

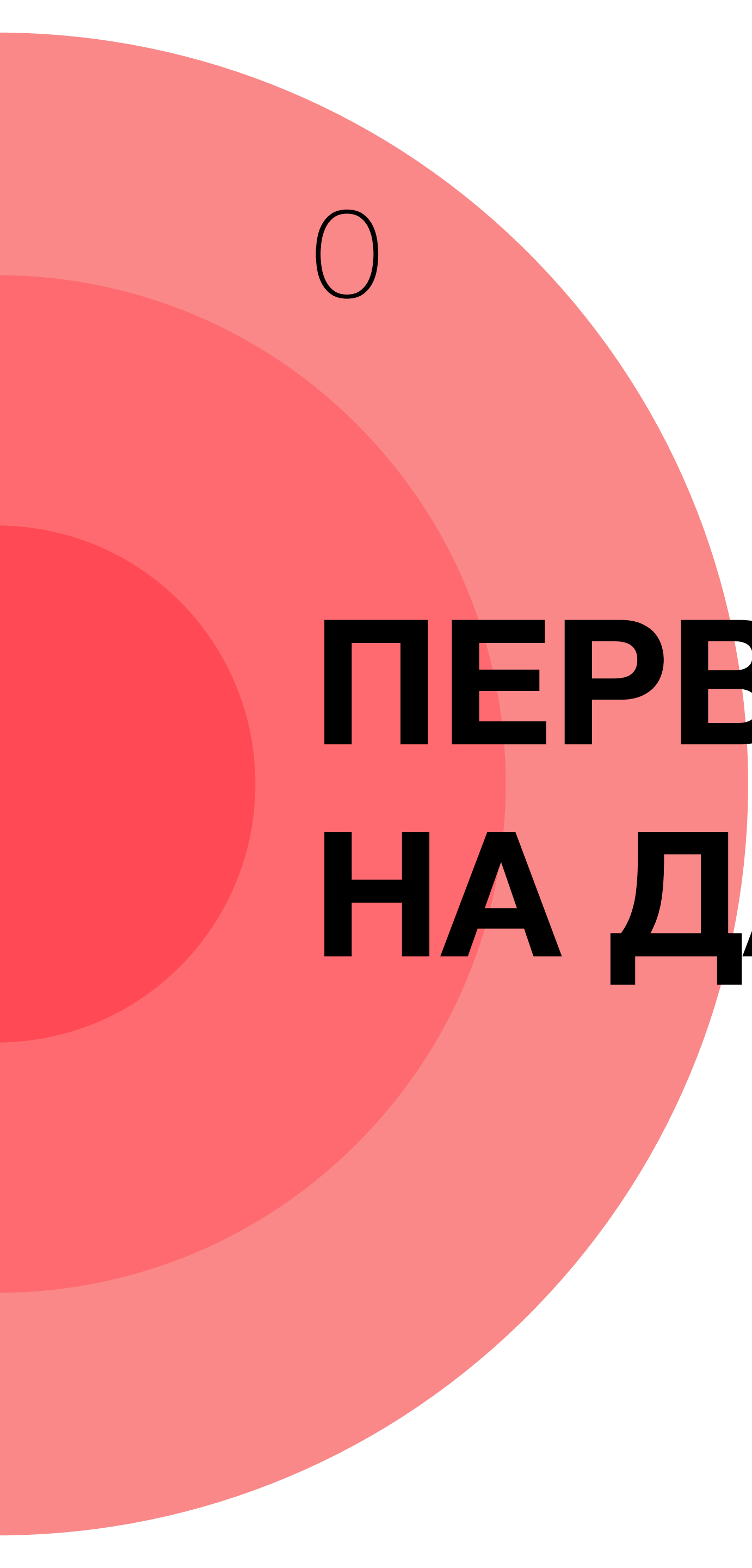
ПРИЧИНЫ НЕЯВКИ ПАЦИЕНТОВ



Colab link: <https://colab.research.google.com/drive/1zkdYSpFPxwI91W-NoGFnJkxfPclst-oc>

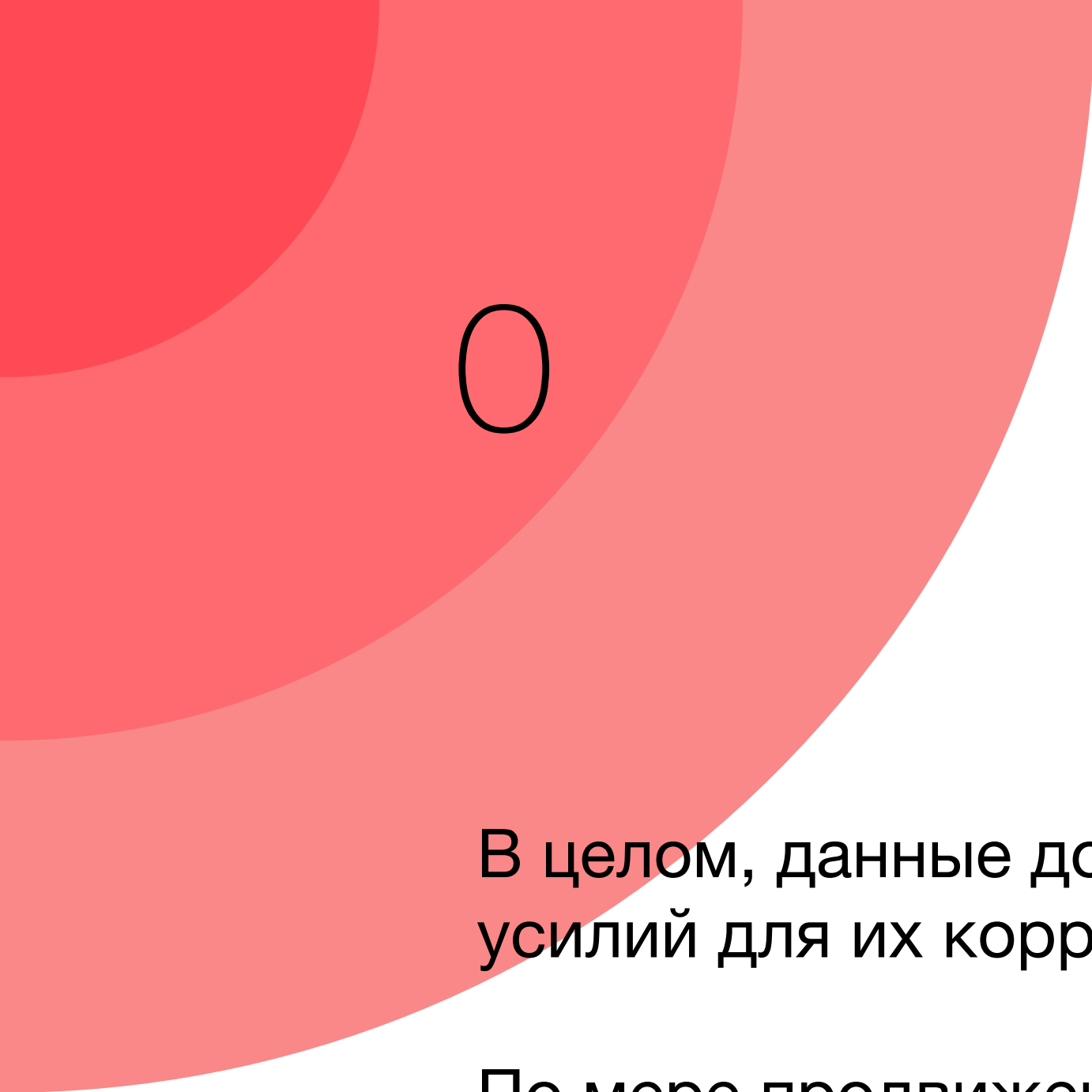
Датасет для анализа:

<https://www.kaggle.com/joniarroba/noshowappointments>



0

ПЕРВЫЙ ВЗГЛЯД НА ДАННЫЕ



0

В целом, данные достаточно чистые, особых усилий для их коррекции не требуется.

По мере продвижения исследования возможно потребуется добавить производные данные.

