

# Segmenting the potential market for the XYZ bespoke software development company

Alaa Mahjoub

June 2020

## 1. Introduction

### 1.1 Background

The XYZ company is a software development company specialized in developing bespoke software management systems for hotels, coffee shops and restaurants. The company develops and maintains three software products:

- A- The Hotels Management Software (HMS)
- B- The Coffeeshops Management Software (CMS)
- C- The Restaurants Management Software (RMS)

The company Marketing Department includes three Sections: a hotel management software marketing section, a coffeeshop management software marketing section and a restaurant management software marketing section.

The company is planning to expand its business by identifying potential overseas customers in the relevant national capitals across the whole world.

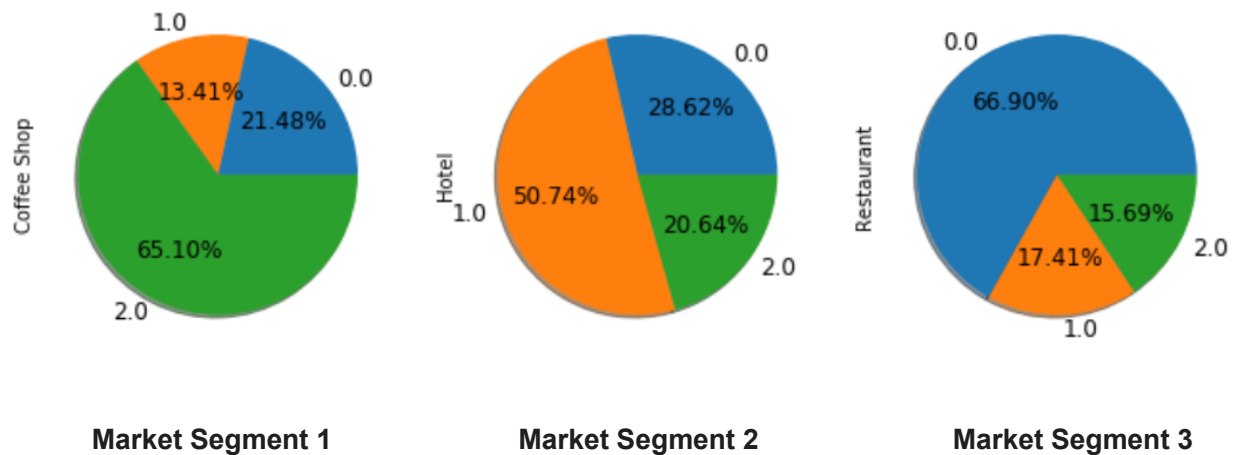
### 1.2 Problem

In order to expand its market, the company adopted a data driven approach and formulated a new market development strategy based on geo-demographic market segmentation. The data which will contribute to determining the market segments includes the national capitals, their geographical coordinates and the relative number of potential customers in each national capital.

This project is a data clustering project that aims to segment the national capitals into three marketing segments based on this data. It also aims to identify the potential customers in each market segment. According to this segmentation, each Marketing Section will lead the new market development efforts in one of the three market segments.

Figure 1 below depicts an example of the potential market segmentation.

**Figure 1 - An example of potential market segmentation**



### 1.3 Interest

Obviously, the XYZ company marketing department would be very interested to know which national capitals will be assigned to each marketing section. It would be interested (as well) in knowing the potential number of customers in each market segment (see Table -1 below for an example). Others who care about this problem include the XYZ company senior managers as well as the company's shareholders.

