Міністерство освіти і науки України

Львівський національний університет імені Івана Франка

Факультет електроніки та комп’ютерних технологій

Кафедра системного проектування

Звіт

Про виконання лабораторної роботи №1

З курсу «Основи аналізу даних»

### **Вступ в Python для аналізу даних.**

Виконав:

Студент групи ФЕІ-42

Прізвище Ім’я

Перевірив:

Асист. Азаров І.В.

Львів 2024

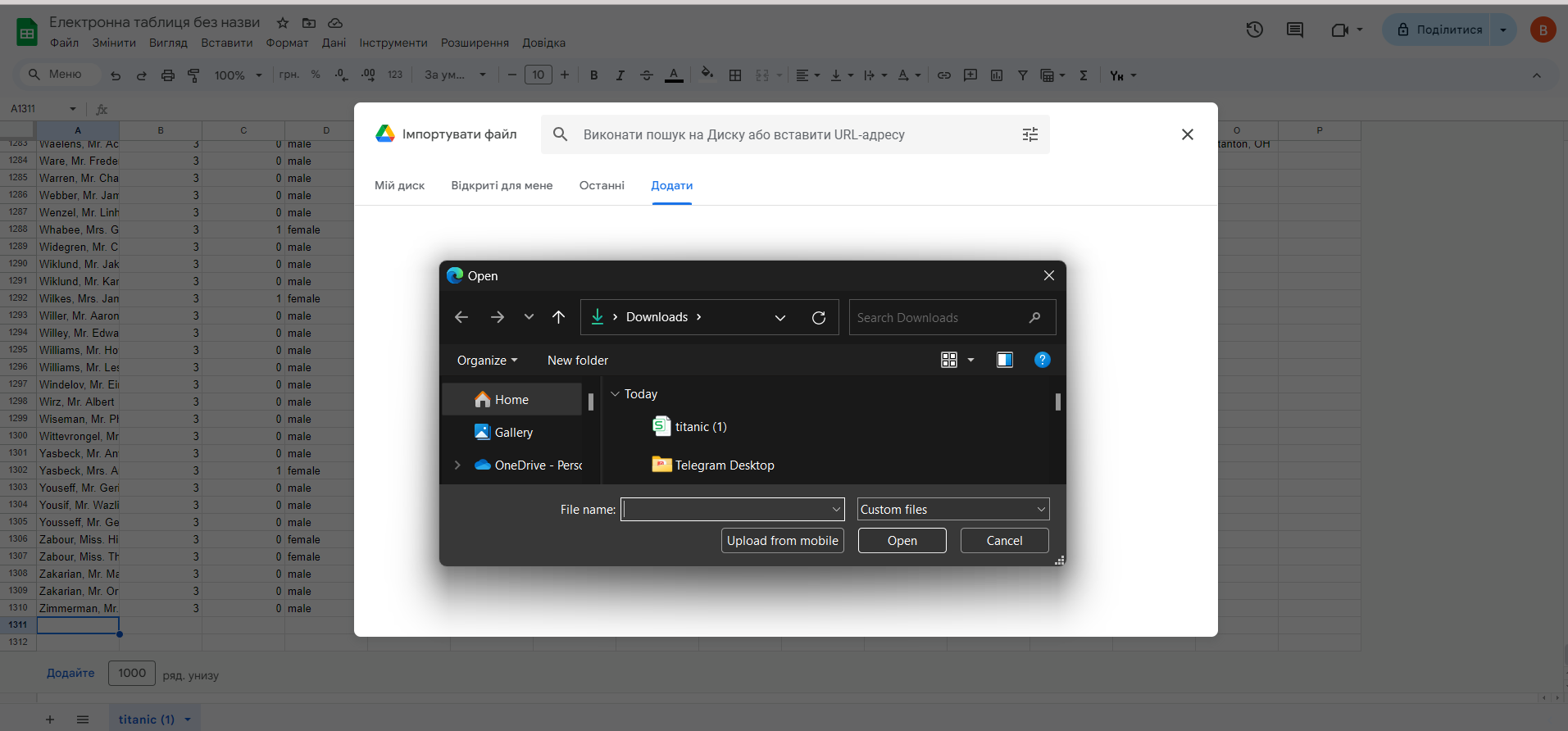
**Мета:** Ознайомитися з бібліотеками Python, що використовуються для аналізу даних.

