**Міністерство освіти і науки України**

**Національний технічний університет України «Київський політехнічний**

**інститут імені Ігоря Сікорського"**

**Факультет інформатики та обчислювальної техніки**

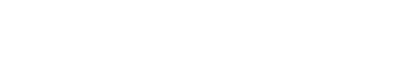
**Кафедра інформатики та програмної інженерії**

# Звіт

з лабораторної роботи № 1 з дисципліни

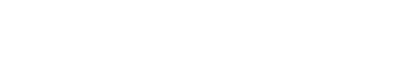
«Прикладні задачі машинного навчання»

«Введення в data science»



**Виконав** *ІП-15, Дацьо Іван*

(шифр, прізвище, ім'я, по батькові)



**Перевірив**



прізвище, ім'я, по батькові



*Нестерук А.*

Київ 2023

# Завдання:

1. На сайті http://www.ukrstat.gov.ua/ обрати дані які для Вас є цікавими, можна використати будь-який ресурс з відкритими даними.

2. Знайти математичне сподівання, медіану, моду, дисперсію, середньоквадратичне відхилення (поясніть їх зміст)

3. Візуалізувати завантажені дані за допомогою гістограми

4. Для цих даних проробити всі дії з пункту колекції Series і DataFrame бібліотеки pandas

5. Прочитати набір даних катастрофи «Титаніка»

6. Завантажити набір даних катастрофи «Титаніка» за URL-адресою

7. Переглянути рядки набору даних катастрофи «Титаніка»

8. Налаштувати назви стовпців

9. Провести простий аналіз даних

10. Побудувати гістограму віку пасажирів

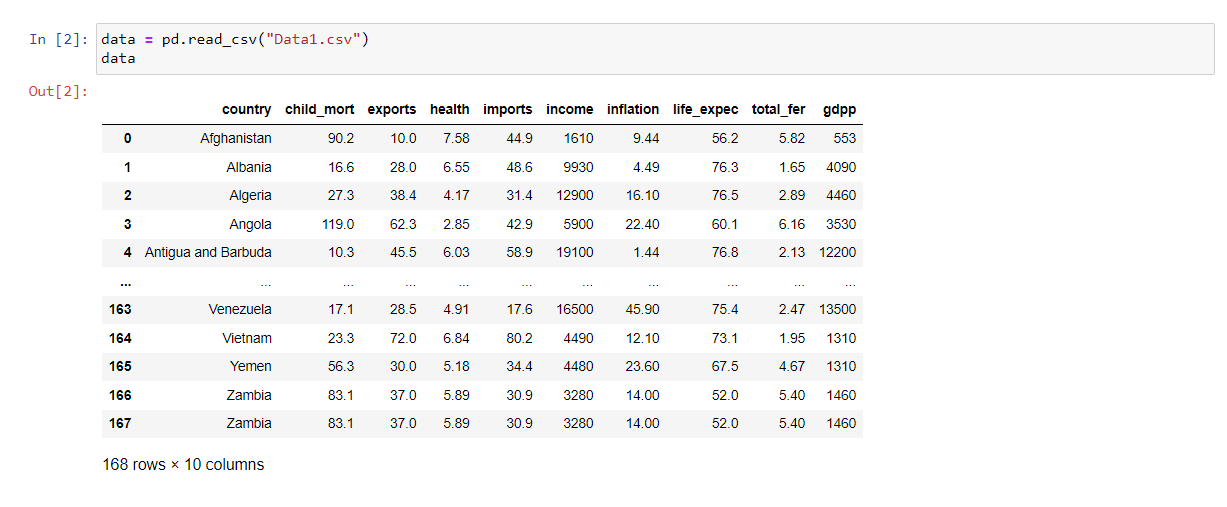
# Виконання

1. Імпортуємо необхідні пакети:

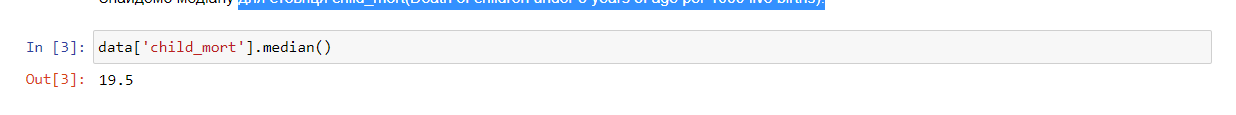


Для виконання лабораторної роботи було використано набір даних, для дослідження країн світу. Завантажимо дані та виведемо дані

:



1. Знайдемо медіану для стовпця child\_mort(Death of children under 5 years of age per 1000 live births), тобто це буде значення яке стоїть по середині даних:



Знайдемо математичне сподівання для стовпця child\_mort те, що очікується в середньому :

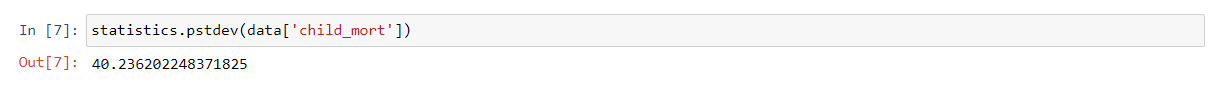
Знайдемо мода для стовпця child\_mort(Death of children under 5 years of age per 1000 live births), тобто це буде найчастіше число смертей дітей на 1000 народжувань:



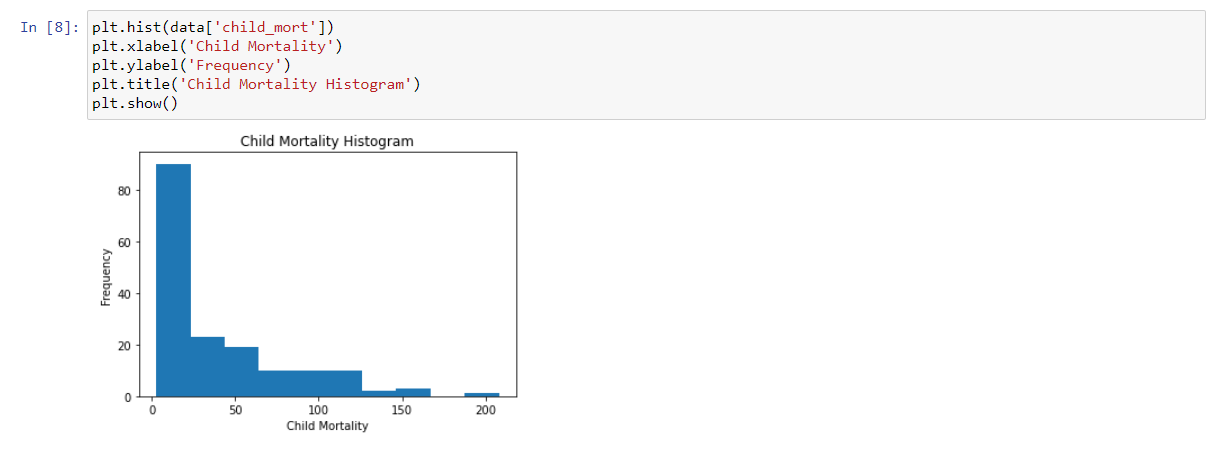
Знайдемо дисперсію (наскільки далеко значення в даних розташовані від середнього значення)



