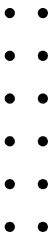


Штрафы ГИБДД

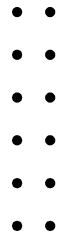
ВШЭстером





01

ДАННЫЕ

- Датасет
 - Основные показатели
 - Обработка выбросов
- 



Датасет

Статистика
штрафов ГИБДД

Записи содержат
информацию по
клиентам и их авто

Данные
предоставлены Т-
банком

- Период с 28 апреля по 28 мая 2024 г.
- Одна запись - один уникальный штраф ГИБДД

Основные показатели

- **age** - возраст водителя
- **gender_cd** - пол водителя
- **offencetime** - время совершения правонарушения
- **typesofday** - время суток
- **engine_type** - рабочий объем всех цилиндров двигателя автомобиля
- **engine_power** - мощность двигателя автомобиля
- **auto_year** - год выпуска авто
- **severity** - тяжесть правонарушения
- **person_monthly_income_amt** - зарплата

Чистка данных

97307



БД почищена от пустых значений



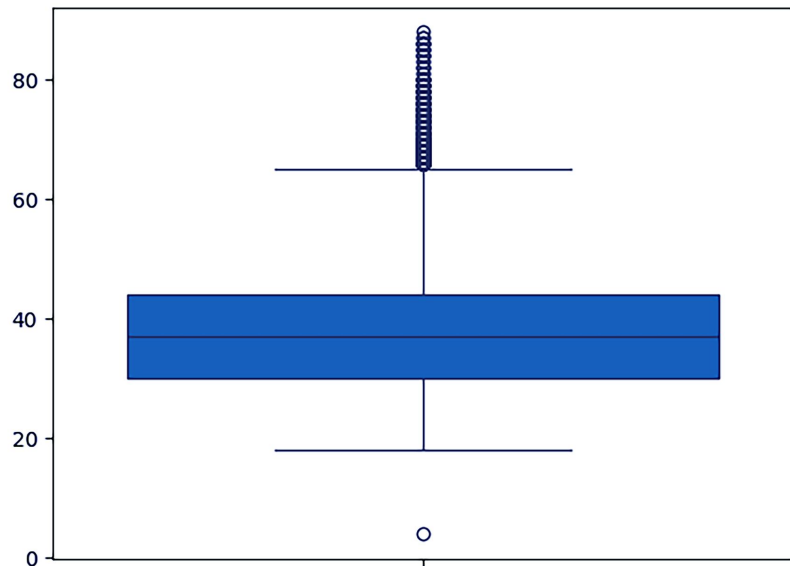
Очистили от выбросов и ошибок



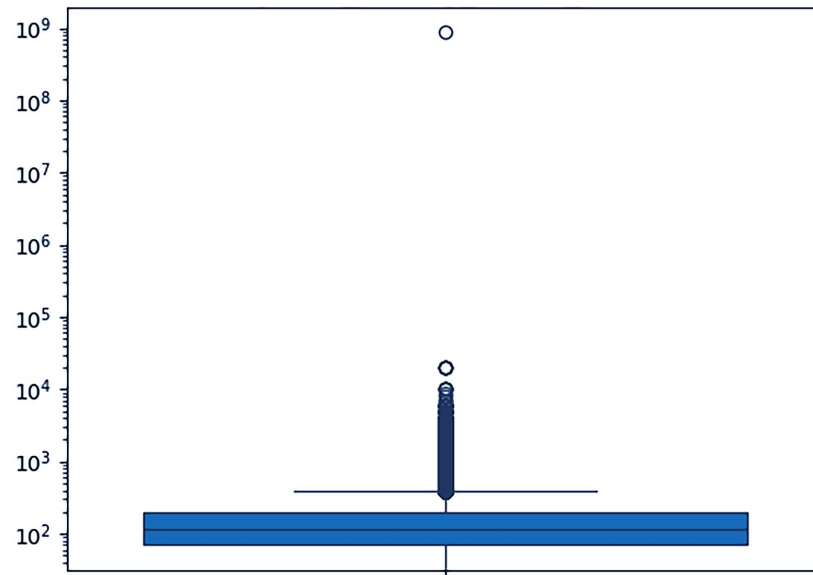
95820

Характеристики клиентов

Возраст

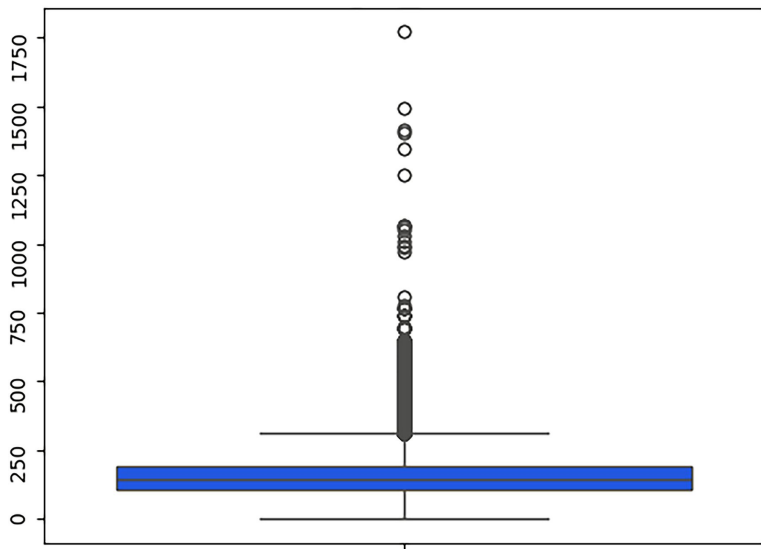


Зарплата

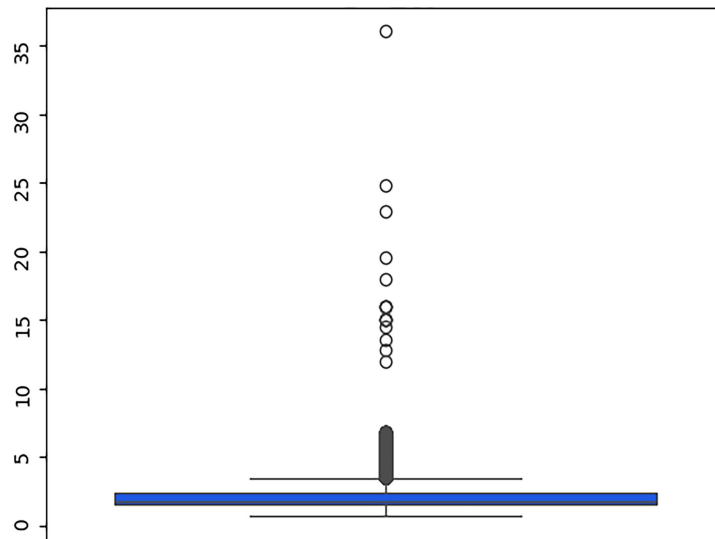


Характеристики авто

Мощность двигателя (л.с)

