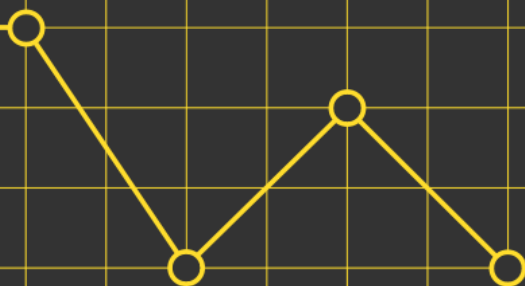
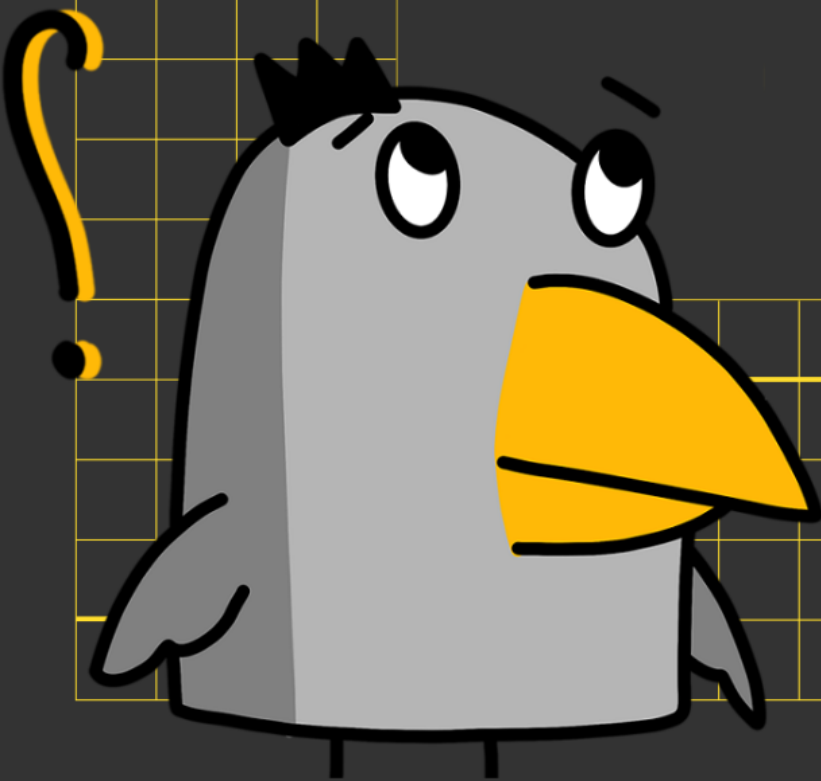


DANC

# Команда №7

Ментор – Дмитрий Захаров



ВЫСШАЯ ШКОЛА  
ЭКОНОМИКИ



# Обработка выбросов

person\_monthly\_income\_amt  
869128596

car\_price gender\_cd  
;M 32

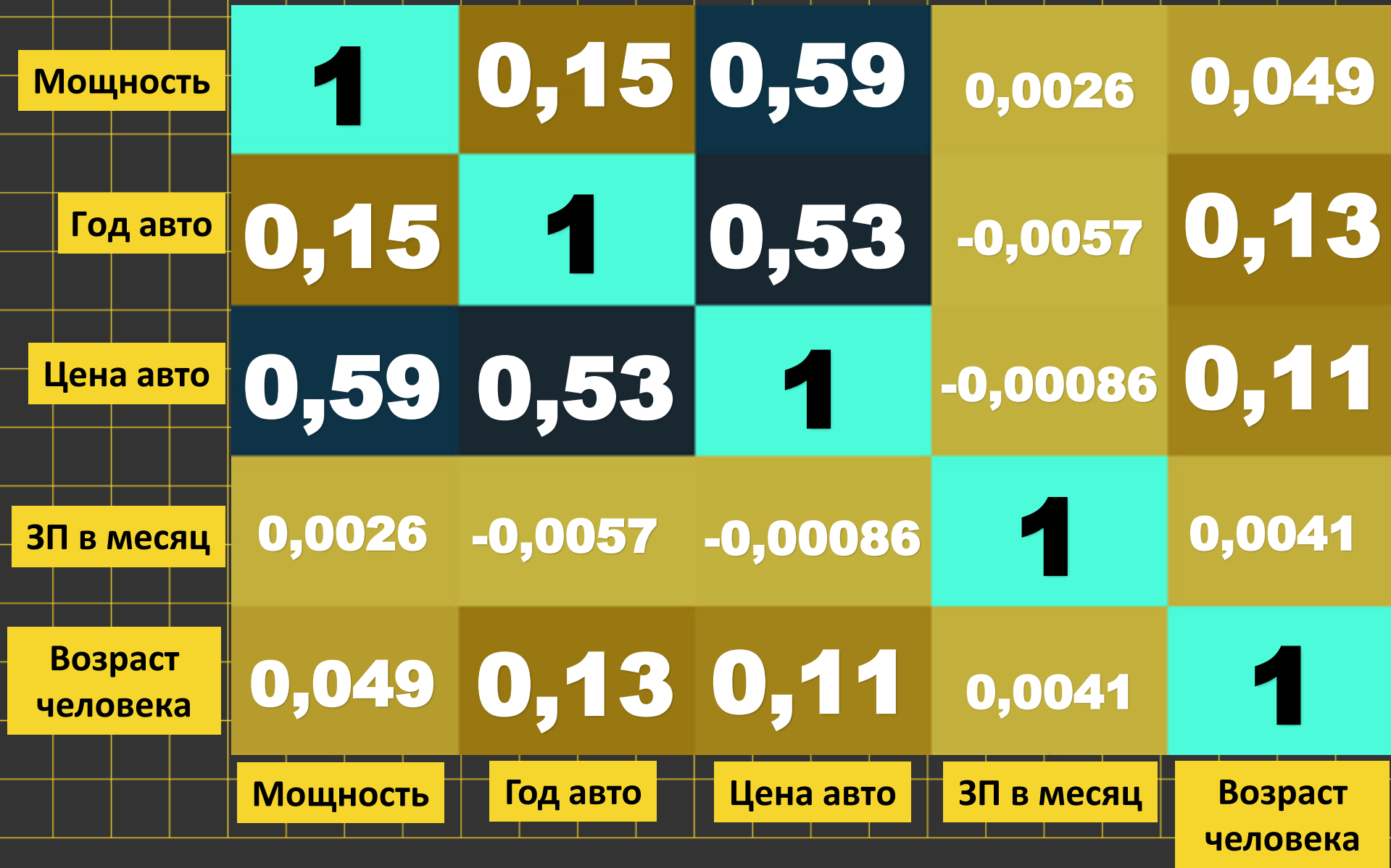
engine\_power  
0.00  
0.00  
0.00  
0.00  
0.00  
0.00  
0.00  
0.00  
0.00

1251,00	Универсал (5 дв.)	VOLVO
1346,00	Бортовой грузовик (2 дв.)	УАЗ
1401,00	Универсал (5 дв.)	FORD
1414,00	Универсал (5 дв.)	SSANGYONG
1495,00	Универсал (5 дв.)	SSANGYONG
1770,00	Седан (4 дв.)	AUDI

# Основные переменные

Ключ физ лица	party_rk
Регион	region
Какое нарушение	offenceshortstatement
Дата совершения нарушения	offencedate
Время совершения в часовом поясе региона	offencetime
Тип двигателя	engine_type
Мощность двигателя	engine_power
Тип кузова	body_type
Марка автомобиля (BMW, ВАЗ)	auto_mark
Год автомобиля	auto_year
Марка автомобиля (X7, 2107)	auto_model
Цвет	color
коробка передач	gear_type
Цена авто	car_price
Пол нарушителя	gender_cd
Возраст	age
Семейный статус	marital_status_cd
Кол-во детей	children_cnt
Образование	education_level_cd
Месячный доход	person_monthly_income_amt
День совершения правонарушения	day_of_week
Является ли выходным	public_holiday

# Корелляция переменных



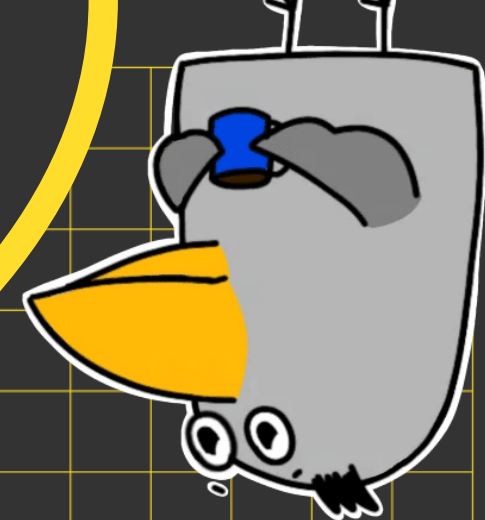
## Выводы:

1. В данных могут содержаться ошибки, нужно внимательно подходить к обработке данных
2. Мы завершили предобработку и можем начинать искать зависимости
3. Для нас главные переменные – мощность автомобиля, пол человека, возраст, его месячный доход и совершенные нарушения



# Гипотеза

Водители с более мощным автомобилем и высоким доходом более склонны к нарушению скоростного режима



## Механизм

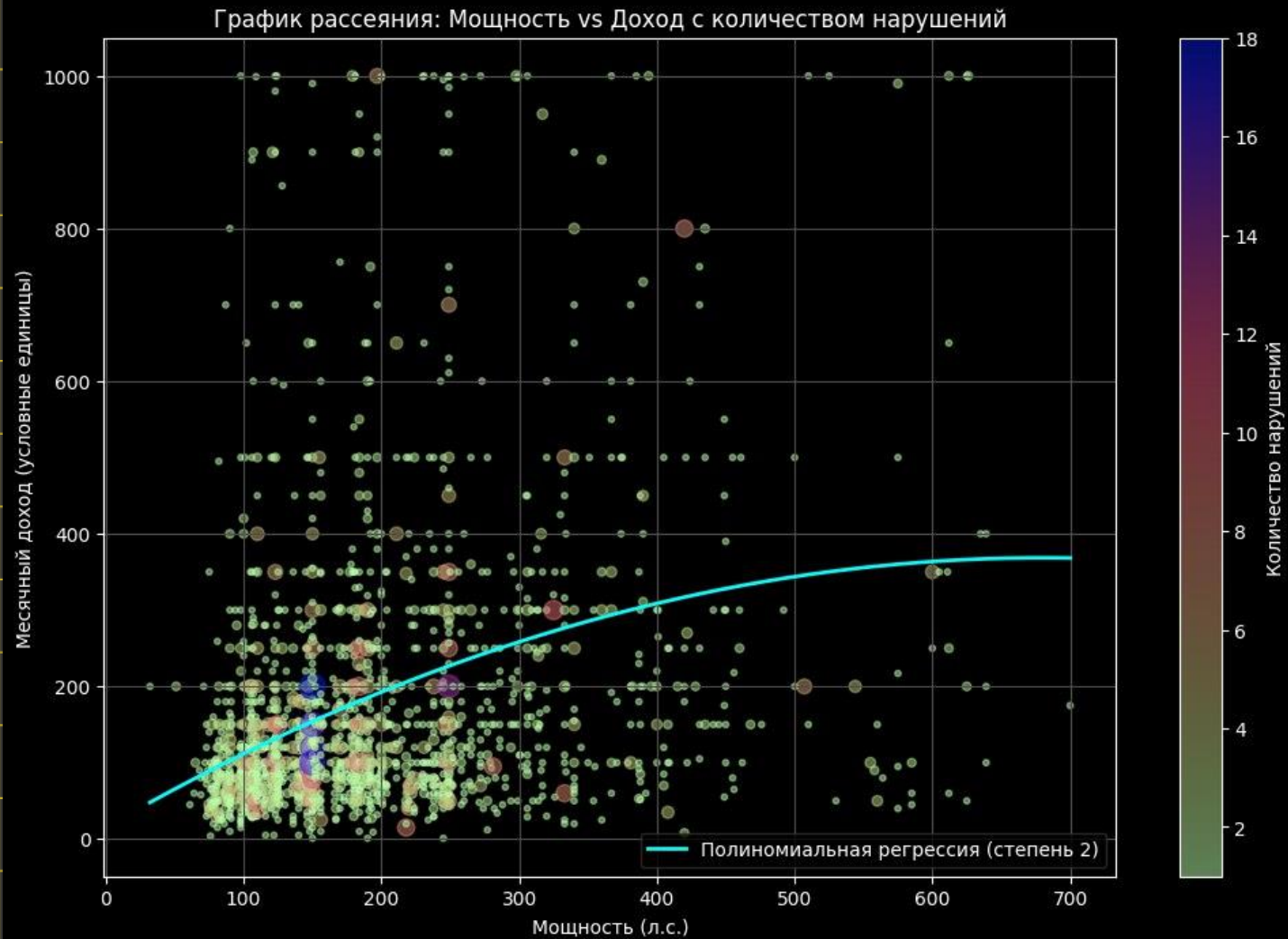
Люди, у которых большой доход, не особо переживают по поводу оплаты штрафа за какое-то незначительное нарушение, в отличии людей с низким доходом.

## Мат модель исследования

T-тест — это статистический метод для сравнения средних значений двух групп и проверки гипотезы о том, что эти средние значения существенно отличаются.

Используется, чтобы понять, случайна ли разница между группами или она действительно существует.

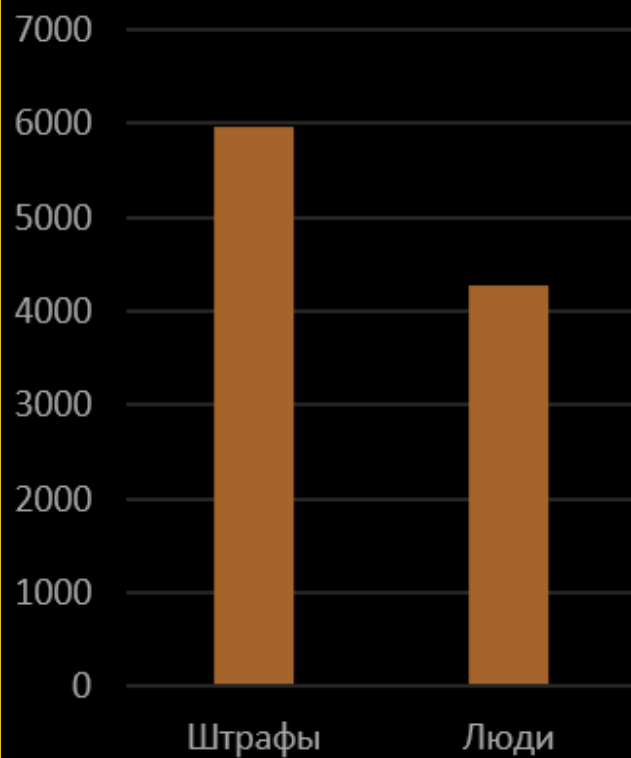
# Мат модель исследования



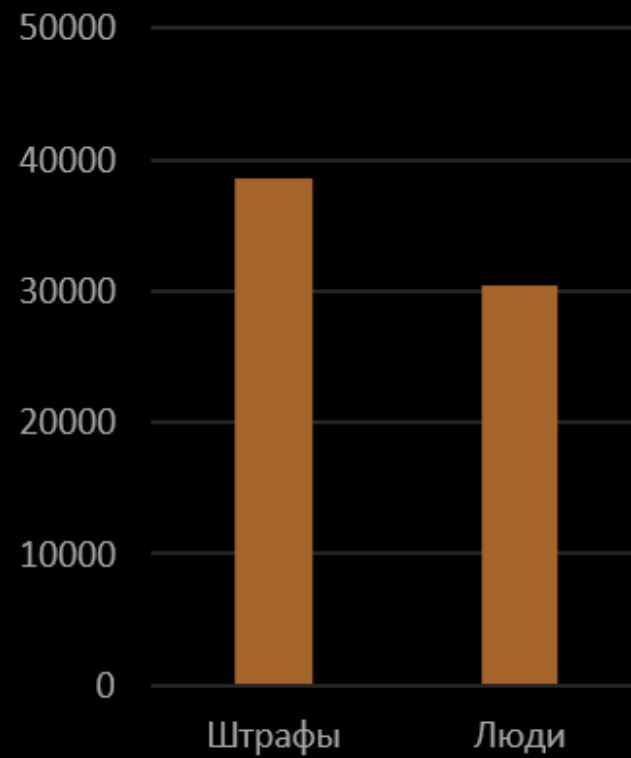
# Визуализация результата

**P-value: 1.452e-11**

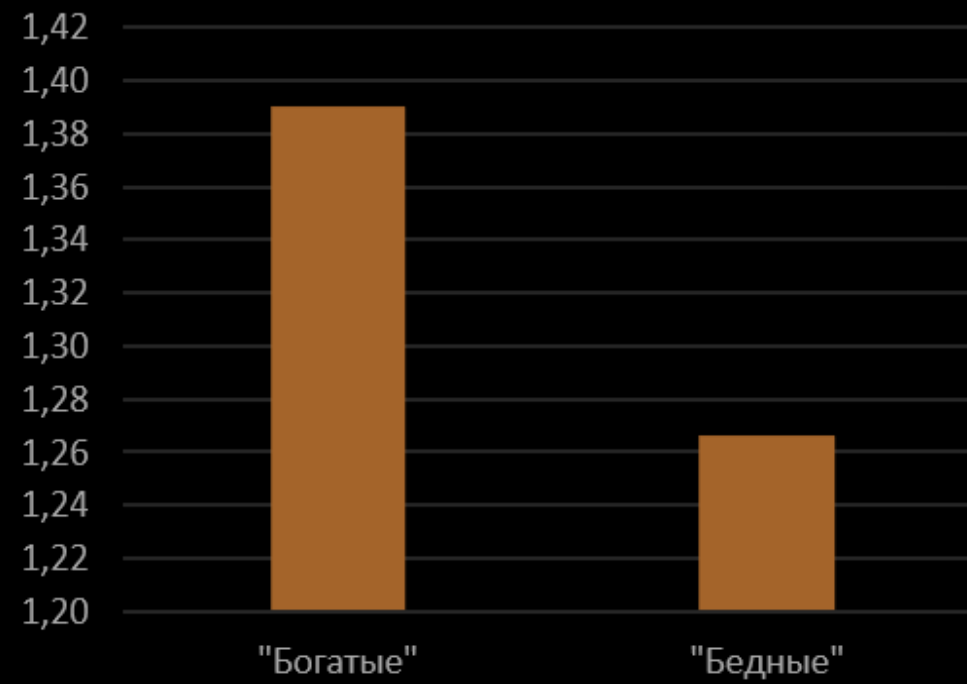
"Богатые"



"Бедные"



Среднее количество штрафов  
на человека на 1 человека

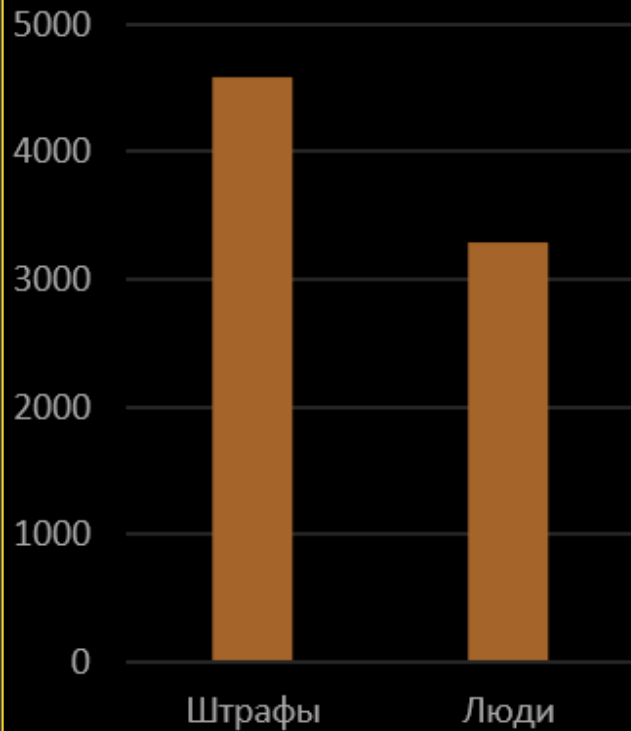




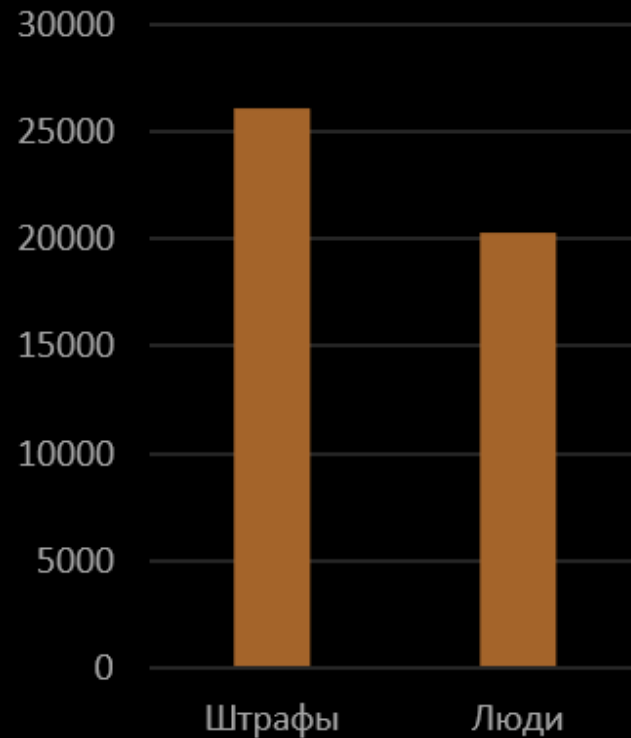
# Проверка работоспособности на подвыборках

