

# Listings\_recommandation\_sy stem

File created on: 2023-10-13 11:12:19 a.m. EDT

# Lighthouse Labs Capstone Project

## ABOUT ME

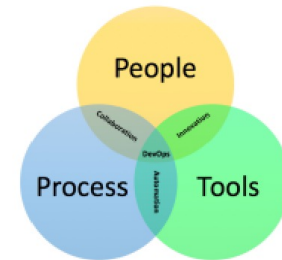


Mechanical engineer  
(24 years)

Business Aircraft  
Interior



Process and Tools  
Efficiency and Quality  
Oriented



Transition to  
Data Analyst



## CUSTOMER

You've got mail: one more listing !!! :)  
You've got mail: one more listing !!! :)  
You've got mail: one more listing !!! :)  
You've got mail: one more listing !!! :)  
You've got mail: one more listing !!! :)

500+

## THE SYSTEM

I HAVE DATA

## BROKER

## I HAVE AN ALGORITHM !

OVERWHELMED  
UNCERTAIN  
DOUBTFUL



UNDER-VALUED      FAIR PRICE      OVER-PRICED

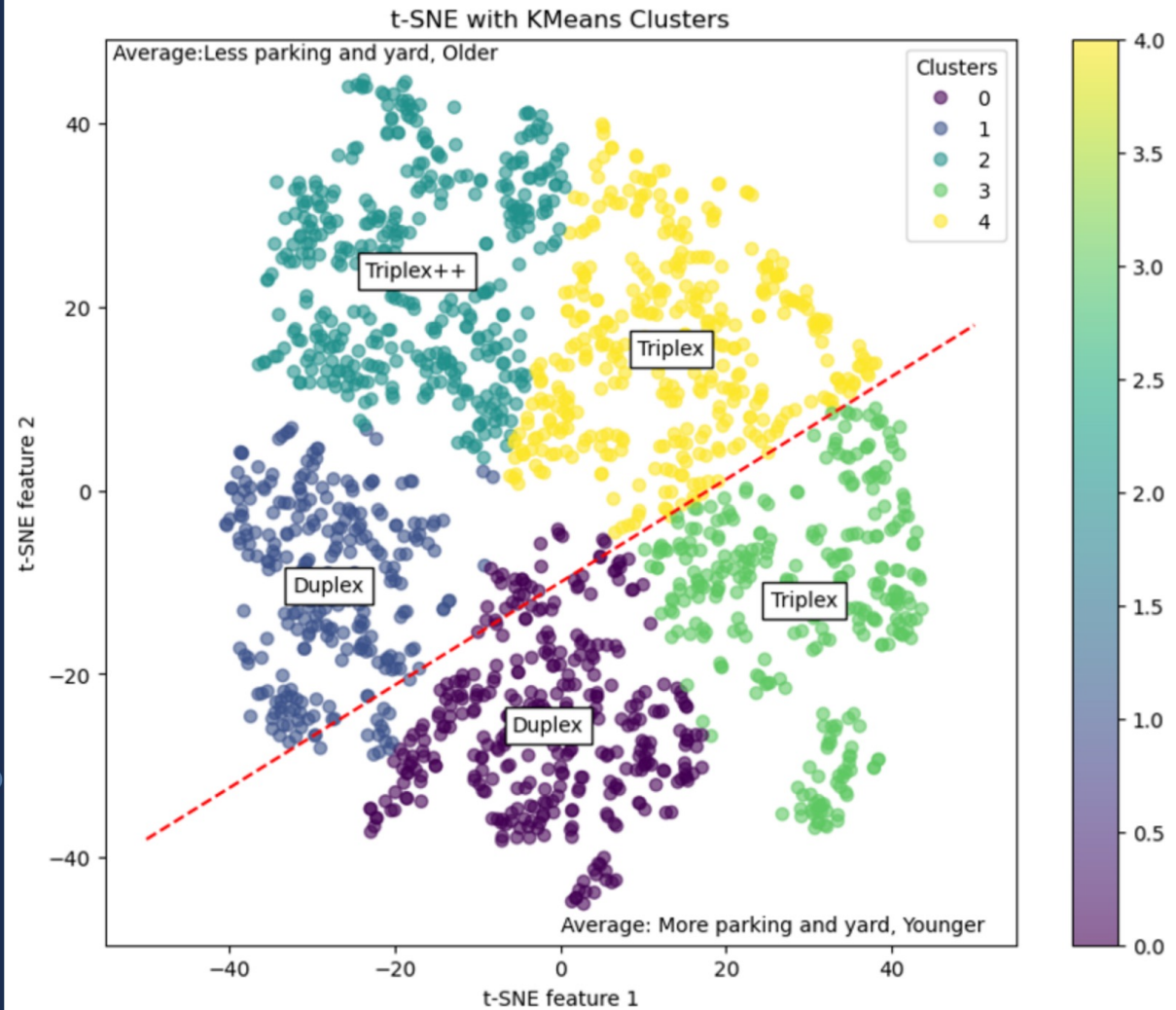
I HAVE REQUIREMENTS

IMPRECISE  
INCOMPLETE  
PRESSED FOR TIME

Welcome to the investment **opportunity of your dreams** or the ideal home for an owner-occupant. This spacious 4-bedroom, 2-bathroom duplex spans two floors and includes a garage, a driveway, and a charming yard. The second floor has been recently renovated and offers a comfortable living space. **Don't miss this incredible opportunity** to own a spacious duplex with excellent income potential. Whether you're looking for a smart investment or a cozy home near a beautiful park, **this property has it all**. Contact us to arrange a tour and discover it for yourself.

## PROCESS CHALLENGES RESULTS

DATA EXTRACTION  
+  
DATA EXPLORATION  
+  
TRAIN A MODEL  
+  
PREDICT PRICES  
+  
RANK HOUSES  
+  
CREATE DASHBOARD



## PROCESS CHALLENGES RESULTS

DATA EXTRACTION  
+  
DATA EXPLORATION  
+  
TRAIN A MODEL  
+  
PREDICT PRICES  
+  
RANK HOUSES  
+  
CREATE DASHBOARD

### CHALLENGES

#### - DATA EXTRACTION:

Dynamic web page issue  
(content of the page changes but not the URL)

50% of the project

### TACKLED WITH:



#### ≡ - FEATURE SELECTIC

✗ Multicollinearity issue

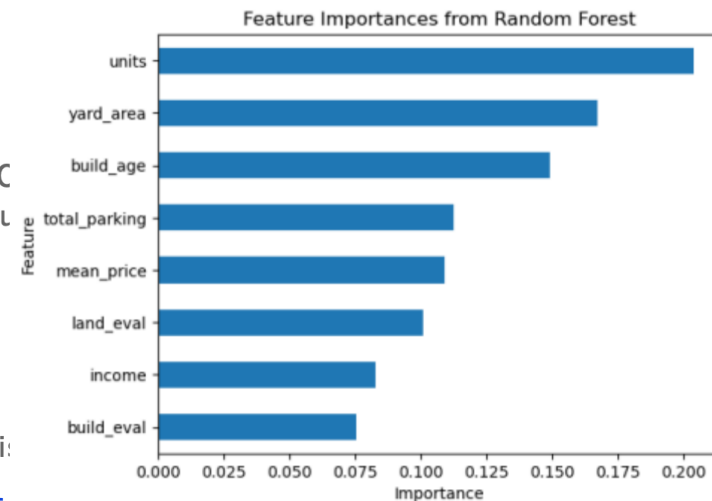
≡

#### - MODELLING:

Sub-optimal model i:

Final predominant features:

- Units
- Yard area
- Building age
- Mean price (District)
- Income
- Building evaluation



LINEAR

REGRESSION:

- OLS, statsmodel
- SCI-KIT learn

## PROCESS CHALLENGES RESULTS

DATA EXTRACTION  
+  
DATA EXPLORATION  
+  
TRAIN A MODEL  
+  
PREDICT PRICES  
+  
RANK HOUSES  
+  
CREATE DASHBOARD

### CHALLENGES

#### - DATA EXTRACTION:

Dynamic web page issue  
(content of the page changes but not the URL)

50% of the project

### TACKLED WITH:



#### ≡ - FEATURE SELECTIC

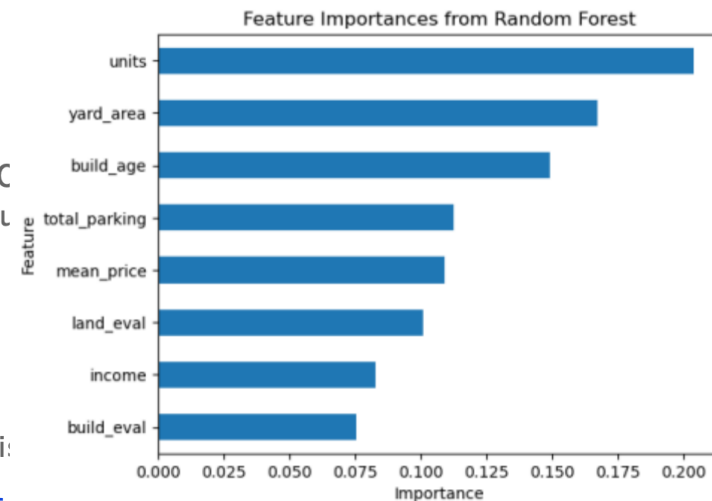
✗ Multicollinearity issue

#### - MODELLING:

Sub-optimal model i

Final predominant features:

- Units
- Yard area
- Building age
- Mean price (District)



LINEAR

REGRESSION:

- OLS, statsmodel
- SCI-KIT learn

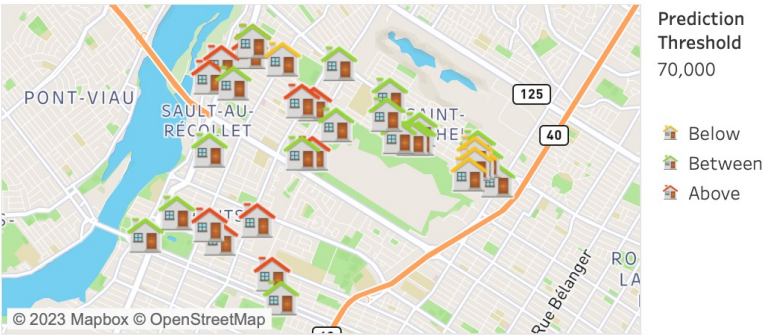
**GOAL:** RANK the listings  
ABS(Actual Price - Predicted Price)



Customer requirements

Units (Duplex, Triplex,...)	Build Type	Price Lower Bound	Price Upper Bound	Income (at least)	Building age (less than)	Yard Area (at least)	His near water	Has a pool	Has a fireplace
2	All	700,000	850,000	30,000	70	2000	False	False	False

Need a Duplex in Ahuntsic near College Regina Assumpta. Not older than 70 years old with at least a revenu of 30k. They require at least 2k square foot of bac..



"THE BARGAIN"  
(hidden gem)

Score	Price	Abs Diff	Income	
22	619,000	84,042	43,800	🏠
	489,000	194,462	32,400	🏠
14	599,900	95,196	37,200	🏠

"THE MAINSTREAM"  
(ideal choice)

Score	Price	Abs Diff	Income	
32	780,000	54,699	50,400	🏠
	799,000	63,869	39,540	🏠
	795,000	64,387	37,200	🏠
	739,000	40,496	33,900	🏠

"THE PREMIUM"  
(aspirational option)

Score	Price	Abs Diff	Income	
32	829,900	100,272	41,400	🏠
	825,000	105,790	33,960	🏠
	849,999	150,204	34,980	🏠
24	849,000	165,496	14,100	🏠

Further Developments

Deep-dive more in Machine Learning:

**Natural Language Processing**

Collaborate with expert in:

**Data science**

**Web development**







**THANK YOU !!!**

<https://github.com/Christian-Leclerc/Capstone-Project-LHL>