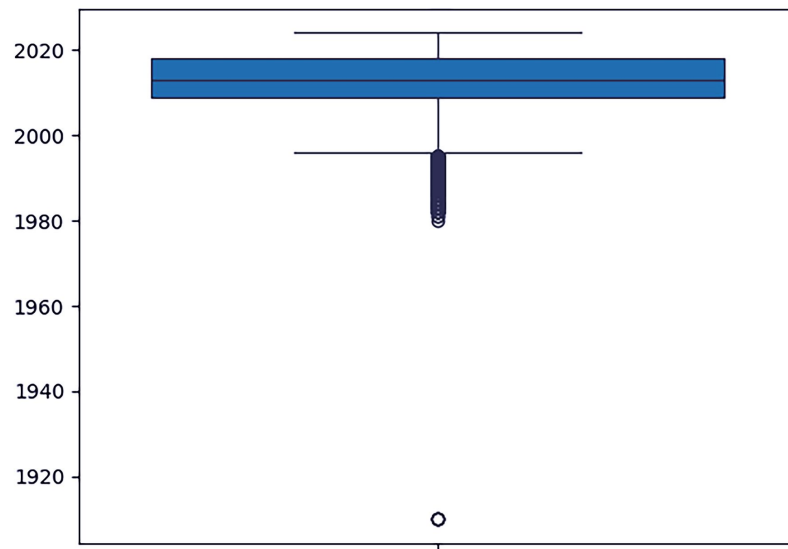


Объем двигателя (литры)



Характеристики авто

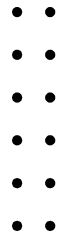
Год выпуска авто





02

ИССЛЕДОВАНИЕ

- Исследовательский вопрос
 - Исследование взаимосвязей
 - Гипотеза
 - Механизм
- 

