

# **COURSERA CAPSTONE**

## **Opening a Movie Theater in Madrid (Spain)**

**ANTONIO ROLDÁN**

**IBM – PROFESSIONAL CERTIFICATE IN DATA SCIENCE**

# BUSINESS PROBLEM

- The industry of Movie Theaters and Entertainment in Spain has an average size of 600 million € per year despite the current trend of rising VOD platforms worldwide.
- Madrid gathers about 23% of the total Box Office of the national territory
- Try to find the best location (if any) to open a new Movie Theater in the city of Madrid (Spain)

# DATA

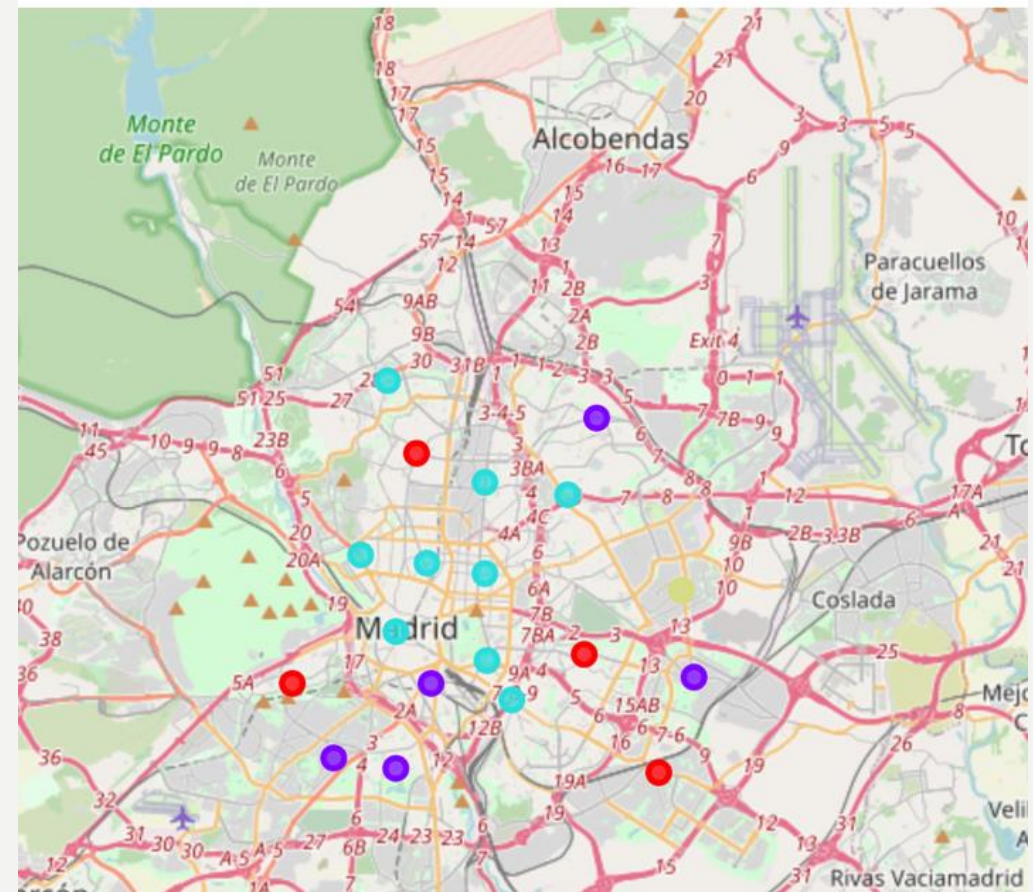
- Data Required
  - List of Neighbourhoods in Madrid
  - Coordinates of Neighbourhoods
  - Venue Data (Movie theaters) nearby each district in Madrid.
- Sources of Data
  - Wikipedia: [https://en.wikipedia.org/wiki/Districts\\_of\\_Madrid](https://en.wikipedia.org/wiki/Districts_of_Madrid).
  - Geocoder Package
  - Foursquare API
  - IBM Watson Studio + Python + Jupyter Notebook

# METHODOLOGY

- Data Scrapping from Wikipedia with Python Pandas
- Cleaning and sorting the data
- Getting location coordinates with Geopy Pandas
  - (In our case: create CSV manually with correct coordinates)
- Creation of a Map with the Neighbourhoods of Madrid with Folium Map
- Obtaining data of Movie Theaters venues from FOURSQUARE API
- Analysing the data with One-Hot Analysis: frequency of occurrence by District.
- Sorting the districts in clusters using K-means
- Creation of a Map with the Neighbourhoods of Madrid divided into the K-means categories.
- Obtaining the results

# RESULTS

- Districts divided into four categories:
  - Cluster 0: District with the highest number of movie theaters. Red color
  - Cluster 1: Low to Moderate number of Movie Theaters. Purple.
  - Cluster 2: Moderate to high number of Movie Theaters. Green.
  - Cluster 3: Low or no number of Movie Theaters occurrence. Yellow



# RECOMMENDATIONS AND CONCLUSIONS

- Concentration of cinemas at the center of the city (cluster 2).
- One market opportunity on the Cluster 3 (Yellow) on the District of San Blas-Canillejas with a population of over 150K citizens.
- Recommendations: consider this first approach as an initial finding to cross more data to interpret the results: average income of the district, habits, social class, etc