



HOUSING PREDICTION MODEL
PREVIEW

A GUIDE TO REAL ESTATE IN TORONTO

Prepared by Arnold Wang

WHAT IS YOUR PROBLEM STATEMENT?

Whether you're looking for a place to settle down, investing in assets for the future or even just looking for a place to live, navigating the real estate market has never been easy to maneuver. That last statement gets amplified with Toronto. With prices soaring and talks of a bubble burst, how can we keep ourselves financially safe in making wise decisions with where we live?

WHAT IS YOUR VALUE ADD?

The AW Model targets and aims to not necessarily make the decisions but guide consumers by providing an accurate yet reasonable estimate via the magics of machine learning.

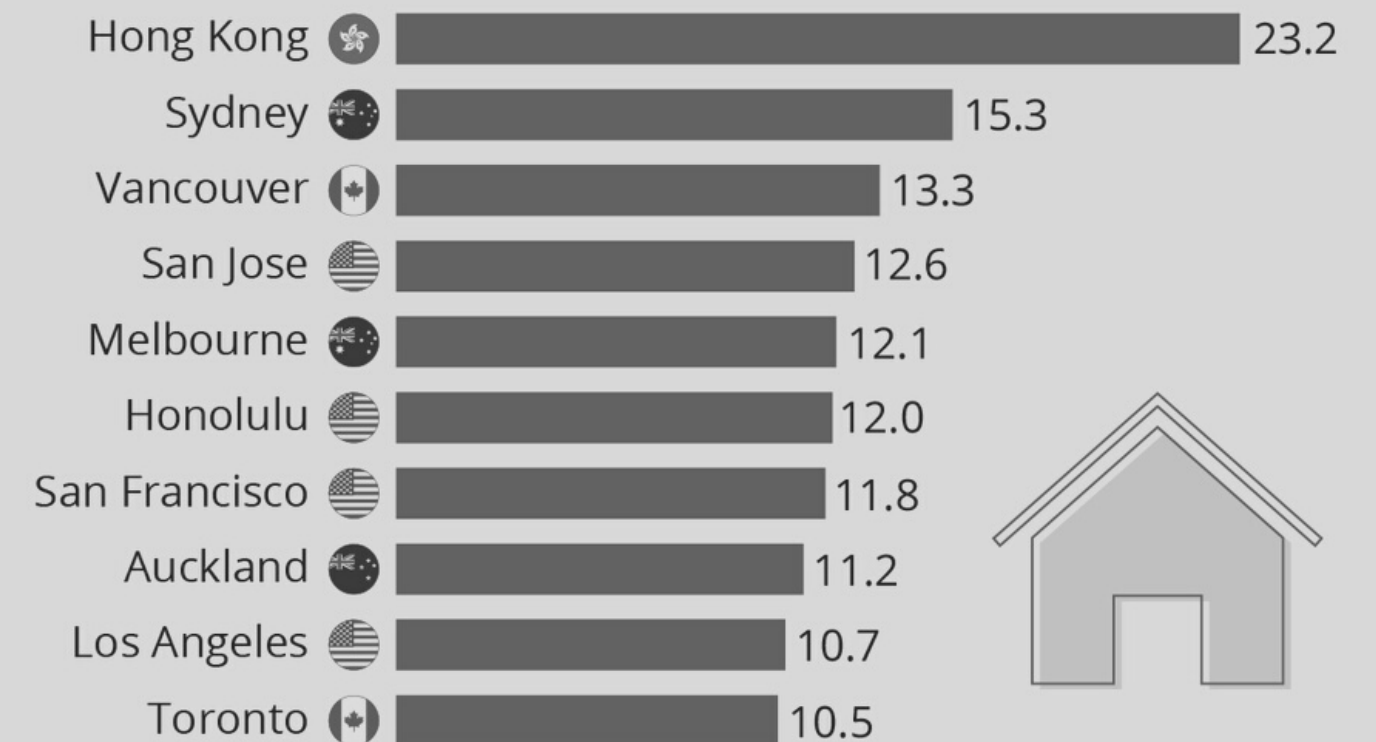
Note: This model IS NOT FINANCIAL ADVICE

Model has a 100% no money back guarantee and serves as a tool as predictions are never absolute



Where It Is Hardest to Afford a Home

Least affordable international housing markets ranked by house price to annual income ratio (2022)*



Data Collection

WHERE DID YOU GET YOUR DATA?

Data was provided via Slava Spirin a former BS student who utilized webscraping on Zoocasa.

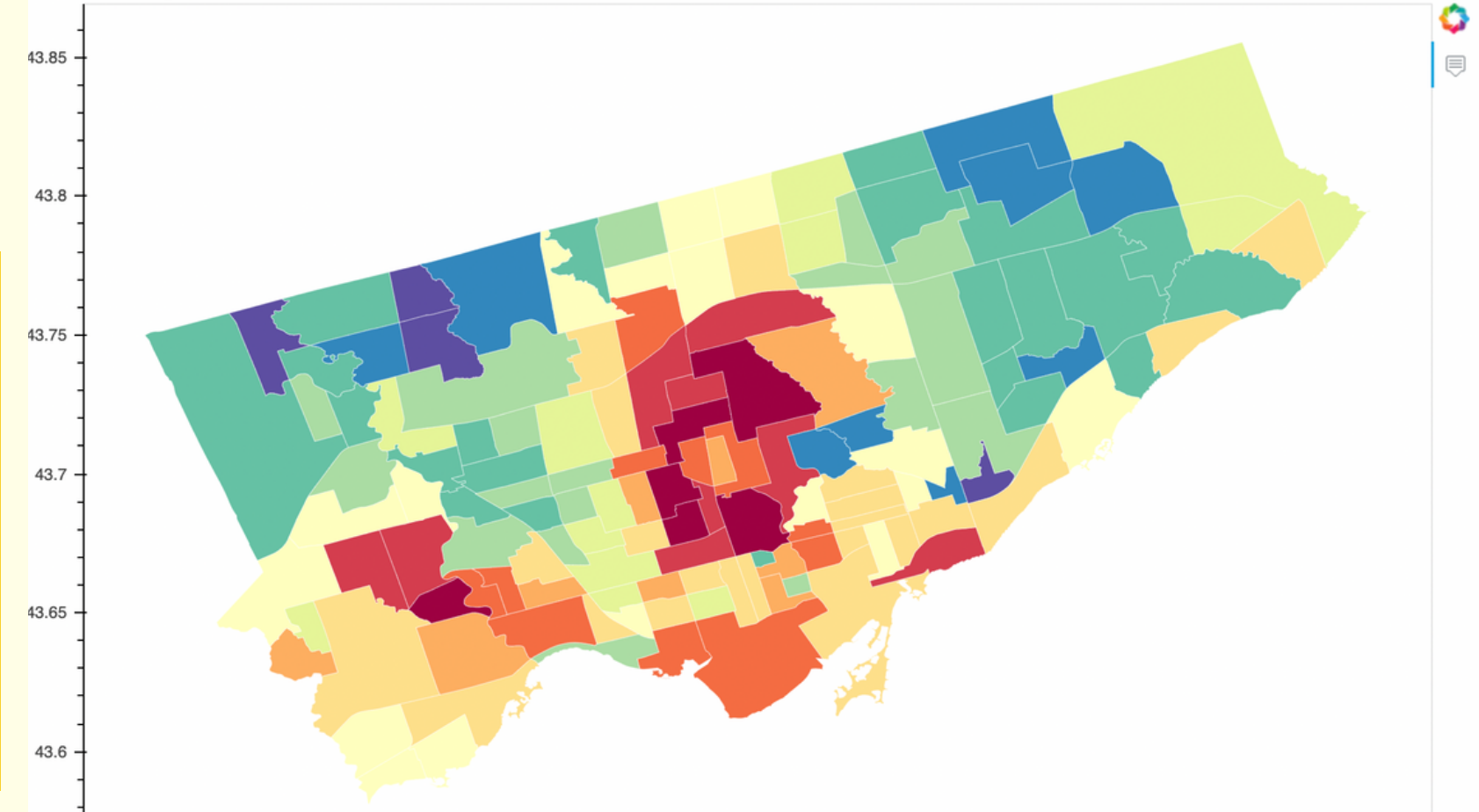
HOW DID YOU COLLECT IT?

Went to the github and 'borrowed' the Dataset

POSSIBLE DRAWBACKS

- Market for homes often vary year over year
- New developments can affect pricing as can demolition of old homes
- No features involving government intervention
- No features involving date

GTA Neighbourhood Income



FUTURE IMPROVEMENTS

- Find sold dates for data for seasonal correlations
- Re-scrape websites such as Zoocasa / Zillow / HouseSigma / Realtor / etc for more updated prices
- Attain older prices possibly via Stats Canada or other Gov websites
- Have new features related to economics (interest rates and stock market pricing)

DATA AND MODELLING

FROM PLANNING TO PROFIT

DATA DESCRIPTION

Each datapoint represents a home that has been listed as sold. Columns are meant to describe the house such as sqft, bedrooms, washrooms, etc

DATA WRANGLING

Created dummy variables and went from 21 columns to 50

CLEAN-UP STEPS

Fill null cells with valid values based on analysis and careful planning.

Check Colinearity among columns removing duplicates.

TARGET

Each model will seek to pinpoint a number that best represents a home's value based on the columns provided

MODELS ATTEMPTED













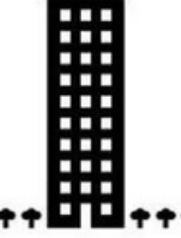







Created a linear regression model with the sold price being the target. Saw strong coef in listed price and sqft. More advanced models will follow upon

Toronto (GTA) Real Estate Market 10 Year Challenge (Avg. Prices)

2011

Presented By:
 HOMING.CA

2021

	 \$450,000 Detached Home		 \$1,240,632 Detached Home
	 \$360,500 Semi-Detached Home		 \$829,437 Semi-Detached Home
	 \$355,000 Row/Town Home (Freehold)		 \$768,474 Row/Town Home (Freehold)
	 \$285,000 Condo Apartment		 \$600,840 Condo Apartment
	 \$277,500 Condo Townhouse		 \$669,204 Condo Townhouse

SETTING EXPECTATIONS

What is your plan for the next two weeks?

- Perform a GridCV on 3 models
- Research more bokeh and plotly visuals
- Implement new features
- Get a job at amazon

Which models do you plan on fitting?

- XGBoost
- LogReg
- Rdm Forest
- *use mean abs error*

THANK YOU!

HERE'S TO TORONTO AND ITS
RIDICULOUS HOUSING MARKET