Domain

I am going to scrap data from the Zillow.com page. First, I plan to use Selenium to iterate over the houses for sale in Seattle, WA. For each page that pops up after I click on each individual house photo, I consider applying BeautifulSoup to scarp detailed data of each house for analysis. If it is not achievable, I would just download some data from its Zillow.com/research/data session and demonstrate my web scrapping skills in a demo.

I am going to predict the price of a newly listed house based on its features of number of bedrooms, bathrooms, year built, total sqft, lot sqft, zip code, days on Zillow, number of saves, number and rating of nearby schools, price history, tax history etc.

Data

Variable	Type	Description
num_of_bed	int	Number of bedrooms
num_of_bath	float	Number of bathrooms
year_bulit	int	The year it was built
total_sqft	int	Total square feet
lot_sqft	int	Lot square feet
zipcode	str	Zipcode
days_on	int	Days since posted

num_of_saves	int	Number of users who saved it
nearby_school	list	Info about nearby schools including distance and ratings
nearby_transportation	list	Info about nearby transportations
price_history	list	Price history in the past 5 years
tax_history	list	Tax history in the past 5 years

Known Unknowns
- The web scrapping part is more complicated than I expect. I would figure a way out before I run out of time.
- Some metrics need to be defined surrounding the information of nearby schools, transportations, price history and tax history.