

# Site selection for a new Peruvian restaurant in Chicago city



# Why is this study important?

- **Where?** It is the critical question to answer
- **Aim:** to identify potential areas for new Peruvian restaurants
- Includes relevant **factors** as: income, unemployment, density population, Latino population, crime and restaurant competitors to understand where an entrepreneur should open a restaurant
- **Target:** Peruvian entrepreneurs



# Study area

---

- 206 census tracts
- Census tracts within the Central Business District (CBD) and tracts to 5 km. from the CBD boundary
- CBD is the commercial and business area of Chicago

# What data do we need?

To accomplish the aim of this study, we first need Census data at census tract level estimated to the year 2018

Explore: <https://data.census.gov/cedsci/>

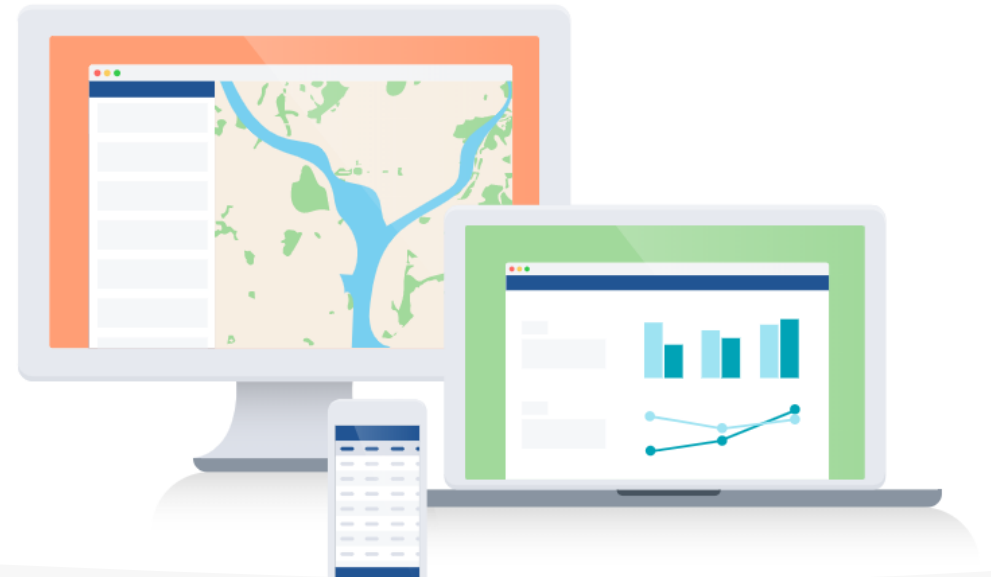


## Explore Census Data

The Census Bureau is the leading source of quality data about the nation's people and economy.

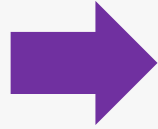
🔍 I'm looking for ...

[Advanced Search](#)



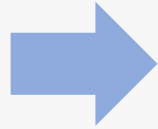
# What Census data?

Population



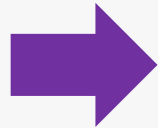
Key variable to build other variables such as: density population, Latino population rate and crime rate which will be used for modeling

Income



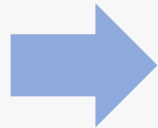
Entrepreneurs will always be interested in investing money in areas where households have high income

