Contents

10. References And Appendix
1. Introduction

HDB flats are the most common type of subsidised public housing in Singapore, currently accommodating an estimated 81 percent of Singapore's population (HDB, 2019). The demand for resale flats has also experienced an upward trend, with a 6.7% increase in applicants registered for resale flats from 2018 to 2019, and a further 17.6% increase in 2020 (Yong, 2020).

2. Business Problem – A Case for A Data-Driven Approach

Consider the case for a **Real Estate Consulting Firm** aiming to provide consulting services to those who intend to buy or sell resale flats that maximise their benefits based on the valuation of current resale flat properties**.** HDB resale flats have been an attractive market for real estate investors and agents, which has been seeing appreciating prices and getting more attention from investors and agents (OrangeTee, 2019). Like most markets, the HDB resale flat market also sees fluctuations in prices over time. To profit from market conditions, it is crucial to be able to accurately value properties and decide if current market prices are suitable for entry or exit. For agents or homebuyers with intention of using real estate as an investment, it is paramount to predict a fair value for their properties, which depend on a myriad of factors, ranging from the region or district the flat is in, floor area, flat model, remaining lease period.

As a complex market, the widely available market data on HDB resale flats provides the perfect opportunity for a data-driven approach to valuation of these real estate properties to provide a concrete, evidence-based metric to complement investors and agents in their decision-making process, that is, to determine with current market conditions, if a property is worth buying or selling. Here, we will explore a quantitative way to do so.

3. Data Source and Feature Engineering

3.1 Overview of Data Set

The preliminary dataset resale-flat-price.csv is freely available from data.gov.sg. This dataset includes key attributes such month, resale price, floor area, flat type and model, remaining lease and the storey level from 2017 to 2020. There is a total of **68161** data points, with initial variables "month", "town", "flat_type", "block", "street_name", "storey_range", "floor_area_sqm", "flat_model", "lease_commencement_date", "remaining_lease", and one variable "resale_price", that will be our target variable.

3.2 Data Wrangling and Feature Engineering

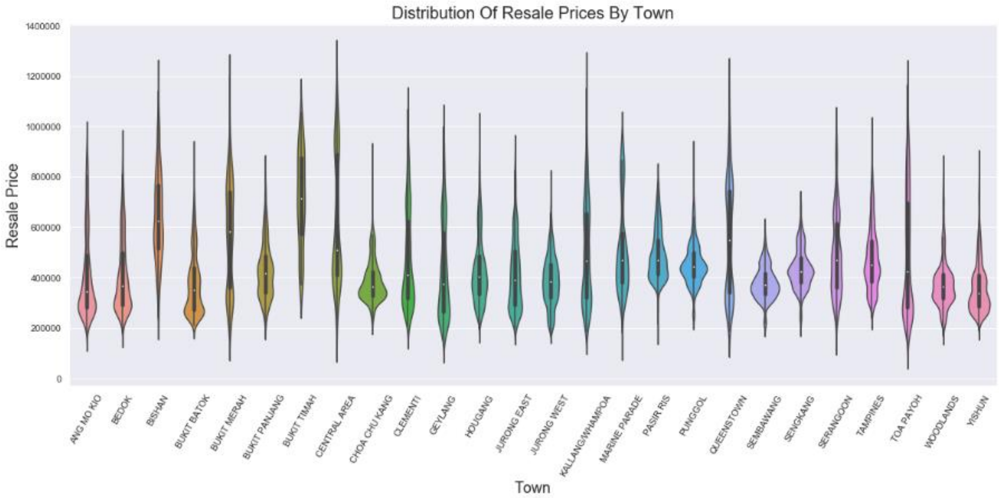
We converted the string attribute "remaining_lease" to obtain a new numerical variable "months_remaining_lease". Next, we grouped the 26 different towns under "town" according to their geographical region (e.g. North, East etc) with reference to the dataset "towns_data.csv" and reclassified the HDBs by their regions in "region". The original data on storey level is given in a range in string, thus we found the mean storey level (e.g. 10 TO 12 → 11) and stored the information under "mean_storey_range". Lastly, we extracted the numerical block numbers from "block" to obtain a new variable "block_number" by removing the alphabets that follow the block numbers (e.g. 100B → 100). The updated primary data source is stored in a new dataset file "*processed-resale-flat-price.csv*".

However, these data attributes are insufficient for a robust predictive model as it does not include other important factors, such as proximity to schools, malls or MRTs. To enhance the dataset, we included the proximity to the nearest MRT station, malls, and schools for each data point. This required the coordinates of each HDB flat, which was retrieved using OneMap.sg's free geocoding Application Programming Interface (API). Using these coordinates, we queried for the nearest MRT based on these coordinates using Google Places API. We also queried the coordinates on a comprehensive list of malls obtained from Wikipedia, and school data from data.gov.sg. Subsequently, we used these data to determine the nearest malls and schools and combined all our data into a final dataset, "*finalDataSet.csv".*

4. Preliminary Findings

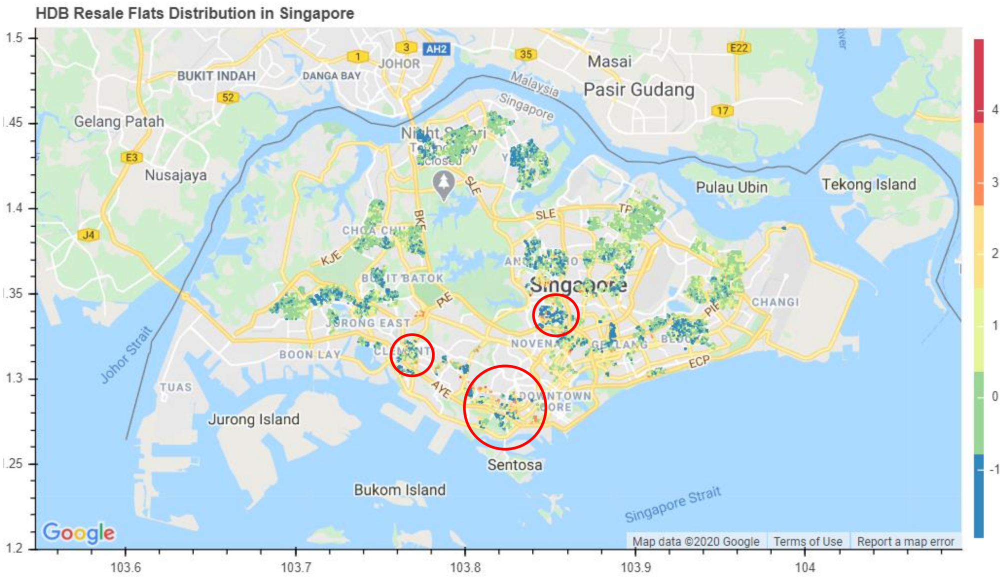
To derive a meaningful, robust model, a visualization on factors which may affect valuation must be understood. To rigorously select these variables and understand their effects individually, a preliminary visualization will help to guide our analysis.

