HDB Resale Analysis for Young Couples

A comprehensive and interactive look into pricing trends in Singapore over the past decades

Group Number: 21
Project Members:
Chen Tingting e0403910@u.nus.edu
Jacob Josiah Santiago Alvarez e0545170@u.nus.edu
Leong Zhong Wei Nicholas e0673944@u.nus.edu

Project Objective

This project aims to provide accessible visualization and valuable insights into the HDB resale market for prospective buyers to understand what truly drives the price of resale housing in Singapore over the past 20 years. We wish to tease apart the various factors determining price trends in the resale market and allow prospective buyers and sellers to estimate a flat's value.

Compared to other existing housing visualizations that already exist, we are targeting young couples (age between 25-35) such as ourselves who wish to quickly get a house and start a family given the recent Covid-19 crisis delaying BTO flats. This target group is stuck between a rock and a hard place with BTO flats being delayed and not having enough income to purchase a condominium. Thus, the best option they have is to purchase a flat from the resale market.

Although numerous articles and databases exist exploring the resale market in Singapore, most lack the ability for a user to query the data without having the requisite programming skills. In addition to discovering and annotating trends in the resale market, we provide an interactive platform for users to search or explore potential areas to purchase, along with the ability to query and compare resale flats across the island.

In addition to analyzing just flat data, young couples interested in resale flats would likely be interested in the school landscape around the area, hence we have also obtained data of various schools in Singapore as a supplement to their flat choice. In Singapore, balloting for a good primary school is often enough justification for parents to buy a particular flat. This is mainly because of how the registration process works. If your chosen primary school has more registrants than vacancies, priority admission will be given based on your child's citizenship and the home-to-school distance (which is in turn based on the address used for registration).

Priority admission is given in this order:

1. Singapore Citizens (SC) living within 1km of the school.
2. SCs living between 1km and 2km of the school.
3. SCs living outside 2km of the school.
4. Permanent Residents (PR) living within 1km of the school.
5. PRs living between 1km and 2km of the school.
6. PRs living outside 2km of the school.

Queries

Based off what we understand from a young couple picking a resale flat, a list of priorities and queries would be:

- Typical flat considerations such as number of rooms, floor, size and how they affect price. This group would be more concerned with value for money purchases since they start off with lower income and total assets. They may also be concerned with value retention of the flat or information about what factors affect price. These are answered in the tableau stories showcasing our top insights from the housing market.
  - What is the overarching principle to understand about what drives resale price?
  - In general, how much would a 3-room, 4-room or 5-room flat cost?
  - Are newer flat sizes getting smaller?
  - Are higher floors significantly more expensive?
  - Which towns are markedly underpriced compared to others?
  - Which areas should I pick? Which are the most expensive or cheapest areas for resale purchases?
  - Can we determine the influence of each factor on the final price?
- Queries for schools surrounding the chosen flat area or flats around chosen schools for their child. These are answered in the exploratory dashboard.
  - Which areas are the best value for a chosen school while still guaranteeing my priority admission? Based on my former school, how do the flat prices look like around the area?
  - In a particular chosen town, what kind of schools are available?
  - What do the historical transactions look like for my chosen flat?
  - I only have a certain budget / want a certain lease to remaining / want a certain amount of floor size / want to only look at recent transactions / only want a particular flat type. What are my options available?

Datasets Used

Data Sources

Main Dataset - *final_df.csv*	
HDB resale transactions data	https://data.gov.sg/dataset/resale-flat-prices
As the original HDB resale dataset does not contain postal code and longitude/latitude information, we made use of the OneMap API to extract them based on ‘block’ and ‘street_name’	https://www.onemap.gov.sg/docs/
Complementary Datasets - *schools.csv*	
Many families with young children also use schools in the vicinity as a filter for their house search	https://data.gov.sg/dataset/school-directory-and-information
As the original school’s information dataset does not contain longitude/latitude information, we made use of the OneMap API to extract them based on ‘postal’.	https://www.onemap.gov.sg/docs/

Data Attributes

final_df.csv - a combined dataset of HDB resale data with spatial data from onemap.gov.sg

Attribute / Alias	Attribute Type	Attribute Description
_id	Ordinal	Unique transaction ID from the respective datasets
town / Town	Categorical	Name of the town (e.g., Ang Mo Kio)
flat_type / Flat Type	Categorical	The type of flat (e.g., 4-room flat)
floor_area_sqm / Floor Area Sqm	Quantitative	Size of the flat in square meters
block / Block	Categorical	Block number of the flat
street_name / Street Name	Categorical	Street name of the flat (e.g., Ang Mo Kio Ave 1)
resale_price / Resale Price	Quantitative	Price the unit was sold at in SGD
remaining_lease / Remaining Lease	Categorical	The number of years and months left on the lease at time of sale
lease_commence_date / Lease Commence Date	Ordinal	The start year of lease of the flat
storey_range / Storey Range	Categorical	The 3-storey range of the unit sold (e.g., 10 TO 12)
computed_remaining_lease / Computed Remaining Lease (computed via python)	Discrete	The number of months left on the lease at time of sale
year_sold / Year Sold (computed via python)	Ordinal	The year of sale in YYYY
month_sold / Month Sold (computed via python)	Ordinal	The month of sale in MM
postal / Postal (OneMap)	Categorical	Postal code of unit
LATITUDE / Latitude (OneMap)	Quantitative	Latitude
LONGITUDE / Longitude (OneMap)	Quantitative	Longitude

schools.csv - a combined dataset of HDB resale data with spatial data from onemap.gov.sg

Attribute / Alias	Attribute Type	Attribute Description
school_name / School Name	Categorical	Name of the school
gifted_ind / Gifted Ind	Categorical	Gifted Programme Status of the school
autonomous_ind / Autonomous Ind	Categorical	Autonomous/Independent School Status of the school

ip_ind / Ip_Ind	Categorical	Integrated Programme Status of the school
dgp_code / Area	Categorical	Town in which the school is located (e.g., Ang Mo Kio)
mainlevel_code / Level	Categorical	Educational level of school (e.g., Primary, Secondary, Mixed etc.)
nature_code / Type	Categorical	The nature of the school (e.g., Boys school, Mixed School etc.)
missionstatement_desc / Mission Statement	Categorical	Mission statement of the school
visionstatement_desc / Vision Statement	Categorical	Vision statement of the school
url_address / Website	Categorical	Website of the school
email_address / Email Address	Categorical	Email address to contact the school
Postal	Categorical	Postal code of the school
latitude / Latitude1 (OneMap)	Quantitative	Latitude of the school
longtitude / Longtitude (OneMap)	Quantitative	Longitude of the school

Data Processing

Both the HDB and school's data were first imported from data.gov.sg using API through python. Manual computations using python were also done for more features such as extracting the remaining number of years left on the lease, year sold, and month sold separately. Next, we made use of python to extract the spatial data (postal code, longitude and latitude) for schools and HDB resale price from the OneMap API. Finally, we merged both data.gov.sg and OneMap API data to obtain the CSV files and upload them onto Tableau.

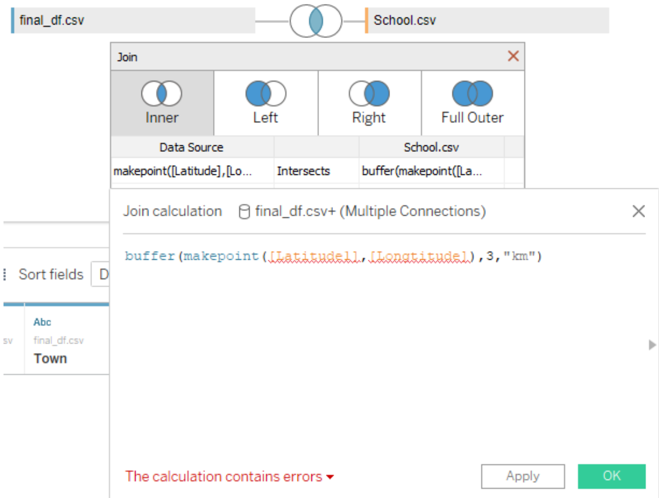
```
# Returns OneMap data based on block_address field and a list of addresses with errors

def get_postal_code(addresslist, extract='n', savefile='onemapresults.csv'):
    # Onemap API server
    SERVER = 'https://developers.onemap.sg/commonapi/search?searchVal='

    # Required variables for API call
    VARIABLES = '&returnGeom=Y&getAddrDetails=Y'

    onemap_results = pd.DataFrame()
    errors = []

    # For loop to pull and append detailed address data including postal code based on HDB street address
    for address in addresslist:
        try:
            url = SERVER + address + VARIABLES
            resp = requests.get(url)
            result = resp.json()
            result = result['results']
            result[0]['SEARCH']=address
            onemap_results = onemap_results.append(result)
        except:
            errors.append(address)
            print('Error with ' + address)

    # Some OneMap results have multiple entries so we need to fill in the blanks as well
    onemap_results = onemap_results.fillna(method='ffill')

    if extract == 'y':
        onemap_results.to_csv(savefile)
    else:
        pass

    # Checksum to make sure all values have been put through the API
    if len(addresslist)-len(errors) == 0:
        print('Dataframe is empty')
    elif len(addresslist)-len(onemap_results.SEARCH.unique())-len(errors) == 0:
        print('All values have been put through the API')
    else:
        print('Not all values have been put through the API')

    return onemap_results, errors
```

