

# Assignment Guidance: Web Scraping

Data Science 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 Cleaning Data (Regex) jika diperlukan kemudian disimpan dalam bentuk dataframe untuk Persiapan Penyimpanan, terjemahkan regex pada bagian materi kelas sesuai dengan yang diminta
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 tentang akomodasi tempat penginapan di Indonesia dari website (**Oyo atau website lainnya, disarankan silahkan bisa di Oyo**). Ikuti saran task yang telah diberikan dibawah.

## 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 akomodasi tempat penginapan / tempat penginapan dari saran website. Data yang harus dikumpulkan meliputi setidaknya: nama tempat penginapan, harga, rating, dan lokasi. Tampilkan hasil scraping Anda dalam bentuk tabel atau dataframe.
3. Setelah data berhasil di-scrape, lakukan transformasi yang diperlukan seperti pembersihan data (misalnya menghapus simbol-simbol tertentu dari harga), normalisasi data (misal pada satuan harga), atau penggabungan data (concat) jika diperlukan. Jelaskan setiap langkah yang Anda ambil dan bagaimana hal itu mempengaruhi integritas data.
4. Buatlah database di PostgreSQL dan tabel untuk menyimpan data hasil scraping.
5. 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.

6. Sebagai tambahan, terjemahkan makna code regex yang diberikan di kelas

## Tools

Google Collaboratory / 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)	60
3		Terjemahan regex code di kelas, Pembersihan dan Concat Data	20
4		Proses Penyimpanan ke database & Query sederhana	10

### 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