**P-value: 1.037e-06**

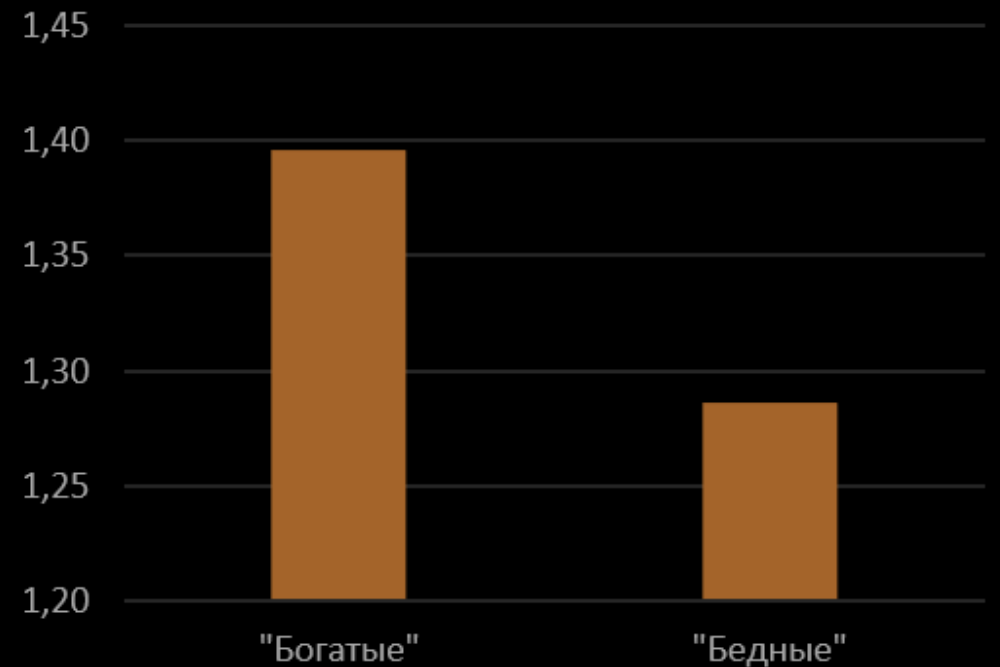
"Богатые" мужчины



"Бедные" мужчины



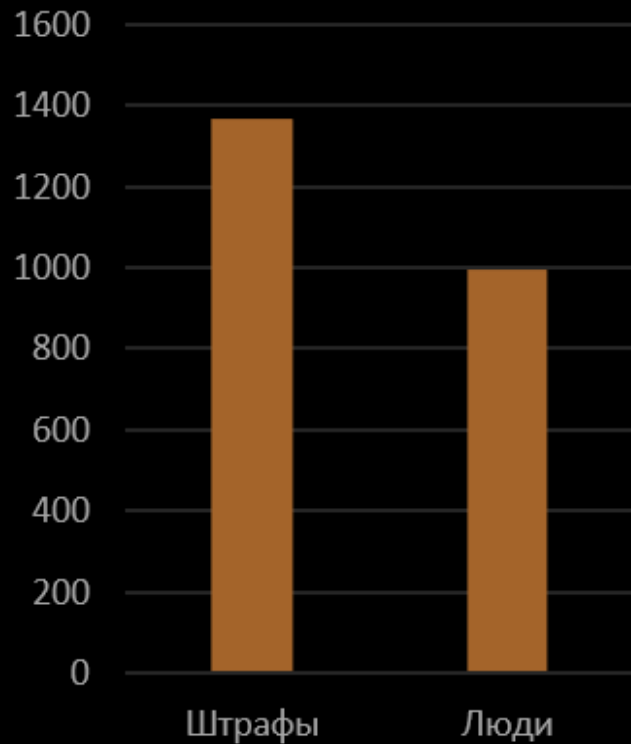
Среднее количество штрафов на человека на 1 мужчину



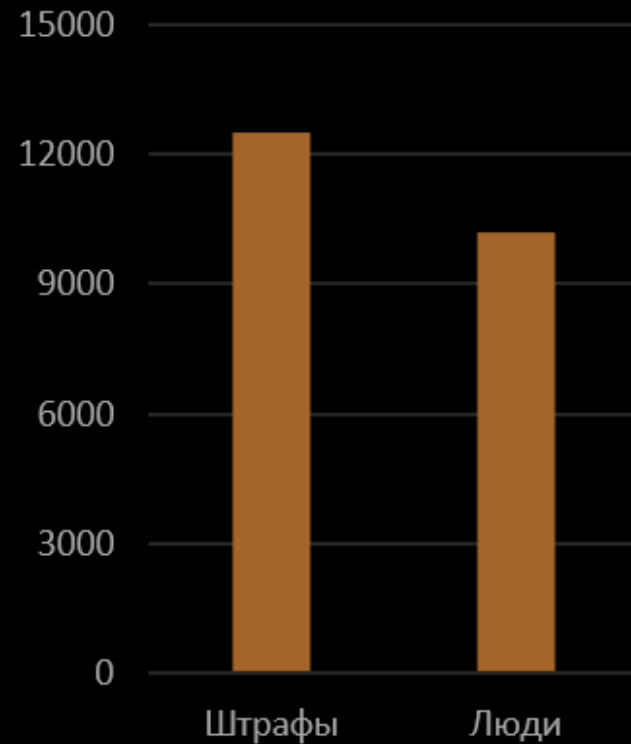
# Проверка работоспособности на подвыборках

**P-value: 6.685e-06**

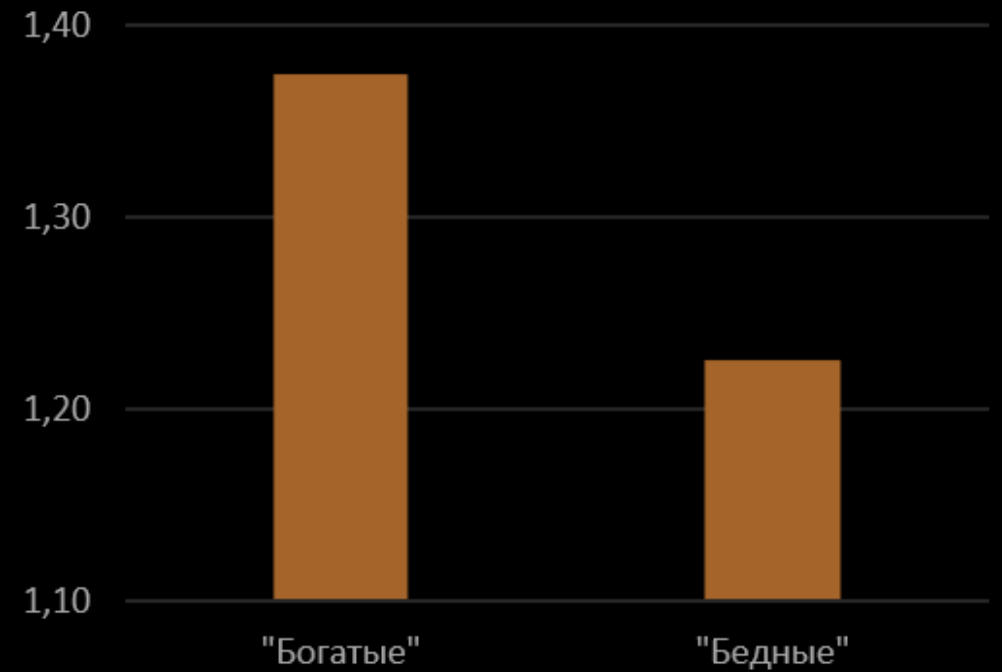
"Богатые" женщины



"Бедные" женщины



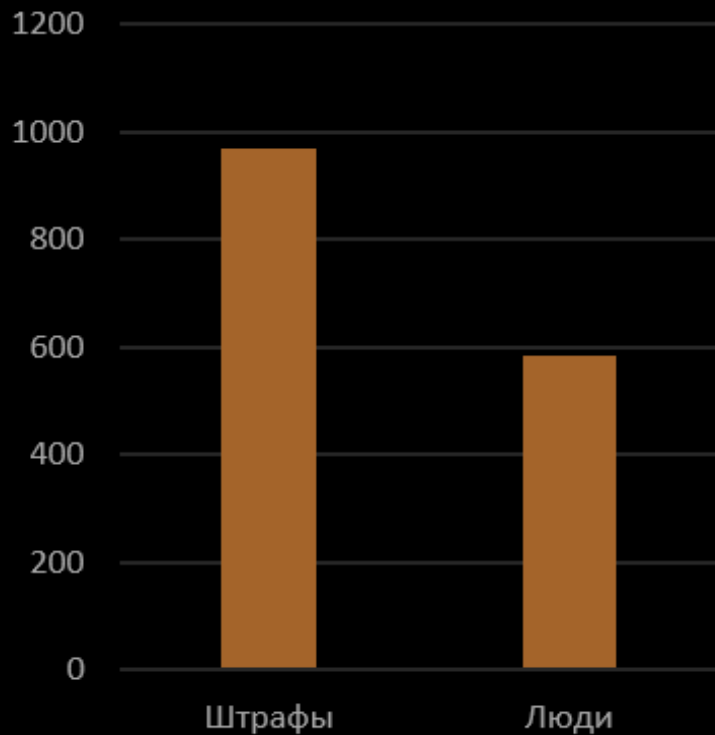
Среднее количество штрафов на человека на 1 женщину



