Nikolina Doksanovic
20 June 2020

Zagreb vs. Split

1. Introduction

1.1. Background

Croatia is a country in Southeast Europe on the coast of the Adriatic sea. It has an area of 56,594 square kilometers (21,851 square miles) and a population of 4.07 million.

- Zagreb is the capital and the largest city of Croatia. It is located in the northwest of the country, along the Sava river, at the southern slopes of the Medvednica mountain. Zagreb lies at an elevation of approximately 122 m (400 ft) above sea level. The estimated population of the city in 2018 was 820,678. The population of the Zagreb urban agglomeration is 1,086,528, approximately a quarter of the total population of Croatia.
- Split is the second-largest city of Croatia and the largest city of the region of Dalmatia, with about 250,000 people living in its urban area. It lies on the eastern shore of the Adriatic Sea and is spread over a central peninsula and its surroundings. An intraregional transport hub and popular tourist destination, the city is linked to the Adriatic islands and the Apennine peninsula.

1.2. Problem

In this report we are going to analyze and compare Zagreb and Split in order to establish which of the cities might be more suitable for a new Café opening.

1.3. Interest

This research might be of a great importance to potential investors looking to start a new business or maybe expand their existing one by opening an additional branch etc.

2. Data acquisition and analysis

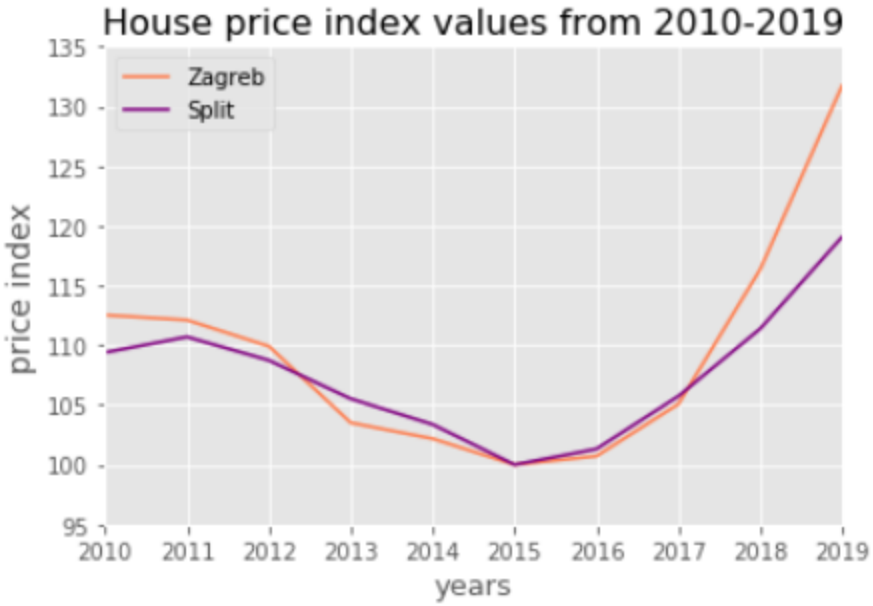
2.1. Data sources

For the purpose of this research we have obtained a file containing housing price indexes from Croatian Bureau of Statistics where, due to lack of explicit values for the city of Split, we used data for that region, therefore, this report is not accurate and shouldn't be used as a parameter for any business decision making. Its sole purpose is to showcase the knowledge and skills obtained throughout the course "IBM Data Science" in practice. Furthermore, we have retrieved 2019 tourism statistics from Croatian National Tourist Board which display visitor capacity throughout the year for Zagreb and Split

respectively and utilized Google Maps to form a database containing all bigger cities distance from the seashore in order to determine whether the vicinity of the seashore is correlated to the numbers of visitors. Once we ran our parameters and determined which city would be more suitable for starting a business, we have made calls to Foursquare API to determine which areas of the city has the highest number of cafés and restaurants and their ratings.

2.2. Data cleaning and analysis

From Croatian Bureau of Statistics we have obtained a 'CVS' file containing house pricing indexes from about 2005 until 2019 for all bigger regions of Croatia including the city of Zagreb. We've had a lot of missing values so we cleaned some rows and columns in order to get cleaner and more straightforward data frame. Since we are comparing Split to Zagreb, and there were no explicit values for Split, but instead the whole region where Split is the biggest city, we've used this data to represent approximate values for the city of Split. Years from 2005 until 2009 have been missing data, so we have decided to use data from 2010 until 2019 in our comparison to represent the trend of housing prices.



In the line graph above we can see the housing prices had the lowest index in 2015 and both cities had same values, however the lateral continuous growth shows higher index for Zagreb, which is not very surprising considering it is the capital of Croatia.

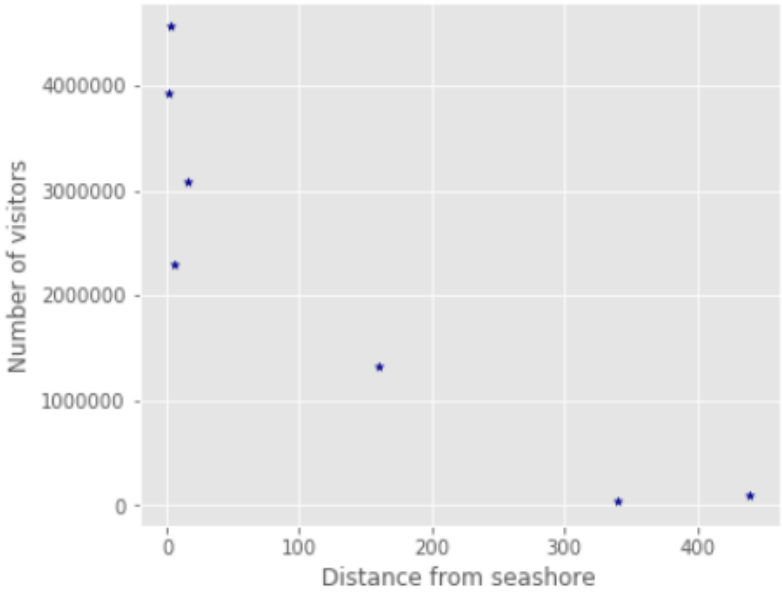
Our next data source was yet another 'CSV' file obtained from Croatian National Tourist Board which contains numbers of visitors throughout the year. This data reflects true values for all Croatian counties and city of Zagreb, so we've assigned to Split the value of its respected county. The dataset contains number of local and foreign visitors, as well as total number and a percentage, but we have used only total number of visitors (local + foreign) for our comparison.

Nikolina Doksanovic
20 June 2020

city	total visitors
Split	3,923,328
Zagreb	1,323,305

After removing all unnecessary data from the file, the result is showing that city of Split had more visitors in 2019 than Zagreb. Since we know Split is located on the coast of Croatia, and Zagreb is in the central part of the country, not in approximate vicinity of the shoreline, we have decided to obtain a data on distance from the shoreline for all the big cities in Croatia and compare them to numbers of visitors to see if there is any correlation between the two. We have used Google Maps to calculate average distances of the cities from the coast, added a column which contained numbers of visitors and came out with the following data frame:

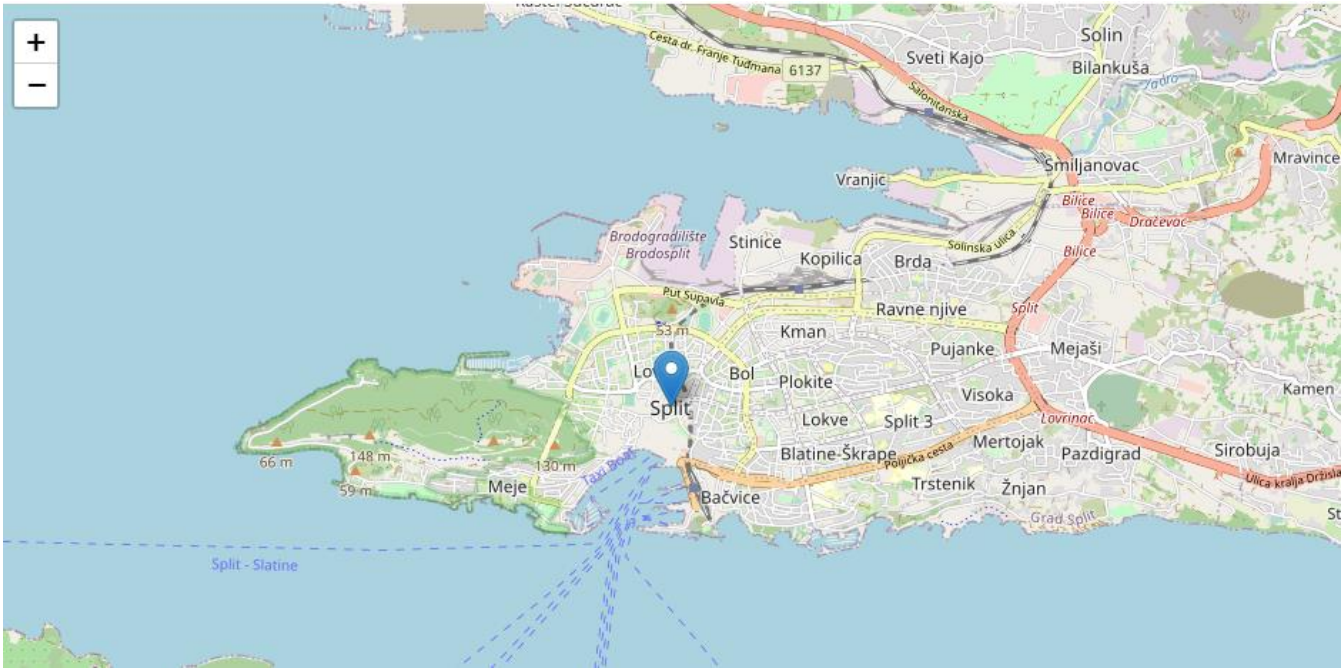
City	Distance	Visitors
Zagreb	160km	1 323 305
Split	1.4km	3 923 328
Pula	2.6km	4 561 303
Rijeka	16km	3 087 281
Dubrovnik	5.3km	2 299 853
Osijek	440km	104 159
Slavonski Brod	340km	35 386



