

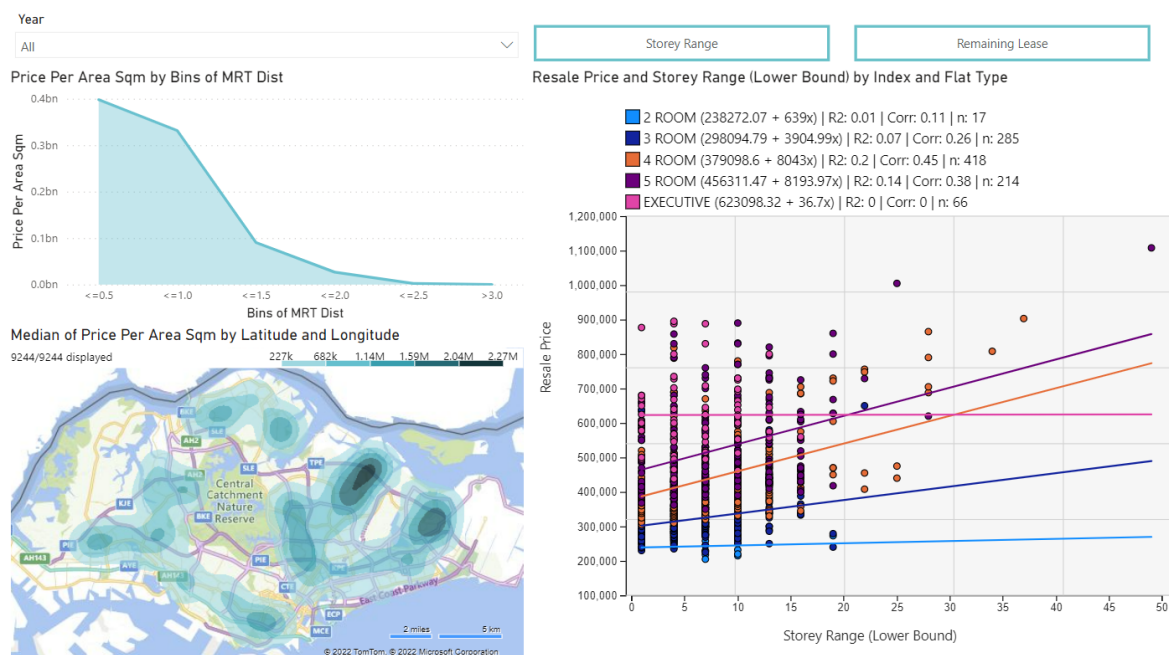
Documentation

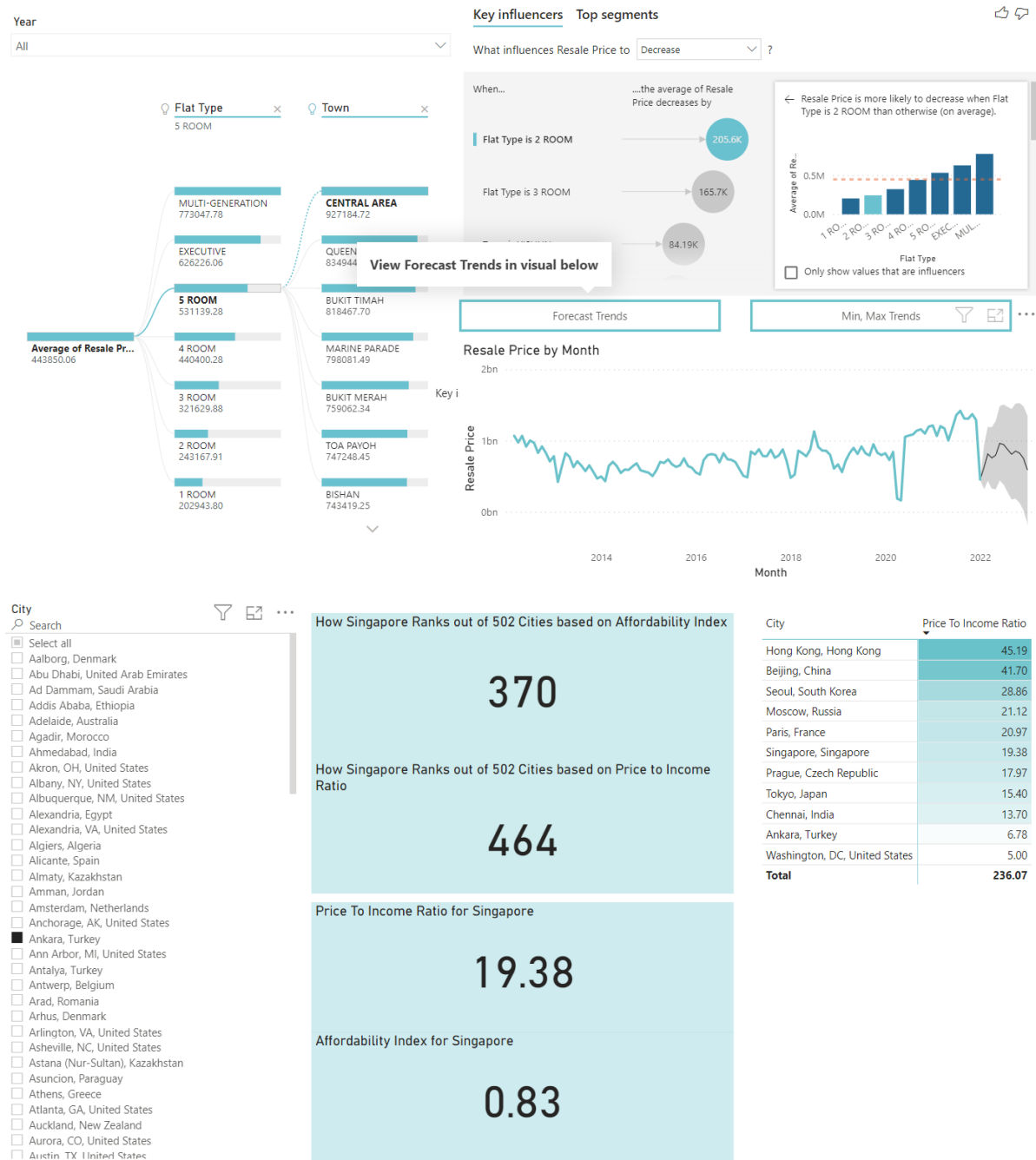
Table Of Contents

Demo	1
Data Preparation & Modelling Processes	2
Scope of datasets	2
Data Preparation	4
Data Cleaning and Transformation	4
Data Modelling	6
Data Visualisation	7
Report 1	7
Report 2	9
Report 3	13
Statistical Modelling	14

Demo

Video Demo: <https://vimeo.com/user122233488/review/678118497/1c3104daa4>





Data Preparation & Modelling Processes

Scope of datasets

Datasets are all not in common.

