2. Data Acquisition and Cleaning

2.1 Data sources

In the project the Foursquare API and Geocoders API are used to find location coordinates and venues around the coordinates. The Foursquare API is used to find the venues around a given location (in latitude & longitude), more information can be found here https://developer.foursquare.com/developer/. The Geocoders API will be used to find the 'latitude & longitude coordinates for the desired locations, more information can be found here https://geocoder.readthedocs.io/. Then, datasets for the population and crime in Chicago were downloaded from https://data.cityofchicago.org/.

2.2 Collecting & Cleaning data

First thing to do is get the coordinates (latitude & longitude) for the neighborhood in Denver and for all the postal codes in Chicago. To get the coordinates for the neighborhood in Denver the Geocoder API was used. Then, to get the coordinates for each postal code in Chicago, all the postal codes need to be collected. The postal codes in Chicago can be found in the dataset "Chicago Population Counts" found at https://data.cityofchicago.org/. In the dataset there are 60 entries and 20 features, there are two features of importance; the 'Population – Total', and 'Record ID'. The postal codes will be used to get the coordinates for the areas in Chicago, and the population will be used later in the analysis.

Population and Postal Codes in Chicago

	Geography Type	Year	Geography	Population - Total	Population - Age 0-17	Population - Age 18-29	Population - Age 30-39	Population - Age 40-49	Population - Age 50-59	Population - Age 60-69	Popul. - Ag
0	Citywide	2018	Chicago	2705988	548999.0	552935.0	456321.0	336457.0	312965.0	262991.0	155;
1	ZIP Code	2018	60601	14675	NaN	NaN	NaN	NaN	NaN	NaN	
2	ZIP Code	2018	60602	1244	NaN	NaN	NaN	NaN	NaN	NaN	
3	ZIP Code	2018	60603	1174	NaN	NaN	NaN	NaN	NaN	NaN	
4	ZIP Code	2018	60604	782	NaN	NaN	NaN	NaN	NaN	NaN	

Then using the Geocoders the coordinates for each postal code was found, and a new data frame was built:

Chicago Postal Codes, Population and Coordinates

	Postal Code	Population	Latitude	Longitude
0	60601-2018	14675	41.885967	-87.624265
1	60602-2018	1244	41.882993	-87.629251
2	60603-2018	1174	41.880925	-87.628176
3	60604-2018	782	41.878284	-87.628010
4	60605-2018	27519	41.870417	-87.627331

Next, to find the venues around each location the Foursquare API is used. By running the Foursquare query all the most popular venues around a given location are returned, however the free version of the API is used so that a maximum of 100 venues can be returned for each query. The API returns a .json file which is parsed to create a dataframe that combined with the dataframe above with the features to describe the center of the location with the coordinates, the venue with the coordinate, and the distance from the venue and the center location.

Venue Dataframe

	Center	Lat_center	Lng_center	Category	Name	Lat_venue	Lng_venue	Distance (m)
0	Lincoln Park	39.733138	-105.005241	Brewery	Renegade Brewing Company	39.730616	-104.999292	581.849369
1	Lincoln Park	39.733138	-105.005241	Japanese Restaurant	Domo Japanese Country Foods Restaurant	39.738100	-105.005650	551.998957
2	Lincoln Park	39.733138	-105.005241	Arts & Entertainment	Santa Fe Art District	39.730636	-104.998669	628.149538
3	Lincoln Park	39.733138	-105.005241	Steakhouse	The Buckhorn Exchange	39.732205	-105.005067	104.710877
4	Lincoln Park	39.733138	-105.005241	Café	The Molecule Effect	39.735386	-104.998781	607.382310

Lastly, the crime data in Chicago is used to find the crimes around each location. The crime data "Crimes _ 2019" was downloaded from https://data.cityofchicago.org/, and the dataset contains 260,000 entries with 22 rows.

Raw Chicago Crime Data

	ID	Case Number	Date	Block	IUCR	Primary Type	Description	Location Description	Arrest	Domestic	...
0	11937949	JC566932	12/31/2019 08:35:00 PM	064XX S WESTERN AVE	031B	ROBBERY	ARMED - OTHER FIREARM	COMMERCIAL / BUSINESS OFFICE	False	False	...
1	11666065	JC237357	04/25/2019 01:49:00 AM	068XX S JUSTINE ST	051A	ASSAULT	AGGRAVATED - HANDGUN	RESIDENCE - PORCH / HALLWAY	False	False	...
2	11613029	JC173331	02/28/2019 04:00:00 PM	035XX W 55TH ST	1582	OFFENSE INVOLVING CHILDREN	CHILD PORNOGRAPHY	SCHOOL - PUBLIC BUILDING	False	False	...
3	11596162	JC152837	02/14/2019 09:03:00 PM	072XX S UNIVERSITY AVE	0498	BATTERY	AGG. DOMESTIC BATTERY - HANDS, FISTS, FEET, SE...	RESIDENCE	False	True	...
4	11936658	JC565320	12/29/2019 04:00:00 PM	097XX S INGLESIDE AVE	1020	ARSON	BY FIRE	RESIDENCE - GARAGE	False	False	...

The crime data was then used to count the crimes around each location.