**Investigate Hotel Business using Data Visualization**

**Overview** :

Selamat datang di Mini Project: *Investigate Hotel Business using Data Visualization*. Pada Mini Project ini, Anda berperan sebagai anggota dari tim Data Scientist di sebuah perusahaan hotel. Sebagai anggota dari tim Data Scientist, Anda mendapatkan tanggung jawab untuk memberikan insight-insight yang berhubungan dengan performa bisnis hotel. Insight tersebut dapat dicari dengan eksplorasi data, seperti menganalisis bagaimana perilaku para pelanggan dalam memesan tiket hotel ataupun mencari faktor-faktor yang mempengaruhi pembatalan pemesanan tiket hotel. Kemudian sajikan insight yang Anda dapatkan menggunakan visualisasi dan data story telling.

**Skill level** : Beginner

**Keterampilan yang dibutuhkan** :

* Python Data Visualization
* Python Data Preprocessing
* Business Understanding
* Analytical Thinking

**Learning output** :

* Memahami cara menyajikan data dengan baik
* Mampu menerjemahkan visualisasi menjadi suatu insight yang bermanfaat
* Mengenal dan memahami beberapa metrik bisnis dalam industri perhotelan.

**Estimasi waktu** : 2-4 jam

**Dataset** :

https://drive.google.com/file/d/1Cq4cHMSpnGml7a2PwDz8Dbp2W3pMbLbY/view?usp=sharing

**Tools yang digunakan** : Jupyter notebook / Google Colabs

**Tugas 1**

**Data Preprocessing**

**Deskripsi singkat** :

Sebelum memulai pemrosesan data, tahap paling awal yang harus dilakukan adalah mempersiapkan data mentah menjadi data yang bersih dan siap diolah. Kamu harus melakukan data preprocessing, yakni menangani berbagai permasalahan data seperti data yang kosong, data yang tidak sesuai, hingga mengidentifikasi data-data yang tidak dibutuhkan.

**Tujuan tugas**  : Membuat data mentah menjadi data yang siap diolah

**Kemampuan yang akan dikuasai** :

1. Kemampuan menangani permasalahan data null
2. Kemampuan mengidentifikasi nilai data yang tidak sesuai
3. Kemampuan memilih metode yang tepat untuk membersihkan data mentah

**Langkah - langkah** :

1. Download file "Dataset"
2. Mengatasi data null  
   *Hint*: Perhatikan jumlah data null dan tipe dari setiap kolom, untuk memilih metode yang tepat.
3. Mengganti value yang tidak sesuai

*Hint*: Perhatikan kolom "meal"

1. Membuang data yang tidak diperlukan  
   *Hint*: Perhatikan total jumlah tamu untuk setiap pemesanan

**Resource**  :

1. https://www.geeksforgeeks.org/data-preprocessing-machine-learning-python/

**Hasil** :

1. Data bersih yang siap digunakan untuk analisis bisnis Hotel

**Template** : [Template Tugas 1 - Investigate Business Hotel using Data Visualization](https://docs.google.com/presentation/d/1cV4DF6dIWSlki-goC5R9PlqLQc_uc5zmDdjroQrMKTQ/edit?usp=sharing)

**Tugas 2**

**Monthly Hotel Booking Analysis Based on Hotel Type**

**Deskripsi singkat** :

Dalam bisnis perhotelan, tentu saja sangat berkaitan erat dengan pelanggan. Semakin banyak pelanggan yang memesan, maka semakin banyak pula pemasukan untuk perusahaan. Oleh karena itu, menganalisa perilaku para pelanggan dalam memesan hotel sangatlah penting. Misalnya, kita dapat mengidentifikasi tipe hotel apa yang paling banyak diminati oleh pelanggan, dan kita juga dapat mengaitkannya dengan kondisi musim ketika hotel tersebut dipesan. Sehingga pada tahap ini, kamu akan mencari perbandingan jumlah pemesanan hotel setiap bulan berdasarkan pada tipe hotel.

**Tujuan tugas**  :

Melihat dan menganalisis pada bulan apa (atau musim apa) terjadi kenaikan dan penurunan dari jumlah pemesanan hotel untuk setiap jenis hotel.

**Kemampuan yang akan dikuasai** :

1. Kemampuan untuk melakukan agregasi pada tabel
2. Kemampuan untuk memilih dan membuat plot yang tepat sesuai dengan informasi yang ingin disampaikan
3. Kemampuan untuk menginterpretasi plot untuk menjadi suatu insight yang informatif

**Langkah - langkah** :

1. Buatlah tabel agregat yang menunjukan perbandingan jumlah pemesanan hotel setiap bulannya berdasarkan jenis hotel  
   *Hint*: Perhatikan data tahun kedatangan
2. Lakukan normalisasi data   
   *Hint*: Perhatikan data bulan September dan Oktober
3. Urutkan data berdasarkan bulan  
   *Hint*: Perhatikan penulisan nama bulan untuk mempermudah visualisasi
4. Buatlah plot yang dapat menunjukan perubahan kenaikan maupun penurunan dari jumlah pemesanan hotel setiap bulannya berdasarkan jenis hotel  
   *Hint*: Perhatikan jenis plot yang digunakan
5. Tuliskan interpretasi terhadap hasil plot pada langkah sebelumnya  
   *Hint*: Perhatikan hari libur di Indonesia

**Resource**  :

1. <https://realpython.com/pandas-groupby/>
2. <https://www.machinelearningplus.com/plots/top-50-matplotlib-visualizations-the-master-plots-python/>
3. <https://www.methodspace.com/blog/tips-on-interpreting-data-visualizations>

**Hasil** :

1. Insight (Plot dan Interpretasinya) yang menunjukan pada bulan apa (atau musim apa) terjadi kenaikan dan penurunan dari jumlah pemesanan hotel untuk jenis hotel yang berbeda

**Template** : [Template Tugas 2 - Investigate Business Hotel using Data Visualization](https://docs.google.com/presentation/d/1mS2kNNdcq2AHrWJgZ-B16cFr4PhYxtD7-2nRRQWE3Y0/edit?usp=sharing)

**Tugas 3**

**Impact Analysis of Stay Duration on Hotel Bookings Cancellation Rates**

**Deskripsi singkat** :

Selain menganalisis perilaku pelanggan dalam memesan hotel, untuk mengukur keberhasilan suatu bisnis perhotelan dapat kita lihat dari tingkat pembatalan pemesanan. Jika banyak pelanggan yang membatalkan pemesanan-nya, maka hal tersebut akan berpengaruh buruk terhadap performa bisnis hotel. Oleh karena itu, kamu perlu mencari tahu hal-hal apa saja yang mempengaruhi pembatalan pemesanan. Pada tahap ini kamu akan menyelidiki bagaimana durasi menginap bisa mempengaruhi tingkat pembatalan pemesanan hotel.

**Tujuan tugas**  :

Menganalisis bagaimana korelasi antara durasi menginap terhadap tingkat pembatalan pemesanan hotel

**Kemampuan yang akan dikuasai** :

1. Kemampuan untuk melakukan agregasi pada tabel
2. Kemampuan untuk memilih dan membuat plot yang tepat sesuai dengan informasi yang ingin disampaikan
3. Kemampuan untuk menginterpretasi plot untuk menjadi suatu insight yang informatif

**Langkah - langkah** :

1. Buatlah kolom baru yang berisi total durasi menginap.  
   *Hint*: Perhatikan kolom-kolom yang memuat informasi tentang durasi menginap
2. Kelompokkan nilai-nilai dari kolom baru pada langkah sebelumnya supaya lebih signifikan

