Project Proposal

Team “Something Creative”

Goal: A predictive algorithm for real estate price listings based on neighborhood/district characteristics in Chicago

Final Output: An interactive website where

1) users can find out the average real estate price for a simulated community with living standard variable values that correspond to user’s input
2) Predict future real estate price in specific community (by zipcode and year users specify)

Data Source:

1) City of Chicago - .csv formatted data ready to be used (data on crime rate/”hardship index”/etc.--i.e. Explanatory variables for regression)
2) Trulia - use web-scraping to gather data on real estate listings by neighborhood and year

Required work:

1) Scraping - web-scraping real estate prices from Trulia
2) Modeling - set up a regression model for real estate pricing; check assumptions have been met, ensure model will be statistically correct
3) Visualization - predictions will be shown through an interactive website; creation of such a web page, and necessary visualization of data with Python

Expected Problems

- Missing Data: In Trulia data, some neighborhoods have no history of homes listed. We need to predict such neighborhoods’ home price using home prices in nearby neighborhoods

Tentative Timeline

Jan 23 ~ Jan 29 (4th week): Scraping Trulia’s Chicago Home Prices Data into a csv file/ Proposal Presentation
Jan 30 ~ Feb 05 (5th week): Using scraped data and setting up a regression model, check for validity
Feb 06 ~ Feb 12 (6th week): Implementing model and data into Python program/ Progress Check-in
Feb 13 ~ Feb 19 (7th week): Creating a database with scraped data and predicted values
Feb 20 ~ Feb 26 (8th week): Building a webpage (1)/ Progress Check-in
Feb 27 ~ Mar 03 (9th week): Building a webpage (2)
Mar 04 ~ Mar 10 (10th week): Debugging, quality checks
Mar 11 ~ Mar 14: Wrap-up, Program Submission & Presentations

Responsibilities

Members meet on Monday and Friday at 3PM in Crerar to divide each week’s work by individual, and update the progress.