S/N	Dataset	Excluded/Included	Description/Reason
-----	---------	-------------------	--------------------

1	https://data.gov.sg/dataset/7a339d20-3c57-4b11-a695-9348adfd7614 HDB Resale Prices	Included	<p>Resale Flat Prices based on Registration date, with important columns which I will be using such as Month, Town, Flat type, Floor area sqm, Remaining Lease, Resale price which could impact Price of Housing.</p> <p>I use this dataset for comparing remaining lease and storey range to resale price specifically and also to see resale price from 2012 to 2017 as it has more data than dataset 2.</p>
2	Dataset from teachers, scraped from property guru, HDB Resale Data	Included	<p>Columns Used: Town, Floor Area per Square Sqm, Latitude, Longitude, Year-Month, Nearest Hawker School and MRT distance, Resale Price.</p> <p>Resale Flat Prices with additional important columns like Distance from Nearest Mrt which would be useful in allowing target users to see how distance from amenities affect housing price.</p>
3	https://www.numbeo.com/property-investment/rankings.jsp?title=2021	Included	<p>Columns used: City, Price to Income Ratio, Affordability Index</p> <p>Used to compare affordability of housing in Singapore with other cities using columns like "Price To Income Ratio" and "Affordability Index".</p> <p>Makes it easier to compare between specific cities to see how Singapore ranks in terms of affordability of housing, which will help target users decide if they're happy with housing prices in Singapore.</p>
4	https://data.gov.sg/dataset/price-range-of-hdb-flats-offered	Excluded	Maximum and Minimum Price according to the Town, I decided not to use it as I could just summarise values as maximum and minimum in my visualisations based on the dataset given by teachers instead so this isn't needed.
5	https://data.gov.sg/dataset/hdb-resale-price-index	Excluded	Not enough variables to compare. Since they only show the resale price index per quarter, there are no other variables to be compared.
6	https://data.gov.sg/dataset/hdb-property-information	Excluded	<p>The dataset shows HDB information like Block number, Street, Maximum Floor, Year completed.</p> <p>However, columns like market and hawker tags data is only yes or no instead of showing distance/name of hawker/market, hence not enough information to compare HDBs.</p>

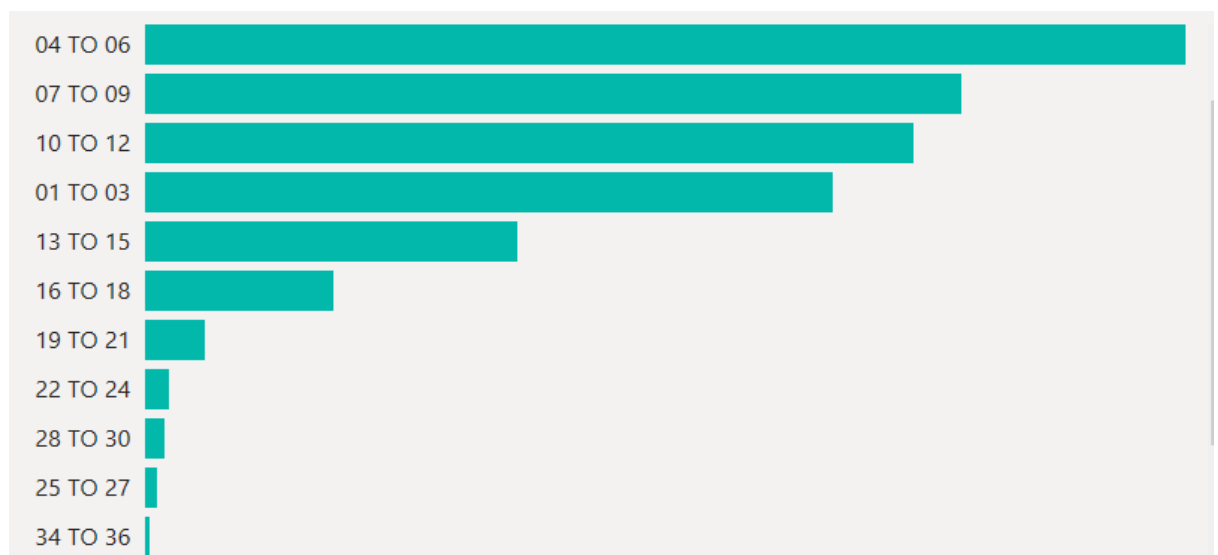
Data Preparation

For the third dataset, I had to scrape the site with python in order to get the csv file. I will only be using the data for 2021. I will be using the 3 datasets I mentioned above.

Data Cleaning and Transformation

Dataset 1 (Resale Flat Prices based on Registration date)

- Removed Columns: Block, lease_commence_date there is a column "Remaining lease date" which will be more useful as a point of comparison for what I need
- Renamed Columns: Capitalise first letters, removed underscores
- Corrected Types of all columns
- First data set was in 3 different separate excel sheets so I appended all 3 tables into one
- Renamed title of the new table to "HDB Resale Data 1"
- Hide initial 3 tables since the appended table has everything
- Added index column
- Remaining lease to text format then created new column in query editor and inserted a custom column with "Text before Delimiter" to get only the year (eg. from 64 years 5 months to 64) so that it's easier to compare remaining lease years to see how that impacts price of housing
- Remaining lease column data type changed to Whole Numbers
- Changed data types in Model view: City to City datatype
- Storey Range used to make a new column with only the lower bound numbers as all ranges had the same number of storeys (eg. '10 TO 12' to 10) and renamed the column to 'Storey Range (Lower Bound)'. Data type is also changed to Whole Numbers. This is done to see if there's a relation between storey range and resale price in statistical modelling



Dataset 2 (Teacher's copy)

- Removed Columns: Town (groups), Nearest Hawker, Nearest MRT and Nearest Primary School as those are not numerical values and are harder to compare, in this case I won't need it
- Calculated Column: Bins of MRT distance to categorise them to visualise them easily and Price per area sqm for better visualisations that are not affected by different sizes of housing

Bins of MRT Dist = `IF('HDB Resale Data 2'[Mrt Dist] <= 0.5, "<=0.5",`

`IF('HDB Resale Data 2'[Mrt Dist] <= 1.0, "<=1.0",`

`IF('HDB Resale Data 2'[Mrt Dist] <= 1.5, "<=1.5",`

`IF('HDB Resale Data 2'[Mrt Dist] <= 2.0, "<=2.0",`

`IF('HDB Resale Data 2'[Mrt Dist] <= 2.5, "<=2.5",`

`IF('HDB Resale Data 2'[Mrt Dist] <= 3.0, "<=3.0", ">3.0"))))))`

Price Per Area Sqm = `DIVIDE('HDB Resale Data 2'[Resale Price], 'HDB Resale Data 2'[Floor Area Sqm])`

- Changed data types in Model view: City to City datatype

Dataset 3 (Property Price Index By City 2021)

- Removed Columns: Row Count, Gross Rental Yield City Centre, Gross Rental Yield Outside of Centre
- Reloaded this data twice into 2 new tables with Python Scripts
 - These 2 tables will be used to show where Singapore stands in comparison to other countries when ordered by Affordability Index and by Income to Ratio Index. The smaller the index, the more affordable it is. (eg. In this case, we rank quite badly for the Price To Income Ratio at 464 out of 502 countries.)
 - First table sorted by Affordability Index with new index column

```
import matplotlib.pyplot as plt
import pandas as pd

IndexSortedByAffordability= pd.read_csv('C:\\Users\\jolen\\OneDrive\\Desktop\\School
Stuffs\\VAP project\\PropertyPriceIndexByCity_2021.csv')

IndexSortedByAffordability= IndexSortedByAffordability.sort_values('Affordability Index')
IndexSortedByAffordability['Index'] = range(0,len(IndexSortedByAffordability))
IndexSortedByAffordability.drop(['Unnamed: 0','Mortgage As A Percentage Of Income','Price
To Rent Ratio City Centre','Price To Rent Ratio Outside Of City Centre','Gross Rental Yield City
Centre','Gross Rental Yield Outside of Centre','Price To Income Ratio','Affordability
```

```
Index'],axis=1,inplace=True)
print(IndexSortedByAffordability.info())
```

- Second table sorted by Price to Income Ratio with new index column

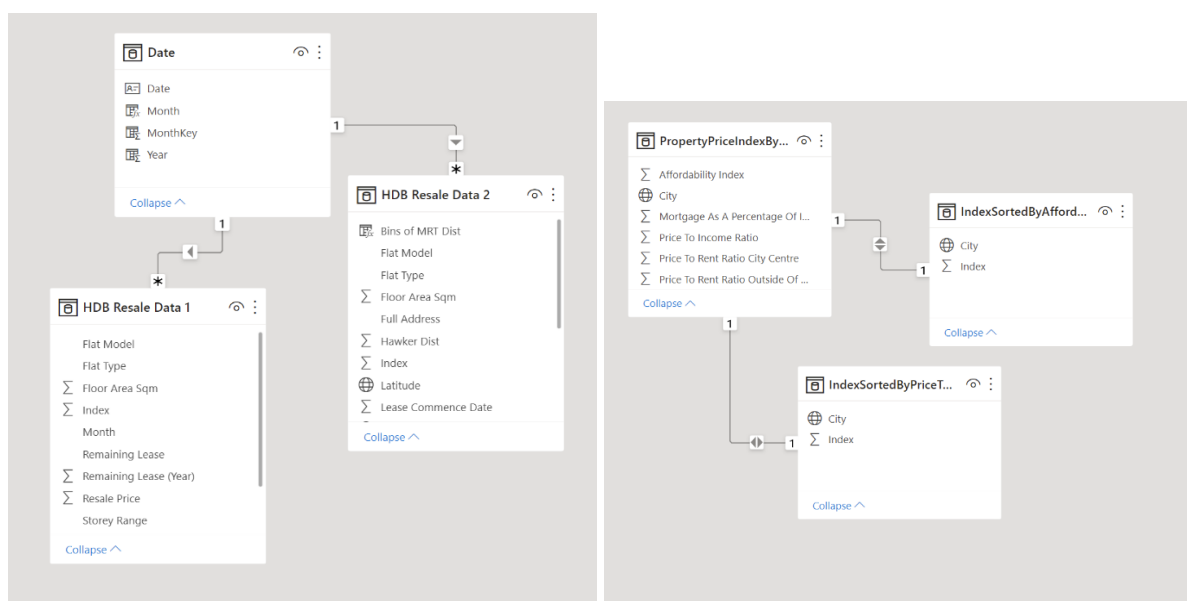
```
import matplotlib.pyplot as plt
import pandas as pd

IndexSortedByPriceToIncome= pd.read_csv('C:\\Users\\jolen\\OneDrive\\Desktop\\School
Stuffs\\VAP project\\PropertyPriceIndexByCity_2021.csv')

IndexSortedByPriceToIncome= IndexSortedByPriceToIncome.sort_values('Price To Income
Ratio')
IndexSortedByPriceToIncome['Index'] = range(0,len(IndexSortedByPriceToIncome))
IndexSortedByPriceToIncome.drop(['Unnamed: 0','Mortgage As A Percentage Of Income','Price
To Rent Ratio City Centre','Price To Rent Ratio Outside Of City Centre','Gross Rental Yield City
Centre','Gross Rental Yield Outside of Centre','Price To Income Ratio','Affordability
Index'],axis=1,inplace=True)
print(IndexSortedByPriceToIncome.info())
```

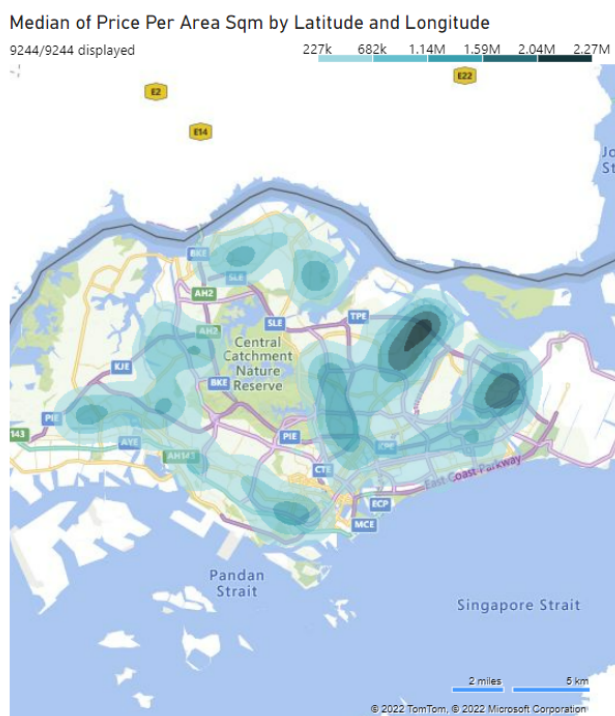
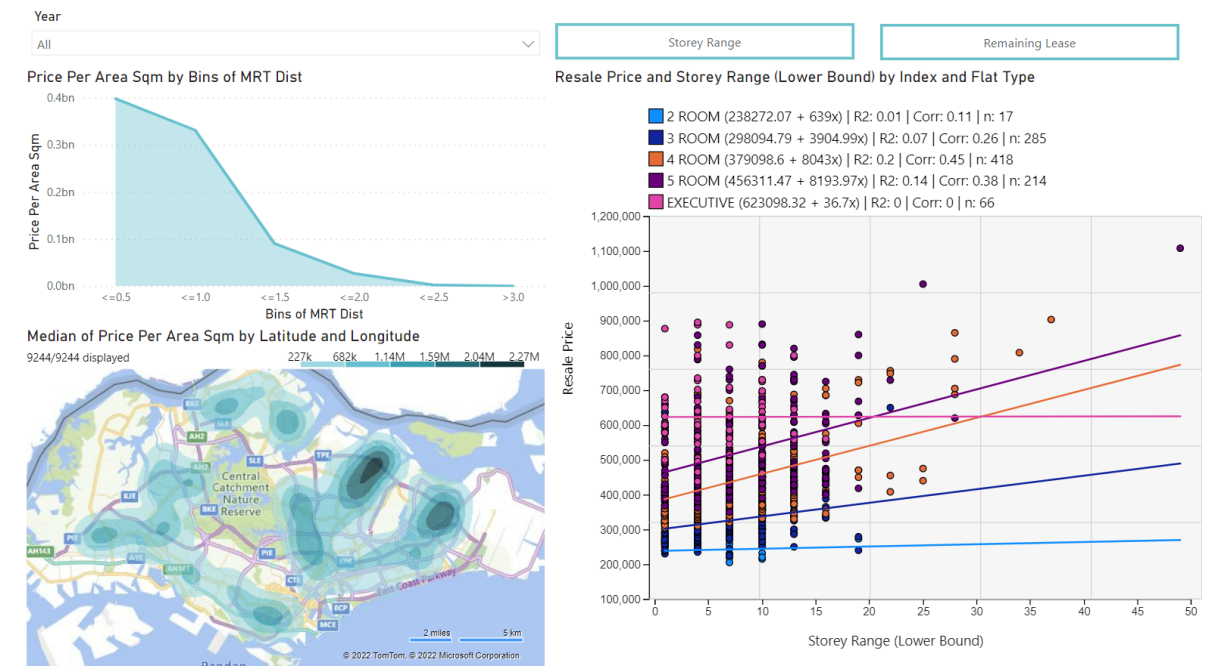
Data Modelling

- Created a new date table and marked it as a date table
- Joined the tables Date with HDB Resale Data 1 and 2
- Joined the index tables for Property Price Index



Data Visualisation

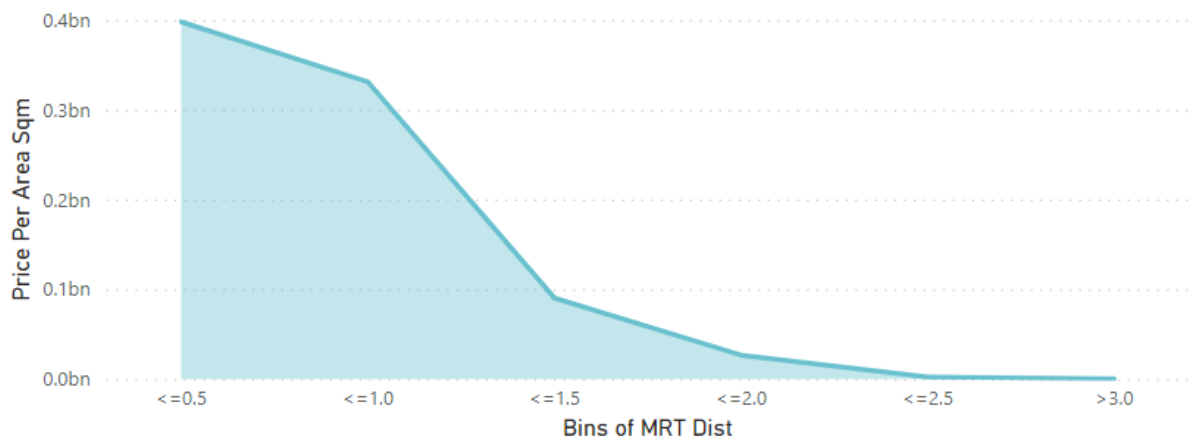
Report 1



Map visualisation to showcase the media price per area sqm by location. I decided to use Price per area sqm so that bigger units would not impact the prices, and using median to avoid anomalies.

This shows that the Punggol-Sengkang area, West and Central have more expensive housing on average compared to other regions of Singapore.

Price Per Area Sqm by Bins of MRT Dist

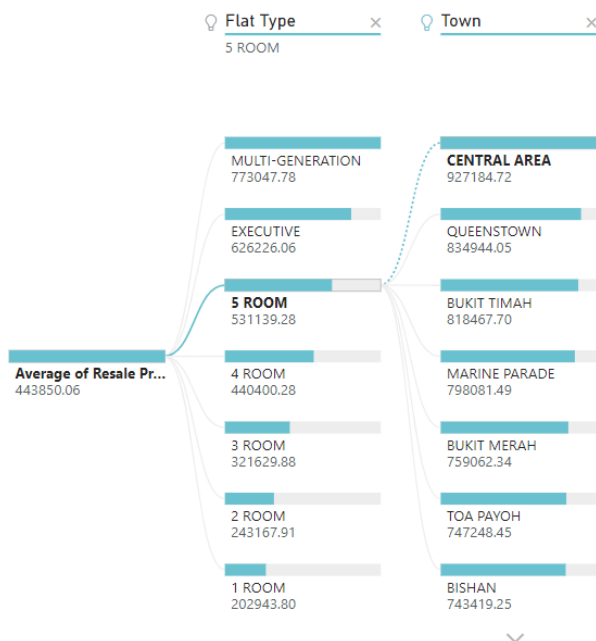
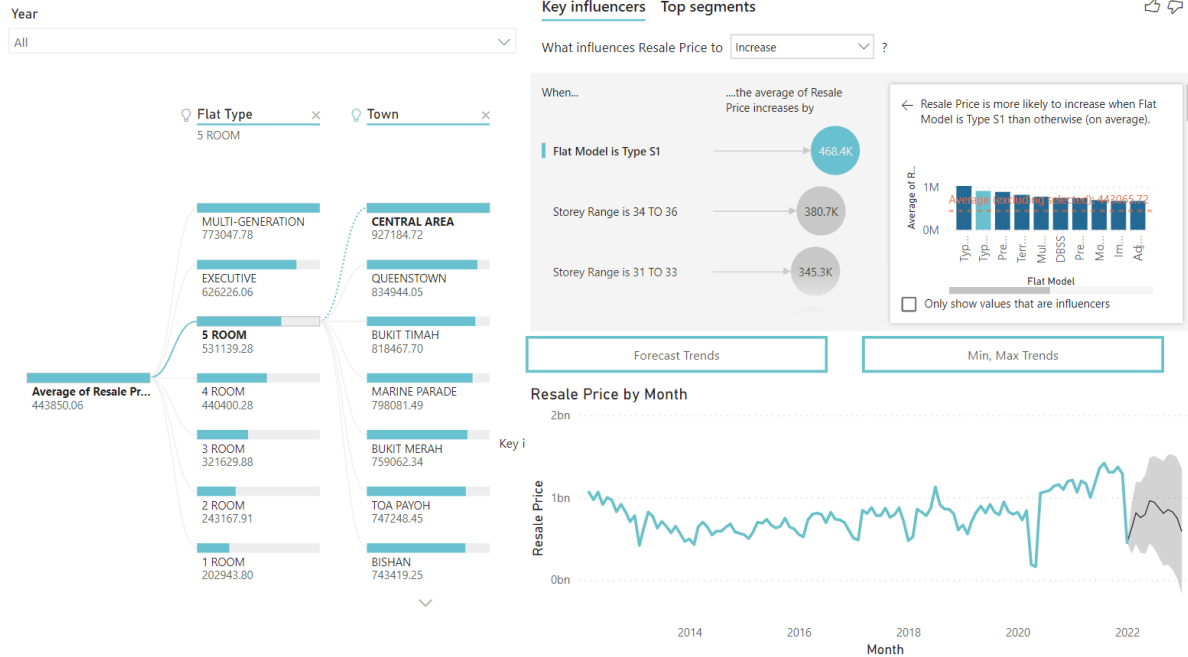


This Line graph shows how Price Per Area Sqm of Housing decreases for flats further from MRTs. In other words, flats with MRT stations nearby would be more expensive. This will help show target users the reason behind the more pricey units as there's quite a significant difference as distance from MRT increases.

