NANYANG TECHNOLOGICAL UNIVERSITY
SINGAPORE

Predicting HDB Resale Prices in Singapore

Thomas ten Hacken, Maxime Kayser & Mei-Jun Yeh

April, 2018

Introduction

- Previous studies suggest the use of neural networks in open markets
- HDB is not a true open market
- Dataset from SG government: resale prices from 1990 – jan 2018

What predictive technique is most suitable for the HDB resale market in SG?

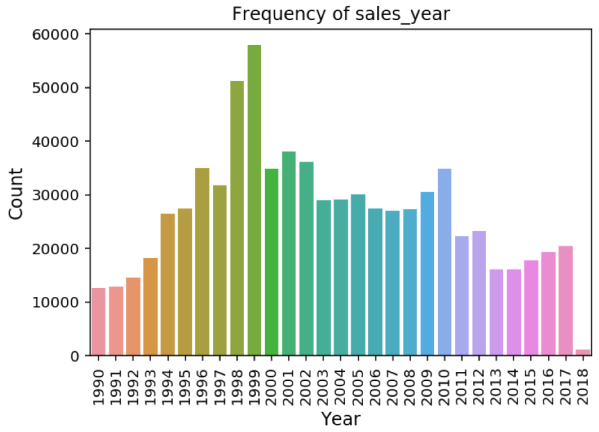
month	town	flat_type	block	street_name	storey_range	floor_area_sqm	flat_model	lease_commence_date	resale_price
1990-01	ANG MO KIO	1 ROOM	309	ANG MO KIO AVE 1	10 TO 12	31.0	IMPROVED	1977	9000

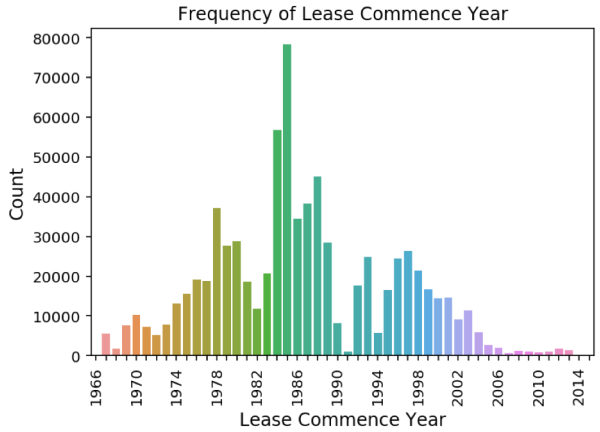
```
data.shape(768629,10)
```

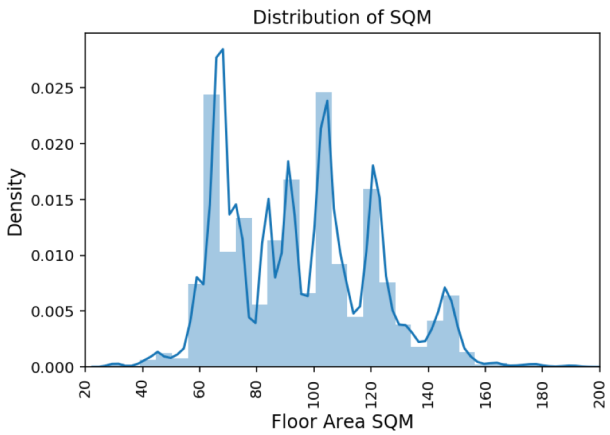
NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Data Exploration

