in Singapore and how can we predict the housing prices in next 5 years? (SQ3)

2.3 Business Justification

Pandemic has not affected all the countries the same. Government reactions and public perceptions were also not consistent across countries. Consequently, the nature and extent of fluctuations of housing prices in different countries are observed to be unique to their own contexts (Fritsche, 2021). Therefore, stakeholders (e.g., businesses as well as current and future house owners) are interested in having better insight on pandemic's impact on real estate to use these insights to predict the impact of continued events on the housing prices.

3. Methodology

3.1 Time Series

If we assume that real estate prices, similar to stock or any common financial, follows a typical time series, it implies that we can find seasonality, be it weekly, monthly, or yearly. This will allow us to decompose housing prices down to the trend line, further allowing us to analyze the trend specifically. Model availability is high, with several powerful packages like ARIMA, prophet, or we can attempt to build our own versions.

These models can then be compared using K-fold cross validations, with a time series split, easily provided by scikit-learn. This will provide us with RMSE/MAPE values for each time series model, allowing us to evaluate the strength of each model.

Assuming we use something like a Random Forest Regressor, hyper-parameter optimization can be done using GridSearchCV, again a common scikit-learn function, allowing us to tune for the best hyper-parameters that produce the best RMSE or R2 values.

We used the *prophet* package (from Meta) for time series analysis.

3.2 CUSUM Filter

CUSUM, or Cumulative Sum, is a type of anomaly detection method designed to detect any large spikes or changes in a time series (e.g., Page, 1954; Granjon, 2013). This method will allow us to detect the specific changepoints talked about in our introduction and see if we can attribute these changepoints to specific global events that have occurred, such as COVID, the Russia-Ukraine war, and so on.

4. Overview of Data Source

4.1 Data Sources

Our analysis and prediction focus on private residential property transactions from Urban Redevelopment Authority in Singapore and the data link is shown below:

https://www.ura.gov.sg/realEstateIIWeb/transaction/search.action.

The e-Service website comprises private residential property transactions with caveats lodged or options issued within the last 60 months. Caveats are legal documents lodged by purchasers with the Singapore Land Authority to register their legal interest in the property. Caveats are usually lodged by purchasers after the Option-to Purchase is exercised or the Sales and Purchase agreement is signed.

4.2 Data Description

A screenshot of the dataset is shown below in Figure 1. The columns are Project Name, Street Name, Type, Postal District, Market Segment, Tenure, Types of Sale, No. of Units, Price, Nett Price, Area, Type of Area, Floor Level, Unit Price and Date of Sale. Price variable refers to the purchase price stated in the Option and S&P agreement and would have already deducted upfront direct discounts (e.g. X% early bird discount), if any. And Nett Price refers to the purchase price of a unit after deducting both upfront direct discounts and amount or value of any benefit or benefits given or agreed to be given of units sold.

Project Name	Street Name	Type	Postal District	Market Segment	Tenure	Type of Sale	No. of Units	Price ($)	Nett Price ($)	Area (Sqft)[1]	Type of Area[2]	Floor Level	Unit Price ($psf)[3]	Date of Sale[4]
KENT RIDGE HILL RESIDENCES	SOUTH BUONA VISTA ROAD	Apartment	05	RCR	99 yrs lease commencing from 2018	New Sale	1	1,889,088	-	947	Strata	01 to 05	1,994	Sep-22
V ON SHENTON	SHENTON WAY	Apartment	01	CCR	99 yrs lease commencing from 2011	Resale	1	4,300,000	-	1,765	Strata	46 to 50	2,436	Sep-22
ONE BERNAM	BERNAM STREET	Apartment	02	CCR	99 yrs lease commencing from 2019	New Sale	1	1,389,000	-	463	Strata	21 to 25	3,001	Sep-22
STIRLING RESIDENCES	STIRLING ROAD	Apartment	03	RCR	99 yrs lease commencing from 2017	Sub Sale	1	2,200,000	-	980	Strata	26 to 30	2,246	Sep-22
ICON	GOPENG STREET	Apartment	02	CCR	99 yrs lease commencing from 2002	Resale	1	1,350,000	-	786	Strata	06 to 10	1,718	Sep-22
THE ANCHORAGE	ALEXANDRA ROAD	Condominium	03	RCR	Freehold	Resale	1	2,550,000	-	1,421	Strata	11 to 15	1,795	Sep-22

