

Expanding into new Markets in Banking

Patrik Vida

November 23rd, 2019

Introduction: Business Problem

As retail banking builds on the physical presence on the high street, location is paramount. To find a new location can be a challenging task. Especially when expanding into a new market. In the first section I will describe the background and description of the problem.

Background

Retail banking focusing on acquiring consumers through branch presence and in recent times also digital precedence. Branch presence is still considered an important part of broadening customer reach. Besides physical and digital precedence, appropriate financial products are key to reach the target customers. To broaden the customer base new markets need to be entered. One such new market would be expanding into a new country. The focus of this report is to expanding into a new country and open a new bank branch in a new city in a foreign country. To achieve this an appropriate country, with a comparable financial system need to be found. After selecting the country, an appropriate location needs to be found in the financial center of the selected country. In the next section the problem is described in more detail.

Problem

The project aims to find an optimal location for a new bank branch in a new city in a foreign country. To reduce friction to the Spanish client, the financial system of the country of interest should be close to that of Spain. The region of interest is in Central and Eastern Europe.

Interest

I will try to compile a set of recommended locations for the new branch in the capital city or other financial center of the foreign country. The compiled list should include addresses, which fulfill additional two conditions, first proximity to the city center. As a second condition is, to have low number of bank branches in the vicinity of the suggested location of the new bank branch.

Structure

The structure of this report is the following. The report has five parts, the first is Introduction: Business Problem, with an overview of the background and description of the problem, with a short description of the main interested of the project. After that comes the second section with the title Data, where I describe the data that will be used to solve the problem and the source of the data. The third part is Methodology, this is the main component of the report where I discuss and describe the exploratory data analysis. This is followed by the Results and Discussion section. In the Results

and Discussion section where I discuss the results and observations and on the results. The final fifth section is the Conclusion, where I conclude the report.

Data

This section gives an overview of the data used to solve the problem and the source of the data. Accessing the data sources is described shortly and also how could the data contribute to solve the problem.

Data sources

Three different data sources were used during this project. The first data source is the financial data provided by the World Bank in the form of the DataBank. The second source is the Geocoding API provided by Google. To get more in-depth information about the neighborhoods, as the third data source on venue details Foursquare was used. A description of the three data sources follows.

Financial data

As the first step, I was comparing and grouping the financial systems of countries based on the data provided from the World Bank. Since there can be reasonable differences in the financial systems of neighboring countries, I have clustered the financial systems based on four characteristics (financial depth, access, efficiency, and stability)¹. To access the data on the servers of the DataBank I used the *wbdata* package. With this approach I could access the data for the indicators through python. In total, I used 8 indicators from the World Bank DataBank, two indicators for each of the four characteristic. One of the two indicators is covering the financial institutions and the other is covering the financial markets. For a better overview the indicators are summarized in the table 1 below.

Tab. 1: Financial indicators

	Financial Institutions	Financial Markets
DEPTH	Private sector credit to GDP (FS.AST.PRVT.GD.ZS)	Stocks traded to GDP Stocks traded, total value (% of GDP)(CM.MKT.TRAD.GD.ZS)
ACCESS	Account ownership at a financial institution or with a mobile-money-service provider, primary education or less (% of population ages 15+) (FX.OWN.TOTL.PL.ZS)	Percent of market capitalization outside of top 10 largest companies Market capitalization of listed domestic companies (% of GDP)(CM.MKT.LCAP.GD.ZS)
EFFICIENCY	Lending-deposits spread Interest rate spread (lending rate minus deposit rate, %) (FR.INR.LNDP)	Turnover ratio (turnover/capitalization) for stock market Stocks traded, turnover ratio of domestic shares (%) (CM.MKT.TRNR)
STABILITY	Asset quality ratios Bank liquid reserves to bank assets ratio (%) (FD.RES.LIQU.AS.ZS)	External debt stocks, total (DOD, current US\$) (DT.DOD.DECT.CD)

1 Cihak, Martin; Demirguc-Kunt, Asli; Feyen, Erik; Levine, Ross. 2012. *Benchmarking financial systems around the world (English)*. Policy Research working paper ; no. WPS 6175. Washington, DC: World Bank.
<http://documents.worldbank.org/curated/en/868131468326381955/Benchmarking-financial-systems-around-the-world>

In this project, I was using Google's Geocoding API in Reverse geocoding and also in the geocoding direction. Geocoding is the process of converting addresses (like a street address) into geographic coordinates (like latitude and longitude). Reverse geocoding is the process of converting geographic coordinates into a human-readable address. The Geocoding API was used to make a list of neighborhoods.

