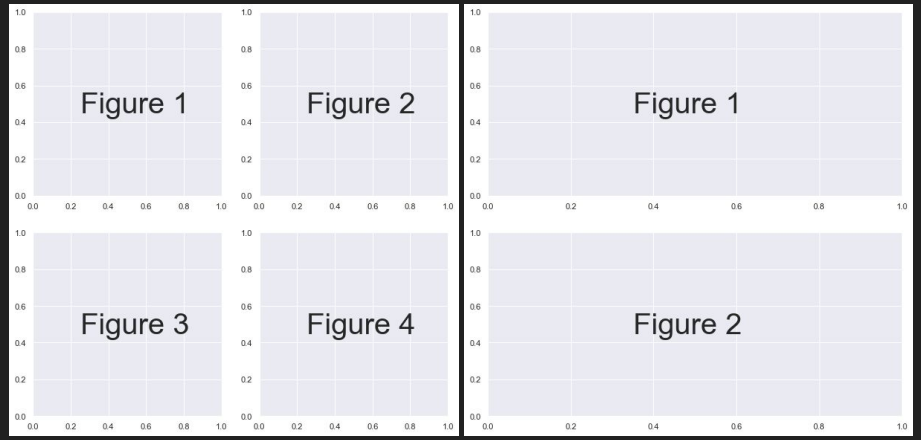
ST1502 : DATA VISUALISATION CA2 assignment
Done By: Lim Chuan Hao(1922264)

Datasets used:

1. Resale flat prices based on approval date from 1990 onwards. (Dataset 1)
   a. Dataset taken from blackboard
2. HDB Property Information (Dataset 2)
   a. Link
3. Land Area and Dwelling Units by Town (Dataset 3)
   a. Link

Note:
The way I will be referring to each plot in the chart is as follows. (Figure numbers are shown in the dummy charts below)
Also a more detailed analysis will be in the notebook.



Objectives:

1. What flat should you buy when starting a new family (Dataset 1)
   a. Context
      i. They will be living in the house for around 30 years
      ii. Hopefully the house will be able to act as an investment as well
   b. Sub-Objectives
      i. How does the flat type, town, storey and number of lease years remaining affect the resale price of a flat.
2. Analysis of the situation of residential dwelling and amenities in Singapore (Dataset 2 & 3)
   a. Context
      i. This would be like a report to HDB
      ii. Residential dwelling refers to both the residential property as well as the residential dwelling units in Singapore
      iii. Amenities will be further defined when explain the datasets
   b. Sub-Objectives
      i. How are residents distributed across Singapore
      ii. Does each region have enough amenities

Explaining the nature, cleaning and manipulation of the dataset. (Dataset 1)

Nature of the dataset

- Columns
  - ‘month', 'town', 'flat_type', 'block', 'street_name', 'storey_range', 'floor_area_sqm', 'flat_model', 'lease_commence_date' and 'resale_price'
  - 10 columns
- Rows
  - 769807 rows

This dataset is mainly about the resale prices of flats. However the dataset is split by both the street and town the flat is in. The flat is also split by flat model and we are also only given the lease commencement date and not how old the flat is.

Cleaning and manipulating the dataset

- Cleaning
  - The flat types had a slight duplicate entry of ‘MULTI-GENERATION’ and ‘MULTI GENERATION’. Thus I first converted all entries of ‘MULTI-GENERATION’ to ‘MULTI GENERATION’
  - I then did more cleaning by creating a new columns ‘regions’ mapping the towns that belonged to each region. The regions are ‘North’, ‘East’, ‘South’, ‘West’ and Central’. (Look at the notebook for the specific mapping)
  - I also new columns, ‘lease_years_left’ and ‘price_per_sqm’ which are calculated.
    - ‘lease_years_left’ are the number of years left on a lease of the flat. We assumed that all flats had a 99 years lease.
    - As for column ‘price_per_sqm’, I simply took the resale price divided by the ‘floor_area_sqm’ of the flat.
- Manipulation
  - For this, I mainly grouped by whatever variable/factor I was investigating.
  - This was specific to the graph I would create and I will mention it when analysing the charts.

Analysis of resale prices by region and town

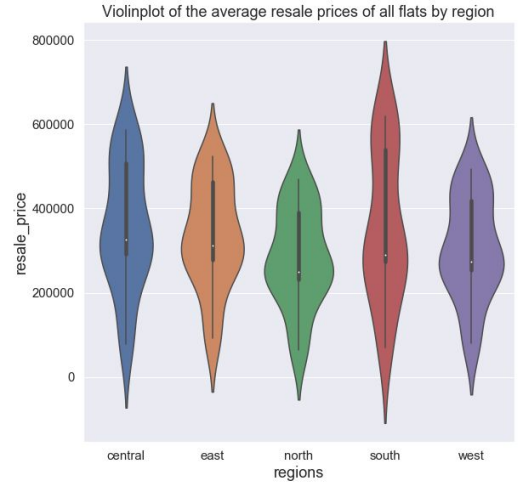
Manipulation:
For the figures 1 and 2, I grouped the resale prices by their regions. For figures 3 and 4, I then grouped by the town in the South Region.