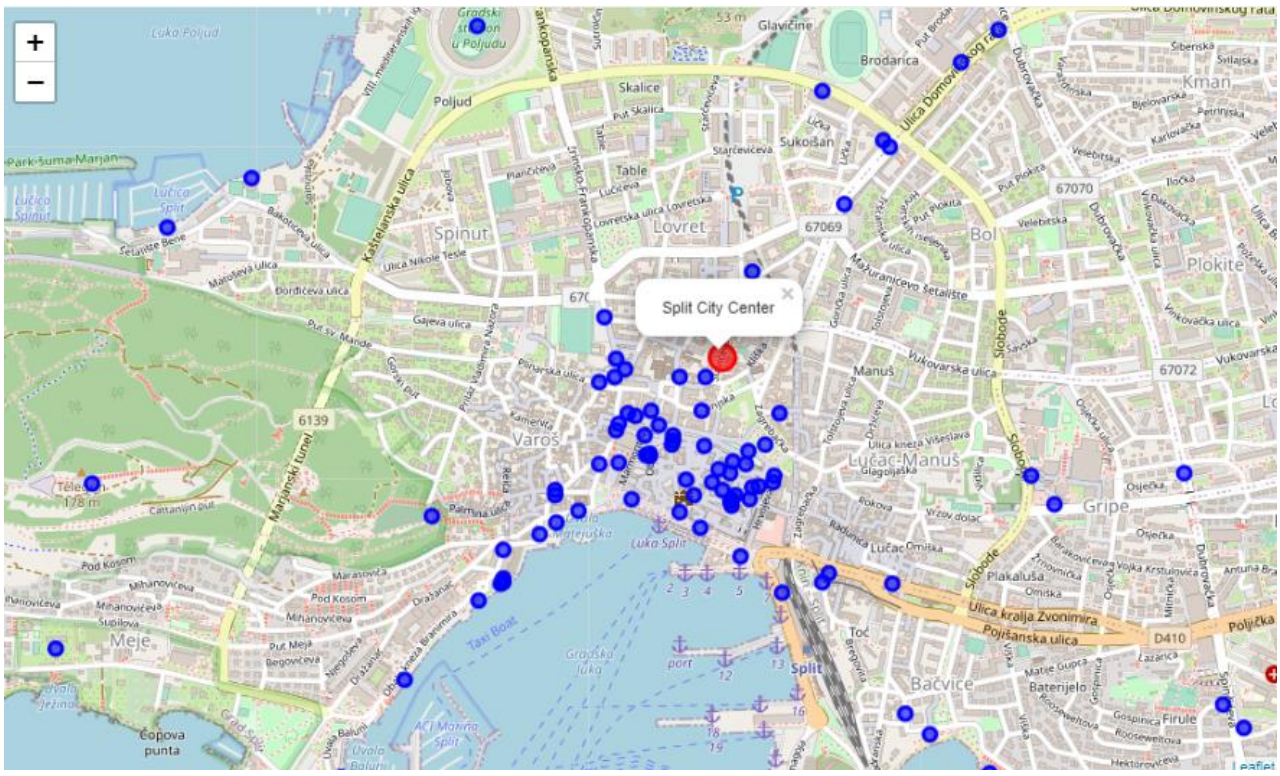
The visualization graph proves there is a correlation between the numbers of visitors and the vicinity of the coastline. Since the cities that are closer to the coastline have higher numbers in visitors, we can see the trend amongst the tourist and could predict which city has more potential for business growth, so at this stage we will disregard Zagreb as an option, since it has higher house price index, is far away from the sea and has significantly lower record of visitors.