Unemployment



Areas with high unemployment rates are avoided by entrepreneurs

Latinos

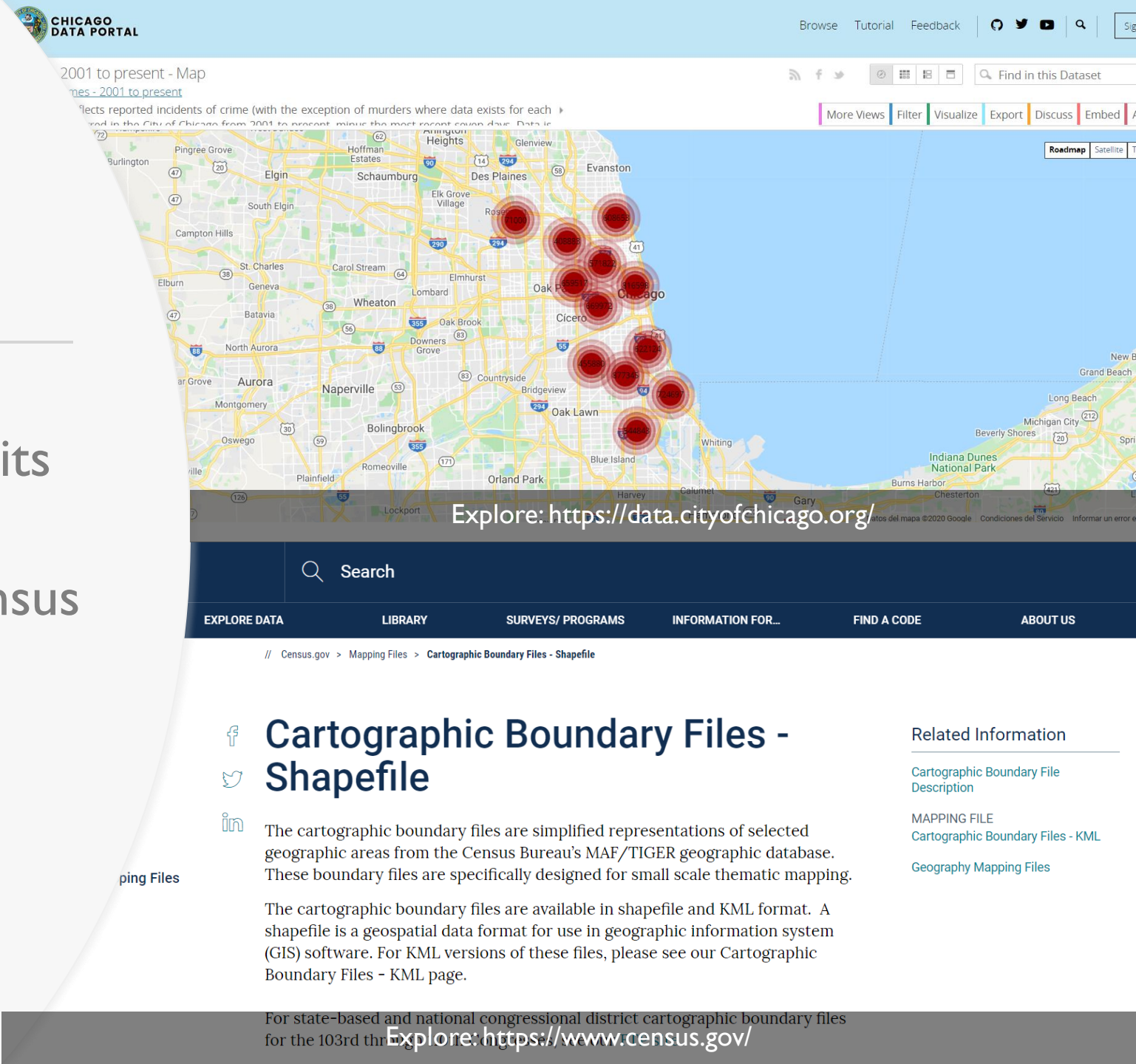


Areas with high Latino population rate could accept the Peruvian cuisine faster in comparison to other customs from Europe, Asian, and so on



# ...and geospatial data

- Census tract boundaries and its area in hectares
- Centers of Population by Census Tract
- CBD boundary
- Crime



# A hard work to get cleaned dataset

Social, economic and cultural dimensions

Business dimension

Id census tract	Income	Unemployment	Density population	Latino population rate	Crime rate	Latitude	Longitude	# Restaurants	1st most common restaurant	2nd most common restaurant
I400000US17031051300	...	...	...	...	...	...	...	...	...	...
I400000US17031051400	...	...	...	...	...	...	...	...	...	...
I400000US17031060900	...	...	...	...	...	...	...	...	...	...
