ANALISIS CLUSTERING HOTEL DI BALI BERDASARKAN HARGA, RATING, JUMLAH ULASAN, DAN FASILITAS

Presented By Kelompok F - SD A1

OUR TEAM

- AQILA HANA WINANGGORO - 164221010
- HERMAN EFFENDI - 164221056
- TIKA DIAN PANGASTUTI - 164221061
- ARKAN SYAFIQ AT'TAQY - 164221062
- NISRINA KHAIRUNISA 164221097

LATAR BELAKANG

Pulau Bali sebagai destinasi utama dengan wisatawan internasional

Pertumbuhan pariwisata meningkatkan permintaan akomodasi berkualitas

Kebutuhan Informasi terperinci tentang akomodasi khususnya hotel masih terbatas

ALASAN

UTAMA

Untuk menjawab tantangan dari ketersedian informasi akomodasi yang terbatas

SAMPINGAN

Kemudahan dalam memperoleh data sehingga dapat mempersingkat waktu pengerjaan projek.

RUMUSAN MASALAH

Bagaimana cara memetakan dan menganalisis klaster hotel di Bali berdasarkan harga, rating, review, dan fasilitas?

Bagaimana pengaruh variabel-variabel tersebut terhadap klaster hotel di Bali?

Apa saja karakteristik masing-masing klaster hotel yang dihasilkan dari analisis tersebut?

TUJUAN

01

Menyelidiki pola-pola klaster hotel di Bali berdasarkan variabel harga, rating, review, dan fasilitas.

02

Mengidentifikasi perbedaan karakteristik antara klaster hotel yang dihasilkan.

03

Menyediakan informasi penting kepada wisatawan mengenai karakteristik hotel yang ada.

DATA

Dataset yang digunakan berasal dari hasil scraping pada website Traveloka (https://www.traveloka.com/), untuk mengambil informasi tentang hotel-hotel di Bali diambil pada tanggal 25 Mei 2024 hingga 26 Mei 2024.

	nama	harga	reviews	rating	facilities
0	ASTON Kuta Hotel & Residence	Rp 734.194	(2.4K reviews)	8.6 (2.4K reviews)	Hot tub\nKids club\nMassage\nFitness center\nP...
1	Brits Hotel Legian	Rp 422.004	(2.1K reviews)	8.6 (2.1K reviews)	Top picked by Family with children\nPay at Hot...
2	Harper Kuta by ASTON	Rp 650.826	(2.9K reviews)	8.6 (2.9K reviews)	Pay at Hotel\nMassage\nAirport transfer\nSauna...
3	Atanaya Kuta Bali	Rp 670.000	(10K reviews)	8.6 (10K reviews)	Babysitting\nPoolside bar\nWheelchair accessib...
4	Wyndham Garden Kuta Beach Bali	Rp 1.167.769	(1.4K reviews)	8.1 (1.4K reviews)	Pay at Hotel\nBabysitting\nBicycle rental\nCar...
...	...	...	...	...	...
259	Hotel Karthi	Rp 289.256	(693 reviews)	8.4 (693 reviews)	Massage
260	Bali Bungalo	Rp 590.909	(58 reviews)	8.1 (58 reviews)	Bicycle storage\nClothes dryer\nBicycle rental...
261	Cattleya Pool Suite - Seminyak by Marbella	Rp 910.301	(219 reviews)	8.2 (219 reviews)	Pay at Hotel\nBanana boat\nBeach volleyball\nB...
262	The Akasha Luxury Villas	Rp 3.910.050	(37 reviews)	8.7 (37 reviews)	Surfing\nHorse riding\nMassage\nKitchenette\nE...
263	Interconnection Hostel Kuta	Rp 460.542	(70 reviews)	8.8 (70 reviews)	Car rental

264 rows × 5 columns

METODOLOGI

1. Data Prepocessing

Pengolahan data dimulai dengan preprocessing yang mencakup beberapa langkah penting untuk memastikan data siap digunakan dalam analisis lanjutan. Pertama, penanganan missing value dilakukan menggunakan metode .isna() dari Pandas untuk mengidentifikasi dan mengatasi data yang hilang. Selanjutnya, label encoding diterapkan untuk mengubah nilai kategorik pada variabel "facilities" menjadi numerik, sehingga dapat dianalisis secara statistik. Penanganan outlier dilakukan pada variabel "harga" dan "reviews" menggunakan Robust Scaler, yang memanfaatkan median dan interquartile range (IQR) untuk mengurangi dampak nilai ekstrem. Terakhir, feature scaling diterapkan menggunakan Standard Scaler dan Min-Max Scaler pada variabel "rating" dan "total_facilities" untuk memastikan bahwa semua variabel memiliki skala yang seimbang, sehingga tidak ada variabel yang mendominasi dalam analisis.

METODOLOGI

2. Exploratory Data Analysis (EDA)

Setelah preprocessing, langkah selanjutnya adalah Exploratory Data Analysis (EDA) untuk mendapatkan pemahaman yang lebih mendalam tentang data. Dalam tahap ini, korelasi heatmap dibuat untuk memvisualisasikan hubungan antara variabel numerik, membantu dalam mengidentifikasi pasangan variabel yang memiliki korelasi kuat atau lemah. Selain itu, boxplot digunakan untuk mendeteksi outlier dan memvisualisasikan distribusi data secara grafis, memberikan gambaran yang jelas tentang penyebaran nilai dalam dataset. Distribusi plot dibuat untuk melihat frekuensi nilai dalam setiap variabel, sementara pair plot membantu mengidentifikasi pola hubungan antar variabel dengan menampilkan scatter plot untuk setiap pasangan variabel, diberi pewarnaan berdasarkan jumlah total fasilitas.

