

Guideline Portfolio Project Bootcamp Data Analyst with Python & SQL Batch 1

Tentang Project

DQLab sebagai perusahaan data mendapatkan permintaan klien yang berasal dari Singapura. Calon owner (pemilik) property tersebut akan menyewakan properti di marketplace Airbnb dan tentunya ingin mendapatkan keuntungan maksimal dari penyewaan properti.. Sebelumnya telah diketahui bahwa tempat penyewaan yang tepat berpengaruh terhadap tingginya kegiatan penyewaan. Sebagai seorang data analyst DQLab yang ditugaskan mengolah data properti Airbnb Singapura, analisis seperti apa yang akan kamu berikan untuk membantu client agar client tersebut bisa mendapatkan keuntungan terbaik dari hasil kegiatan penyewaan yang akan dilakukan?

Dalam final project kali ini akan dilakukan analisis terhadap data penginapan airbnb di Singapore. Terdapat 3 jenis data (tersedia dalam excel dan csv) yang diberikan kepada peserta. Data yang diberikan mencakup data detail per listings, data historis penyewaan per listings, dan mapping neighbourhood_group untuk setiap neighborhood.

Diberikan juga kamus data untuk setiap dataset yang diberikan. Dalam kamus tersebut, setiap field diberi deskripsi singkat mengenai arti dari field tersebut.

Dataset Glossarium

Data 1: DQLab Airbnb Listing Dataset

Data pertama berisi mengenai daftar listing Airbnb di Singapore.

File name : **DQLab_listings(22Sep2022).csv**

DQLab_listings(22Sep2022).xlsx

Kamus data untuk field dataset listings adalah seperti berikut :

No	Field Name	Description
1	id	Listing's unique Airbnb id
2	name	Listing's name/ display in the Airbnb web
3	host_id	Listing host's unique id
4	host_name	Listing host's name
5	neighborhood	Listings's neighborhood name
6	latitude	Listing's earth latitude location
7	longitude	Listing's earth longitude location

8	room_type	Listing's room type
9	price	Listing's price
10	minimum_nights	indicator of minimum stay length
11	availability_365	indicator of the total number of days the listing is available for during the year

Data 2 : Listings Reviews History

Data kedua berisi mengenai data historis order/ rent per listing di Singapore dari 1 Januari 2018 sampai dengan 22 September 2022.

File name : **DQLab_reviews(22Sep2022).csv**

DQLab_reviews(22Sep2022).xlsx

Kamus data untuk field dataset reviews adalah seperti berikut :

No	Field Name	Description
1	listing_id	Listing's unique Airbnb id
2	date	Date when specific listing was rented

Data 3 : Neighborhood Mapping

Data ketiga berisi mengenai mapping region Singapore berdasarkan neighbourhoodnya

File name : **DQLab_neighbourhood(22Sep2022).csv**

DQLab_neighbourhood(22Sep2022).xlsx

Kamus data untuk field dataset neighborhood adalah seperti berikut :

No	Field Name	Description
1	Neighborhood_group	Singapore region (North/East/South/West/North-East)
2	neighborhood	Singapore neighborhoods name

Link dataset :

<https://drive.google.com/drive/folders/1kplf-V22uhghGP-zQ1kXsX7b9VmhyURy?usp=sharing>

Analisa Rekomendasi

Beginner

- Bagaimana distribusi harga listing airbnb di Singapura?
- Bagaimana trend harga pada Neighborhood ?
- Apakah room_type berpengaruh terhadap harga?
- Bagaimana trend penyewaan dari 01-01-2018 sampai dengan 22-09-2022 ?

Intermediate

- Berapa rata-rata dan median harga listings secara keseluruhan ?
- host_id dan host_name apa yang mempunyai jumlah listing terbanyak dan paling sedikit ?
- Bagaimana perubahan trend penyewaan listing dari tahun ke tahun ? Mana yang paling tinggi dan paling rendah ? Kenapa bisa terjadi demikian ?

Semi-advanced

- id dan name listing apa yang mempunyai number of reviews paling banyak? Mengapa bisa terjadi demikian?
- Bagaimana rata-rata harga listings per neighbourhood_group?
- Apakah room_type berpengaruh pada ramai atau tidaknya (number of reviews) suatu listing?
- Bagaimana tingkat aktivitas penyewaan setiap Neighbourhood ? dan berapa harga listings termahalnya ?
- Region mana yang memiliki aktivitas tertinggi (number of reviews) ? Bagaimana hubungannya dengan number of listing tiap region?

Advanced

- Ada berapa listings yang memiliki reviews diatas rata-rata per masing-masing neighbourhood_group ?
- Berdasarkan neighbourhood yang paling tinggi aktivitasnya, berapa jarak terdekat dari listings teraktif ke pesisir ?
- Bagaimana kondisi industri penginapan setelah 23 Januari 2020 ? Ada berapa banyak listings yang masih aktif setelah tanggal tersebut? Ada kejadian apa pada tanggal tersebut ?
- Bagaimana percentage accumulation jumlah reviews dari tahun ke tahun (mulai dari 2018 – 2022)? Berapa kenaikan paling tajam dan penurunan paling tajam ?
- Pada bulan dan tahun berapa aktivitas penyewaan listing paling sedikit dan paling banyak ?

Tips Mengerjakan :

Terdapat beberapa tips yang dapat kamu lakukan dalam melakukan analisis untuk final project ini, diantaranya :

- Dalam memulai proses pemahaman data, data dapat dilakukan join satu dataset dengan yang lainnya sesuai kebutuhan. Proses join bisa dapat dilakukan menggunakan *unique key values*
- Dapat dilakukan pengelompokan data sesuai dengan variabel yang ingin diujikan, misalnya data dapat dikelompokkan berdasarkan neighbourhood menggunakan syntax *groupby*
- Untuk dapat mendukung serta membayangkan proses analisis dan menjawab pertanyaan, dapat dibuat visualisasi terlebih dahulu sesuai dengan materi yang sudah diajarkan.
- Berdasarkan analisis dan visualisasi tersebut, dapat dirancang dashboard yang dapat menyampaikan findings yang ditemukan

Ketentuan Submission

Upload hasil portfolio-mu di LinkedIn, Tag DQLab, Mentor & Asisten Mentor DQLab