NH Real Estate (ctd)

Sept. 6, 2017

Announcements

1. Xin has her office hours Sundays 7-9PM. Mark your calendars.

2. Count of who's taking the class. Not everyone has completed the survey, so my headcount from day 1 is outdated. 5 of you are shopping both 530 and 625.

Today

1. Quick check on csv submission formats.

2. We'll spend the time reconciling submissions.

Homework due Next Wednesday (end of class)

This homework will require scraping many more variables than either of the earlier two iterations, so do try to get a head start. In the same groups as your in-class exercise, collectively scrape the following new variables.

- 'address': Owner address
- 'saledate1', 'saleprice1', 'saleowner1', 'saledate2', 'saleprice2', 'saleowner2', etc.: sale history (up to the 5 most recent; most recent first)
- 'sqft': living area in square feet
- 'replcost': replacement cost
- 'pctgood': building percent good
- 'style', 'model', 'grade', 'occupancy', 'actype', 'bedrooms', 'bathrooms, 'halfbaths', 'bathstyle', 'kstyle': style, model, grade, occupancy, AC Type, bedrooms, bathrooms, half baths, bath style, kitchen style
- 'exval': the sum of the value of any extra features
- 'acres': land size in acres
- 'zone': land use zone (e.g. RS1, RS2, . . .)
- 'neighborhood': neighborhood (like 0200, which should not be a number)
- 'landval': appraised value of the land
- 'garagesqft': area of garage in square feet

As you identify edge cases in the process of your scrape, continue to use Piazza to try to reach a consensus. And feel free to add your collective decisions (once you have them) to our pinned post.