

**LAPORAN PENUGASAN**  
**"Python For Data Wrangling"**



**DISUSUN OLEH:**

Muhammad Aryasatya Nugroho (22083010085)

**DOSEN PENGAMPU:**

Kartika Maulida Hindrayani, S.Kom., M.Kom (NIP. 20219920909201)

**PROGRAM STUDI SAINS DATA**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN" JAWA TIMUR**

**2023**

## EXERCISE WEEK 2

### SOAL NOMOR 1

Definisi String, List, Dictionary

1. **String** adalah tipe data yang digunakan untuk merepresentasikan teks atau karakter dalam Python. Untuk membuat sebuah string, Dapat menggunakan tanda kutip tunggal ( ' ') atau ganda ( " ").
2. **List** adalah tipe data yang digunakan untuk menyimpan kumpulan nilai atau objek dalam satu variabel. Untuk membuat sebuah list, Dapat menggunakan tanda kurung siku ( []).
3. **Dictionary** adalah tipe data yang digunakan untuk menyimpan pasangan key-value atau kunci-nilai dalam satu variabel. Untuk membuat dictionary, Dapat menggunakan tanda kurung kurawal ( {}) dan memisahkan setiap pasangan key-value dengan tanda titik dua ( : ).

Penggunaan beberapa metode pada String, List, Dictionary

#### 1. String

- a) `upper()`: untuk mengubah semua karakter dalam string menjadi huruf kapital.
- b) `lower()`: untuk mengubah semua karakter dalam string menjadi huruf kecil.
- c) `replace()`: untuk mengganti suatu karakter atau teks dalam string dengan karakter atau teks lain.
- d) `split()`: untuk memisahkan string menjadi beberapa bagian berdasarkan separator yang ditentukan.

Contoh Implementasi dalam program python:

```
1  string = "Ini merupakan sebuah string."
2
3  # BEBERAPA METODE YANG DAPAT DILAKUKAN PADA STRING:
4
5  # 1. Mengubah semua karakter dalam string menjadi huruf kapital
6  string_kapital = string.upper()
7  # 2. Mengubah semua karakter dalam string menjadi huruf kecil
8  string_kecil = string.lower()
9  # 3. Memisahkan string menjadi beberapa bagian berdasarkan separator yang ditentukan
10 string_memisah = string.split(", ")
11 # 4. Mengganti suatu karakter dalam string dengan karakter lain
12 string_mengganti = string.replace("Halo", "Hi")
```

-Aryasatya-

#### 2. List

- a) `append()`: untuk menambahkan sebuah objek ke dalam list.
- b) `remove()`: untuk menghapus sebuah objek dari list.
- c) `sort()`: untuk mengurutkan objek dalam list secara ascending atau descending.

Contoh Implementasi dalam program python:

```

1 list_angka = [3,1,4,5,2]
2 list_buah = ['apel', 'jeruk', 'mangga', 'pisang']
3
4 # BEBERAPA METODE YANG DAPAT DIGUNAKAN PADA LIST
5
6 # 1. Menambahkan sebuah objek ke dalam list
7 list_angka.append(6)
8 # 2. Menghapus sebuah objek dari list
9 list_buah.remove('jeruk')
10 # 3. Mengurutkan objek dalam list secara ascending
11 list_angka.sort()
12 # 4. Mengurutkan objek dalam list secara descending
13 list_angka.sort(reverse=True)

```

-Aryasatya-

### 3. Dictionary

- keys(): untuk mengambil semua kunci (key) dalam dictionary.
- values(): untuk mengambil semua nilai (value) dalam dictionary.
- items(): untuk mengambil semua pasangan kunci-nilai (key-value) dalam dictionary.

Contoh Implementasi dalam program python:

```

1 data_mahasiswa = {'nama': 'aku', 'nim': '987654321', 'prodi': 'Sains Data'}
2 data_buku = {'judul': 'Belajar Python', 'penulis': 'Jacqueline', 'tahun_terbit': 2022}
3
4 # BEBERAPA METODE YANG DAPAT DILAKUKAN PADA DICTIONARY:
5
6 # 1. Mengambil semua kunci dalam dictionary
7 keys_mahasiswa = data_mahasiswa.keys()
8 # 2. Mengambil semua nilai dalam dictionary
9 values_buku = data_buku.values()
10 # 3. Mengambil semua pasangan kunci-nilai dalam dictionary
11 items_mahasiswa = data_mahasiswa.items()

```

-Aryasatya-

## SOAL NOMOR 2

### 1. Membuat list sebanyak 100 nomor random.

Dapat dilakukan dengan menggunakan modul random dalam Python sehingga dapat memudahkan kita untuk tidak merandom angka secara manual. Contoh Implementasi kode yang digunakan:

```

1  # Untuk membuat list 100 nomor random, dapat menggunakan modul random dari Python:
2  import random
3
4  list_angka = []
5
6  for i in range(100):
7      angka = random.randint(0, 100)
8      list_angka.append(angka)
9
10 print(list_angka)