To make the hdb database and school database appear together in the same visual in Tableau, we used an inner join method in Tableau with the following method:

- For each school location: *makepoint(latitude,longtitude) - schools.csv*
- And each HDB location: *makepoint(latitude,longtitude) - final_df.csv*

- We made both of them intersect within 3km with the buffer formula.

The result is that whenever we view any school on the map, we are able to filter and view houses within 0.5-3km of the range based on the Tableau parameter we have defined “Distance from School (KM)”.



Certain categories of data were also put into their hierarchical order:
final_df.csv - Town, Street, Block
final_df.csv - Year Sold, Month Sold
schools.csv - Town, School Name

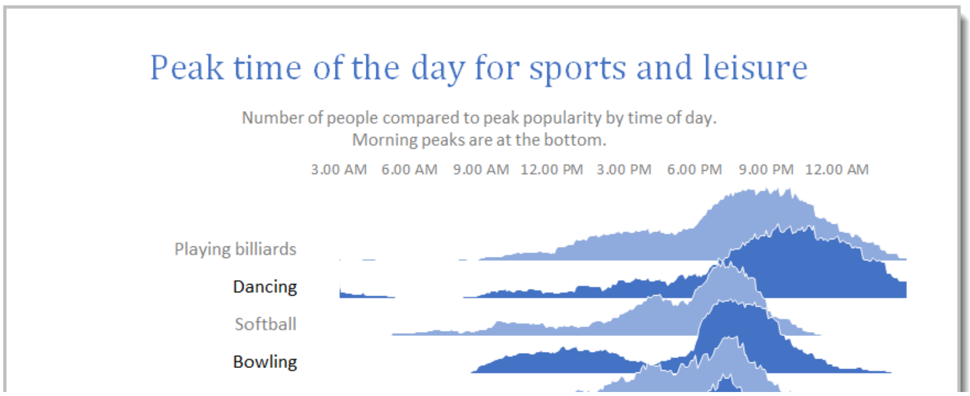
To display certain visuals, custom calculations were also created in Tableau such as:
SchoolLocation: An actual coordinate on the map of a school, *makepoint(school latitude, school longitude)*
HouseLocation: An actual coordinate on the map of a house, *makepoint(hdb latitude, hdb longitude)*
Distance from School (KM): a parameter between 0.5 and 3km in 0.5km steps for how far to search from a school.
DistanceSchoolHouse: the calculated distance between a *HouseLocation* and *SchoolLocation*
HouseInDistance: a filter for whether a house was within the *DistanceSchoolHouse* defined.
AUT IF/GIFTED IF/IP IF: Converting “Yes/No” data into “1/-1” to format into shapes for visuals.
Independent School? /Gifted Program? /Integrated Program? Getting the sign of the IFs to format into shapes for visuals

```r
library(tidyverse)
packages <- c("ggplot2", "mgcv", "magrittr", "vtreat", "MASS", "statmod",
              "dplyr", "tidyr")
sapply(packages, library, character.only = TRUE, logical.return = TRUE)

dat <- read_csv(
  "icloud/NUS_PhD/CS5346/Project/jacob/data/updated_final_df.csv"
) %>% select(-X1)

range01 <- function(x){(x-min(x))/(max(x)-min(x))}
t <- dat %>%
  dplyr::select(-flat_model,-street_name,-lease_commence_date,
         -remaining_lease,-block,-block_address,
                    -Blend,-`_id`,-month) %>%
  mutate(storey=as.integer(str_extract(storey_range,'[0-9]+')),
         storey=range01(storey),
         town=as.factor(town))

t$price <- t$resale_price/1000
property <- t %>% mutate_if(is_character,as_factor)
property_dens <- data.frame(price = t$price)
m <- mean(t$price)
v <- var(t$price)

gamma_param <- fitdistr(x = t$price, densfun = "gamma")[[1]]
invgauss_param <- fitdistr(x = t$price, densfun = dinvgauss,
                           start = list(mean = m, shape = m ^ 3 / v))[[1]]

property_dens <- property_dens %>%
  mutate(
    gamma = dgamma(x = price, gamma_param[1], gamma_param[2]),
    invgauss = dinvgauss(x = price, invgauss_param[1], invgauss_param[2])
  ) %>%
  gather(
    key = distr, value = dens, gamma, invgauss
  )
```

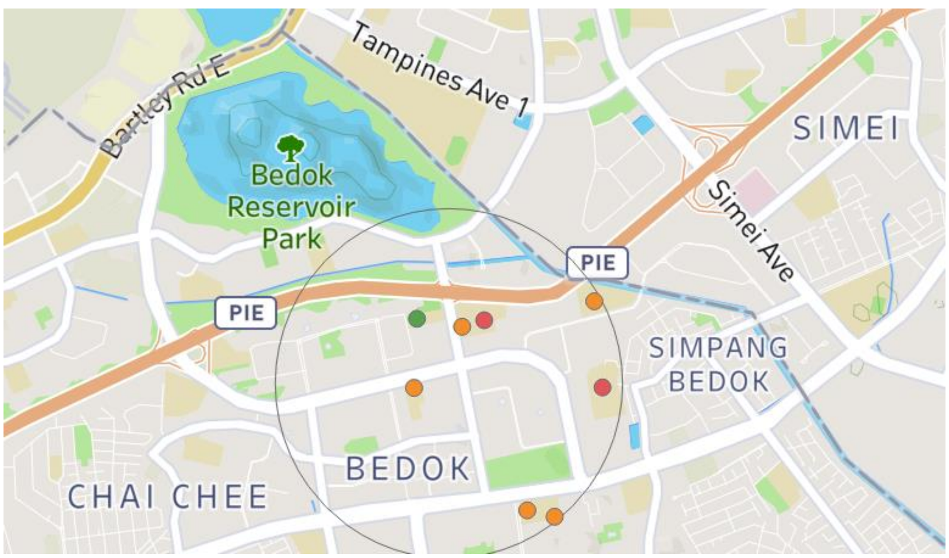
To carry out Generalized Additive Modelling in R, the final_df.csv was read in using the function *reader::read_csv*, and pre-processing was done to convert storey into a standardized numerical value using the scaling function *range01*, and all character type columns into factors. Prices were divided by 1000 to improve convergence of the model, and a gamma distribution was fit to the price distribution using *stats::fitdistr*.

Visualization Ideas

The main theme of the project was to discover trends in resale pricing over the years, thus the choices of visualizations needed to best portray changes over time. Line graphs, area graphs and ridgeline plots were considered the best way to show trends over time.