I400000US17031061500	...	...	...	...	...	...	...	...	...	...
I400000US17031061800	...	...	...	...	...	...	...	...	...	...

Used to call Foursquare API and built the business variables

# My tools used were:

## Tasks



Selection of census tracts within the CBD and 5 kms. to it. It allowed to calculate area in hectares of census tracts and the obtain density population. Export to 'json' file. Geocoding of crimes and impute id of census tract to each crime.



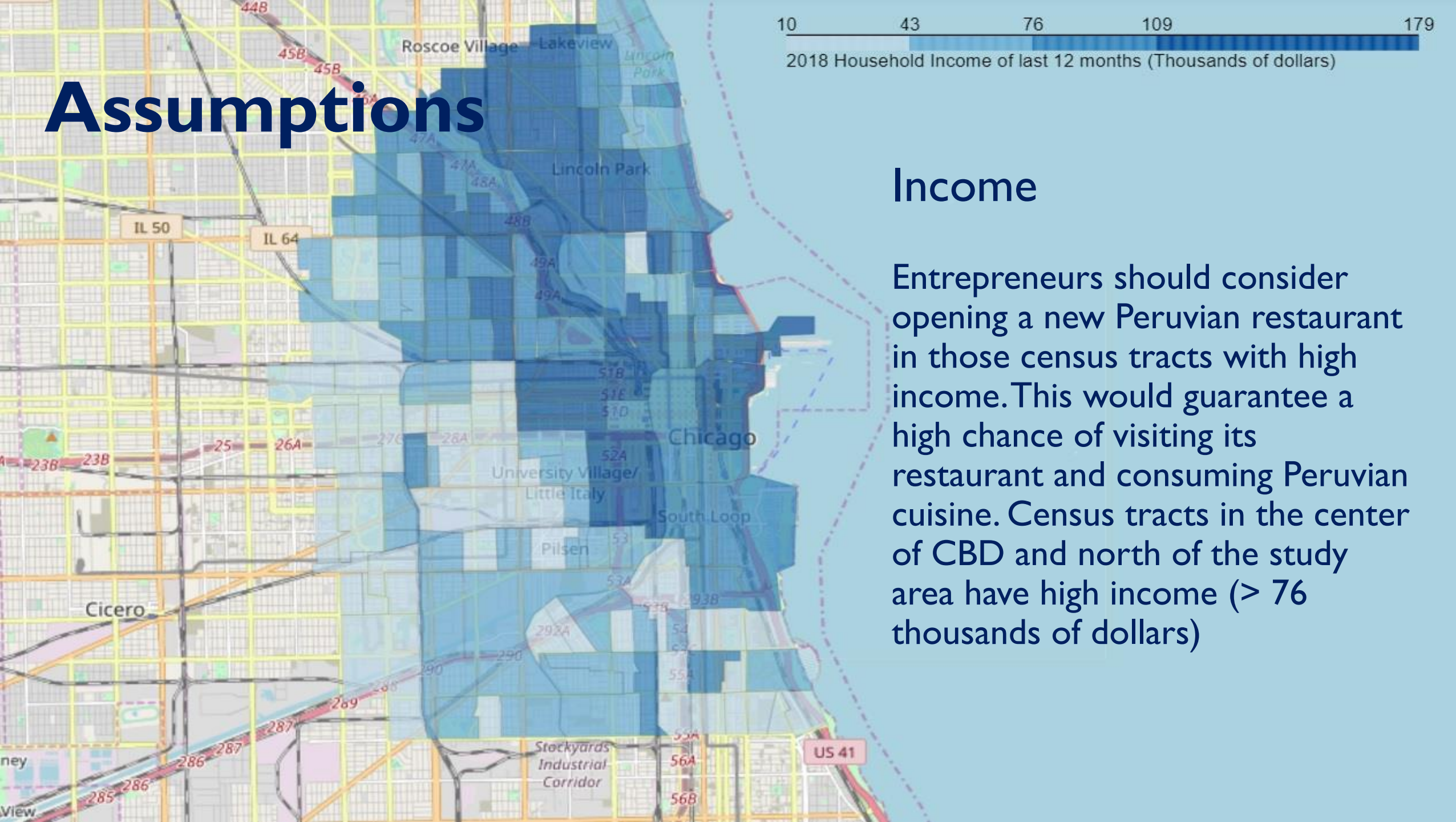
Used from Anaconda Navigator. Read and merge different tables. Data pre-processing and cleaning to get the cleaned dataset. Build code to call Foursquare API and get restaurants venues around each coordinate of census tracts (latitude and longitude) and then build the 3 business variables from the cleaned dataset. Map visualization. Modeling for scoring.



