

Battle of Neighborhoods

Opening a Hungarian Restaurant in a Hungarian Neighborhoods somewhere in North America

Csilla Gates

IBM Data Science

Professional Certificate

March 15, 2020

Author Notes

Csilla Gates [Github repository: //github.com/Csillag61/Coursera_Capstone](https://github.com/Csillag61/Coursera_Capstone)

Correspondence concerning this article should be addressed to Csilla Gates.

Email: fake@gmail.com

Phone: 888/888-8888

Table of Contents

Summary	3
Introduction	4
Data	6
Work Flow	8
Results Part I	9
Results Part 2	14
Results Part3	15
Conclusion	18

This project is about whether to open a new Hungarian Restaurant in a Hungarian neighborhood considering the Hungarian American population in the time of the COVID-19 outbreak.

Introduction

Large-scale Hungarian immigration to the US began more than 40 years ago. Cleveland, OH has been a focal point for thousands of Hungarian newcomers. The influx of Hungarians was so great at times that Cleveland was considered the city with the second largest Hungarian population, after Budapest, the capital of Hungary. Their number was 203,417 in 2010 and 183,593 in 2014.

Hungarian cuisine surely has some delights including hearty soups, stews and game dishes, simple but tasty casseroles and luscious cakes and pastries. Hungarian cuisine is a synthesis of ancient Uralic components mixed with West Slavic, Balkan, Austrian, and German. The food of Hungary can be considered a melting pot of the continent, with a culinary base formed from its own, original Magyar (Hungarian) cuisine.

Today we really started to feel the effects of the deadly Wuhan virus outbreak.

Coronaviruses are a large family of viruses which may cause illness in animals or humans. In humans, several coronaviruses are known to cause respiratory infections ranging from the common cold to more severe diseases such as Middle East Respiratory Syndrome (MERS) and Severe Acute Respiratory Syndrome (SARS). The most recently discovered coronavirus causes coronavirus disease COVID-19.

We will explore data to help us visualize and even predict the outcome that might or might not influence our business decision to open a Hungarian cuisine restaurant in the near future.

It is true, a great Hungarian invented the Rubik cubes and every good Hungarian has one, but the majority including myself can't solve it. This project is aiming to solve the burning problem:

- Is a living Hungarian neighborhood still out there today?
- Is today the right time to open a restaurant so every one of us can enjoy it's famous dishes?

Datasets and API

1. The first part of my project is looking for parts of US where the largest number of the Hungarians are. I used the available census data, narrowed down to one state, carefully observed and analyzed the ranks and the population of cities.

Data sources for this part of my projects are:

- factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_10_1YR_DP02&prodType=table (Please note: Although the American Community Survey (ACS) produces population, demographic and housing unit estimates, it is the Census Bureau's Population Estimates Program that produces and disseminates the official estimates of the population for the nation, states, counties, cities, and towns and estimates of housing units for states and counties.)
- web.archive.org/web/20131213043218/http://www.hhrf.org/restitution/hungariansintheus.htm
- zipatlas.com/us/oh/zip-code-comparison/percentage-hungarian-population.htm

2. The second part of my project's objective is

- using the Four-Square API (Beautiful Soup, http request) as its prime data gathering source to collect the neighborhood's top trending venues. This

API provides the ability to perform location search, location sharing and details about a business.

- Forming neighborhood clusters based on venue categories using unsupervised k-mean clustering algorithm (sklearn)
- Identifying and understanding which neighborhood should be the chosen one.

3. The 3th and last part of my project relies on the 2019 Novel Coronavirus COVID-19 Data Repository by Johns Hopkins CSSE://github.com/CSSEGISandData/COVID-19.

This data repository for the 2019 Novel Coronavirus Visual Dashboard operated by the Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE). Also, Supported by ESRI Living Atlas Team and the Johns Hopkins University Applied Physics Lab (JHU APL).

This dataset has information from 50 US states and the District of Columbia at daily level.

In this part of the project I analyze the pandemic spread, by using data from all around the world and several demographic data for each country. I focus on the increase rate.

