

Using machine learning and visualization  
to identify the geolocation patterns of  
customer preferences in Cyprus

## Contents

<b>1. Introduction .....</b>	<b>3</b>
<b>2. Aim and objectives .....</b>	<b>3</b>
<b>3. Data and tools.....</b>	<b>4</b>
<b>4. Libraries Which are Used to Develop the Project:.....</b>	<b>4</b>

# **1. Introduction**

Cyprus is one of the 27 members of the European Union. The island currently has around 1 million population, that is distributed in 4 main cities and many villages, with individuals having minimum knowledge of the internet. In addition, due to the old infrastructure used by the country, it is very difficult to acquire relevant data for a business problem and further the accuracy of the data might be questioned. With the two aforementioned issues, the Cypriot community struggles to assess issues using analytics and statistical models.

With the use of recent software, such as Google maps APIs, 2GIS and Foursquare, it is more convenient to generate specific geographical information that will provide the opportunity for further analysis using analytical tools and software.

To that instance, this project will aim to gather all the areas in Cyprus, including main cities, as well as villages, and deploy a visualization of the island, providing crucial information such as the most visited venues of the whole island.

## **2. Aim and objectives**

The aim of this project is to generate data through the utilization of methods and software that will be used to provide analytical insights for geolocation-based data.

1. Locate available data that can be used as a basis for geolocation
2. Using APIs such as Google Maps API, 2GIS API or Foursquare, to generate more specific data such as longitude and latitude of each village and other areas in Cyprus.
3. Develop visualization of geographical data that will be used as complementary findings to stakeholders for better understanding
4. Identify patterns in customer preferences and provide suggestions to potential stakeholders regarding the findings

### **3. Data and tools**

For the development of this project, the following data are required:

- Basic geolocation data for Cyprus (Areas, Regions)
- Specific geolocation
- Acquire statistical data regarding customer preferences

### **4. Libraries Which are Used to Develop the Project:**

Pandas: For creating and manipulating Dataframes.

Folium: Python visualization library would be used to visualize the neighbourhoods cluster distribution of using interactive leaflet map.

Scikit Learn: For importing k-means clustering.

JSON: Library to handle JSON files.

XML: To separate data from presentation and XML stores data in plain text format.

Geocoder: To retrieve Location Data.

Beautiful Soup and Requests: To scrap and library to handle http requests.

Matplotlib: Python Plotting Module.

Google API: To acquire geolocation data