

Ryerson University
CKME136 Winter 2017
March 20, 2017.
Interim Report
Name: Leung Ming Kevin TAM
Student ID: 500209198

Data Analysis of Vancouver, BC property tax trend recent years

A research of data analysis will be conducted on the topic of the property taxes of the City of Vancouver, B.C. This research will be looking for the answers of how and what the property tax of Vancouver will be in coming years, which type of properties will have the fastest/slowest growing rate in term of tax. The open data (CSV/XLS files) from the City of Vancouver website will be used and the technique of Classification & Regression and Recommender System with the tools of R will be proposing applied/used.

Literature Review

This project is a research of data analysis for the topic of the property taxes of the City of Vancouver, B.C. A few related papers/articles have been reviewed. Before the implementation of 15 percent property transfer tax, which targeting foreign buyers, by the provincial government on August, 2016, Vancouver enjoyed the strong home sales growth as well as the price of the properties.

The city has been facing redefining changes since past decades as the result of continued wave of immigration from Hong Kong, China and South Asia, the growth and maturation of employment cluster outside the city centre and the redevelopments of downtown (CBD).

The housing markets in Canada became next economic bubble after the dot-com bubbles in 1999 and the financial crisis in 2008-2009. Foreign Direct Investments (FDI) also became one of major source. The buy-low investments was gradually becoming buy-high risks during the last decade.

References:

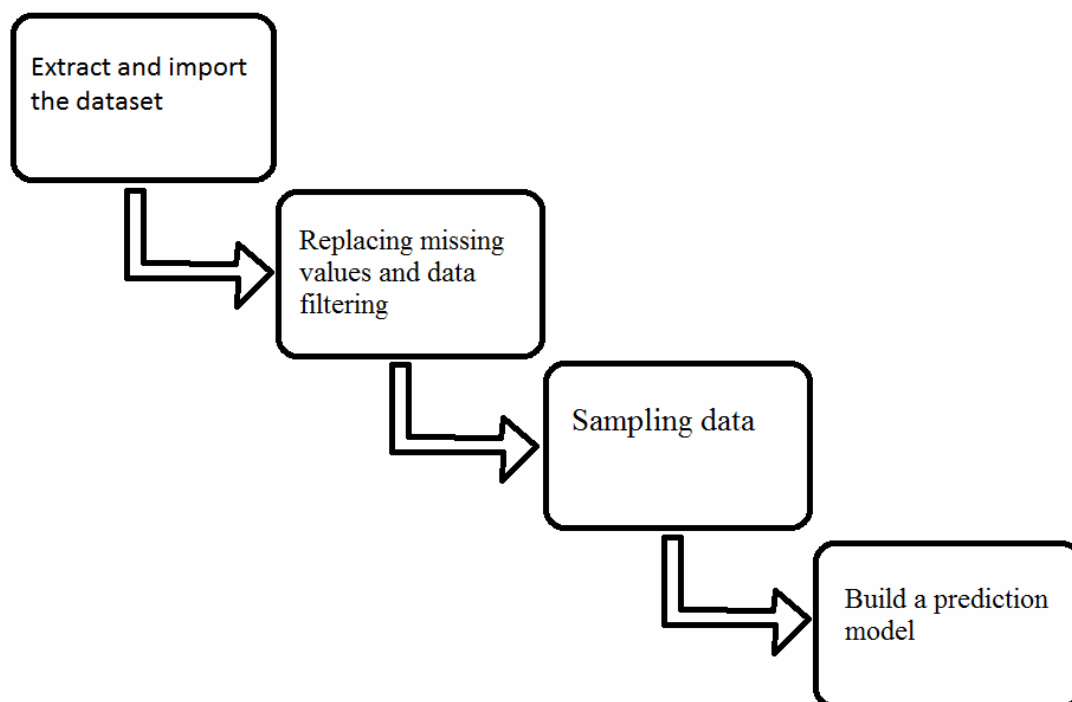
Cultures and Globalization: Cities, Cultural Policy and Governance, Helmut K Anheier, Yudhishtir Raj Isar, 2012, London, UK

When the Bubble Bursts: Surviving the Canadian Real Estate Crash, Hilliard MacBeth 2015, Toronto, Canada

Dataset

As mentioned in the section of abstract, the source of the dataset for this project is the open data from the City of Vancouver. The URL is <http://data.vancouver.ca/datacatalogue/propertyTax.htm>. The dataset contains a number of demographic data in terms of identifier (by authorities), geographic (zone, street, postal code, neighborhood), narrative (description) and numbers (years, values, taxes) . The attributes of ZONE_NAME, ZONE_CATEGORY, PROPERTY_POSTAL_CODE, CURRENT_LAND_VALUE, CURRENT_IMPROVEMENT_VALUE, TAX_ASSESSMENT_YEAR, PREVIOUS_LAND_VALUE, PREVIOUS_IMPROVEMENT_VALUE, YEAR_BUILT and BIG_IMPROVEMENT_YEAR are used for this project, and the attributes of PID, LEGAL_TYPE, FOLIO, LAND_COORDINATE, LOT, BLOCK, PLAN, DISTRICT_LOT, FROM_CIVIC_NUMBER, TO_CIVIC_NUMBER, STREET_NAME, NARRATIVE_LEGAL_LINE1, NARRATIVE_LEGAL_LINE2, NARRATIVE_LEGAL_LINE3, NARRATIVE_LEGAL_LINE4, NARRATIVE_LEGAL_LINE5, TAX_LEVY and NEIGHBOURHOOD_CODE will not be used. At the point to interim of this project, only the attributes of CURRENT_LAND_VALUE and TAX_ASSESSMENT_YEAR are being used due to the inaccessibility to the other attributes.

Approach



Step 1: Extract and import the dataset

Dataset files were downloaded from previously mentioned website, unzipped and imported into R application.

Step 2: Replacing missing values and data filtering

Since, for unknown reason, some records were missing values for some fields especially the attribute of TAX_ASSESSMENT_YEAR. Fortunately, the dataset files were stored yearly, the attribute can be replaced on each dataset. For the rest attributes, records with missing values will be omitted. At the point to interim of this project, records with empty CURRENT_LAND_VALUE will be omitted.

Step 3: Sampling data

Due to the size of migrated datasets (approximately 2.3 million of records), one tenth of the records is used as test set and the rest is used as training set. At the point to interim of this project, since only two attributes are being used for prediction and R can handle such number of records, this step is no longer valid.

Step 4: Build a prediction model

Multivariate linear regression is used to build to prediction model for providing future result of property prices. At the point to interim of this project, first prediction formula has been built as the following:

$$\text{Predicted CURRENT_LAND_VALUE} = -161818411 + (\text{YEAR} * 80907)$$