**Хід роботи:**

Завдання

**1. Робота з таблицями**

* Імпортувати наявні дані



* Пояснити до якої шкали належать різні змінні в таблиці

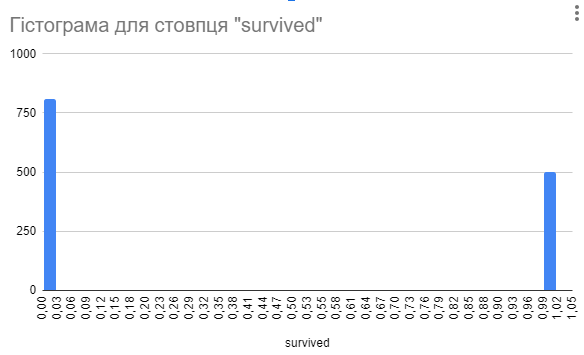
Змінні name,ticket,home.dest належать до якісних решта до кількісних

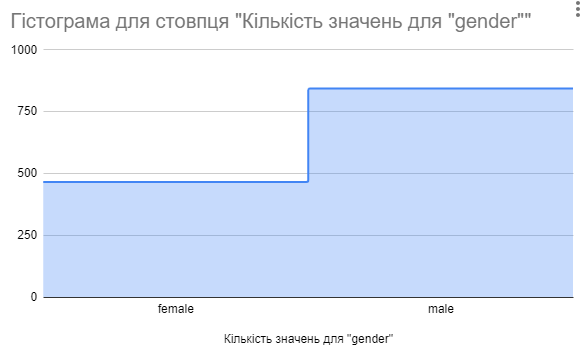
* Додайте вигадану особу з своїм прізвищем та збережіть зміни



* Побудувати гістограми для кожного стовпця

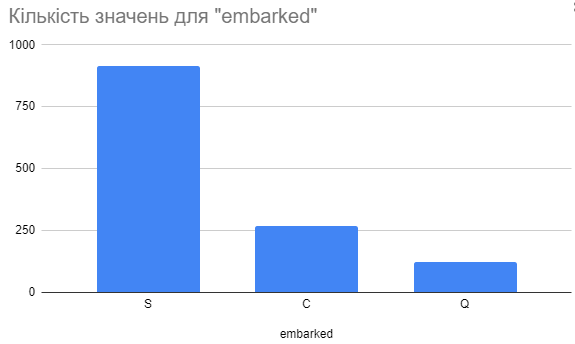








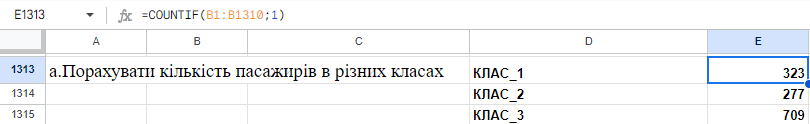




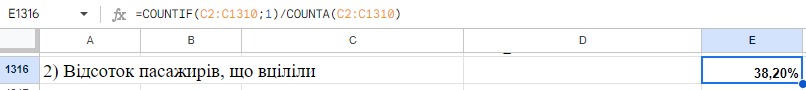


* Фільтрація даних за певними критеріями

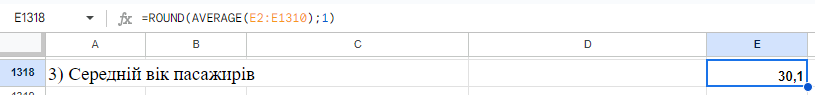
1. Порахувати кількість пасажирів в різних класах



1. Відсоток пасажирів, що вціліли

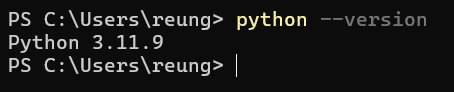


1. Середній вік пасажирів



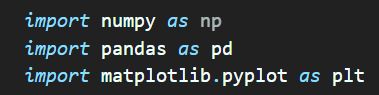
1. **Підготовка середовища**

* Встановити IDE Python



* Встановити NumPy, pandas, matlibplot

Всі бібліотеки були встановлені раніше.

