**Практика: вправа для Аналітиків даних - реалізація трансформації**

**Дані:**

1. data.csv – анкетні дані

|  |  |
| --- | --- |
| **Колонка** | **Опис** |
| FREQ.id | Частота даних |
| EDRPOU.id | ЄДРПОУ код (id фірми-респондента) |
| WASTE.id | Вид відходу |
| HAZARD.id | Клас небезпеки відходу |
| MATERIAL.id | Матеріал відходу |
| 2020 | Об'єм утворених відходів у 2020 звітному році |
| 2021 | Об'єм утворених відходів у 2021 звітному році |

1. edrpou.csv – список респондентів, що брали участь у обстеженні

|  |  |
| --- | --- |
| **Колонка** | **Опис** |
| EDRPOU | ЄДРПОУ код (id фірми-респондента) |
| KATOTTG | Код місцевості за класифікацією КАТОТТГ ( [wiki](https://uk.wikipedia.org/wiki/%D0%9A%D0%BE%D0%B4%D0%B8%D1%84%D1%96%D0%BA%D0%B0%D1%82%D0%BE%D1%80_%D0%B0%D0%B4%D0%BC%D1%96%D0%BD%D1%96%D1%81%D1%82%D1%80%D0%B0%D1%82%D0%B8%D0%B2%D0%BD%D0%BE-%D1%82%D0%B5%D1%80%D0%B8%D1%82%D0%BE%D1%80%D1%96%D0%B0%D0%BB%D1%8C%D0%BD%D0%B8%D1%85_%D0%BE%D0%B4%D0%B8%D0%BD%D0%B8%D1%86%D1%8C_%D1%82%D0%B0_%D1%82%D0%B5%D1%80%D0%B8%D1%82%D0%BE%D1%80%D1%96%D0%B9_%D1%82%D0%B5%D1%80%D0%B8%D1%82%D0%BE%D1%80%D1%96%D0%B0%D0%BB%D1%8C%D0%BD%D0%B8%D1%85_%D0%B3%D1%80%D0%BE%D0%BC%D0%B0%D0%B4) ) |
| KVED | Вид економічної діяльності за КВЕД |

1. SSSU\_CL\_WASTE\_HAZARD\_CLASS(2.0.1).json - Класифікатор груп та класів небезпеки відходів (ієрархічний)

Атрибути:

* **annotations – анотації до коду класифікатора**
* **id –** id елемента класифікатора
* **name** – ім'я елемента класифікатора
* **names –** імена елемента в альтернативних мовах
* **parent –** батьківський елемент класифікатора в ієрархії (група класу небезпеки, до якої належить елемент)

**Технічні вимоги:**

Для виконання вправи потрібно мати встановленими:

* Мова програмування Python[Python 3.10.0](https://www.python.org/downloads/release/python-3100/)
* Середа програмування Jupyter notebook <https://jupyter.org/install>
* Бібліотека [Pandas](https://pypi.org/project/pandas/)

**Прикладна реалізація трансформації:**

Опираючись на структури та методи бібліотеки Pandas та використовуючи Jupyter Notebooks як середу розробки коду, попрактикуйтесь в реалізації наступних трансформацій даних та дайте відповідь на питання:

1. В даних із файла data.csv створити колонку ‘TIME\_PERIOD’, до якої записати значення відповідних часових періодів в даних. Об'єм утворених відходів записати у стовпчик ‘OBS\_VALUE’
2. Як змінився рівень заповненості звітності в 2021 у порівнянні з 2020?
3. Додати в таблицю, створену у пункті 1, атрибути КВЕД та КАТОТТГ з файлy edrpou.csv
4. На базі таблиці з пункту 3, розрахувати Об'єм утворених відходів для кожного класу небезпеки, матеріалу відходу, виду відходу, року звітності на рівні області КАТОТТГ (Перші 2 знаки в коді КАТОТТГ, деталі за посиланням [wiki](https://uk.wikipedia.org/wiki/%D0%9A%D0%BE%D0%B4%D0%B8%D1%84%D1%96%D0%BA%D0%B0%D1%82%D0%BE%D1%80_%D0%B0%D0%B4%D0%BC%D1%96%D0%BD%D1%96%D1%81%D1%82%D1%80%D0%B0%D1%82%D0%B8%D0%B2%D0%BD%D0%BE-%D1%82%D0%B5%D1%80%D0%B8%D1%82%D0%BE%D1%80%D1%96%D0%B0%D0%BB%D1%8C%D0%BD%D0%B8%D1%85_%D0%BE%D0%B4%D0%B8%D0%BD%D0%B8%D1%86%D1%8C_%D1%82%D0%B0_%D1%82%D0%B5%D1%80%D0%B8%D1%82%D0%BE%D1%80%D1%96%D0%B9_%D1%82%D0%B5%D1%80%D0%B8%D1%82%D0%BE%D1%80%D1%96%D0%B0%D0%BB%D1%8C%D0%BD%D0%B8%D1%85_%D0%B3%D1%80%D0%BE%D0%BC%D0%B0%D0%B4))
5. На базі таблиці з пункту 3, розрахувати Об'єм утворених відходів для кожного класу небезпеки, виду економічної діяльності КВЕД, року звітності на рівні області КАТОТТГ (Перші 2 знаки в коді КАТОТТГ, деталі за посиланням [wiki](https://uk.wikipedia.org/wiki/%D0%9A%D0%BE%D0%B4%D0%B8%D1%84%D1%96%D0%BA%D0%B0%D1%82%D0%BE%D1%80_%D0%B0%D0%B4%D0%BC%D1%96%D0%BD%D1%96%D1%81%D1%82%D1%80%D0%B0%D1%82%D0%B8%D0%B2%D0%BD%D0%BE-%D1%82%D0%B5%D1%80%D0%B8%D1%82%D0%BE%D1%80%D1%96%D0%B0%D0%BB%D1%8C%D0%BD%D0%B8%D1%85_%D0%BE%D0%B4%D0%B8%D0%BD%D0%B8%D1%86%D1%8C_%D1%82%D0%B0_%D1%82%D0%B5%D1%80%D0%B8%D1%82%D0%BE%D1%80%D1%96%D0%B9_%D1%82%D0%B5%D1%80%D0%B8%D1%82%D0%BE%D1%80%D1%96%D0%B0%D0%BB%D1%8C%D0%BD%D0%B8%D1%85_%D0%B3%D1%80%D0%BE%D0%BC%D0%B0%D0%B4))
6. На базі таблиці з пункту 3, розрахувати Об'єм утворених відходів для кожного року звітності на рівні області КАТОТТГ.   
   В якому з регіонів України утворилась найбільша кількість відходів (у відсотковому вираженні на рівні усієї України)?
7. В таблицю з пункту 3 додати атрибути **name**, **parent** з файлу SSSU\_CL\_WASTE\_HAZARD\_CLASS(2.0.1).json.
8. Для кожного року звітності обрахувати Об'єм утворених відходів для груп класів небезпеки: 0000 (I-IV. Відходи в цілому), 0123 (I-III. Небезпечні відходи), 0124 (IV. Не небезпечні відходи). Значення на рівні групи рахується як сума відповідних класів, що входять до даної групи. Приналежність класу до групи визначається значенням атрибута **parent.**
9. **Результати аналізу та код для виконання завдання збережіть як *ipynb file* (файл створений у середовищі програмування** **Jupyter notebook), відправте результати у відповідь Спеціалісту, що надіслала(в) вам завдання, з назвою ‘*Прізвище*\_*Ім’я*\_Реалізація\_трансформації’.**   
     
   ***У вас точно усе виде!***