**Міністерство освіти і науки України**

**Національний технічний університет України**

**«Київський політехнічний інститут імені Ігоря Сікорського»**

**Факультет інформатики та обчислювальної техніки**

**Кафедра обчислювальної техніки**

**Лабораторна робота №3**

з дисципліни

«Аналіз даних з використанням мови Python»

Виконала: Викладач:

студентка групи ІП-03 Тимофєєва Ю.

Величко Ірина Євгенівна

Київ 2022

Завдання

Варіант №4 (frogs.csv)

Створити програму, яка за даними файлу відповідно до варіантів

лабораторної №2, виконує наступні завдання:

1. Виділити один зі стовпців (на вибір) з файлу як об’єкт Series, виділити з нього підмасив. Задати назви індексів цього об’єкту. Виділити підмасиви за допомогою прямої та непрямої індексацій.

2. До об’єкту DataFrame, в який записано вміст файлу, додати новий стовпець, що є результатом операцій над іншими стовпцями. Також продемонструвати додавання та видалення рядків, видалення стовпців.

3. Встановити один зі стовпців індексом. Визначити основні статистичні характеристики та типи даних всіх стовпців. Змінити тип даних для одного з стовпців. Згрупувати дані за одним зі стовпців, застосувати кілька агрегуючих функцій, виділити підмасив за певними ознаками.

4. Створити декілька власних об’єктів DataFrame за такою ж тематикою, що й файл. Наприклад, якщо тема файлу – жаби, можна створити об’єкти, що містять розміри жаб, вагу, стать, кількість особин в популяції і т.д. Використати описані в теоретичних відомостях параметри методів merge та concat для різних видів злиття та об’єднання даних цих об’єктів.

Код програми:

import pandas as pd

import numpy as np

data = pd.read\_csv('frogs.csv')

col1 = pd.Series(data['NoOfPools'])

a = col1[4:8]

index = ['I', 'II', 'III', 'IV']

a.index = index

print(a)

print()

print(data.iloc[:4, :4])

print()

print(data.loc[:5,:'altitude'])

data['ratioAD'] = data['altitude']/(data['distance']) #add column

print(data)

newrow = pd.DataFrame([[121, 1098, 1666, 2380]],columns=['northing', 'easting', 'altitude','distance'])

data = pd.concat([data, newrow]) #add row

print(data)

data.drop(1) #delete row

del data['Unnamed: 0'] #delete column

print(data)

data = data.set\_index(['NoOfSites'])

print(data)

print(data.describe())

print(data.dtypes)

data['easting'] = data['easting'].astype('float64')

print(data.dtypes)

print(data.groupby('pres.abs').mean())

print(data.groupby('pres.abs').agg({'avrain': max,'NoOfPools': np.median}))

print(data[(data.distance>600)&(data.distance<1000)]) #subarray

'''New DataFrame'''

body\_length = np.random.randint(7, 15,size=(30))

body\_weight = np.random.randint(50, 100, size=(30))

gend\_dist = np.random.randint(0, 2, size=(30))

frogs = pd.DataFrame({'Length':body\_length, 'Weight':body\_weight})

gender = pd.DataFrame({'Gender':gend\_dist})

print(frogs)

frogs = pd.concat([frogs, gender], axis=1)

print(frogs)

df2 = pd.DataFrame({'Length':np.random.randint(5, 13,size=(30)), 'NoOfPools': np.random.randint(100, 200,size=(30))})

print(pd.merge(frogs, df2, on='Length'))

Результати:











