



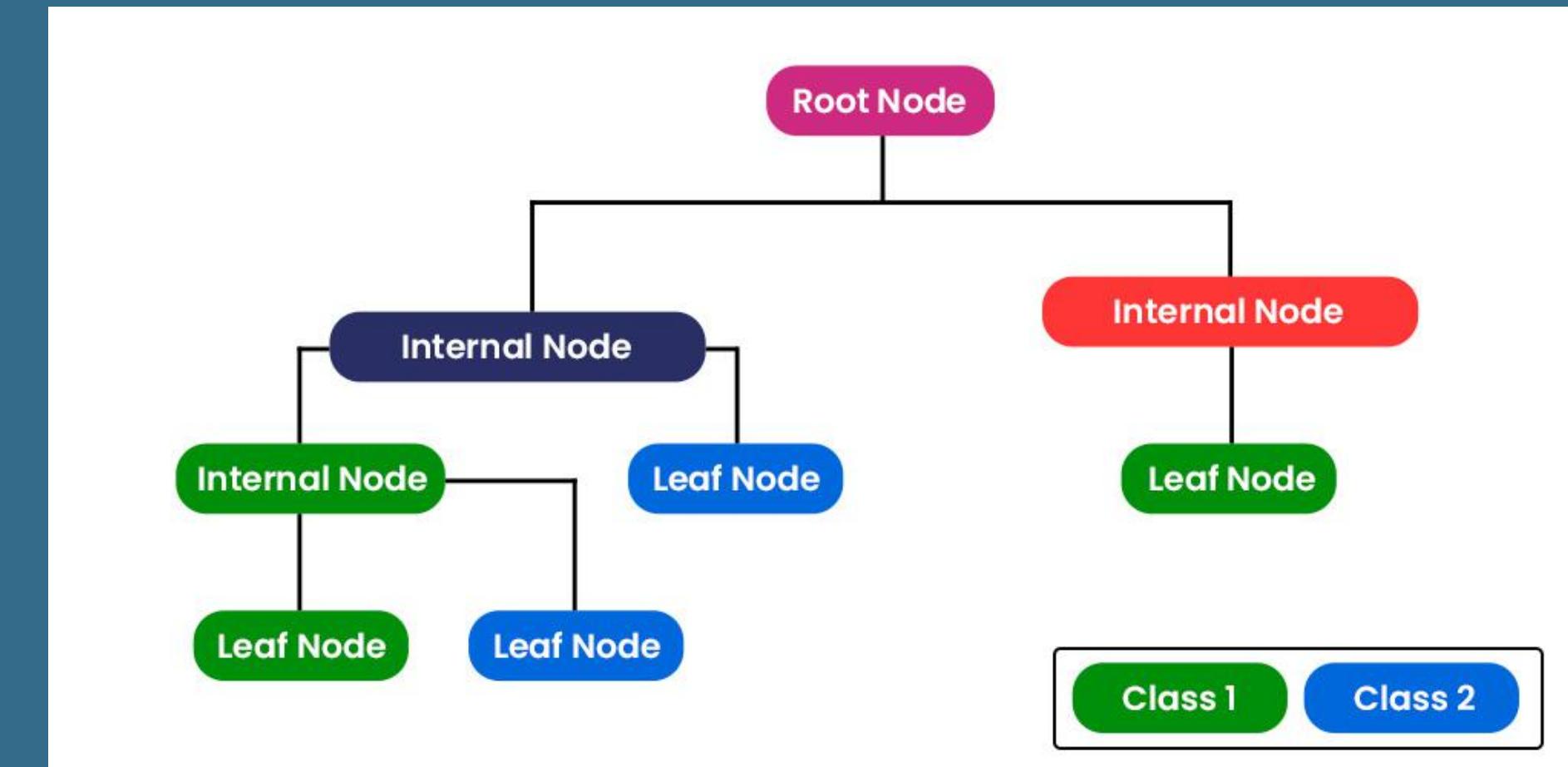
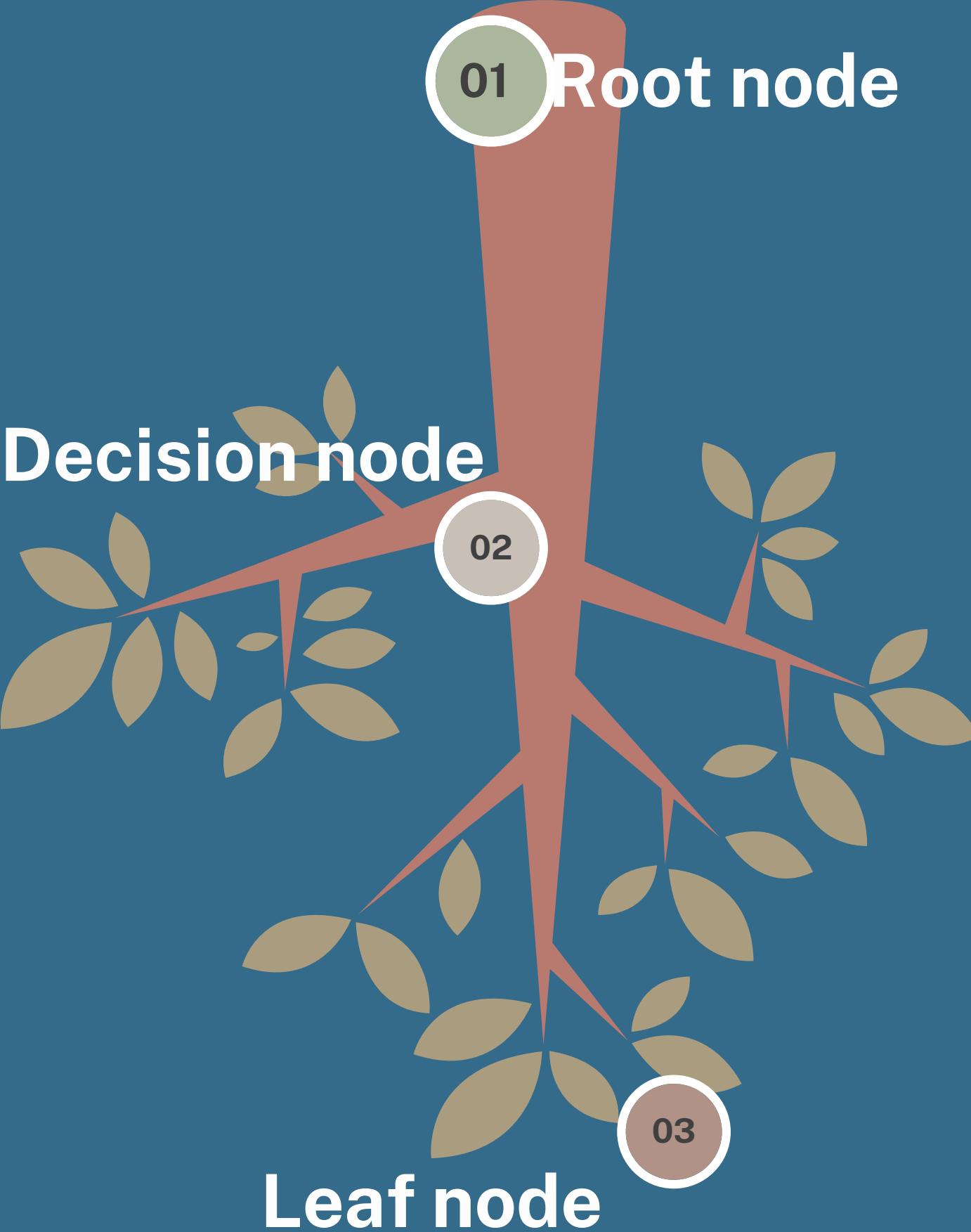
# Vancouver housing price prediction based on Decision Tree Regression

---

Zhou Chen, Xie Xueyi, Lyu Bochen

# Agenda

- Decision Tree and Project Overview
- Data Overview and Preparation
- Analysis and Results

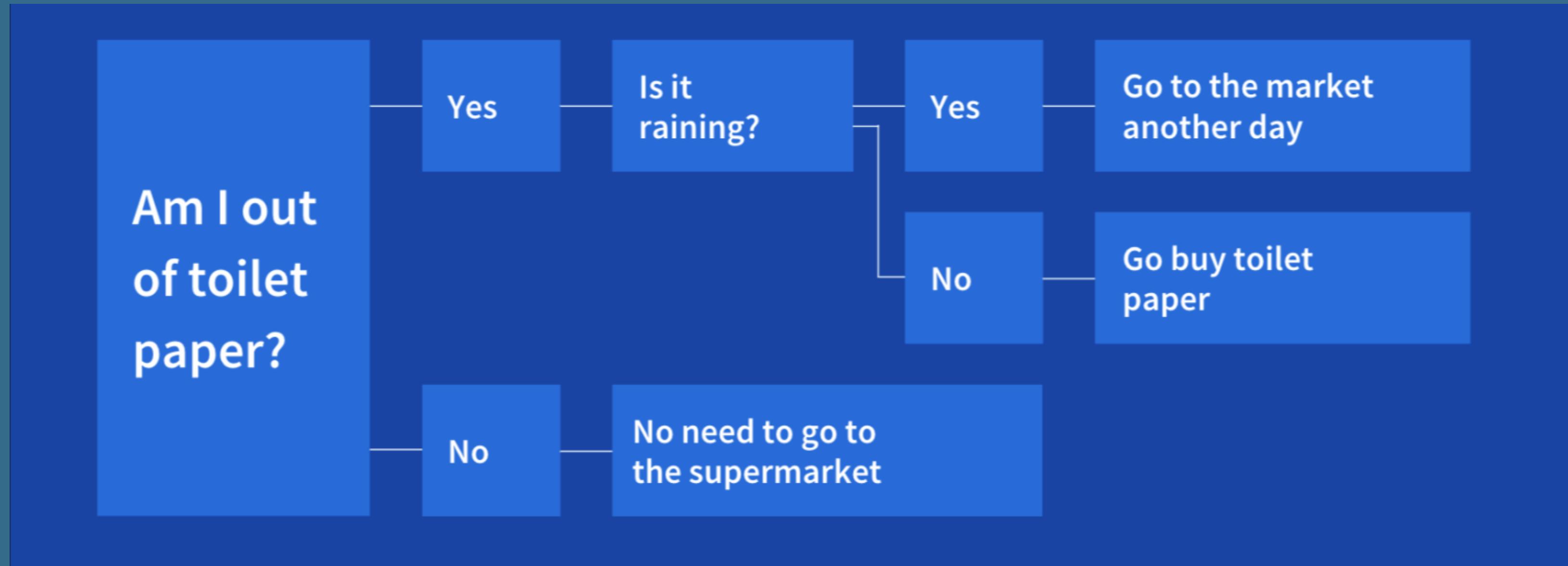


Two types of Decision Trees:

- Decision Tree Classifier
- Decision Tree Regressor

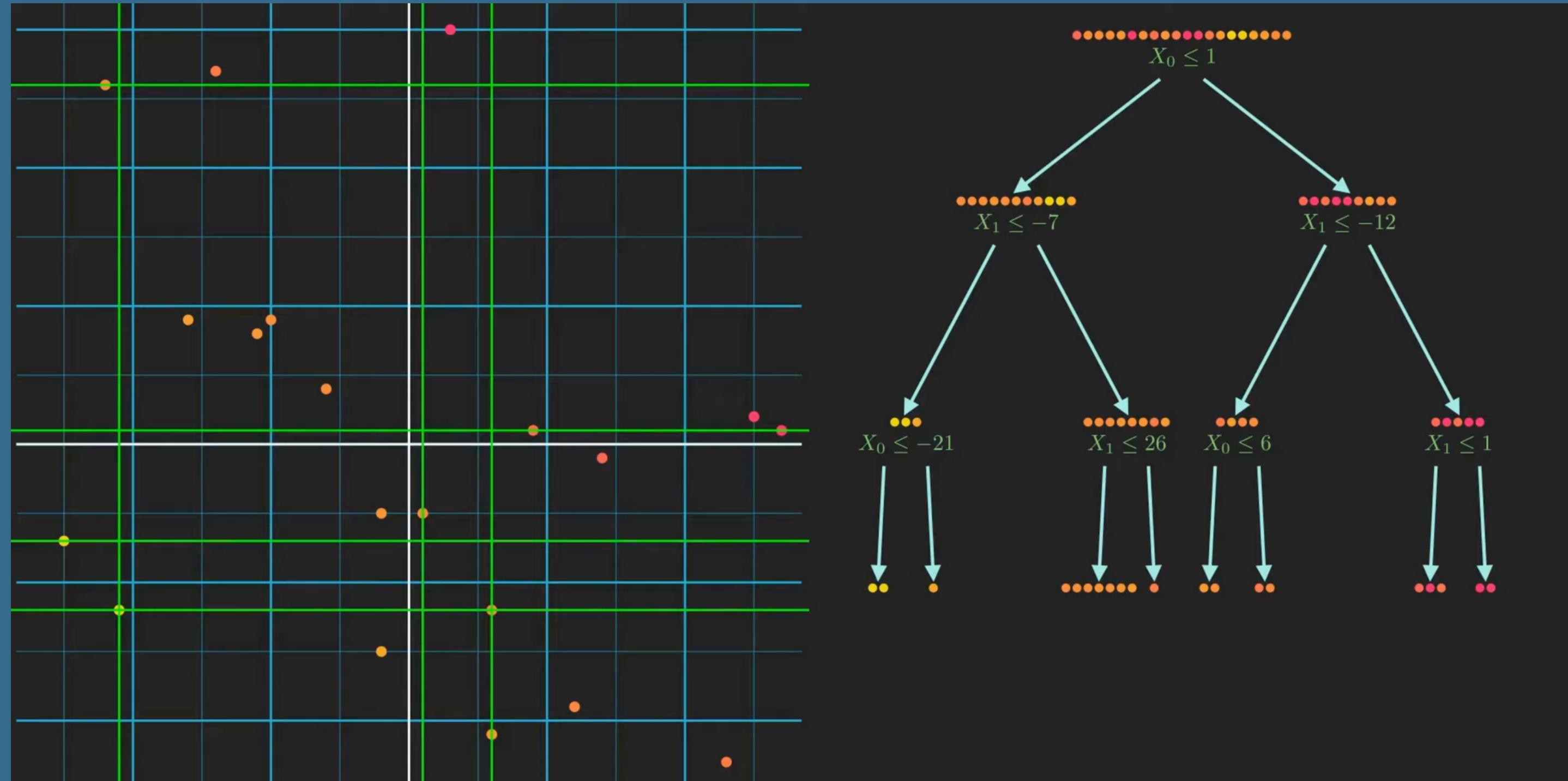
# 1. Classification trees

Classification trees determine whether an event happened or didn't happen.  
Mainly used to predict categorical data.



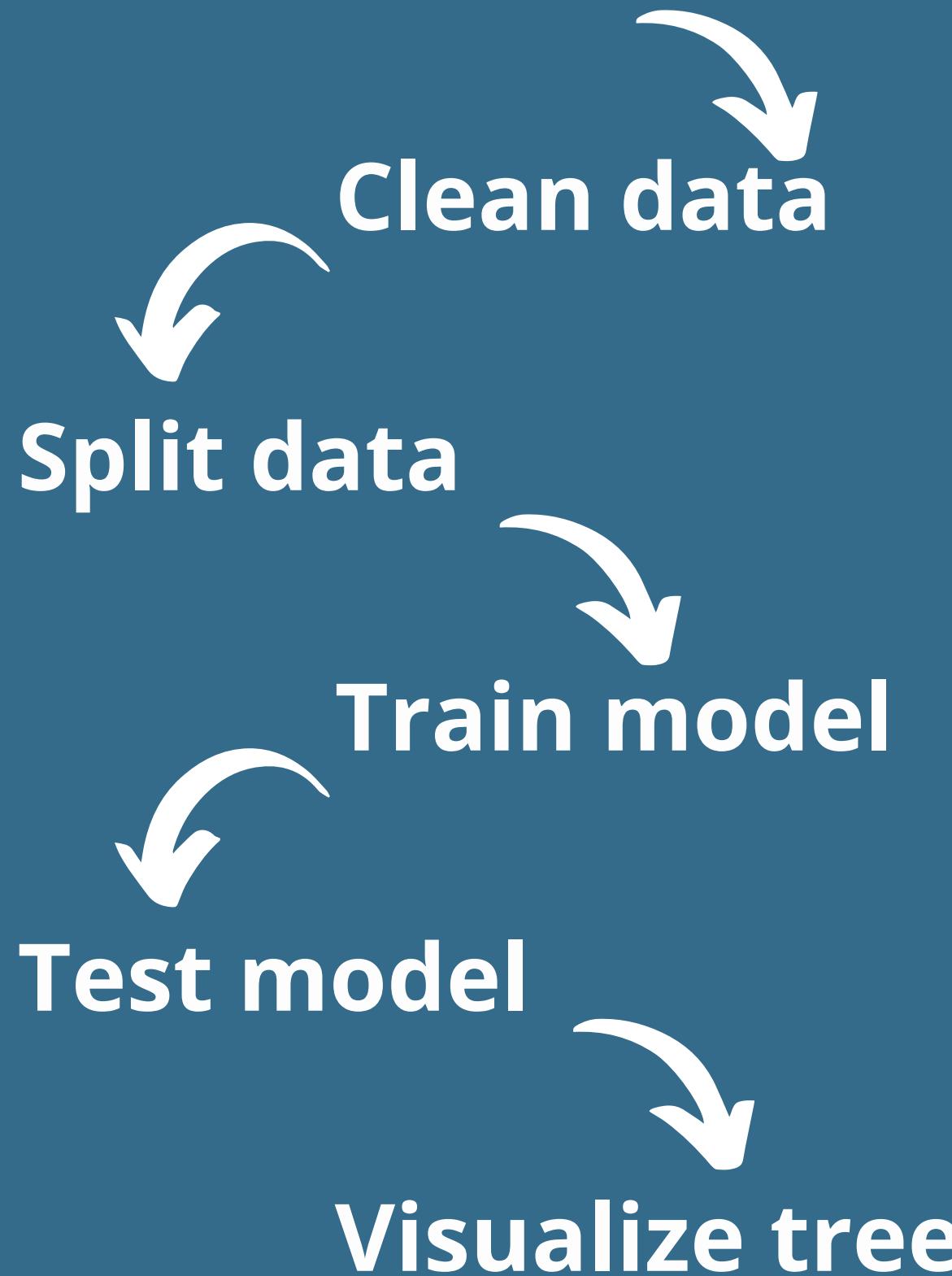
## 2. Regression trees

Regression trees predict what is likely to happen, given previous behavior or trends.  
Mainly used to predict numerical data.



# Project Scope

Obtain data

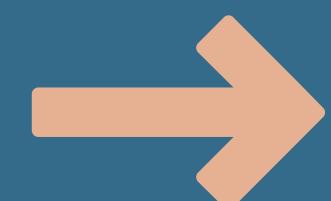


```
{"columns":[],"rows":[[{"R2739628","49.27395","-123.12865","11/22/2022","","# 1701 550 PACIFIC STREET","Yaletown","1249.000","Wow factor! Water view home at Aqua at the Park! Located in the lovely & coveted Beach district neighborhood at the south edge of Yaletown. This seldom available water view unit has floor to ceiling windows to showcase the protected southerly views of George Wainborn Park, False Creek and Granville Island. The building provides amazing facilities including 24 hr concierge, theatre room, gym, indoor pool, hot tub, steam room, sauna, squash court and guest suites. Steps to Sea Wall, shopping, fine dining, fresh market grocery and more. Pets and rentals allowed. Perfect level to embrace the greenery and water view in every room. Parking and storage locker too.", "Apartment", "Corner Unit", "17", "2", "1070", "", "", "88West Realty", "1701", "Keller Williams Realty VanCentral", "", "", "", "", "0.000", "", "2999.93", "1249.000", "0", "View (WATER&#44; PARK&#44; CITY), Central Location, Golf Course Nearby, Marina Nearby, Paved Road, Recreation Nearby, Shopping Nearby, Bike Room, Club House, Elevator, Exercise Centre, Guest Suite, In Suite Laundry, Playground, Pool; Outdoor, Sauna\Steam Room, Concierge, Clothes Washer \ Dryer \ Refrigerator \ Stove \ Dishwasher, Drapes \ Window Coverings, Garage Door Opener, Hot Tub, Intercom, Microwave, Pantry, Smoke Alarm, Sprinkler - Fire", "None", "1230.000", "11/25/2022", "11/28/2022", "V6Z 3G2", "Vancouver", "37", "2022", "0", "https://yout.be/P1WvifdLQ-c", "2022-11-28T18:49:09", "262761255", "1.104", "Other", "Baseboard, Electric", "City\Municipal", "DD", "Living Room\13'5 x 13'6\Main", "Dining Room\6'11 x 10'1\Main", "Foyer\4'10 x 3'10\Main", "Kitchen\8'2 x 7'8\Main", "Pantry\4'11 x 7'11\Main", "Office\6'8 x 5'7\Main", "Bedroom\9'5 x 12'5\Main", "Primary Bedroom\11'2 x 13'2\Main", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "", "Balcony(s)", "Garage Under Building, Visitor Parking (1)", "", "", "Pets Allowed with Restrictions, Rentals Allowed with Restrictions", "661.72", "0", "Brick, Concrete, Concrete Block", "Concrete Perimeter", "", "Concrete, Glass, Mixed", "0", "1070", "0", "0", "1", "Partly", "Side", "3 pc.\Main", "3 pc.\Y\Main", "", "", "", "", "", "City\Municipal", "026469880", "49.27369, -123.12855, 49.27369, -123.1286, 49.2737, -123.1287, 49.27371, -123.12875, 49.27371, -123.12881, 49.27392, -123.12915, 49.27426, -123.12864, 49.27395, -123.12816, 49.27369, -123.12855", "2497", "753", "361", "", "", "", "", "", "", "1", "BCS1570", "3.255% ON THE FIRST $100,000 AND 1.1625% ON THE BALANCE", "V", "", "", "", "", "", "", "", "Vancouver West", "", "Freehold Strata", "", "\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\0\0\0\67ac49e0e53d520c945b4ab9a0c6aa48\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\1\0\0\3472b9791c8dd679a03bd978b28e1af7\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-1.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\2\0\0\1e2f33cd1a1412157a466dd4ecd92df5\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-2.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\3\0\0\4a0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-3.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\4\0\0\5a0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-4.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\5\0\0\6a0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-5.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\6\0\0\7a0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-6.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\7\0\0\8a0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-7.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\8\0\0\9a0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-8.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\9\0\0\aa0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-9.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\10\0\0\ba0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-10.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\11\0\0\ca0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-11.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\12\0\0\da0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-12.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\13\0\0\ea0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-13.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\14\0\0\fa0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-14.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\15\0\0\ga0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-15.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\16\0\0\ha0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-16.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\17\0\0\ia0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-17.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\18\0\0\ja0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-18.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\19\0\0\ka0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-19.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\20\0\0\la0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-20.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\21\0\0\ma0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-21.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\22\0\0\oa0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-22.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\23\0\0\qa0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-23.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\24\0\0\ra0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-24.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\25\0\0\sa0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-25.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\26\0\0\ta0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-26.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\27\0\0\ua0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-27.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\28\0\0\va0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-28.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\29\0\0\za0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-29.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\30\0\0\aa0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-30.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\31\0\0\ba0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-31.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\32\0\0\ca0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-32.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\33\0\0\da0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-33.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\34\0\0\ea0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-34.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\35\0\0\fa0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-35.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\36\0\0\ga0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-36.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\37\0\0\ia0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-37.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\38\0\0\ja0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-38.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\39\0\0\ka0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-39.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\40\0\0\la0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-40.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\41\0\0\ma0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-41.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\42\0\0\oa0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-42.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\43\0\0\qa0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-43.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\44\0\0\ra0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-44.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\45\0\0\sa0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-45.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\46\0\0\ta0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-46.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\47\0\0\ua0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-47.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\48\0\0\va0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-48.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\49\0\0\za0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-49.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\50\0\0\aa0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-50.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\51\0\0\ba0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-51.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\52\0\0\ca0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-52.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\53\0\0\da0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-53.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\54\0\0\ea0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-54.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\55\0\0\fa0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-55.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\56\0\0\ga0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761255-56.JPG|\cdnparap130.paragonrels.com\ParagonImages\Property\p13\BCRES\262761255\57\0\0\ia0e40c371c743f7efc8fa7062f129ca\16\ e5021f0ca7cba4090390d66ac1c2bd74\262761
```



## Heatmap

The lighter the color, the bigger the correlation magnitude.



```

Sqft <= 1900.5
squared_error = 1401917846944.578
samples = 376
value = 1384825.734
  
```

## Sqft

is the root node.

```
import pandas as pd
from scipy.interpolate._ppoly import evaluate
from sklearn.tree import DecisionTreeRegressor
from sklearn.metrics import mean_absolute_percentage_error
from sklearn.model_selection import train_test_split
from sklearn import tree
import matplotlib.pyplot as plt
import numpy as np
```

DecisionTreeRegressor from scikit-learn to train our model for home price prediction

```
# Load the csv file
main_file_path = 'final_data_test.csv'
data = pd.read_csv(main_file_path)

# Prepare data for analysis
price_variable = data.Sold_price
data_predictor = list(data.dtypes[data.dtypes == "int64"].index)
data_predictor = data_predictor[:-1]
predicting_factors = data[data_predictor]

# Split data in half for training and testing
x_train, x_test, y_train, y_test = train_test_split(predicting_factors, price_variable, test_size=0.2, random_state = 42)
```

Prepare the data for machine learning,  
split the data by 80/20 to generate training data and testing data

- OVERFITTING

When we divide the houses into many leaves, there are also fewer houses on each leaf. Leaves with few houses will make predictions that are very close to the actual value of those houses, but with new data, they may make very unreliable predictions (since each prediction is based on only a few houses).

- UNDERFITTING

If we make the tree very shallow, it will not divide the house into very distinct groups. The final prediction will probably be poor even in the training data (for the same reason it will be poor in validation).

We want to find the sweet spot between underfitting and overfitting.

```
# Function to test the accuracy of the model with different max depth
def test_max_depth():
    for maximum_depth in [2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 50, 100]:
        # Fit the model with the data
        model = DecisionTreeRegressor(max_depth = maximum_depth, random_state = 42, splitter='best')
        model.fit(x_train, y_train)
        # Evaluate model accuracy
        predicted_price = model.predict(x_test)
        print('max_depth:', maximum_depth)
        print('mean_absolute_error:', mean_absolute_error(y_test, predicted_price))
        print('mean_absolute_percentage_error: {:.0.4f}%'.format(mean_absolute_percentage_error(y_test, predicted_price)*100))
        print('Accuracy: {:.0.4f}%'.format(100 - mean_absolute_percentage_error(y_test, predicted_price)*100))
```

Use mean percentage error to find the best max depth

- **Max\_depth**

This indicates how deep the tree can be. The deeper the tree, the more splits it has and it captures more information about the data.

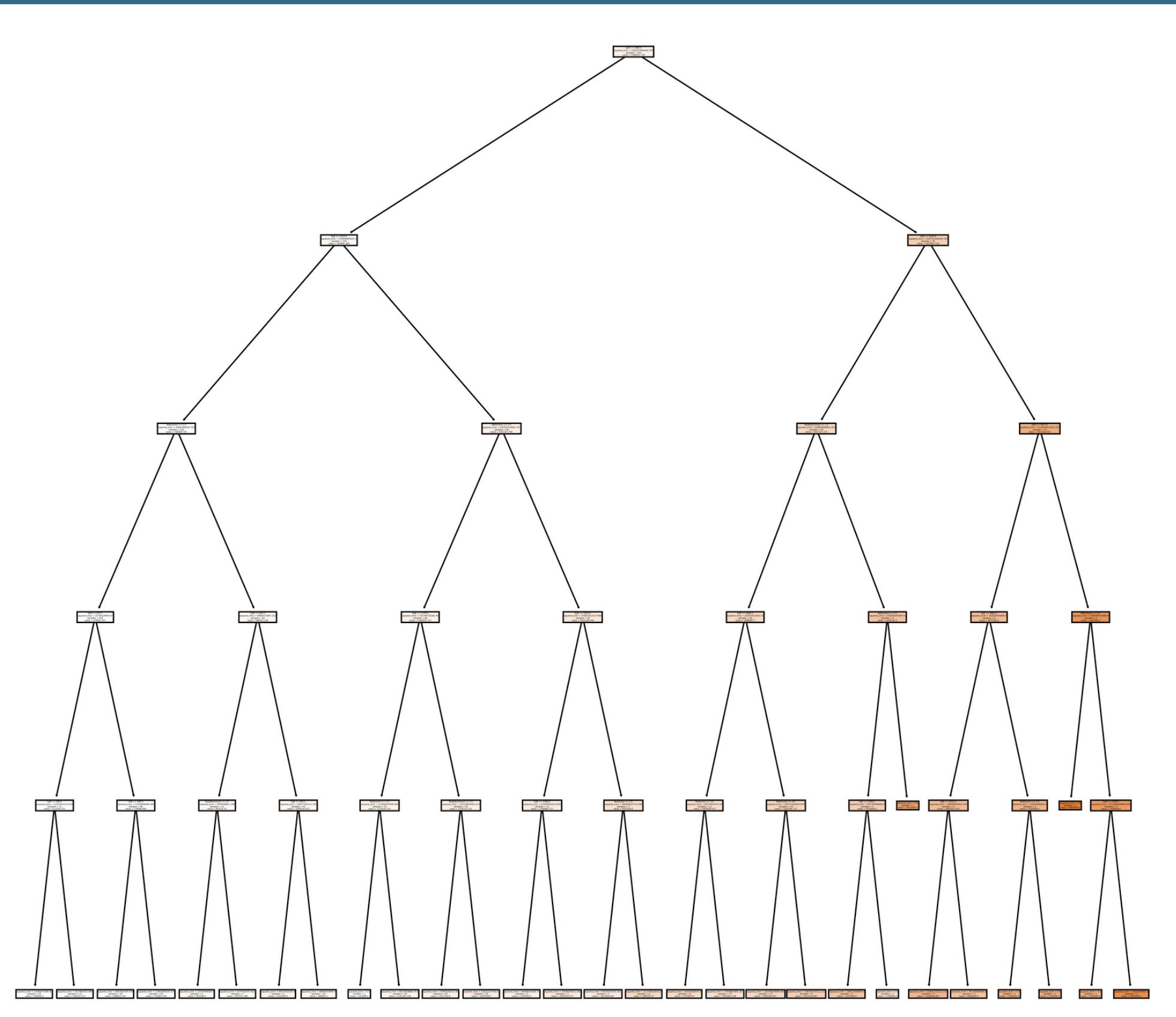
- **Mean percentage error**

The computed average of percentage errors by which forecasts of a model differ from actual values of the quantity being forecast.

max_depth	2	3	4	5	6	7
Accuracy	77.2357%	81.6043%	82.6461%	84.4449%	83.9333%	83.3470%
max_depth	8	9	10	20	50	100
Accuracy	81.1205%	80.1597%	79.2664%	79.2146%	79.2146%	79.2146%

```
# Function to draw the final decision tree
def draw_decision_tree(model_provided, data_predictor_provided):
    fig, axes = plt.subplots(nrows=1, ncols=1, figsize = (15, 15), dpi=500)
    tree.plot_tree(model_provided, feature_names = data_predictor_provided, filled = True)
    fig.savefig('imagename.png')
```

Finally, we draw the decision tree with built in functions



**Best model performance:  
Max\_depth = 5  
Accuracy = 84.4449%**

An aerial photograph of a coastal city at dusk. In the foreground, a large marina is filled with many sailboats and small boats. A prominent suspension bridge with illuminated towers spans the water. The city skyline features numerous skyscrapers and modern apartment buildings, some with lights on. The sky is a mix of blue and pinkish-orange hues, suggesting sunset. A dark, irregular shape containing the text "THANK YOU!" is overlaid in the center.

THANK YOU!