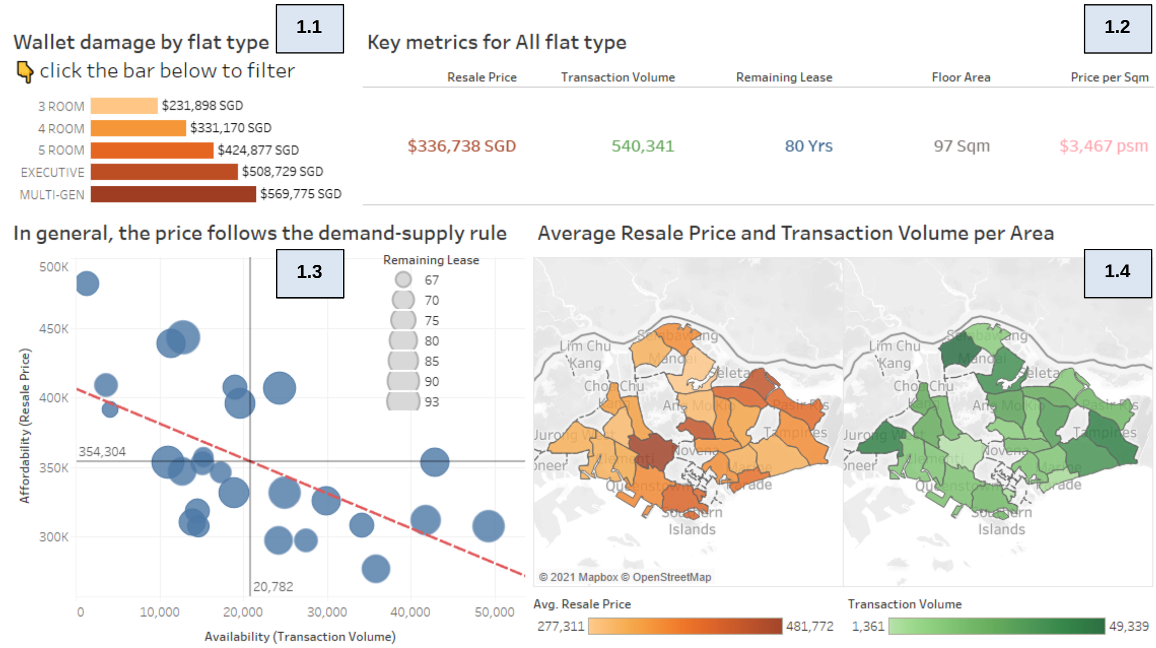
Information was also very spatially rich, which were best shown with maps. This required the location information of the townships and the units, which for this project was embedded in their latitude and longitude locations.



Lastly, relationships between multiple factors were of high interest. Dot plots were the best way to do this for continuous variables, while heatmaps also served an important role in showcasing price differences between two categorical variables. Dot plots also allow extended visual encoding of additional information such as remaining lease of flats in the area – important with highly complex and nested data.

Visualizing Trends in the Resale Market

Visual 1 A General Overview



Visual 1.1
A colored bar chart to encode the average resale price based on the different flat types.

Visual 1.2
A simple text visualization to highlight key metrics of the historical data. Throughout the visualization we make use of these hues to showcase the different factors: resale price, transaction volume, remaining lease, floor area, price per sqm.

Visual 1.3
A bubble chart is used to highlight the general trend of increasing supply against decreasing price along with a linear regression line. This plot is made at the detail level of each town. Size is used to encode the remaining lease for each town.

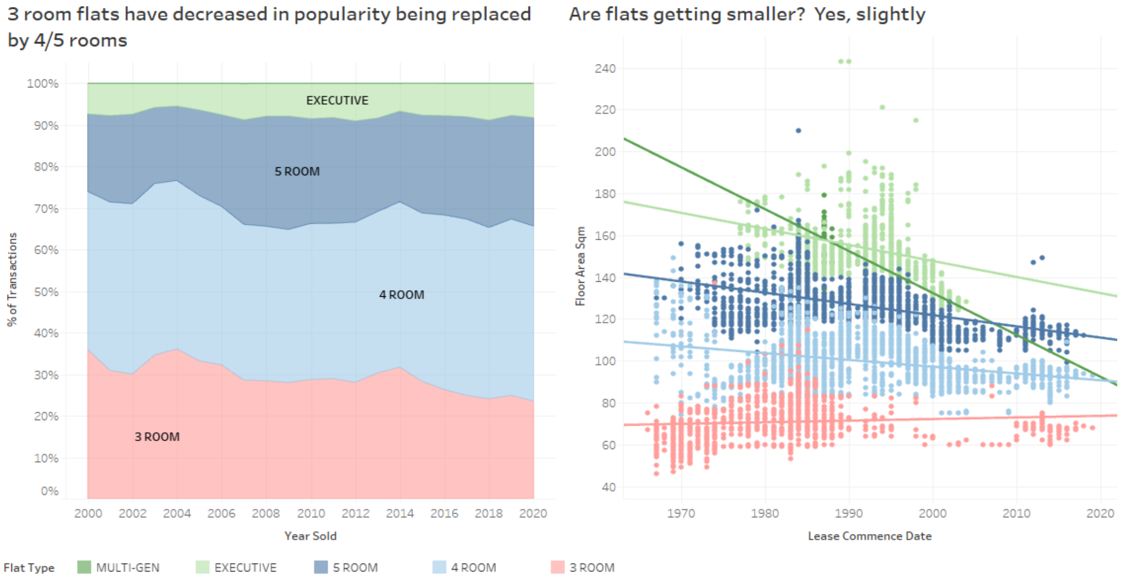
Visual 1.4
A spatial representation of visual 1.3. A geographical heat map using saturation to show average resale prices and transaction volume of each region encoded by position. Darker regions with higher resale prices also tend to have the lowest transaction volume around the central region of Singapore. Places with more transactions near the outskirts of Singapore tend to have lower average prices.

Initially we had wanted to break down the flats across Singapore individually into blocks instead of towns. However, we realized there was too much noise in 1 visual that prevented us from making any valuable insights. Instead, we give this level of detail in our future exploratory dashboard in Visual 6 where users can explore each town more in-depth instead after actually making a decision on which area, they might be interested in. This could be from our above insights of the general market or by selecting their alumni primary school to begin exploring around in detail.

Visual 2 The trade-off between size and lease remaining

2.1

2.2



Visual 2.1

Stepwise Guide #1

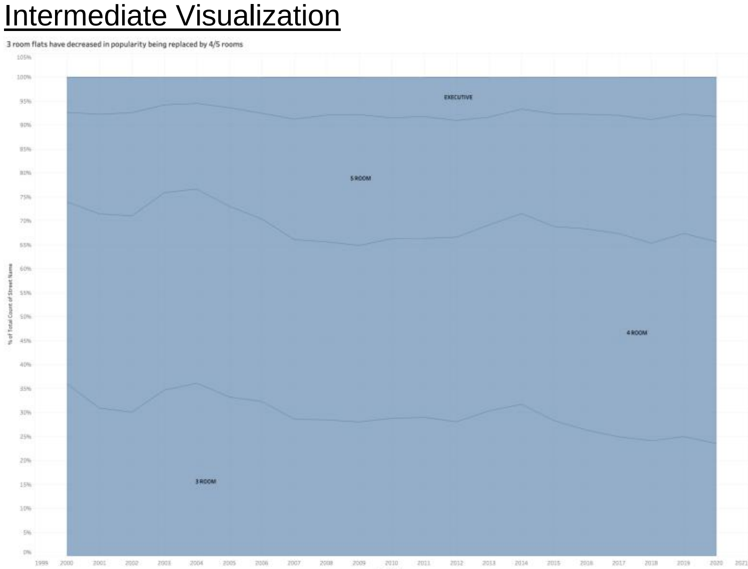
Options used in Tableau:

Rows: Count of units; *Columns: Year Sold*

Color: Flat Type (Computed); *Text: Flat Type*

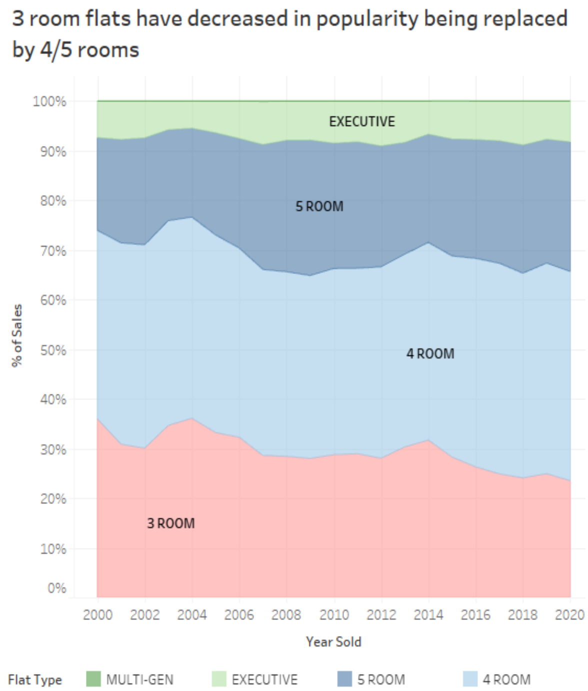
Filters: Year Sold > 2000

Intermediate Visualization



Stacked Area chart (Area marks) generated by plotting Year Sold (Dimension) against Count of StreetName (Continuous) using percent of total computed using Table(down) with Flat Type as labels, filtered by the years sold 2000 and onwards. Hue was included (encoding the flat types) afterward to allow differentiation between the flat types.

Final Visualization



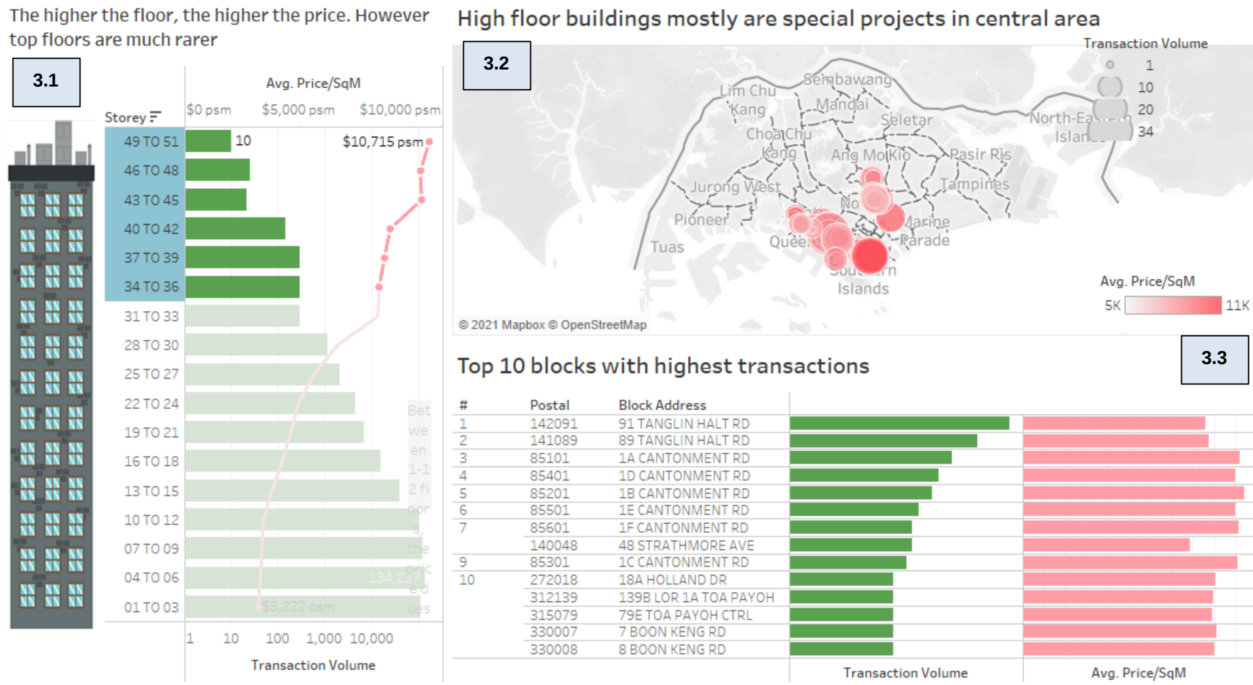
Visual Encoding and Rationale

A stacked area chart was used to describe the percentage share of each flat type for each year, using hue to encode the different flat types. This allowed for comparison of the share of each flat type in the resale market over the years as compared to a line plot with the absolute numbers. The x-axis was filtered to show flats sold from 2000 onwards to zoom into a more recent time window. Hue was used to encode the flat types, with a slightly reduced opacity to show gridlines.

Visual 2.2

A linear regression line was used to make sense of the noisy scatter plot of floor size over time released. Hue was used again to encode the different flat types. This in conjunction with visual 2.1 showed that not only

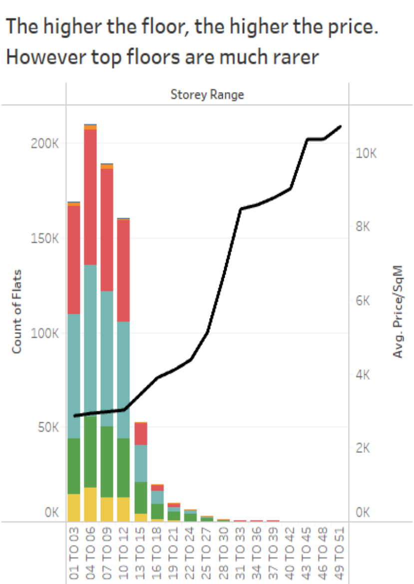
Visual 3 The trade-off between floor and price



Visual 3.1

A dual-axis combined graph to show how transaction volume and price inversely varied with floor. The size of the bar chart is used to encode the transaction volume while the position of the line graph encoding the average price per sqm. We modified the scale for the transaction volume to a logarithmic one as the scale difference between the highest and lowest amount varied significantly. Visually putting the storey range on the y-axis alongside a building graphic helped the user understand the chart faster. The hue follows from visual 1 where volume is represented by green and average price per sqm by pink.

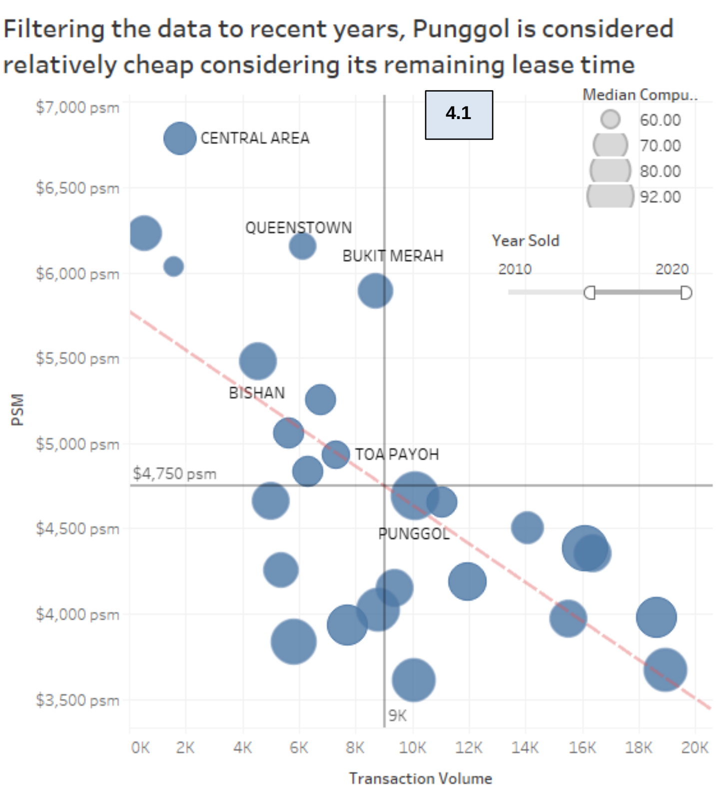
Our initial design had wanted to break down by flat types, but this was not necessary as it offered no additional insight. Furthermore, adding on a log scale helped put the transaction volume into clearer perspective, otherwise people might assume there were no transactions at the higher floors.