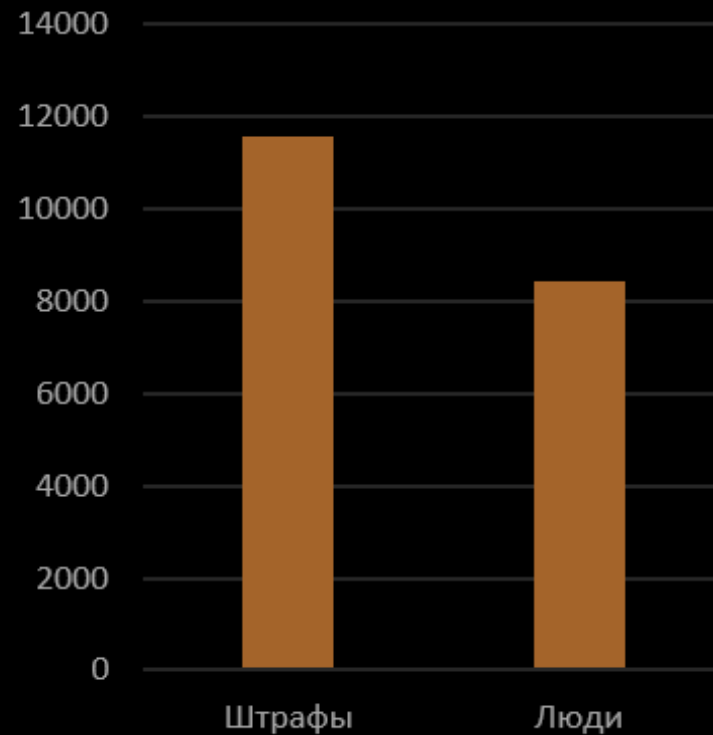
# Проверка работоспособности на подвыборках

**P-value: 1.387e-06**

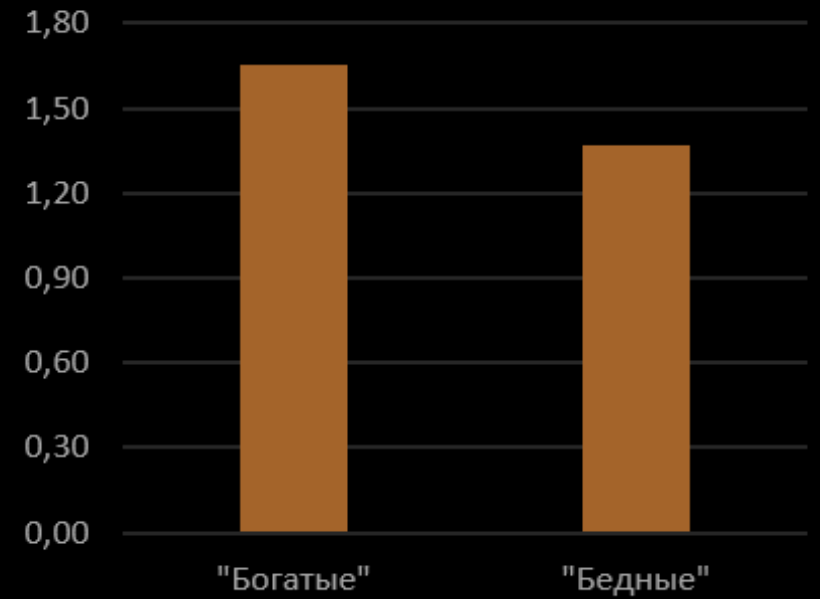
"Богатые" 18-29 лет



"Бедные" 18-29 лет



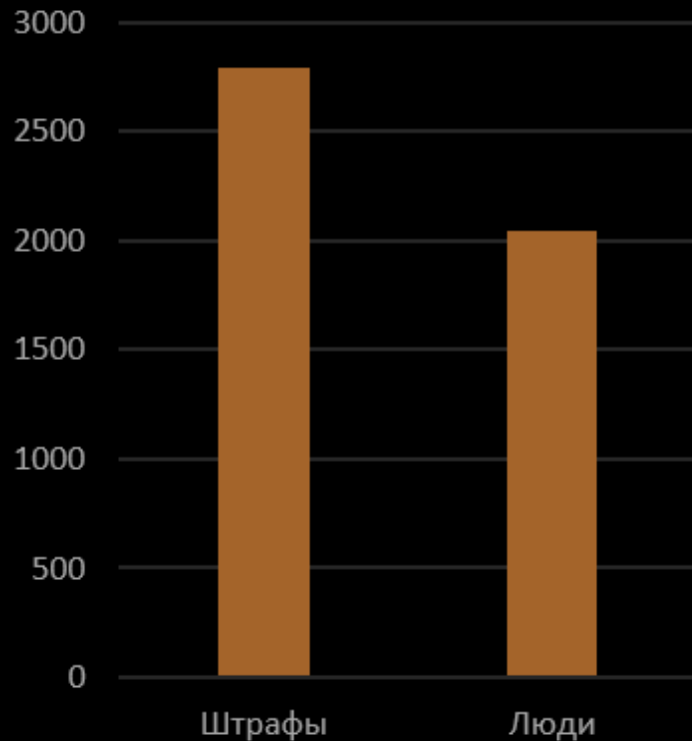
Среднее количество штрафов на человека 18-29 лет



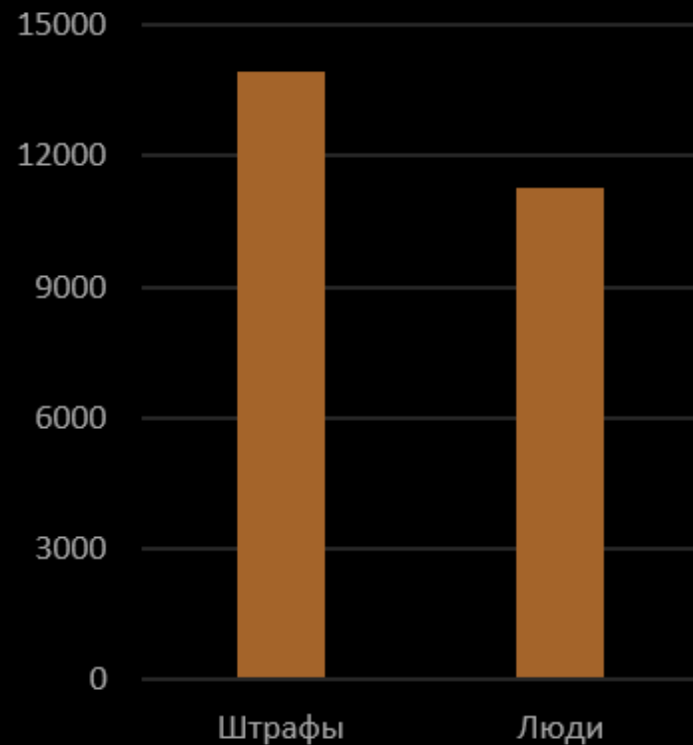
# Проверка работоспособности на подвыборках

**P-value: 3.802e-07**

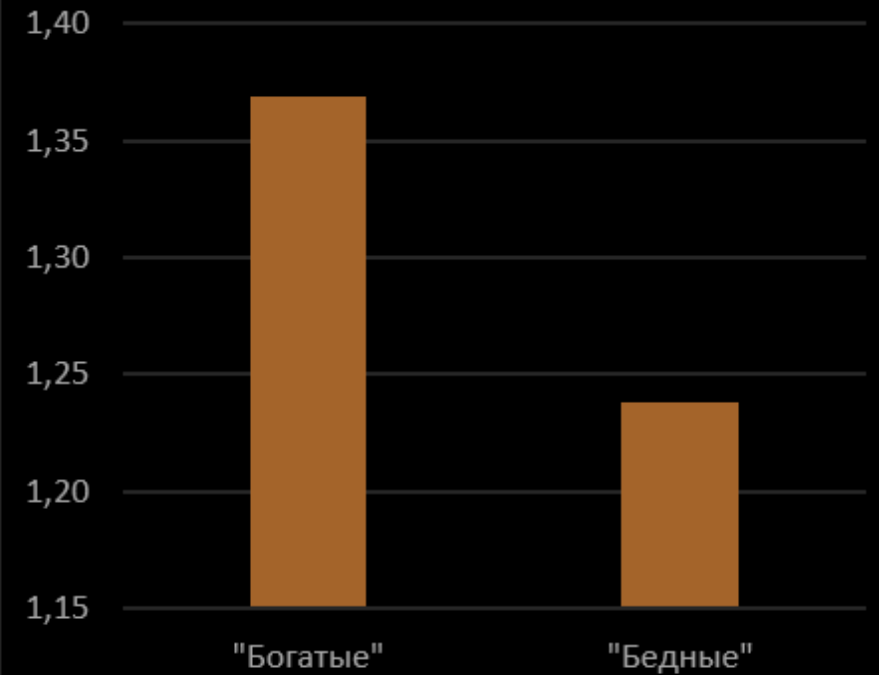
"Богатые" 30-39 лет



"Бедные" 30-39 лет



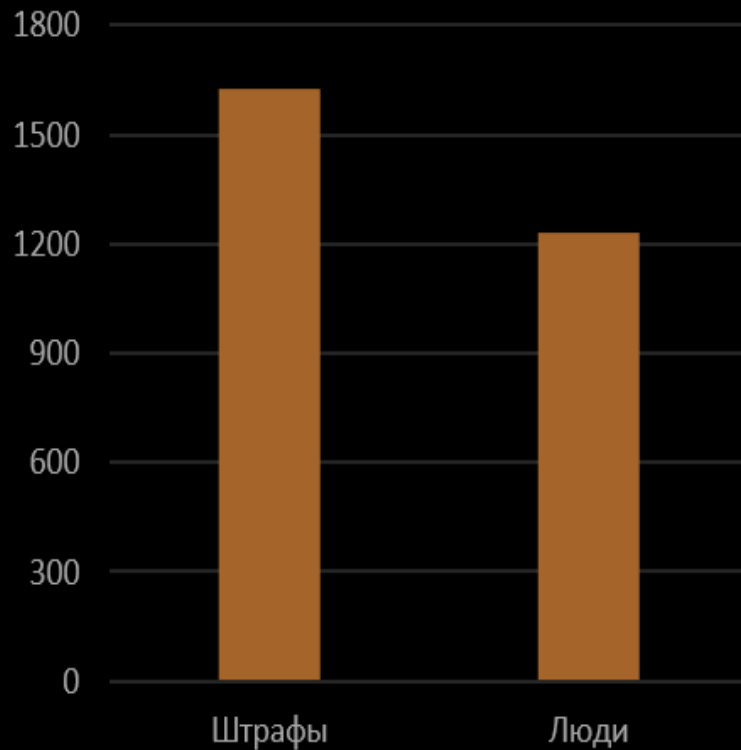
Среднее количество штрафов на человека 30-39 лет



