Daniel Reiff
Capstone 1 Proposal
9/5/19

How can I improve the recreation.gov product through a demographic analysis of their customers?

To answer this question, I will analyze recreation.gov reservation data from 2018 (1,048,575 total reservations) and 2017 census data. Specifically, I will relate both databases using the customer's zip code. My plan is to start with one campsite and create a census info database based on each customer's zip code. I envision having a database that looks roughly like this:

HistoricalReservationID	FacilityID	Park	FacilityState	FacilityLongi	FacilityLatitu	CustomerZIP	CustomerLat	CustomerLng	Census Tract	County	State	Total households	Households with 65 years and over	Females 15 years and over
2976269875	251937	Klondike Gol	AK	-135.31389	59.4583333	32344	30.4	-83.9	2502	Jefferson Co	Florida	1903	785	178
2976269875	251937	Klondike Gol	AK	-135.31389	59.4583333	92064								
2976269875	251937	Klondike Gol	AK	-135.31389	59.4583333	98006								
2976269875	251937	Klondike Gol	AK	-135.31389	59.4583333	48118								

Campsite Data
Customer Data
Census Data

I have access to over 1300 census variables. I foresee a couple of strategies with their usage:

- I can use all variables and select ones that are extreme for each customer zip (vs. national average) and compare among all customer zips.
- I can use all variables and select ones that have similar profiles (for example % of households with 65 or older residents) among all the customer zips
- I can select variables that I think would correlate with national park use, select them for each customer zip and compare among all customer zips.

I also have coordinate data for both the camp site and customer location, with which I will create some type of visual map to show customer routes. Once, I have determined my census variable usage strategy and a created route map for one campsite, I will work on the repeatability of the process, and repeat the exercise for more campsites.

Overall, the reservation data is very clean with no Nan values for customer zip codes. However, I will have to eliminate international customers.

MVP: I hope to create a pipeline that creates database from unstructured data (census api) and ties it with structured data (reservation data). With this pipeline, I will create a route map for one campsite and identify common variable profiles among the site's customers.

(Maximum)VP: Use the pipeline to identify common variable profiles among all sites in Colorado. Tie the customer route map to transportation routes (air, road, rail, etc.)