110 527

Строк в датасете

0% NaN

Отсутствуют пропуски
в данных

20.2%

Неявка по записи



1

ПОСТАНОВКА ЗАДАЧИ, ГИПОТЕЗЫ

1

Целью исследования является анализ данных, для попытки понять, почему существенная часть запланированных посещений не происходит. Для этого:

- Рассмотрим структуру посещения в разрезе записей на пациента и их связь с неявкой
- Рассмотрим как соотносятся время записи/посещения и неявка
- Рассмотрим влияние половозрастных характеристик на неявку
- Рассмотрим влияние сопутствующих заболеваний на неявку
- Рассмотрим как наличие или отсутствие социального пособия влияет на явку
- Рассмотрим как оказывает влияние на неявку СМС информирование

Так же попробуем применить методы машинного обучения для определения степени влияния различных факторов на неявку пациентов.



2

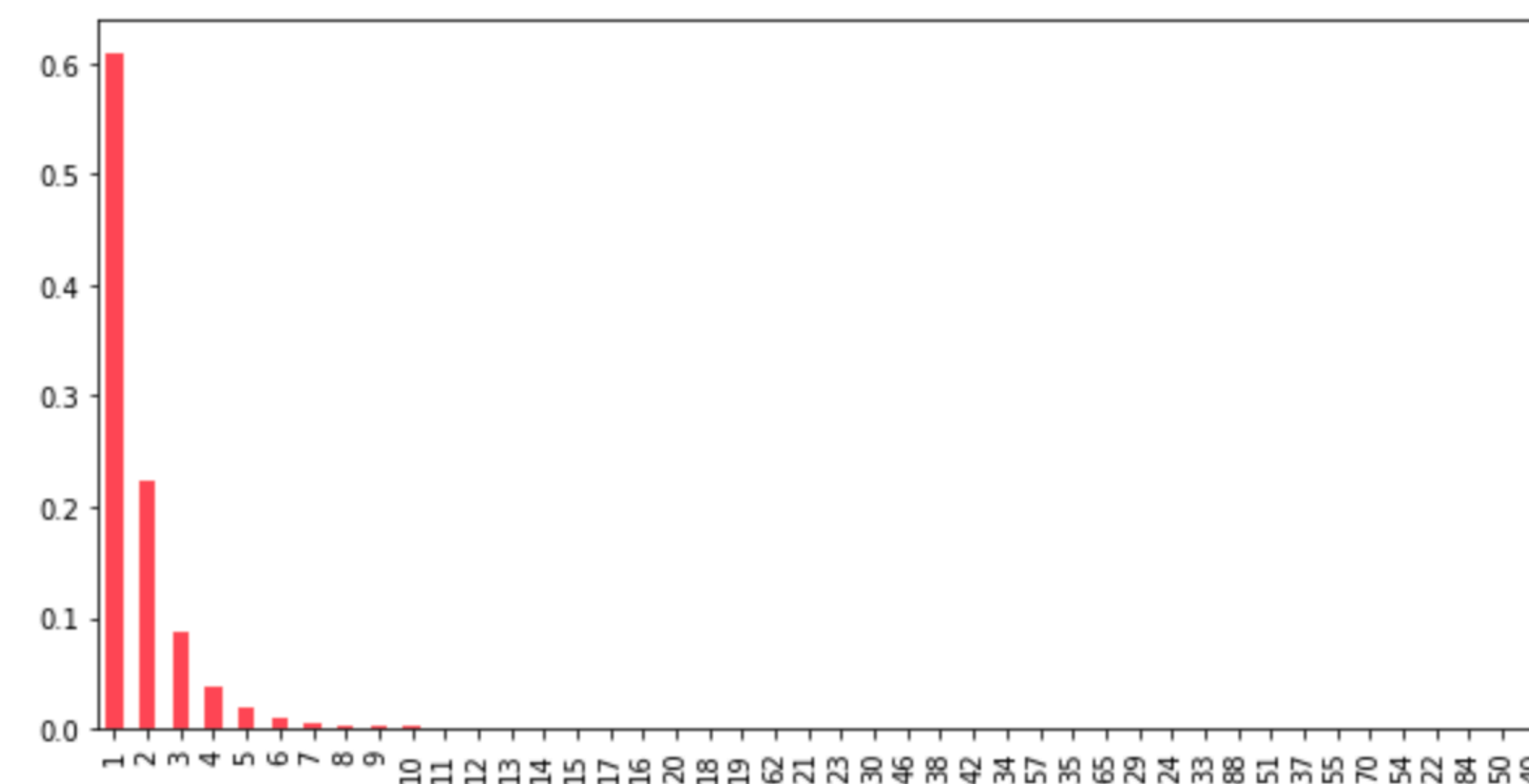
число посещений

2

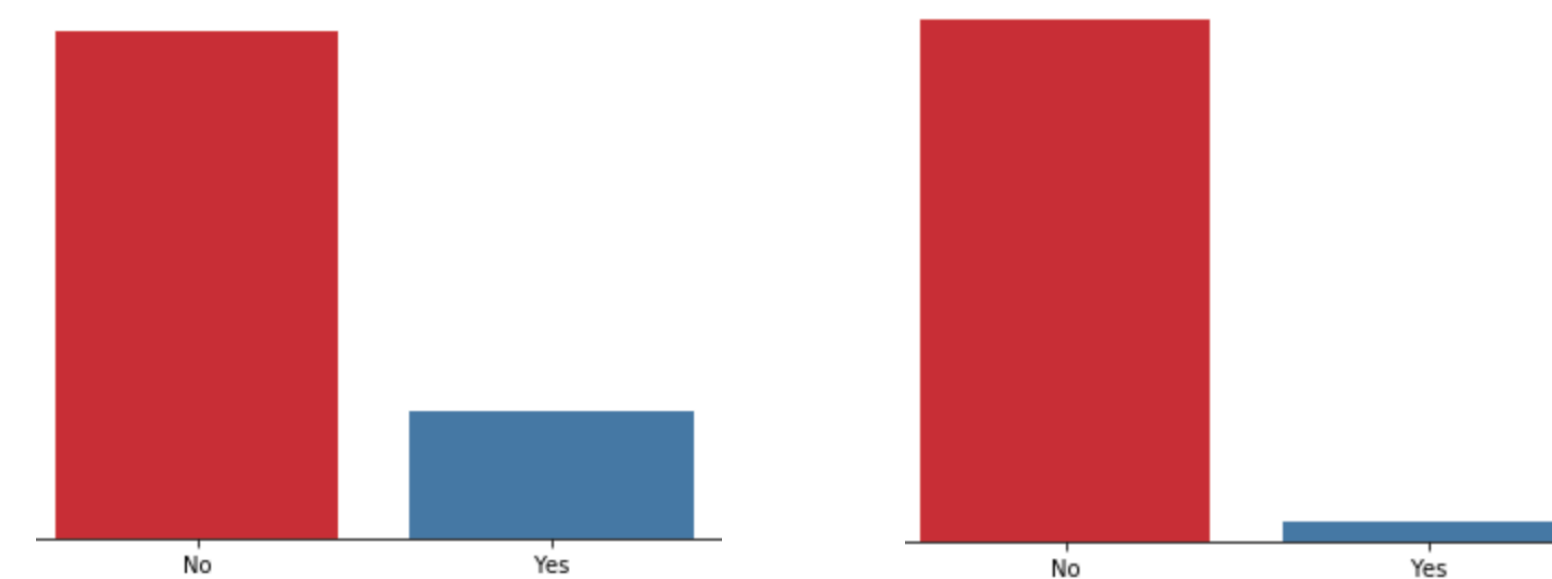
Как видно из графика, количество посещений (записей на посещение) на одного пациента не однородно.

В группах пациентов с менее чем 15 посещениями и пациентов с 30 и более посещениями, наблюдается существенное различие в количестве неявок.

Учитывая что данные предоставлены за не очень большой промежуток времени, вероятно, пациенты с большим количеством записей и малой долей неявок, находились на стационарном лечении и им было сложно пропускать приемы.



Распределение пациентов по количеству посещений



От 1 до 5 посещений

Более 30 посещений

3

ВРЕМЯ ПОСЕЩЕНИЯ И ЗАПИСИ



3

Рассматривая время записи и время прихода на прием, можно отметить что значительное количество приемов происходит в день обращения.

