

PANDUAN PRAKTIK MANDIRI

Program Menganalisis Data bagi Desainer Sistem TI (Data Analyst)

Halo, Teman Karier.mu!

Selamat kamu telah menyelesaikan serangkaian aktivitas dalam program “Menganalisis Data bagi Desainer Sistem TI (Data Analyst)”. Kami harap kamu belajar banyak dari program ini. Agar dapat lebih memahami materi yang telah diberikan, kami memberikan tugas praktik mandiri ini. Di bawah ini terdapat langkah-langkah pengerjaan praktik mandiri ini. Mohon baca dengan baik dan ikuti instruksi yang diberikan.

Selesaikan analisis data berikut secara berurutan dengan menggunakan *tools* yang telah kita pelajari pada program yaitu SQLite, Python, dan Tableau.

1. SQLite

- a. Buka new notebook di Google Colab
<https://colab.research.google.com/> sesuai dengan akun email gmail kamu
- b. Upload database chinook.db ke dalam folder *Files*
- c. Buat syntax untuk restore database dan melihat tabel-tabel yang ada di dalam database chinook.db. Tunjukkan salah satu isi dari tabel yang ada di dalam database ini dengan menggunakan gabungan syntax SQLite dan query berbasis SQL. Contoh isi query: “select * from <nama_table>”
- d. Jawablah pertanyaan-pertanyaan berikut dengan menggunakan syntax query SQLite di Google Colab.
 - i. Apa 5 genre paling banyak jumlah lagunya (petunjuk: gunakan table tracks dan genres)
 - ii. Apa 3 jenis media yang paling banyak dipakai? (petunjuk: gunakan table tracks dan media_type)
 - iii. 5 lagu apa yang durasinya paling lama? (petunjuk: gunakan table tracks)

- iv. Rata-rata setiap genre itu durasi lagunya berapa lama? genre mana yang durasinya paling lama dan durasi yang paling sedikit? (petunjuk: gunakan table tracks dan genres)
- v. Customers paling banyak membeli lagu di bulan apa? (gunakan tabel invoices)
- e. Screenshot syntax dan output

2. Python

- a. Simpan tabel invoices ke dalam sebuah pandas dataframe. Table ini akan jadi sumber pengolahan data untuk tugas Python. Jangan lupa simpan data ini ke-dalam data dengan ekstensi <nama_file>.csv. Download untuk dipakai dalam tugas C.
- b. Keluarkan angka mean, median, Q1, Q3, min, dan max untuk kolom Total, untuk setiap tahun.
- c. Plot perbandingan antara total invoice tadi ke dalam sebuah boxplot yang berbeda untuk masing-masing tahun. Lakukan perbandingan rata-rata dan persebaran data, dan tuliskan kesimpulan kamu.
- d. Pada kolom Total, temukan nilai outliers dengan menggunakan prinsip IQR (interquartile Range). Manakah invoice_id yang merupakan outliers? Tunjukan dengan grafik histogram
- e. Berapakah jumlah missing values (pengamatan yang tidak ada nilainya) masing-masing pada kolom Billing State dan BillingPostalCode? Berapa persen missing values yang terjadi?

Untuk tugas 1 (SQLite) dan 2 (Python) jawaban bisa dijelaskan dalam file .ipynb (notebook). Save dan Download file tersebut dan kumpulkan pada aktivitas “Pengumpulan Praktik Mandiri” di pertanyaan nomor 1 dan 2

3. Tableau

- a. Load data Invoices yang sudah didownload dari tugas B dengan menggunakan software Tableau Public. Perhatikan apakah tipe-tipe data yang dideteksi oleh Tableau Public sudah tepat?
- b. Bikin sebuah trend line chart yang menunjukkan jumlah customers dari waktu ke waktu. Setiap garis menunjukkan tahun yang berbeda.

- c. Bikin grafik batang yang menunjukkan total invoices di masing-masing negara. 5 Negara apakah yang memiliki jumlah Total invoices tertinggi?
- d. Berapakah rata-rata Total Invoices dari 5 negara tersebut? Tampilkan dalam bentuk tabel.
- e. Buat satu visualisasi apapun sesuai kreasi kamu, jelaskan mengapa kamu ingin membuat visualisasi ini. Apa cerita yang ingin kamu sampaikan?
- f. Buat sebuah dashboard dengan visualisasi-visualisasi di atas. Kamu bisa bebas tambahkan dengan visualisasi lain yang kamu inginkan.
- g. Screenshot grafik-grafik di sheet dan di dashboard yang sudah kamu buat dan simpan di dalam 1 file dan kumpulkan pada aktivitas “Pengumpulan Praktik Mandiri” di pertanyaan nomor 3

Selamat Mengerjakan!