

Questions to consider to help you get started

- **What is the problem you want to solve?**

- Lack of specific current vacancy rate data hampers real estate investors ability to definitively know a key factor in their decision for where to invest.
- Currently many investors rely on local knowledge of an area - ie. local real estate agents, property managers etc. for this knowledge - so the info is hard to get and hard to tell if accurate.

- **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?**

- BiggerPockets - Real Estate Investing Website, can share this data with all their clients and subscribers
- Real estate investors - this would help them to decide which areas to invest in or avoid
- Real estate agents/property managers - this would help them steer their clients towards/away from specific areas to invest or purchase
- Banks - this would help them with lending decisions, especially commercial loans which are based more on the potential profitability of a building vs. the individual applicant's ability to repay the loan
- Insurance Agents - vacancy also is a key indicator of potential vandalism or property theft/damage, knowing this may help insurance agencies adjust their premiums accordingly
- The real estate market at large - knowing vacancy rate would add knowledge that may adjust pricing in some areas

- **What data are you using? How will you acquire the data?**

- American Housing Survey has vacancy rate and population data that can be filtered by zip code, up to the year 2018
- arcgis has unemployment data by zip code by zip code up to the year 2018
- FRED has a lot of current economic data
- I have already found these data sets

Problem Statement Worksheet (Hypothesis Formation)

What is the current vacancy rate in a certain zip code?

H

1 Context

<Why are you working on this problem?>

difficult to find vacancy rate data.

- Investors have to rely on vacancy rate info from talking with property managers, real estate agents in the area.

very important factor for RE investors

2 Criteria for success

<What is the key criteria that will deem this work successful?>

Being able to model what vacancy rates are by zip code for the current year with RMSE values between 0.2 and 0.5 and R-squared value of >75%

3 Scope of solution space

<What is the focus of this business initiative? I.e. What are you specific items will you focus on exclusively?>

US Residential Vacancy rate data by zip code

4 Constraints within solution space

<What constraints exist that may prevent this business initiative from succeeding?>

Only have vacancy rates from 2018 (seems to be a 1-2 year lag in data)

More national economic statistics available, less by zip code

5 Stakeholders to provide key insight

<Who are the key stakeholders that need to be involved in this project? Where will you source your data from and who will you present your recommendation to once you have identified a solution?>

Nadia E - Real Estate Data Blogger

Kevin Glynn - SpringBoard Mentor

Dave Meyer - Data Scientist from BiggerPockets

6 Key data sources

<What are the key pieces of data you need to answer the questions related to the problem you are trying to solve?>

FRED Data for economic indicators

ACH Data for zipcode by data

Zillow data for rent/home prices by zipcode

H

D

E

I

P

Briefly outline how you'll solve this problem.

Your approach may change later, but this is a good first step to get you thinking about a method and solution.

- **What are your deliverables?**

- A GitHub repo containing the work you complete for each step of the project, including:
 - A slide deck
 - A project report