

Взаимосвязь характеристик клиентов и их автомобилей с характеристиками правонарушений.

```
>root@team25
Команда 25: Яшков Дмитрий,
Садков Андрей, Агафонов Иван,
Лебедев Максим

honorable mention: Арбуз Арбуз
```

Гипотеза

Чем выше покупательная способность человека, тем выше количество повторных правонарушений, которые он совершает.

Исследовательский вопрос

3

Взаимосвязаны ли характеристики клиентов и их автомобилей с характеристиками правонарушений (время/дата совершения правонарушения, статья правонарушения) и если да, то как?

Механизм

4

Человек, который может позволить себе больше машин за свою зарплату, будет более часто нарушать правила дорожного движения, так как не боится повредить своё транспортное средство.

Описание переменных

Нам были даны переменные связанные с водителем:

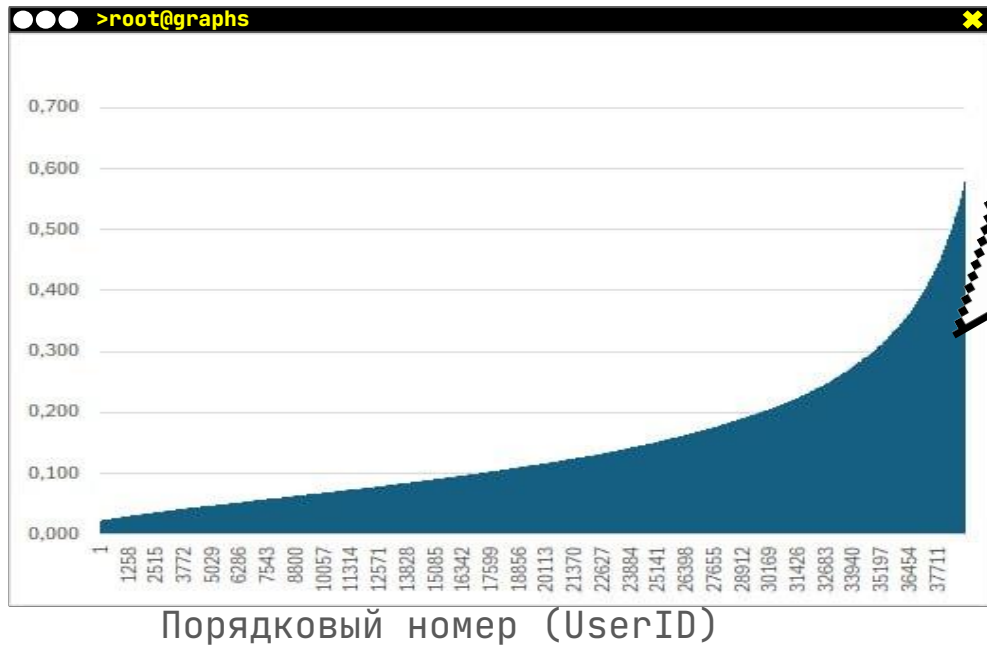
- party_rk - идентификатор физического лица;
- gender_cd - пол водителя;
- age - возраст водителя;
- marital_status_cd - семейный статус водителя;
- children_cnt - количество детей у водителя;
- educational_level_cd - образование водителя;
- person_monthly_income_amt - размер месячного дохода водителя;

И с его машиной:

- engine_type - объём двигателя автомобиля;
- engine_power - мощность двигателя (в лошадиных силах);
- body_type - тип кузова автомобиля;
- auto_mark - марка автомобиля;
- auto_year - год выпуска автомобиля;
- auto_model - модель автомобиля;
- color - цвет автомобиля;
- gear_type - тип коробки переключения передач;
- car_price - стоимость автомобиля;

