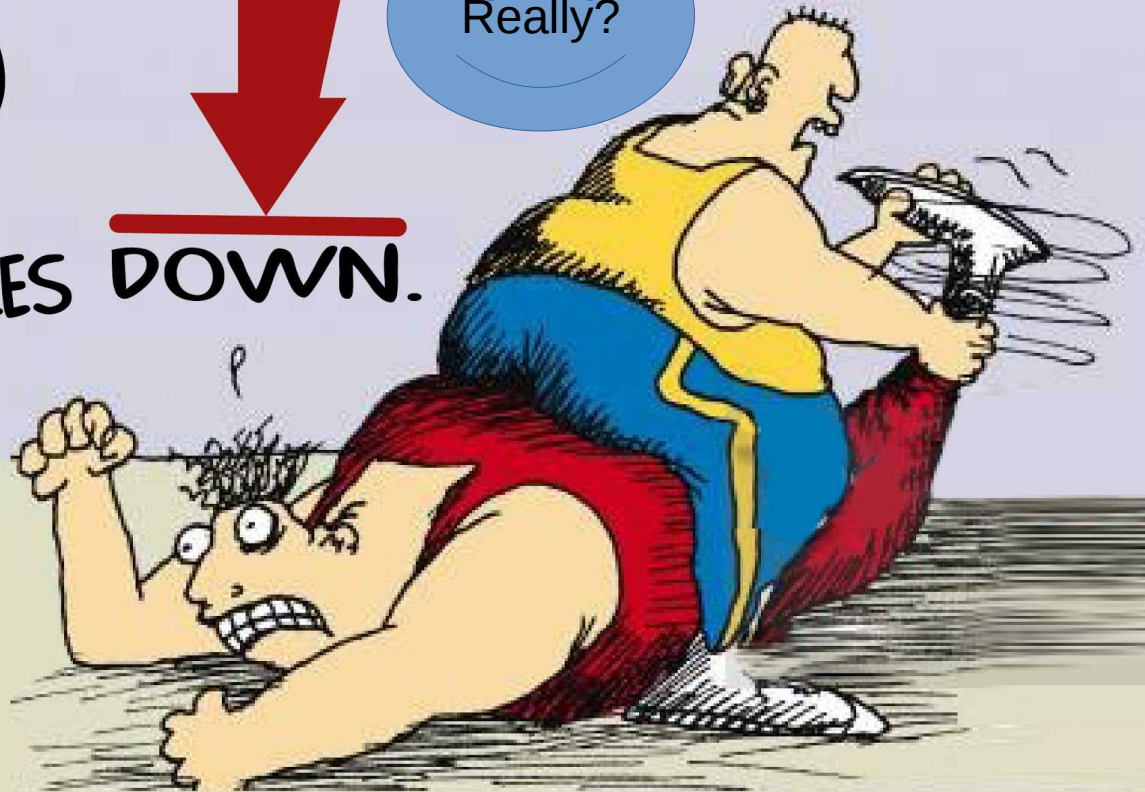


Resale HDB prices

BTO
KEEPS **HDB** PRICES DOWN.



