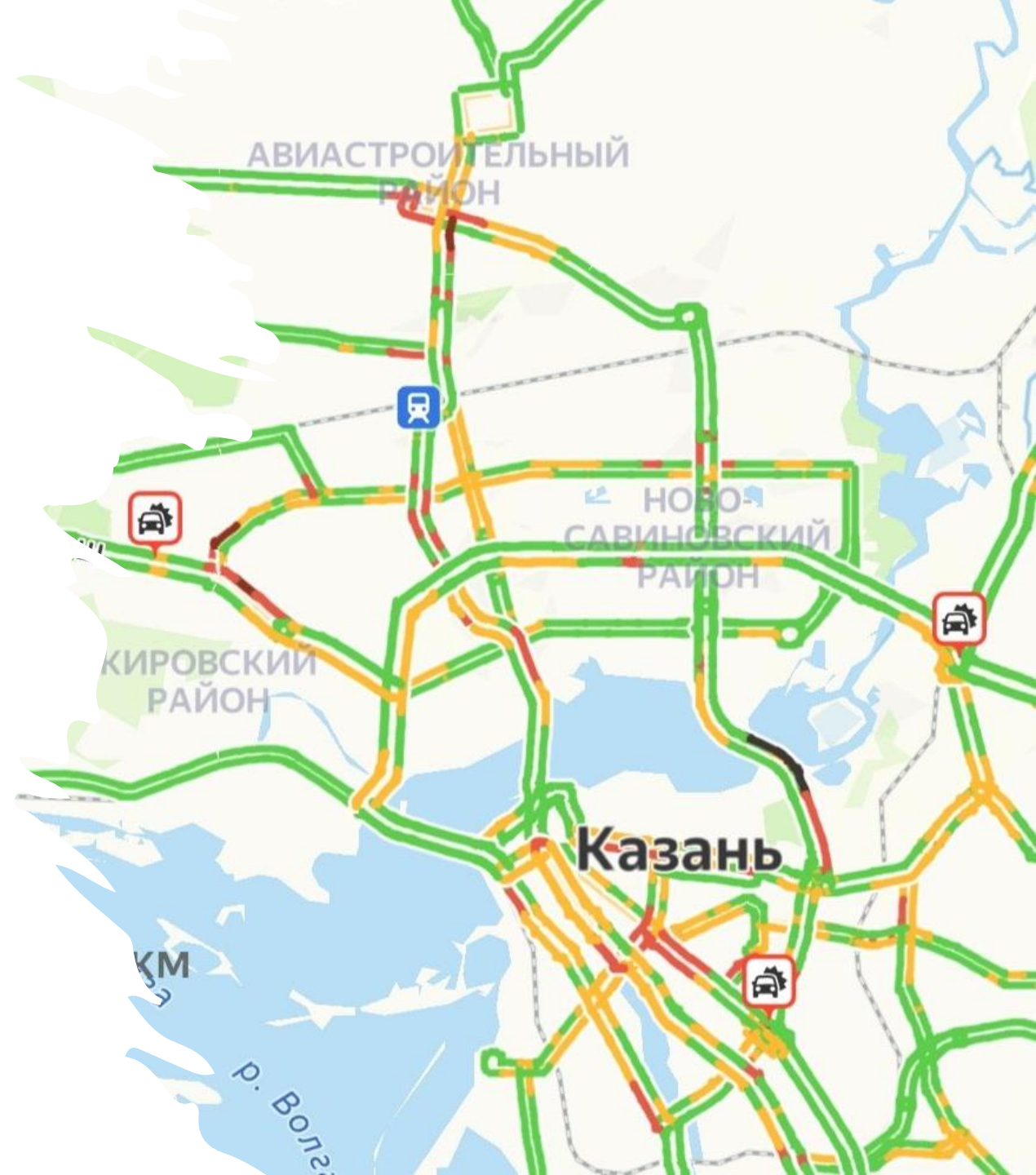


Аналитика ДТП за период с 2015 по 2024 годы.

