

**!! Catatan:** Diharapkan seluruh pengerjaan Assignment tidak sepenuhnya mengandalkan penggunaan AI **!!**

*"Proses belajar ibarat menanam pohon. Jika hanya mengandalkan AI tanpa memahami esensinya, yang berkembang bukan kompetensimu, melainkan ketergantungan yang melemahkan."*

*- Learning Design Dibimbing*

# Assignment Guidance: Web Scraping

Data Science dan Data Analyst Bootcamp

## Periode Pembelajaran

Web Scraping

## Objectives

1. Menentukan Objective Penyelesaian dari Case Study Web Scraping
2. Mengumpulkan Data dari Web Menggunakan BeautifulSoup ataupun Selenium
3. Melakukan proses data cleaning jika diperlukan beserta alasannya, lalu lakukan data visualisasi terhadap data yang anda ambil.
4. Simpan ke dalam PostgreSQL (sesuai dengan materi SQL yang telah diajarkan)

## Deskripsi Assignment

Dalam assignment ini, Anda diminta untuk melakukan web scraping pada sebuah website menggunakan BeautifulSoup untuk mengumpulkan data dari **job street**. Ikuti saran task yang telah diberikan dibawah. Anda diharuskan memiliki tujuan analisis dari data yang ingin diambil, serta anda diwajibkan untuk mengambil data yang lebih detail selain mengambil data pada link general.

Clue :

- <https://id.jobstreet.com/id/data-analyst-jobs> -> API link general (data general)
- <https://id.jobstreet.com/id/data-analyst-jobs?jobId=87262587&type=standard> -> API link detail (disesuaikan dengan sebanyak data yang terambil dengan cara yang sama seperti melakukan pengambilan API URL href yang telah diajarkan di kelas + looping pada API URL href tersebut yang telah disimpan kedalam dataframe)
- Datanya nanti ada 2 dataframe yaitu data general dan data detail

## Detail Assignment

Task yang diberikan :

1. Tentukan data apa yang ingin Anda kumpulkan dari halaman salah satu website di atas dan bagaimana data tersebut dapat digunakan dalam konteks aplikasi

dunia nyata. Jelaskan mengapa data tersebut penting dan apa manfaatnya jika disimpan dalam database PostgreSQL.

2. Gunakan BeautifulSoup untuk mengumpulkan data job street. Tampilkan hasil scraping Anda dalam bentuk tabel atau dataframe baik dari data general dan data detail.
3. Setelah data berhasil di-scraping, lakukan transformasi yang diperlukan seperti pembersihan data (jika diperlukan beserta alasannya), atau penggabungan data (concat) jika diperlukan. Jelaskan setiap langkah yang Anda ambil dan bagaimana hal itu mempengaruhi integritas data.
4. Lakukan visualisasi EDA pada data yang sudah anda ambil serta berikan insightnya dari data yang anda ambil.
5. Buatlah database di PostgreSQL dan tabel untuk menyimpan data hasil scraping berupa data general dan data detail.
6. Setelah data disimpan, lakukan query sederhana untuk menarik data dari PostgreSQL dan tampilkan beberapa contoh hasil query tersebut dalam bentuk Screen Shot yang bisa anda simpan di drive beserta file code scrapping yang telah anda kerjakan. Jelaskan bagaimana data ini dapat digunakan untuk analisis lebih lanjut atau aplikasi lainnya.

## Tools

Jupyter Notebook

PostgreSQL Database

Python (BeautifulSoup)

## Pengumpulan Assignment

### Details :

**Dikumpulkan dalam link google drive yang memuat (file code scrapping, screenshot query dan database tempat menyimpan file scrapping)** dan berikan **izin akses** ke dalam drive tersebut, lakukan secara INDIVIDU, di LMS

### Notes :

Bilamana terdapat **salah pengumpulan file, private link pada google drive dengan alasan apapun tidak diterima dan akan diberikan nilai 0**, selain itu hal ini berlaku jika jawaban yang diberikan tidak sesuai dengan yang diminta pada parameter indikator penilaian / pada file template assignment.

## Indikator Penilaian

| No. | Aspek Penilaian | Parameter   | Bobot Maksimal |
|-----|-----------------|---|----------------|
| 1   | Web Scraping    | Menentukan Objective dari Case Study  | 10             |
| 2   |                 | Mengumpulkan Data yang Relevan (Web Scrapping) untuk data general dan data detail | 40             |
| 3   |                 | Pembersihan, Visualisasi data dan insightnya, Concat Data                         | 40             |
| 4   |                 | Proses Penyimpanan ke database & Query sederhana                                  | 10             |

### Sanksi Penggunaan AI:

Apabila student terdeteksi 100% menggunakan AI, maka hasil assignment akan diberikan skor 0.

### Ketentuan Pencapaian Nilai:

Nilai minimum Lulus Penyaluran Kerja: 75

Nilai minimum Lulus Bootcamp: 65

### Ketentuan Penilaian:

Mengumpulkan Assignment tepat waktu: Sesuai dengan nilai yang diberikan mentor

Mengumpulkan Assignment 12 jam setelah deadline: - 3 dari nilai yang diberikan mentor

Mengumpulkan Assignment 1 x 24 Jam setelah deadline: - 6 dari nilai yang diberikan mentor

Mengumpulkan Assignment 2 x 24 Jam setelah deadline: - 12 dari nilai yang diberikan mentor

Mengumpulkan Assignment 3 x 24 Jam setelah deadline: - 18 dari nilai yang diberikan mentor

Mengumpulkan Assignment 4 x 24 Jam setelah deadline: - 24 dari nilai yang diberikan mentor

Mengumpulkan Assignment 5 x 24 Jam setelah deadline: - 30 dari nilai yang diberikan mentor

Mengumpulkan Assignment 6 x 24 Jam setelah deadline: - 36 dari nilai yang diberikan mentor

Mengumpulkan Assignment 7 x 24 Jam setelah deadline: - 42 dari nilai yang diberikan mentor