



PIZZA HUNTERS

They see me rollin'

Команда

Владислав Денышев

Павел Савинов

Максим Садиков

Владимир Сиванев

О базе данных

Данные о полученных клиентами Т-банка штрафах ГИБДД с 28.04.24 по 28.05.24 с характеристиками водителя и автомобиля

97K+

Наблюдений

22

Столбца

10543

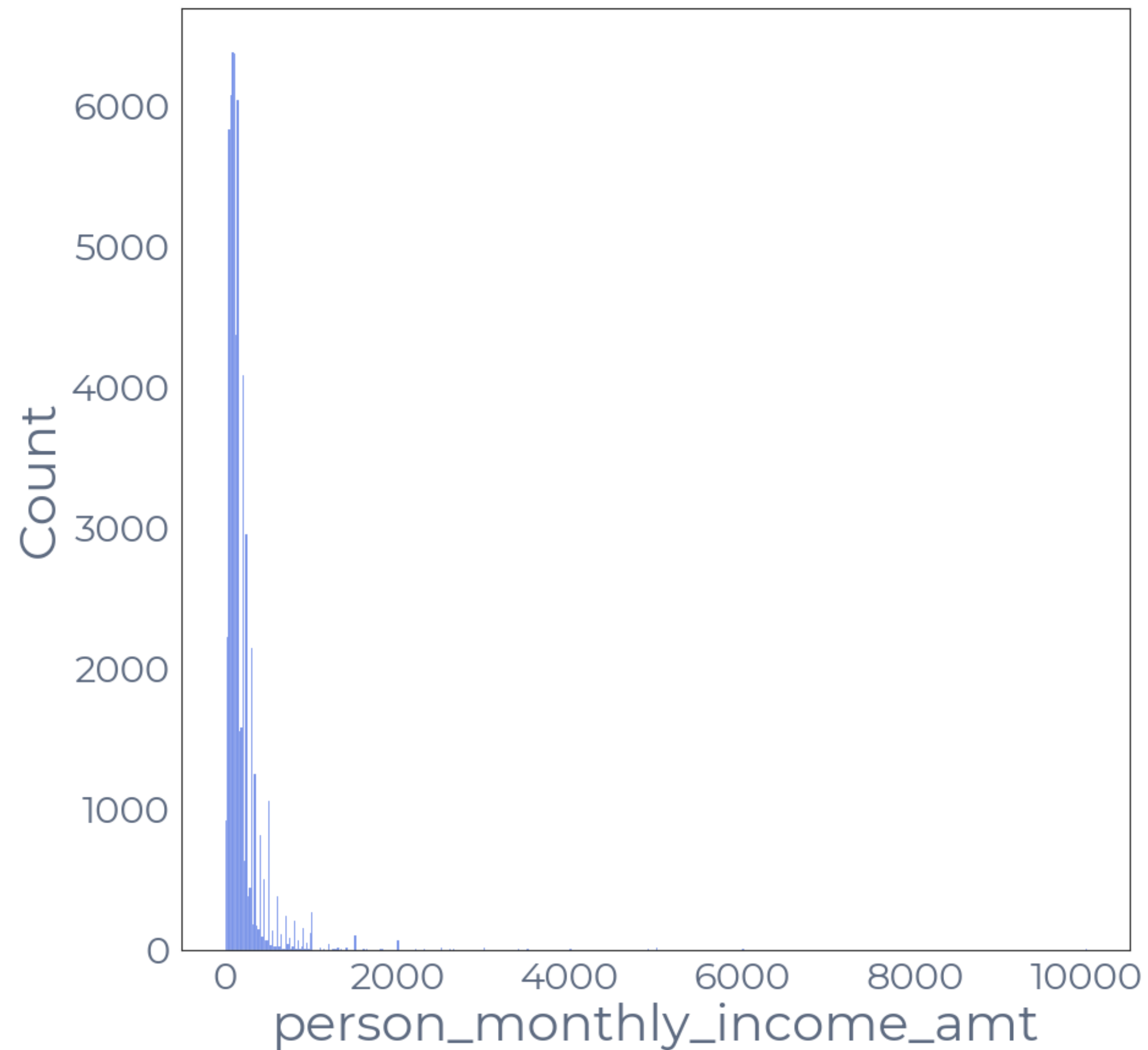
Средний
доход клиента
в УЕ

37

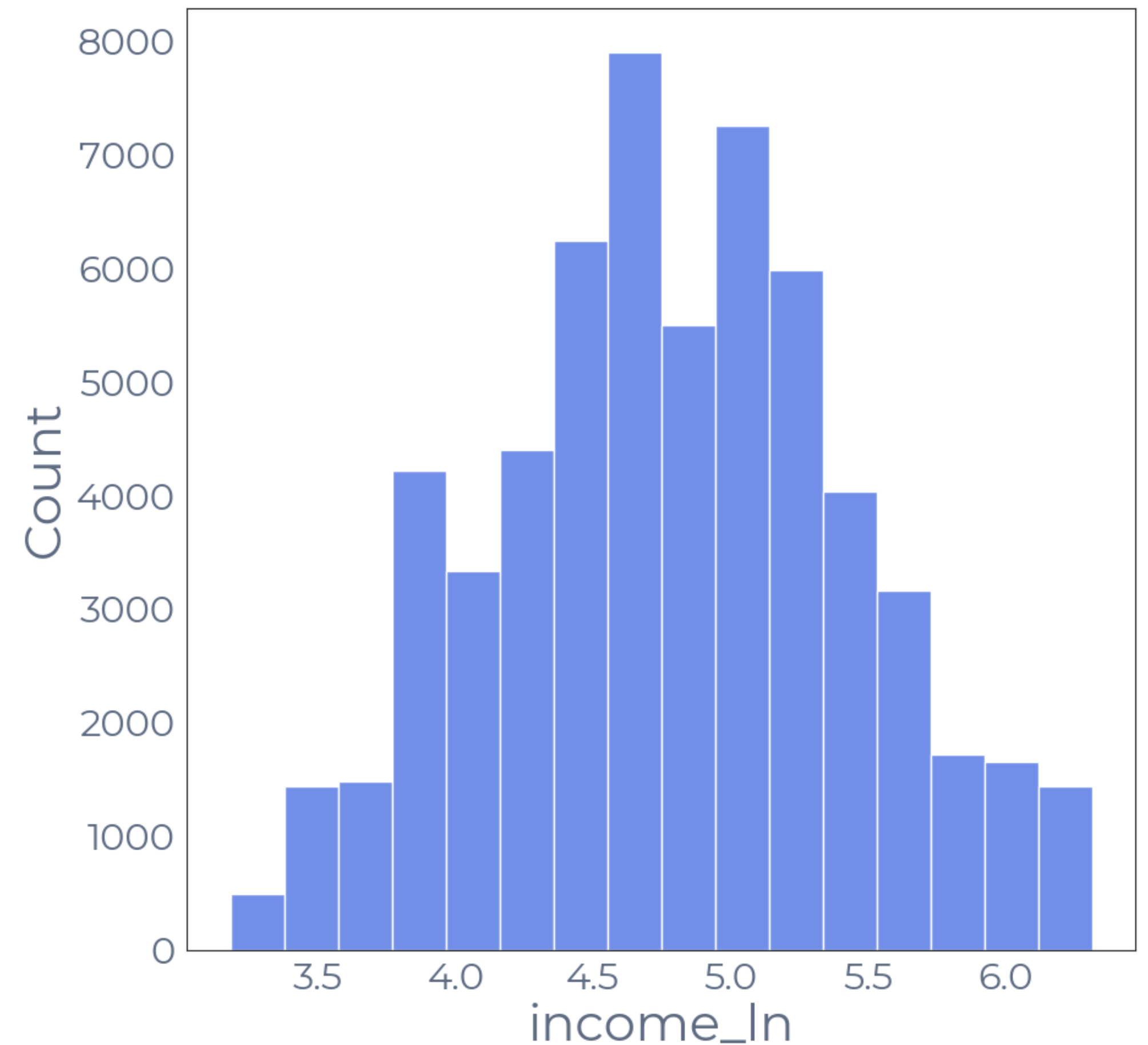
Медианный
возраст
клиента



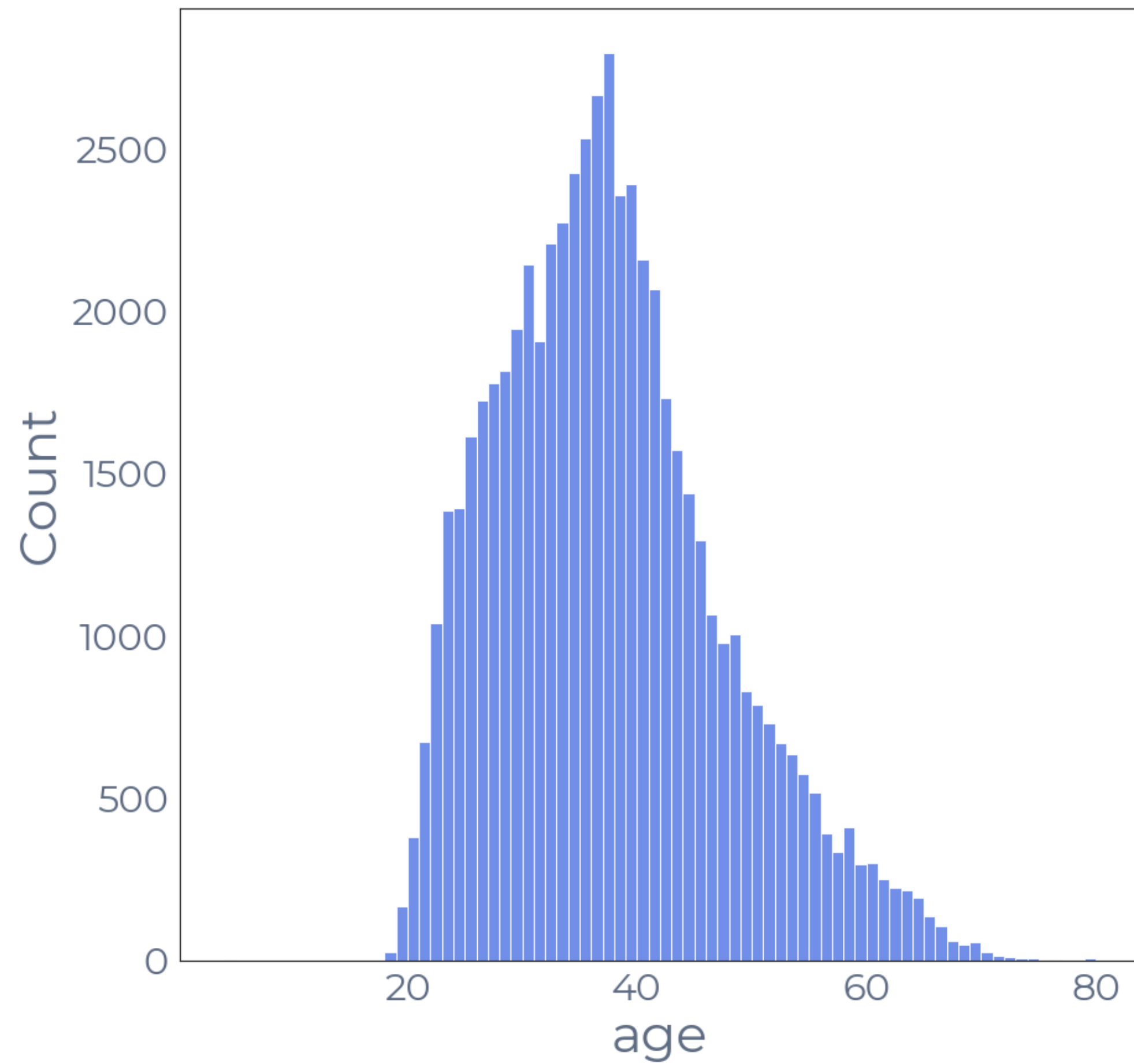
Распределение доходов



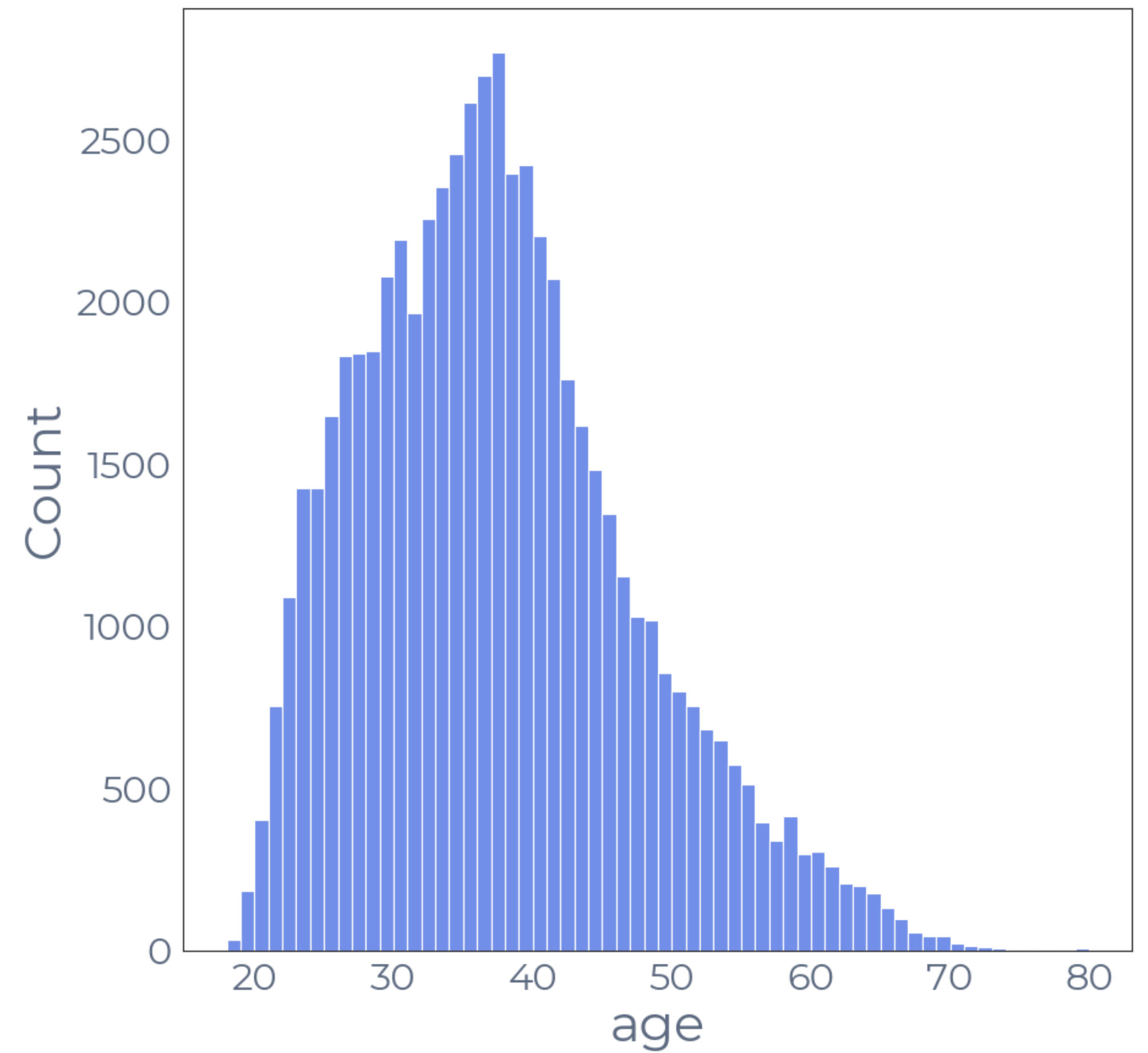
Тест Колмогорова-Смирнова
p-value < 0.05: ненормальное распределение



Распределение возраста



Тест Колмогорова-Смирнова
p-value < 0.05: ненормальное распределение



Матрица корреляций

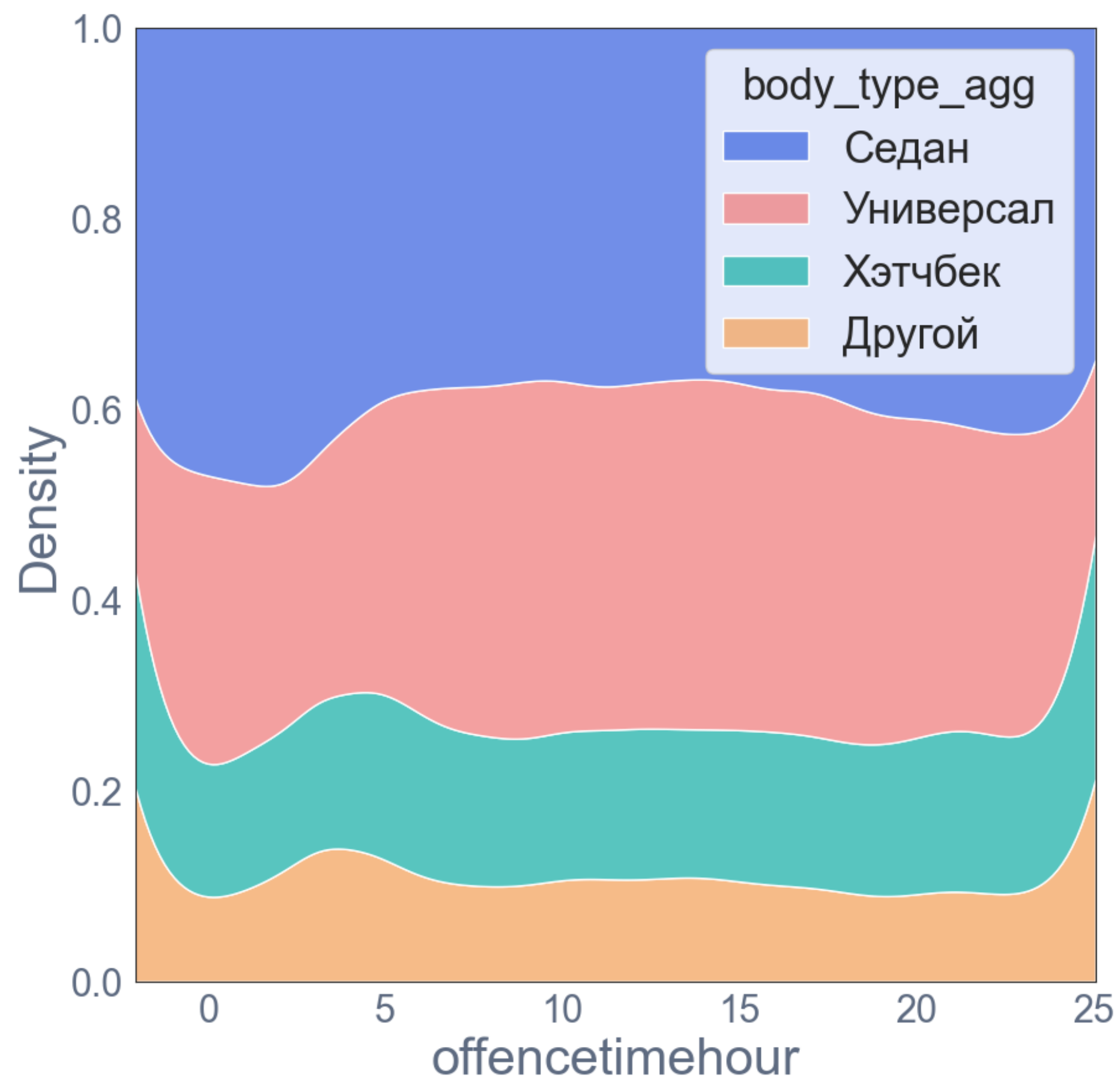
engine_type	1	0.17	-0.021	0.023	0.0053	-0.0015	0.012
engine_power	0.17	1	0.13	0.57	0.051	0.009	0.25
auto_year	-0.021	0.13	1	0.54	0.14	0.038	0.22
car_price	0.023	0.57	0.54	1	0.12	0.016	0.29
age	0.0053	0.051	0.14	0.12	1	0.15	0.019
children_cnt	-0.0015	0.009	0.038	0.016	0.15	1	0.014
income_In	0.012	0.25	0.22	0.29	0.019	0.014	1
	engine_type	engine_power	auto_year	car_price	age	children_cnt	income_In

Тест ANOVA на наличие взаимосвязи между доходами и характеристиками

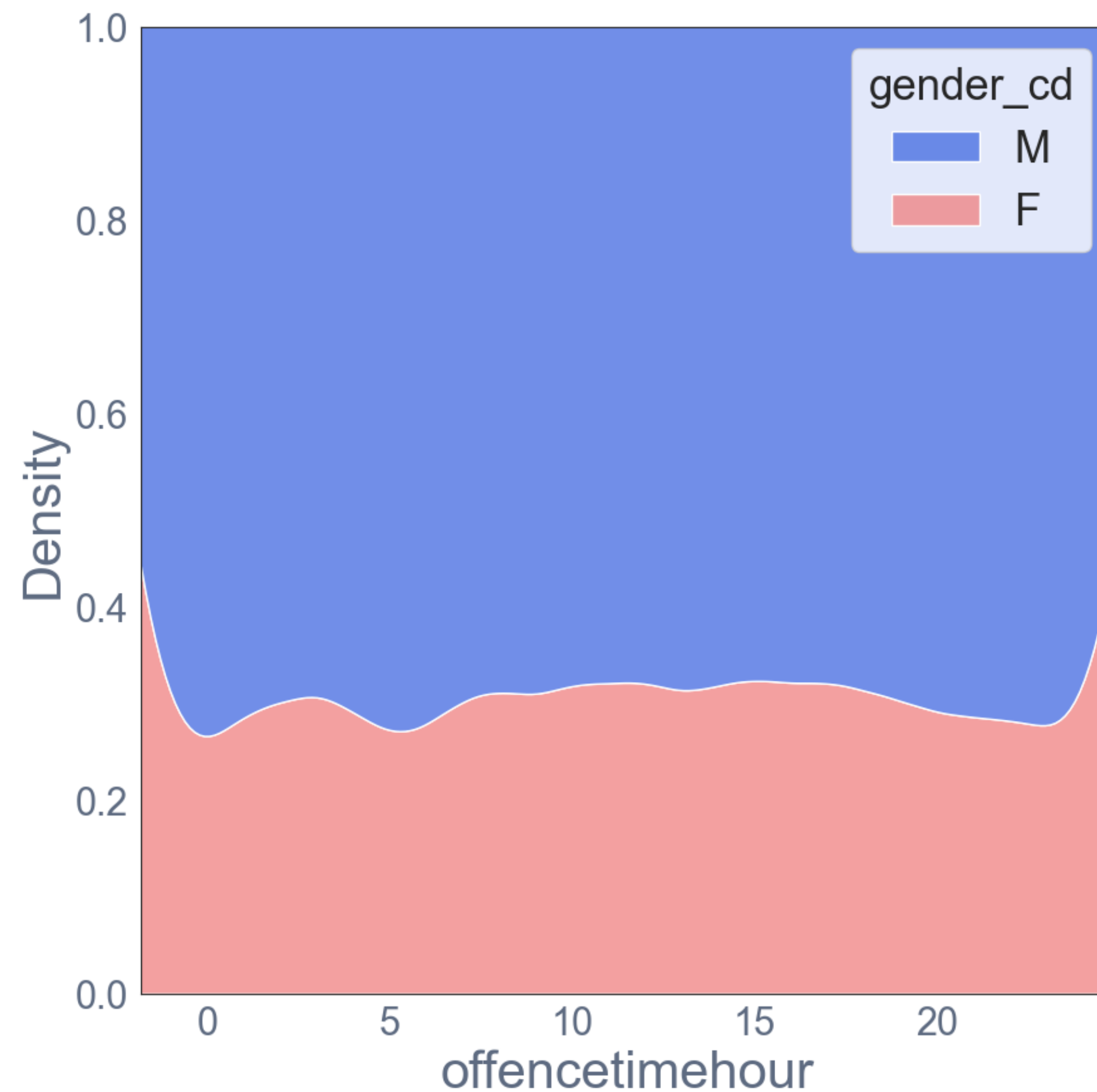
Характеристика	ANOVA p-value	
Марка автомобиля	<0.01	✓
Тип кузова	<0.01	✓
Пол	<0.01	✓
Семейный статус	<0.01	✓
Образование	<0.01	✓



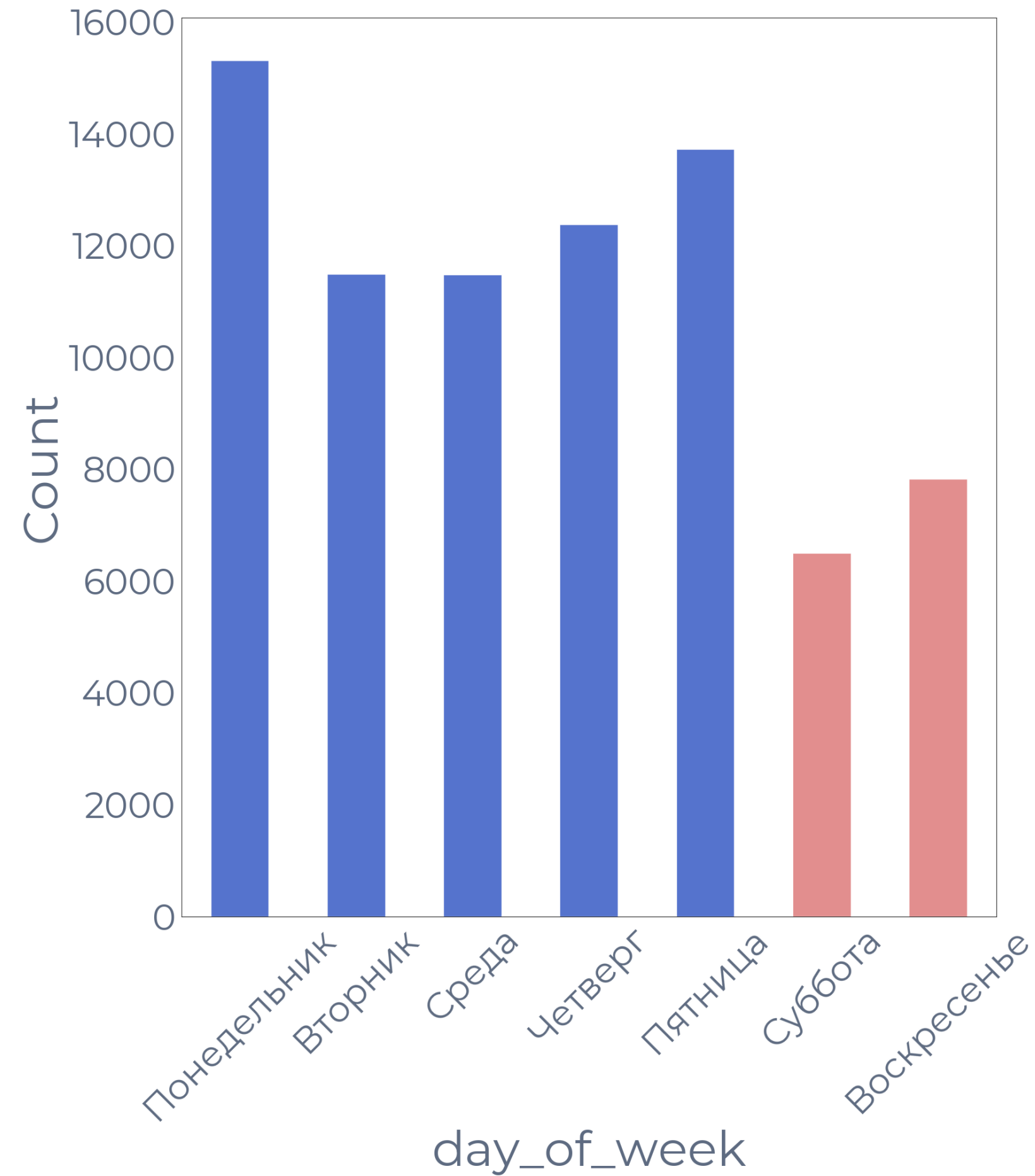
Распределение типов автомобиля по времени нарушения