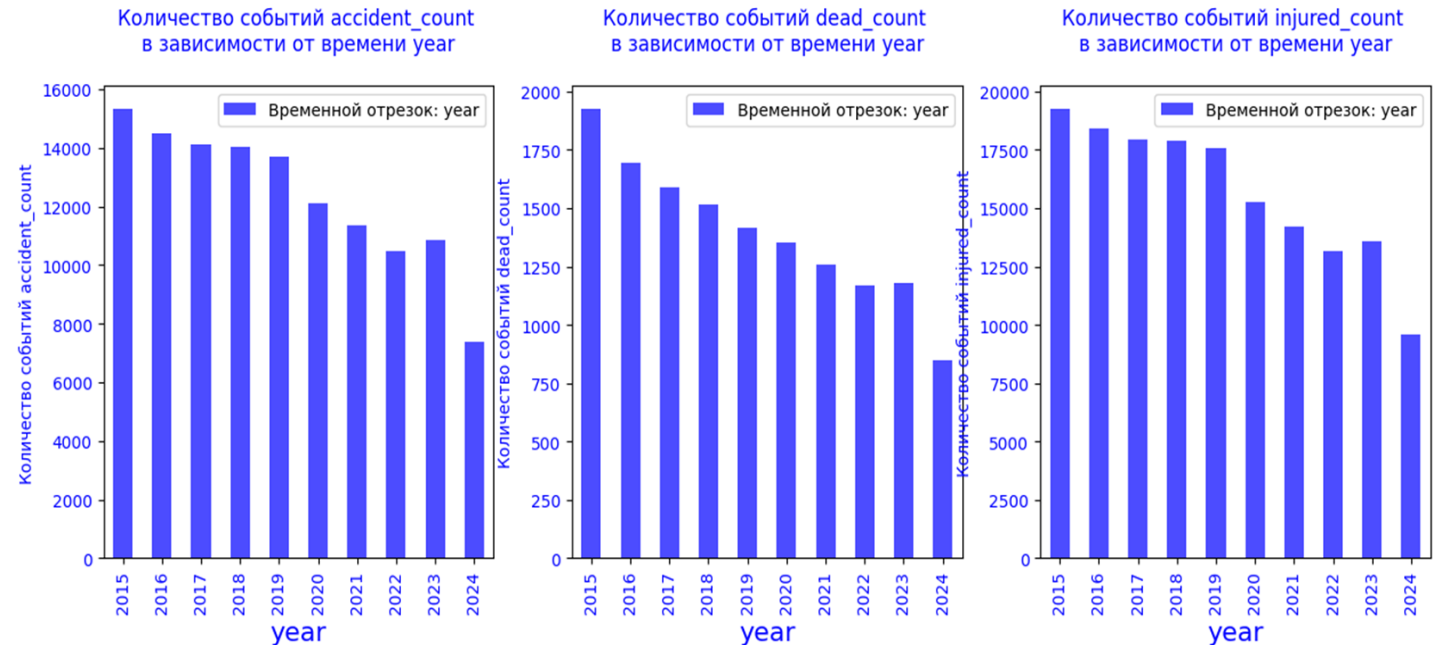
- Исследование данных о дорожно – транспортных происшествиях .
- Формулирование и проверка гипотез.
- Построение дашборда.

Александр Черезов
08 апреля 2024 года.
Казань.



В исследовании. В Части 1 мы провели:

- Загрузку данных из предоставленных файлов
 - Чтение и обработку данных в столбцах в форматах json и в списках
 - Формирование новых датафреймов, содержащих:
 - данные о транспортных средствах, участвовавших в ДТП
 - о водителях, пассажирах и пешеходах, участвовавших в ДТП
 - Обработку данных, изменение типов данных для уменьшения размера занимаемой оперативной памяти
-
- Формирование трех сводных датасетов, объединяющих в себе данные по всем регионам и содержащие в себе:
 - данные о ДТП
 - данные о транспортных средствах, участвовавших в ДТП
 - о водителях, пассажирах и пешеходах, участвовавших в ДТП
 - Выгрузку на диск сводных датасетов в виде файлов формата .pickle



В исследовании. В части 2 провели:

- исследовательский анализ данных
- формирование и проверку гипотез:
 - В большинстве случаев виноваты пьяные водители, нужно усилить контроль
 - Безопасность на дорогах растет для каждого из участников
 - В личном автомобиле ехать безопаснее чем в автобусе
 - Если пешеходы не будут нарушать, то не будут гибнуть
 - Низкая скорость не влияет на тяжесть ДТП
- построение дашборда средствами DataLens

При работе с гипотезами использовали корреляции Спирмена и Пирсона, провели z- и t- тесты



Столкновение и наезд являются основными причинами ДТП.

При примерном равенстве количества таких ДТП количество участников и количество раненых при столкновении больше, чем при наезде - а значит, больше и доли и столкновения более травмоопасны.

А вот количество погибших - и доли - примерно одинаковы для этих ДТП.

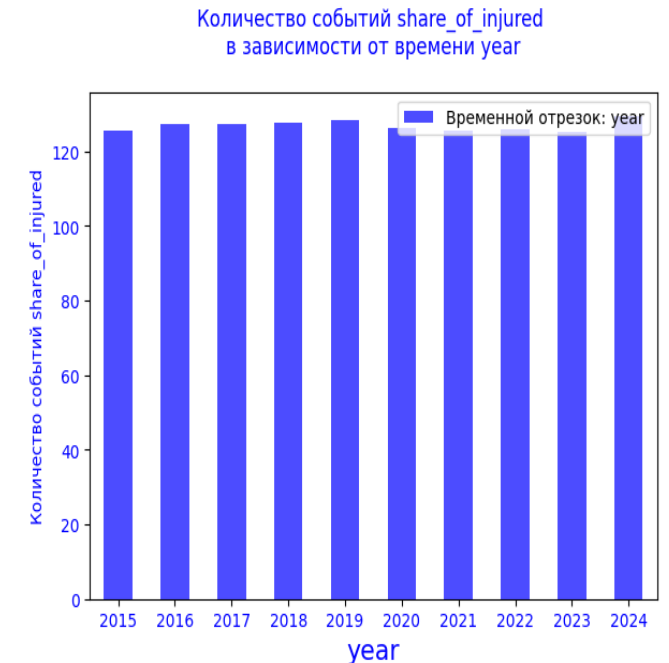
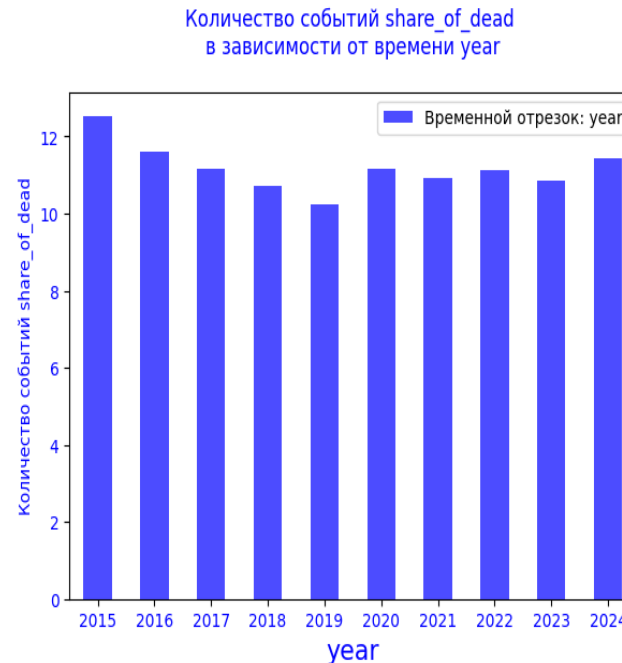
Гипотеза: «Безопасность на дорогах растет для каждого из участников»

Произвели сравнение ДТП за три последних года с ДТП за три предшествующих года.

Тесты отвергают нулевую гипотезу о том, что Среднее число погибших в ДТП в 2021-2023 годах не больше , чем среднее число погибших в ДТП в 2018-2020 годах

То есть люди в 2021-2023 годах погибают в ДТП НЕ РЕЖЕ, чем в 2018-2020 годах

- Проводя z- и t- тесты для пострадавших в ДТП мы обнаружили различные результаты.
- Несмотря на то, что правосторонний t-тест Стьюдента показывает нам, что среднее пострадавших в ДТП в 2021-2023 годах не больше (МЕНЬШЕ), чем в 2018-2020 годах, мы должны отдать предпочтение параметрическому z-тесту.
- Проведенные тесты отвергают нулевую гипотезу о том, что Среднее число получивших повреждения и ранения в ДТП в 2021-2023 годах не больше , чем среднее число получивших повреждения и ранения в ДТП в 2018-2020 годах **То есть люди в 2021-2023 годах получают травмы НЕ РЕЖЕ, чем в 2018-2020 годах**



Гипотеза: «В личном автомобиле ехать безопаснее чем в автобусе»

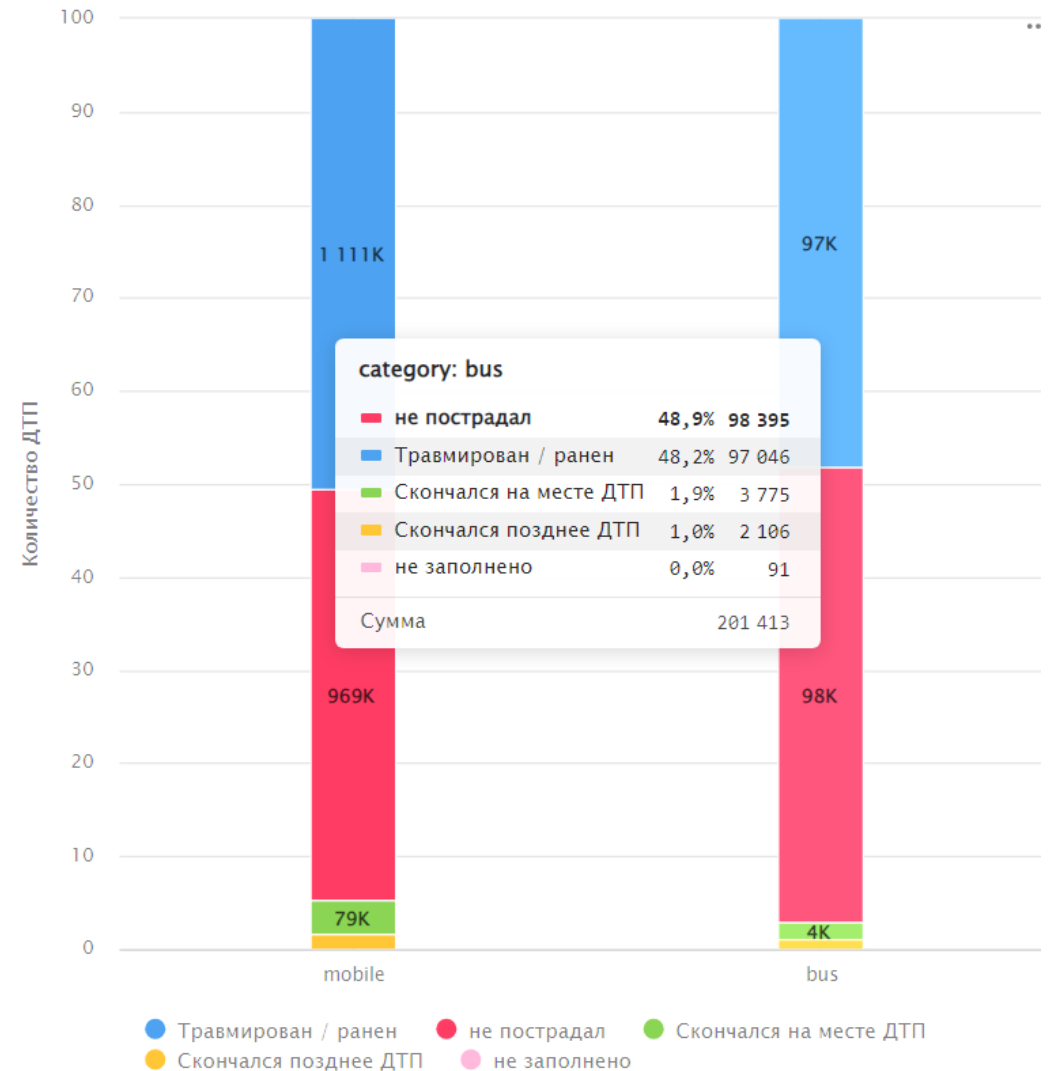
Процент ДТП с автобусом с погибшими почти вдвое меньше, чем аналогичный показатель с автомобилями (где транспортные средства - только автомобили)

Проведенные тесты отвергают нулевую гипотезу о том, что Среднее погибших в ДТП с автомобилями не больше, чем среднее погибших в ДТП с автобусами. **Мы не можем подтвердить, что в автомобиле ехать безопаснее, чем в автобусе.**

Вновь мы видим расхождение в результатах z и t тестов. Отдадим приоритет параметрическому z-тесту.

