

Real Estate Prices in Mexico City

KEISUKE YOKOTA, JESUS RAMIREZ, ALEISTER MONTFORT

03-09-18

Contents

1. Motivation
2. Data
3. Methodology
4. Results
5. Questions

Motivation

Develop a tool to estimate average real-estate prices in Mexico City at the neighborhood level, and display it together with other relevant information.

The goal is to :

- Create a database with accurate information on real-estate prices at the neighborhood level
- Integrate real-estate information with other relevant information
- Produce visualizations of the spatial distribution of prices with other variables

Data

The sources of our data are multiple:

- Real-estate prices:
 - It comes mainly from 4 different websites
 - Century 21st, Icasa, Metros Cubicos, Inmuebles 24
 - ~ 80 K raw observations
 - Data was collected between 2nd-3rd week of February
 - ~ 50 K of clean records were collected
 - Challenges
- Official Sources:
 - Office of National Statistics (INEGI) for shapefiles, population, services, neighborhoods shapefile
 - Open Data Sources (CDMX) for earthquake damages
 - Other public sources (for crime)
 - Using shapefiles, we analyzed data on temperature, public services, crime and other demographics; estimated population at the neighborhood level, and normalized data when necessary (crime).

Libraries and methods used

Web scraping using ‘beautiful soup’:

- Real estate prices for houses and apartments were averaged at the neighborhood level (median was used to display information in maps)

Record Linkage using ‘fuzzy wuzzy’ (similar to ‘Jellyfish’)

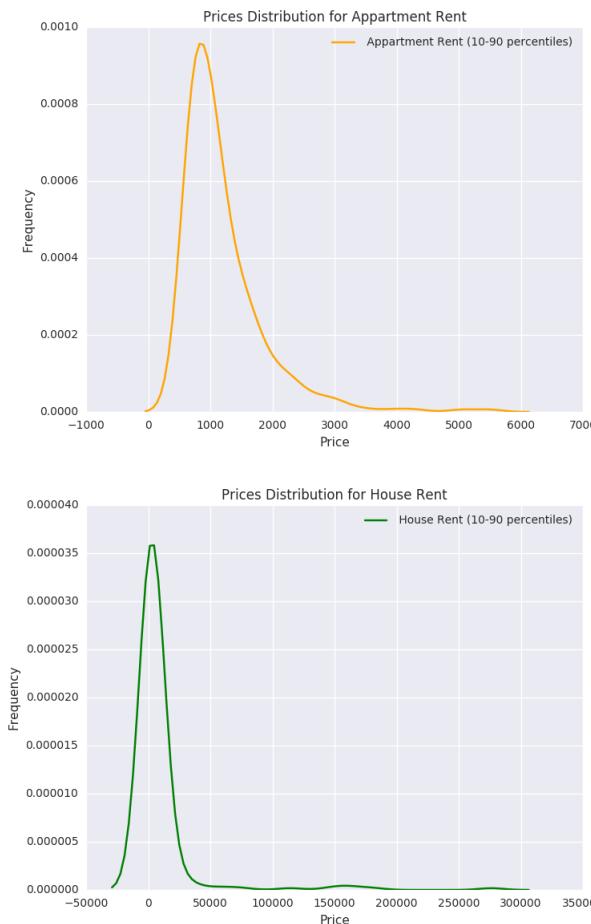
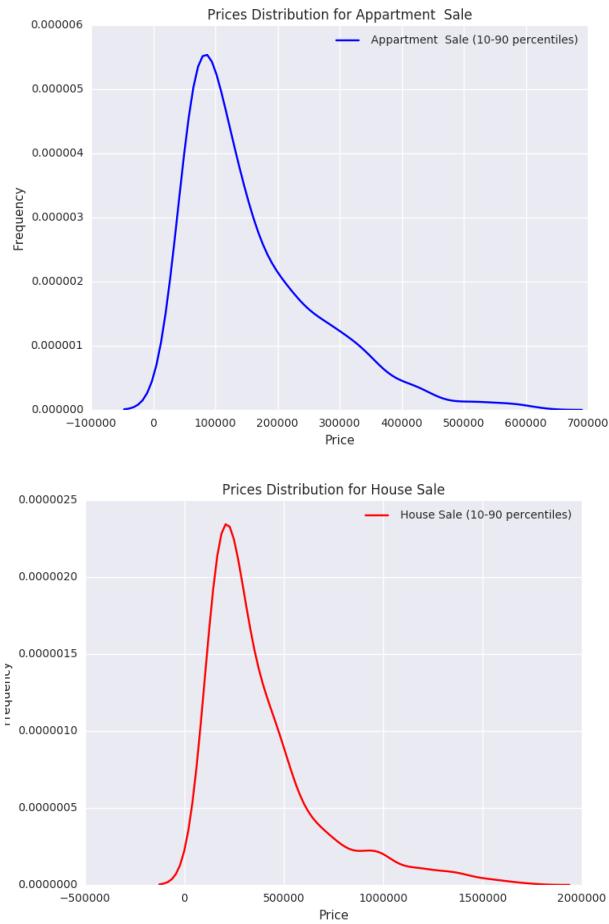
Folium (Leaflet) and GeoJson for creating map visualizations

Geopandas for computing statistics at the neighborhood level

Seaborn for distribution plots

Other common libraries in Python (Pandas, Numpy, etc)

Results (1/2)



	Count	Median	Mean
Apartment Rent	12,137	\$1,382	\$57,409
Apartment Sale	20,784	\$236,660	\$449,358
House Sale	13,625	\$446,334	\$993,024
House Rent	2,716	\$3,719	\$20,656

Results: examples

1) Earthquake Damages and Houses for sale

<file:///User/aleistermontfort/Downloads/'house','%20'sale','%20'%20earthquake'.html>

2) Robberies and Apartments for sale

<file:///Users/aleistermontfort/Downloads/'appartment','%20'sale','%20'Number%20of%20robberies'.html>

3) Number of Schools and Homes for sale

<file:///Users/aleistermontfort/Downloads/'house','%20'rent','%20'Number%20of%20Schools'.html>

Issues in the future

Improve data quality and increase samples (+ web scraping, time series)

- Improve efficiency of web scraping
- Uniform way of collecting data from multiple websites

Different information display at the neighborhood level

- Friendly user
- Use of more buttons
- Markers, selection/combinations of search criteria (SQL)

Increase rate of survival of observations

- More accurate text analysis

Questions

Annex

