

## Вариант 7

### А. Этапы аналитики данных

Небольшой транспортной компании (пассажирское такси) из Перу исполняется 1 год. В честь праздника руководство компании запланировало премировать лучших водителей, а также постоянных и наиболее ответственных клиентов. Вы работаете аналитиком данных в этой компании и вам поручили подготовить данные для принятия решений:

- определить критерии выбора лучших водителей и пассажиров с учетом имеющихся оценок и количества поездок;
- отфильтровать и подготовить необходимые данные для анализа;
- придумать и рассчитать численные метрики для решения поставленной задачи;
- подготовить необходимые графики для лиц принимающих решения.

Вместе с этим руководство компании обратилось к вам с просьбой дать ответ на ряд дополнительных вопросов:

1. Есть ли корреляция между типом заказа и тем, как клиент оценивает водителя?
2. Есть ли корреляция между ответственностью клиента (оценками, которые выставляют ему водители) и платформой, с которой он делает заказ.
3. Есть ли зависимость между продолжительностью поездки и типом заказа?

Для решения поставленных задач Вы располагаете данными, выгруженными из автоматической системы учета поездок, которые хранятся в файле `taxi_peru.csv`

Расшифровка колонок:

- `journey_id` – уникальный id поездки
- `user_id` – id пользователя
- `driver_id` – id водителя
- `taxi_id` – id машины
- `icon` – тип поездки
- `start_type` – тип заказа (`asap`, `reserved`, `delayed`)
- `start_at` – время начала поездки
- `start_lat` – исходное местоположение пользователя, широта
- `start_lon` – исходное местоположение пользователя, долгота
- `end_at` – время окончания поездки
- `end_lat` – итоговое местоположение, широта
- `end_lon` – итоговое местоположение, долгота
- `end_state` – состояние заказа
- `driver_start_lat` – исходное местоположение водителя, широта
- `driver_start_lon` – исходное местоположение водителя, долгота
- `arrived_at` – время прибытия водителя
- `source` – платформа, с которой сделан заказ
- `driver_score` – оценка водителя клиентом
- `rider_score` – оценка клиента водителем

Возможно, для удобства расчетов времени поездок, вам потребуется считать данные колонок `start_at`, `end_at`, `arrived_at` в формате даты (см. параметр `parse_dates` функции `read_csv`).