Figure 1: Screenshots of Private Residential Property dataset

4.3 Key Variables

In our time series model, we focused on housing price, which is the variable Price in Private Residential Property dataset. We performed prediction on the dataset using regression model as well. In regression model, house price (Price in dataset) will be considered as dependent variables and other variables are independent variables. The independent variables contain some basic features of each residential property including floor area, floor level, type and so on. Before our analysis, we hypothesize that the most important variable will possibly contain Floor Level, Type, and Area.

4.4 Data preparation

Datasets are cleansed and processed (e.g., date formatting, normalization, text-to-numeric conversion, cleaning of missing data) before we fit them into time series and regression models. We created average housing price of all residential types on a monthly basis for the time series model. In addition, we created dummy variables for the independent variables in the regression model.

5. Results

5.1 Supporting Question 1 & 2

What are the potential change-points in the housing price trend in Singapore during the last 5 years, and what are their causes?

- After decomposing our time series into its individual components, we can extract the main trendline of the average housing price in Singapore. We find the changepoints of the gradient shifting in the image below, shown by the dotted red lines.

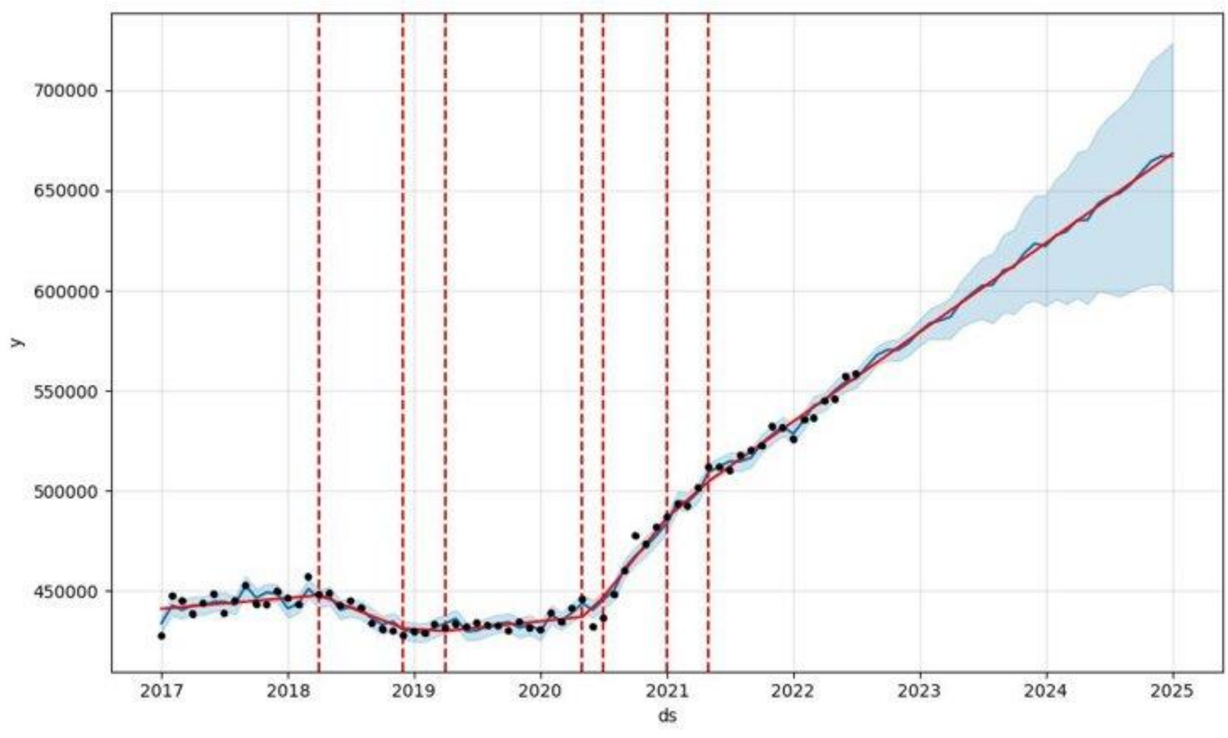


Figure 2: Identification of change points

