

Міністерство освіти і науки України  
НТУУ «КПІ ім. Ігоря Сікорського»  
Навчально-науковий інститут атомної та теплової енергетики  
Кафедра цифрових технологій в енергетиці

Лабораторна робота №1  
з дисципліни «Вступ до інтелектуального аналізу даних»  
Тема «Завантаження даних у Python.  
Приклад блокнота Jupyter»  
Варіант №19

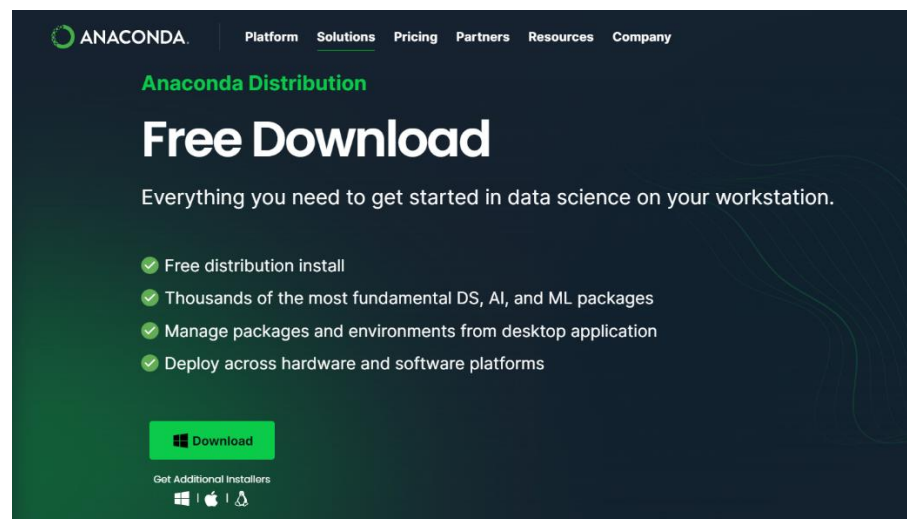
Студента 3-го курсу НН ІАТЕ гр. ТР-12  
Ковальова Олександра  
Перевірив: д.т.н., проф. Путренко В. В.

**Мета:** Ознайомитись з прикладом створення блокноту Jupyter, зчитування даних з SQLite, задачі найпростішого аналізу даних.

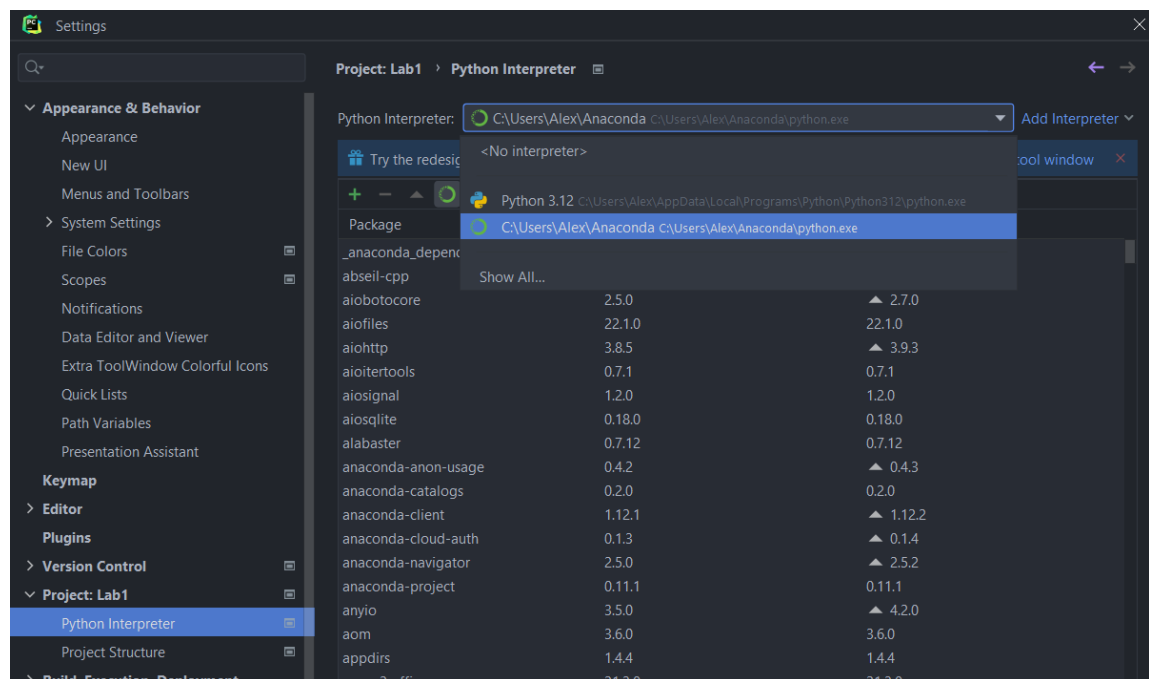
## Хід роботи

Задача – сформувати блокнот, який завантажує базу даних злаків і їх різних властивостей, а потім обчислює та відображає середній вміст цукру (у грамах) у крупах кожного виробника. Потрібно покроково ознайомитися з прикладом коду та дізнатись про функції кожного блоку.

