Лабораторна робота №3

Бібліотека pandas

Цей зошит призначений для навчання основним прийомам роботи з типом даних DataFrame бібліотеки pandas. Під час навчання необхідно використати способи доступу до даних, які уже відомі вам з лекційного курсу. Навчання будемо проводити на конкретному прикладі.

Умова задачі

Ми зупинили десять людей на вулиці і запитали їх, які у них домашні тварини. Ми також зафіксували стать та вік людини. В результаті сформували структуру даних pets.

```
In [1]:
       import numpy as np
       import pandas as pd
       'pets': np.array([['cat', 'dog'],
                                       ['hamster'],
                                       ['cat', 'gerbil'],
['fish', 'hamster', 'gerbil'],
                                       ['cat'],
                                       ['dog'],
                                       ['dog'],
                                       ['cat'],
                                       ['rabbit', 'cat'],
                                       ['dog']], dtype=object)})
```

Використовуючи зібрані дані, необхідно проаналізувати їх і на основі нашого опитування дати відповідь на такі запитання:

- Якої статі був наймолодший респондент?
- У якому віці була людина з найбільшою кількістю домашніх тварин?
- Яка тварина є найпопулярнішим вихованцем?
- Який був середній вік власників собак?

По-перше, давайте просто розглянемо дані. Наш набір даних не дуже великий, тому насправді навіть не потрібно використовувати head(). Одержимо вивід даних, запустивши на виконання код Python у наступній комірці.

Нагадування: Запуск коду - натиснути "Run" або "Shift+Enter".

```
In [2]:
             pets
Out[2]:
                sex age
                                              pets
                                         [cat, dog]
                  M
                        21
            1
                  Μ
                        45
                                         [hamster]
            2
                  F
                        23
                                       [cat, gerbil]
                             [fish, hamster, gerbil]
            3
                  M
                        56
                                              [cat]
            4
                  F
                        47
            5
                   F
                        70
                                             [dog]
            6
                  F
                        34
                                             [dog]
            7
                  M
                        30
                                              [cat]
                                       [rabbit, cat]
            8
                  F
                        19
            9
                  Μ
                        62
                                             [dog]
```

Зверніть увагу також і на те, як в зошиті гарно представлено дані за замовчуванням для типу DataFrames . Крім того, ми можемо таке представлення редагувати для своїх потреб! Проте ми не будемо виконувати ці дії в рамках даної практичної роботи. Відповімо по черзі на поставлені запитання.

Підказка. Тут ви можете ефективно застосувати метод . loc . Подумайте над тим, як представити

Якої статі був наймолодший респондент

це завдання у вигляді двох умов: створення булевого індексу, який відповідає значенням True , де значення у віковій колонці, що дорівнює мінімуму стовпця віку. Потім для вибраного запису проаналізуйте колонку статі. Код запишіть в комірці нижче.

```
# one line of code
         pets.loc[pets.age == pets.age.min()].sex
Out[3]: 8
             F
        Name: sex, dtype: object
```

У якому віці була людина з найбільшою кількістю домашніх тварин? Підказка: спочатку створіть новий стовпець в структурі DataFrame і назвіть це «num_pets», який

Якщо ви правильно записали код, то наймолодший респондент - жінка (F).

міститиме число тварин у даного респондента. Знайдіть максимум у цьому стовпці і виведіть вік відповідного респондента. Код запишіть у комірці нижче. # task: create new column 'num_pets' which contains the number of pets

```
# one line of code here:
         pets['num_pets'] = pets['pets'].apply(lambda ps: len(ps))
         # view the DataFrame again to check our new column is there
         pets.loc[pets['num_pets'] == max(pets['num_pets']), 'age']
Out[4]: 3
             56
        Name: age, dtype: int64
        Правильна відповідь: людині з найбільшою кількістю домашніх тварин було 56 років.
```

Це дуже цікаве запитання, оскільки наші дані про тварин групуються не по колонці тварин, а по

each person had (hint: this is the length of eachlist in the pets column)

In [3]:

In [4]:

In [5]:

In [6]:

Яка тварина є найпопулярнішим вихованцем?

респондентах. Нам зараз потрібно ввійти в колонку з домашніми тваринами, щоб підрахувати кожен тип тварин. Для виконання цього завдання потрібно відійти від традиційного програмістського

підходу і використати потужний механізм групування бібліотеки pandas. Початковий код цієї технології наведено в наступній комірці. Нижче наведіть завершальний код. pet_series = pets['pets'].apply(pd.Series).stack().reset_index(drop=True) pet_series # task: produce an ordered count of each animal

```
pet_series.value_counts()
                    5
Out[5]: cat
                    4
         dog
        hamster
                    2
         gerbil
        rabbit
                    1
```

fish dtype: int64 Відповідь: кішка - найпопулярніша домашня тварина.

Який був середній вік власників собак?

one line of code here:

False . Код запишіть в комірці нижче. # example ('dog' in ['dog', 'cat'], 'dog' in ['rabbit'])

task: use a lambda function to test whether 'dog' is contained in each list of anima

Підказка: знову ж таки, вам може бути корисно використовувати функцію лямбда для створення булевого індексу, який дорівнює True , якщо респондент сказав, що він має собаку, а в іншому -

```
# extract the age column and then chain the mean method to calculate the average age.
         # one line of code here:
         pets.loc[pets.pets.apply(lambda owned_pets: 'dog' in owned_pets)].age.mean()
Out[6]: 46.75
```

Висновок

Ми маємо зрозуміти, що бібліотека pandas може обробляти табличні дані, де кожен стовпець може мати різний тип (наприклад, стать - це символ, а вік - це число). Крім того, бібліотека pandas має потужні методи для отримання зрізів та вибору даних. Це дає можливість одержати відповідь на

```
складні питання, використовуючи порівняно невеликий код.
In [ ]:
```