Introduction to Big Data

- Developed by Dr. Keungoui KIM
- https://awekim.github.io/portfolio/

Lecture 3. Data structure

List

```
my_list = [1, 1, 4, 5, "Yes", [2, 4]]
my_list
```

```
list("Yes")
```

```
type(my_list) # type of data structure
```

```
len(my_list) # size of the list
```

```
my_list.append("HGU")
my_list
```

```
len(my_list) # size of the list
```

```
my_list[0]
```

```
#list to dataframe
import pandas as pd
pd.DataFrame(my_list, columns=['values'])
```

packing and unpacking

```
num_list = [1, 2, 3, 4]
num_list
```

```
aa, bb, cc, dd = num_list
```

```
print(aa, bb, cc, dd)
```

```
aa, bb, cc = num_list
print(aa, bb, cc)
```

```
list_list = [[1,2,3,4],[5,6,7,8]]
alist, blist = list_list
print(alist)
print(blist)
```

```
midterm_score = [84, 23, 92, 17, 88]
assignment_score = [88, 87, 90, 88, 89]
final_score = [71, 99, 78, 89, 82]
```

```
total_score = [midterm_score, assignment_score, final_score]
total_score
```

```
total_score[2][3]
```

```
# final_score[3] = 22
# total_score

#list to dataframe
import pandas as pd
total_score_df = pd.DataFrame(total_score,
                              columns=['John','James','BoA','Tanaka','Kim'])
total_score_df.index = ['Midterm','Assignment','FinalExam']
total_score_df

#total_score_df['score_type'] = ['Midterm','ASsignment','FinalExam']
#total_score_df

import pandas as pd
total_score_df = pd.DataFrame(total_score,
                              columns=['John','James','BoA','Tanaka','Kim'])
total_score_df.index = ['Midterm','Assignment','FinalExam']
total_score_df
```

Tuple

```
# list
list_sample = [1, 3, 5]
print(type(list_sample))
print(list_sample)

list_sample[0] = 'one'
list_sample.append(7)
list_sample

# tuple
tuple_sample = (1, 3, 5)
print(type(tuple_sample))
print(tuple_sample)

tuple_sample[0] = 'one'

tuple_sample.append(7)
```

Dictionary

```
my_dict = {'Name':'Kim',
           'Nationality':'Korea',
           'Age':21,
           'BigData':'A+'}
print(my_dict)

type(my_dict)

len(my_dict)

my_dict.keys()

my_dict.values()

my_dict['Name']
```

```
my_dict

my_dict['gender']='M'
print(my_dict)

my_dict['Gender']='F'
print(my_dict)

my_dict['BigData']='B+'
print(my_dict)

del my_dict['gender']
print(my_dict)

my_dict = {'Name':'Kim', 'Nationality':'Korea', 'Age':21, 'BigData':'A+'}
my_dict['gender']='M'
my_dict['BigData']='B+'
print(my_dict)

# Nested dictionary - multiple sample
people = {1: {'name': 'John', 'age': '27', 'sex': 'Male'},
          2: {'name': 'Marie', 'age': '22', 'sex': 'Female'}}
print(people)

# Dictionary to dataframe
import pandas as pd
pd.DataFrame.from_dict(people, orient='index')
```

Review

- My favorite baseball player is Ohtani Shohei. He's one of the most popular Japanese baseball player in the world. He's a pitcher, designated hitter, and outfielder for LA Angels of MLB.

- 1) Refering to information above, create a dictionary called ohtani, which includes the information of name, nationality, position, team, and league.

- 2) Write down a Python code that checks the number of elements in ohtani.

- 3) Ohtani's age is 28. Add this information to ohtani.

- 4) Write down a Python code that checks ohtani's number of positions

```
# 1
ohtani = {'name':'Ohtani Shohei','nationality':'Japan',
          'position':['pitcher','designated hitter','outfielder'],
          'team':'LA Angels','league':'MLB'}
ohtani

# 2
len(ohtani)

# 3
ohtani['age'] = 28
ohtani

# 4
len(ohtani['position'])
```

Review

```
Mydict = {'Team':['GSW','LAL','CHI','BKY'],
          'Win':[32,28,22,31],
          'Lose':[3,7,13,4],
          'KeyPlayer':['Curry','Lebron','Jordan','Durant']}
```

```
Mydict.keys()
```

```
Mydict['KeyPlayer'][1]
```

```
len(Mydict)
```

```
Mydict
```

```
Mydict['US'] = ['Yes','Yes','Yes','Yes']
Mydict
```

```
Mydict['Win'] = [0,0,0,0]
Mydict
```

Indexing & Slicing

List's indexing and slicing

```
my_list = ["Yes","No", 1, 1, 4, 5, [2, 4]]
my_list
```

```
my_list[0]
```

```
my_list[0] == my_list[-0]
```

```
my_list[1]
```

```
my_list[0] + my_list[1]
```

```
my_list[2] + my_list[3]
```

```
my_list[0:2]
```

```
my_list[:2]
```

```
my_list[2:]
```

```
my_list[:]
```

```
my_list[-1]
```

```
my_list[-1][0]
```

```
ChristmasGift = ['MacBook', 'iPhone', 'Bitcoin',
                 'Tesla', '1', 2, '3', 4]

print(ChristmasGift[:2])
print(ChristmasGift[4:6])
print(ChristmasGift[0:3])
print(ChristmasGift[:])
```

Conditional Statement

```
a = 1
if a == 1:
  print("Hello")
  print("I'm Kim")

a=2
if a ==1:
  print("Hello")
  print("I'm Kim")

a=2
if a ==1:
  print("Hello")
print("I'm Kim")

3%2

2%2

0==0

# odd or even detector

x = 1
if x%2 == 0:
  print( x, "is even.")
else:
  print( x, "is odd.")

# Grade detector

score = int(input("What do you want to expect from this class? "))
if score >= 90:
  print("Your grade is A.")
elif score >=80:
  print("Your grade is B.")
elif score >=70:
  print("Your grade is C.")
else:
  print("Your grade is D.")

temperature = int(input("What is your body temperature? "))

if temperature >= 37.5:
  print("Need PCR test.")
else:
  print("You are safe.")
```

```
x=11
if x>0 and x<10:
  print("x is greater than 0 and less than 10.")

x=11
if x>0 or x<10:
  print("x is greater than 0 and less than 10.")
```

Iterative Statement

while

```
i = 0
while i < 10:
  print(i)
  i = i + 1
print('last value: ', i)

value = 100
while 0 < value:
  value = value-5
  print(value)
print("last value=",value)

value = 100
while 0 < value:
  print(value)
  value = value-5
print("last value=",value)

while True:
  print("You can exit this loop by pressing Ctrl+c.")

# infinite loop
i = 1
while i < 10:
  print(i)
  i=i-1
print("last i", i)
```

for

```
for i in [1,2,3,4,5]:
  print(i)

for i in range(5):
  print(i)

for i in range(10):
  print(i)

for i in range(1,10):
  print(i)
```

```
for i in range(1,10,-1):
  print(i)

for i in ['Apple','Samsung','Google','LG']:
  print(i)
  print('i')

for i in ['Apple','Samsung','Google','LG']:
  print(i)
print('i')
```

iteration in list

Given list_temp = [0, 1, 2, 3, 4], create a new list called sqrt_temp, which converts all elements into squared form [0, 1, 4, 9 , 16]. (HINT: use .append() method)

```
sqrt_temp = []
for i in list_temp:
   sqrt_temp.append(i**2)
  #  print(sqrt_temp)
sqrt_temp

sqrt_temp = [i ** 2 for i in list_temp]
sqrt_temp

names = ['kim','lee','ki','park','li','son']
[x.upper() for x in names if len(x)>=3]

tuples = [(1,2,3), (4,5,6), (7,8,9)]

[x for tup in tuples for x in tup]

temp = []
for tup in tuples:
  for x in tup:
    temp.append(x)
temp

items = [0, 5, 10, 15, 20]
for i in range(len(items)):
  print(i, items[i])
```