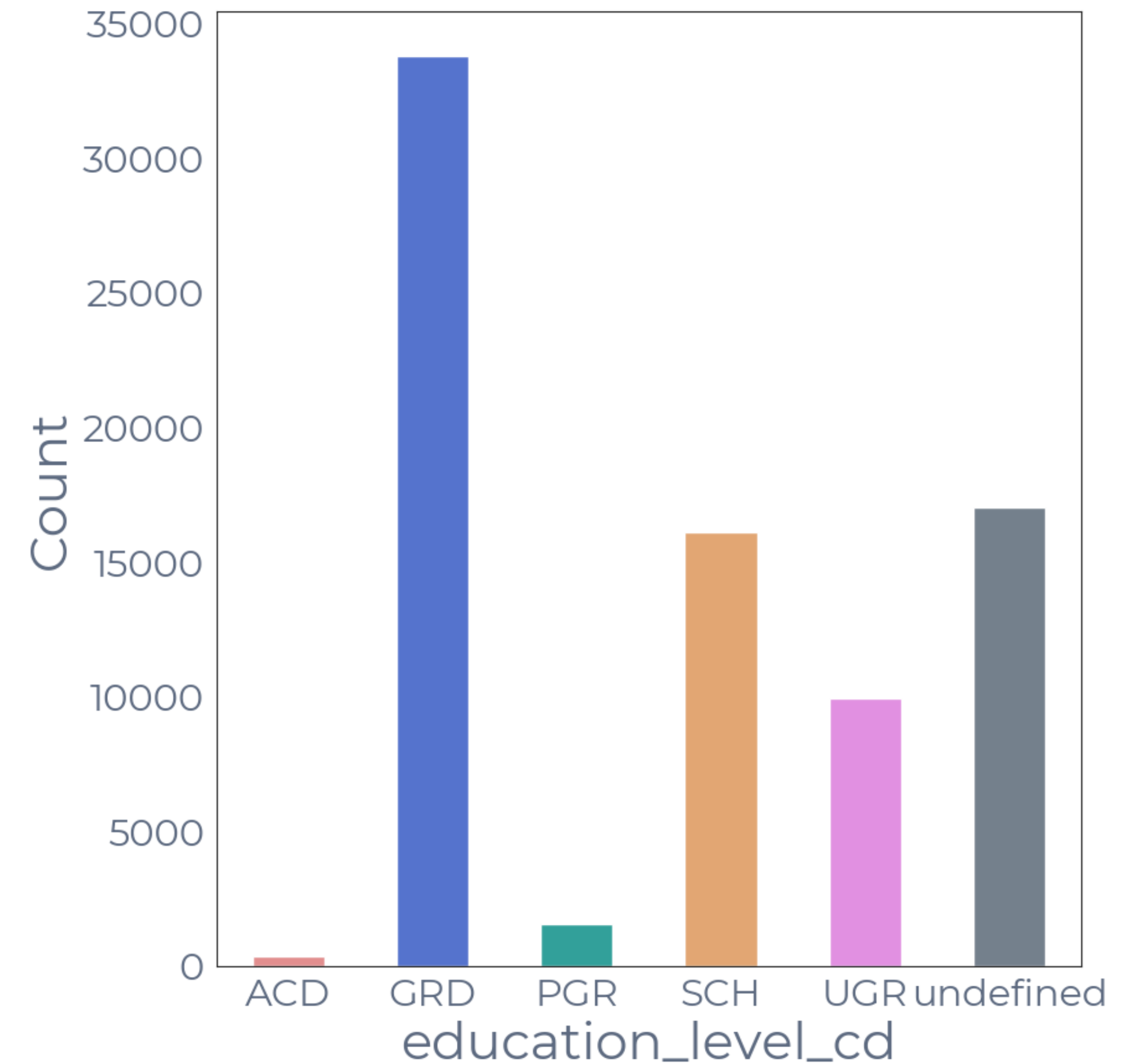
Распределение гендеров по времени нарушения



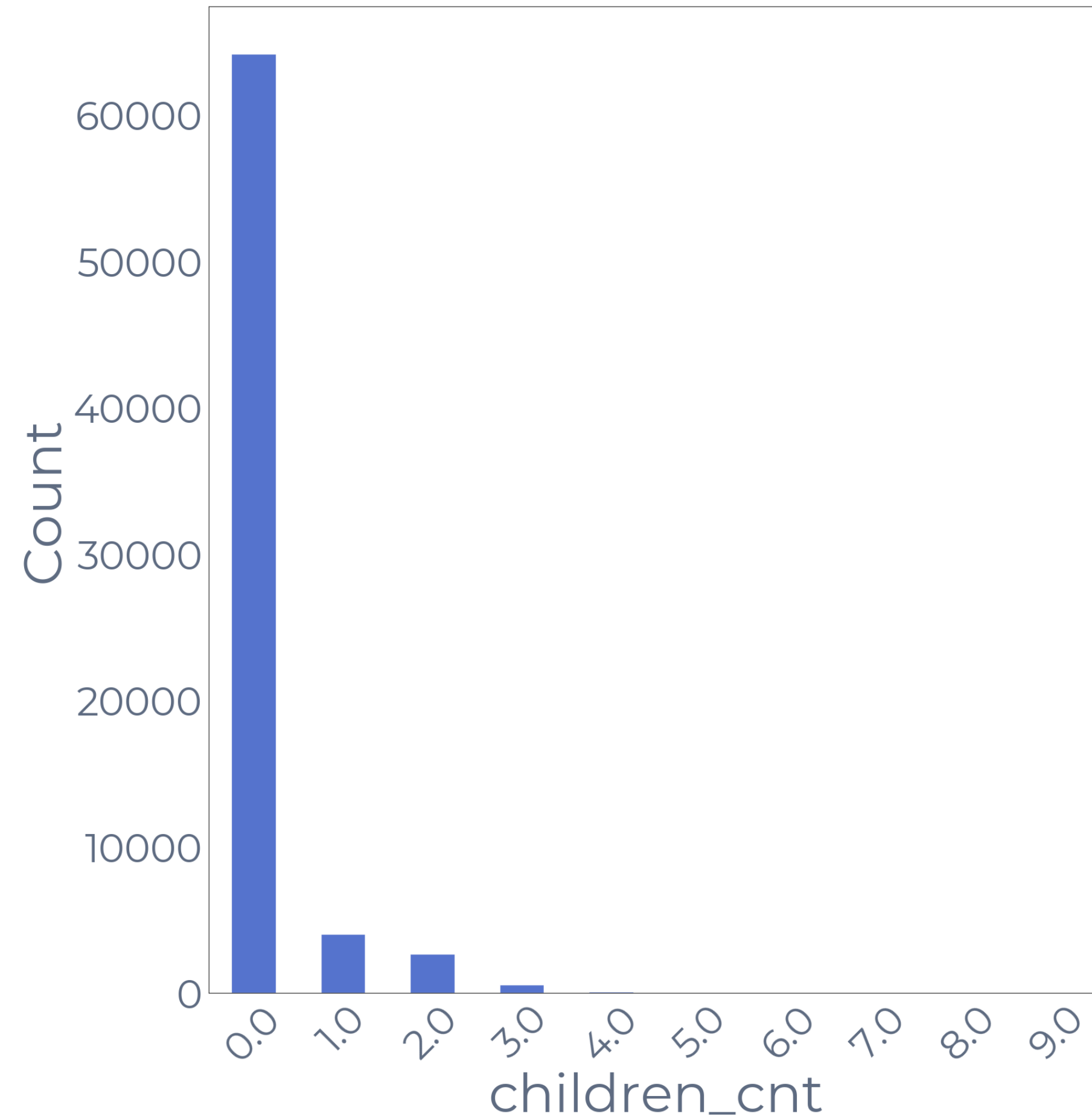
Распределение количества нарушений по дням недели



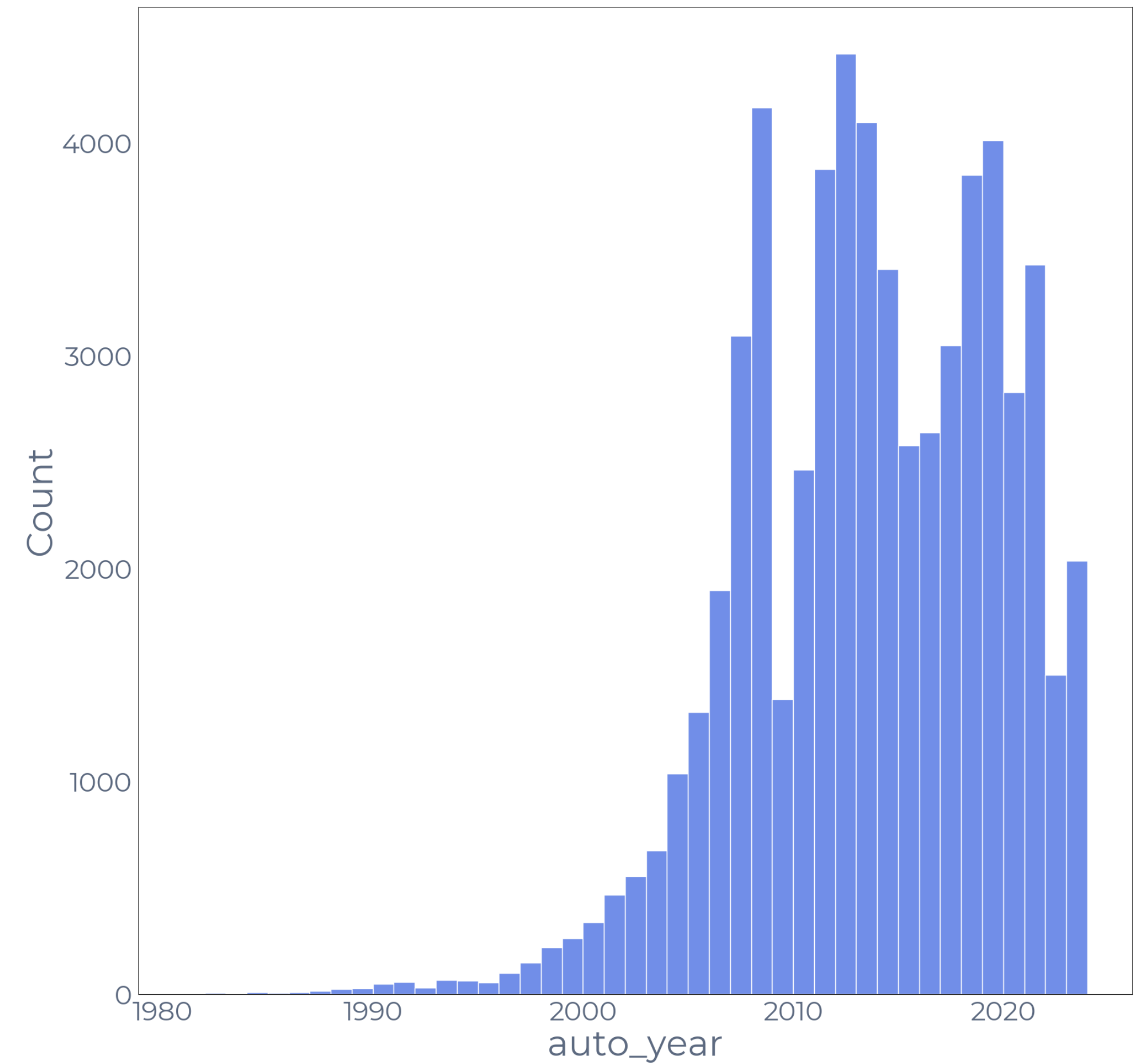
Распределение количества нарушений по типам образования



