



SAMBOIS

PROPERTY SALES OF MELBOURNE

OUR TEAM



Aidil Fahmi Zuhdi



Harun Jeynaro



Sahrur Ramadhan



CONTENT

01

Latar Belakang

02

Tujuan Analisis

03

Tentang Dataset



SAMBOIS

LATAR BELAKANG

Melbourne City merupakan salah satu kawasan paling dinamis di Australia, dengan **pertumbuhan penduduk, aktivitas ekonomi, dan gaya hidup urban yang terus berkembang**. Sebagai pusat bisnis, pendidikan, dan budaya, wilayah ini menarik banyak penduduk lokal maupun internasional untuk tinggal, bekerja, maupun berinvestasi.

Pasar properti di Melbourne City dikenal memiliki variasi harga yang cukup tajam, tergantung pada lokasi (suburb), kedekatan dengan pusat kota (CBD), akses transportasi, hingga fasilitas publik seperti sekolah, rumah sakit, dan taman kota. Perbedaan ini membuat analisis data properti menjadi penting untuk memahami tren pasar, sekaligus memberikan gambaran bagaimana masyarakat berinteraksi dengan ruang perkotaan.

Dengan adanya data penjualan properti, kita dapat menelusuri:

- Bagaimana harga rumah dan unit tersebar di area Melbourne City.
- Faktor apa saja yang paling berpengaruh terhadap harga jual di kawasan metropolitan ini.
- Pola distribusi harga berdasarkan suburb dan karakteristik properti.

TUJUAN

Proyek ini berfokus pada analisis data penjualan properti di Melbourne City, dengan tujuan:

- 01** Mengidentifikasi tren harga properti di pusat kota dan suburb sekitarnya.
- 02** Menemukan hubungan antara karakteristik properti (jumlah kamar, luas tanah, tipe bangunan) dengan harga jual.
- 03** Memberikan wawasan yang relevan bagi pembeli, penjual, investor, maupun pengembang properti di Melbourne.

DATASET

Dataset yang digunakan berfokus pada penjualan properti di Melbourne City, meliputi informasi seperti:

01 Harga jual properti (Price)

02 Suburb di kawasan Melbourne City

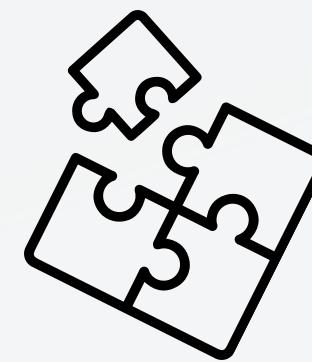
03 Tanggal penjualan serta agen properti

04 Jenis properti (House, Unit, Townhouse)

05 Jumlah kamar tidur, kamar mandi, dan garasi

06 Tanggal penjualan serta agen properti

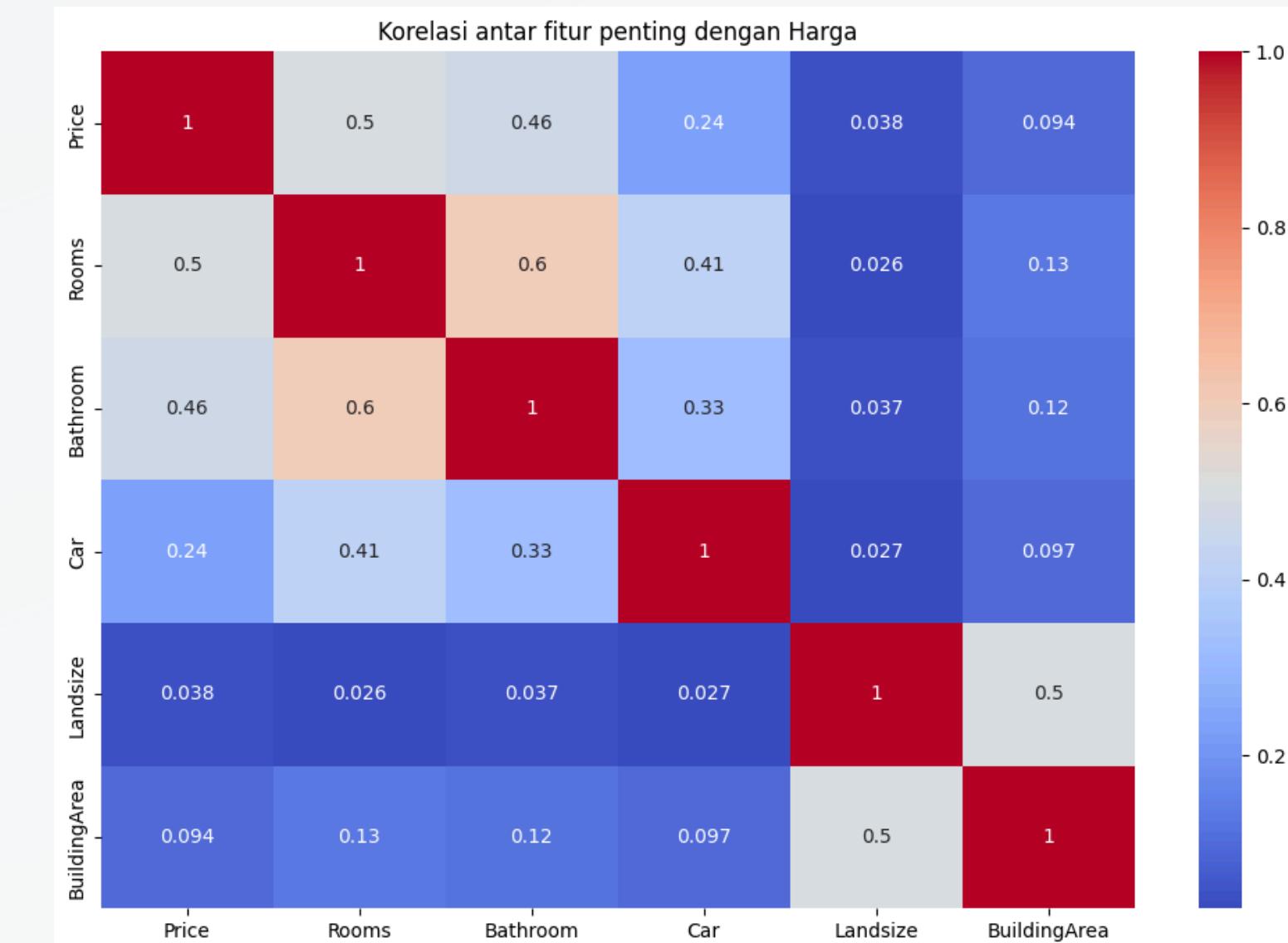
KORELASI ANTAR FITUR



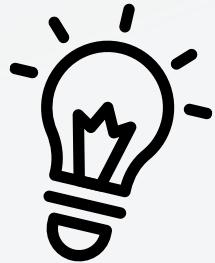
Berdasarkan hasil analisis korelasi, dapat kita lihat bahwa harga rumah paling banyak dipengaruhi oleh **jumlah kamar (Rooms)** dan **jumlah kamar mandi (Bathroom)**, dengan nilai korelasi masing-masing sebesar **0.50** dan **0.46**. Artinya, **semakin banyak jumlah kamar maupun kamar mandi, maka harga rumah cenderung lebih tinggi**. Faktor lain seperti **jumlah garasi (Car)** memiliki **pengaruh yang lebih kecil** dengan korelasi 0.24, sehingga tidak terlalu menentukan harga.

Sementara itu, **luas tanah (Landsize)** dan **luas bangunan (BuildingArea)** hanya menunjukkan **korelasi yang sangat rendah terhadap harga**, yaitu di bawah 0.10. Hal ini mengindikasikan bahwa **ukuran tanah atau bangunan tidak secara langsung meningkatkan harga rumah**, kemungkinan besar karena faktor lokasi lebih berperan dalam menentukan nilai properti.

Selain itu, kita juga bisa melihat adanya keterkaitan antarfitur, misalnya **rumah dengan jumlah kamar yang lebih banyak** biasanya juga **memiliki lebih banyak kamar mandi dan garasi**. Dengan demikian, dapat disimpulkan bahwa **jumlah kamar dan kamar mandi adalah faktor utama yang memengaruhi harga rumah** berdasarkan analisis korelasi ini.



MODEL YANG DIGUNAKAN



Random Forest dan XGBoost dipilih karena keduanya termasuk **model berbasis pohon keputusan** yang efektif untuk menangani data dengan hubungan non-linear. **Random Forest bekerja dengan membangun banyak pohon keputusan secara acak, kemudian menggabungkan hasilnya untuk menghasilkan prediksi yang stabil dan lebih tahan terhadap overfitting.**

Pendekatan ini membuat Random Forest mampu menangkap pola yang kompleks dan memberikan hasil yang konsisten.

Sementara itu, **XGBoost** menggunakan teknik boosting, yaitu membangun pohon secara bertahap untuk memperbaiki kesalahan dari pohon sebelumnya. Hal ini membuat **XGBoost lebih akurat, efisien, dan mampu memberikan hasil prediksi yang lebih baik pada data besar**. Dengan kombinasi keduanya, analisis prediksi harga rumah menjadi lebih kuat sekaligus memberikan wawasan tentang faktor-faktor apa saja yang paling memengaruhi nilai properti.

```
from sklearn.ensemble import RandomForestRegressor

rf_model = RandomForestRegressor(
    n_estimators=200,      # jumlah tree
    max_depth=None,
    random_state=42
)

# Latih Random Forest
rf_model.fit(X_train, y_train)
```

RandomForestRegressor
RandomForestRegressor(n_estimators=200, random_state=42)

```
from xgboost import XGBRegressor

xgb_model = XGBRegressor(
    n_estimators=300,      # jumlah boosting round
    learning_rate=0.1,     # kecepatan belajar
    max_depth=6,           # kedalaman tree
    random_state=42,
    n_jobs=-1
)

# Latih XGBoost
xgb_model.fit(X_train, y_train)
```

XGBRegressor
XGBRegressor(base_score=None, booster=None, callbacks=None,
 colsample_bylevel=None, colsample_bynode=None,
 colsample_bytree=None, device=None, early_stopping_rounds=None,
 enable_categorical=False, eval_metric=None, feature_types=None,
 feature_weights=None, gamma=None, grow_policy=None,
 importance_type=None, interaction_constraints=None,
 learning_rate=0.1, max_bin=None, max_cat_threshold=None,
 max_cat_to_onehot=None, max_delta_step=None, max_depth=6,
 max_leaves=None, min_child_weight=None, missing=nan,
 monotone_constraints=None, multi_strategy=None, n_estimators=300,
 n_jobs=-1, num_parallel_tree=None, ...)



EVALUASI MODEL



Evaluasi model dilakukan menggunakan tiga metrik utama, yaitu **MAE (Mean Absolute Error)**, **RMSE (Root Mean Squared Error)**, dan **R² (R-squared)**. MAE digunakan untuk **mengukur rata-rata selisih absolut antara nilai prediksi dengan nilai aktual**, sehingga memberikan gambaran seberapa jauh prediksi model dari kenyataan dalam satuan harga. RMSE juga mengukur **perbedaan antara prediksi dan nilai aktual**, tetapi memberikan penalti lebih besar pada kesalahan yang besar karena menggunakan kuadrat error sebelum diakarkan. Sementara itu, R² mengukur **seberapa baik variabel independen** dalam menjelaskan variasi pada harga rumah, dengan nilai mendekati 1 menunjukkan model mampu menjelaskan data dengan baik.

Hasil evaluasi menunjukkan bahwa:

- Random Forest memiliki,
 - MAE = 0.27, RMSE = 0.46, dan R² = 0.7795,
 - artinya model ini cukup baik dengan rata-rata kesalahan kecil dan mampu menjelaskan sekitar 78% variasi data.
- XGBoost juga memiliki
 - MAE = 0.27, RMSE = 0.45 dan R² = 0.7913.
 - Hal ini menandakan bahwa XGBoost sedikit lebih akurat dan lebih baik dalam menangkap pola data dibanding Random Forest.

==== Random Forest ===

MAE : 0.27

RMSE : 0.46

R² : 0.7795

==== XGBoost ===

MAE : 0.27

RMSE : 0.45

R² : 0.7913



SAMBOIS

REKOMENDASI BISNIS

01

Untuk Investor :

- Fokus investasi di suburb yang sedang berkembang di sekitar CBD dengan harga menengah (500K–1M AUD), karena punya potensi pertumbuhan lebih tinggi dibanding suburb elit yang sudah jenuh.
- Investor bisa membeli di pertengahan tahun dan melepas properti di akhir tahun untuk memaksimalkan margin capital gain.

02

Untuk Agen Properti

- Tonjolkan jumlah kamar tidur dan kamar mandi dalam promosi, karena itu faktor utama pembeli.
- Segmentasi pasar bisa difokuskan:
 - Keluarga → properti dengan 3–4 kamar tidur.
 - Pasangan muda → properti dekat CBD dengan akses transportasi meskipun ukuran tanah kecil

03

Untuk Pengembang (Developer)

- Tidak perlu fokus pada tanah yang luas, tetapi optimalisasi tata ruang (lebih banyak kamar dalam bangunan kompak) akan lebih menarik pasar.
- Pertimbangkan pembangunan unit apartemen dengan banyak kamar tidur di suburb sekitar CBD, karena demand-nya tinggi.



TERIMA KASIH