Процент ДТП с автобусом с пострадавшими лишь незначительно отличается (и превышает) от аналогичного показателя с автомобилями (где транспортные средства - только автомобили)

Проведенные тесты отвергают нулевую гипотезу о том, что Среднее пострадавших в ДТП с автомобилями не больше, чем среднее пострадавших в ДТП с автобусами. **Мы не можем подтвердить, что в автомобиле ехать безопаснее, чем в автобусе.**

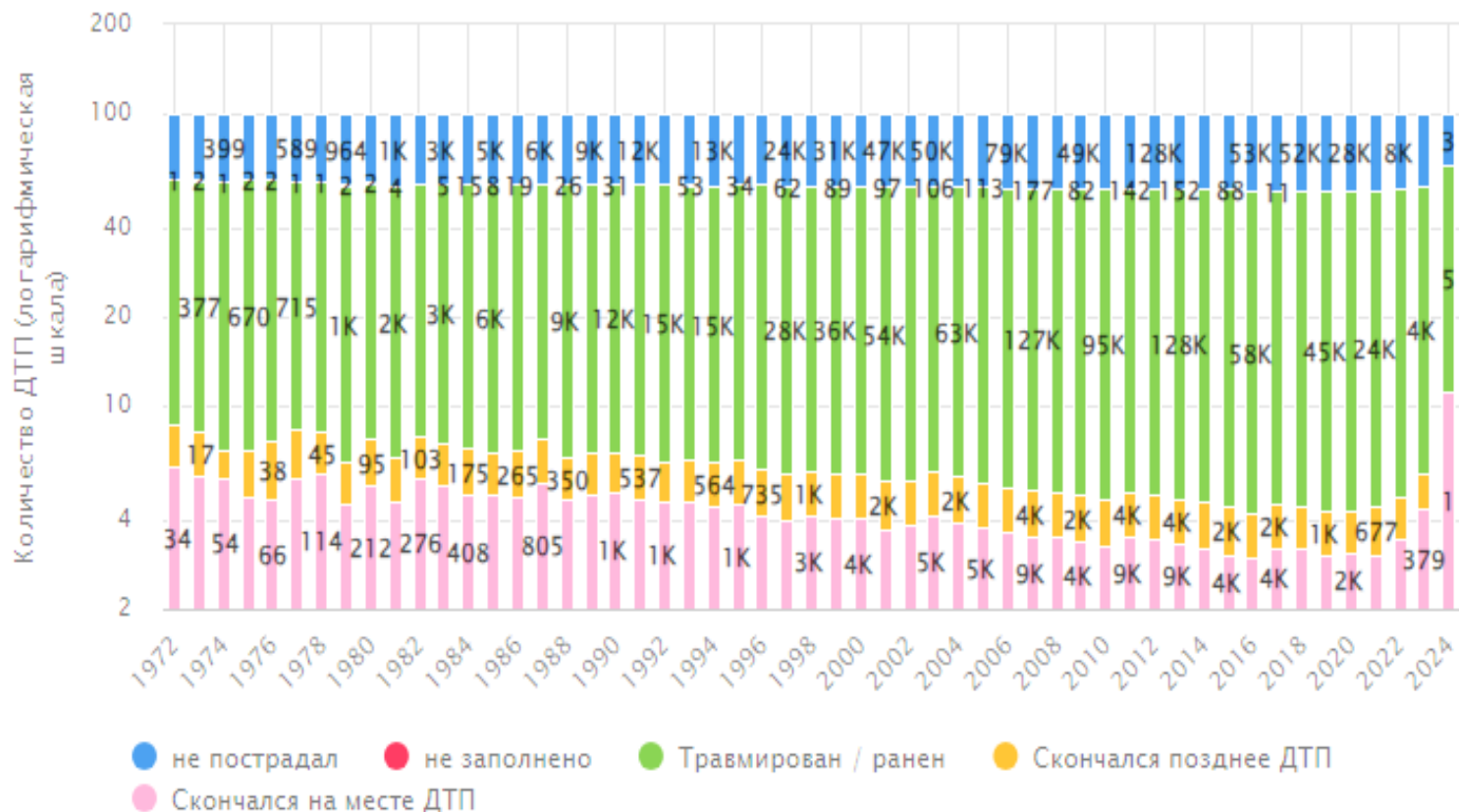


Визуализация: «Год выпуска транспортного средства и вред здоровью в ДТП»

*Использована логарифмическая шкала для удобства визуализации малых значений. Сиреневого цвета и желтая зоны - количество (доли для каждого года) скончавшихся во время и после ДТП - увеличиваются в обратном хронологическом порядке.

