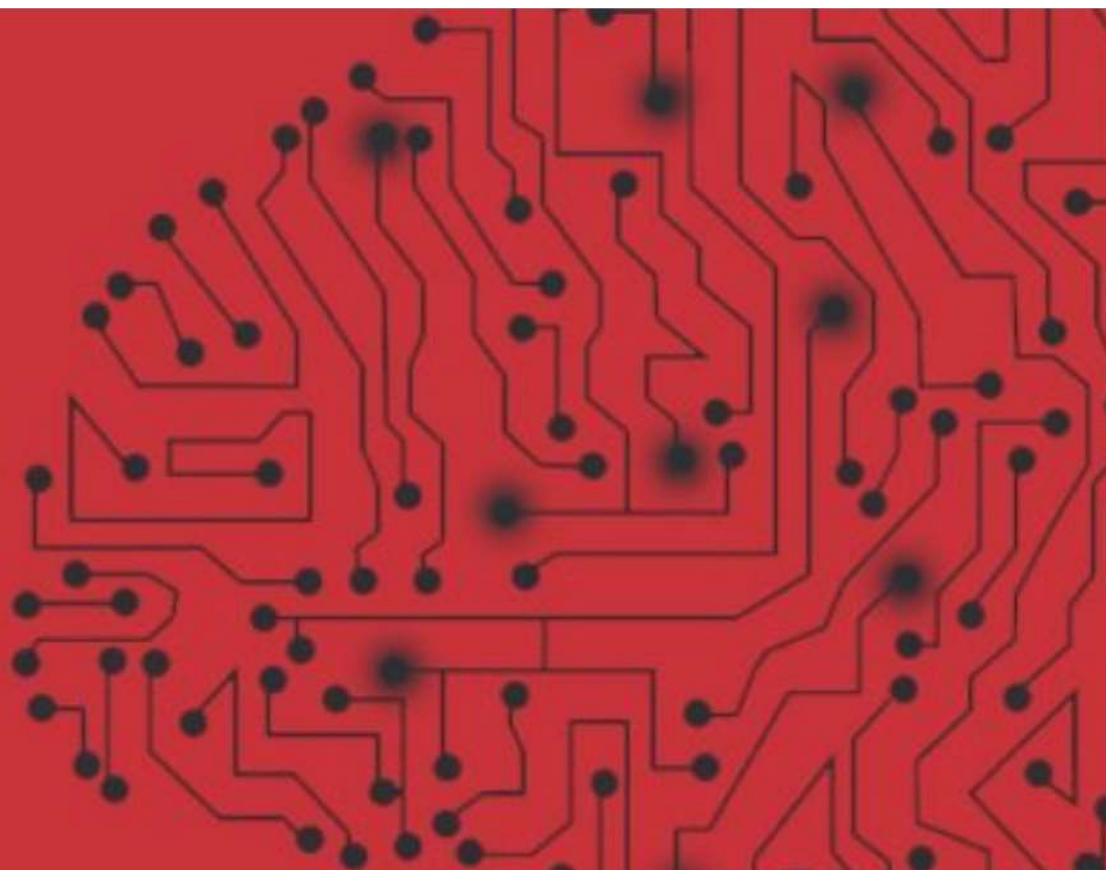


# MINIS B



## Аналитик - тестовое задание

🕒 Created April 23, 2022 2:30 PM

🏷 Tags **Work**

📄 Version v1.2 - SG

### Задание

Требуется провести анализ медицинских данных и рассчитать ряд средних показателей, описывающих поведение застрахованных

**Определение:** Медицинский кейс - последовательность медицинских услуг, полученных одним застрахованным в одной клинике с интервалом между любыми соседними по дате услугами не более 14 дней. Т.е. между любыми 2-мя услугами в кейсе не должно быть разрыва более 14 дней.

**Определение:** Визит - уникальное сочетание даты оказания услуги, застрахованного и клиники. Т.е. один визит - это один день, в который застрахованный побывал в клинике. Если застрахованный за один день был в 2-х клиниках, то это два визита. Если был в одной клинике 2 раза, но в разные дни, то это два визита. Если получал услуги (в любом количестве) в одной клинике в один день, то это один визит.

**Определение:** Диагноз кейса - самый распространенный диагноз в данном кейсе. Если в кейсе было 3 услуги, по первой услуге диагноз = J06.9, по второй = M54.4, по третьей = M54.4, то диагноз кейса = M54.4, т.к. он встречается чаще всего. Если в кейсе несколько диагнозов с одинаковой частотой, то берем первый из них в алфавитном порядке

**Определение:** Группа заболеваний - укрупненные группы диагнозов согласно кодировке МКБ-10. За обозначение группы диагнозов отвечает первая буква кода диагноза. Т.е., группа диагноза J06.9 = J

**Что нужно посчитать:**

1. Количество медицинских кейсов
2. Количество визитов в клинику
3. Средняя длительность медицинского кейса в днях
4. Среднее количество визитов на один кейс
5. Средняя сумма выплат на один кейс
6. Средняя сумма выплат на один визит

Указанные показатели требуется посчитать:

- a. в целом по выборке

- b. в разрезе по ТОП-10 страхователям
- c. в разрезе по укрупненным группам по полу и возрасту (Дети - М, Дети - Ж, Взрослые - М, Взрослые - Ж, Пожилые - М, Пожилые - Ж)  
Дети - от 0 до 17 лет , взрослые - от 18 до 59 лет , пожилые - после 60 лет
- d. В разрезе по группам заболеваний на основе показателя “Диагноз кейса”.  
При вычислении показателя диагноз кейса предусмотреть чистку некорректных данных.

### Описание данных

- row\_id - уникальный идентификатор строки данных
- insured\_id - уникальный идентификатор застрахованного
- date\_of\_birth - дата рождения застрахованного
- gender - пол застрахованного
- employer\_id - уникальный идентификатор страхователя
- medical\_provider\_code - идентификатор медицинской клиники
- date\_of\_service - дата оказания услуги
- service\_name - наименование услуги
- service\_quantity - количество услуг
- icd\_code - код диагноза по МКБ-10
- service\_amount - сумма выплат (стоимость оказанных услуг в рублях)

Данные можно скачать по [ссылке](#)

### Требования к результату

1. Инструмент - формулы в excel или python-скрипт с использованием pandas
2. Результат должен содержать рассчитанные показатели, указанные в задании + скрипты по их расчету (или формулы в excel)
3. Оцениваться будет корректность расчетов показателей и оформление отчета