4.1 Statistical Distribution of Prices



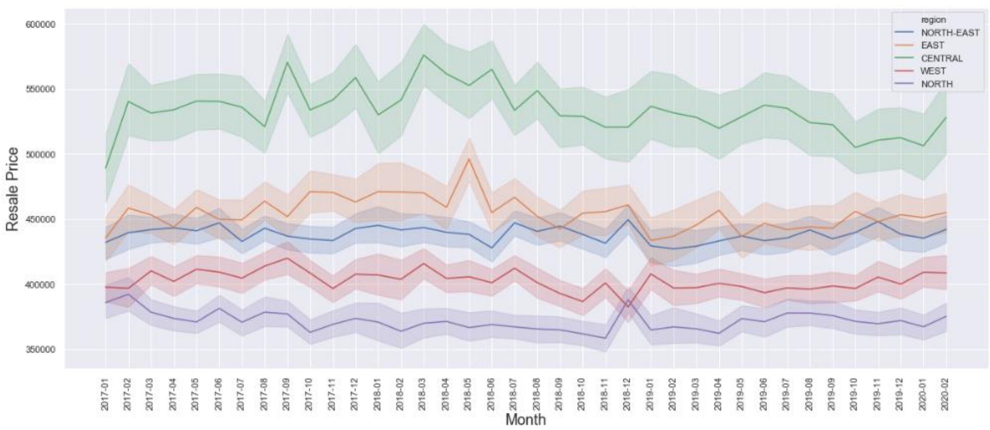
Using Seaborn's violinplot, important statistical information can be visualized – the median, upper and lower quartiles, range and the kernel density function (frequency) by their resale prices. The Central Region has the widest range of prices while Bukit Timah has the highest median resale price. Choa Chu Kang also has the most "centered" resale prices slightly below $400,000

4.2 Geographical Distribution and Prices



Using Google Maps API, *bokeh* and our queried geocoded coordinates, we can visualize how HDB resale flats are distributed geographically, as well as their relative prices, marked by their intensity of colour. Evidently, the more expensive flats lie in the Downtown Core, and perhaps surprisingly, pockets of expensive HDB flats are also found in the Ang Mo Kio and Jurong Region (See circled points). Please refer to Notebook *3.2* for the interactive plot.

4.3 Resale Price Trends Over Time

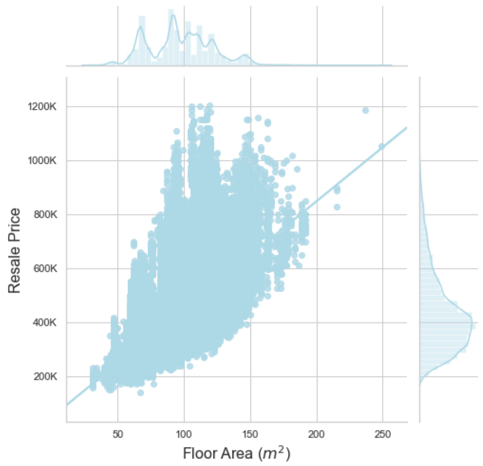


Using Seaborn's new lineplot method, we can visualize price trends over time, with its confidence bands, over the various regions. The Central region maintains an exclusive upper band of prices. The HDB resale price of units in the East is the 2nd most expensive on average in Singapore, and units in the North as valued the lowest. There does not seem to be a clear seasonality in terms of prices based on this visualization.

5. Analysis On Selected Features

A full comprehensive visualization can be found in our Notebook. We have selected a few key features based on our visualizations which have exhibited clear relationships with resale prices, analysed below.

Key Attributes



5.1 Floor Area

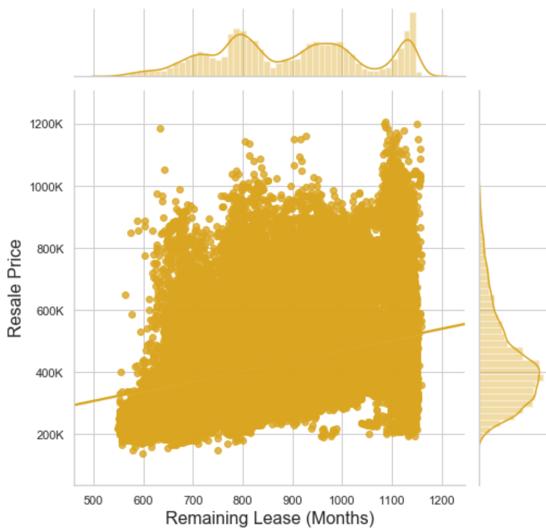
Rationale: Widely known, on average, keeping all other factors constant, the larger the floor area, the more expensive a property as one would have to pay more to own more land.

Analysis: Using Seaborn's jointplot, analysis shows a clear positive linear relationship between floor area and price, pearsonr coefficient of 0.632. Hence, we expect floor area to be an important determinant in valuating HDB resale flats. We also note from the distribution plots of floor area that among the floor areas around 90 to 130 square metres, there is a large range of resale prices, indicating the possibility that there are other factors which may be significant factors to resale prices beyond simply the floor area. Potential buyers could also use this empirical data to decide if an offered price could be negotiated lower. We also note the presence of 2 points with a floor area of around $250m^2$. To investigate this point, we take advantage of bokeh's HoverPlot (see Notebook *3.6*). These points are revealed to be that of 65 Jln Ma'Mor and 41 Jln Bahagia, part of HDB's 1959 project of landed properties, accounting for its high floor area and resale price. Removing these 2 points, we get a new pearsonr coefficient of again 0.632, which still maintains that there is a strong relationship between these 2 variables. Assuming a linear relationship between these 2 variables, the regression coefficient stands at 4003, that is, every additional metre square cost $4003. Due to the importance of this factor on resale price, we will be normalizing resale prices with their respective floor area, that is, their price per square metre, in subsequent analysis.

5.2 Average Storey Level

Rationale: Usually, higher floors are priced higher even by HDB (HDB, n.d, 1). This could be due to a majority of homeowners having a preference for better views from higher floors, better air ventilation, or a increased sense of privacy from passer-by on the ground level.

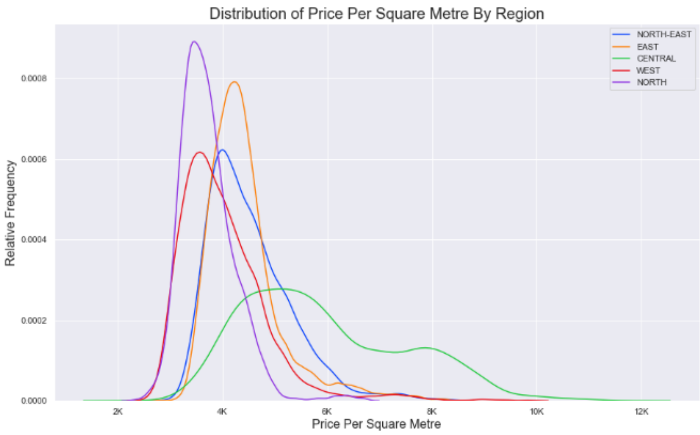
Analysis: This plot shows the relationship between the storey level and the normalized resale price by floor area (price per m^2). The correlation coefficient is found to be 0.47 (see below), indicating a moderate positive linear relationship between these 2 factors, though it is apparent that at least some of the difference can be explained by its geographical region, as we observe the rough "bands" of colours in this strip plot.

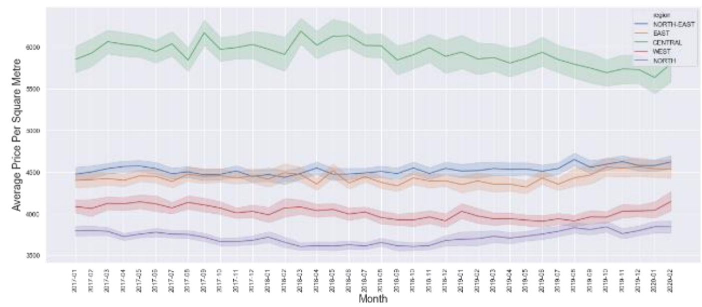


5.3 Remaining Lease

Rationale: All HDB flats have a lease of 99 years (HDB, n.d, 2), after which it is returned to HDB. Therefore, the longer the remaining lease, the more we expect the unit to be valued. At the same time, a longer lease implies that the unit is newer, and hence is valued at a higher price. However, there is a conflicting factor of floor area, that is, that HDB flats have been getting smaller over time (Chan & Lim, 2019). This would entail paying less for a smaller land area but paying a premium for a newer flat. This will be accounted for in our model construction later.