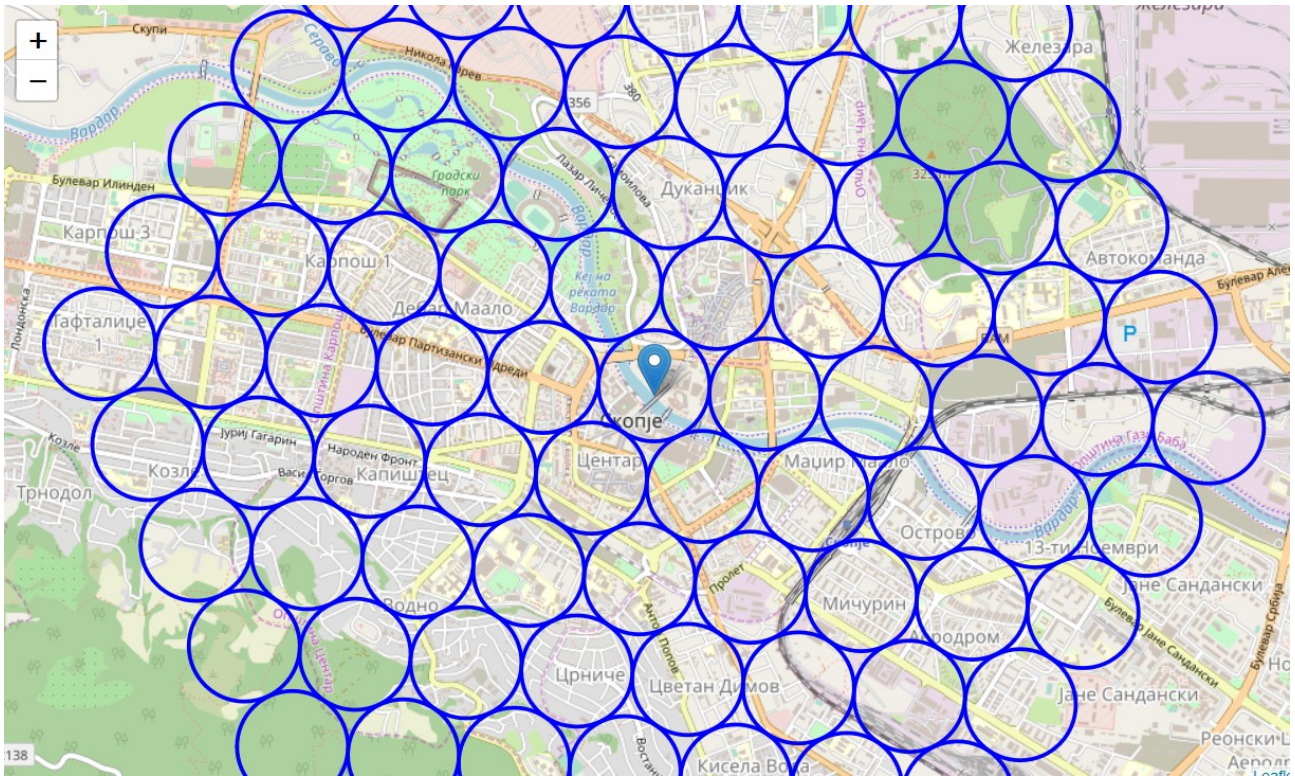


Fig. 1: Use of Geocoding API, to obtain a grid of the city of interest , here Skopje

From the latitude and longitude of the centers of the calculated grid, I could get the addresses with the help of the Geocoding API. Based of the generated data a *folium* map generated to visualize the grid, shown in Fig. 1.

The final list in a *pandas* data-frame of neighborhoods, with the addresses, latitude and longitude and distance from center is has shape as in table 2 below.