Распределение количества детей у нарушителей



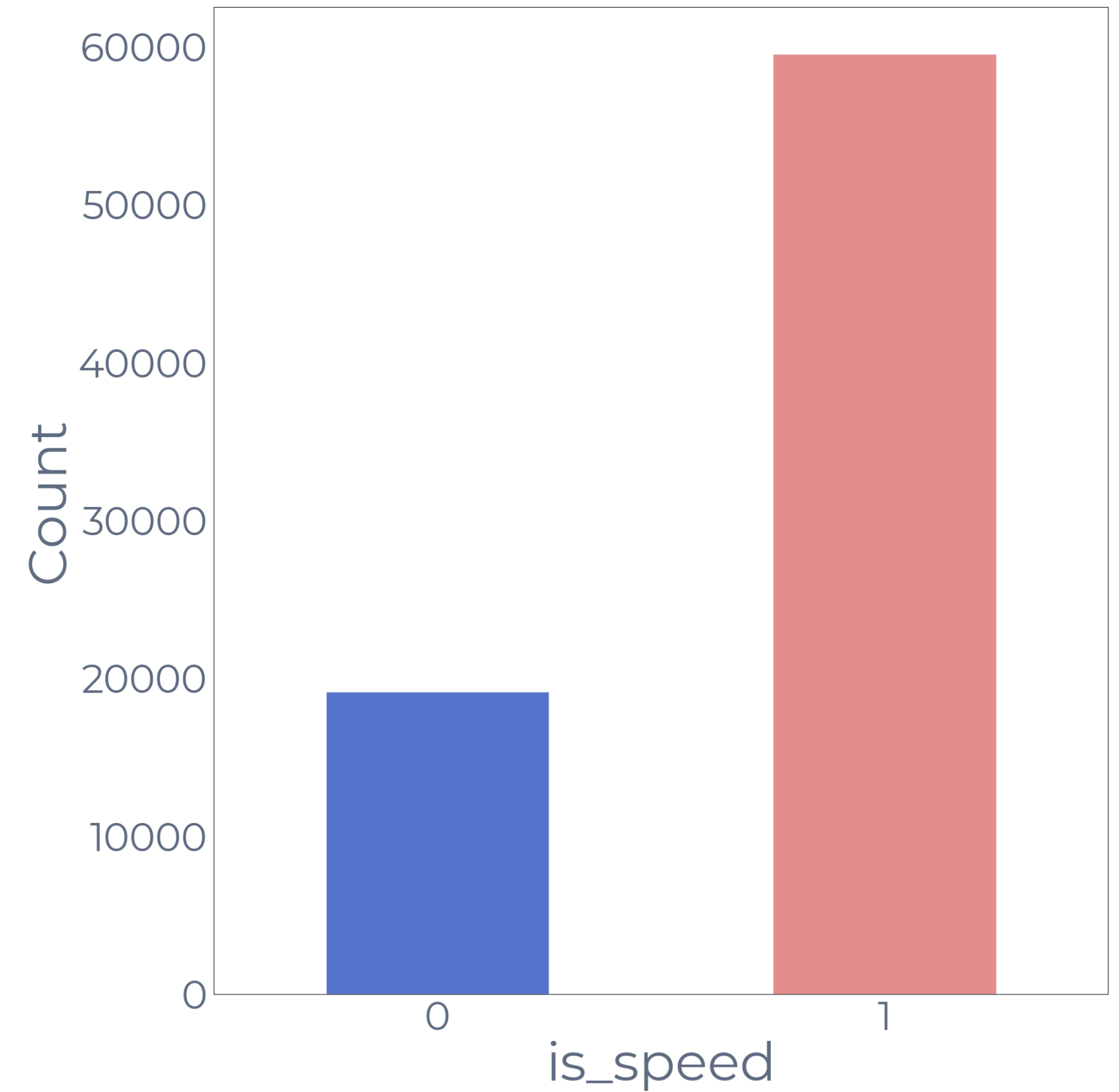
Распределение года производства авто у нарушителей



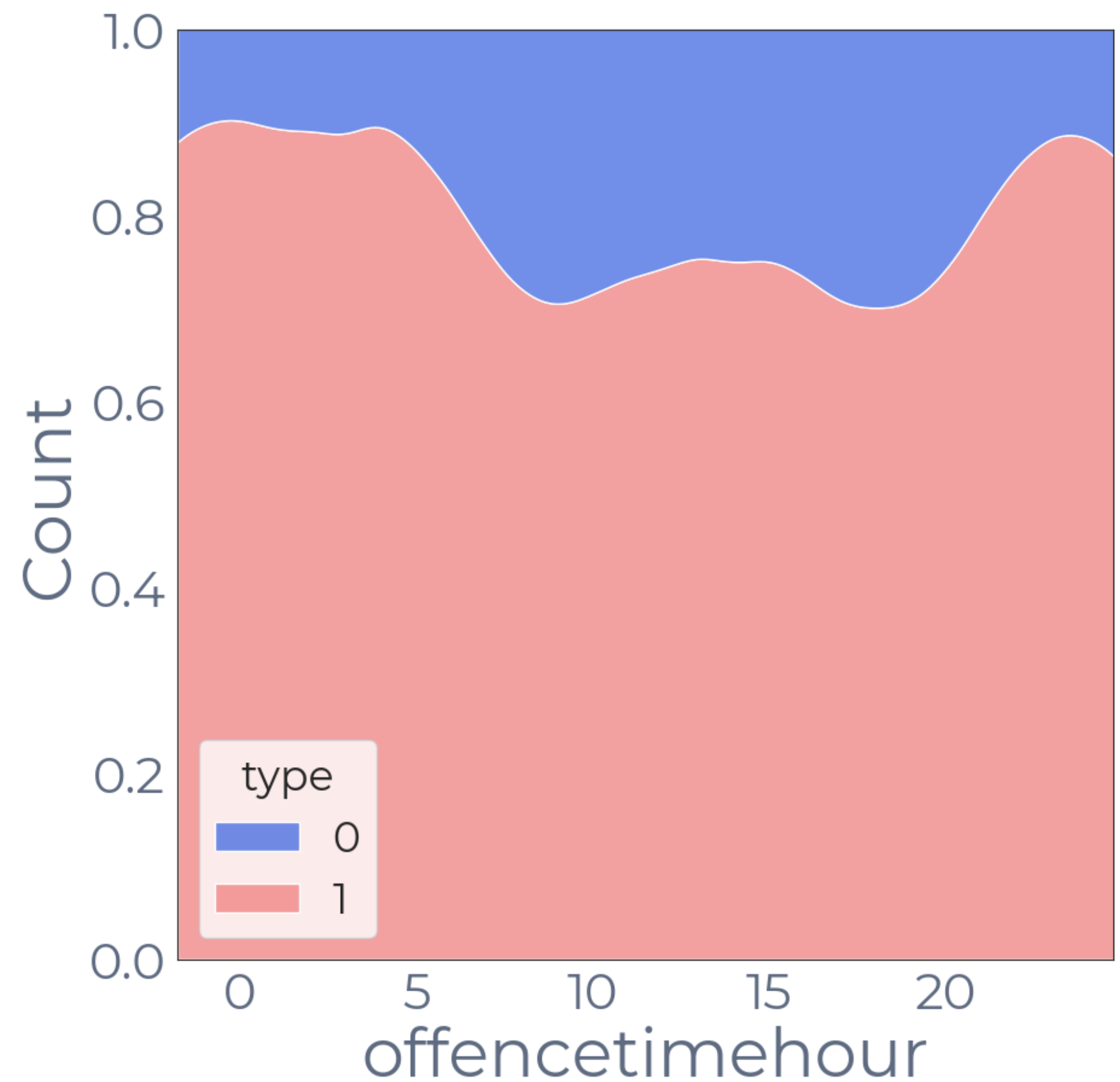
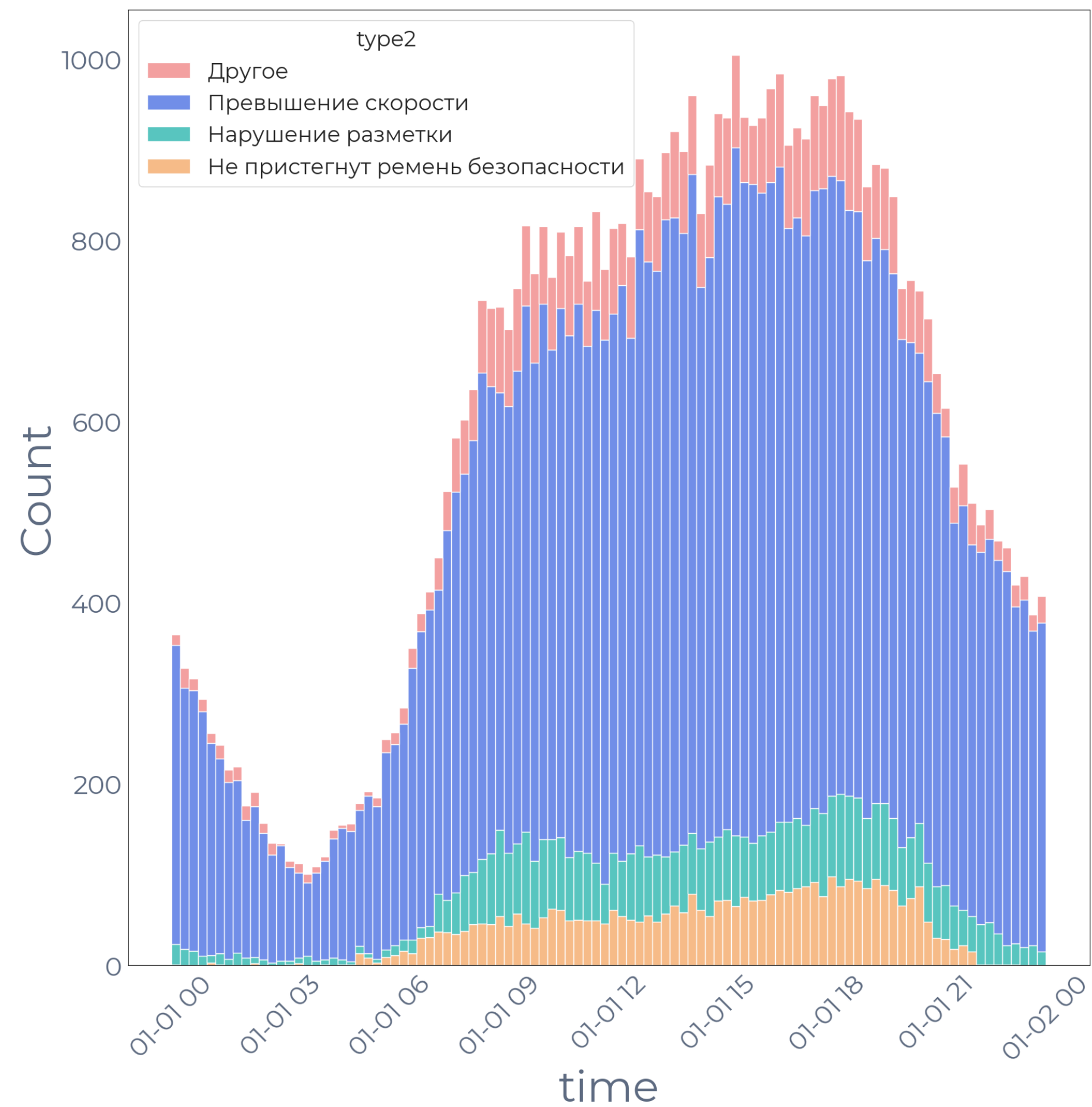
Самые популярные нарушения