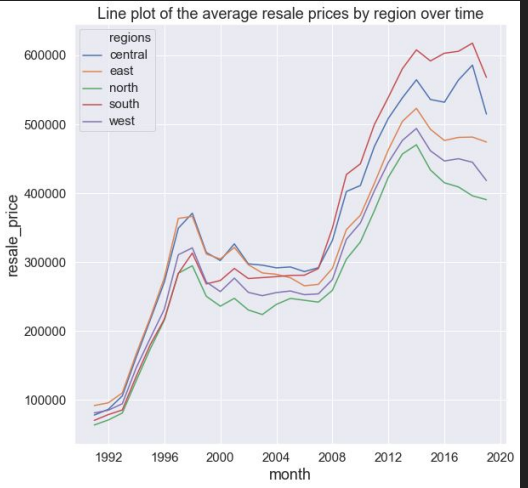
Analysis:

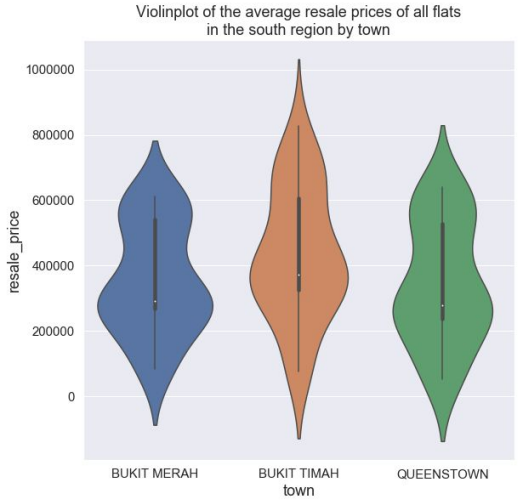
- Figure 2
  - All regions tend to have same trend
  - South over time clearly has a higher resale price
- Figure 1
  - South has the longest violinplot
  - Rest has a longer tail, due to low prices in the past
- Figure 4
  - Bukit Timah has the highest resale price
  - Trend is increasing
  - Average resale price in South are all quite high
- Figure 3
  - We see that the violinplot for Bukit Timah has is also the longest, likely due to it having the highest price

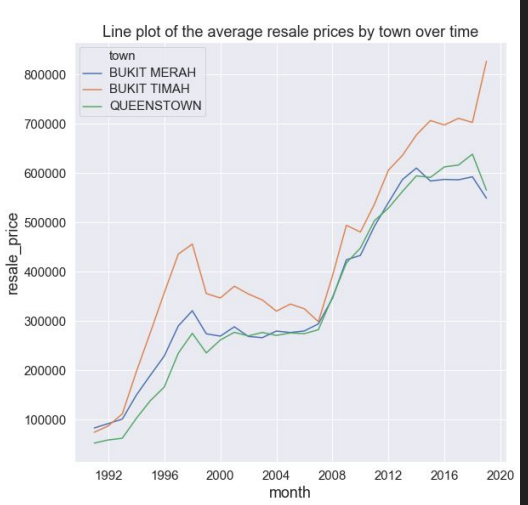
Conclusion:

- South has the highest demand
- Better to buy a flat in the South as all tend to have the same trend.
- Bukit Timah has the highest demand in the South
- Prices for all town in South seem to be quite high, indicates demand is high in general
- Prices would tend to remain high or even increase
- Buy a flat in Bukit Timah









Analysis of resale prices by flat type in Bukit Timah

Manipulation:

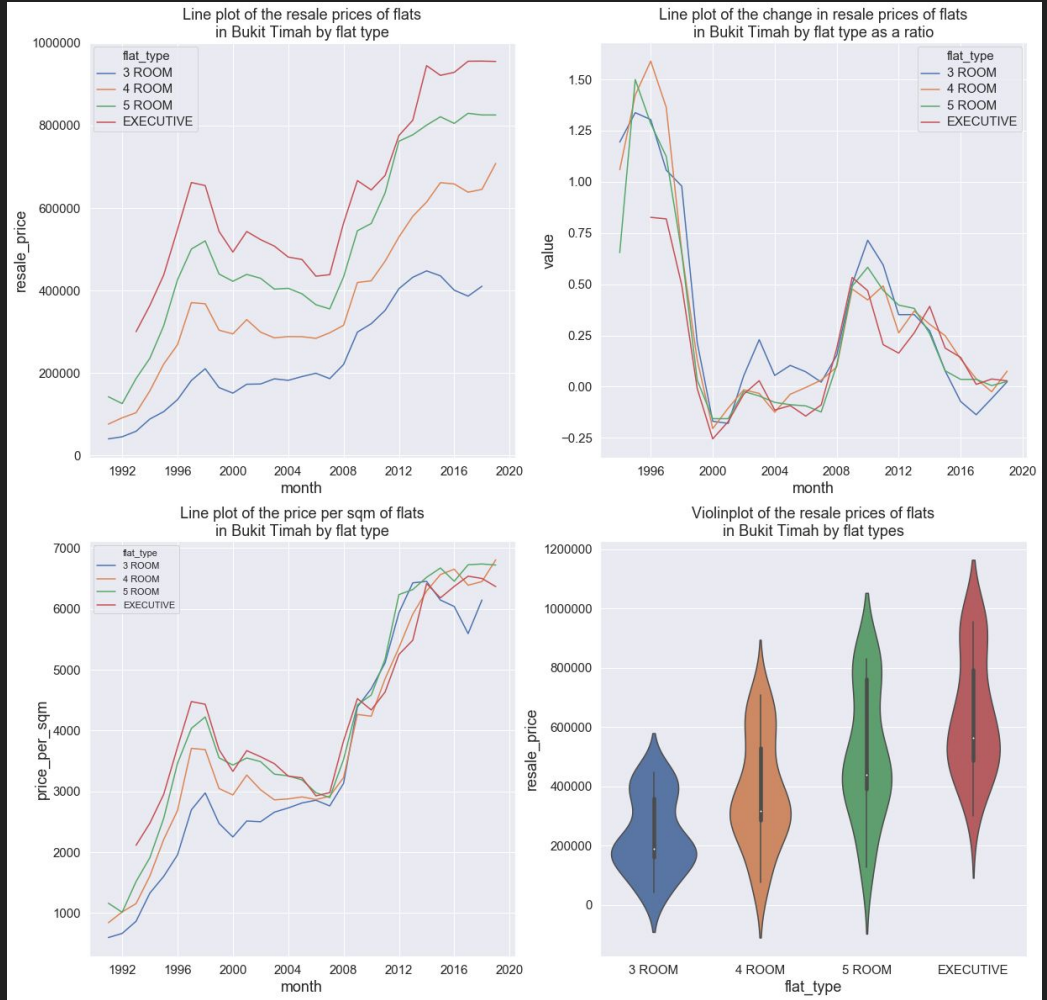
- Filter data to only flats in Bukit Timah
- Grouped by flat type
- Calculated the change in resale prices as a ratio(percentage)

Analysis:

- Figure 1
  - All flat types also seem to follow the same trend
  - Difference in prices is due to how large the flat is
  - However there is a change in recent years
  - Note how 5 rooms were almost as expensive as executive flats
  - The gap of 4 and 5 room flats decreased as well.
- Figure 2
  - We can see that the lines do somewhat follow the same trend until recently
  - Executive flats resale prices changes are also more volatile
- Figure 3
  - This is more looking at the worth of a flat
  - We see that now, the worth of a 4, 5 room and executive flat are around the same.
  - But a 3 room flat is much cheaper, but the prices are increasing
- Figure 4
  - This mainly reinforces figure 2 and 3, where the resale prices of 4, 5 and executive flats have the same shape.

Conclusions:

- 4, 5 and executive flats are worth around the same
- 3 Room flats are cheaper but worth lesser.
- 3 room flats also tend not to vary, not much as an investment
- Not considering executive, as it is mainly too expensive upfront
- Thus we are looking to buy a 4 or 5 Room flat in Bukit Timah



Analysis of number of lease years remaining affecting resale prices of 4 and 5 room flats in Bukit Timah

Manipulation:

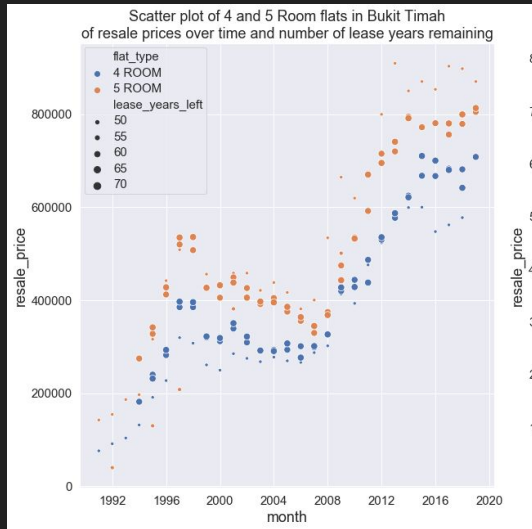
- Filter data to only 4 and 5 Room flats in Bukit Timah
- Grouped by lease years left

