

# Задача хакатона: ML-модель оценки стоимости квартир

## Бизнес-проблема:

Необходимо предсказывать рыночную цену квартиры на основе реальных объявлений. Это поможет определить адекватную стоимость и принимать решения в сфере недвижимости.

---

## Цель

Создать **машинную модель**, которая на основе характеристик квартиры (площадь, этаж, район, состояние и др.) будет предсказывать адекватную цену.

---

## Данные

Участникам предоставляется датасет с объявлениями о продаже квартир (например, с Somon.tj).

Каждое объявление содержит:

- Площадь квартиры ( $\text{м}^2$ )
  - Количество комнат
  - Этаж / Этажность
  - Район / Локация
  - Состояние (ремонт / новостройка / без ремонта)
  - Другие доступные параметры
  - Фактическая цена
- 

## Задачи

1. **Предобработка данных**
  - Обработка пропусков
  - Работа с категориальными признаками
  - Обнаружение и обработка выбросов
  - Нормализация и масштабирование
2. **Построение модели**
  - Разделение данных на train/test (hold-out или кросс-валидация)

- Обучение **как минимум двух моделей регрессии** (например, Linear Regression + RandomForest/XGBoost/Gradient Boosting)
  - Оценка качества модели по метрикам: MAE, RMSE, R<sup>2</sup>
3. **Анализ признаков**
- Определение влияния признаков на цену
  - Визуализация предсказанных и реальных цен
4. **Документация**
- Скрипты/ноутбуки с подготовкой данных, обучением и тестированием моделей
  - README с инструкцией для запуска и воспроизведения результатов
- 

## **Ожидаемый результат**

- **GitHub-репозиторий** с:
    - Кодом / ноутбуками
    - Подготовкой данных
    - Обучением моделей и тестированием
    - Визуализациями
  - **Отчет или презентация**:
    - Какие модели были протестированы и почему выбран финальный подход
    - Метрики оценки качества
    - Анализ ошибок
    - Визуализация предсказаний vs. фактических цен
    - Важность признаков
- 

## **Критерии оценки**

1. **Качество прогноза**
  - Низкая ошибка предсказания (MAE/RMSE)
  - Высокий R<sup>2</sup>
2. **Работа с данными**
  - Корректная обработка
  - Обоснование выбора признаков
3. **Презентация и анализ**
  - Четкие выводы
  - Графики и пояснения
  - Обоснование выбора модели
4. **Реплицируемость**
  - Четкие инструкции для запуска
  - Возможность воспроизвести все результаты