38%

Посещений
происходят в тот же день

35%

Посещений
оформлены «задним числом»*

178 дней

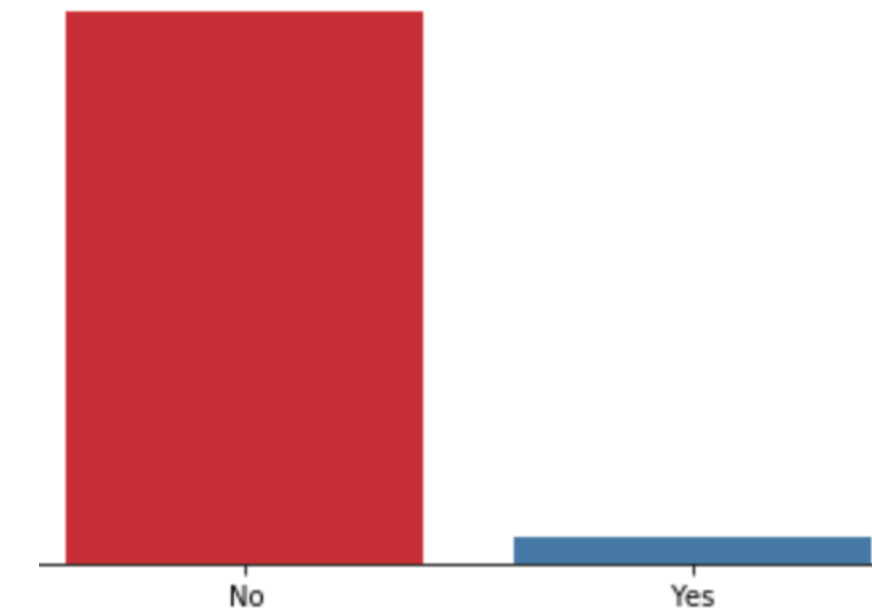
Максимальный период
от записи до посещения

** Похоже что если время записи позже времени прихода на прием, пациента оформляют «задним числом»*

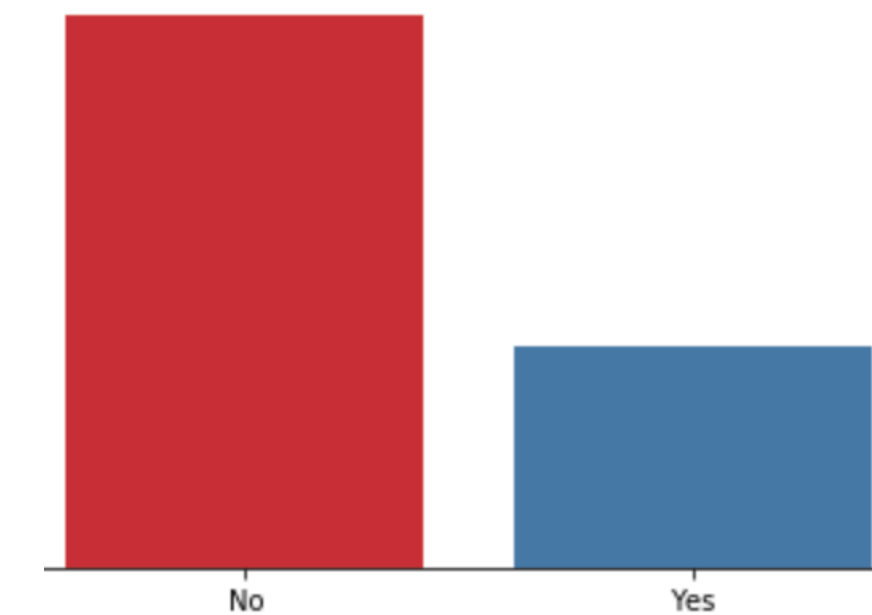
3

Если рассмотреть группы пациентов с разницей явки в -1 день и больше чем -1 день, становится очевидной существенная разница в доле неявки.

Похоже гипотеза об оформлении «задним числом» верна.



Неявка среди оформленных «задним числом»

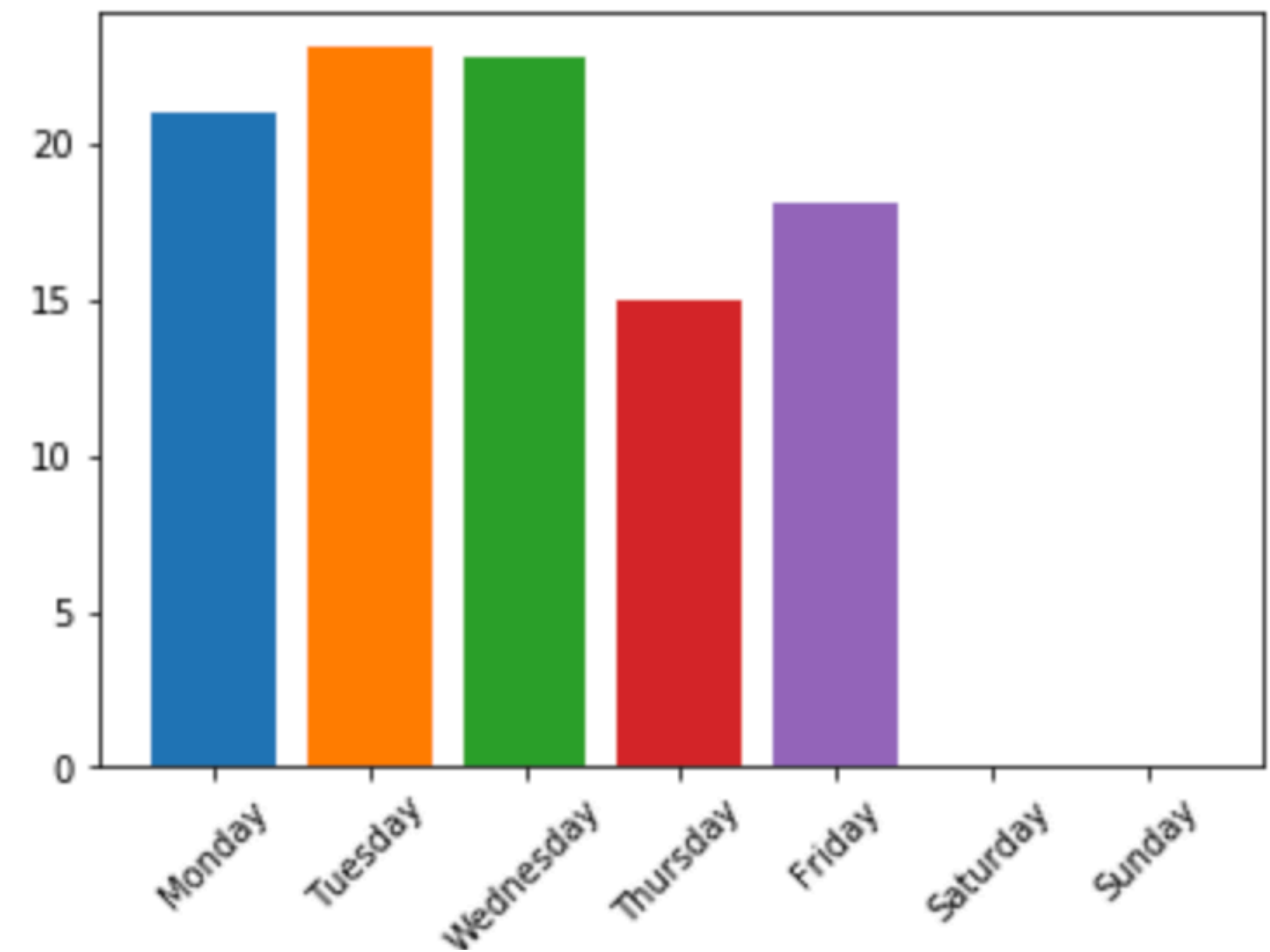


Неявка среди приходящих на прием через какое то время после записи

3

Рассмотрим так же, распределение неявки по дням недели.

Заметно что наибольшая явка происходит по четвергам, а во вторник и среду, она наименьшая.



