Directions: Any problem marked with asterisk (*) should be completed using IPython Notebook (Jupyter) and can be uploaded at https://dropitto.me/m50f15 .

1*. **Real Estate Sales**

Data Description:
The city tax assessor was interested in predicting residential home sales prices in a mid- western city as a function of various characteristics of the home and surrounding property. Data on 522 arms-length transactions were obtained for home sales during the year 2002. Each line of the data set has an identification number and provides information on 12 other variables. The 13 variables are:

Variable Number	Variable Name	Description
1	Identification number	1-522
2	Sales price	Sales price of residence (dollars)
3	Finished square feet	Finished area of residence (square feet)
4	Number of bedrooms	Total number of bedrooms in residence
5	Number of bathrooms	Total number of bathrooms in residence
6	Air conditioning	Presence or absence of air conditioning: 1 if yes; 0 otherwise
7	Garage size	Number of cars that garage will hold
8	Pool	Presence or absence of swimming pool: 1 if yes; 0 otherwise
9	Year built	Year property was originally constructed
10	Quality	Index for quality of construction: 1 indicates high quality; 2 indicates medium quality; 3 indicates low quality
11	Style	Qualitative indicator of architectural style
12	Lot size	Lot size (square feet)
13	Adjacent to highway	Presence or absence of adjacency to highway: 1 if yes; 0 otherwise

Data Source:
https://netfiles.umn.edu/users/nacht001/www/nachtsheim/Kutner/Appendix%20C%20Data%20Sets/APPENC07.txt

Refer to Real estate sales data set described above. Obtain a random sample of 200 cases from the 522 cases in this data set. Using the random sample, build a regression model to predict sales price (Y) as a function of finished square feet (X). The analysis should include an assessment of the degree to which the key regression assumptions are satisfied. If the regression assumptions are not met, include and justify appropriate remedial measures. Use the final mode to predict sales price for two houses that are about to come on the market: the first has $X = 1100$ finished square feet and the second has $X = 4900$ finished square feet. Assess the strengths and weaknesses of the final model.

Reference: *Kutner et. al. "Applied Linear Regression Models", Ed. 5 (problem no 3.31 on pages 153).*