**Table 1- An example of potential market segmentation**

|                       | Restaurant | Hotel | Coffee Shop |
|-----------------------|------------|-------|-------------|
| <b>Cluster Labels</b> |            |       |             |
| <b>0.0</b>            | 388.0      | 366.0 | 165.0       |
| <b>1.0</b>            | 101.0      | 649.0 | 103.0       |
| <b>2.0</b>            | 91.0       | 264.0 | 500.0       |

## 2. Data acquisition and cleansing

### 2.1 Data sources

Table 2 below describes the data sets and their data sources:

**Table 2- the data sets and their data sources**

| No | Data set                                   | Description  | Data Source   |
|----|--|--|---|
| 1  | List of world-wide national capitals       | Data fields include City, Country and Notes. See Appendix I for an <u>example of this data set.</u>  | I scraped the following Wikipedia site to obtain this data<br><br><a href="https://en.wikipedia.org/wiki/List_of_national_capitals">https://en.wikipedia.org/wiki/List_of_national_capitals</a> |
| 2  | Geo-Location data of each national capital | Data fields include the longitude and latitude coordinates of each national capital. See Appendix II for an <u>example if this data set.</u> | I obtained this data using the Python geocoding web services API.   |
| 3  | Potential customers' data                  | Data fields include the venue name, category, longitude, latitude, See Appendix III for an <u>example if this data set.</u>                  | I obtained this data by exploring the national capitals venues using the Foursquare API   |
| 4  | The world map GIS data                     | Data of world map with the national capitals across the world. See Appendix IV for an <u>example if this data set.</u>                       | I obtained this data using the Folium API   |

## 2.2 Data Cleansing

Data of national capitals are scraped from the Wikipedia page through python libraries. There were some missing data records which I discovered during searching the location data of the national capitals. After investigation I discovered that the missing data were due to some comments that were included in the Wikipedia page and put between round parentheses. So, I removed the parentheses and all data within them using Pandas, and then I used this data to search the locations of the national capitals again. This time I got no missing data. However, I included some code to remove the missing (NaN) values such that it can be used in future in case of any update may take place on the Wikipedia page.

Also, I have notices that the column names in the Wikipedia page are not put in standard naming convention, and column names such as City/Town are used, and this may jeopardize the Python program code. So, I modified the column name to comply with the standard naming convention.

Then I combined the lat. and long coordinate columns structure to the data frame structure of the table received from the Wikipedia page to allow me to include the coordinates data.

I then obtained the national capitals' coordinates data using the geocoding web services. While doing that, I discovered that there are very few missing data and I treated this by displaying exemption messages in the data acquisition software module, and then I drop the rows with nan values in latitude or longitude fields (if any).

I then used Folium to create a World Map with all national capitals superimposed on top, and used this map to visually verify the correctness of acquired data on the map.

## 2.3 Feature Selection

After data cleansing, there were 16,702 samples and to know the total number of features (i.e. the number of venue categories of the national capitals), I calculated the number of unique categories curated from all the returned natural capital venues. They were 522 unique venue categories, however, in this market segmentation problem, we need only three features of these features. These are the features marked as 'Kept' in the Feature selection Table -3 below:

**Table 3. Feature selection during data cleaning**

| No | Feature   | Type of variable | Kept/Dropped | Reason   |
|----|---|------------------|--------------|--|
| 1  | Hotel Category Venue  | Categorical      | Kept         | We need it to build our market segmentation cluster          |
| 2  | Coffeeshop Category venue   | Categorical      | Kept         | We need it to build our market segmentation cluster          |
| 3  | Restaurant category venue   | Categorical      | Kept         | We need it to build our market segmentation cluster          |
| 4  | All other categorical variables such as , Auto Workshop, Supplement Shop, Women's Store, etc. | Categorical      | Dropped      | We do NOT need them to build our market segmentation cluster |

