

Coursera Capstone Project for IBM Data Science Professional Certificate - Battle of sporty neighbourhoods - 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 french region "Pays de la Loire" :

- Nantes, regional capital
- Angers, second city by size of the region

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 two levels of practice for each sport : Leisure and Competition.

So the initial problem will be to find in which district of a city a sporty person should settle, 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 but i will use only 25)

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) - gives both Nantes and Angers sports equipments.

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

INPUT_CSV_3 : For Angers postal codes : <https://www.data.gouv.fr/es/datasets/base-officielle-des-codes-postaux/>

INPUT_CSV_4 : Demographic infos on Loire Atlantique, department of Nantes and on Maine et Loire, department of Angers metropolises :

<https://www.insee.fr/fr/statistiques/4989704?sommaire=4989761>

SOURCE DATAS (2/2)

MAIN_CLUBS_1 - Main sport clubs of Nantes and Angers :

FOOT/NANTES : https://fr.wikipedia.org/wiki/Football_Club_de_Nantes at Training center of La Jonelière
District La Chapelle-sur-Erdre / GPS : 47.26003625074568, -1.5431419374263136

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

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

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

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

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

METHODOLOGY

In this study, i will do for each of the 2 city :

- Data cleaning, data exploration
- A K-Mean clustering
- Another data exploration for the most interesting cluster of the city
- Discussion on city results
- Conclusion for the city

At the end i will also do a global conclusion on both cities.

The goal is to determine, where a sporty person should settle, depending on his/her favorite sport(s) and level of practice.

Technically, I will use Anaconda 3.8 with python 3.8, jupyter notebooks 6.1.4 and Folium library for the map. Folium is using OpenStreetMap.

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)