Assignment

The Battle of Neighbourhoods

Moscow

The Battle of multicultural gastronomic variety

Report

Introduction

Discuss the business problem and who would be interested in this project.

Moscow Metro has more than 250 stations and due to constant road problems and massive traffic jams is the best way of transportation so if you decide to go out metro is often the best way to get to a place. But which metro station is the best if you want to have a nice restaurant experience with diverse chose of restaurants with many different national and international cuisines?

The main purpose of this project is to determine which neighbourhoods surrounding which metro stations in Moscow offers the biggest number of restaurants with the largest variety of national and international cuisines. In other words we are looking for a metro station that can offer us the most multicultural gastronomic diversity experience and choice when deciding on a restaurant.

This information can be used by both tourists and locals to chose a metro station from where they will have the biggest number of choices for a restaurant.

Another purpose of this project is on the contrary to determine Moscow neighbourhoods with the smallest diversity and/or lack of certain ethic foods and restaurants representation.

This information can assist business people in opening a new restaurant and deciding which national or international cuisine to go for and where to open it.

We will use Foursquare location data with basic descriptive statistics as well as some unsupervised machine learning with cluster analysis to classify stations and derive the information and conclusions we need.

All restaurants must be located within just 500 meters walk from the metro station so it could be an easy walk.

Data

Describe the data that you will be using to solve the problem or execute your idea and the source of the data. Remember that you will need to use the Foursquare location data to solve the problem or execute your idea.

The base list of metro stations and their geographical coordinates are scraped from this Wikipedia page: https://en.wikipedia.org/wiki/List_of_Moscow_Metro_stations

	L	English transcription	Russian Cyrillic	Transfer	Opened	Elev.	Type	Coordinates	Pic.
0	NaN	Bulvar Rokossovskogo	Бульвар Рокоссовского	<>	1990-08-01	−8 m	column, triple-span	55°48′53″N 37°44′03″E / 55.8148°N 37.7342°E	NaN
1	NaN	Cherkizovskaya	Черкизовская	{}	1990-08-01	−9 m	single-vault, shallow	55°48′14″N 37°44′41″E / 55.8038°N 37.7448°E	NaN
2	NaN	Preobrazhenskaya Ploshchad	Преображенская площадь	NaN	1965-12-31	−8 m	column, triple-span	55°47′47″N 37°42′54″E / 55.7963°N 37.7151°E	NaN
3	NaN	Sokolniki	Сокольники	NaN	1935-05-15	−9 m	column, triple-span	55°47′20″N 37°40′49″E / 55.7888°N 37.6802°E	NaN
4	NaN	Krasnoselskaya	Красносельская	NaN	1935-05-15	−8 m	column, double-span	55°46′48″N 37°40′02″E / 55.7801°N 37.6673°E	NaN

Then using Foursquare API the base list of venue types is received:

```
4d4b7104d754a06370d81259 Arts & Entertainment
4d4b7105d754a06372d81259 College & University
4d4b7105d754a06373d81259 Event
4d4b7105d754a06374d81259 Food
4d4b7105d754a06376d81259 Nightlife Spot
4d4b7105d754a06377d81259 Outdoors & Recreation
4d4b7105d754a06375d81259 Professional & Other Places
4e67e38e036454776db1fb3a Residence
4d4b7105d754a06378d81259 Shop & Service
4d4b7105d754a06379d81259 Travel & Transport
```

We can see that `4d4b7104d754a06370d81259` represents Food and that is what we are going to be using to filter only food venues data from Foursquare.

We will use that category ID to first get all food places per each metro station:

```
radius = 500 # define radius
categoryId = '4d4b7105d754a06374d81259' # category ID for "Food"

url = 'https://api.foursquare.com/v2/venues/search?
&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&categoryId={}&limit={}'.format(
    CLIENT_ID,
    CLIENT_SECRET,
    VERSION,
    moscow_metro_station.Latitude,
    moscow_metro_station.Longitude,
    radius,
    categoryId,
    LIMIT)
```

This Foursquare request will result in something like this:

```
{'id': '5baca878f427de002cfcbc17',
 'name': 'Мистер Круассан',
 'location': {'address': 'Большая Черёмушкинская улица, 1',
  'lat': 55.690278,
  'lng': 37.601879,
  'labeledLatLngs': [{'label': 'display',
    'lat': 55.690278,
    'lng': 37.601879}],
  'distance': 198,
  'postalCode': '117447',
  'cc': 'RU',
  'city': 'Москва',
  'state': 'Москва',
  'country': 'Россия',
  'formattedAddress': ['Большая Черёмушкинская улица, 1',
   '117447, Москва',
   'Россия']},
 'categories': [{'id': '4bf58dd8d48988d1e0931735',
```

```
    'name': 'Coffee Shop',
    'pluralName': 'Coffee Shops',
    'shortName': 'Coffee Shop',
    'icon': {'prefix': 'https://ss3.4sqi.net/img/categories_v2/food/coffeeshop_',
     'suffix': '.png'},
    'primary': True}],
  'referralId': 'v-1582192970',
  'hasPerk': False},... JSON of all food venues for each metro station;
```

We will later clean-up this data to only have a list of ethnic type restaurants with clear national/international cuisine.

We will need only food_venue_item['categories'][0]['name'] attributes values for our filter.

The goal is then to merge the two data sources and to get them into a pandas DataFrame which would show all metro stations with their latitude and longitude from the Wikipedia page and for each metro station show all associated food venues like that:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bulvar Rokossovskogo	55.8148	37.7342	Burger King	55.814026	37.733659	Fast Food Restaurant
1	Bulvar Rokossovskogo	55.8148	37.7342	Broker Coffee	55.813895	37.733526	Coffee Shop
2	Bulvar Rokossovskogo	55.8148	37.7342	шаурма	55.816166	37.730582	Shawarma Place
3	Bulvar Rokossovskogo	55.8148	37.7342	Китайская Кухня "Лотос"	55.813580	37.732730	Chinese Restaurant
4	Bulvar Rokossovskogo	55.8148	37.7342	Подсолнухи Art&Food	55.816065	37.736457	Food Court

That DataFrame would then, as already mentioned, be filtered down to only ethnic type restaurants with clear national/international cuisine and will exclude things like Fast food places, Coffee Shops, Food courts etc. like so:

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Bulvar Rokossovskogo	55.8148	37.7342	Китайская Кухня "Лотос"	55.813580	37.732730	Chinese Restaurant
Bulvar Rokossovskogo	55.8148	37.7342	Фо & Ролл	55.815955	37.736421	Vietnamese Restaurant
Bulvar Rokossovskogo	55.8148	37.7342	Суши Wok	55.814660	37.731430	Asian Restaurant
Bulvar Rokossovskogo	55.8148	37.7342	El Taco	55.815503	37.737244	Tex-Mex Restaurant
Bulvar Rokossovskogo	55.8148	37.7342	СушиStore	55.814618	37.730911	Sushi Restaurant
Bulvar Rokossovskogo	55.8148	37.7342	академия плова	55.816015	37.736523	Asian Restaurant
Bulvar Rokossovskogo	55.8148	37.7342	Churros House	55.816125	37.736488	Mexican Restaurant
Bulvar Rokossovskogo	55.8148	37.7342	Ksusha Kitchen	55.816019	37.736674	Italian Restaurant

This resulting DataFrame would allow us to achieve the goad which we set in the introduction section, which is to quantify and classify ethnic cuisine diversity available within 500 meters reach of each Moscow metro station.