- From observation, we see that at the start of 2018 the housing prices had a gradient shift from increasing to decreasing, most likely due to the explosive growth of US tech stocks at that point in time. Next, in 2019 with the COVID situation impacting most countries, we find that most Asian economies experienced their slowest rate of growth since the

2008 financial crisis, reflected in the stagnation of the housing prices for the entirety of 2019.

- We notice however that at the start of 2020 the housing prices in Singapore start to spike upwards, jumping from 450k to around 500k. This is most likely due to the labor shortage caused by the COVID-19 pandemic, causing a drastic decrease in the availability of new houses being built for the majority of 2020-2021. Many young Singaporeans might end up having to wait years for their BTO (Build to Order public housing) to get completed.
- This is further exacerbated by the beginning of the Russia-Ukraine war, as then began a global shortage of steel, resulting in the prices of houses globally increasing once again.
- All in all, according to the model, it appears that the housing prices in Singapore are not set to come down anytime soon.

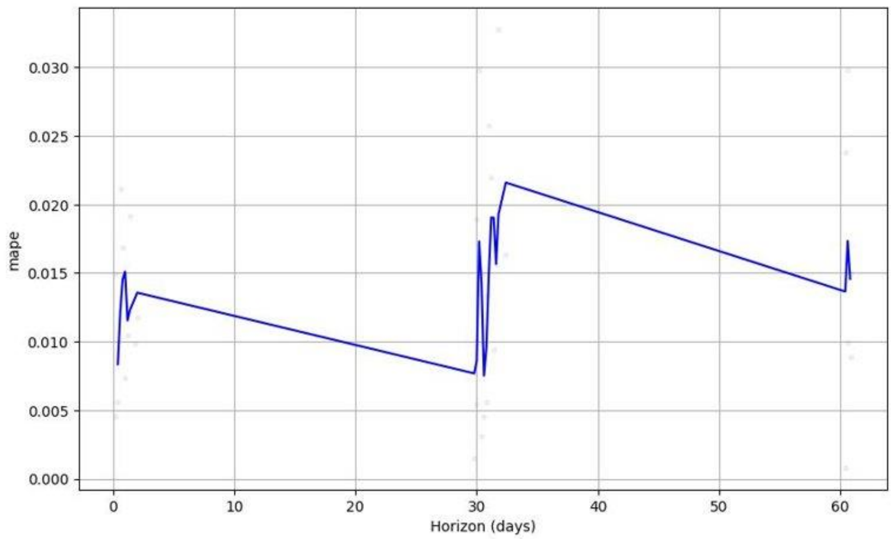


Figure 3: MAPE Cross Validation plot

- As a side note, we also cross validated the models in order to make sure the models are accurate. According to the mean average percentage error (MAPE) plot (Figure 3), the prediction ability of our model for the next 2 months is around 98%. This is verified against 8 cross validated runs, trained on about 3 years of data, and then iteratively predicted against the next 2 months.

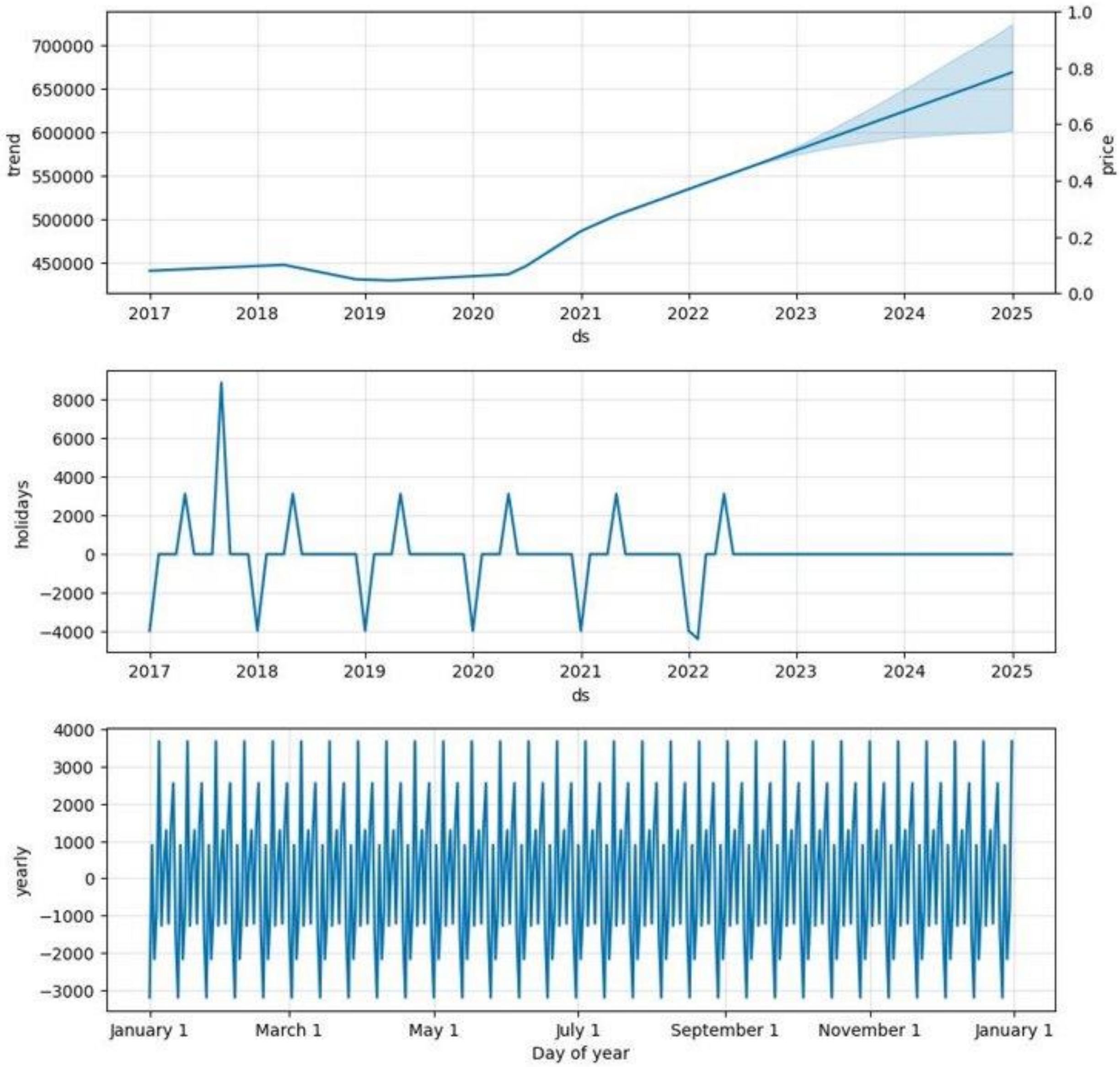


Figure 4: Seasonality Plots

- Seasonalities were also extracted using our time series model, although the implications of this plot are not as useful, as they are aggregated on a monthly basis.

5.2 Supporting Question 3

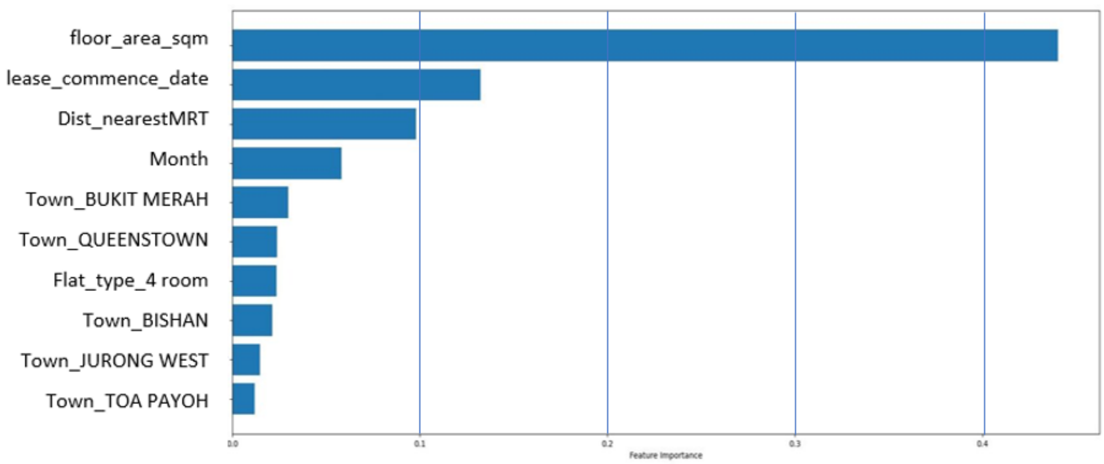
What factors were the most dominant in driving house prices due to global events in Singapore and how can we predict the housing prices in next 5 years? (SQ3)

- In order to address the SQ3, we have implemented five regression models namely Lasso regression, Ridge regression, linear regression, MLP Regressor and Random Forest Regression. We will ascertain the best performing model using some metrics like MSE, MAE, MSLE, MAPE and R values.

	MSE	MAE	MSLE	MAPE	R	time_taken
Lasso	0.016180038	0.099184065	0.009513734	8779326970	-4.98E-05	1.701666594
Ridge	0.002309747	0.037025302	0.001337749	4370918164	0.857236301	1.358635902
LinearRegression	0.002309686	0.037029291	0.001337659	4368224162	0.857240106	4.241663933
MLPRegressor	0.001002557	0.022874973	0.000577019	2362377176	0.938029819	176.3257101
RandomForestRegressor	0.000533181	0.016153401	0.000304953	2524498973	0.967046289	515.4346607

For each model, we conducted 3 repetitions of seeded 5-fold cross validation and obtained the results in Figure 5. We have also recorded the time taken for the models to run to measure efficiency of the models.

- From the results, we concluded that Random Forest Regressor provided the best across all the metrics used.



- From Figure 6, we observed that floor area is the most important feature. This means that the bigger the house the more expensive it is, and this could be the singular most important factor for house prices in Singapore.
- Next, the distance to public transport is the third most important feature. This is a phenomenon we often see in Singapore as houses located nearer to public transport, especially train stations, will provide greater convenience.
- We also see several categorical variables in the feature importance list. This includes the town in which the houses are located and some features of the house. We then plotted several focused feature importance charts to investigate.

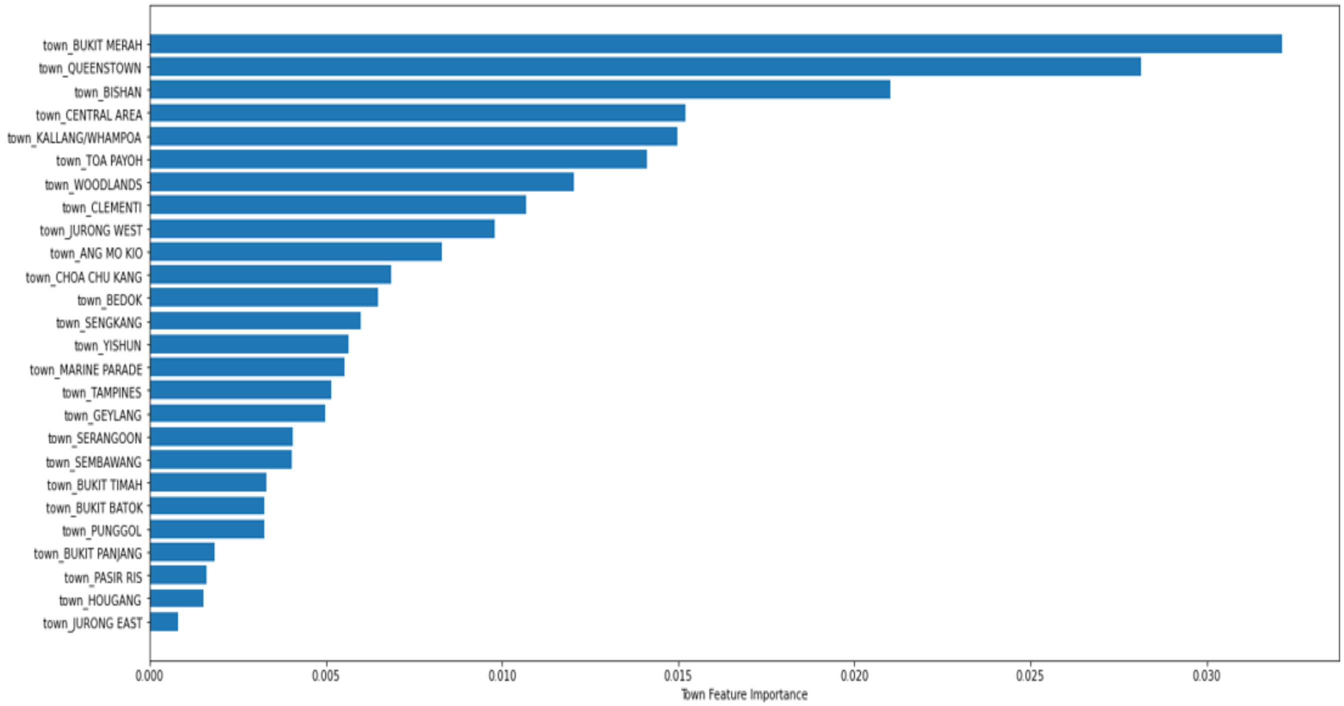


Figure 7: Town Feature Importance

- We plotted the feature importance in Figure 7, limiting to only the town area categorical variables. From the chart, we observed that the towns that contribute the most to the house prices are either the central or the sub-urban locations. Towns that are further away from the city center will have little impact on house prices.

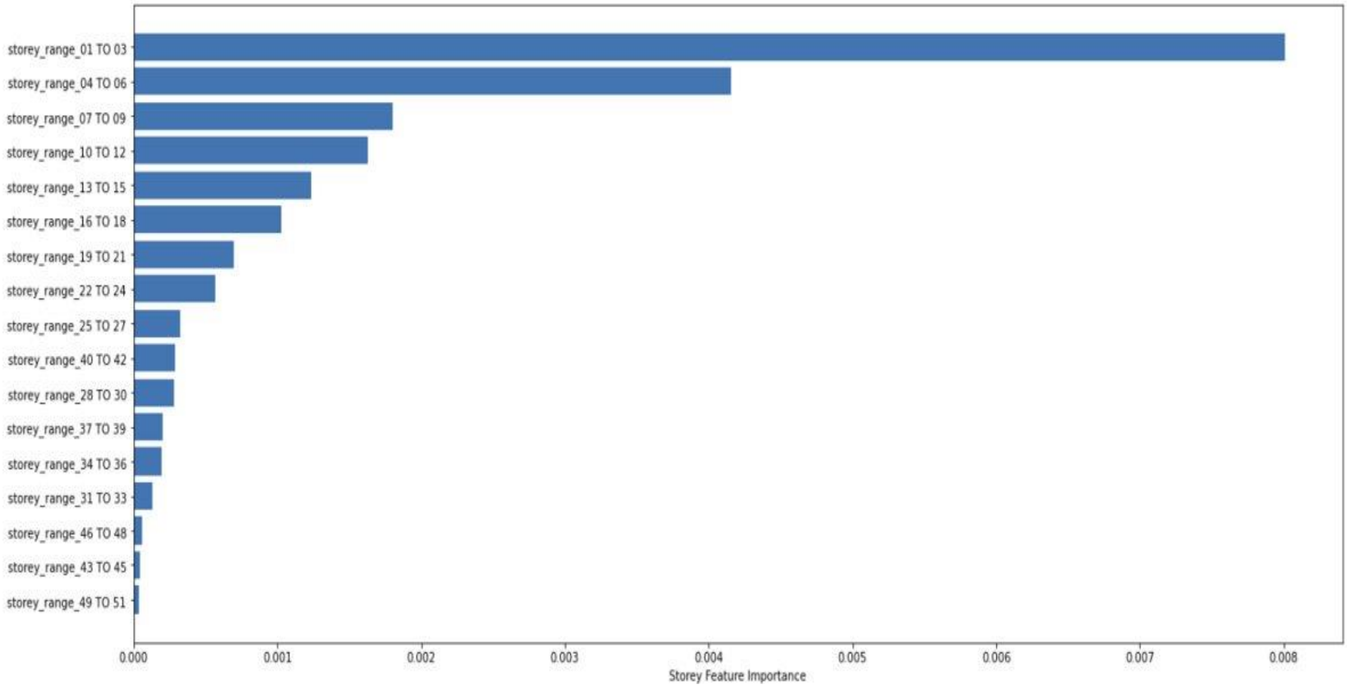


Figure 8: Storey Feature Importance

- Next, we also want to find out if the level of the houses will affect their price. Hence, we plotted a Storey feature importance chart in Figure 8.
- From the chart, we observed that lower floors will have a greater impact on house prices. We understand that Singaporeans tend to prefer houses located at higher floors. Hence, based on the results from the chart, we can reasonably suspect that the lower floor categorical variables are important in identifying the cheaper house prices, keeping all other variables constant

Based on the analysis done on feature importance from our RandomForestRegressor model, we found that the most important variable in predicting the house prices in Singapore is the floor area. Other factors that play a smaller role include distance from public transport, the town which the house is located in and the storey of the house.

We opted to keep the time series model and the regression model separate. The time series model will be used to find certain the optimum time to buy assets or to detect the time that buyers should avoid as the market is not predicted to be well-performing. Regression model, on the other hand, will then be used to predict the price based on dependent variables. Integrating both models into one is found not to be well performing. This combination of models will give us macro (i.e., seasonality) as well as micro level (i.e., understanding if a certain house is over or underpriced) understanding of house prices.

6. Discussion and Business Impact

The Singaporean property market, despite its steady growth in price when economic expansion, is highly sensitive to mega negative events, including the “black swans” (Covid-19, the Great Recession in 2008, etc.) and macroeconomic conditions (geopolitical conflict between Ukraine and Russia). Negative events and their subsequent impacts both influence the property market but may in different time spans and possibly different ways in price. Covid-19 and the labor shortage from relevant shutdowns, for example. The former provides a short-term strike while the latter generates a long-term upward spike. Once the market price increases, even driven by abnormal inflation, it will never be adjusted back. This once again proves that houses are more unaffordable to the public whereas the mega negative events may come from wrong decisions from a few people.

Many micro-level features our team first believed to be highly impactful to the property market turned out to be not. Distance from the closest subway stations, for example, is not considered important while subway stations are the primary measure of transportation for Singaporeans. In fact, features providing information about property location, especially in the suburban area, take most of the importance in the model. It suggests suburbanization in Singapore, though more towards the physical landscapes and built greeneries than directly moving major property sources to the outskirts, still shows a positive impact on the price as cities like New York and London. Also, there might be more effective factors that influence the property market on the micro-level, despite our team’s hypothesis from a user point of view.

The model identifies the growth momentum, in turn provides investors a leading indicator to join the investment opportunities. It also identifies the potential loss in prices, informing investors to leave the market at an early stage or make short positions.

7. Limitations and Future Work

We identified some limitations of our work and look forward to further exploring this topic. First, time series model can be expanded to markets in other countries or regions. For example, a dataset from Zillow can be tested: https://www.zillow.com/research/data/. This dataset contains typical value for homes within the 65th to 95th percentile range for regions in the United States. Mega negative events typically have global impacts. Therefore, testing our model on dataset from another market will be meaningful.