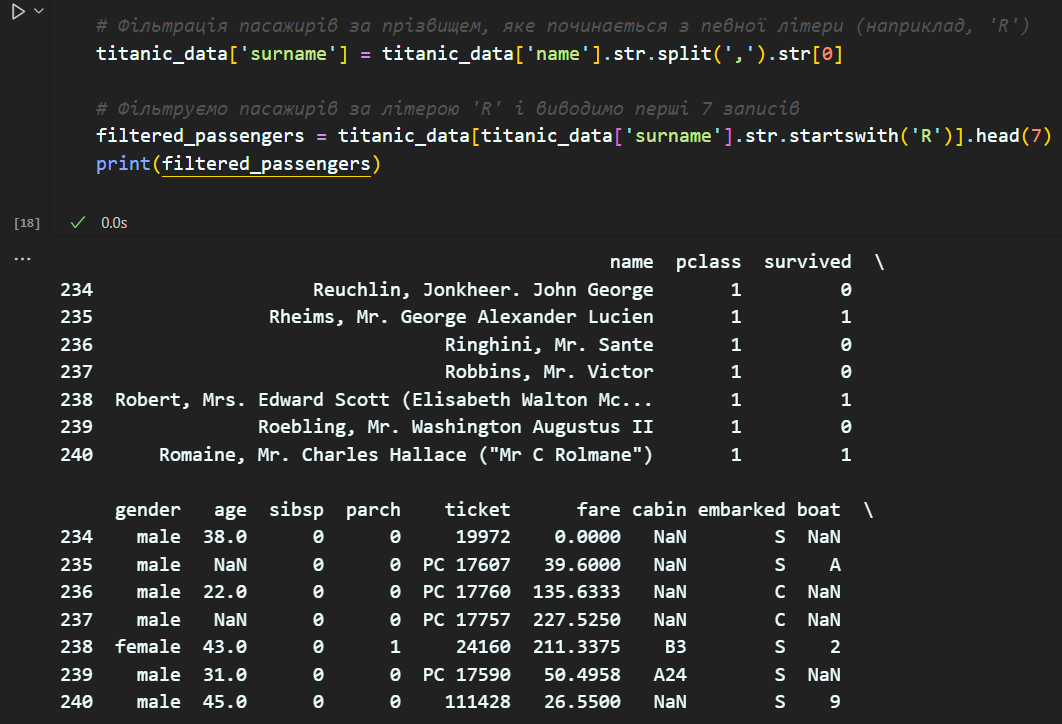


1. **Виконання базових операцій з даними за допомогою Python (завантаження, фільтрація, нормалізація):**

* Завантажити дані з CSV-файлу.

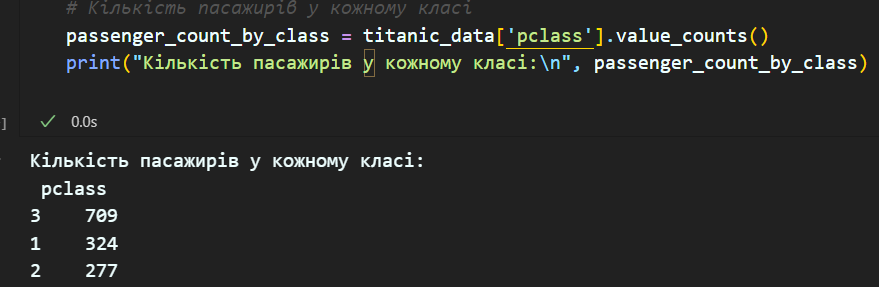


* Продемонструвати перших 7 пасажирів, перша літера прізвища яких співпадає з прізвищем автора

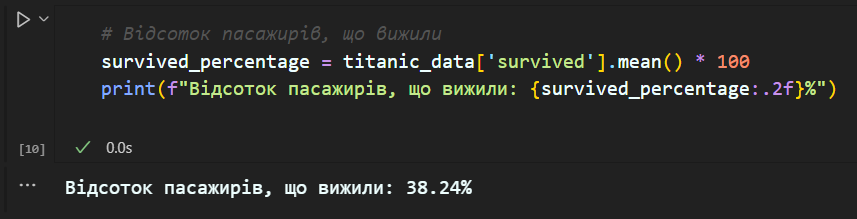


* Фільтрація даних за певними критеріями

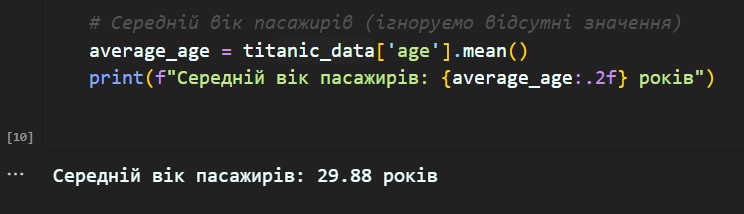
1. Порахувати кількість пасажирів в різних класах



1. Відсоток пасажирів, що вціліли

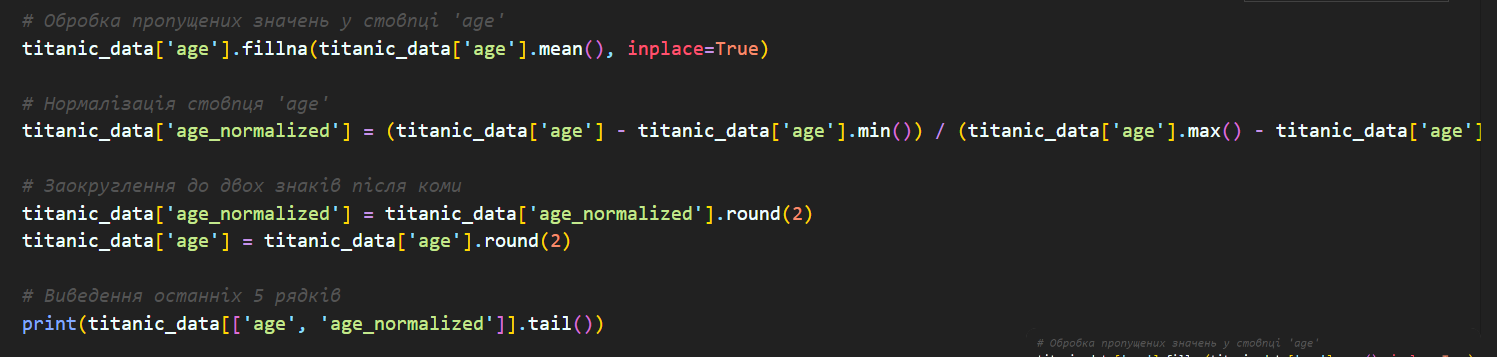


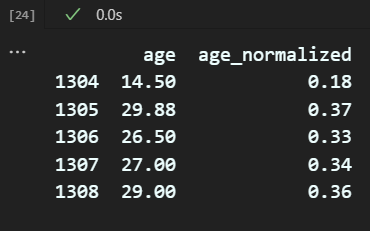
c. Середній вік пасажирів



* Нормалізація даних.

1. Виконати нормалізацію





1. Пояснити зміни в даних

Як правило, ми **нормалізуємо** дані, коли робимо якийсь тип аналізу, у якому ми маємо кілька змінних, виміряних у різних масштабах, і ми хочемо, щоб кожна зі змінних мала однаковий діапазон.

Це запобігає надмірному впливу однієї змінної, особливо якщо вона вимірюється в різних одиницях (тобто якщо одна змінна вимірюється в дюймах, а інша – у ярдах).

З іншого боку, ми зазвичай **нормалізуємо** дані, коли хочемо знати, скільки стандартних відхилень має кожне значення в наборі даних від середнього.

**Нормалізований набір даних** завжди матиме значення від 0 до 1. У нашому випадку, я нормалізував значення у колонці age, адже як можна побачити решта числових значень у таблиці в межах від 0 до 1.

**Висновок:** У процесі лабораторної роботи з аналізу даних було:

1. **Знайомство з Google Sheets:** було досліджено використання формул для фільтрації даних та побудова графіків (гістограм).
2. **Ознайомлено з бібліотеками Python** для роботи з даними (pandas, matplotlib).
3. **Використання Python:** імпортовано CSV-файл, додано вигадану особу, здійснено фільтрацію за класом пасажира, відсотком виживання та середнім віком

Отримані результати допомогли краще зрозуміти дані та підготувати їх для подальшого аналізу.