

# Нижегородцы Team #8

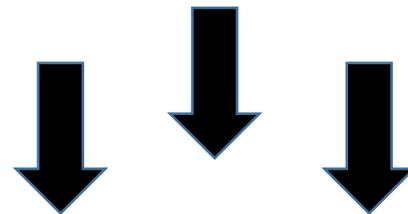


# СТРУКТУРА БАЗЫ ДААННЫХ

5 количественных

16 качественных

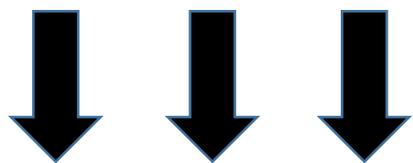
**ПЕРЕМЕННЫХ**



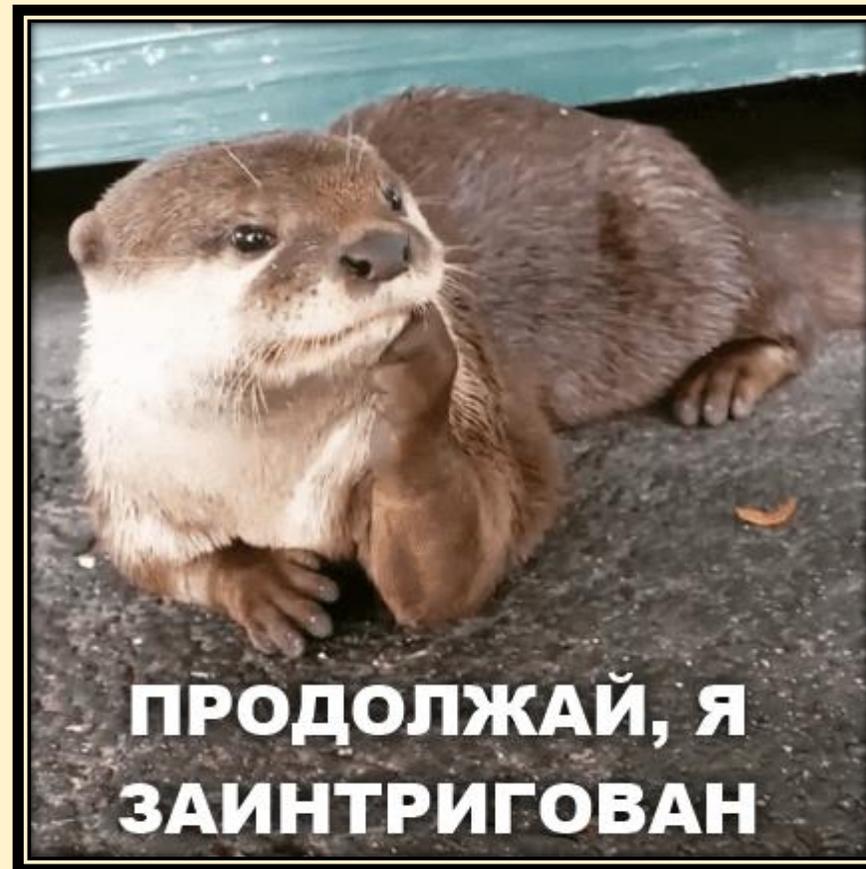
97307 строк

# Предобработка

- Обработали пропуски
- Рассмотрели штрафы только за превышение скорости
- Убрали выбросы
- Ввели новые переменные
- Разделили людей на подвыборки

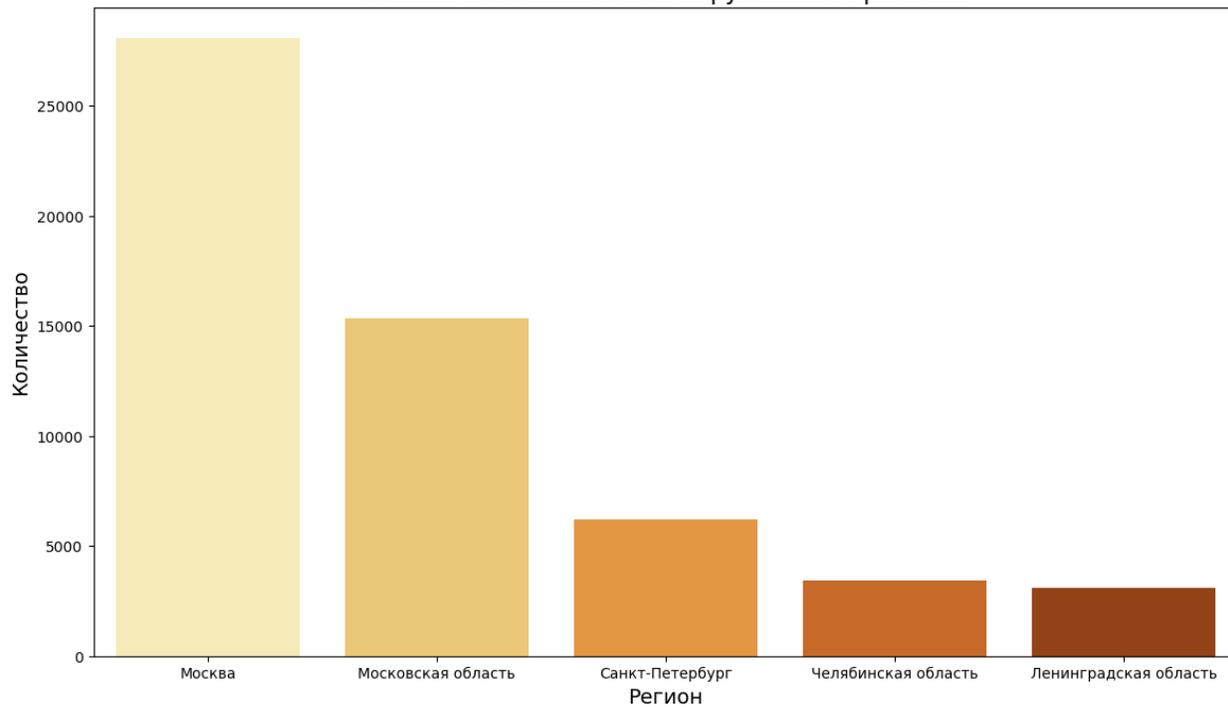


69237 строк

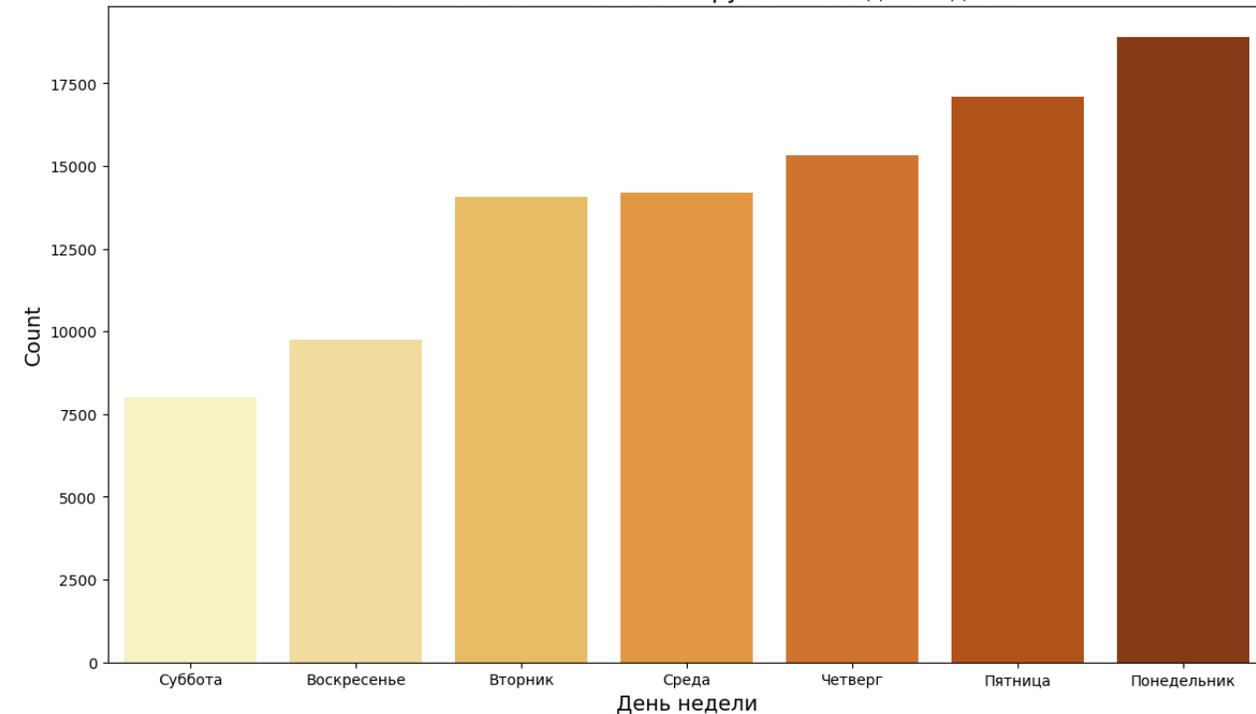


# Предварительный анализ данных

Зависимость количества нарушений от региона

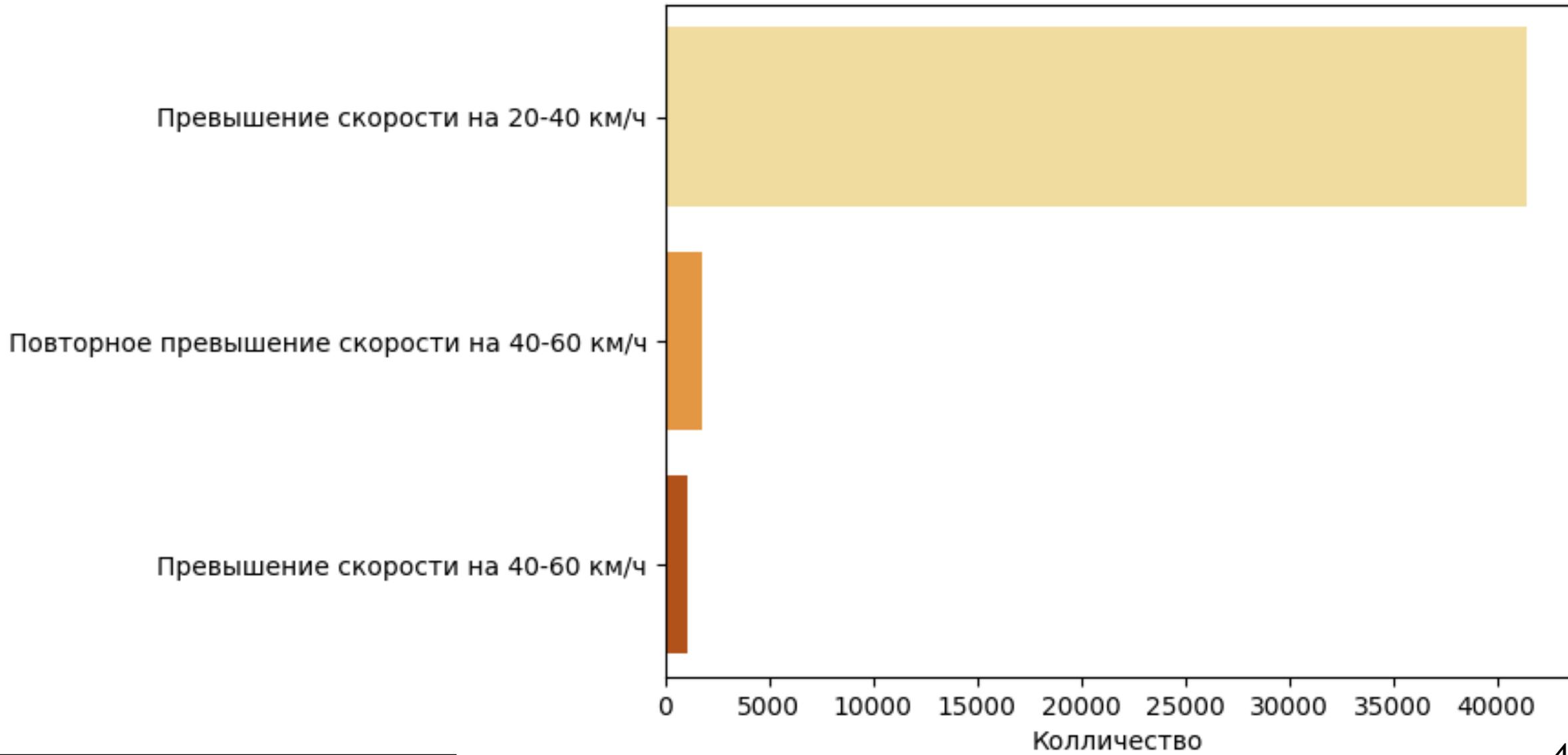


Зависимость количества нарушений от дня недели

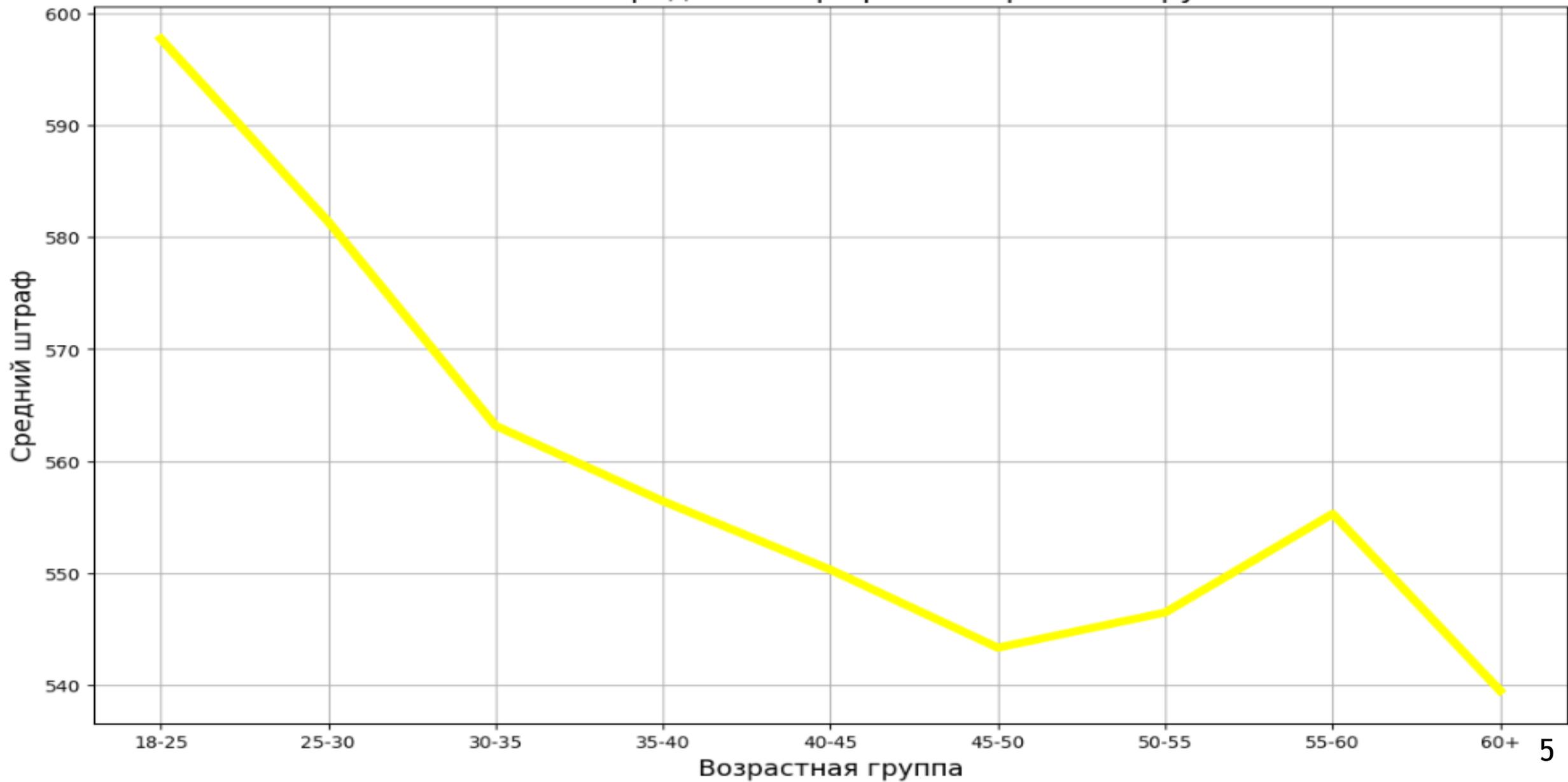


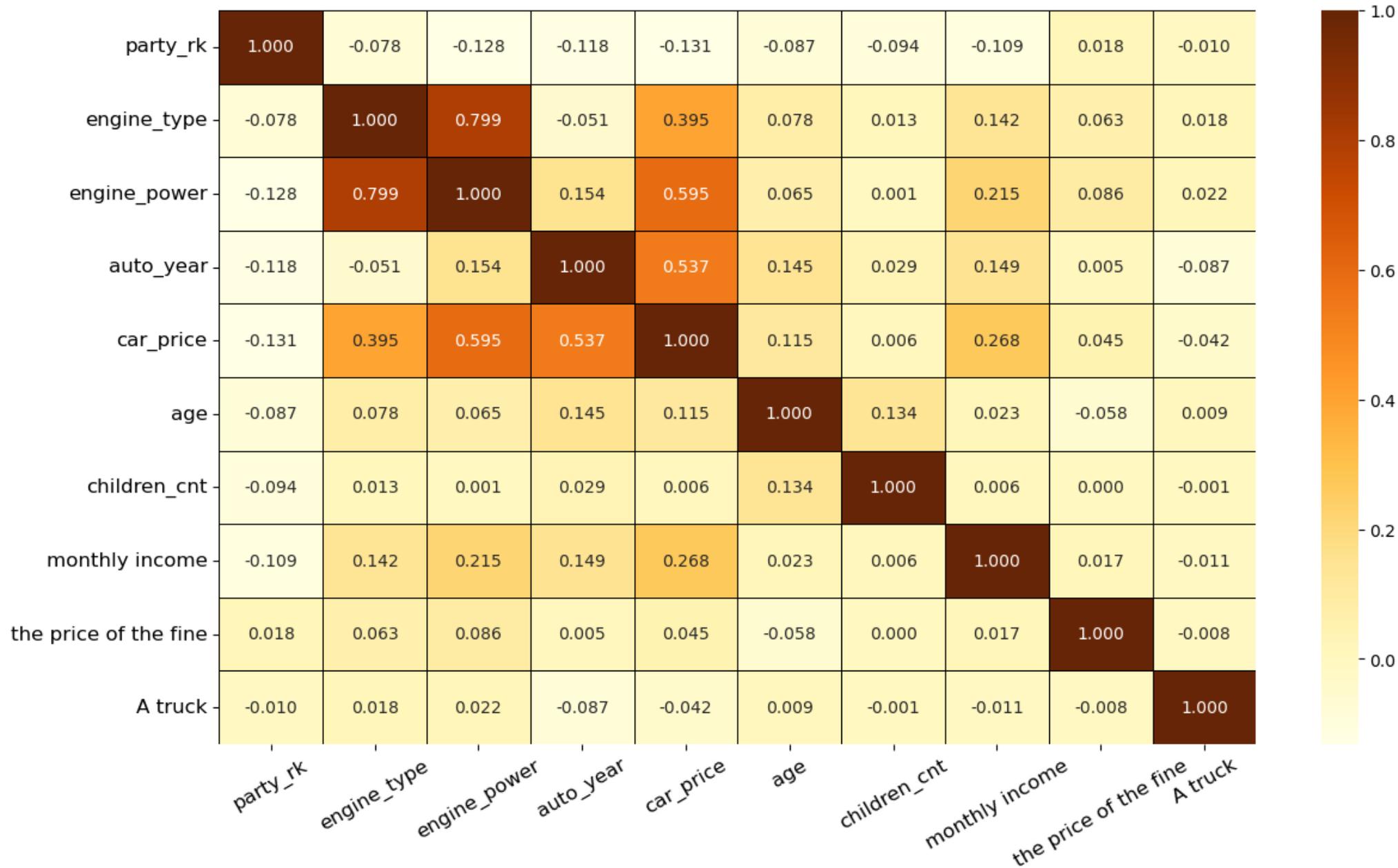
### Распределение штрафов за превышение скорости

Типы штрафов за превышение скорости



Зависимость среднего штрафа от возрастной группы





## Создание новых переменных

truck

\*Переменная обозначает, является ли автомобиль грузовым:

0 - не является,  
1 - является

67035-легковые

2202-грузовые

price\_of\_the\_fine

\*Переменная обозначает сумму штрафа за определенное нарушение:

“Информацию о суммах штрафов брали с открытых источниках информации”

# Исследовательский вопрос

Взаимосвязаны ли характеристики клиентов и их автомобилей с характеристиками правонарушений и если да, то как?

# Наша гипотеза

Нулевая гипотеза (H0) :

С возрастом у людей, вне зависимости от вида транспорта, не меняется средний штраф.

Альтернативная гипотеза (H1) :

С возрастом у людей, вне зависимости от вида транспорта, меняется средний штраф.