4. Python packages and Dependencies:

- Pandas - Library for Data Analysis
- NumPy – Library to handle data in a vectorized manner
- JSON – Library to handle JSON files
- Geopy – To retrieve Location Data
- Requests – Library to handle http requests

- Matplotlib – Python Plotting Module
- Sklearn – Python machine learning Library
- Folium – Map rendering Library

WORK FLOW

for each part of the project

- Web Scraping and Data Wrangling
 - get the necessarily data with Beautiful Soup
 - Data Visualization with seaborn, pie chart, bar plots to get a better understanding of the numbers. Folium map grouped the cities into different clusters by national ranks and population.

- Top Trending Places Extraction and Clustering

HTTP requests would be made to Foursquare API server using zip codes of the Cleveland city neighborhoods to pull the location information (Latitude and Longitude).

Folium- Python visualization library would be used to visualize the neighborhoods cluster distribution of Cleveland city over an interactive leaflet map.

Unsupervised machine learning algorithm K-mean clustering would be applied to form the clusters of places. (Elbow curve, Silhouette analysis used for selecting the number of clusters)

These clusters would be analyzed individually collectively and comparatively to derive the conclusions.

- Decision Making based on the data analysis of Hungarian population, clustered neighborhoods, COWID-19 prediction.

RESULTS, PART 1

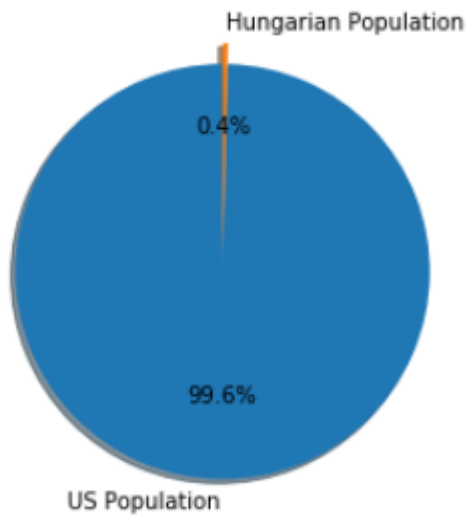
First I research on the Hungarian American population in the US as a whole country. Based the census factfinder most recent data , it turns out, there was 1,373.503 Hungarian Americans in the US, which is 0.4% of the whole US population.

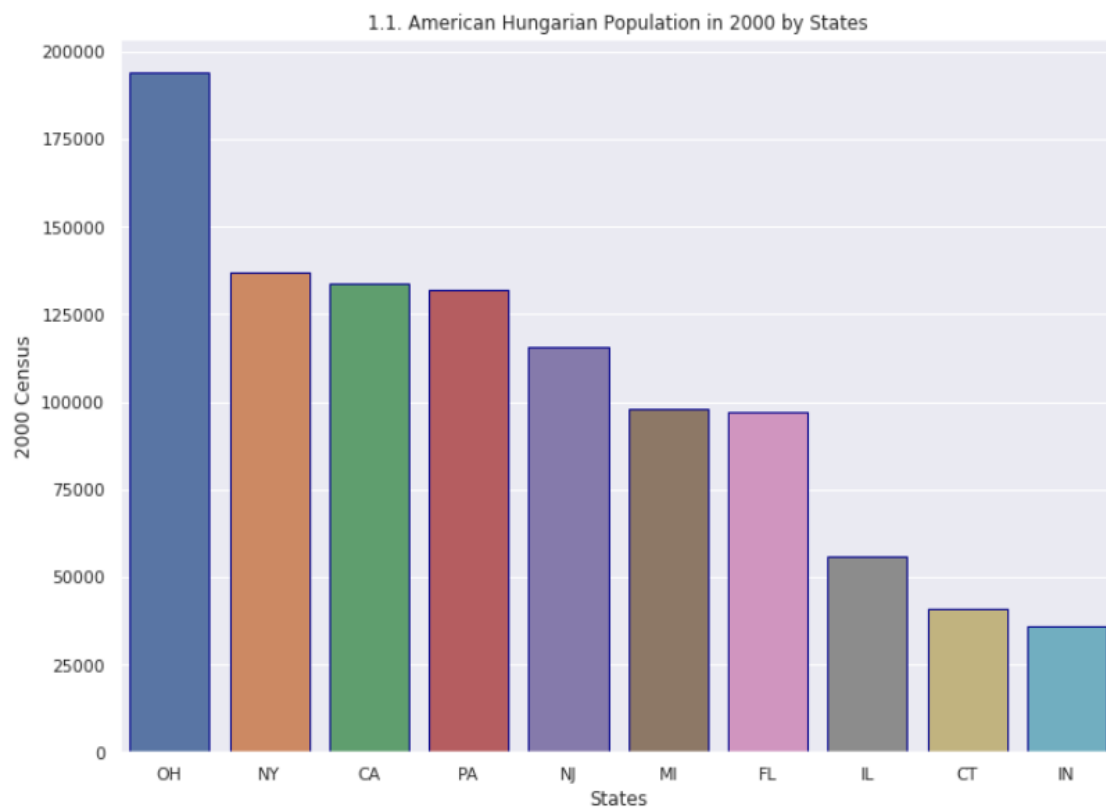
	GEO.display-label	Hungarian Population	Percent of US Population
0	Geography	Estimate in 2017	%
1	United States	1373503	0.4