А зона синего цвета - количество (доли) не пострадавших в ДТП - наоборот - уменьшается в обратном

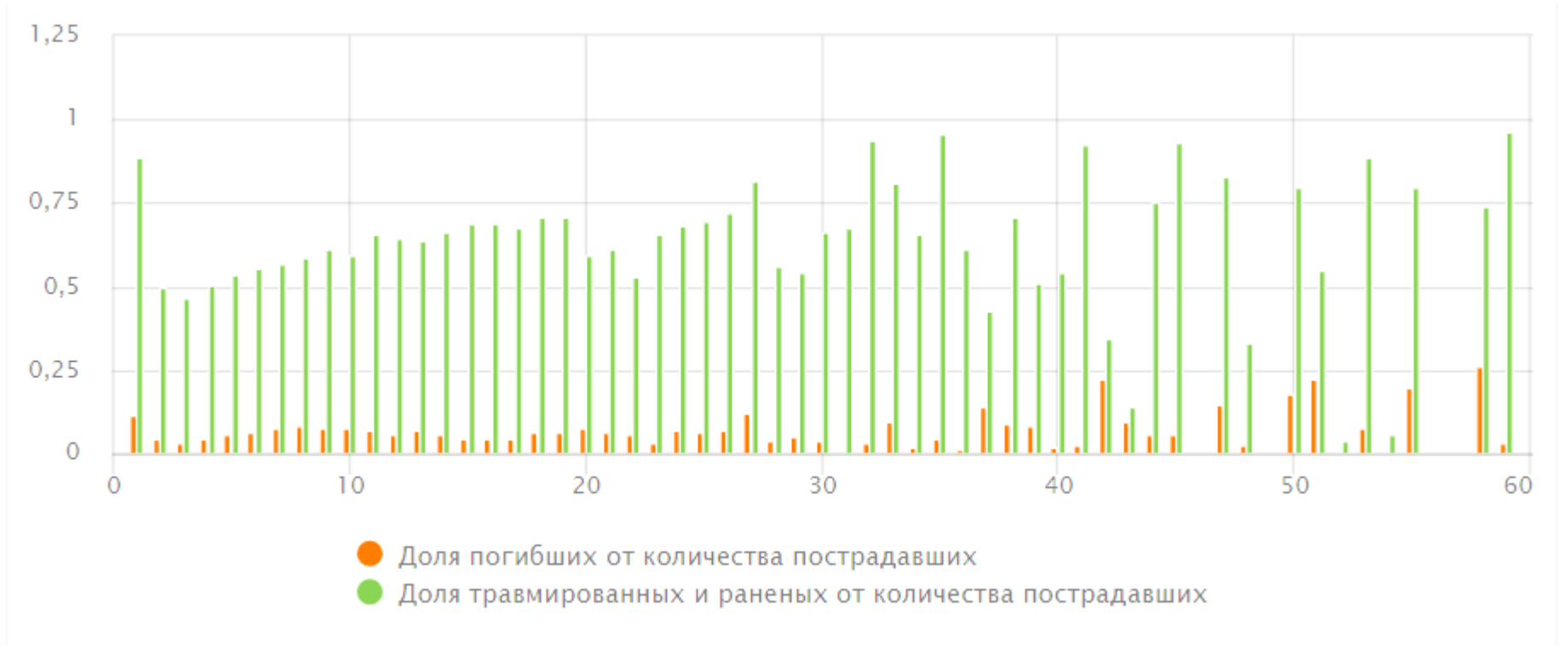
хронологическом порядке. В совокупности это говорит о том, что более старые ТС ведут к большей аварийности по сравнению с новыми.



Визуализация: «Доля погибших и пострадавших от количества участников ДТП»

Обращает на себя внимание более высокая доля раненых и травмированных для тех ДТП, где количество участников составляет более 20 человек, а также 1 человек.

Это наблюдение может быть предметом отдельного исследования.