Year

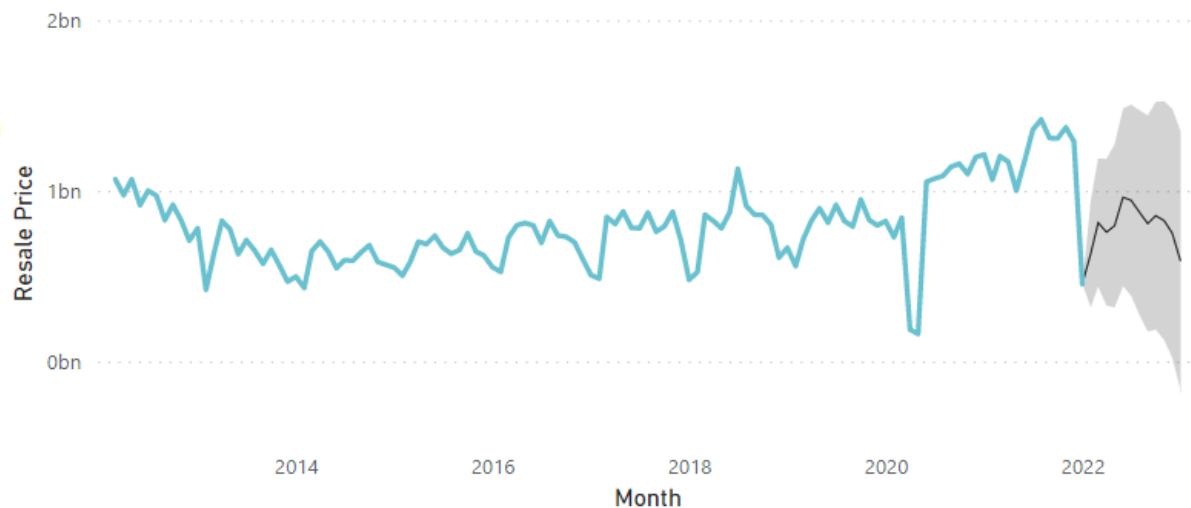
Year slicer to see how data changes across the years, whether the decrease seen in price of housing as MRT distance increases is the same in different years and whether the more expensive sectors in Singapore is the same in different years.

Report 2



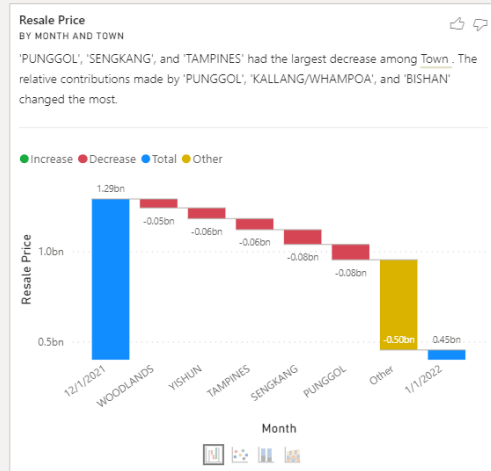
With the help of a decomposition tree, we can see which types of units tend to be more expensive and for those types of units, which town would be more expensive. For example, for 5 Room Flats, the Central Area is the most expensive area. This will help target users see which locations they should consider for different types of flats, for example they could look at the least expensive town for 4 room flats if they're looking to get a 4 room flat.

Resale Price by Month



Resale Price Trends over the years from 2012 to 2021 and predicted data for 2022.

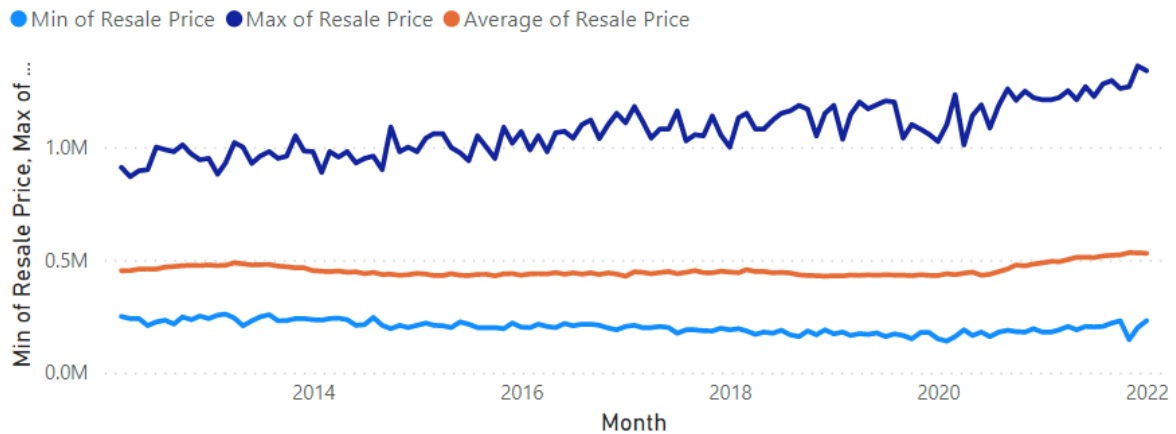
Here's the analysis of the 64.76% decrease in Resale Price between 12/1/2021 and 1/1/2022



There was a huge decrease of 64.76% in December of 2021 to 2022. As seen from the analysis on the left, Punggol, Sengkang and Tampines had the largest decrease among all Towns, which might be due to people getting less housing in those areas that month compared to the following year since relative contributions made by Punggol was one of the few that changed the most as well.

With the help of forecasting trends chart target users would be able to see that there has been a gradual increase from 2015 to 2021. However, the forecast says that we can expect resale prices to stay relatively constant in comparison to the years 2020-2021. The forecast of resale prices increases from \$454236665 to \$962151699 and then decreases again to \$588147201 within 2022. With this information, target users can decide when it is most suitable to get a house if they're planning to get one in 2022.

Min of Resale Price, Max of Resale Price and Average of Resale Price by Month

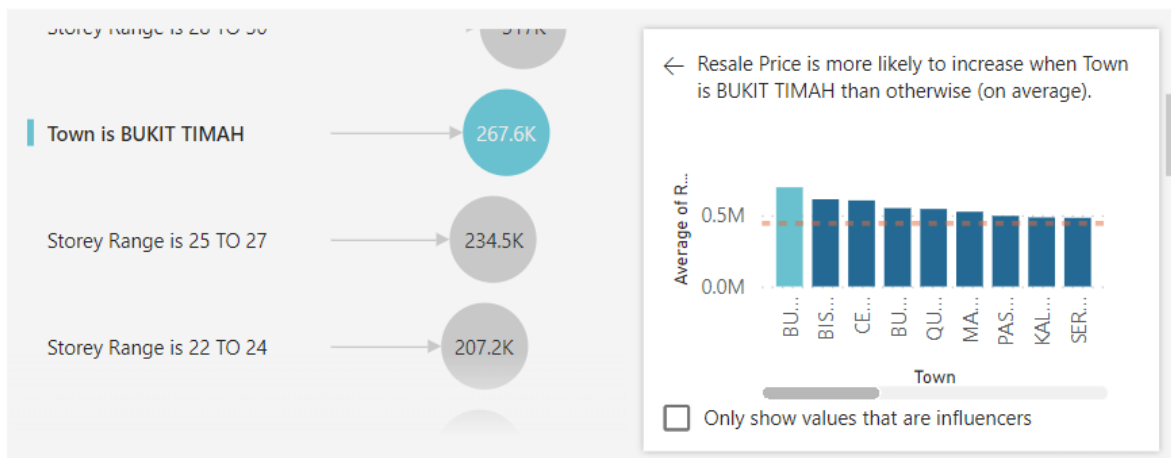


When users click on the button to show “Min, Max Trends”. They would be able to see this line graph. This line graph shows that Housing Prices on average have not changed that much over the years in comparison to Maximum Housing Prices which shows a gradual increase from 2016 to 2021.