Create a free account to get a CLIENT ID and call the API in Jupyter Notebook



# Assumptions



## Income

Entrepreneurs should consider opening a new Peruvian restaurant in those census tracts with high income. This would guarantee a high chance of visiting its restaurant and consuming Peruvian cuisine. Census tracts in the center of CBD and north of the study area have high income ( $> 76$  thousands of dollars)



# Assumptions

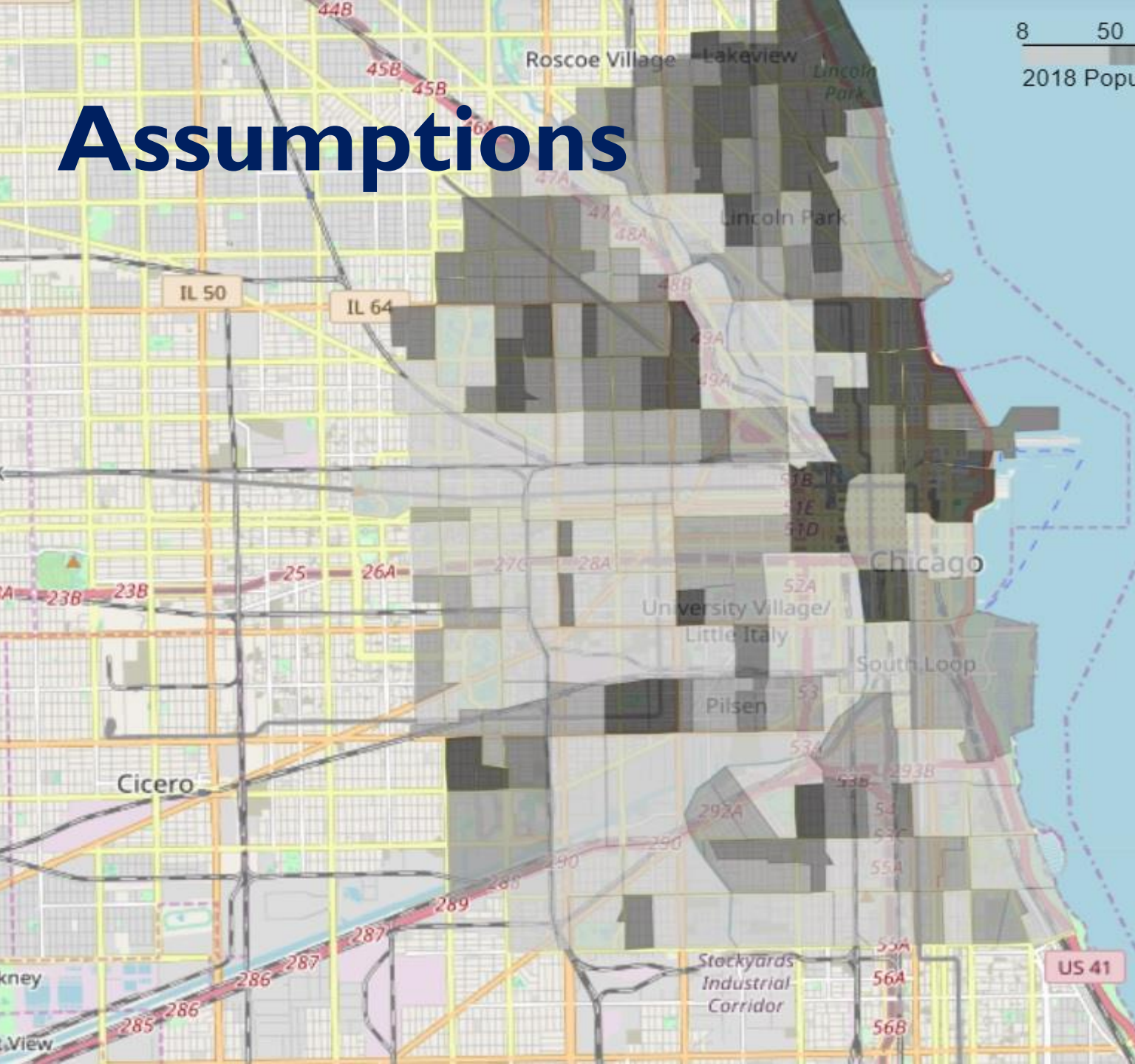


## Unemployment

This is an important variable since the entrepreneur looks for areas where people are working and have a stable employment. So, entrepreneurs want to avoid areas with high rates of unemployment as they can be mainly seen in the west and south



# Assumptions



# Density population

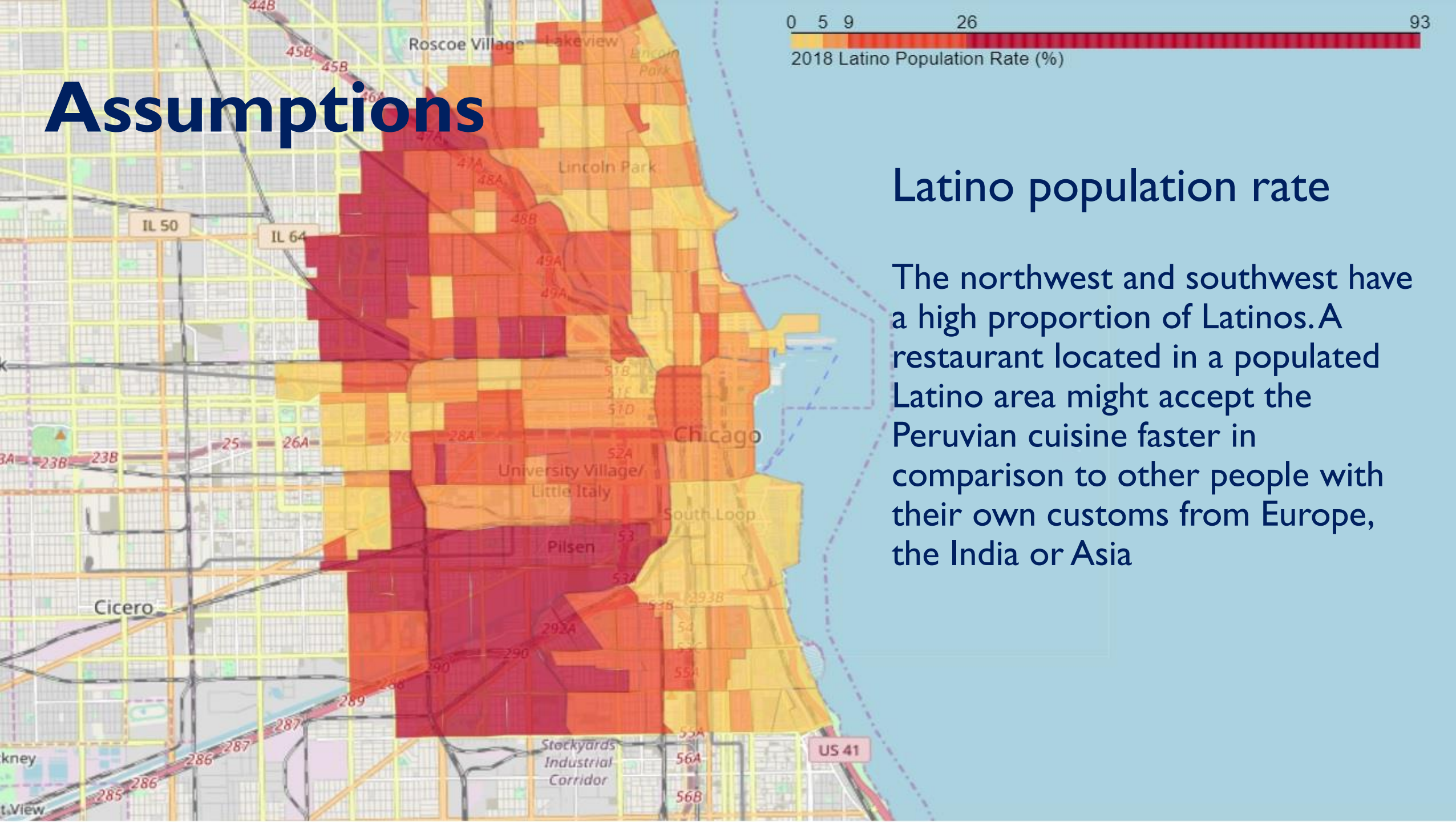
Populated areas might guarantee a high chance of visiting a commercial store as restaurants. Thus, entrepreneurs are always looking for populated areas which were identified in the north and center

# Density population

Populated areas might guarantee a high chance of visiting a commercial store as restaurants. Thus, entrepreneurs are always looking for populated areas which were identified in the north and center



