#Исследование поездок пользователей на самокатах

###В последнее время электрические самокаты стали все чаще появляться на улицах: они относительно дешевые, быстрые и не требуют водительских прав для катания. Проанализировав популярный сервис аренды самокатов GoFast, мы сможем лучше понять поведение клиентов как с подпиской, так и без нее. Правильный анализ этих данных позволит составить примерный портрет пользователя и укажет на возможные изменения, которые необходимо совершить для более комфортного существования сервиса.

## План проекта

### Шаг 1. Подготовка данных

Загрузка и изучение общей информации из предоставленных датасетов

### Шаг 2. Предобработка данных

Обработка дубликатов и пропущенных значений

### Шаг 3. Исследовательский анализ данных

Описание и визуализация общей информации о пользователях и поездках:

- частота встречаемости городов;

- соотношение пользователей с подпиской и без подписки;

- возраст пользователей;

- расстояние, которое пользователь преодолел за одну поездку;

- продолжительность поездок.

### Шаг 4. Объединение данных

Объединение данных о пользователях, поездках и подписках в один датафрейм

Создание ещё двух датафреймов из общего:

- c данными о пользователях без подписки;

- с данными о пользователях с подпиской.

Визуализация информации о расстоянии и времени поездок для пользователей обеих категорий.

### Шаг 5. Подсчёт выручки

- создание датафрейма с агрегированными данными о поездках на основе датафрейма с объединёнными данными из шага 4: суммарное расстояние, количество поездок и суммарное время для каждого пользователя за каждый месяц.

- добавление столбца с помесячной выручкой, которую принёс каждый пользователь

###Шаг 6. Проверка гипотез

- тратят ли пользователи с подпиской больше времени на поездки?

- можно ли сказать, что среднее расстояние, которое проезжают пользователи с подпиской за одну поездку, не превышает 3130 метров?

- будет ли помесячная выручка от пользователей с подпиской по месяцам выше, чем выручка от пользователей без подписки?

###Шаг 7. Распределения

- отделу маркетинга GoFast поставили задачу: нужно провести акцию с раздачей промокодов на один бесплатный месяц подписки, в рамках которой как минимум 100 существующих клиентов должны продлить эту подписку. То есть по завершении периода действия подписки пользователь может либо отказаться от неё, либо продлить, совершив соответствующий платёж.

Эта акция уже проводилась ранее и по итогу выяснилось, что после бесплатного пробного периода подписку продлевают 10% пользователей. Какое минимальное количество промокодов нужно разослать, чтобы вероятность не выполнить план была примерно 5%?

- отдел маркетинга рассылает клиентам push-уведомления в мобильном приложении. Клиенты могут открыть его или не открывать. Известно, что уведомления открывают около 40 %получивших клиентов. Отдел планирует разослать 1млн уведомлений. С помощью аппроксимации построю примерный график распределения и оценю вероятность того, что уведомление откроют не более 399,5тыс. пользователей.

###Шаг 8. Вывод

## Используемые инструменты

scipy, numpy, pandas, matplotlib