Нарушение	Количество
Превышение 20-40	55333
Нарушение разметки	6304
Не пристегнут ремень	5310
Повторное превышение 40-60	1463
Движение по обочине	1103

Является ли нарушение скоростным



Распределение количества нарушений по часам





Взаимосвязаны ли **характеристики клиентов** и их автомобилей с **характеристиками правонарушений**?



В ночное время **чаще** превышают **скорость** в сравнении с другими нарушениями



В ночное время меньше участников дорожного движения и отвлекающих факторов



**Меньше
обстоятельств для
прочих нарушений**



**Больше
возможностей для
превышения скорости**



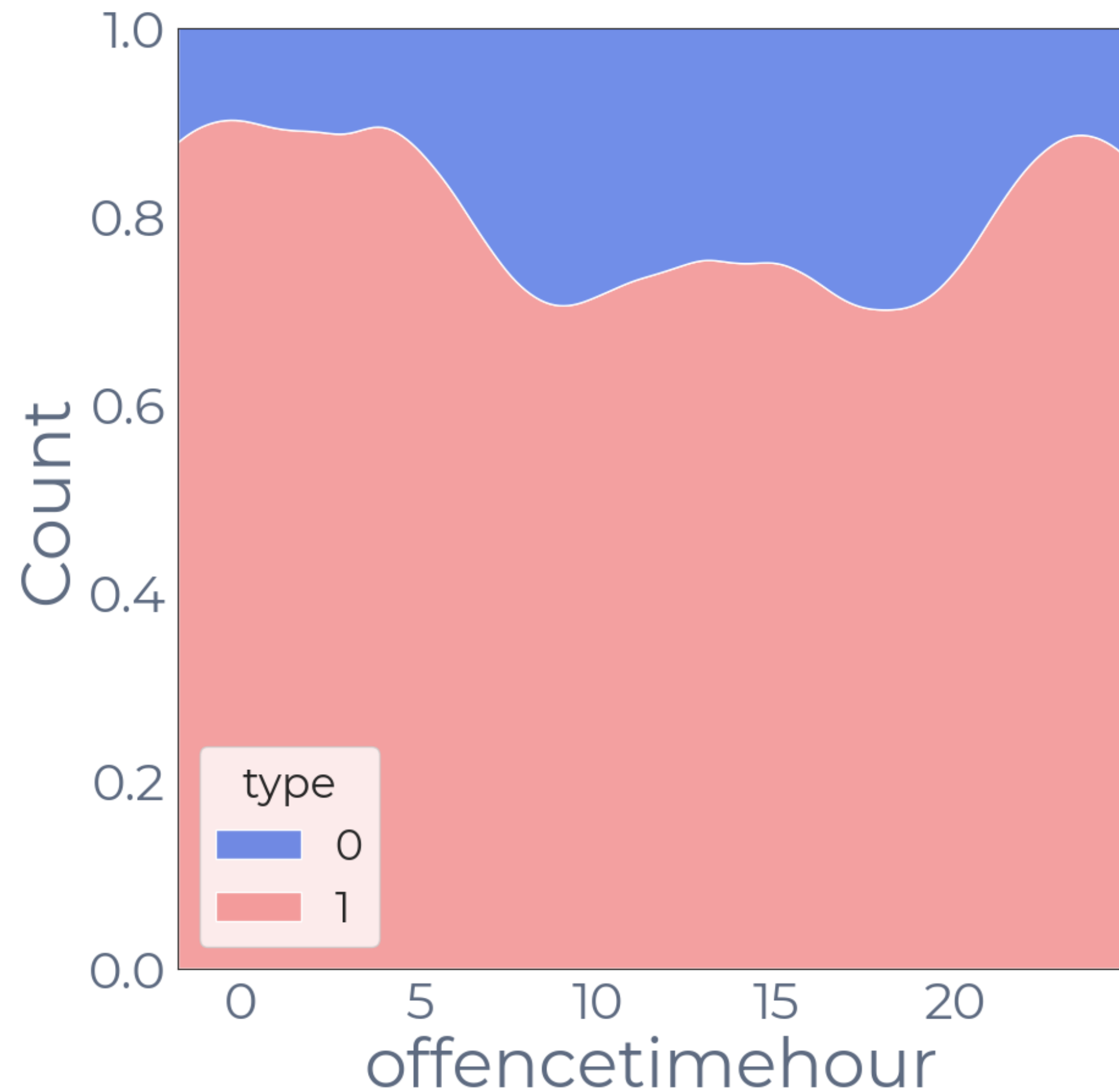
**Уменьшается доля
прочих нарушений**



**Увеличивается доля
превышений скорости**



Распределение количества нарушений по часам



Тест Манна-Уитни
p-value < 0.01

H0

Нет взаимосвязи

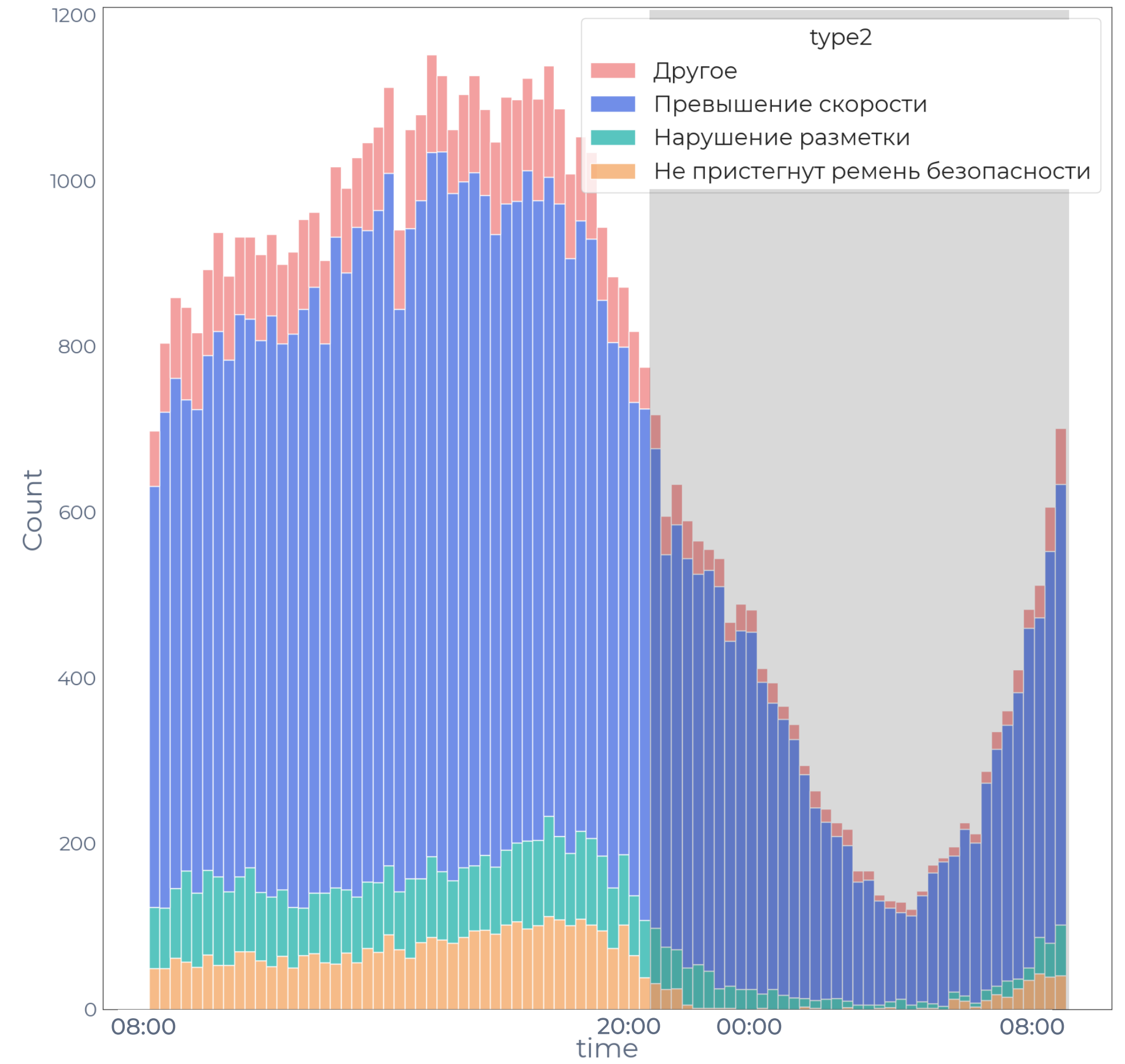
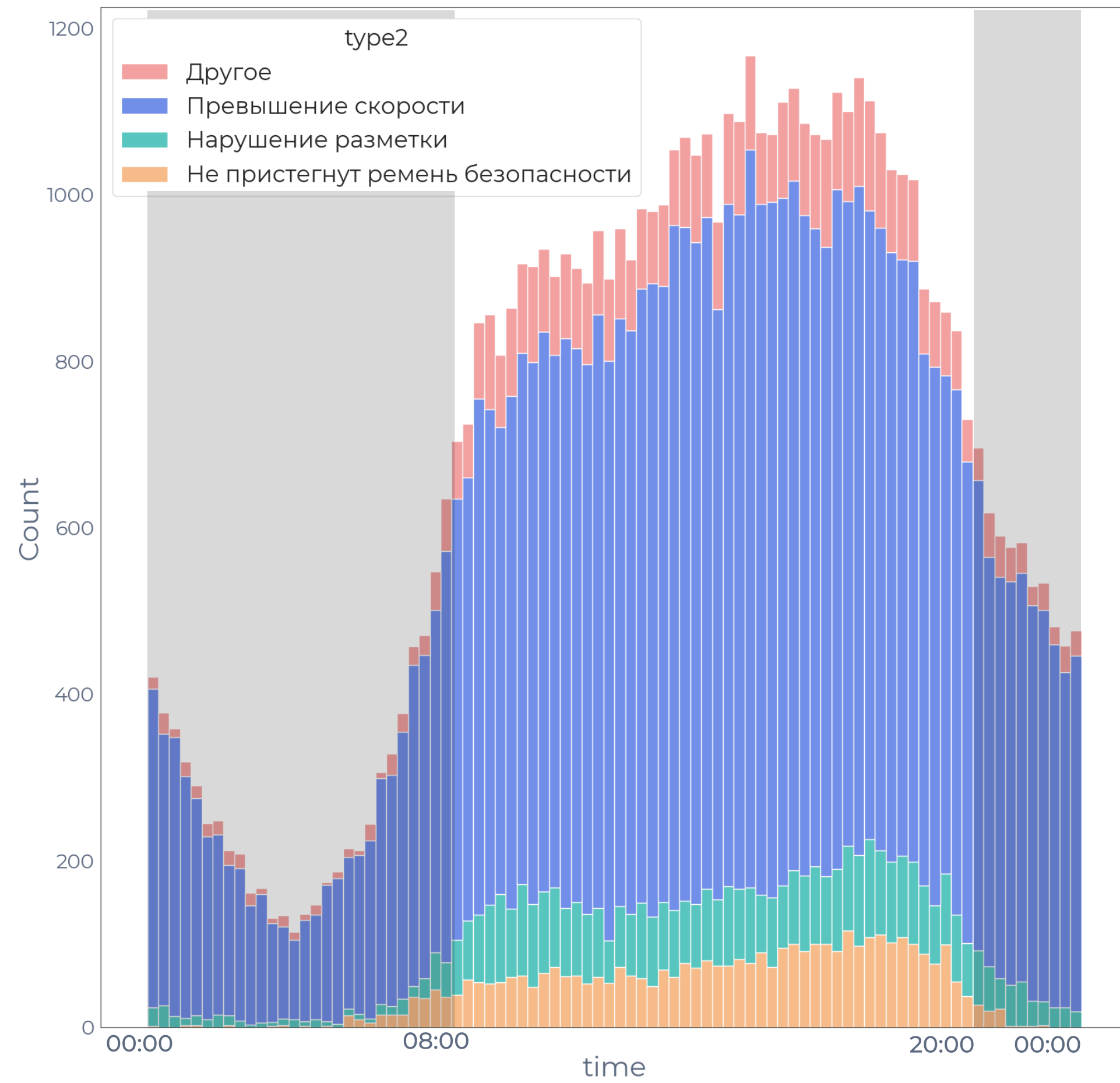