Tab. 2: Obtained geographical data

	Address	Latitude	Longitude	X	Y	Distance from center
0	Hiking Trail, 1000	41.975493	21.412811	1.031415e+06	4.666989e+06	2949.184880
1	Hiking Trail, 1000	41.975089	21.420008	1.032015e+06	4.666989e+06	2694.010292
2	Hiking Trail, 1000	41.974684	21.427204	1.032615e+06	4.666989e+06	2556.890974
3	Milan Mijalkovik 9, Skopje 1000	41.974280	21.434401	1.033215e+06	4.666989e+06	2556.890974
4	Georgi Dimitrov 21, Skopje 1000	41.973874	21.441597	1.033815e+06	4.666989e+06	2694.010292
5	Hristo Tatarchev 47g/62, Skopje 1000	41.973469	21.448793	1.034415e+06	4.666989e+06	2949.184880
6	Unnamed Road, 1000	41.980345	21.409681	1.031115e+06	4.667509e+06	2705.336527
7	Unnamed Road Skopje 1000, 1000	41.979941	21.416879	1.031715e+06	4.667509e+06	2349.222366

To obtain information about the venues, in this case of the banks, was done with the help of the Foursquare API. Based on the location data of the grid of the city shown in Fig 1., banks were explored in the proximity with the Foursquare API. The banks were explored in a radius of 500 meters, from the middle point of grid generated. With this approach a list of 173 banks was generated, based on the provided data from Foursquare.

As you can see in the figure above, all banks from the list are represented in the proximity from the center of Skopje. The location of the banks is represented with red dots on the figure Fig. 2.

As described in the problem section, I want to find the set of countries with a financial system closest that of Spain. From this set I try to select smaller set of countries from Central and Eastern Europe. From the Central and Eastern Europe set I selected the country with the lowest percentage of account ownership at a bank regarding the countries' population. With the assumption, that low account ownership represents an unmet need of the population and turn them to potential customers. From this prospective a lower percentage, points to a potentially relatively larger market. This is rough estimate, other factors could be incorporated. But I used as a guidance. From this approach North Macedonia turned out to be the target for opening new bank branch. Since the largest city, capital and financial center of the country is Skopje, I focused my attention to this city.

As described above I focused my efforts on detecting areas of Skopje that have low bank density and still be in area ~3 km around city center.

In the first step I have collected the required data: location of every bank within 3km from Skopje center (Macedonia Square).

Second step in the analysis will be calculation and exploration of 'bank density' across different areas of Skopje. Here I will use heat maps to identify a few promising areas close to center with low number of banks in general and focus our attention on those areas.

In third and final step. I will focus on most promising areas and within those create clusters of locations that meet some basic requirements by the stakeholders: I will take into consideration locations with no more than two banks in radius of 250 meters. I will present a map of all such locations, but also create clusters (using k-means clustering) of those locations to identify general addresses. The generated addresses, can serve as a starting point for final 'street level' exploration and search for optimal new bank branch location.

Exploratory Data Analysis

In this section I describe the two steps of the data analysis. First the analysis of financial data to select a country to be the target of expanding. Second to generate a list of addresses, that can serve as a starting point for final exploration and search for optimal new bank branch location.

Financial data

The financial data was obtained for the four characteristics (financial depth, access, efficiency, and stability) of the financial systems. As described above two indicators were used to cover the financial institutions and the financial markets' aspect of the four financial characteristics. This yield a need for at least 8 indicators. With the help of the *wbdata* package from the World Bank DataBank. The extensive data set of the 8 indicators (described in Tab. 1) for more than 250 countries and regions for the last 60 years was obtained. Based on the obtained data and with the help of *scipy.cluster.hierarchy* a dendrogram was created.

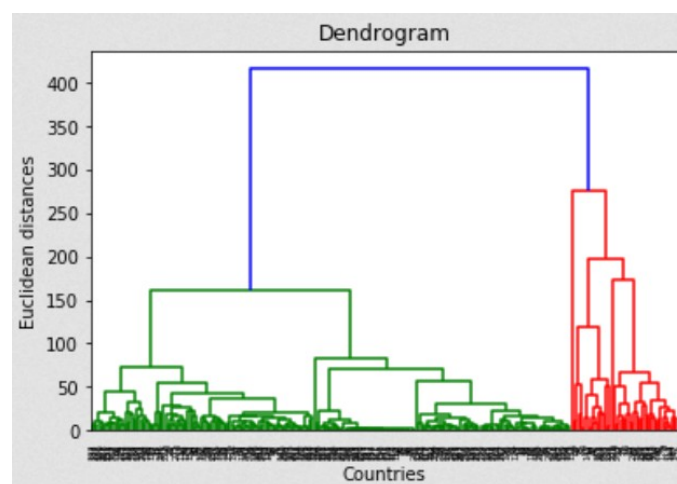


Fig. 3: Dendrogram of the financial characteristics of countries based on Word Bank data

Based of the dendrogram, the European Countries were selected, that ware in the same group as Spain. From this group the North Macedonia was selected, since has the 76.57% bank account

ownership. This number suggests the largest proportion of the population with no bank account, and a potential market for new bank customers.