SRX Property



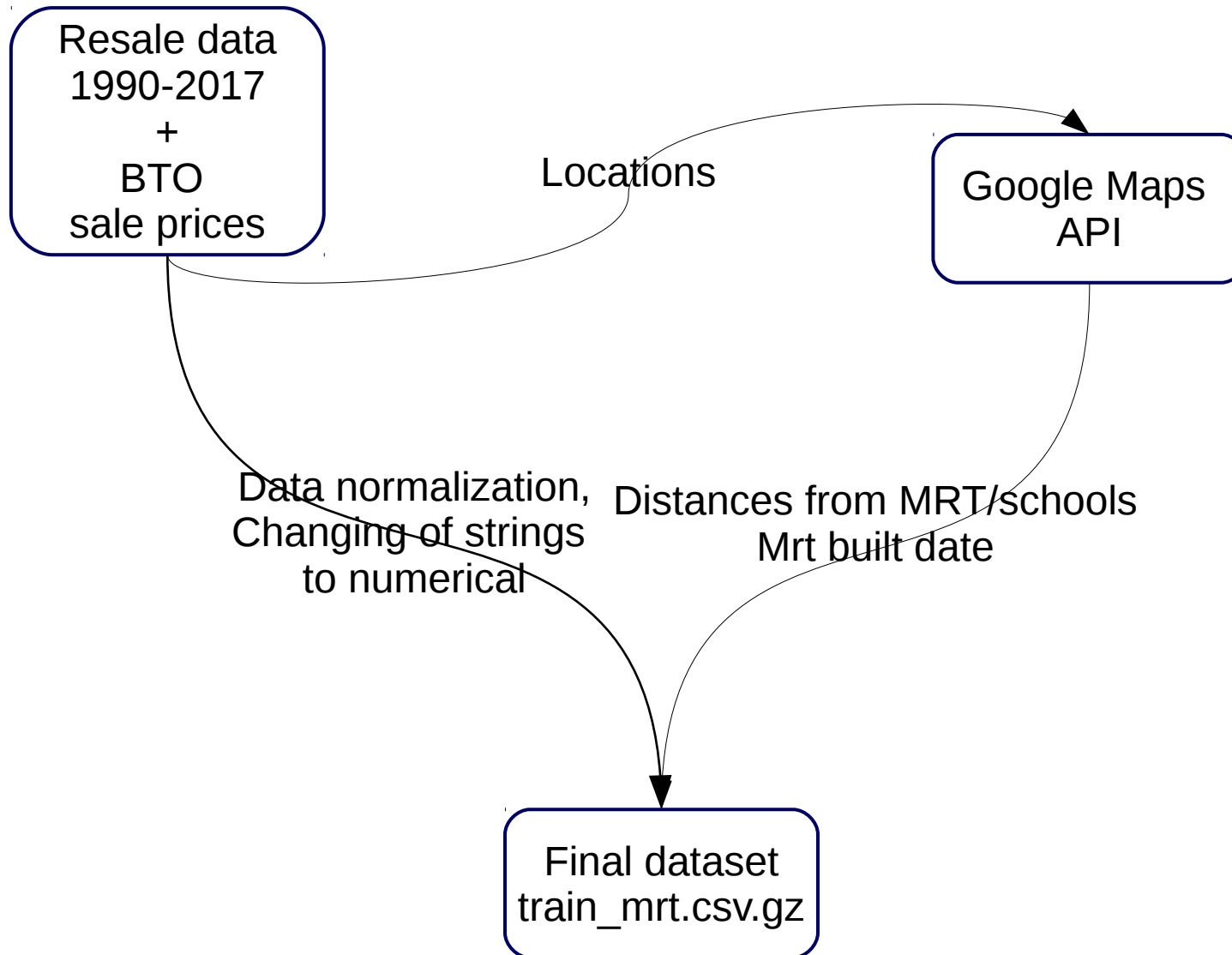
Why study HDB prices

- HDB is the largest expense in many people's life
- Study how prices of HDB changes over time
- Learn about the overall trend of HDB prices over the past 40 years

File list

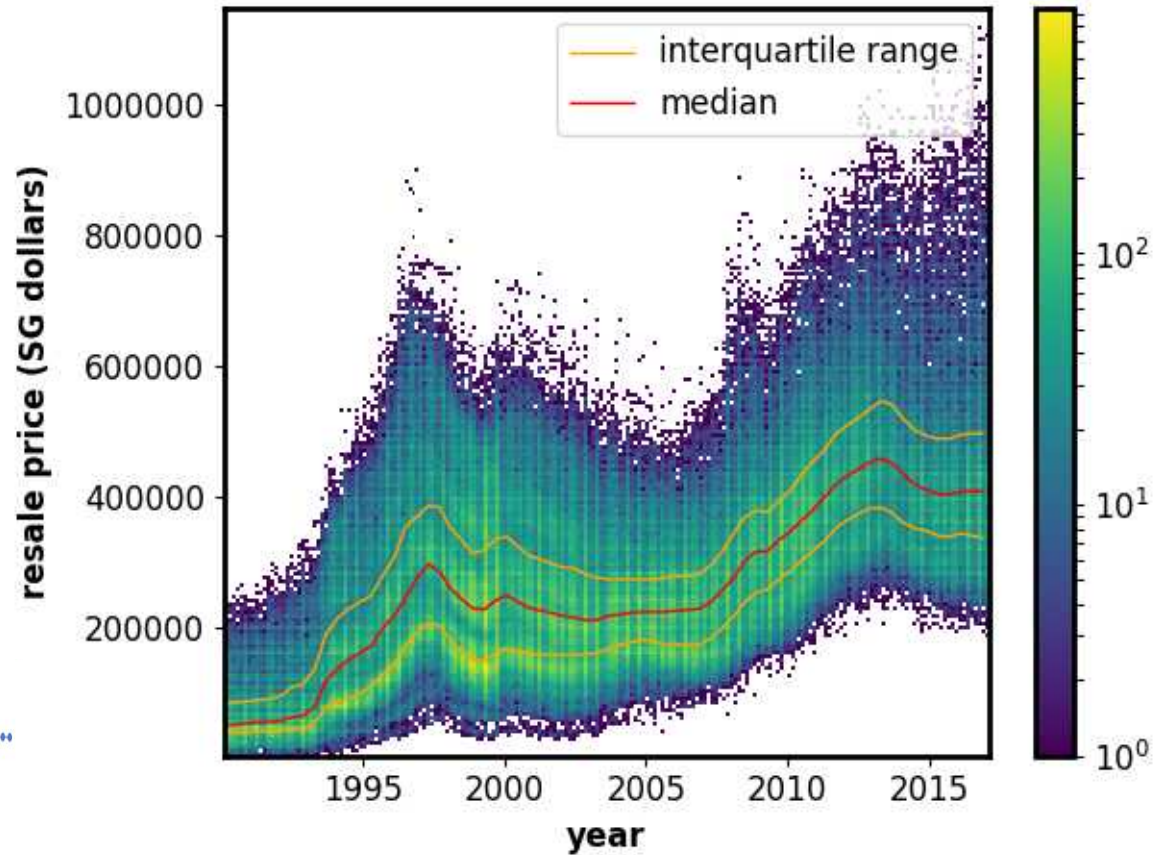
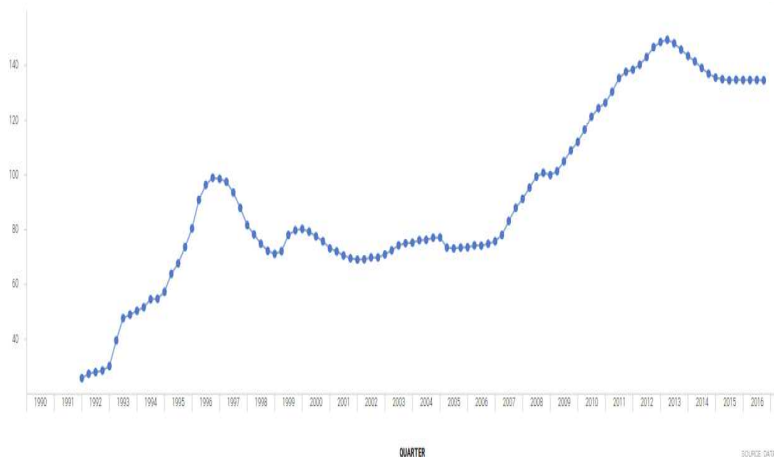
- Initial data file
resale-flat-prices-based-on-approval-date*.csv
- MRT/Primary school location files
Primary_school, mrt_date.csv, where.data_mrt
- Preprocessed files
train_mrt.csv.gz
- Scripts
notebook_preprocess.ipynb
notebook_analysis.ipynb

