

Заочные задания на позицию аналитика данных в Аналитический Центр ДОМ.РФ

Задание 1.

В базе данных есть таблица **sandpit.test_table**. Она содержит сведения о строящихся на текущий момент многоквартирных домах. Поля этой таблицы:

Наименование поля	Тип	Описание
obj_id	int8	Идентификатор жилого дома. Уникальный ключ таблицы
obj_square	float8	Площадь жилого дома, тыс. кв. м
region_id	int4	Идентификатор субъекта РФ, в котором строится жилой дом
region_name	text	Наименование субъекта РФ, в котором строится жилой дом
dev_id	text	Идентификатор девелопера, строящего жилой дом
dev_name	text	Наименование девелопера, строящего жилой дом

Подготовьте один sql-запрос, который выгружает таблицу с полями:

- Рейтинг региона по объему жилищного строительства
- Наименование региона
- Объем жилищного строительства в регионе, в тыс. кв. м
- Количество девелоперов, занимающихся строительством в регионе
- В т.ч. количество девелоперов, строящих более 1 жилого дома в регионе
- Наименование девелопера с наибольшим объемом жилищного строительства в регионе
- В т.ч. доля его объем жилищного строительства в регионе, в %

Формат результата: текстовый файл в кодировке UTF-8 с расширением .sql

Задание 2.

На [сайте Центрального Банка](#) публикуются показатели рынка ипотечного кредитования в региональном разрезе. Используя эти данные подготовьте ежемесячную динамику среднего ипотечного кредита на первичном рынке (под залог прав ДДУ) за 2021-2022 гг. в разрезе регионов. Допустимо использовать любой удобный способ автоматизации при подготовке данных (python, tableau, power query + DAX, макросы и пр.) Контрольный вопрос по этому заданию: в каком регионе произошёл самый высокий прирост среднего ипотечного кредита за 2021 год (янв. 2022 г. к янв. 2021 г.)?

Формат результата: эксель-файл, ссылка на tableau-public, ссылка на google-sheets или любой другой удобный формат, который будет доступен для просмотра через интернет. Предпочтительный формат: эксель-файл

Задание 3.

Подготовить python-скрипт извлечения данных из файла 01_02_flowvariables.xlsx. Скрипт должен извлечь данные со всех листов кроме листа «01_02_00». С каждого листа следует извлечь табличные данные без сносок и пустых строк, также следует извлечь наименование листа и наименование показателя (ячейка A2). Горизонтальная структура таблиц должна быть трансформирована в вертикальную, таким образом, чтобы заголовки колонок-дат образовали новую вертикальную атрибут-колонку. Полученный результат сохранить в excel-файле в виде таблицы со следующим перечнем колонок:

1. Region – наименование региона
2. date – отчётный месяц показателя
3. value – значение показателя
4. val_name – наименование показателя
5. sheet_name – наименование листа

Формат результата: python-скрипт в utf-8 кодировке, excel-файл с выгрузкой данных