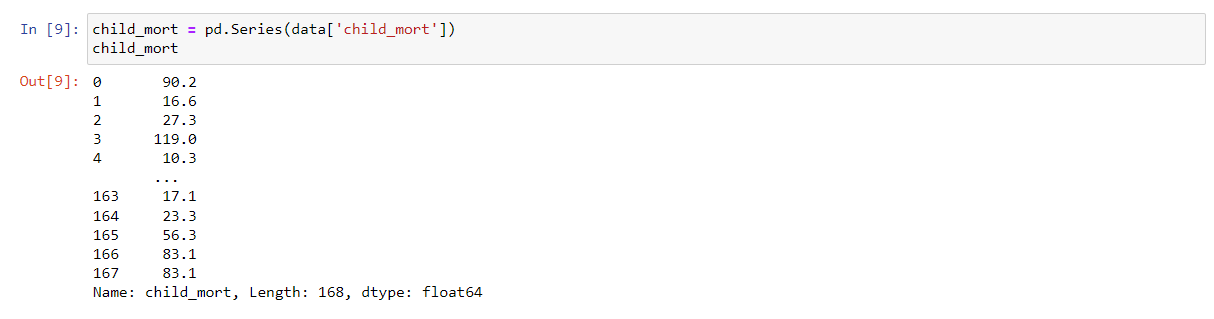
Знайдемо середньоквадратичне відхилення(наскільки сильно значення розкидані відносно середнього)



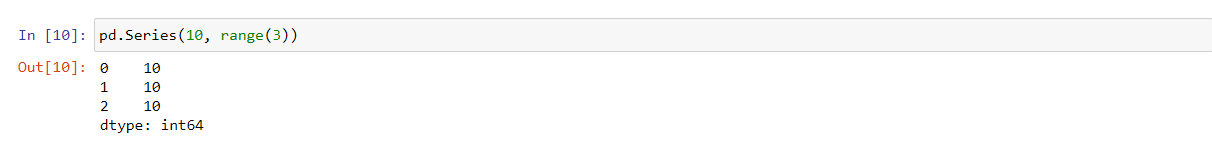
1. Візуалізуємо завантажені дані за допомогою гістограми



1. Створюємо Series на основі даниих із минулого завдання



Використання range()



