1. What is the problem you want to solve?

I want to make it easier for the customers to find the house the best house in the market.

What’s your hypothesis? Come up with a hypothesis about which factors matter the most

First, I want to examine the factors (top 5-7) that affect or influence the buying/renting decisions of the pool of customers in the database (factors would be based on data available i.e., location, house characteristics, house description, number of words used, etc.). –come with hypothesis (list of factors)

The next goal is to identify the houses with these characteristics so the company can add a badge like “Zillow’s Choice or Popular Choice” to the listings that meet the criteria.

1. Who is your client and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn’t have otherwise?

This will help the buyers to easily identity the popular choice in the market and they can start checking that house first before looking at other houses. It saves the buyers some browsing time and they will be happy with the company.

1. What data are you going to use for this? How will you acquire this data?

Airbnb (data available) or Zillow (via API).

1. In brief, outline your approach to solving this problem (knowing that this might change later).

- Start by exploring the data (how do you analyze the text?)

- Run regressions and text analysis to check which factors matter.

1. What are your deliverables? Typically, this would include code, along with a paper and/or a slide deck.

Code and Paper