

# Child Care Deserts and Venue Data in the State of Colorado

Michelle Salvador

## A. Introduction

A child care desert (CCD) is a place where child care for children under five is not readily available. This project focuses on exploring CCDs in Colorado and their relationship with businesses in the same location. Providing child care itself is a business and learning what variables and predictors affect their site (if any) such as location and proximity to other business types is beneficial. For people living in CCDs this can become a problem as parents will encounter difficulties finding someone to look after their child while they have to go to work. In addition to the high cost of child care, having to commute (possibly by public transport) can worsen a person's economic hardship by having to choose between work and caring for their child.

## B. Background

### The problems Child Care Deserts present

The following section outlines the problems CCDs cause to parents, guardians, caregivers, and policy makers.

#### B.1. Problems for Parents or Guardians (and Children)

Living in a **CCD** can present a problem to families with children under the age of 5. Without someone to care for their children, a 2015 poll conducted by the Washington Post showed that 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.

According to CAP, 51% of people in the US live in a CCD, with low income people being at 55%. 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! This leaves many parents unable to pay anything greater and places an additional economic stress to many families. To make the situation worse, 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. Research by [5 in 3] shows that when early childcare is available the child's social, economic, health, and education outcome improves.

#### B.2. Problems for Child Care Providers

In most situations the demand for a service increases supply of that service in the market. However, childcare presents a series of problems to providers in terms of business economics and liability. Since parents are often unable to afford higher childcare costs, providers make low profits. Caring for infants is also highly liable as young children are fragile and mistakes in care can have mortal consequences. One example being Sudden Infant Death Syndrome (SIDS) where babies can die without warning if they are placed to sleep incorrectly. Infant-care guidelines and complex regulations are put in place by the government to reduce these dangers but they cause an uninviting business situation along with the low profit margins. The report by CAP [3] has seen an increase of childcare providers in affluent suburbs where parents are willing to pay greater fees, but demand continues to outpace childcare options in all other places.

### **B.3. Problem for Policy Makers**

With the problems outlined above for parents, guardians, and childcare providers, CCDs are a nonpartisan problem for policy thinkers [4]. As roads and bridges are thought of as infrastructure to support the American economy, investing in childcare is a necessary service that is needed by citizens that is not being met.. Further understanding the cause for lack of childcare can provide insight into policy solutions and funding in areas most needed by the population.

## **C. 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? For example I hypothesise that ZIP codes with high amounts of adult businesses (liquor stores, marijuana dispensaries, adult venues etc) will have a higher likelihood to exist in CCDs. However, I believe it is out of the scope of this project knowing if the existence of these types of businesses cause CCD.
2. Is there a correlation with low business density to CCD? This could indicate as to how easy it is to start a business at that location or the population density that works in that ZIP code. I hypothesise that ZIP codes with low business density will be more likely to be a CCD.
3. Can a model be created using foursquare business data to predict if a census tract will be a CCD? The various types of machine learning techniques learned in this

course can be applied to attempt the highest accuracy. I predict a model can be built with accuracy greater than random coin toss (50%).

## D. Data

In order to tackle the goals of this project, first we must define what CCD is as there two definitions available [5][3]. One using ZIP code and another census tract. As the data for Colorado and geo-location was most readily available and up-to-date for the ZIP code definition, I will use the ZIP code to define CCD.

### D.1. Child Care Deserts: 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. 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.

This definition is given since children under the age of five are usually not enrolled in school and according to census averages, one third of children are “regularly in nonrelative care.” [8 lynda]

### D.2. Data Sources and Cleansing

The Colorado Information Marketplace was a wealth of data. From it, I got the **Census Zip Codes in Colorado 2017** (*Census*) [7] data for **Colorado Licensed Child Care Facilities Report** (*Facilities*) [8]. Using **Foursquare** I gathered the venues within an approximate radius equal to the ZIP code area. In an effort to use the most recent data I choose the datasets available with the most recent versions available.

#### D.2.i. Census Zip Codes in Colorado 2017

This dataset contains demographic information of each ZIP code in colorado. The relevant information I used was the ZIP code (**zipcode**), total population (**pop**), the population of children under 5 (**ageless5**), and multipolygon data. All other columns were dropped. ZIPs with no population were dropped. These are P.O. box ZIPs.

Using the multipolygon data and ArcGIS [9], an online geographic information system software, I calculated the centroid of each polygon and found the **latitude** and **longitude** for each ZIP as these were unavailable online.

Learning how to use the geojson area package, I found the total area of the polygon contained by the ZIP and calculated the radius (**radius\_meters**) of a circle that would be enclosed in a square having the same area of the ZIP polygon to the lowest hundred. Though overlap would be present with other ZIPs, this radius variable solved the problem of need for an area variable to calculate business density. Since there are ZIPs with areas as small as a few hundred meters and as large as over 20,000 meters - using a constant radius in Foursquare would not be acceptable.

The resulting pandas data frame looked like this (508 rows x 6 columns):

	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

#### D.2.ii. Colorado Licensed Child Care Facilities Report

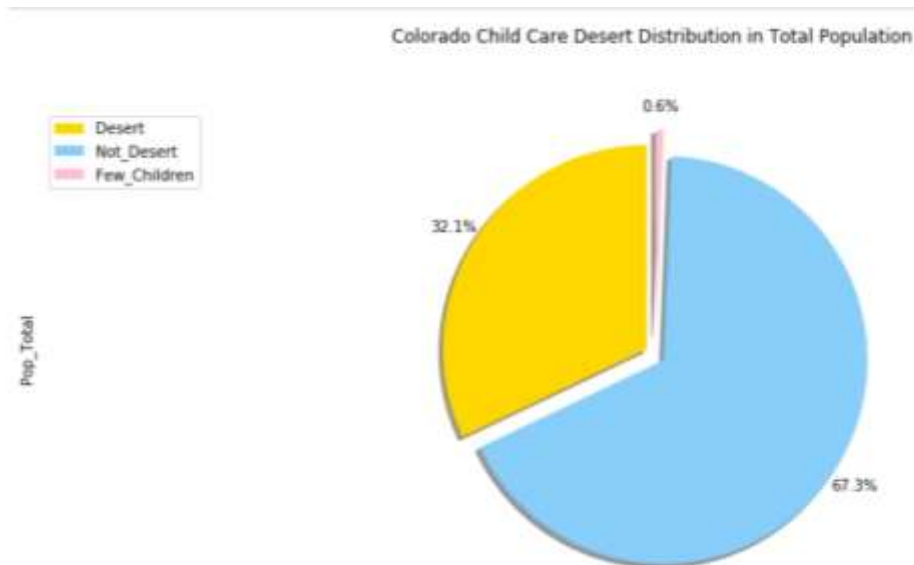
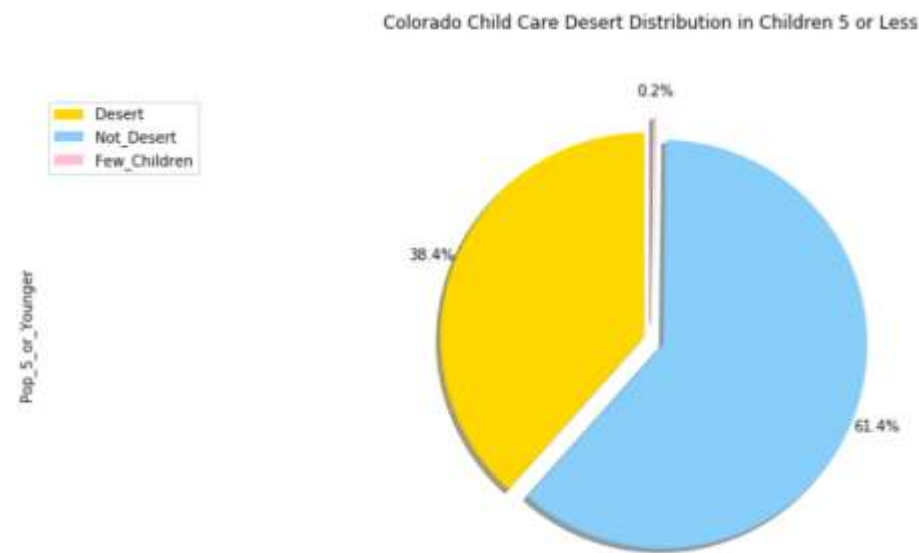
This dataset contained the names of all licensed child care facilities in Colorado with their address, type of service provided, and child capacity. Facilities that only provided 'School-Age Child Care Center' were dropped since we are targeting children under 5 who are not school aged. I then grouped the facilities by ZIP and found the sum of their capacity (**child capacity**). I merged this data with the previous data frame and using the definition of CCDs labeled each ZIP as **Desert (label 1)** (met definition of CCD), **Few\_Child (label 2)**(if there where less than 30 children), and **Not\_Desert (label 3)** (over 30 children and enough capacity requirements).

