Using the California data set, see slides [10-exploration-notes.ipynb](https://elearning.mines.edu/courses/52392/files/5459246?wrap=1)[Download 10-exploration-notes.ipynb](https://elearning.mines.edu/courses/52392/files/5459246/download?download_frd=1)for an example with a similar data set from Boston Housing Market and [08-scikit-learn-basics.ipynb](https://elearning.mines.edu/courses/52392/files/5459124?wrap=1)[Download 08-scikit-learn-basics.ipynb](https://elearning.mines.edu/courses/52392/files/5459124/download?download_frd=1)for a reminder of these steps and more code examples.

For this activity, you are going to download [california\_regression.ipynb](https://elearning.mines.edu/courses/52392/files/5459247?wrap=1" \o "california_regression.ipynb" \t "_blank)[Download california\_regression.ipynb](https://elearning.mines.edu/courses/52392/files/5459247/download?download_frd=1)and complete the below activity. Make an initial guess or hypothesis about which features best represent the California dataset. You are first going to perform a regression of just a subset of the features you choose, then you are going to perform a regression with all of the features.  The steps for performing a linear regression with sciKit learn are as follows:   
 1. Load Data (Which includes separating the data into X and y)  
 2. Do some initial visualizations (e.g., scatter plots or visualizations with 2 or 3 features or a feature vs. the target, maybe some histograms).  
 3. Split data using train\_test\_split()  
 4. Create lr object  
 5. Train the lr model  
 6. Test the lr model  
   
Note: You do NOT need to do final visualizations like when we used simulated data.

View Rubric