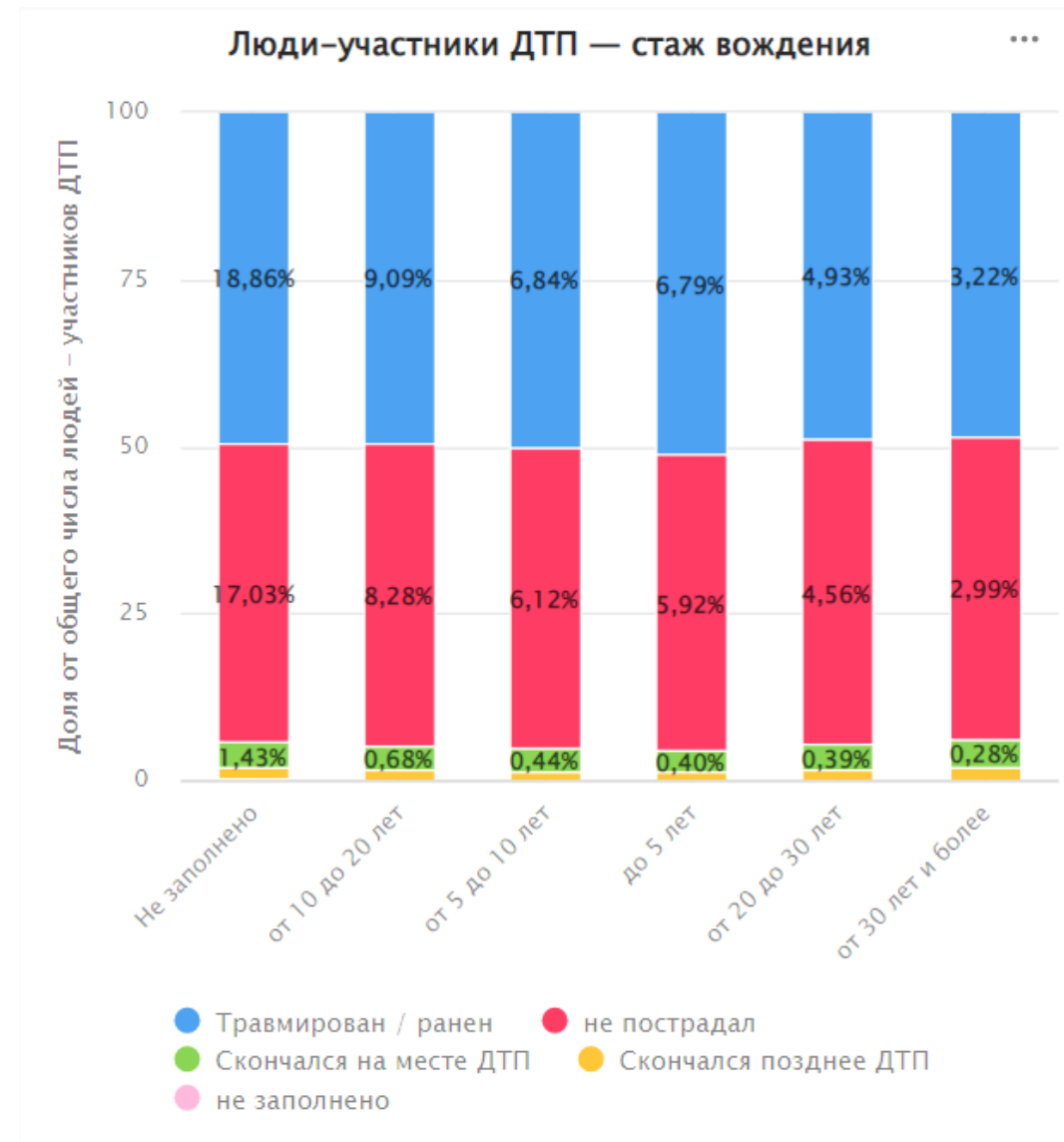


Визуализация: «Зависимость тяжести ДТП от стажа вождения»

Вопреки распространённому мнению водители с небольшим стажем – до 5 лет – не самые частые участники ДТП. Лидирует здесь группа со стажем от 10 до 20 лет. Возможно, причина здесь – в активности этой группы населения.

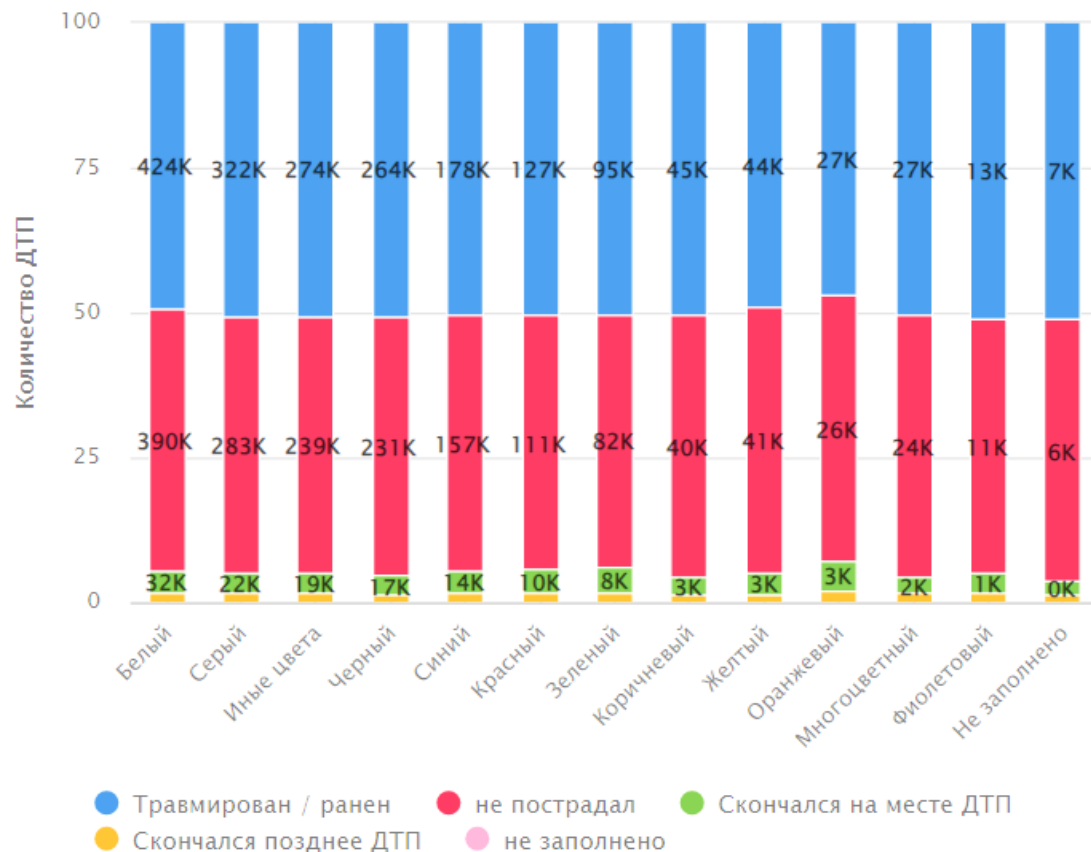
А две последние позиции довольно ожидаемо занимают водители с наибольшим стажем вождения – сказывается и опыт и меньшая активность, меньше частота поездок.

Диаграмма позволяет увидеть, что доля травмированных в двух последних колонках несколько меньше, чем в других категориях, а вот доля смертельных случаев – наоборот – выше.