Data processing



General Prices of HDB

- General trend similar to one reported by gov.data.sg (below)



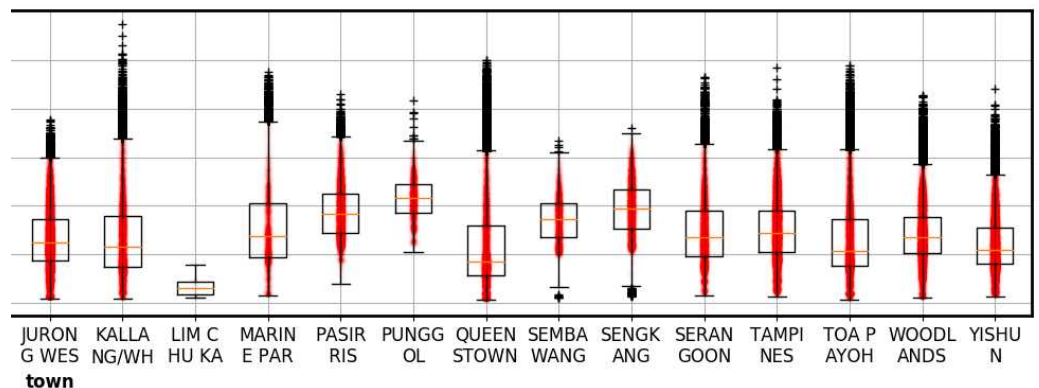
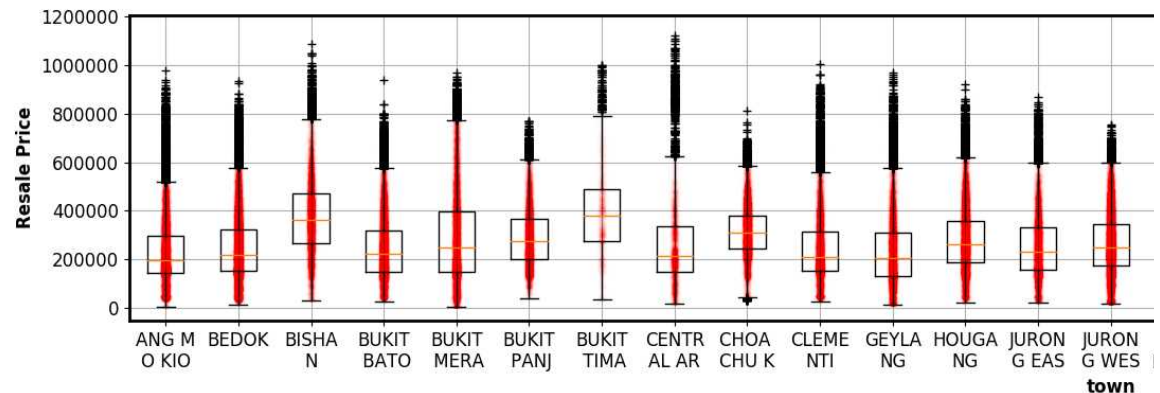
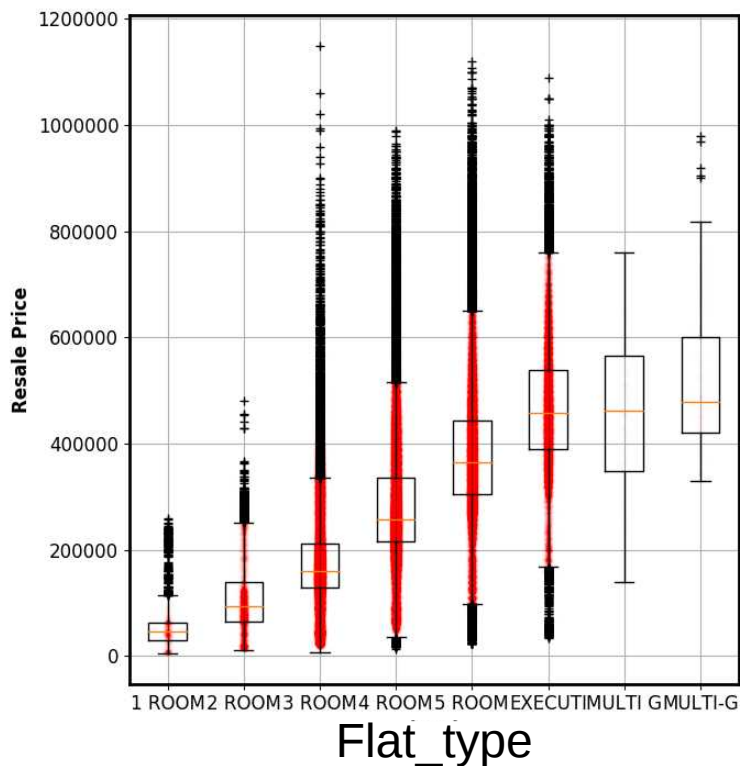
Main predictors of HDB flats

