# Python-разработчик. Тестовое задание кандидатам

# Задание 1  
# имеется текстовый файл f.csv, по формату похожий на .csv с разделителем |

"""  
lastname|name|patronymic|date\_of\_birth|id  
Фамилия1|Имя1|Отчество1 |21.11.1998   |312040348-3048  
Фамилия2|Имя2|Отчество2 |11.01.1972   |457865234-3431  
...  
"""

# 1. Реализовать сбор уникальных записей  
# 2. Случается, что под одинаковым id присутствуют разные данные - собрать такие записи

# Задание 2  
# в наличии список множеств. внутри множества целые числа  
# посчитать   
#  1. общее количество чисел  
#  2. общую сумму чисел  
#  3. посчитать среднее значение  
#  4. собрать все числа из множеств в один кортеж  
m = [{11, 3, 5}, {2, 17, 87, 32}, {4, 44}, {24, 11, 9, 7, 8}]  
# \*написать решения в одну строку

# Задание 3  
# имеется список списков  
# a = [[1,2,3], [4,5,6]]  
# сделать список словарей  
# b = [{'k1': 1, 'k2': 2, 'k3': 3}, {'k1': 4, 'k2': 5, 'k3': 6}]  
# \*написать решение в одну строку

# Задание 4  
# Имеется папка с файлами  
# Реализовать удаление файлов старше N дней

# Задание 5\*  
Имеется текстовый файл с набором русских слов(имена существительные, им.падеж)  
Одна строка файла содержит одно слово.  
Написать программу которая выводит список слов, каждый элемент списка которого - это новое слово,  
которое состоит из двух сцепленных в одно, которые имеются в текстовом файле.  
Порядок вывода слов НЕ имеет значения

Например, текстовый файл содержит слова: ласты, стык, стыковка, баласт, кабала, карась

Пользователь вводмт первое слово: ласты  
Программа выводит:  
ластык  
ластыковка

Пользователь вводмт первое слово: кабала  
Программа выводит:  
кабаласты  
кабаласт

Пользователь вводмт первое слово: стыковка  
Программа выводит:  
стыковкабала  
стыковкарась

# Задание 6\*  
Имеется банковское API возвращающее JSON  
{  
    "Columns": ["key1", "key2", "key3"],  
    "Description": "Банковское API каких-то важных документов",  
    "RowCount": 2,  
    "Rows": [  
        ["value1", "value2", "value3"],  
        ["value4", "value5", "value6"]  
    ]  
}  
Основной интерес представляют значения полей "Columns" и "Rows",  
которые соответственно являются списком названий столбцов и значениями столбцов

Необходимо:  
    1. Получить JSON из внешнего API  
        ендпоинт: GET https://api.gazprombank.ru/very/important/docs?documents\_date={"начало дня сегодня в виде таймстемп"}

        (!) ендпоинт выдуманный  
    2. Валидировать входящий JSON используя модель pydantic  
        (из ТЗ известно что поле "key1" имеет тип int, "key2"(datetime), "key3"(str))  
    2. Представить данные "Columns" и "Rows" в виде плоского csv-подобного pandas.DataFrame  
    3. В полученном DataFrame произвести переименование полей по след. маппингу  
        "key1" -> "document\_id", "key2" -> "document\_dt", "key3" -> "document\_name"  
    3. Полученный DataFrame обогатить доп. столбцом:  
        "load\_dt" -> значение "сейчас"(датавремя)

\*реализовать п.1 с использованием Apache Airflow HttpHook

#  
Решение задач со \* не обязательно, но как плюс