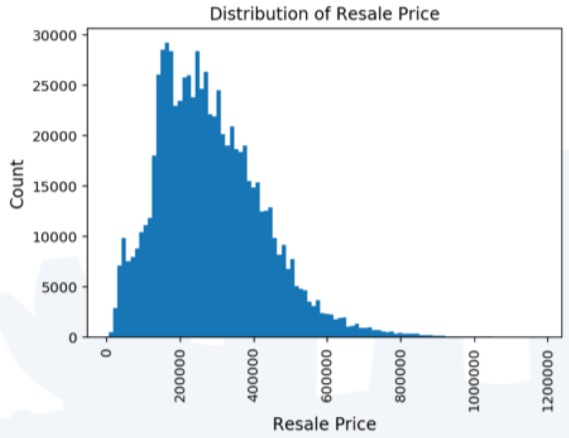
Features Exploration

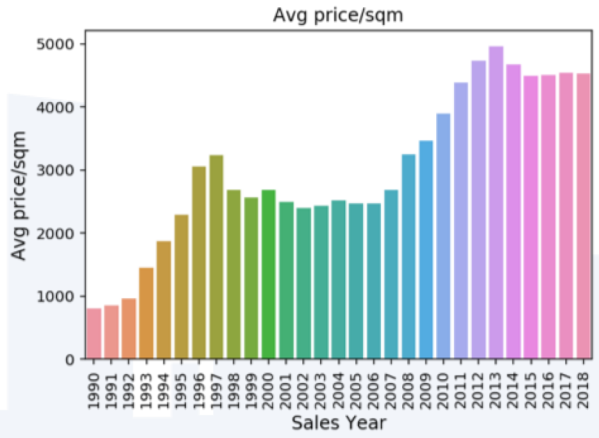


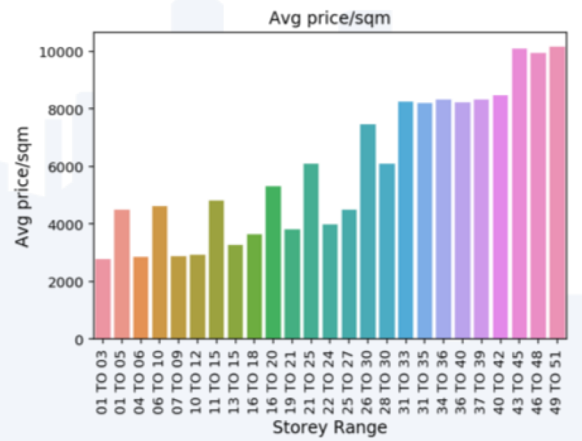




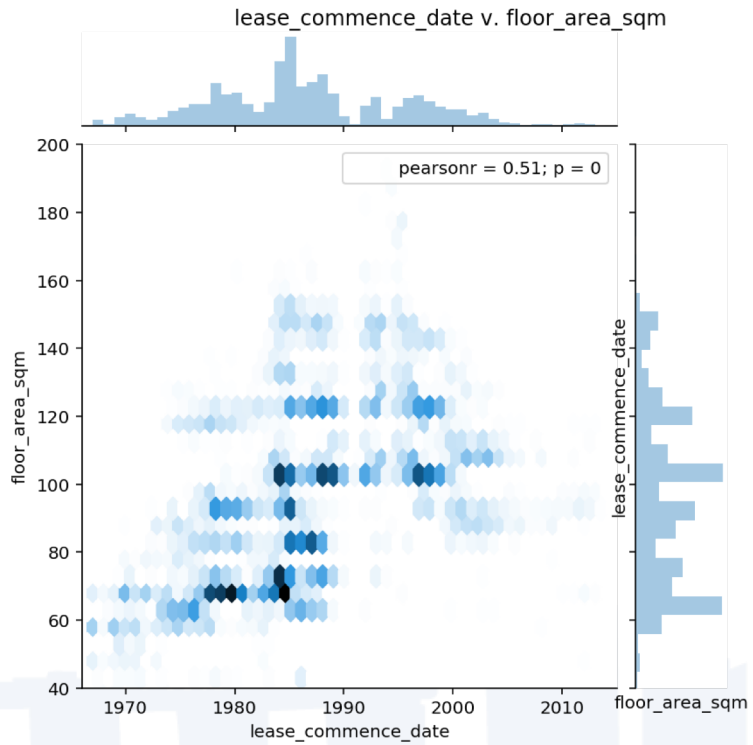
Output

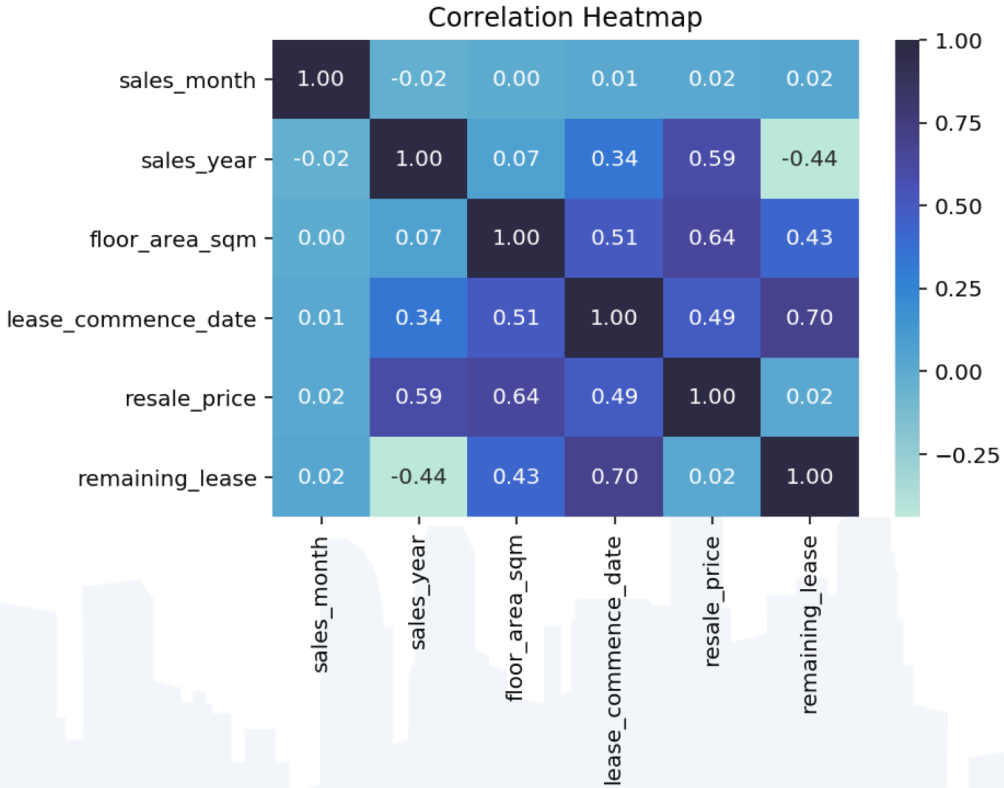






Data Exploration: Correlation





NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Data Analysis: MAE

	Linear Regression	Random Forest	Gradient Boost Regressor	XGBoost	Neural Network	Ada Boost
Encoding	54199	15928	28375	32849	71294	78036
One-Hot Encoding	47967	15837	28961	32464	60630	79460
+ Remaining lease year	47969	15860	28812	32985	-	79226
+ Area	47966	15722	26157	29446	-	77135
+ GDP per Capita	41710	15748	26074	29181	-	-
+ Land mass, - GDP per Capita	42056	15705	26157	29446	-	-
+ Demand, - Land Mass	46704	15732	26143	29479	-	-
Data Normalization	47972	15706	26157	29446	-	75663
Max-Min Normalization	47968	15717	26157	29446	-	77430
One-Hot Encoding Month & Years	29559	16045	36117	44152	-	9675

Conclusions

- Random Forest provides best prediction
- Scenario: adding two features, namely GDP per capita and land mass
- R squared of 97 %

- Cross validation showed no overfitting

Limitations & Future Research

Limitations

- Geocoder (Google API requests)
- Governmental influence

Future Research

- Data features (e.g. distance to MRT)
- Private market comparison
- Similar studies in other countries

… also available on GitHub!

Click here