Лабораторна робота 2.3. Бібліотека Pandas.

Коротка теорія:

pandas добре підходить для багатьох різних типів даних:

- ✓ Табличні дані з неоднорідно типізованими стовпцями, як у таблиці SQL або електронній таблиці Excel;
- ✓ Упорядковані та невпорядковані (не обов'язково з фіксованою частотою) дані часових рядів;
- ✓ Довільні матричні дані з мітками рядків і стовпців;
- ✓ Будь-яка інша форма наборів спостережень/статистичних даних.

Основними структурами даних, які використовуються в Pandas, є:

Series та DataFrame.

Series - це одномірний масив даних з індексами, який можна використовувати для зберігання та обробки даних, що мають один тип.

DataFrame - це двовимірна таблична структура даних, що складається з рядків та стовпців, яку можна використовувати для зберігання та обробки даних, що мають різні типи.

Основні методи та функції, що використовуються в Pandas:

Читання та запис даних:

read_csv() - читання даних з CSV файлу

read_excel() - читання даних з Excel файлу

to_csv() - запис даних до CSV файлу

to_excel() - запис даних до Excel файлу

Робота з даними:

head() - виведення перших декількох рядків таблиці

tail() - виведення останніх декількох рядків таблиці

shape - виведення кількості рядків та стовпців в таблиці

describe() - виведення статистичних характеристик таблиці

sort_values() - сортування даних за певними стовпцями

groupby() - групування даних за певними стовпцями

drop() - видалення рядків або стовпців з таблиці

replace() - заміна значень в таблиці

Обробка даних:

apply() - застосування функції до даних таблиці

fillna() - заповнення пропущених значень в таблиці

merge() - об'єднання даних з різних таблиць

pivot_table() - створення зведеної таблиці на основі вихідних даних.

Вибір певних даних з DataFrame:

loc[] - вибір даних за міткою рядка або стовпця.

Приклад:

```
df.loc[3] # вибір стовпця за міткою
df.loc[:, 'column_name']# вибір рядків та стовпців за мітками
df.loc[3:6, 'column_name_1':'column_name_2']
```

iloc[] - вибір даних за номером рядка або стовпця.

Приклад:

```
df.iloc[3] # вибір рядка за номером
df.iloc[:, 0] # вибір стовпця за номером
df.iloc[3:6, 0:2] # вибір рядків та стовпців за номерами
```

at[] - вибір конкретного значення за міткою рядка та стовпця.

```
df.at[3, 'column_name']
```

query() - вибір рядків, які відповідають певній умові.

```
df.query('column_name > 10')
```

filter() - вибір стовпців за певними умовами (назва стовпця, регулярний вираз тощо).

```
df.filter(regex='_id$')
```

Групування даних та застосування агрегаційних функцій:

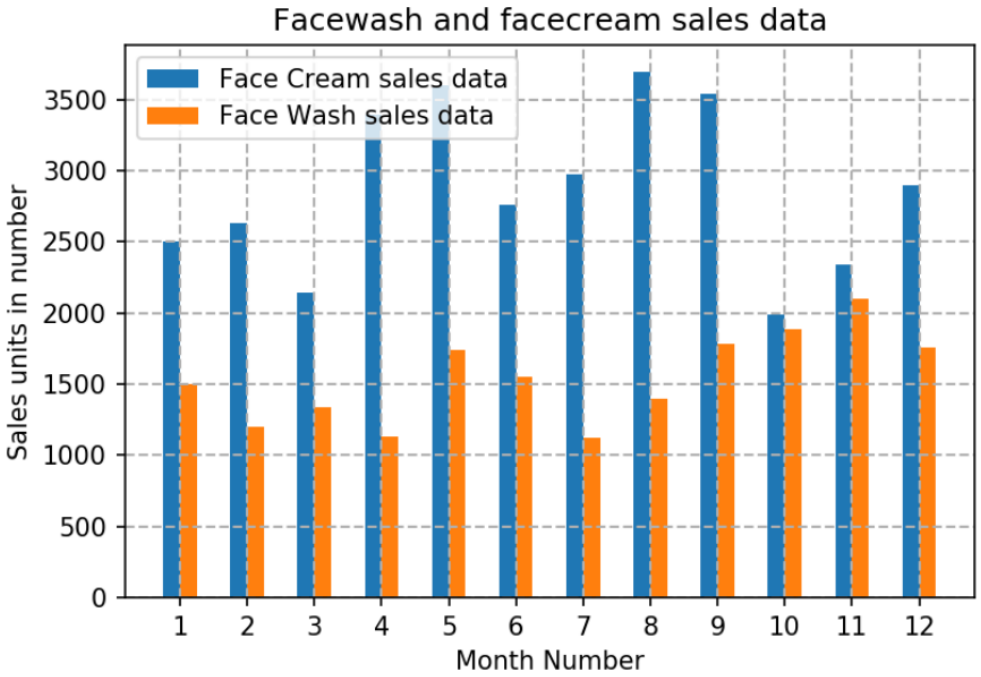
grouped_df = df.groupby('column_name').mean()['value'] # групування за значенням у стовпці 'column_name' та знаходження середнього значення в стовпці 'value'

Основна інформація по Pandas в документації: https://pandas.pydata.org/docs/

Завдання 1. Використовуйте файл company_sales_data.csv.

Прочитайте дані про продажі кремів для обличчя та засобів для вмивання та відобразіть їх за допомогою гістограми.

Приклад, як має виглядати ваша гістограма. Кольори довільно.



Завдання 2. Використовуйте файл "voltage.xlsx"

- Побудувати залежності, у має бути представлений в координатах через log(10)(y від x)
- **abs(col(Current11)) від Voltage_1 (1)від - Омічна провідність**
- **col(Iabs)/col(Uabs) від 1/col(Uabs) -інжекційна провідність**
- **col(Iabs)/col(Uabs)^2 від 1/col(Uabs) - провідність за Фаулера-Нордгеймом**
- **col(Iabs)/col(Uabs від col(Uabs)^0.5 – провідність Пула-Френкеля**
- **Знайти найбільше і найменше значення струму і напруги.**

Завдання 3

Розв’яжіть систему рівнянь з чотирма невідомими з використанням numpy залежно від величини параметра «a», який набуває цілих значень від -30 до +30 включно. Створіть Data Frame, у якому назвами стовпців є імена змінних, а назвами рядків – значення параметра «a», та внесіть в отриману таблицю результати обчислень.

$$\begin{cases} x_1 - 4x_2 + 3x_3 - x_4 = a \\ x_1 + 2x_2 + 5x_3 - x_4 = 2 \\ 2x_1 - 3x_2 + 4x_4 = 5 \\ -x_1 - 2x_2 - 3x_3 - 4x_4 = -5 \end{cases}$$