# Описание кейса «Прогнозирование оттока клиентов»

**Суть бизнес задачи:**

Каждый год компания пролонгирует полисы Каско клиентов - физических лиц. Для оптимизации работы со списками на пролонгацию необходимо прогнозировать с какой вероятностью каждый из клиентов пролонгируется и какие факторы на это влияют. В зависимости от этого расставляются приоритеты операторам колл-центра, которые обрабатывают список (осуществляют обзвон клиентов), а также принимаются решения по дополнительной мотивации клиентов к пролонгации.

**Описание массива:**

Дана выборка полисов на пролонгацию в формате CSV-файла с набором полей, характеризующих сам полис, клиента (и его историю страхования) и транспортное средство. Полис на пролонгацию – это полис, период действия заканчивается и который нужно пролонгировать.

Прогнозируемая переменная – факт пролонгации полиса «POLICY\_IS\_RENEWED», где «1» – клиент пролонгировался, «0» - клиент не пролонгировался.

Массив случайным образом разбит на 2 части: 80% данных – тренировочная выборка, 20% данных – тестовая выборка.

**Проверка и оценка результата:**

Для целей моделирования и проверки данных предоставляется тренировочная и тестовая выборки. В тестовой выборке значения прогнозируемой переменной обнулены.

**Требования к оформлению:**

1. Каждый шаг описать любым удобным способом (word, excel, pp, jupyter и т.д.), чтобы была возможность проверить проделанную работу.

2. Итоговый слайд оформить наглядно и понятно.

3. Прислать код на Python

4. Ответы на тестовой выборке предоставить в виде CSV файла вида:

Формат таблицы передачи результата:

|  |  |  |
| --- | --- | --- |
| POLICY\_ID | POLICY\_IS\_RENEWED | POLICY\_IS\_RENEWED\_PROBABILITY |

POLICY\_IS\_RENEWED\_PROBABILITY – вероятность пролонгации (если метод прогнозирования предполагает её расчёт), параметр не обязателен для заполнения.