Milestone Project

Taylor Hughes

4/29/2018

# Milestone Report

## Introduction

My project will attempt to learn what makes a good demographic customer profile for Coworking Spaces and apply that model to the Market Area of Santa Barbara and Ventura Counties in order to determine if that area has good customer potentiality for a new boutique Coworking Space situated in the small beach community of Carpinteria, CA.

## What data is important?

For this project, we use US Census Data of a 1% slice of the Santa Barbara/Ventura areas at the “person” level, which has over 150 different fields per person observed. We needed to whittle that down to only the variables represented by the Global Coworking Survey in order to blend our Census data with constructed Coworker member profiles. I will construct these member profiles by taking the percentages given from the Global Coworking Survey and extrapolating and shuffling them over 100 observations. Please see below the variables utilized from the Census:

* Mode of Transportation: “JWTR”
* Travel Time to Work: We need to create this variable using Google Maps API using “JWAP” and “PUMA”
* Age: “AGEP”
* Job Type: “COW”
* Occupation: “OCCP”
* Gender: “SEX”
* Marital Status: “MAR”
* Number of Children: Unfortunately this data is not available.
* Health Insurance Status: “HICOV”
* Current Workspace: “JWTR” contains Home Office information
* Education information: “SCHL”
* Income Information: “PINCP” adjusted by “ADJINC”

Please read the full paper to fully understand the formats and data transformations required to blend the Census data with the constructed Coworker Member profiles.

## Problems with the Project

### Geographical Problems for Commute Time

The largest problem to solve with this data set is how to find a connection between Commute times in the Market Area to the proposed Coworking location and the Commute Times to Coworking Spaces in the training data, which is clearly not based in Santa Barbara. A major limitation on this was that the US Census did not provide any information on where any of the respondents lived. Another limitation was that Google Maps API only allows 2,500 API Calls per day. How we solved these problems:

* Market Data
* Pulled information about the Population, Average Income, and Geographic Coordinates of the centroid point of every Census Tract in Santa Barbara and Ventura Counties.
* Calculated driving, biking, walking, transit times between every Census Tract centroid point and the proposed Coworking Location using Google Maps API and CRAN Package “gmapsdistance”.
* Created an algorithm that assigns every person in the US Census for SB/Ventura Counties to the Census Tract with the closest average income to their total income. Once a Census Tract is proportionally full relative to the rest of the Market Area (as dictated by the real-life proportions of Census Tract Population to County Population), it is removed from consideration and people will be assigned to the next closest one. That way we get a real distribution of Census Tract Population.
* Based off of Transportation Mode, observations were assigned Commute Times to the proposed location based off of their assigned Census Tract.
* If people lived within a 20-minute Commute Time, they were assigned a value of “Yes”. Otherwise they were assigned a value of “No”.
* Training Data
* Pulled information about the Population, Average Income, and Geographic Coordinates of the centroid point of every Census Tract in Los Angeles County.
* Pulled information, partnering with Coworker.com and their API team, about the location of every Coworking Space within a 150 kilometer radius of the center point of Los Angeles County.
* Since there are 2,307 census tracts in Los Angeles County and we can only call Google Maps API 2,500 times per day, I created an algorithm which calculates the closest Coworking Space to the center point of all 2,307 census tracts in Los Angeles County. This was achieved by creating vectors for all census tracts of the euclidean distance between its center point and every Coworking space. I then used the function “which.min” to identify the closest geographical Coworking Space to each Census Tract, assuming that it would at least be close to the shortest driving distance.
* Calculated driving times between every Census Tract centroid point and the geographically closest identified Coworking Location using Google Maps API and CRAN Package “gmapsdistance”.
* Created an algorithm that assigns every person in the US Census for Los Angeles County to the Census Tract with the closest average income to their total income. Once a Census Tract is proportionally full relative to the rest of the Area (as dictated by the real-life proportions of Census Tract Population to County Population), it is removed from consideration and people will be assigned to the next closest one. That way we get a real distribution of Census Tract Population.
* Observations were assigned Commute Times to the proposed location based off of their assigned Census Tract.
* If people lived within a 20-minute Commute Time, they were assigned a value of “Yes”. Otherwise they were assigned a value of “No”.

### Linear Regression Scores Problem

Another problem that was observed