

Визуализация открытых данных

Тема №50

Вступление

В данной теме надо было выбрать любые российские открытые данные и сделать красивую, информативную визуализацию с выводами. В качестве данных я взял информацию о дорожно-транспортных происшествиях на автодорогах федерального, регионального и межмуниципального значения с ресурса бездороги.рф.

Основная часть

К сожалению, на данном ресурсе отсутствовала информация за 2020 и 2019 годы, поэтому я решил взять за основу данные за 2017 и 2018 годы. Для анализа данных я решил использовать Python вместе с библиотеками `pandas`, `pumPy`, а также `ipyleaflet` для создания карты с помощью `openstreetmap` и `matplotlib` для визуализации данных.

Данные возможно было получить в формате `csv` или `json`, я выбрал формат `json`, так как этот формат мне больше знаком.

Для начала я решил получить агрегированную информацию по общему количеству аварий (209,165), пострадавших (282,887) и жертв (28,499), для этого я сделал функцию `get_meta`. Далее, я сделал функцию `show_vics_count_histogram`, которая рисует гистограмму, на которой изображено количество аварий и жертв в каждый из месяцев года, на которой видно, что больше всего аварий приходится на осень, а меньше всего весной и функцией `show_count_daily_histogram`, которая рисует гистограмму с распределением аварий в течение дня, на которой видно, что значительная часть ДТП происходит в промежутке 17:00-20:00 (что неудивительно, т.к. это часы пик и дороги загружены больше всего в это время), а затем снижается и

достигает минимума в 05:00, после чего опять начинает постепенно расти, с большим скачком в 07:00.

Далее, мне стало интересно посмотреть на распределение ДТП по регионам России, для этого я сделал функцию `show_region_count_histogram`, которая рисует гистограмму и показывает количество ДТП в разных регионах, на которой видно, что основная часть ДТП происходит в Москве, Московской области и Санкт-Петербурге, так как в этих регионах самая большая плотность населения и наибольшее число автомобилей.

После чего, я решил посмотреть на соотношение разных типов ДТП, для этого я сделал функцию `show_crash_type_pie`, которая рисует круговую диаграмму, на которую нанесены разные типы ДТП, из которой можно сделать вывод, что наиболее частый тип ДТП – это столкновение и наезд на пешехода и функцию `show_accidents_with_vic_perc_pie`, которая рисует круговую диаграмму, на которой изображено соотношение аварий, где есть жертвы и где их нет, в 2017 году в 91% аварий не было жертв, а в 2018 году 82% аварий обошлись без жертв.

Далее, я сделал функцию `show_map`, которая наносит на карту России ДТП, на которой видно, что подавляющее число аварий происходит в европейской части России.

Заключение

Ежегодно на дорогах погибает огромное количество людей, сопоставимое с населением небольшого города. Выводы можно сделать очень банальные и простые, автомобиль – это транспорт повышенной опасности и халатное отношение к технике безопасности при вождении может привести к непоправимым последствиям.

Источники

- [безопасныедороги.рф](#)
- [pandas.pydata.org](#)

- `pyleaflet.readthedocs.io`