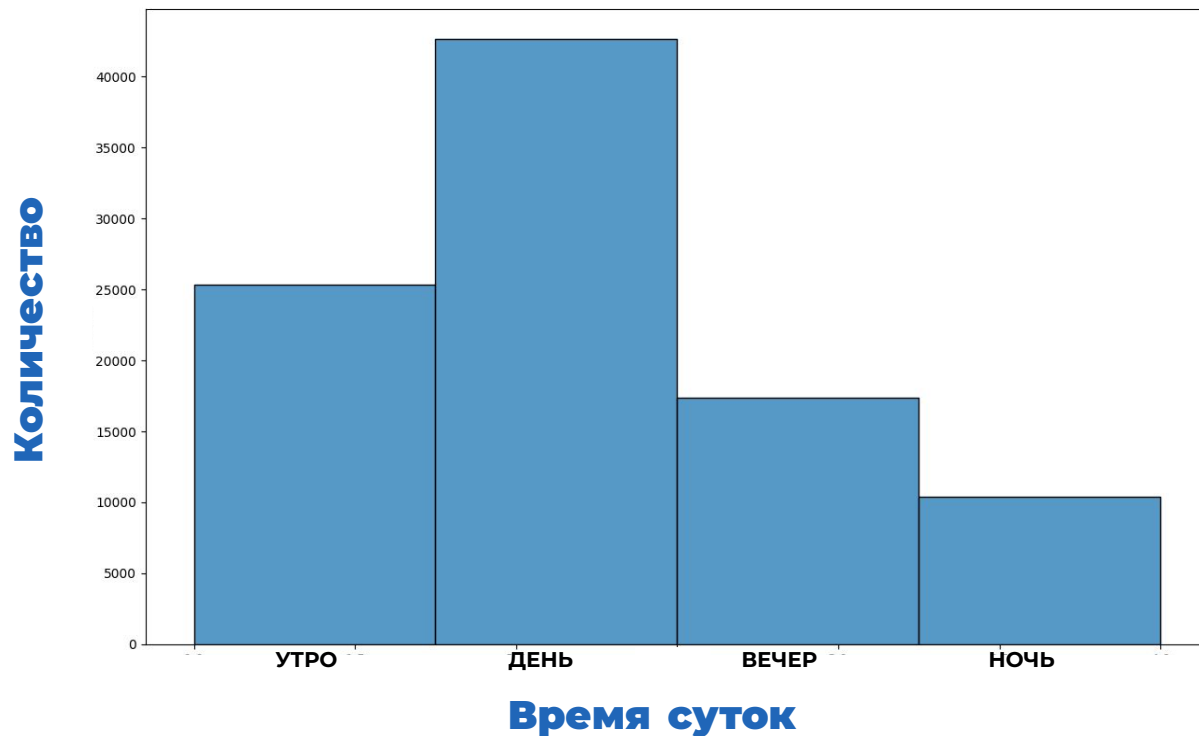




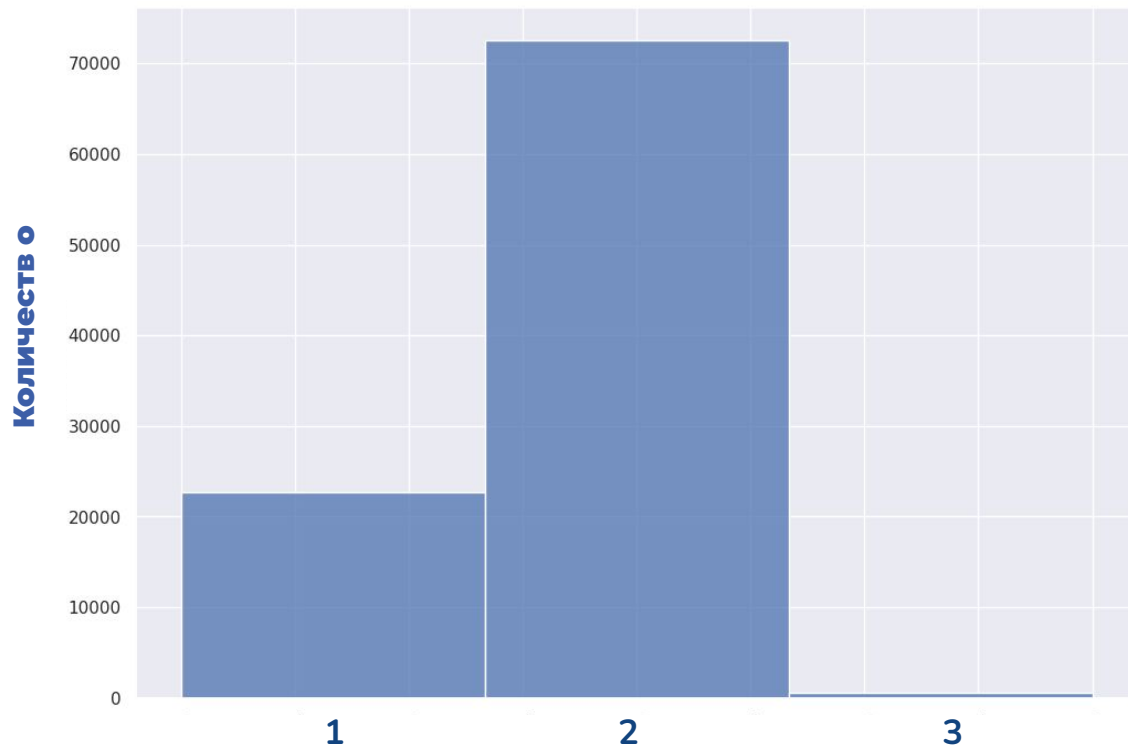
Исследовательский вопрос

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

Распределение по времени суток



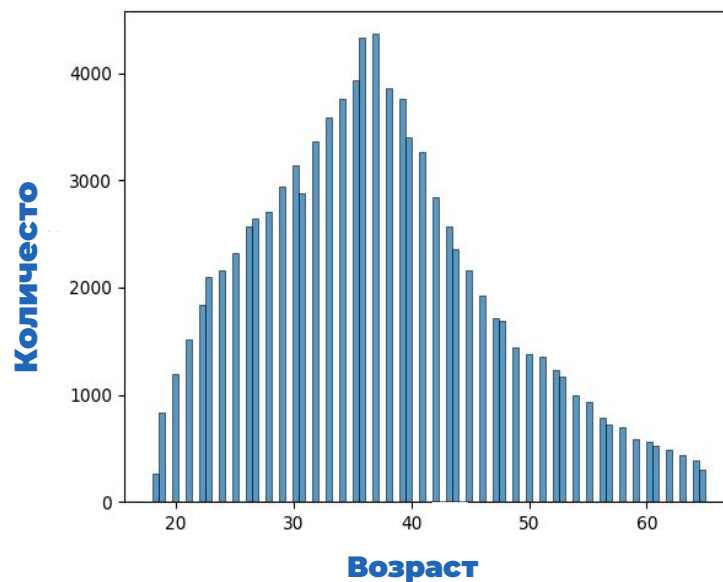
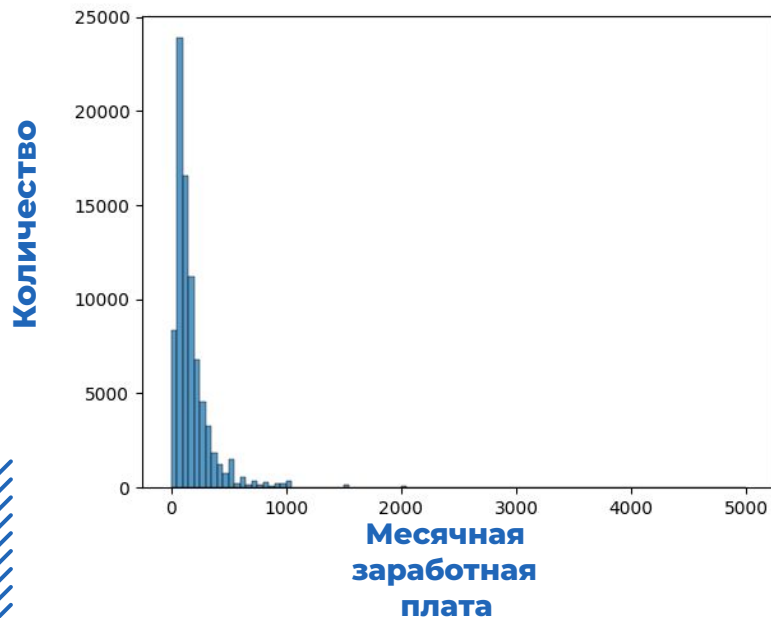
Распределение по типам нарушений



Тип правонарушения:

1. Управление автомобилем
2. Скоростной режим
3. Остановка, стоянка

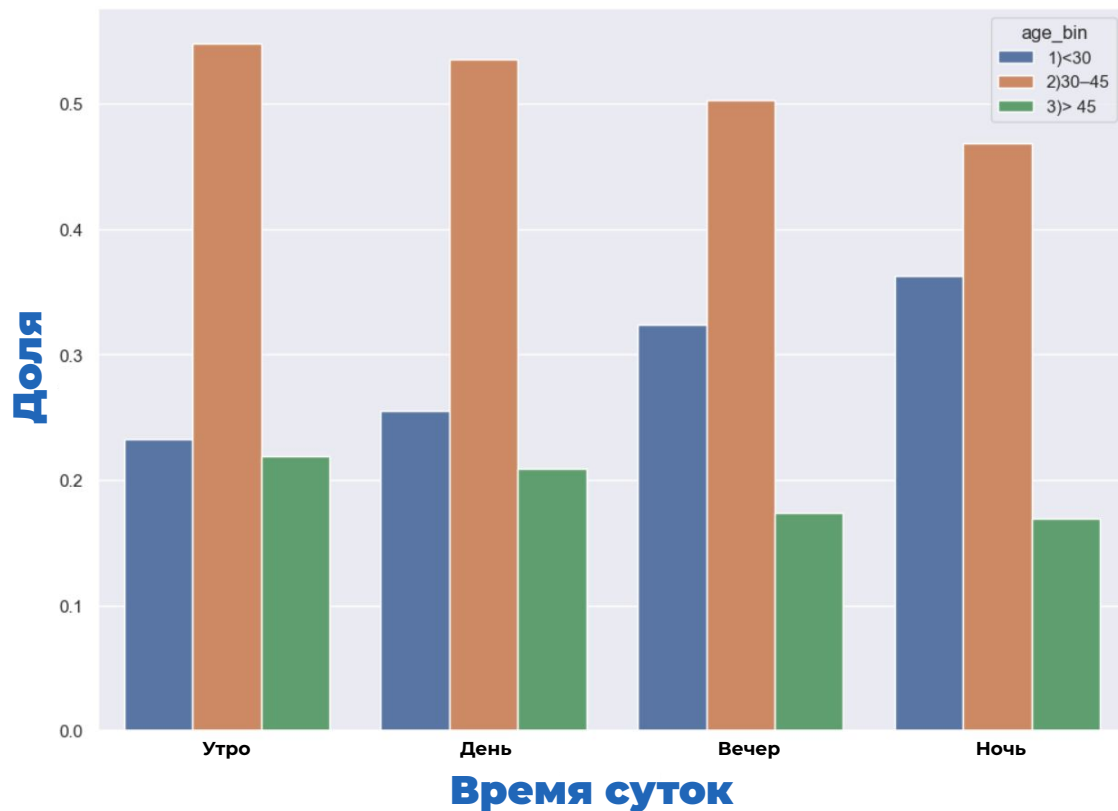
Распределения по зарплате и возрасту

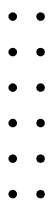


Корреляции



Связь возраста и времени суток





Гипотеза

Люди, совершающие правонарушения ночью, в среднем моложе, чем люди, совершающие правонарушения в остальные времена суток



