



# Prediksi Harga Properti dan Segmentasi Tipe Properti

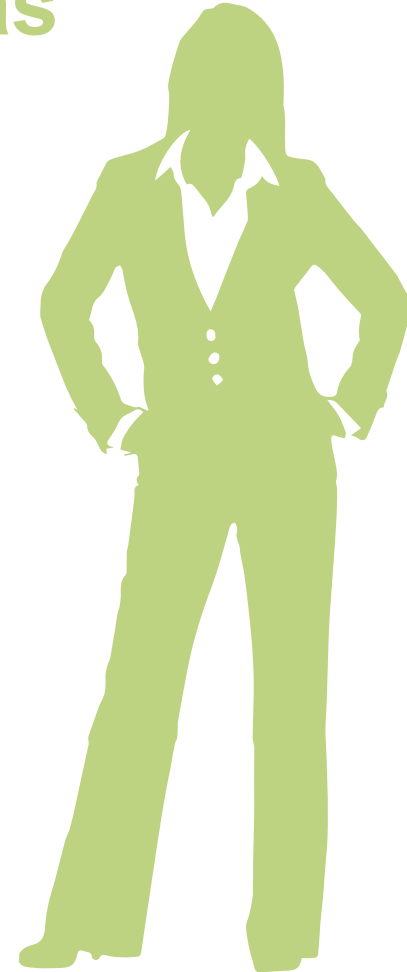
**Mardi Kurnianto  
Nicholas Halasan  
Taara Mona Theresia Sidabutar  
Zaky Ramdhany**

**Final Project – Data Science  
FTDS-HCK-008-GROUP-2**

# Overview

1. Latar Belakang
2. Pembagian Tugas
3. Timeline Project
4. Dataset

5. DAG
6. Deploy
7. Challenges
8. Rekomendasi





# Latar Belakang

Properti merupakan salah satu kebutuhan pokok manusia. Harga properti yang semakin meningkat menjadi salah satu tantangan yang dihadapi oleh para pelaku industri properti . Oleh karena itu, penting untuk memahami faktor-faktor yang mempengaruhi harga properti agar dapat memprediksi harga properti dengan lebih akurat.

Untuk menghadapi tantangan tersebut, Kami perkenalkan *PropertyPrognosis*.

*PropertyPrognosis* lahir berkat semakin tingginya permintaan customer dalam menuntut cepatnya waktu pelayanan, efektifitas dan efisiensi. Dengan *PropertyPrognosis* pelaku usaha dapat dengan cepat melakukan prediksi harga properti serta melakukan klasterisasi / segmentasi tipe properti.



# Property Prognosis

Adalah sebuah aplikasi yang diciptakan menggunakan teknologi Machine Learning, bertujuan untuk memprediksi dengan akurat harga properti berdasarkan sejumlah faktor serta mengelompokkan tipe properti ke dalam klaster atau segmentasi yang tepat.

Dengan kombinasi prediksi harga yang akurat dan pemahaman yang lebih baik terkait preferensi pembeli, PropertyPrognosis menjadi alat bantu yang bermanfaat bagi para pelaku industri properti.





# Pembagian Tugas Tim



Zaky Ramdhany

**Data Engineer**

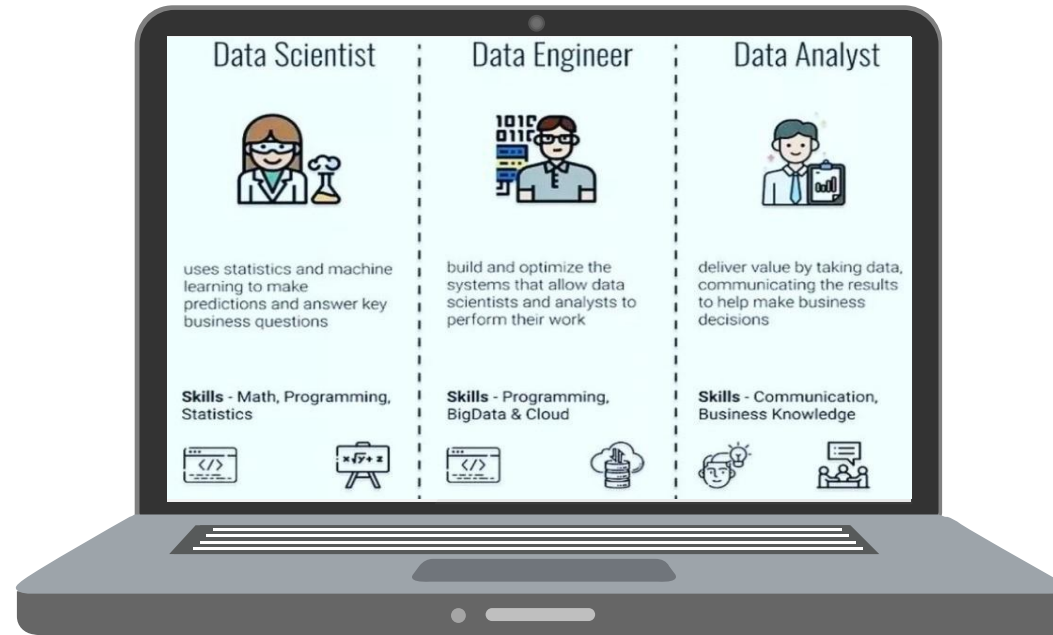
Memegang peran penting dalam mengotomatisasi proses dan memastikan bahwa data yang diolah oleh data analyst dan data scientist akurat dan tepat.



Mardi Kurnianto

**Data Analyst**

Bertugas memberikan rekomendasi bisnis berdasarkan insight yang ditemukan.



Nicholas Halasan  
Taara Mona Theresia Sidabutar

**Data Scientist**

Bertanggung jawab untuk membuat model prediksi regresi harga properti dan clustering tipe properti

# Dataset

Proyek ini menggali analisis mendalam tentang pasar property di Melbourne. Dalam dataset terdiri dari berbagai atribut property serta harga yang sesuai. Proyek ini dibagi menjadi lima fase utama, masing-masing fase mengekstraksi wawasan berharga dari dataset dan menciptakan model prediktif untuk harga rumah.

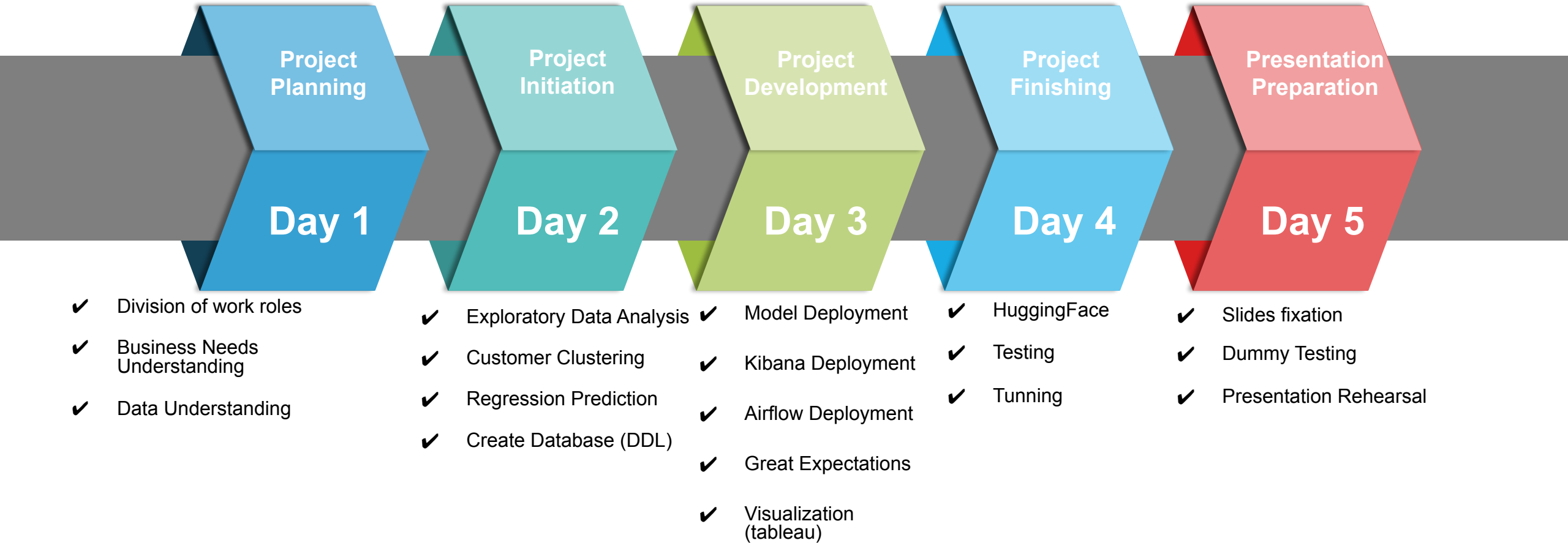
Ref:

<https://www.kaggle.com/datasets/anthonypino/melbourne-housing-market/data>





# Timeline Project

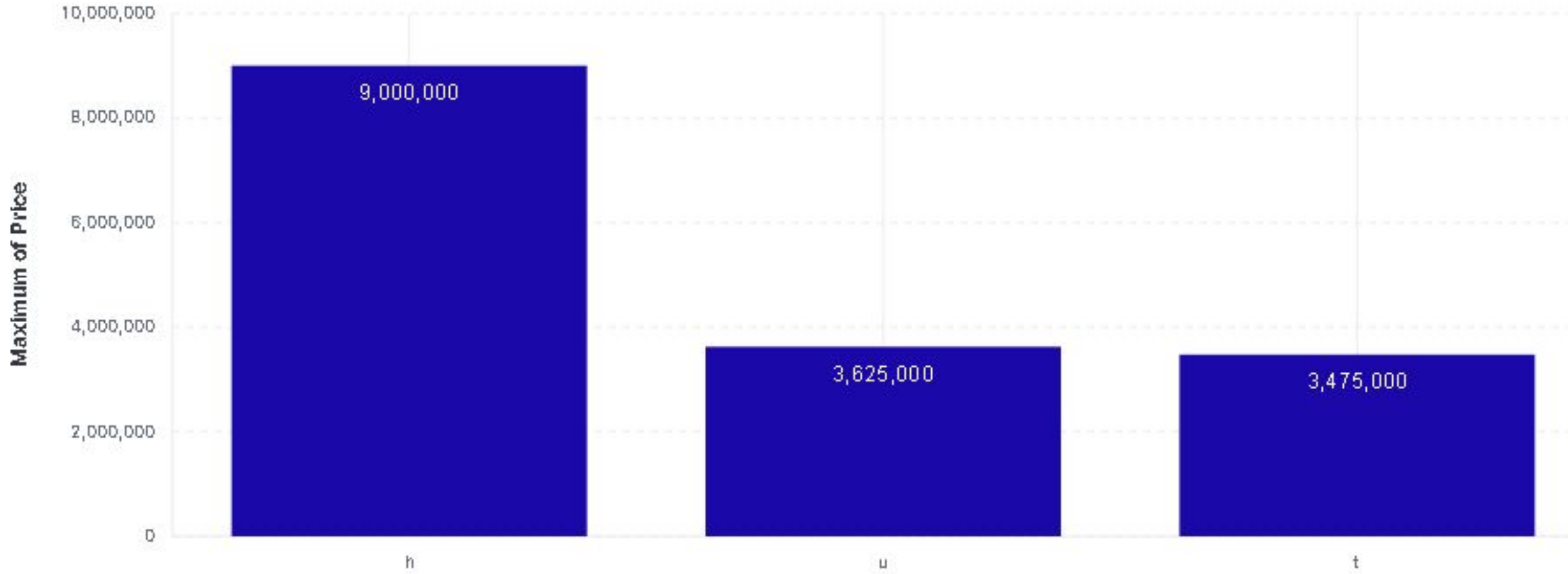




# Infographic



Max Prices by Property Type



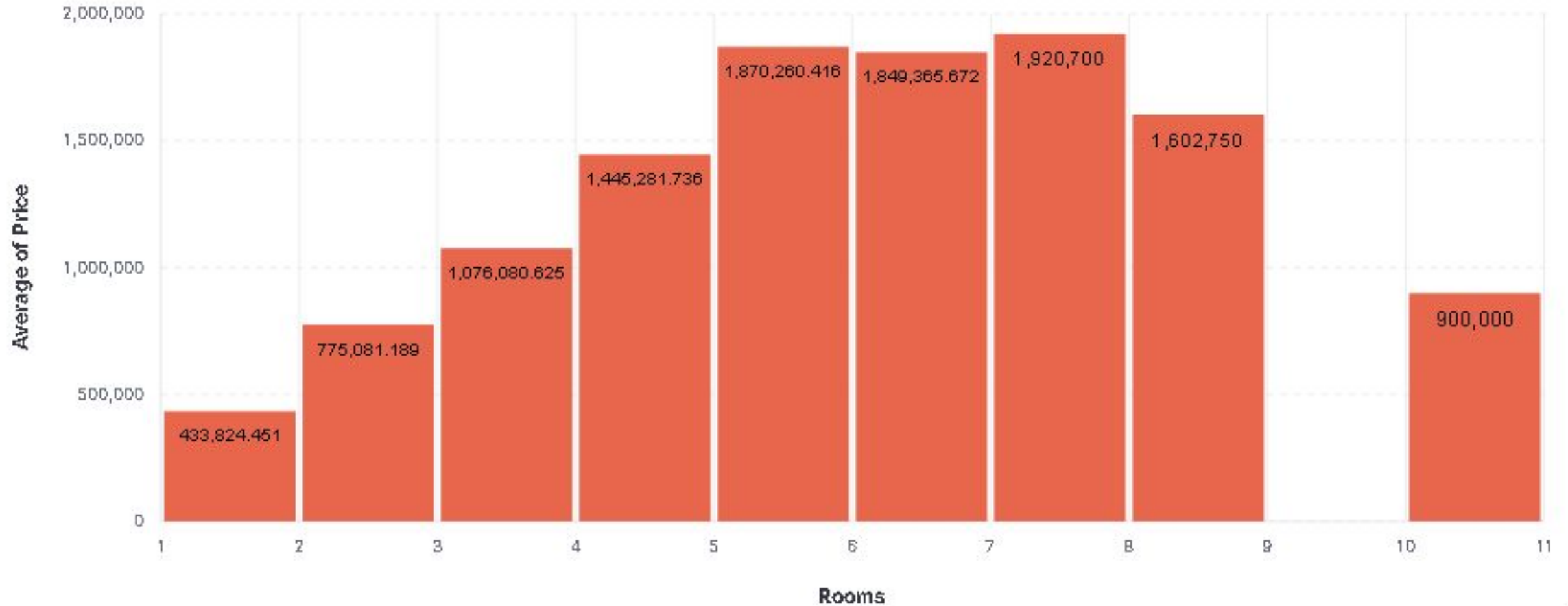
Top 5 values of Type.keyword

Tipe:  
h - rumah, cottage, villa, semi, teras;  
u - unit, duplex;  
t - rumah susun;



# Infographic

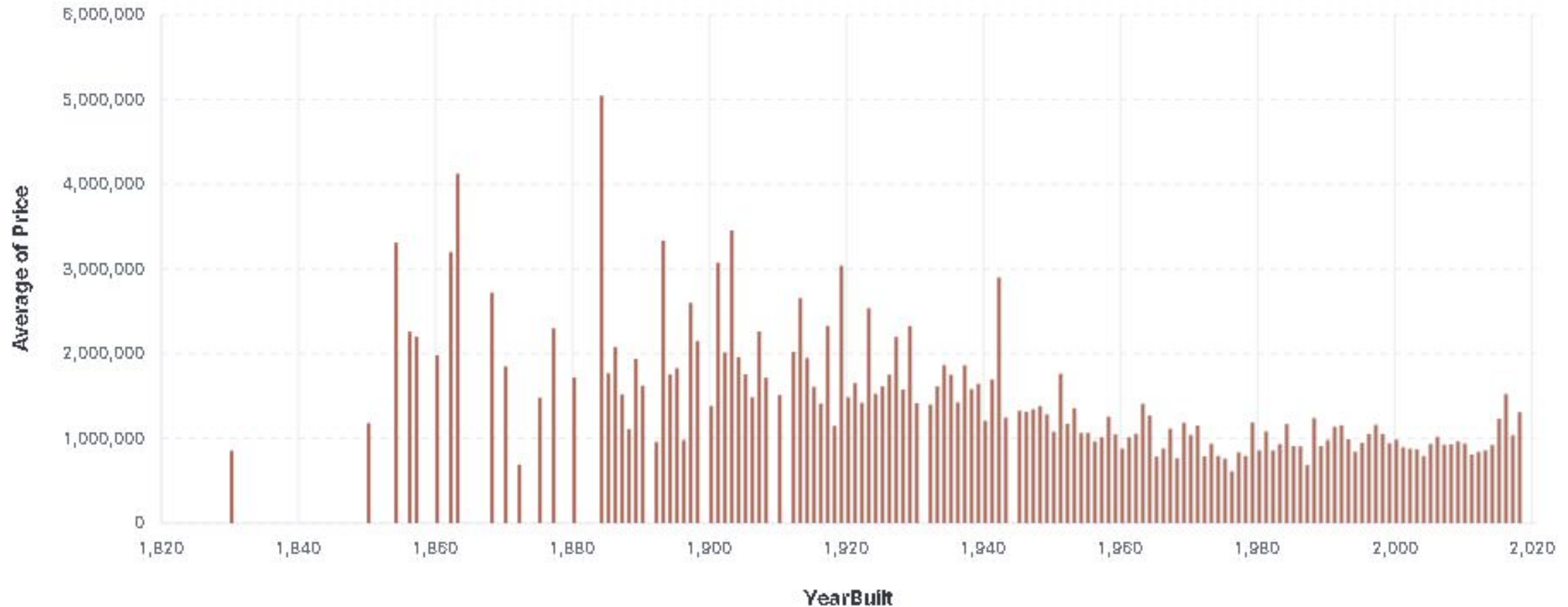
Average Price by Room





# Infographic

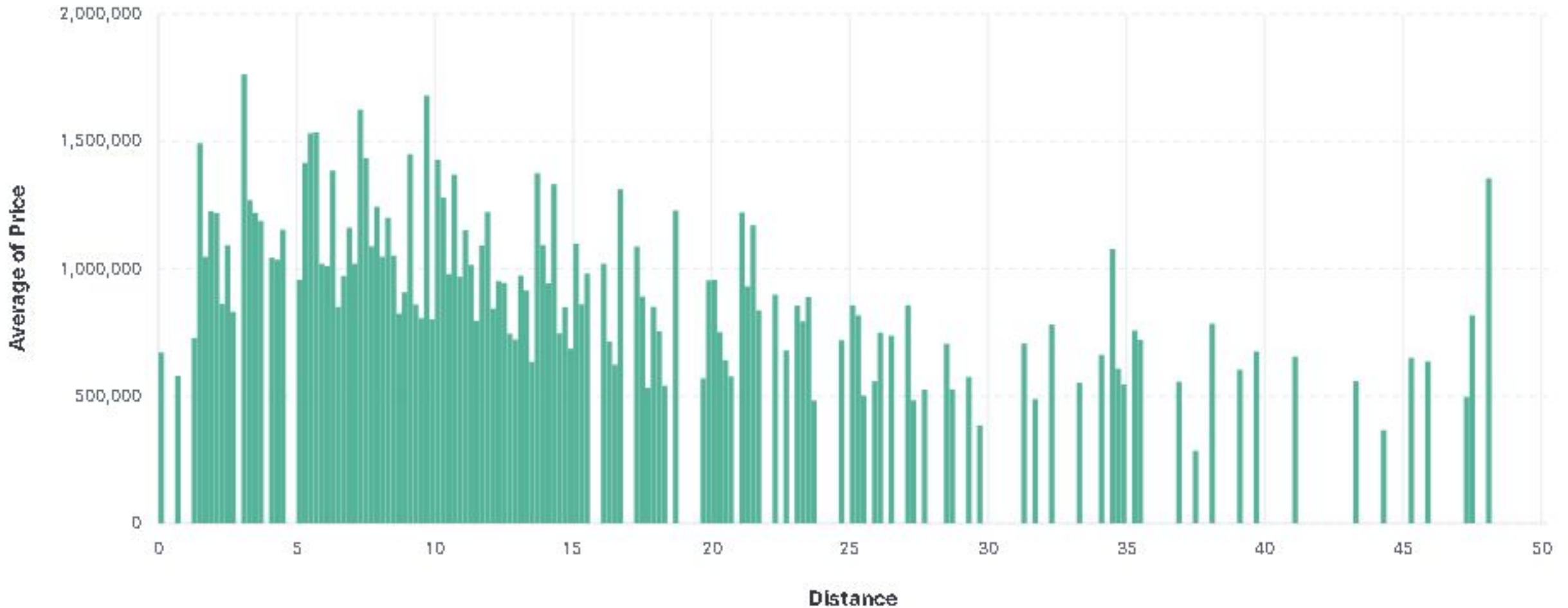
Average Price by Year Built





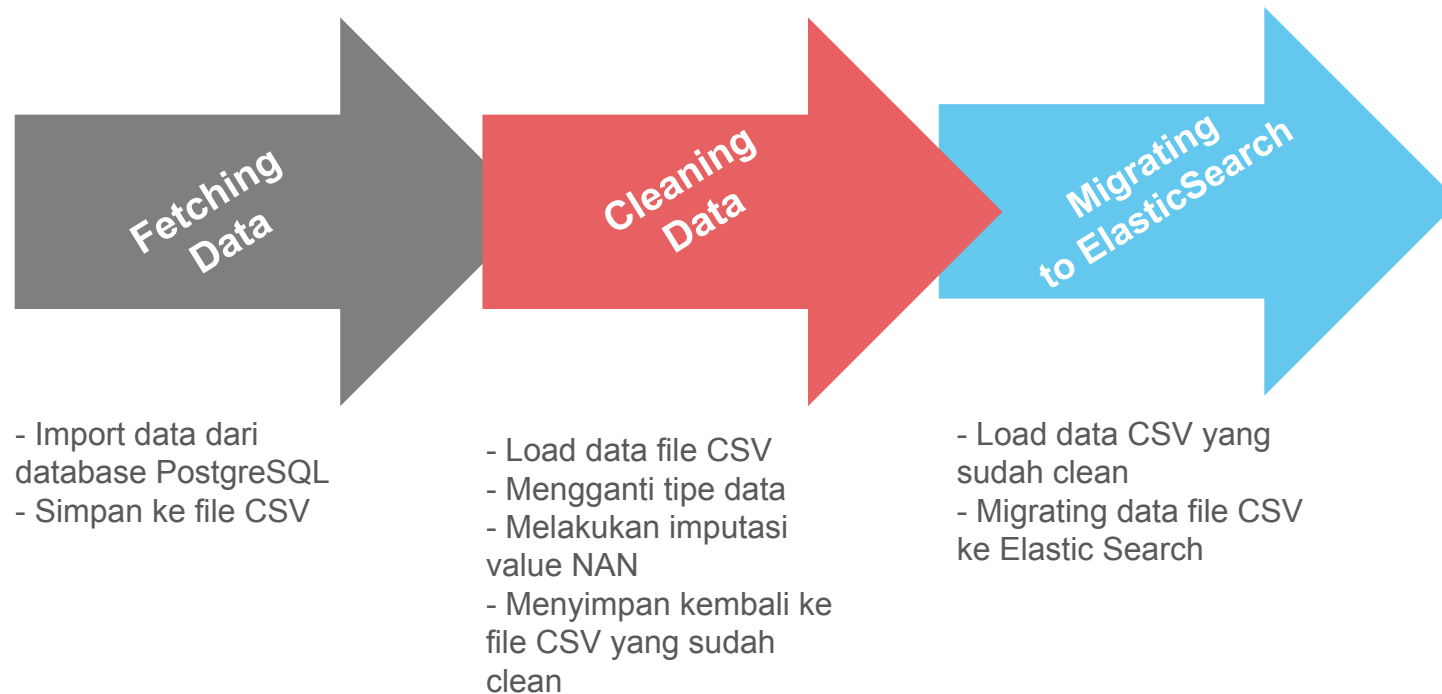
# Infographic

Average Price by Distance



# DAG

(DAG) *Directed Acyclic Graph* adalah representasi visual yang memetakan urutan tugas atau operasi yang akan dilakukan pada data. DAG sering digunakan untuk merepresentasikan hubungan antar elemen dalam berbagai bidang, seperti perencanaan proyek, aliran data dalam pemrograman paralel dan distribusi, serta dalam analisis dependensi dalam pemodelan sistem.





# Method



01



## Prediksi Harga Properti

Dalam membuat prediksi harga properti, kami melibatkan penanganan missing values serta melakukan feature engineering, seleksi fitur berdasarkan korelasi, dan perhitungan VIF untuk mengidentifikasi multicollinearity. Beberapa model seperti Linerar Regression, KNN, Decision Tree, Random Forest, dan SGD dipertimbangkan. Cross-validation digunakan untuk evaluasi model, dan model terbaik dipilih untuk tuning hyperparameter, khususnya pada Random Forest, guna meningkatkan performa. Evaluasi model dilakukan dengan metrik seperti MAE, MSE, dan R-squared untuk memilih model terbaik untuk prediksi harga properti.

02



## Clustering

Dalam membuat clustering tipe properti, kami menggunakan PCA untuk mereduksi dimensi data, kami menggunakan Metode Elbow, Silhouette Plot dan Silhouette Score untuk menentukan jumlah cluster yang optimal. Setelah jumlah cluster ditentukan, K-Means digunakan untuk melatih model clustering. Hasilnya memberikan pemahaman yang lebih mendalam tentang tipe properti yang dikategorikan ke dalam 4 cluster berdasarkan kesamaan karakteristik dan perilaku.





# Regression Result



# Base Model

	Linear Regression	KNN	Decision Tree	Random Forest Regressor	SGD Regressor
Train	45%	66%	99%	94%	45%
Test	46\$	53%	25%	63%	46%

# Cross Validation

	Linear Regression	KNN	Decision Tree	Random Forest Regressor	SGD Regressor
Mean	0,45	0,48	0,19	0,59	0,45
Standar Deviasi	0,01	0,02	0,06	0,02	0,01



# Perbandingan Nilai R2 Sebelum dan Sesudah Hyperparameter Tuning

	Model Default RFR	Model RFR Setelah Hyperparameter Tuning
train	94%	78%
test	63%	62%

# Model Conclusion



- Model default cenderung lebih baik dalam memprediksi data latih (tinggi  $R^2$  dan rendahnya error metrics), namun kemungkinan overfitting pada data uji.
- Terjadi peningkatan pada generalisasi model, namun performa pada data uji menunjukkan sedikit penurunan yang menandakan tuning mungkin belum mengoptimalkan model secara keseluruhan.
- Diperlukan evaluasi lebih lanjut, mungkin dengan eksplorasi hyperparameter lain atau menambah beberapa fitur.





# Property Segmentation



# Model Information

## Model

### K - Means

- Efisiensi
- Dataset dengan Sampel besar
- Dapat diinterpretasikan dengan mudah dan jelas

## Hyper Parameter

- Jumlah Cluster
- Inisialisasi Centroid (init)
- Jumlah Iterasi Maksimum

## Metriks

- Inertia (within - cluster sum of squares)
- Silhouette Score

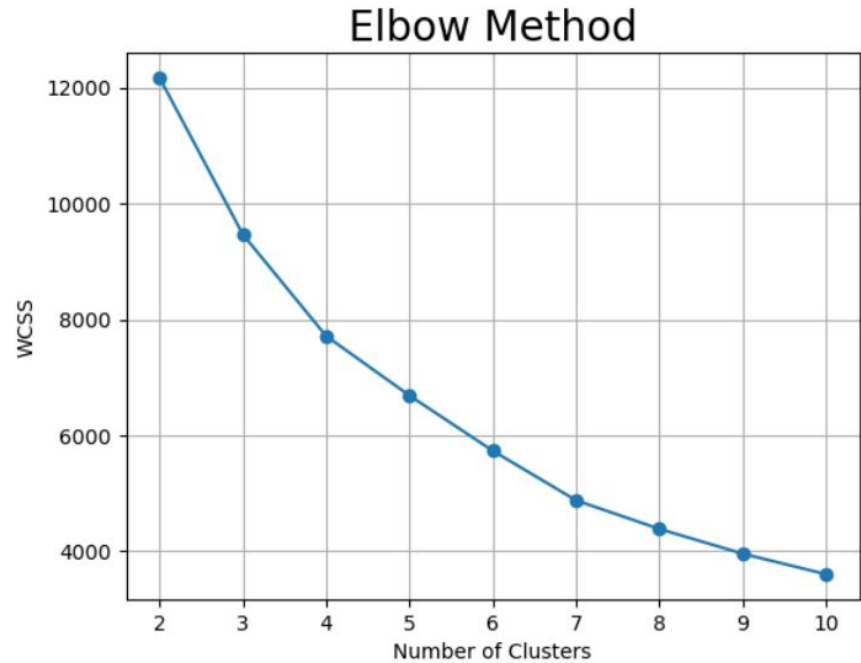
## PCA

- Dimensionality Reduction
- 90% Sebagai Treshold
- didapatkan nilai komponen = 9

# Penentuan Jumlah Cluster

## Elbow Method

Siku terletak pada n-cluster = 4.



## Silhouette Score

3 Silhouette Score tertinggi

Jumlah Cluster	Silhouette Score
3	0.38
4	0.43
5	0.46

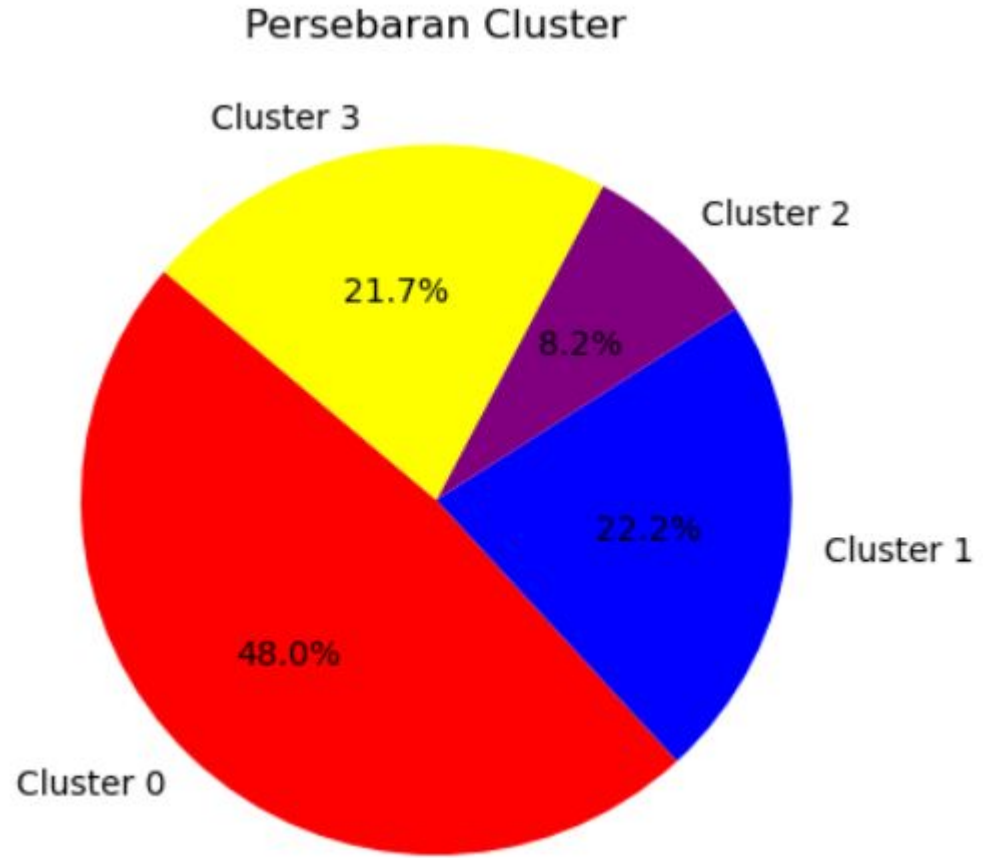


## Final Cluster

Jumlah Cluster yang akan dipakai : 4.



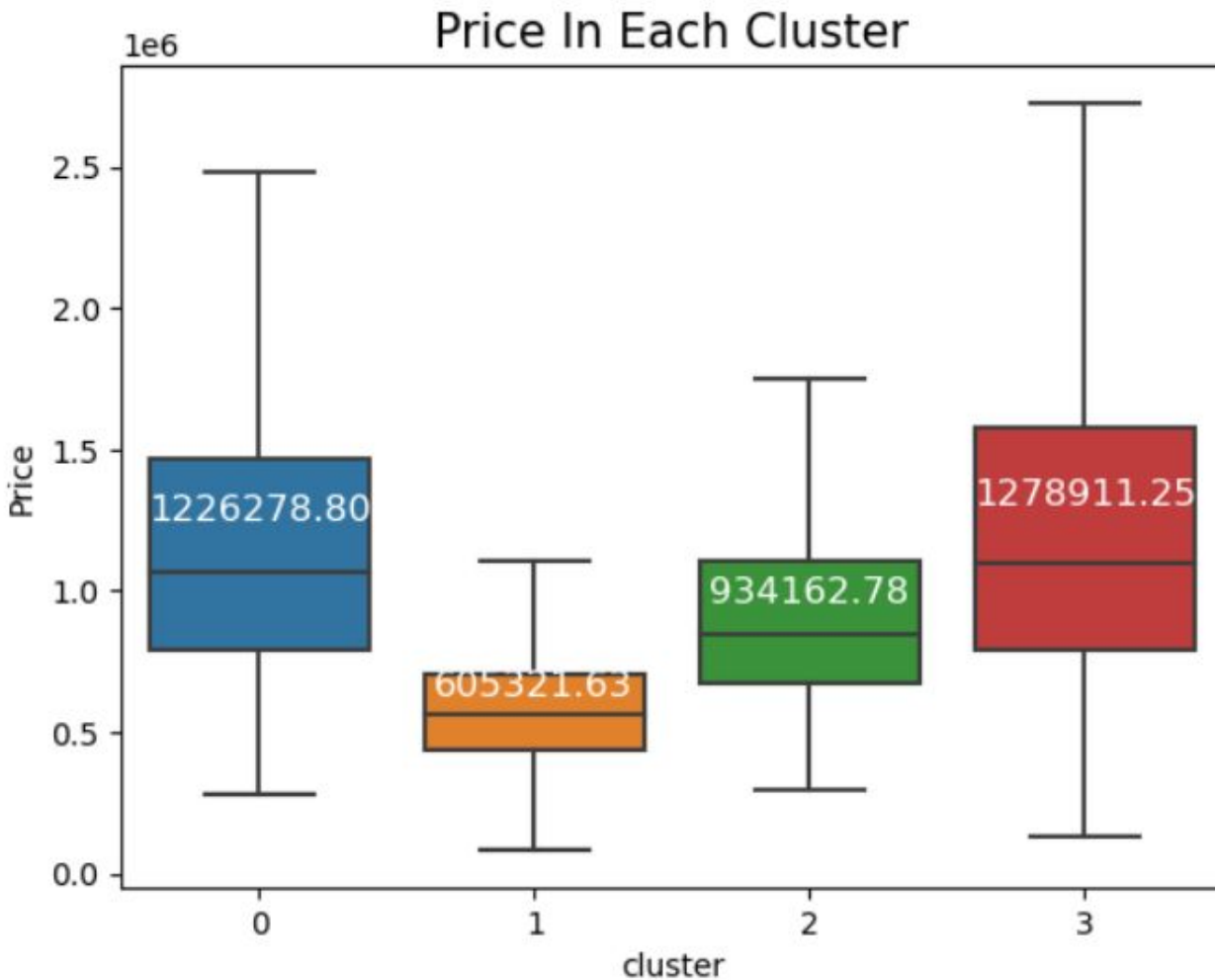
# Clustering Result



CLUSTER	JUMLAH
0	6503
1	3010
2	2936
3	1111



# Cluster Characteristic



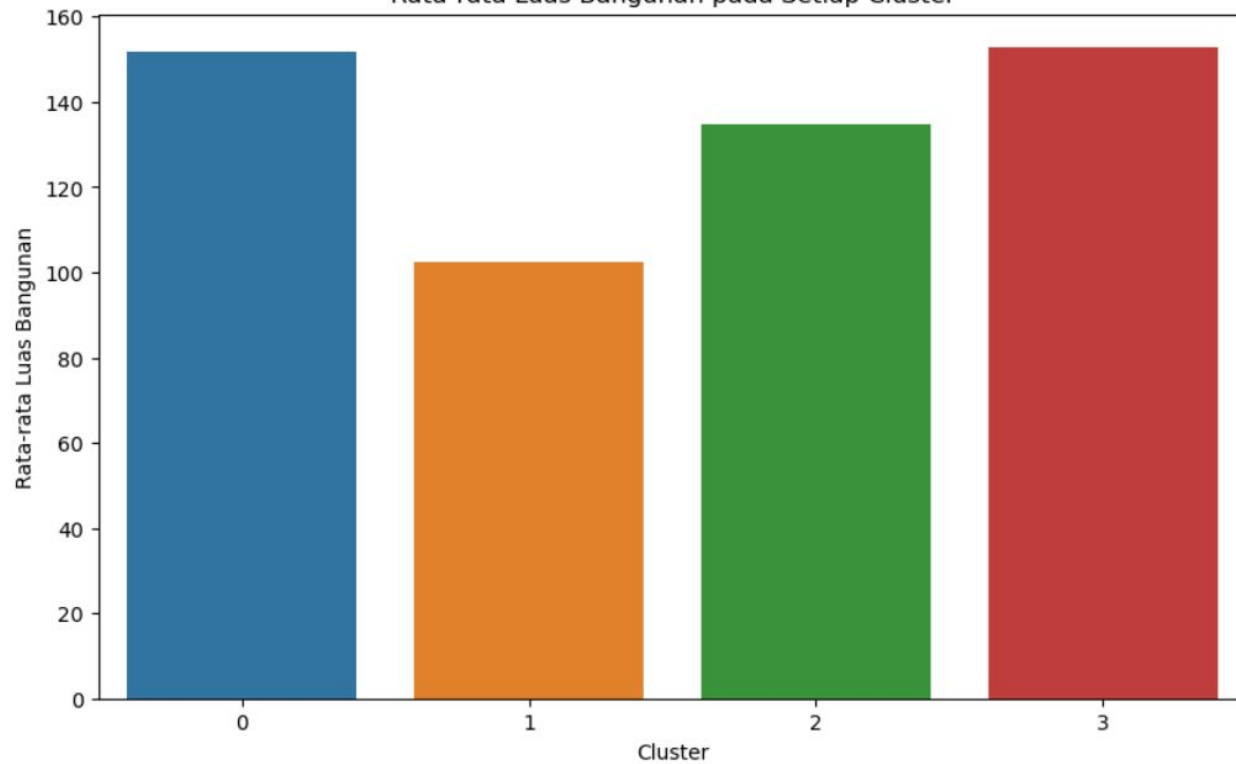
CLUSTER	AVERAGE PRICE	MINIMUM PRICE	MAXIMUM PRICE
0	1.226.279	280.500	7.650.000
1	605.321	85.000	3.625.000
2	934.162	300.000	3.745.000
3	1.278.911	131.000	9.000.000





# Cluster Characteristic

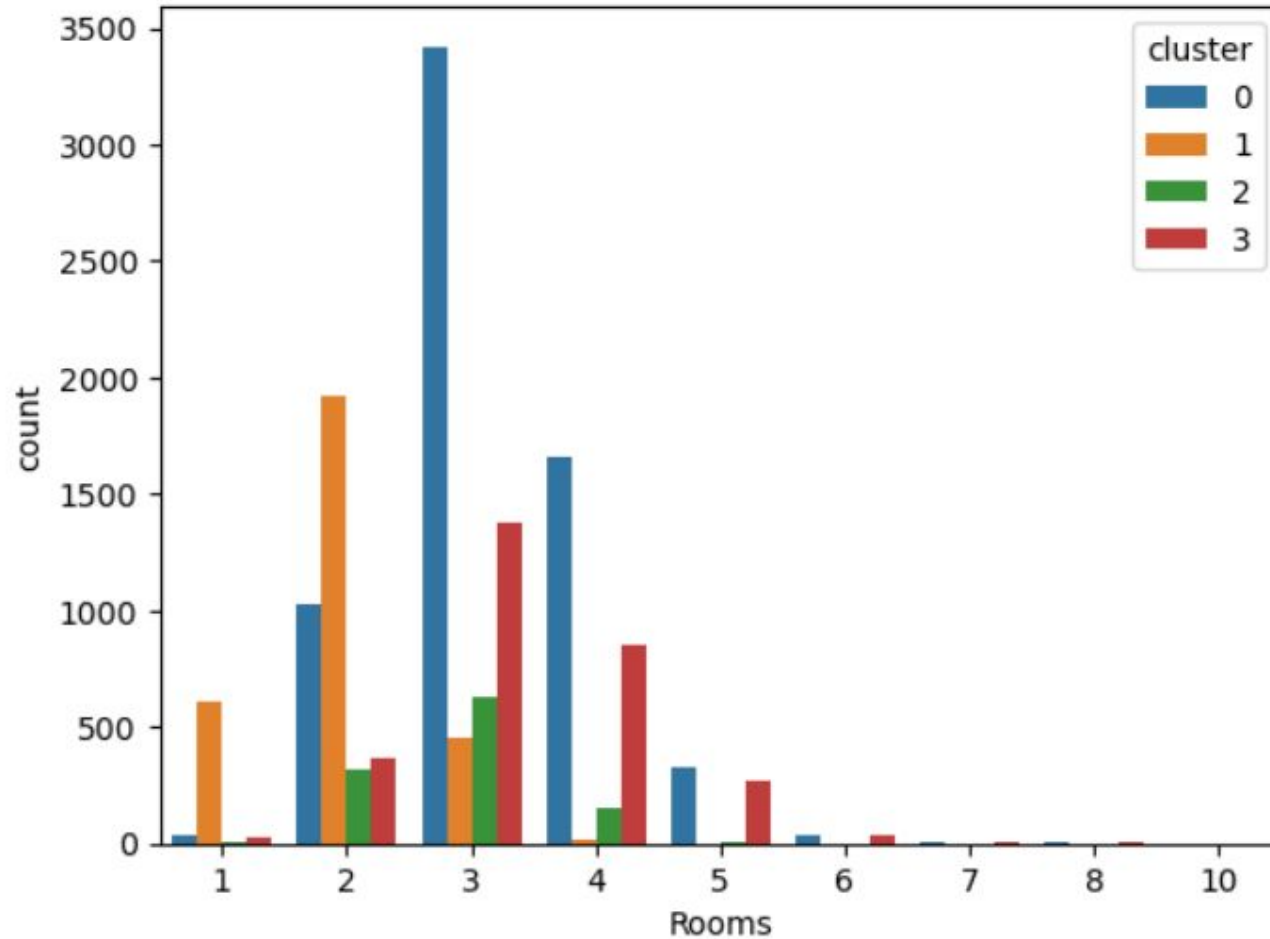
Rata-rata Luas Bangunan pada Setiap Cluster



CLUSTER	RATA-RATA LUAS BANGUNAN (Meter Persegi)
0	152.81
1	102.276
2	134.698
3	152.913

# Cluster Characteristic

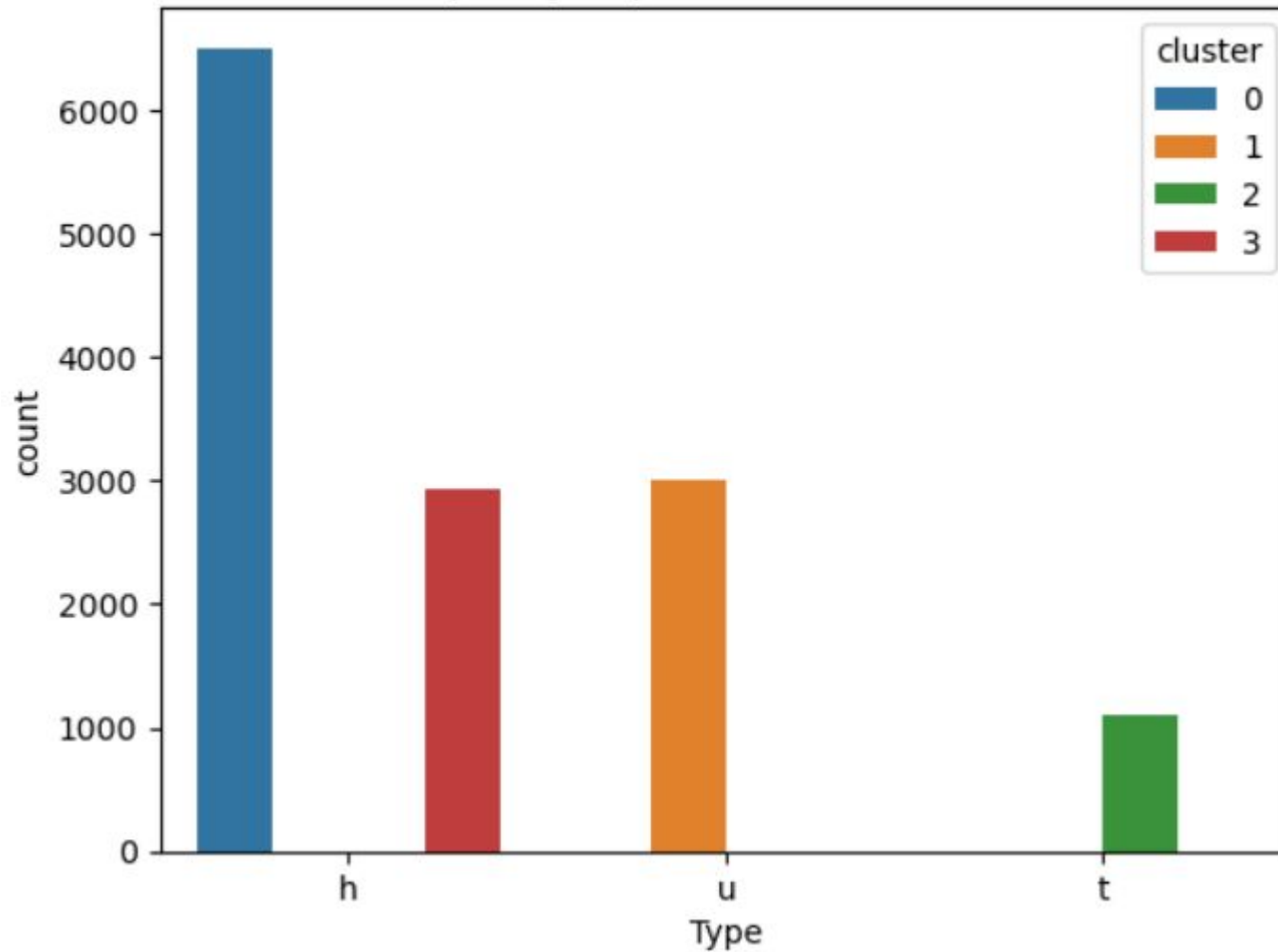
Rooms In Each Cluster



CLUSTER	Rata - Rata Jumlah Ruangan
0	3
1	1
2	2
3	3

# Cluster Characteristic

Property Tipe In Each Cluster

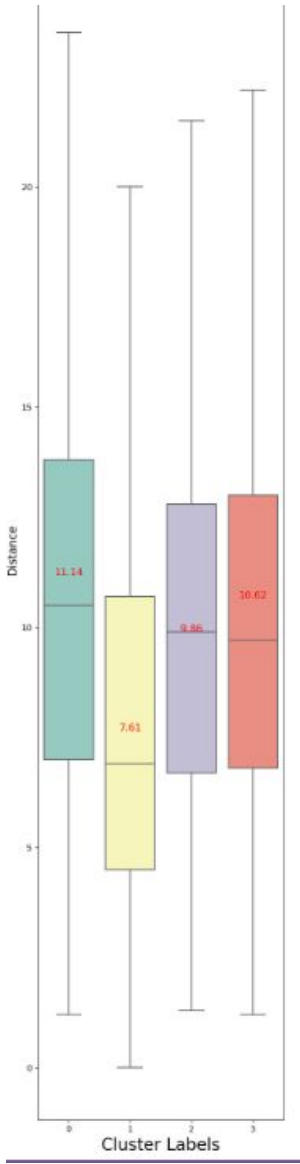


Penjelasan :

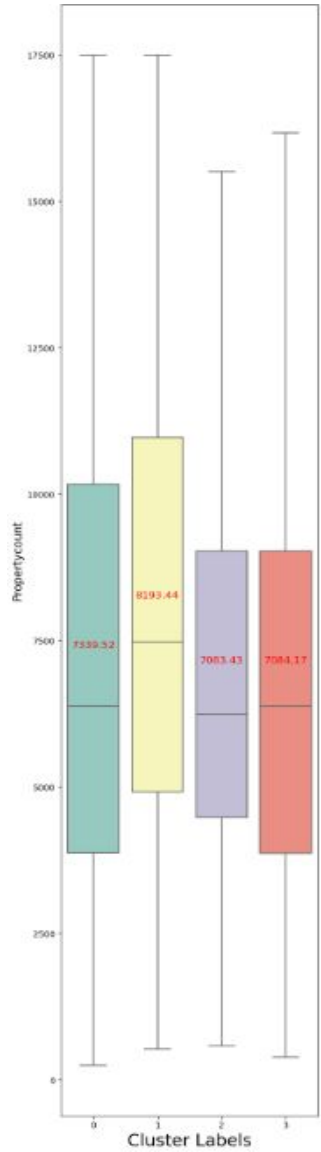
h : house,cottage,villa,terrace

u : unit, duplex

t : townhouse



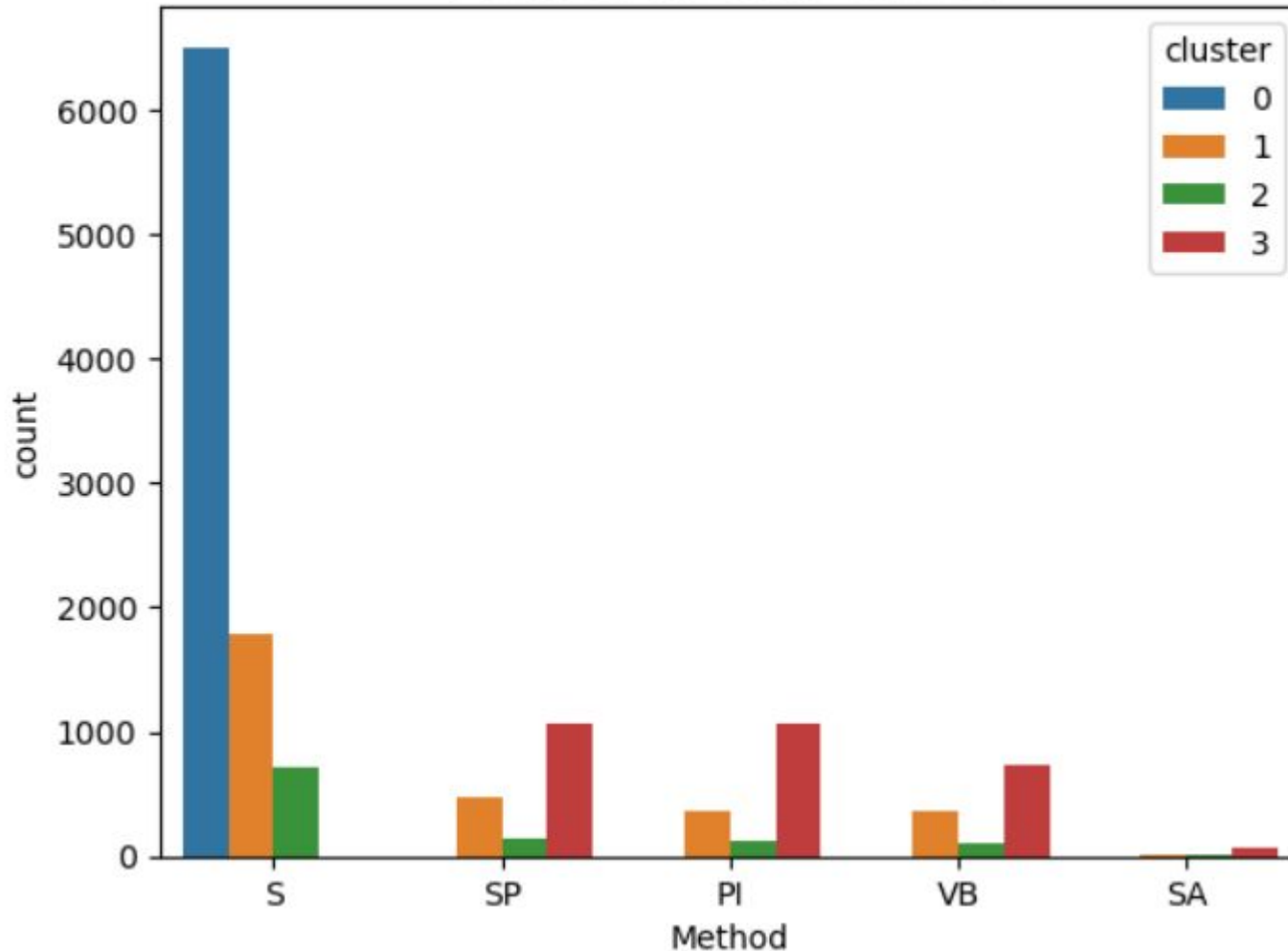
CLUSTER	AVG JARAK DENGAN CBD (Miles)
0	11.14
1	7.61
2	9.86
3	10.62



CLUSTER	AVG PROPERTY COUNT
0	7339
1	8193
2	7083
3	7084

# Cluster Characteristic

Method In Each Cluster



Penjelasan :

S = Sold - Properti Terjual Di Lelang

SP = Sold Prior - Properti Terjual Sebelum Lelang dimulai

PI = Property Passed In - Properti tidak terjual karena harga tidak mencapai harga reserve

VB = Vendor Bid - Penawaran yang diajukan Vendor untuk menaikkan harga

SA = Sold After Auction - Properti Terjual Setelah lelang selesai.





# Cluster Characteristic



## Cluster 0 :

Cluster ini memiliki rata-rata harga properti yang sangat tinggi , Luas Bangunannya merupakan yang paling tinggi bersama dengan cluster 3, Rata-rata ruangan di rumah ini berjumlah 3 ruangan , Biasanya property pada cluster ini bertipe house,cottage,villa,terrace, memiliki jarak terjauh dari CBD , Kepadatan lingkungan propertynya cukup padat dan bervariasi.

## Cluster 1 :

Cluster ini memiliki rata-rata harga properti , luas area, luas bangunan terendah dibanding 3 cluster lainnya , Hanya memiliki rata-rata 1 ruangan karena pada cluster ini biasanya tipenya Unit / Apartemen . Jaraknya Paling dekat dengan CBD . Rata-rata kepadatannya tertinggi , letaknya di tengah kota.

## Cluster 2:

Cluster ini memiliki rata-rata harga dan luas bangunan yang tergolong menengah , diantara 3 cluster lainnya , rata-rata properti dengan 2 ruangan , Tipe propertinya Townhouse, Jaraknya dengan CBD tidak terlalu jauh, dan kepadatan properti di sekitarnya tidak terlalu padat.

Cluster 3= Cluster ini memiliki karakteristik yang hampir sama dengan cluster 0 , yang membedakan dari Metode penjualannya , tidak seperti cluster 0 , cluster 3 ini biasanya tidak terjual diluar lelang , baik sebelum atau sesudah , dan Properti tidak terjual karena harga tidak mencapai harga reserve , kepadatan property di sekitarnya juga tidak setinggi cluster 0.

# Challenges

Membuat model prediksi harga properti melibatkan sejumlah tantangan yang perlu diatasi agar model dapat memberikan hasil yang akurat dan bermanfaat.

## Keterbatasan Data

Ketersediaan data yang terbatas atau data yang tidak lengkap dapat menjadi tantangan. Informasi yang kurang dapat mempengaruhi keakuratan model.

## Faktor Eksternal yang Sulit Diprediksi

Perubahan dalam kebijakan pemerintah, kondisi ekonomi global, atau peristiwa alam dapat sulit diprediksi dan dapat memiliki dampak signifikan pada harga properti.

## Pemodelan Tren Pasar yang Dinamis

Pasar properti dapat mengalami perubahan tren yang cepat. Menangkap dinamika ini dengan model yang statis bisa menjadi tantangan.

## Sensitivitas Terhadap Data Outlier

Keberadaan outlier dalam data bisa memengaruhi hasil prediksi. Menangani outlier dengan benar tanpa kehilangan informasi penting adalah tantangan.

## Heteroskedastisitas

Variabilitas harga properti yang tidak konstan (heteroskedastisitas) dapat mempengaruhi kinerja model. Menangani heteroskedastisitas adalah tantangan tersendiri.



## Tren Teknologi

Perkembangan teknologi dalam industri properti, seperti platform daring dan teknologi konstruksi baru, dapat mempengaruhi nilai properti

## Mengatasi Tantangan

Diperlukan pendekatan yang hati-hati dalam pemilihan data, pengembangan model, dan validasi model menggunakan data yang belum pernah dilihat sebelumnya. Selain itu, iterasi dan peningkatan berkelanjutan terhadap model juga diperlukan untuk memastikan keterampilan prediktif yang optimal.



# Rekomendasi Bisnis



Membuka bisnis properti bisa menjadi usaha yang menarik, terutama dengan semakin meningkatnya minat orang dalam investasi properti.

Untuk meningkatkan peluang properti Anda terjual dengan cepat, Anda dapat mempertimbangkan beberapa strategi bisnis. Berikut adalah beberapa rekomendasi:

## Pemahaman Pasar

Sebelum berinvestasi, lakukan riset pasar menyeluruh. Pahami kebutuhan dan preferensi calon pembeli di wilayah tersebut. Identifikasi tren pasar dan faktor-faktor yang mempengaruhi permintaan properti.

## Harga yang Kompetitif

Tentukan harga yang kompetitif berdasarkan nilai pasar. Jangan mematok harga terlalu tinggi yang dapat membuat calon pembeli enggan. Penetapan harga yang rasional dapat meningkatkan daya tarik properti Anda.

## Tawarkan Fasilitas atau Bonus

Pertimbangkan untuk menawarkan bonus atau fasilitas tambahan kepada pembeli potensial. Ini bisa berupa perabotan, peralatan dapur, atau peningkatan energi yang efisien.

## Pembiayaan Kreatif

Menawarkan opsi pembiayaan kreatif, seperti skema cicilan atau kerjasama dengan lembaga keuangan untuk memberikan opsi pembayaran yang lebih fleksibel kepada calon pembeli.

## Layanan Profesional

Dapatkan bantuan profesional dari agen properti yang berpengalaman atau pemasar properti. Mereka dapat memberikan saran tentang harga yang tepat, strategi pemasaran, dan membantu Anda menjangkau calon pembeli.

## Properti yang Sesuai dengan Pasar

Pilih jenis properti yang sesuai dengan kebutuhan pasar setempat. Misalnya, jika ada kebutuhan besar untuk rumah keluarga, fokuslah pada investasi properti perumahan.





# THANK YOU



**Zaky Ramadhani**

**FTDS-HCK-008**  
**Data Engineer**

**0896 4669 5368**  
**zakynamdhani857@gmail.com**



**Mardi Kurnianto**

**FTDS-HCK-008**  
**Data Analyst**

**0821 1666 6621**  
**mardi.kurnianto@gmail.com**



**Nicholas Halasan**

**FTDS-HCK-008**  
**Data Scientist**

**0812 8068 7117**  
**nicholashg20@gmail.com**



**Taara Mona**

**FTDS-HCK-008**  
**Data Scientist**

**0812 9597 2161**  
**taarasidabutar@gmail.com**

