Utilizing Data Science to Satisfy Food Cravings

Peer-graded Assignment: Capstone Project The Battle of Neighborhoods

Christian Kong

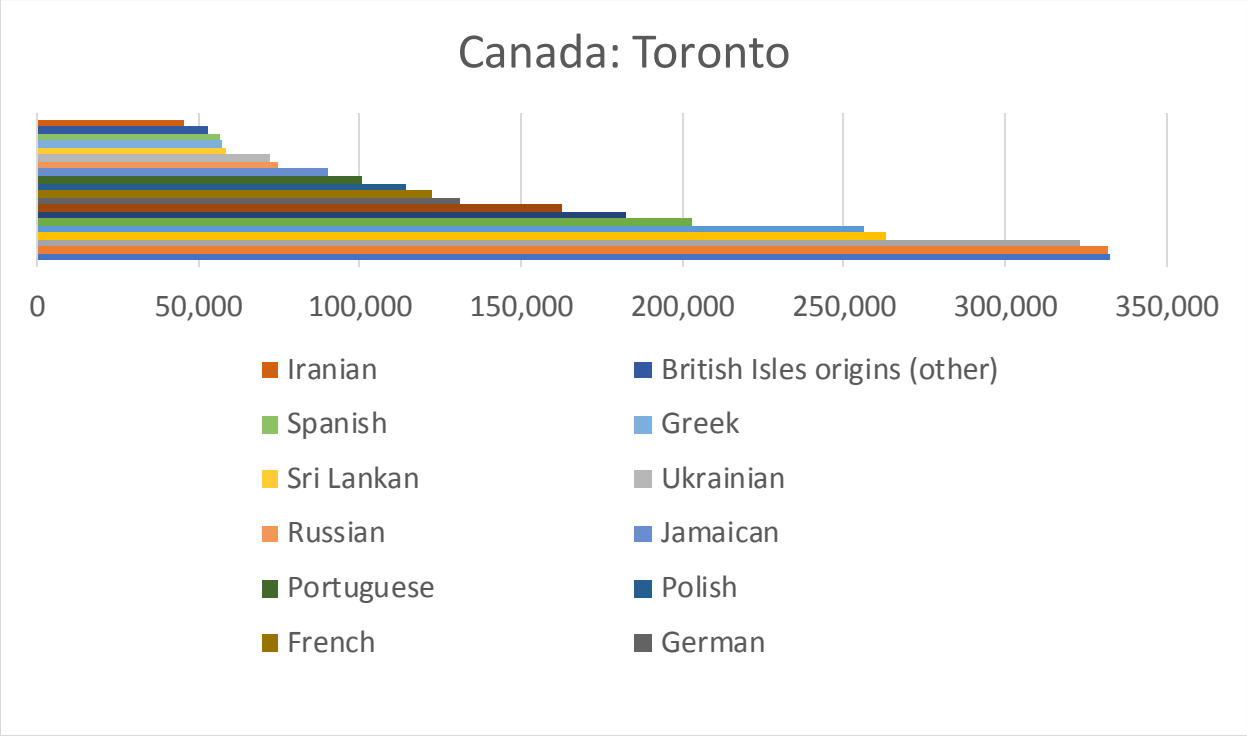
Introduction

It happens in an instant, you or your loved ones are hungry and don't know where to eat. Is that really the problem though? Is it you don't know where to eat or is it really you don't know what to eat due to not knowing the food selection available? Enter the digital age where food delivery services and "online restaurants" place the food at your fingertips, all you have to do is eat. Millions of consumers everyday order food, and food delivery companies need to satisfy their each individual cravings, providing wide selection of choice. Large cities naturally have all sorts of cultural centers, thus providing a wide variety of food choice, but where exactly are the optimal locations to setup these food delivery areas? Utilizing data science, this report will examine 5 of the major cities and their population demographic to provide data food delivery companies focused in North America may find useful. Extreme foodies may also be interested in the data and travel to such locations to get the widest selection of foods

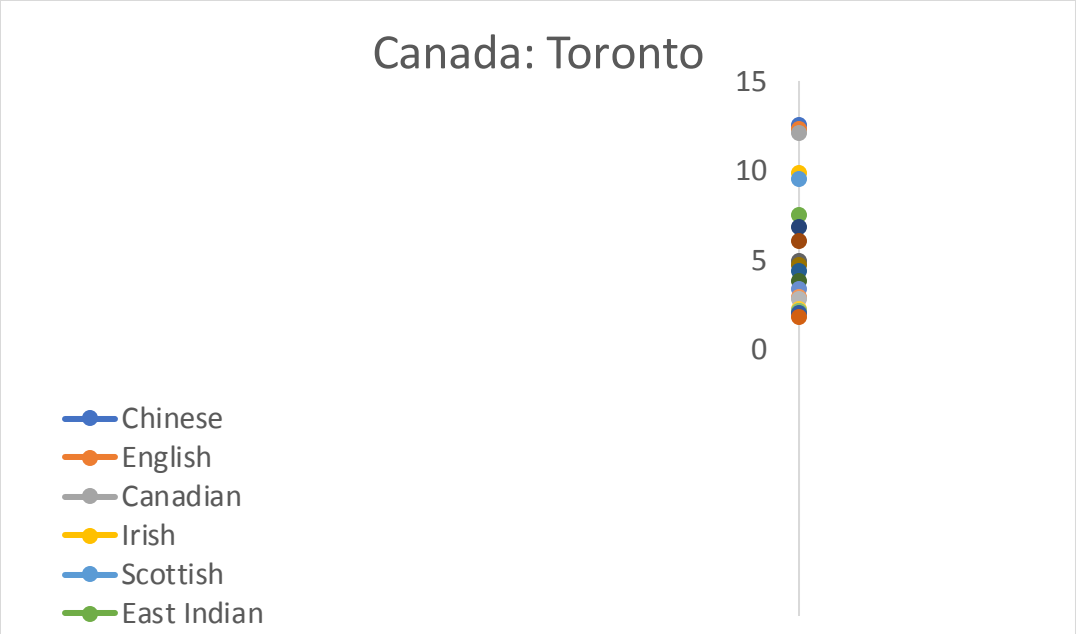
Data & Methodology

Wikipedia incorporates multiple up-to-date sources including "Statistics Canada - Census Subdivision" for Canada and "U.S. Census Bureau" for America, into a clean data table that will be scraped. This report will specifically analyze the demographics of the largest cities in North America and provide the data in a singular table. Unfortunately, Wikipedia provides a lot more information on city demographics than we need, including table categories such as ages and genders which serve no use to us. This report will be deleting and disregarding those sets of data during the final presentation. This removes 8 data sets and a total of 371 individuals data, leaving us with 63 samples. Unlike other data sets with outliers, this report will not be discriminating against certain races thus all races will be included even those with smaller populations.

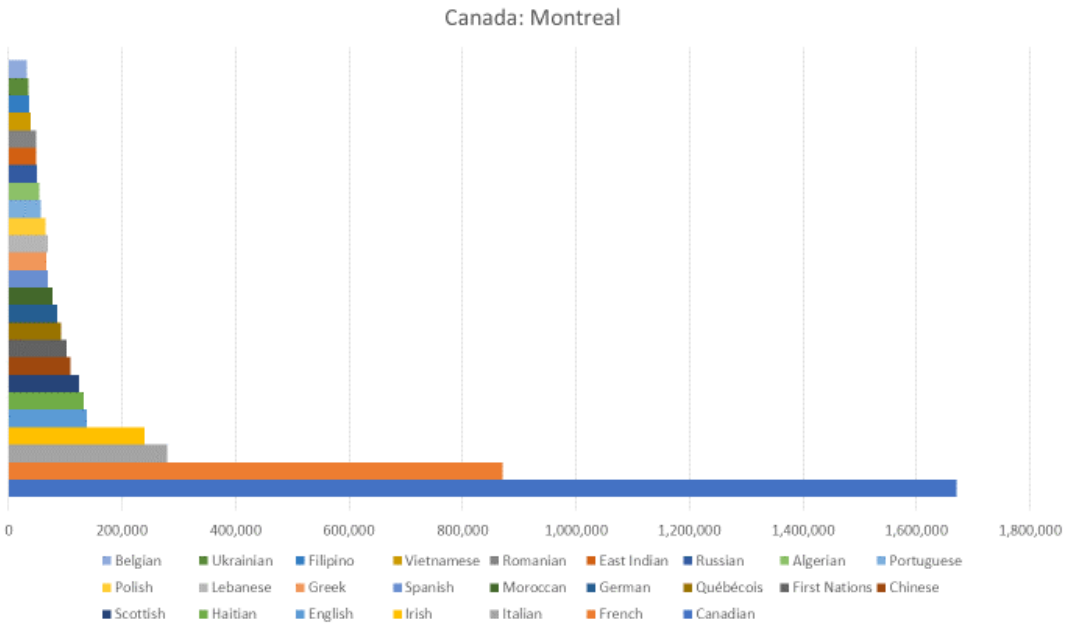
Results & Discussion



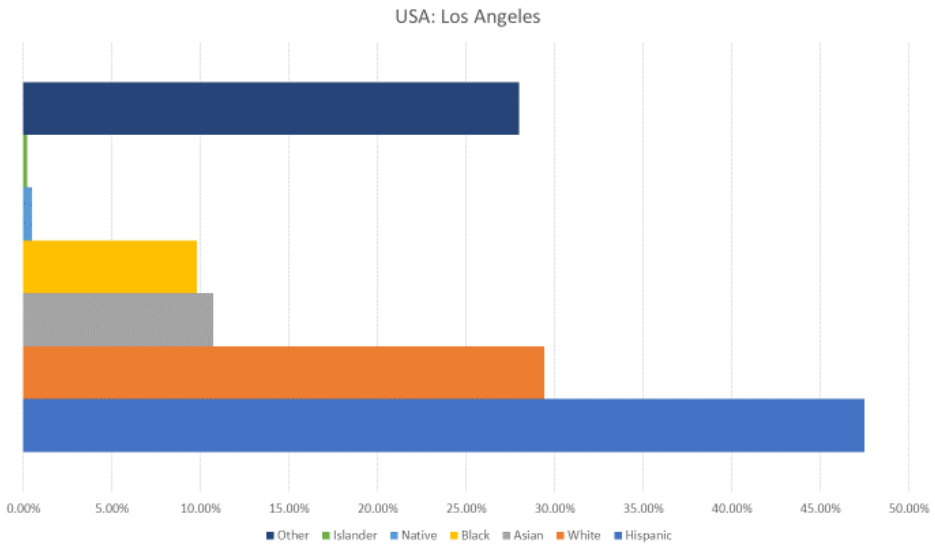
The above chart represents the city of Toronto in Canada, displaying 12 different demographic data, the median demographic being Spanish with +/- 200, 000 individuals. It is clearly shown here that Toronto provides a wide selection of possible food options, although it may be skewed towards some of the larger demographics such as Caucasian or Asian as seen in the relationship below.



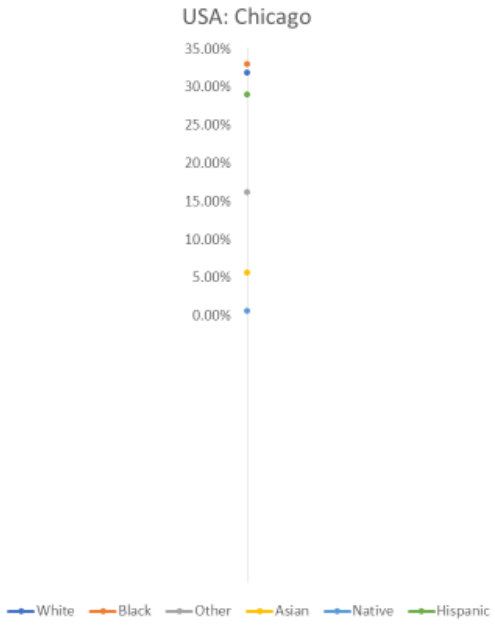
With a more focused relationship on the lower spectrum ranging from +/- 2 - +/- 8 it clearly states although an uneven demographic, the selection of food is larger compared to the upcoming data in the USA.



The above data from Montreal in Canada again shows a wide demographic, however it can be seen the data is has two clear large “outliers” skewing the data with French and Canadian. Although the French is considered the median at approximately +/- 850,000 the mean is more towards the smaller demographic.



Now compared to Canada, USA provides a more focused set of data with less food choices. Comparing Los Angeles to the relationship data below in Chicago. Those under 30% can be seen as the majority-minor outlier.



It can be seen that unlike the relationship in Toronto, USA has a larger spread of almost 12% per data set with larger demographics nearing 35%.

I have noticed that although Toronto has a larger selection of demographics, it may not lead to more selection of foods compared to Chicago due to the clusters of data in +/- 4. With the tight clusters of data, it can be assumed that competition and space will be an issue for some restaurants from opening and competition with one another, resulting in less available food to choose from.

Conclusion

In conclusion, the data may be useful to certain upcoming restaurants in choosing where to setup. Restaurants with "exotic" food may wish not to open in Canada due to the clustering of minor demographics, and Caucasian restaurants may not wish to open in USA due to the overabundance of existing competition. This is the exact opposite of what I would have predicated, as Canada has the wider selection but more difficult for new foods to enter.