```

-Aryasatya-

Penjelasan mengenai kode diatas:

- 1) Pertama, Import modul random dengan sintaks 'import\_random'.
- 2) Selanjutnya buat list kosong menggunakan 'list\_angka = []'
- 3) Kemudian, gunakan loop for unruk mengulangi random nomor sebanyak 100 kali dengan 'for i in range(100):'.
- 4) Pada setiap iterasi loop, gunakan metode randint() dari modul random untuk merandom angka dari 0 hingga 100 dengan 'angka = random.radint(0,100)'.
- 5) Setelah mendapat nomor random, kita tambahkan ke list\_angka menggunakan metode append() dengan sintaks 'list\_angka.append(angka)'.
- 6) Setelah loop selesai, kita mencetak isi dari llist\_angka menggunakan sintaks 'print(list\_angka)'.

Kode diatas akan menghasilkan output:

```

[64, 62, 33, 100, 3, 54, 14, 18, 83, 27, 26, 46, 57, 99, 61, 35, 6, 36, 4, 42, 59, 47, 26, 45, 47, 36, 20, 52, 0, 64, 75, 78, 75, 55, 87,
71, 65, 58, 61, 63, 61, 21, 83, 67, 30, 12, 13, 24, 7, 64, 41, 64, 11, 57, 83, 59, 91, 87, 11, 3, 64, 51, 92, 14, 40, 26, 5, 46, 92, 96, 3
5, 40, 2, 25, 46, 85, 65, 33, 27, 4, 14, 33, 19, 93, 2, 89, 72, 82, 34, 5, 55, 80, 90, 34, 57, 14, 0, 23, 43, 19]

```

## 2. Membuat list baru dari list acak, dengan angka yang habis dibagi 3.

Untuk membuat list baru dari list acak sebelumnya dengan angka yang habis dibagi 3, dapat menggunakan loop for untuk mengecek setiap angka dalam list dan memasukkan hanya angka yang habis dibagi 3 ke dalam list baru. Contoh implementasi kode yang digunakan:

```

1  # Untuk membuat list baru dari list acak sebelumnya dengan angka yang habis dibagi 3, dapat menggunakan
2  # loop for mengecek setiap angka dalam list dan memasukkan angka yang habis dibagi 3 ke dalam list baru
3
4  list_angka_baru = []
5
6  for angka in list_angka:
7      if angka % 3 == 0:
8          list_angka_baru.append(angka)
9
10 print(list_angka_baru)

```

-Aryasatya-

Penjelasan mengenai kode diatas:

- 1) Pertama, buat sebuah list kosong untuk menyimpan hasil angka yang habis dibagi 3 dengan sintaks 'list\_angka\_baru = []'.

- 2) Selanjutnya, gunakan loop for untuk mengecek setiap angka dalam list\_angka menggunakan sintaks 'for angka in list\_angka:'.
- 3) Pada setiap iterasi loop, dapat menggunakan kondisi if untuk mengecek apakah angka tersebut habis dibagi 3 dengan menggunakan 'if angka % 3 == 0:'.
- 4) Jika angka habis idbagi 3, maka kita menambahkannya ke dalam list\_angka\_baru menggunakan metode append() dengan menggunakan sintaks 'list\_angka\_baru.append(angka)'.
- 5) Setelah loop selesai, kita mencetak isi dari list\_angka\_baru menggunakan sintaks 'print(list\_angka\_baru)'.

Kode diatas akan menghasilkan output:

```
[33, 3, 54, 18, 27, 57, 99, 6, 36, 42, 45, 36, 0, 75, 78, 75, 87, 63, 21, 30, 12, 24, 57, 87, 3, 51, 96, 33, 27, 33, 93, 72, 90, 57, 0]
```

### 3. Menghitung Panjang kedua list dan menyimpan selisihnya dalam variabel baru.

Untuk menghitung Panjang dari dua list, dapat menggunakan fungsi len() pada masing-masing list. Setelah itu, dapat dihitung selisihnya dan menyimpannya dalam sebuah variabel baru. Contoh implementasi kode yang digunakan:

```
1 # Untuk menghitung panjang dari dua list, dapat menggunakan fungsi len() pada masing-masing list.
2 # Selain itu, dapat dihitung selisihnya dan menyimpannya dalam sebuah variabel baru.
3
4 panjang_list_angka = len(list_angka)
5 panjang_list_angka_baru = len(list_angka_baru)
6 selisih_panjang = abs(panjang_list_angka - panjang_list_angka_baru)
7
8 print("Panjang list angka:", panjang_list_angka)
9 print("Panjang list angka baru:", panjang_list_angka_baru)
10 print("Selisih panjang:", selisih_panjang)
```

-Aryasatya-

Penjelasan mengenai kode diatas:

- 1) Pertama, gunakan fungsi len() untuk menghitung Panjang dari list\_angka dan list\_angka\_baru masing masing dan menyimpannya ke dalam variabel panjang\_list\_1 dan panjang\_list\_2.
- 2) Selanjutnya, hitung selisih panjang kedua list dengan menggunakan sintaks 'selisih\_panjang = abs(panjang\_list1 - panjang\_list\_2)'.
  - Menggunakan fungsi abs() agar selisihnya selalu positif.
- 3) Setelah itu, cetak panjang kedua list dan selisih panjang dengan menggunakan sintaks print("Panjang list pertama:", panjang\_list\_1), print("Panjang list kedua:", panjang\_list\_2), print("Selisih panjang:", selisih\_panjang).

Kode diatas akan menghasilkan output:

```
Panjang list angka: 100
Panjang list angka baru: 35
Selisih panjang: 65
```

#### 4. Melakukan loop sebanyak 10 kali dengan langkah 1,2, dan 3.

Kita dapat melakukannya dengan Contoh implementasi kode yang digunakan:

```
1  # Melakukan looping for sebanyak 10 kali
2
3  import random
4
5  for j in range(10):
6      list_angka = []
7
8      for i in range(100):
9          angka = random.randint(0, 100)
10         list_angka.append(angka)
11
12         print("List acak ke", j+1, ":", list_angka)
13
14         list_angka_baru = []
15
16         for angka in list_angka:
17             if angka % 3 == 0:
18                 list_angka_baru.append(angka)
19
20         print("List baru ke", j+1, ":", list_angka_baru)
21
22         panjang_list_angka = len(list_angka)
23         panjang_list_angka_baru = len(list_angka_baru)
24         selisih_panjang = abs(panjang_list_angka - panjang_list_angka_baru)
25
26         print("Panjang list angka ke", j+1, ":", panjang_list_angka)
27         print("Panjang list angka baru ke", j+1, ":", panjang_list_angka_baru)
28         print("Selisih panjang ke", j+1, ":", selisih_panjang)
29         print("=====")
```

Penjelasan mengenai kode diatas:

- 1) Pertama, Impor modul random untuk menghasilkan angka acak.
- 2) Selanjutnya, gunakan loop for sebanyak 10 kali dengan menggunakan range(10) dan menyimpan setiap iterasinya ke dalam variabel j.
- 3) Di dalam loop for, buat list kosong bernama list\_angka.
- 4) Selanjutnya, menggunakan loop for lagi untuk menghasilkan 100 angka acak dengan menggunakan fungsi randint() dari modul random dan menambahkannya ke dalam list\_angka.
- 5) Setelah itu, cetak list\_angka dan memberikan label "List acak ke j"
- 6) Kemudian membuat list kosong bernama list\_angka\_baru.
- 7) Menggunakan loop for bertujuan untuk mengecek setiap angka dalam list\_angka dan memasukkan hanya angka yang habis dibagi 3 ke dalam list\_angka\_baru.
- 8) Kemudian, cetak list\_angka\_baru dan memberikan label "List baru ke j".
- 9) Selanjutnya, gunakan fungsi len() untuk menghitung panjang dari list\_angka dan list\_angka\_baru masing-masing dan menyimpannya ke dalam variabel panjang\_list\_angka dan panjang\_list\_angka\_baru.
- 10) Kita menghitung selisih panjang kedua list menggunakan fungsi abs() dan menyimpannya ke dalam variabel selisih\_panjang.

11) Terakhir, kita mencetak panjang kedua list dan selisih panjang dengan memberikan label yang sesuai dengan setiap iterasi dan membuat garis pemisah antar setiap iterasi menggunakan tanda “=” sebanyak 40 kali.

