

# Capstone Project Proposal - Carpinteria Coworking Space

## Introduction - What are we trying to solve?

I was speaking to a dear friend recently, and he proposed to me an idea he had for opening a boutique Coworking Space in a very nice work/live loft in the small beachside community of Carpinteria, California. He explained to me that he thought that commuting patterns between the cities of Ventura, California and Santa Barbara, California would play well into having a coworking space situated between the two cities due to existing traffic patterns. Other assets of the area include a beautiful and peaceful beach scene, strong showing of local restaurants and cafes, strong sense of local community, and convenient transportation options. Intrigued, I accepted his offer to go in with him and begin analyzing the feasibility of opening this business as a first step.

I decided that this was a good opportunity to combine real life with the Springboard capstone assignment. As part of my project, I will run an analysis of the local Santa Barbara and Ventura markets to get a sense of our potential customers, if the area has the right demographics to start a local coworking space, and what market areas of Santa Barbara have the highest scores of probable customers that can be targeted for a marketing campaign for our potential business.

## Who am I working for, anyway?

It's nice to know that in this problem, I'm working for #1 - me! This poses some advantages - I can understand my analysis with the kind of granularity where I can instantly answer my level of comfort with certain problems that come up in analysis. I know all assumptions. I don't require meetings with a business sponsor because I am the business sponsor. I do need to deliver status reports to my colleague, but we have a working relationship where trust in each other comes very naturally in terms of division of labor. However, if I were to perform this type of analysis for others, there would likely need to be some level of communication established: obviously timelines would be important, some sort of periodic status report on progress, any follow-up meetings required to answer questions and gauge level of comfort with the client regarding any assumptions or problems in the analysis.

In terms of the outcome of this problem - if we can see evidence that our idea would work successfully, we will continue to prepare to open our business. If we see marginal evidence that our project could be successful, we may re-scope our objective to something smaller. If we see that there's an oversaturated market or else there's not much going for us in terms of a potential customer base, we will get jobs and settle our lives into the 9-5 workplace until we think of another idea.

In addition, instead of doing a more expensive wide-cover marketing campaign, if we can pinpoint sectors of Santa Barbara and Ventura with the most potential for customers based on commute time, job type, age, etc. we can perform geographically targeted marketing. This will likely result in both quicker, more successful, and cheaper marketing.

## What data will I be working with? How will I obtain this data?

In order to perform a successful data analysis, I will primarily be using census data called PUMS data. This data takes a 1% sample of the entire population of the United States and is at the granularity of "person". I will whittle this data down to the Santa Barbara South Coast area, Ventura area, Oxnard area, and I'll throw in Moorpark area for good measure. This data will contain our profession, demographic, age, household data.

Unfortunately this PUMS data does not contain any geographical information on the individuals listed in order to pinpoint commute times, which is likely the biggest factor to consider in determining feasibility. As a result of this, I will be compiling population data on individual census tracts within my market areas, and then I will assign corresponding population ratios to the PUMS data. This should give a rough simulation of geographic distribution. From there, I will include driving, biking, and public transit times from Google Maps API to the PUMS Data in order to get as much of an insight as possible into how our market feasibility is affected by commute times.

I will also need to get information on current Coworking customers. This will be furnished by the 2017 Global Coworking Survey, which shows several demographic tendencies of coworking customers, what fields they are in, commute times, etc. I will also attempt to get granular data on the individual level from the authors of the report in order to combine with my PUMS data to get a more detailed and robust analysis. If I can't get this data, I will use the publicly available report to perform my analysis.

## Approach to solving the problem

The first step is to actually get the data - more difficult than it seems from large data sets on government websites. We will then randomize the assignment of census tract on these individuals in the same ratio as the real population. Once we have the data compiled into one final PUMS data frame, we should have the demographic information, job type information, a simulated geographical location, income information, drive/bike/transit time information for that geographical location all located in one place.

Step two is to create a training and test set. I will likely have to go through my data manually and decide which folks are most likely to be coworking customers based on the information provided by the 2017 Global Coworking Survey. It is from this that I can derive a logistic regression model on my data in order to run it with the other data and provide scores (from 0-1) to every individual of relative likelihood of becoming a customer of our coworking space.

Step three is to figure out a benchmark system to determine if my area is a high-scoring area or a relatively low-scoring area. This has some kinks left to be worked out but may be accomplished by comparing these scores to what are already known as high-potential coworking areas and already known low-potential coworking areas to see how our scores size up.

Step four is to run a total sum score and average score analysis on each census tract in order to determine what, if any, areas are "hot spots" for marketing and bringing in customers. Once we know this, we will move forward with our most aggressive marketing in these areas as part of our Marketing Plan in our preparation phase for opening this business.

There are challenges in each step of this process, but the most challenging in my opinion is determining a proper benchmark against our model to determine if we actually have something here.

## Deliverables of the project

Deliverables of this project include:

- R Markdown Document with scripts to run data analysis and explain process
- Slideshow to present findings
- R Project Code