**Федеральное государственное бюджетное образовательное учреждение**

**высшего образования**

**«Московский государственный технический**

**университет имени Н.Э. Баумана**

**(национальный исследовательский университет)» (МГТУ им. Н.Э. Баумана)**

Факультет «Информатика и вычислительная техника»

Кафедра ИУ5 «Системы обработки информации и управления»

Курс «Технология машинного обучения» Отчет по лабораторной работе №1

«Разведочный анализ данных. Исследование и визуализация данных.»

|  |  |
| --- | --- |
| Выполнил: | Проверил: |
| студент группы ИУ5-61Б | преподаватель каф.  ИУ5 |
| Рысьева Е. А. | Гапанюк Ю.Е. |
| Подпись и дата: | Подпись и дата: |

Москва, 2023 г.

# Цель лабораторной работы:

Изучение различных методов визуализация данных.

# Описание задания:

* Выбрать набор данных (датасет)
* Для первой лабораторной работы рекомендуется использовать датасет без пропусков в данных

Для лабораторных работ не рекомендуется выбирать датасеты большого размера.

* Создать ноутбук, который содержит следующие разделы:

1. Текстовое описание выбранного Вами набора данных.
2. Основные характеристики датасета.
3. Визуальное исследование датасета.

## Задача

Сеть фитнесс-центров "Культурист-датасаентист" разрабатывает стратегия взаимодестввия с пользователями на основе аналитических данных. Распространённая проьоема фитнес-клубов и других сервисов - отток клиентов. Как понять, что пользователь больше не с вами? Можно записать в отток тез, кто попросил закрыть договор или удалил аккаунт. Однако клиенты не всегда уходят демонстративно: чаще перестают пользоваться сервисом тихо. Индикаторы оттока зависят от специфики отрасли. Когда пользователь редко, но стабильно закупается в интернет-магазине - не похоже, что он "отвалился". А вот если две недели не заходит на канал с ежедневно обновляемым контентом, дела плози: подписчик заскучал и, кажется, оставил вас. Для фитнес-центра можно считать, что клиент попал в отток, если за последний месяц ни разу не посетил спортзал. Конечно, не исключено, что он уехал на Бали и по приезде обязательно продолжит ходить на фитнес. Однако чаще бывает наоборот. Если клиент начал новую жизнь с понедельника, немного походил в спортзал, а потом пропал - скорее всего, он не вернется. Чтобы бороться с оттоком, отдел по работе с клинтами "Культуриста-датасаентиста" перевёл в электронный вид множество анкет пользователей.

**Задача - провести анализ и подготовить план действий по удержанию клиентов. А именно:**

* Научиться прогнозировать вероятность оттока (на уровне следующего месяца) для каждого клиента;
* Сформировать типичные портреты пользователей: выделить несколько наиболее ярких групп и охарактеризовать из основные свойства
* Проанализировать основные признаки, наиболее сильно влияющие на отток
* Cформулировать основные выводы и разработать рекомендации по повышению качества работы с клиентами:
  + Выделить целевые группы клиентов
  + Предложить меры по снижению оттока
  + Определить другие особенности заимодейстия с клиентами.

## Спецификация

* 'churn' - факт оттока в текущем месяце (целевая переменная)
* 'gender' - пол
* 'near\_location' - проживание или работа в районе, где находится фитнес-центр
* 'partner' - сотрудник компании-партёра клуба
* 'promo\_friends' - факт первоначальной записи в рамках акции "приведи друга"
* 'phone' - наличие контактного телефона
* 'age' - возраст
* 'lifetime' - время с первого обращения в фитнесс-центр
* 'contract\_period' - длительность текущего действующего аобнемента (месяц, 3 месяца, 6 месяцев, 1 год)
* 'month\_to\_end\_contract' - срок до окончания текущего действующего абонемента (в месяцах)
* 'group\_visits' - факт посещения групповых занятий
* 'avg\_class\_frequency\_total' - средняя частота посещений в неделю за все время с начала действия абонемента
* 'avg\_class\_frequency\_current\_month' - средняя частота посещений в неделю за предыдущий месяц
* 'avg\_additional\_charges\_total' - суммарная выручка от других услуг фитнес-центра: кафе, спорт-товары, косметический и массажный салон