Kode diatas akan menghasilkan output:

```
List acak ke 1 : [37, 77, 15, 34, 37, 65, 9, 80, 5, 46, 81, 90, 24, 60, 45, 0, 9, 97, 31, 45, 40, 36, 79, 90, 31, 61, 81, 73, 39, 66, 84, 66, 6, 55, 96, 18, 93, 38, 83, 4, 75, 57, 55, 41, 71, 48, 25, 11, 99, 78, 69, 86, 3, 84, 63, 32, 69, 53, 76, 76, 85, 48, 28, 17, 90, 86, 6, 0, 93, 32, 90, 33, 60, 2, 14, 36, 62, 66, 1, 9, 15, 81, 88, 25, 69, 99, 64, 75, 96, 34, 88, 46, 61, 92, 31, 8, 86, 69, 13, 12, 32]
List baru ke 1 : [15, 9, 81, 90, 24, 60, 45, 0, 9, 45, 36, 81, 39, 66, 84, 66, 6, 96, 18, 93, 75, 57, 48, 99, 78, 69, 3, 84, 63, 68, 48, 9, 0, 60, 93, 90, 33, 60, 36, 66, 9, 15, 81, 69, 99, 75, 96, 69, 12]
Panjang list angka ke 1 : 100
Panjang list angka baru ke 1 : 48
Selisih panjang ke 1 : 52
=====
List acak ke 2 : [30, 93, 94, 68, 53, 16, 93, 8, 30, 5, 72, 12, 30, 86, 97, 49, 82, 15, 76, 97, 47, 47, 49, 20, 74, 71, 93, 65, 27, 44, 7, 9, 53, 19, 41, 30, 17, 73, 96, 60, 64, 3, 61, 89, 96, 17, 20, 26, 55, 34, 72, 16, 91, 9, 65, 18, 36, 76, 51, 30, 30, 85, 58, 72, 26, 12, 2, 7, 22, 77, 87, 95, 16, 44, 88, 5, 62, 43, 1, 55, 29, 69, 91, 56, 33, 75, 7, 97, 91, 71, 77, 93, 96, 80, 10, 13, 28, 87, 18, 39, 59, 88]
List baru ke 2 : [30, 93, 93, 30, 72, 12, 30, 15, 93, 27, 30, 96, 60, 3, 96, 72, 9, 18, 36, 51, 30, 30, 72, 12, 27, 87, 33, 75, 93, 96, 8, 7, 18, 39]
Panjang list angka ke 2 : 100
Panjang list angka baru ke 2 : 33
Selisih panjang ke 2 : 67
=====
List acak ke 3 : [36, 62, 79, 23, 36, 63, 81, 10, 96, 26, 56, 96, 2, 37, 52, 43, 5, 84, 81, 98, 36, 58, 58, 71, 53, 16, 82, 54, 44, 61, 7, 4, 18, 53, 46, 48, 93, 99, 61, 63, 30, 35, 30, 79, 61, 14, 85, 60, 34, 0, 91, 77, 8, 77, 95, 2, 83, 59, 89, 97, 55, 13, 40, 73, 11, 38, 9, 2, 80, 79, 15, 9, 2, 64, 96, 7, 55, 12, 73, 10, 33, 22, 15, 61, 71, 64, 36, 28, 16, 99, 85, 41, 45, 67, 29, 49, 86, 93, 27, 94, 90, 99]
List baru ke 3 : [36, 36, 63, 81, 96, 96, 84, 81, 36, 54, 18, 48, 93, 99, 63, 30, 30, 60, 0, 15, 9, 96, 12, 33, 15, 36, 99, 45, 93, 27, 9, 0, 99]
Panjang list angka ke 3 : 100
Panjang list angka baru ke 3 : 32
Selisih panjang ke 3 : 68
=====
List acak ke 4 : [34, 41, 67, 53, 94, 71, 20, 7, 53, 13, 98, 24, 88, 31, 68, 35, 91, 13, 13, 10, 98, 37, 6, 54, 11, 72, 10, 39, 29, 28, 7, 8, 36, 40, 5, 9, 24, 53, 92, 28, 46, 34, 80, 54, 16, 66, 80, 25, 79, 30, 92, 58, 99, 24, 62, 19, 86, 92, 33, 50, 28, 42, 82, 46, 4, 79, 5, 9, 40, 38, 37, 39, 50, 85, 20, 44, 98, 78, 95, 39, 20, 0, 88, 51, 93, 61, 2, 10, 40, 69, 82, 45, 98, 21, 56, 4, 13, 98, 8, 56, 100, 61]
List baru ke 4 : [24, 6, 54, 72, 39, 78, 36, 9, 24, 54, 66, 30, 99, 24, 33, 42, 39, 78, 39, 0, 51, 93, 69, 45, 21]
Panjang list angka ke 4 : 100
Panjang list angka baru ke 4 : 25
Selisih panjang ke 4 : 75
=====
List acak ke 5 : [61, 27, 40, 92, 99, 10, 58, 80, 28, 32, 44, 80, 35, 67, 90, 24, 42, 85, 76, 20, 73, 100, 50, 19, 49, 50, 23, 50, 46, 57, 92, 56, 6, 92, 0, 20, 50, 36, 97, 23, 40, 40, 43, 11, 90, 1, 96, 56, 25, 65, 26, 32, 68, 99, 83, 17, 54, 6, 11, 82, 20, 99, 33, 21, 89, 2, 5, 91, 82, 98, 36, 11, 77, 5, 61, 13, 85, 15, 59, 90, 63, 97, 34, 93, 74, 89, 20, 29, 30, 70, 37, 6, 53, 53, 88, 86, 94, 13, 29, 22, 71]
List baru ke 5 : [27, 99, 90, 24, 42, 57, 6, 0, 36, 90, 96, 99, 54, 6, 99, 33, 21, 36, 15, 90, 63, 93, 30, 6]
Panjang list angka ke 5 : 100
Panjang list angka baru ke 5 : 24
Selisih panjang ke 5 : 76