# Как это работает?

Человек, вне зависимости от возраста и вида транспорта, аккуратно водит, так как считает важным соблюдение пдд

Получает штрафы редко и только по неосторожности

С возрастом мнение человека не меняется

Не меняется средняя стоимость штрафа

Человек не аккуратно водит, так как считает соблюдение пдд не важным

Человек чаще получает штрафы и игнорирует их

С возрастом мнение человека меняется

Меняется средняя стоимость штрафа

# МАТ.МОДЕЛЬ – Критерий Манна – Уитни

## Условия применения критерия:

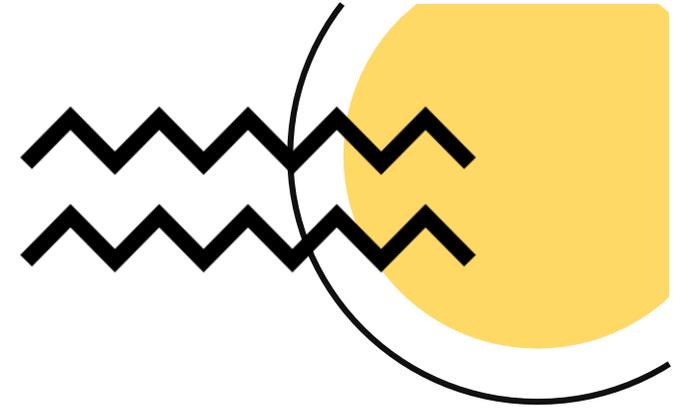
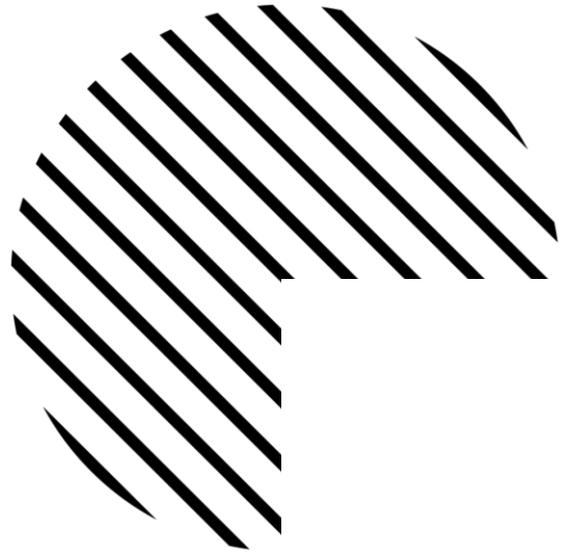
- Распределение ненормальное
- Выборки независимы (две группы никак не связаны между собой)



# Преобразование данных



ДААННЫЕ	МЕТОДЫ	РЕЗУЛЬТАТЫ	ОБСУЖДЕНИЕ	ПОДВАЛ
Возрастная группа	P-value	P-value (Мск,проверка устойчивости)	P-value (Ленинградская область,проверка устойчивости)	P-value (Мск,проверка устойчивости)
18-25	≈0.25	≈0.15	≈0.12	≈0.17
25-30	≈0.3	≈0.41	≈0.15	≈0.2
30-35	≈0.22	≈0.04	≈0.15	≈0.04
35-40	≈0.53	≈0.29	≈0.43	≈0.37
40-45	≈0.4	≈0.44	≈0.49	≈0.32
45-50	≈0.32	≈0.21	≈0.26	≈0.49
50-55	≈0.62	≈0.39	≈0.2	≈0.11
55-60	≈0.5	≈0.39	≈0.14	≈0.33

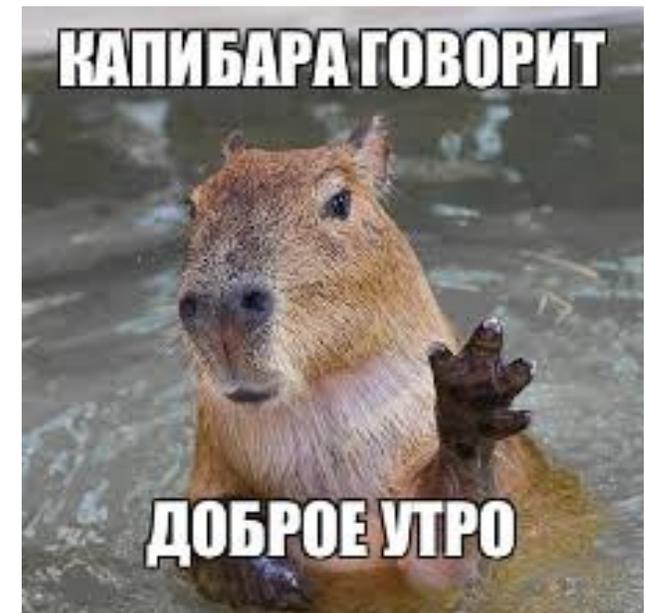


# Гипотеза опровергнута



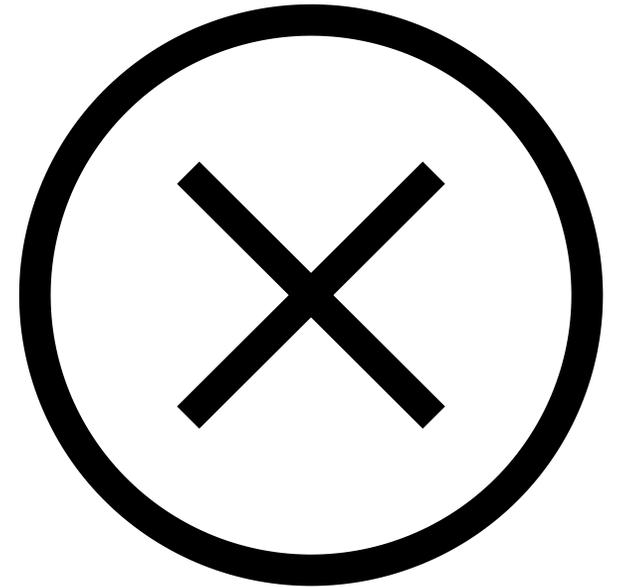
# Policy implication

1. Введение рейтинга водителя позволит понимать, насколько водитель безопасен на дорог.
2. Увеличение суммы повторного штрафа для того, чтобы стимулировать водителя не повторять нарушение.



# ОГРАНИЧЕНИЯ

1. Данные за 1 месяц.
2. Всего 0.5 % данных от всех.
3. Данные одного банка.

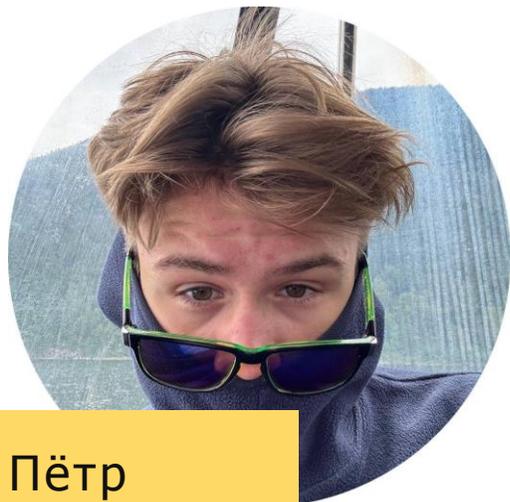


# Перспективы

- 1.Время, спустя которое был оплачен штраф.
- 2.Собирать информацию о стаже вождения.



# Нижегородцы



Пётр



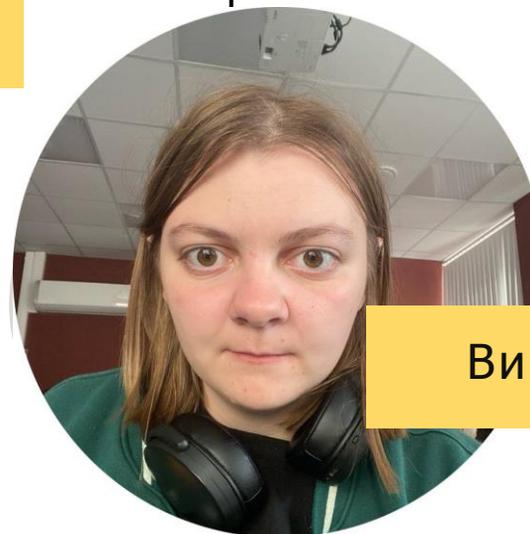
Владимир



Глеб

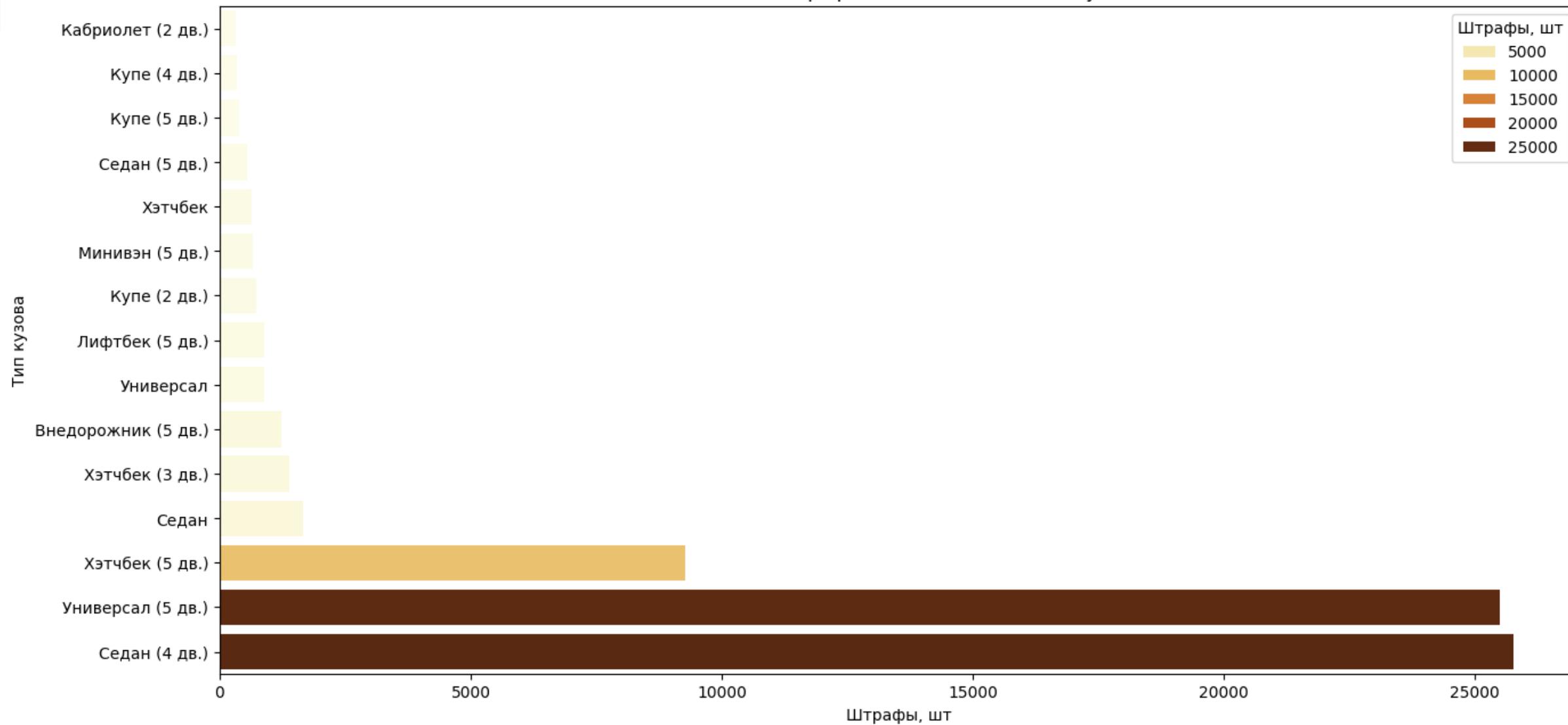


Кирилл

