ANTONIO J. ROLDÁN CASTRO

IBM Professional Certificate on Data Science

Coursera Capstone

Opening a Multiplex or Movie Theater in Madrid, Spain

Introduction

Despite the current trend of appearing new VOD platforms in Spain and worldwide, such as Netflix, Disney +, HBO, etc., the business of entertainment and theaters still generates an average of Box Office of 600 million € per year in Spain with an upward trend since few years ago (without contemplating the current Covid situation).
The city of Madrid gathers itself approx. the 20% of the total Box Office of films in the national territory.

Business Problem

With the mentioned situation, we'll intent to realize an initial approximation to analyze the best (if any) location of a new profitable Multiplex or Movie Theaters in Madrid (Spain) using data science tools and methodologies and machine learning techniques such as clustering data.

Target of the project

This report will be useful as a first study to analyze the different neighborhoods of Madrid in terms of density of theater nearby each one of them, to identify the best initial position to open a new theater.

Data and Documentation Needed

- List of neighborhoods in Madrid, Spain.

- Location of each neighborhood: latitude and longitude.

- Venue data (theaters and Multiplexes) to enumerate the amount of theaters around each node (neighborhood).

Sources of data and software

- Wikipedia pages to extract information of the Neighborhoods of Madrid:

    o https://en.wikipedia.org/wiki/Districts_of_Madrid.

- Python, Jupyter Notebooks and packages related to:

    o Python Geocoder (to extract coordinates)

    o Plotting maps (Folium)

    o Reading of websites (html)

    o K-means and Cluster analysis

- Foursquare API

    o To get the information and location of the theaters