

# Jabodetabek Housing Price Analysis

Bootcamp Data Analyst with SQL &  
Python using Google Platform – Batch 19

Muhammad Ilham



# Latar Belakang

- Di kehidupan sehari-hari saya, seringkali kepemilikan rumah pribadi menjadi bahan obrolan
- Kenaikan harga tanah dan properti di kota-kota besar seperti di Jakarta, seringkali tidak sebanding dengan pendapatan Masyarakat sehingga harga hunian menjadi tidak terjangkau.

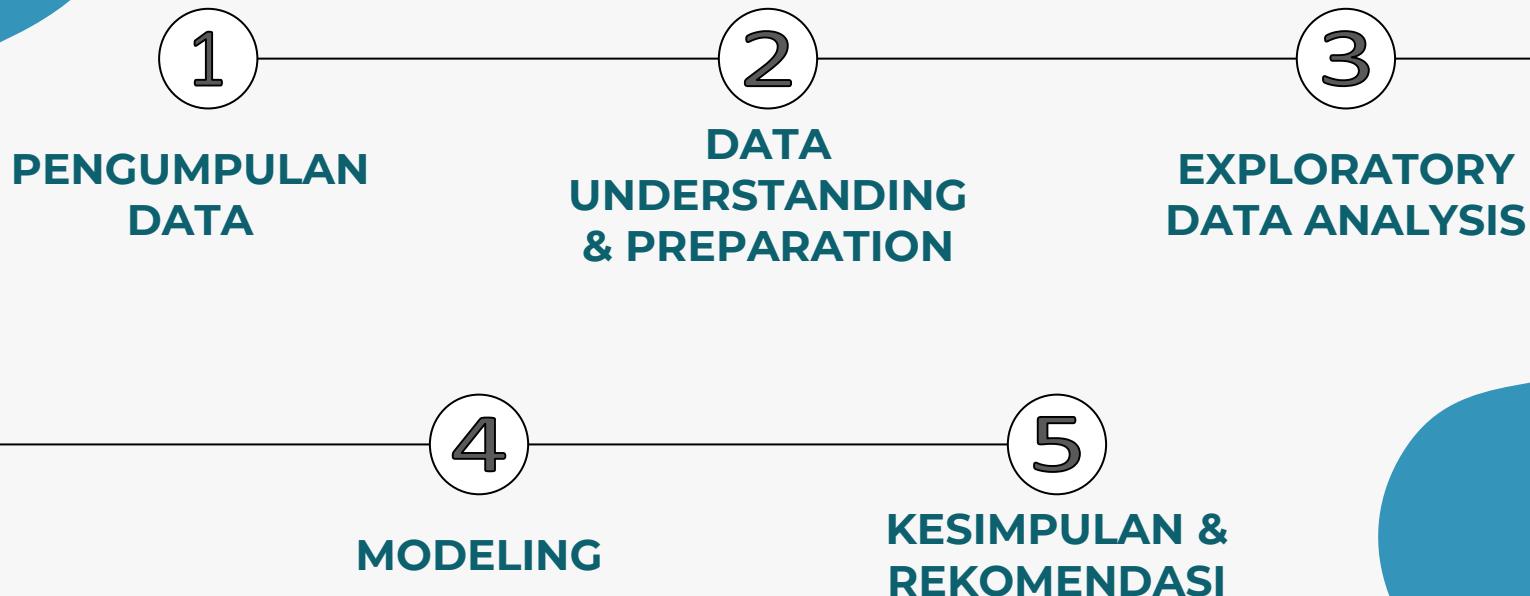


**Bagaimanakah kondisi pasar property di Indonesia ?**

Hal ini akan dianalisis lebih lanjut, khususnya **untuk property rumah di sekitaran Jabodetabek**



# Alur Analisis



# Pengumpulan Data

The screenshot shows the rumah123 website interface. At the top, there's a navigation bar with links for Dijual, Disewa, Properti Baru, Aset Bank, KPR, and Lainnya. Below the navigation is a search bar with the placeholder 'Jabodetabek' and a 'Cari' button. Underneath the search bar are filter options: Properti Baru, Harga, Luas Tanah, Luas Bangunan, and Lokasi. The main content area displays a grid of property listings. One listing is highlighted with a yellow 'Official Developer' badge. Two specific properties are shown in larger detail: one for Rp 2,9 Miliar (Tipe T5, LT: 97 m², LB: 70 m²) and another for Rp 4,03 Miliar (Tipe T6, LT: 142 m², LB: 86 m²). Each listing includes a small thumbnail image, price, type, and dimensions.



The screenshot shows the Instant Data Scraper tool window. At the top, there are buttons for 'Locate "Next" button', 'Start crawling', 'CSV', 'XLSX', and 'COPY ALL'. To the right, it displays statistics: Pages scraped: 530, Rows collected: 10616, Rows from last page: 20, and Working time: 8505s. Below these are sections for 'Infinite scroll' settings (Min delay: 15 sec, Max delay: 60 sec), a message 'Table not found', a note about the crawl, a preview message, and a log of URLs being crawled. A green arrow points from the Instant Data Scraper window to the Microsoft Excel icon.



- Data rumah dijual diambil dari website <https://www.rumah123.com/>
- Lokasi dibatasi : **Jabodetabek**
- Pengambilan data menggunakan metode scraping dengan tools : **Instant Data Scraper**
- Data disimpan dalam file .xlsx

# Data Understanding

live  
Class



| attribute-text | attribute-text 2 | attribute-text 3 | attribute-info 2 | attribute-info 4   | name               | card-featured_middle-section_price | card-featured_middle-section_href | card-featured_middle-section_2  |
|----------------|------------------|------------------|------------------|--------------------|--------------------|------------------------------------|-----------------------------------|---|
| 3247           | 2                | 1                | 2                | 60 m <sup>2</sup>  | 40 m <sup>2</sup>  | Angel                              | Rp 330 Juta                       | <a href="https://www.rumah123.com/properti/bogor-parungpanjang/rumah-murah-full-renovasi-cluster-the-river-siap-huni-dekat-stasiun-parung-panjang-hos39848306/">https://www.rumah123.com/properti/bogor-parungpanjang/rumah-murah-full-renovasi-cluster-the-river-siap-huni-dekat-stasiun-parung-panjang-hos39848306/</a> |
| 7408           | 2                | 1                | 1                | 77 m <sup>2</sup>  | 53 m <sup>2</sup>  | Dimas                              | Rp 600 Juta                       | <a href="https://www.rumah123.com/properti/bekasi-harapan-baru/rumah-murah-dalam-komplek-harapan-baru-kota-bekasi-hos16671673/">https://www.rumah123.com/properti/bekasi-harapan-baru/rumah-murah-dalam-komplek-harapan-baru-kota-bekasi-hos16671673/</a>   |
| 8552           | 3                | 3                | NaN              | 96 m <sup>2</sup>  | 112 m <sup>2</sup> | Tinus Infinite                     | Rp 1,85 Miliar                    | <a href="https://www.rumah123.com/properti/tangerang-suvarna-sutera/dijual-rumah-cluster-grandura-lavon-swancity-hos39769908/">https://www.rumah123.com/properti/tangerang-suvarna-sutera/dijual-rumah-cluster-grandura-lavon-swancity-hos39769908/</a>   |
| 140            | 3                | 3                | 1                | 66 m <sup>2</sup>  | 95 m <sup>2</sup>  | Sinarmas Land Project              | Rp 2,2 Miliar                     | <a href="https://www.rumah123.com/properti/tangerang-bsd-puspita-loka/jual-rumah-strategis-2-lantai-di-bsd-nyaman-punya-3-kamar-hos40220156/">https://www.rumah123.com/properti/tangerang-bsd-puspita-loka/jual-rumah-strategis-2-lantai-di-bsd-nyaman-punya-3-kamar-hos40220156/</a>                                     |
| 3252           | 5                | 8                | 4                | 644 m <sup>2</sup> | 850 m <sup>2</sup> | Jenny Kam                          | Rp 18 Miliar                      | <a href="https://www.rumah123.com/properti/tangerang-bsd/rumah-super-luas-dan-bagus-di-telaga-golf-bsd-tangerang-harga-bu-hos18707936/">https://www.rumah123.com/properti/tangerang-bsd/rumah-super-luas-dan-bagus-di-telaga-golf-bsd-tangerang-harga-bu-hos18707936/</a>   |

Dataset merupakan **hasil web scraping**, dari hasil preview & telaah hanya beberapa variable yang bermanfaat untuk diolah lebih lanjut.

Variable tersebut antara lain :

- attribute-text : **Jumlah kamar tidur**
- attribute-text 2 : **Jumlah kamar mandi**
- attribute-text 3 : **Kapasitas garasi (mobil)**
- attribute-info 2 : **Luas Tanah**
- attribute-info 4 : **Luas Bangunan**
- name : **Nama Penjual**
- card-featured\_middle-section\_price : **Harga Rumah**
- card-featured\_middle-section\_href : **Detail URL**
- card-featured\_middle-section\_2 : **Lokasi**

# Data Preparation & Cleaning

## Menanangani missing values

| # | Column        | Non-Null Count | Dtype           |
|---|---------------|----------------|-----------------|
| 0 | kamar_tidur   | 10524          | non-null object |
| 1 | kamar_mandi   | 10485          | non-null object |
| 2 | garasi        | 7679           | non-null object |
| 3 | luas_tanah    | 10616          | non-null object |
| 4 | luas_bangunan | 10602          | non-null object |
| 5 | penjual       | 10616          | non-null object |
| 6 | harga         | 10540          | non-null object |
| 7 | url           | 10540          | non-null object |
| 8 | lokasi        | 10540          | non-null object |

## Mengubah type/format data

| #  | Column         | Non-Null Count | Dtype            |
|----|----------------|----------------|------------------|
| 0  | kamar_tidur    | 10409          | non-null int64   |
| 1  | kamar_mandi    | 10409          | non-null int64   |
| 2  | garasi         | 10409          | non-null int64   |
| 3  | luas_tanah     | 10409          | non-null float64 |
| 4  | luas_bangunan  | 10409          | non-null float64 |
| 5  | penjual        | 10409          | non-null object  |
| 6  | harga          | 10409          | non-null float64 |
| 7  | url            | 10409          | non-null object  |
| 8  | lokasi         | 10409          | non-null object  |
| 9  | harga_nilai    | 10409          | non-null float64 |
| 10 | satuan         | 10409          | non-null object  |
| 11 | harga_satuan   | 10409          | non-null int64   |
| 12 | lokasi_kawasan | 10409          | non-null object  |
| 13 | lokasi_kota    | 10409          | non-null object  |



- Data yang tidak mencantumkan **harga** tidak dapat digunakan sehingga perlu dihapus
- Data yang tidak mencantumkan jumlah **kamar\_tidur** dan **kamar\_mandi** juga tidak dapat digunakan
- **garasi** dengan nilai NaN diasumsikan memiliki kapasitas 1 kendaraan

- Ubah Dtype kolom **kamar\_tidur**, **kamar\_mandi** dan **garasi** ke **integer**
- Split dan Ubah Dtype kolom **harga**, **luas\_tanah** dan **luas\_bangunan** ke **float**
- Split kolom **lokasi** ke kolom baru **lokasi\_kawasan** dan **lokasi\_kota**

# Data Preparation & Cleaning

Before

| kamar_tidur | kamar_mandi | garasi | luas_tanah | luas_bangunan      | penjual            | harga                        | url           | lokasi  |
|-------------|-------------|--------|------------|--------------------|--------------------|------------------------------|---------------|---|
| 10          | 3           | 3      | 1          | 120 m <sup>2</sup> | 140 m <sup>2</sup> | Primary Projects by Rumah123 | Rp 2,8 Miliar | <a href="https://www.rumah123.com/properti/bogor-cibubur/rumah-baru-mewah-luas-dan-nyaman-di-cibubur-dengan-3-kamar-tidur-harga-2800000000">https://www.rumah123.com/properti/bogor-cibubur/rumah-baru-mewah-luas-dan-nyaman-di-cibubur-dengan-3-kamar-tidur-harga-2800000000</a> |

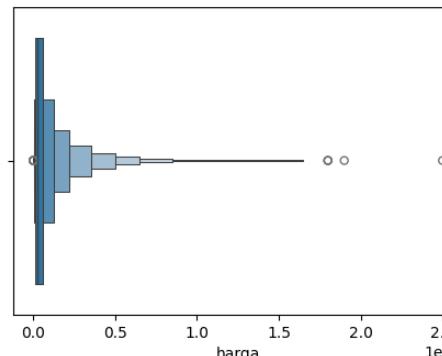
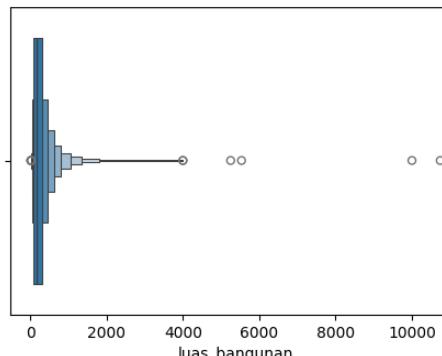
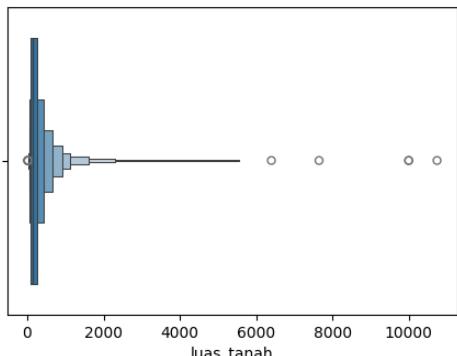
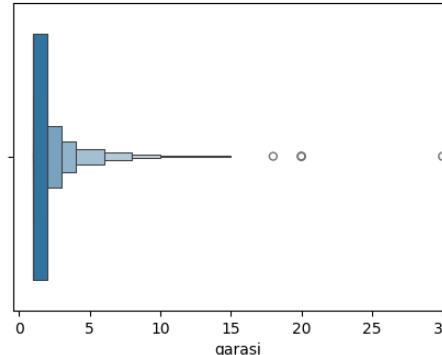
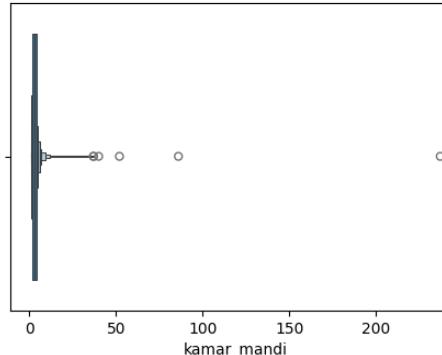
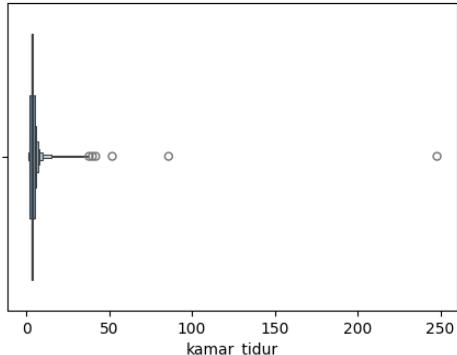


After

| kamar_tidur | kamar_mandi | garasi | luas_tanah | luas_bangunan | penjual | harga                        | url          | lokasi  | harga_nilai    | satuan | harga_satuan | lokasi_kawasan | lokasi_kota |       |
|-------------|-------------|--------|------------|---------------|---------|------------------------------|--------------|---|----------------|--------|--------------|----------------|-------------|-------|
| 10          | 3           | 3      | 1          | 120.0         | 140.0   | Primary Projects by Rumah123 | 2.800000e+09 | <a href="https://www.rumah123.com/properti/bogor-cibubur/rumah-baru-mewah-luas-dan-nyaman-di-cibubur-dengan-3-kamar-tidur-harga-2800000000">https://www.rumah123.com/properti/bogor-cibubur/rumah-baru-mewah-luas-dan-nyaman-di-cibubur-dengan-3-kamar-tidur-harga-2800000000</a> | Cibubur, Bogor | 2.8    | Miliar       | 1000000000     | Cibubur     | Bogor |

# Exploratory Data Analysis

## Sebaran nilai per variable - Outlier

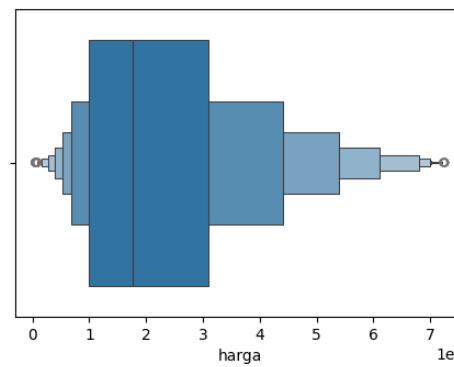
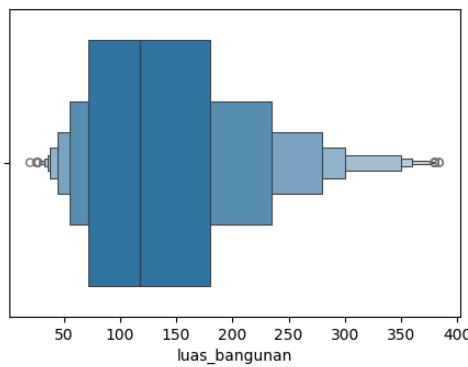
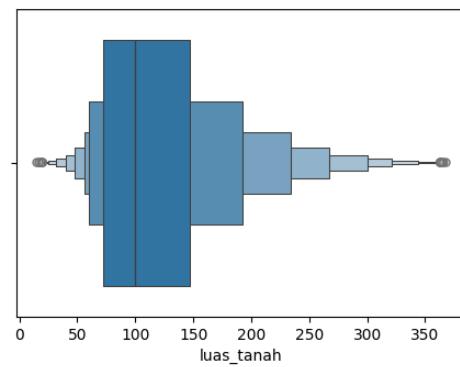
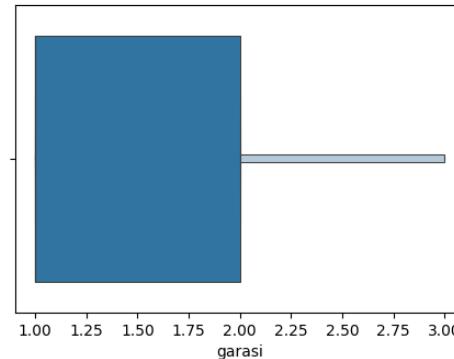
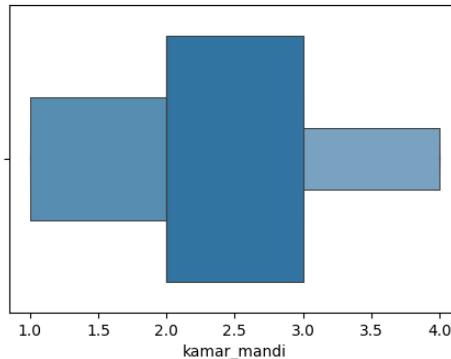
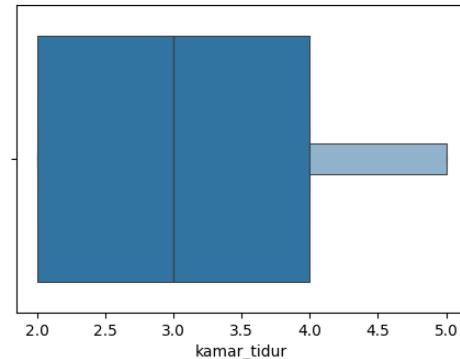


### Catatan

- Masih banyak terdapat outlier yang akan menimbulkan **misrepresentasi data**

# Exploratory Data Analysis

## Sebaran nilai per variable - Drop Outlier

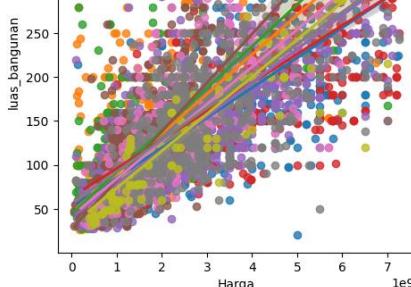
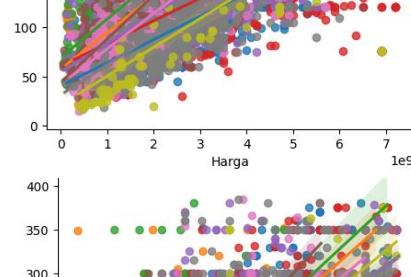
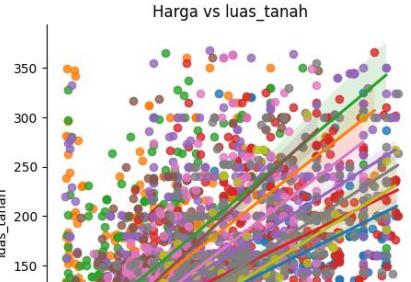
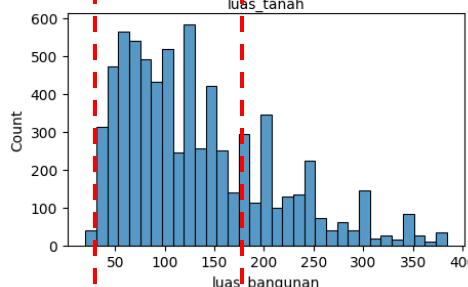
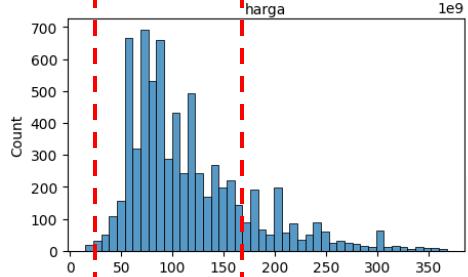
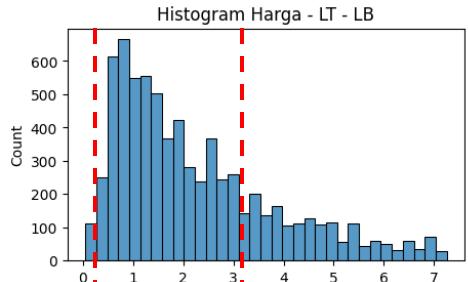


### Catatan

- Outlier sudah dihilangkan, sebaran data cenderung ke kiri / *left skewed*

# Exploratory Data Analysis

## Histogram per variable

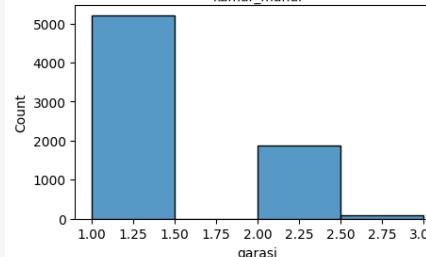
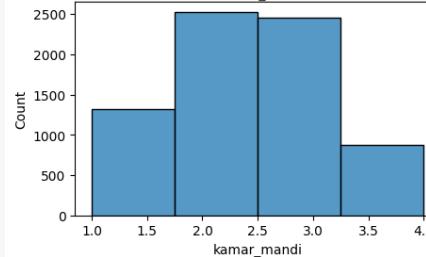
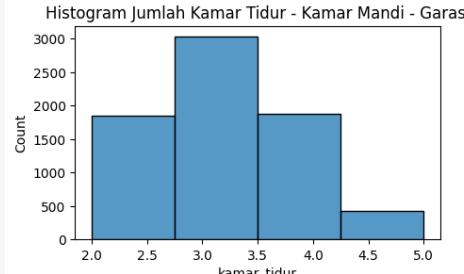


lokasi\_kota

- Jakarta Utara
- Bekasi
- Bogor
- Jakarta Selatan
- Tangerang
- Depok
- Jakarta Timur
- Jakarta Barat
- Jakarta Pusat

lokasi\_kota

- Jakarta Utara
- Bekasi
- Bogor
- Jakarta Selatan
- Tangerang
- Depok
- Jakarta Timur
- Jakarta Barat
- Jakarta Pusat



## Catatan

- Terlihat bahwa **harga** sekilas memiliki kemiripan sebaran dengan **luas tanah** dan **luas bangunan**
- Sehingga, memungkinkan memiliki korelasi yang kuat
- Kamar **mandi**, **kamar tidur**, dan **garasi** memiliki sebaran yang beragam / belum konklusif

# Exploratory Data Analysis

live  
Class

## Sebaran Lokasi

| Column         | Unique Count |
|----------------|--------------|
| lokasi         | 559          |
| lokasi_kawasan | 536          |
| lokasi_kota    | 9            |

| Kota            |
|-----------------|
| lokasi_kota     |
| Tangerang       |
| Bekasi          |
| Jakarta Barat   |
| Depok           |
| Jakarta Utara   |
| Jakarta Selatan |
| Bogor           |
| Jakarta Timur   |
| Jakarta Pusat   |

| lokasi_kota     | kamar_tidur | kamar_mandi | garasi   | luas_tanah | luas_bangunan | harga        |
|-----------------|-------------|-------------|----------|------------|---------------|--------------|
| Jakarta Selatan | 3.536982    | 2.875740    | 1.505917 | 144.997041 | 183.659763    | 3.690203e+09 |
| Jakarta Utara   | 3.371901    | 2.808540    | 1.345730 | 123.061983 | 171.619835    | 3.522729e+09 |
| Jakarta Barat   | 3.328194    | 2.746696    | 1.332599 | 121.618943 | 163.854626    | 2.969405e+09 |
| Jakarta Pusat   | 3.179688    | 2.484375    | 1.218750 | 99.890625  | 145.429688    | 2.697289e+09 |
| Tangerang       | 3.133290    | 2.444083    | 1.289337 | 121.916775 | 128.364759    | 2.291721e+09 |
| Jakarta Timur   | 3.224907    | 2.511152    | 1.384758 | 117.934944 | 131.715613    | 2.115095e+09 |
| Depok           | 2.843750    | 2.059152    | 1.148438 | 101.948661 | 100.046875    | 1.314474e+09 |
| Bekasi          | 2.896898    | 2.031022    | 1.238139 | 103.601277 | 101.862226    | 1.244995e+09 |
| Bogor           | 2.755952    | 1.889881    | 1.154762 | 117.550595 | 103.261905    | 1.237264e+09 |

### Catatan

- Berdasarkan Lokasi, hanya **Lokasi berdasarkan Kota** yang layak dibreakdown lebih lanjut, selebihnya terlalu banyak
- Secara rata-rata, **Jakarta Selatan** memiliki harga rumah tertinggi, **Bogor** memiliki harga rumah terendah, dan **Tangerang** memiliki harga menengah

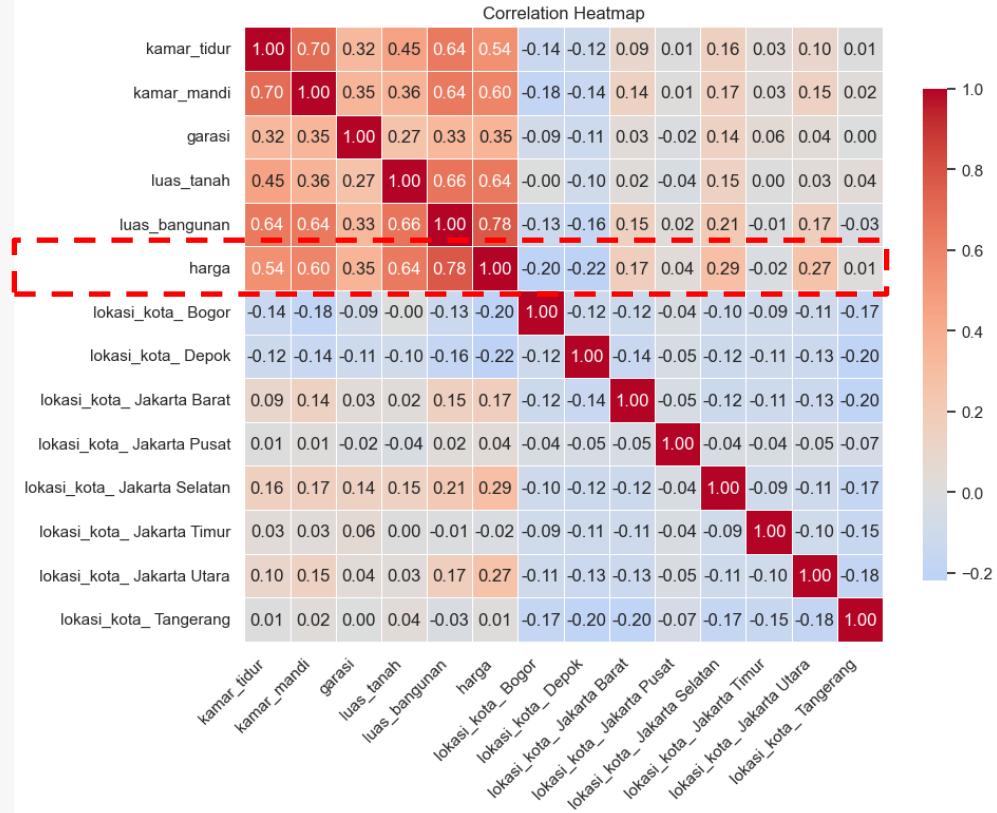
# Exploratory Data Analysis

live  
Class

## Korelasi antar variable

### Catatan

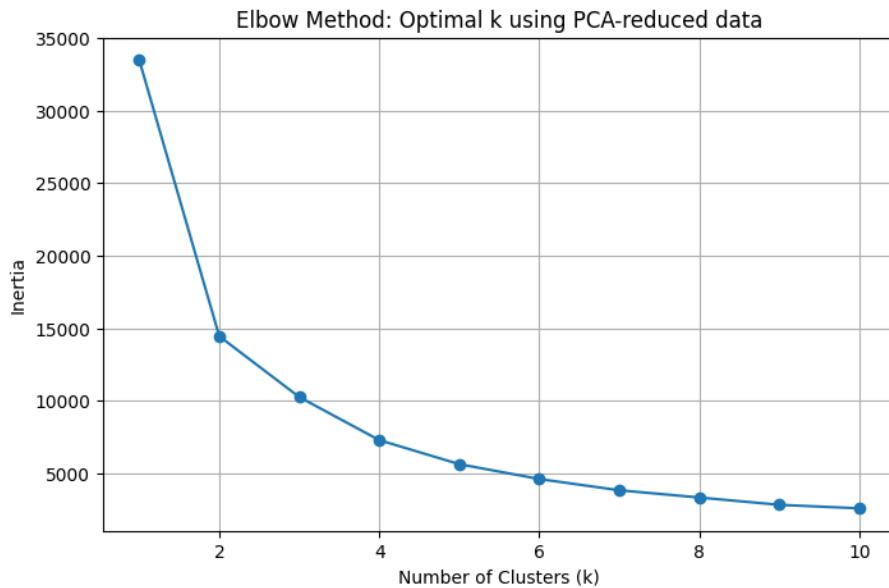
- Dengan menggunakan Pearson Correlation, diketahui bahwa **Harga** berkorelasi kuat dengan **Kamar Tidur, Kamar Mandi, Luas Tanah, dan Luas Bangunan**



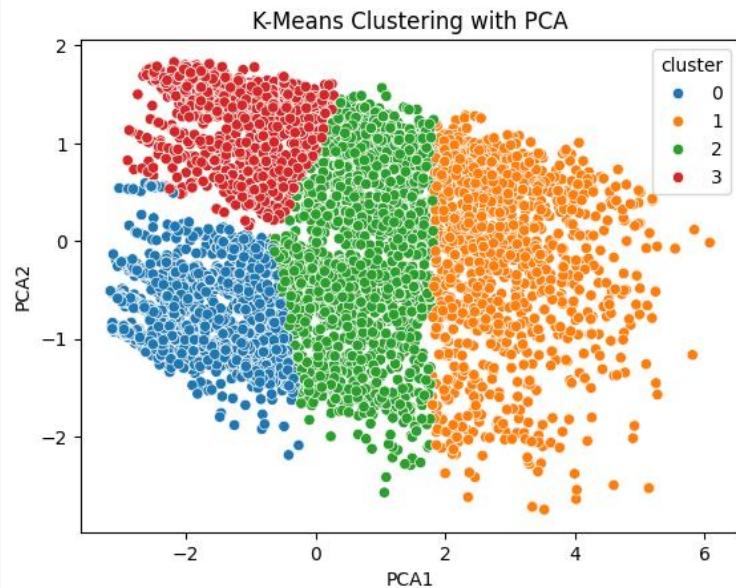
# ML Modeling

live  
Class

Clustering menggunakan metode K-Means & PCA

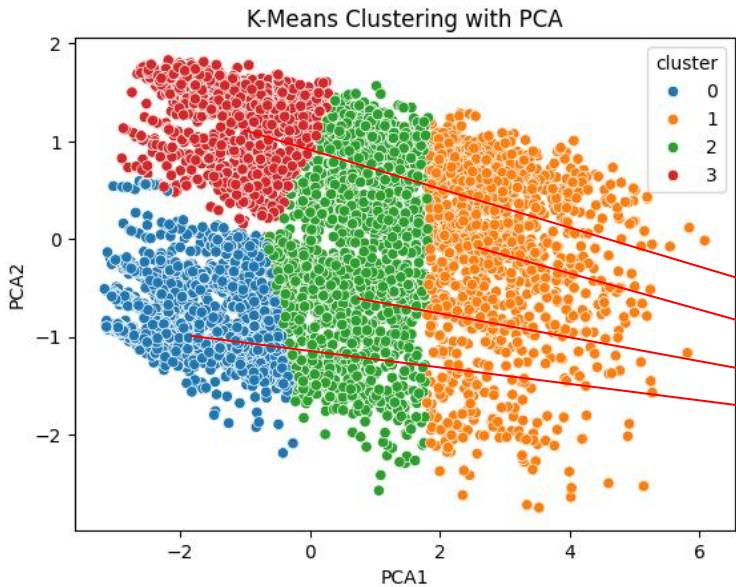


- Dengan elbow method didapat jumlah k optimum = 4 cluster



- Gambaran cluster dengan PCA

# Kesimpulan & Rekomendasi



|   | lokasi_kota_original | lokasi_kota_encoded |
|---|----------------------|---------------------|
| 0 | Bekasi               | 0                   |
| 1 | Bogor                | 1                   |
| 2 | Depok                | 2                   |
| 3 | Jakarta Barat        | 3                   |
| 4 | Jakarta Pusat        | 4                   |
| 5 | Jakarta Selatan      | 5                   |
| 6 | Jakarta Timur        | 6                   |
| 7 | Jakarta Utara        | 7                   |
| 8 | Tangerang            | 8                   |

| clusters | kamar_tidur | kamar_mandi | garasi   | luas_tanah | luas_bangunan | harga        | lokasi_kota |
|----------|-------------|-------------|----------|------------|---------------|--------------|-------------|
| 0        | 2.538035    | 1.686454    | 1.101017 | 86.020977  | 74.586704     | 8.192047e+08 | 1.180684    |
| 1        | 4.161173    | 3.516346    | 1.664200 | 185.471278 | 237.054280    | 4.357276e+09 | 5.342061    |
| 2        | 3.377359    | 2.672897    | 1.382678 | 134.816416 | 158.732507    | 2.755536e+09 | 4.316332    |
| 3        | 2.597421    | 1.986328    | 1.066237 | 74.422708  | 81.306989     | 1.567330e+09 | 7.032109    |

- Berdasarkan clustering, didapatkan 4 cluster :
- **Cluster 0** : rata-rata Rp.819jt, Type 70, berlokasi sekitaran Bogor
- **Cluster 1** : rata-rata Rp. 4.35M, Type 120 2 lantai, berlokasi sekitaran Jaksel
- **Cluster 2** : rata-rata Rp. 2.75M, Type 80 2 lantai, berlokasi sekitaran Jakpus
- **Cluster 3** : rata-rata Rp.1.56M, Type 80, berlokasi sekitaran Jakut
- Jika diurutkan berdasarkan penghasilan dan kemampuan beli maka :
- **Cluster 0 & 3** merupakan kategori yang paling affordable, namun dengan jumlah kamar dan luasan yang lebih kecil, serta lokasi yang kurang strategis
- **Cluster 1 & 2** merupakan kategori yang kurang terjangkau/mahal, namun luasan dan jumlah kamar yg memadai, serta lokasi yang strategis

# Referensi

- Sumber Data : <https://www.rumah123.com/> diambil per 3 November 2025, digunakan hanya untuk kepentingan pembelajaran
- <https://medium.com/data-science/using-weighted-k-means-clustering-to-determine-distribution-centres-locations-2567646fc31d>
- <https://medium.com/data-science/rfm-segmentation-in-e-commerce-e0209ce8fcf6>

# Repository

- <https://github.com/ilham-muhammad/jabodetabek-housing-price>

# TERIMA KASIH

Thanks!