Analysis: The relationship between resale prices and remaining lease is relatively strong, with a correlation coefficient of 0.325. After normalizing for floor area, the correlation coefficient drops to 0.258. This suggests that the remaining lease remains an important factor for valuation, though the lease period may also share a relationship with floor area (see Notebook *5.2*). This may be due to financing options from banks or HDB become more limited when lease becomes shorter (Wong, 2018). Further, flats with less than 60 years' lease may start to depreciate as prospective buyers are limited in using Central Provident Funds (CPF) to finance the purchase (Wong, 2018), which may affect the demand for these resale flats and hence ability to command higher prices. Interestingly, we can also note this effect from the frequency distribution of remaining lease displayed by the joint plot, with those with extremely low remaining leases having low volume of transactions as compared to those with longer leases. There also seems to be a large variation in price at every lease level, which could suggest that there are other factors which also contribute to its price.

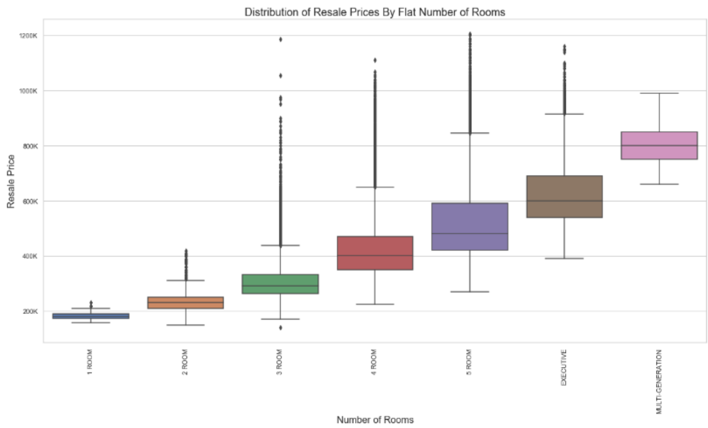




5.4 Geographical Region

Rationale: The location of a property is a key factor in its value due to its proximity to town, reputation and status. As a status-conscious society (Paulo & Low, 2018), property is an important status symbol (Goh, 2005). Hence, the more "expensive" a region is, the higher the cost per square foot of the property.

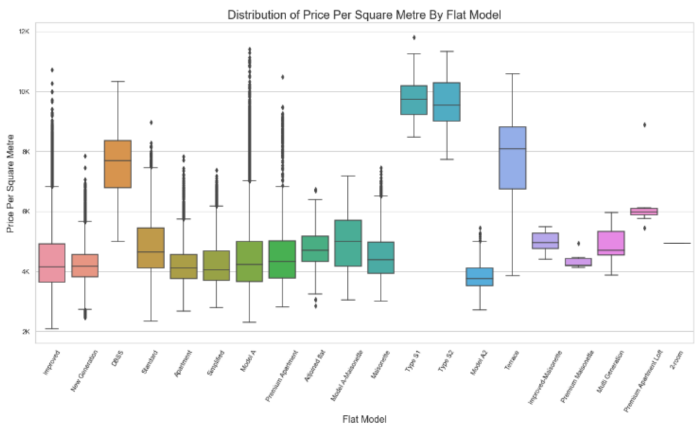
Analysis: As expected, properties in the Central Region command on average higher prices, but its distribution is relatively even, that is, there is broad but almost equal frequency of properties prices from $5K to $8K per m^2. A buyer looking for cheaper housing would find most in the North and West, with most of its values around $3K to $4K per m^2. Properties in the East have more expensive prices but exhibits a tighter spread of prices compared to the North-East, which exhibits "fatter" tails. Analysing the price per m^2 over time, unlike the total resale price examined above, properties in the North-East seem to be more expensive on average compared to the East. This may be attributed to the "fat-tails" of the North-East distribution of prices. It is also likely that properties in the East have larger floor areas on average, accounting for the higher overall resale price despite having lower per m^2 price than the North-East.



5.5 Flat Type (Number of Rooms)

Rationale: We would expect that the higher the number of rooms, the higher the resale price.

Analysis: It appears that this assumption seems to be empirically true, with median resale prices increasing over time, as well as their interquartile ranges. While this may also be attributed to the increase in floor area, we note that the marginal increase in floor area with every increase in room number diminishes (HDB, n.d, 4). Hence, we do observe some premium effect with the increase in number of rooms.

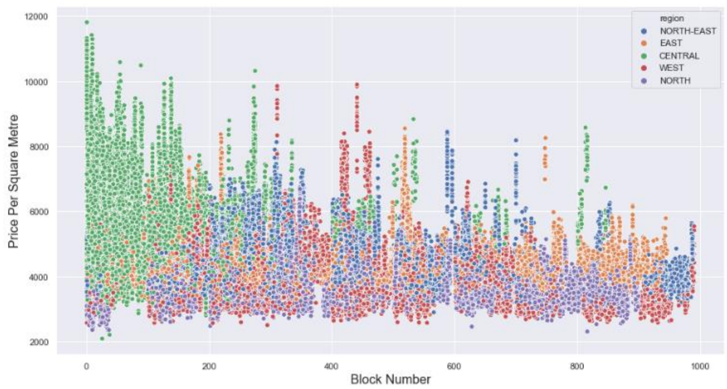


5.6 Flat Model

Rationale: HDB has been constantly upgrading its designs to improve the perceived quality of its public housing. Popular new initiatives such as Build-To-Order aim to provide condominium-like finishings to HDB homeowners at a fraction of the cost (Ang, 2019). We do expect that more polished, luxurious designs and quality be worth a higher price.

Analysis: Evidently, different flat models incorporate different floor areas (HDB, n.d, 3). We hence look at the price per m^2 instead of the absolute prices to determine the effect of flat model itself on prices (Visualization on absolute prices found in Notebook *3.9*). While most of the flat models have around the same median price per m^2, some exceptions are present such as DBSS models, Type S1, S2, and Terrace models. These units are often seen as "premium units" (Ong, 2017). These flats are usually found in prime areas, and feature unique architectural features (Choo, 2019), which hence command for the premium visualized here. Hence, the flat models itself, independent of its floor area, may be an important metric in valuation.

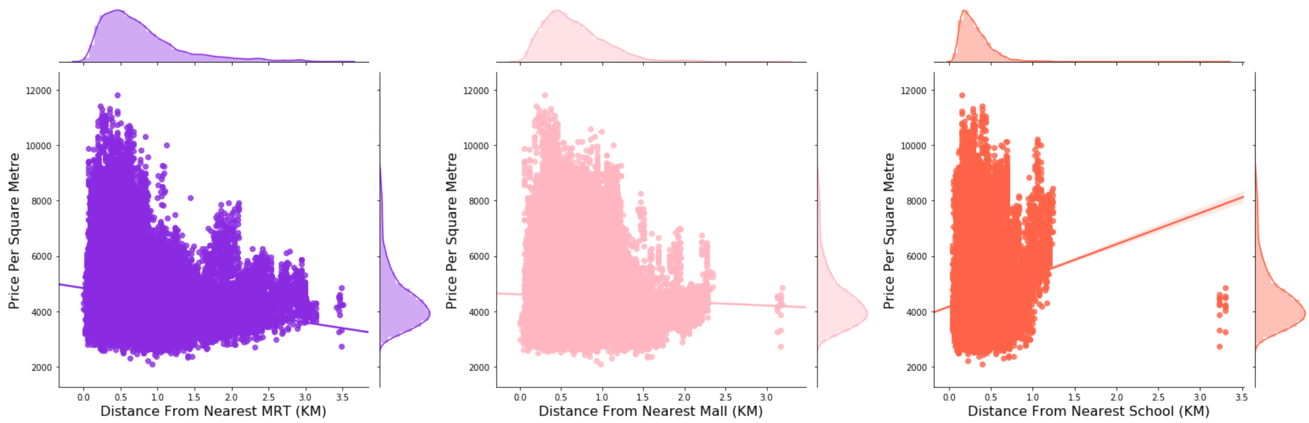
Other Attributes



5.7 Block Number

Analysis: Surprisingly, block numbers played a non-trivial effect on valuation prices even after accounting for floor area through normalization, having a correlation coefficient of -0.339. We note that smaller block numbers are found disproportionally in the Central Region, which may account for its higher price per m^2. Interestingly, this is also due to HDB's block numbering policy, where the first digit in HDB's "3-digit numbering system" denotes the neighbourhood (Teo, n.d), while 2-digits usually denotes a Central area. This may account for the variation in prices due to block number.