# Assumptions

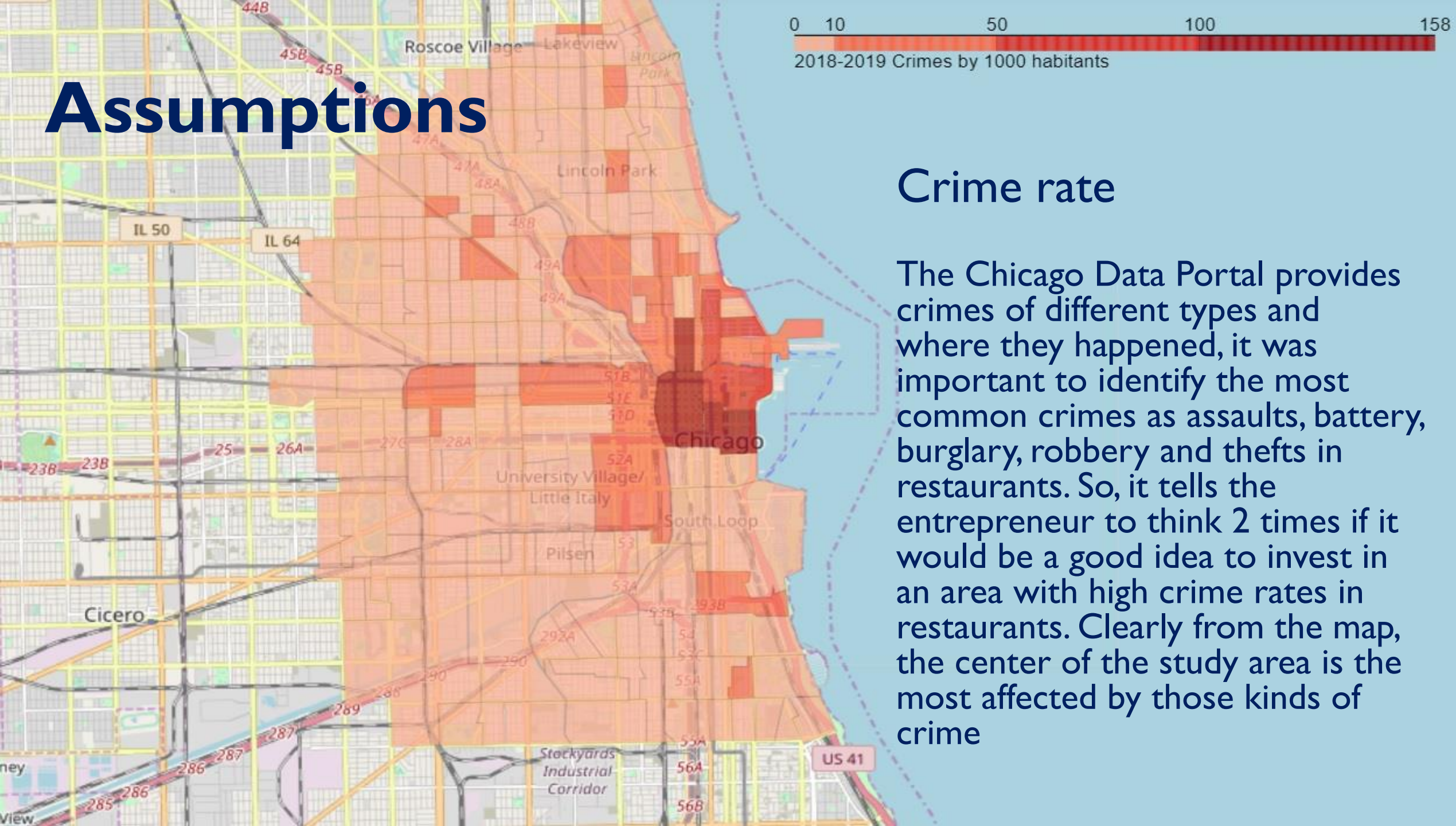


## Latino population rate

The northwest and southwest have a high proportion of Latinos. A restaurant located in a populated Latino area might accept the Peruvian cuisine faster in comparison to other people with their own customs from Europe, the India or Asia