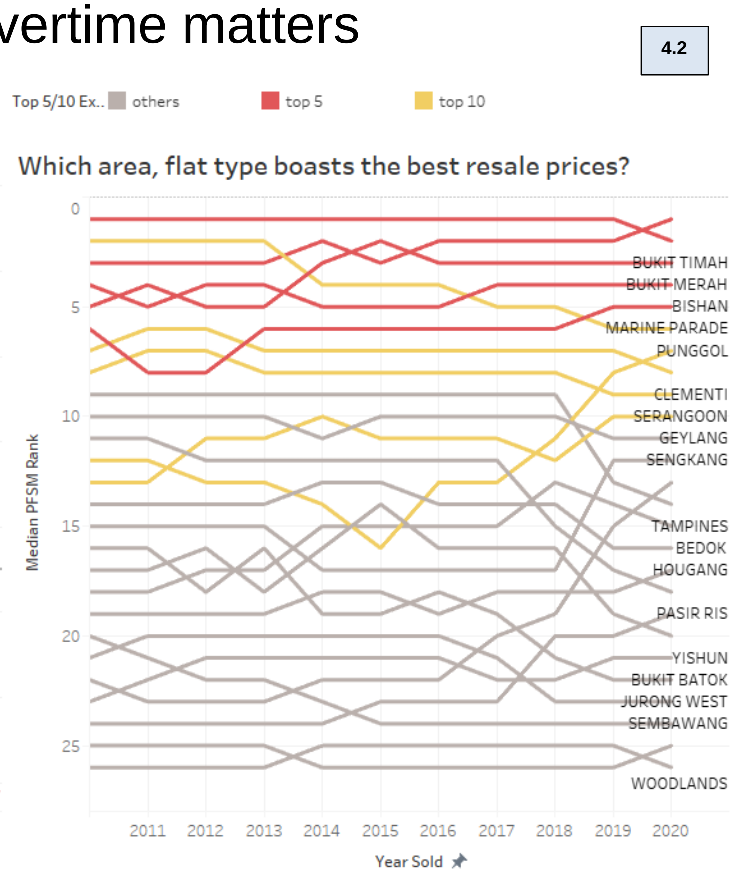


Visual 3.2
A density plot was used to show how highly priced sales from higher floors are typically associated with only a few flats along the central area. In fact, the difference between the highest and lowest are almost $5000 per sqm.

Visual 3.3
A bar chart was used to show how the data from the line graph skewed towards the higher end at higher floors because of the number of transactions in the central areas e.g., Cantonment Road, thus confirming visual 3.2. The hue follows from visual 1 where volume is represented by green and average price per sqm by pink.

Visual 4 Supply and demand overtime matters

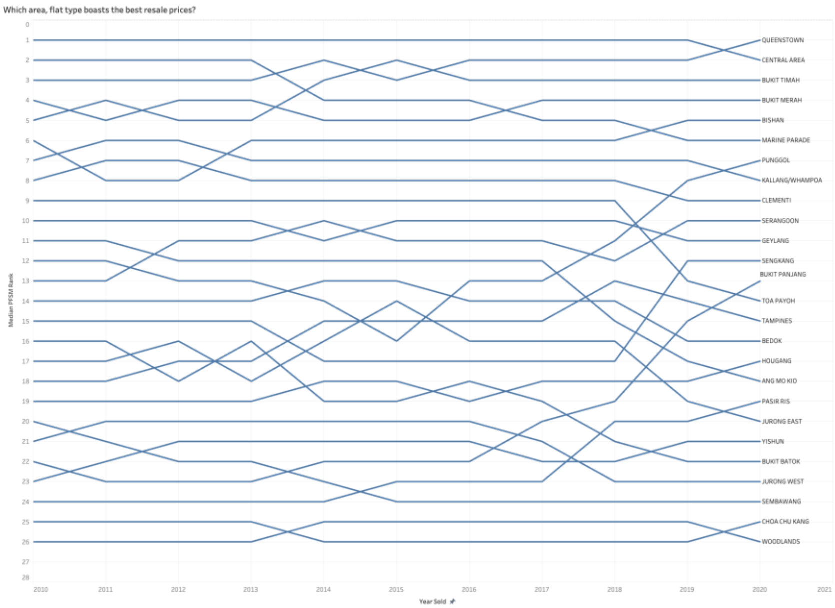




Visual 4.1
A bubble chart is used to highlight the general trend of increasing supply against decreasing price along with a linear regression line. This plot is made at the detail level of each town. Size is used to encode the remaining lease for each town. The same bubble chart is used as per Visual 1 but now we provide a time filter to remove the noise of past transactions.

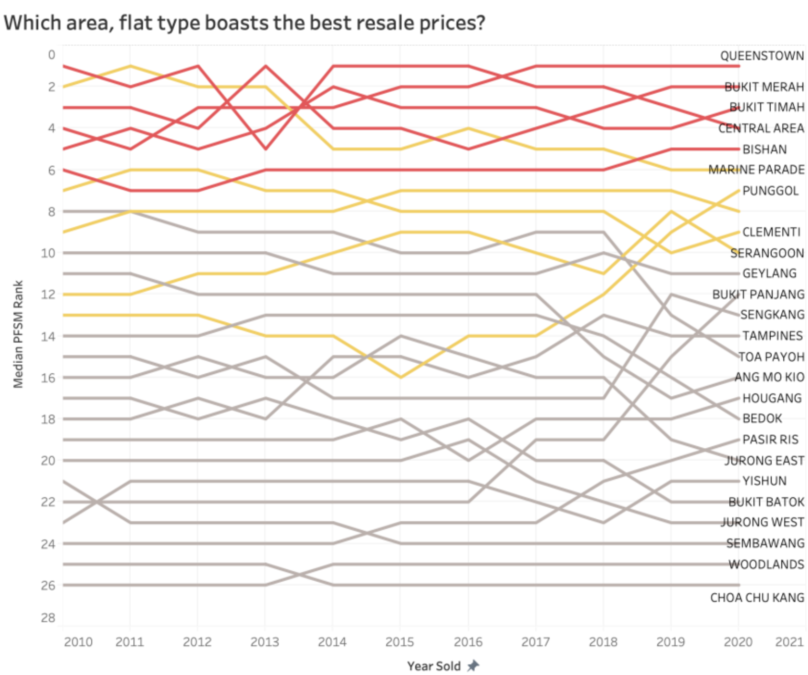
Visual 4.2
<u>Stepwise Guide #2</u>
Options used in Tableau:
Rows: Median Price Per Square Foot Rank (Calculated); *Columns: Year Sold*
Color: Top 5/10 Expensive Towns/Others (Computed); *Text: Town*
Filters: Year Sold > 2010

Intermediate Visualization



Bump chart (Line marks) generated by plotting Year Sold (Dimension) against Median PFSM Rank (Continuous) with Town as labels, filtered by the years sold 2010 and onwards. Hue was included (encoding top 5/10 in red/orange) afterward to draw focus on the most expensive towns.

