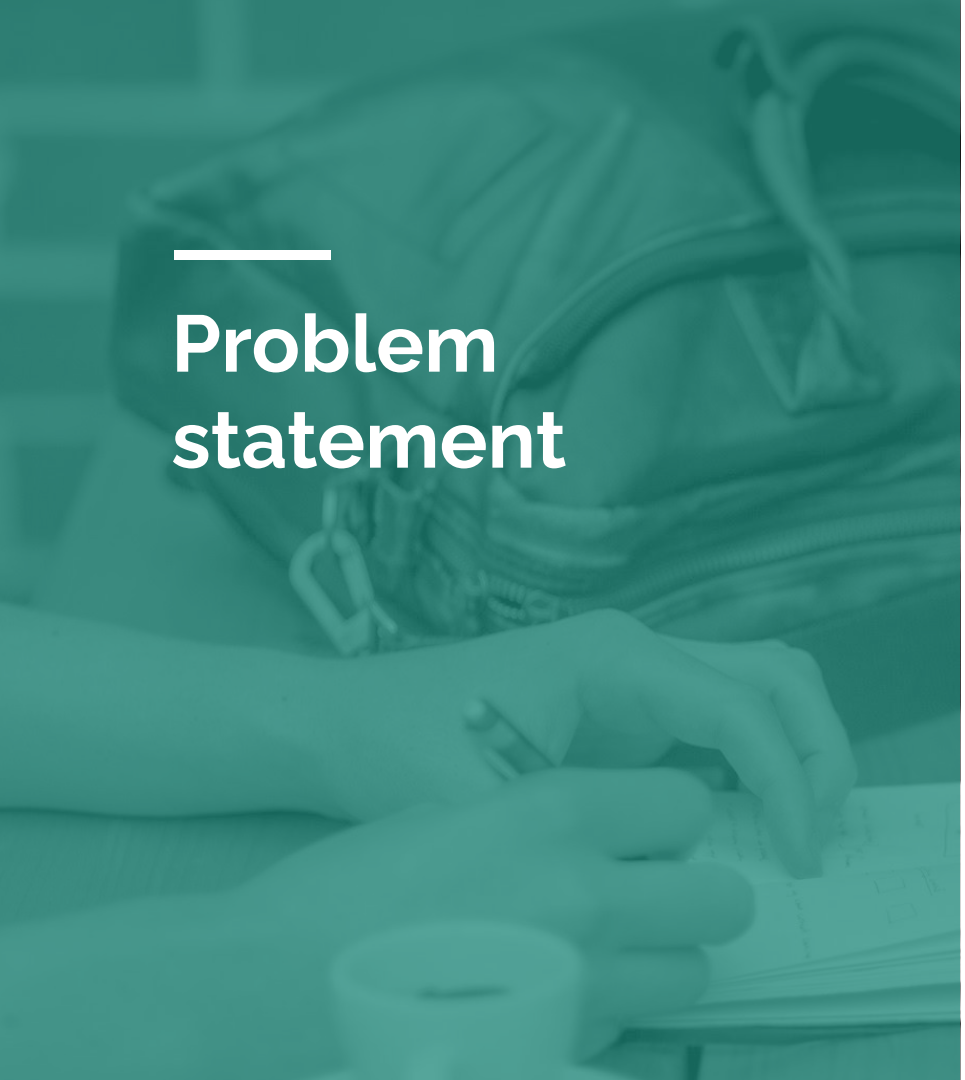
Measuring Culinary Diversity

How to quantify food diversity

Girish Ganesan

Outline


The Problem



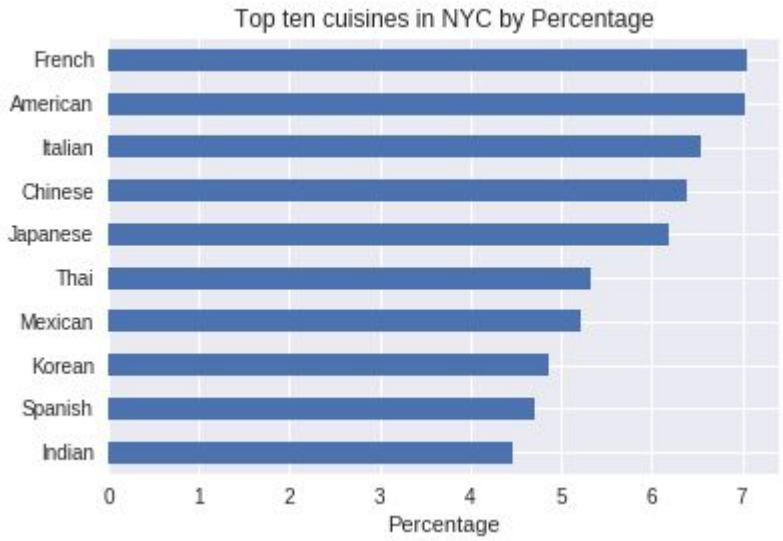
Problem statement

How to quantify “lots of good restaurants” and “many dining options”

We need a quantitative metric to measure diversity; an absolute metric that can measure culinary diversity.

