Data-driven Recipe to New Cuisine Business at Toronto, ON, Canada - Data

Gilberto da Silva Guimarães Filho

April 23, 2021

1. Data acquisition and cleaning

1.1 Data sources

To have a look at the above said, following data assets could be used:

- Wikipedia's "List of Postal Code of Canada: M": this is a website that give us data about boroughs and neighborhoods with postal codes from Toronto, ON, Canada.
- Foursquare API: this tool allows interested people get location data and other resources by RESTfull API calls. Using this API, I will get some data like categories of business in Toronto and specificities of cuisine entrepreneurship at this city. Beyond this I will use this API to get locations and Toronto restaurant's information.
- For geolocation data: Geopy, a Python library, will be used. This library is especially good in retrieving longitude and latitude data. But, if necessary, some aditional tools will be used as well.
- If geolocation data Python packages does not work well for our purposes, Coursera's Toronto geospatial data will be used to get latitude and longitude at Toronto by neighborhood's postal codes.

1.2 Data cleaning and exploratory data analysis

After web scraping some geospatial data of Toronto from Wikipedia website, the data will be cleaned and then Exploratory Data Analysis will be done by data wrangling. K-means clustering machine learning algorithm will be used to try to find some insights about cuisine entrepreneurship at Toronto.

There are different methods to gather data. This two methods will be used because were studied in this certification process. Beside that Python libraries, packages and community have a lot of material for a faster process. As an example, if you wish to learn more complicated cases of web scraping using the BeautifulSoup Python package, you can start your journey here: http://beautiful-soup-4.readthedocs.io/en/latest/

In these work it will be used some hints form Cousera's course, like hints for web scraping. Pandas, another Python library, or the BeautifulSoup package, or any other way of transforming the data can be used as well.

1.3 Feature selection

In this work the focus will be on international cuisine. Therefore, during data processing, the selection of features will be made, hypothetically, with the choices of restaurants of interest by the customer. In this way, we will restrict our analysis according to its objective, namely: to provide insights for new entrepreneurs in the international restaurant industry in Toronto, ON, Canada.

Kept features	Dropped features
'Caribbean Restaurant', 'Chinese Restaurant', 'Indian Restaurant', 'African Restaurant', 'Filipino Restaurant', 'Greek Restaurant', 'German Restaurant', 'Japanese Restaurant', 'Italian Restaurant', 'Mexican Restaurant', 'Korean BBQ Restaurant', 'French Restaurant', 'South Indian Restaurant', 'Vegetarian / Vegan Restaurant', 'Thai Restaurant', 'Asian Restaurant', 'Hakka Restaurant', 'Halal Restaurant', 'Sushi Restaurant', 'Korean Restaurant', 'Middle Eastern Restaurant', 'Sri Lankan Restaurant', 'Falafel Restaurant', 'American Restaurant', 'Bangladeshi Restaurant', 'Pakistani Restaurant', 'Vietnamese Restaurant', 'North Indian Restaurant', 'Kebab Restaurant', 'Afghan Restaurant', 'Seafood Restaurant', 'Iraqi Restaurant', 'Cantonese Restaurant', 'Hong Kong Restaurant', 'Malay Restaurant', 'Taiwanese Restaurant', 'Dumpling Restaurant', 'Szechuan Restaurant', 'Ramen Restaurant', 'Udon Restaurant', 'Mediterranean Restaurant', 'Dim Sum Restaurant', 'Eastern European Restaurant', 'New American Restaurant', 'South American Restaurant', 'Persian Restaurant', 'Doner Restaurant', 'Tapas Restaurant', 'Cuban Restaurant', 'Brazilian	'Coffee Shop', 'Sandwich Place', 'Fast Food Restaurant', 'Fried Chicken Joint', 'Restaurant', 'Bakery', 'Burger Joint', 'Breakfast Spot', 'Fish & Chips Shop', 'Café', 'BBQ Joint', 'Pizza Place', 'Shawarma Place', 'Donut Shop', 'Grocery Store', 'Salad Place', 'Sports Bar', 'Juice Bar', 'Snack Place', 'Wings Joint', 'Ice Cream Shop', 'Food Truck', 'Deli / Bodega', 'Steakhouse', 'Diner', 'Tea Room', 'Gastropub', 'Cajun / Creole Restaurant', 'Harbor / Marina', 'Bistro', 'Hot Dog Joint', 'Bubble Tea Shop', 'Noodle House', 'Food Court', 'Internet Cafe', 'Event Space', 'Bar', 'Frozen Yogurt Shop', 'Burrito Place', 'Chaat Place', 'Pub', 'Cafeteria', 'Buffet', 'Food Stand', 'Cha Chaan Teng', 'Bagel Shop', 'Pastry Shop', 'Cupcake Shop', 'Butcher', 'Gourmet Shop', 'Lounge', 'Factory', 'Office', 'Irish Pub', 'Hookah Bar', 'Beer Bar', 'Brewery', 'Wine Bar', 'Gelato Shop', 'Comfort Food Restaurant', 'Soup Place', 'Food & Drink Shop', 'Supermarket', 'Poke Place', 'Convenience Store', 'Gay Bar', 'Chocolate Shop', 'Market', 'Gluten-free Restaurant', 'Bowling Alley', 'Taco Place', 'Hotel', 'Cheese Shop', 'Pie Shop', 'Shopping Mall', 'Comic Shop', 'Theme

Restaurant', 'Modern European Restaurant', 'Caucasian Restaurant', 'Latin American Restaurant', 'Turkish Restaurant', 'Portuguese Restaurant', 'Russian Restaurant', 'Bosnian Restaurant', 'Moroccan Restaurant', 'Ethiopian Restaurant', 'Venezuelan Restaurant', 'Egyptian Restaurant', 'Syrian Restaurant', 'Tibetan Restaurant', 'Belgian Restaurant', 'Kosher Restaurant', 'Lebanese Restaurant', 'Dutch Restaurant', 'Peruvian Restaurant', 'Colombian Restaurant', 'Empanada Restaurant', 'Indian Chinese Restaurant', 'Argentinian Restaurant', 'Hungarian Restaurant', 'Burmese Restaurant', 'Polish Restaurant', 'Scandinavian Restaurant', 'Swiss Restaurant', 'Cajun / Creole Restaurant'	Restaurant', 'Dessert Shop', 'Food', 'Hotpot Restaurant', 'Poutine Place'

Table 1 – Feature selection