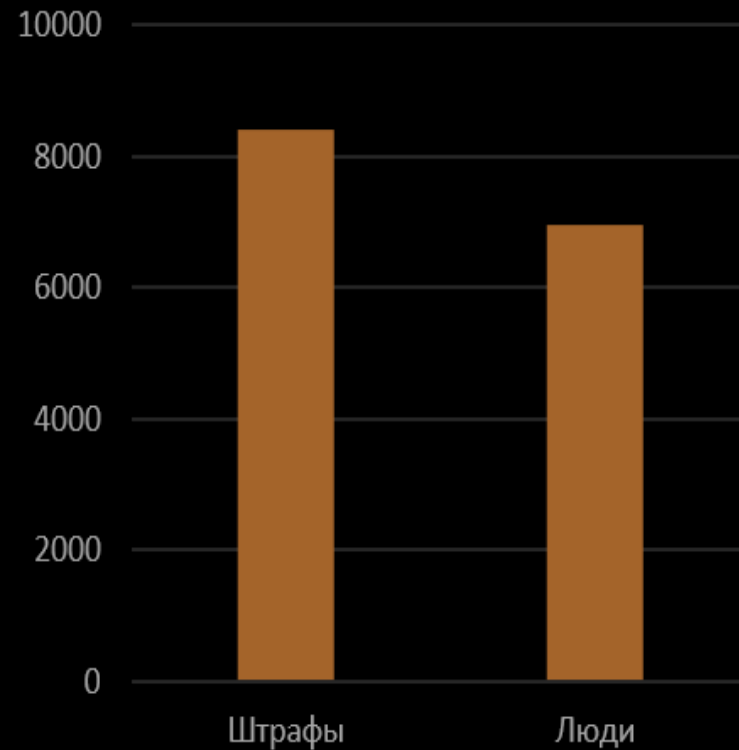
# Проверка работоспособности на подвыборках

**P-value: 0.000326**

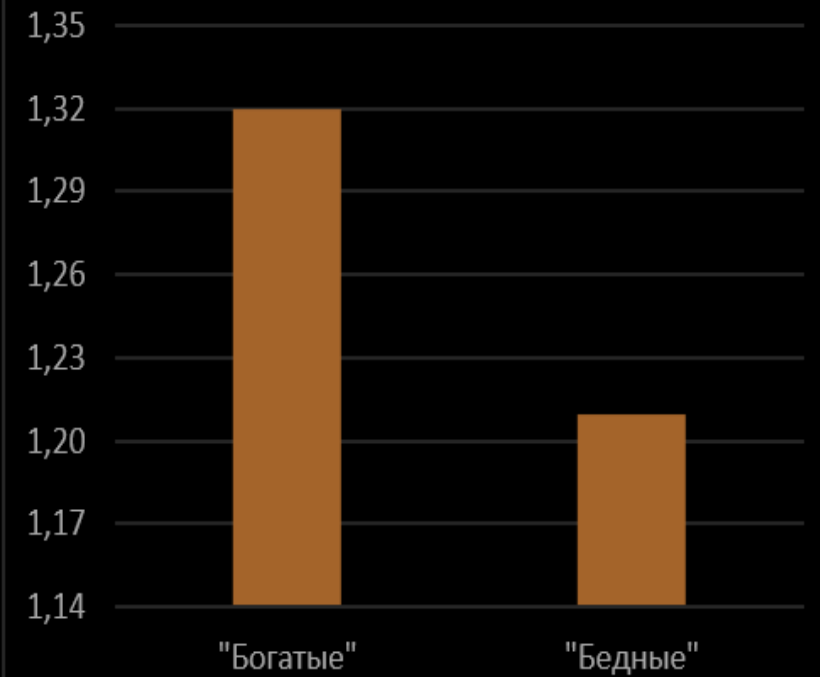
"Богатые" 40-49 лет



"Бедные" 40-49 лет



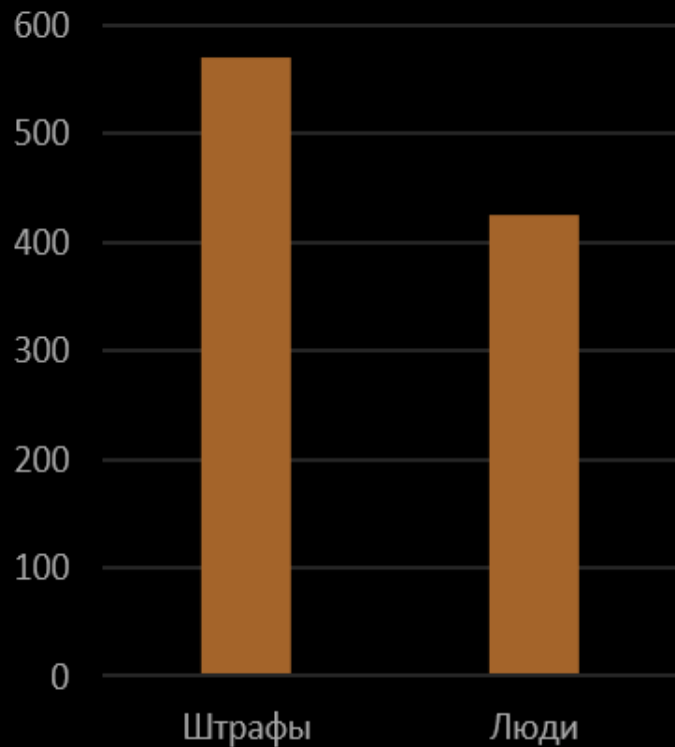
Среднее количество штрафов на человека 40-49 лет



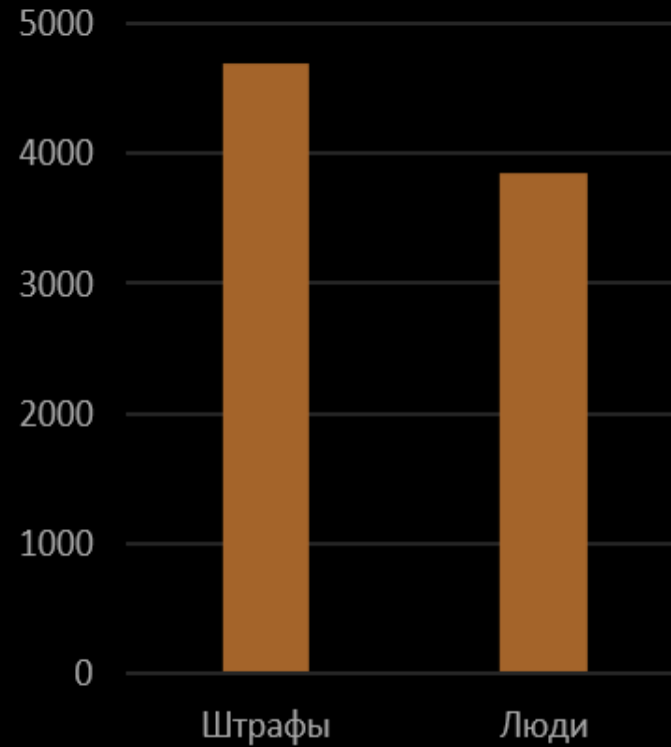
# Проверка работоспособности на подвыборках