- Variables are month of sale, HDB town, flat type, block, street name, level of flat, floor area, lease commence date, nearest MRT, consumed lease length, distance to MRT, area,
- Correlation of numerical variables show that floor area and month of sale are most important numerical factors with pearson coefficient of 0.65 and 0.58 respectively.

```
>>> train[['month', 'storey_range', 'floor_area_sqm', 'dist_nearestMRT', 'Time_sinceMRTbuilt', 'lease_length', 'resale_price']].corr()['resale_price']
month                0.579766
storey_range         0.174023
floor_area_sqm       0.650346
dist_nearestMRT      0.021082
Time_sinceMRTbuilt    0.295540
lease_length         -0.033021
resale_price          1.000000
Name: resale_price, dtype: float64
```

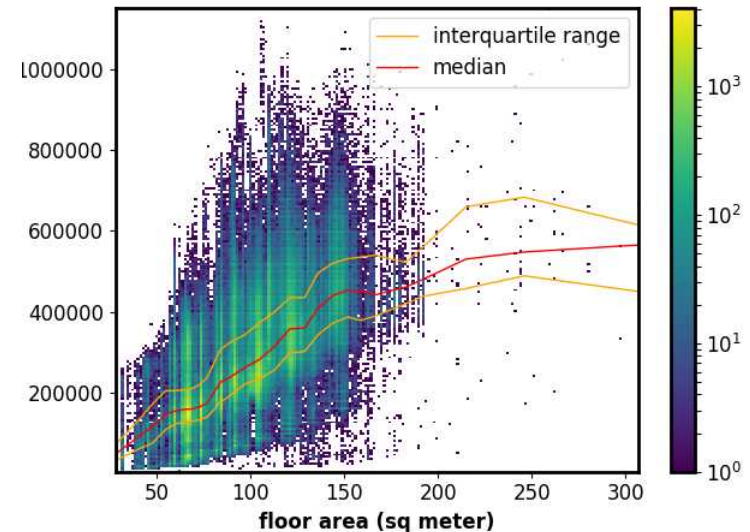
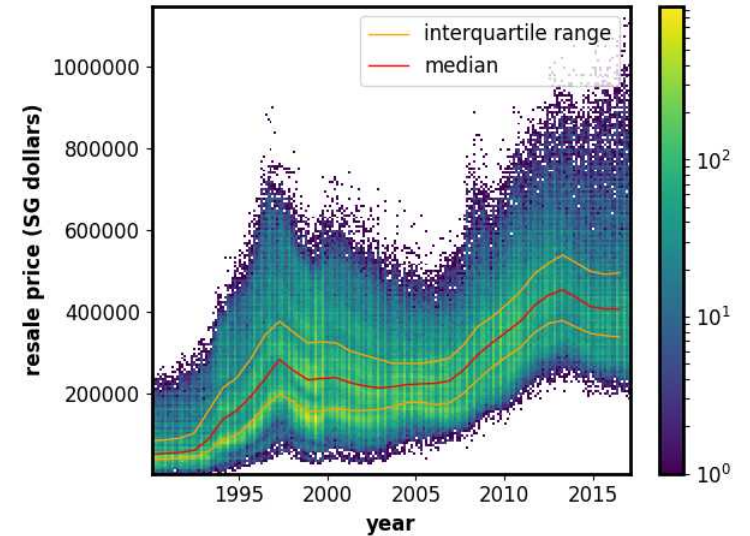
Most important categorical variables

- Flat type and town were import variables
- I will ignore flat type for this study and focus on town



Normalizing the price across years and floor area

- Prices are dependent on inflation and economic climates, we need to normalize across different years.
- Normalized price by area and median resale price per month to get relative price across years