=====
```

```
=====
List acak ke 6 : [26, 91, 84, 34, 1, 88, 90, 63, 98, 64, 10, 97, 41, 32, 85, 32, 7, 43, 71, 38, 31, 17, 96, 23, 43, 33, 65, 75, 50, 45, 6, 8, 6, 5, 24, 23, 31, 56, 22, 57, 38, 33, 43, 92, 51, 22, 5, 63, 53, 91, 56, 79, 94, 16, 5, 55, 65, 33, 10, 98, 59, 77, 60, 30, 81, 21, 39, 99, 17, 63, 49, 25, 99, 82, 54, 62, 29, 6, 16, 16, 83, 70, 27, 26, 86, 24, 52, 92, 25, 57, 30, 6, 30, 35, 86, 0, 60, 56, 15, 49, 86]
List baru ke 6 : [84, 90, 63, 96, 33, 75, 45, 6, 24, 57, 33, 51, 63, 33, 60, 30, 81, 21, 39, 99, 63, 99, 54, 6, 27, 24, 57, 30, 6, 30, 0, 60, 15]
Panjang list angka ke 6 : 100
Panjang list angka baru ke 6 : 33
Selisih panjang ke 6 : 67
=====
List acak ke 7 : [88, 53, 85, 9, 99, 82, 44, 44, 64, 65, 53, 40, 83, 12, 50, 84, 87, 53, 97, 92, 13, 39, 18, 82, 69, 57, 0, 18, 61, 95, 1, 7, 35, 53, 2, 27, 1, 66, 87, 46, 91, 9, 84, 27, 51, 69, 11, 47, 82, 25, 31, 81, 75, 89, 4, 22, 14, 26, 44, 12, 87, 100, 47, 39, 62, 1, 15, 75, 9, 10, 16, 86, 48, 35, 9, 55, 50, 85, 29, 30, 25, 14, 95, 57, 3, 97, 45, 94, 45, 3, 53, 53, 78, 32, 0, 41, 21, 46, 43, 74, 83]
List baru ke 7 : [9, 99, 48, 12, 84, 87, 39, 18, 69, 57, 0, 18, 27, 66, 87, 9, 84, 27, 51, 69, 81, 75, 12, 87, 39, 15, 75, 9, 48, 9, 30, 5, 7, 3, 45, 45, 3, 78, 0, 21]
Panjang list angka ke 7 : 100
Panjang list angka baru ke 7 : 39
Selisih panjang ke 7 : 61
=====
List acak ke 8 : [77, 34, 66, 74, 37, 87, 17, 32, 40, 68, 12, 72, 76, 61, 4, 44, 78, 39, 89, 74, 6, 47, 89, 63, 39, 74, 77, 100, 39, 12, 1, 9, 18, 32, 40, 41, 0, 50, 76, 86, 51, 53, 41, 98, 62, 37, 38, 27, 11, 65, 33, 55, 33, 17, 73, 38, 9, 35, 66, 76, 80, 44, 82, 24, 27, 99, 6, 9, 88, 22, 17, 84, 26, 30, 44, 52, 31, 57, 51, 82, 46, 4, 32, 20, 35, 94, 58, 55, 71, 10, 11, 16, 97, 13, 87, 29, 19, 50, 35, 95, 78, 40]
List baru ke 8 : [66, 87, 12, 72, 78, 39, 6, 63, 39, 39, 12, 18, 0, 51, 27, 33, 33, 9, 66, 24, 27, 99, 69, 84, 30, 57, 51, 87, 78]
Panjang list angka ke 8 : 100
Panjang list angka baru ke 8 : 29
Selisih panjang ke 8 : 71
=====
List acak ke 9 : [30, 77, 64, 65, 21, 78, 79, 30, 32, 37, 18, 46, 26, 1, 50, 68, 59, 98, 98, 44, 42, 31, 88, 26, 50, 59, 81, 45, 60, 27, 8, 6, 89, 15, 96, 47, 12, 22, 22, 18, 94, 74, 12, 9, 63, 67, 92, 82, 21, 47, 64, 52, 0, 0, 59, 72, 83, 9, 100, 60, 97, 35, 57, 75, 66, 15, 1, 5, 10, 63, 40, 95, 67, 47, 76, 97, 90, 58, 41, 88, 25, 44, 43, 70, 80, 40, 18, 51, 81, 16, 61, 17, 45, 79, 52, 62, 83, 32, 28, 87, 14, 86]
List baru ke 9 : [30, 21, 78, 30, 18, 42, 81, 45, 60, 27, 15, 96, 12, 18, 12, 9, 63, 21, 0, 0, 72, 9, 60, 57, 75, 66, 15, 15, 63, 90, 18, 51, 81, 45, 87]
Panjang list angka ke 9 : 100
Panjang list angka baru ke 9 : 35
Selisih panjang ke 9 : 65
=====
List acak ke 10 : [0, 85, 42, 2, 14, 63, 29, 47, 64, 27, 26, 80, 71, 0, 4, 50, 17, 89, 1, 42, 25, 22, 10, 52, 38, 41, 99, 57, 50, 15, 30, 33, 93, 76, 54, 35, 22, 37, 95, 12, 53, 91, 27, 4, 33, 30, 86, 81, 20, 9, 92, 10, 1, 26, 10, 23, 35, 12, 58, 72, 52, 1, 74, 9, 6, 57, 63, 99, 5, 17, 23, 71, 84, 98, 90, 73, 46, 20, 34, 21, 6, 23, 93, 58, 79, 35, 40, 41, 29, 11, 14, 28, 49, 5, 100, 64, 32, 28, 10, 72]
List baru ke 10 : [0, 42, 63, 27, 0, 42, 99, 57, 15, 30, 33, 93, 54, 12, 27, 33, 30, 81, 9, 12, 72, 9, 6, 57, 63, 99, 84, 90, 21, 6, 93, 7, 2]
Panjang list angka ke 10 : 100
Panjang list angka baru ke 10 : 32
Selisih panjang ke 10 : 68
=====
```



## 5. Menemukan rata-rata aritmatika.

Untuk menemukan rata-rata aritmatika dari `list_angka`, dapat menggunakan fungsi `sum()` untuk menjumlahkan seluruh elemen dalam list dan kemudian membaginya dengan panjang list (`len()`).

Berikut contoh implementasi kode yang digunakan:

```
1 # Menentukan rata-rata aritmatika dari list_angka
2 # Menggunakan fungsi sum() untuk menjumlahkan seluruh elemen
3 # Kemudian membaginya dengan panjang list
4
5 import random
6
7 list_angka = []
8
9 for i in range(100):
10     angka = random.randint(0, 100)
11     list_angka.append(angka)
12
13 print("List acak:", list_angka)
14
15 rata_rata = sum(list_angka) / len(list_angka)
16
17 print("Rata-rata:", rata_rata)
```

-Aryasatya-

Penjelasan mengenai kode diatas:

- 1) Pertama-tama, impor modul `random` untuk menghasilkan angka acak.
- 2) Selanjutnya, buat list kosong bernama `list_angka`.
- 3) Menggunakan loop `for` untuk menghasilkan 100 angka acak dengan menggunakan fungsi `randint()` dari modul `random` dan menambahkannya ke dalam `list_angka`.
- 4) Setelah itu cetak `list_angka`.
- 5) Menggunakan fungsi `sum()` untuk menjumlahkan seluruh elemen dalam `list_angka` dan kemudian membaginya dengan panjang list menggunakan `len()` untuk menemukan rata-rata aritmatika.
- 6) Terakhir, jangan lupa untuk mencetak rata-rata aritmatika tersebut.

Kode diatas akan menghasilkan output:

```
List acak: [94, 24, 36, 16, 53, 55, 96, 85, 1, 17, 56, 94, 74, 37, 100, 29, 22, 12, 27, 18, 36, 66, 5, 100, 1, 54, 58, 89, 90, 66, 86, 37,
25, 20, 56, 18, 14, 6, 85, 65, 91, 3, 26, 90, 3, 33, 49, 93, 63, 25, 71, 58, 39, 57, 69, 40, 54, 62, 72, 13, 63, 67, 8, 82, 30, 59, 70, 2
5, 93, 42, 86, 82, 86, 16, 49, 81, 49, 86, 27, 55, 80, 85, 28, 21, 56, 37, 13, 20, 30, 3, 12, 13, 5, 19, 71, 45, 42, 98, 93, 88]
Rata-rata: 49.49
```



## EXERCISE WEEK 3

### 1. Menampilkan data dari HTML ke data frame agar dapat ditampilkan sesuai dengan yang diinginkan.

Untuk menampilkan data dari HTML ke dataframe, kita dapat memanfaatkan library pandas untuk membaca tabel dari sebuah halaman web HTML. Disini kita mencobanya dengan link web HTML <https://pbpython.com/pandas-html-table.html>. Berikut implementasi kode yang digunakan:

```
1 import pandas as pd
2
3 # URL yang akan dibaca
4 url = 'https://pbpython.com/pandas-html-table.html'
5
6 # membaca tabel dari URL
7 tables = pd.read_html(url)
8
9 # Loop melalui semua tabel dan print DataFrame
10 for i, df in enumerate(tables):
11     print(f'Table {i}:')
12     print(df)
```

-Aryasatya-

Penjelasan mengenai kode script diatas:

- 1) Pertama, import library pandas sebagai pd untuk membaca tabel dari sebuah halaman web HTML.
- 2) Kemudian, URL halaman yang ingin dibaca sebagai variabel url.
- 3) Selanjutnya, metode read\_html() pada pandas digunakan untuk membaca semua tabel dari halaman web yang ditentukan oleh url itu sendiri. Hasil dari metode read\_html() adalah list dari pandas dataframe, dimana setiap dataframe merepresentasikan sebuah tabel pada halaman web.
- 4) Baris selanjutnya, loop dilakukan setiap dataframe dalam list 'tables'. Selama loop, indeks diatur sebagai variabel 'i', dan setiap data frame akan dicetak menggunakan fungsi print()
- 5) Dengan demikian kode tersebut akan membaca tabel dari halaman web HTML.
  - Perlu diingat bahwa metode read\_html() pada library panda ini mungkin menghasilkan beberapa kesalahan atau tidak dapat mengenali tabel pada halaman web yang kompleks atau tidak terstruktur dengan baik.