# Assumptions

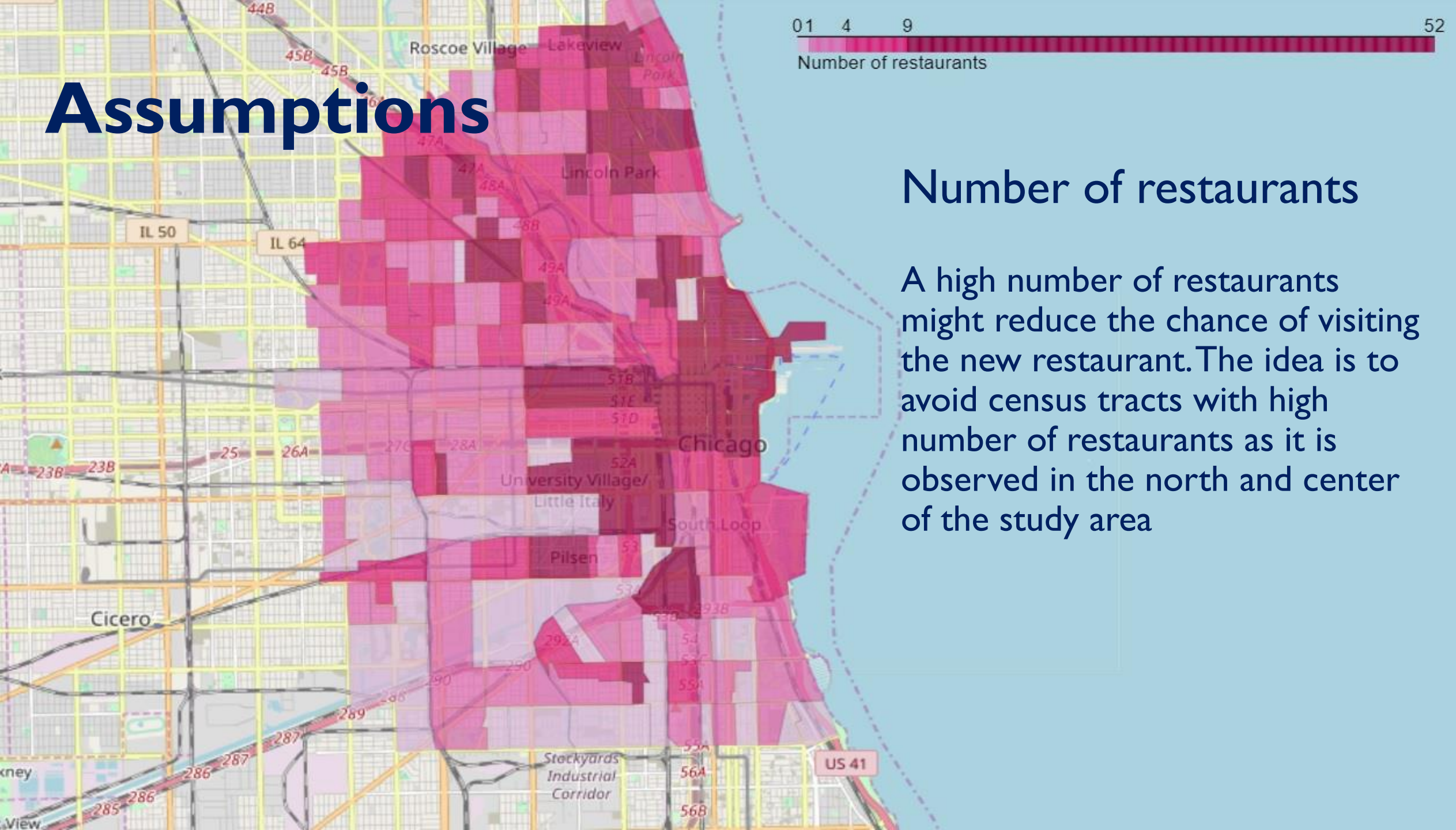


## Crime rate

The Chicago Data Portal provides crimes of different types and where they happened, it was important to identify the most common crimes as assaults, battery, burglary, robbery and thefts in restaurants. So, it tells the entrepreneur to think 2 times if it would be a good idea to invest in an area with high crime rates in restaurants. Clearly from the map, the center of the study area is the most affected by those kinds of crime



# Assumptions





# Modeling

- **A single score** to find potential areas to open a new Peruvian restaurant