H1

Есть взаимосвязь



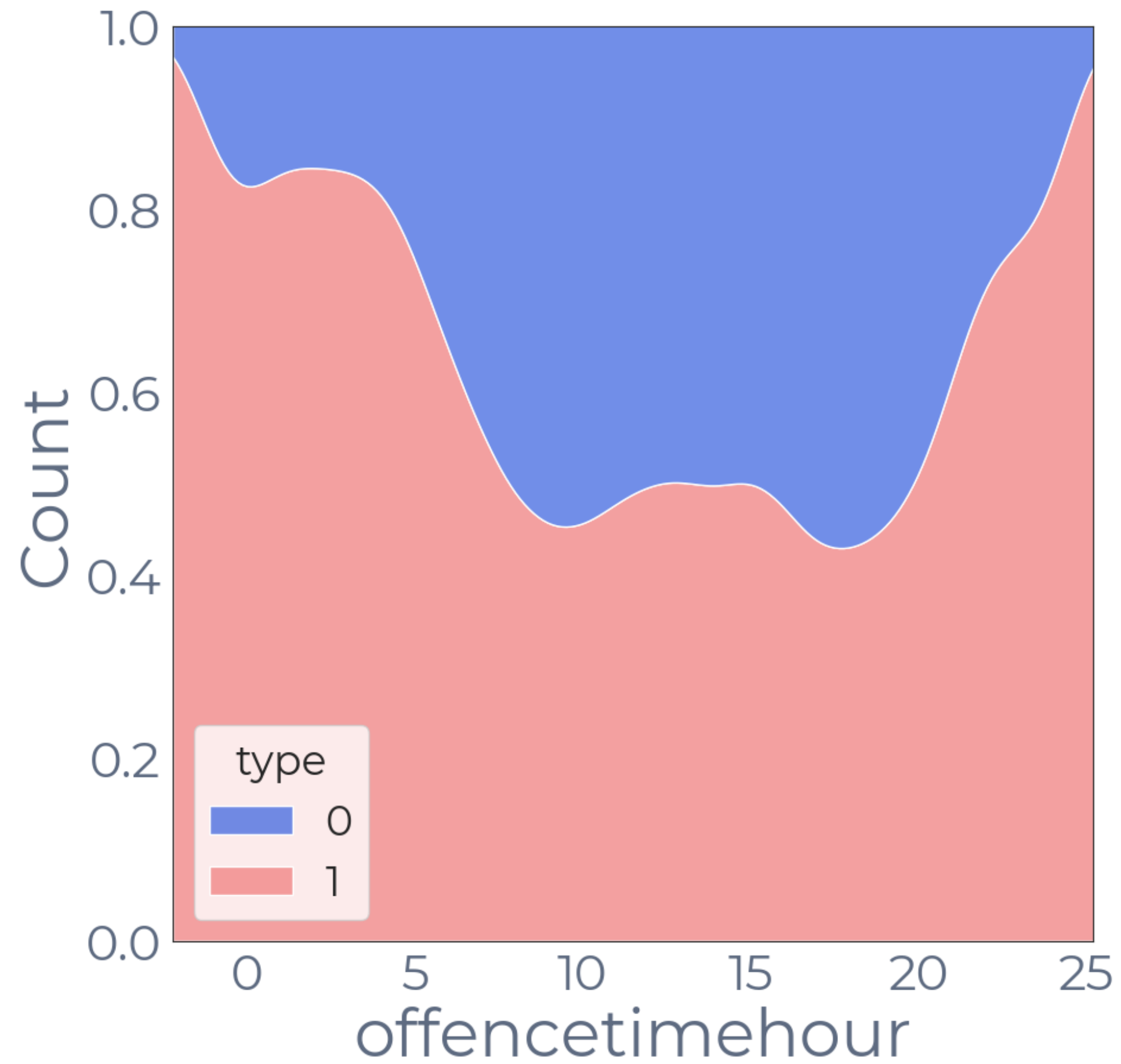
Распределение количества нарушений по часам



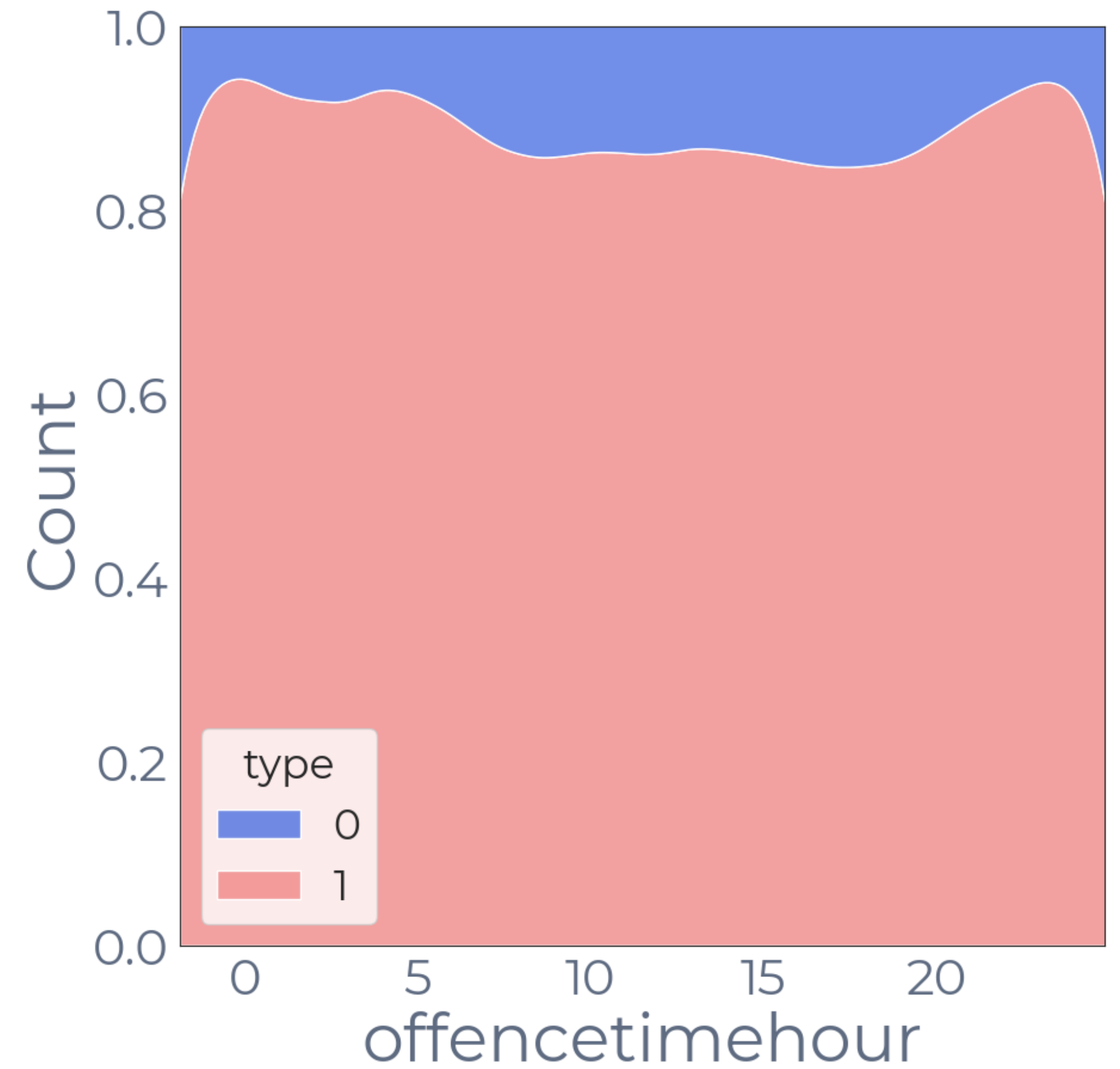
Распределение количества нарушений по часам

