

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

# Data Cleansing

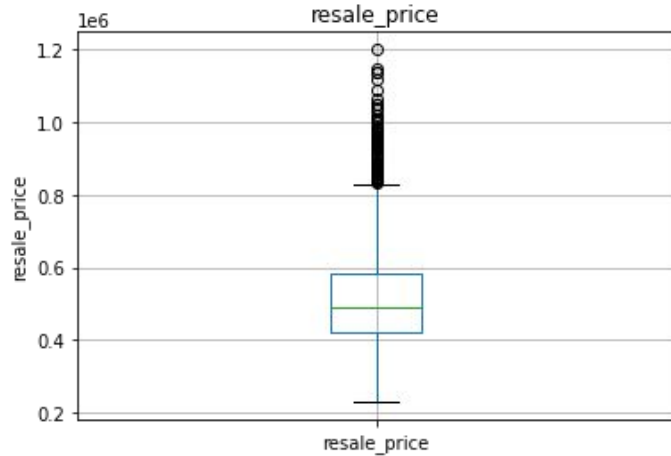
```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1187 entries, 30680 to 80366
Data columns (total 11 columns):
#   Column                Non-Null Count  Dtype
---  -
0   month                 1187 non-null   object
1   town                  1187 non-null   object
2   flat_type             1187 non-null   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

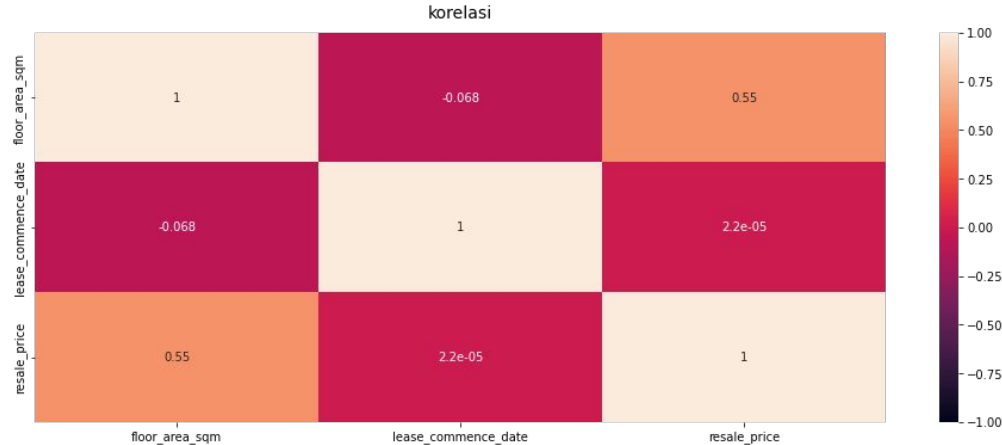


Berdasarkan boxplot di samping, dapat dikatakan bahwa:

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

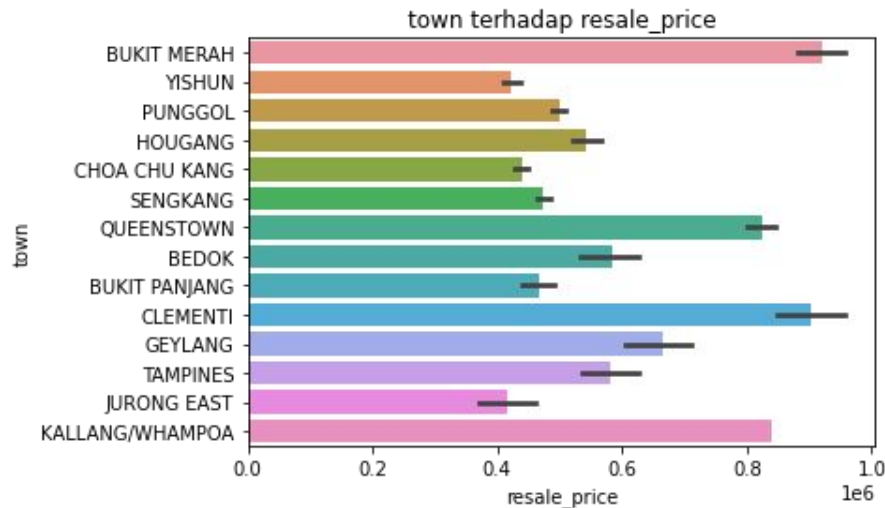
Outlier pada data 'df' ini tidak dilakukan penanganan, Karena data outlier yang berada pada variabel `'resale_price'` merupakan hal yang wajar untuk nominal resale price of flat yang tinggi.

# Exploratory Data Analysis



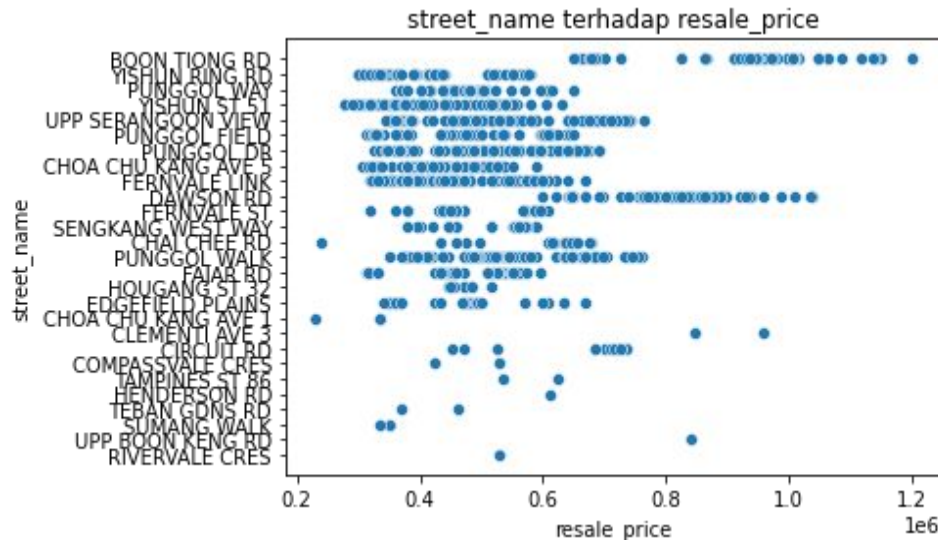
Berdasarkan nilai korelasi untuk variabel numerik, terlihat bahwa ``resale_price`` dengan ``lease_commence_date`` memiliki nilai korelasi yang searah atau positif dengan nilai korelasi sebesar 0.55. Artinya ``resale_price`` dengan ``lease_commence_date`` memiliki hubungan yang kuat.

# Exploratory Data Analysis



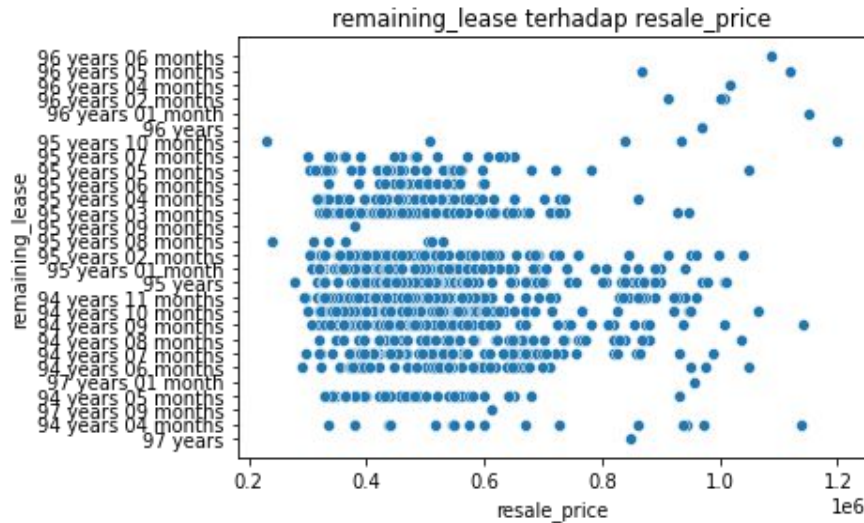
Berdasarkan Barplot ``resale_price`` dengan ``town``, ``resale_price`` tertinggi berada di kota BUKIT MERAH dan ``resale_price`` terendah berada di kota JURONG EAST.

# Exploratory Data Analysis



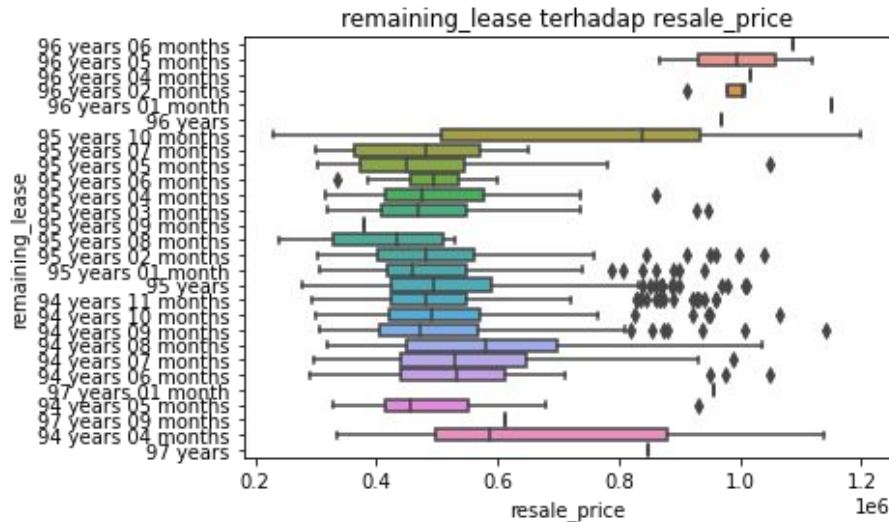
Berdasarkan Scatterplot disamping terlihat bahwa persebaran data `street\_name` terhadap `resale\_price` tidak beraturan dan tidak berpola

# Exploratory Data Analysis



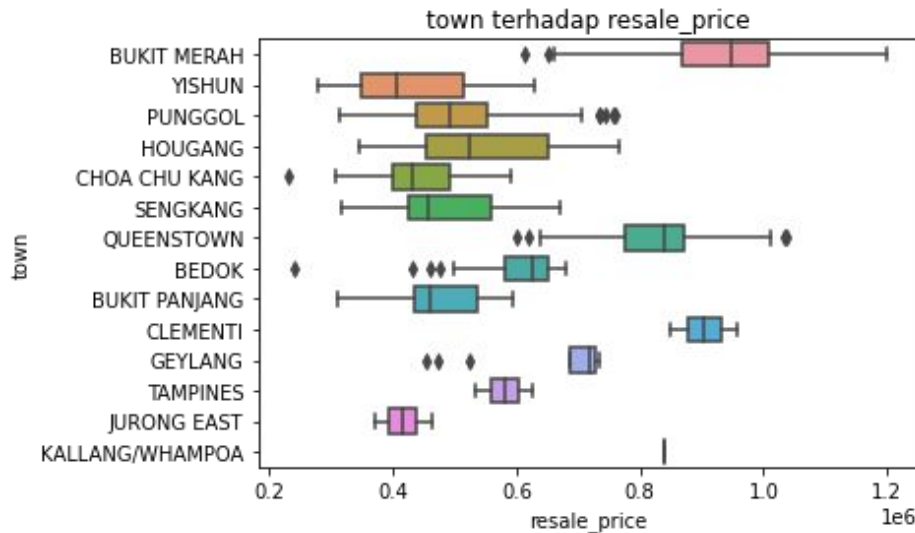
Berdasarkan Scatterplot disamping terlihat bahwa persebaran data `remaining\_lease` terhadap `resale\_price` tidak beraturan dan tidak berpola

# Exploratory Data Analysis



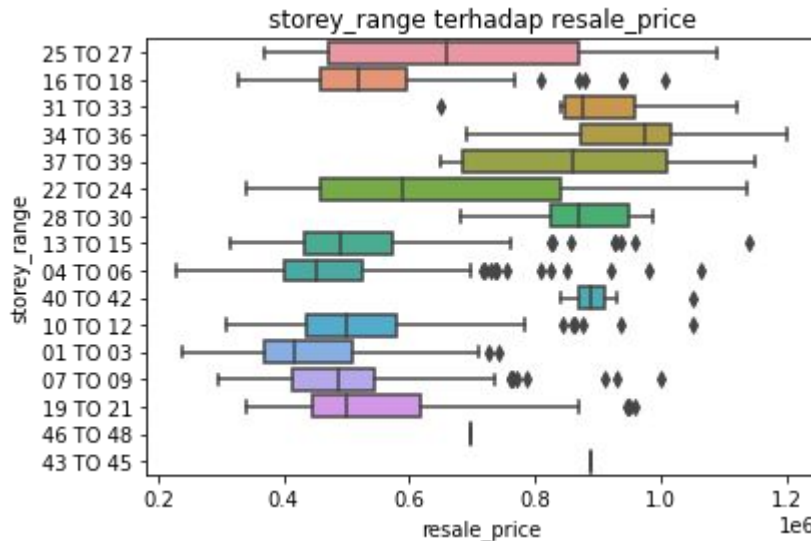
Pada boxplot disamping terlihat bahwa data `remaining_lease` terhadap `resale_price` cenderung saling tumpang tindih sehingga tidak menunjukkan persebaran yang signifikan. Oleh karena itu, tidak cocok untuk dijadikan sebagai variabel prediktor.

# 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

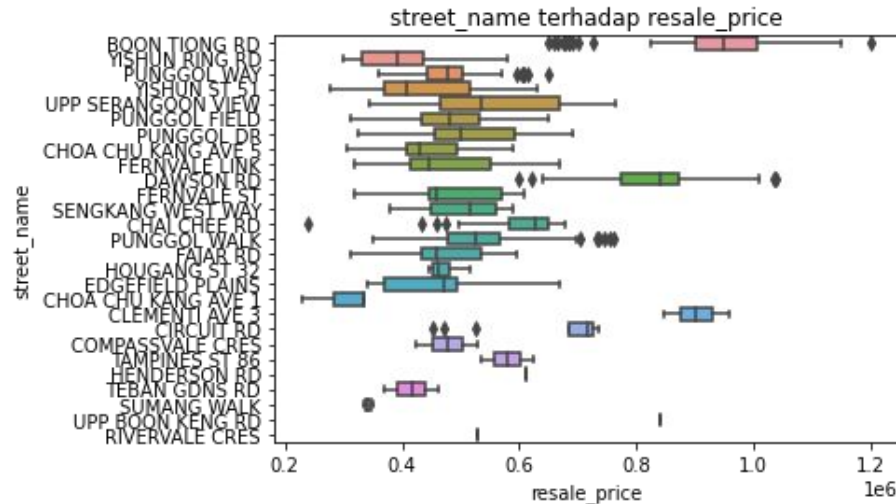


Pada boxplot diatas terlihat bahwa data `storey\_range` terhadap `resale\_price` memiliki pola boxplot berbeda.

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.

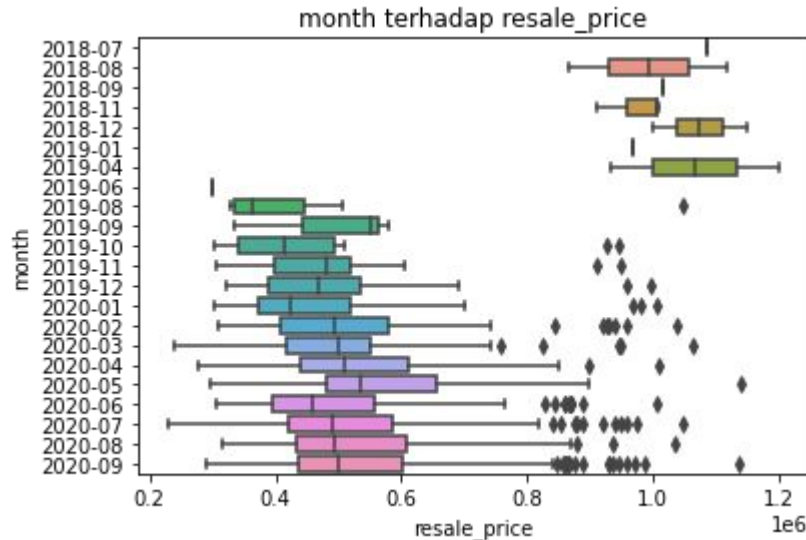


# Exploratory Data Analysis



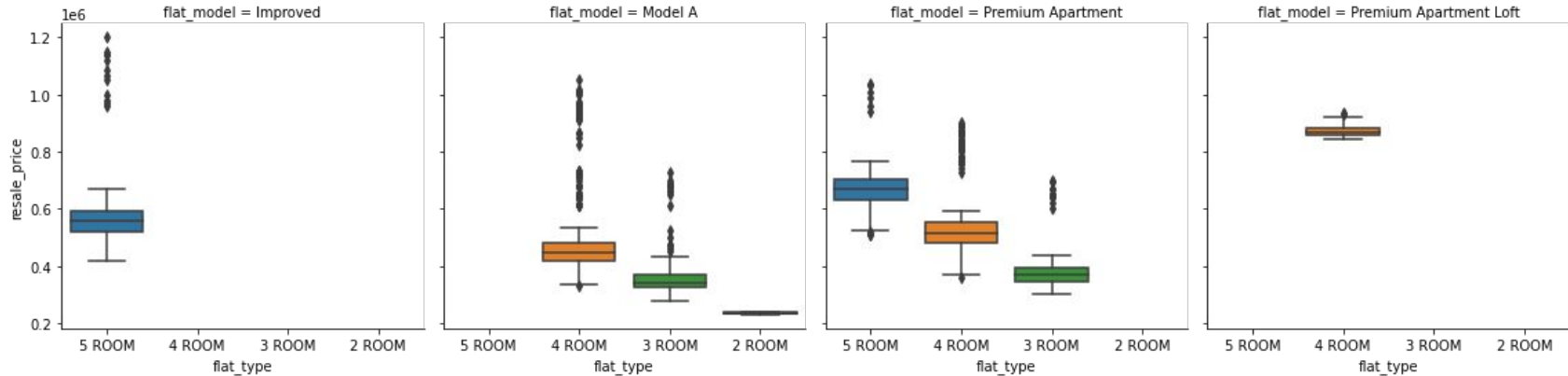
Pada boxplot diatas terlihat bahwa variabel `'street_name'` terhadap `'resale_price'` cenderung saling tumpang tindih dan tidak menunjukkan persebaran yang signifikan sehingga tidak cocok untuk dijadikan sebagai variabel prediktor

# 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 Regression - Data 1

Model cukup baik karena nilai rata-rata kuadrat dari jumlah kesalahan pada model prediksi lebih kecil dari nilai rata-ratanya.



	floor_area_sqm	lease_commence_date	resale_price
count	1185.000000	1185.000000	1.185000e+03
mean	93.607595	2016.011814	5.237517e+05

```
#MAE
```

```
23189.426966292136
```

```
# RMSE
```

```
32563.752714888444
```

```
# R2 score
```

```
0.9634351182562396
```



Diperoleh nilai R-Square sebesar 0.96.3 artinya sebesar 96% variabel independen mempengaruhi variabel dependen.

# Multivariate Regression - Data 2

Model cukup baik karena nilai rata-rata kuadrat dari jumlah kesalahan pada model prediksi lebih kecil dari nilai rata-ratanya.

	floor_area_sqm	lease_commence_date	resale_price
count	1185.000000	1185.000000	1.185000e+03
mean	93.607595	2016.011814	5.237517e+05

#MAE

23315.47191011236

# RMSE

32605.714069431575

# R2 score

0.9633408231956968

Diperoleh nilai R-Square sebesar 0.963 artinya sebesar 96.3% variabel independen mempengaruhi variabel dependen.

# Ridge Regression - Data 1

```
#MAE
```

```
23546.388023022766
```

```
# RMSE
```

```
32369.798245759324
```

```
# R2 score
```

```
0.9638693926488003
```

Diperoleh nilai R-Square sebesar 0.964 artinya sebesar 96.4% variabel independen mempengaruhi variabel dependen.

Model cukup baik karena nilai rata-rata kuadrat dari jumlah kesalahan pada model prediksi lebih kecil dari nilai rata-ratanya.



	floor_area_sqm	lease_commence_date	resale_price
count	1185.000000	1185.000000	1.185000e+03
mean	93.607595	2016.011814	5.237517e+05



# Ridge Regression - Data 2

```
#MAE
```

```
23413.9164729176
```

```
# RMSE
```

```
32299.736938182272
```

```
# R2 score
```

```
0.96402562582243
```

Diperoleh nilai R-Square sebesar 0.964 artinya sebesar 96.4% variabel independen mempengaruhi variabel dependen.

Model cukup baik karena nilai rata-rata kuadrat dari jumlah kesalahan pada model prediksi lebih kecil dari nilai rata-ratanya.



	floor_area_sqm	lease_commence_date	resale_price
count	1185.000000	1185.000000	1.185000e+03
mean	93.607595	2016.011814	5.237517e+05

# Model Final - Multivariate Regression - Data 1

**Resale\_price** = -2.921 + 2.921town\_BEDOK + 2.921town\_BUKIT MERAH + 2.921town\_BUKIT PANJANG + 2.921town\_CHOACHU 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

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 berkoefisien positif yang berarti semakin besar **floor\_area\_sqm** maka semakin besar **resale\_price**.

# 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