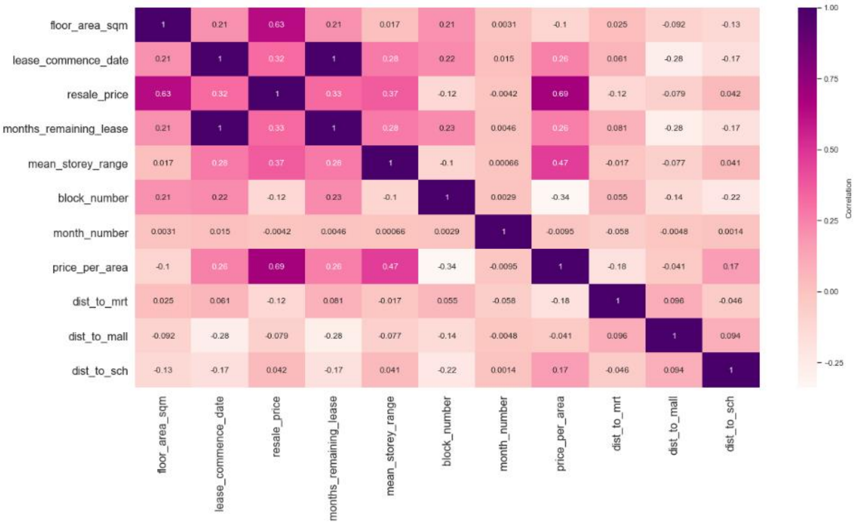
5.8 Distances To Nearest Key Facilities



Rationale: Proximity to key facilities such as to MRT stations, malls and schools implies higher convenience, which would hence be expected to be worth a premium. In fact, MRT is a common mode of transport, with 60% of Singaporeans taking MRT to work (Lee, 2016). It is therefore no surprise that on average, properties near MRT stations are valued around 10-15% more than those further away from MRT stations (Navaratnarajah, 2015). We also expect a property to be more expensive when it is nearer to a school since it is more convenient for families with children. This would be especially true for parents with children seeking entry to primary schools, where their residential address must be within a 1km radius to gain priority in balloting phase for a place in the primary school (MOE, n.d).

Analysis: Likewise, to remove the effect of floor area on resale prices, we normalized the prices by their floor area.While our hypothesis held for the first 2 attributes (MRT and malls), with a correlation coefficient of -0.185 and -0.041 respectively, distance to schools showed a positive correlation (0.168). We note that all three exhibit a few points which are much further away from all 3 facilities than most, which corresponed to properties on Changi Village Road. After accounting for these, the correlation coefficients are rougly the same, at -0.185, -0.078 and 0.175 respectively. This may be due to our dataset having only public school data and not private schools (e.g international schools) which may be in the area. The correlation between schools and prices may also be attributed to the fact that properties near schools have significantly higher noise pollution and congestion in the morning (Lagman, 2019), which may lead to lower prices when units are closer to schools. We nonetheless see that these factors are also contributing attributes to the valuation of HDB resale units.

6. Summary Statistics – Correlation Heatmap Of Attributes



After conducting visualizations on some key expected vairables, we can summarize our findings using Seaborn's pairplot function (see Notebook *4.1*) and Pandas' correlation method to visualize quantitatively, the relationships between attributes and our target variable, that is, resale price, while also analysing the matrix to identity potential attributes which may pose a potential multicollinearity problem in our subsequent model construction.

Upon inspection, we observe that floor area, remaining lease and storey level have higher positive correlations to resale price, with values **0.63, 0.32** and **0.37** respectively. These are expected as they are key primary determinants of property prices, as explained above. Price per m^2 shows the highest positive correlation to resale price at **0.69.** Distances to MRT and block number exhibit a weak correlation with resale price of **-0.12.** We also observe possible collinear attributes, some of which are expected. For instance, lease_commence_date and months_remaining_lease has a perfect positive correlation, as expected since the latter was derived from the former. The mean storey level and remaining lease also exhibit some positive linear correlation, which is expected, since newer projects by HDB are usually built taller due to increasing land constraints (Neo, 2020), as well as HDB's recent endavours to built impressive high-rise buildings such as *Pinnacle@Duxton*. Block numbers also show positive linear correlation to remaining lease due to HDB's relatively recent "3-digit numbering" system, which was only introduced in the 1970s (Teo, n.d).

7. Modelling Methodology

7.1 Assumptions

- **Linearity of Model**: Our models use a general linear form of $Y = b_0 + b_1x_1 + b_2x_2 + \cdots$ with variables transformed where appropriate to the model. We assume that the relationship between the attributes (transformed or otherwise) and the target variable is linear.
- **No Perfect Collinearity**: This is seen from the correlation matrix that no attribute is perfectly correlated with any other attribute (less derived attributes mentioned above which will be removed during model construction).

```
const                     71.078723
mean_storey_range          1.124972
block_number               1.156580
month_number               1.003549
dist_to_mrt                1.029977
dist_to_mall               1.110564
dist_to_sch                1.080683
floor_area_sqm             1.080240
months_remaining_lease     1.298170
dtype: float64
```

Checking for Multicollinearity

Multicollinearity may influence the interpretation of our coefficients in the final model (Frost, n.d). We investigate the possibility of multicollinearity in our model using *Variance Influence Factor (VIF)*. In practice, a VIF that exceeds 5 would be considered a problematic level of collinearity. We observe that all our numerical variables do not exhibit this level of collinearity. We will hence retain all these variables in our preliminary model construction.

7.2 Model Construction

In the preliminary model, we used the baseline model:

$$\begin{aligned}\boldsymbol{resale\ price} = {} & const + street\ name\ premium + town\ premium + flat\ model\ premium \\ & + flat\ type\ premium + month\ premium + region\ premium + \beta_{storey\ level}(storey\ level) \\ & + \beta_{floor\ area}(floor\ area) + \beta_{remaining\ lease}(remaining\ lease) + \beta_{block\ number}(block\ number) \\ & + \beta_{dist\ to\ mrt}(dist\ to\ MRT) + \beta_{dist\ to\ school}(dist\ to\ school) + \beta_{dist\ to\ mall}(dist\ to\ mall)\end{aligned}$$

```
==============================================
Adj. R-squared:                0.935
AIC:                           1635781.9548
BIC:                           1641022.3611
Log-Likelihood:                -8.1732e+05
F-statistic:                   1705.
Prob (F-statistic):            0.00
Scale:                         1.5361e+09
----------------------------------------------
```

7.2.1 Transformations

In the initial model, using Ordinary Least Squares method, we note a good fit of $Adjusted\ R^2 = 0.935$, and an AIC of 1635781.9548. We conjectured that this may be due to the large range of resale prices. We hence moved to transform the target variable.

```
=========================================
Adj. R-squared:              0.947
AIC:                         -158116.0819
BIC:                         -152875.6756
Log-Likelihood:              79632.
F-statistic:                 2122.
Prob (F-statistic):          0.00
Scale:                       0.0057073
```

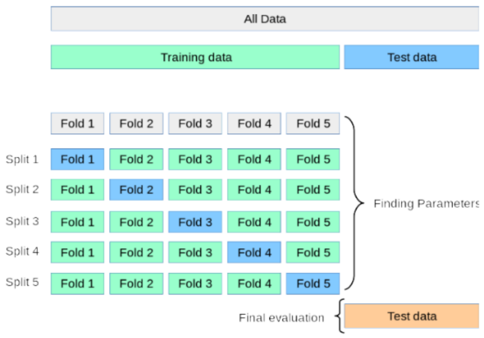
After applying a logarithmic transformation on resale price, our preliminary model's fit improved with a new $Adjusted\ R^2 = 0.947$, with a new AIC of -158116. We will hence use the natural log of resale prices in subsequent steps of model construction.

```
Retaining dist_to_sch
Retaining region
Retaining flat_model
Retaining month_number
Retaining floor_area_sqm
Retaining months_remaining_lease
Retaining block_number
Retaining dist_to_mrt
Retaining street_name
Retaining dist_to_mall
Retaining flat_type
Retaining town
Retaining mean_storey_range
Best model has an r-squared value of 0.947, AIC: -158116.08194722625 and Attributes: {'dist_to_sch', 'region', 'flat_model', 'm
onth_number', 'floor_area_sqm', 'months_remaining_lease', 'block_number', 'dist_to_mrt', 'street_name', 'dist_to_mall', 'flat_t
ype', 'town', 'mean_storey_range'}
```

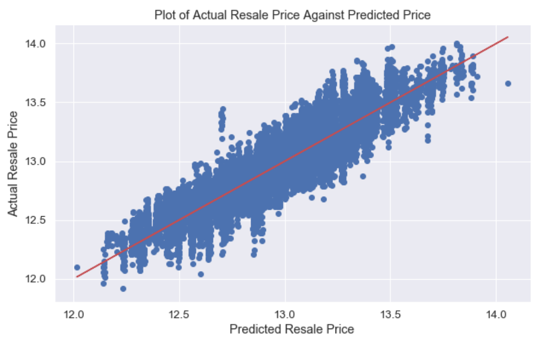
7.2.2 Attribute Selection – Backward Stepwise Regression

Using our current attribute set, there is a need to prevent the possibility of overfitting, which may occur when there are too many attributes such that the learned hypothesis from the model may fit the "training set" (here, all the data in our dataset) but cannot be generalized for uses outside of our dataset. We therefore use a backward stepwise regression method to remove unnecessary variables. We can compare the relative degree of over or underfitting by comparing the *Akaike Information Criterion (AIC)* between the candidate models during the iteration of backward stepwise regression. The smaller the number, the lesser the degree of overfitting. Of course, the $Adjusted\ R^2$ should not be compromised. We hence used both criteria in attribute selection (see Notebook *6.3*). This results in a final model with a $Adjusted\ R^2 = 0.947$. We will use the resulting model in our subsequent validation.

7.3 Model Validation – Train-Test Splits and Cross Validation



We use scikit-learn packages to conduct train-test splits, that is, to separate our dataset into "training data", which we will feed the model to "learn" , and use the "test data" to test how "accurate" the model is able to predict resale prices. As with standard practice, we used a 75-25% train-test split in this process. We also used cross validation to ensure that the results of our model are not obtained by chance of a good set of train-test data splits. This idea is best explained by this diagram retrieved from scikit-learn.



Cross Validation Results

The final model after using the training data returned an $Adjusted\ R^2 = 0.809712$, and a $Root\ mean\ squared\ error\ (RMSE) = 0.14211$. However, we also note that the final model included "street_name" as an attribute, which could be too specific for potential users of our model. We attempted to remove the attribute and retested the model using the Train-Test splits. There was a slight improvement in the $Adjusted\ R^2 = 0.809715$ and

$RMSE = 0.14211$. We hence have confidence in accepting this final model even if without the street name attribute, that is, the remaining attributes are sufficient for a reliable valuation of resale prices. To further verify this, we used 5-fold cross validation, which returned a similar $Adjusted\ R^2 = 0.80952$, $RMSE = 0.14213$. We can visualize the difference between the predicted resale price, after removing the street name attribute, and the empirical data in this scatter plot above, which shows a respectable predicted resale price compared to the empirical data from our dataset.

8. Final Predictive Model – Interpretations

$$\begin{aligned}\boldsymbol{ln(resale\ price)} = 10.4072 &+ town\ premium + flat\ model\ premium + flat\ type\ premium + region\ premium \\ &+ 0.0082 * storey\ level + 0.0081 * floor\ area + 0.0009 * months\ remaining\ lease \\ &- 0.0001 * (block\ number) - 0.0008 * (month\ number) - 0.07(dist\ to\ MRT, km) \\ &+ 0.0206(dist\ to\ school, km) - 0.0639 * (dist\ to\ mall, km)\end{aligned}$$

Or approximately,

$$\begin{aligned}\boldsymbol{Resale\ price} = -202196.1733 &+ town\ premium + flat\ model\ premium + flat\ type\ premium + region\ premium \\ &+ 4437.4498 * storey\ level + 3726.6492 * floor\ area + 403.3752 * months\ remaining\ lease \\ &- 51.0047 * (block\ number) - 253.9495 * (month\ number) - 32769.8752(dist\ to\ MRT, km) \\ &- 27985.1580 * (dist\ to\ mall, km) + 11665.3730(dist\ to\ school, km)\end{aligned}$$

The premiums for each categorical variable can be found in the appendix.

From our validated model, we can estimate that each floor up is worth approximately $4400 and the average. Convenience seems to be an important factor, with each km nearer to the MRT be worth around $32K and each km nearer to the mall to be worth $27K. Some towns are also valued more highly than others. For instance, Bedok is worth approximately a premium of approximately $10K (see appendix), while Bukit Batok enjoys a discount (or negative premium) of 16K. This is as expected due to different regions and districts having different prices as explain in our analysis above.

9. Concluding Remarks

While our model has tried, to the best of our ability, to incorporate salient factors of valuation, there may be other factors which affect the valuation, such as amenities and features the estate offers (Benson et al., 1998). For instance, having a scenic view amenity tends to add considerably residential property values, these includes amenity view to ocean, lake and mountains (Benson et al., 1998). More locally, unquantifiable factors may also affect valuation figures. For instance, *Feng Shui* could be a significant factor on residential prices (So, 2009). With a predominantly ethnically Chinese population, these investors may be influenced by their beliefs and traditions, including *Feng Shui*. Studies proclaim that bad *Feng Shui* may arise as a negotiating factor for buyers to negotiate for the final prices of units (So, 2009). Other factors such as the quality or reputation of nearby primary schools may also affect resale prices of nearby units. However, this data is not easily obtained. Existing interior designs by current owners and possible historical significance of certain units may also affect the value of a unit.

These factors, though important, are at this stage, hard to incorporate. Nonetheless, the current model provides a quantitative method to estimate a benchmark for an acceptable valuation to guide individual buyers or agents to make their investment or purchase decisions.

