

Lab 1: Pandas practice & Data Exploration

Due at 11:59PM 6/13

Notes:

This lab is worth 100 points total. You must show code to back up your results i.e. if you report just a number, you will get zero credit. Submit your code in a Jupyter Notebook. Include your name in the name of the code file. Clearly label each section of your code that responds to a particular question. There should be a labeled section for Question 1, 2, ...etc. For example, you could call the first section "Question 1: Reading in data".

Background:

In this lab, you will practice working with Python's pandas package using a dataset from Zillow.

Problems:

1. Read in the Zillow dataset. Report the number of rows and columns **(10 points)**
2. Report the columns with missing values. **(15 points)**
3. For homes with 3 or more bedrooms, what is the mean number of bathrooms? **(20 points)**
4. Provide a count of records by assessment year **(20 points)**
5. Perform data analysis on the poolcnt variable. What do you notice about this variable? How do you think you should handle this if you were to use poolcnt as a variable in a predictive model? **(25 points)**
6. Calculate the correlation matrix on the dataset. What observations can you make when looking at this matrix? **(10 points)**