

## СЕССИЯ 2

- |                               |                                |
|-------------------------------|--------------------------------|
| 1) Data.zip                   | (Результаты предыдущей сессии) |
| 2) Машинное обучение – C2.pdf | (Инструкция к 2 сессии)        |

### Результаты работы:

- |                                    |                              |
|------------------------------------|------------------------------|
| 1) Data.zip                        | (Предобработанные данные)    |
| 2) Report_C2.html, Report_C2.ipynb | (Отчет о проделанной работе) |
| 3) Readme.txt                      | (Дополнительные комментарии) |

## ВВЕДЕНИЕ

В этой сессии вы продолжаете работать с данными, подготовленными в предыдущей сессии. Предстоит провести разведочный анализ для изучения имеющихся данных (EDA) и заняться построением признаков.

Какая-либо работа, обусловленная задачами предыдущей сессии, выполненная в ходе текущей, оцениваться не будет, поэтому проделывайте её только в случае необходимости.

## ЗАДАНИЕ

### 2.1 Конструирование признаков (Feature Engineering)

Добавьте новые временные признаки, которые на ваш взгляд позволят улучшить качество выбранной модели. Дополните набор данных гео-признаками. Опишите приемы генерации новых атрибутов и результаты работы.

### 2.2 Кластеризация

Добавьте новые признаки путем кластеризации координат посадки и высадки. Выполните визуализацию пространственных данных на карте с отображением полученной кластерной структуры и дайте интерпретацию полученным результатам.

### 2.3 Исследование целевой переменной

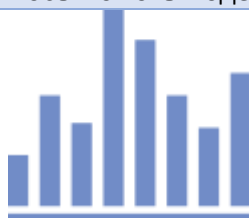
Выполнить проверку нормальности распределения продолжительности поездки. Для определения отклонений от нормы провести графическую и статистические оценки. Обосновать выбор методов оценки. На основе полученных результатов сделать вывод о типе распределения и мере скошенности.

Провести преобразования для изменения формы распределения.

### 2.4 Визуальный анализ данных

Используя программные средства, визуализируйте зависимости атрибутов в наборе данных. Визуализация должна отражать влияние атрибутов на целевую переменную (не менее трех). Приведите интерпретацию полученных результатов.

Выполнить корреляционный анализ для количественной оценки взаимосвязи атрибутов в наборе данных. Проведите выбор признаков, т.е. удалите часть коллинеарных признаков, чтобы помочь модели лучше обобщать новые данные для повышения её качества.



## 2.5 Подготовка отчета

Подготовьте отчет о проделанной работе по итогам сессии, в котором будут представлены результаты, выводы и обоснования выбора по каждому разделу задания. Отчет должен включать следующие пункты:

- Выбор способов визуализации
- Результаты визуализации
- Выбор и обоснование дополнения выборки новыми признаками
- Анализ визуального исследования новых признаков

Результаты работы должны состоять из отчетов в формате .html и исходников с возможностью перекомпиляции. 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.

