Child Care Deserts and Venue Data in the State of Colorado

Michelle Salvador

The problems Child Care Deserts (CCD) present

- Over 75% of mothers and 50% of fathers all over the United States had to decline a work opportunity and change or quit a job because of the scarcity of child care or paid leave.

- A report by the New America think tank found that child care costs on average $9,589 - more than the average in-state college tuition of $9,410!

- For those who live in CCDs this means having to travel much longer distances to find childcare they can afford on top of the struggle to find any at all.

Project Goal Description

- The aim of this project is to study the CCD found in Colorado and how they relate to the types of businesses that are present in the corresponding ZIP code.

- Learning more about the characteristics of CCD when it comes to why this highly demanded business is not thriving can be of interest to parents, childcare business providers, and policy makers wishing to improve the circumstances.

- I choose Colorado since this is the state I reside in and according to a CAP report [5], 45% of residents live in a CCD using 2014 census estimate data, by the ZIP code definition.

Using foursquare, census data, and a childcare database provided by the Colorado government I wish to explore the following:

1. Are there certain types of businesses that occur with higher frequency in Child Care Deserts? If so, what type?

2. Is there a correlation with low business density to CCD?

3. Can a model be created using foursquare business data to predict if a census tract will be a CCD?

Child Care Deserts (CCD): A Definition

According to the Center for American Progress (CAP) a Child Care Desert (CCD) [5] is a ZIP code defined as having both the following characteristics:

1. At least 30 children under the age of 5

   AND

2. 2. Has either no child care centers or so few centers that there are more than three times as many children under age 5 as there are spaces in centers.

Data Sources:

1. Census Zip Codes in Colorado 2017

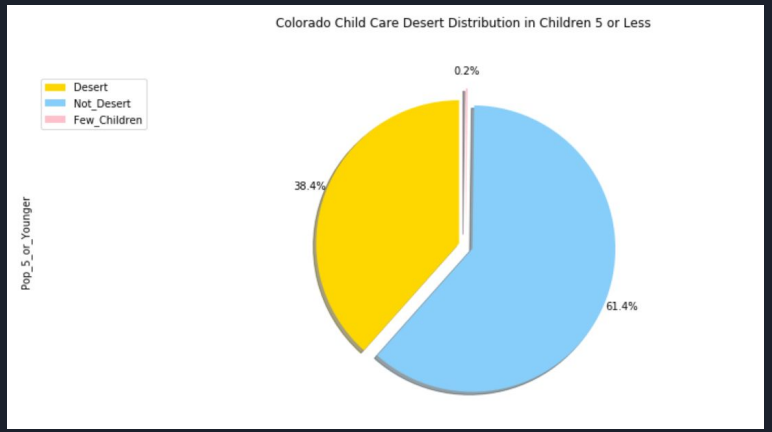
	zipcode	ageless5	pop	Latitude	Longitude	radius_meters
0	80476	11	196	39.695974	-105.731550	6100
1	80477	0	146	40.348242	-106.926910	100
2	80478	27	1625	40.001670	-105.868600	5200
3	80479	0	5	40.041288	-106.855700	8500
4	80480	89	1342	40.621621	-106.244570	31400
5	80481	6	601	40.106130	-105.480440	6800

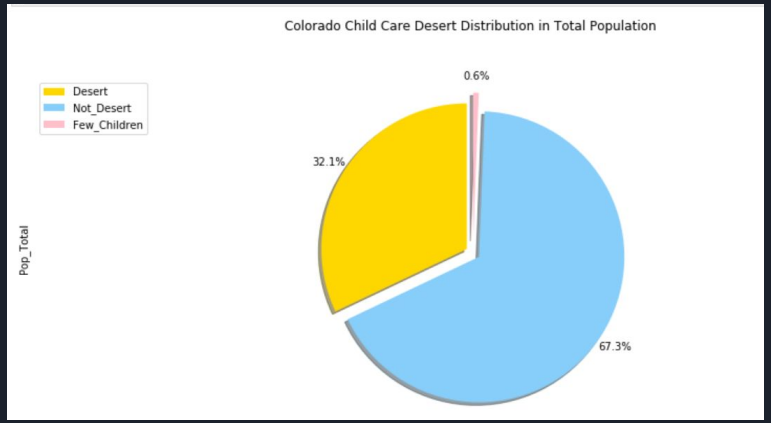
2. Foursquare API

80033	42	42	42	42	42	42
80045	13	13	13	13	13	13
80102	20	20	20	20	20	20
80103	9	9	9	9	9	9
80104	82	82	82	82	82	82
80105	7	7	7	7	7	7
80106	6	6	6	6	6	6
80107	23	23	23	23	23	23
...	...	...	...	...	...	...
81503	100	100	100	100	100	100
81504	41	41	41	41	41	41

3. Colorado Licensed Child Care Facilities Report

Child Care Desert Distribution in Colorado





Data Exploration

Comparing variable distributions across the CCD classes.

