

Predicting The Type & Location Of a New Venue

PRESENTATION AGENDA

1

Introduction/Business Problem

2

Data Acquisition

3

Data Analysis

4

Conclusions

1. Introduction/Business Problem



1.1 Background

Decision what type of place to open, and where it should be located, is crucial when preparing business plan for a new project. It is hard to keep track and understand which places have the biggest financial potential. Each year new or different businesses might be more popular than the others. Without proper analysis of the trends over the years, it is almost impossible to be sure that some particular type of venue is potentially more profitable than the others. It is also extremely important to know, where the new venue, which we would like to invest in, should be located. Because it is impossible to have an access to the financial records of each venue, it could be assumed that the number of visitors at some particular place (popularity) corresponds to the income. Also by utilisation of location data, the optimal area in reference to city centre could be chosen.

1.2 Problem

Data that might contribute to determine what business should be opened, in the particular city, based on number of citizens, might include number of visitors and its location in reference to the city centre and and other main areas such as train stations etc.. This project aims to predict what type of venue investor should invest in.

1.3 Interest

Everyone who thinks about investing into new venue, would be highly interested to see, what types of places are the most popular/profitable, and what is their ideal location.

2. Data Acquisition



2.1 Data sources

In different cities different venue types might be popular. This is why, in order to predict what is the best investment in the city of our interest, cities with similar size/population will be chosen for statistical analysis. Information about their population, is extracted from [simplemaps.com](#) (worldcities.csv). Because this file does not include data about size of the cities, it will be assumed that size correlates with population. Next the most popular venues will be selected based on their popularity. Such information can be extracted from [foursquare.com](#). Additionally the optimal location will also be chosen based on the coordinates. This test will be run for several years in order to see what is the trend of the most popular venues.