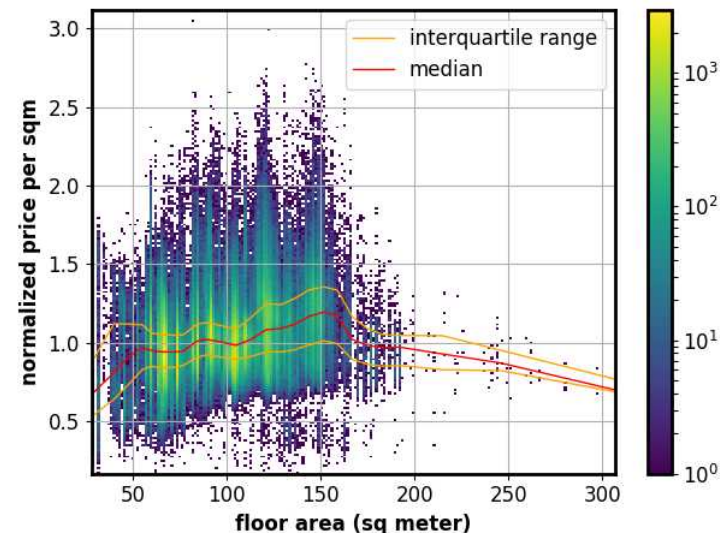
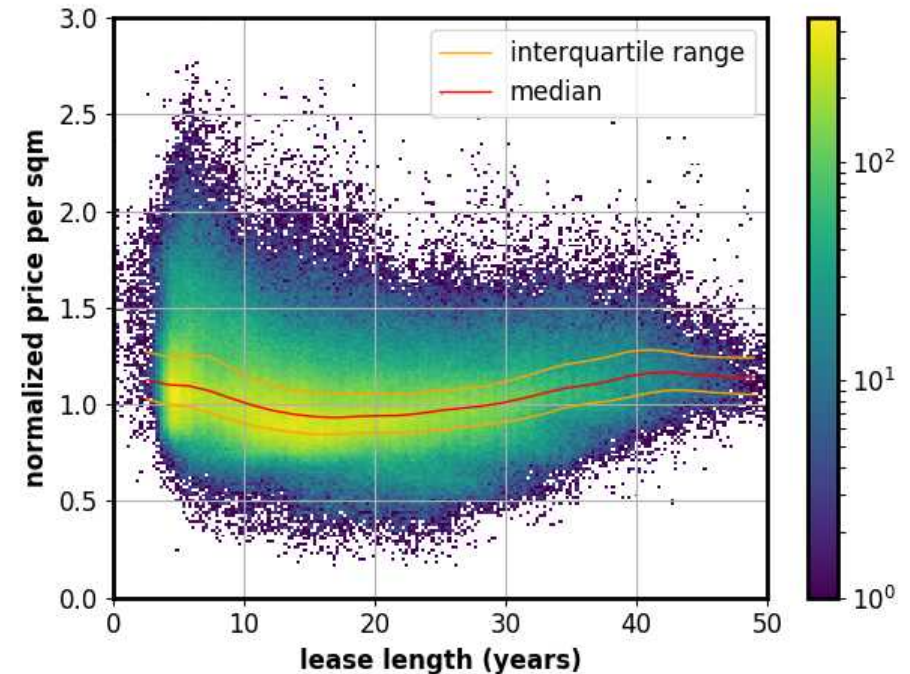


$$\text{Normalized Price}_{\text{month}} / \text{sqm} = \frac{\text{resale value}}{\text{floor area} * \sum_{\text{month}} \text{resale value} / n}$$