- **Score 1.** Based on numerical data types (census variables, crime and one variable built from Foursquare). A weighted mean of normalized variables.

$$Score\ 1 = \frac{(5 * Income) - (3 * Unemployment) + (5 * Density) + (2 * Latinos) - (3 * Crime) - (2 * Restaurant)}{20}$$

# Modeling

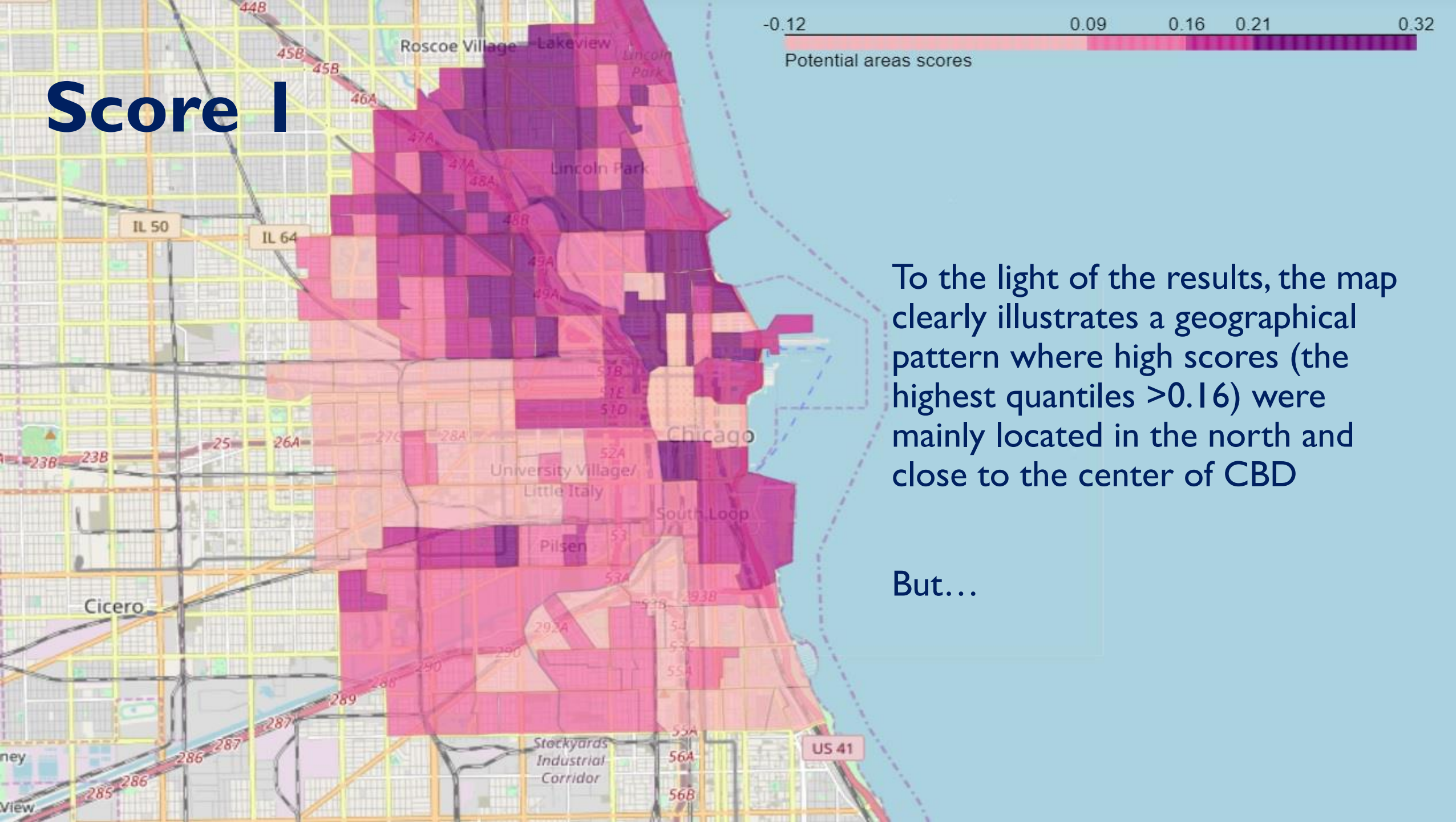
- **Score 2.** Based on string data types. It is also meaningful to research what kind of food predominates in census tracts. If Latin American cuisine is the most common, an entrepreneur would not be very interested in investing in those areas
  - If Latin American restaurants become the first or second most common restaurant in a census tract, then a value of 1 is given.
  - Otherwise, a value of 3 is given.