***The following two pages include:***

**Appendix I – Example of data set 1, the Wikipedia List of world-wide national capitals**

**Appendix II – Example of data set 2 - Geo-Location data of each national capital from the geocoding web services**

**Appendix III – Example of data set 3 - National capitals important venues from the Foursquare API.**

**Appendix IV – Example of data set 3 - The world map GIS data from Folium**

## Appendix I – Example of data set 1, the Wikipedia List of world-wide national capitals

| City/Town ↕   | Country/Territory ↕  | Notes ↕   |
|---|--|---|
| Abidjan (former capital; still has many government offices) |  <a href="#">Ivory Coast</a>          |   |
| Yamoussoukro (official)                                     |  |   |
| Abu Dhabi   |  <a href="#">United Arab Emirates</a> |   |
| Abuja   |  <a href="#">Nigeria</a>              | <a href="#">Lagos</a> was the capital from 1914 to 1991.  |
| Accra   |  <a href="#">Ghana</a>                |   |
| Adamstown   |  <a href="#">Pitcairn Islands</a>     | <a href="#">British Overseas Territory</a> .  |
| Addis Ababa   |  <a href="#">Ethiopia</a>             |   |
| Aden (de facto, temporary)                                  |  <a href="#">Yemen</a>                | <a href="#">Sana'a</a> has been occupied by <a href="#">Houthis</a> rebels since February 2015. Aden is Yemen's acting capital. See also: <a href="#">Yemeni Civil War (2015–present)</a> . |
| Sana'a (de jure)  |  |   |
| Algiers   |  <a href="#">Algeria</a>              |   |
| Alofi   |  <a href="#">Niue</a>                 | Self-governing in <a href="#">free association</a> with <a href="#">New Zealand</a> .   |
| Amman   |  <a href="#">Jordan</a>               |   |
| Amsterdam (official)  |  | The Dutch constitution refers to Amsterdam as the " <a href="#">capital</a> ".  |

## Appendix II – Example of data set 2

### Geo-Location data of each national capital from the geocoding web services

|    | City         | Country              | lat      | lng      |
|----|--------------|----------------------|----------|----------|
| 0  | Abidjan      | Ivory Coast          | 5.32036  | -4.01611 |
| 1  | Yamoussoukro | Ivory Coast          | 6.80911  | -5.27326 |
| 2  | Abu Dhabi    | United Arab Emirates | 24.4748  | 54.3706  |
| 3  | Abuja        | Nigeria              | 9.06433  | 7.4893   |
| 4  | Accra        | Ghana                | 52.4934  | 4.80368  |
| 5  | Adamstown    | Pitcairn Islands     | -25.0667 | -130.1   |
| 6  | Addis Ababa  | Ethiopia             | 9.01079  | 38.7613  |
| 7  | Aden         | Yemen                | 12.8333  | 44.9167  |
| 8  | Sana'a       | Yemen                | 15.3539  | 44.2059  |
| 9  | Algiers      | Algeria              | 36.7754  | 3.06019  |
| 10 | Alofi        | Niue                 | -19.0534 | -169.919 |

**Appendix III – Example of data set 3**  
**National capitals important venues from Foursquare API.**

|   | World Capital | Latitude | Longitude | Venue                        | Venue Latitude | Venue Longitude | Venue Category      |
|---|---------------|----------|-----------|------------------------------|----------------|-----------------|---------------------|
| 0 | Abidjan       | 5.320357 | -4.016107 | Sofitel Abidjan Hôtel Ivoire | 5.327097       | -4.004801       | Hotel               |
| 1 | Abidjan       | 5.320357 | -4.016107 | Norima                       | 5.363668       | -3.992067       | American Restaurant |
| 2 | Abidjan       | 5.320357 | -4.016107 | Cap Sud                      | 5.298763       | -3.987246       | Shopping Mall       |
| 3 | Abidjan       | 5.320357 | -4.016107 | Bao Café                     | 5.348778       | -3.996881       | Coffee Shop         |
| 4 | Abidjan       | 5.320357 | -4.016107 | Pink Club                    | 5.305360       | -3.988696       | Nightclub           |
| 5 | Abidjan       | 5.320357 | -4.016107 | Nice Cream                   | 5.291398       | -3.982492       | Ice Cream Shop      |
| 6 | Abidjan       | 5.320357 | -4.016107 | Lifestar                     | 5.324086       | -4.015354       | Nightclub           |
| 7 | Abidjan       | 5.320357 | -4.016107 | Des Gateaux & Du Pain        | 5.360270       | -3.989671       | Bakery              |
| 8 | Abidjan       | 5.320357 | -4.016107 | Di Sorrento                  | 5.288542       | -3.987629       | Italian Restaurant  |

**Appendix IV – Example of data set 3**  
**The world map GIS data from Folium**

