# Проект по Машинному обучению

\* - не обязательное (индивидуально для каждого датасета)

### Задание:

1. Описать набор данных (ЛР1 ИБД)
   1. Про что он
   2. Кем, когда создан
   3. Объем
   4. Количество строк
   5. Задача (Классификация/Регрессия), полный ответ
   6. Описать каждый признак (Название, тип данных, тип шкалы [Количественные, Порядковые, Категориальные, Бинарные], Пропуски, Оценка востребованности)
2. Подготовка данных (ЛР3 ML)
   1. Получить список категориальных и числовых признаков
   2. Использовать для предобработки данных визуализацию
   3. \*Балансировка класса (ЛР7-8 ML)
   4. \*Заменить пропуски
   5. \*Кодировка категориальных значений
   6. \*Нормализировать числовые данные
   7. \*Отбор по перцентилю
   8. \*Удалить из набора объекты с редкими категориями
   9. Сформировать минимум 3 набора данных, используя различные способы подготовки
3. Выбор и оценка модели
   1. Задача
   2. Выбрать метрики оценки результата (модели)
   3. Подготовка обучающего и тестового наборов (X, y, train, test)
   4. Выбор моделей (BaseLine + одиночные + ансамблевые). Не менее трех одиночных и трех ансамблевых.
   5. Обучить модели и подобрать гиперпараметры (GridSearchCV)
   6. Для подобранных моделей (Best estimator) найти значения метрик (из пункта 3.2)
   7. Повторить пункты 3.5 и 3.6 для всех наборов данных из пункта 2
4. Анализ Результатов:
   1. Визуализация по метрикам
   2. Какая модель проявила себя лучше (из пункта 3.4)
   3. Объяснение результатов