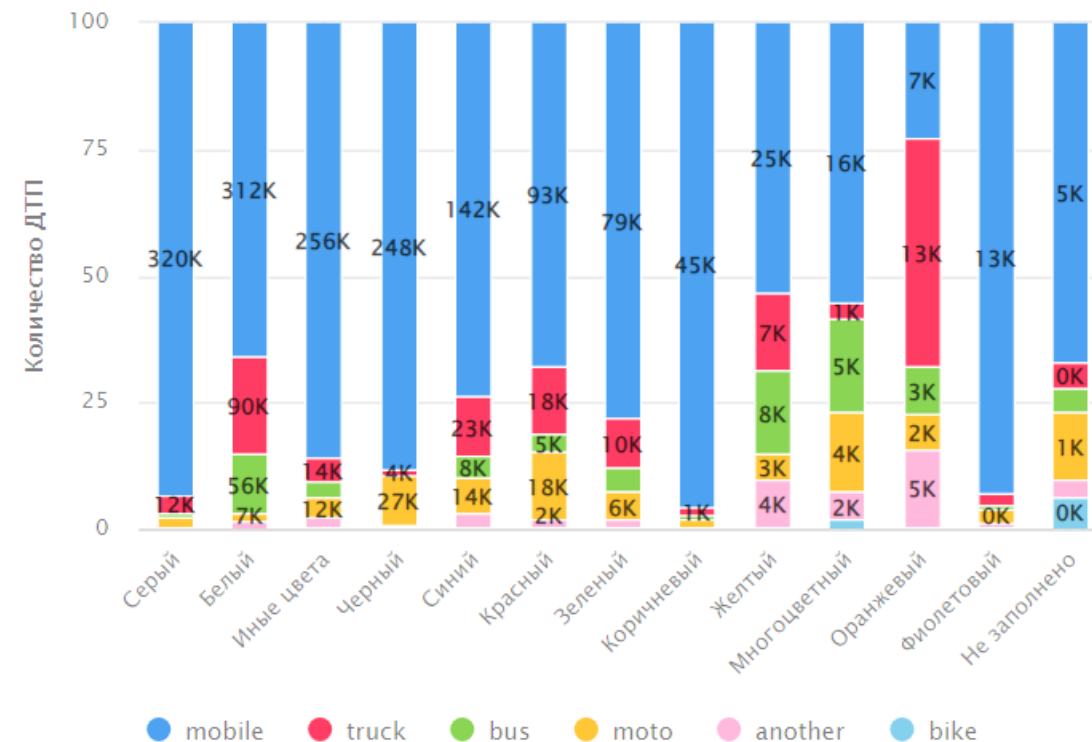
Визуализация: «Последовательный анализ средствами DataLens»

Цвет ТС и вред здоровью



С некоторым удивлением можно увидеть, что ТС ярких цветов - оранжевого и желтого, а также зеленого - имеют большую долю смертельных случаев. Интересно увидеть - к каким категориям относятся эти ТС?

Цвет ТС и категория ТС



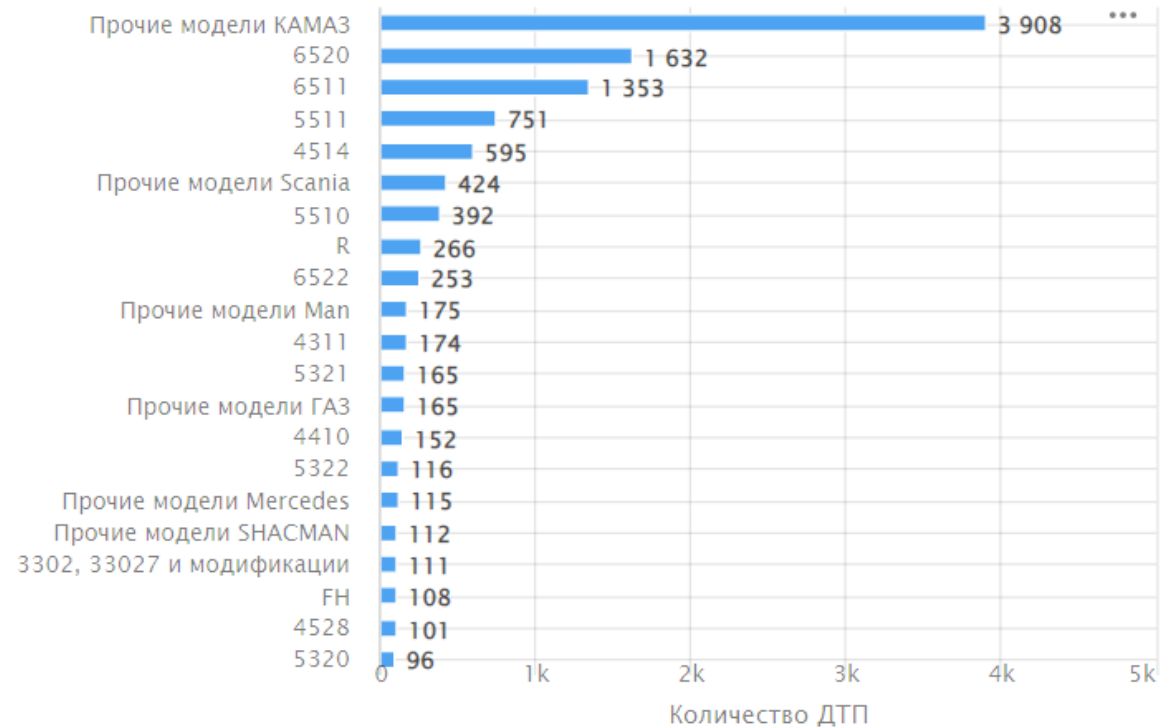
Видим, что в оранжевом цвете, как ни в каких иных - очень большая доля грузовых автомобилей.