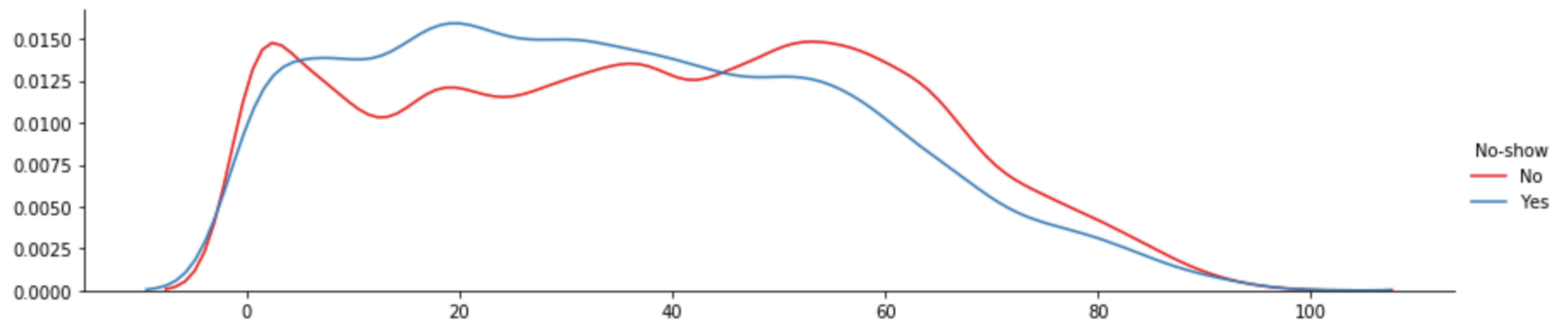
4

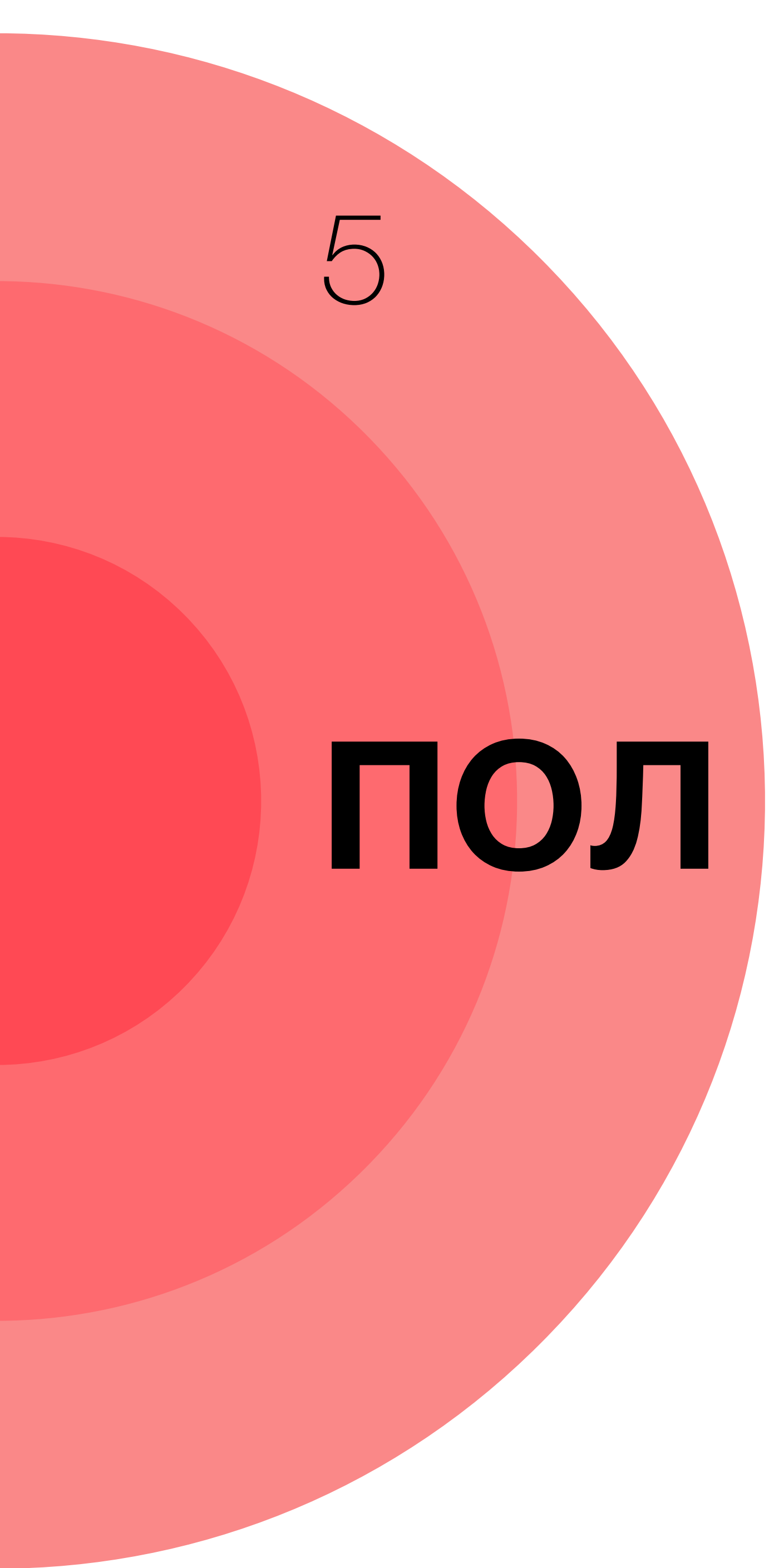
ВОЗРАСТ

4

Явка по возрасту распределена таким образом, что младенцы и маленькие дети, а так-же пожилые люди не появляются на приеме реже средних подростков, молодежи средних возрастных групп.

Вероятно для младенцев и стариков участие в их жизни семьи и общества заставляет более ответственно относиться к приходу ко врачу.





