

zenius

Kampus
Merdeka
INDONESIA JAYA

Final Project Presentation

Nomor Kelompok: 2

Nama Mentor: Diajeng Ciptaning Ayu

Nama:

- Yusuf Kamal
- Sinta Asanah

Machine Learning Class

Program Studi Independen Bersertifikat
Zenius Bersama Kampus Merdeka



Petunjuk

- Waktu presentasi adalah 5 menit (tentatif, tergantung dari banyaknya kelompok yang mendaftarkan diri)
- Waktu tanya jawab adalah 5 menit
- Silakan menambahkan gambar/visualisasi pada slide presentasi
- Upayakan agar tetap dalam format poin-poin (ingat, ini presentasi, bukan esai)
- Jangan masukkan *code* ke dalam slide presentasi (tidak usah memasukan screenshot jupyter notebook)

1. Latar Belakang
2. Explorasi Data dan Visualisasi
3. Modelling
4. Kesimpulan

Latar Belakang

Latar Belakang Project

Sumber Data:

<https://www.kaggle.com/datasets/teyang/singapore-hdb-flat-resale-prices-19902020>

Problem: **regression**

Tujuan:

- Untuk mengetahui faktor apa yang paling mempengaruhi tinggi rendahnya harga properti dan pergolakan trend serta volume transaksi properti dari tahun 2016 sampai tahun 2020.

Explorasi Data dan Visualisasi

Business Understanding

HDB (Lembaga Perumahan dan Pembangunan) adalah sebuah lembaga negara yang berada di bawah Kementerian Pembangunan Negara yang bertugas di bidang perumahan publik di Singapura. Menurut HDB, pengembangan perumahan ini berjasa mengubah Singapura yang pada tahun 1960-an dipenuhi gubuk-gubuk dan penghuni liar menjadi perumahan yang teratur dan berkepadatan tinggi

Warga dengan pendapatan keluarga melebihi batas tertentu tidak boleh memiliki rumah HDB (misal, pembeli apartemen 2-kamar atau 3-kamar standar tidak boleh memiliki pendapatan bulanan di atas S\$5.000). Tujuannya adalah memprioritaskan perumahan publik untuk warga yang lebih membutuhkan.

Lebih dari 80% populasi penduduk di Singapura tinggal di flat HDB. Flat atau rumah susun HDB di Singapura dijual dengan perjanjian sewa 99 tahun. Ada beberapa lokasi rumah susun HDB di Singapura, salah satunya adalah di kawasan Yishun. Dikutip dari laman propertyguru.com.sg, harga rumah susun di kawasan Yishun tergantung dari besaran unit dan lokasi bloknya.

Sumber : Wikipedia

grid.id : (www.grid.id/read/04147790/ternyata-segini-lho-harga-rumah-susun-yang-ditinggali-presiden-terpilih-singapura-halimah-yacob?page=all)

