МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

Західноукраїнський Національний Університет

Кафедра

Інформаційно обчислювальних

Систем і управління

Лабораторна робота № 2

Тема: ДОСЛІДНИЦЬКИЙ АНАЛІЗ ДАНИХ У PYTHON.

Виконав:

Ковальковський В. В.

м. Тернопіль

2023 рік

ТЕМА 2. ДОСЛІДНИЦЬКИЙ АНАЛІЗ ДАНИХ У PYTHON.

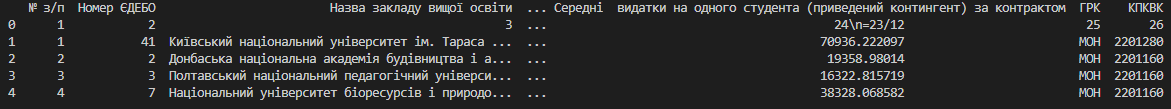
ЗАВДАННЯ



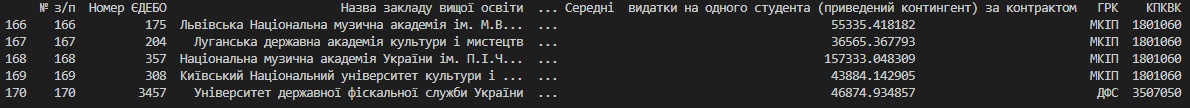


ВИЗНАЧЕННЯ НАБОРУ ДАНИХ:

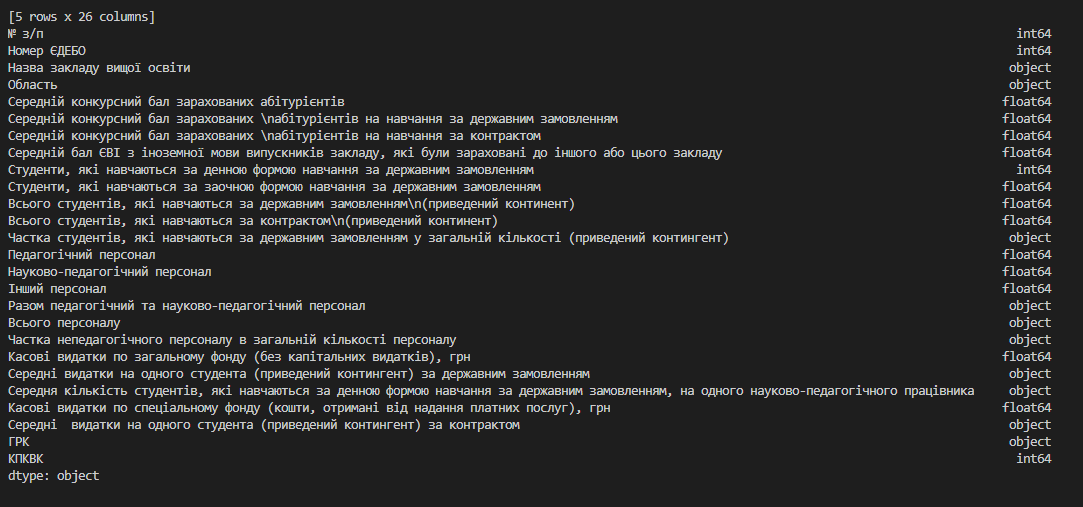






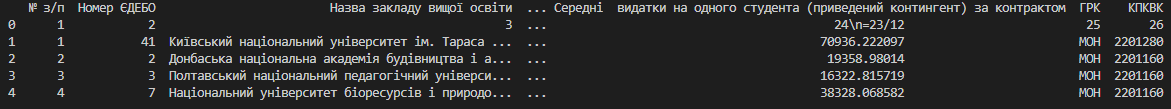


Типи даних:

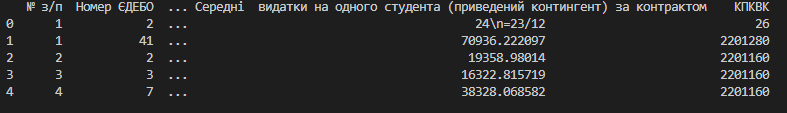


ВИДАЛЕННЯ НЕАКТУАЛЬНИХ СТОВПЦІВ

Df.head() перед видаленням:

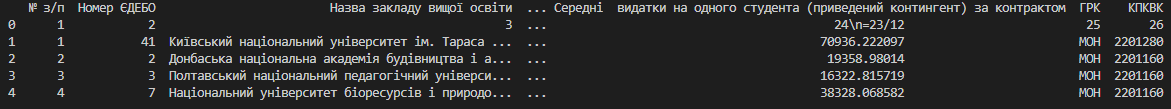


Df.head() після видалення:

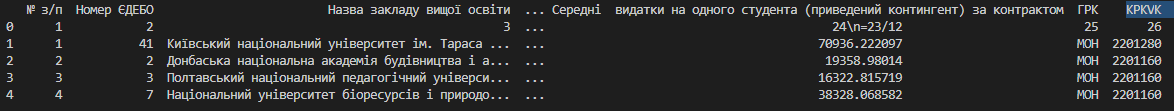


ПЕРЕЙМЕНУВАННЯ СТОВПЦІВ

Df.head() перед перейменуванням:



Df.head() після перейменування:

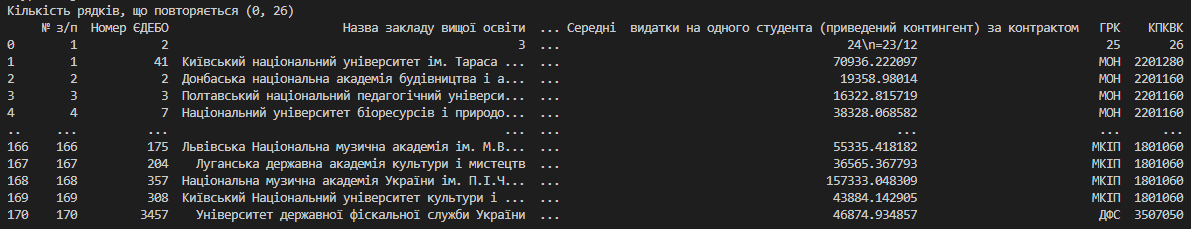


ВИДАЛЕННЯ ПОВТОРЮВАНИХ РЯДКІВ



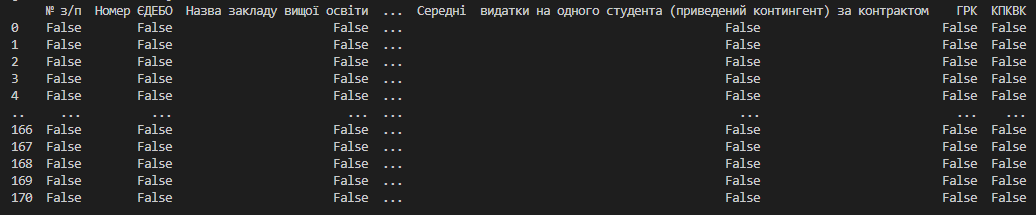






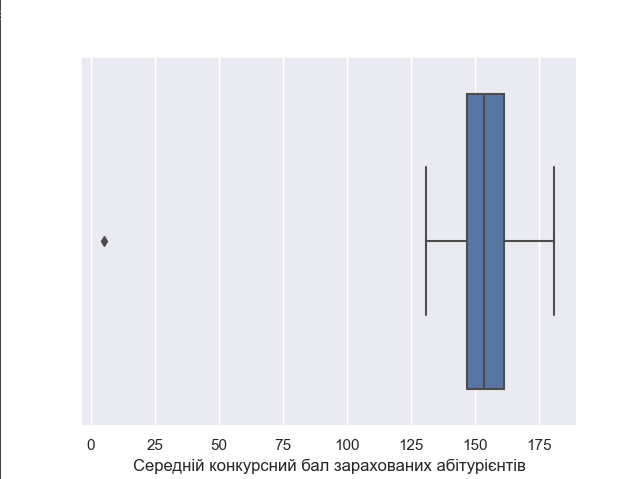
ВИДАЛЕННЯ ПРОПУЩЕНИХ АБО НУЛЬОВИХ ЗНАЧЕНЬ.

print(df.isnull())



ВИЯВЛЕННЯ ВИКИДІВ

print(sns.boxplot(x=df['Середній конкурсний бал зарахованих абітурієнтів']))



Повний лістинг:

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

sns.set(style="darkgrid")

df = pd.read\_excel('ВНЗ-2019.xlsx')

number\_donetsk = []

ser\_zarah = []

ser\_zarah\_derz = []

edbo\_number = []

state = []

for i in range(len(df)):

    if df.loc[i, "Область"] == "Донецька область":

        number\_donetsk.append(df.loc[i, "№ з/п"])

        ser\_zarah.append(df.loc[i, "Середній конкурсний бал зарахованих абітурієнтів"])

        ser\_zarah\_derz.append(df.loc[i, "Середній конкурсний бал зарахованих \nабітурієнтів на навчання за державним замовленням"])

print(ser\_zarah)

print(ser\_zarah\_derz)

t = df.drop(['ГРК'], axis = 1)

v = df.rename(columns={"КПКВК" : "KPKVK"})

print(v.head())

print(df.tail())

print(df.dtypes)

print("Кількість рядків, що повторяється", df[df.duplicated()].shape)

deleted\_df = df.drop\_duplicates()

print(deleted\_df)

print(df.isnull())

print(sns.boxplot(x=df['Середній конкурсний бал зарахованих абітурієнтів']))

plt.figure(figsize=(10, 6))

plt.plot(range(1, len(number\_donetsk) + 1), ser\_zarah, marker='o', label='Середній бал абітурієнтів')

plt.plot(range(1, len(number\_donetsk) + 1), ser\_zarah\_derz, marker='x', label='Середній бал за держзамовленням')

plt.xlabel('Номер університету')

plt.ylabel('Середній бал')

plt.title('Середній бал абітурієнтів та за держзамовленням в Донецькій області')

plt.xticks(range(1,11),number\_donetsk)

plt.legend()

plt.grid(True)

plt.show()

Висновок: під час виконання лабораторної роботи, опрацював роботу з наборами даних та можливостями взаємодії з ними.