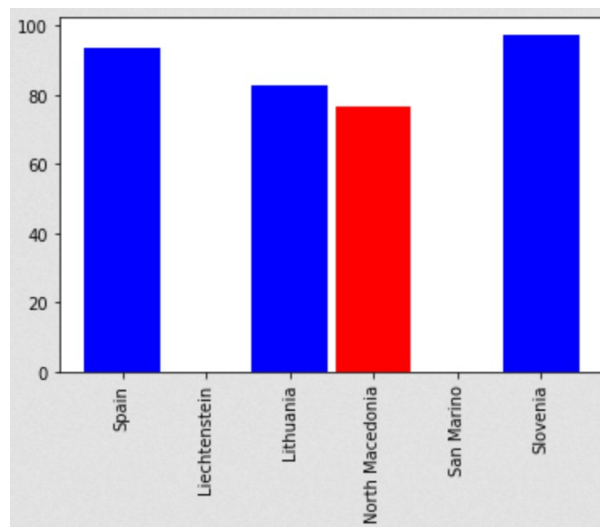


Fig. 4: Account ownership at a bank in % of population in the selected region

For the further analysis of possible locations for the new bank branch Skopje was chosen. Skopje is the financial and economic center of North Macedonia.

Venue Data

Based on the data from Foursquare the list of 173 banks is visualized as a heat map in figure 5. From that the highest density of bank branches can be deduced. The south and the south-west part of the city center is hot. This indicates a high bank density there. The neighborhoods just north from the center show a low bank density. The Čair Municipality with large low bank density north from city center is an example. However, this borough is less interesting to stakeholders as it's mostly residential, with less economic activity. Possible locations in proximity of the city center are the neighborhoods Aerodrom and Kisela Voda south and south-east from the center.



Fig. 5: Heat map of bank branch location in Skopje

Based on the heat map I moved the focus my of the analysis south-east of the center as show in the figure 6 below.