Next step to find, where is this little piece of the big pie located.

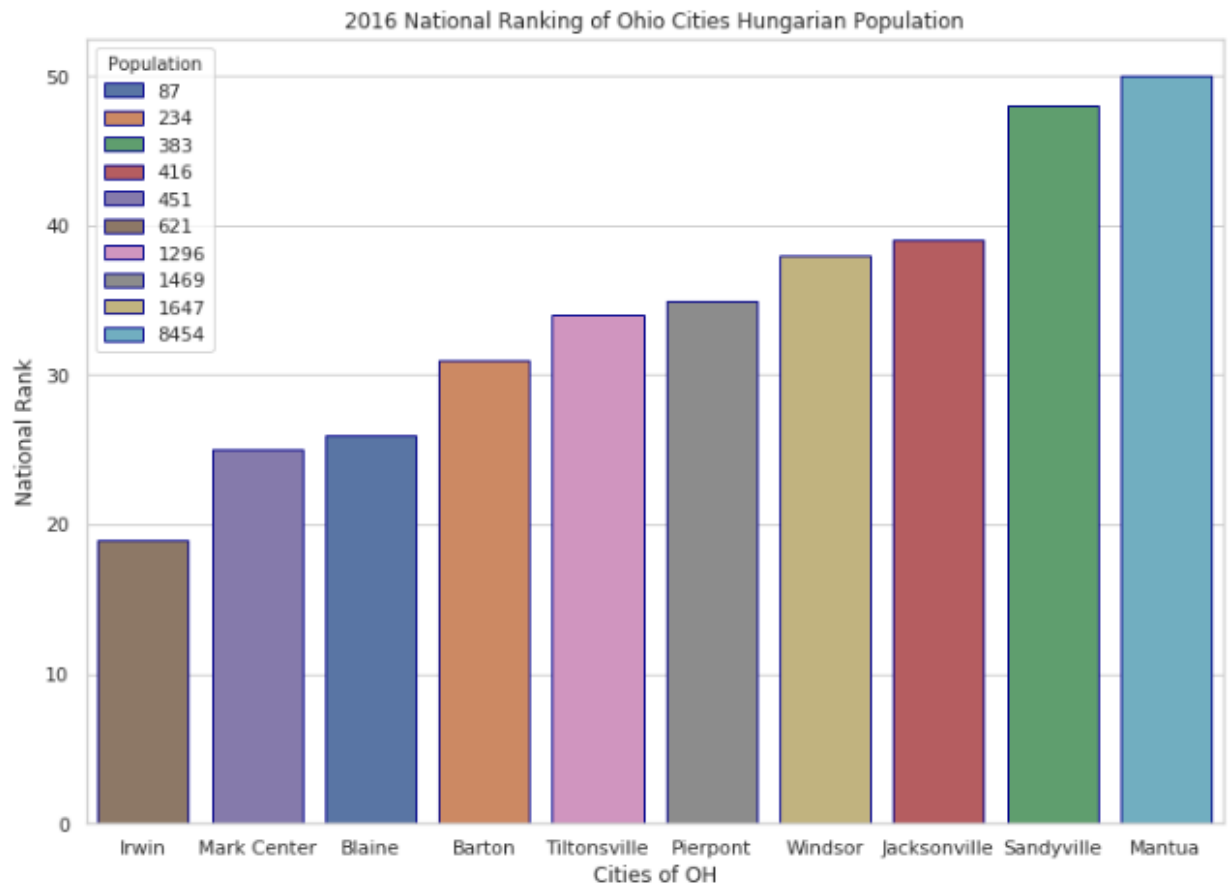
I scraped again census data, that again had a big data to clean and observe.

