Національний технічний університет України «КПІ ім. Ігоря Сікорського»

Факультет Інформатики та Обчислювальної Техніки

Кафедра інформаційних систем та технологій

Лабораторна робота №1

з дисципліни «Прикладні задачі машинного навчання»

на тему

«Введення в data science»

15 варіант

Виконав:

студент групи ІС-21 Костюк А. С.

Викладач:

Нестерук А.О.

Київ – 2024

**Зміст**

[**1. Постановка задачі** 3](#_Toc161134903)

[**2. Виконання** 4](#_Toc161134904)

[**3. Висновок** 20](#_Toc161134905)

**1. Постановка задачі**

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

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. Виконати первинну обробку даних.

6. Прочитати набір даних катастрофи «Титаніка».

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

8. Переглянути рядки набору даних катастрофи «Титаніка».

9. Налаштувати назви стовпців.

10. Провести простий аналіз даних.

11. Побудувати гістограму віку пасажирів.

12. Зробити звіт про роботу.

**2. Виконання**

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

Обрані дані, а саме середня заробітня плата штатних працівників у 2010-2023 роках:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 |
| **2250** | **2648** | **3041** | **3282** | **3480** | **4195** | **5183** | **7104** | **8865** | **10497** | **11591** | **14014** | **14847** | **17442** |

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

Зображення, що містить текст, знімок екрана, програмне забезпечення, Мультимедійне програмне забезпечення

Автоматично згенерований опис

3) **Візуалізувати завантажені дані за допомогою гістограми.**

Зображення, що містить текст, знімок екрана, Шрифт, програмне забезпечення

Автоматично згенерований опис

Зображення, що містить текст, знімок екрана, Графік, схема

Автоматично згенерований опис

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

Зображення, що містить текст, знімок екрана, програмне забезпечення, Мультимедійне програмне забезпечення

Автоматично згенерований опис

Зображення, що містить текст, знімок екрана, програмне забезпечення, монітор

Автоматично згенерований опис

Зображення, що містить текст, знімок екрана, програмне забезпечення, Шрифт

Автоматично згенерований опис

Зображення, що містить текст, знімок екрана, програмне забезпечення, Мультимедійне програмне забезпечення

Автоматично згенерований опис

Зображення, що містить текст, знімок екрана, програмне забезпечення, Шрифт

Автоматично згенерований опис

Зображення, що містить текст, знімок екрана, Шрифт, програмне забезпечення

Автоматично згенерований опис

Зображення, що містить текст, знімок екрана, програмне забезпечення

Автоматично згенерований опис

Зображення, що містить текст, знімок екрана, програмне забезпечення, Мультимедійне програмне забезпечення

Автоматично згенерований опис

Зображення, що містить текст, знімок екрана, Шрифт, програмне забезпечення

Автоматично згенерований опис

Зображення, що містить текст, знімок екрана, програмне забезпечення, Мультимедійне програмне забезпечення

Автоматично згенерований опис

Зображення, що містить текст, знімок екрана, програмне забезпечення, Мультимедійне програмне забезпечення

Автоматично згенерований опис

Зображення, що містить текст, знімок екрана, програмне забезпечення, Мультимедійне програмне забезпечення

Автоматично згенерований опис

Зображення, що містить текст, знімок екрана, програмне забезпечення, Мультимедійне програмне забезпечення

Автоматично згенерований опис

Зображення, що містить знімок екрана

Автоматично згенерований опис

Зображення, що містить текст, знімок екрана, програмне забезпечення, Шрифт

Автоматично згенерований опис

Зображення, що містить текст, знімок екрана

Автоматично згенерований опис

Зображення, що містить текст, знімок екрана, програмне забезпечення, Шрифт

Автоматично згенерований опис

Зображення, що містить текст, знімок екрана, програмне забезпечення, Шрифт

Автоматично згенерований опис

Зображення, що містить знімок екрана, текст, програмне забезпечення, Шрифт

Автоматично згенерований опис

5) **Виконати первинну обробку даних.**

Зображення, що містить текст, знімок екрана, програмне забезпечення, Мультимедійне програмне забезпечення

Автоматично згенерований опис

Зображення, що містить текст, знімок екрана, програмне забезпечення, Мультимедійне програмне забезпечення

Автоматично згенерований опис

Зображення, що містить текст, знімок екрана, програмне забезпечення, Шрифт

Автоматично згенерований опис

6) **Прочитати набір даних катастрофи «Титаніка».**

Набір даних катастрофи «Титаніка» належить до числа найпопулярніших наборів даних машинного навчання і доступний в багатьох форматах, включаючи CSV:

https://vincentarelbundock.github.io/Rdatasets/datasets.html

7) **Завантажити набір даних катастрофи «Титаніка» за URL-адресою.**

Зображення, що містить текст, знімок екрана, програмне забезпечення, Шрифт

Автоматично згенерований опис

8) **Переглянути рядки набору даних катастрофи «Титаніка».**

Зображення, що містить текст, знімок екрана, програмне забезпечення, Шрифт

Автоматично згенерований опис

9) **Налаштувати назви стовпців.**

Зображення, що містить текст, знімок екрана, програмне забезпечення, Шрифт

Автоматично згенерований опис

10) **Провести простий аналіз даних.**

Зображення, що містить текст, знімок екрана, програмне забезпечення, Шрифт

Автоматично згенерований опис

Зображення, що містить текст, знімок екрана, Шрифт, дизайн

Автоматично згенерований опис

Зображення, що містить текст, знімок екрана, програмне забезпечення, Шрифт

Автоматично згенерований опис

Зображення, що містить текст, знімок екрана, програмне забезпечення, Мультимедійне програмне забезпечення

Автоматично згенерований опис

Зображення, що містить текст, знімок екрана, програмне забезпечення, Комп’ютерна піктограма

Автоматично згенерований опис

Зображення, що містить текст, знімок екрана, програмне забезпечення, Мультимедійне програмне забезпечення

Автоматично згенерований опис

11) **Побудувати гістограму віку пасажирів.**

Зображення, що містить текст, знімок екрана, схема, Графік

Автоматично згенерований опис

**3. Висновок**

В даній лабораторній роботі було розглянуто різні способи обробки та візуалізації даних за допомогою Python та його бібліотек Matplotlib, statistics, numpy, seaborn та pandas. Отримані знання були використані для обробки даних катастрофи Титаніка, а результати представлені в лабораторній.