As a Real Estate Consulting firm aiming to provide consulting services to those who seek to buy or sell resale flats at the greatest worthiness in terms of its cost and price, it is imperative that we are able to predict reasonably, the value of a unit before offering appropriate advice to clients. As with all predictive models, it is important for users to be wary and cognizant of other peripheral factors affecting the value of real estate property. These include amenities and features offered by the HDB, 'Feng Shui', neighbourhood characteristics as well as ever-changing geopolitical landscape and changes in the Government's policies. We should also be cognizant of individual preferences which may cause individuals to value a property differently. Concluding, we should use our predictive models as the beginning of the valuation, not as an end.

10. References

Ang, R. (2019). New HDB flats to come with condo-like fittings. *The Straits Times.* Retrieved From:

https://www.straitstimes.com/singapore/housing/new-hdb-flats-to-come-with-condo-like-fittings

Chan, J. & Lim, J. (2019). 4 Questions About Owning A HDB Flat That You Probably Always Had In Your Mind. Retrieved From: https://dollarsandsense.sg/4-questions-owning-hbd-probably-always-mind/

Choo, C. (2019). $1 Million HDB Flats: Here's What You Need To Know. Retrieved From: https://blog.seedly.sg/1-millon-dollar-hdb-flats-singapore-heres-what-you-need-to-know/

Choo, C. (2019). HDB to build more new flats next year to meet greater demand. *Channel News Asia*. Retrieved From: https://www.channelnewsasia.com/news/business/hdb-to-build-more-new-flats-next-year-to-meet-greater-demand-12188602

Frost, J (n.d). Multicollinearity in Regression Analysis: Problems, Detection, and Solutions. Retrieved From: https://statisticsbyjim.com/regression/multicollinearity-in-regression-analysis/

Goh, R.B.H (2005). *Contours of Culture: Space and Social Difference in Singapore*. HK: Hong Kong University Press. Retrieved From: https://books.google.com.sg/books?id=ZFD1GxJRx98C&pg=PA150&lpg=PA150&dq=property+is+a+status+symbol+singapore&source=bl&ots=_HRkdB53e3&sig=ACfU3U32F0z9v08iuNBAXBchQIl8apURAg&hl=en&sa=X&ved=2ahUKEwiSy4DNxpnpAhV8wTgGHUyXAXMQ6AEwGXoECAoQAQ#v=onepage&q=property%20is%20a%20status%20symbol%20singapore&f=false

Hastie, T., James, G., Witten, D & Tibshirani, R. (2013). *An Introduction to Statistical Learning with Applications in R*. UK: Springer

HDB (2019). Key Statistics, HDB Annual Report 2018/2019. Retrieved From: https://services2.hdb.gov.sg/ebook/AR2019-keystats/html5/index.html?&locale=ENG&pn=9

HDB (n.d, 1) DIFFERING PRICES FOR BTO FLATS IN THE SAME TOWN. Retrieved From: https://www.hdb.gov.sg/cs/infoweb/hdbspeaks/differing-prices-for-bto-flats-in-the-same-town

HDB (n.d, 2). Do buyers of HDB flats own their flat? Retrieved From: (https://www.hdb.gov.sg/cs/infoweb/hdbspeaks/hdb-flat-buyers-own-their-flats)

HDB (n.d, 3). Types of Flats. Retrieved From: https://www.hdb.gov.sg/cs/infoweb/residential/buying-a-flat/resale/types-of-flats

HDB (n.d, 4). Sales Launch. Retrieved From: https://esales.hdb.gov.sg/bp25/launch/20feb/bto/20FEBBTOSB_page_0687/pricing.html

Lagman, M (2019). Dear Parents, Is Living Near Your Child's School Really Worth It? Retrieved From: https://www.99.co/blog/singapore/dear-parents-is-living-near-your-childs-school-really-worth-it/

Lee, P. (2016). More Singaporeans take bus, MRT to work: Government survey. *The Straits Times*. Retrieved from https://www.straitstimes.com/singapore/more-singaporeans-take-bus-mrt-to-work-government-survey

MOE (n.d). Understand how balloting works. Retrieved From: https://beta.moe.gov.sg/primary/p1-registration/understand-balloting/

Navaratnaragjah, R (2015). The MRT Effect: How It Will Affect Your Property's Value. Retrieved From: https://www.propertyguru.com.sg/property-management-news/2015/2/85301/the-mrt-effect-how-it-will-affect-your-propertys-value

Neo, X. (2020). Evolution of HDB designs. *The Straits Times*. Retrieved From: https://www.straitstimes.com/singapore/housing/evolution-of-hdb-designs

Ong, K.S (2017). Why are DBSS flats achieving high prices in the HDB resale market? Retrieved From: https://www.99.co/blog/singapore/dbss-flats-high-prices-hdb-resale/

Orange Tee (2020). HDB Market Pulse – Real Estate Data Trend Q4 2019. Retrieved From: https://blog.orangetee.com/market-analysis-news/hdb-market-pulse-real-estate-data-trend-q4-2019/

Paulo, A.D & Low, M. (2018). Class – not race nor religion – is potentially Singapore's most divisive fault line. *Channel News Asia*. Retrieved From: https://www.channelnewsasia.com/news/cnainsider/regardless-class-race-religion-survey-singapore-income-divide-10774682

So, C. F. (2009). An examination of the Effect of Feng Shui on residential property price in HongKong, HongKong: The University of HongKong.

Teo, A. (n.d). HDB Interesting facts and records. Retrieved From: https://www.teoalida.com/singapore/hdbrecords/

Wong, P. (2018). The Big Read: No easy answers to HDB lease decay issue, but public mindset has to change first. *Channel News Asia* Retrieved from https://www.channelnewsasia.com/news/singapore/big-read-hdb-lease-decay-public-mindset-change-homeownership-10361572

Yong, C. (2020). More HDB resale flats sold in March, prices down 0.3%. *The Straits* Times. Retrieved From: https://www.straitstimes.com/singapore/more-hdb-resale-flats-sold-in-march-prices-down-03

Appendix

Ordinary Least Squares Regression Model Report, Target Variable: np.log(resale_price)

Results: Ordinary least squares

Model:	OLS	Adj. R-squared:	0.906
Dependent Variable:	np.log(resale_price)	AIC:	-119379.4787
Date:	2020-05-08 03:20	BIC:	-118849.9603
No. Observations:	68161	Log-Likelihood:	59748.
Df Model:	57	F-statistic:	1.147e+04
Df Residuals:	68103	Prob (F-statistic):	0.00
R-squared:	0.906	Scale:	0.010151

	Coef.	Std.Err.	t	P>|t|	[0.025	0.975]
Intercept	10.4072	0.0684	152.1924	0.0000	10.2732	10.5412
C(town)[T.BEDOK]	0.2430	0.0021	116.9297	0.0000	0.2390	0.2471
C(town)[T.BISHAN]	0.9216	0.0062	148.6568	0.0000	0.9095	0.9338
C(town)[T.BUKIT BATOK]	0.0840	0.0020	41.0248	0.0000	0.0799	0.0880
C(town)[T.BUKIT MERAH]	0.9214	0.0059	157.0992	0.0000	0.9099	0.9329
C(town)[T.BUKIT PANJANG]	-0.0098	0.0021	-4.7341	0.0000	-0.0138	-0.0057
C(town)[T.BUKIT TIMAH]	1.0710	0.0087	122.4992	0.0000	1.0539	1.0882
C(town)[T.CENTRAL AREA]	0.9725	0.0074	131.5329	0.0000	0.9580	0.9870
C(town)[T.CHOA CHU KANG]	-0.1280	0.0020	-63.4738	0.0000	-0.1320	-0.1241
C(town)[T.CLEMENTI]	0.3522	0.0026	137.8339	0.0000	0.3472	0.3572
C(town)[T.GEYLANG]	0.7870	0.0061	129.1071	0.0000	0.7750	0.7989
C(town)[T.HOUGANG]	-0.1596	0.0027	-58.9058	0.0000	-0.1649	-0.1543
C(town)[T.JURONG EAST]	0.1664	0.0025	65.5025	0.0000	0.1614	0.1713
C(town)[T.JURONG WEST]	0.0044	0.0017	2.6598	0.0078	0.0012	0.0077
C(town)[T.KALLANG/WHAMPOA]	0.8320	0.0060	138.5702	0.0000	0.8203	0.8438
C(town)[T.MARINE PARADE]	1.1631	0.0074	158.0910	0.0000	1.1487	1.1776
C(town)[T.PASIR RIS]	0.0682	0.0024	28.7987	0.0000	0.0636	0.0728
C(town)[T.PUNGGOL]	-0.2355	0.0029	-80.7878	0.0000	-0.2412	-0.2298
C(town)[T.QUEENSTOWN]	0.9162	0.0061	150.9034	0.0000	0.9043	0.9281
C(town)[T.SEMBAWANG]	0.0235	0.0024	9.9964	0.0000	0.0189	0.0282
C(town)[T.SENGKANG]	-0.3132	0.0027	-115.0341	0.0000	-0.3185	-0.3078
C(town)[T.SERANGOON]	0.0153	0.0034	4.5208	0.0000	0.0086	0.0219
C(town)[T.TAMPINES]	0.1880	0.0020	95.3023	0.0000	0.1842	0.1919
C(town)[T.TOA PAYOH]	0.7957	0.0059	133.7963	0.0000	0.7840	0.8073
C(town)[T.WOODLANDS]	0.0716	0.0019	37.2004	0.0000	0.0679	0.0754
C(town)[T.YISHUN]	0.2130	0.0020	108.1198	0.0000	0.2092	0.2169
C(flat_type)[T.2 ROOM]	0.1000	0.0189	5.3015	0.0000	0.0630	0.1370
C(flat_type)[T.3 ROOM]	0.2511	0.0187	13.4021	0.0000	0.2144	0.2878
C(flat_type)[T.4 ROOM]	0.3327	0.0191	17.4052	0.0000	0.2952	0.3701
C(flat_type)[T.5 ROOM]	0.3534	0.0197	17.9621	0.0000	0.3148	0.3919
C(flat_type)[T.EXECUTIVE]	0.3467	0.0204	16.9729	0.0000	0.3067	0.3867
C(flat_type)[T.MULTI-GENERATION]	0.2873	0.0381	7.5325	0.0000	0.2126	0.3621
C(flat_model)[T.Adjoined flat]	0.1527	0.0719	2.1234	0.0337	0.0118	0.2937
C(flat_model)[T.Apartment]	0.1330	0.0715	1.8601	0.0629	-0.0071	0.2731
C(flat_model)[T.DBSS]	0.2202	0.0714	3.0831	0.0020	0.0802	0.3602
C(flat_model)[T.Improved]	0.0632	0.0713	0.8860	0.3756	-0.0766	0.2031
C(flat_model)[T.Improved-Maisonette]	0.3812	0.0767	4.9722	0.0000	0.2309	0.5314
C(flat_model)[T.Maisonette]	0.1746	0.0715	2.4416	0.0146	0.0344	0.3147
C(flat_model)[T.Model A]	0.0684	0.0713	0.9585	0.3378	-0.0714	0.2082
C(flat_model)[T.Model A-Maisonette]	0.2419	0.0720	3.3595	0.0008	0.1008	0.3830
C(flat_model)[T.Model A2]	0.0609	0.0714	0.8534	0.3934	-0.0790	0.2009

C(flat_model)[T.Multi Generation]	0.2873	0.0381	7.5325	0.0000	0.2126	0.3621
C(flat_model)[T.New Generation]	0.0996	0.0714	1.3957	0.1628	-0.0403	0.2395
C(flat_model)[T.Premium Apartment]	0.0970	0.0713	1.3596	0.1740	-0.0428	0.2368
C(flat_model)[T.Premium Apartment Loft]	0.1888	0.0798	2.3672	0.0179	0.0325	0.3452
C(flat_model)[T.Premium Maisonette]	0.0536	0.0810	0.6620	0.5080	-0.1051	0.2124
C(flat_model)[T.Simplified]	0.0951	0.0714	1.3317	0.1830	-0.0448	0.2350
C(flat_model)[T.Standard]	0.0764	0.0714	1.0706	0.2843	-0.0635	0.2164
C(flat_model)[T.Terrace]	0.7305	0.0732	9.9825	0.0000	0.5871	0.8739
C(flat_model)[T.Type S1]	0.1439	0.0721	1.9945	0.0461	0.0025	0.2852
C(flat_model)[T.Type S2]	0.1324	0.0727	1.8217	0.0685	-0.0101	0.2748
C(region)[T.EAST]	0.4993	0.0043	115.9612	0.0000	0.4908	0.5077
C(region)[T.NORTH]	0.3082	0.0043	71.4226	0.0000	0.2998	0.3167
C(region)[T.NORTH-EAST]	0.7499	0.0059	126.1594	0.0000	0.7383	0.7616
C(region)[T.WEST]	0.4692	0.0049	96.0172	0.0000	0.4596	0.4788
floor_area_sqm	0.0081	0.0001	113.1330	0.0000	0.0080	0.0083
months_remaining_lease	0.0009	0.0000	182.5345	0.0000	0.0009	0.0009
mean_storey_range	0.0082	0.0001	109.2927	0.0000	0.0081	0.0084
block_number	-0.0001	0.0000	-39.9598	0.0000	-0.0001	-0.0001
month_number	-0.0008	0.0001	-6.6485	0.0000	-0.0010	-0.0005
dist_to_mrt	-0.0700	0.0008	-87.0537	0.0000	-0.0716	-0.0685
dist_to_mall	-0.0639	0.0013	-51.0544	0.0000	-0.0664	-0.0615
dist_to_sch	0.0206	0.0024	8.6123	0.0000	0.0159	0.0253

Omnibus:	1224.409	Durbin-Watson:	1.195
Prob(Omnibus):	0.000	Jarque-Bera (JB):	2056.949
Skew:	0.160	Prob(JB):	0.000
Kurtosis:	3.788	Condition No.:	10189999488328168

* The condition number is large (1e+16). This might indicate strong multicollinearity or other numerical problems.

Ordinary Least Squares Regression Model Report, Target Variable: resale_price

Results: Ordinary least squares

Model:	OLS	Adj. R-squared:	0.884
Dependent Variable:	resale_price	AIC:	1674355.4701
Date:	2020-05-08 03:22	BIC:	1674884.9886
No. Observations:	68161	Log-Likelihood:	-8.3712e+05
Df Model:	57	F-statistic:	9134.
Df Residuals:	68103	Prob (F-statistic):	0.00
R-squared:	0.884	Scale:	2.7257e+09

