Pacmann

Housing Price Analytics

Statistic for Business - Sekolah Data Pacmann

© 2022 – Pacmann AI

Outline

- Introduction and Background
- Datasets
- Problem Statements
- Statistical Test
- Regression Model
- Conclusion and Recommendation
- References

© 2022 – Pacmann AI

Introduction and Background

© 2022 – Pacmann AI

Introduction

- Properti adalah salah satu kebutuhan dasar manusia. Penentuan harga properti (rumah) memerlukan perhitungan banyak faktor. Memahami faktor-faktor yang memengaruhi harga properti adalah kunci bagi pembeli, penjual, dan investor untuk membuat keputusan yang tepat.

© 2022 – Pacmann AI

Background

“Bagaimana hubungan harga rumah dengan luas area, airconditioning dan furniture status?”

© 2022 – Pacmann AI

Background

- Untuk menjawab masalah ini, kita dapat melakukan beberapa analisis, seperti:
  - **Memprediksi harga berdasarkan area**
  - **Memprediksi harga berdasarkan area dan apakah ada AC atau tidak**
  - **Memprediksi harga berdasarkan area dan furnishing status**

© 2022 – Pacmann AI

Dataset

Data yang digunakan dalam penelitian kali ini berasal dari situs Kaggle, yaitu Housing Prices Dataset.

Data ini memiliki kolom-kolom sebagai berikut:

- price
- area
- bedrooms
- bathrooms
- stories
- mainroad
- guestroom
- basement
- waterheating
- airconditioning
- parking
- prefarea
- furnishingstatus

© 2022 – Pacmann AI

Statistical Test

1. Apakah harga rata-rata dari properti di area ini lebih dari 6.000.000?

- Hipotesis Nol (H0): μ <= 6,000,000 (Harga rata-rata sama dengan 6,000,000)
- Hipotesis Alternatif (H1): μ > 6,000,000 (Harga rata-rata lebih dari 6,000,000)

Uji statistik menggunakan uji t-statistic dan alpha = 0.05.

```
Jumlah data: 545
------------------------------------------------------------
rata-rata sampel: 4766729.2477
standar deviasi sampel: 1870439.6157
------------------------------------------------------------
alpha: 0.05
nilai uji t-statistic: -15.3926
nilai critical: 1.6477
------------------------------------------------------------
Tidak cukup bukti untuk menolak H0. Rata-rata harga properti sama dengan atau kurang dari 6,000,000.
```

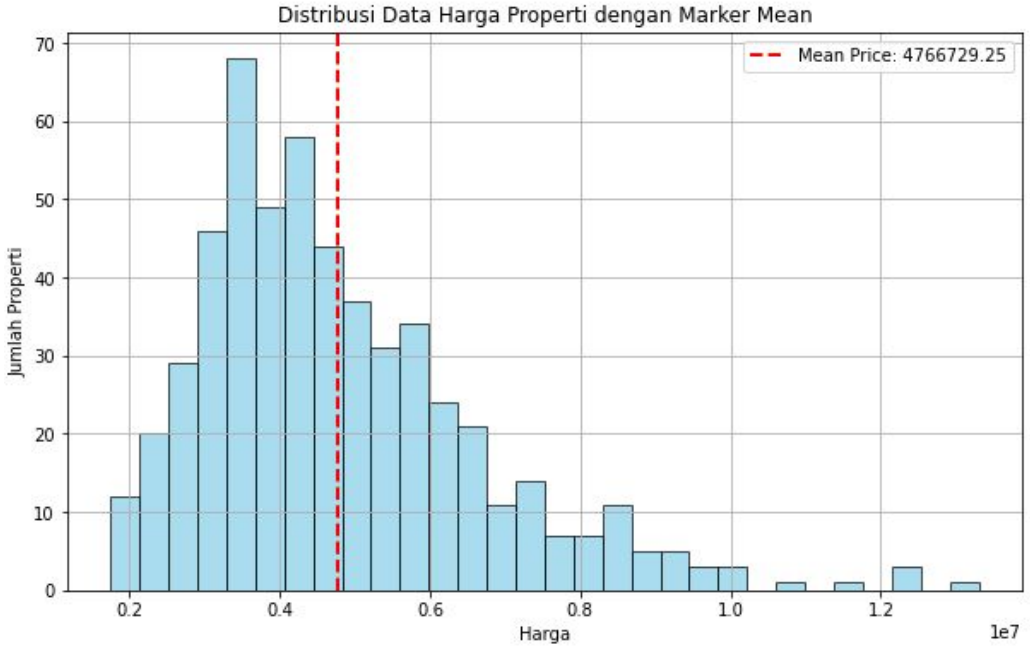

© 2022 – Pacmann AI

Statistical Test

1. Apakah harga rata-rata dari properti di area ini lebih dari 6.000.000?

- Hipotesis Nol (H0): μ <= 6,000,000 (Harga rata-rata sama dengan 6,000,000)
- Hipotesis Alternatif (H1): μ > 6,000,000 (Harga rata-rata lebih dari 6,000,000)

Uji statistik menggunakan uji t-statistic dan alpha = 0.05.



Statistical Test

1. Apakah harga rata-rata dari properti di area ini lebih dari 6.000.000?

- Hipotesis Nol (H0): μ <= 6,000,000 (Harga rata-rata sama dengan 6,000,000)
- Hipotesis Alternatif (H1): μ > 6,000,000 (Harga rata-rata lebih dari 6,000,000)

Uji statistik menggunakan uji t-statistic dan alpha = 0.05.

Jadi, dari tes statistik dan visualisasi disimpulkan bahwa harga rata-rata dibawah 6,000,000.

© 2022 – Pacmann AI

Statistical Test

2. Apakah proporsi properti yang memiliki AC (Air Conditioning) lebih dari 50%?

- Hipotesis Nol (H0): p <= 0.50 (Proporsi properti dengan AC sama dengan 50%)
- Hipotesis Alternatif (H1): p > 0.50 (Proporsi properti dengan AC lebih dari 50%)

Uji statistik menggunakan Uji Z dan alpha = 0.05

```
rata-rata sampel: 0.3156
standar deviasi: 0.0199
------------------------------------------------------------
alpha: 0.05
nilai uji z-statistics: -9.2629
nilai z-critical: 1.6449
------------------------------------------------------------
Tidak cukup bukti untuk menolak H0. Proporsi properti dengan AC kurang dari atau sama dengan 50%.
```

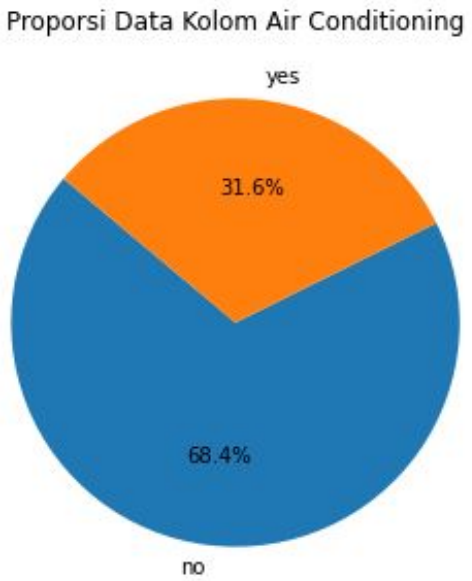

© 2022 – Pacmann AI

Statistical Test

2. Apakah proporsi properti yang memiliki AC (Air Conditioning) lebih dari 50%?

- Hipotesis Nol (H0): p <= 0.50 (Proporsi properti dengan AC sama dengan 50%)
- Hipotesis Alternatif (H1): p > 0.50 (Proporsi properti dengan AC lebih dari 50%)

Uji statistik menggunakan Uji Z dan alpha = 0.05



© 2022 – Pacmann AI

Statistical Test

2. Apakah proporsi properti yang memiliki AC (Air Conditioning) lebih dari 50%?

- Hipotesis Nol (H0): p <= 0.50 (Proporsi properti dengan AC sama dengan 50%)
- Hipotesis Alternatif (H1): p > 0.50 (Proporsi properti dengan AC lebih dari 50%)

Uji statistik menggunakan Uji Z dan alpha = 0.05

Jadi, dari tes statistik dan visualisasi disimpulkan bahwa proporsi properti AC kurang dari atau sama dengan 50%.

© 2022 – Pacmann AI

Statistical Test

3. Apakah harga rata-rata properti dengan dan tanpa AC berbeda signifikan?

- Hipotesis Nol (H0): μ1 = μ2 (Tidak ada perbedaan harga rata-rata antara properti dengan AC dan tanpa AC)
- Hipotesis Alternatif (H1): μ1 ≠ μ2 (Ada perbedaan harga rata-rata antara properti dengan AC dan tanpa AC)

Uji statistik menggunakan t-statistic dan alpha = 0.05

```
jumlah data dengan AC: 172
rata-rata sampel dengan AC: 6013220.5814
standar deviasi dengan AC: 1998149.4750
------------------------------------------------------------
jumlah data tanpa AC: 373
rata-rata sampel tanpa AC: 4191939.6783
standar deviasi tanpa AC: 1493711.7610
------------------------------------------------------------
alpha: 0.05
nilai uji t-statistic: 10.65924416552892
nilai t-critical: 1.9643423968425016
------------------------------------------------------------
Tolak H0. Ada perbedaan harga rata-rata antara properti dengan AC dan tanpa AC.
```


© 2022 – Pacmann AI

Statistical Test

3. Apakah harga rata-rata properti dengan dan tanpa AC berbeda signifikan?

- Hipotesis Nol (H0): μ1 = μ2 (Tidak ada perbedaan harga rata-rata antara properti dengan AC dan tanpa AC)
- Hipotesis Alternatif (H1): μ1 ≠ μ2 (Ada perbedaan harga rata-rata antara properti dengan AC dan tanpa AC)

Uji statistik menggunakan t-statistic dan alpha = 0.05

© 2022 – Pacmann AI

Statistical Test

3. Apakah harga rata-rata properti dengan dan tanpa AC berbeda signifikan?

- Hipotesis Nol (H0): μ1 = μ2 (Tidak ada perbedaan harga rata-rata antara properti dengan AC dan tanpa AC)
- Hipotesis Alternatif (H1): μ1 ≠ μ2 (Ada perbedaan harga rata-rata antara properti dengan AC dan tanpa AC)

Uji statistik menggunakan t-statistic dan alpha = 0.05

Jadi, dari tes statistik dan visualisasi disimpulkan bahwa perbedaan rata-rata harga antara properti dengan AC dan tanpa AC

© 2022 – Pacmann AI

Statistical Test

4. Apakah harga rata-rata properti tiap furnishing status berbeda signifikan?

- Hipotesis Nol (H0): Tidak ada perbedaan harga rata-rata antara ketiga kelompok furnishing status (“Furnished”, “Semi-Furnished”, dan “Unfurnished”).
- Hipotesis Alternatif (H1): Ada perbedaan harga rata-rata antara setidaknya satu pasang kelompok furnishing status.

Uji statistic menggunakan ANOVA, alpha = 0.05

```
F-statistic: 28.27
p-value: 0.0000
alpha: 0.05
----------------------------------------------------------
Tolak H0. Ada perbedaan harga rata-rata yang signifikan antara kelompok furnishing status.
```

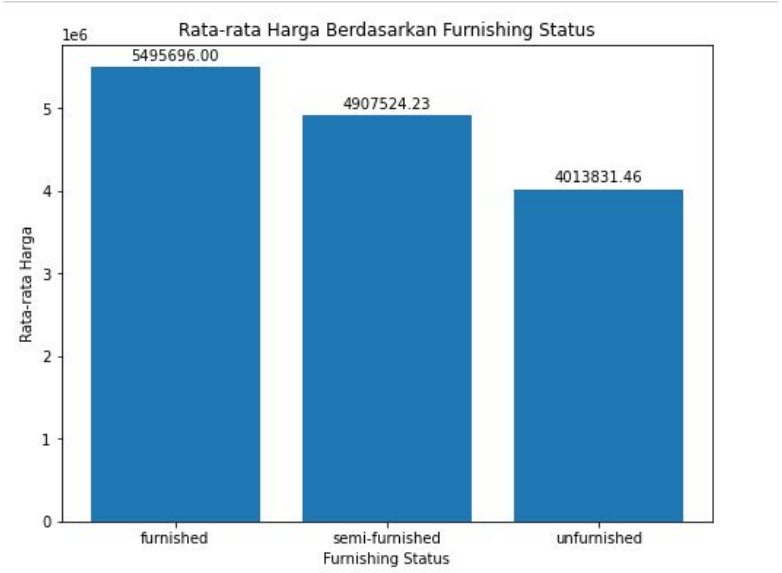

© 2022 – Pacmann AI

Statistical Test

4. Apakah harga rata-rata properti tiap furnishing status berbeda signifikan?

- Hipotesis Nol (H0): Tidak ada perbedaan harga rata-rata antara ketiga kelompok furnishing status (“Furnished”, “Semi-Furnished”, dan “Unfurnished”).
- Hipotesis Alternatif (H1): Ada perbedaan harga rata-rata antara setidaknya satu pasang kelompok furnishing status.

Uji statistic menggunakan ANOVA, alpha = 0.05



© 2022 – Pacmann AI

Statistical Test

4. Apakah harga rata-rata properti tiap furnishing status berbeda signifikan?

- Hipotesis Nol (H0): Tidak ada perbedaan harga rata-rata antara ketiga kelompok furnishing status (“Furnished”, “Semi-Furnished”, dan “Unfurnished”).
- Hipotesis Alternatif (H1): Ada perbedaan harga rata-rata antara setidaknya satu pasang kelompok furnishing status.

Uji statistic menggunakan ANOVA, alpha = 0.05

Jadi, dari tes statistik dan visualisasi disimpulkan bahwa ada perbedaan rata-rata harga antara yang signifikan antara furnishing status.

© 2022 – Pacmann AI

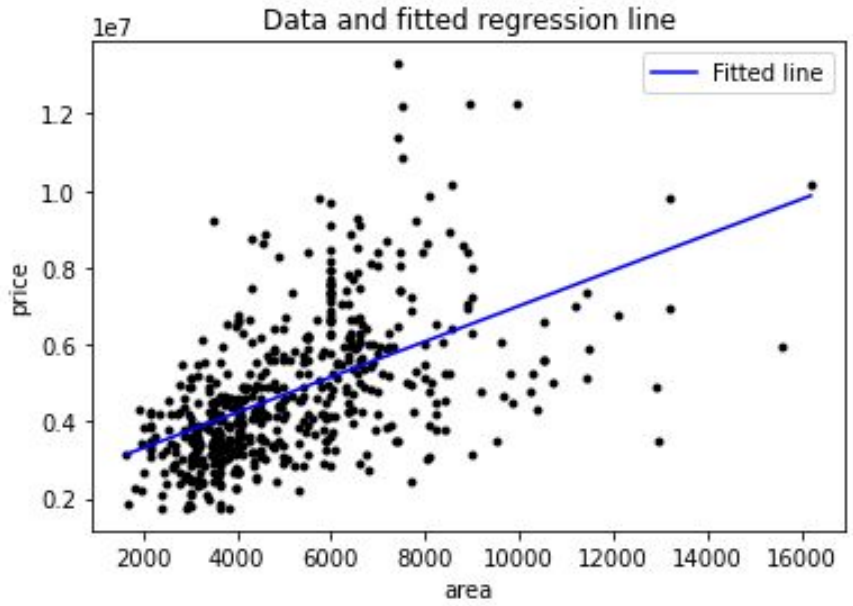
Regression Model

- Setelah melakukan uji statistik, selanjutnya kita bisa melakukan pemodelan dengan linear regression menggunakan prediktor yang sudah kita uji. Pada kesempatan kali ini saya hanya akan menggunakan kolom **area**, **airconditioning**, dan **furnishingstatus**.

© 2022 – Pacmann AI

Regression Model

1. Prediksi harga berdasarkan area (Linear Regression dengan satu prediktor.)



Hasil visualisasi regression line pada kolom price dan kolom area.

© 2022 – Pacmann AI

Regression Model

1. Prediksi harga berdasarkan area (Linear Regression dengan satu prediktor.)

	coef	std err
Intercept	2.387308e+06	174497.798084
area	4.619749e+02	31.225636

$$y = 2387308 + 461x$$

Interpretasi:

- Rata-rata harga rumah ketika luas areanya 0 adalah 2.387.308

- Rata-rata perbedaan harga antara dua rumah dengan perbedaan satu satuan area adalah 461.974

Nilai harga rumah dalam jutaan agak sulit dibaca, mari kita ubah nilai pada kolom price ke bentuk ribuan.

© 2022 – Pacmann AI

Regression Model

1. Prediksi harga berdasarkan area (Linear Regression dengan satu prediktor.)

	coef	std err
Intercept	2387.308482	174.497798
area	0.461975	0.031226

$$y = 2387 + 0.461x$$

Interpretasi:

- Rata-rata harga rumah ketika luas areanya 0 adalah 2387 ribu dolar

- Rata-rata perbedaan harga antara dua rumah dengan perbedaan satu satuan luas area adalah 0.46 ribu dolar

© 2022 – Pacmann AI

Regression Model

1. Prediksi harga berdasarkan area (Linear Regression dengan satu prediktor.)

```
# R-squared before scalling
results.rsquared
```

```
0.2872931546811468
```

```
# R-squared after scalling
results_scaled.rsquared
```

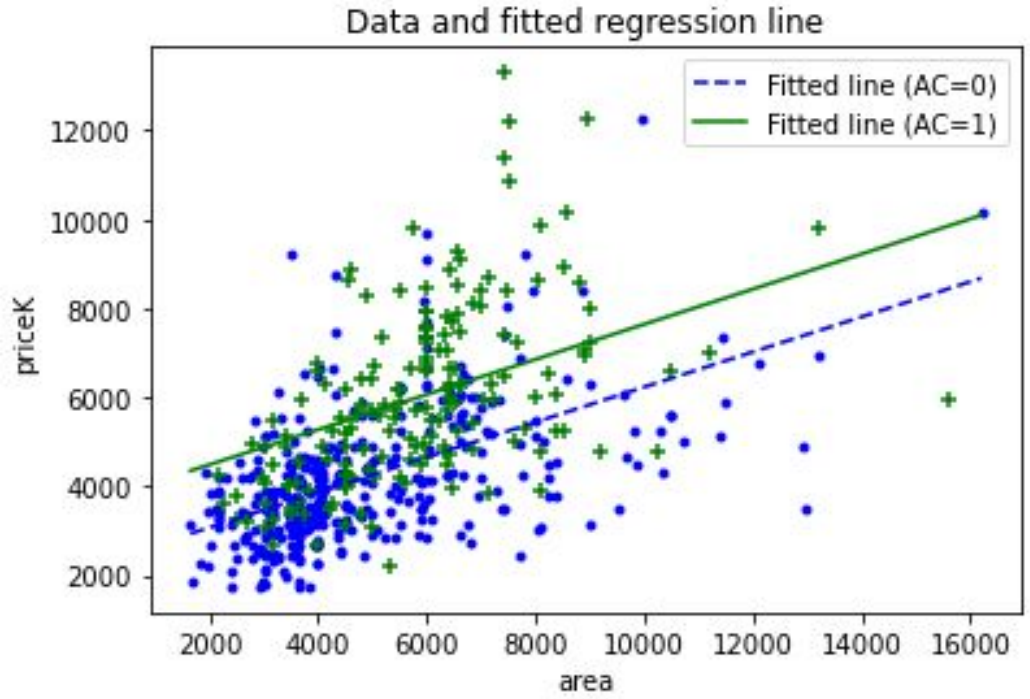
```
0.2872931546811468
```

Performa model sebelum dan sesudah dilakukan scalling tetap sama, tidak ada perbedaan sama sekali. Namun scalling dapat mempermudah menginterpretasi persamaan model

© 2022 – Pacmann AI

Regression Model

2. Prediksi harga berdasarkan luas area dan apakah ada AC atau tidak. (Linear Regression dengan multiple prediktor.)

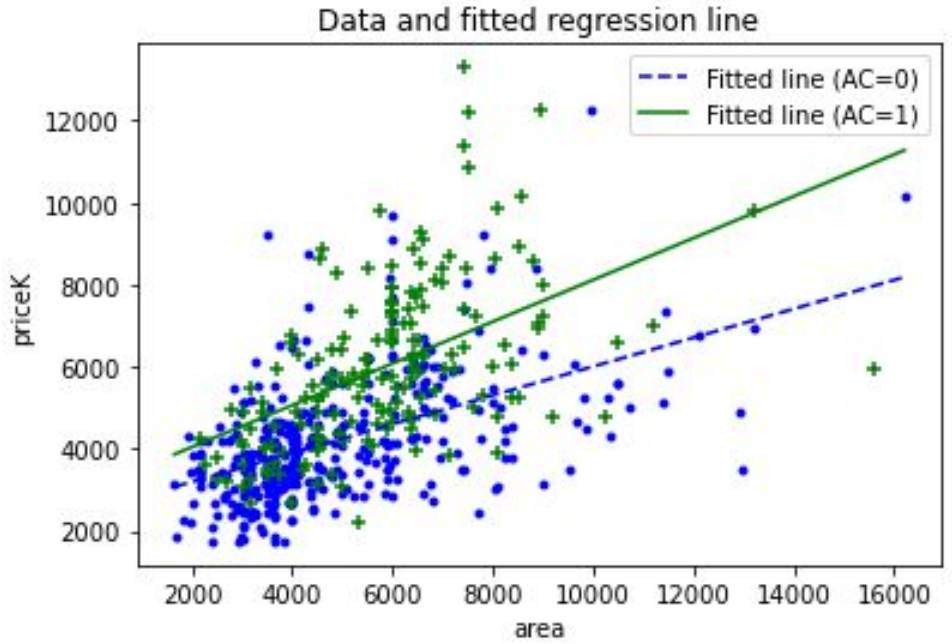


Hasil visualisasi regression line pada kolom price, area, dan airconditioning.

© 2022 – Pacmann AI

Regression Model

2. Prediksi harga berdasarkan luas area dan apakah ada AC atau tidak. (Linear Regression dengan multiple prediktor.)



Hasil visualisasi regression line pada kolom price, area, dan airconditioning. Setelah dilakukan Iteration

© 2022 – Pacmann AI

Regression Model

2. Prediksi harga berdasarkan luas area dan apakah ada AC atau tidak. (Linear Regression dengan multiple prediktor.)

	coef	std err
Intercept	2497.870807	181.269487
area	0.351240	0.034259
airconditioning	530.746328	388.928791
airconditioning:area	0.158025	0.065347

$$\text{price} = 2497 + 0.35\ \text{area} + 530\ \text{airconditioning} + 0.16\ \text{aircondotioning*area}$$

$$\text{AC} = 0,\ \text{price} = 2497 + 0.35\ \text{area}$$

- Estimasi perbedaan harga rumah yang tidak ada AC tetapi berbeda satu satuan luas area adalah 0.35 ribu dollar atau 350 dollar

$$\text{AC} = 1,\ \text{price} = 3027 + 0.51\ \text{area}$$

- Estimasi perbedaan harga rumah yang ada AC tetapi berbeda satu satuan luas area adalah 0.51 ribu dollar atau 510 dollar

© 2022 – Pacmann AI

Regression Model

3. Prediksi harga berdasarkan area dan furnishing status Linear Regression dengan melakukan transformasi pada data.

	coef	std err
Intercept	3050.661639	216.999265
C(furnishingstatus)[T.semi-furnished]	-363.892474	165.034605
C(furnishingstatus)[T.unfurnished]	-1060.393894	175.263337
area	0.429851	0.030653

Pada model ini, **furnishing status = furnished** menjadi baseline

- **Intercept 3050** (ribu) adalah harga rata-rata properti dengan **furnished** dan memiliki area = 0 (not a meaningfull scenario)
- Koefisien **semi-furnished**, prediksi perbedaan harga **furnished** dan **semi-furnished** yang memiliki area = 0, adalah **-363 ribu** dollar (lebih murah 363 ribu)
- Koefisien **unfurnished,** prediksi perbedaan harga **furnished** dan **unfurnished** yang memiliki area = 0, adalah **-1060 ribu** dollar (lebih murah 1060 ribu)

© 2022 – Pacmann AI

Regression Model

3. Prediksi harga berdasarkan area dan furnishing status Linear Regression dengan melakukan transformasi pada data.

	coef	std err
Intercept	5264.625954	130.229825
C(furnishingstatus)[T.semi-furnished]	-363.892474	165.034605
C(furnishingstatus)[T.unfurnished]	-1060.393894	175.263337
z_area	932.836864	66.521237

Hasil setelah dilakukan standarisasi pada kolom area. koefisien pada z-area berubah, sedangkan pada furnishing status masih sama.

Interpretasi
- Intercept 5264 (ribu) adalah harga furnishing status = furnished dan memiliki area = rata-rata.
- Koefisien z_area, presiksi perbedaan harga furnishing status = furnished dengan perbedaan 1 standar deviasi adalah 932 (ribu).

© 2022 – Pacmann AI

Conclusion

- **Prediksi harga berdasarkan area.**
  Harga berkorelasi positif dengan area. Semakin tinggi (luas) area, harganya juga mengalami kenaikan.

- **Prediksi harga berdasarkan area dan apakah ada AC atau tidak.**
  Harga berkorelasi positif dengan area dan adanya fasilitas AC. Jika ada AC maka harganya akan semakin naik.

- **Prediksi harga berdasarkan area dan furnishing status**
  Harga berkorelasi positif dengan area dan furnitured. Rumah yang masih semi-furnitured dan unfurnitured harganya akan semakin murah.

© 2022 – Pacmann AI

Recommendation

- Pembeli bisa memprediksi harga rumah berdasarkan area, airconditioning, dan furnishing status. Agar bisa membeli properti dengan harga yang sesuai.
- Penjual juga bisa menentukan harga yang adil bagi penjual dan pembeli agar sama-sama untung.
- Melakukan penelitian lebih lanjut untuk melihat hubungan harga dengan fitur-fitur lainnya.

© 2022 – Pacmann AI

Reference

- Statistic for Business—Pacmann.io
- Housing Prices Dataset—Kaggle

© 2022 – Pacmann AI

Thank You

© 2022 – Pacmann AI