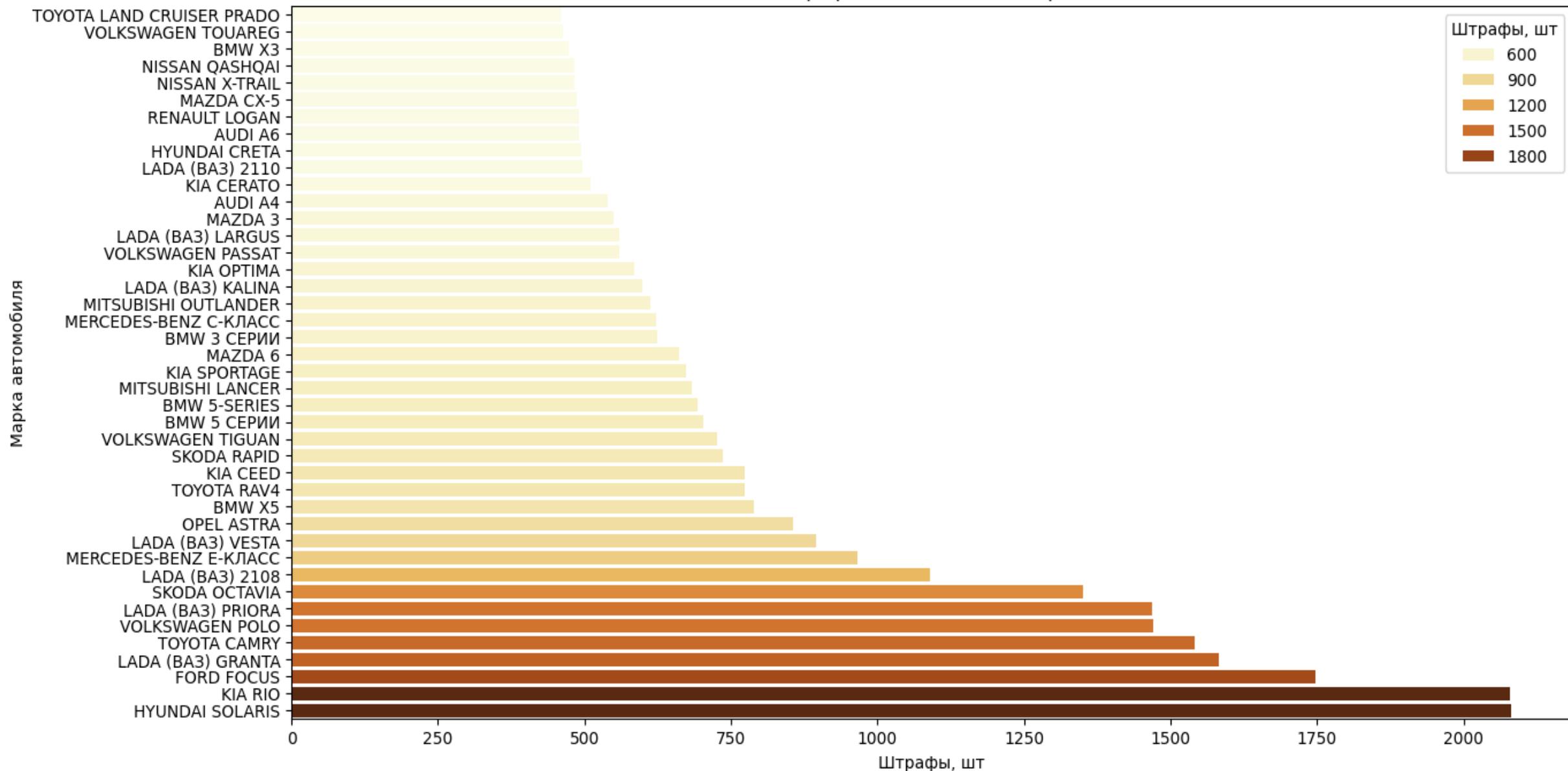


Виктория

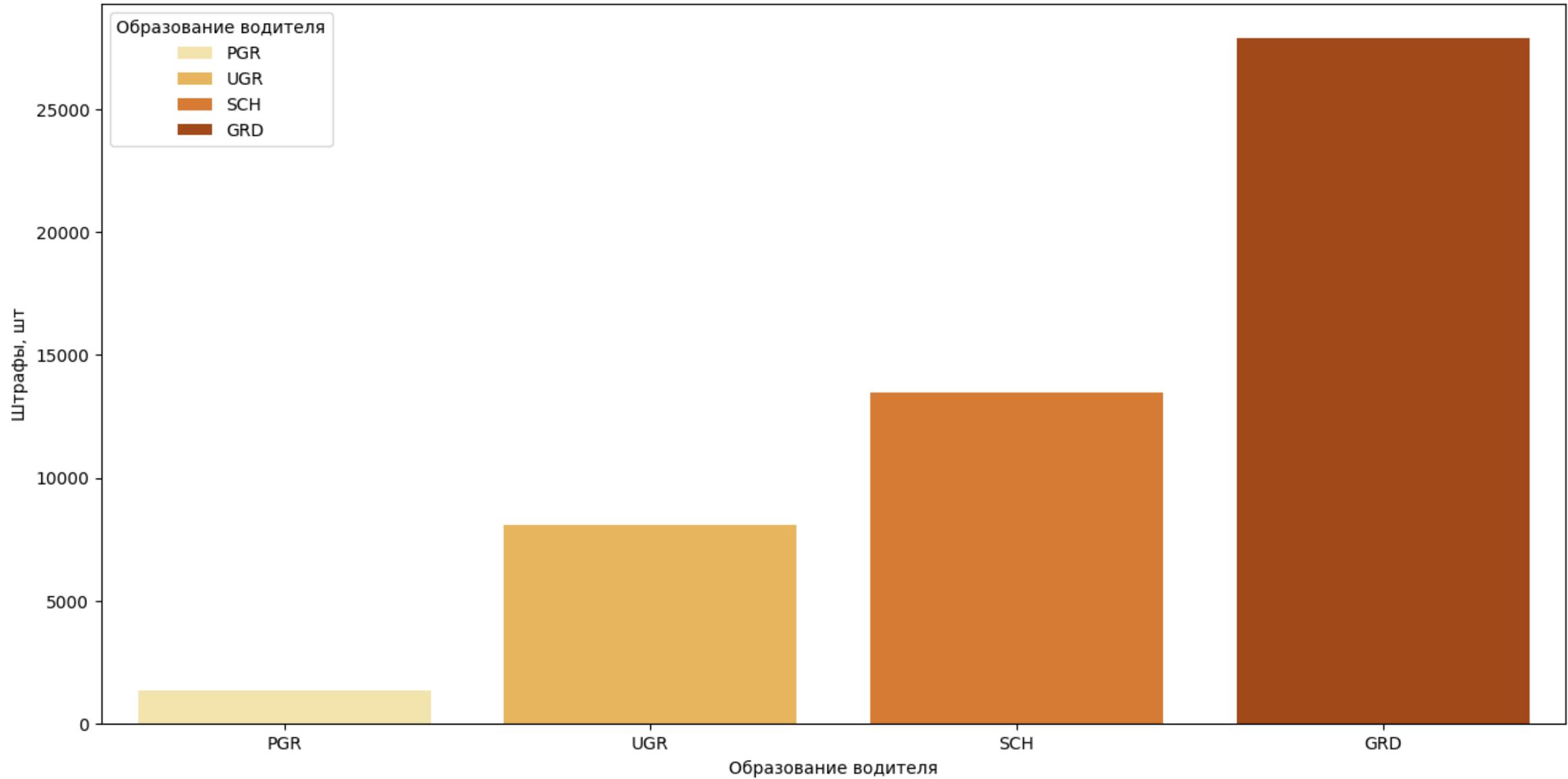
Количество штрафов относительно типа кузова