	Coef.	Std.Err.	t	P>|t|	[0.025	0.975]
Intercept	-202196.1733	35433.9362	-5.7063	0.0000	-271646.6465	-132745.7002
C(town)[T.BEDOK]	13620.6741	1077.0656	12.6461	0.0000	11509.6268	15731.7215
C(town)[T.BISHAN]	32671.2206	3212.4860	10.1701	0.0000	26374.7518	38967.6893
C(town)[T.BUKIT BATOK]	-16907.6449	1060.4560	-15.9437	0.0000	-18986.1374	-14829.1523
C(town)[T.BUKIT MERAH]	25625.3302	3039.0358	8.4321	0.0000	19668.8237	31581.8368
C(town)[T.BUKIT PANJANG]	-68863.5048	1068.5106	-64.4481	0.0000	-70957.7843	-66769.2254
C(town)[T.BUKIT TIMAH]	119390.0984	4530.4436	26.3528	0.0000	110510.4344	128269.7624
C(town)[T.CENTRAL AREA]	14018.2740	3831.2728	3.6589	0.0003	6508.9837	21527.5642
C(town)[T.CHOA CHU KANG]	-119643.2264	1045.0592	-114.4846	0.0000	-121691.5412	-117594.9115
C(town)[T.CLEMENTI]	109487.8271	1324.1493	82.6854	0.0000	106892.4960	112083.1582
C(town)[T.GEYLANG]	-39432.9943	3158.6369	-12.4842	0.0000	-45623.9188	-33242.0698
C(town)[T.HOUGANG]	-79197.0279	1403.9957	-56.4083	0.0000	-81948.8577	-76445.1980
C(town)[T.JURONG EAST]	14791.7377	1316.0696	11.2393	0.0000	12212.2428	17371.2326
C(town)[T.JURONG WEST]	-54806.4242	861.8989	-63.5880	0.0000	-56495.7450	-53117.1034
C(town)[T.KALLANG/WHAMPOA]	-25402.3491	3111.3923	-8.1643	0.0000	-31500.6744	-19304.0239
C(town)[T.MARINE PARADE]	119576.6203	3812.4296	31.3649	0.0000	112104.2627	127048.9779
C(town)[T.PASIR RIS]	-75610.3909	1227.0654	-61.6189	0.0000	-78015.4378	-73205.3441
C(town)[T.PUNGGOL]	-125668.4287	1510.5328	-83.1948	0.0000	-128629.0712	-122707.7861
C(town)[T.QUEENSTOWN]	23472.9942	3146.2210	7.4607	0.0000	17306.4046	29639.5837
C(town)[T.SEMBAWANG]	-97580.3668	1220.3497	-79.9610	0.0000	-99972.2508	-95188.4828
C(town)[T.SENGKANG]	-162803.8325	1410.7059	-115.4059	0.0000	-165568.8145	-160038.8506
C(town)[T.SERANGOON]	1630.4105	1748.9870	0.9322	0.3512	-1797.6020	5058.4230
C(town)[T.TAMPINES]	-20411.7010	1022.3783	-19.9649	0.0000	-22415.5613	-18407.8407
C(town)[T.TOA PAYOH]	-28706.9889	3081.4785	-9.3160	0.0000	-34746.6831	-22667.2947
C(town)[T.WOODLANDS]	-66057.1564	997.8192	-66.2015	0.0000	-68012.8808	-64101.4320
C(town)[T.YISHUN]	-1592.8498	1021.0023	-1.5601	0.1187	-3594.0130	408.3134
C(flat_type)[T.2 ROOM]	3840.9663	9777.1295	0.3929	0.6944	-15322.1959	23004.1285
C(flat_type)[T.3 ROOM]	25495.6771	9709.2343	2.6259	0.0086	6465.5892	44525.7649
C(flat_type)[T.4 ROOM]	39282.8169	9903.8492	3.9664	0.0001	19871.2840	58694.3497
C(flat_type)[T.5 ROOM]	53030.0603	10194.0649	5.2021	0.0000	33049.7052	73010.4155
C(flat_type)[T.EXECUTIVE]	50899.8275	10584.2289	4.8090	0.0000	30154.7513	71644.9037
C(flat_type)[T.MULTI-GENERATION]	107669.0068	19764.7906	5.4475	0.0000	68930.0406	146407.9731
C(flat_model)[T.Adjoined flat]	81466.1652	37263.8556	2.1862	0.0288	8429.0522	154503.2783
C(flat_model)[T.Apartment]	60869.8084	37039.3376	1.6434	0.1003	-11727.2497	133466.8664
C(flat_model)[T.DBSS]	142741.2368	37008.4061	3.8570	0.0001	70204.8045	215277.6690

```
C(flat_model)[T.Improved]                 16321.6452 36971.7402    0.4415 0.6589  -56142.9219   88786.2124
C(flat_model)[T.Improved-Maisonette]     187821.4947 39723.9074    4.7282 0.0000  109962.6831  265680.3063
C(flat_model)[T.Maisonette]               94247.0324 37047.2947    2.5440 0.0110   21634.3786  166859.6862
C(flat_model)[T.Model A]                  18587.9631 36964.7010    0.5029 0.6151  -53862.8072   91038.7333
C(flat_model)[T.Model A-Maisonette]      138487.9449 37306.7499    3.7121 0.0002   65366.7591  211609.1306
C(flat_model)[T.Model A2]                 26008.7809 37006.7469    0.7028 0.4822  -46524.3992   98541.9611
C(flat_model)[T.Multi Generation]        107669.0068 19764.7906    5.4475 0.0000   68930.0406  146407.9731
C(flat_model)[T.New Generation]           32903.6806 36976.1209    0.8899 0.3735  -39569.4727  105376.8340
C(flat_model)[T.Premium Apartment]        29803.4517 36968.8603    0.8062 0.4201  -42655.4708  102262.3742
C(flat_model)[T.Premium Apartment Loft]  176282.4400 41335.1534    4.2647 0.0000   95265.5882  257299.2918
C(flat_model)[T.Premium Maisonette]       71202.1659 41973.7185    1.6964 0.0898  -11066.2729  153470.6046
C(flat_model)[T.Simplified]               38034.7056 36986.9235    1.0283 0.3038  -34459.6207  110529.0320
C(flat_model)[T.Standard]                 30234.8742 36992.0087    0.8173 0.4137  -42269.4193  102739.1676
C(flat_model)[T.Terrace]                 373097.0005 37919.9502    9.8391 0.0000  298773.9429  447420.0580
C(flat_model)[T.Type S1]                 188682.4312 37377.2069    5.0481 0.0000  115423.1499  261941.7126
C(flat_model)[T.Type S2]                 232502.1731 37653.6828    6.1748 0.0000  158700.9993  306303.3469
C(region)[T.EAST]                        -82401.4178  2231.0386  -36.9341 0.0000  -86774.2508  -78028.5848
C(region)[T.NORTH]                      -165230.3730  2236.1107  -73.8919 0.0000 -169613.1474 -160847.5986
C(region)[T.NORTH-EAST]                  -59835.3523  3080.2548  -19.4255 0.0000  -65872.6480  -53798.0566
C(region)[T.WEST]                       -135941.2355  2532.0842  -53.6875 0.0000 -140904.1175 -130978.3534
floor_area_sqm                             3726.6492    37.2567  100.0263 0.0000    3653.6262    3799.6723
months_remaining_lease                      403.3752     2.4757  162.9317 0.0000     398.5227     408.2276
mean_storey_range                          4437.4498    39.0690  113.5798 0.0000    4360.8746    4514.0250
block_number                                -51.0047     1.0572  -48.2432 0.0000     -53.0769     -48.9325
month_number                               -253.9495    58.8779   -4.3132 0.0000    -369.3501    -138.5488
dist_to_mrt                              -32769.8752   416.8423  -78.6146 0.0000  -33586.8855  -31952.8648
dist_to_mall                             -27985.1580   648.5906  -43.1476 0.0000  -29256.3948  -26713.9212
dist_to_sch                               11665.3730  1238.4814    9.4191 0.0000    9237.9509   14092.7950
---------------------------------------------------------------------------------------------------------
Omnibus:                        5053.864            Durbin-Watson:                      1.105
Prob(Omnibus):                  0.000               Jarque-Bera (JB):                   7962.797
Skew:                           0.586               Prob(JB):                           0.000
Kurtosis:                       4.195               Condition No.:                      10189999488328168
=========================================================================================================
* The condition number is large (1e+16). This might indicate          strong multicollinearity or other
numerical problems.
```