

# Clustering Russian cities to determines food tastes in different regions.

Alexander V. Ivanov

11.08.2020

## ***1. Introduction***

### 1.1 Background

Russia or the Russian Federation is one of the largest country in the world area. Its territory extends from Baltic Sea in the west to the Pacific Ocean in the east.

Russia is the most populous nation in Europe. It is a multi-national state with over 185 ethnical groups, the population of those groups very enormously, from millions (Russian and Tatars) to under 10.000(Samis and Eskimos)

### 1.2 Problem

Because Russia's multi-national and wide-spread territory Russian's people tastes can vary from one region to another. So if your company is big and you what to expand your restaurant busines you need to know what kind of food/kitchen people prefer in desirable region.

## ***2 Data acquisition and cleaning***

### 2.1 Data sources

In order to provide information about Russian people tastes in different regions we need collect data containing Russian cities including theirs coordinates and we need information about food points in these cities such as restaurants, café, etc.

And we need to know type of these point, what kind of food people prefer.

For Russian cities we will use information from Wikipedia page:

[http://en.wikipedia.org/wiki/List\\_of\\_cities\\_and\\_towns\\_in\\_Russia\\_by\\_population](http://en.wikipedia.org/wiki/List_of_cities_and_towns_in_Russia_by_population)

We will read it into pandas data frame. To determine cities coordinates we will use Nominatim API.

All data related to the food points will be obtained via the FourSquare API.

So we build pandas dataframe containing cities, its coordinates and food points in these cities. After we build dataframe, we can cluster and visualize all this information to determine food tastes in different Russian cities or regions.

### 2.2 Data cleaning

In Wikipedia page we need only City Name. So we need to drop all other columns.

And we need to correct some city coordinates, such as Saint Petersburg, because derived from Nominatim coordinates do not pointing to the city center.