5

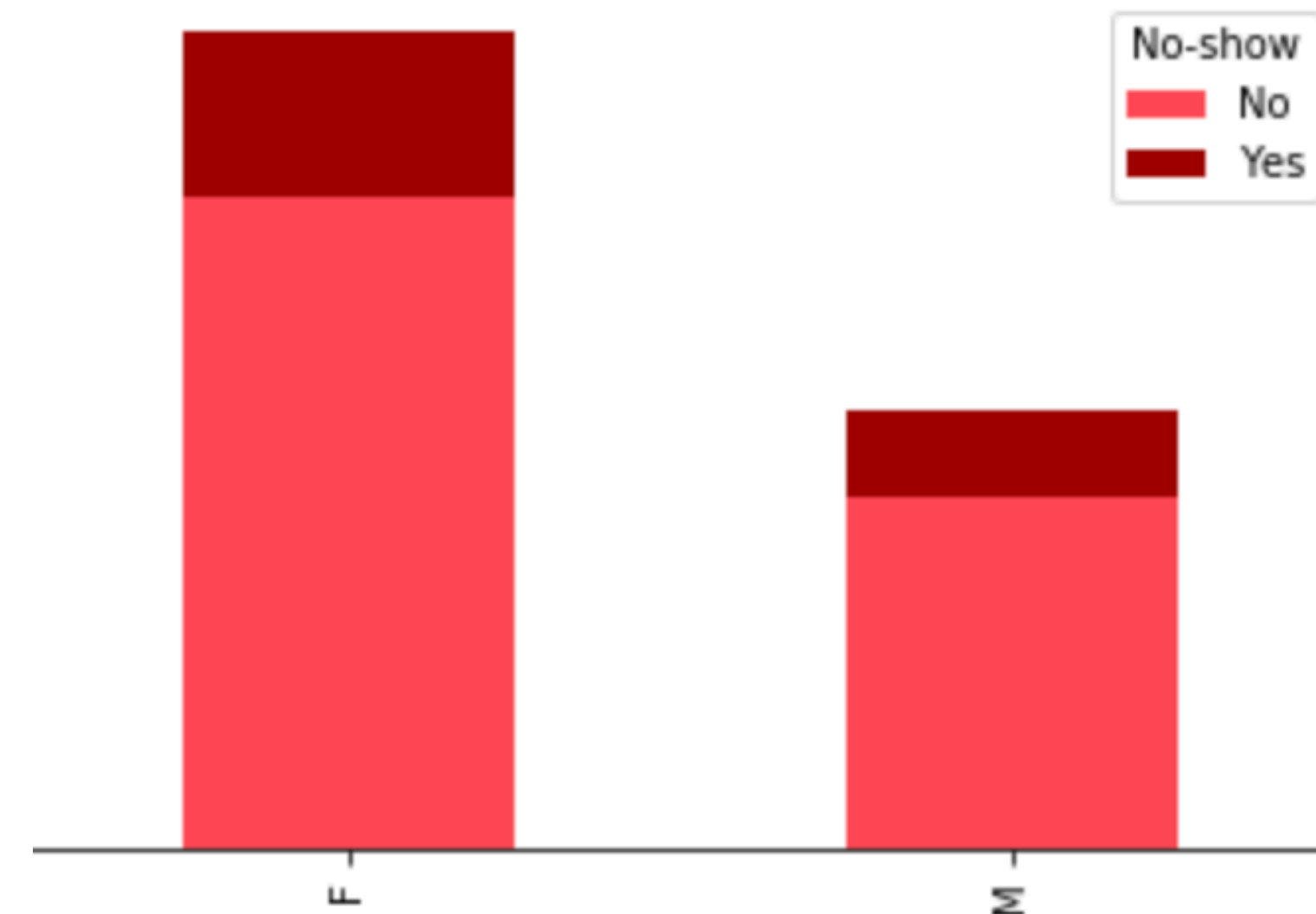
пол

5

Обратим внимание, выборка не сбалансирована по полу.

По видимому это связано с тем что женщины чаще обращаются за медицинской помощью.