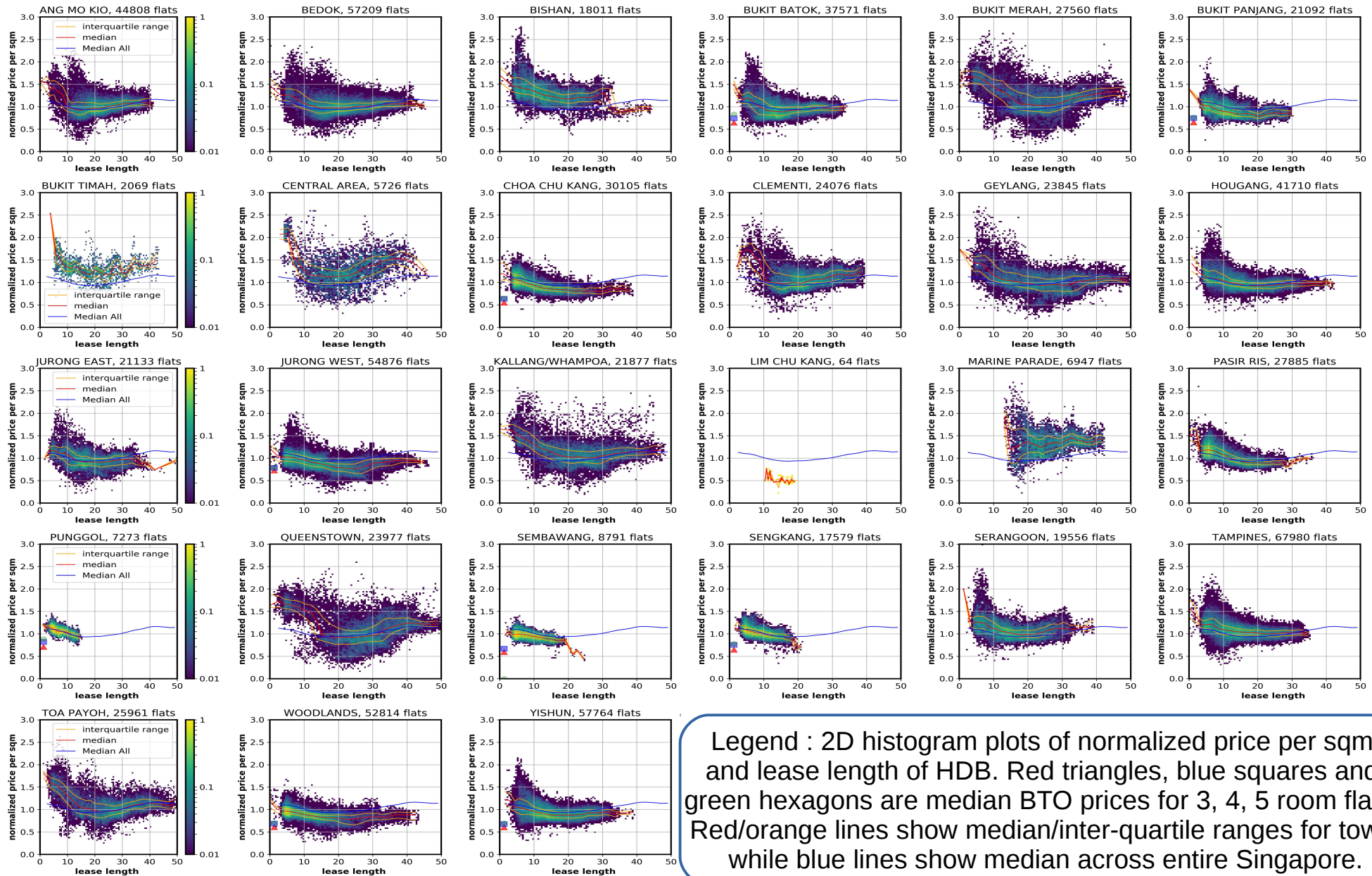
Average resale price for Month (or median)

Normalized price per sqm over time

- Peak prices : 5 year, lowest prices: 20 year accounting for inflation and floor area.
- Older flats above 40 years are sold for higher prices than newer flats
- Larger flats are more expensive per sqm, but I have decided to concentrate on other factors instead.



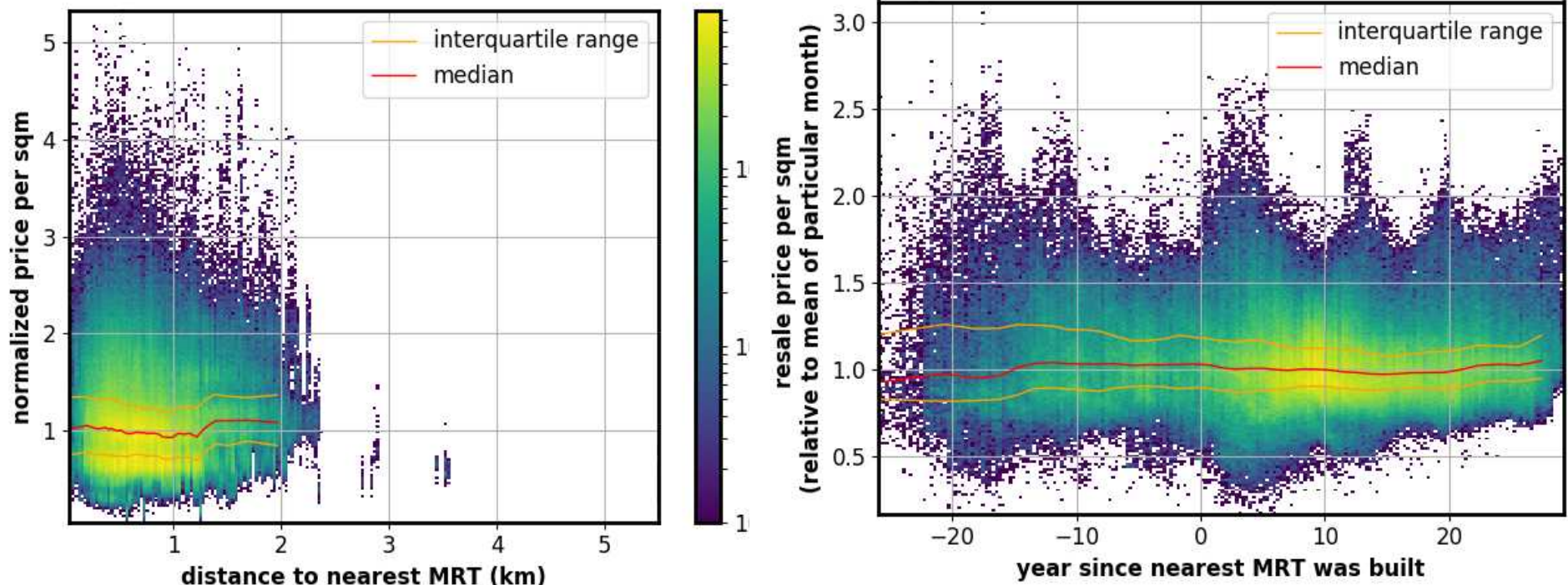
Normalized price per sqm over time (by town)