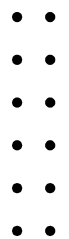
Механизм

1. Чувство свободы и безнаказанности молодых людей при низком трафике в ночное время суток
2. Нежелание более взрослого население сталкиваться с риском, спокойный образ жизни



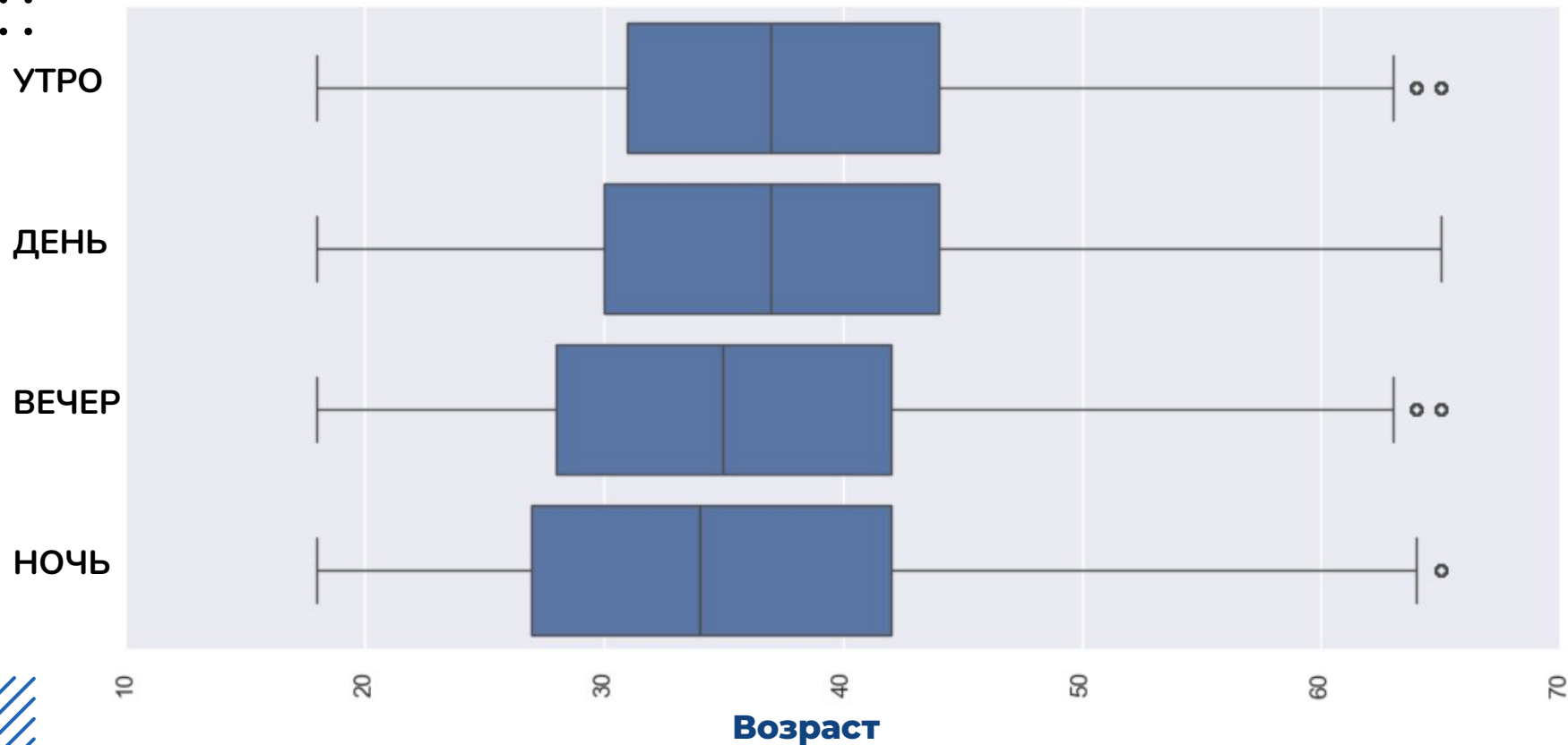
03

ПРОВЕРКА ГИПОТЕЗЫ

- Математическая модель
 - Проверка устойчивости
 - Интерпретация
- 



Возраст и время суток



Проверка распределения

- Тест Шапиро-Уилка для распределения выборочных средних возраста

P-value = 0.24 (Для всей базы)

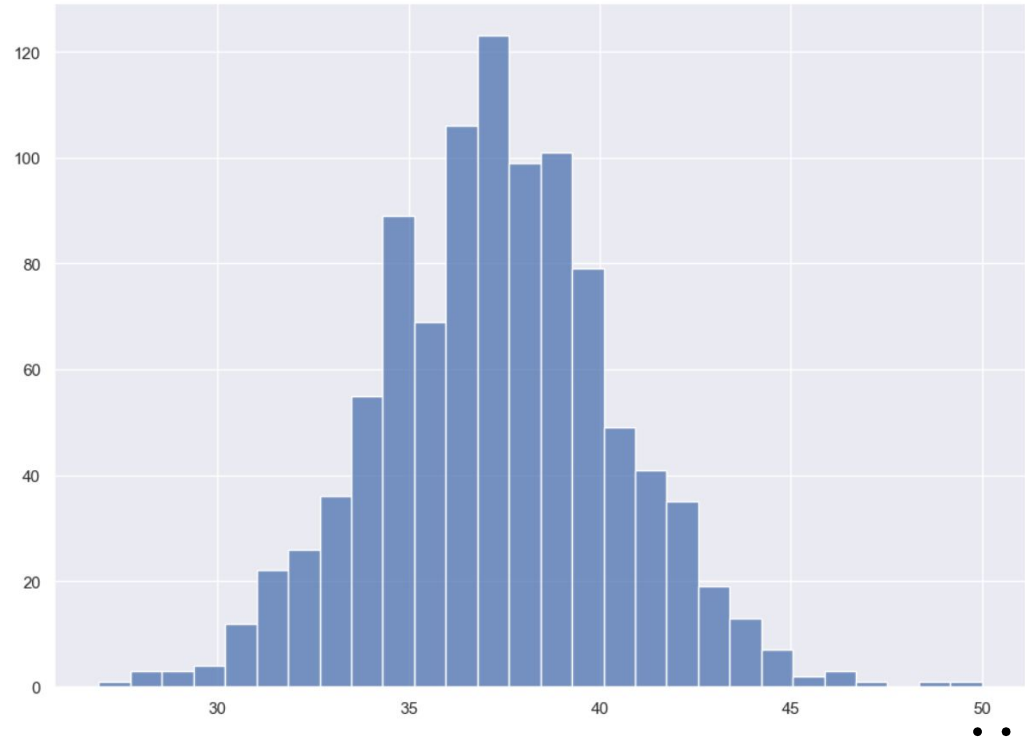
P-value = 0.33 (Ночь)