Но так же, **неявка среди женщин выше чем среди мужчин.**



7%

Неявка среди мужчин

13.2%

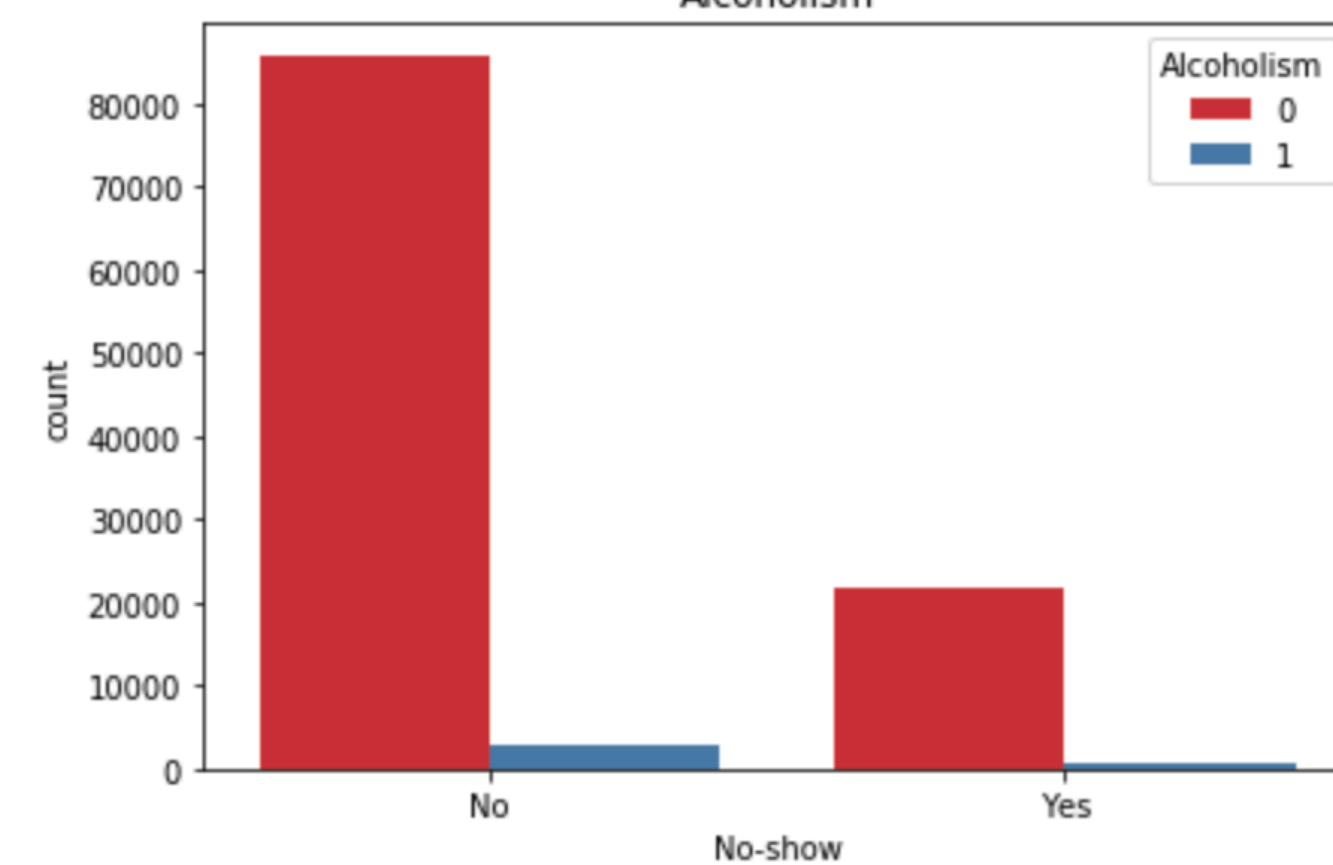
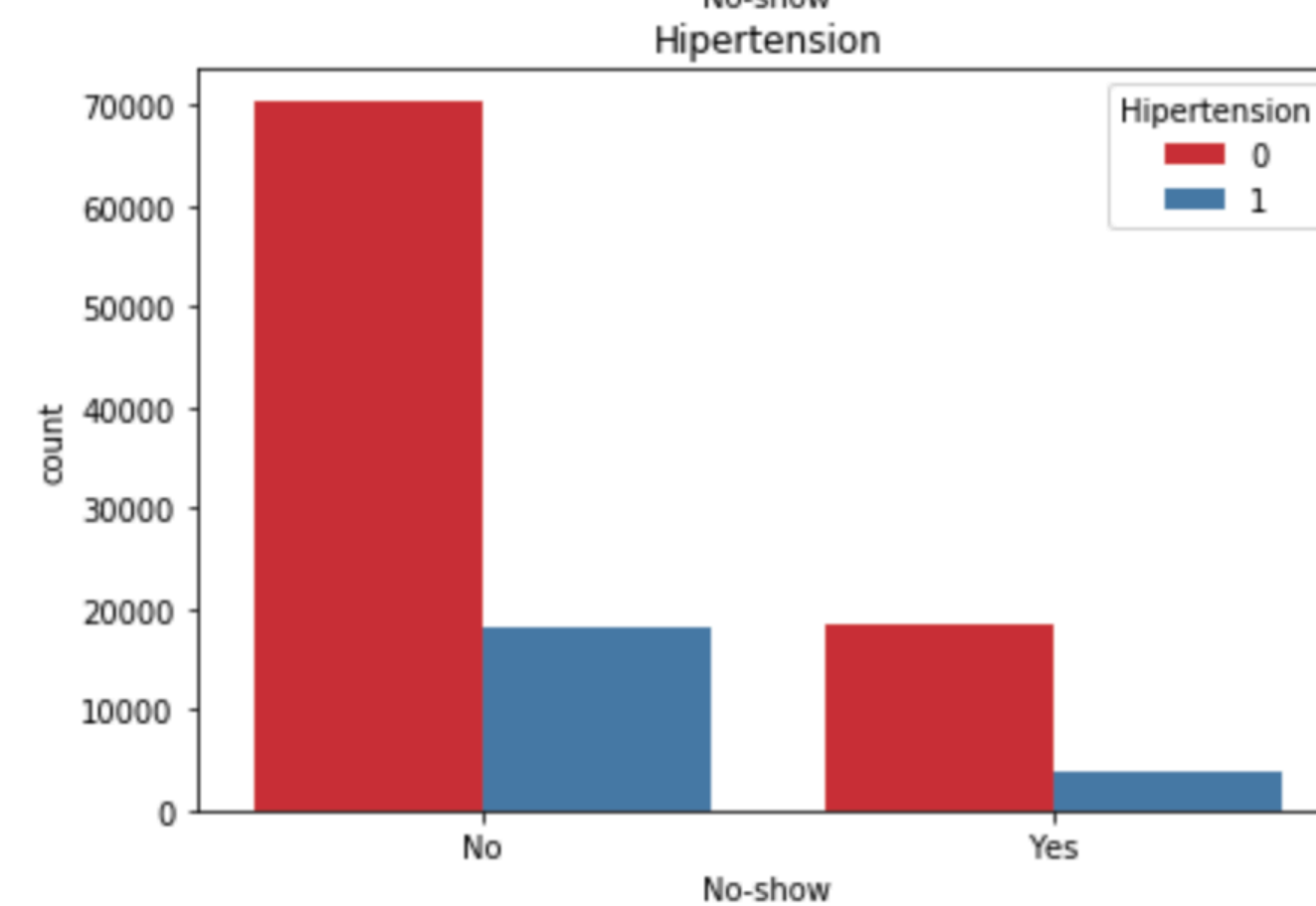
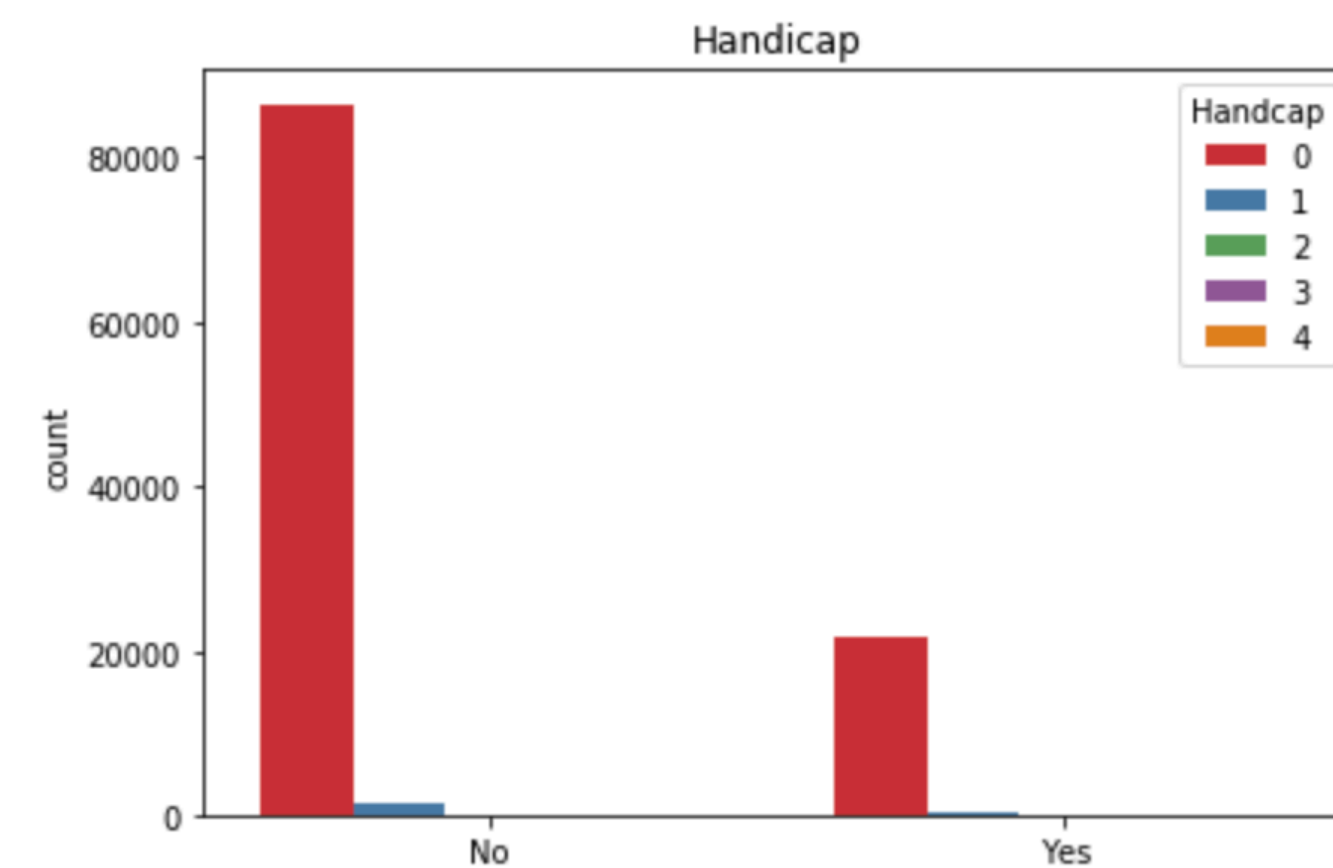
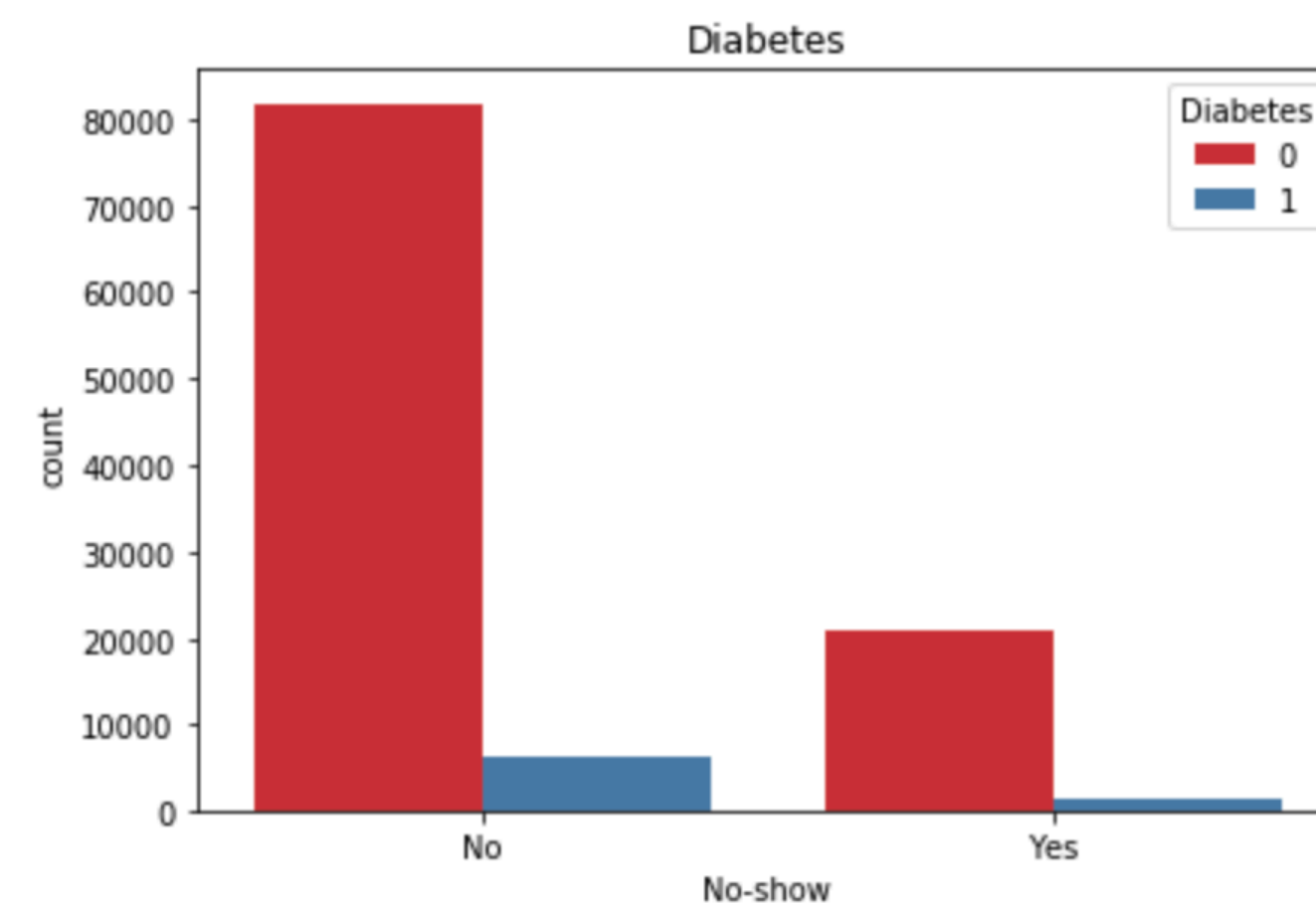
Неявка среди женщин

6

СОПУТСТВУЮЩИЕ ЗАБОЛЕВАНИЯ

6

В целом сопутствующие заболевания не оказывают значительного влияния на пропуск приема.

