

# Proyek Data: Pengembangan Hotel Search Engine untuk DQLab Travel

## Background:

Kamu adalah seorang data scientist di Marketplace 'DQLab Travel' yang ditempatkan pada divisi Hotel. Untuk memberikan pengalaman yang baik kepada pelanggan DQLab Travel dalam mencari hotel yang tepat, Bosmu, Kroma, manager DQLab Travel memberikan sebuah challenge untuk mengembangkan Search Engine guna mencari hotel yang relevan dengan harga terbaik.

Kamu akan mempresentasikan model search engine dan mendapatkan feedback dari Senja tanggal 20 Juni 2020, senior Data Analyst yang bekerja di DQLab Travel.

## Reference:

Menggunakan pengalaman yang kamu dapatkan sebelumnya dari latihan project:

- Building Recommender System using Python  
(<https://academy.dqlab.id/main/package/project/212>)
- Building Recommender System using Similarity Function in Python  
(<https://academy.dqlab.id/main/package/project/214>)

## Case Study:

Kamu diberikan dua studi kasus oleh Senja, rekan kerjamu, dimana calon pelanggan (user) ingin mencari hotel dengan harga terbaik dan terjangkau, dengan spesifikasi:

- *Kasus pertama*
  - *Servis: Family Room*
  - *Provinsi: Jawa Timur*
  - *Kota: Surabaya*
  - *Lokasi: any*
  - *Diurutkan berdasarkan harga terendah dan rating terbaik*
- *Kasus kedua*
  - *Provinsi: DI Yogyakarta*
  - *Kota: Jogja*
  - *Hotel apapun yang relevan dengan "Santika hotels"*

## Contoh Output:

hotel_id	hotel_name	hotel_description	hotel_province	hotel_city	hotel_address	service_provided	Price_per_night (IDR)	overall_avg_rating
bcd	Hotel BCD	Very amusing hotel, international environment	Jawa Timur	Surabaya	Jl. Basuki Rachmat	[Executives Suites, Standard Room, Family Room]	[700000, 200000, 350000]	4.2

## Challenge Details: (max 100 pts)

*!Jangan lupa untuk membuat "requirements.txt" yang isinya semua library dan versinya (jika memungkinkan) yang digunakan dalam coding nya untuk menyelesaikan permasalahan ini*

1. Buatlah langkah yang dapat dieksekusi untuk mendapatkan tujuannya, lakukan ETL pada data tersebut, siapkan data yang sudah "bersih" untuk langkah selanjutnya, "bersih" di sini adalah asumsi masing-masing. (up to 10 pts)
2. Apakah nilai review berguna untuk model yang kamu buat? Bagaimana hal itu bisa berguna? Akan sangat dihargai jikalau kamu dapat menyebutkan alasannya dan langkah-langkah yang diperlukan untuk membuktikan hal tersebut. (up to 20 pts)
3. Buatlah analisis data untuk eksplorasi (Exploratory Data Analysis/EDA) untuk berbagai data spesifik, dan sebutkan berbagai informasi unik yang dapat ditarik dari langkah tersebut. (up to 30 pts)
4. Putuskan fitur-fitur yang mana saja yang dapat dimasukkan ke dalam pembentukan model/formula, sebutkan alasannya. (up to 10 pts)
5. Buatlah model/formula dari sistem rekomendasi. (up to 15 pts)
6. Dapatkah kamu mengevaluasi sistem rekomendasi kamu sudah seberapa akurat? Tolong tunjukkan caranya. (up to 5 pts)
7. Bagaimana kamu dapat meningkatkan sistem rekomendasi kamu? Tolong sebutkan caranya, alasannya dan keuntungannya jikalau kamu dapat melakukan hal-hal tersebut. (up to 10 pts)



**Bonus:** Jika kamu dapat melakukan pembentukan ulang dengan menggunakan pendekatan/model/formula lain untuk kasus spesifik seperti challenge ini dan kamu meningkatkan akurasi, tolong sebutkan alasannya, tunjukkan pembentukan modelnya dengan setiap langkahnya dan perbandingannya dengan model yang telah kamu buat di langkah sebelumnya. (100 pts, sedangkan kalau ada langkah yang hilang atau ketidakjelasan pembentukan modelnya maka akan mendapat 0 pts)

## Submission Details:

1. Publikasikan hasil data modelling dan analisismu di LinkedIn (dalam bentuk pdf).
2. Pastikan kamu menjawab points yang ada pada challenge details di atas.
3. Tag akun LinkedIn DQlab, Karl Christian, dan tambahkan hashtag #MulaiBelajarData
4. Periode pengerjaan: 02 - 10 Juni 2020
5. Periode upload pekerjaan: 11 Juni 2020, pukul **23:59 WIB**
6. Pengumuman member DQLab yang terpilih untuk sesi mentoring: 17 Juni 2020
7. Periode mentoring: 20 Juni 2020

## Reward:

Top 5 Submissions akan mendapatkan reward, berupa:

1. Feedback dari Karl Christian sebagai mentor dalam proyek ini.
2. Sertifikat Proyek yang ditandatangani oleh Karl Christian
3. Kesempatan untuk membangun Proyek Data pertamamu!
4. FREE Tambahan Waktu belajar (DQLab Premium Membership) dengan durasi 6 bulan!

Top 10 Submissions akan mendapatkan reward, berupa:

1. Sertifikat Proyek yang ditandatangani oleh Karl Christian
2. Kesempatan untuk membangun Proyek Data pertamamu!
3. FREE Tambahan Waktu belajar (DQLab Premium Membership) dengan durasi 3 bulan!

**GOOD LUCK!**