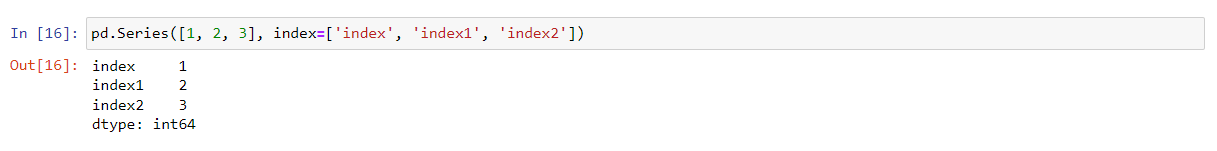
Звертання до елементу



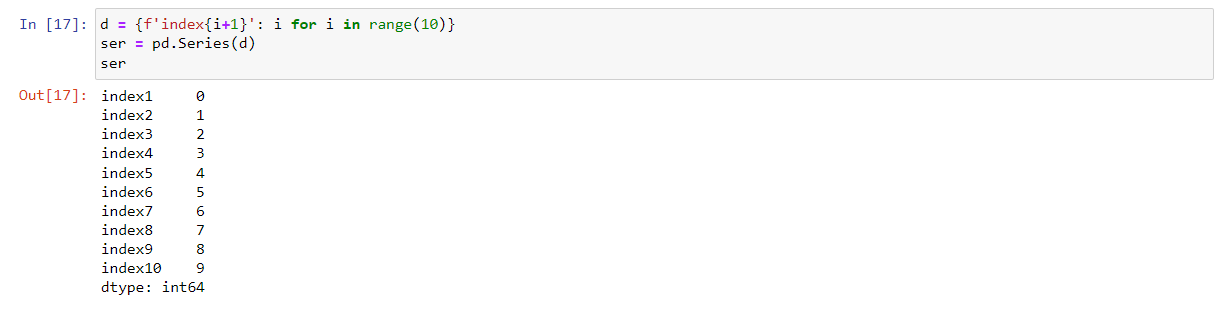
Обчислення описових статистик



Створення колекції з не типових індексів



Використання як інікалізатора словник

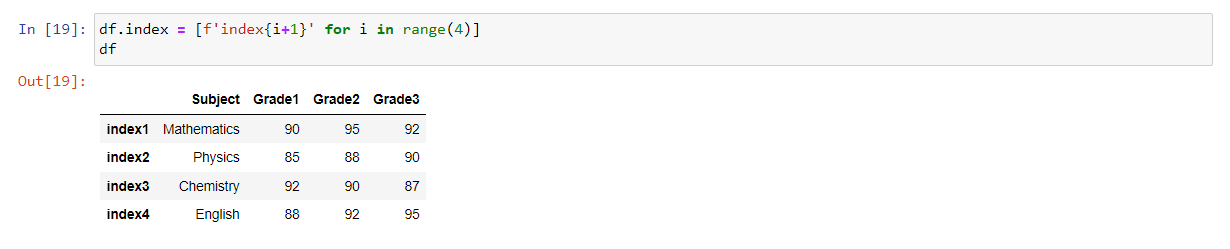


Для Dataframe:

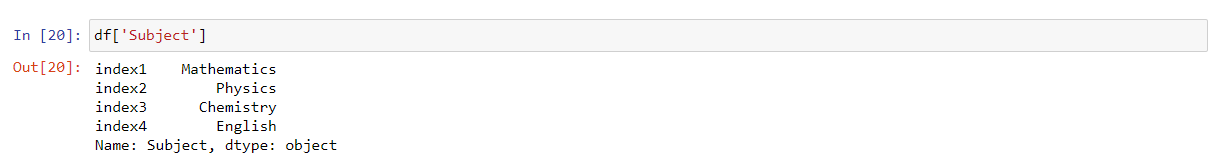
Створюємо dataframe використовуючи словник:



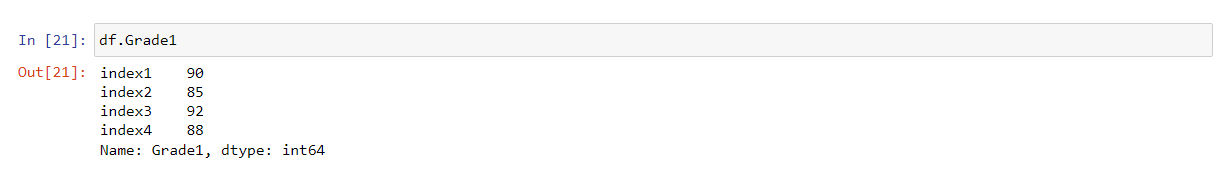
Змінимо індекси:



Звернемося до стовпця `Subject`:



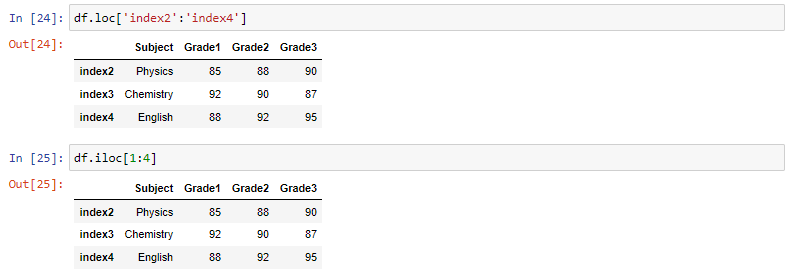
Звернемося до значень першої оцінки Grade1:



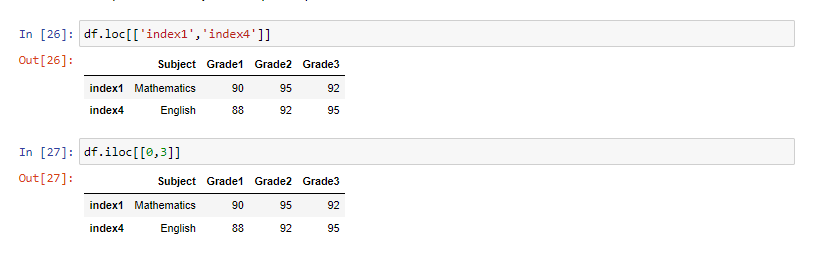
Виберемо рядки використовуючи loc та iloc:



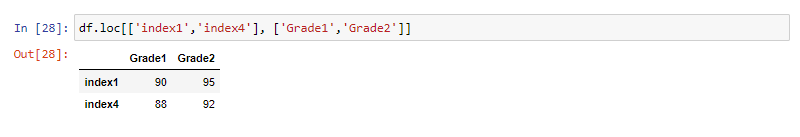
Використаємо як індекс сегмент:



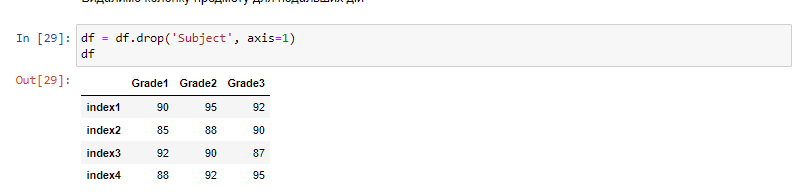
Для конкретних рядків використаємо список:



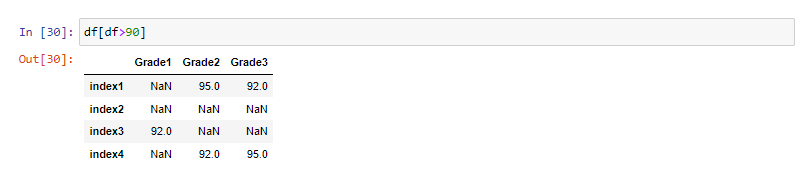
Виберемо підмножину рядків та стовпців:



Видалимо колонку предмету для подальших дій:



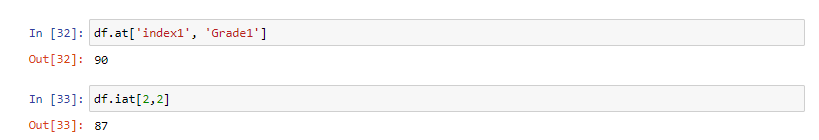
Виберемо усі значення які більші за 90:



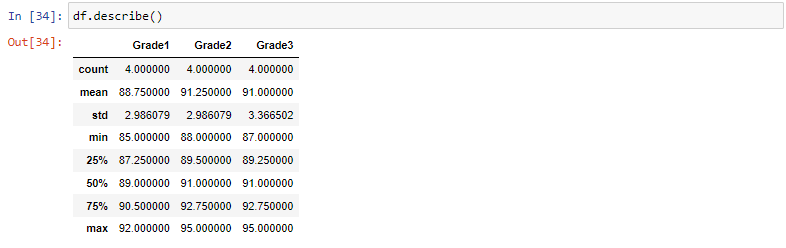
Усі значення які більші за 85 але менші за 90:



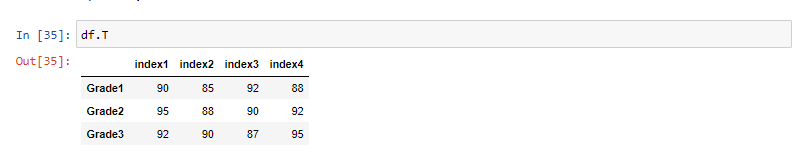
Виведення елементу по назві рядку та стовпця:



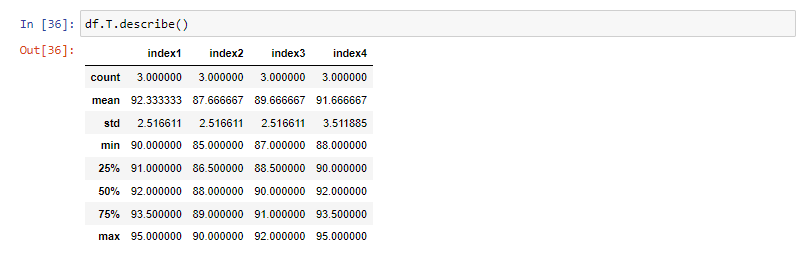
Виведемо повну описову статистику:



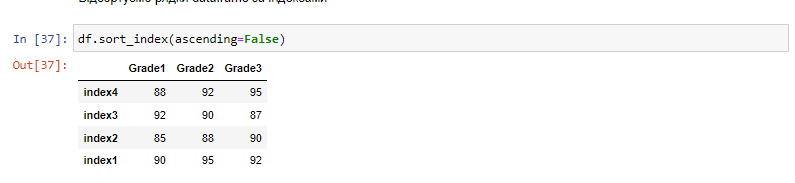
Транспонуємо даний dataframe:



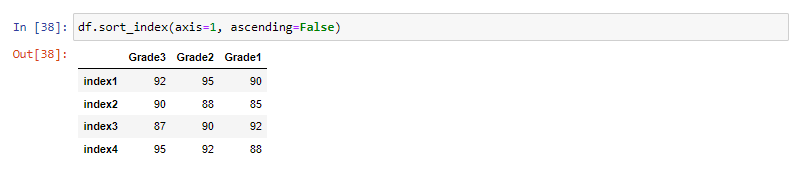
Виведемо описову статистику для транспонованого dataframe:



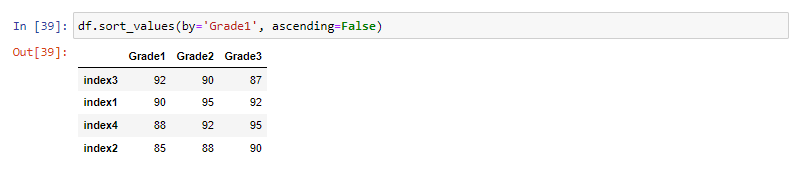
Відсортуємо рядки dataframe за індексами:



та за стовпцями:

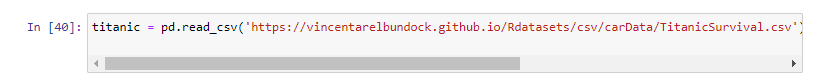


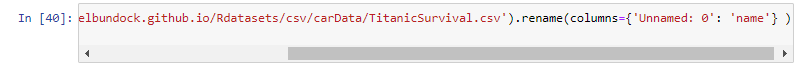
Сортуємо dataframe за значеннями Grade1:



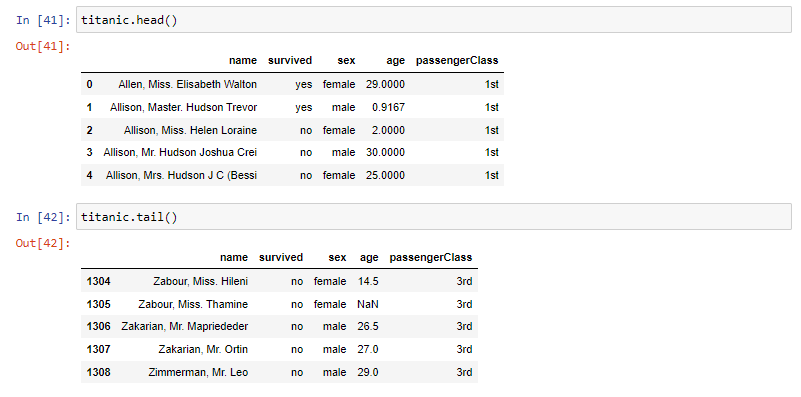
1. 6. 7.

Прочитаємо, редагуємо та завантажимо набір даних катастрофи «Титаніка»:

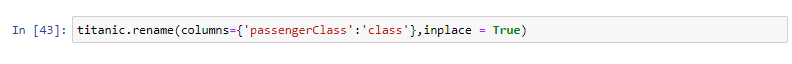




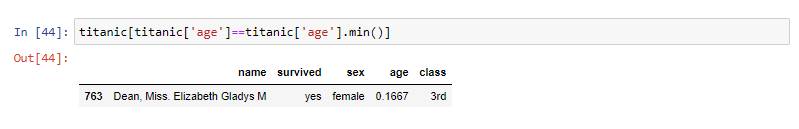
1. Виведемо дані задопомогою методів head і tail:



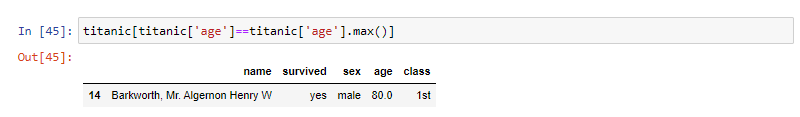
1. Скоротимо 'passengerClass' до 'class':



1. Визначимо наймолодшого пасажира що вижив:



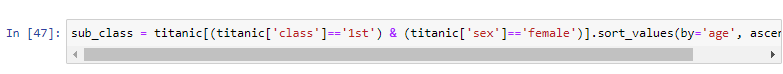
Найстаршого:

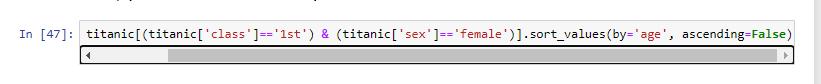


Середній вік тих хто вижив:

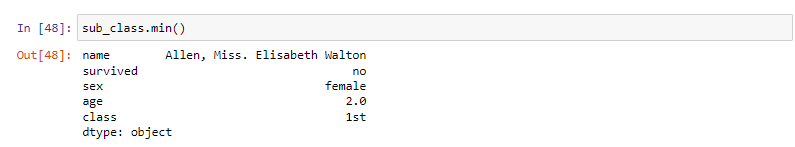


Відсортуємо всіх жінок з кают 1-го класу:

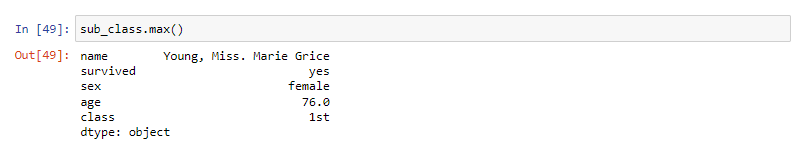




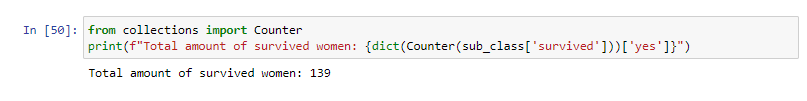
Наймолодша жінка яка вижила із першого класу:



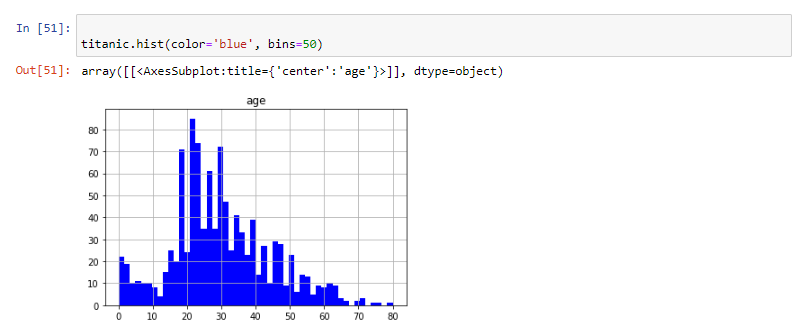
Найстарша жінка яка вижила із першого класу:



Загальна кількість жінок що вижили:



1. Побудуємо гістограму віку пасажирів:



**Висновок**

Виконавши дану лабораторну я ознайомився із такими бібліотеками як pandas та matplotlib. Були досліджено загальну інформацію про країни, а саме смерть дітей (віком до 5 років) на 1000 народжень. Для цих даних було знайдено: математичне сподівання, медіану, моду, дисперсію, середньоквадратичне відхилення. Проведена візуалізація даних та пророблені усі пункти над колекціями Series і DataFrame бібліотеки pandas. Також був проведений аналіз набору даних катастрофи лайнера “Титаніка”.