Final Visualization



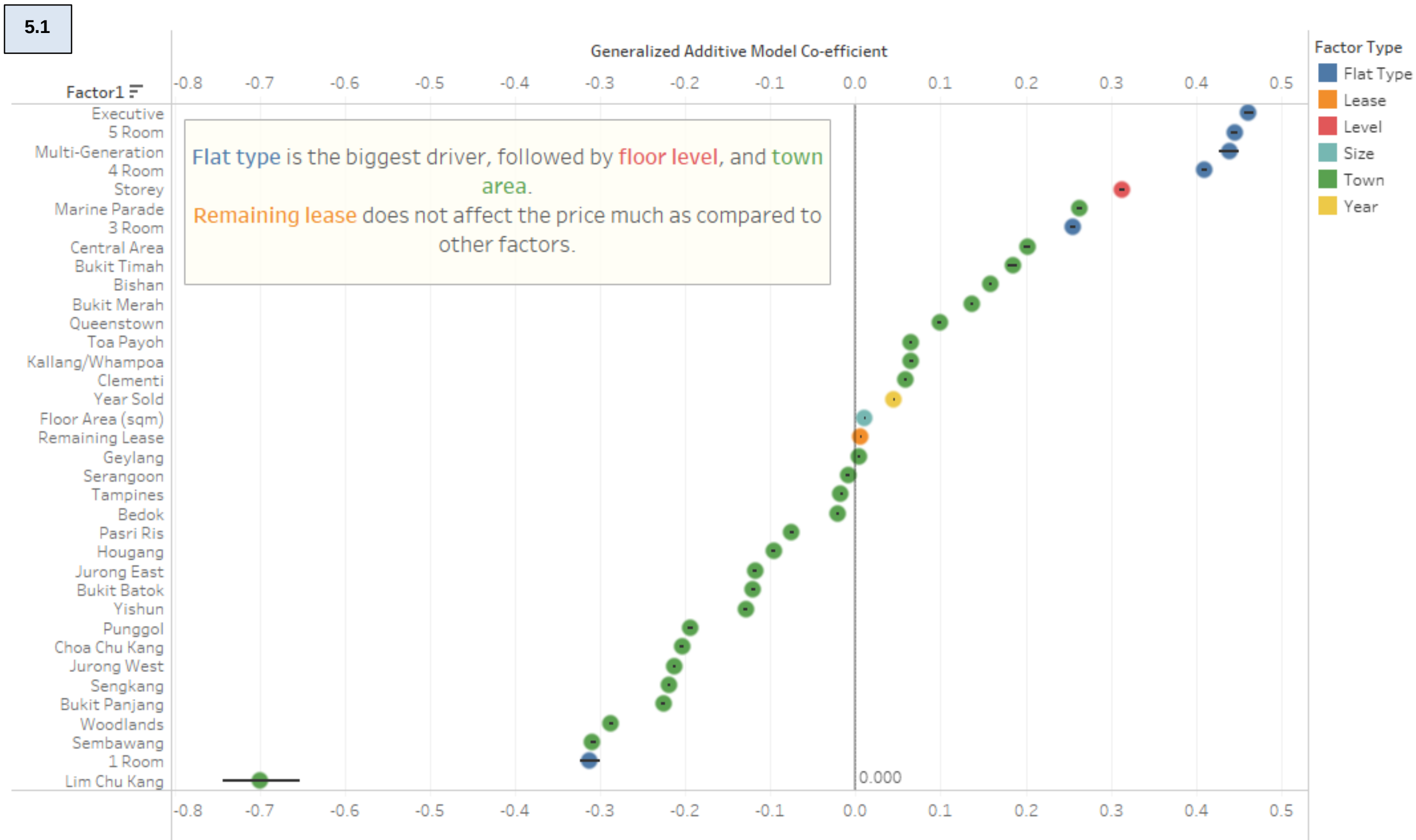
Visual Encoding and Rationale

A bump chart was chosen to allow comparisons between towns over the years, and towns were ranked across years to facilitate easier comparisons between towns over the years. The choice of a bump chart also allowed visualization of the volatility of price ranks across the years. Per square foot pricing was calculated to compare across flat types. The x-axis displays only the filtered years from 2010 to 2020 to focus on only the most recent changes. The top 10 ranked towns in 2020 were encoded by the hues red and orange to minimize visual complexity, and each line was labelled at the most recent year. Mouse-over detailed views also allow for users to inspect each rank for each year, and area highlighting allows for increased contrast of relevant lines against the rest of the towns.

The bump chart helps to highlight Punggol again as one of the trending towns in the top 10 (the other being Sengkang but it's outside of the top 10).

Visual 5 The Main Driving Factors

5.1

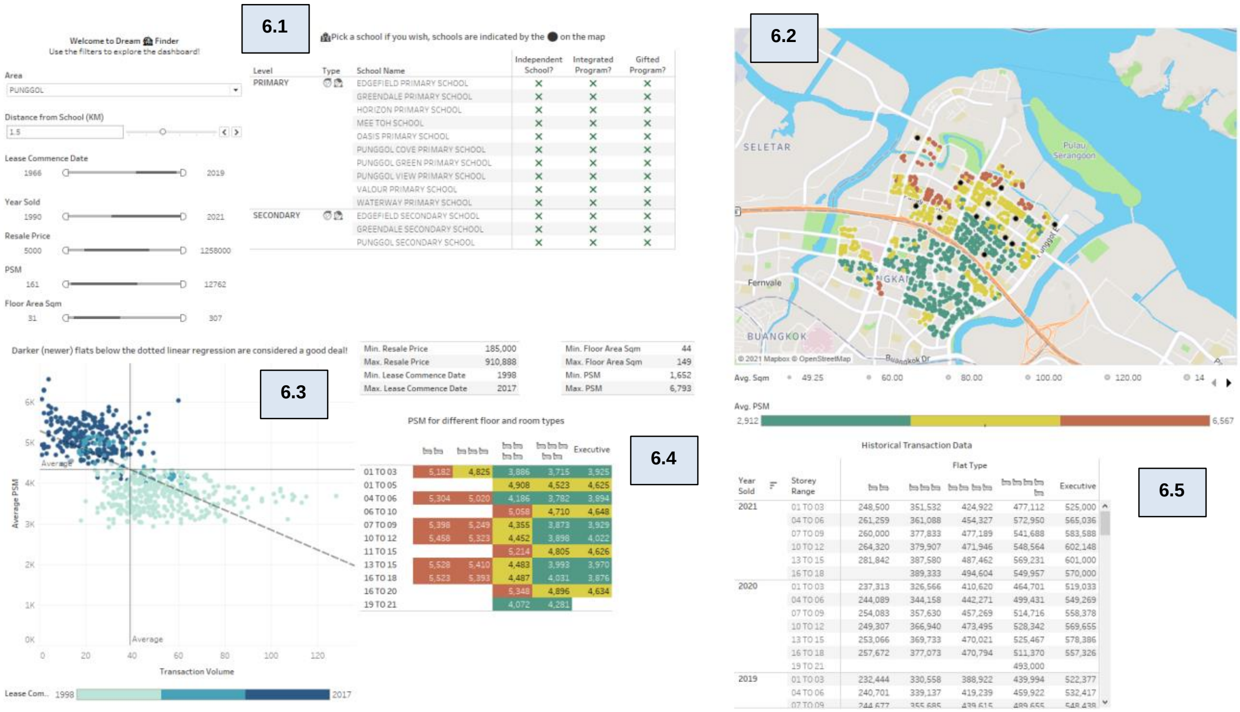


Visual 5.1

Position is used to show the coefficient while hue is used to categorize the different factors. A constant line is added to show where factors start diverging from positive coefficients to negative coefficients.

Visual 6 An Interactive Dashboard for Users

To provide a rich, and flexible tool for users to use in their hunt for the perfect resale flat, the following interactive dashboard was created in Tableau, incorporating school vicinity information for all areas in Singapore. This would allow users to find the pricing trends for their custom-tailored needs and queries about the housing market. All the visuals have also conveniently linked one another for ease of filtering and zooming in on the perfect flat. Each visual also provides information to the user for them to pick and choose which filters to apply.

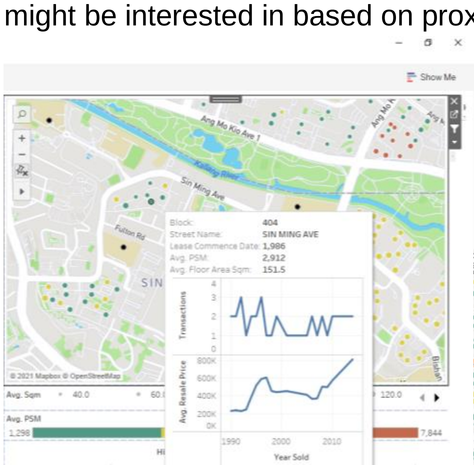


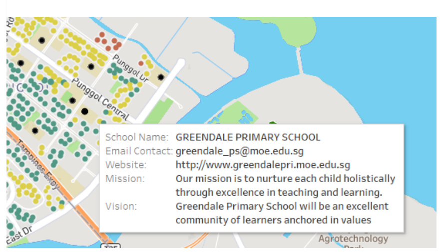
Visualization 6.1

We utilize shapes (e.g., 👦👧) to show the type of school and also whether certain programmes are offered at respective schools. Hue is used to further solidify and contrast the yes (green) and no (red). The schools are also organized in the respective levels (e.g., Primary, Secondary) for ease of filtering and browsing. Here the user is able to pick particular schools they want in the area based on the categories.

Visualization 6.2

We utilize hue to show the different bins of housing based on average price per sqm of a particular block. Because we wanted to have the background map to show some detail as well, darker contrasting hues were used to clear spot the flats. Using simpler traffic light colors, we subconsciously let the user now to look for green flats since they would be cheaper in the same area. Although not apparent, when zoomed in, there is also size encoded to show the average floor area (sqm) of a particular block. Schools are encoded with a distinct black dot. Scrolling over the tooltip for each block reveals 2 line graphs that show temporal data of the block's number of transactions and absolute average resale price. Scrolling over the tooltip for each school reveals contact information for the school as well as some basic information such as its vision and mission statement. This visual helps users pick flats they might be interested in based on proximity to the school and compare prices around the immediate vicinity.

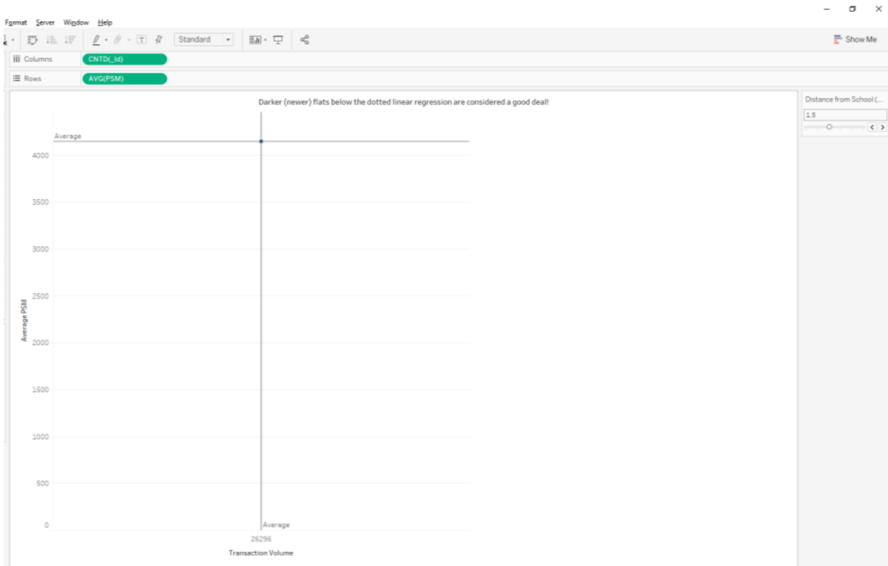




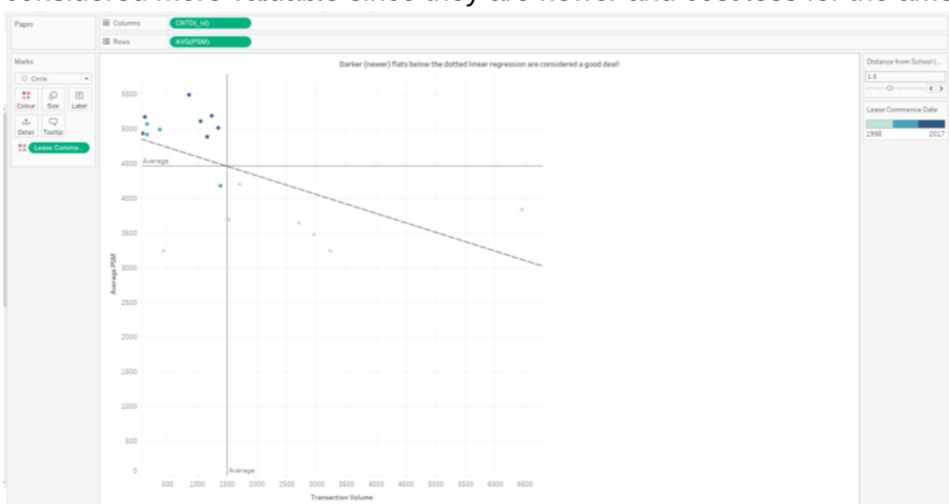
Visual 6.3

Stepwise Guide #3

To recreate the bottom left graphic, we first plotted 2 measures. A distinct count of transactions and the average resale price. This is because we're interested to see how price is affected by the supply of flats which we know to be inversely proportional based on earlier analysis in Visual 4. This leaves us with only one mark which is the average of all the data.



Next, we want to add in more details such as how the data would differ across the flats which commence earlier and later and encode 3 categories through a color bar. We want the darker colors to be more apparent to the user and hence encode them to be of a later date. At this stage, since we have more data, we can choose to add a linear regression plot which would help our user find value-for-money flats. Darker plots below the line would be considered more valuable since they are newer and cost less for the amount of supply and demand in the market.

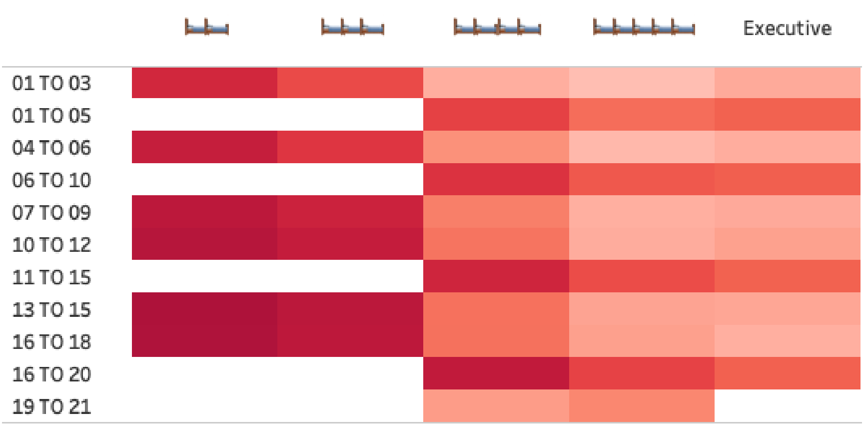


Lastly, because this visualization is viewed in the context of a chosen school or town, we can afford to include a street name and block detail without causing too much noise. This allows the user to zoom in immediately on a particular block in the filtered area that is considered "value-for-money" to begin exploring. It is also easier to spot outliers as we expect darker spots to be higher compared to lighter colored spots.

Visualization 6.4

Stepwise Guide #4

To create the heat map in the interactive dashboard, a heatmap (square marks) generated by plotting Flat Type (Dimension) against Storey Range (Dimension), with the hue of each square encoding the average price per square meter.



Hue was then used (via the color marks) altered to utilize a stepped divergent color scheme for the price to allow for better readability. Text labels were also used to encode the average price per square meter for easier comparison between cells.

PSM for different floor and room types

					Executive
01 TO 03	5,182	4,825	3,886	3,715	3,925
01 TO 05			4,908	4,523	4,625
04 TO 06	5,304	5,020	4,186	3,782	3,894
06 TO 10			5,058	4,710	4,648
07 TO 09	5,398	5,249	4,355	3,873	3,929
10 TO 12	5,458	5,323	4,452	3,898	4,022
11 TO 15			5,214	4,805	4,626
13 TO 15	5,528	5,410	4,483	3,993	3,970
16 TO 18	5,523	5,393	4,487	4,031	3,876
16 TO 20			5,348	4,896	4,634
19 TO 21			4,072	4,281	

A heatmap was used to visualize the best pricing per square meter by flat types, showcasing that the larger the flat and higher the storey range – the higher the resale price. The colors also make it easy to spot outliers. If we notice green colored squares in the higher levels, we might potentially want to filter out those blocks to take a look.

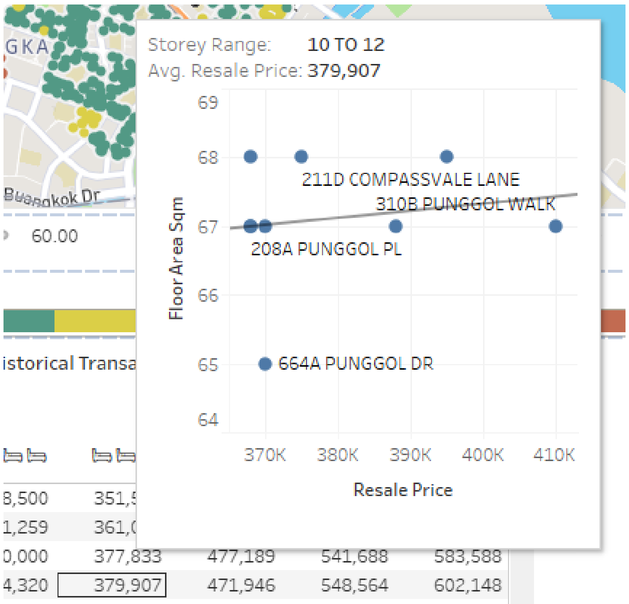
Initially we had picked the heatmap to be used for selecting a town to filter down. However, the area to explore after selecting a town was still too large and it didn't help the user in narrowing down particular blocks or streets to look at. Hence, we made use of other factors to select a town in Visual 4 instead.

Which area, flat type boasts the best resale prices?

