# **🧠 Учебная задача: Анализ сайта «СберАвтоподписка»**

## **🎯 Цель проекта**

Компания «СберАвтоподписка» хочет увеличить эффективность сайта: улучшить пользовательский опыт, повысить конверсию, сделать рекламные кампании более результативными. Для этого вам предстоит **создать модель, которая предсказывает вероятность того, что пользователь совершит целевое действие (оставит заявку, закажет звонок и пр.)** на сайте.

Эта модель поможет:

* Оценивать эффективность каналов привлечения трафика;
* Адаптировать рекламные кампании;
* Улучшать UX сайта за счет анализа поведения пользователей.

## **🏢 О компании «СберАвтоподписка»**

**СберАвтоподписка** — это сервис долгосрочной аренды автомобилей для физических лиц. Клиент платит фиксированную сумму в месяц и получает авто в пользование сроком от 6 месяцев до 3 лет.

В стоимость включены:

* Страхование (КАСКО, ОСАГО, ДСАГО);
* ТО и ремонт;
* Сезонная смена и хранение шин;
* Круглосуточная поддержка.

Дополнительно можно заказать **консьерж-сервис** — доставка автомобиля в сервис и обратно.

Сервис выступает альтернативой покупке авто или автокредиту — машину выгоднее арендовать, а средства инвестировать.



## **🚘 О продукте**

* В каталоге представлено 20+ моделей.
* Возможна покупка авто под конкретного клиента.
* Пользователь выбирает:  
  + Срок аренды (6–36 мес.);
  + Пакет пробега (ограничение по км).

Процесс оформления:

1. Пользователь выбирает авто → нажимает «Оформить подписку»;
2. Вводит контактные данные;
3. Менеджер связывается с пользователем, помогает подобрать машину;
4. Проводится скоринг;
5. После одобрения — оформление договора, выдача авто (в Москве или с доставкой в регионы);
6. В дальнейшем клиент взаимодействует через email и телефон.

## **📊 Данные**

Вам предоставлен лог пользовательских действий на сайте и дополнительные данные по сессиям.

**Некоторые атрибуты**:

* utm\_\*, geo\_\*, device\_\* — характеристики визитов;
* ga\_hits.hit\_page\_path — страницы сайта;
* ga\_sessions.utm\_medium/source — источники трафика.

## **📚 Глоссарий**

| **Термин** | **Значение** |
| --- | --- |
| **Целевое действие** | Любое ключевое событие на сайте: Оставить заявку, Заказать звонок, Открыть диалог, и др. (ga\_hits.event\_action in [перечень ключей]). |
| **CR (Conversion Rate)** | Конверсия — доля визитов, в которых было совершено хотя бы одно целевое действие. |
| **Органический трафик** | Визиты с utm\_medium in ('organic', 'referral', '(none)'). |
| **Платный трафик** | Все остальные типы трафика. |
| **Соцсети** | Визиты с источников из набора utm\_source (перечень зашифрованных ID). |

## **🛠 План работы**

### **1. Подготовка данных**

* Прочитайте датасет и изучите описание атрибутов.
* Оцените полноту, чистоту и структуру.
* Приведите данные к удобному виду (типизация, обработка NaN, удаление дубликатов и пр.).

### **2. Разведочный анализ (EDA)**

* Исследуйте распределения ключевых признаков (histograms, boxplots).
* Постройте тепловую карту корреляций.
* Добавьте комментарии с интерпретацией: какие связи вы наблюдаете?

### **3. Модель машинного обучения**

* Постройте модель, предсказывающую вероятность совершения целевого действия.
* Ориентир по метрике качества: **ROC-AUC ~ 0.65**.
* Подходите к фиче-инжинирингу осмысленно:  
  + создайте новые фичи;
  + проверьте линейную зависимость;
  + интерпретируйте значимость фичей.

### **4. Упаковка модели**

* Сделайте API или скрипт, который:  
  + принимает данные по визиту (utm, geo, device и т.д.);
  + возвращает 0 или 1 (будет ли целевое действие).

## **📁 Что сдавать**

### **1. Аналитический отчет (Jupyter Notebook):**

* Чтение и очистка данных;
* EDA с графиками и комментариями;
* Подход к ML;
* Выводы.

### **2. Модель и API:**

* .py-файл с моделью + инструкция по запуску;
* или web app, доступный по адресу localhost.

## **🧮 Критерии оценки**

### **EDA**

* Обработка пропусков, дубликатов, типизация;
* Визуализация распределений и корреляций;
* Интерпретация связей.

### **ML**

* Сравнение метрики модели с бейзлайном;
* Качество фичей:  
  + инжиниринг;
  + интерпретируемость;
  + значимость;
* Выбор подходящей модели.

### **API**

* Корректность;
* Масштабируемость;
* Время ответа ≤ 3 секунды.