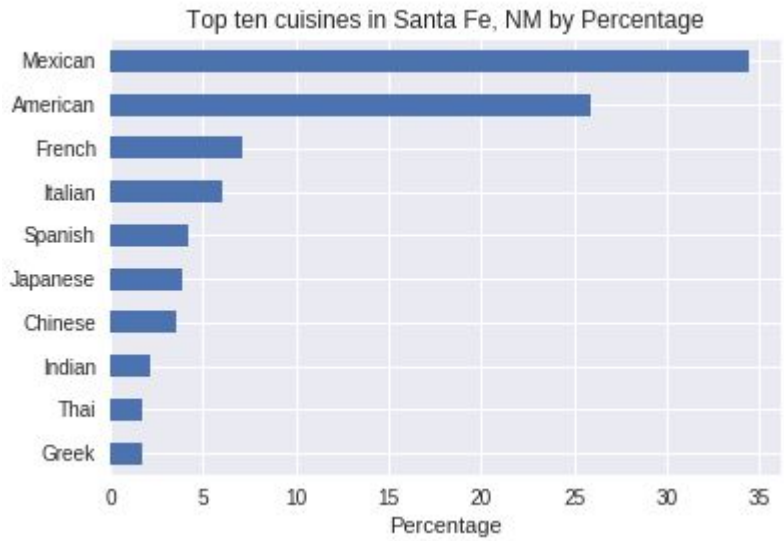
Percentage of Restaurants by Cuisine in New York

Use Foursquare data to calculate the percentage of restaurants by cuisine in New York.



Santa Fe, NM, USA

Smaller town. How do we compare this mix of restaurants with that in New York?



Requirements

- ➔ If cuisines are equally distributed (ideal case) score should be higher
- ➔ Larger the number of cuisines, the higher the score

Solution Proposal

Solution description

Based on entropy derive the Culinary Diversity Index. If the cuisine c_k has a fraction p_k of restaurants:

$$\text{CDI} = -\sum p_k \log_2(p_k)$$

A more interpretable measure is based on the maximum number of cuisines considered (sixt five):

$$\text{Percentage CDI} = 100 * -\sum p_k \log_2(p_k) \,/\, -\log_2(1/65)$$

Advantages

This give a quantitative metric to compare cities.

Measures where the city stands in terms of “absolute best”. (Highest score is 100%).

Results

Data and Methodology

Four cities were considered: New York, San Francisco, Toronto and Santa Fe, NM

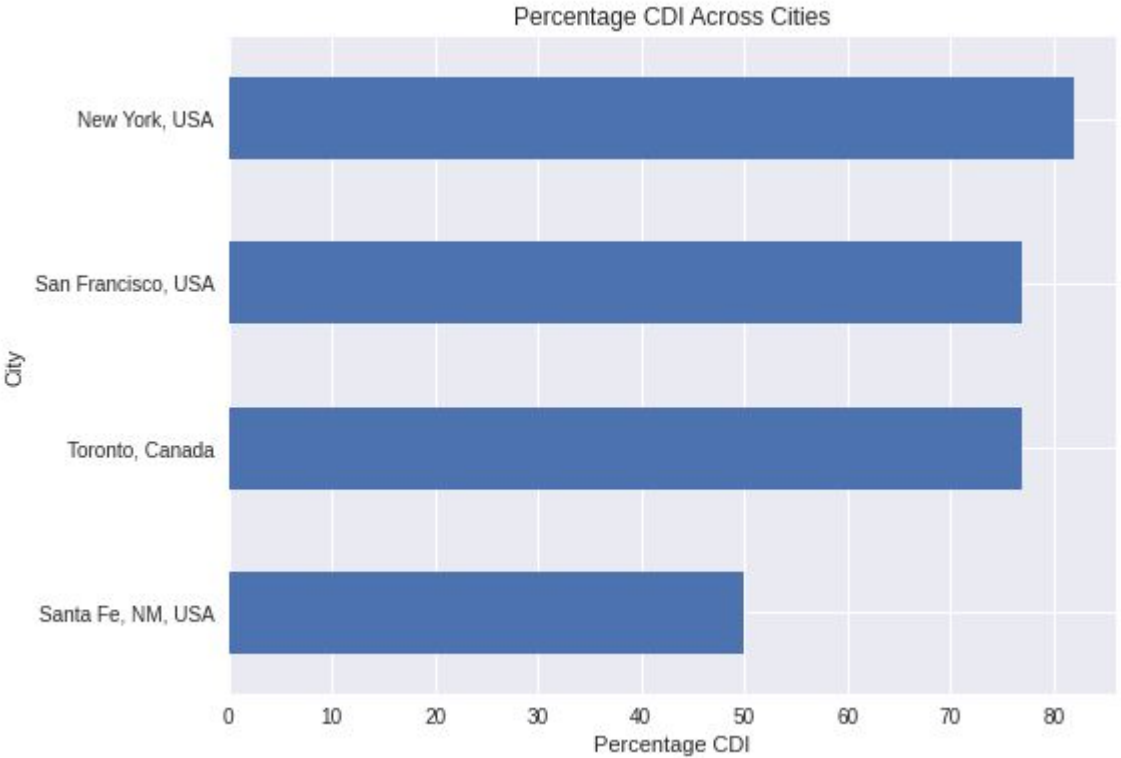
Data based on Foursquare API.

Cuisines are coded based on description from Foursquare developers guide.

Python notebook has the details of cuisines and codes.

Sub-cuisines (like Indian / South Indian) were not considered.

Comparison of Percentage CDI



Insights

New York has the biggest diversity.

San Francisco and Toronto are similar.

Santa Fe, NM has the smallest.

Even though New York is highest, it has room to grow.


Next Steps

What next?

- ➔ Explore if we can find a *utility function* that is a monotonic function of Percentage CDI to measure consumer preferences for diversity.
- ➔ Need to model consumer behaviour.
- ➔ Parameters would probably be estimated via experimentation.