Identify The Next Trendy Neighborhood in Chicago Three Years in Advance

Problem Statement

In large cities, old and neglected neighborhoods can quickly transform into popular and trendy neighborhoods with high real estate demand and sharp increases in real estate prices. Some signs that can show a neighborhood is transforming and becoming more popular are: an increase in the number of home renovations and new constructions (older houses and buildings are replaced by new condos and apartments), an increase in its population, an increase in the number of restaurants, coffee shops, and bars in the area, an increase in real estate prices in the adjacent neighborhoods, and a decrease in crime. These are only a few of the factors that show a neighborhood is becoming more trendy and real estate prices are going to rise.

If we detect these neighborhoods a few years in advance, before the prices start to increase drastically, these areas can be good real estate investment opportunities. The goal of this project is to identify the next trendy neighborhood three years in advance. To do this, we will predict the future median housing price in each Chicago zip code and recommend the top ten zip codes which would have the maximum increase in value over the next three years.

Clients

The results of the model would be of interest to real estate investors and future home buyers.

Objectives

- Identify available and relevant data.
- Clean and analyze the data, and engineer new features that can help better predict housing prices.
- Build different models, compare their performance, and ensure that the models perform better than a baseline (dummy) model.
- Provide recommendations to clients.

Data Sources

Only a portion of the mentioned data is publicly available. The data sources below will be used in this project.

1. Historical Housing Price, 1996 to 2020: Information on housing data is obtained from Zillow in csv format. The data includes historical monthly housing price data from 1996 to 2020 for all zip codes located in the United States. The data is seasonally adjusted. The data includes information on 54 Chicago zip codes.

2. Construction and Renovation Permits Data: A building permit is required before beginning most construction, demolition, and repair work in Chicago. Permits issued by the Department of Buildings in the city of Chicago from 2006 to the present and the type of the permit is available from the City of Chicago Website. Here, I retrieve the data using Socrata Open Data API (SODA). I only used the data with permit type that showed more significant alteration such as "New Construction and Renovation" which includes new projects or rehabilitations of existing buildings, "Wrecking/Demolition" which includes private demolition of buildings and other structures, and "Renovation/Alteration". I did not include minor repairs such as electric wiring because that probably is not a sign of change in a neighborhood. The data also includes the latitude and longitude of the construction site for each issued permit which I used to find the zip code.

3. Issued Restaurants license Data: Records of business licenses issued by the Department of Business Affairs and Consumer Protection in the city of Chicago from 2002 to the present are available at Chicago Data Portal. I filtered and retrieved the data using Socrata Open Data API to only include valid business licenses with license code 1006 (Retail Food Establishment). A Retail Food Establishment License is required any time perishable food is prepared, served or sold to the public and includes restaurants, cafés, taverns, grocery stores, convenience stores and more. Increase in the number of restaurants and cafés in a neighborhood can be a good sign and showing that the neighborhood is becoming more popular.

4. Crime data: The crime data was obtained from the City of Chicago Data Portal. This dataset reflects reported incidents of crime (with the exception of murders where data exists for each victim) that occurred in the City of Chicago from 2001 to present, minus the most recent seven days. Data is extracted from the Chicago Police Department's CLEAR (Citizen Law Enforcement Analysis and Reporting) system. The data includes location of the crime (latitude and longitude) and crime type.

5. Zip code related data: Some Zip code information was scraped from CYBO, such as all zip codes' population change from 1975 to 2015, and population change from 2000 to 2015, and neighborhoods close to each zip code.

Deliverables

The deliverables for this project will be Ipython notebooks showing the data collection and wrangling steps, exploratory data analysis, modeling and evaluation as well as reports and slides presenting the project and the final results -- all accessible in a Github repository.