

## СЕССИЯ 1

### Исходные файлы:

- |                               |   |
|-------------------------------|---|
| 1) train_first_part.xlsx      | (Первая часть обучающего набора данных) |
| 2) train_second_part.json     | (Вторая часть обучающего набора данных) |
| 3) weather.csv                | (Набор данных о погоде)                 |
| 4) Машинное обучение – C1.pdf | (Инструкция к 1 сессии)                 |

### Результаты работы:

- |                                    |                              |
|------------------------------------|------------------------------|
| 1) Data.zip                        | (Предобработанные данные)    |
| 2) Report_C1.html, Report_C1.ipynb | (Отчет о проделанной работе) |
| 3) Readme.txt                      | (Дополнительные комментарии) |

## ВВЕДЕНИЕ

На этом чемпионате вам предстоит изучить информацию об движения транспорта в исследуемой локации и решить задачу прогнозирования времени поездки. Задача является актуальной для оператора компании. Прогнозирование времени поездки позволит оператору оптимизировать свою работу и улучшить финансовые показатели.

С точки зрения учебных целей эта работа также представляет большой интерес. В задаче предстоит работать с различными типами данных, в том числе с геораспределенными. Эти данные позволяют генерировать большое количество новых признаков. Что выделяет одну из особенностей задачи - возможность построения и исследования большого количества новых признаков.

Набор данных представлен в виде csv, xlsx и json-файлов для train и test-выборки.

Целевая переменная Trip duration - продолжительность поездки в секундах.

На этой сессии необходимо выполнить подготовку данных к анализу и построению моделей.

## ЗАДАНИЕ

### 1.1 Подготовка обучающей выборки

Необходимо выполнить подготовку данных обучающей выборки для дальнейшего описательного анализа и построения прогнозных моделей. Следует выполнить загрузку всех необходимых данных по перевозкам. Требуется выполнить объединение двух частей обучающего набора данных по перевозкам. Данные необходимо очистить от неинформативных признаков.

### 1.2 Форматирование - приведение данных к приемлемому формату

Данные выборки привести к приемлемому формату. В данных могут присутствовать значения, являющиеся выбросами. Требуется выявить такие значения и обработать их.



### 1.3 Дополнение набора данных

Следует выполнить загрузку и преобразование данных по погоде в регионе. Дополнить исходный набор данных по поездкам информацией о погодных условиях в соответствии с временными параметрами. Полученные наборы сохранить для дальнейшего исследования.

### 1.4 Подготовка отчета

Подготовьте отчет о проделанной работе по итогам сессии, в котором будут представлены результаты, выводы и обоснования выбора по каждому разделу задания. Результаты работы должны состоять из отчетов в формате .html и исходников с возможностью перекompilляции. Архив Data.zip должен содержать все результаты выполнения модуля, а также все необходимые файлы для запуска и проверки участков кода. В файле Readme.txt необходимо описать содержимое результирующих файлов архива Data.zip.