METODOLOGI

3. PRINCIPAL COMPONENT ANALYSIS (PCA)

Tahap selanjutnya adalah penerapan Principal Component Analysis (PCA) untuk melakukan reduksi dimensi data. PCA membantu dalam memilih komponen utama yang signifikan, yang dapat menjelaskan setidaknya 95% dari total variansi data. Dengan mengurangi jumlah dimensi, PCA membantu dalam mengurangi kompleksitas data sambil mempertahankan informasi yang paling penting. Hasil dari penerapan PCA menghasilkan variabel baru, yaitu PC1 dan PC2, yang kemudian digunakan dalam analisis lebih lanjut untuk clustering dan visualisasi.

METODOLOGI

4. IMPLEMENTASI K-MEANS CLUSTERING

Implementasi K-Means Clustering dilakukan dengan beberapa tahap. Pertama, jumlah cluster optimal ditentukan menggunakan metode elbow, yang melibatkan plot jumlah cluster versus sum of squared distance (inertia) dan mencari nilai elbow dimana penurunan inertia mulai melambat. Evaluasi cluster dilakukan menggunakan silhouette score, yang mengukur seberapa baik cluster terbentuk dengan menghitung nilai silhouette untuk setiap sampel dan memvisualisasikannya dalam silhouette plot. Proses clustering menggunakan K-Means melibatkan variabel "harga", "rating", dan "total_facilities", berdasarkan heatmap korelasi yang menunjukkan hubungan yang kuat. Hasil clustering divisualisasikan dalam plot 2D untuk membantu memahami struktur dan distribusi cluster dalam data.

METODOLOGI

5. IMPLEMENTASI DBSCAN

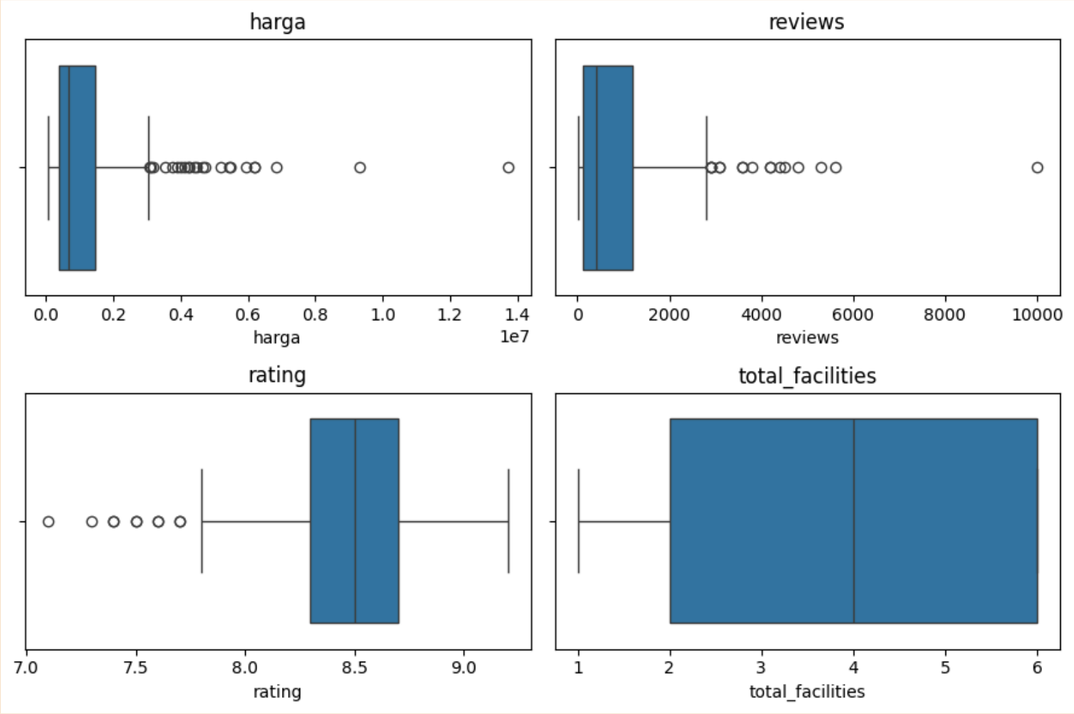
Selain K-Means, implementasi DBSCAN juga dilakukan untuk mengelompokkan data berdasarkan kepadatan. Penentuan nilai eps optimal dilakukan dengan k-distance graph menggunakan algoritma Kneedle untuk menemukan elbow point pada grafik jarak k-terdekat. DBSCAN kemudian mengelompokkan titik data berdasarkan kepadatan, dengan titik yang tidak cukup dekat dengan titik lain dan tidak memenuhi jumlah minimum sampel dianggap sebagai noise. Jumlah cluster ditentukan dari label unik yang dihasilkan oleh DBSCAN, dikurangi satu jika terdapat label -1 yang menunjukkan noise. Implementasi DBSCAN membantu dalam mengidentifikasi pola dan anomali dalam data, memberikan wawasan yang lebih mendalam tentang berbagai aspek hotel di Bali, termasuk harga, popularitas, dan fasilitas yang ditawarkan.

STATISTIK DESKRIPTIF

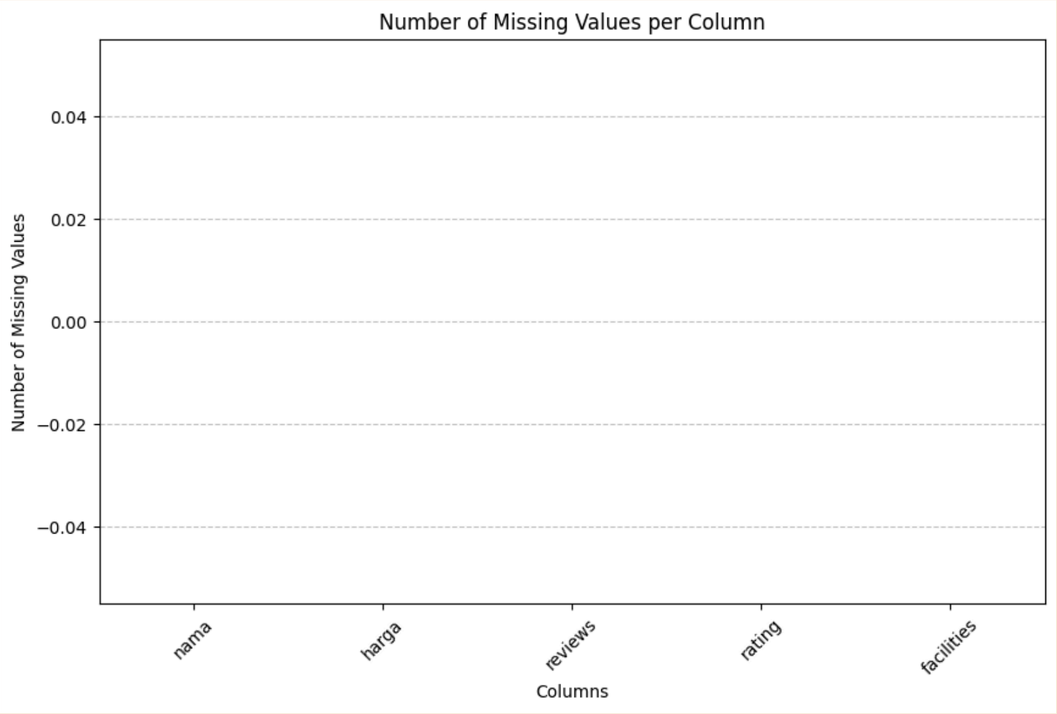
	Nama Hotel	Harga	Reviews	Rating	Fasilitas
Jumlah	264	264	264	264	264
Nilai Unik	255	243	187	242	223
Terbanyak	Coast Boutique Hotel	Rp247.934	(1.5K reviews)	8.5 (54 reviews)	Pay at Hotel
Frekuensi	2	5	7	2	8

EDA

CHECK OUTLIER

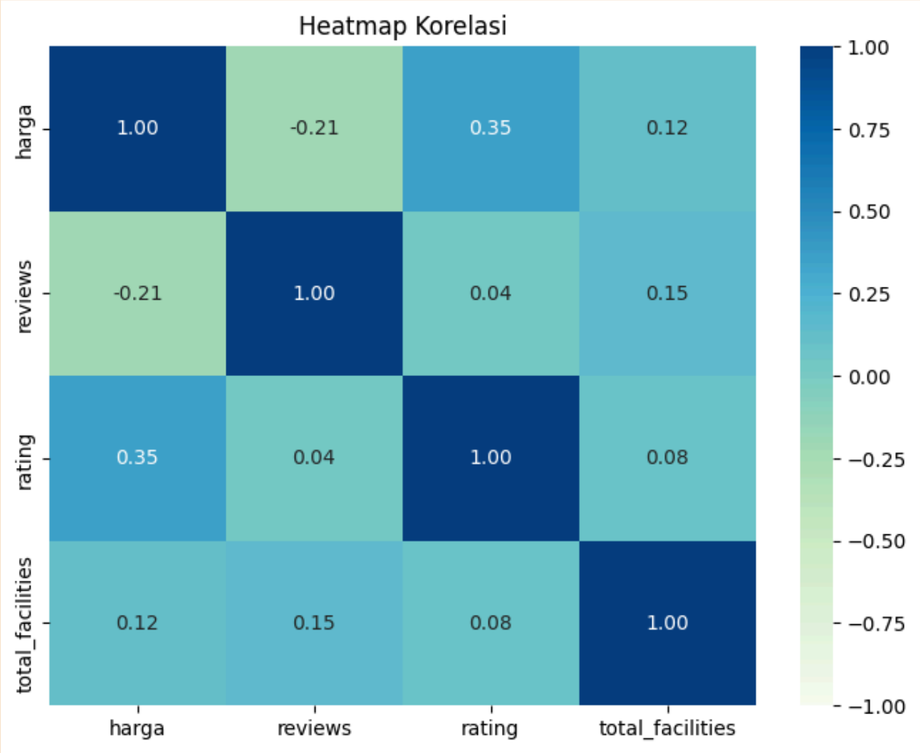


CHECK MISSING VALUE

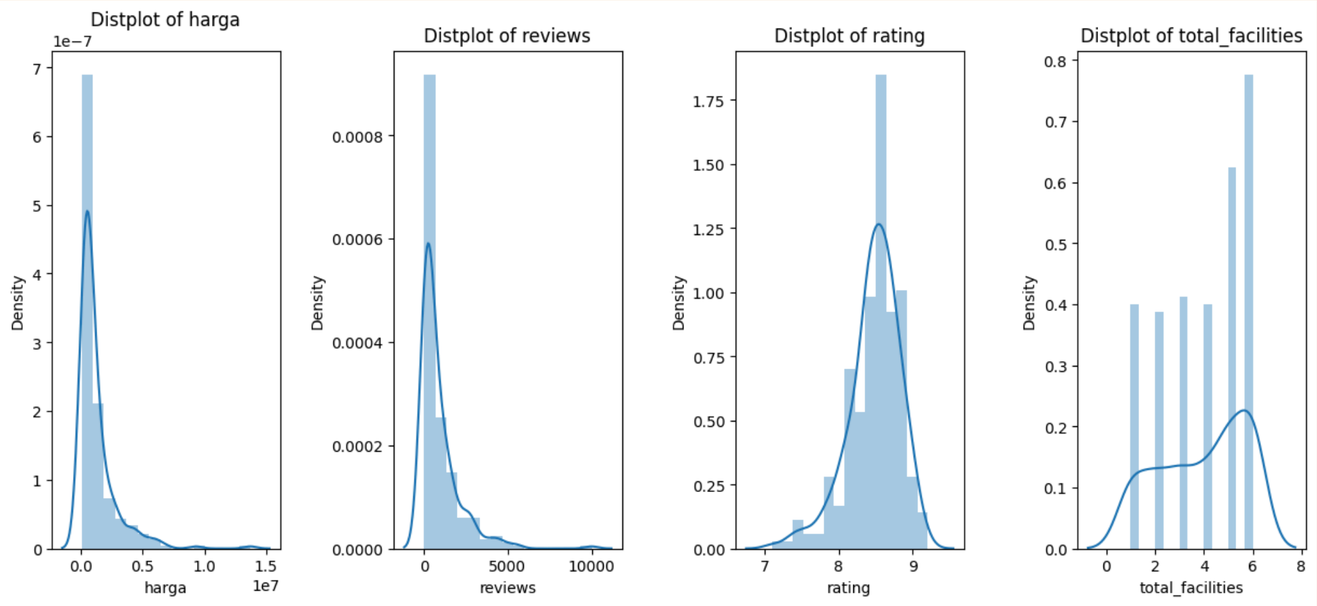


EDA

HEATMAP CORRELATION



DISTRIBUTION PLOT



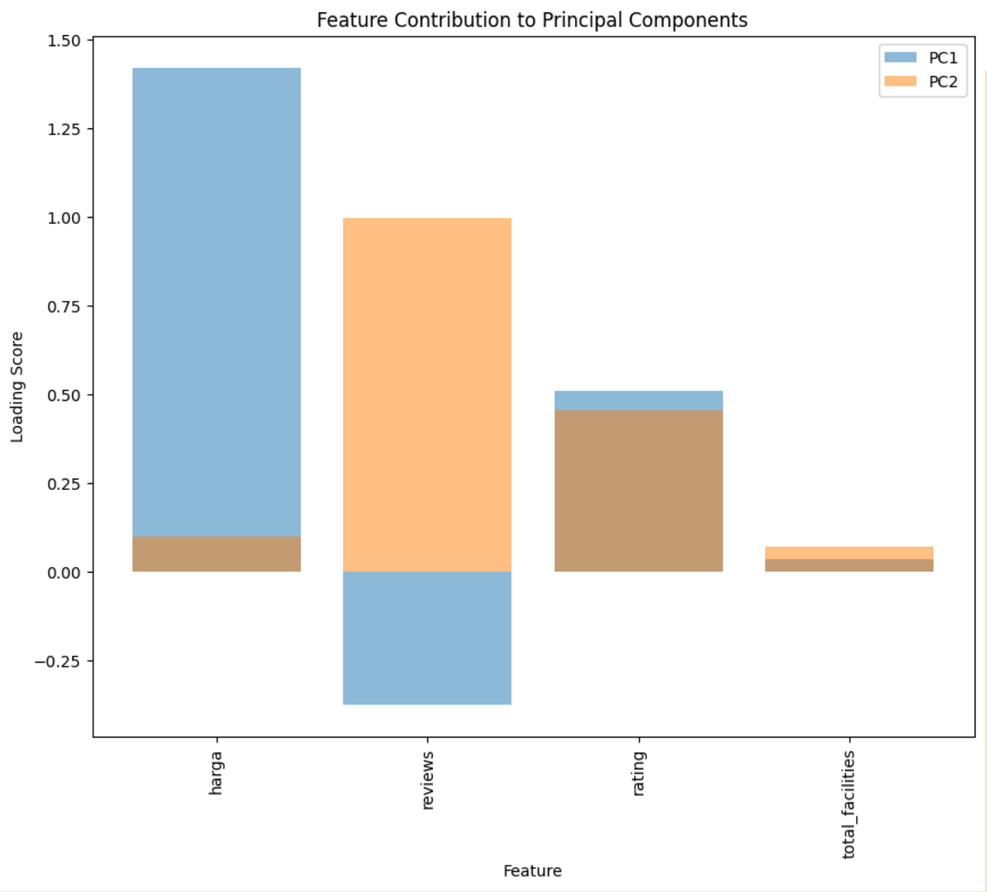
FEATURE SCALING

- Melakukan Robust Scaler untuk variabel yang memiliki outlier, yaitu harga dan reviews
- Melakukan Min-Max Scaler untuk variabel total fasilitas
- Melakukan Standard Scaler untuk variabel rating

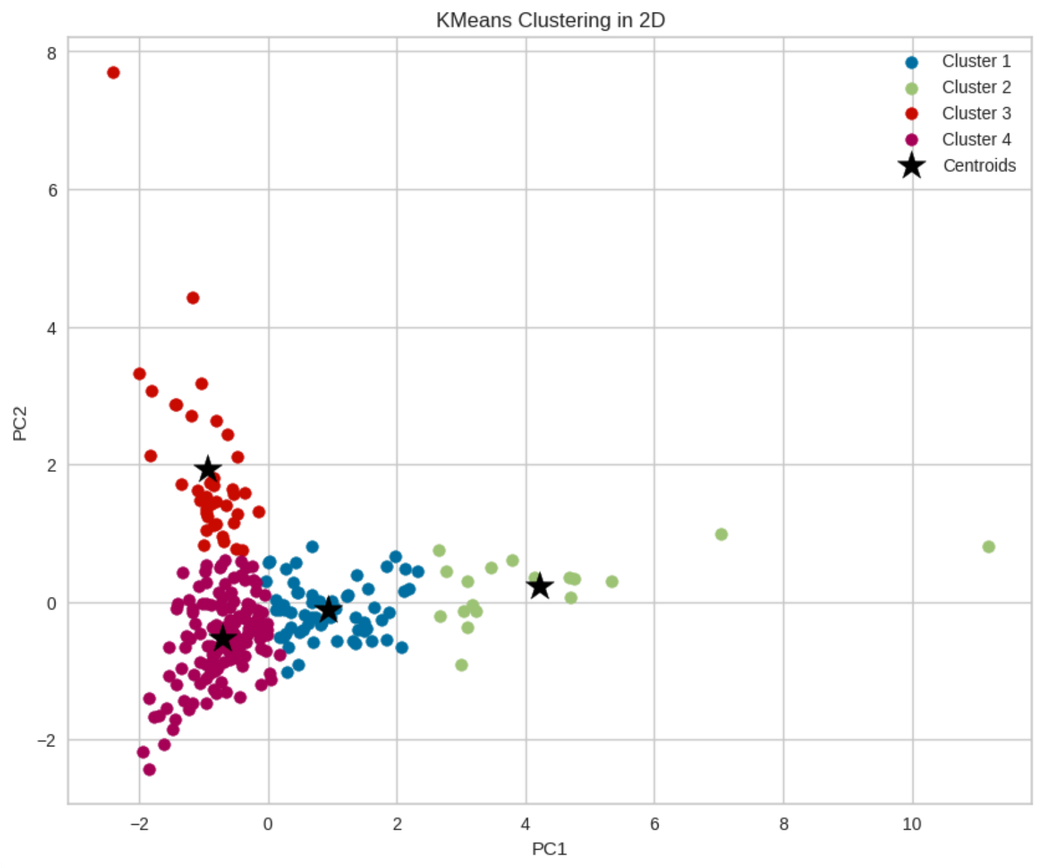
	harga	reviews	rating	total_facilities
0	0.065847	1.837263	0.345972	1.0
1	-0.224777	1.559871	0.345972	1.0
2	-0.011762	2.299584	0.345972	0.8
3	0.006087	8.864540	0.345972	0.4
4	0.469471	0.912621	-1.086218	1.0

PCA

Melakukan PCA untuk mereduksi dimensi menjadi 2 dimensi. Selain itu juga dapat memperoleh indeks ringkasan yang mewakili keragaman data



K-MEANS

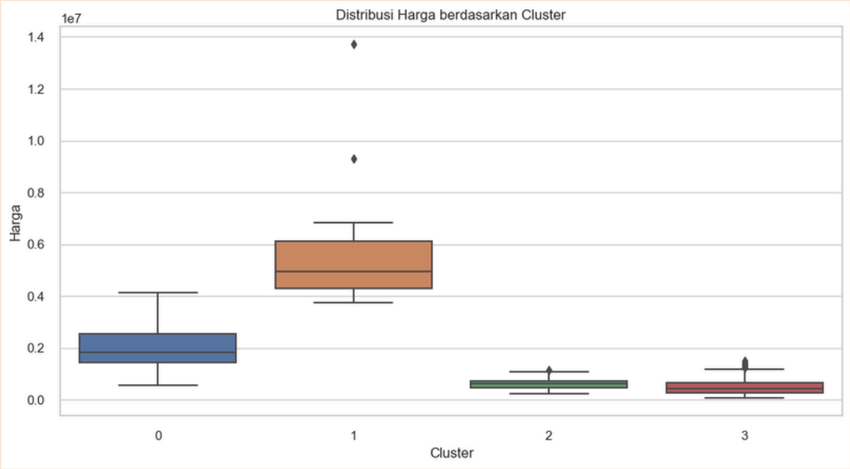


Dihasilkan cluster sejumlah 4, sesuai dengan yang telah ditentukan untuk metode K-Means. Terdapat titik pusat atau centroid yang ditandai dengan bintang hitam. Centroid ini menunjukan pusat dari masing-masing klaster dan membantu dalam memahami karakteristik utama dari setiap klaster.

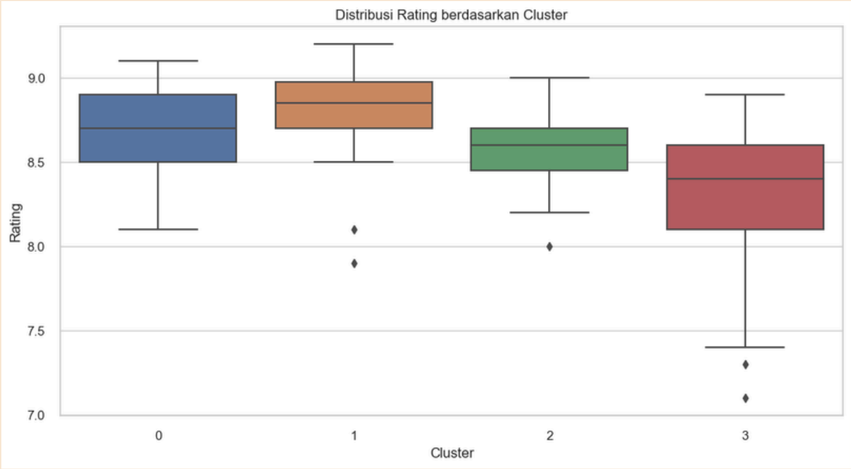
Dengan demikian, analisis clustering ini memberikan wawasan yang jelas mengenai segmentasi hotel di Pulau Bali berdasarkan harga, rating, jumlah review, dan fasilitas.

HASIL K-MEANS

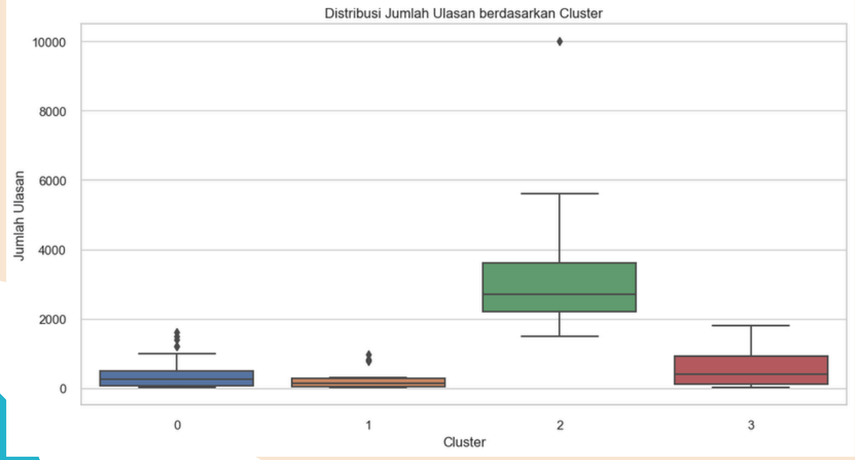
Harga



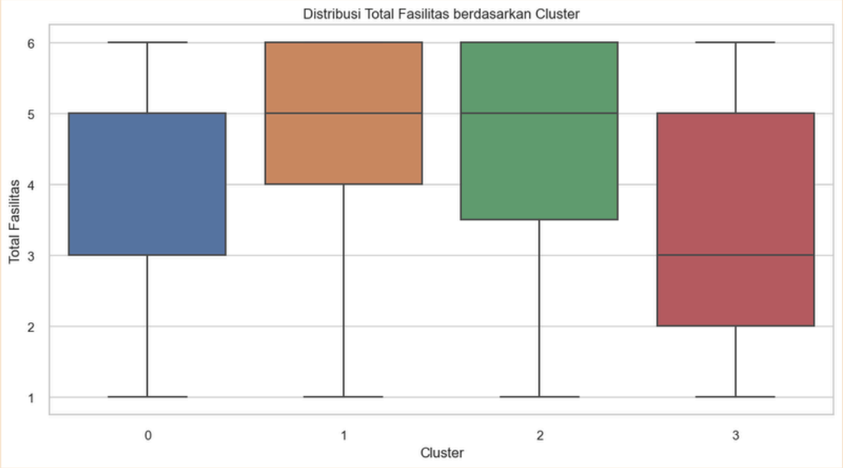
Rating



Jumlah Ulasan



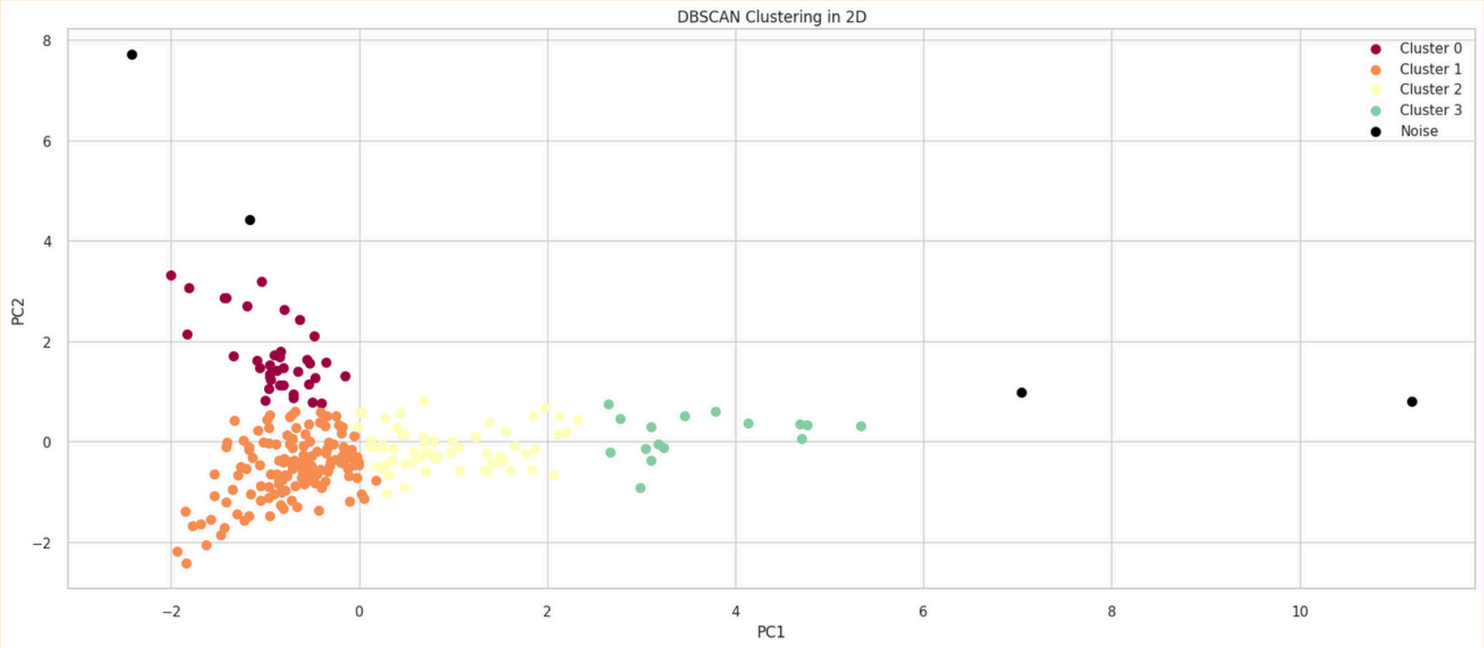
Total Fasilitas



HASIL K-MEANS

Klaster	Harga	Reviews	Rating	Total Fasilitas	Anggota
1	Rp1.976.745,26	369	8.69	4	61
2	Rp5.708.495,44	271	8.78	5	18
3	Rp614.669,00	3110	8.57	5	39
4	Rp513.285,24	559	8.32	4	137

DBSCAN

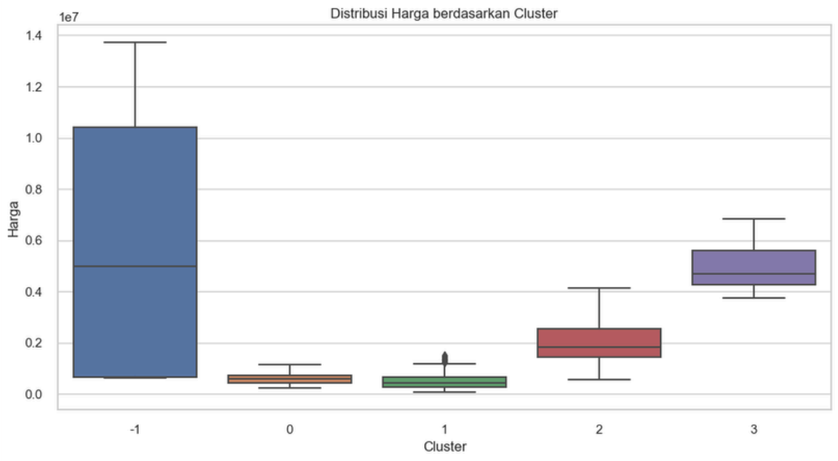


Didapatkan jumlah cluster yang dihasilkan dengan metode DBSCAN sebanyak 4 cluster. Terdapat juga titik hitam yang tidak masuk ke dalam 4 cluster manapun yang disebut titik noise berjumlah 4.

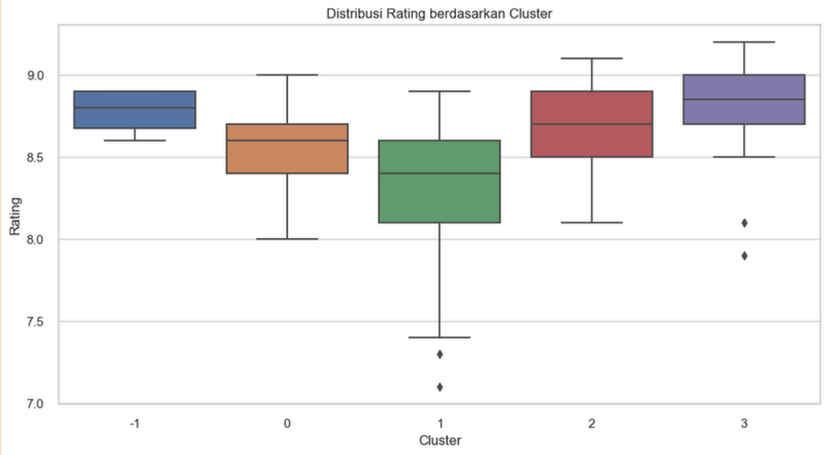
Secara keseluruhan, clustering dengan metode DBSCAN maupun K-Means menghasilkan cluster yang cukup mirip, yang membedakan hanya titik noise saja.

HASIL DBSCAN

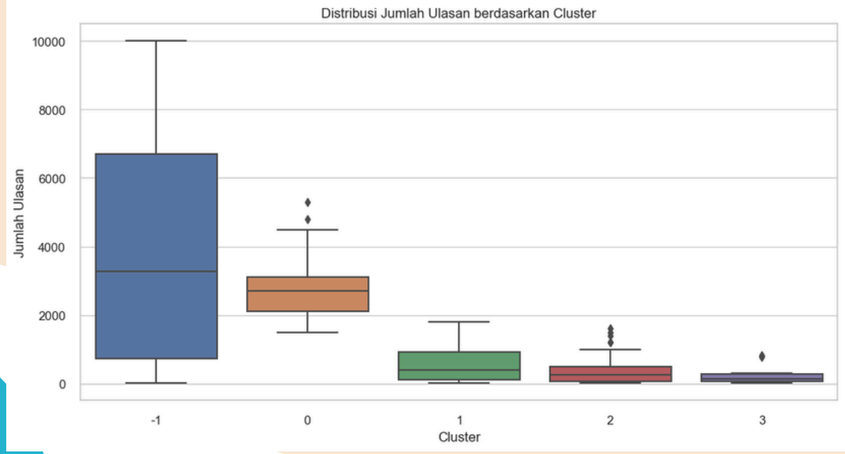
Harga



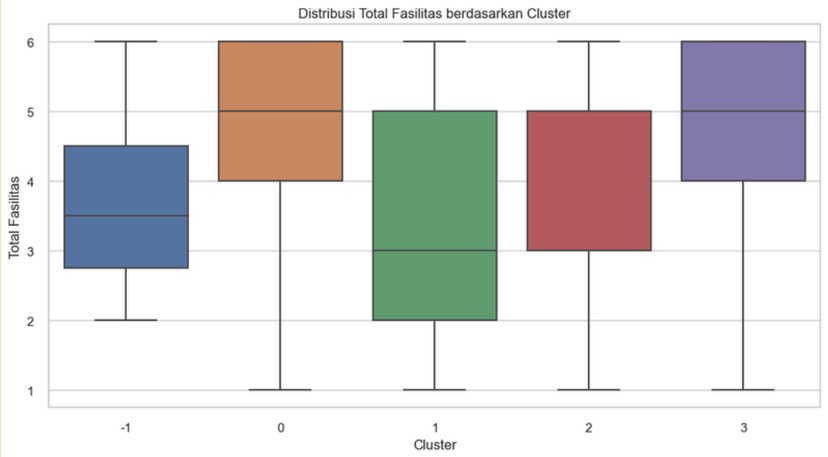
Rating



Jumlah Ulasan



Total Fasilitas



HASIL DBSCAN

Klaster	Harga	Reviews	Rating	Total Fasilitas	Anggota
0	Rp612.447,62	2857	8.56	5	37
1	Rp513.285,24	559	8.32	4	137
2	Rp1.976.745,26	369	8.69	4	61
3	Rp4.983.063,62	244	8.78	5	16

KESIMPULAN DAN SARAN

KESIMPULAN

K-Means

- Klaster 1: Hotel harga menengah dengan rating cukup baik, jumlah review moderat, dan fasilitas memadai
- Klaster 2: Hotel harga lebih tinggi dengan rating sangat baik, dan fasilitas lengkap
- Klaster 3: Hotel harga terjangkau dengan rating lebih rendah, jumlah review sedikit, dan fasilitas dasar
- Klaster 4: Hotel dengan harga, rating, jumlah review, dan fasilitas yang beragam

DBSCAN

- Klaster 0: Hotel harga terjangkau, rating beragam, jumlah reviewer sedang, dan fasilitas memadai
- Klaster 1: Hotel harga lebih tinggi dengan rating baik, dan fasilitas lebih lengkap
- Klaster 2: Hotel harga terjangkau dengan rating lebih rendah, jumlah review lebih sedikit, dan fasilitas dasar
- Klaster 3: Hotel dengan kombinasi harga, rating, jumlah review, dan fasilitas yang beragam namun memiliki kesamaan tertentu

KESIMPULAN DAN SARAN

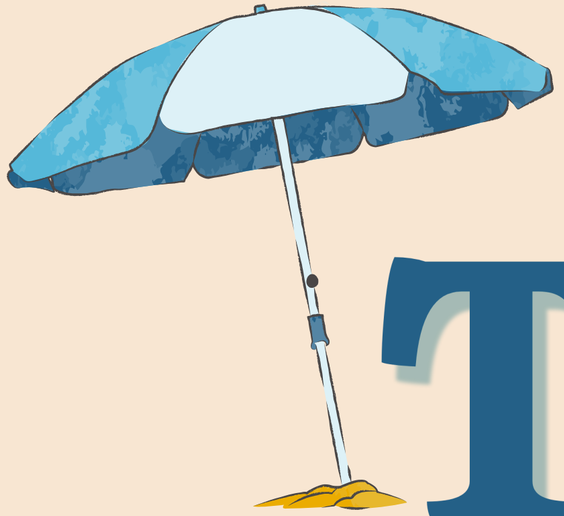
SARAN

Untuk Pemangku Kepentingan Pariwisata

- Dapat menggunakan informasi ini untuk merencanakan strategi pemasaran yang lebih efektif, menargetkan segmen pasar spesifik sesuai dengan karakteristik setiap klaster
- Dapat membantu dalam pengembangan infrastruktur dan promosi wisata yang lebih terfokus sesuai dengan prefrensi wisatawan

Untuk wisatawan

- Dapat menggunakan informasi ini untuk memilih hotel yang sesuai dengan prefrensi dan anggaran



THANK YOU