Data Cleansing

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1187 entries, 30680 to 80366
Data columns (total 11 columns):
#   Column              Non-Null Count  Dtype
---  -
0   month               1187 non-null   object
1   town                1187 non-null   object
2   flat_type           1187 non-null   object
3   block               1187 non-null   object
4   street_name         1187 non-null   object
5   storey_range        1187 non-null   object
6   floor_area_sqm      1187 non-null   float64
7   flat_model          1187 non-null   object
8   lease_commence_date 1187 non-null   int64
9   remaining_lease     1187 non-null   object
10  resale_price        1187 non-null   float64
dtypes: float64(2), int64(1), object(8)
memory usage: 111.3+ KB
```

Dari **informasi data**, terlihat tidak ada data dengan non-null. Selain itu, data terdiri dari 3 tipe data :

1. Tipe data float64 yaitu kolom `floor_area_sqm` dan `resale_price`
2. Tipe data int64 yaitu kolom `lease_commence_date`
3. Tipe data object terdapat 8 kolom

Data Cleansing

month	22
town	14
flat_type	4
block	140
street_name	27
storey_range	16
floor_area_sqm	25
flat_model	4
lease_commence_date	4
remaining_lease	28
resale_price	393

Dari informasi mengenai data unik di atas, dapat diartikan bahwa variabel ``month`` memiliki 22 kategori, variabel town memiliki 14 kategori, variabel ``block`` memiliki 140 kategori dst

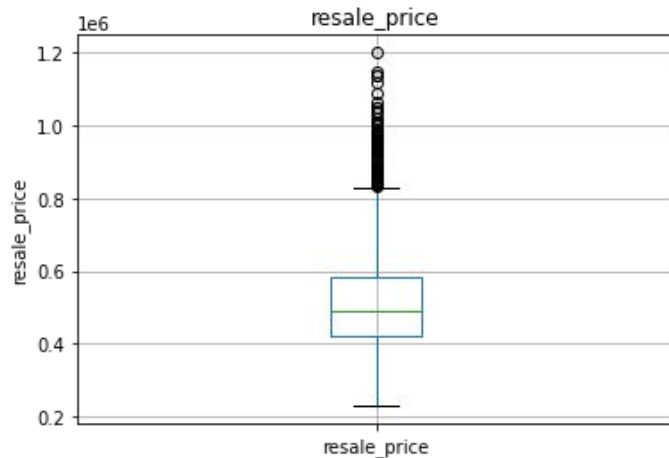
Data Cleansing

	floor_area_sqm	lease_commence_date	resale_price
count	1185.000000	1185.000000	1.185000e+03
mean	93.607595	2016.011814	5.237517e+05
std	15.219302	0.141882	1.582689e+05
min	47.000000	2016.000000	2.300000e+05
25%	92.000000	2016.000000	4.200000e+05
50%	93.000000	2016.000000	4.900000e+05
75%	112.000000	2016.000000	5.850000e+05
max	117.000000	2019.000000	1.200000e+06



Statistika deskriptif untuk variabel numerik

Data Cleansing

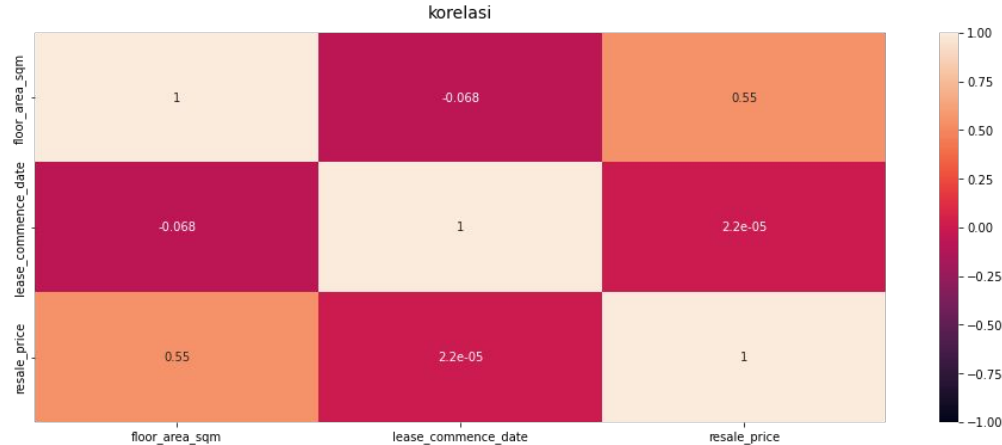


Berdasarkan boxplot di atas, dapat dikatakan bahwa:

1. Boxplot `'floor_area_sqm'` terdapat satu outlier bawah
2. Boxplot `'lease_commence_date'` terdapat beberapa outlier atas
3. Boxplot `'resale_price'` terdapat beberapa outlier atas

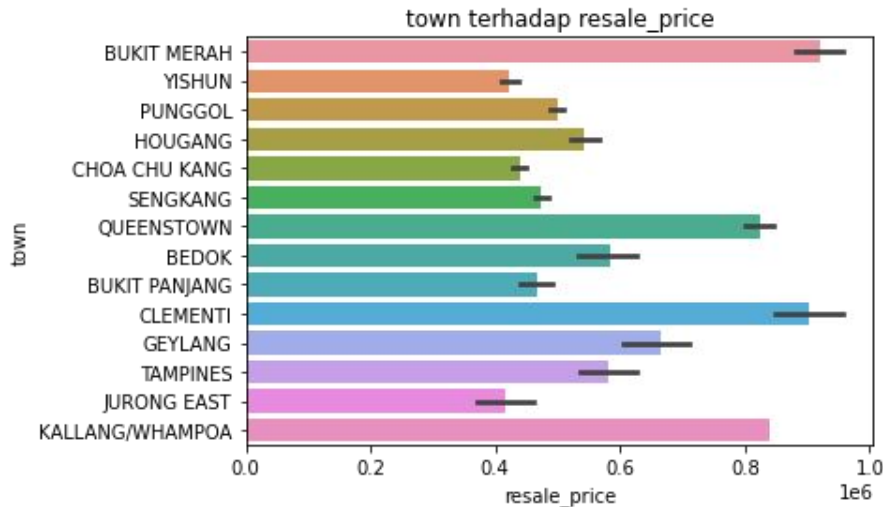
Outlier pada data 'df' ini tidak dilakukan penanganan, khususnya untuk variabel `'resale_price'`. Karena data outlier yang berada pada variabel `'resale_price'` merupakan hal yang wajar untuk nominal resale price of flat yang tinggi.

Exploratory Data Analysis



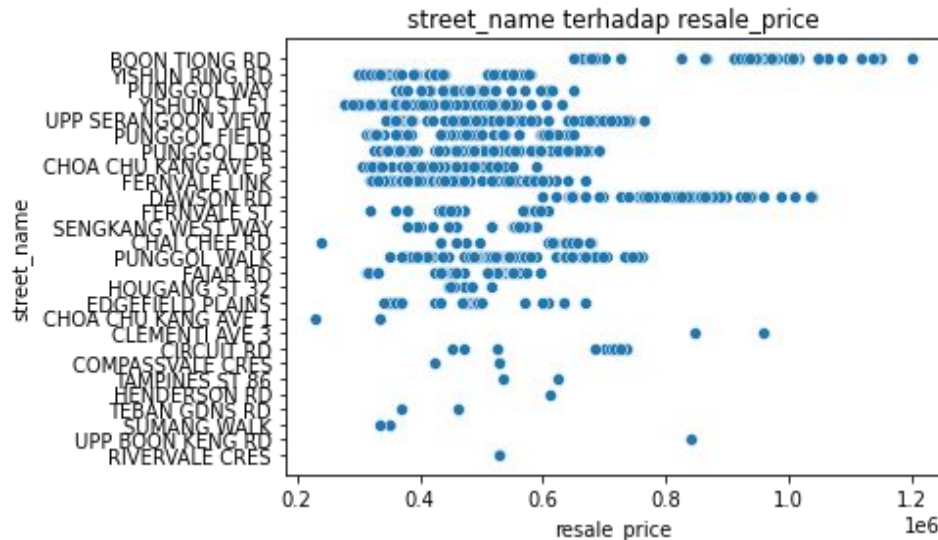
Berdasarkan nilai korelasi untuk variabel numerik, terlihat bahwa ``resale_price`` dengan ``lease_commence_date`` memiliki nilai korelasi yang searah atau positif dengan nilai korelasi sebesar 0.55. Artinya ``resale_price`` dengan ``lease_commence_date`` memiliki hubungan yang kuat.