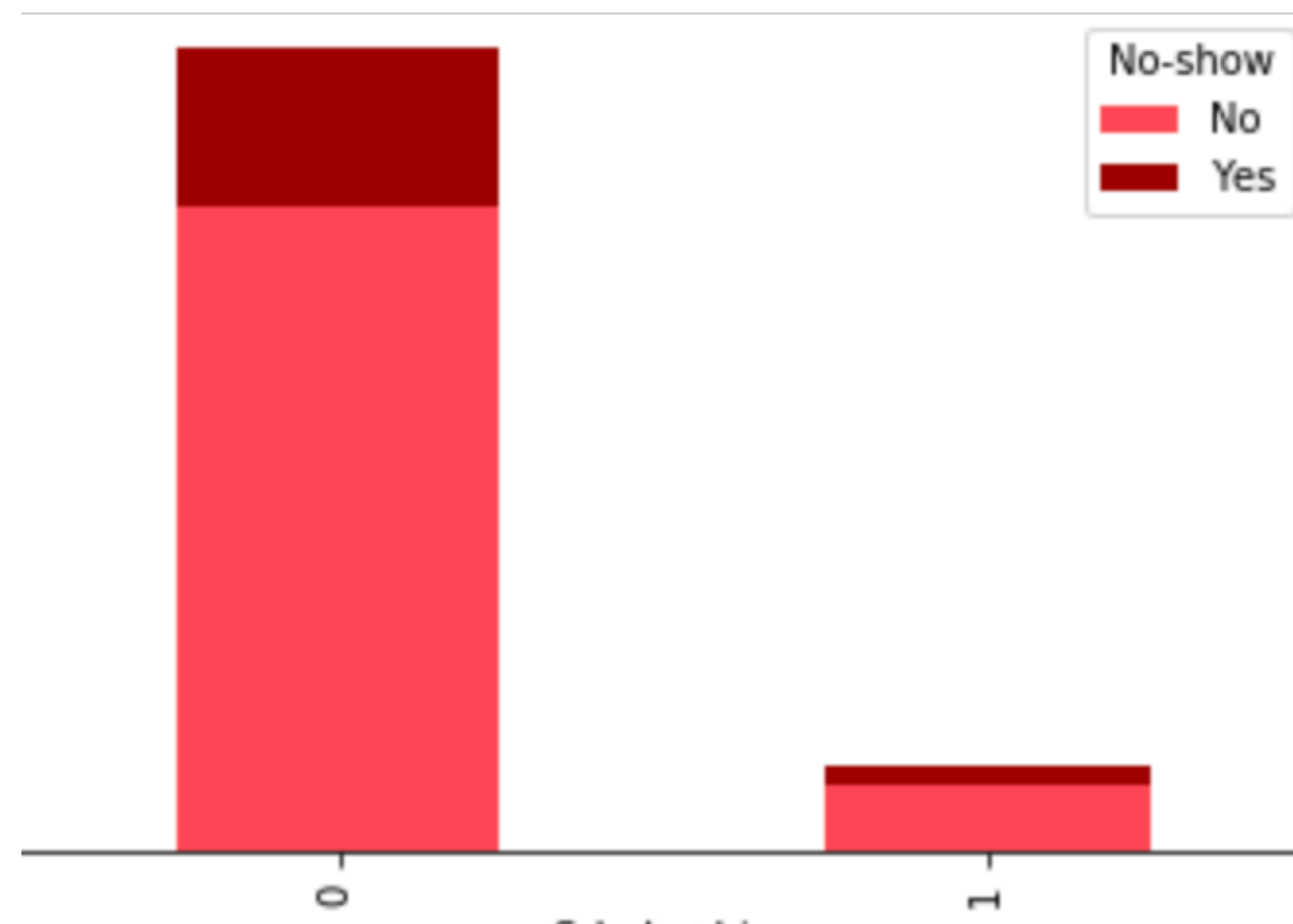


7

СОЦИАЛЬНОЕ ПОСОБИЕ

7

Факт получения пособия так же не
влияет пропуск приема.