Количество штрафов относительно марки автомобиля



Количество штрафов относительно уровня образования водителя



ДАННЫЕ

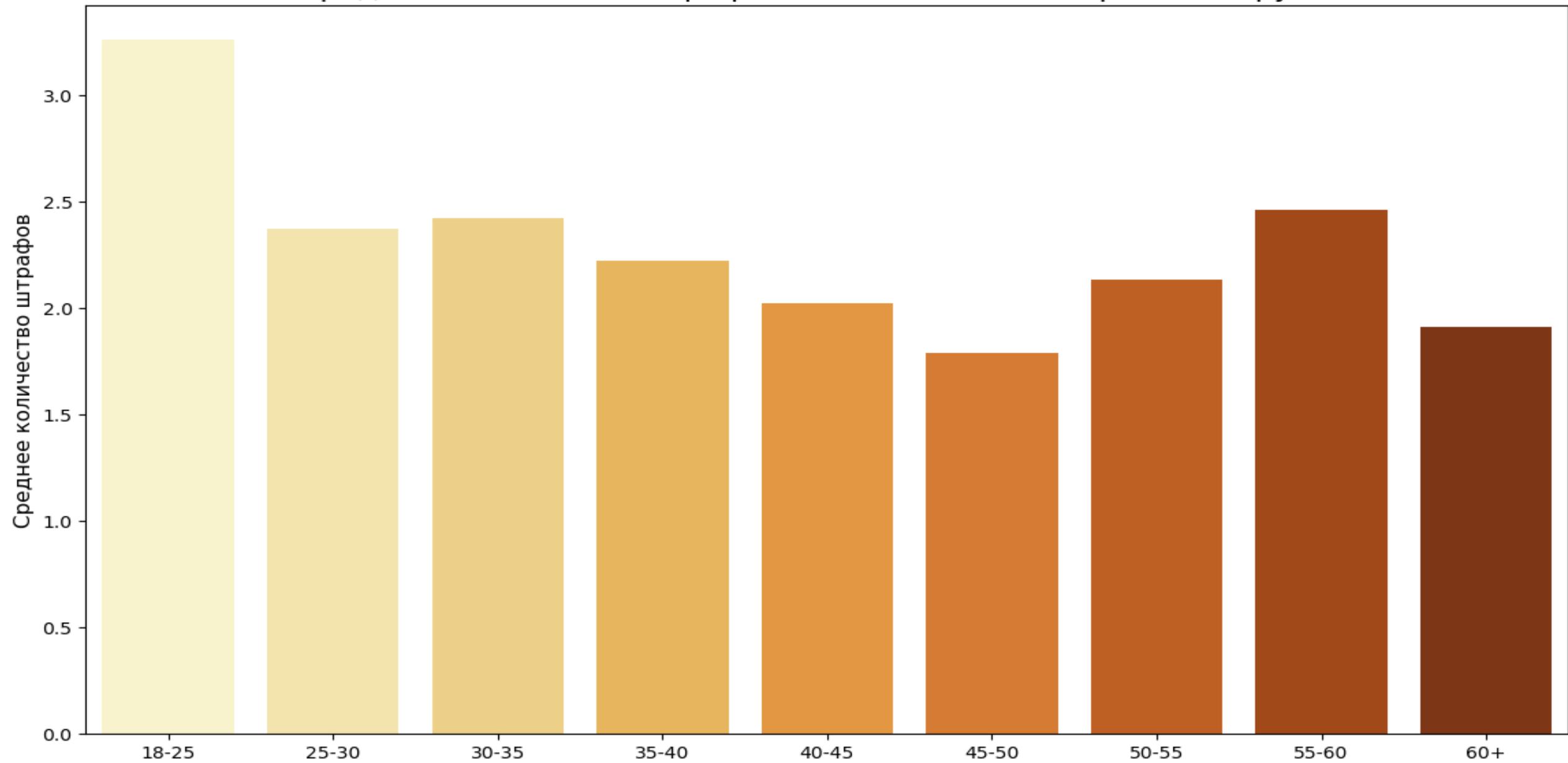
МЕТОДЫ

РЕЗУЛЬТАТЫ

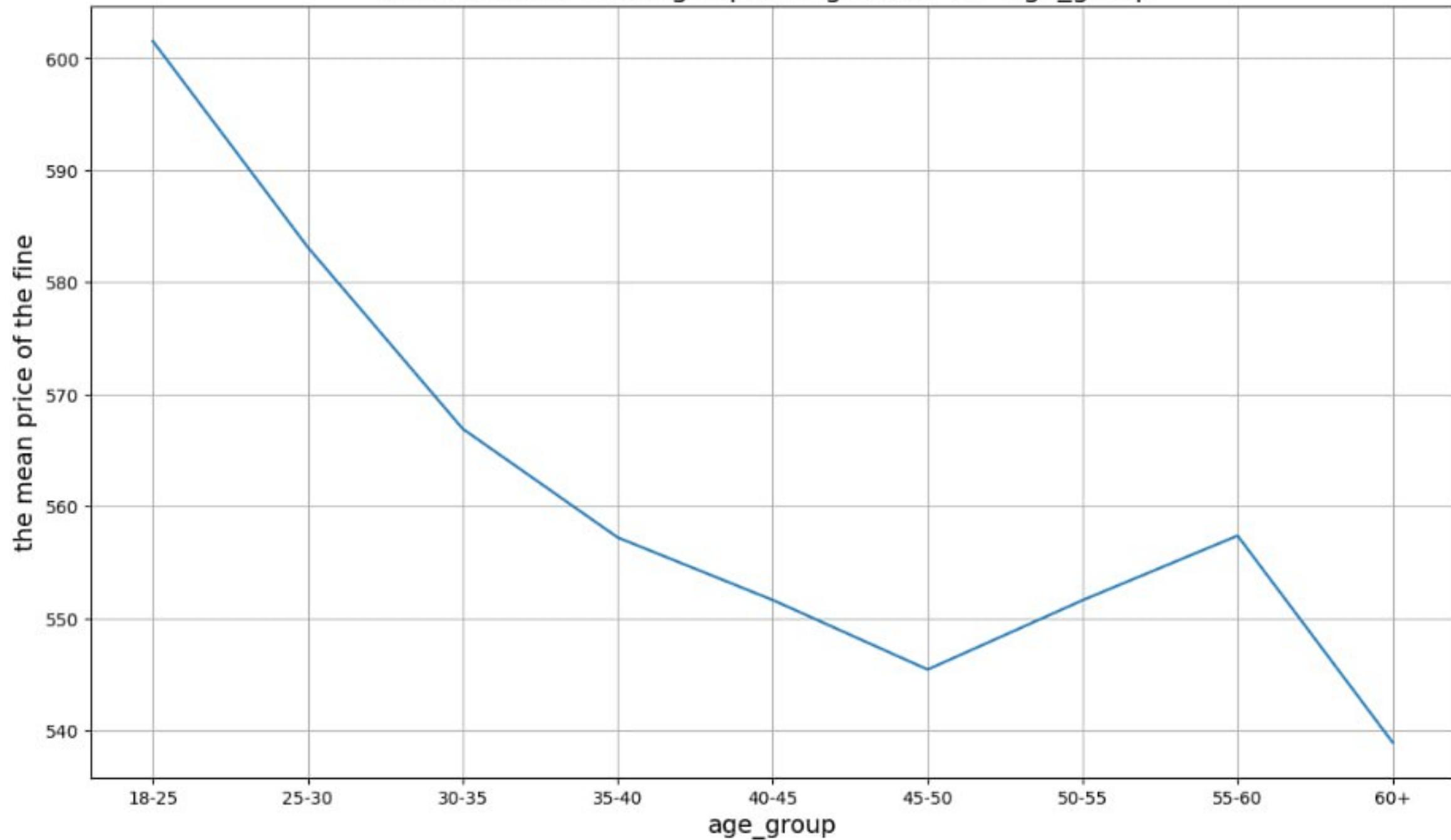
ОБСУЖДЕНИЕ

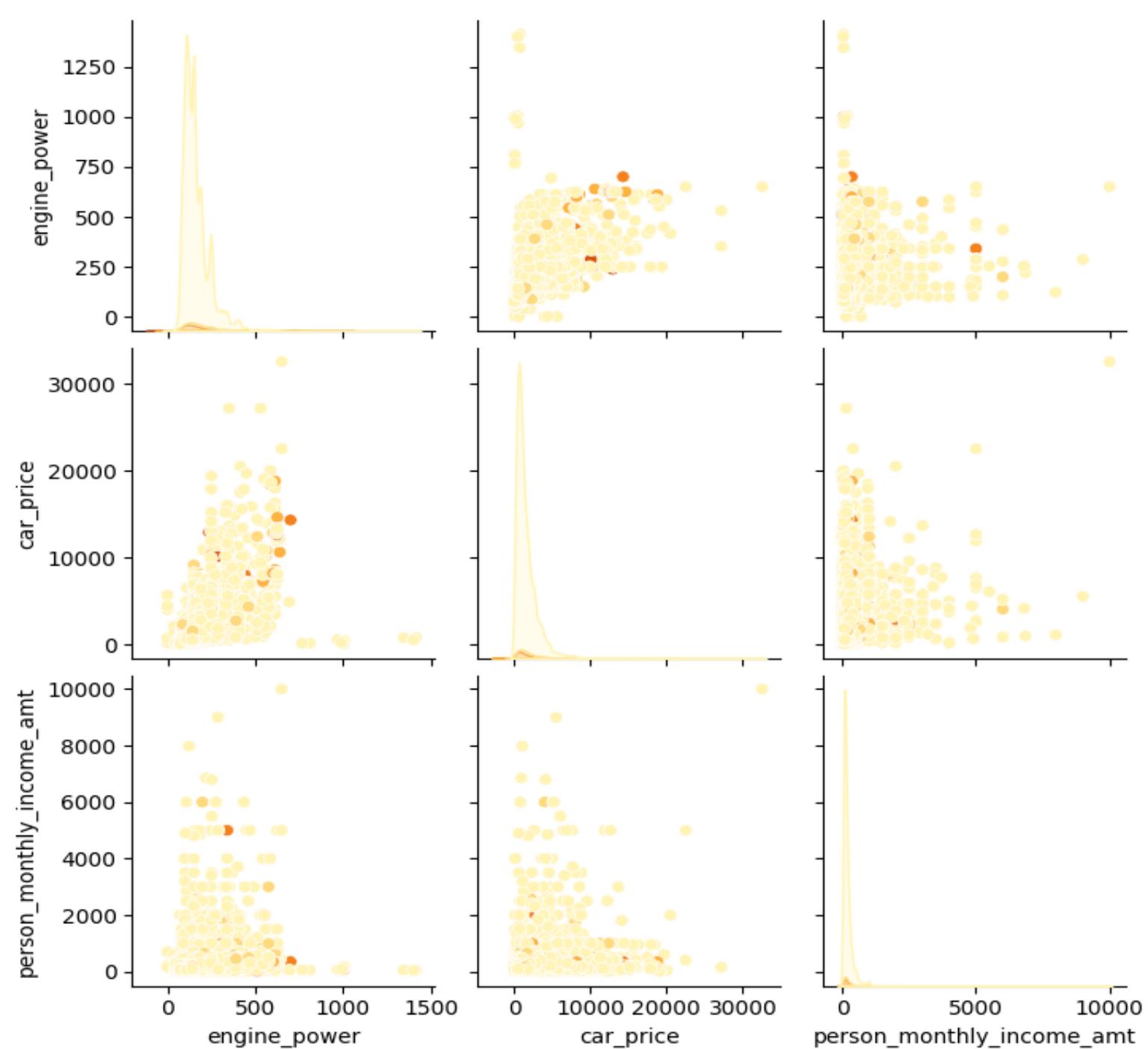
ПОДВАЛ

## Среднее количество штрафов на человека от возрастной группы



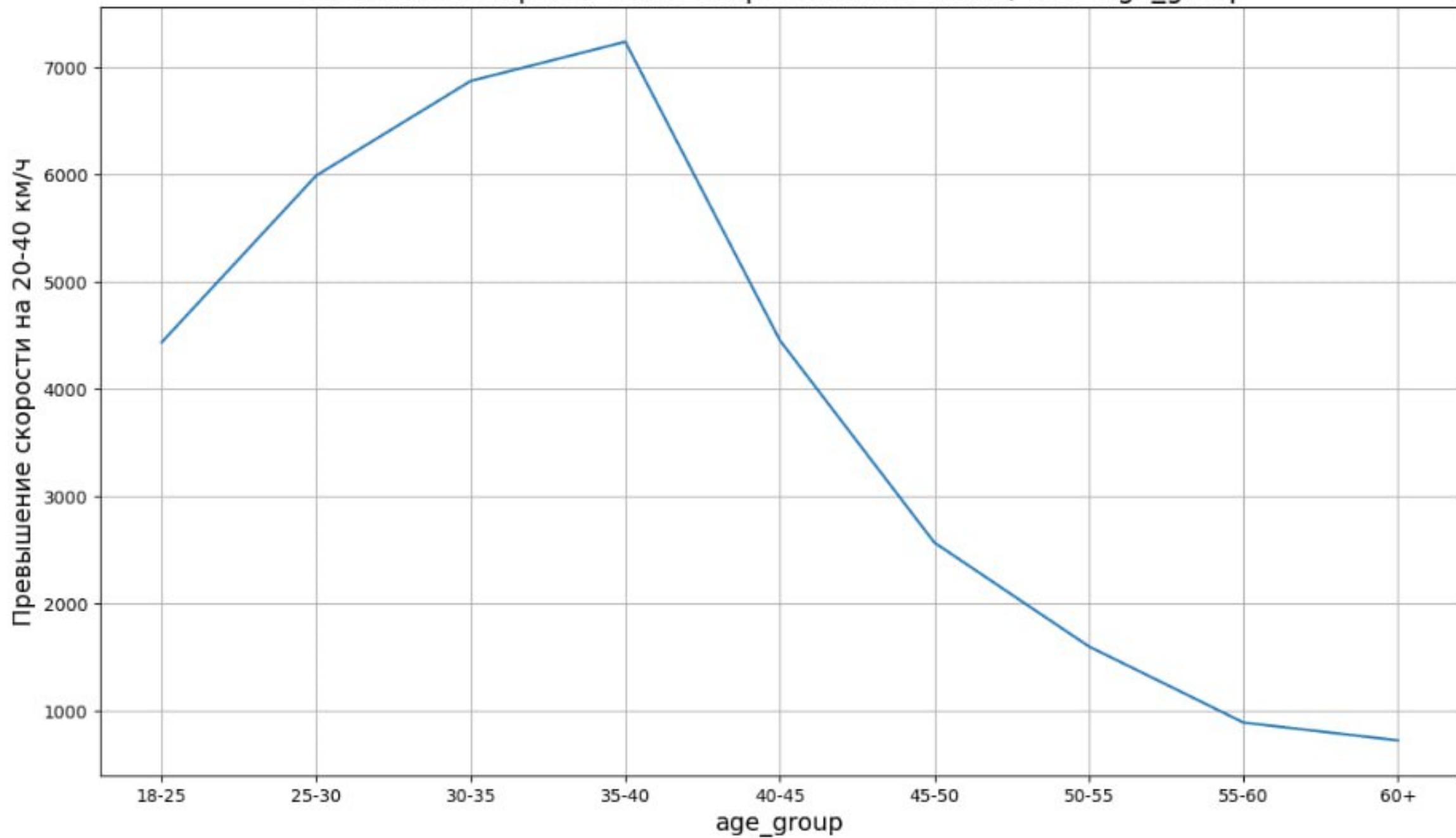
Зависимость "average speeding ticket" от "age\_group"





- offenceshortstatement
- Превышение скорости на 20-40 км/ч
  - Повторное превышение скорости на 40-60 км/ч
  - Повторное превышение скорости более чем на 60 км/ч
  - Превышение скорости на 40-60 км/ч
  - Превышение скорости на 60-80 км/ч
  - Превышение скорости более чем на 80 км/ч

Зависимость Превышение скорости на 20-40 км/ч от "age\_group"



Количество штрафов относительно заработной платы

