Project Context

Our client is a large Real Estate Investment Trust (REIT).

- They invest in houses, apartments, and condos within a small county in New York State.
- As part of their business, they try to **predict the fair transaction price** of a property before it's sold. They do so to calibrate their internal pricing models and keep a pulse on the market.

Current Solution

The REIT currently uses a third-party appraisal service. Appraisers are professionals who visit a property and estimate a fair price using their own expertise.

- Unfortunately, the skill levels of individual appraisers vary greatly. During a trial run, the REIT compared appraiser estimates to actual transaction prices. The REIT found that the estimates given by **inexperienced appraisers** were off by $70,000, on average!

Our Role

The REIT has hired us to find a data-driven approach to valuing properties.

- They currently have an untapped dataset of transaction prices for previous properties on the market. Our task is to build a real-estate pricing model using that dataset.
- If we can build a model to **predict transaction price** with an average error of under $70,000, then our client can replace inexperienced appraisers with our model.

Problem Specifics

It's always helpful to scope the problem before starting.

`Deliverable:` **Trained model file**
`Machine learning task:` **Regression**
`Target variable:` **Transaction Price**
`Win condition:` **Avg. prediction error < $70,000**

Data Dictionary

For this project:

- The dataset has **1883** observations in the county where the REIT operates.
- Each observation is for the transaction of one property only.
- Each transaction was between $200,000 and $800,000.

We have the following features:

Target variable

- `'tx_price'` - Transaction price in USD

Public records for the property

- `'tx_year'` - Year the transaction took place
- `'property_tax'` - Monthly property tax
- `'insurance'` - Cost of monthly homeowner's insurance

Property characteristics

- `'beds'` - Number of bedrooms
- `'baths'` - Number of bathrooms
- `'sqft'` - Total floor area in squared feet
- `'lot_size'` - Total outside area in squared feet
- `'year_built'` - Year property was built
- `'basement'` - Does the property have a basement?

Location convenience scores

- `'restaurants'` - Number of restaurants within 1 mile
- `'groceries'` - Number of grocery stores within 1 mile
- `'nightlife'` - Number of nightlife venues within 1 mile
- `'cafes'` - Number of cafes within 1 mile
- `'shopping'` - Number of stores within 1 mile
- `'arts_entertainment'` - Number of arts and entertainment venues within 1 mile
- `'beauty_spas'` - Number of beauty and spa locations within 1 mile
- `'active_life'` - Number of gyms, yoga studios, and sports venues within 1 mile

Neighborhood demographics

- `'median_age'` - Median age of the neighborhood
- `'married'` - Percent of neighborhood who are married
- `'college_grad'` - Percent of neighborhood who graduated college

Schools

- `'num_schools'` - Number of public schools within district
- `'median_school'` - Median score of the public schools within district, on the range 1 - 10