The resulting data frame was as follows (376 rows x 11 rows)

zipcode	child_capacity	ageless5	pop	Latitude	Longitude	radius_meters	Desert	Few_Child	Not_Desert	Classification
80477	203.0	0	146	40.348242	-106.92691	100	0	2	0	2
80478	39.0	27	1625	40.001670	-105.86860	5200	0	2	0	2
80480	15.0	89	1342	40.621621	-106.24457	31400	1	0	0	1
80481	486.0	6	601	40.106130	-105.48044	6800	0	2	0	2
80483	36.0	56	660	40.149432	-106.90681	7500	0	0	3	3

Interestingly there were ZIPs with no children yet a large capacity for childcare.

To visualize the amount of children and total population who lived in CCD I created the following pie charts.



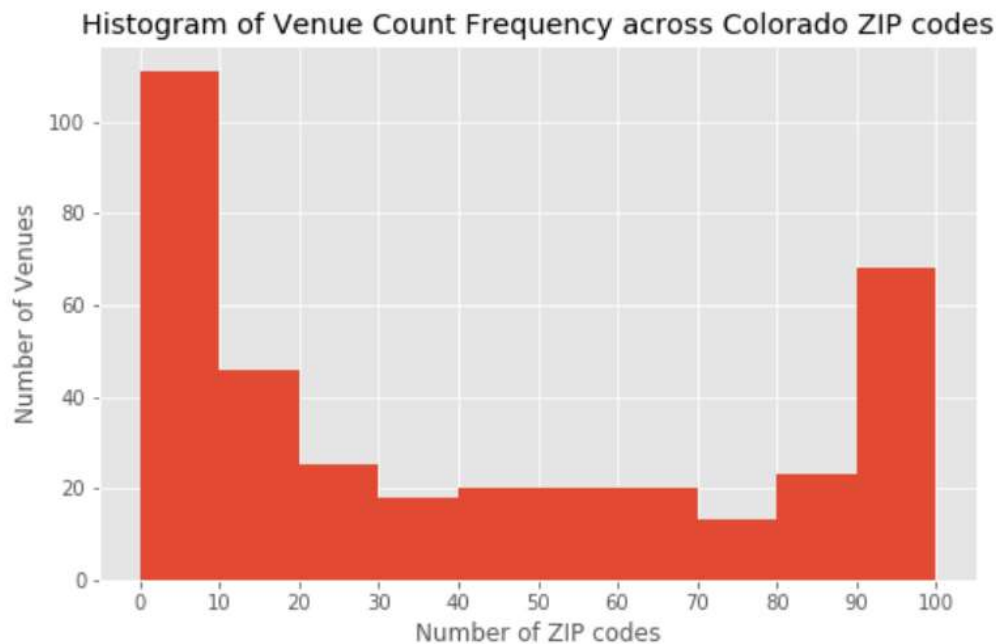
### D.2.iii. Foursquare API

Using the Foursquare API and the radius value of each ZIP I calculated from the census data, I collected the number of venues in a ZIP and the top categories in each. Though I did place an upper limit of 100 venues, I did not see much “clipping”. By this I mean that if

the venue count for the ZIPs where 100 for most, then this means I would need to increase the venue limit count. I did not see this.

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

432 unique categories where found. A wide range. 364 ZIP in total. Just to make sure not too many ZIPs were maxing at 100, I created a Histogram showing the number of ZIPs that had a certain number of venue counts.



## Sources

- [1] <https://childcaredeserts.org/index.html>
- [2] <https://www.newamerica.org/in-depth/care-report/introduction/>
- [3] <https://www.americanprogress.org/issues/early-childhood/reports/2017/08/30/437988/mapping-americas-child-care-deserts/>

- [4] <https://www.npr.org/sections/health-shots/2017/01/03/506448993/child-care-scarcity-has-very-real-consequences-for-working-families>
- [5] <https://www.americanprogress.org/issues/early-childhood/reports/2016/10/27/225703/child-care-deserts/>
- [6] <https://data.colorado.gov/>
- [7] <https://data.colorado.gov/Demographics/Census-Zip-Codes-in-Colorado-2016/rwak-e74e>
- [8] <https://data.colorado.gov/Early-childhood/Colorado-Licensed-Child-Care-Facilities-Report/a9rr-k8mu>
- [9] <https://www.arcgis.com/home/index.html>