P-value = 0.11 (Вечер)

P-value = 0.09 (День)

P-value = 0.12 (Утро)

Тест пройден



Математическая модель

Используем Т-критерий Стьюдента

H_0 - Среднее значение возраста нарушителей в ночное время не отличается от их среднего возраста в другие времена суток

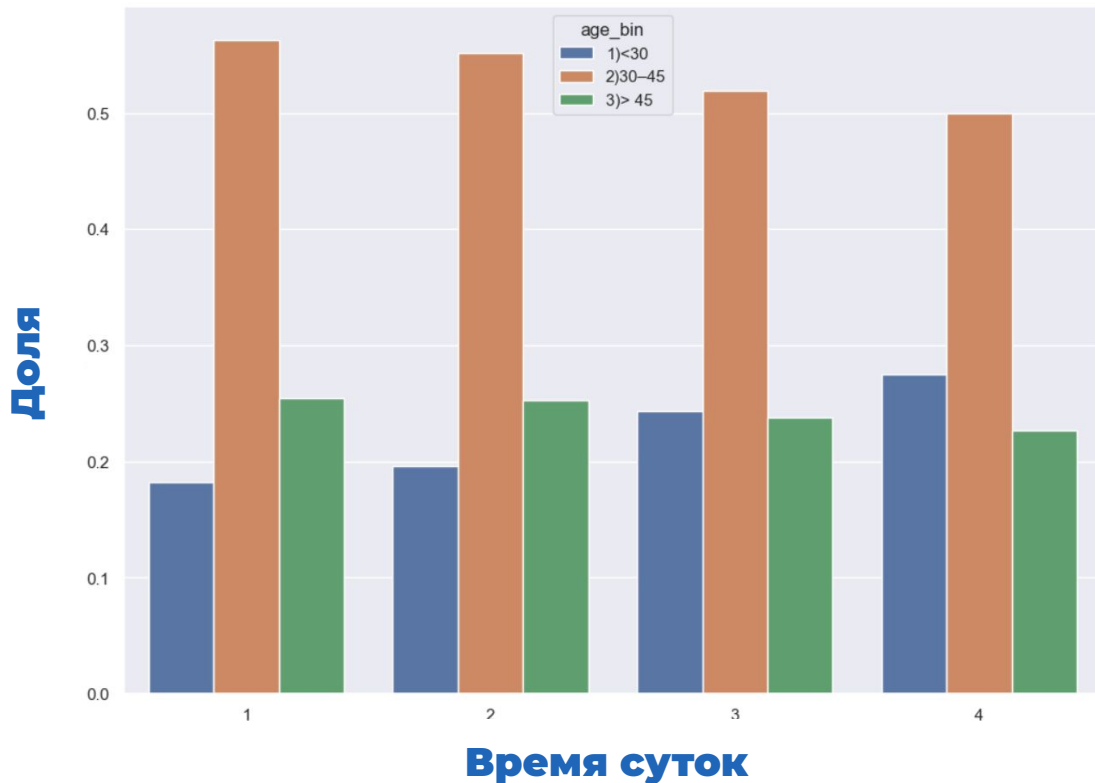
P-value < 0.001 (Ночь и утро)

P-value < 0.001 (Ночь и день)

P-value = 0.002 (Ночь и вечер)