How do flats fare in pricing

- BTOs are lower than most resale flats even comparing with rock bottom prices at 20 year lease.
- Newer estates like Punggol and Sembawang Sengkang show much less variability compared to most mature estates like Ang Mo Kio, Toa Payoh and Queenstown.

MRT effect on HDB prices



- Building of MRT and distance to MRT does not impact prices per sqm.
- Houses further from MRT are generally a bit more expensive

Applications of Study

- Investigate trends of data using a normalized measure taking account inflation and floor area.
- This metric found that relative prices are lowest at twenty years and increases henceforth.
- Future resale prices can be forecasted by the formula below after getting an expert to estimate the median resale price in the future.

$$\text{Normalized Price}_{\text{month}} / \text{sqm} = \frac{\text{resale value}}{\text{floor area} * \sum_{\text{month}} \text{resale value} / n}$$

Median/Average resale price can be estimated by experts. This is affected by property market forces

Forecasting Data

- Previous slides give an intuition of prices, now for forecasting.

1)xgboost model

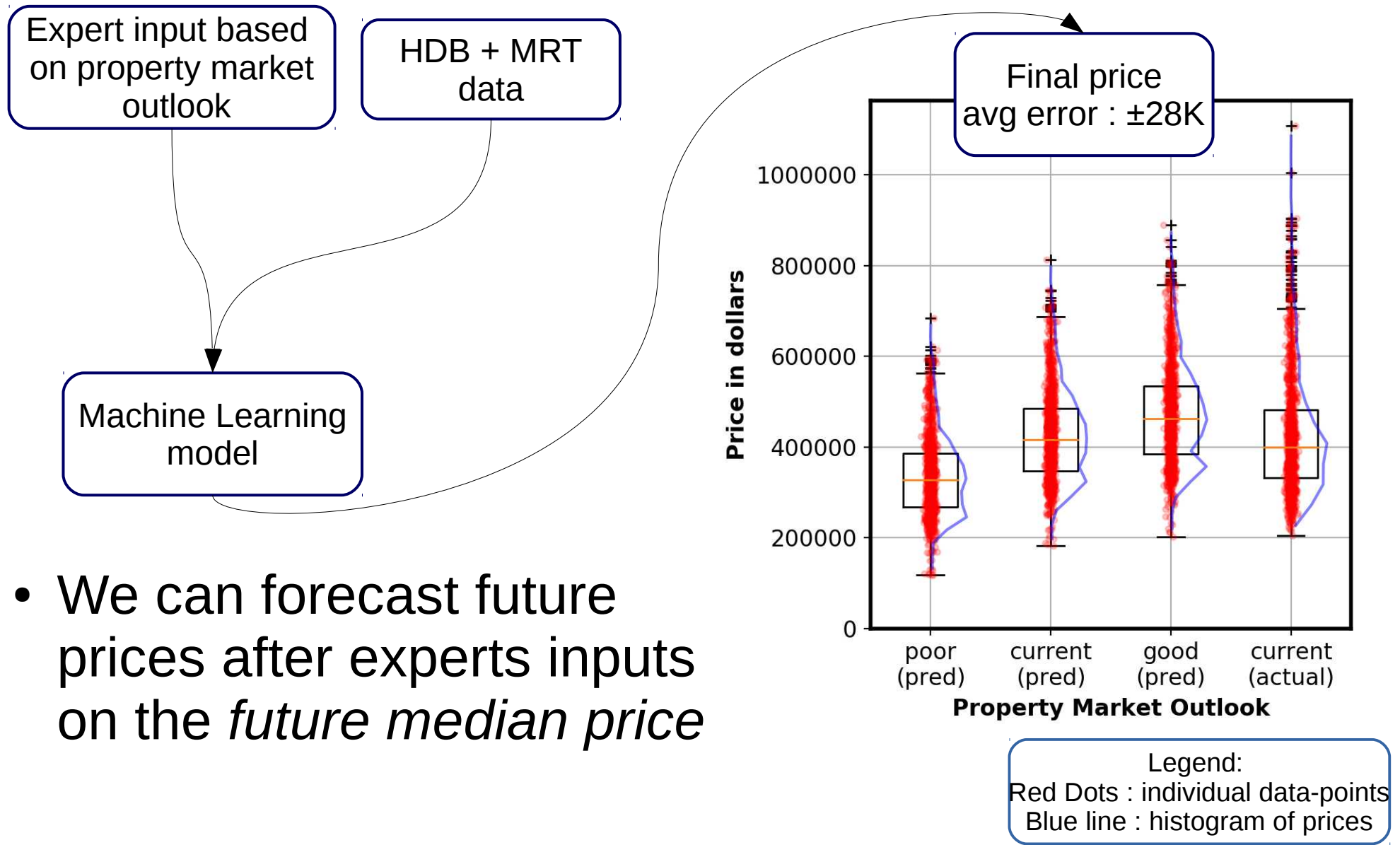
2)Train data :1990-2010

3)Test data :2010-2017

4)Features:

month_of_sale, town, flat_type, storey_range,
floor_area_sqm, lease_commence_date, lat, lng, nearestMRT,
dist_nearestMRT, MRTbuilt, Time_sinceMRTbuilt, month_mean,
lease_length

