Prerequisites:

Must be within 15 miles of the midpoint of Kamila & my potential job.

Kamila and I must decide on how much we’re going to save towards a house:

* Put down at least 20% of the house so that there are no mortage insurance monthly fees
  + Ex: if we want to save $50,000, we have a max range of houses that cost $250,000.
  + Questions to ask before making a decision:
    - How much will saving $50,000 take?
    - Are we going to do a 30 year or 15-year mortgage?
    - Go on Zillow to get a zestimate of all these expenses.

Will do this by Grabbing a shapefile of Connecticut which will include all towns/cities.

I will then create an intersect (or is it Buffer) that essentially shows a spatial query of all the towns/cities within a specific midpoint

QA: How will I create a midpoint between my and kamilas house??

A couple of things that my project will touch:

* Webscrapping : I will webscrape all of houses on Zillow for sale or that have sold within this buffer. I will use pythons beautiful soup
* Creation of NoSQL and SQL databases.
* I will query information from these databases using python
* EDA: I will expore variables within my data. See what has strong correlations and have an overlay of price for these variables
* Predictive Model:
  + I will talk to kamila and come up with certain attributes in which we would want in a house.
  + We will add these attributes to our model and see how much our ‘dream house’ would be given these variables (regression)
* Classification Model:
  + Like my DS project for MAT 342: I will create a decision tree model (CART/Decision trees/etc..) that wil show me the attributes associated with the price range we want.

Title Page:

"Making Informed Decisions: A Data Science Approach to Home Buying"

Background Information:

Definition of data science

Overview of web scraping

Explanation of SQL and NoSQL databases

Importance of data analysis and visualization in decision making

Overall Objective:

1 slide summarizing the goal of the project:

"The objective of this project is to use data science techniques to gather information about homes available for purchase, and make informed decisions by analyzing and visualizing the data. This project will help prepare the presenter for their BI&A internship at Travelers this summer."

Project Details:

Slides describing specific steps/objectives in the project:

Web scraping data from Zillow's website

Creation of SQL and NoSQL databases

Data pipeline and export

Data analysis and visualization

Making informed decisions based on the analyzed data

Timeline:

A slide indicating the intended milestones and accomplishments each week:

Week ending on February 22nd: Data collection and web scraping

Week ending on March 8th: Database creation

Week ending on March 29th: Data pipeline and export

Week ending on April 12th: Data analysis and visualization

Week ending on April 26th: Informed decision making and presentation preparation

Questions for the Class:

How can I make sure my web scraping is effective and efficient?

What are some recommended methods for visualizing and presenting my findings to non-technical stakeholders?