The next two bar plots tell us, that out of the top ten states we have OH on the first place, with close to 200.000 Hungarian population.

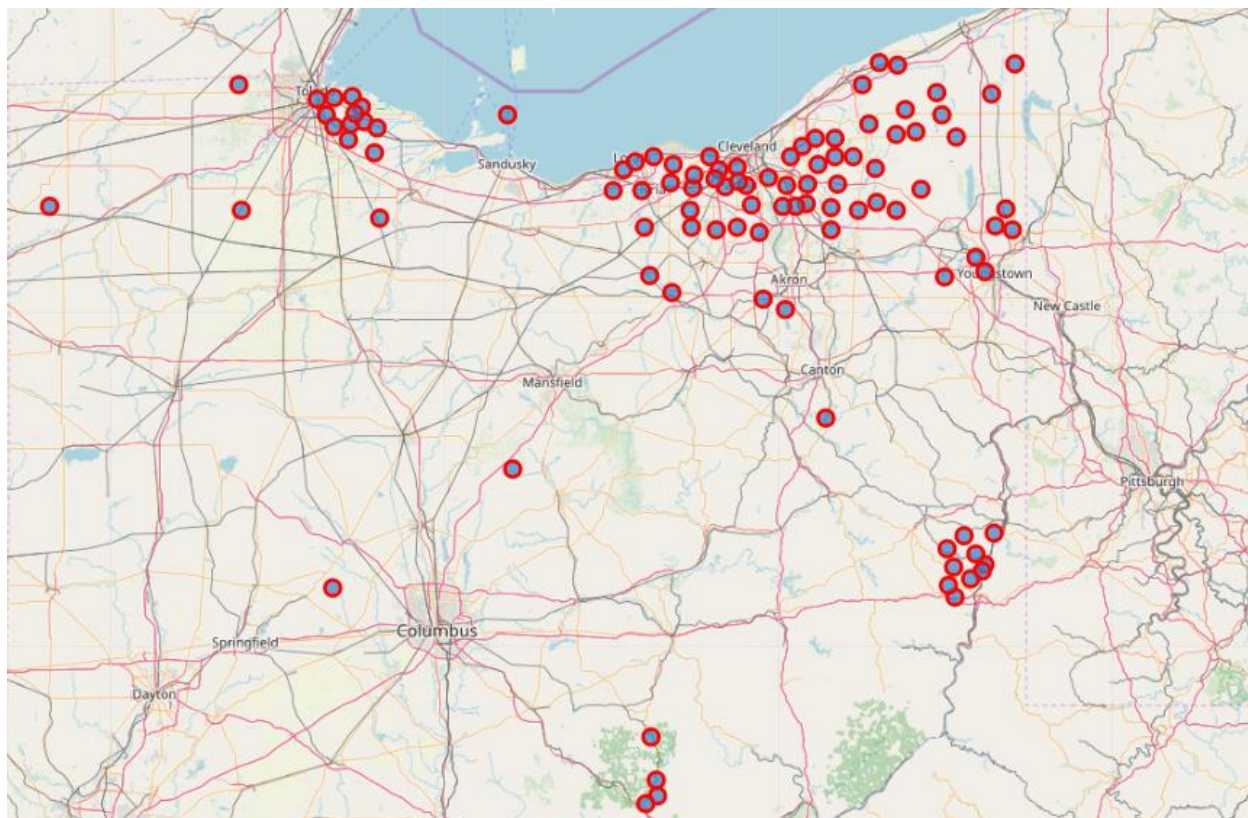




