**ACME’s new offices:**

*Identify and recommend the best neighbourhood for a corporate expansion in Europe*

**Paul Beauregard**

**March 2020**

**1. Introduction**

* 1. **Background**

The conglomerate ACME, major player on the U.S. market wants to extend its presence in Europe, and open 3 brand new offices. The targeted markets are the Portuguese, Spanish and Dutch one.

I have been contacted to undertake an analysis of the different areas that would be the most suitable for the new ACME offices in Lisbon, Madrid and Amsterdam.

**1.2 Problem**

ACME's head office is located in the neighbourhood of SoHo in Manhattan and is composed of over a thousand employees. It has been voted the most enjoyable place to work in the U.S two years in a row, and location was a major player amongst others.

It has been internally approved to establish the new premises in the respective European capitals, but the HR department at ACME fears that the initial culture shock might affect negatively the employees' productivity. Reason is that the teams transitioning from the ACME head office in New-York, to Europe, are composed of U.S. citizens, unfamiliar to European culture.

The objective of this analysis, stated by ACME, is to find an adequate neighbourhood in each destination that will match the standards and atmosphere that SoHo provides.

**2. Data Acquisition and Cleaning**

The datasets I am using are a list of Lisbon, Madrid and Amsterdam neighbourhoods found on Wikipedia.

Once scraped with the BeautifulSoup package, the datasets were combined into one pandas data frame, to which I appended Soho Manhattan in order to have a future reference.

Madrid’s data set came with the city name, neighbourhood name, population index, as well as the area dimensions. Population index and neighbourhood surface were dropped as they are not essential to obtain the neighbourhoods’ postcodes. I then merged the city and neighbourhood names to prepare the next step.

After using the geopy package to retrieve the latitude and longitudes, I have been parsing each entry into a API call using the free tier Foursquare API, to retrieve the top 100 venues, in each neighbourhood, in a radius of 500 meters.

Next step was to calculate the frequency of each venue category for each neighbourhood and define a top 10, last step to prepare our dataset for our K-Means algorithm.

**3. Clustering with K-Means**

**3.1. Clusters**

In order to find the ideal number of clusters to pass into our K-Means clustering algorithm, I used the Elbow Method to plot the different values of K errors, for different number of clusters. As show in *fig1*, the elbow was not very easy to determine, and was finally identified at 13.

A screenshot of a cell phone

Description automatically generated

**3.2. K-Means**

Now that I have selected the number I want to go with (13), I just ran my algorithm to assign a cluster to each of the neighbourhoods in my dataset. Next step is to identify the cluster where SoHo is in, and list other neighbourhood in that same cluster, to provide a list of matching areas.

**4. Conclusion**

My recommendation to ACME is to focus their searches on these areas, in order to better fit the expectation of their moving employees:

A screenshot of a cell phone

Description automatically generated