Coursera Capstone Project for IBM Data Science Professional Certificate - Battle of sporty neighbourghoods - Presentation

By COIFFET Jean-sébastien

Introduction and problematic

Considering that France will organise summer olympic games 2024, a craze for sport (at all levels of practice) could be expected in France in the next years. This project aim on modeling a neighborhood recommender for someone who want to make sport in one of the two main cities of the "Pays de la Loire" french region :

Nantes metropole (department 44) which is the regional capital and Angers metropole (department 49).

The study will focus on the following 3 sports : Football, Basketball and Swimming.

I've chosen these sports because first they are olympic sports, second because they need equipments and third because i estimate them interesting for the reader and/or for me.

Football and basketball are the two more practiced sports in the region (see Bibliography/Input_PDF_1).

I will consider 2 levels of practice for each sport : Leisure and Competition.

So the initial problem will be to find where a sporty person should settle in a given perimeter (Nantes alone, Angers alone and finally Nantes+Angers), depending on his/her favorite sports and the level of practice.

SOURCE DATAS (1/2)

INPUT_CSV_1 : Nantes Métropole official postal code database (544 records)

https://data.nantesmetropole.fr/explore/dataset/244400404_base-codes-postaux-nantes-metropole/information/?disjunctive.code_commune_insee&disjunctive.nom_de_la_commune&disjunctive.code_postal&disjunctive.ligne_5

INPUT_CSV_2 : NANTES Metropole Data Portal/Region_Sport_Equipments (8981 records) - 234400034_004-009_activites-des-fiches-equipements-rpdl_CL (Competition/Leisure)

https://data.nantesmetropole.fr/explore/dataset/234400034_004-009_activites-des-fiches-equipements-rpdl%40paysdelaloire/information/?disjunctive.comlib&disjunctive.actlib&disjunctive.actnivlib

SOURCE DATAS (2/2)

Main sport clubs of Nantes and Angers :

FOOT/NANTES : https://fr.wikipedia.org/wiki/Football_Club_de_Nantes at Training center of La Jonelière (GPS : 47.26003625074568, -1.5431419374263136)

FOOT/ANGERS : https://fr.wikipedia.org/wiki/Angers_sporting_club_de_l'Ouest at Training center of La Baumette (GPS : 47.46022228358844, -0.5687972904023035)

BASKET/NANTES : https://fr.wikipedia.org/wiki/Nantes_Basket_Hermine at Sports Complex of Rézé (GPS : 47.1804447592872, -1.5716645075361122)

BASKET/ANGERS : https://fr.wikipedia.org/wiki/%C3%89toile_Angers_Basket at Sports Complex Jean Bouin (GPS : 47.460179020489846, -0.5313707516520269)

SWIMMING/NANTES : Nantes Natation at Leo Lagrange swimming pool (GPS : 47.210138455645044, -1.5594551310482154)

SWIMMING/ANGERS : Angers Natation at Jean Bouin swimming pool (GPS : 47.459135380533176, -0.5321586598738427)

BIBLIOGRAPHY

Input_PDF_1 : SPORTS IN PAYS DE LA LOIRE :

   https://www.sraenutrition.fr/wp-content/uploads/2020/09/Fiche-sport-Pays-de-la-Loire-2020.pdf

Input_PDF_2 : SPORTS EQUIPMENTS IN FRANCE :

   https://injep.fr/wp-content/uploads/2020/07/FR49_lieux_pratiquesportive.pdf (2018)

METHODOLOGY

In this study, i will first do data cleaning, data exploration and then use K-Mean clustering to determine where a sporty person should settle in Nantes and then in Angers, depending on his/her favorite sport(s) and level of practice.

I may also consider the situation of a family including two sporty people wanting each 1 of the 3 studied sports but i will not consider the situation of a family including three or more sporty people wanting each differents sports.