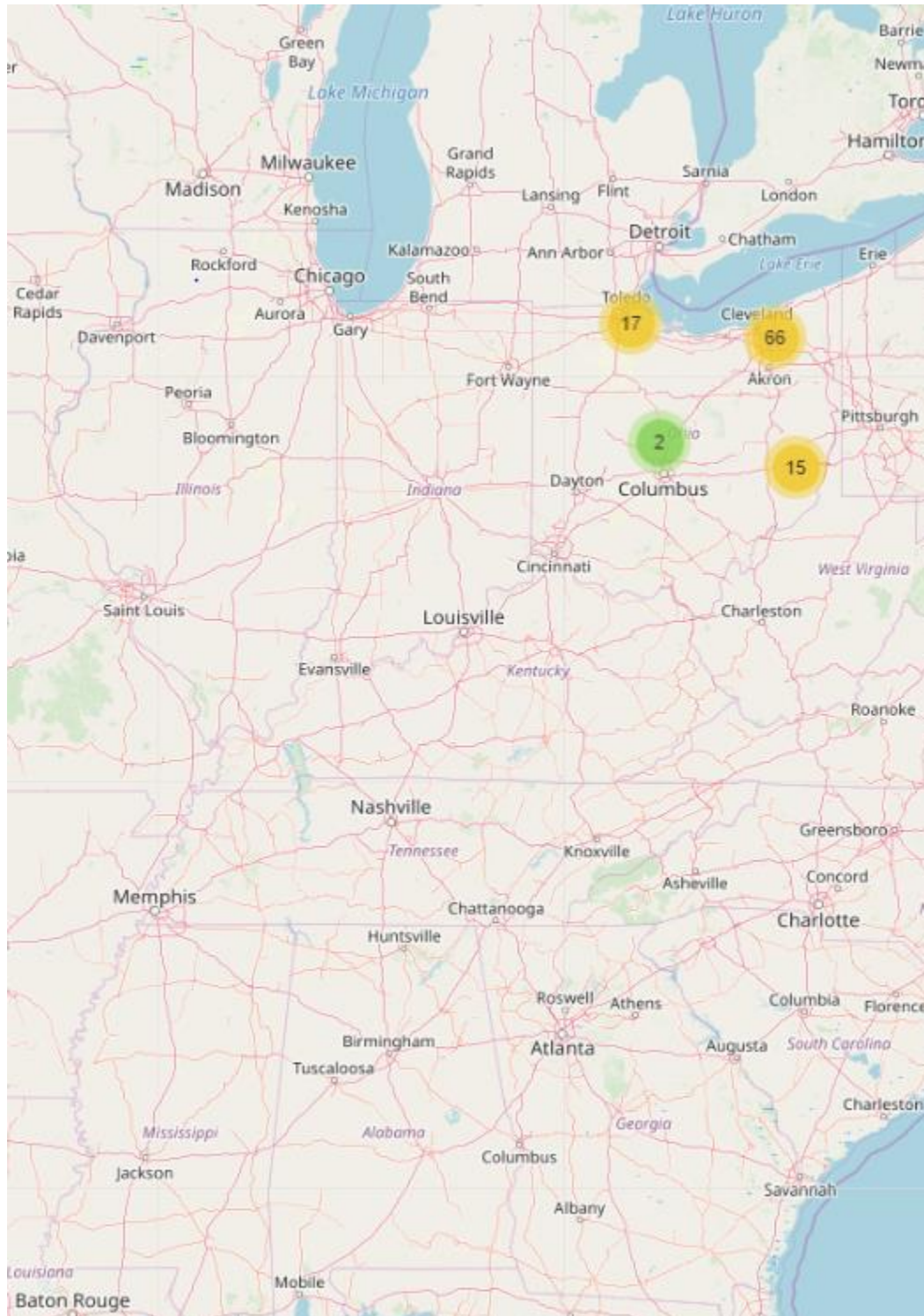
The next dataset gave me the final answer. I located the first 100 cities, that had the highest Hungarian % to compare to the American population.