Для початку треба встановити Anaconda – вільно та відкрито розповсюджуваний дистрибутив мов програмування Python та R. Платформа спеціалізується на «наукових обчисленнях». Використання платформи має на меті спрощення управління пакетами та їх розгортання. Версіями пакунків керує система управління пакетами Conda. Одним з рекомендованих кроків встановлення Jupyter Notebook є завантаження Anaconda.



Після завантаження та інсталяції, потрібно встановити інтерпретатор в IDE (інтегроване середовище розробки) PyCharm:



Імпортуємо пакет `sqlite3` для роботи з БД SQLite. У змінну записуємо шлях до файлу з БД:

```
import sqlite3

db_filename = 'data/cereals-1.db'
```

Під'єднуємось до БД. Отримуємо об'єкт курсор, який потрібен для роботи зі запитами:

```
conn = sqlite3.connect(db_filename)
c = conn.cursor()
```

За допомогою курсору виконуємо запит до БД – отримуємо всіх виробників злаків без повторень.

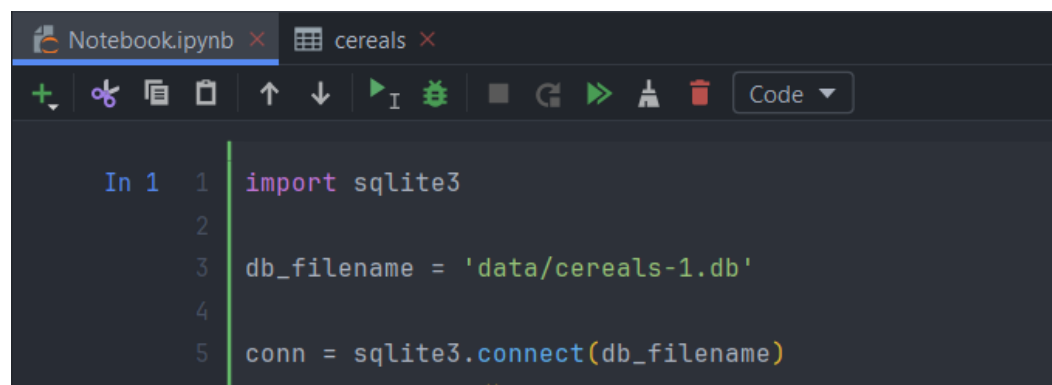
```
c.execute('''
    SELECT DISTINCT Manufacturer FROM cereals
''')
```

Для кожного виробника виконуємо запит – знаходимо середнє значення цукру в грамах. Після цього, друкуємо результат у форматі «Виробник: Число».

```
for manufacturer in c.fetchall():
    c.execute('''SELECT avg(Sugars)
        FROM cereals WHERE
        Manufacturer='{0}'
        '''.format(manufacturer[0]))

    print("{0}: {1:.2f}"
        .format(manufacturer[0], c.fetchall()[0][0]))
```

Весь код потрібно вставити в комірку у новоствореному файлі `Notebook.ipynb`.



The screenshot shows a Jupyter Notebook window with two tabs: 'Notebook.ipynb' and 'cereals'. The 'Notebook.ipynb' tab is active. The code editor shows the following code:

```
In 1 1 import sqlite3
      2
      3 db_filename = 'data/cereals-1.db'
      4
      5 conn = sqlite3.connect(db_filename)
```

## Результат виконання коду:

```
Executed at 2024.02.08 20:34:36 in 11ms  
✓ G: 7.95  
K: 7.57  
N: 1.83  
P: 8.78  
Q: 5.25  
R: 6.12  
A: 3.00
```

**Висновок:** Під час виконання лабораторної роботи були набуті практичні навички для роботи з задачами, де потрібно зробити найпростіший аналіз даних (у даному випадку – знайти середнє відносно даних з одного стовбця таблиці). Було налаштоване робоче середовище для роботи зі задачами зі сфери аналізу даних.

### Програмний код

#### *Notebook.ipynb:*

```
import sqlite3  
  
db_filename = 'data/cereals-1.db'  
  
conn = sqlite3.connect(db_filename)  
c = conn.cursor()  
  
c.execute('''  
    SELECT DISTINCT Manufacturer FROM cereals  
''')  
  
for manufacturer in c.fetchall():  
    c.execute('''SELECT avg(Sugars)  
        FROM cereals WHERE  
        Manufacturer='{0}'  
        '''.format(manufacturer[0]))  
  
    print("{0}: {1:.2f}"  
        .format(manufacturer[0], c.fetchall()[0][0]))
```