Exploratory Data Analysis



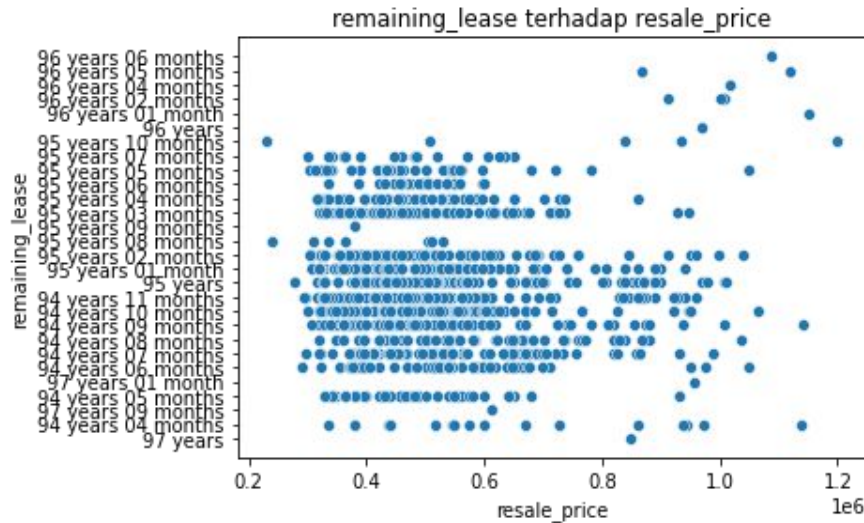
Berdasarkan Barplot ``resale_price`` dengan ``town``, ``resale_price`` tertinggi berada di QUEENSTOWN dan ``resale_price`` terendah berada di JURONG EAST.

Exploratory Data Analysis



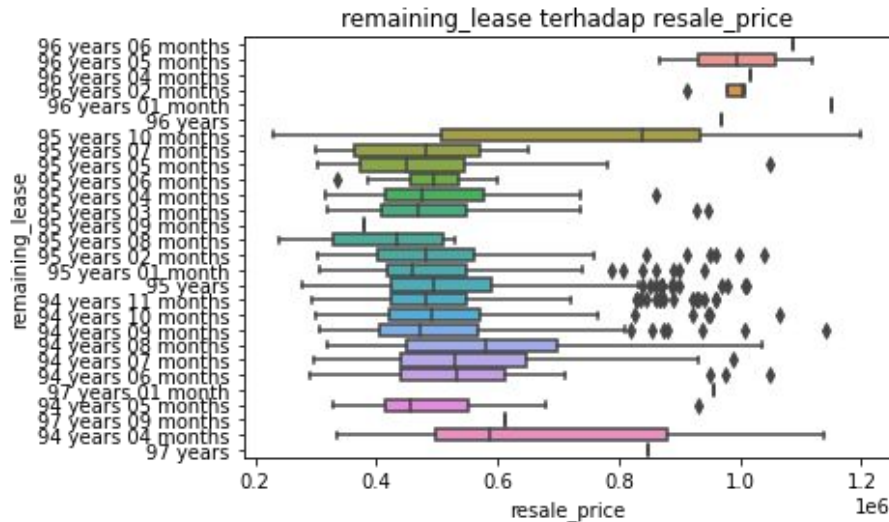
Berdasarkan Scatterplot disamping terlihat bahwa persebaran data `street_name` terhadap `resale_price` tidak beraturan dan tidak berpola

Exploratory Data Analysis



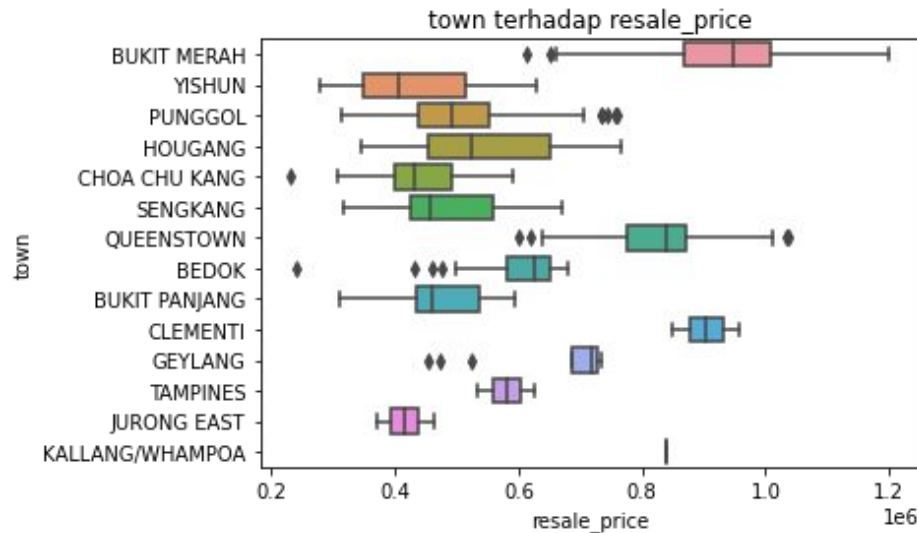
Berdasarkan Scatteplot disamping terlihat bahwa persebaran data `remaining_lease` terhadap `resale_price` tidak beraturan dan tidak berpola

Exploratory Data Analysis



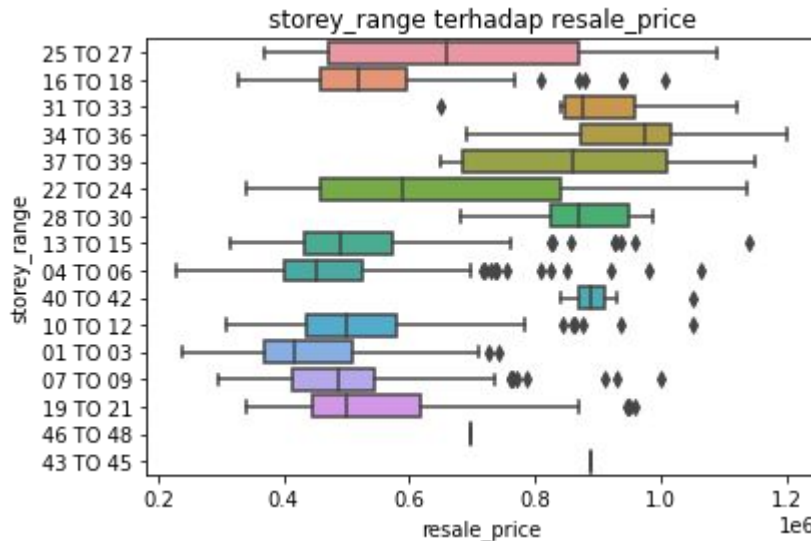
Pada boxplot disamping terlihat bahwa data `remaining_lease` terhadap `resale_price` cenderung saling tumpang tindih sehingga tidak menunjukkan persebaran yang signifikan. Oleh karena itu, tidak cocok untuk dijadikan sebagai variabel prediktor.

Exploratory Data Analysis



Pada boxplot diatas terlihat bahwa variabel `'town'` terhadap `'resale_price'` memiliki pola yang beberapa boxplotnya cukup berbeda dengan boxplot yang lain sehingga beberapa boxplotnya tidak saling tumpang tindih. Oleh karena itu, variabel `'town'` dapat dijadikan sebagai variabel prediktor.

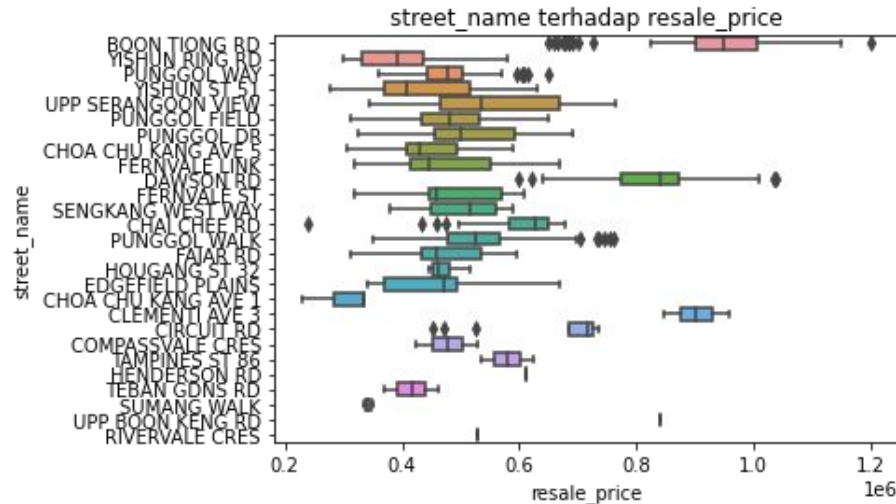
Exploratory Data Analysis



Pada boxplot diatas terlihat bahwa data `storey_range` terhadap `resale_price` memiliki pola boxplot berbeda.