**P-value: 0.0172**

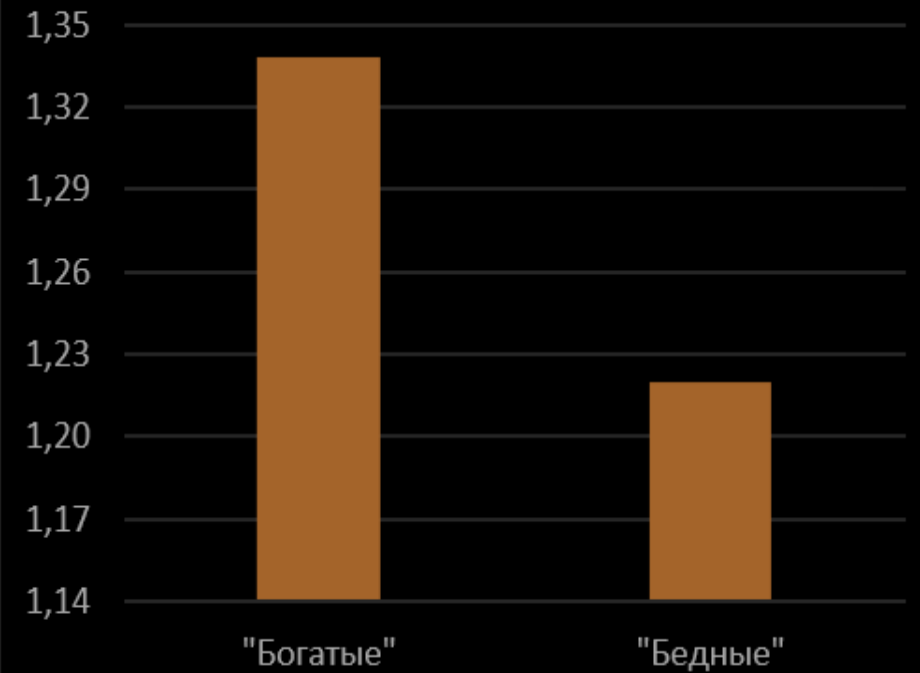
"Богатые" 50+ лет



"Бедные" 50+ лет

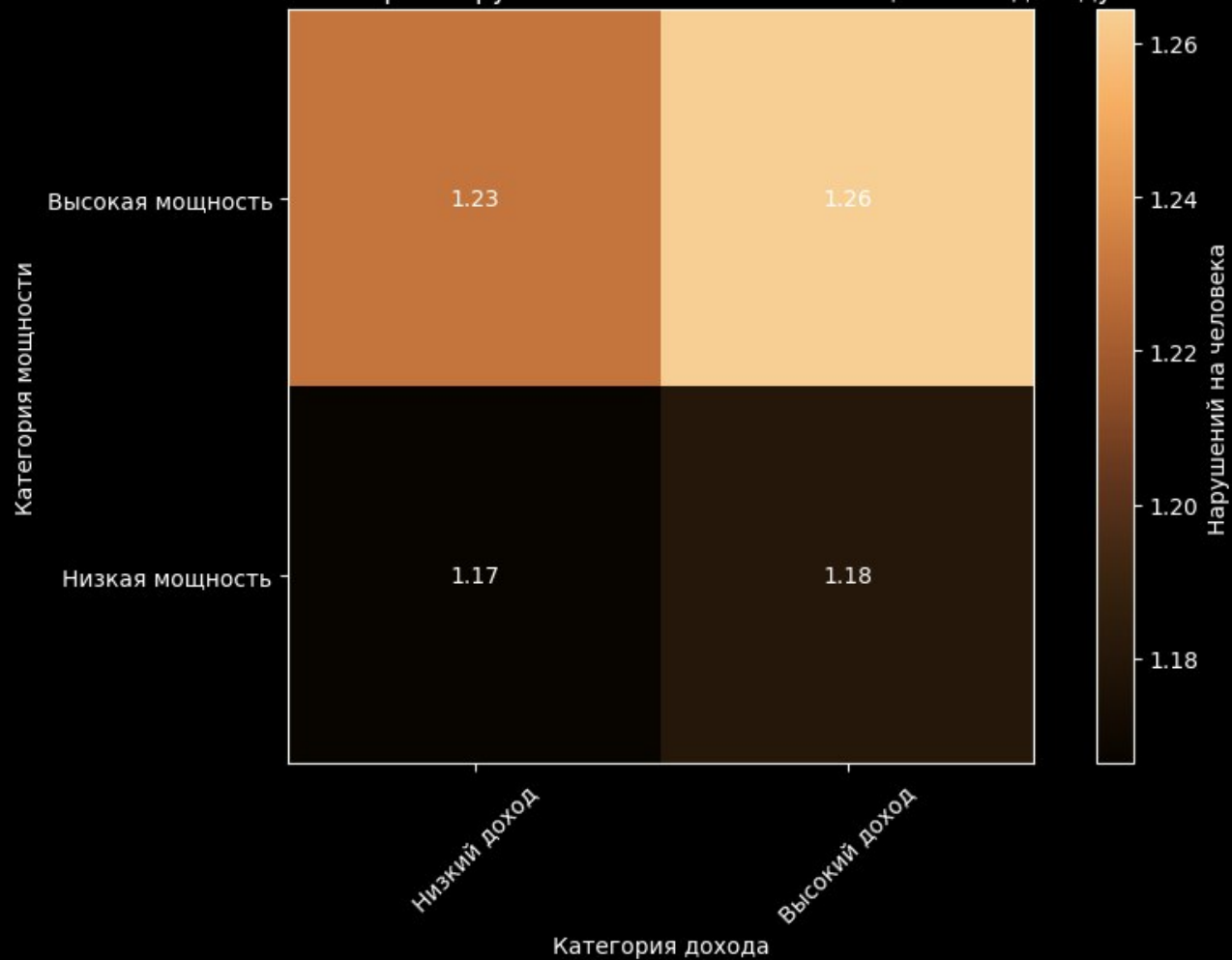


Среднее количество штрафов на человека 50+ лет

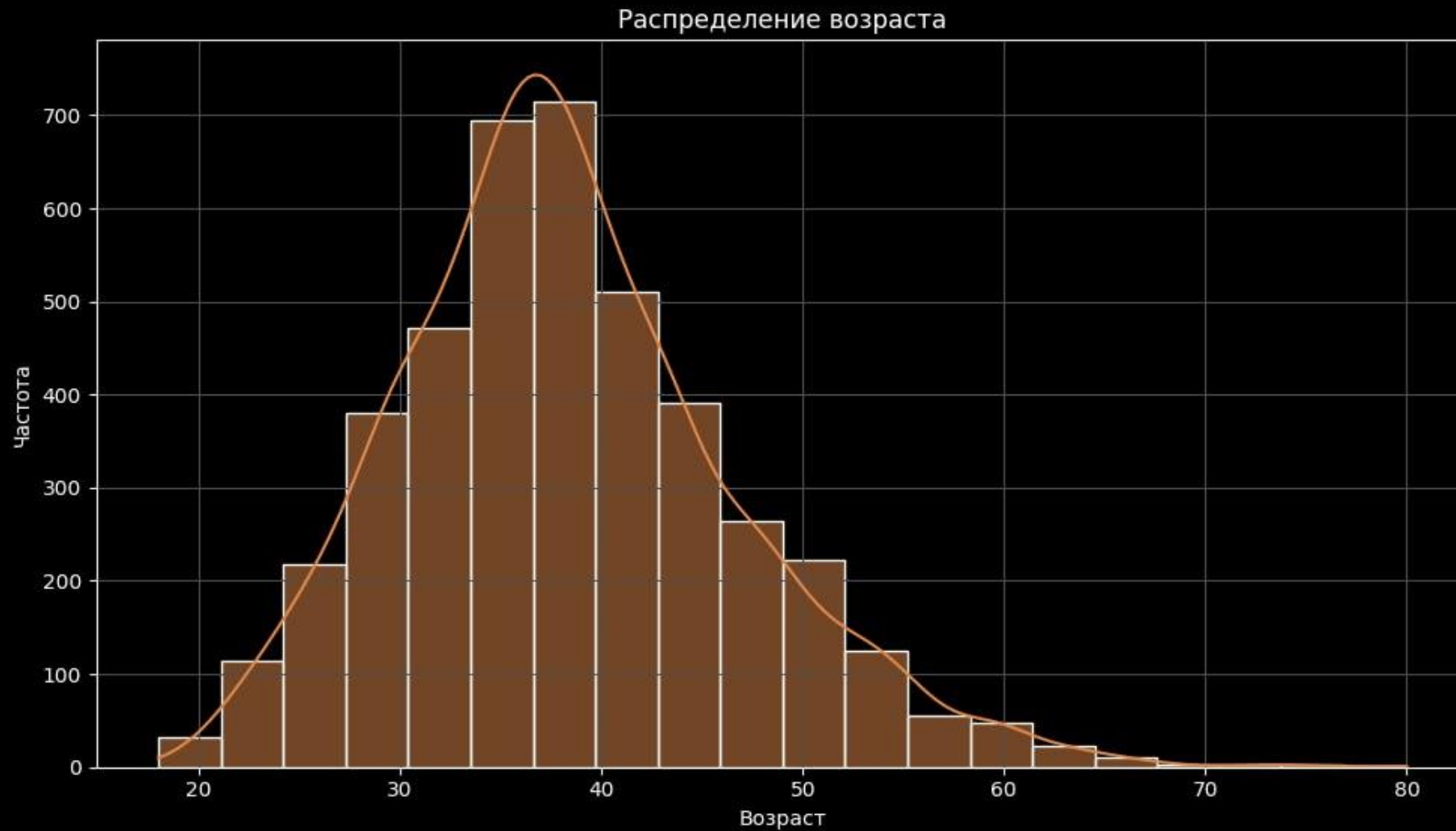


# Визуализация результата

Тепловая карта нарушений на человека по мощности и доходу



# Визуализация результата





# Интерпретация

Можно сказать, что более мощные автомобили и высокий доход связаны с более высоким уровнем превышения скорости, что может указывать на определённые поведенческие тенденции среди этих водителей.

## Практическая польза

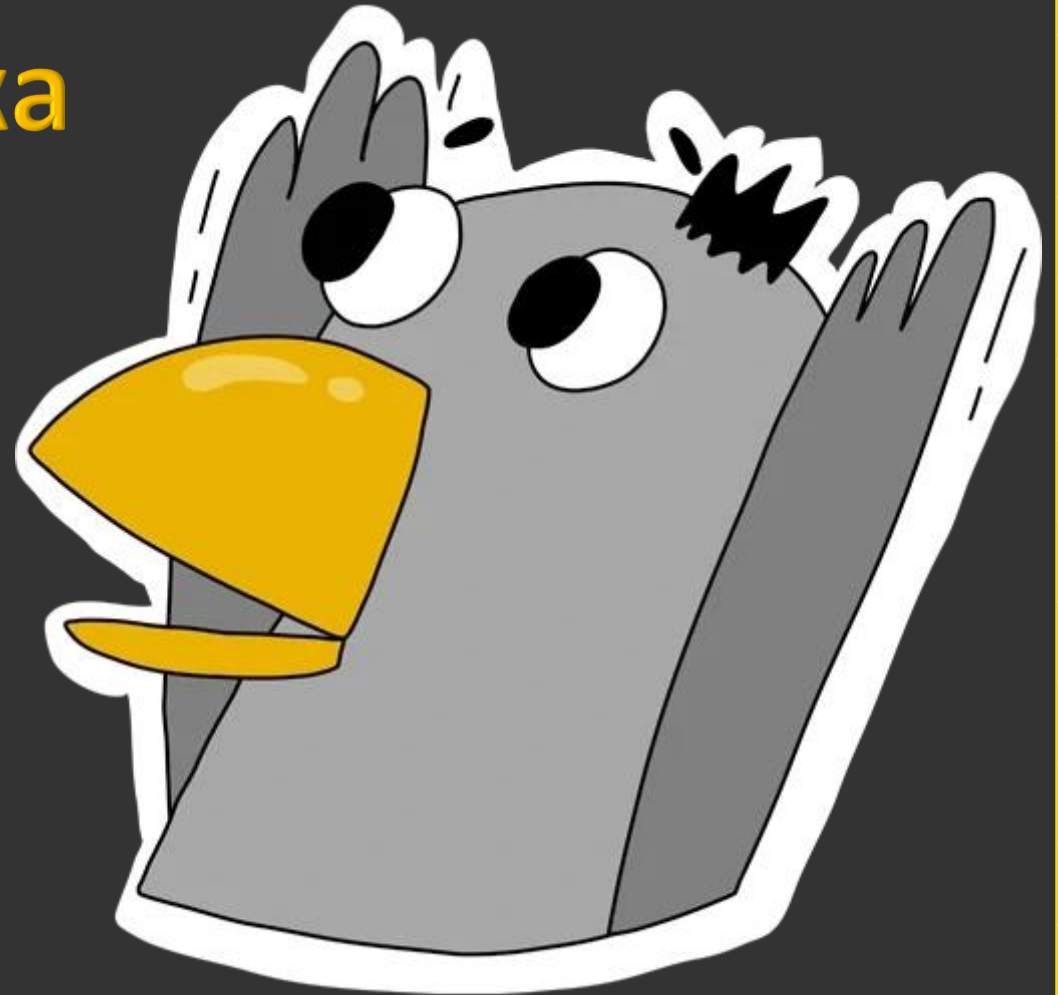
ФНС могут повышать налогообложение для тех людей, которые обладают высокой заработной платой или мощным автомобилем.

