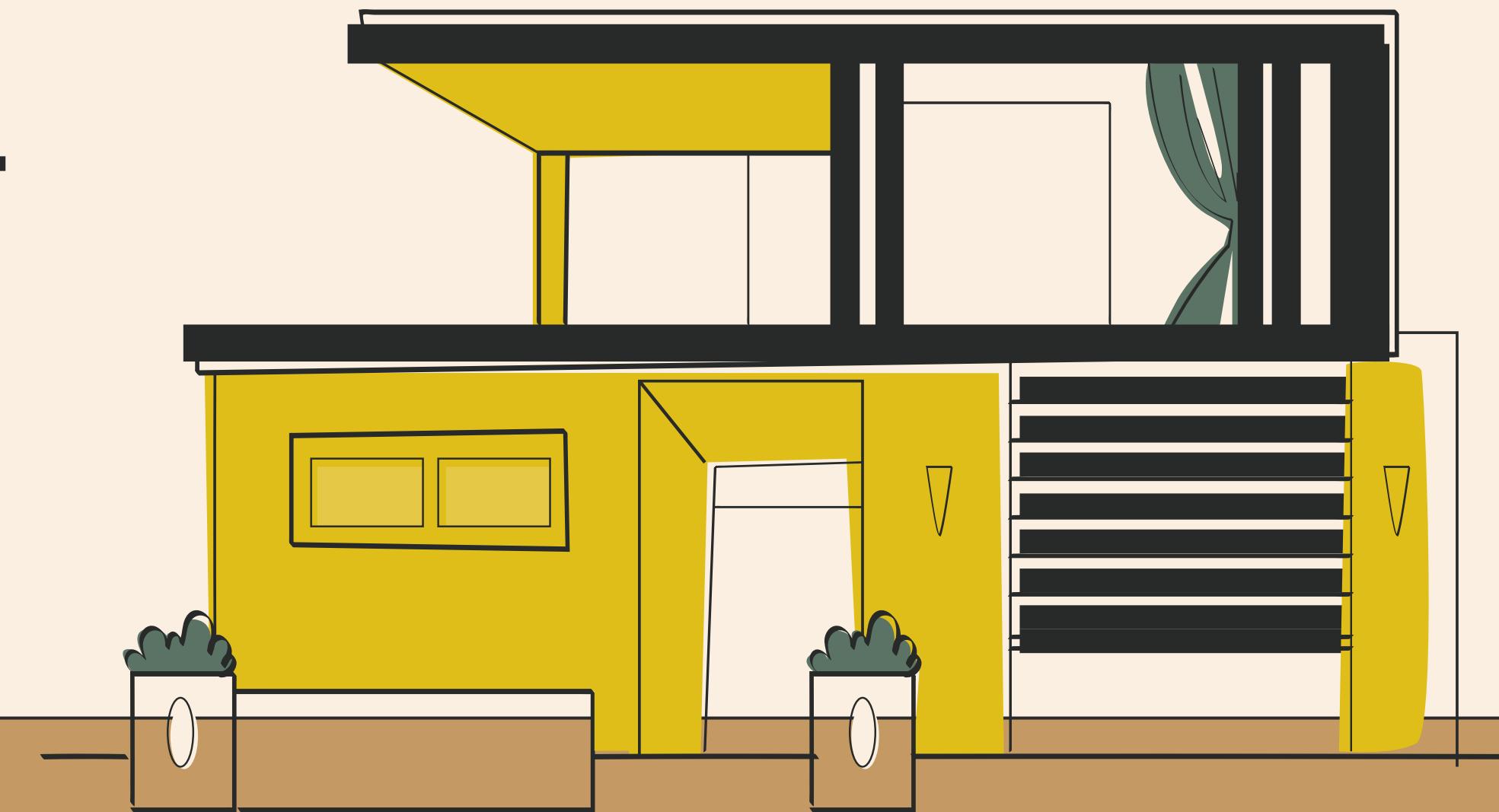




CS5800 Algorithms Final Project Report

Decision Tree Analysis of Vancouver's Real Estate Market



Group Members:

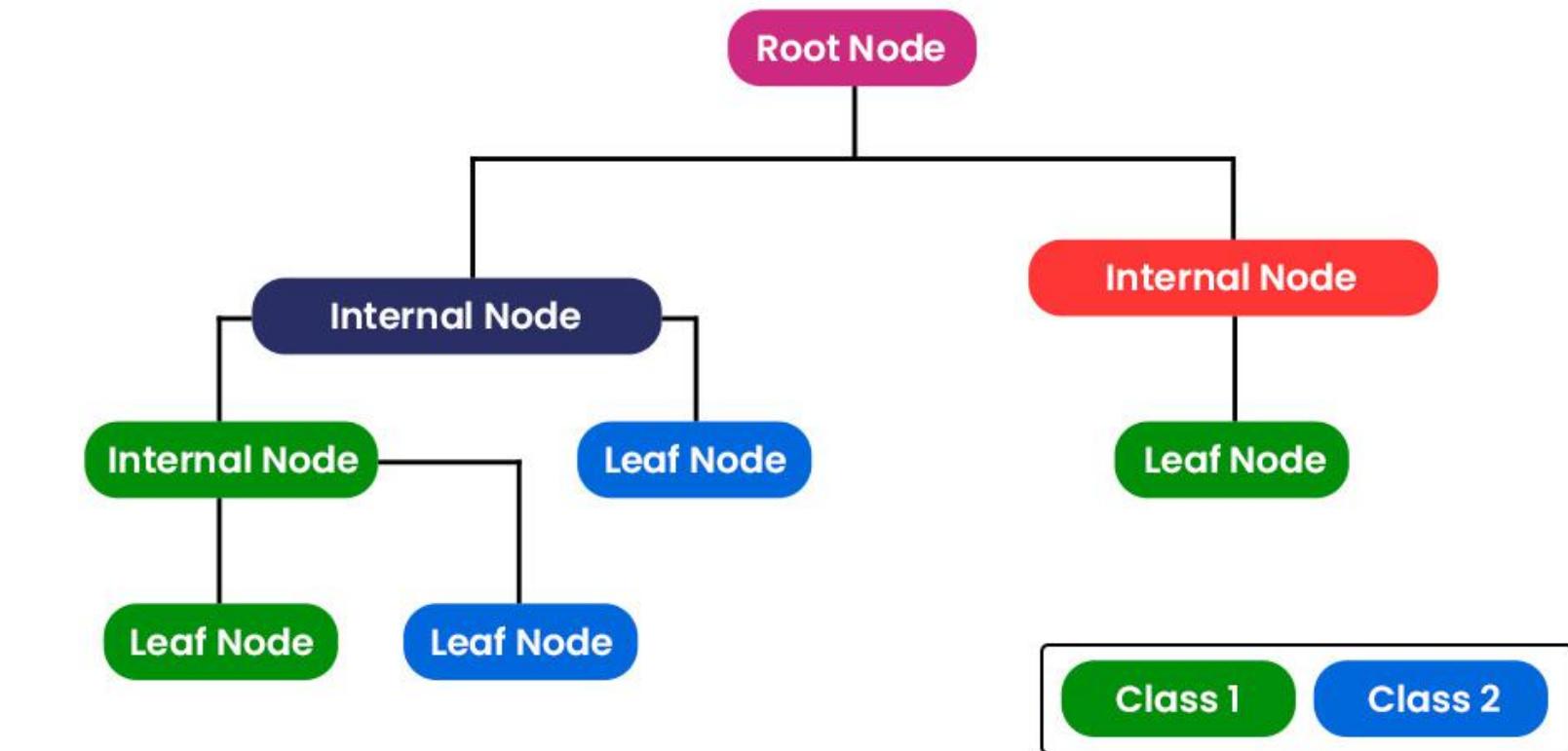
Chen Zhou, Tianzhi Liao, Zhi Liu, Nethra Narayanan

Introduction

Combine binary trees and greedy algorithms to solve real word problems



Decision Tree Algorithm



01

Feature Interpretability:

- Clearly identifies crucial factors affecting house prices.
- Essential for stakeholders to understand valuation drivers.

02

Handling Non-linear Relationships:

- Effectively captures complex relationships between features and prices.
- Ideal for real estate data.

03

Heterogeneous Data Handling:

- Seamlessly processes mixed data types (numerical and categorical).
- Example features: Sqft, Bedrooms, Bathrooms, Neighbourhood, House Type.

Data Preparation and Examination

Titles and Descriptions of the Six Columns in Prepared Dataset

Column No.	Title	Description
1	Neighbourhood	Integer representation: 1 for Vancouver West, 2 for Downtown
2	Type_of_house	Integer representation: 1 for Apartment, 2 for Detached House, 3 for Townhouse
3	Bedrooms	Number of bedrooms in the house
4	Bathrooms	Number of bathrooms in the house
5	Sqft	Size of the house in square feet
6	Sold_price	Sale Price of the house

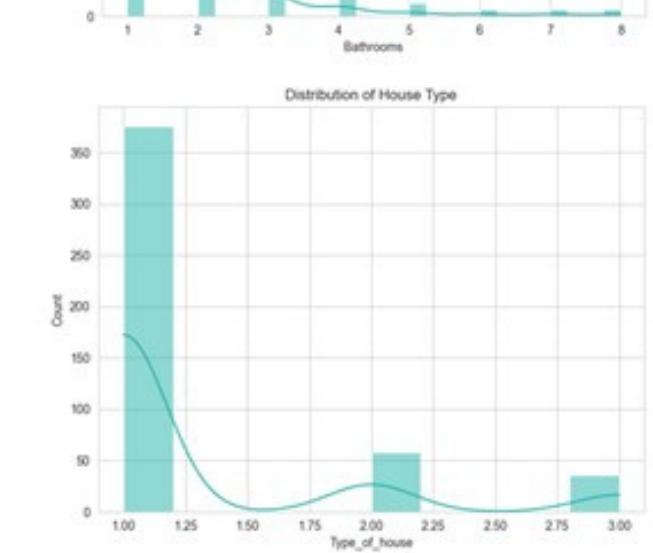
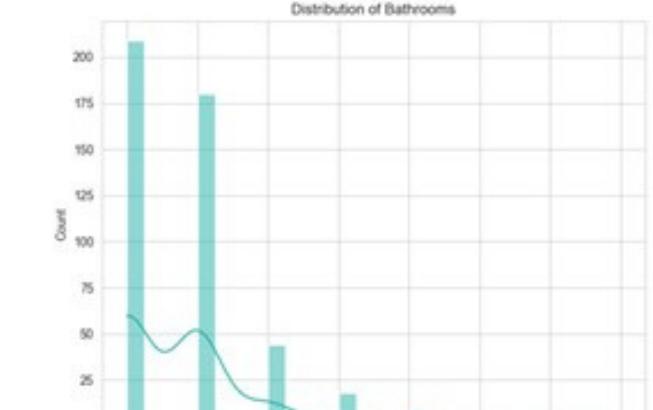
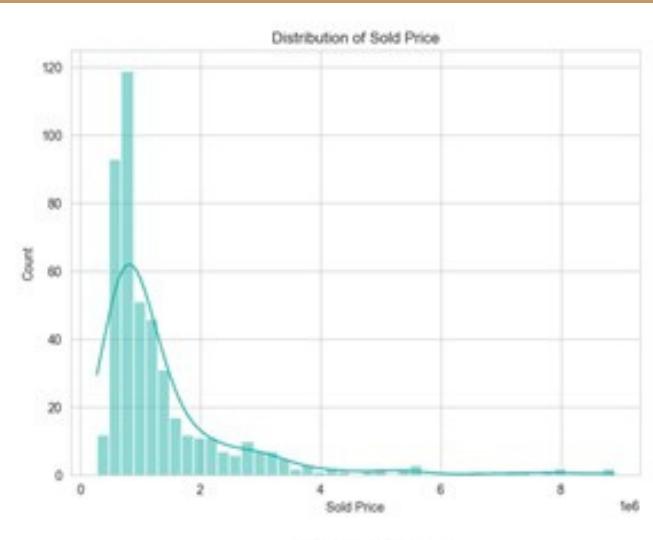
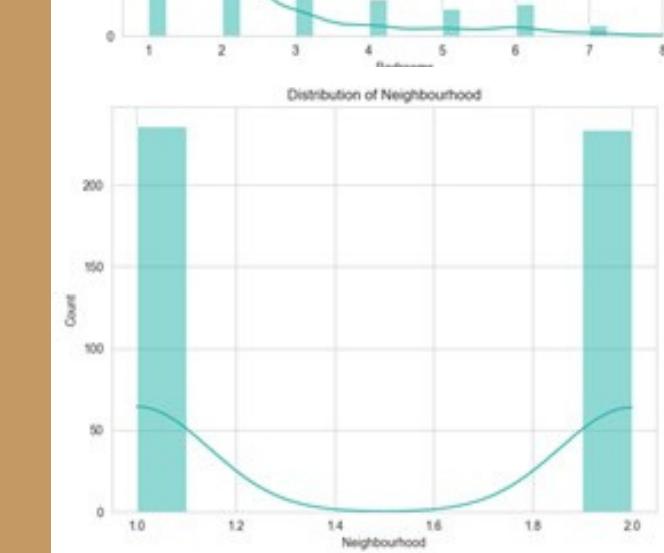
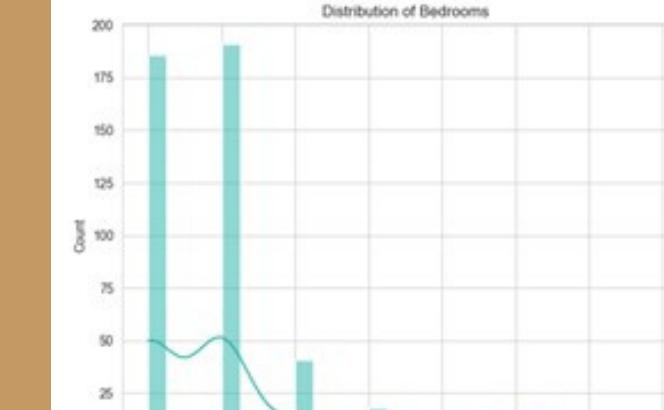
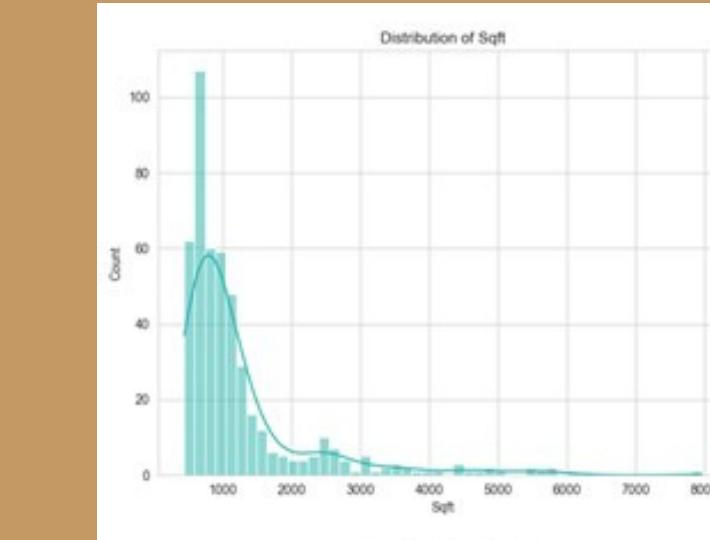
```
{"columns":[],"rows":[[{"R2739628","49.27395","-123.12865","11\22\2022","","# 1701 550 PACIFIC STREET","Yaletown","1249.000","Wow factor! Water view home at Aqua at the Park! Located in the lovely & coveted Beach district neighborhood at the south edge of Yaletown. This seldom available water view 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","2022","0","https://youtu.be/PlWwifdLQ-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","1070","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\ /e5021
```

data
cleaning

	Neighbourhood	Type_of_house	Bedrooms	Bathrooms	Sqft	Sold_price
0	Downtown Area	Apartment	2	2	1070	1230000
1	Vancouver West	House	4	2	1709	1598000
2	Vancouver West	Townhouse	2	3	1266	1220000

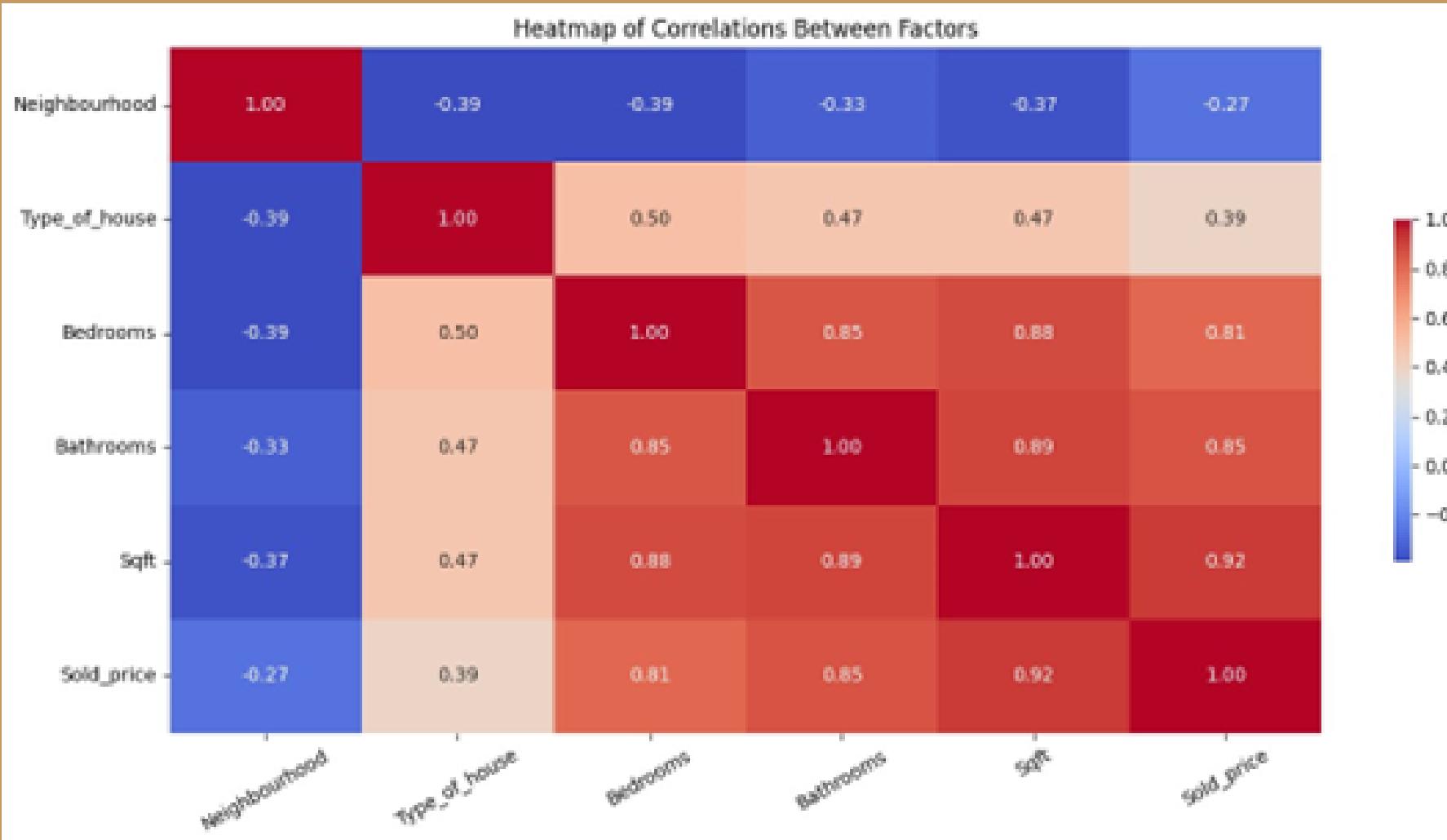
data
processing

	Neighbourhood	Type_of_house	Bedrooms	Bathrooms	Sqft	Sold_price	
0		2	1	2	2	1070	1230000
1		1	2	4	2	1709	1598000
2		1	3	2	3	1266	1220000
count	470.000000	470.000000	470.000000	470.000000	470.000000	4.700000e+02	
mean	1.497872	1.276596	2.044681	1.897872	1236.091489	1.388927e+06	
std	0.500528	0.595009	1.319092	1.213901	1006.818520	1.262850e+06	
min	1.000000	1.000000	1.000000	1.000000	441.000000	2.800000e+05	
25%	1.000000	1.000000	1.000000	1.000000	668.000000	6.950000e+05	
50%	1.000000	1.000000	2.000000	2.000000	899.000000	9.390000e+05	
75%	2.000000	1.000000	2.000000	2.000000	1263.750000	1.483750e+06	
max	2.000000	3.000000	8.000000	8.000000	7955.000000	8.888900e+06	



data
visualization

Histograms of Six Factors in Housing Dataset



Heatmap

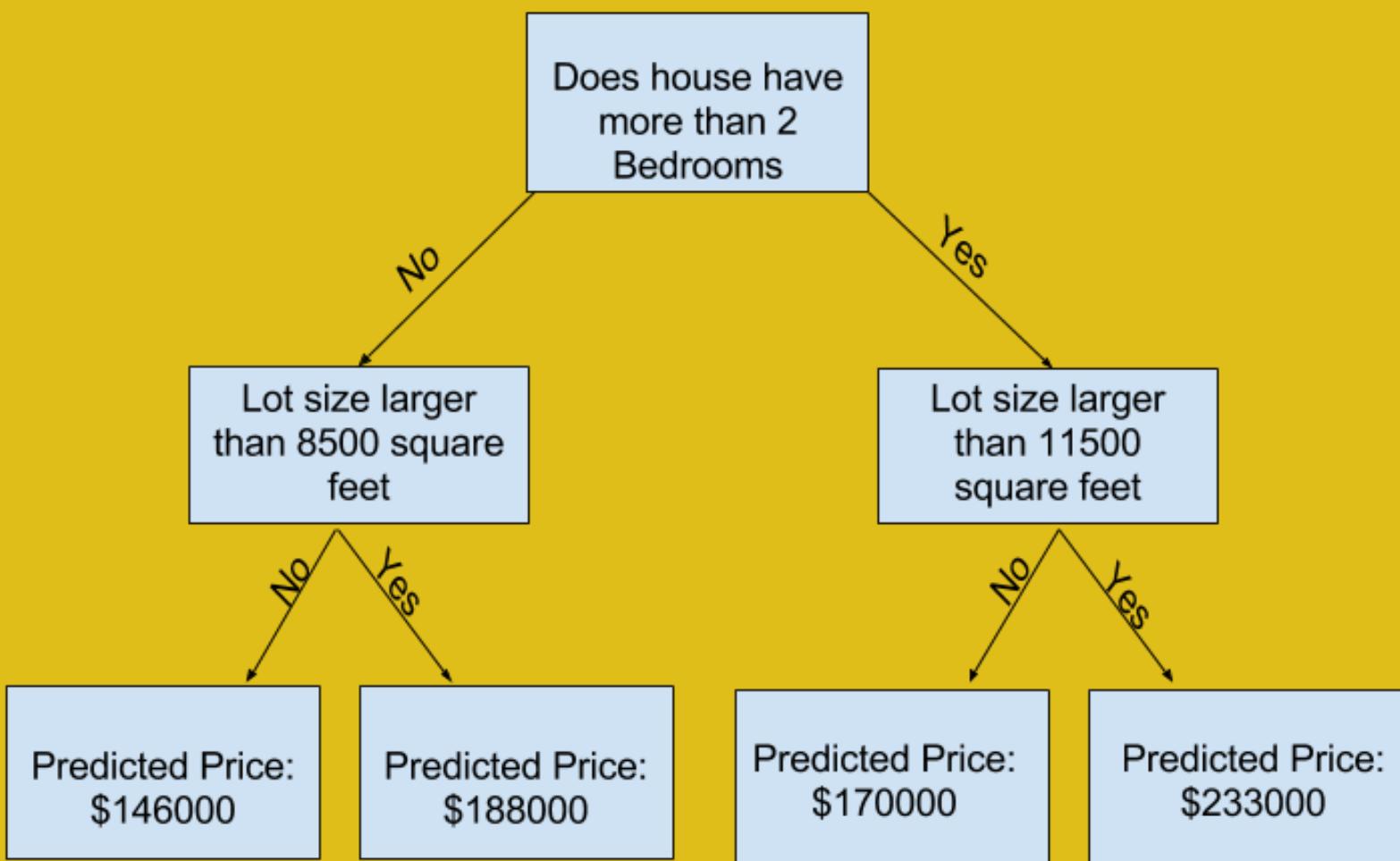


Scatterplot of Sqft vs. Sold Price

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

Sqft is the root node

Algorithm Process



01 Start at the Root

02 Check for Stopping Conditions

03 Feature Selection

04 Data Splitting

05 Iterative Branching

06 Repetition Until Stopping Conditions

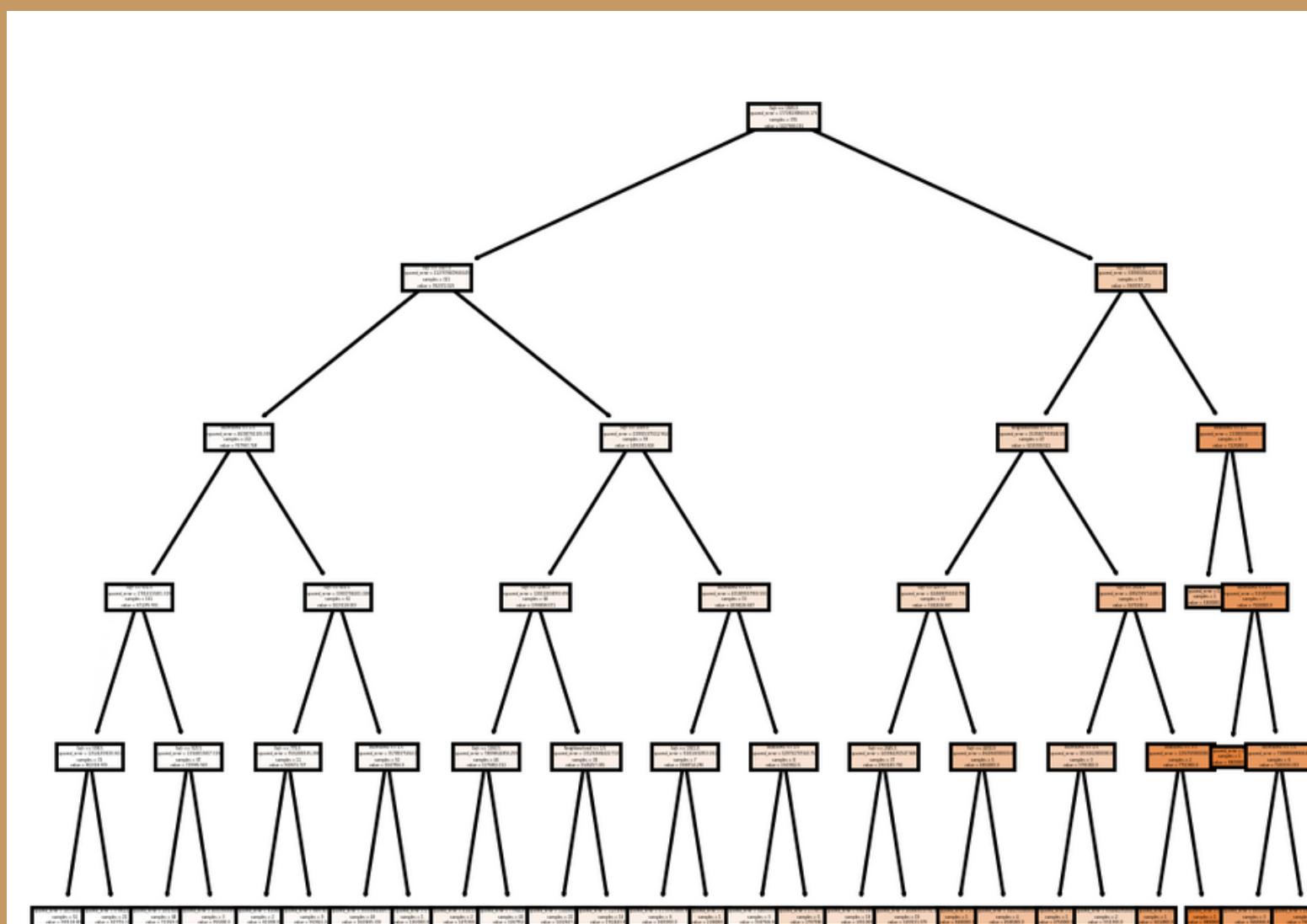
07 Prediction Process

08 Get a Decision

Algorithm Implementation

- Data loading
- Training the model
- Evaluating its performance
- Tuning
- Deploy in Streamlit app

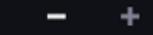
Max Depth	MSE	MAE	MPE (%)	R^2 Score
1	4.102912e+11	473352.569506	42.798050	0.508843
2	2.377993e+11	318882.752837	25.014898	0.715332
3	2.332779e+11	295609.798324	23.716562	0.720744
4	1.037571e+11	221416.602314	21.159266	0.875793
5	1.145356e+11	230373.455929	21.916200	0.862890
6	1.345227e+11	241205.747159	22.294640	0.838964
7	1.578671e+11	259147.979291	22.731553	0.811018
8	1.314571e+11	246417.572699	22.138458	0.842633
9	1.021335e+11	226160.972023	21.579243	0.877736
10	1.634360e+11	263486.184670	24.212617	0.884352



X

Neighbourhood (1 for Vancouver West, 2 for Downtown)

2



Type of House (1 for Apartment, 2 for House, 3 for Townhouse)

3



Number of Bedrooms

3

1

10

Number of Bathrooms

3

1

10

Size of the house (sqft)

2031

500

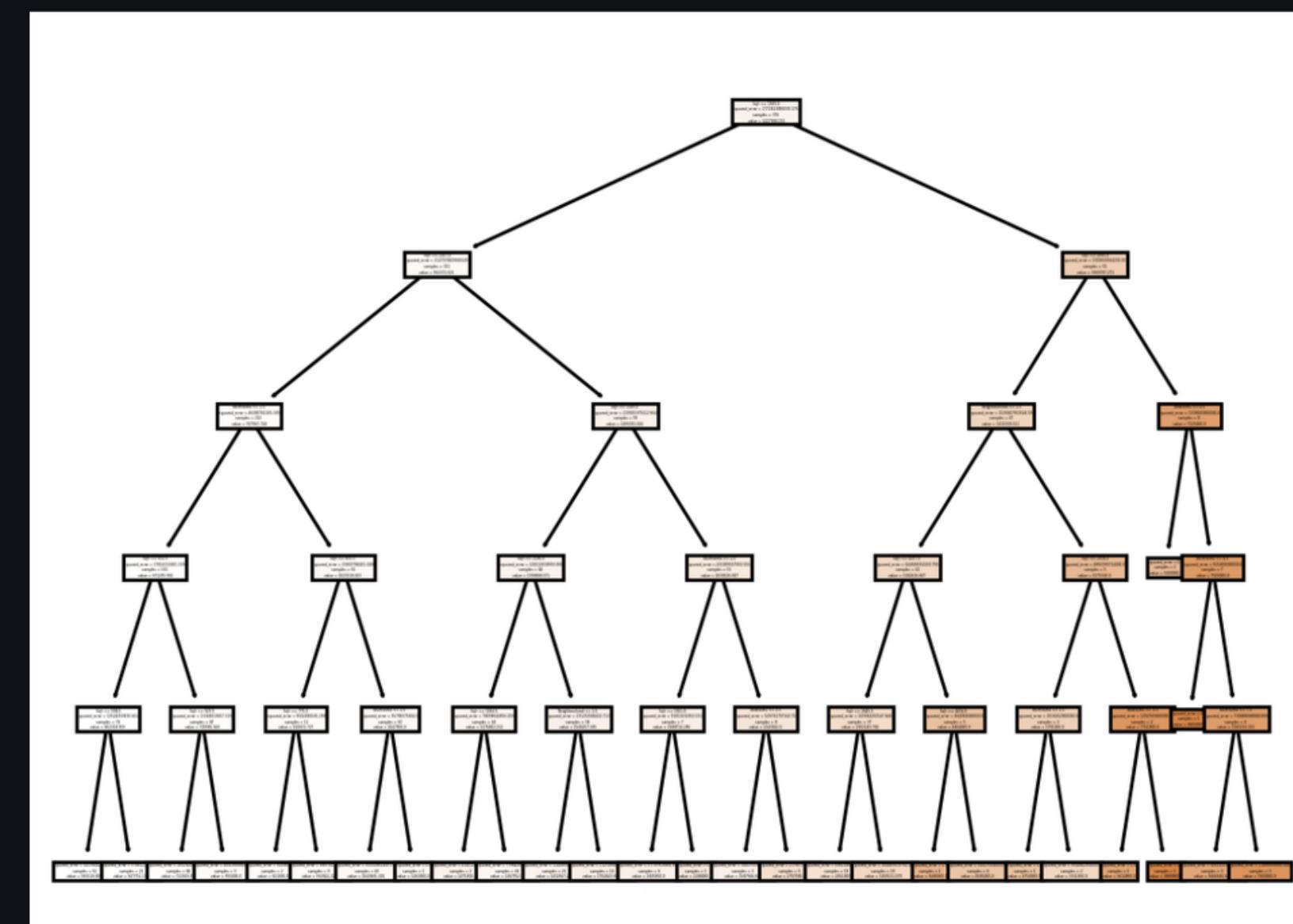
10000

st.cache is deprecated. Please use one of Streamlit's new caching commands, st.cache_data or st.cache_resource.

More information [in our docs](#).

Estimated Home Price in CAD: 3311500.0

Show Decision Tree



Run Streamlit

Conclusion



01

Expand dataset

02

Explore advanced modeling techniques

03

Integrate economic indicators

04

Incorporate a more rigorous analysis

Thank you for listening!