Key influencers Top segments



What influences Resale Price to ?

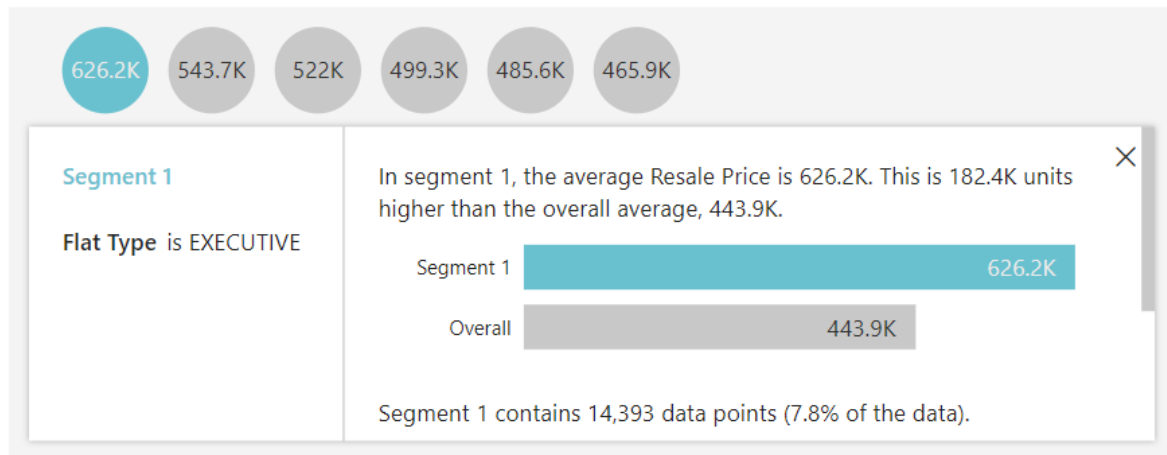


As seen from the Key Influencers Analysis, Bukit Timah causes the biggest increase in resale prices, followed by Bishan and Central Area. Hence with this information, if target users are looking to get housing in these areas, they would have to expect housing to be more expensive.

Key influencers Top segments



When is Resale Price more likely to be High ?

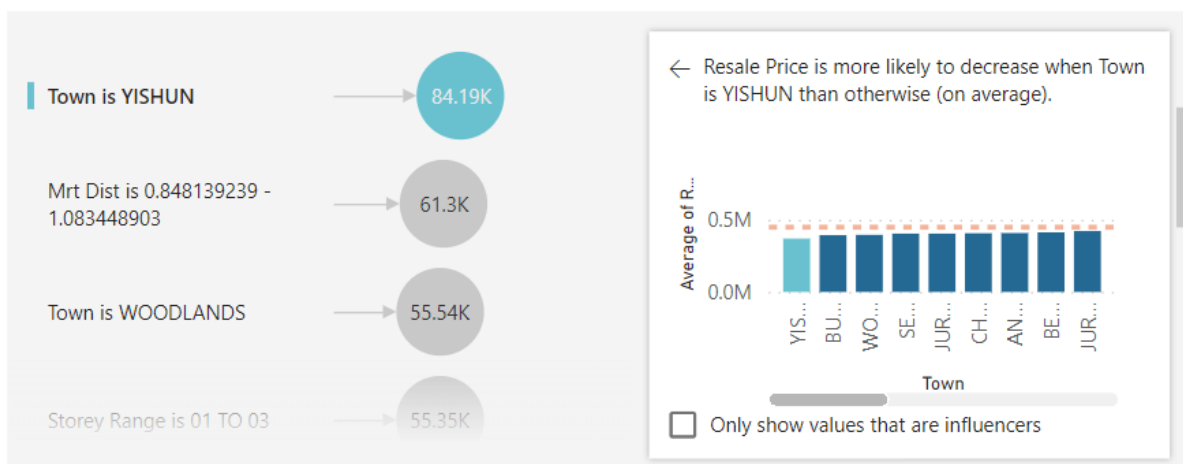


This Top Segments Analysis shows that Resale Prices are also more likely to increase when Flat Type is Executive as the biggest segment has all the same flat type of Executive. The average Resale price is \$626.2k which is a lot higher than the overall average in other kinds of flat types.