RegionID	SizeRank	RegionName	RegionType	StateName	12/31/17	1/31/18	2/28/18	3/31/18	4/30/18	5/31/18	6/30/18
102001	0	United States	country		257933	256300	258267	263267	271267	276300	279667
394913	1	New York, NY	msa	NY	489300	489300	494333	503000	513000	521300	526300
753899	2	Los Angeles, CA	msa	CA	703332	702998	711333	721333	735000	743333	750000
394463	3	Chicago, IL	msa	IL	274933	271600	276267	284600	294600	300600	302267
394514	4	Dallas, TX	msa	TX	316333	317333	320667	322997	328497	331797	332800
394692	5	Houston, TX	msa	TX	287633	289133	291433	294467	297933	299633	299933
395209	6	Washington, DC	msa	VA	404667	395667	395667	405633	419967	433267	439933

Figure 9: Screenshots of Zillow dataset (LIST and SALE PRICES)

Furthermore, we can use house price dataset from other cities or areas to check and compare the key predictors of house price in the local market across different countries. Table 1 listed some data sources from different countries.

Table 1: Datasets of housing price

City or Area	**Data Link**
Melbourne, Australia	https://www.kaggle.com/datasets/dansbecker/melbourne-housing-snapshot
Boston, US	https://www.kaggle.com/datasets/vikrishnan/boston-house-prices
King county, US	https://www.kaggle.com/datasets/harlfoxem/housesalesprediction
Chennai, India	https://www.kaggle.com/datasets/amaanafif/chennai-house-price
Bangalore, India	https://www.kaggle.com/datasets/saipavansaketh/pune-house-data

In 2018, Singapore released its Land Transport Master Plan (LTMP) 2040 and envisions 20-minute towns and 45-munites city, where people can reach nearest neighborhood center within 20 minutes.

To reflect this concept in housing price prediction, we can create a few new variables include number of shopping malls in postal district, number of hospitals in district, availability of good schools, distance to nearest tourist attraction, and number of hawker centers in our model and identify more key indicators in our prediction model for housing price. These variables are assumed to reflect the convenience and quality of life in each house.

Lastly, we would like to find a way to make the regressors work within the time series itself instead of using two separate models. And we will try Pytorch neural networks to see if it will outperform the model we created in this project.

Due to time limit, we were not able to incorporate every thought in our current working paper. However, we will continue exploring this topic to make it more solid and comprehensive.

8. References

Fritsche, C. (2021). *The expected impact of COVID-19 on the housing market* (No. 5). Streiflicht VWL.

Granjon, P. (2013). The CuSum algorithm-a small review. 2013. hal-00914697

HA, N. H. P. (2021). Factors affecting real estate prices during the COVID-19 pandemic: An empirical study in Vietnam. *The Journal of Asian Finance, Economics and Business*, *8*(10), 159-164.

Latif, N. S. A. ., Rizwan, K. M. ., Rozzani, N. ., & Saleh, S. K. . (2020). Factors Affecting Housing Prices in Malaysia: A Literature Review. *International Journal of Asian Social Science*, *10*(1), 63–67. https://doi.org/10.18488/journal.1.2020.101.63.67

Lau, J. C. Y. (2011). Spatial mismatch and the affordability of public transport for the poor in Singapore's new towns. *Cities*, *28*(3), 230-237.

Li, X., & Zhang, C. (2021). Did the COVID-19 pandemic crisis affect housing prices evenly in the US?. *Sustainability*, *13*(21), 12277.

Padawangi, R. (2010). The planned suburbanization of a city-state: Singapore's new towns. In *Suburbanization in global society*. Emerald Group Publishing Limited.

Page, E. S. (1954). Continuous inspection schemes. *Biometrika*, *41*(1/2), 100-115.

Subaşı, S. Ö., & Baycan, T. (2022). Impacts of the COVID-19 pandemic on private rental housing prices in Turkey. *Asia-Pacific Journal of Regional Science*, 1-17.

Wong, G. (2008). Has SARS infected the property market? Evidence from Hong Kong. *Journal of Urban Economics*, *63*(1), 74-95.

Yang, M., & Zhou, J. (2022). The impact of COVID-19 on the housing market: evidence from the Yangtze river delta region in China. *Applied Economics Letters*, *29*(5), 409-412.