

# Лабораторна робота № 1.

## Аналіз даних про доходи населення

**\*\*В завданні пропонується за допомогою Pandas відповісти на декілька питань за даними репозиторія UCI [\[Adult\]](https://archive.ics.uci.edu/ml/datasets/Adult)(<https://archive.ics.uci.edu/ml/datasets/Adult>) (качати дані не потрібно - вони вже є в директорії роботи).**

Унікальні значення ознак (більше інформації за посиланням вище):

- age: continuous.
- workclass: Private, Self-emp-not-inc, Self-emp-inc, Federal-gov, Local-gov, State-gov, Without-pay, Never-worked.
- fnlwgt: continuous.
- education: Bachelors, Some-college, 11th, HS-grad, Prof-school, Assoc-acdm, Assoc-voc, 9th, 7th-8th, 12th, Masters, 1st-4th, 10th, Doctorate, 5th-6th, Preschool.
- education-num: continuous.
- marital-status: Married-civ-spouse, Divorced, Never-married, Separated, Widowed, Married-spouse-absent, Married-AF-spouse.
- occupation: Tech-support, Craft-repair, Other-service, Sales, Exec-managerial, Prof-specialty, Handlers-cleaners, Machine-op-inspct, Adm-clerical, Farming-fishing, Transport-moving, Priv-house-serv, Protective-serv, Armed-Forces.
- relationship: Wife, Own-child, Husband, Not-in-family, Other-relative, Unmarried.
- race: White, Asian-Pac-Islander, Amer-Indian-Eskimo, Other, Black.
- sex: Female, Male.
- capital-gain: continuous.
- capital-loss: continuous.
- hours-per-week: continuous.
- native-country: United-States, Cambodia, England, Puerto-Rico, Canada, Germany, Outlying-US(Guam-USVI-etc), India, Japan, Greece, South, China, Cuba, Iran, Honduras, Philippines, Italy, Poland, Jamaica, Vietnam, Mexico, Portugal, Ireland, France, Dominican-Republic, Laos, Ecuador, Taiwan, Haiti, Columbia, Hungary, Guatemala, Nicaragua, Scotland, Thailand, Yugoslavia, El-Salvador, Trinidad&Tobago, Peru, Hong, Holand-Netherlands.
- salary: >50K,<=50K

```
In [1]: import pandas as pd
```

```
In [2]: data = pd.read_csv('data/adult.data.csv')
data.head()
```

Out[2]:

	age	workclass	fnlwgt	education	education-num	marital-status	occupation	relationship	race	se
0	39	State-gov	77516	Bachelors	13	Never-married	Adm-clerical	Not-in-family	White	Mal
1	50	Self-emp-not-inc	83311	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Mal
2	38	Private	215646	HS-grad	9	Divorced	Handlers-cleaners	Not-in-family	White	Mal
3	53	Private	234721	11th	7	Married-civ-spouse	Handlers-cleaners	Husband	Black	Mal
4	28	Private	338409	Bachelors	13	Married-civ-spouse	Prof-specialty	Wife	Black	Femal



1. Скільки чоловіків і жінок (ознака *sex*) представлено в цьому наборі даних?

```
In [4]: # Ваш код тут
```

2. Який середній вік (ознака *age*) жінок?

```
In [5]: # Ваш код тут
```

3. Яка частка громадян Німеччини (ознака *native-country*)?

```
In [6]: # Ваш код тут
```

\*4-5. Які середні значення і середні відхилення віку тих, хто отримує більше 50K в рік (ознака *\*salary*) і тих, хто отримує менше 50K в рік? \*\*

```
In [7]: # Ваш код тут
```

6. Чи правда, що люди, які отримують більше 50k, мають як мінімум вищу освіту? (ознака *education* – *Bachelors*, *Prof-school*, *Assoc-acdm*, *Assoc-voc*, *Masters* чи *Doctorate*)

```
In [8]: # Ваш код тут
```

7. Вивести статистику віку для кожної раси (ознака *race*) і кожної статі. Використовуйте *groupby* і *describe*. Знайдіть таким чином максимальний вік чоловіків раси *Amer-Indian-Eskimo*.

In [9]: `# Ваш код тут`

8. Серед кого більша частка заробляючих багато (>50K): серед одружених чи холостих чоловіків (ознака *marital-status*)? Одруженими вважаємо тих, у кого *marital-status* починається з *Married* (*Married-civ-spouse*, *Married-spouse-absent* чи *Married-AF-spouse*), решту вважаємо холостими.

In [10]: `# Ваш код тут`

9. Яку максимальну кількість годин людина працює за тиждень (ознака *hours-per-week*)? Скільки людей працюють таку кількість годин і який серед них відсоток заробляючих багато?

In [11]: `# Ваш код тут`

10. Підрахуйте середній час роботи (*hours-per-week*) заробляючих мало і багато (*salary*) для кожної країни (*native-country*).

In [12]: `# Ваш код тут`

In [ ]: