

Взаимосвязаны ли характеристики клиентов и их автомобилей с характеристиками правонарушений?

Выполнили участники 10 группы



Предварительный анализ

Матрица корреляций всех числовых значений

	Год выпуска автомобиля	Возраст человека	Цена машины (усл. ед.)	Количество детей у водителя	Размер месячного дохода водителя
Год выпуска автомобиля	1	0,131	0,532	0,031	-0,006
Возраст человека	0,131	1	0,106	0,138	0,004
Цена машины (усл. ед.)	0,532	0,106	1	0,005	-0,001
Количество детей у водителя	0,031	0,138	0,005	1	0,008
Размер месячного дохода водителя	-0,006	0,004	-0,001	0,008	1

Предварительный анализ

Распределение количества людей от их возраста

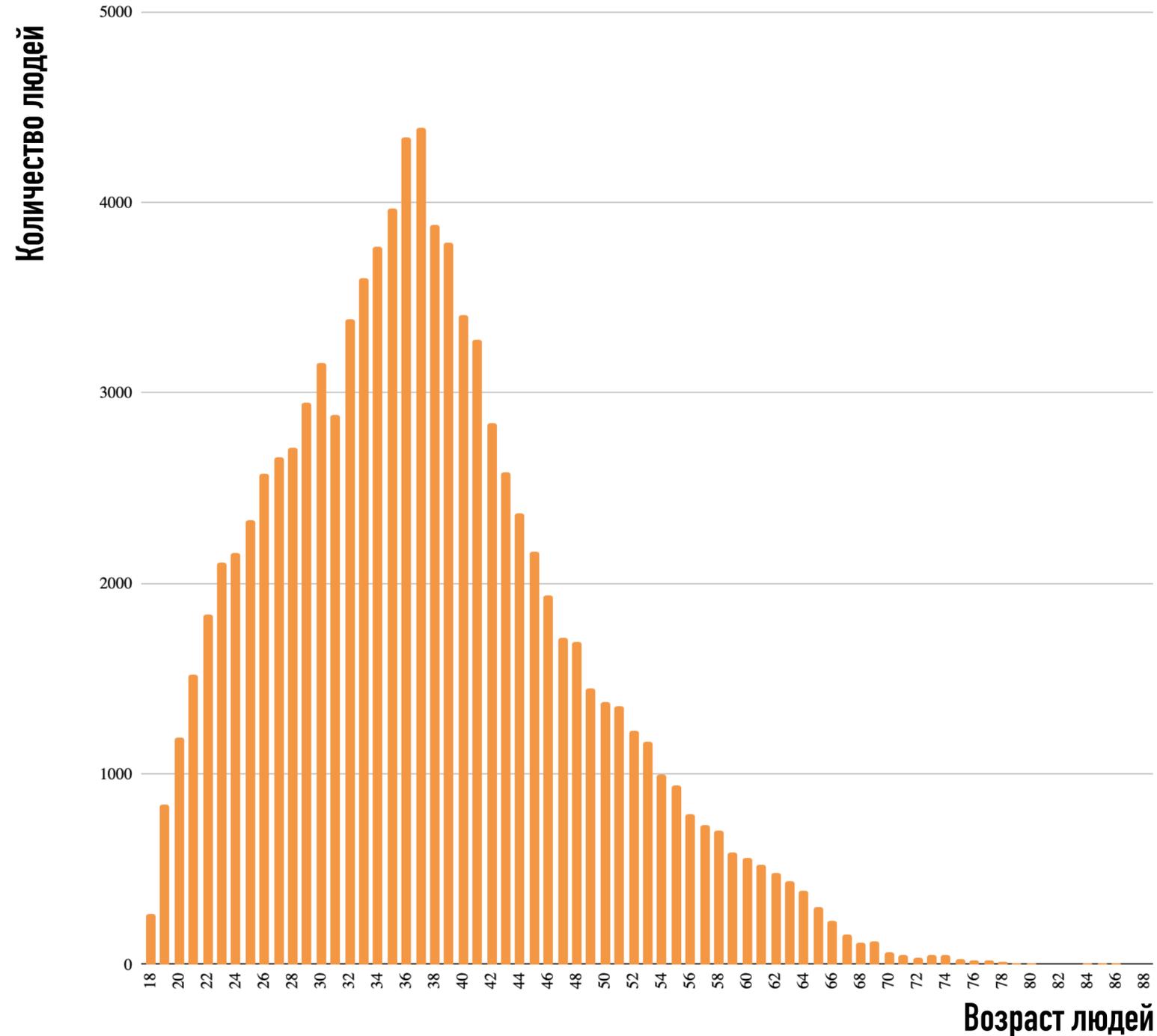


Таблица распределения количества штрафов от количества людей их совершивших

Кол-во штрафов	Кол-во человек
1	61645
от 2 до 5	11042
от 6 до 10	640
от 11 до 15	119
от 16 до 20	56
от 21 до 25	21
>26	20

Убрали выбросы и почистили данные

Введение

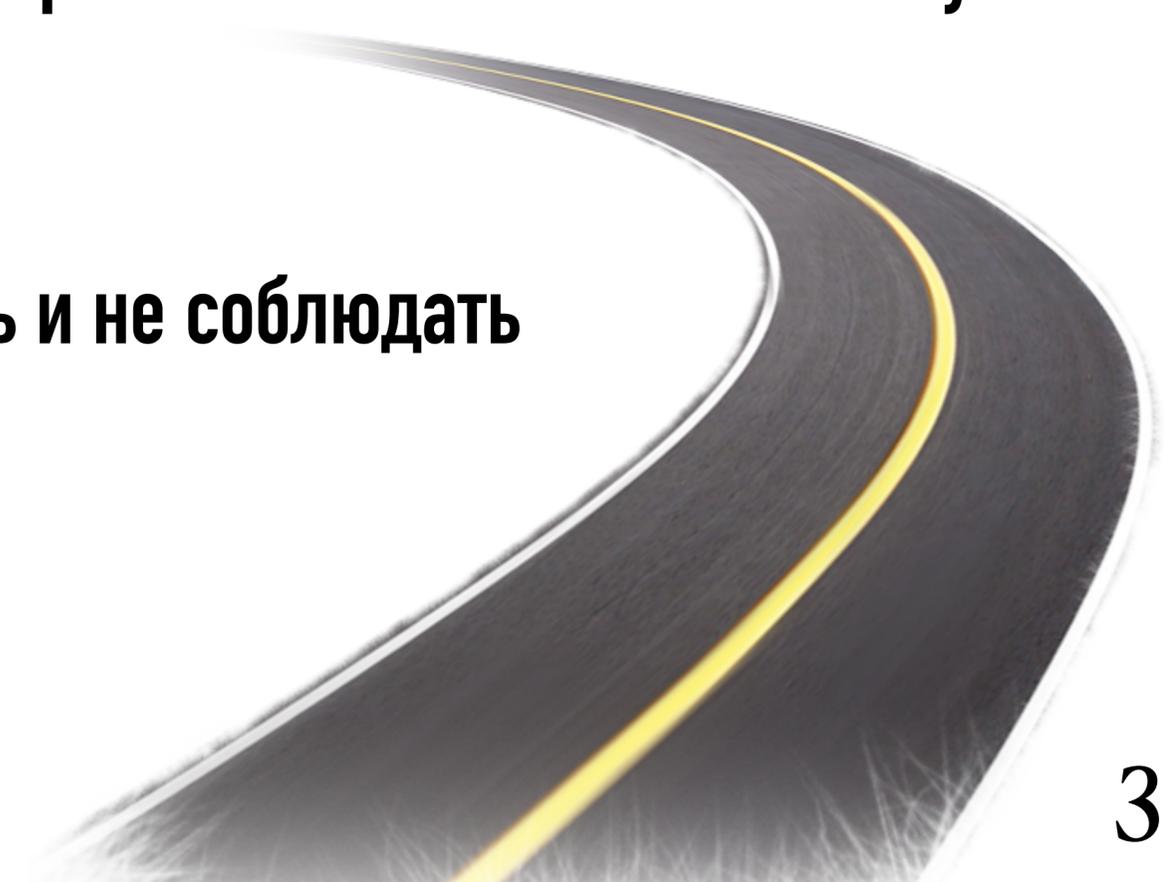
Гипотеза:

Чем младше человек, тем больше нарушений он совершает.

Механизм:

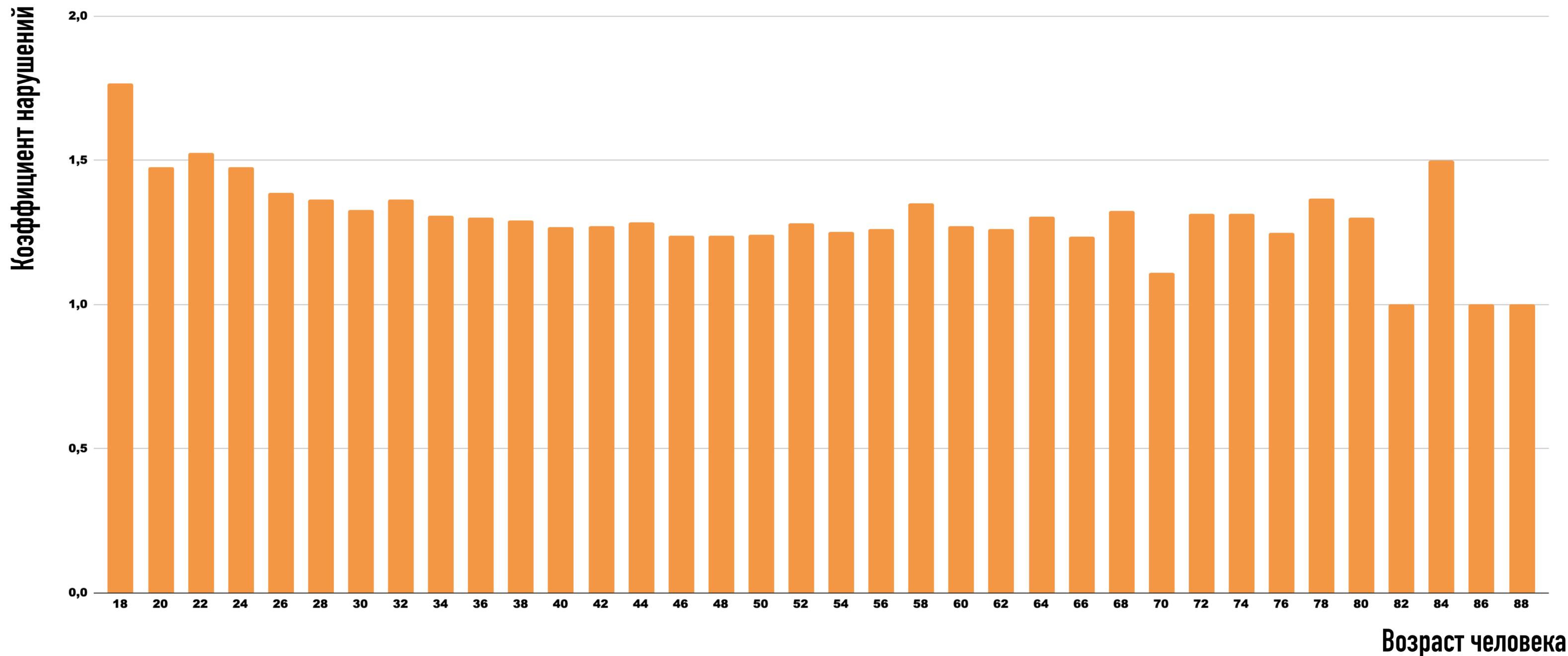
Люди младшего возраста имеют меньший опыт в вождении, совершают больше ошибок и могут что-то пропустить, тем самым получают больше штрафов.

К тому же, есть некоторые молодые люди, которые любят гонять и не соблюдать правила.



Визуализация результата

Коэффициент совершенных человеком нарушений от его возраста



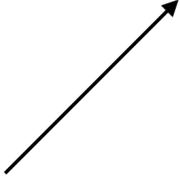
Коэффициент нарушений – количество штрафов для опр. категории / кол-во людей в определённой категории.

Математическая модель

Матрица корреляций для результирующей таблицы

	штрафы	возраст	кол-во людей	Коэффициент нарушений
штрафы	1	-0,695	0,998	0,158
возраст	-0,695	1	-0,668	-0,377
кол-во людей	0,998	-0,668	1	0,127
Коэффициент нарушений	0,158	-0,377	0,127	1

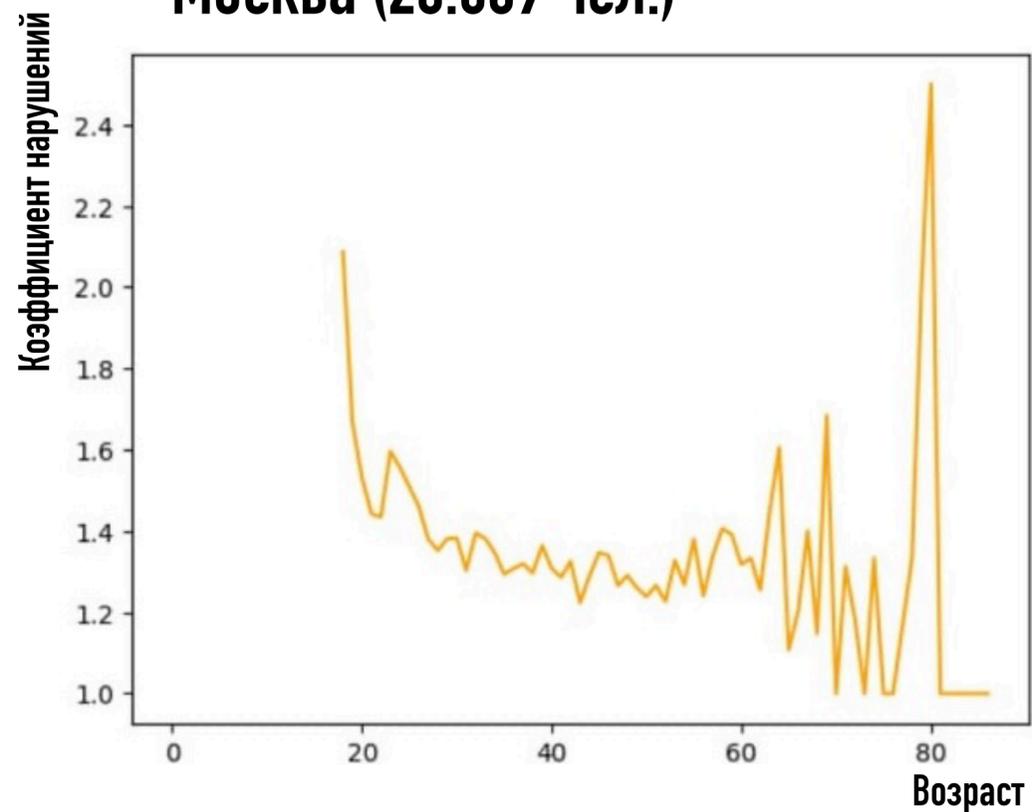
Искомая корреляция



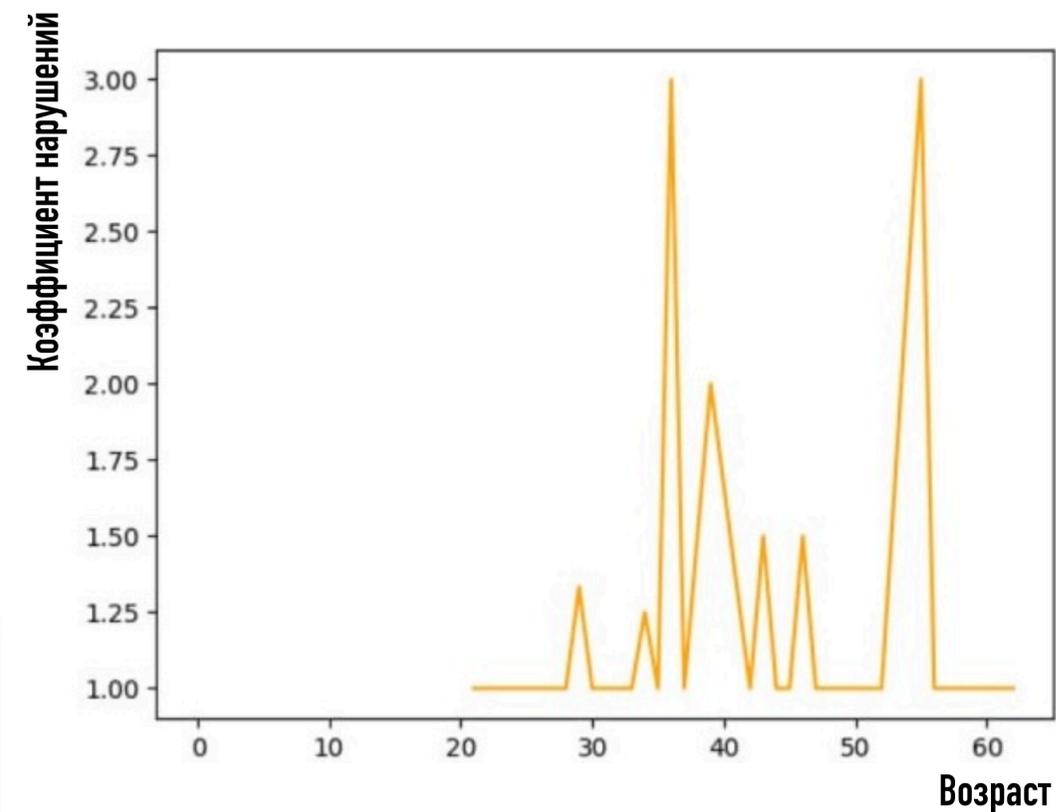
Коэффициент нарушений – количество штрафов для опр. категории / кол-во людей в определённой категории.

Проверка устойчивости результатов

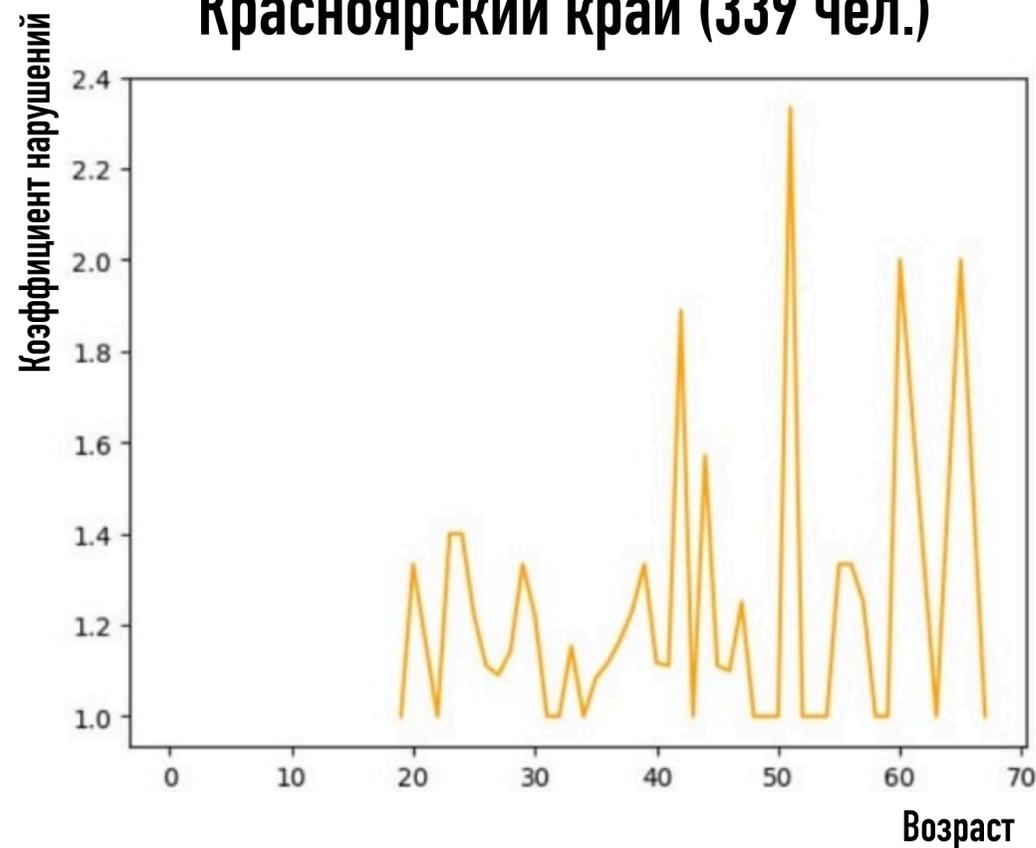
Москва (20.889 чел.)



Приморский край (48 чел.)

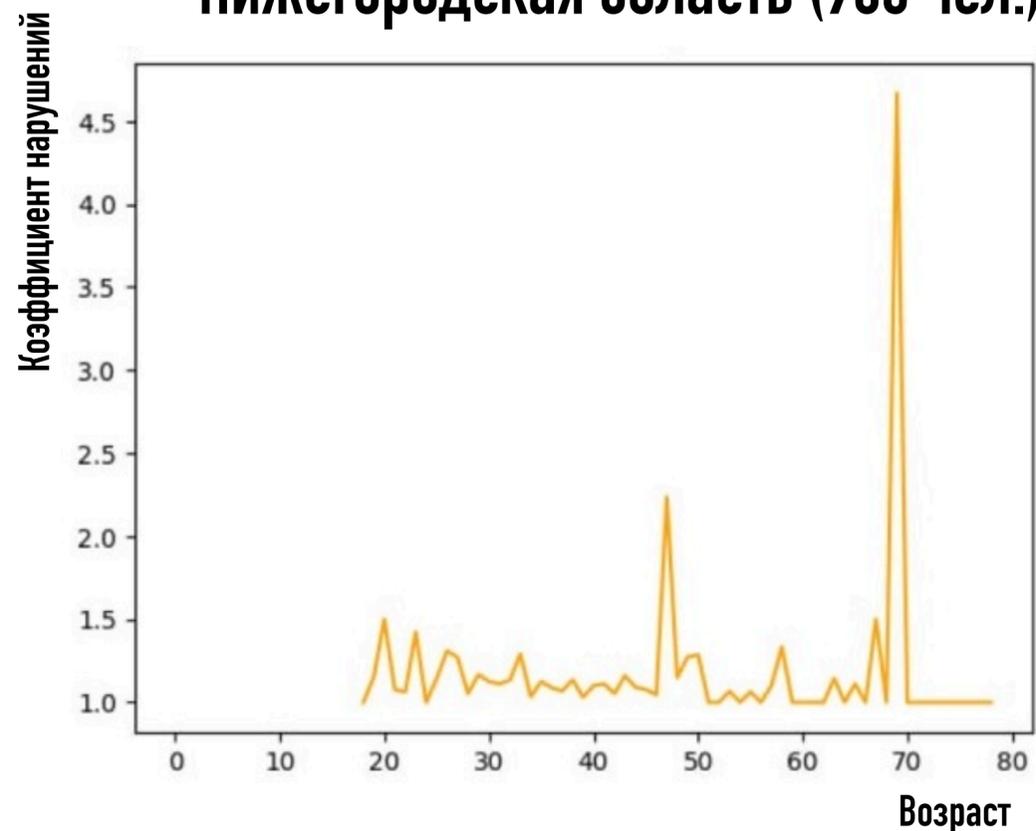


Красноярский край (339 чел.)

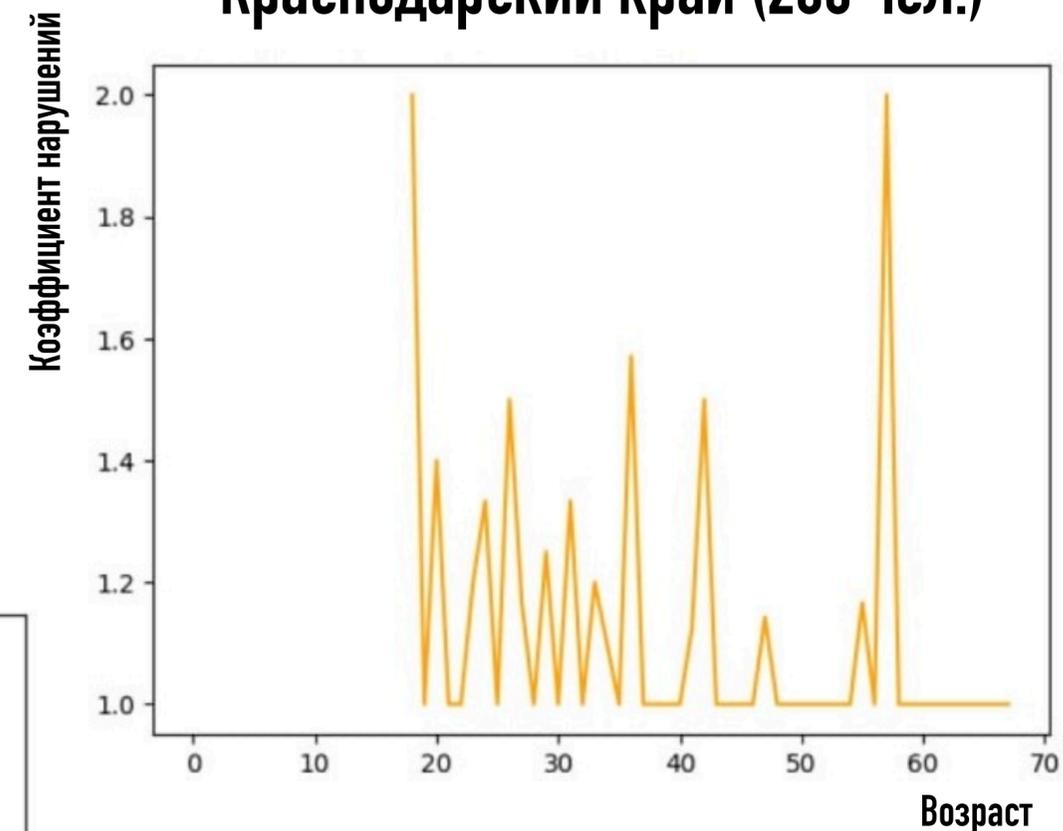


Проверка устойчивости результатов

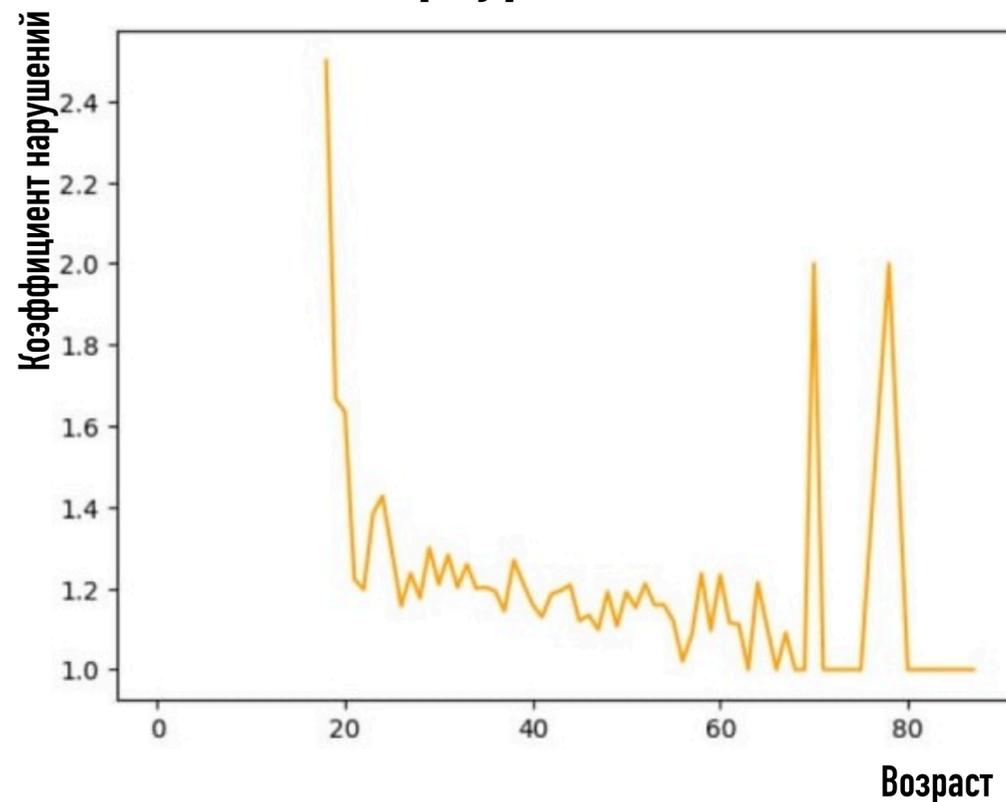
Нижегородская область (963 чел.)



Краснодарский край (206 чел.)



Санкт-Петербург (5.146 чел.)



Выводы

Мы опровергли поставленную гипотезу: возраст водителя не влияет на кол-во полученных штрафов.

Использование: предвзятое отношение к водителям младшего возраста — стереотип, не отражающий реального положения дел. Для более здоровой обстановки в обществе необходимо искоренять этот стереотип.

Ограничения: такую модель будет некорректно применять в других странах, также эта база только клиентов Т-банка и у других источников данных могут отличаться значения.

Благодарим за внимание!



Данные

Описание используемых переменных:

Party_rk – уникальный ключ физического лица количество

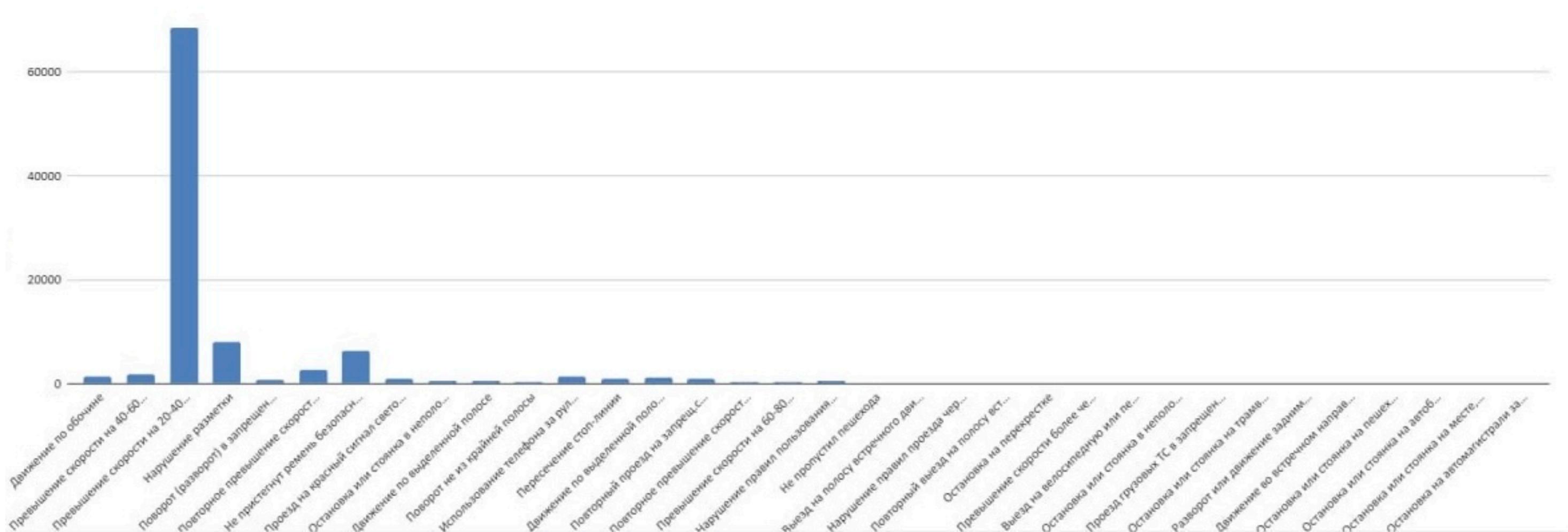
age – Возраст водителя

index – количество штрафов для опр. категории / кол-во людей в определённой категории. Категории – группы по доходу

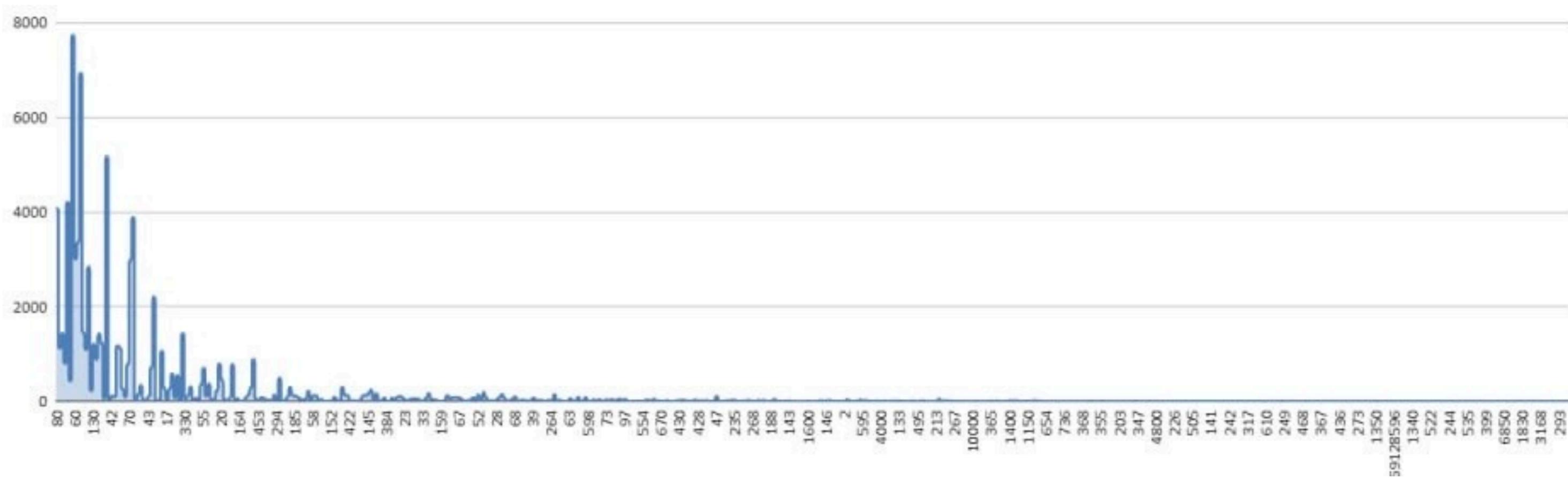
region – регион, в котором совершено правонарушение)

dept–количество штрафов

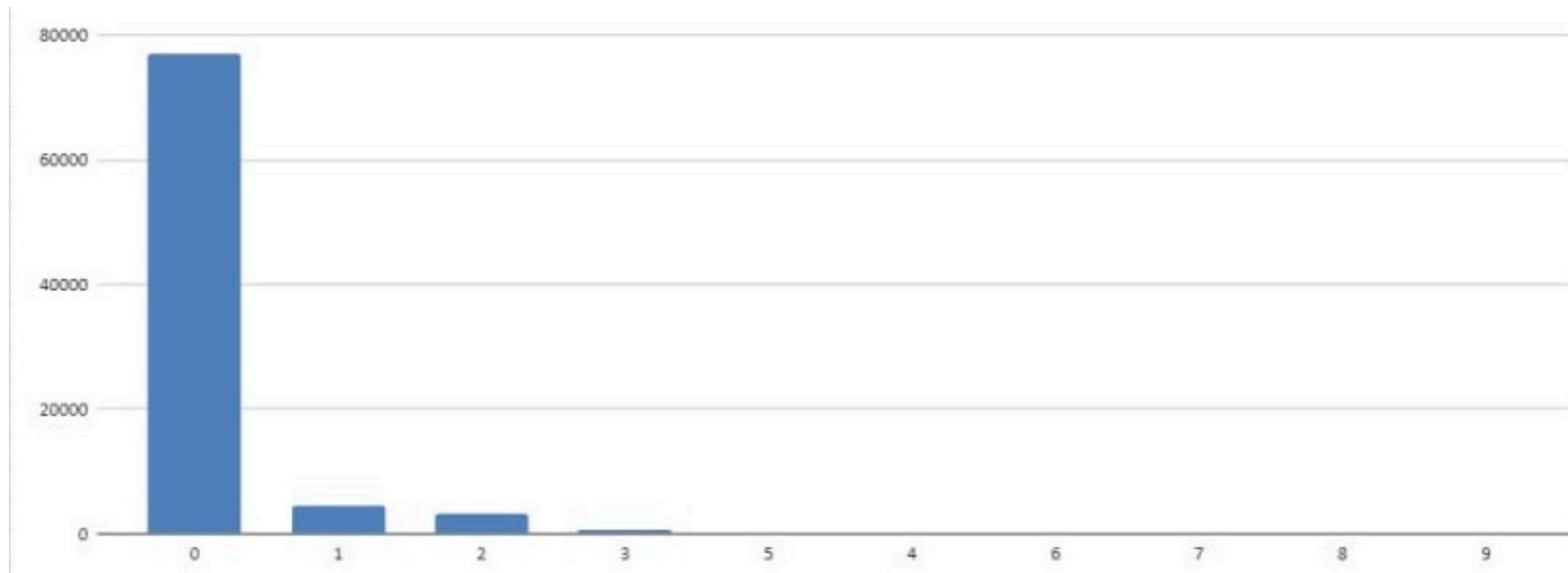
Виды нарушений



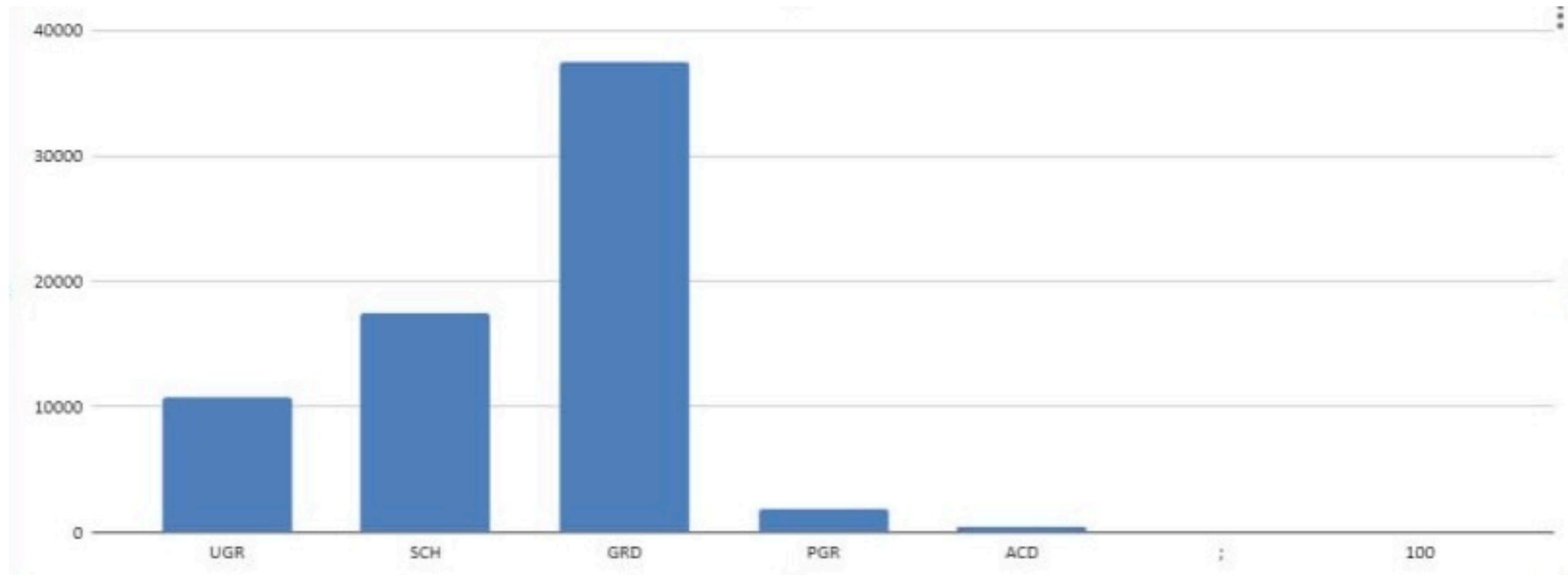
Доход



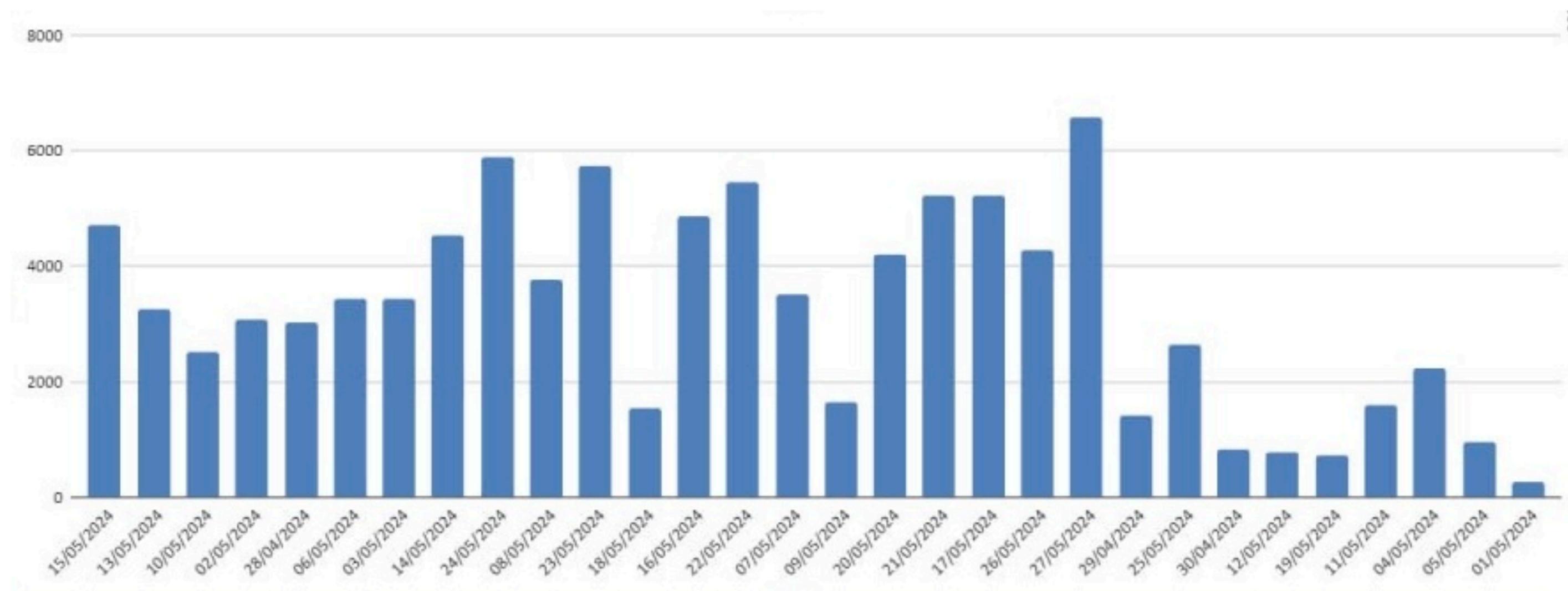
Дети



Образование



Даты



Регионы