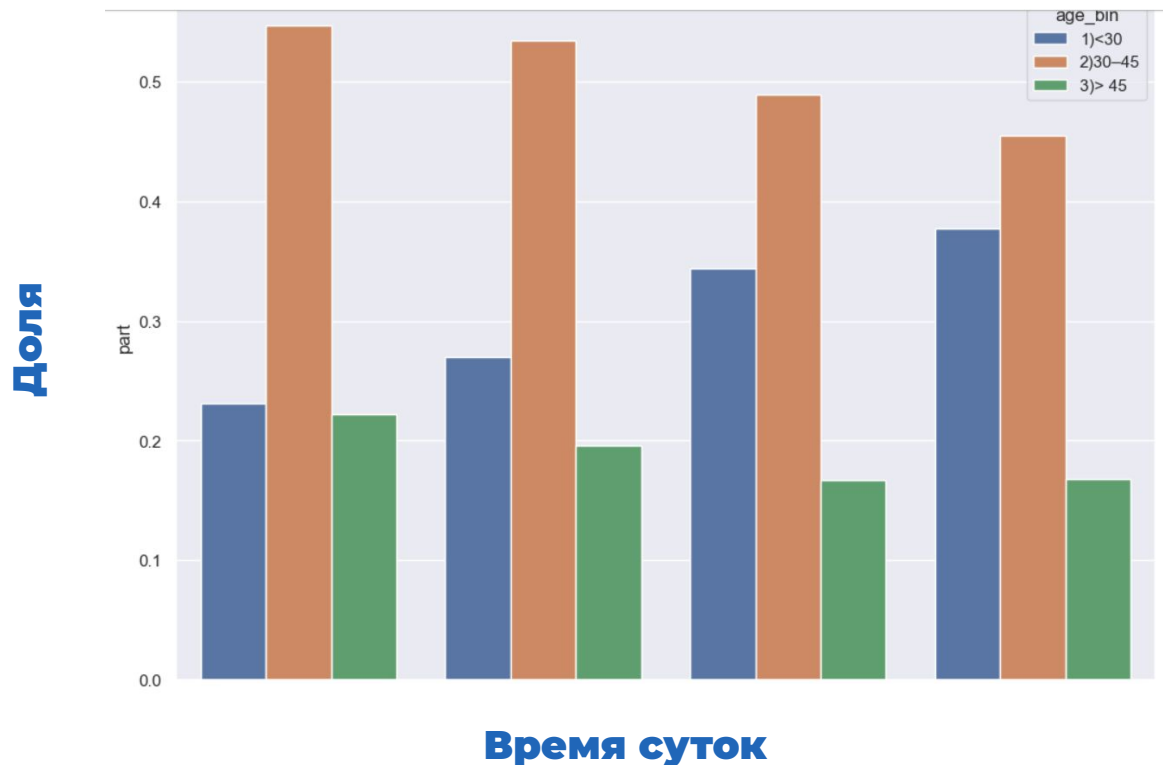
H_0 отвергается

Устойчивость: женщины



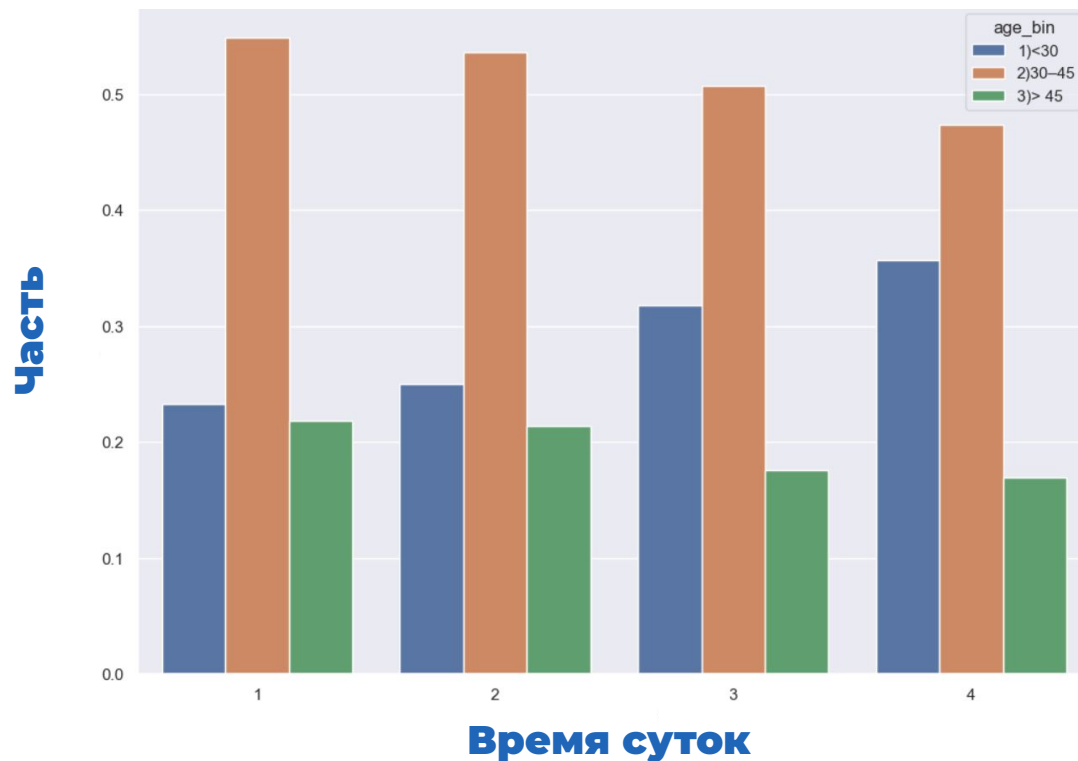
1 - Утро
2 - День
3 - Вечер
4 - Ночь

Устойчивость: выходные и праздники



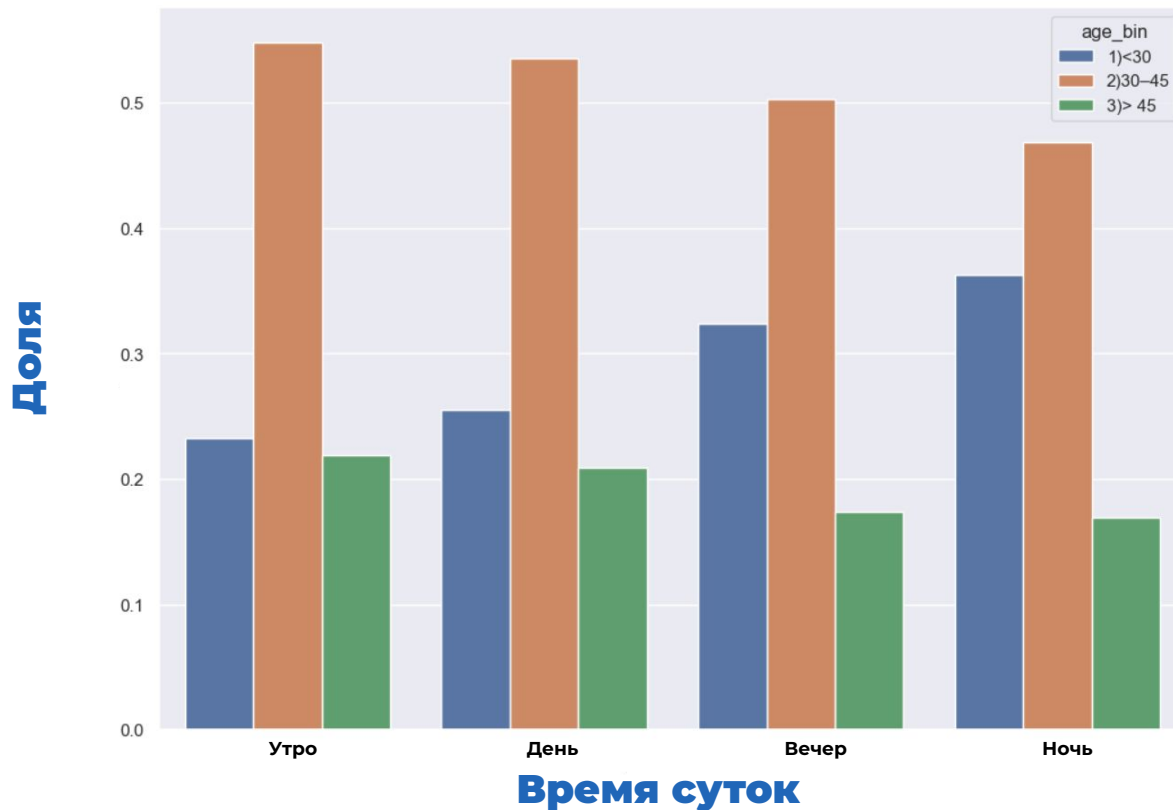
1 - Утро
2 - День
3 - Вечер
4 - Ночь

Устойчивость: будни



- 1 - Утро**
- 2 - День**
- 3 - Вечер**
- 4 - Ночь**

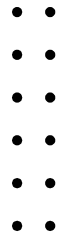
Устойчивость: мужчины





04

ВЫВОДЫ

- Перспективы
 - Ограничения
 - Применение
- 



ПЕРСПЕКТИВЫ



Сбор данных в другие времена года



Водительский стаж



Форма административного наказания



Место совершения правонарушения

ОГРАНИЧЕНИЯ



Короткий период



**Физическое и эмоциональное
состояние водителя**



**Невозможность проследить
тенденцию**

ПРИМЕНЕНИЕ

Анализ активности молодых водителей в
позднее время, транслирование
таргетированной рекламы и кроссел



ВШЭстером

Николай Матвеев

Тимлид

Полина Жуйкова

Аналитик

Захарий Димидов

Программист

Денис Игнатов

Программист

Полина

Мирошниченко

Аналитик



Андрей Чагин

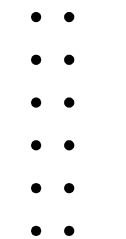
Аналитик





-1

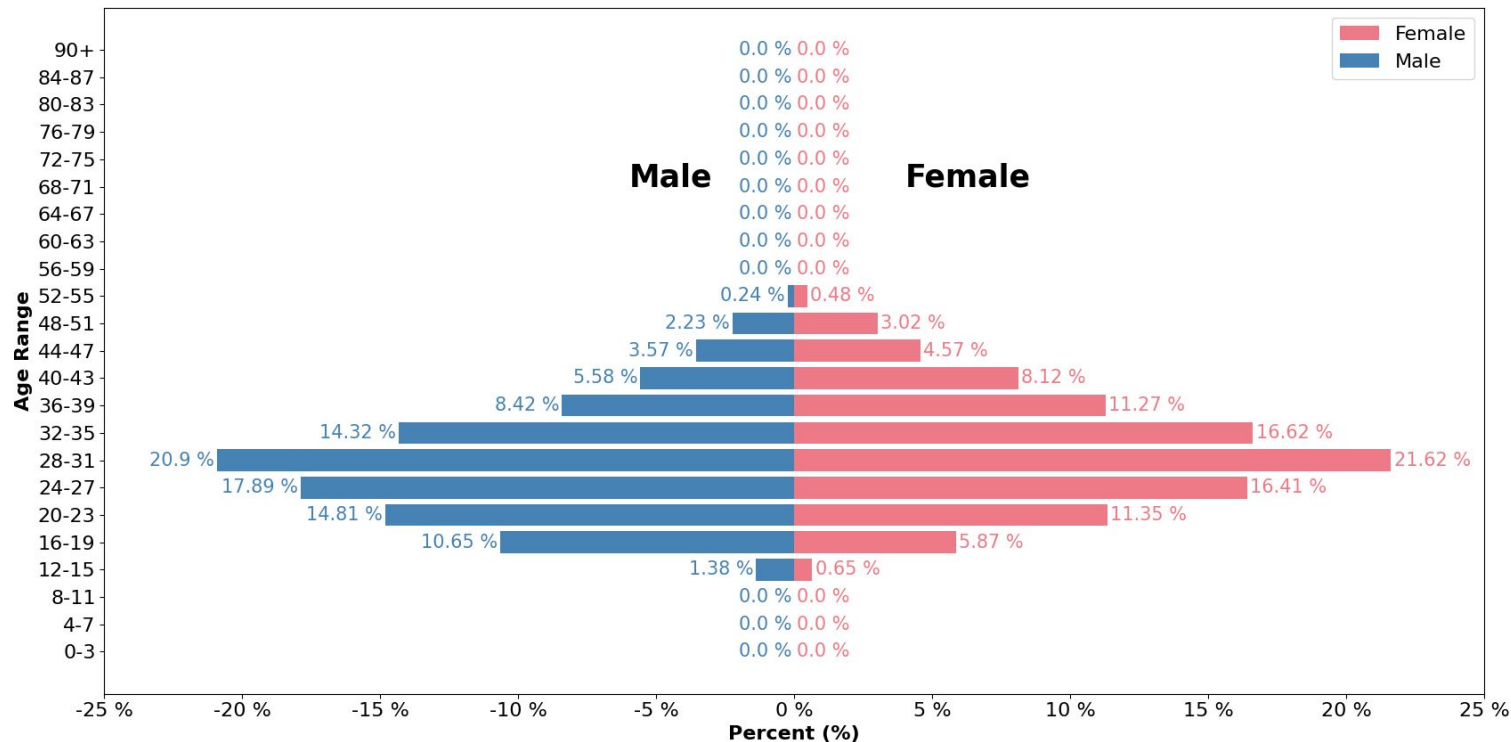
ПОДВАЛ

- Возрастно-половая пирамида
 - Матрица корреляций
 - Дополнительные вычисления
- 



