

Міністерство освіти і науки, молоді та спорту України

Національний технічний університет України

“Київський політехнічний інститут”

Фізико-Технічний інститут

Курсова робота

Аналіз даних за допомогою Python та бібліотеки Pandas,matplotlib

**Роботу виконав:**

Студент 2 курсу ФТІ

групи  ФI-92

Плахтій Гліб

**Приймали:**

                                                       Шуміло Леонід Леонідович

Шелестов Андрій Юрійович

Київ – 2020

**Мета роботи**

Отримати досвід використання основних методів та засобів аналізу та візуалізації даних у середовищі розробки IPython Notebook на реальних наборах даних.

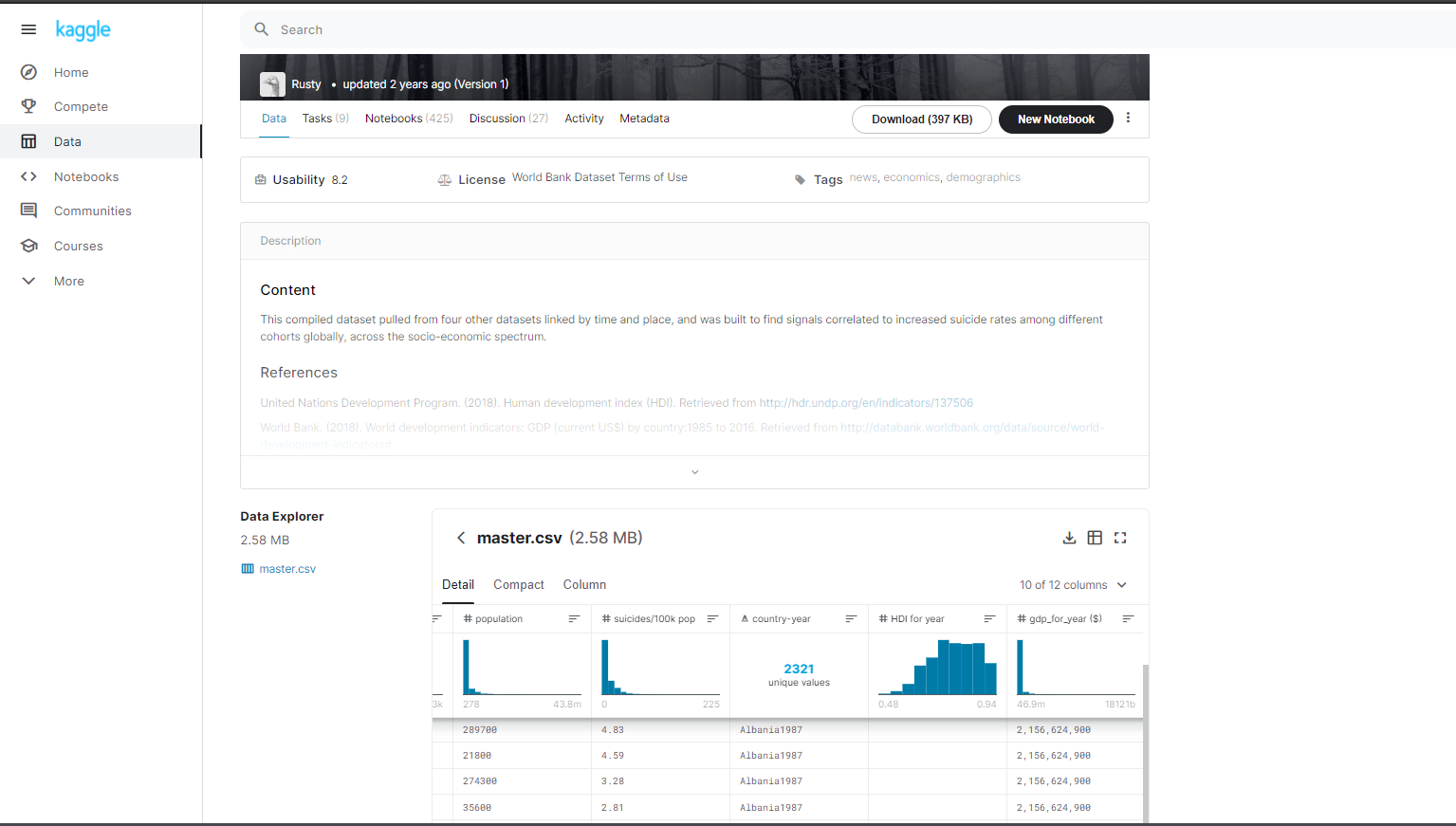
**Хід роботи**

1. Встановити anaconda, Ipython Notebook разом з anaconda та створити новий notebook python 3/2
2. Проаналізувати набір даних на наявність пропущених значень та NaN. У випадку наявності таких значень їх потрібно замінити за допомогою методу ковзного вікна.
3. Відкрити набір даних через пандас. Представити його у матричному вигляді. Відобразити 10 елементів для 4-ох ознак Виділити зеленим кольором максимальні елементи в таблиці, а червоним мінімальні для цих 10-ти елементів.
4. Реалізувати 4 функції для вибору елементів датафрейму за логічними правилами, в яких обов’язково необхідно використати такі опції:
5. Вибір елемента за парністю/не парністю індексу
6. Використати і, або та логічне заперечення, менше, більше (не обов’язково в одній функції всі одночасно)
7. Не менше ніж одна функція має використовувати loc
8. Не менше ніж одна функція має використовувати iloc

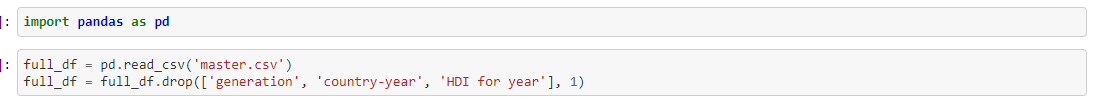
**Структура даних**

**Виконання роботи**

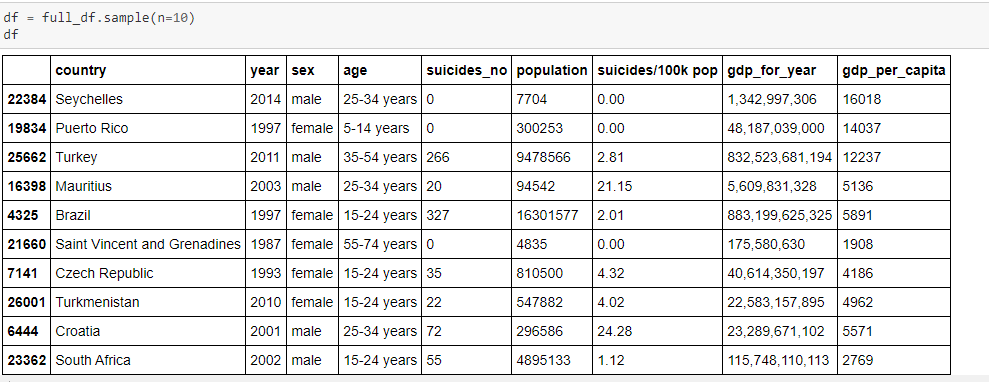
1. Обрав датасет з Kaggle про комп’ютерні ігри. Завантажив дані, встановив Anaconda3 та необхідні бібліотеки pandas, numpy , matplotlib



1. Спочатку імпортував бібліотекки pandas для майбутньої роботи з даними. Потім завантажив ці дані.



1. Взяв 10 рядків с датасету для подальшої роботи



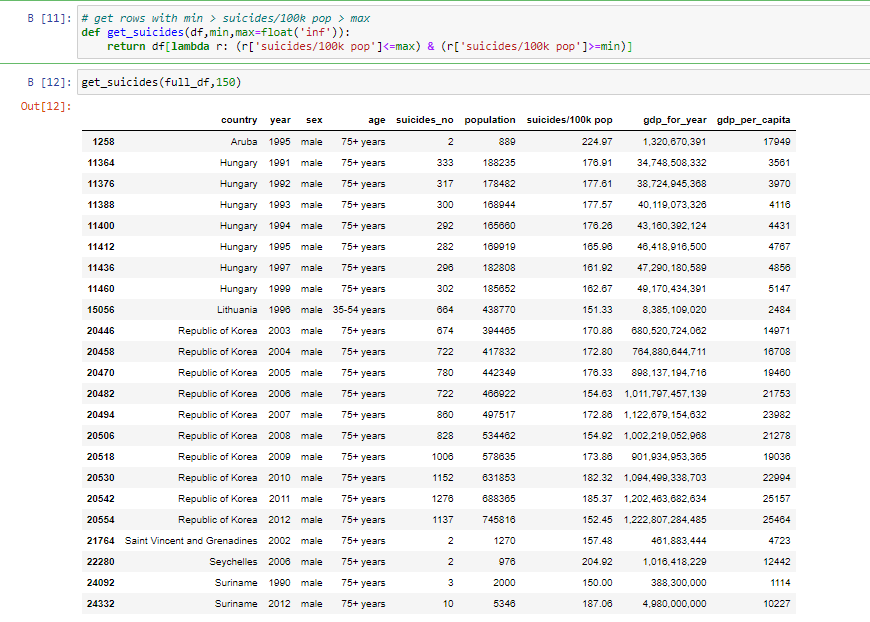
1. Створив функцію, яка виділяє найменьше і найбільше значаення з стовбчика

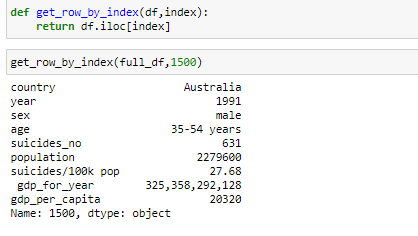


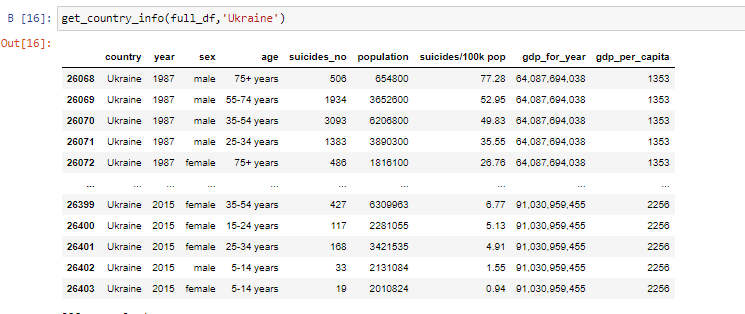
1. Створив функцію, яка виводить парні і непарні рядки



Створив функцію, яка повертає рядки в яких кількість самогубств на 10 000 знаходить в діапазоні між min i max



1. Написав функцію, яка повертає рядок за вказаним індексом 
2. Створив функцію, яка повертає рядки з вказаною країною



1. Імпортував модулі для побудови графіків



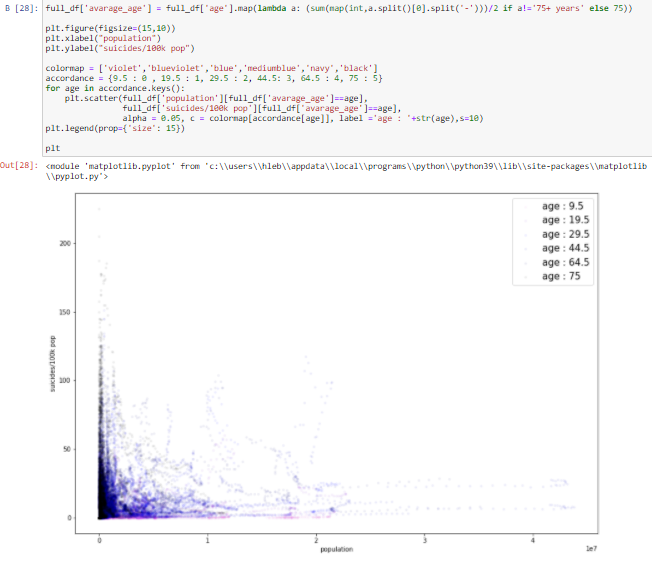
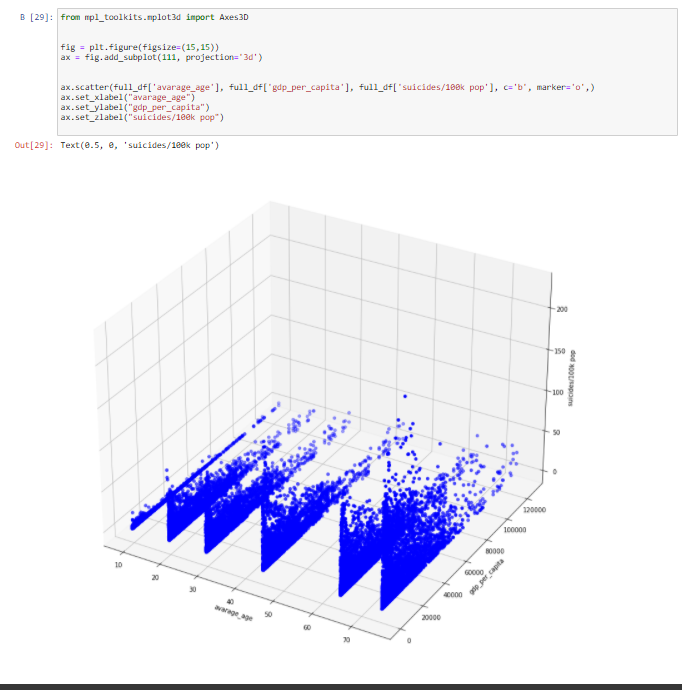
1. Побудував графік залежності кількості самогубств на 100000 людей від віку



1. Створив графік залежності кількості самогубств на 100 000 людей в залежності від ВВП на душу населення

(червоне – жінки ; сине - чоловіки)



1. Побудував графік залежності самогубств на 100 000 людей в залежності від населення ()
2. Отримав 3D графік залежності від кількості самогубств на 100 000 людей від віку і ВВП на людину 

**Висновки**

Після проходження цього курсу я навчився працювати з різними наборами даних та відображати результати у вигляді: списків, таблиць, діаграм та графіків. У цій роботі я застосовував бібліотеки pandas, numpy, matplotlib. Проаналізував датасет та зробив деякі статистичні результати. Висновки з цих даних можна практично застосувати для аналізу ринку комп’ютерних ігор для пошуку найкращої з них на будь-якій платформі.

**Посилання на джерела**

Dataset Kaggle:

https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016