*Hint*: Perhatikan persebaran datanya

1. Buatlah tabel agregat yang menunjukan perbandingan jumlah pemesanan hotel yang dibatalkan terhadap durasi menginap untuk setiap jenis hotel  
   *Hint*: Perhatikan proporsi dari pemesanan yang dibatalkan
2. Buatlah plot yang dapat menunjukan rasio pembatalan pesanan terhadap durasi menginap untuk setiap tipe hotel  
   *Hint*: Perhatikan jenis plot yang digunakan
3. Tuliskan interpretasi terhadap hasil plot pada langkah sebelumnya  
   *Hint*: Perhatikan pengaruh dari durasi menginap terhadap tingkat pembatalan pemesanan hotel

**Resource**  :

1. https://realpython.com/pandas-groupby/
2. https://www.machinelearningplus.com/plots/top-50-matplotlib-visualizations-the-master-plots-python/
3. https://www.methodspace.com/blog/tips-on-interpreting-data-visualizations

**Hasil** :

1. Insight (Plot dan Interpretasinya) yang menunjukan ada atau tidaknya hubungan antara durasi menginap terhadap tingkat pembatalan pemesanan hotel

**Template** : [Template Tugas 3 - Investigate Business Hotel using Data Visualization](https://docs.google.com/presentation/d/1eCN4Dk5JLLbQCL7gbLBT1j_jBvlcMdluuHDF_Ta_3nM/edit?usp=sharing)

**Tugas 4**

**Impact Analysis of Lead Time on Hotel Bookings Cancellation Rate**

**Deskripsi singkat** :

Bisnis hotel biasanya memperbolehkan pelanggan memesan hotel sebelum hari kedatangannya. Jarak waktu yang terjadi pun bervariasi, ada yang hanya beberapa hari, dan ada pula pelanggan yang memesan hotel sampai beberapa bulan sebelum hari kedatangan. Saat ini, kamu diminta untuk mengecek apakah jarak waktu antara pemesanan hotel dengan hari kedatangan pelanggan memiliki pengaruh terhadap tingkat pembatalan pemesanan hotel.

**Tujuan tugas**  :

Menganalisis bagaimana korelasi antara jarak waktu pemesanan hotel terhadap tingkat pembatalan pemesanan hotel

**Kemampuan yang akan dikuasai** :

1. Kemampuan untuk melakukan agregasi pada tabel
2. Kemampuan untuk memilih dan membuat plot yang tepat sesuai dengan informasi yang ingin disampaikan
3. Kemampuan untuk menginterpretasi plot untuk menjadi suatu insight yang informatif

**Langkah - langkah** :

1. Buatlah kolom baru yang berisi kategorisasi dari kolom jarak waktu pemesanan

*Hint*: Perhatikan interval dalam membuat kategorisasi

1. Buatlah tabel agregat yang menunjukan perbandingan jumlah pemesanan hotel yang dibatalkan terhadap jarak waktu pemesanan untuk setiap jenis hotel  
   *Hint*: Perhatikan proporsi dari pemesanan yang dibatalkan
2. Buatlah plot yang dapat menunjukan rasio pembatalan pesanan terhadap jarak waktu pemesanan untuk setiap tipe hotel  
   *Hint*: Perhatikan jenis plot yang digunakan
3. Tuliskan interpretasi terhadap hasil plot pada langkah sebelumnya  
   *Hint*: Perhatikan pengaruh interval waktu

**Resource**  :

1. https://realpython.com/pandas-groupby/
2. https://www.machinelearningplus.com/plots/top-50-matplotlib-visualizations-the-master-plots-python/
3. https://www.methodspace.com/blog/tips-on-interpreting-data-visualizations

**Hasil** :

1. Insight (Plot dan Interpretasinya) yang menunjukan ada atau tidaknya hubungan antara jarak waktu pemesanan terhadap tingkat pembatalan pemesanan hotel

**Template** : [Template Tugas 4 - Investigate Business Hotel using Data Visualization](https://docs.google.com/presentation/d/1g-84mMHoiRYCKVtkd_apj6yNHexhAKkaua6GknxFS5k/edit?usp=sharing)