Может помочь в дальнейших исследованиях по Data-Science по штрафам ГИБДД



# Ограничения

- Данные одного банка
- Один месяц



# Перспективы

- Нет стажа вождения
- Нет пола ребёнка

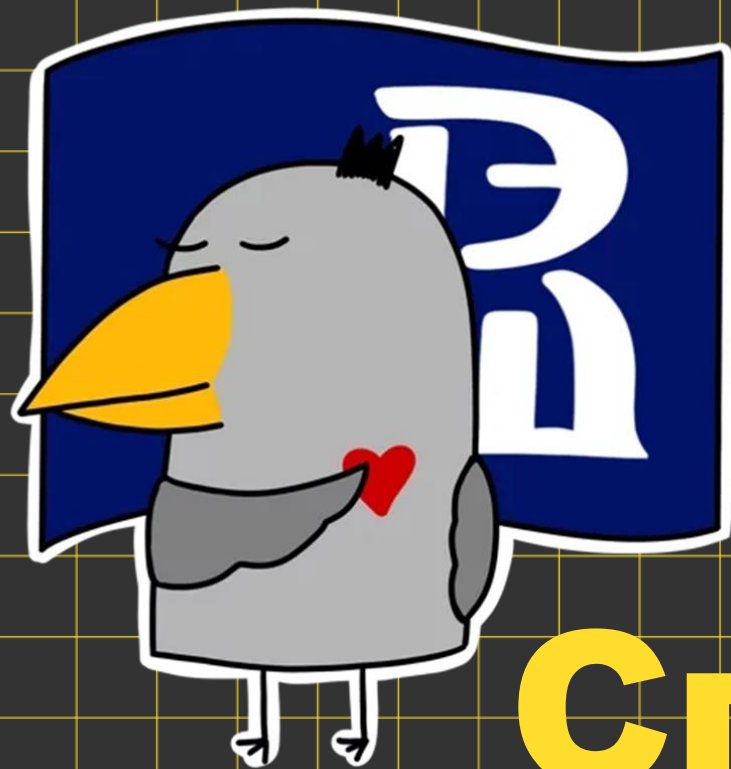


# Минусы модели

# Модель



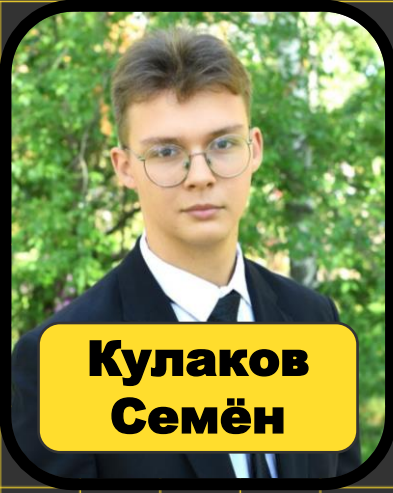
# ИДЕАЛЬНА



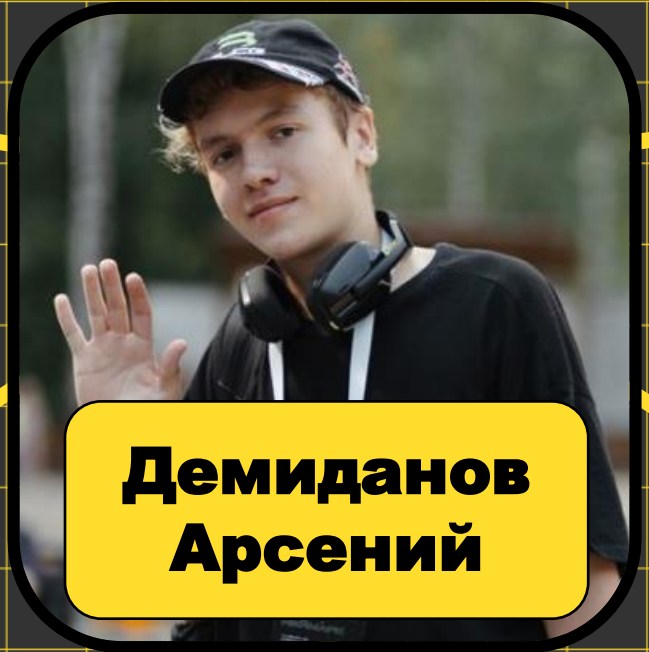
**Спасибо!**  
**Спасибо....**

© Кар-Карыч





**Кулаков  
Семён**



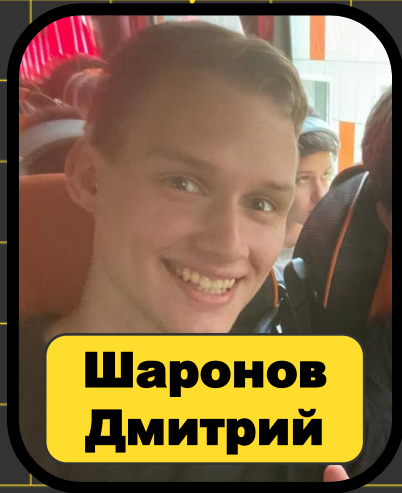
**Демиданов  
Арсений**



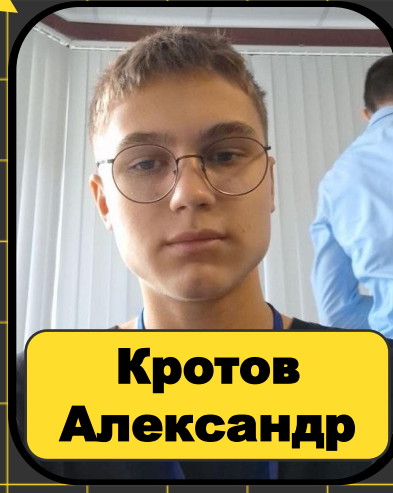
**Евстратов  
Илья**



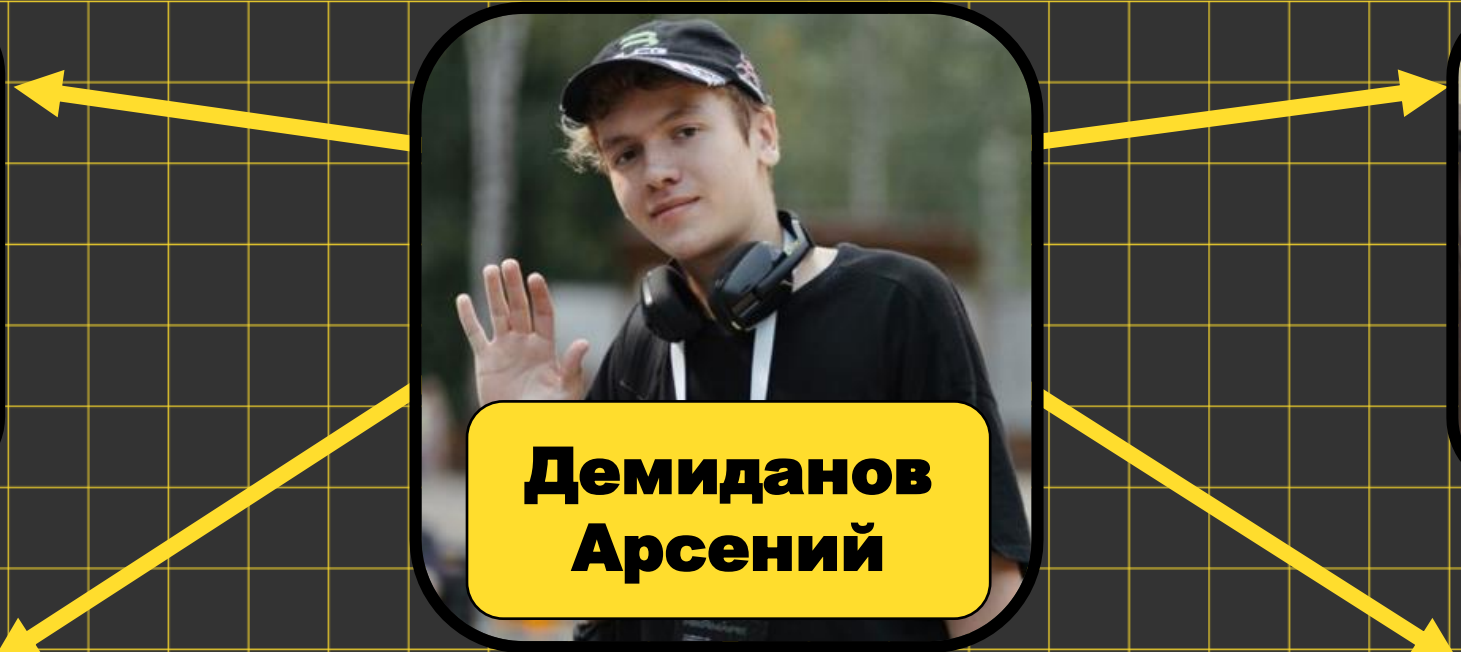
**Парамонов  
Даниил**



**Шаронов  
Дмитрий**



**Кротов  
Александр**

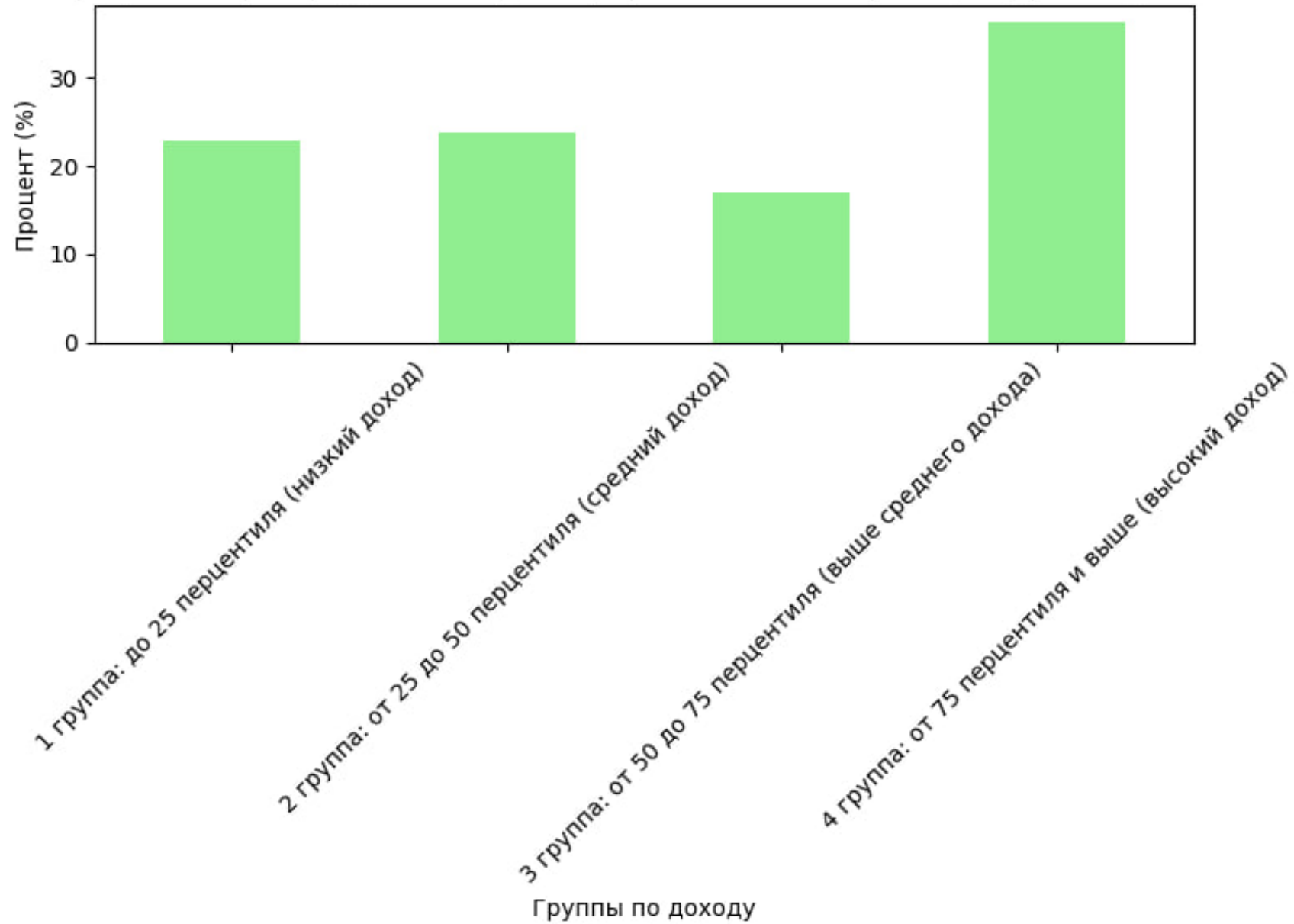




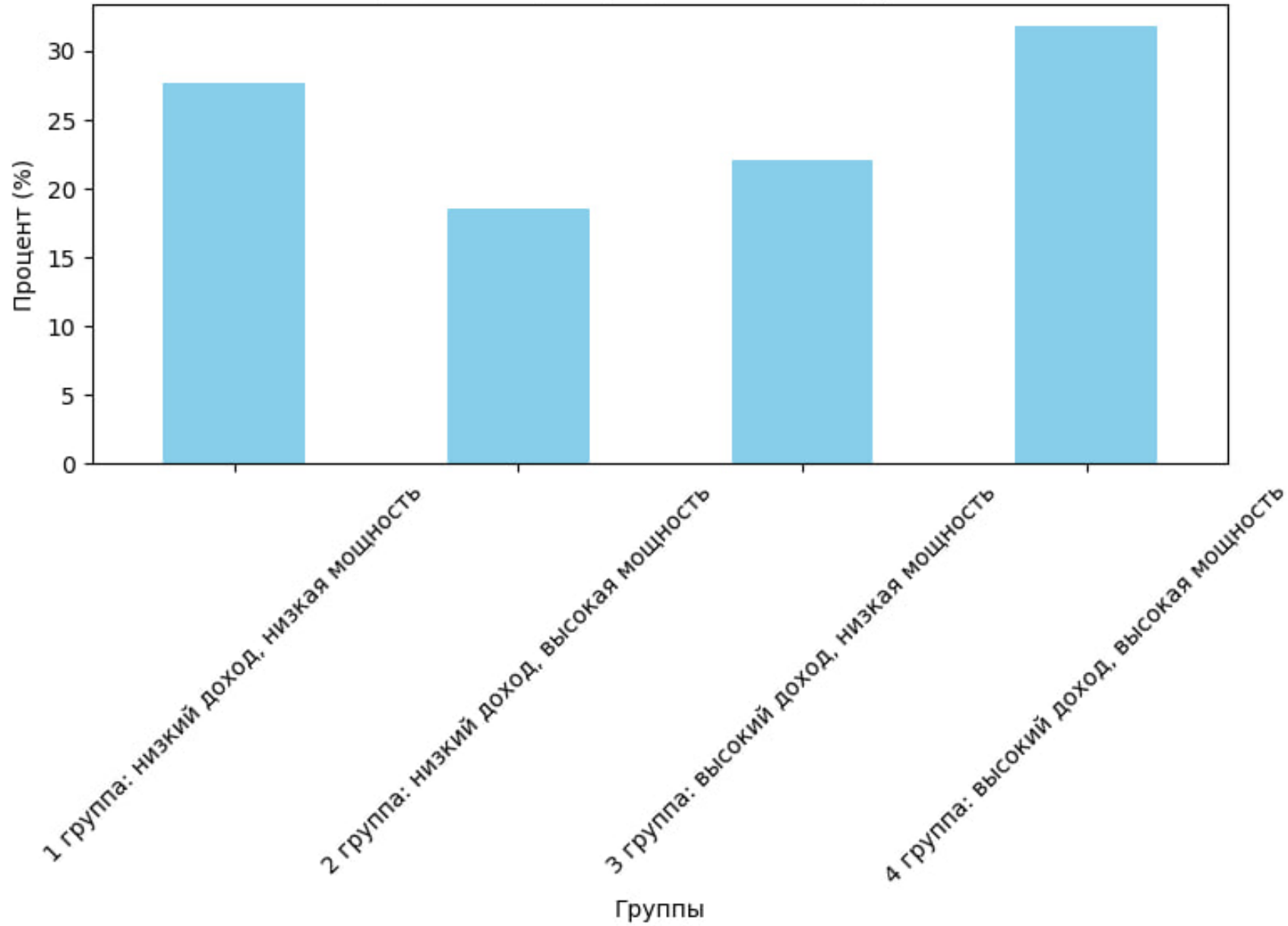
Подвал



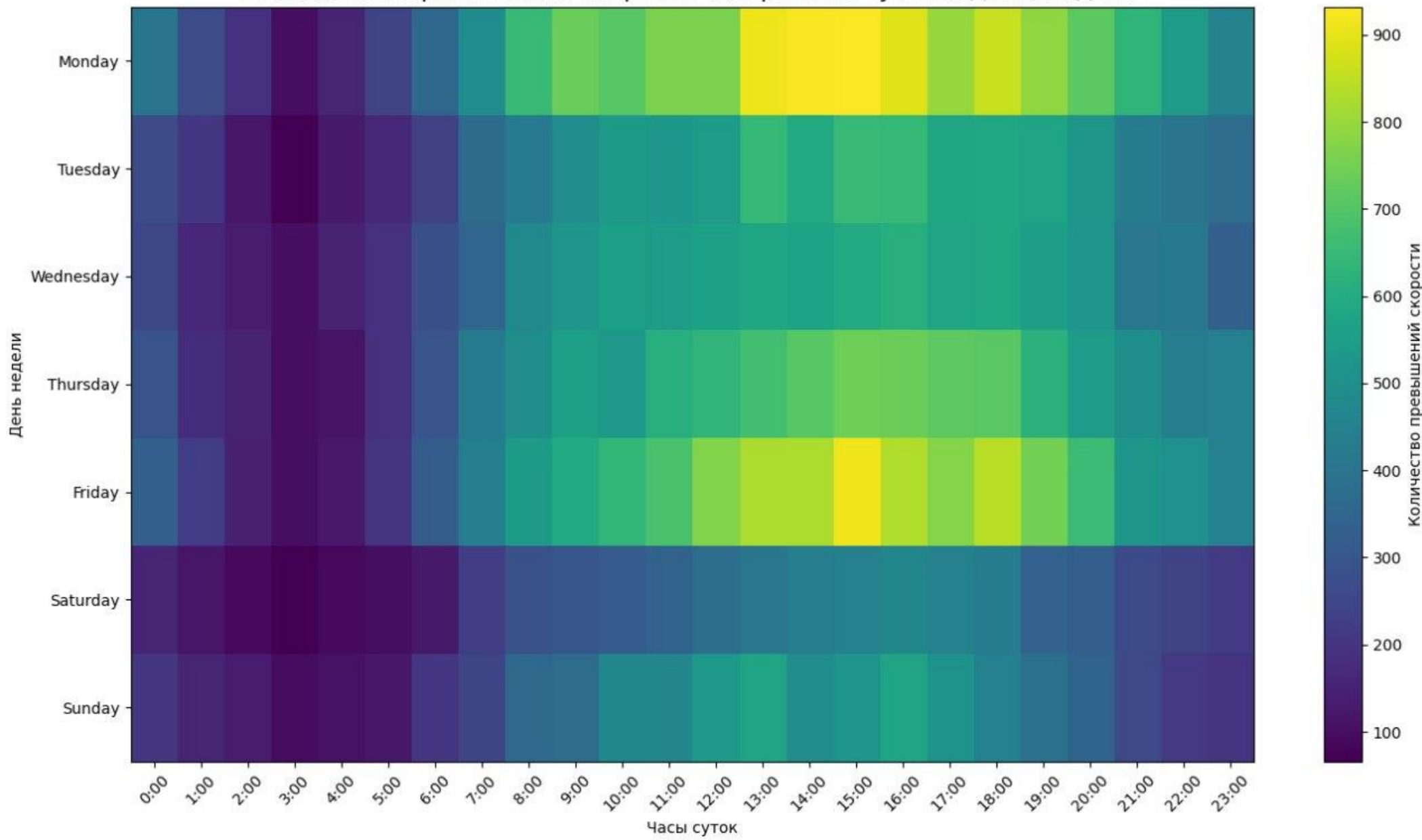
Процентное распределение мужчин, превысивших скорость, по уровню дохода



Процентное распределение М по группам



Количество превышений скорости по времени суток и дням недели



Процентное соотношение правонарушений по дням недели по категориям