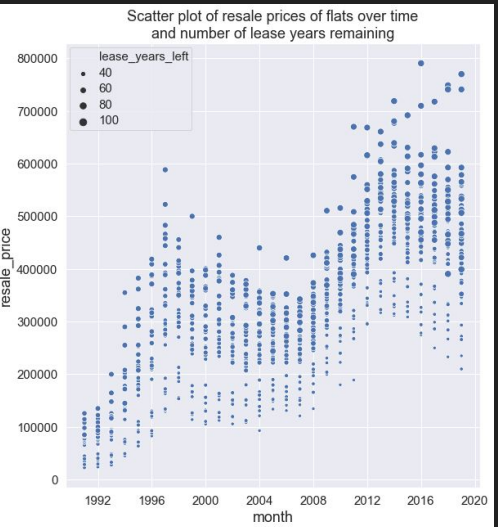
Analysis:

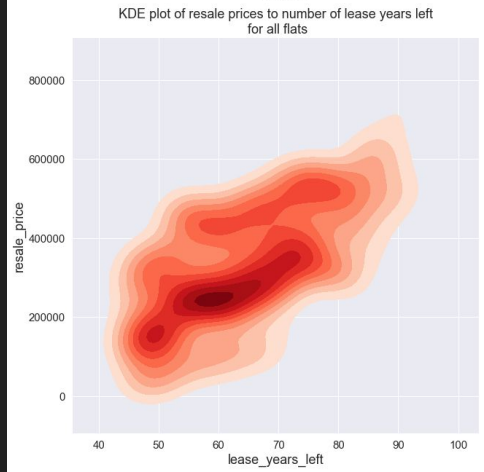
- Figure 1
  - We can see that for 5 Room flats, a flat with more lease years remaining is cheaper than a flat with lesser lease years remaining
  - The reverse is true for 4 Room flats where flats with more lease years are more expensive
  - Gap in high number of lease years left of 4 and 5 Room is not that much
- Figure 2
  - Generally, flats with a longer number of lease years left are more expensive. (Larger dots are higher)
- Figure 3
  - With a KDE plot of number of lease years left with resale price, we can see that flats with a smaller number of lease years left are generally cheaper.
  - Seen from shaded region tending towards a downwards left.
- Figure 4
  - KDE for 4 and 5 Rooms flats in Bukit Timah shows a shaded region being a lower right for the 5 Room flats, showing higher number of lease years are cheaper for 5 Room flats

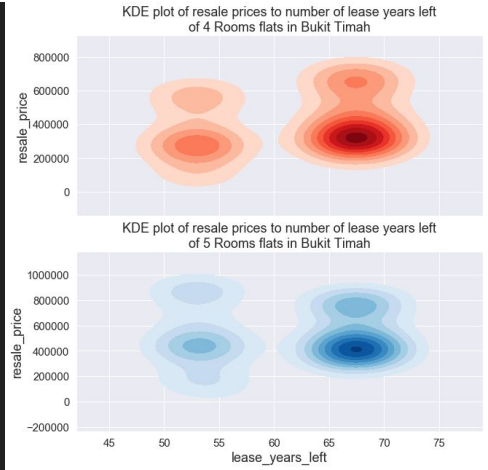
Conclusions:

- Cheaper to buy a younger 5 room flat.
- Compared to trends generally, this flat would be worth more, even if it is not worth as much in Bukit Timah.
- The old 5 room flat that is more expensive may be an outlier.
- Thus we would want to buy a younger 5 Room Flat in Bukit Timah.









Analysis of the storey of a flat affecting its resale price

Manipulation:

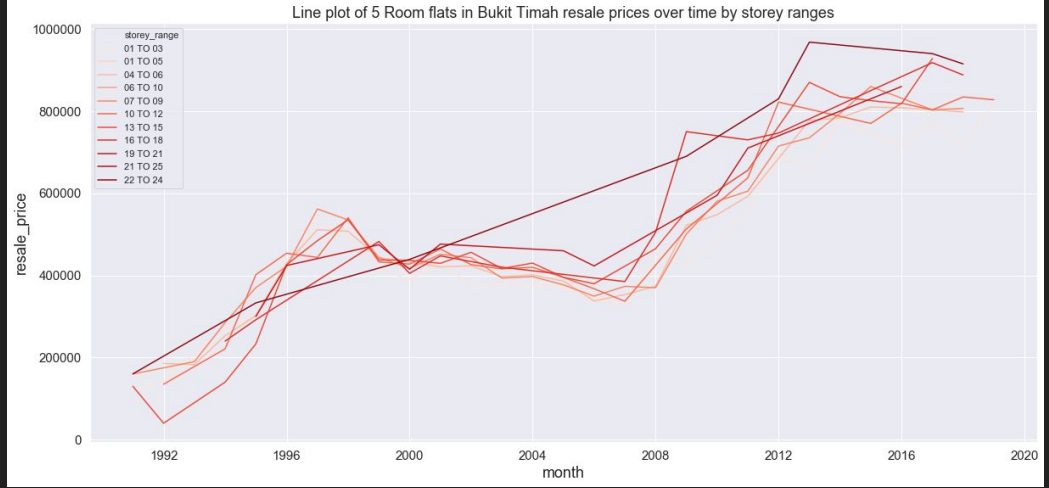
- Filter data to only 5 Room flats in Bukit Timah

Analysis:

- Figure 1
  - We can see a very clear trend that the higher the flat storey is, the more expensive it tends to be across Singapore.
  - However we can see that there are ‘groups’ of storeys that clump with each other.
- Figure 2
  - This trend is also reflected in 5 Room flats in Bukit Timah as well
  - However besides the outliers of the higher storey and lowest storey, the other storey seem to have generally the same prices
  - Also we see that the prices tend to really stick to the same trend, thus we probably would not expect much in a flats price to come from the storey.

Conclusions:

- The storey of the 5 Room flat would both come down to preference and how much more the family would be willing to pay.
- It might be a better investment to invest in the higher floors, but the data in Bukit Timah suggest they are around the same price, which could be that they are in the same storey ‘group’
- Thus we would pick 10-12 storey for us.



Final conclusion and recommendation for objective 1

Objective 1: What flat should you buy when starting a new family (Dataset 1)
Final recommendation: The family should buy a young 5 Room flat in Bukit Timah in th 10-12 storey range.

Reason from chart:

1. Bukit Timah in the South region seems to have the highest demand out of all regions
    a. As they all tend to have the same trend, a flat in Bukit Timah would be better with the higher demand.
2. Executive, 4 and 5 room flats tend to have the same worth
    a. Executive is excluded as most families cannot afford to buy a resale executive flat.
    b. Executive is also more volatile
    c. Thus worth of 4 and 5 Room is about the same
3. We can see that younger 5 Room flats are better than the rest
    a. Younger 5 room flats are cheaper than older ones
    b. For 4 Room flats, we can see that the gap to young 5 Room flats is much smaller
    c. General trend for all flats, is that younger flats are valued more, thus in the future it would not depreciate as much. Also may become valued more in Bukit Timah
4. Storey range in Bukit Timah does not seem to matter much
    a. Besides outliers the value and resale price of 5 Room flats in Bukit Timah does not seem to vary much
    b. They seem to be in the same group
    c. However, across Singapore, higher storeys are preferred, thus we will still get a high flat.

Sub-Objective: How does the flat type, town, storey and number of lease years remaining affect the resale price of a flat

1. It seems that the flat type and storey of a house are mainly categorical variables, where if the flat is in a particular category, their price will vary
2. As for the town and number of lease years remaining, this mainly depends on the demand in the town for these types of flats

Explaining the nature, cleaning and manipulation of the dataset. (Dataset 2 & 3)

Nature of the dataset 2

- Columns
  - 'blk_no', 'street', 'max_floor_lvl', 'year_completed'
  - Columns of residential units information
  - Columns of amenities information
  - 24 columns
- Rows
  - 12200 rows

Nature of the dataset 3

- Columns
  - 'financial_year', 'town', 'total_land_area', 'residential_land_area', 'Dwelling_units_under_management', 'projected_ultimate_dwelling_units'
  - 6 columns
- Rows
  - 266 rows

Cleaning and manipulating the dataset 2

- Cleaning
  - Changed the area code of ‘bldg_contract_town’ to the strings of the town
  - Used similar mapping from dataset 1 for creating regions
  - Changed the ‘Y’ and ‘N’ in the residential and amenities to 1 and 0
- Manipulation
  - Counted the sum and cumulative sum of amenities and residents by region and year
  - Calculated the ratio of amenities to residents
  - Grouped the data based on charts

Cleaning and manipulating the dataset 3

- Cleaning
  - Made all the strings in town to be upper case before creating the regions column
  - Fixed the null values in columns ‘total_land_area’ and ‘dwelling_units_under_management’ like ‘-’ and ‘na’
- Manipulation
  - Created a land use ratio and dwelling units to projected total dwelling units ratio

Analysis of dwelling units over time by regions (Dataset 3)

Manipulation:

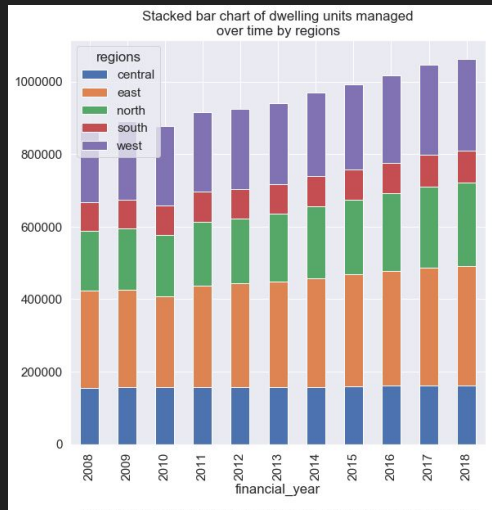
- Grouped by year and regions

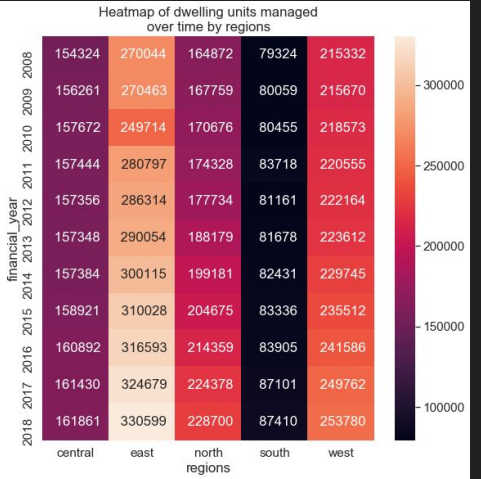
Analysis:

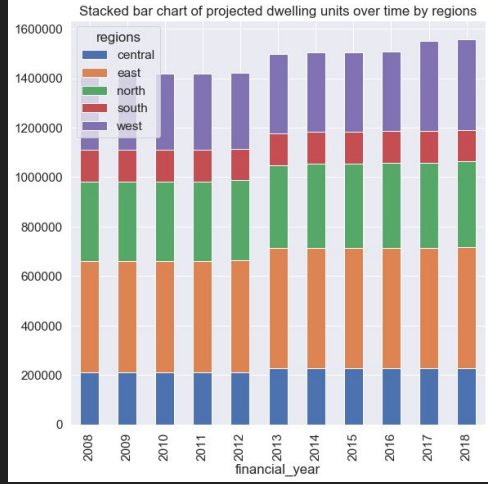
- Figure 1
  - Total number of people living in Singapore is increasing
  - We can see that the regions causing this are the, west, north and the east regions
- Figure 2
  - This confirms figure 1 analysis, where for west, north and east regions the color gradients get lighter downwards
  - This shows they had a greater increase
  - Based on the numbers we can see that the number of dwelling units across all regions also increased
- Figure 3
  - We see that the only increase in the projected number of dwelling units is for the east and west regions.
  - This happened in 2013 and 2017 respectively
- Figure 4
  - As for the ratio of number of dwelling units to projected number, we can see that there was an outlier of east and the north region, with a very small ratio
  - However over time, we notice the gradients for all regions turn to the same gradient of orange, indicating they had the same ratio (0.7) in the end.

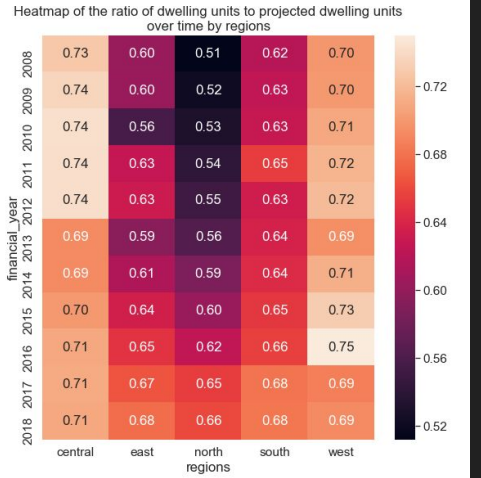
Conclusions:

- Overall, more people are choosing to live in Singapore as time goes on
- However, our projected number of people living in each region, although more than the population of people dwelling in each region, is not increasing.
- Also we see that the government wants to generally have the same ratio of people dwelling in each region to the projected number.
- Thus in future, they would also try and equally distribute the demand for dwelling units across the regions
- But what led to the increase in projection?









Analysis of land area used for residential property in Singapore by regions (Dataset 3)

Manipulation:

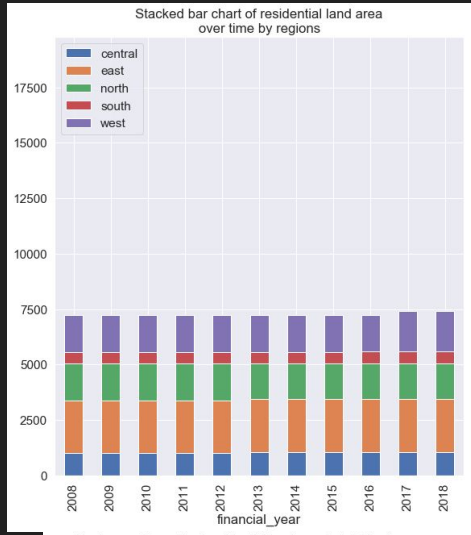
- Grouped by year and regions

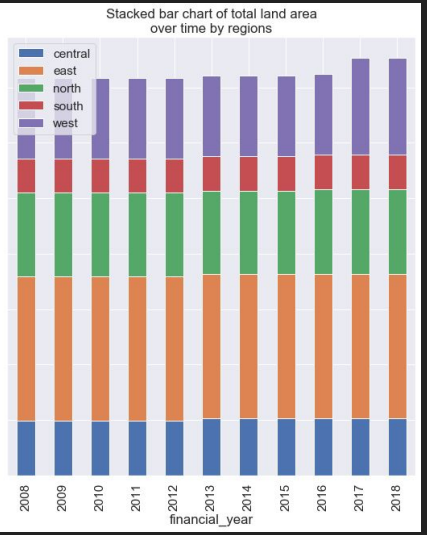
Analysis:

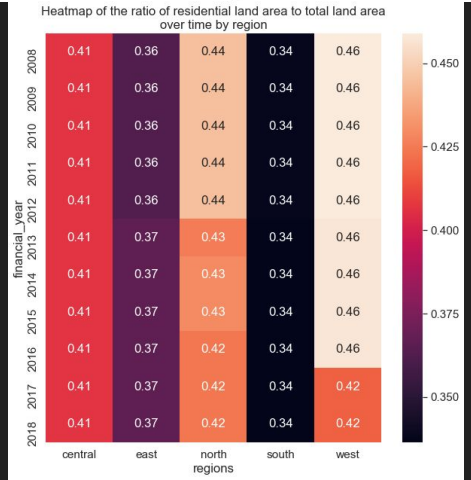
- Figure 1
  - We can see that the land used for residential property only increased for the west region
- Figure 2
  - This is also accompanied by the increase in total land area increase for the west region in the same years of 2017 as well
- Figure 3
  - We can see that the ratio of land used for residential property either remains the same or is decreasing.
  - This can be seen from north and west regions color gradient getting dark and the rest remaining the same

Conclusions:

- We can see that as the years goes by the land reserved for residential property will probably not increase any further unless the total area increases
- This means that as more people choose to live in Singapore, we could run out of land to build residential properties.
- We can also see that the increase in land use for residential property is not the reason that led to the increase in the projected number of dwelling units from the previous slide.
- Therefore the change in the ratio of dwelling units to the projected number is probably due to some dwelling units being relocated to other regions.
- This prevention of overpopulation in one region by the government is probably not to drastically raise the demand in any regions.







Analysis of the number of amenities and residential property in Singapore (Dataset 2)

Manipulation:

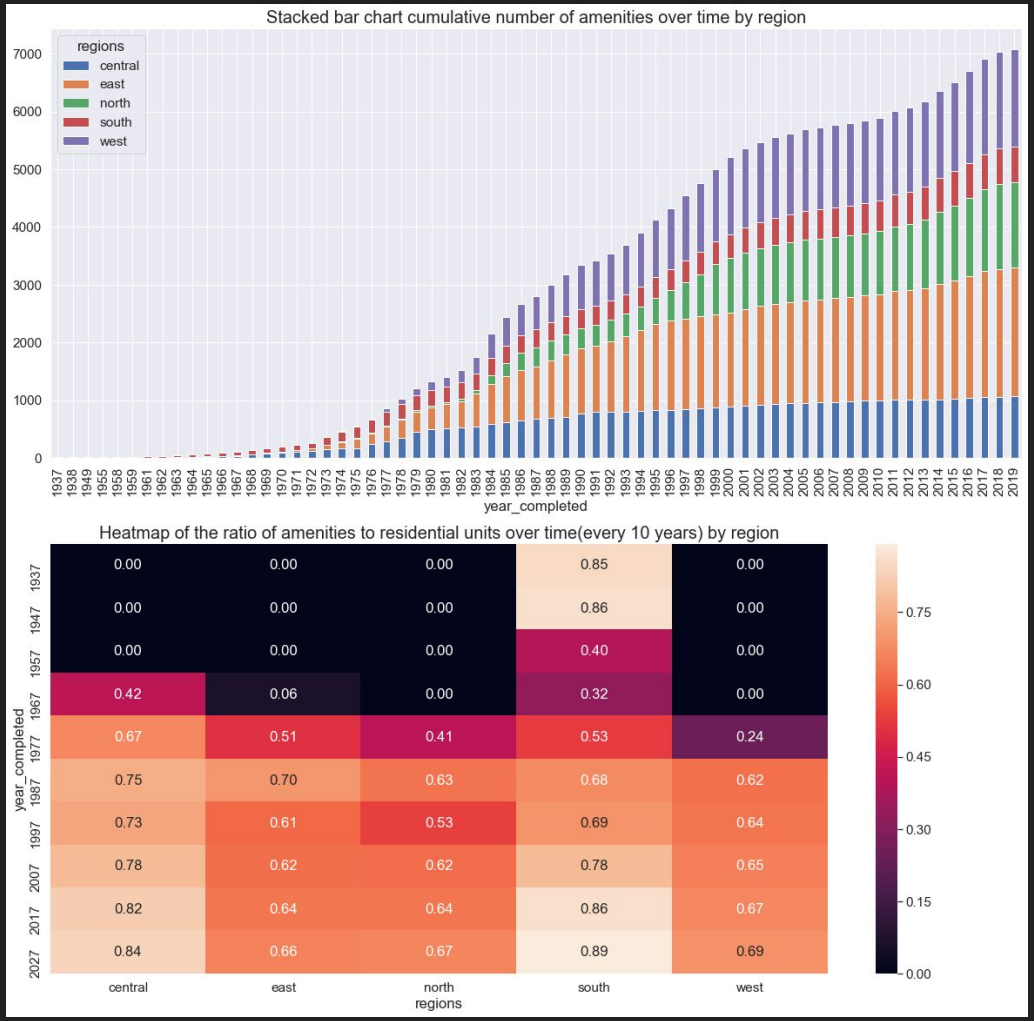
- Grouped by year and regions

Analysis:

- Figure 1
  - We can see that over time, some regions did not have amenities until 1975
  - We can also see that there was more emphasis placed on developing the east, west and north regions, in that order of priority
  - The total height of the bar chart is still increasing, suggesting that the total number of amenities is still increasing till today
- Figure 2
  - As this heatmap is a mean of every 10 years, we can see that the south region started out with amenities, but only after 1977 there were amenities in all regions
  - We can see that today, there are 2 groups of regions.
  - Central and south with a much higher ratio of 0.85 - 0.9
  - The rest, east, north and west with a ratio of 0.7

Conclusions:

- As more residential property and units are made, similarly, more amenities are created, keeping the ratio the same.
- From this we can see that the general ratio of amenities for a region is around 0.7, with outliers in the central and south regions.
- This increase in amenities ratio is probably due to these area being closer to the Central Business District(CBD) of Singapore thus a lot more development has occurred.
- We can also see that there is a very high effort placed to make sure this ratio is around 0.7 - 0.65 for amenities as from the start, the ratios were already quite close to this.
- Thus as the number of people living in Singapore increases, we should try and maintain these ratios



Final conclusion and recommendation for objective 2

Objective 2: Analysis of the situation of residential dwelling and amenities in Singapore (Dataset 2 & 3)
Final recommendation: The number of people choosing to live in Singapore is increasing, although Singapore has planned the residential units for them, we are not increasing the projected land use, thus we are approaching the maximum number of residents able to live in Singapore.
The ratio for amenities to residential property is around 0.7 except for areas near the CBD. Although alarming, we are not exceeding the government's expectation yet, so it is still fine as of now.

Reason from chart:

1. We can see that more people are choosing to live in Singapore
    a. The government seems to be spreading these people out to the different regions, keeping the ratio of residents dwelling to the projected number around the same across regions.
    b. However the projected number of dwelling units does not seem to be increasing with the number of dwelling units under management, suggesting we are approaching a maximum.
    c. Also since the projected number is not really changing, the ratio change is likely due to some people being reallocated to other regions to more equally distribute the ratio
2. The land allocated for residential property is also not increasing
    a. This mean that we do have a maximum amount of land allocated to residential property, only more will be allocated if the total land in the region increases.
3. The number of amenities in Singapore is continually increasing, but the ratio to residential property is still around the same
    a. As the number of residential property and units made and used increases, the number of amenities increase as well to accomodate for them.
    b. This ensures that there is sufficient amenities for the residents and the government is not over developing regions too fast

Sub-Objectives: How are residents distributed across Singapore?, Does each region have enough amenities?

1. Based on the projected number, there was a larger concentration in the central and west, although now it is much more evenly distributed
2. In the past, some regions had not developed yet, however as of now, based on the ratio, each regions should have enough amenities.