# Modeling

- **A final score.** It is the sum of both scores. The higher the score, the more potential an area is to open a Peruvian restaurant in Chicago

$$\textit{Final score} = \textit{score 1} + \textit{score 2}$$

# Score I

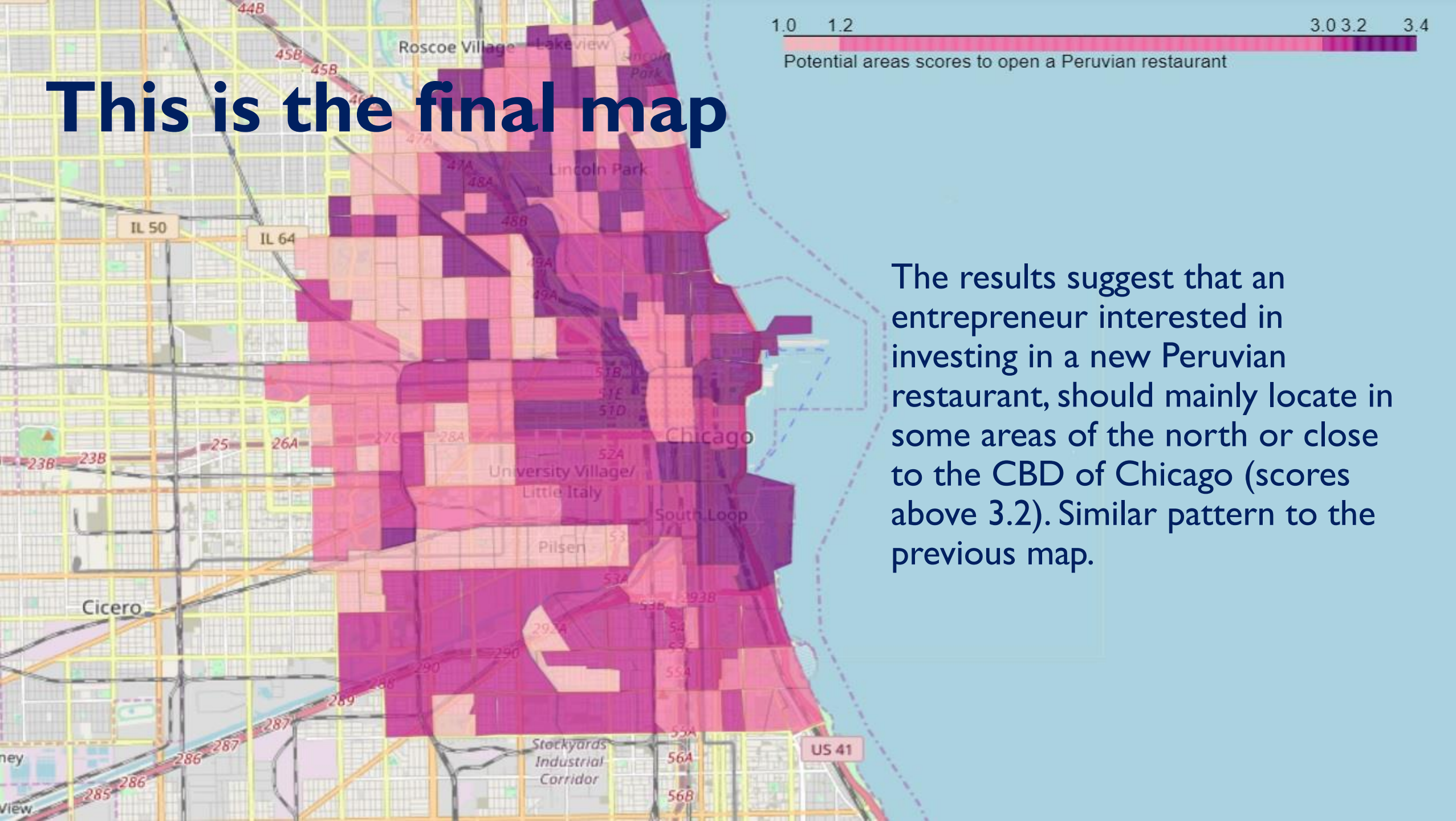


To the light of the results, the map clearly illustrates a geographical pattern where high scores (the highest quantiles  $>0.16$ ) were mainly located in the north and close to the center of CBD

But...



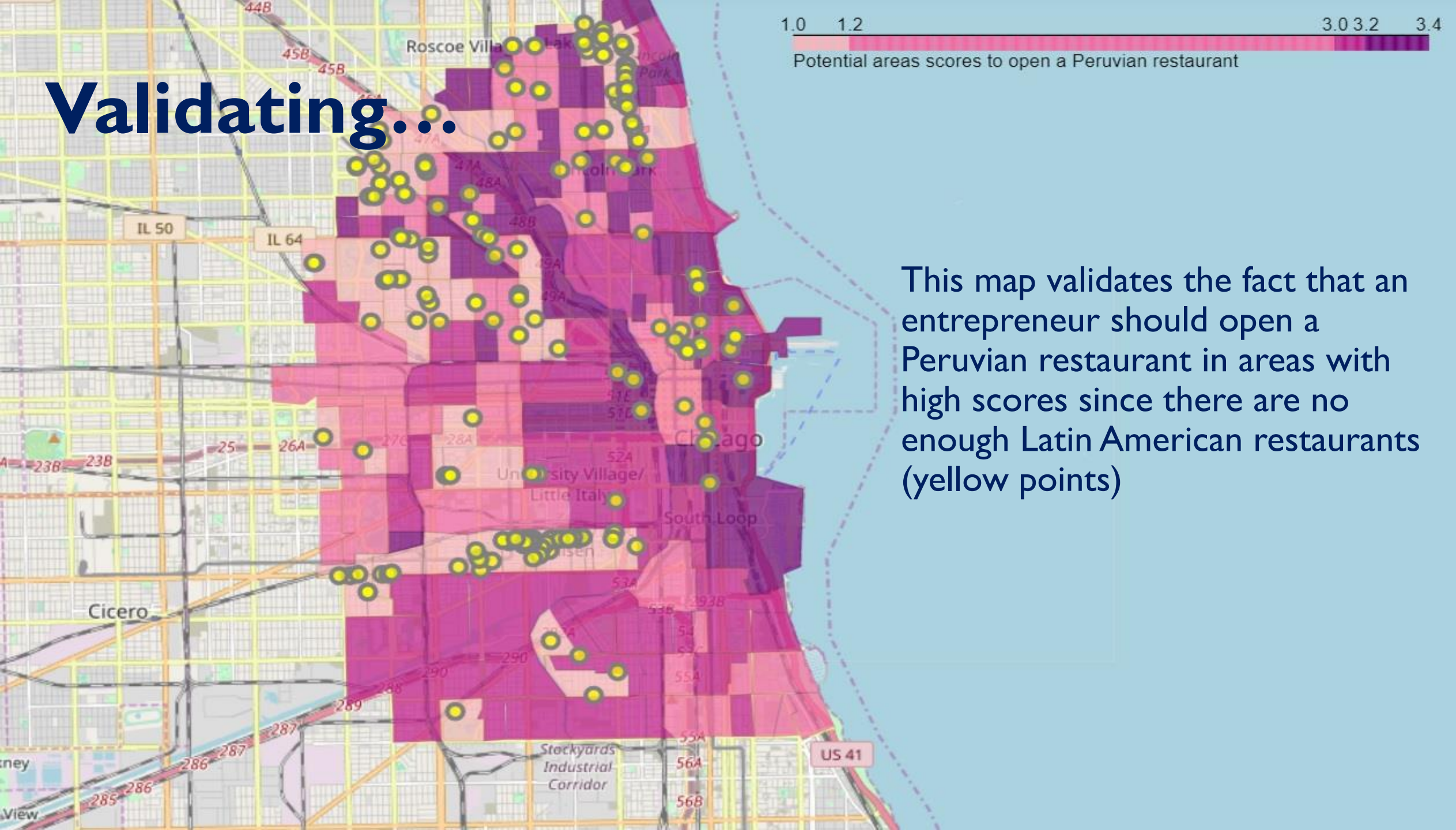
# This is the final map



The results suggest that an entrepreneur interested in investing in a new Peruvian restaurant, should mainly locate in some areas of the north or close to the CBD of Chicago (scores above 3.2). Similar pattern to the previous map.



# Validating...





# Conclusions

- Potential areas in the north and close to the CBD to open a new Peruvian restaurant
- These areas have high income and density population, and low unemployment and crime
- This study is a first approximation to find potential areas and might be improved by considering other variables that can be missing here: food expenses, nearness to Farm markets
- A simple model can tell us a lot!