

## Основные понятия модуля

**EDA** — разведывательный анализ данных — анализ основных свойств данных, нахождение в них закономерностей, характера и свойств данных.

### EDA включает в себя:

- Проектирование признаков
- Отбор признаков
- Кодирование признаков
- Построение и проверка статистических гипотез

### Цикл машинного обучения

Кнопка	Действие
<b>Постановка проблемы</b>	Жизненный цикл нашей модели начинается с определения проблемы бизнеса
<b>Сбор данных</b>	На этапе сбора необходимо собрать как можно больше любых данных
<b>Очистка данных</b>	На этапе очистки данных мы определяем пропущенные значения, аномалии и выбросы в данных. Подготавливаем набор данных к анализу
<b>EDA</b>	Исследование признаков, извлечение полезной информации
<b>Моделирование</b>	Построение предсказательных моделей, проверка качества модели
<b>Продакшен</b>	Вывод модели в продакшен

## Направления профессии

Профессия	Роль
<b>Дата-сайентист</b>	Создает рекомендательные алгоритмы и прогнозные модели
<b>Дата-аналитик</b>	Проводит предварительный анализ данных и их очистку, ищет закономерности в данных и ответы на запросы бизнеса
<b>Дата-инженер</b>	Обеспечивает сбор и хранение данных. Занимается внедрением модели

## Инструменты автоматической визуализации

Библиотека	Установка
<a href="#">Pandas-profiling</a>	<code>pip install pandas-profiling</code>
<a href="#">Sweetviz</a>	<code>pip install sweetviz</code>
<a href="#">D-tale</a>	<code>pip install dtale</code>