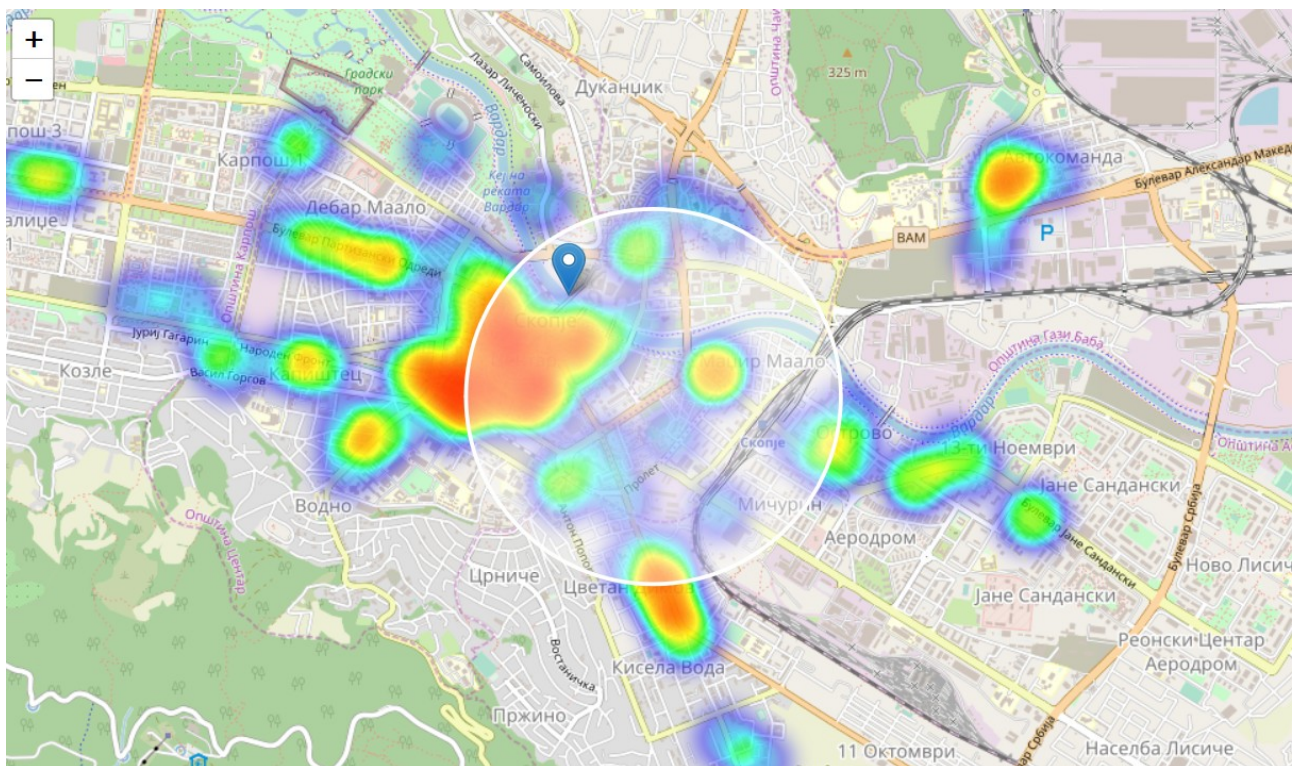


Fig. 6: Define new, narrower region of interest, with a 1km radius

With in the neighborhood in the 1 km radius from the figure above all possible locations were listed. That do not have more the two bank in a 250 m radius. This list of potential location is visualized below in figure 7.



Fig. 7: Visualization of interested locations with no more than two banks in radius of 250 meters

To have a better picture of the candidate locations in relation to the main central point of Skopje Macedonia square, a heat map was created. In the heat map below the potential locations are represented (Fig. 8).



Fig. 8: Heat map of locations fairly close to Macedonia square, and we know that each of those locations has no more than two banks in radius of 250m

To further reduce the number of potential addresses, the potential locations were clustered. The used algorithm is *KMeans* from the package *sklearn.cluster*. As clusters are located at the centers of

zones containing good locations, the result of the clustering effort is represented in the figure below (Fig. 9).



Fig. 9: Clusters of good locations in the proximity of the center with low bank density

To get the approximate addresses of center locations point of the clusters form Fig. 9 the Google Maps API was utilized. The list of recommend addresses has been summarized in the table below (Tab. 3)

Tab. 3: Addresses of centers of areas recommended for further analysis

Address	Distance
Транспортен Центар, Skopje 1000	1.2 km from Macedonia square
Sv. Kliment Ohridski 1, Skopje 1000	0.8 km from Macedonia square
Dimche Mirchev 3, Skopje 1000	1.1 km from Macedonia square
Vardarska 39, Skopje 1000	1.3 km from Macedonia square
Parking Garage 'Smilevski Kongres" , Iljo Vojvoda, Skopje 1000	0.2 km from Macedonia square
3rd Macedonian Brigade Blvd 20, Skopje 1000	1.5 km from Macedonia square
ул, Кеј 13-ти Ноември, бр 2, 1000	1.1 km from Macedonia square
1000, Boulevard Goce Delchev, Skopje 1000	1.1 km from Macedonia square
Frederic Chopin 28, Skopje 1000	1.4 km from Macedonia square
Veniamin Machukovski 17, Skopje 1000	1.6 km from Macedonia square

This concludes the analysis. With the help of the algorithms a list of 10 addresses were created. These addresses representing centers of zones containing locations with low number of banks

nearby, all zones being fairly close to city center (all less than 2km from Macedonia square). Most of the zones are located in Center, Aerodrom and Kisela Voda neighborhoods, which were identified as interesting due to being vivid business areas, fairly close to city center and well-connected by public transport.