Тест Манна-Уитни

Москва, СПб p-value <0.01 ✓



Регионы p-value <0.01 ✓

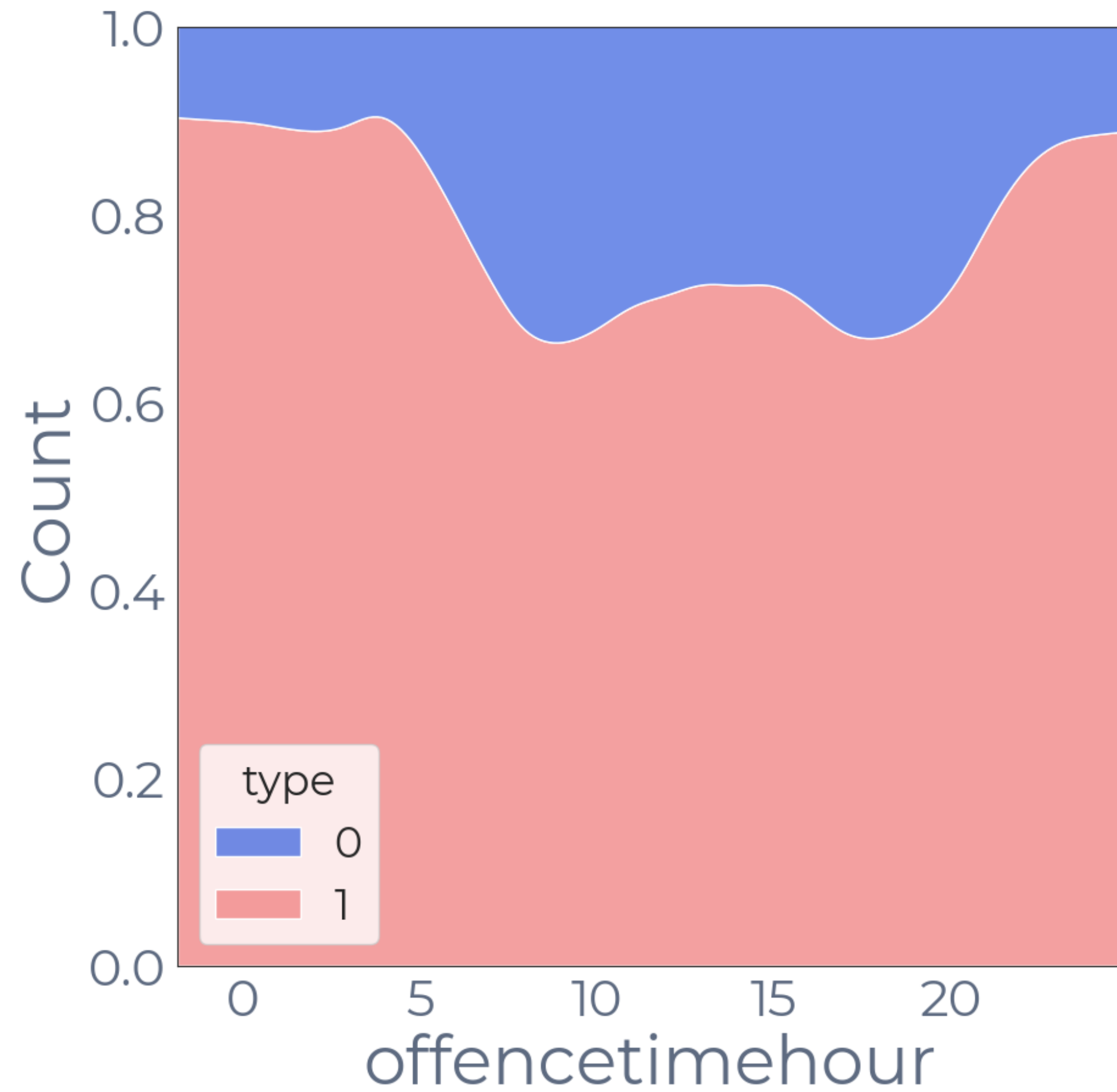


Распределение количества нарушений по часам

Тест Манна-Уитни

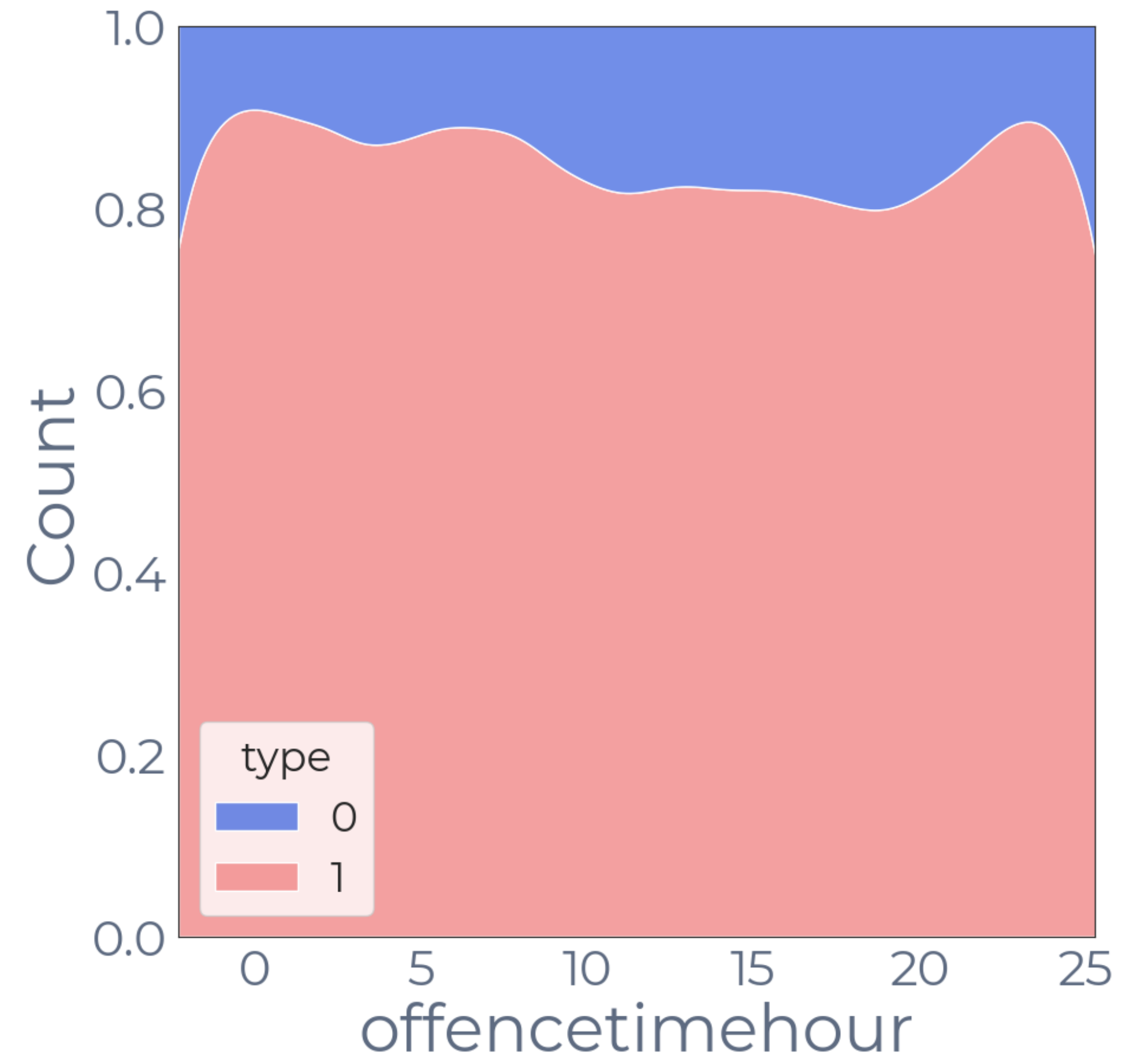
Будни

p-value < 0.01 ✓



Выходные

p-value < 0.01 ✓



Интерпретация

Гипотеза подтвердилась: доля нарушений скоростного режима ночью больше, чем днем

В дневное время нарушение скоростного режима проблематично из-за высокого трафика

Гипотеза подтверждается все зависимости от региона и дня недели



Policy implication



Не отключать светофоры на ночь



Увеличение суммы штрафа за превышение скорости в темное время суток



Ограничение максимальной скорости автомобиля



Ограничения



Возможное влияние других факторов



Данные только для одного месяца



Мало видов нарушений



Перспективы



Более точный анализ при увеличении выборки



Изучить данные обо всех водителях, а не только о нарушителях



Исследовать график работы патрулей ДПС и другие факторы





Владислав Денышев



Павел Савинов



Максим Садиков



Владимир Сиванев



PIZZA HUNTERS