Boomerang

Ask what you want. We will swift through time to deliver the answers to your fingertips like a boomerang.

Extensive Vancouver House Price Analysis

Joanne Yoon & Hye Lim Moon. April 9th, 2018

Introduction

Should I buy this house?
Is it worth selling my current house?
Is the new house in a good neighbourhood?

Investment decision is a big decision for each **family**. Everyone is interested in making good investment.

In another perspective, **banks** will be interested in this analysis because it helps them to determine if they should provide mortgage to a customer or not. Banks can go broke if they constantly provide mortgages to properties with decreasing values. Also **government** can use this analysis to detect abnormal fluctuations and enforce regulations

Using big data, statistics, and machine learning, we will answer your questions from multiple perspectives .

Methods/Pipeline

Collect

Current Listing: Scrape REW.ca.

Historical Property Value: Collect 10-Year property tax report from City of Vancouver and Surrey Open Data

Schools Rating: Scrape Fraser Institute

Clean Integrate

Explore: Understand the meaning and value of each column in different datasets

Merge: Clean each data to the same format and merge on address

Grow: Fill in missing data from a source using data from another source.

Analyze

History: Group by features and contrast value fluctuation

Listing Price: Compare assessed value with listed price using bootstrap and test their significance

School Rating & Distance: Discover school's impact on property values

Produce

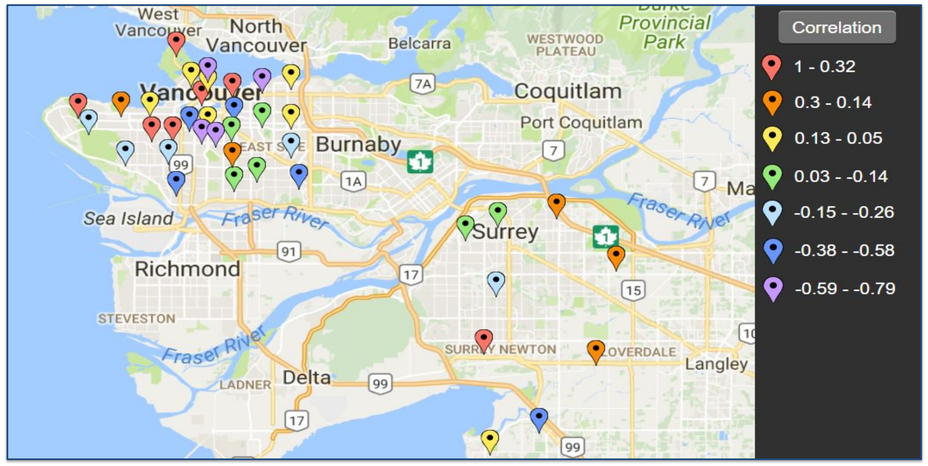
Web Interface: Access from Google cloud platform

All in One: Presents extensive summary

Feature Selection: Get price fluctuations and estimations via Tensorflow

Analysis / Web Deployment

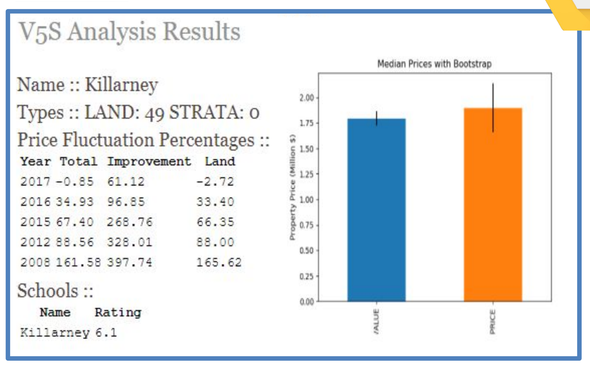
Correlation between Property Value and Distance to Public Secondary School



Google Cloud Platform

We have created a 16.04 Ubuntu instance under the Google cloud platform. It runs jupyter notebook in the background for our web server. Files are transferred using buckets. Static resources(images, csv files) are also uploaded on the cloud.

All-In-One Overview



This overview targets all interest groups to give a general idea of each area in Greater Vancouver. Each area is grouped by the first three digits of postal code. By aggregating our data by area, we can display property value fluctuation, different types of properties, and nearby school ratings

BC PROPERTY REPORT

- Count different property types
- Summarize value fluctuation over 10 years

SCHOOL RATING

- Web scrap distance to schools and their ratings

REW.ca WEB SCRAPE DATA

- Bootstrap to quantify the uncertainty of property values

Evolution of Property Values

USER INPUT

- Enter house attributes and current interest rate

FUTURE

- Predict house prices using DNN (Deep neural network) Regression in Tensorflow
- Integrate machine learning into user-facing serving systems

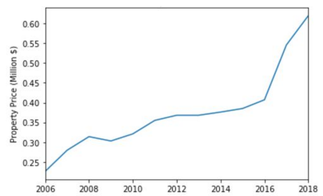
PAST (for Surrey and Vancouver only)

- Integrate 10-Year Property tax report with a realtor site to display price fluctuations over a timeline

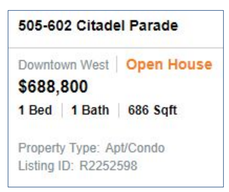
PRESENT

- Crawl the web in real time to show current listings on a realtor site.

Value Fluctuations



Current Listings



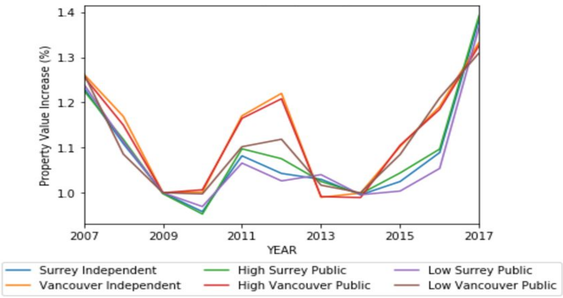
CITY OF VANCOUVER

REW

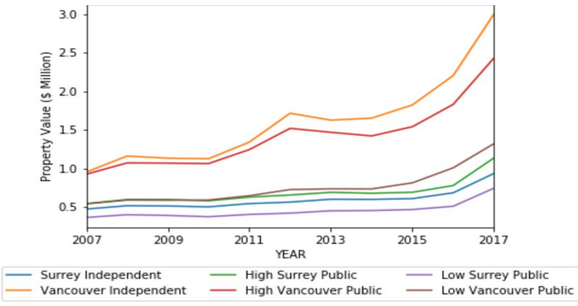
Property Value & School

We only analyzed Vancouver and Surrey properties where we had historical Open Data

Value Increase Depending on School Rating



Value Fluctuation Depending on School Rating



Government Assessed Value Vs. Listing Price

When a property is under the catchment of a school with a higher-than-average rating, the price difference between its assessed value and listing value is significantly lower than that near a lower rating school. This is true for Vancouver and Surrey houses.

Conclusion

SUMMARY

By Web Scraping, we have collected realistic, up-to-data data. Then we merged data from multiple sources, and used machine learning, statistics, and analytics skills to assess the value of each house and area. We displayed our findings on the web using Google Cloud Services. It also includes a prediction tool to estimate a property's future price.

LESSON

Data is money and power. A dataset alone is poor especially in a volatile market. We had to search through many data sources, explore their features, and integrate them to answer vague questions.

FUTURE

If we had more time, we would have continuously scrapped new house listings as they come up in order to richen up our dataset.