Kode diatas akan menghasilkan output:

Table 0:

Unnamed: 0	Year	Office	GOP	DFL	Others
0	2018	Governor	42.4%	53.9%	3.7%
1	2018	Senator	36.2%	60.3%	3.4%
2	2018	Senator	42.4%	53.0%	4.6%
3	2016	President	44.9%	46.4%	8.6%
4	2014	Governor	44.5%	50.1%	5.4%

Table 1:

Unnamed: 0	Year	Office	GOP	DFL	Others
0	2018	Governor	42.4	53.9	3.7
1	2018	Senator	36.2	60.3	3.4
2	2018	Senator	42.4	53.0	4.6
3	2016	President	44.9	46.4	8.6
4	2014	Governor	44.5	50.1	5.4

Table 2:

Unnamed: 0	Year	Nominal GDP(in bil. US-Dollar)	\
0	2020		20234.0
1	2019		21439.0
2	2018		20580.2
3	2017		19519.4
4	2016		18715.0

	GDP per capita(in US-Dollar)	GDP growth(real)	Inflation rate(in percent)	\
0	57589	-5.9	0.62	
1	64674	2.2	1.80	
2	62869	3.0	2.40	
3	60000	2.3	2.10	
4	57878	1.7	1.30	

	Unemployment (in percent)	Budget balance(in % of GDP)[107]	\
0	11.1	NaN	
1	3.5	-4.6	
2	3.9	-3.8	
3	4.4	-3.4	
4	4.9	-3.1	

	Government debt held by public(in % of GDP)[108]	\
0	79.9	
1	78.9	
2	77.8	
3	76.1	
4	76.4	

	Current account balance(in % of GDP)
0	NaN
1	-2.5
2	-2.4
3	-2.3
4	-2.3

Selain dapat membaca isi HTML secara keseluruhan, kita juga dapat mengeluarkan beberapa output yang dapat dihasilkan dari data HTML seperti contoh kita ingin mengetahui jumlah tabel yang ditemukan dalam web, maka kita bisa menggunakan kode seperti berikut:

```

1 # URL yang akan dibaca
2 url = 'https://pbpython.com/pandas-html-table.html'
3
4 # membaca tabel dari URL
5 tables = pd.read_html(url)
6
7 # jumlah tabel yang ditemukan pada halaman web
8 num_tables = len(tables)
9
10 print(f"Number of tables found: {num_tables}")
11

```

Number of tables found: 3

Output yang dihasilkan bahwa ditemukan jumlah tabel pada halaman web sebanyak 3.

Lalu, jika kita ingin menampilkan hanya kolom-kolom pada tabel pertama, maka kita gunakan kode seperti berikut:

```
1 # URL yang akan dibaca
2 url = 'https://pbpython.com/pandas-html-table.html'
3
4 # membaca tabel dari URL
5 tables = pd.read_html(url)
6
7 # menampilkan kolom-kolom pada tabel pertama
8 columns = tables[0].columns
9
10 print(f"Columns in the first table: {columns}")
11
```

Columns in the first table: Index(['Unnamed: 0', 'Year', 'Office', 'GOP', 'DFL', 'Others'], dtype='object')

-Aryasatya-

Terakhir, jika kita ingin mengetahui berapa jumlah baris pada tabel kedua dapat menggunakan kode seperti berikut:

```
1 # URL yang akan dibaca
2 url = 'https://pbpython.com/pandas-html-table.html'
3
4 # membaca tabel dari URL
5 tables = pd.read_html(url)
6
7 # menampilkan jumlah baris pada tabel kedua
8 num_rows = tables[1].shape[0]
9
10 print(f"Number of rows in the second table: {num_rows}")
11
```

Number of rows in the second table: 5

-Aryasatya-

Dengan memanfaatkan variabel `num_rows = tables[1].shape[0]` kita dapat mencetak hasil dan ditemukan jumlah baris pada tabel kedua adalah 5.