Возрастно-половая пирамида

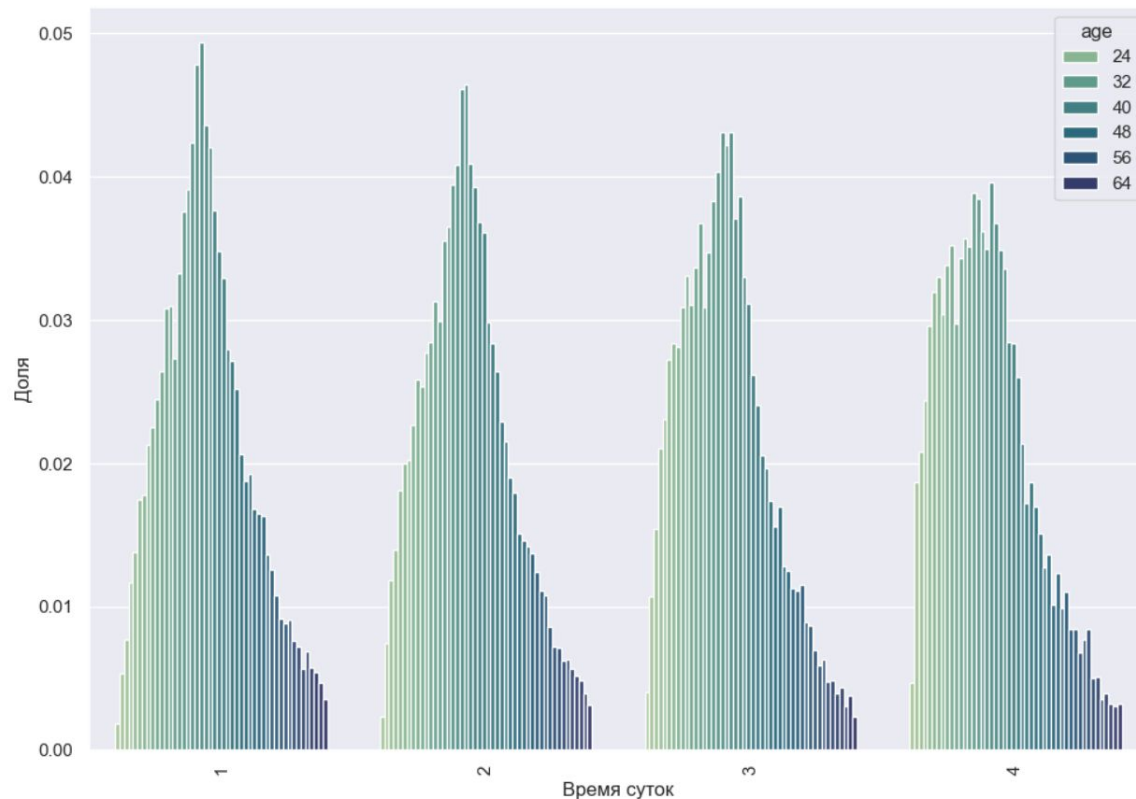
Sample Population Pyramid Chart



Распределение штрафов по видам правонарушения



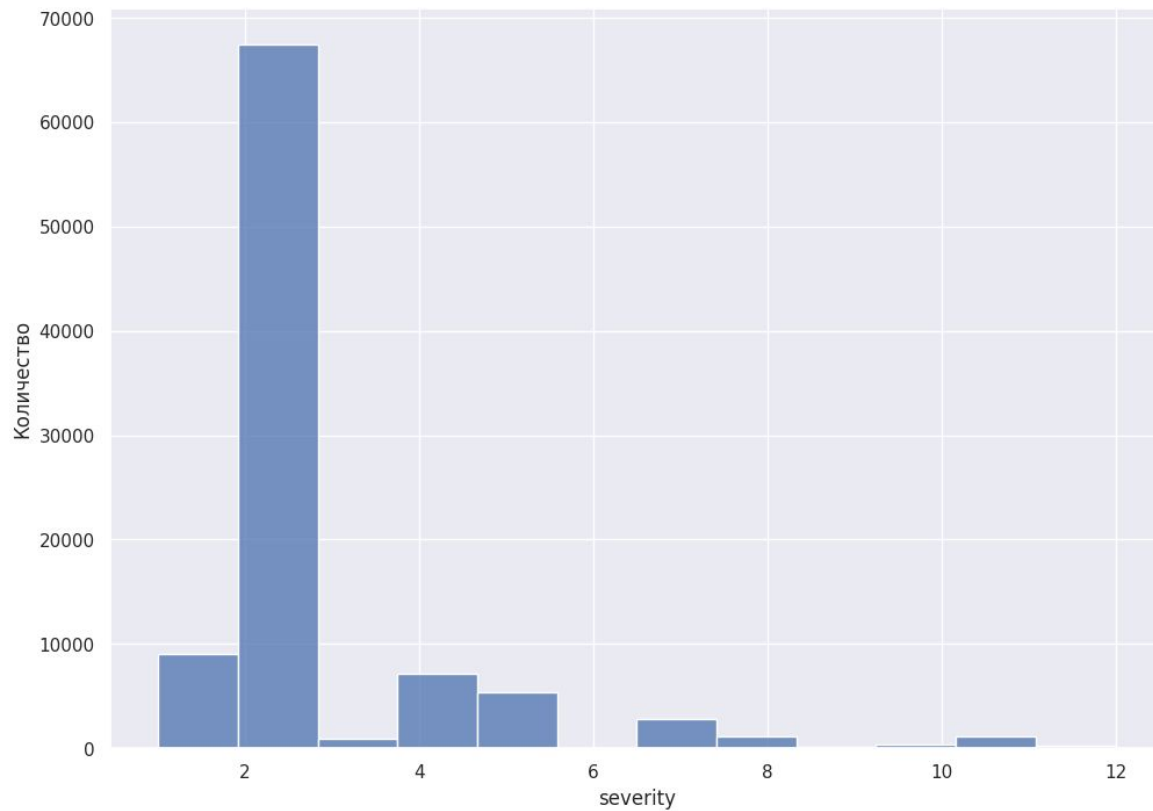
Распределение возраста в группах по времени суток



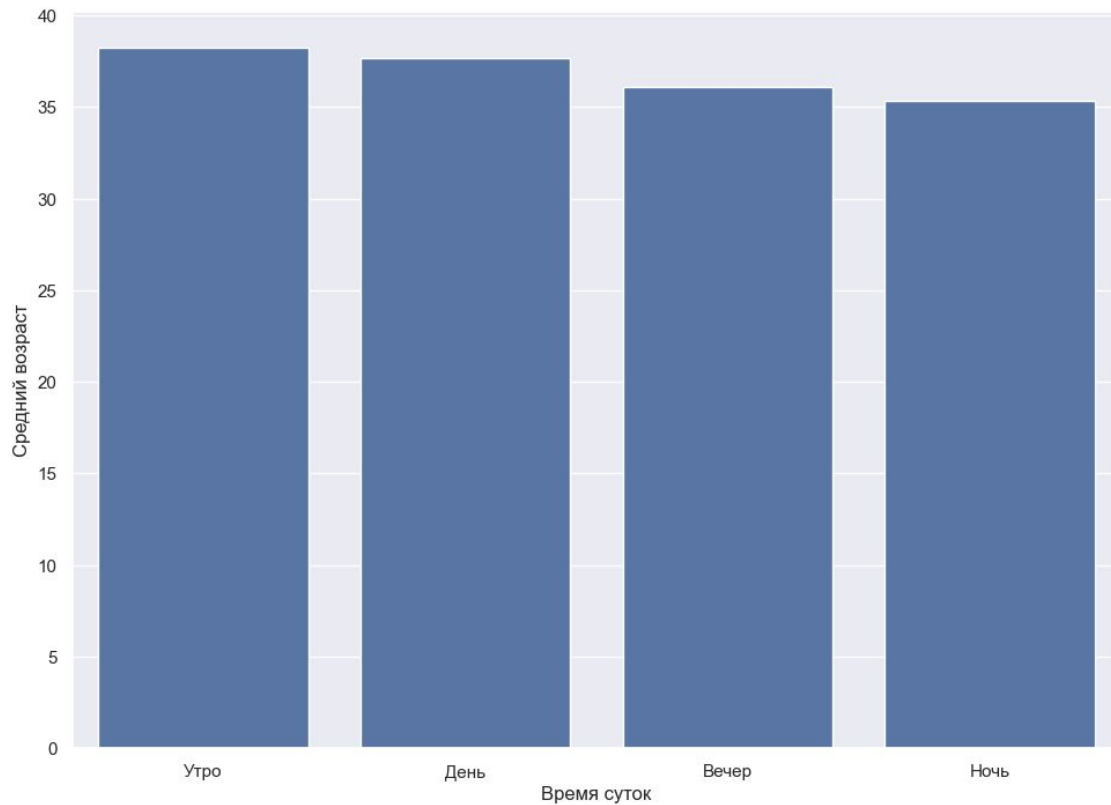
Количество нарушений в подгруппах

typesofday	1	2	3	4
age				
1)<30	5890	10888	5627	3778
2)30–45	13914	22834	8749	4886
3)> 45	5564	8912	3017	1761

Кол-во нарушений разной тяжести



Средний возраст и время суток



Распределение по возрасту

