

Investigate Business Hotel using Data Visualization



Created by:

Okky Darmawan Kostidjan

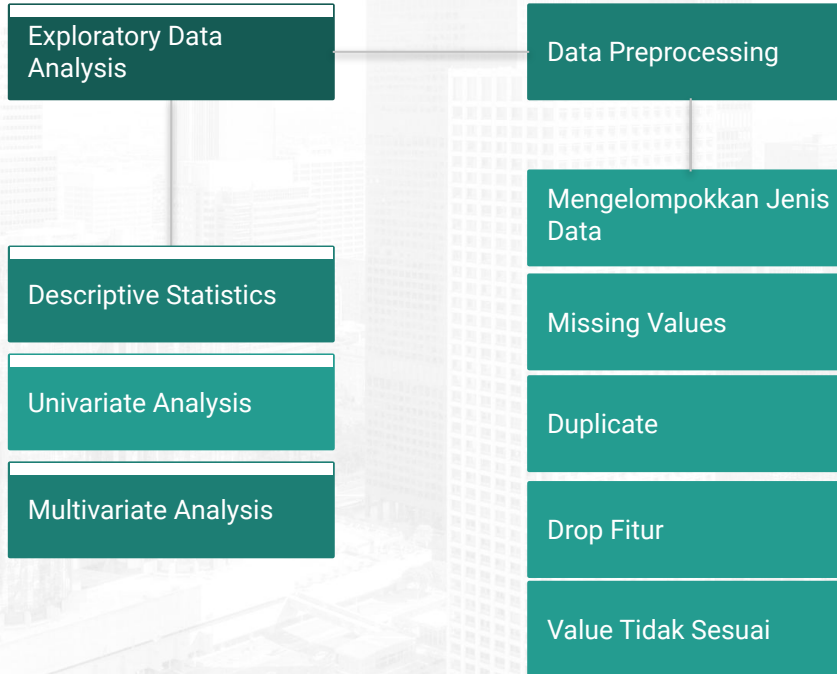
OkkyDarmawanK@gmail.com

<https://www.linkedin.com/in/okkydk/>

“I am a qualified analyst with 5 years of experience in the field of analysis, specifically as an Inventory Analyst. My are of expertise are in maintaining stock availability, analyzing causes of product damage, and generating sales and analytical reports. ”

“Sangat penting bagi suatu perusahaan untuk selalu menganalisa performa bisnisnya. Pada kesempatan kali ini, kita akan lebih mendalami bisnis dalam bidang perhotelan. Fokus yang kita tuju adalah untuk mengetahui bagaimana perilaku pelanggan kita dalam melakukan pemesanan hotel, dan hubungannya terhadap tingkat pembatalan pemesanan hotel. Hasil dari insight yang kita temukan akan kita sajikan dalam bentuk data visualisasi agar lebih mudah dipahami dan bersifat lebih persuasif.”

Langkah Kerja



- Descriptive Statistic :

Tahap ini dilakukan dengan tujuan memberikan gambaran mengenai jumlah fitur, jenis data, dan jumlah baris tiap fitur yang ada pada dataset.

- Univariate Analysis :

Tahap ini dilakukan dengan tujuan untuk mengetahui jenis distribusi, melakukan cek outliers, kepadatan data dan kesesuaian type data (apakah ada data categorical yang menggunakan basis penyimpanan data dengan menggunakan type numeric)

Contoh yang ditemukan adalah fitur `is_canceled`

- Multivariate Analysis :

kamu buat, dapat ditampilkan dan berikan link untuk mengakses file tersebut. Contohnya seperti di pojok kanan bawah.

- Mengelompokkan Jenis Data :

Tahap ini dilakukan untuk memilah jenis data categorical dan numerical pada dataset.

Pada tahap ini diperoleh pembagian data sebagai berikut.

Kolom tipe objek:

['hotel', 'arrival_date_month', 'meal', 'city', 'market_segment', 'distribution_channel', 'deposit_type', 'customer_type', 'reservation_status', 'is_canceled']

Kolom tipe numerik:

['lead_time', 'arrival_date_year', 'arrival_date_week_number', 'arrival_date_day_of_month', 'stays_in_weekend_nights', 'stays_in_weekdays_nights', 'adults', 'children', 'babies', 'is_repeated_guest', 'previous_cancellations', 'previous_bookings_not_canceled', 'booking_changes', 'agent', 'company', 'days_in_waiting_list', 'adr', 'required_car_parking_spaces', 'total_of_special_requests']

- Missing Values

Pada tahap ini dilakukan pengecekan apakah ada missing values dengan hasil cek sebagai berikut

```
▶ kolom_null = df.columns[df.isnull().any()]

# Tampilkan hasil
if len(kolom_null) > 0:
    print('Kolom-kolom yang memiliki nilai null:', list(kolom_null))
else:
    print('Tidak terdapat nilai null pada dataset')
```

Kolom-kolom yang memiliki nilai null: ['children', 'city', 'agent', 'company']

● Missing Values

Setelah diketahui fitur mana saja yang memiliki missing value, dilakukan imputasi atau drop jika memang jumlah baris dari missing value terlalu banyak atau tidak ada nilai yang sesuai untuk di input.

▼ Imputasi

```
#Imputasi nilai mean pada fitur dengan tipe data numerical yang memiliki missing values
#Imputasi missing value type numerical dengan mean
num_miss = ['children', 'agent', 'company']
for i in num_miss:
    mean_fitur = df[i].mean()
    df[i] = df[i].fillna(mean_fitur)

✓ [48] #Imputasi nilai modus pada fitur dengan tipe data categorical yang memiliki missing values
3s cat_miss = ['city']
for i in cat_miss:
    df[i] = df[i].fillna('unknown')
```

Untuk fitur children, agent dan company dipilih nilai mean dan untuk fitur city dilakukan inputasi dengan nilai unknown dikarenakan data yang telah ada sebelumnya masih bisa digunakan untuk pemetaan asal customer.

- Duplicate :

Dikarenakan pada data tersebut tidak memiliki kode booking yang bisa digunakan sebagai identitas pemesanan untuk memastikan apakah pemesanan tersebut benar-benar duplikat atau memang kebetulan memang memiliki value yang sama untuk setiap barisnya sehingga tidak dilakukan penghapusan untuk baris dengan nilai yang memiliki duplikat

- Drop Fitur

Fitur adult, children, dan babies di bisa drop karena datanya tidak relevan, ada baris yang tidak memiliki value pada ketiga fitur tersebut, ada juga baris yang tidak memiliki value pada fitur adults, sedangkan tamu pada hotel tidak memperbolehkan anak kecil atau bayi menginap sendiri tanpa ada orang dewasa yang menemani, dan juga diperlukan tanda pengenal yang hanya dimiliki orang dewasa untuk menginap di hotel.

- Value Tidak Sesuai :

Tahap ini dilakukan untuk pengecekan apakah ada value yang tidak sesuai pada suatu fitur atau tidak, pada dataset ini di fitur meal terdapat value undefined yang kemudian diganti dengan nilai No Meal

```
[123] # Cek nilai pada fitur meal  
df['meal'].unique()
```

```
array(['Breakfast', 'Full Board', 'Dinner', 'No Meal', 'Undefined'],  
      dtype=object)
```

```
✓ [124] # Mengganti nilai "Undefined" menjadi "No Meal" pada kolom "meal"  
s df['meal'] = df['meal'].replace('Undefined', 'No Meal')
```