## 2. Menampilkan data tabel dari PDF ke dataframe agar dapat ditampilkan sesuai dengan yang diinginkan.

Untuk menampilkan data table dari PDF, kita dapat memanfaatkan library 'tabula' untuk membaca tabel dari file PDF "Housing\_data.pdf". Berikut implementasi kode yang digunakan:

```
1 import tabula as tb
2
3 df = tb.read_pdf("Housing_data.pdf", pages="all")
4 print(df)
```

-Aryasatya-

Penjelasan mengenai kode diatas:

- 1) Pada baris pertama, import library tabula sebagai tb. Kemudian, metode read.pdf() digunakan untuk membaca semua halaman pada file PDF dan parameter pages diatur all yang artinya untuk membaca semua halaman.
- 2) Hasil metode read\_pdf() adalah list dari pandas DataFrame, di mana setiap data frame mempresentasikan sebuah table yang ditemukan pada halaman PDF. Selanjutnya, dataframe disimpan ke dalam variabel df.
- 3) Sedangkan pada baris terakhir, variabel df dicetak menggunakan fungsi print().
  - Perlu diingat bahwa metode read.pdf() pada library tabula memiliki keterbatasan dan dapat terbatas pada file PDF yang memiliki format dan struktur yang kompleks.

Kode diatas akan menghasilkan output:

```
[ 0.17004 12.5 7.87 0 0.524 6.004 85.9 6.5921 5 311
0 0.22489 12.5 7.87 0 0.524 6.377 94.3 6.3467 5 311
1 0.11747 12.5 7.87 0 0.524 6.009 82.9 6.2267 5 311
2 0.09378 12.5 7.87 0 0.524 5.889 39.0 5.4509 5 311, 15.2 386.71 17.1 18.9
0 15.2 392.52 20.45 15.0
1 15.2 396.90 13.27 18.9
2 15.2 390.50 15.71 21.7]
```

-Aryasatya-

### 3. Beautiful Soup dan Web Scrapping

Beautiful Soup adalah pustaka Python untuk melakukan web scrapping atau ekstraksi data dari halaman web. Library ini juga memungkinkan kita untuk mengekstrak data dari file HTML dan XML dengan mudah dan efisien. Beberapa fitur dari Beautiful Soup:

- 1) Parsing HTML dan XML: Beautiful Soup dapat mem-parsing dan mengurai dokumen HTML atau XML. Library ini dapat mengubah dokumen yang sulit dibaca dan tidak teratur menjadi struktur data yang dapat diakses dan dimanipulasi.
- 2) Parsing HTML dan XML: Beautiful Soup dapat mem-parsing dan mengurai dokumen HTML atau XML. Library ini dapat mengubah dokumen yang sulit dibaca dan tidak teratur menjadi struktur data yang dapat diakses dan dimanipulasi.
- 3) Parsing HTML dan XML: Beautiful Soup dapat mem-parsing dan mengurai dokumen HTML atau XML. Library ini dapat mengubah dokumen yang sulit dibaca dan tidak teratur menjadi struktur data yang dapat diakses dan dimanipulasi.
- 4) Dukungan Unicode: Beautiful Soup mendukung pengolahan karakter Unicode, sehingga memudahkan dalam mengekstrak data dari halaman web dalam berbagai bahasa.
- 5) Dukungan Unicode: Beautiful Soup mendukung pengolahan karakter Unicode, sehingga memudahkan dalam mengekstrak data dari halaman web dalam berbagai bahasa.

Dengan segala fitur yang dimilikinya, Beautiful Soup menjadi salah satu library yang paling populer dan sering digunakan untuk melakukan web scraping dan pengolahan data dari halaman web. Kelebihan library beautiful soup adalah bekerja lebih baik untuk data berukuran besar, mudah dipelajari bagi pemula, dan menggunakan deteksi encoding otomatis. Satu-satunya kelemahan yang dimiliki library ini adalah lebih lambat jika dibandingkan dengan library lxml pada Python.

Lalu apa itu web scrapping?

Web scrapping masih ada kaitan dengan BeautifulSoup. Beautiful soup merupakan library yang dapat membantu untuk mem-parse dokumen HTML atau XML yang diambil dari website, yang nantinya akan digunakan dalam proses web scrapping.

Web scrapping sendiri dapat didefinisikan sebagai teknik pengambilan data otomatis dari halaman web dengan menggunakan program atau script komputer. Teknik ini mengambil data yang ada di halaman web dan menyimpannya dalam bentuk yang dapat diolah seperti file CSV atau bentuk database. Web scraping umumnya dilakukan dengan menggunakan library atau tool yang tersedia dalam berbagai bahasa pemrograman, seperti Python, PHP, dan JavaScript. Library atau tool ini memudahkan dalam mengekstrak data dari halaman web agar tidak copy paste secara manual.

Dalam melakukan web scrapping, penting untuk diketahui dan memperhatikan etika dan hukum yang berlaku seperti menghormati hak cipta, menghindari pengambilan data yang terlalu banyak/dalam jumlah terlalu besar dan menyebabkan beban pada sever, dan memperhatikan kebijakan privasi, pastikan data yang diambil tidak melanggar privasi pengguna.

Jadi, dapat disimpulkan web scrapping merupakan sebuah teknik yang sangat berguna dalam pengolahan atau analisis data.