Deserts have 1. The lowest venues per person. 2. The most babies per person. 3. Slightly higher population density than the other CCD classes (Desert, Non-Desert, Few Children).

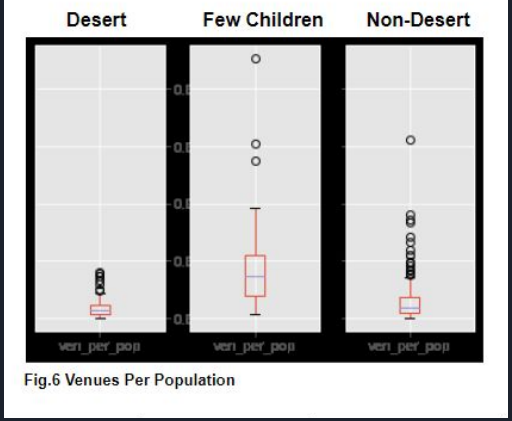


Fig.6 Venues Per Population

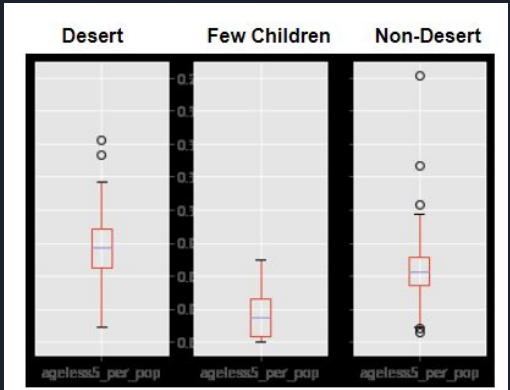


Fig.7 Ageless5 Per Population

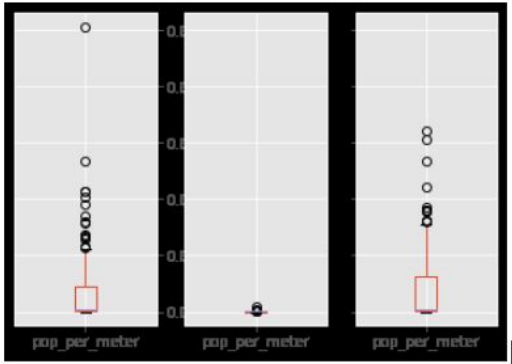


Fig.9 Total Population per Surveyed Square Meter

Studying Top Venue Frequency across CCD Classes

Once again the Legend shows 0 for Desert, 1 for Few Child, and 2 for Non-Desert.

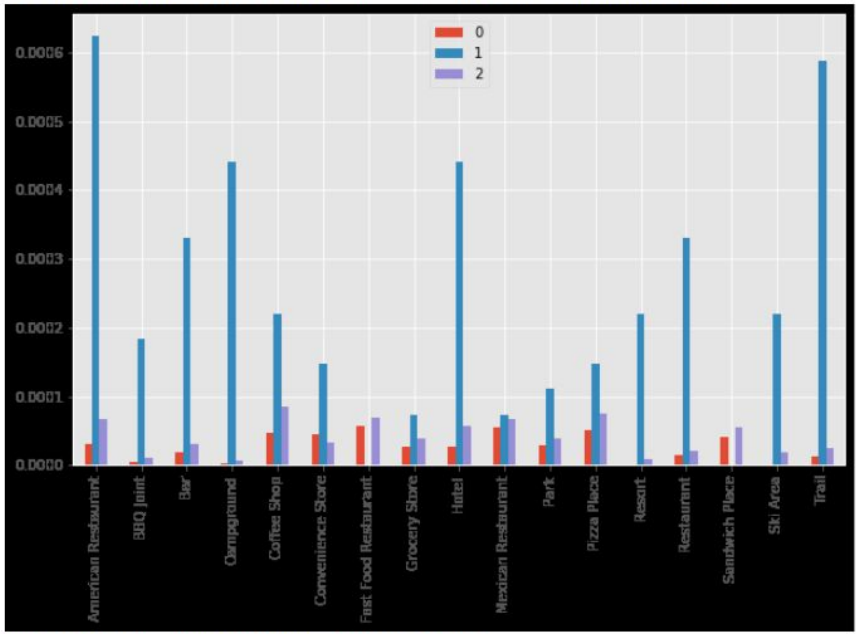


Fig.12 Top Venue Count per person in each CCD classification

The legend shows 0 for Desert and 2 for Non-Desert.

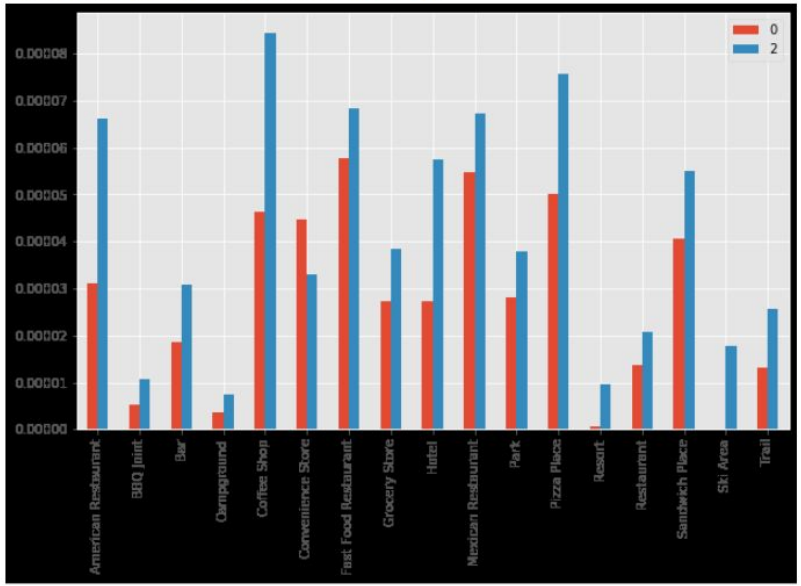


Fig.13 Top Venue Count per person in Desert and Non-Desert Classes

Convenience Stores are the Only Top Venue Most Frequent in Non-Deserts.

The Ski Area ration is infinite (there are no Ski Areas in Deserts) therefore not plotted.

Resort is also 16 x more likely in a Non-Desert and not shown to plot the other venues types closely.

The figure below shows the ratio of venus per person of the Non-Desert areas divided by the Desert minus one to see which venues are more abundant in either class. Those with higher count in the Desert have a negative value. Only Convenience Stores are more abundant in the top venues of both classes.

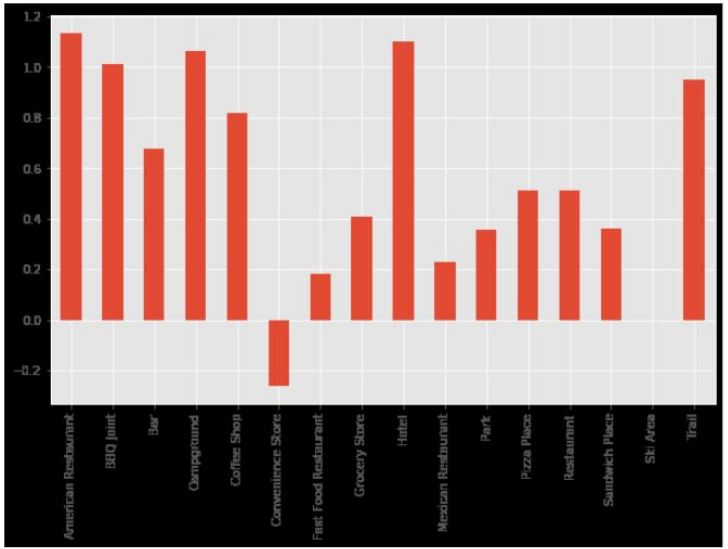


Fig.14 Top Venue Count per person of the Non-Desert divided by the Desert minus one.

Plotting and Modeling

"Manual" and DBSCAN Modeling results

I place a circle around the trends that I notice for each class in the plot. Nonetheless there are many zip codes that overlap in these circles from each class near the origin.

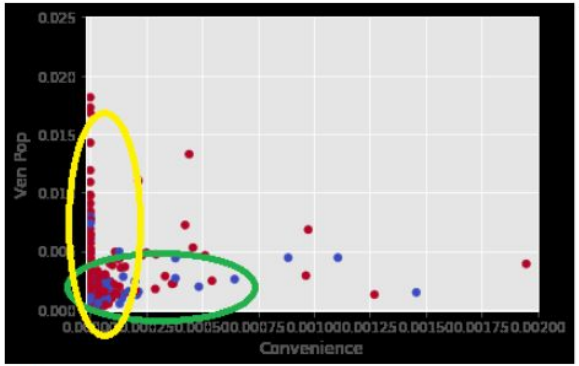


Fig.18

I do the same for Coffee shop.