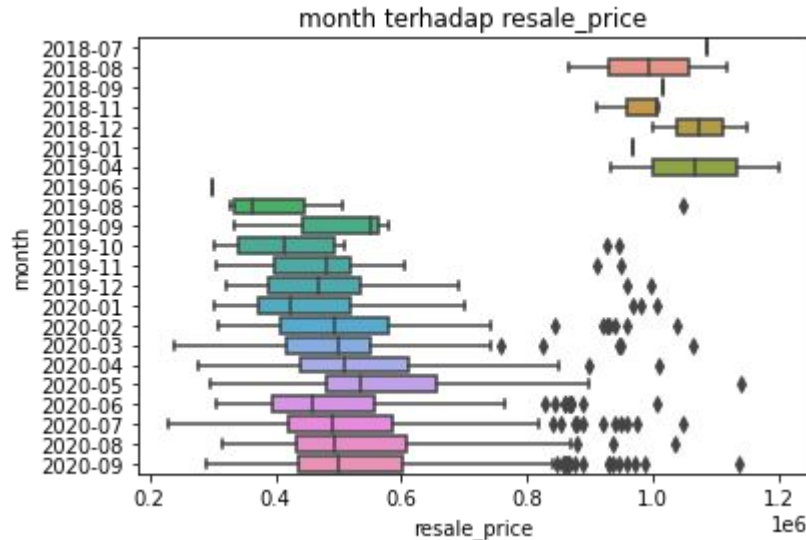
1. Boxplot tidak saling tumpang tindih seperti boxplot `storey_range` (37 to 39) dan (28 to 30).
2. Boxplot yang saling tumpang tindih tetapi data `storey_range` masih cukup untuk dijadikan sebagai variabel prediktor.

Exploratory Data Analysis



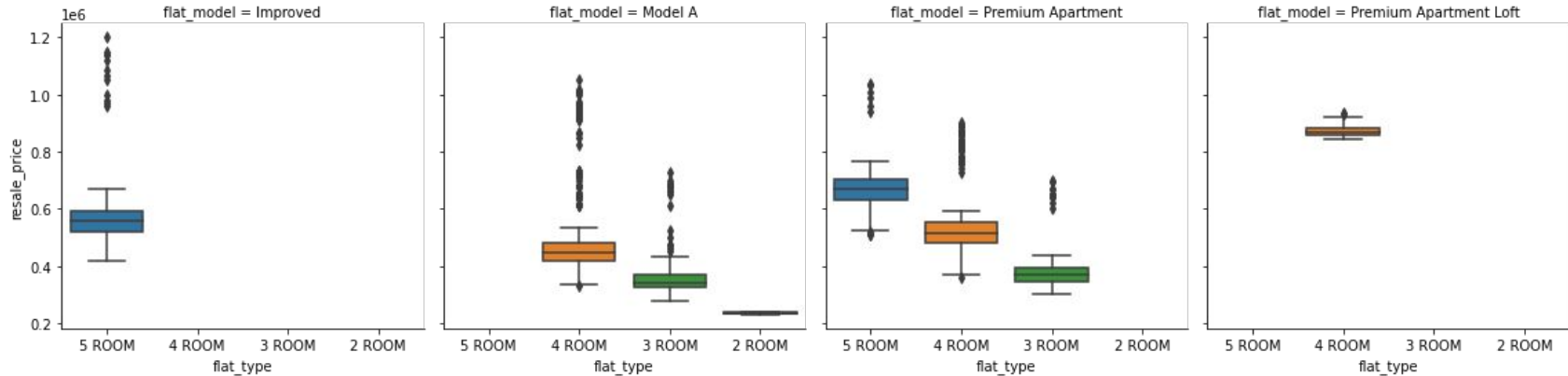
Pada boxplot diatas terlihat bahwa variabel `street_name` terhadap `resale_price` cenderung saling tumpang tindih dan tidak menunjukkan persebaran yang signifikan sehingga tidak cocok untuk dijadikan sebagai variabel prediktor

Exploratory Data Analysis



Pada boxplot disamping terlihat bahwa variabel `month` terhadap `resale_price` cenderung saling tumpang tindih dan tidak menunjukkan persebaran yang signifikan sehingga tidak cocok untuk dijadikan sebagai variabel prediktor.

Exploratory Data Analysis



Pada boxplot diatas terlihat bahwa variabel ``flat_type`` terhadap ``resale_price`` dan ``flat_model`` memiliki boxplot yang cukup berbeda dengan yang lain sehingga variabel ``flat_model`` dan ``flat_type`` cocok dijadikan sebagai variabel prediktor.

Modelling

Variabel Data 1

town, flat_type, storey_range, floor_area_sqm, flat_model dan resale_price.

Variabel Data 2

town, flat_type, storey_range, floor_area_sqm, flat_model, month_f, year_f, block_f dan resale_price.

Multivariate Linear Regression - Data 1

```
Coefficients: [ 2.92124725e+16  2.92124725e+16  2.92124725e+16  2.92124725e+16
 2.92124725e+16  2.92124725e+16  2.92124725e+16  2.92124725e+16
 2.92124725e+16  2.92124725e+16 -2.49452599e+05 -4.19981018e+04
 7.15884320e+04  2.19802462e+05 -6.89447820e+04 -5.14176492e+04
-3.17817372e+04 -2.34202561e+04 -1.87479595e+04 -1.43963667e+04
 5.37130452e+03 -7.29261961e+03  2.56679574e+02 -5.73621286e+03
 1.14653992e+04  4.97550821e+04  1.29107548e+04  6.63551305e+04
 7.59660628e+04  9.76562500e-04 -6.29461250e+04 -2.00340000e+04
 8.56675000e+03  7.65574844e+04  5.72000000e+02]
Intercept: -2.9212472464382304e+16
```

```
Index(['town_BEDOK', 'town_BUKIT MERAH', 'town_BUKIT PANJANG',
'town_CHOA CHU KANG', 'town_CLEMENTI', 'town_GEYLANG', 'town_HOUGANG',
'town_JURONG EAST', 'town_KALLANG/WHAMPOA', 'town_PUNGGOL',
'town_QUEENSTOWN', 'town_SENGKANG', 'town_TAMPINES', 'town_YISHUN',
'flat_type_2 ROOM', 'flat_type_3 ROOM', 'flat_type_4 ROOM',
'flat_type_5 ROOM', 'storey_range_01 TO 03', 'storey_range_04 TO 06',
'story_range_07 TO 09', 'storey_range_10 TO 12',
'story_range_13 TO 15', 'storey_range_16 TO 18',
'story_range_19 TO 21', 'storey_range_22 TO 24',
'story_range_25 TO 27', 'storey_range_28 TO 30',
'story_range_31 TO 33', 'storey_range_34 TO 36',
'story_range_37 TO 39', 'storey_range_40 TO 42',
'story_range_43 TO 45', 'storey_range_46 TO 48', 'flat_model_Improved',
'flat_model_Model A', 'flat_model_Premium Apartment',
'flat_model_Premium Apartment Loft', 'floor_area_sqm'],
dtype='object')
```

Dari koefisien diatas dapat dilihat bahwa data yang cukup mempengaruhi [resale_price](#) yaitu [floor_area_sqm](#), [flat_model](#), [storey_range](#), dan [flat_type](#) dengan melihat besar nilai koefisien dan tanda positif atau negatifnya. Misalnya [floor_area_sqm](#) yang berkofisien positif yang berarti semakin besar [floor_area_sqm](#) maka semakin besar [resale_price](#), begitu sebaliknya atau misalnya variabel [flat_model_Premium Apartment Loft](#) memiliki kofisien bernilai positif juga yang berarti bahwa [flat_model_Premium Apartment Loft](#) memiliki hubungan yang berbanding lurus terhadap [resale_price](#).

Multivariate Linear Regression - Data 1

```
#MAE
```

```
23189.426966292136
```

```
# RMSE
```

```
32563.752714888444
```

```
# R2 score
```

```
0.9634351182562396
```

Multivariate Linear Regression - Data 2

```

Coefficients: [ 2.53136498e+16 2.53136498e+16 2.53136498e+16 2.53136498e+16
2.53136498e+16 2.53136498e+16 2.53136498e+16 2.53136498e+16
2.53136498e+16 2.53136498e+16 -2.51187532e+05 -4.11004693e+04
7.25131755e+04 2.19593241e+05 -6.88501821e+04 -5.09802433e+04
-3.16081413e+04 -2.33196690e+04 -1.88432261e+04 -1.36757786e+04
4.91359955e+03 -7.47659252e+03 -1.78819271e+02 -4.97775356e+03
1.08996911e+04 5.03222277e+04 1.29291980e+04 6.59566644e+04
7.59650605e+04 2.50000000e-01 -6.22806250e+04 -1.99580000e+04
8.25225000e+03 7.55757734e+04 -8.02625000e+02 5.37893750e+03
-4.04750000e+02 -1.56687500e+03 -3.06528125e+03 6.08000000e+02
-7.02812500e+02 0.00000000e+00]
Intercept: -2.5313649818086164e+16

```

```

Index(['town_BEDOK', 'town_BUKIT MERAH', 'town_BUKIT PANJANG',
'town_CHOACHU KANG', 'town_CLEMENTI', 'town_GEYLANG', 'town_HOUGANG',
'town_JURONG EAST', 'town_KALLANG/WHAMPOA', 'town_PUNGOL',
'town_QUEENSTOWN', 'town_SENGKANG', 'town_TAMPINES', 'town_YISHUN',
'flat_type_2 ROOM', 'flat_type_3 ROOM', 'flat_type_4 ROOM',
'flat_type_5 ROOM', 'storey_range_01 TO 03', 'storey_range_04 TO 06',
'story_range_07 TO 09', 'storey_range_10 TO 12',
'story_range_13 TO 15', 'storey_range_16 TO 18',
'story_range_19 TO 21', 'storey_range_22 TO 24',
'story_range_25 TO 27', 'storey_range_28 TO 30',
'story_range_31 TO 33', 'storey_range_34 TO 36',
'story_range_37 TO 39', 'storey_range_40 TO 42',
'story_range_43 TO 45', 'storey_range_46 TO 48', 'flat_model_Improved',
'flat_model_Model A', 'flat_model_Premium Apartment',
'flat_model_Premium Apartment Loft', 'block_f_A', 'block_f_B',
'block_f_C', 'block_f_D', 'block_f_E', 'floor_area_sqm', 'month_f',
'year_f'],
dtype='object')

```

Dari koefisien diatas dapat dilihat bawah variabel yang cukup mempengaruhi **resale_price** yaitu **floor_area_sqm**, **flat_model**, **storey_range**, **flat_type**, **month_f**, dan **block_F** dengan melihat besar nilai koefisien dan tanda positif atau negatifnya. Pada kolom **blok_f_A** memiliki koefisien yang bernilai negatif yang menandakan bahwa hubungan kolom tersebut terhadap **resale_price** berbanding terbalik.

Linear Regression - Data 2

```
#MAE
```

```
23315.47191011236
```

```
# RMSE
```

```
32605.714069431575
```

```
# R2 score
```

```
0.9633408231956968
```

Ridge Regression - Data 1

```
Coefficients: [ 43721.74986947 316564.73795313 -125019.75965512 -156590.50379691
100103.92790488 102248.43862235 -82765.50756028 -82333.07653679
98360.29789189 -93423.61362374 203559.6464263 -133685.26772138
-30364.92864139 -160376.14113253 -72092.98345754 -19087.04084577
8604.88595632 82575.13834691 -70493.32858872 -52138.94799288
-32105.94613583 -24201.1692583 -18638.55537462 -13724.66289527
6858.05451056 1372.42759101 2843.45001075 5873.28630831
38607.10437525 43728.05802969 19564.5148202 53982.84670223
38472.86789759 0. -51079.74152315 -11448.19243874
16422.19951565 46105.73444628 4051.54485584]
Intercept: 231983.12984874687
```

Ridge Regression - Data 1

```
#MAE
```

```
23546.388023022766
```

```
# RMSE
```

```
32369.798245759324
```

```
# R2 score
```

```
0.9638693926488003
```

Ridge Regression - Data 2

```
Coefficients: [ 45179.55681423  317365.76131015 -123499.88322039 -154800.41002034
 97709.68569773  104473.12351018  -82801.13834109  -86417.70216265
 95589.58550533  -91520.55200393  198546.73473008 -132897.88247449
-28189.79964058 -158737.07970434  -72469.13520143  -18210.25707657
 8971.80995607   81707.58232185  -70665.5127694   -51714.93344058
-31810.89673279  -23936.65596787  -18567.854877    -13479.07918759
 6721.27860308   1429.74633494    2564.91767537    6490.73386574
 37325.45306897  44069.96205278    19484.38833124   53661.34639296
 38427.10665015      0.        -50324.93205254  -10994.18443988
 16242.71884334  45076.39764912   -2450.53792527   3214.45432957
 -2780.01586037  -4603.67898626    6619.77844232    4099.0038356
 -600.63483571    0.        ]
Intercept: 231557.124552452
```

Ridge Regression - Data 2

```
#MAE
```

```
23413.9164729176
```

```
# RMSE
```

```
32299.736938182272
```

```
# R2 score
```

```
0.96402562582243
```

Model Final - Multivariate Regression - Data 1

Resale_price = -2.921 + 2.921town_BEDOK + 2.921town_BUKIT MERAH + 2.921town_BUKIT PANJANG + 2.921town_CHOJA CHU KANG + 2.921town_CLEMENTI + 2.921town_GEYLANG + 2.921town_HOUGANG + 2.921town_JURONG EAST + 2.921town_KALLANG/WHAMPOA + 2.921town_PUNGGOL + 2.921town_QUEENSTOWN + 2.921town_SENGKANG + 2.921town_TAMPINES + 2.921town_YISHUN - 2.495flat_type_2 ROOM - 4.200flat_type_3 ROOM + 7.159flat_type_4 ROOM + 2.198flat_type_5 ROOM - 6.894storey_range_01 TO 03 - 5.142storey_range_04 TO 06 - 3.178storey_range_07 TO 09 - 2.342storey_range_10 TO 12 - 1.875storey_range_13 TO 15 - 1.440storey_range_16 TO 18 + 5.371storey_range_19 TO 21 - 7.293storey_range_22 TO 24 + 2.567storey_range_25 TO 27 - 5.736storey_range_28 TO 30 + 1.146storey_range_31 TO 33 + 4.975storey_range_34 TO 36 + 1.291storey_range_37 TO 39 + 6.635storey_range_40 TO 42 + 7.596storey_range_43 TO 45 + 9.765storey_range_46 TO 48 - 6.295flat_model_Improved - 2.003flat_model_Model A + 8.567flat_model_Premium Apartment + 7.656flat_model_Premium Apartment Loft + 5.720floor_area_sqm

Conclusion

Kesimpulan

Harga jual properti yang tersedia pada HDB (Lembaga Perumahan dan Pembangunan) bergantung pada faktor luas ruangan, tingkat lantai, jumlah kamar, dan model dari ruangan/kamar. Sehingga jika ingin membeli properti di HBD disarankan untuk melihat faktor-faktor tersebut, yaitu mempertimbangkan luas ruangan yang diinginkan, model dari ruangan dan jumlah kamar yang dibutuhkan.

Terima kasih!

Ada pertanyaan?

zenius



Kampus
Merdeka
INDONESIA JAYA