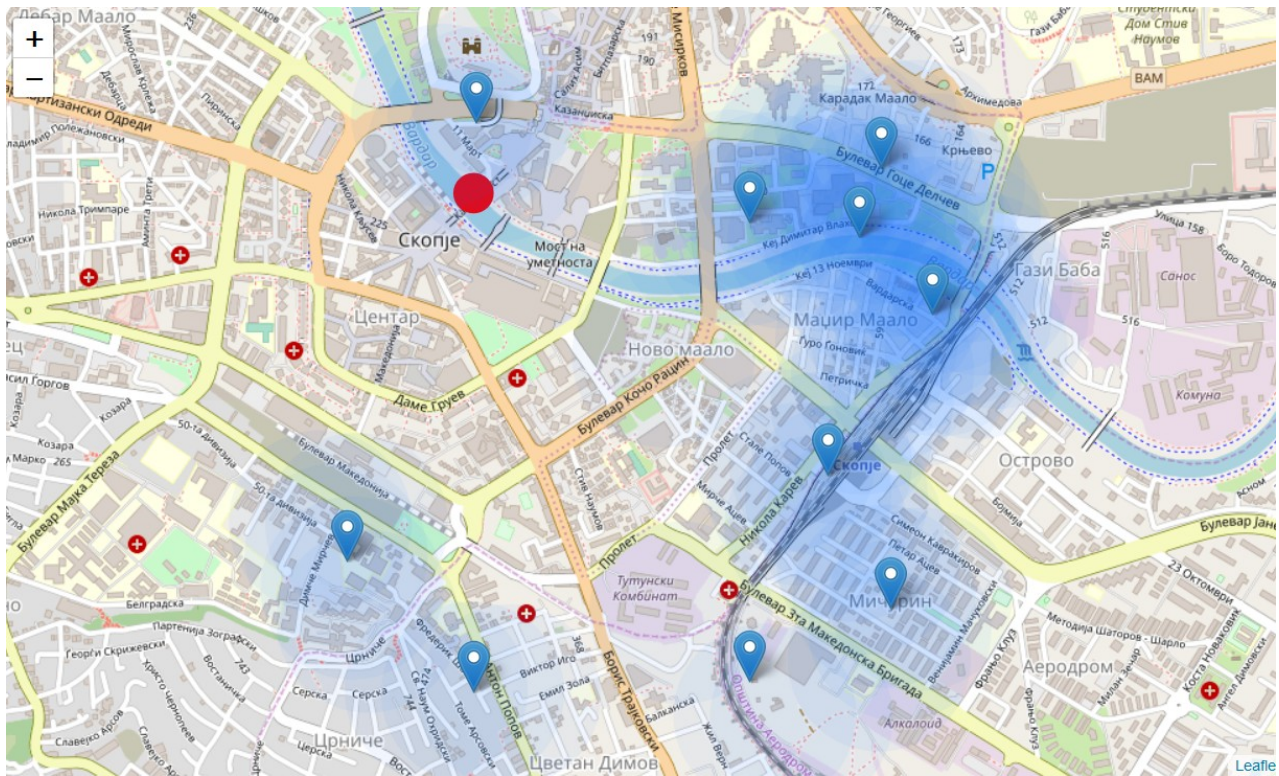


Fig. 10: Addresses of centers of areas recommended for further analysis

In this section, which represents the main component of the report. I discussed and described the conducted exploratory data analysis. In the next section I discuss the results.

Results and Discussion

After the new market and the target countries could be shortlisted. North Macedonia proved to an appropriate country, with a comparable financial system to Spain based on hierarchical clustering. Since Skopje is the largest city of North Macedonia, the focus of opening a new bank branch was set to be in the proximity of center of this city.

This analysis shows that although there are numerous banks in Skopje (173 in our initial area of interest which was 6×6 km around Macedonia square), there are pockets of low bank density fairly close to city center. The highest concentration of banks was detected south and south-west from Macedonia square, so we focused our attention to areas south-east and east, corresponding to neighborhoods Aerodrom and Kisela Voda and south-east corner of the Centar neighborhood. The neighborhoods north from the city center like Čair Municipality, were not considered, since that are more residential areas not suitable for a bank branch.

After directing our attention to this narrower area of interest (covering approx. 2×2 km south-east from Macedonia square) we first created a dense grid of location candidates (spaced 100 m apart); those locations were then filtered so that those with more than two banks in radius of 250 m.

Those location candidates were then clustered to create zones of interest which contain the greatest number of location candidates. Addresses of centers of those zones were also generated using reverse geocoding to be used as starting points for more detailed local analysis based on other factors.

Result of all this is 10 zones containing the largest number of potential new bank locations based on number of and distance to existing venues. This, of course, does not imply that those zones are actually optimal locations for a new bank branches in a foreign country. Purpose of this analysis was only to provide info on areas close to Skopje center, with low density of existing bank branches of other banks. Recommended zones should therefore be considered only as a starting point for more detailed analysis.

Conclusion

Purpose of this project was to identify a country in Central and Eastern Europe with low number bank accounts per citizen. Since the client is a Spanish consumer bank, an additional condition was that the financial system of the country of interest should be close to that of Spain. By hierarchical clustering of the financial characteristics of countries based on World Bank data, North Macedonia was set as an appropriate candidate.

By calculating bank branch density distribution from Foursquare data I have first identified general neighborhoods that justify further analysis (neighborhoods Centar, Aerodrom and Kisela Voda), and then generated extensive collection of locations which satisfy some basic requirements regarding existing nearby banks. Clustering of those locations was then performed in order to create major zones of interest (containing the greatest number of potential locations) and addresses of those zone centers were created to be used as starting points for final exploration by the client.

Final decision on optimal bank branch location will be made by the Spanish client bank based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location, proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood.