

Indonesia Tourism Destination





Outlines

Objectives

EDA

Model Choices

Best Model and Recommendation



Objectives



Problem Statement

- Bagaimana Karakteristik Wisatawan?
- Adanya asumsi bahwa bogor merupakan kota destinasi yang paling favorit karena selalu macet total ketika memasuki waktu liburan
- Adanya asumsi bahwa jumlah orang yang memberikan rating bisa mempengaruhi nilai rating itu sendiri, karena semakin sedikit orang yang memberikan rating akan membuat rating itu tampak tinggi.
- Tim sales marketing sering kewalahan dengan permintaan klien yang memakai jasa travel secara rombongan karena perlu ada penyesuaian dengan keinginan mereka yang kebanyakan akan pergi berwisata ke Yogyakarta.

Objective

- ★ Mengetahui karakteristik wisatawan
- ★ Adanya asumsi bahwa bogor merupakan kota destinasi yang paling favorit karena selalu macet total ketika memasuki waktu liburan
- ★ Menguji kebenaran asumsi bahwa semakin banyak orang yang memberi rating semakin rendah nilai rating
- ★ Membuat sistem rekomendasi tempat wisata di Yogyakarta sehingga Tim Customer Specialist mudah merekomendasikan tempat tempat yang serupa jika terjadi penolakan dalam penyusunan itinerary perjalanan wisata ke Yogyakarta

Data Overview

Terdapat 3 data diantaranya data klien (**user.csv**), tempat wisata yang dikunjungi klien (**tourism_with_id.csv**), dan nilai kesukaan klien terhadap tempat wisata yang telah dikunjungi (**tourism_rating.csv**)

Sample data (user.csv)

	User_Id	Location	Age
0	1	Semarang, Jawa Tengah	20

Sample data (tourism_rating.csv)

	User_Id	Place_Id	Place_Ratings
0	1	179	3

Sample data (tourism_rating.csv)

	Place_Id	Place_Name	Description	Category	City	Price	Rating	Time_Minutes	Coordinate	Lat	Long	Unnamed: 11	Unnamed: 12
0	1	Monumen Nasional	Monumen Nasional atau yang populer disingkat d...	Budaya	Jakarta	20000	4.6	15.0	{'lat': -6.1753924, 'lng': 106.8271528}	-6.175392	106.827153	NaN	1

Data Overview

Summary

- Time_Minutes pada tourism_df sebagian nilai kosong
- Unnamed: 11 pada tourism_df seluruhnya memiliki nilai kosong
- Unnamed: 12 pada tourism_df memiliki nilai yang sama dengan Place_id
- Location pada user_df berisikan Kota dan Provinsi sedangkan data pada tourism_df hanya berisikan kota
- Terdapat perbedaan penulisan kota “Jakarta” pada user_df dan tourism_df

Data Overview

Data setelah dilakukan proses Data Cleaning

Sample data (user.csv)

User_Id	Location	Age
0	1 Semarang	20

Sample data (tourism_rating.csv)

User_Id	Place_Id	Place_Ratings
0	1 179	3

Sample data (tourism_rating.csv)

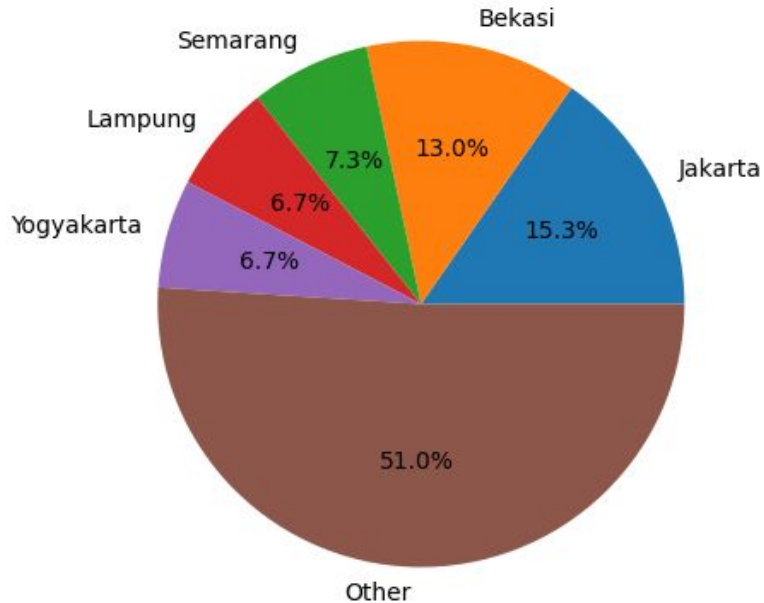
	Place_Id	Place_Name	Category	City	Price	Rating	Time_Minutes	Lat	Long	RatingCount
0	1	Monumen Nasional	Budaya	Jakarta	20000	4.6	15.0	-6.175392	106.827153	18
1	2	Kota Tua	Budaya	Jakarta	0	4.6	90.0	-6.137645	106.817125	25



EDA

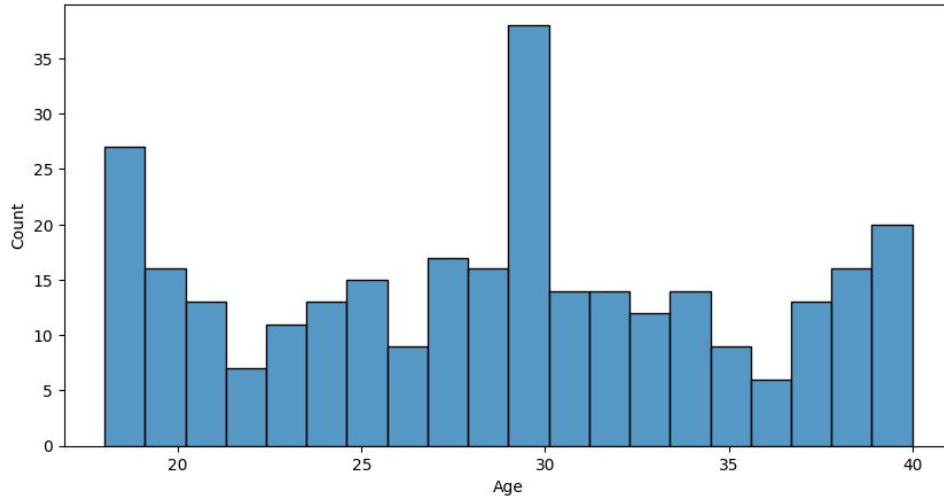


Data Insights - Through Exploration



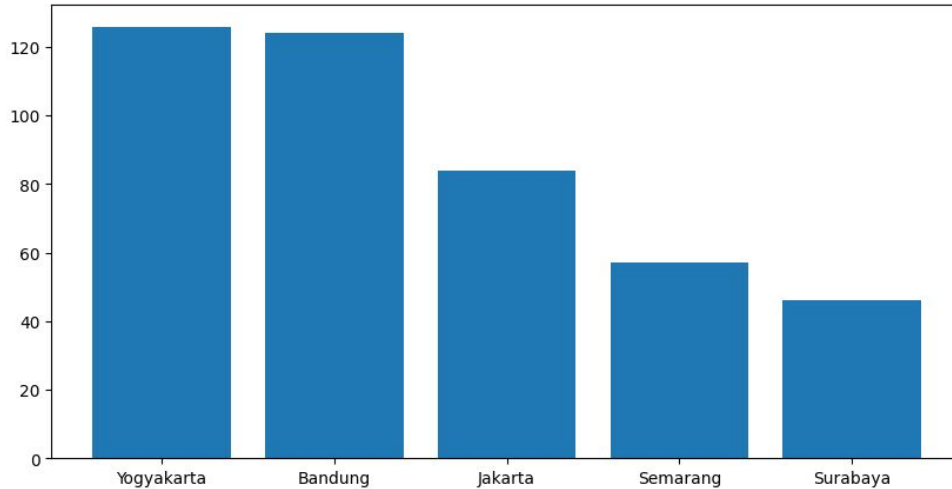
Berdasarkan Pie chart diatas terlihat bahwa klien agensi travel kita kebanyakan berasal dari Kota Jakarta dengan persentase 15.3% dan Bekasi dengan persentase 13% dibandingkan kota lainnya.

Data Insights - Through Exploration



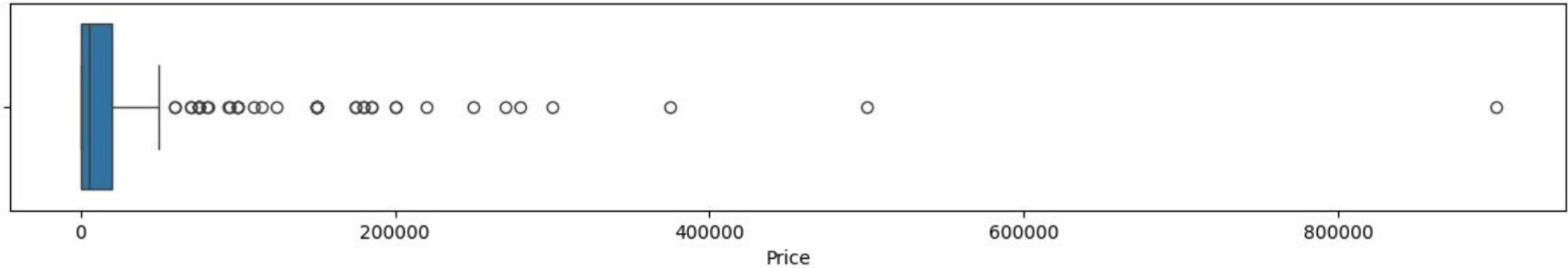
Berdasarkan histogram, dapat diketahui bahwa umur dari klien agensi kita berada pada rentang 6 tahun - 40 tahun dengan yang tertinggi berada di umur 30 tahunan. Lalu yang tertinggi kedua adalah umur klien agensi kita yang kurang dari 10 tahun. Hal ini bisa terjadi karena klien agensi kita yang berumur 30 tahunan selalu membawa anak-anak mereka selama berwisata.

Data Insights - Through Exploration



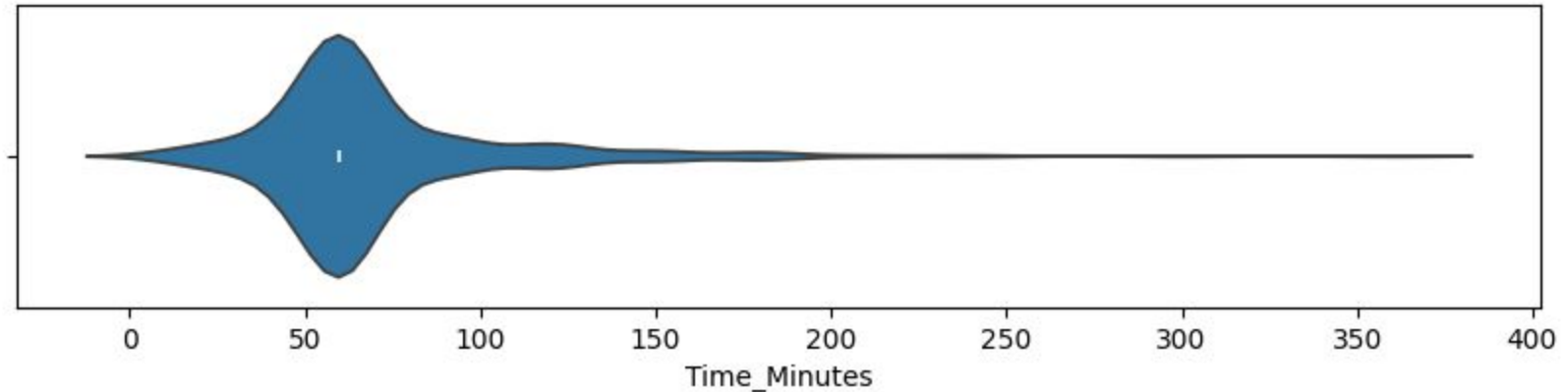
Berdasarkan bar chart terlihat bahwa objek wisata terbanyak berada di Yogyakarta dan Bandung, sehingga dapat diketahui bahwa asumsi tidak benar jika Bogor adalah kota yang paling disukai oleh klien agensi kita.

Data Insights - Through Exploration



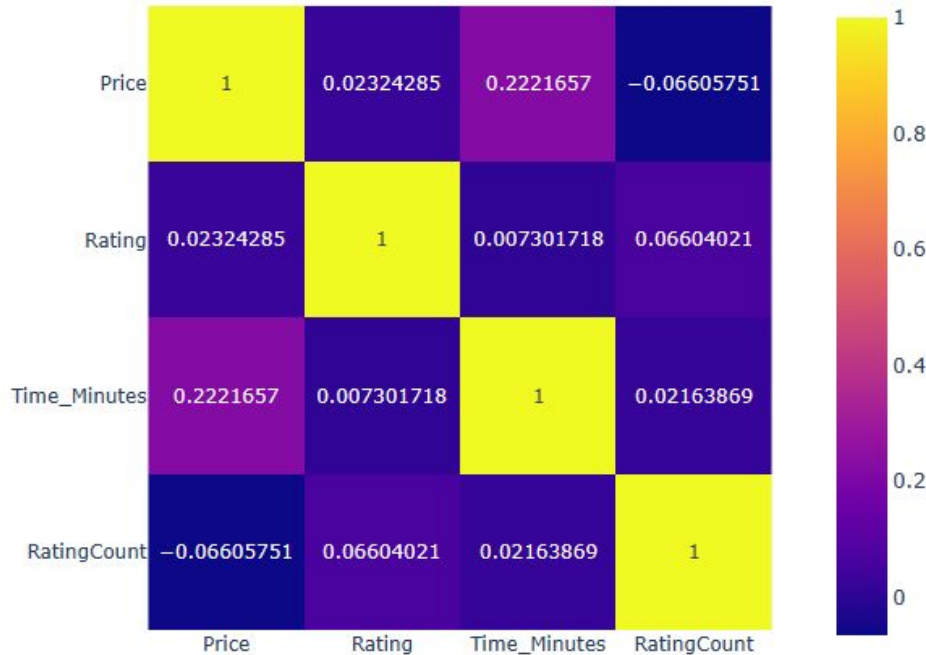
Berdasarkan box plot dan statistik deskriptif dapat diketahui bahwa kunjungan wisata yang disediakan agensi travel kita kebanyakan berkisar pada harga Rp 0 - Rp 5.000

Data Insights - Through Exploration



Berdasarkan violin plot dan statistik deskriptif dapat diketahui bahwa banyak tempat wisata yang disediakan oleh agensi travel kita yang membutuhkan waktu 60 menit untuk menempuhnya dari pusat kota.

Correlation



Berdasarkan heatmap, angka korelasi antara Rating dan RatingCount menunjukkan korelasi positif yang lemah antara keduanya. Artinya semakin tinggi jumlah rating suatu tempat wisata dengan kemungkinan yang rendah rating yang diperoleh juga akan naik



Model Choices



Model Description + Result

Sistem Rekomendasi dengan Content Based Filtering menggunakan algoritma K-NearestNeighbors untuk menghitung seberapa dekat antara satu item dengan item lainnya. Model yang dibangun dengan 3 neighbors ($k=3$) dan Euclidean distance metric.

	0	1	2	Place_Name
0	23	8	25	Taman Pintar Yogyakarta
1	27	39	32	Keraton Yogyakarta
2	31	79	41	Sindu Kusuma Edupark (SKE)
3	27	39	32	Museum Benteng Vredeburg Yogyakarta
4	18	13	19	De Mata Museum Jogja

Hasilnya adalah 3 Place_Id yang paling memiliki kemiripan dengan Place_Name yang diinputkan

Insight

	0	1	2	Place_Name
0	Embung Tambakboyoy	Gembira Loka Zoo	Monumen Serangan Umum 1 Maret	Taman Pintar Yogyakarta
1	Bukit Bintang Yogyakarta	Kauman Pakualaman Yogyakarta	The World Landmarks - Merapi Park Yogyakarta	Keraton Yogyakarta
2	Jurang Tembelan Kanigoro	Pintoe Langit Dahromo	Pantai Goa Cemara	Sindu Kusuma Edupark (SKE)
3	Bukit Bintang Yogyakarta	Kauman Pakualaman Yogyakarta	The World Landmarks - Merapi Park Yogyakarta	Museum Benteng Vredeburg Yogyakarta
4	Tugu Pal Putih Jogja	Taman Pelangi Yogyakarta	Tebing Breksi	De Mata Museum Jogja

Sistem Rekomendasi akan memberikan 3 rekomendasi tempat wisata berdasarkan nama tempat wisata yang diinputkan oleh pengguna dengan merekomendasikan tempat wisata yang memiliki kemiripan kategori dengan tempat wisata yang diinputkan.



Best Model and Recommendation



Evaluation

Evaluasi model menggunakan metrik Precision. Hal yang dipertimbangkan adalah kategori dari setiap rekomendasi

- Jika kategori sama dengan kategori Place Name yang diinputkan user, maka tidak akan dianggap sebagai error
- Jika kategori berbeda dengan kategori Place Name yang diinputkan user, maka akan dianggap error

```
Precision on 1st recommendation: 0.7142857142857143  
Precision on 2st recommendation: 0.8650793650793651  
Precision on 3rd recommendation: 0.8412698412698413
```

Meskipun terdapat variasi dalam tingkat presisi antara rekomendasi pertama, kedua, dan ketiga, namun secara keseluruhan, sistem rekomendasi telah menunjukkan kemampuan untuk memberikan rekomendasi yang relevan, dengan presisi paling tinggi di rekomendasi kedua.

Kesimpulan

- ★ Klien didominasi oleh Warga Jakarta dan Bandung yang berkisar umur 30 tahun dengan banyak membawa anak-anak di bawah umur kurang dari 10 tahun untuk berpergian
- ★ Meskipun Bogor merupakan kota wisata yang paling dekat dengan Jakarta, dimana klien agensi travel kita banyak dari Jakarta, tetapi kota Yogyakarta merupakan wisata yang menjadi favorit klien agensi travel kita.
- ★ Agensi travel kita menawarkan banyak tempat wisata dengan kisaran harga masuk Rp0-5000 dan waktu yang ditempuh dari pusat kota sekitar 60 menit
- ★ Rating dan RatingCount memiliki Korelasi positif lemah yang menunjukkan bahwa adahubungan yang ditemukan antara jumlah orang yang memberikan rating dan nilai rating yang diberikan, meskipun hubungan ini mungkin tidak begitu kuat. Dengan kata lain, meskipun lebih banyak orang memberikan rating, hal itu tidak secara signifikan menjamin bahwa nilai rating akan naik.

Website



<https://sistemrekomendasi-destinasivyogyakarta.streamlit.app/>