Application: Forecasting with expert inputs



Conclusion

- Showed certain trends regarding property prices related to lease, town and MRTs.
- Prediction accuracy of machine learning model was in the error of $\pm 28K$.
- Our model facilitated expert inputs to help predict future prices

Limitations of data

- Current data is based on resale data, which is not truly representation of entire HDB data.
- Size of BTO flat was estimated using medians since it was not given. The prices per sqm meter are thus also estimates.
- It is known that larger HDB flats cost more per sqm meter, this was not taken into account when plotting the prices per sqm per town.
- Forecasting uses mean of all flats sold in the period, which might be better represented by a stratified means of different flats.

Appendix: Have HDB sizes shrunk ?

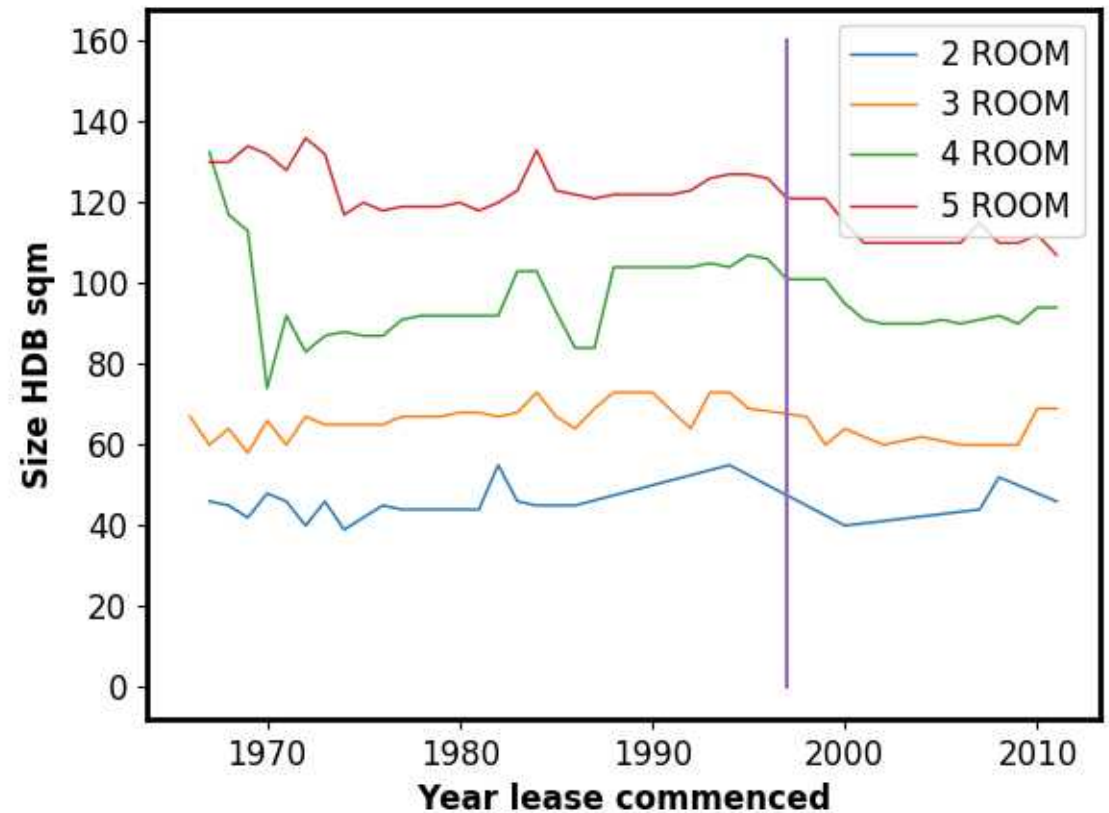
- HDB flat sizes not shrinking: Khaw Boon Wan

“My comment at that dialogue was in response to a question. I was purely stating that HDB plans (flat sizes) based on certain design norms, and as far as I know, it has not changed for the past 15 years,”

Posted by temasektimes on June 13, 2012

Appendix: HDB resale data

- Taking 15 years back, to 1998 (purple line)
- Flat sizes have been decreasing since 1998, with the decrease mainly from 1998 to 2001
- Flat sizes have not decreased for 10 years since 2001-2012



Appendix : Data sources

- <https://data.gov.sg/dataset/resale-flat-prices>
from year 1990-2017
- Google maps API data for geospatial coordinates
- Wikipedia for MRT and primary school data