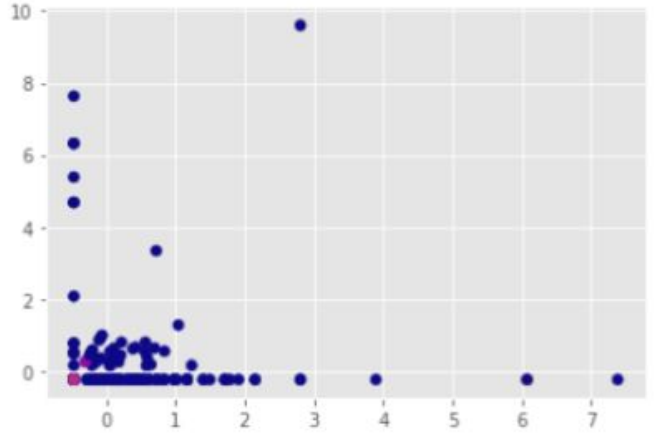


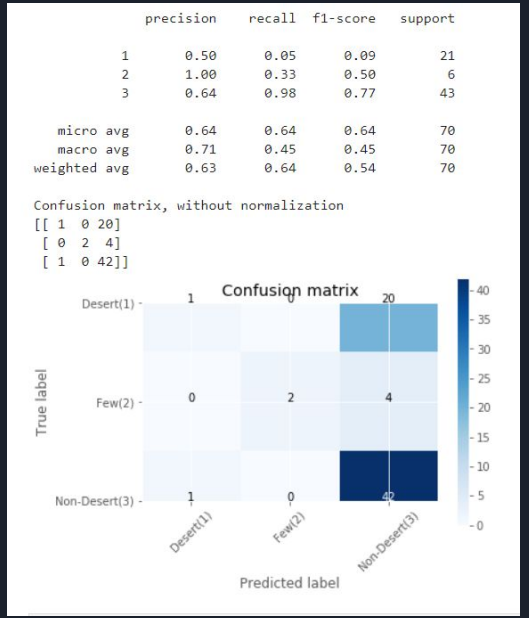
Fig.21 DBSCAN Results of Desert (pink) and Non-Desert (blue) using the top two venues scaled as the axes to plot.

SVM Modeling

For a supervised model, I chose SVM as there are various kernels within the method I can use to see if I can create a “better than random” model. The following shows the kernel methods I used from the sklearn library and their ARI score results:

	rbf	poly	sigmoid	linear
C= 10	0.042	0	0.086	0.103
C= 100	0.016	0	0.127	0.105
C=400	0.208	.061	0.100	0.087

Looking further into SVM results with the rbf kernel at C=400, the confusion matrix is as follows:
Repeating the experiment gave similar results.



	precision	recall	f1-score	support
1	0.50	0.05	0.09	21
2	1.00	0.33	0.50	6
3	0.64	0.98	0.77	43
micro avg	0.64	0.64	0.64	70
macro avg	0.71	0.45	0.45	70
weighted avg	0.63	0.64	0.54	70

Confusion matrix, without normalization
[[1 0 20]
 [0 2 4]
 [1 0 42]]

Conclusion and Discussion

- Although places that lacked child care, Deserts, where on average more densely populated, there where more venues per person in places where there were Few-Children or non-Deserts.
- Not only was there a lower amount of businesses of all types serving the Desert areas, but the people in these areas tended to have more babies to find child care for on average.
- Looking at the types of venues that are most frequently available in Non-Deserts, American Restaurants and Coffee shops were the top two venue types. All other top venues where also most frequently available in Non-Deserts in comparison to Deserts. The only top business type most available in Deserts where Convenience Stores.
- Ski-Areas exist only in Non-Deserts and Few-Children classes. There are none in Deserts.
- Fast food restaurant is the most frequent venue in Deserts but there were still more fast food restaurants in Non-Deserts than Deserts.
- Using the one-hot encoding of the top 10 venues as a way to develop an ML model, SVM produces a model slightly better than random according to the ARI metric. Doing it manually (comparing vs human) gives similar results where it is difficult to separate the classes because there are a large amount of samples of every CCD class that have a scarce number of the top venues.