Nikolina Doksanovic
20 June 2020

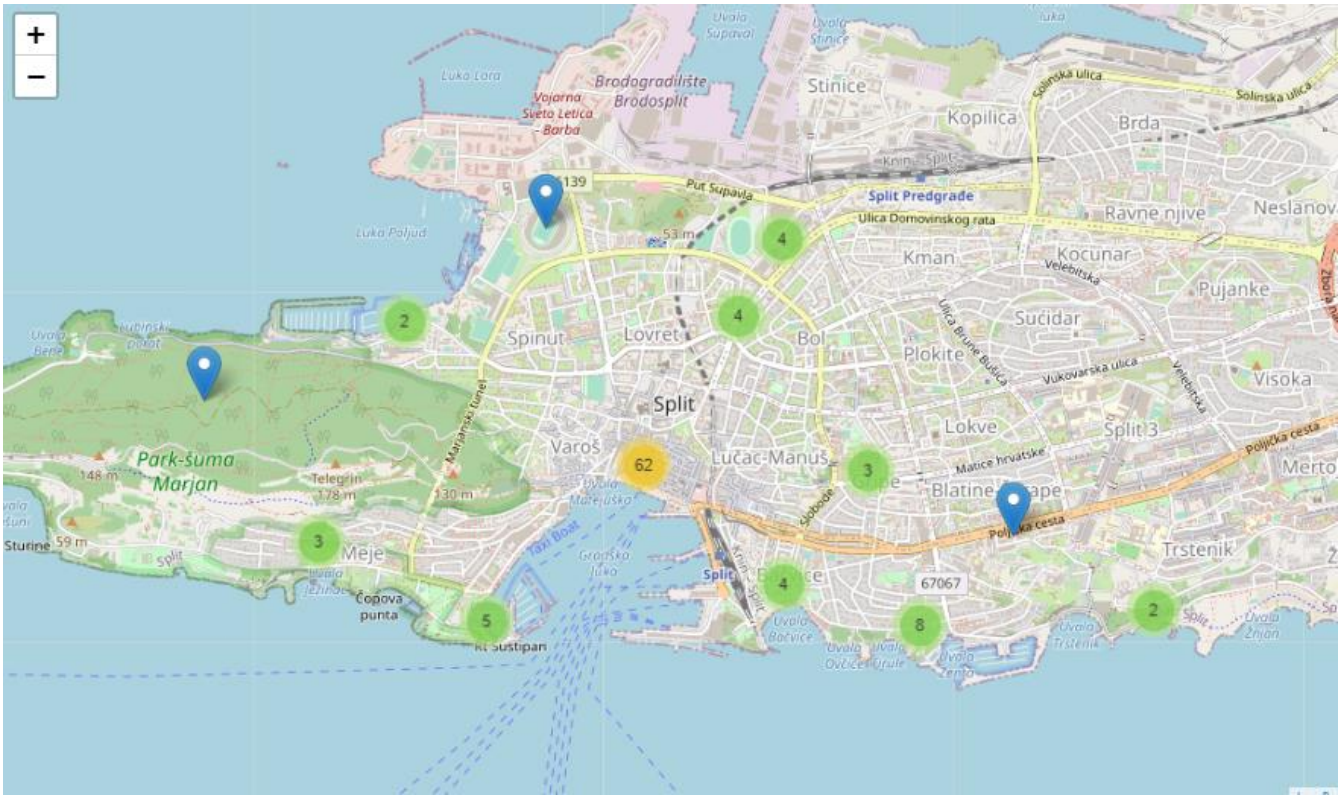
Moving our focus to Split only, we wanted to explore the most popular venues and determine which neighborhood/s might be most suitable ones for our business. We have used geolocator to determine latitude and longitude of the city and then marked it down on the map we've generated with folium library.



We have used Foursquare API the to find out the most popular venues within 5 Km distance from our starting point and generated the following map using folium again. Our coordinates are marked with a red circle, while venues are shown in blue:



On the second map we can clearly see the density of the most popular venues throughout the city. We went further and grouped the places into clusters using Leaflet plugin within the Folium library and got a total of 62 most popular venues within one area of the city:



3. Results

In our research we have used several methods to compare the cities in order to determine which one would be more suitable for starting a new business within hospitality industry. All parameters pointed towards Split as a better choice of location, so we ruled Zagreb out. Running a deeper research into popular venues around the city of Split we have narrowed down the locations within the city to the one specific cluster which had the highest numbers of the most popular venues.

4. Conclusion

In this study, we have compared the house price indexes throughout the span of 9 years to see the trend among the house prices, later we used a regression model to determine whether the distance from the coastline has an impact on the numbers of visitors and to help us predict where will our business most likely have a potential to grow. We have then retrieved the data on the venue popularity and their locations within the city of Split and by simple cluster markers we've determined which area of the city has the biggest potential.