А автобусы бывают желтого цвета чаще, чем любого другого.

Но для зеленого цвета не видно закономерностей или аномалий.

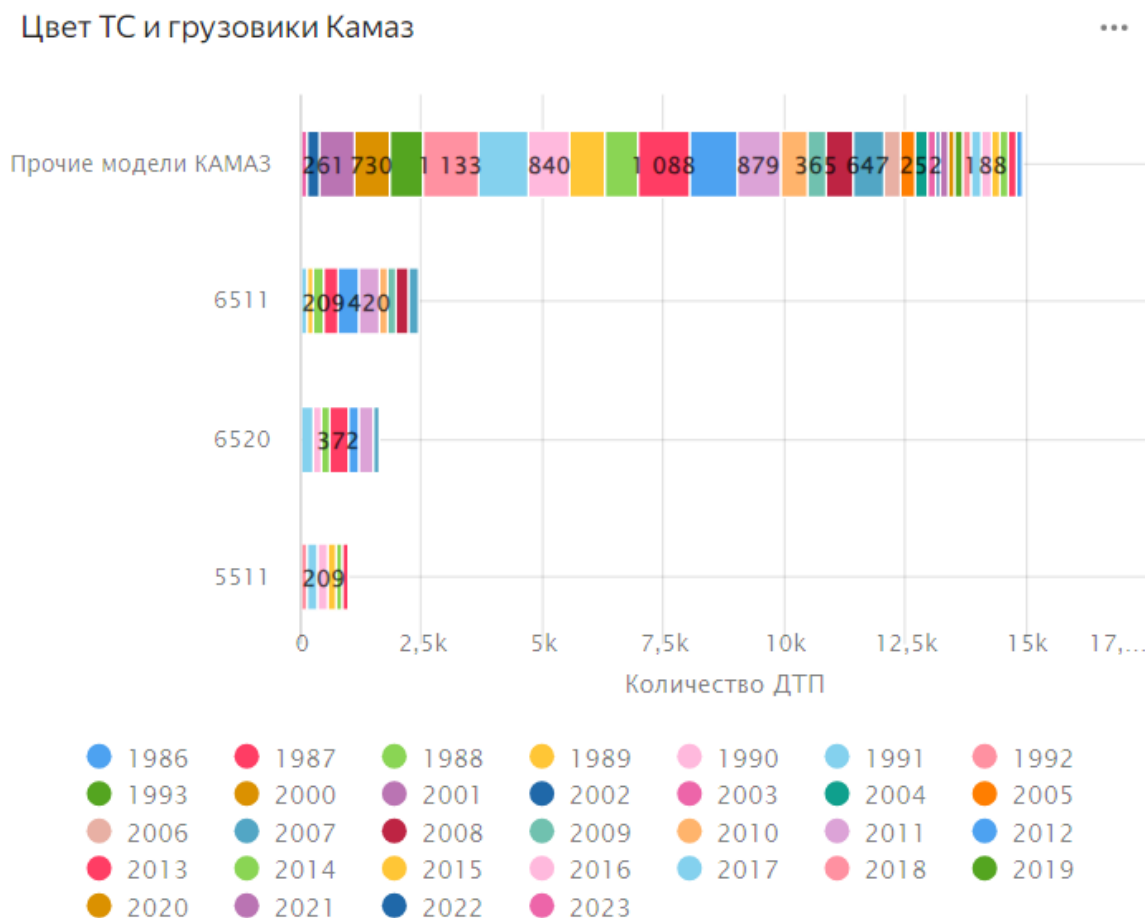
Раскроем исследование и посмотрим марки грузовиков оранжевого цвета

Визуализация: «Последовательный анализ средствами DataLens»



На оси ординат диаграммы позиции, номинированные 4-х значными цифрами - это модели автомобилей.. Те, которые начинаются с 5 и с 6 - это модели Камазов.

мы можем сделать вывод, что именно Камазы вносят основной и доминирующий вклад в смертность при ДТП, но вновь мы должны вспомнить, что Камазов, вероятно, больше всего в автопарках и частном владении среди грузовиков. Но это не отменяет значимости технического обслуживания и контроля за ним для этой марки грузовиков.



И даже с учетом наложенного в чарте ограничения мы видим, что в эксплуатации находится довольно старая техника - которой уже более 30 лет и которая требует либо очень повышенного внимания к ее техническому состоянию, либо к ее замене



Выводы и Бизнес-задачи

1. В исследовании мы проверили несколько гипотез. Они лежали на поверхности и результат проверки ожидаем
2. Целесообразно рекомендовать:
 - использовать встроенный обновляемый классификатор, который автоматически отнести бренд/модель легкового (и иного) автомобиля в соответствующую категорию.
 - Целесообразно унифицировать классификацию автобусов в плане этажности
 - Проверить корректность наименования и заполнения значений категорий участников "Все участники".
3. Мы имеем огромный статистический материал и интересно сформировать и реализовать по-настоящему прикладную задачу. Такой могла бы быть проверка гипотезы, что на загородных дорогах с раздельным движением в разные стороны безопасность (аварии, ранения и смертность) выше, чем на дорогах без раздельного движения. Развитием мог бы быть расчет и экономического эффекта от строительства таких дорог.