Разведывательный анализ

Покупательная способность

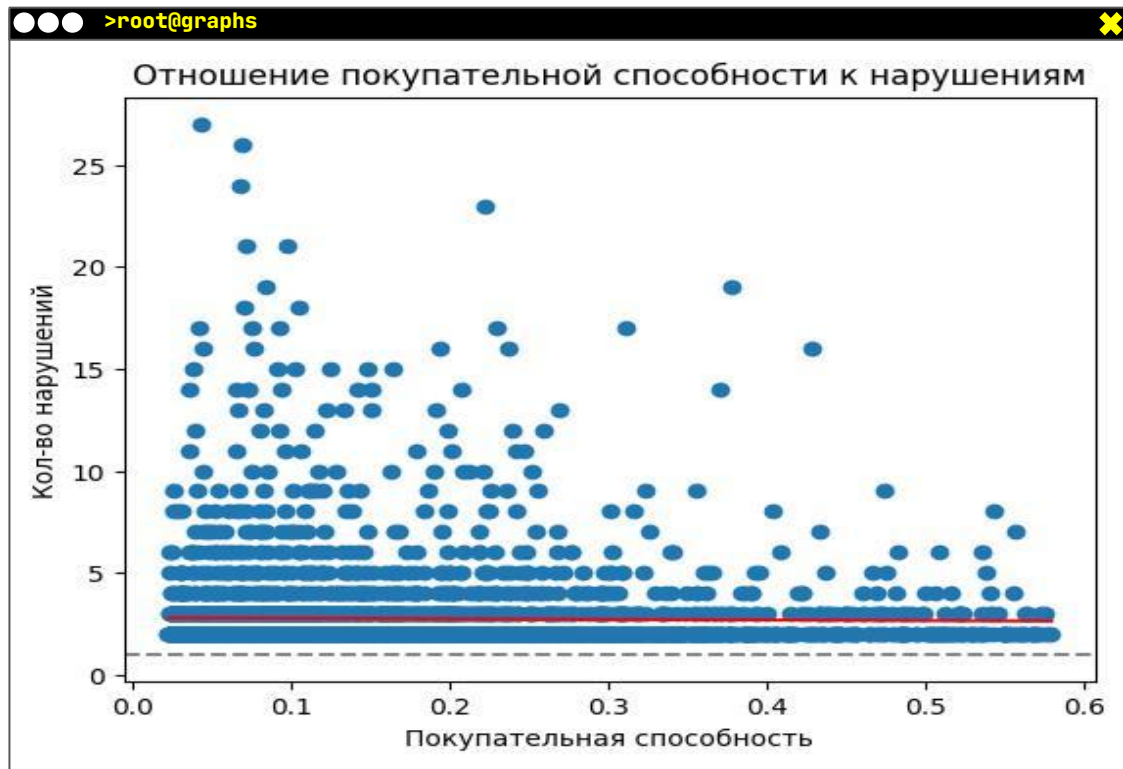


- Данный график мы делали для проверки работы отношения в целом

Разведывательный анализ



Разведывательный анализ



Основные результаты и выводы

Наша гипотеза не получила подтверждения, поскольку по графику мы выявили отсутствие корреляции между переменными, указывая на то, что нет значимой связи между покупательной способностью и количеством нарушений закона.

**Спасибо за внимание
:)**

Практическая польза

Если бы удалось доказать взаимосвязь покупательной способности человека с количеством правонарушений, которые он совершает, это могло бы помочь в выявлении потенциально опасных водителей, к которым стоит применять более строгие меры наказания.

Альтернативные механизмы, которые могли привести

к тем же выводам из модели

Вывод из модели: гипотеза опровергнута, так как математическая модель (линейная регрессия = прямая убывающая линия) показала, что количество ДТП убывает при увеличении покупательной способности, что противоречит формулировке гипотезы.

Альтернативные механизмы: Люди с более высокой покупательной способностью могут владеть большим количеством транспортных средств, чаще садиться за руль или преодолевать большие расстояния, что увеличивает их подверженность ситуациям, в которых они могут совершать нарушения правил дорожного движения.

Другие ограничения применения модели к используемым данным

1. В указанном в таблице Excel временном промежутке есть праздники, во время которых существует различие между поездками у людей с разными покупательными способностями.
2. Качество данных: Недостатки в процессе сбора данных могут привести к неточностям. Например, информация о ДТП может быть неполной или искаженной.

Исходная таблица содержала значительное количество выбросов и дубликатов, которые необходимо было удалить для получения точных и корректных результатов исследования.

Общее количество строк в исходной таблице: 97 307.

Первым шагом было удаление пустых значений с помощью встроенного фильтра Excel. Подробности этого процесса можно найти на листе "ИСХОДНИК", начиная со столбца "X".

Всего было удалено 38 509 строк, содержащих хотя бы одно пустое значение.

Впоследствии я удалил дубликаты, выделив всю таблицу и воспользовавшись функцией Excel "Удалить дубликаты".

В общей сложности получилось: 33.

Таким образом, количество значений, с которыми мы работаем: $97307 - 38509 - 33 = 58765$.

Далее по порядку:

party_rk

Выбросов нет.

region

Была проведена проверка того, что все регионы, указанные в таблице, действительно существуют.

Выбросов нет.

offenceshortstatement

Была проведена проверка того, что все правонарушения, указанные в таблице, действительно существуют.

Выбросов нет.

offencedate

Была проведена проверка на корректность данных о дате. (Например, соблюдается ли условие, что в дате $x:y:z$ $0 < X < 32$ (для мая) $0 < X < 31$ (для апреля) $3 < y < 6$ $z = 2024$). Выбросов нет

offencetime

Проверка на корректность данных о времени (Например, соблюдается ли условие что в ячейке с данными, записанными в виде $a:b:c$ $0 \leq a < 24$ $0 \leq b < 60$ $0 \leq c < 60$) не выявила выбросов.

engine_type

Были удалены текстовые значения, значения, явно превышающие 6,6, а также значения 0,1 и 0,2, поскольку у автомобилей, рядом с которыми были указаны эти значения, на самом деле другие значения.

engine_power

Были удалены текстовые значения, 0, а также числа, не соответствующие реальной информации о количестве лошадиных сил у автомобиля.

body_type Выполнена проверка на наличие несуществующих типов кузова автомобилей. Выбросов нет

auto_mark Выполнена проверка на наличие несуществующих марок автомобилей. Выбросов нет

auto_year Значения от 1981 до 2024. Выбросов нет

auto_model Выполнена проверка на наличие несуществующих моделей автомобилей, а также на их соответствие маркам. Выбросов нет

В подвал

color
Были удалены небуквенные значения цветов, а также повторяющиеся цвета и цвета, которые невозможно определить (Например, "БЕЛЫЙ СИНИЙ ЧЕРНЫЙ ЗЕЛЕНый").

gear_type Выбросов нет

car_price Выбросов нет

gender Всего два. Выбросов нет

age
Было удалены некорректные значения возраста (<18).

marital_status_cd Выбросов нет

children_cnt Отрицательных значений не выявлено)))
Выбросов нет

education_level_cd Выбросов нет

person_monthly_income_amt Не выявлены значения, явно выделяющиеся на фоне остальных.
Выбросов нет

day_of_week Выбросов нет

public_holiday Выбросов нет

В ОБЩЕЙ СЛОЖНОСТИ ВЫБРОСОВ ОКАЗАЛОСЬ: 1349
ИТОГО СТРОК В ТАБЛИЦЕ: 57416