Key influencers Top segments

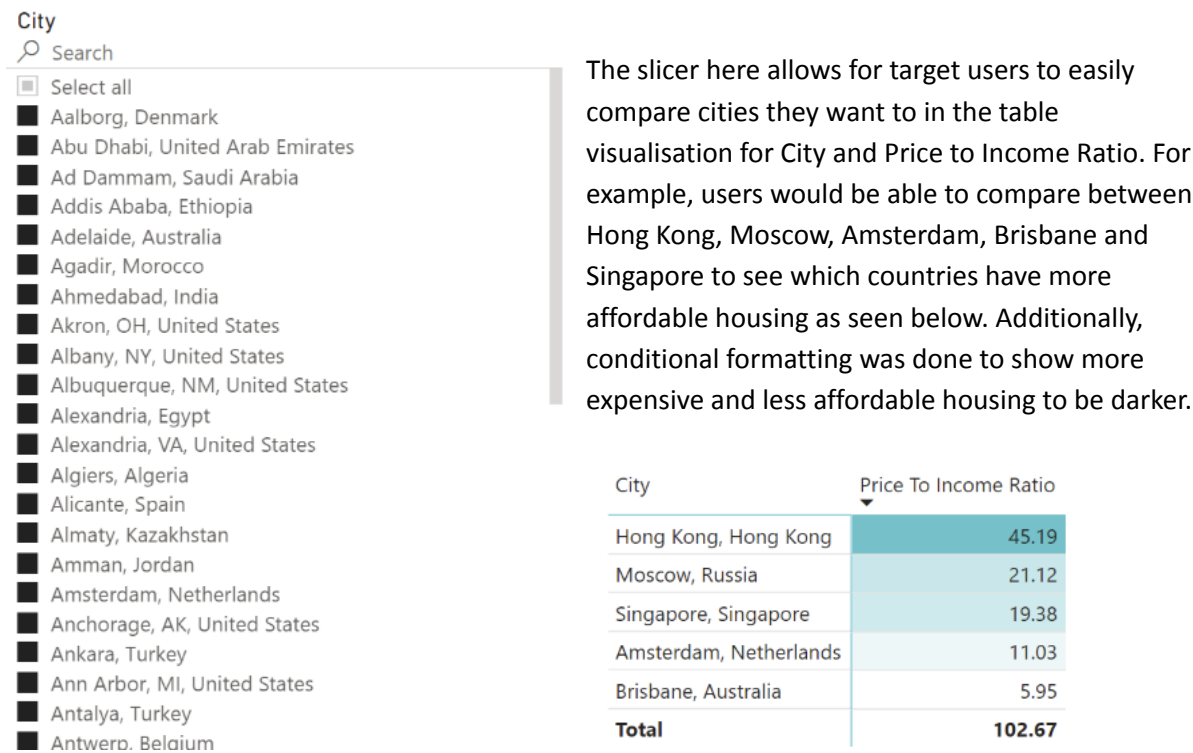
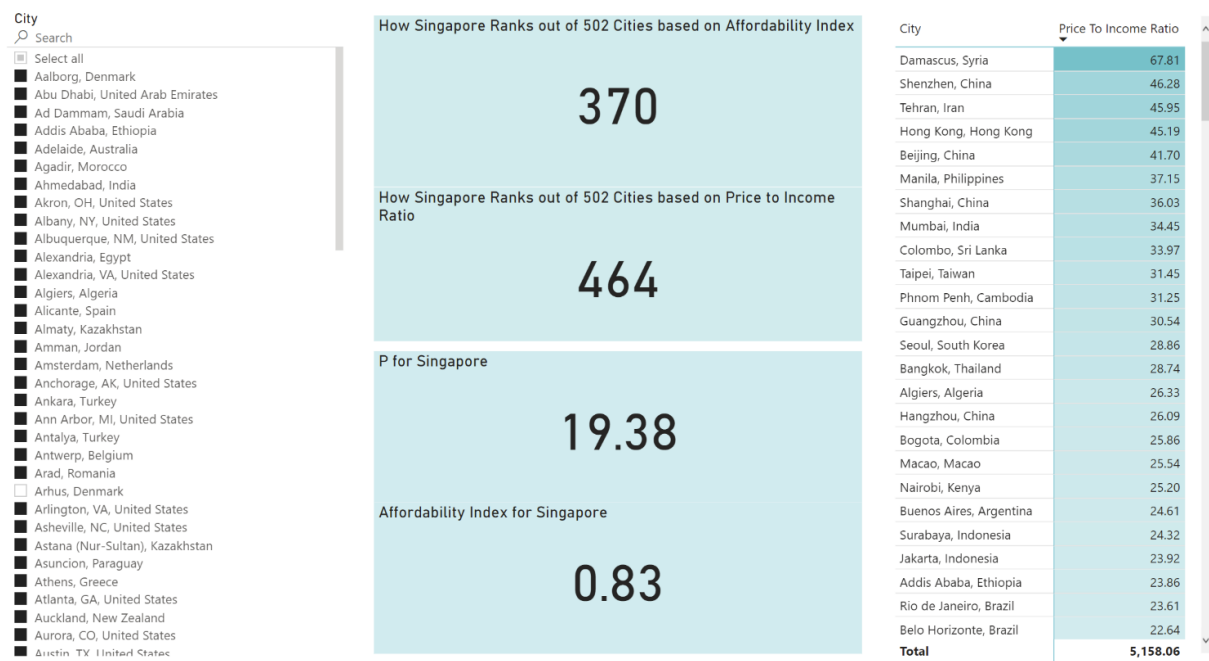


What influences Resale Price to Decrease ?



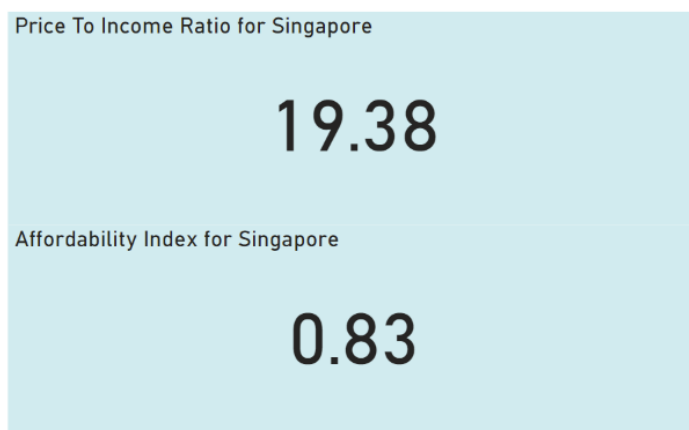
For target users looking for more affordable housing, the Key Influencers Analysis also shows that resale price is more likely to decrease when the Town is Yishun followed by Woodlands, so they could look into finding housing there instead. School Distance and Hawker Distance from flat is also considered in the analysis however it does not prove to influence the data as much as these other factors.

Report 3





This KPI visualisation shows how Singapore Ranks out of 502 Cities based on Affordability Index. This makes use of the dataset created earlier with Python. The smaller the number, the more affordable that country is. Hence, this lets target users know that Singapore isn't very affordable in comparison to the 502 countries in this dataset.

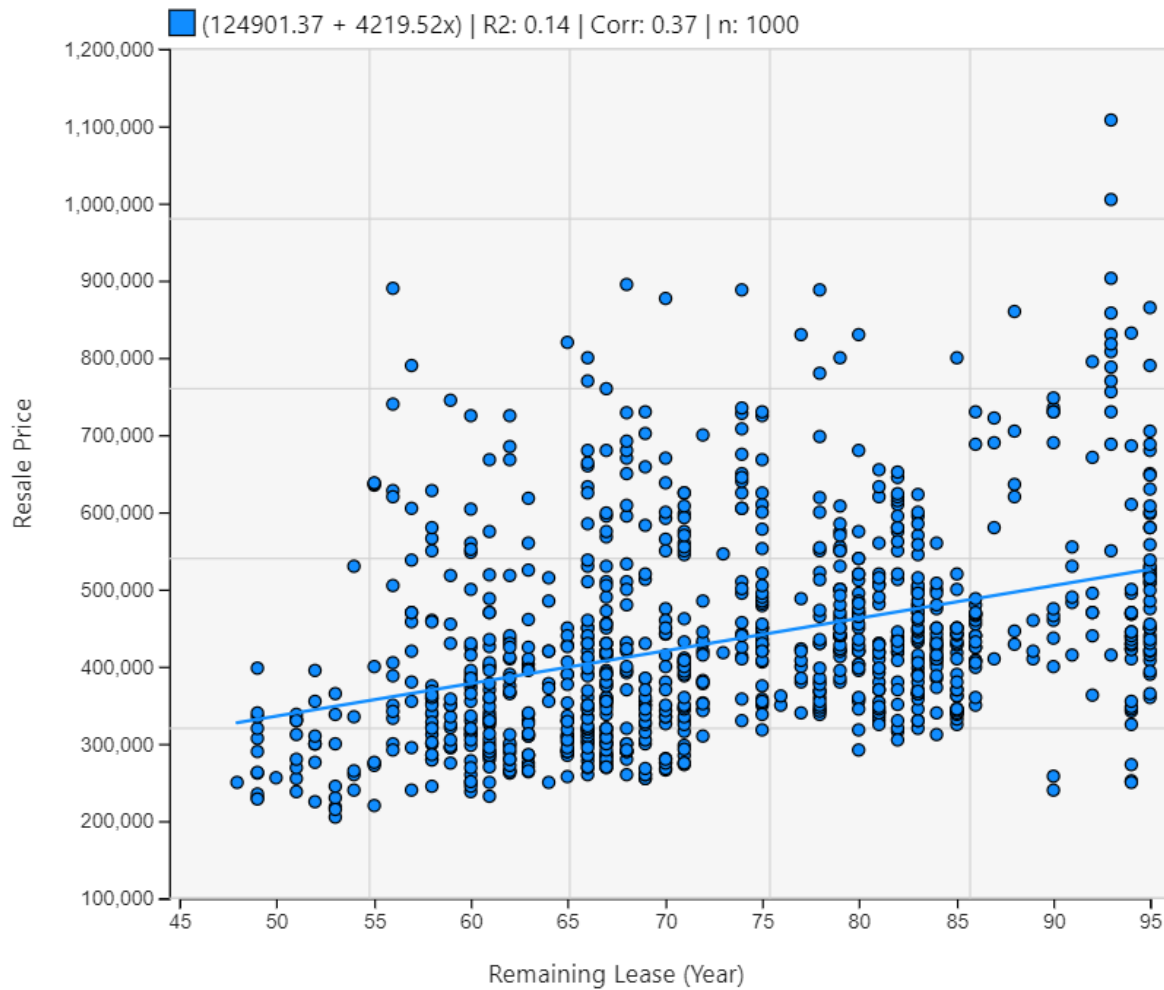


Then, this KPI visualisation shows Singapore's Exact Price to income Ratio and Affordability Index for easy reference. Target users will also be able to make use of filters on this visualisation to choose to see affordability of different countries instead of Singapore.

Statistical Modelling

I will be modelling Resale Price as the Y axis variable against Remaining Lease (in Years) and Resale Price against Storey Range (Lower Bound) to see if they have any relationships between them. This would allow users to look more specifically for housing with longer/shorter lease or flats that are at a higher story if they're looking to get more affordable/expensive housing.

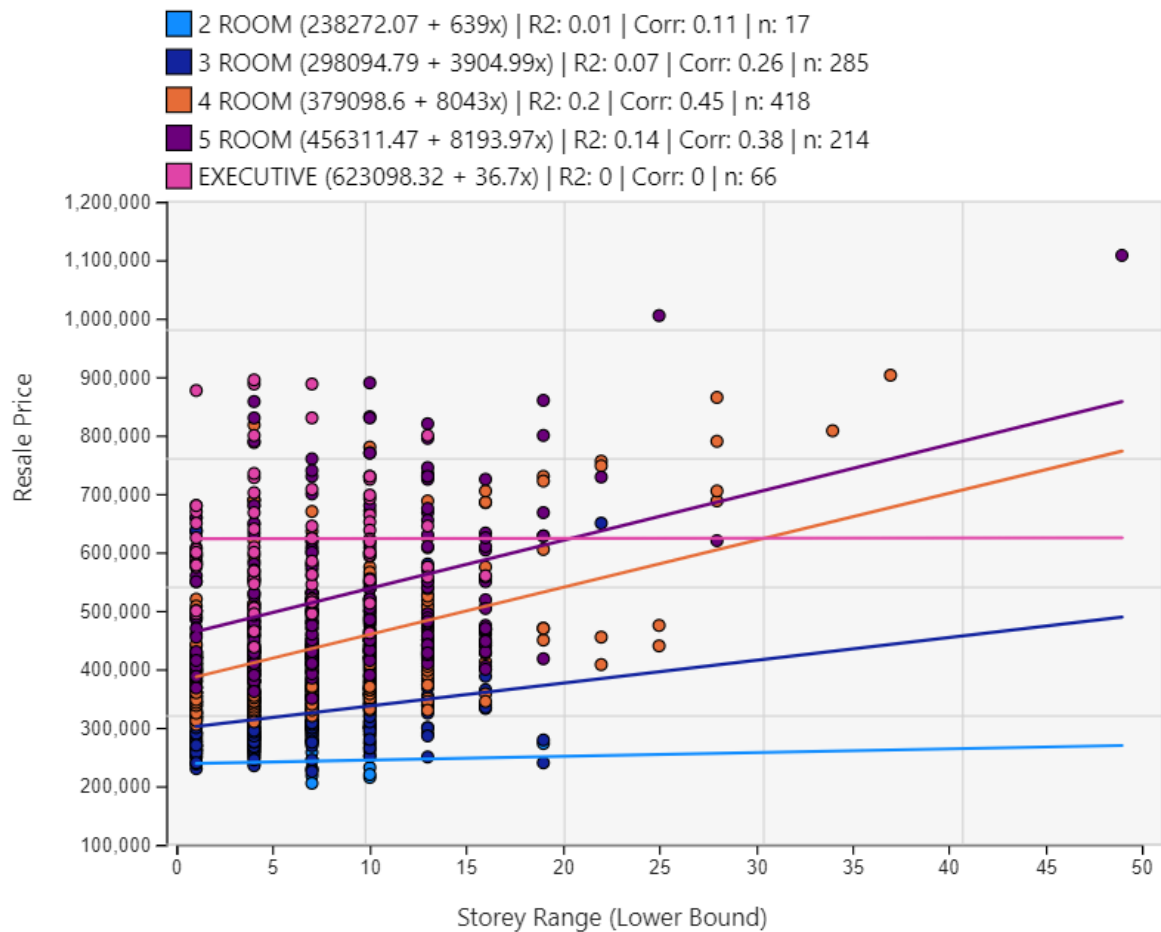
Resale Price by Index and Remaining Lease (Year)



With the help of Linear Regression, I plotted to see the relationship between Resale Prices with the Remaining Lease the flat has. Added an index column to be used for this model. As seen from the R^2 value of 0.14 and increasing trend, the relationship is a weak and positive relationship. This means that remaining lease only plays a very small role in deciding resale price and it is not one of the important influencers.

With this, target users would be able to see that though a higher remaining lease means higher resale price, it doesn't impact resale price as much as most would think it does.

Resale Price and Storey Range (Lower Bound) by Index and Flat Type



The Analysis shows that 4 room and 5 room flats have the strongest relationship as the R2 value is the highest among all flat types. This means that there is a strong positive linear relationship between flat type 3 ROOM, 4 ROOM and 5 ROOM and Storey Range and Resale Price. 2 ROOM, 4 ROOM and 5 ROOM flats increase in resale price as storey range increases. 2 ROOM flats don't show a stronger relationship as its R2 value is 0.01 and it only has 17 units hence not enough data to compare. Executive Flat Type is not affected by storey range as it shows a weak relationship with a low R2 value and is neither positive nor negative meaning that Storey range of executive flats do not affect its resale price much.

With this, target users will be able to see how choosing flats that are at a higher storey would cost more especially if they choose to pick a 3, 4 or 5 room flat.