Segmenting and Clustering potencial pharmacies spots in Los Angeles, CA¶

Introduction and instructions for the tasks

1. Clearly define a problem or an idea of your choice, where you would need to leverage the Foursquare location data to solve or execute. Remember that data science problems always target an audience and are meant to help a group of stakeholders solve a problem, so make sure that you explicitly describe your audience and why they would care about your problem.
2. This submission will eventually become your Introduction/Business Problem section in your final report. So I recommend that you push the report (having your Introduction/Business Problem section only for now) to your Github repository and submit a link to it
3. Describe the data that you will be using to solve the problem or execute your idea. Remember that you will need to use the Foursquare location data to solve the problem or execute your idea. You can absolutely use other datasets in combination with the Foursquare location data. So make sure that you provide adequate explanation and discussion, with examples, of the data that you will be using, even if it is only Foursquare location data.

Introduction

The global pharmaceutical market is expected to overcome $1.5 trillion by 2023, from the $1.2 trillion of 2018, at an annual growth rate of 3-6%. This increase should be paralleled by the increase in the number of launched products, that might reach 54 in 2023.
Competition in biosimilars is also expected to increase by three times, while specialty share of total medicine spending will reach 50% by 2023 in most developed markets. These scenarios are illustrated by the recent IQVIA Institute of Human Data Science "The Global Use of Medicine in 2019 and Outlook to 2023".

Explaining the task

A foreign group would like to install some pharmacies in LA in order to gain customers in not foreseen areas. This task was performed aiming to execute an areal study of new places and then expand pharmacies in Los Angeles.

Some places the number of pharmacies are fewer or almost inexistent.Due this It was performed a areal studies and suggested some conclusions using Machine Learning tools. The term "k-means" was first used by James MacQueen in 1967, [1] though the idea goes back to Hugo Steinhaus in 1956.[2] The standard algorithm was first proposed by Stuart Lloyd of Bell Labs in 1957 as a technique for pulse-code modulation.

k-means clustering is a method of vector quantization, originally from signal processing, that aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean (cluster centers or cluster centroid), serving as a prototype of the cluster. This results in a partitioning of the data space into Voronoi cells. k-means clustering has been used as a feature learning (or dictionary learning) step, in either (semi-)supervised learning or unsupervised learning. Reference
For this case was necessary to download the dataframe of neighborhood and their respective coordinates. Then the search was complemented by the venues rank by the FourSquare API.

Complementary References

[1](MacQueen, J. B. (1967). Some Methods for classification and Analysis of Multivariate Observations. Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability. 1. University of California Press. pp. 281–297. MR 0214227. Zbl 0214.46201. Retrieved 2009-04-07.)

[2](Steinhaus, Hugo (1957). "Sur la division des corps matériels en parties". Bull. Acad. Polon. Sci. (in French). 4 (12): 801–804. MR 0090073. Zbl 0079.16403.)

All this work was performed inspiring the exercicies done which comprehend the Module "Applied Data Science Capstone"from Coursera (Data Science Professional Specialization)