Здравствуйте!  
Перед вами тестовое задание для проверки аналитических навыков.

На выполнение задание отводится 60 минут.

**Описание проекта**

В вашем распоряжении 2 таблицы: данные о приёмах врачей и о населении, прикреплённом к одной из московских поликлиник.

Нужно разобраться, влияет ли пол и возраст пациента на количество посещений медицинской организации.

**Инструкция по выполнению задания**

1. **Откройте файлы с данными и изучите общую информацию о таблицах**
2. **Предобработка данных**

* Определите и изучите пропущенные значения. Помните, что отсутствие значения в датасете – это важная информация, которую необходимо проанализировать.
* Заполните пропуски, где это уместно. Укажите причины, которые могли привести к отсутствию данных.
* Приведите данные к нужным типам.
* Вычислите и добавьте в таблицу «Прикреплённое население» возраст каждого пациента на момент 01.05.2023;

**3) Проведите исследовательский анализ данных и выполните инструкции**

* Изучите следующие параметры: распределение прикреплённого населения по возрасту и по полу. Постройте соответствующие визуализации.
* Изучите время приёма пациентов. Посчитайте среднее и медианное время приёма. Опишите, сколько обычно занимает приём пациента. Когда можно считать, что приём прошёл очень быстро, а когда необычно долго?
* Уберите редкие и выбивающиеся значения. Опишите, какие особенности обнаружили в данных.
* Какие факторы больше всего влияют на количество приёмов? Изучите, зависит ли количество приёмов от возраста и пола пациента. При необходимости постройте визуализацию.
* Выберите 5 пациентов с наибольшим числом приёмов Терапевта. Посещали ли эти пациенты других специалистов в данной медицинской организации?

1. **Напишите общий вывод**

**Варианты оформления:**   
1) в Jupyter Notebook. Заполните программный код в ячейках типа code, текстовые пояснения — в ячейках типа markdown. Примените форматирование и заголовки.

2) в EXCEL. Необходимые вычисления выполняйте на листах документа. На первом листе необходимо предоставить краткую аналитическую справку.

**Описание данных**

Таблица «Прикреплённое население»:

PATIENT\_ID – ID пациента

MO\_ID - ID медицинской организации,

BIRTH\_DATE – дата рождения пациента,

GENDER – пол пациента,

ATTACHMENT\_START\_DATE – дата начала прикрепления,

CREATE\_DATE – дата и время выгрузки данных

Таблица «Приёмы врачей в медицинской организации»

REC\_RCP\_DATE – дата записи,

MO\_ID - ID медицинской организации,

REC\_ID – ID записи,

RCP\_ID – ID приёма,

RCP\_START\_DATETIME – дата и время начала приёма,

RCP\_DURATION – длительность приёма (в минутах),

PATIENT\_ID – ID пациента,

GENDER – пол пациента,

BIRTH\_DATE – дата рождения пациента,

SPEC\_GROUP – специальность врача,

MEJI\_ID – ID врача,

CREATE\_DATE - дата и время выгрузки данных

**Как будут проверять моё задание?**

На мы будем обращать внимание, проверяя задание:

* Как вы описываете выявленные в данных проблемы?
* Какие способы обработки пропусков применяете?
* Как используете срезы данных?
* Решают ли ваши графики определённый вид задач?
* Какие методы построения графиков применяете?
* Выводите финальные данные в сводных таблицах?
* Считаете ли показатели взаимосвязи в данных и правильно ли их объясняете?
* Соблюдаете структуру проекта и поддерживаете аккуратность кода?
* Какие выводы делаете?
* Оставляете комментарии к шагам?

Успехов!