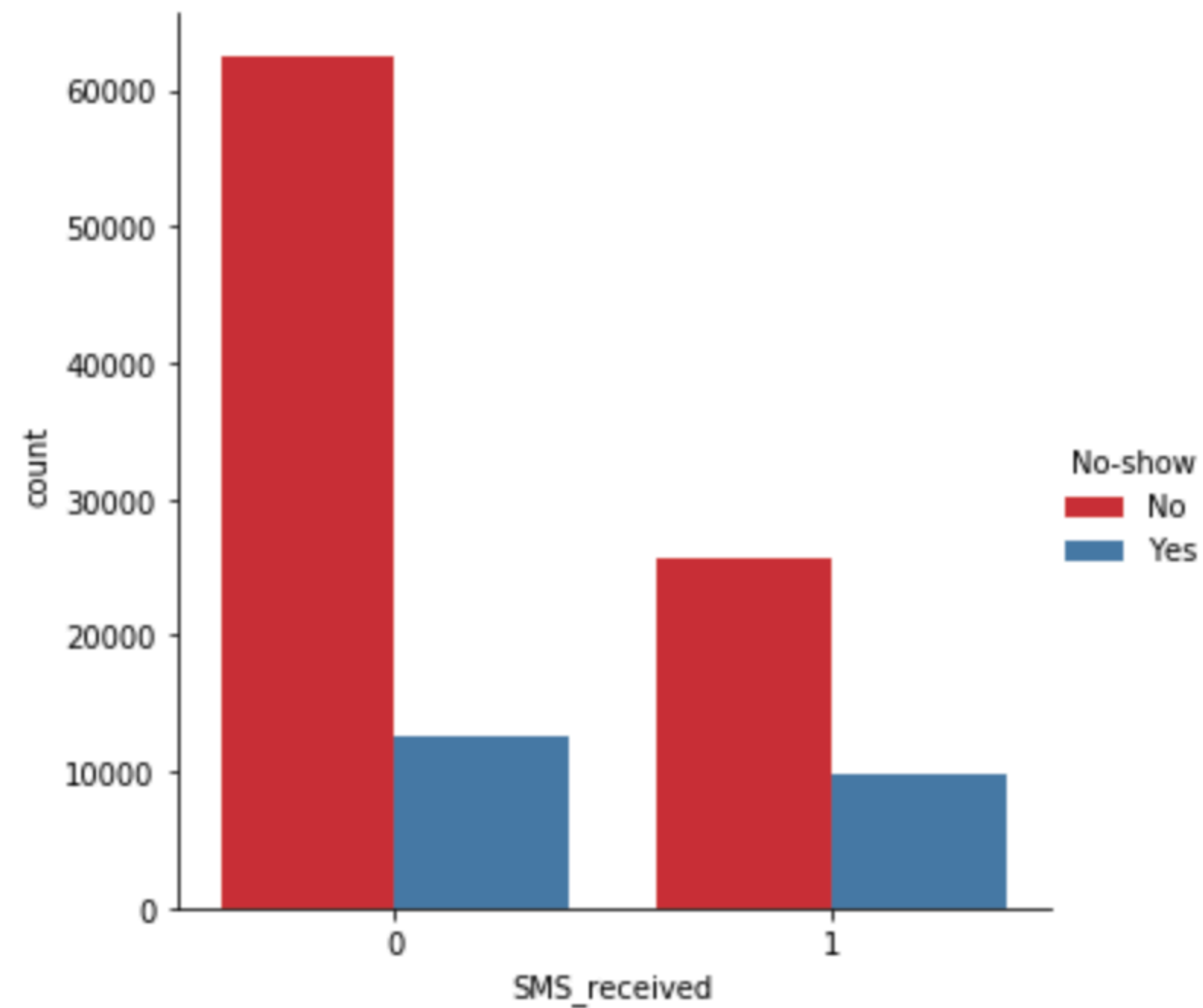


8

SMS оповещение

8

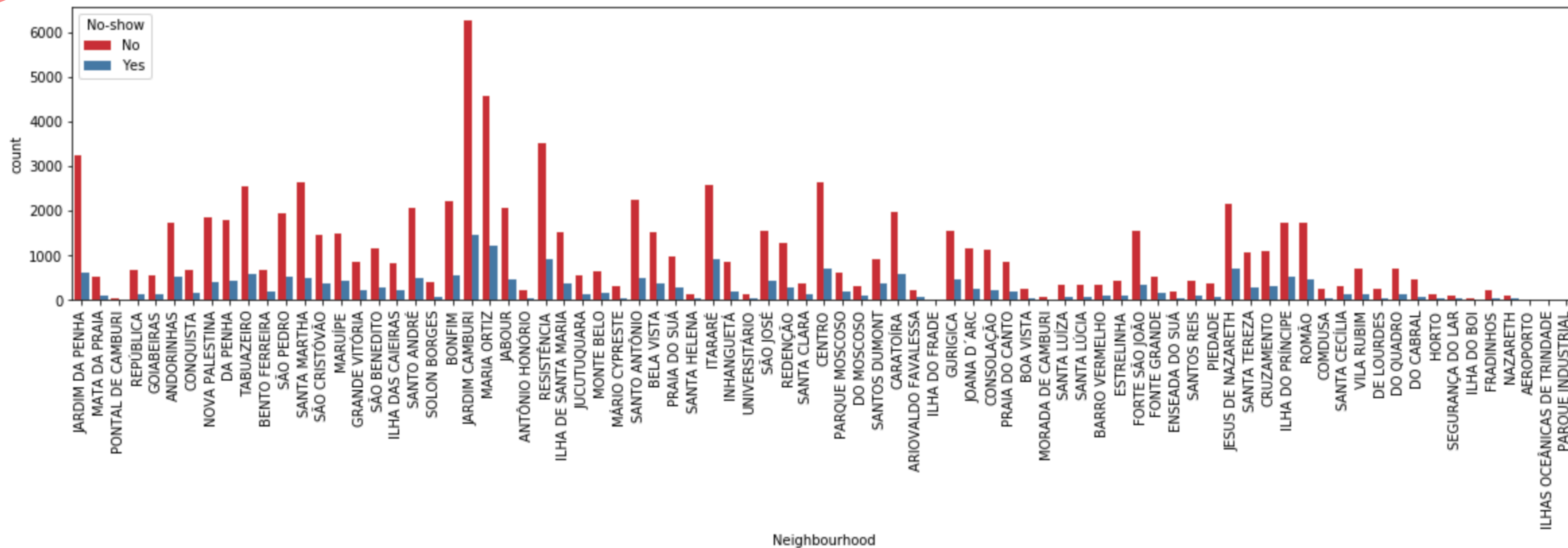
Заметим что 2/3 пациентов не получали SMS, тем не менее они практически так же приходят или не приходят на прием, как и те кто SMS получали



9

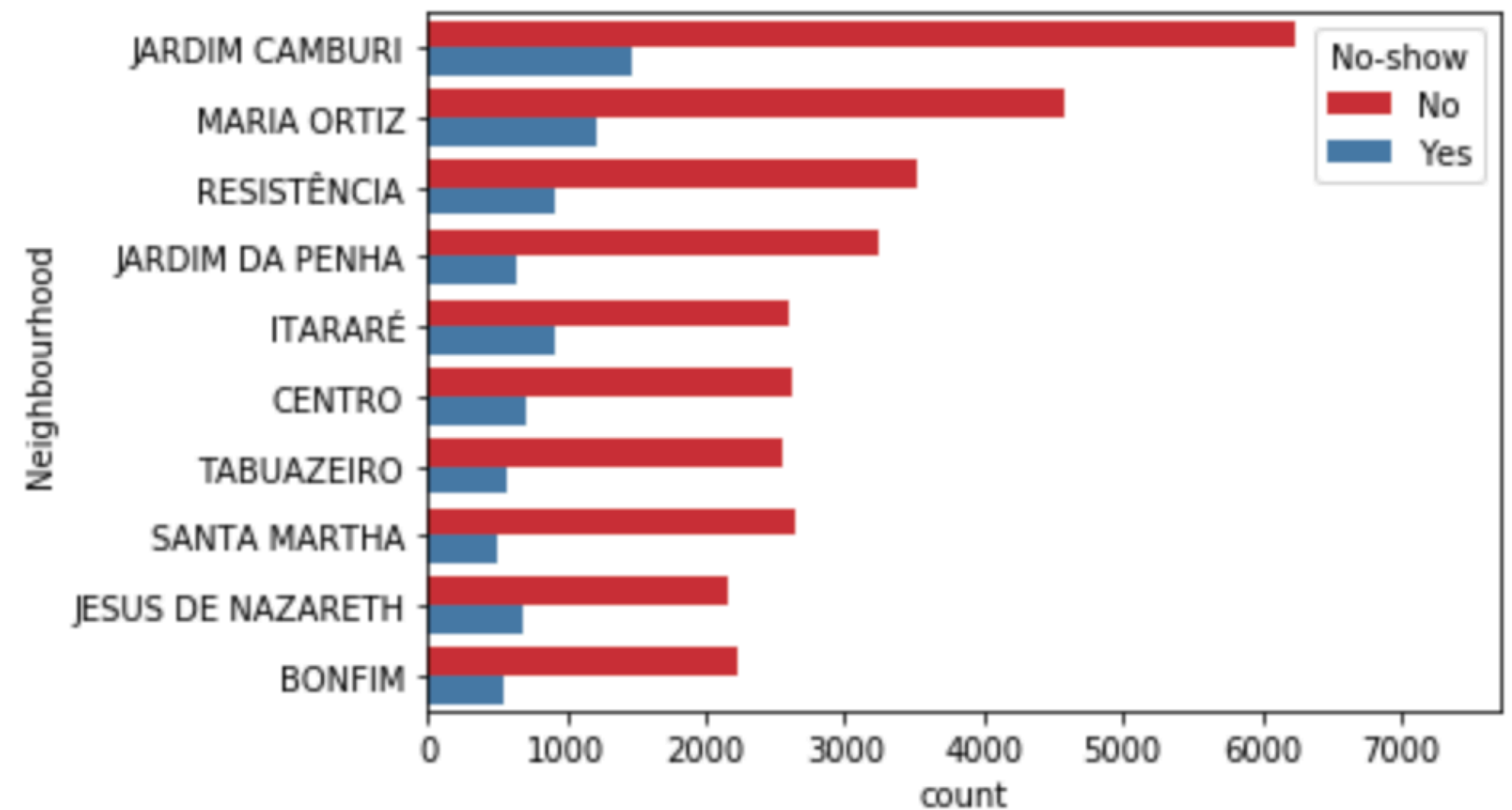
Локации

Посмотрим на распределение посещений/непосещений по локациям



9

Сформируем список десяти самых посещаемых локаций, как мы видим отношение посещений/непосещений напрямую не зависит от общего количества записей



10

Итоги

По итогам анализа можно сделать выводы:

1. Женщины как чаще записываются на приемы, так и чаще их пропускают
2. Наиболее «пунктуальные» посетители это младенцы и пожилые люди
3. Сопутствующие заболевания, не оказывают серьезного влияние на пропуск приемы (что странно, возможно нужна большая выборка)
4. Получение пособия и алкоголизм, хоть и формально влияют на пропуск приема, но как же не оказываю серьезного влияния
5. SMS уведомления минорно влияют на пропуск приема, возможно стоит отказаться от рассылки
6. В четверг наименьший процент пропусков приемов, а в среду и четверг наибольший. В целом чем меньше времени прошло после записи, тем более вероятен приход на прием.
7. Есть ряд закономерностей в расположении локаций, но для их анализа необходима дальнейшая обработка данных.



9

Что можно улучшить

1. Ввести отметку о лечении в стационаре (возможно это отдельная страт. группа)
2. Для корректного изучения временных закономерностей, необходим больший отрезок времени
3. Сгенерировать дополнительные данные по локациям, возможно там будут интересные закономерности

Telegram: @aburakov



Спасибо!