	1 ROOM	2 ROOM	3 ROOM	4 ROOM	5 ROOM	EXECUTI..	MULTI-G..
Grand Total	2,930	2,998	2,782	3,119	3,340	3,323	3,145
ANG MO KIO		3,501	2,652	3,174	3,760	3,885	
BEDOK		2,904	2,692	2,956	3,203	3,412	
BISHAN			3,152	3,680	3,986	4,083	3,606
BUKIT BATOK			2,519	2,778	3,122	3,035	
BUKIT MERAH	2,930	3,522	3,225	4,583	4,485	2,604	
BUKIT PANJA..		5,018	2,615	2,888	3,204	3,321	
BUKIT TIMAH			3,357	3,771	4,110	4,220	
CENTRAL AR..		2,578	3,734	4,943	6,938	3,275	
CHOA CHU K..		5,001	2,454	2,774	2,996	3,148	
CLEMENTI		5,499	2,829	3,479	3,726	3,539	
GEYLANG		2,423	2,888	3,264	3,529	3,285	
HOUGANG		5,078	2,647	2,883	3,162	3,349	
JURONG EAST		5,383	2,550	2,778	2,940	3,163	
JURONG WE..		4,371	2,416	2,732	3,039	3,085	
KALLANG/W..		2,480	3,067	3,837	4,027	3,545	

Visualization 6.5

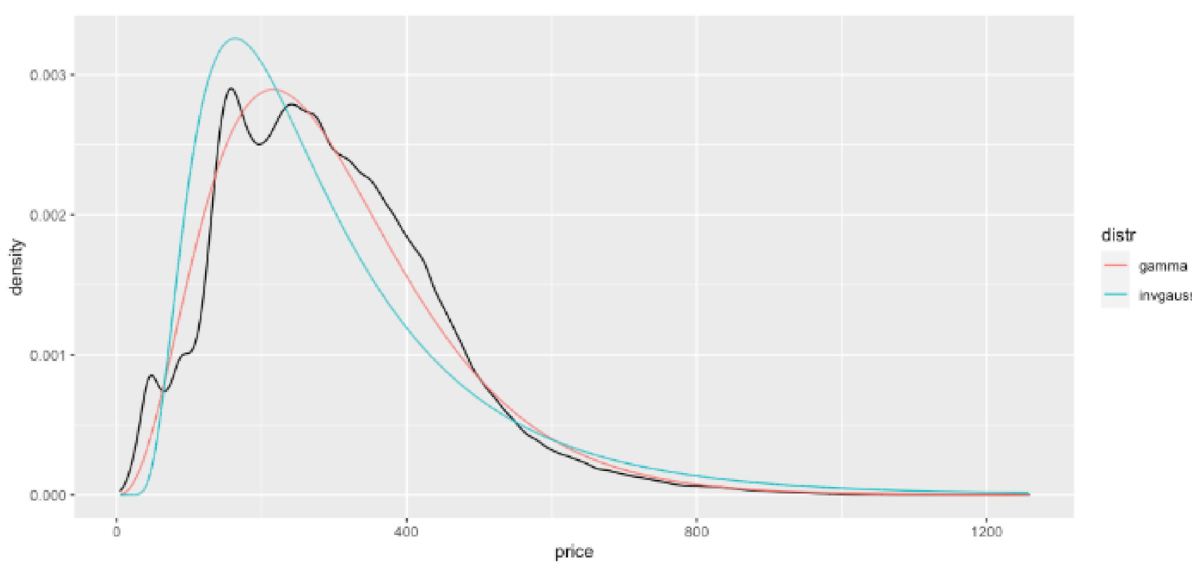
A simple text table breaks down the historical average resale price across storey ranges, year sold, flat type. After all the other visuals have been looked at and the user has finally decided on a particular flat for viewing. We want to give the user a feel of the historical pricing around the area so they can have a benchmark to start off negotiations with the prospective seller and property agent. Hovering over the tooltip also shows different historical transactions for the same flat type and storey range in different areas on the map they may not have considered.



Exploring the Generalized Additive Model

With the number of factors that go into consideration for resale prices, many of which are overlapping, it would be extremely useful for buyers to know exactly how each factor influences the price against other factors. For example, does location have such a big influence on price that a user could be convinced to consider alternative areas – this would allow the buyer to prioritize areas to maximize their given budget and needs.

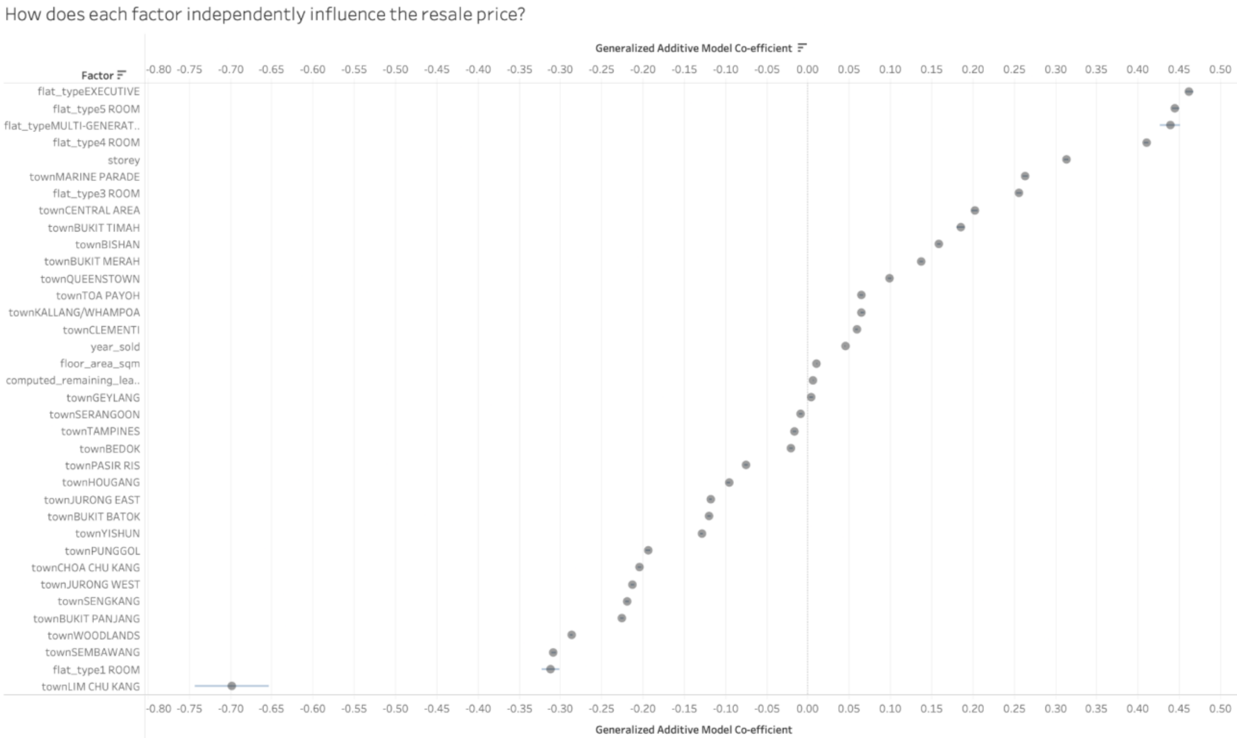
To further add insight to the pricing trends in Singapore, the wealth of data and variables allows us to carry out a Generalized Additive Model taking into account all variables available, before simplifying the model to best predict the pricing data seen. This would allow us to describe the influence of each independent variable in the final resale price seen over the last 20 years. A gamma density distribution was found to better fit the price of all resale transactions than Gaussian or Inverse Gaussian distributions.



A list of potential models was fit against the estimated gamma function. The model with the following formula was found to have the lowest AIC, indicating the best fit, out of all proposed models:

`price ~ floor_area_sqm + storey + year_sold + computed_remaining_lease + flat_type + town`

This analysis was carried out in R using the glm function from the stats package and the resultant coefficients and confidence intervals were then plotted in tableau.



The findings suggest that independent of other variables, flat types have by far the largest influence on the price of a resale flat, more so than the location. Similarly, flats on higher floors command much higher prices, while towns like Marine Parade and Bukit Timah have higher prices than other towns. Another striking observation is that resale prices recorded in Lim Chu Kang tend to have much lower prices than any other town. Overall, the GAM analysis is extremely useful in taking the extensive data and extracting the independent influences that each factor has on resale flat analysis.

Group Contributions

Nicholas	Data Preprocessing, Report, Exploratory Dashboard, Poster
Tingting	Insights for Tableau Stories, Storyboarding
Jacob	Report, Generalized Additive Model, Poster