Based on these findings, I retrieve the coordinates of the cities and put them on the map with Cleveland in the center.



Finally, I grouped the ranked cities into different clusters on a new map, so we can analyze them separately.



As we can see, the cities are forming 4 big clusters, with 2, 15, 17 and 66 cities in them.

The circle with the biggest number (66 cities) is located in the Cleveland neighborhood!

The first part of my project gave us a clear answer to our problem: yes, in the US we still have plenty of Hungarian Americans, and the majority are in the neighborhoods of Cleveland, OH.

RESULTS, PART 2

Now that we know that most of the Hungarian Americans are living in the Cleveland neighborhood, we are focusing on that area.

With web scraping as a main tool, I gathered together 33 neighborhood, and their coordinates. Now we have a larger area to work with. The Foursquare API helped to explore the top 200 venues and narrow down to their unique venues. I checked whether the word 'Hungarian' was among them, and the answer was 'true'. Now I grouped the venues by neighborhood, analyzed each of them, by the mean of the frequency of occurrence of each category. Analyzing the elbow curve and observing the silhouette analysis for KMean clustering on sample data with different n_clusters, I have found the optimum number of clusters. So I clustered the Cleveland neighborhoods into 3 clusters. I have found only 1 Hungarian Restaurant in the Cleveland area, with no other Hungarian competition!

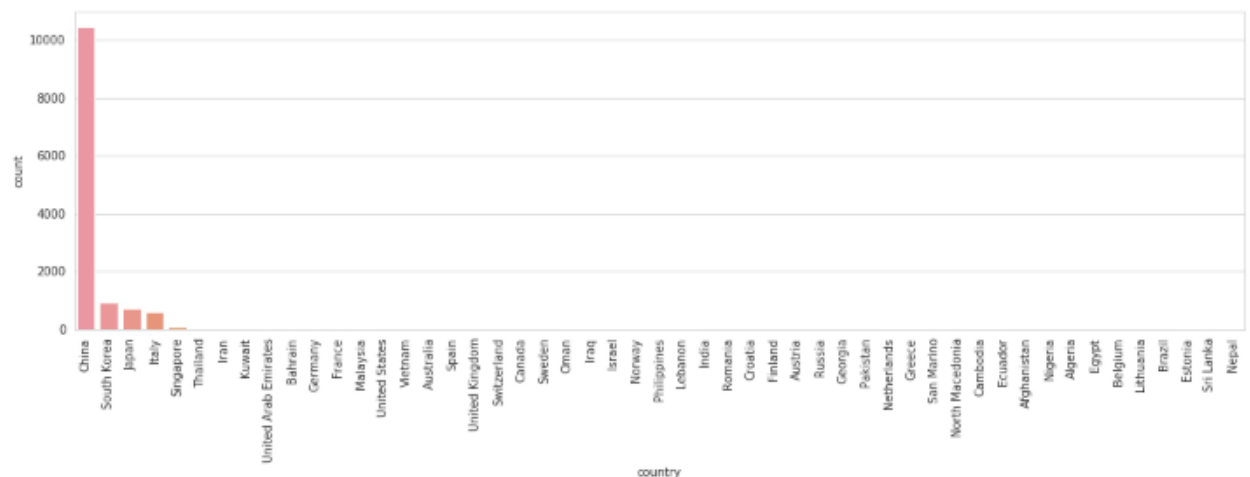
RESULTS, PART 3

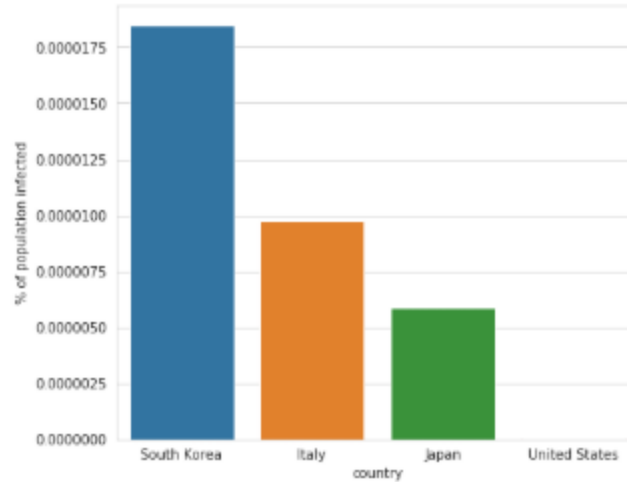
It is true, we have a large Hungarian American population in the Cleveland area and only one Hungarian Restaurant. We can't jump into any conclusion or recommendation before we take time to observe the COVID-19 effect on businesses and everyday life.

The spread of virus across the countries plot shows clearly, that the highest count of the virus 'cases' is in China, South Korea and Japan.

Three of Asia's largest countries has the highest count.

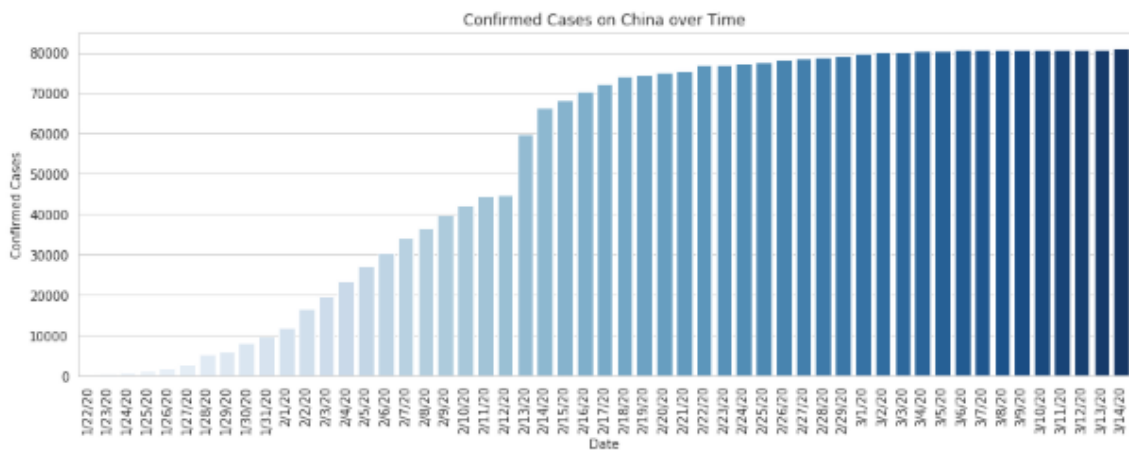
It seems that the vast majority of cases occurred in China. China has at least 10 times more cases than South Korea and Italy combined.





If I compare the percent of the population infected in South Korea, Japan, Italy and the United States, the present for US is not even visible.

However, I am checking on the confirmed cases on China over time, it seems to me, it reached the top and does not grow any further.





Our number in the US jumping up, doubled the number in 2 days.

Looking just at facts, in all over the world, the numbers are really bad:

	Date	Confirmed	Deaths	Recovered	Active
0	2020-03-13 00:00:00	145193	5404	70251	69538

My final conclusion is as follows:

We have a large Hungarian Community around Cleveland, OH. There is very little competition among Hungarian Cousins, it would be ideal to open up a new restaurant. My recommendation is to wait or instead of a restaurant consider a takeout business.