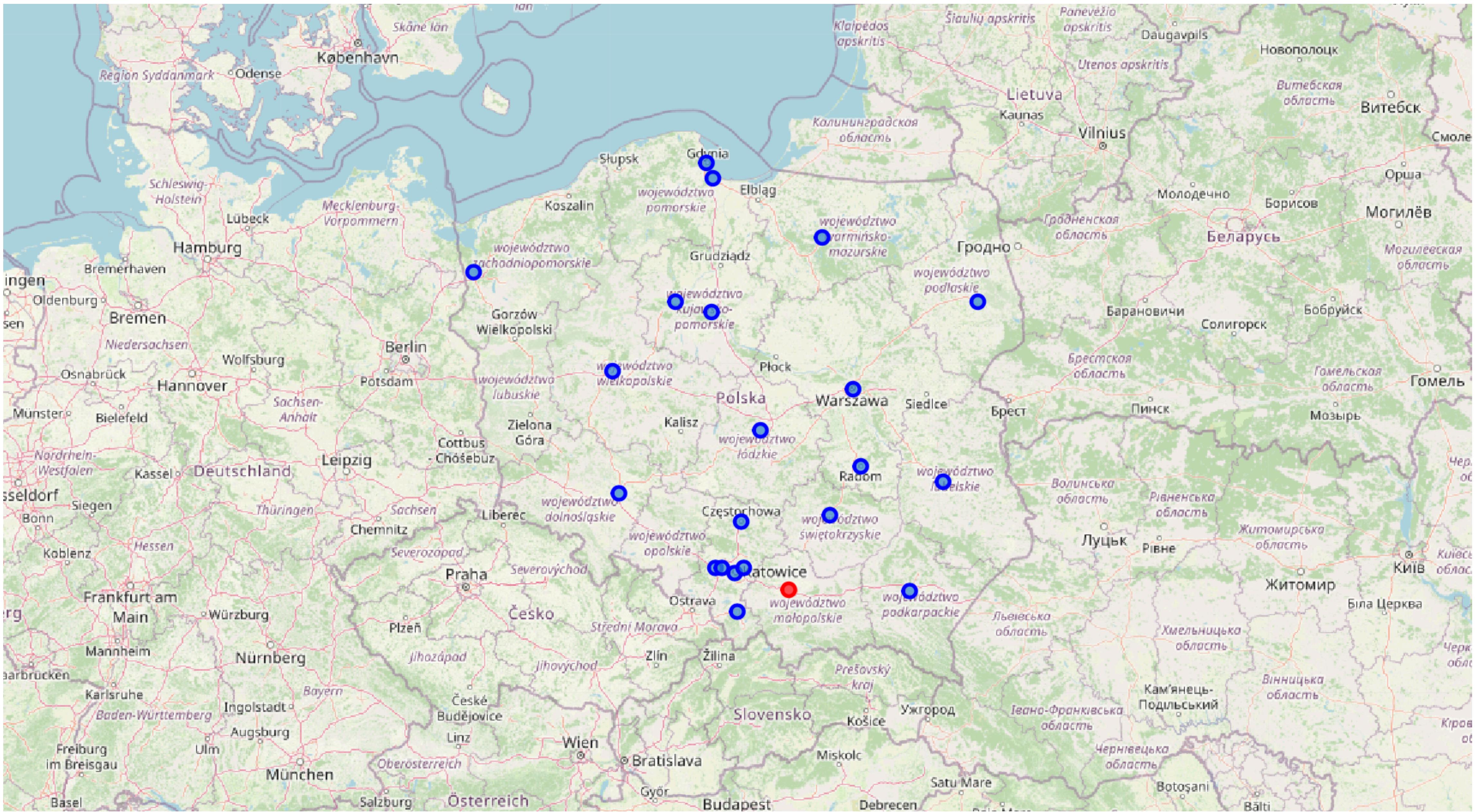
2.2 Data examples

Selecting city we would like to invest in

```
city_name = "Krakow"
```

Finding cities with similar population. For our test we will find 20 cities from the same geographical region/country, in this case Poland, with the closest number of citizens to “city_name” (Krakow)

Depicting all cities on the map (city of interest (Krakow) marked on red):



3. Data Analysis



DATA ANALYSIS

3.1 Data Manipulation

Main issue with data extracted from foursquare.com (basic dev account) is that they do not consist of number of "check ins" into particular venue (I can't see such information in json file). It significantly limits whole analysis. Because we do not have this data, it will be assumed, that most viable option from the business perspective, will correlate with the venue which statistically occurs the most in analysed cities.

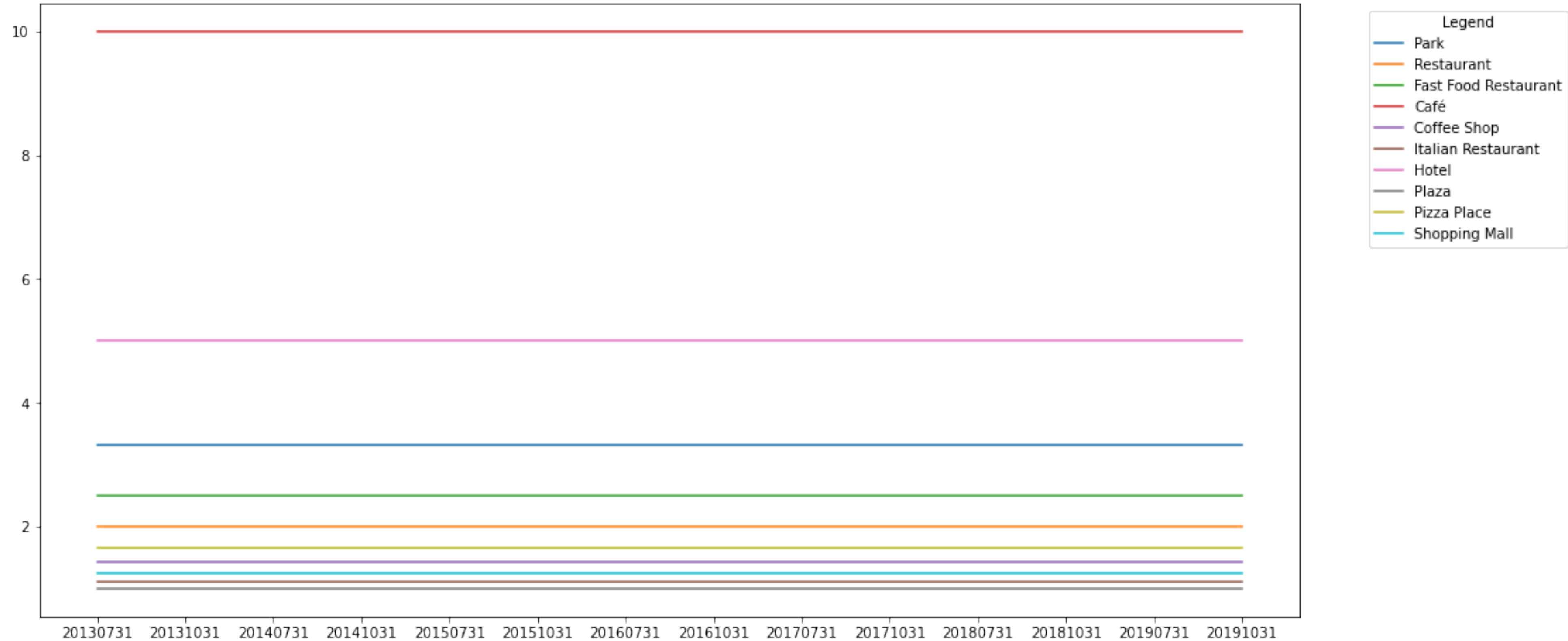
Checking what are statistically the most popular venues, in all analysed cities, for each 'Date'. This way we can see, how trend changes over time

Below table shows to 10 venues for each date

Date	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
20130731	Café	Hotel	Park	Fast Food Restaurant	Restaurant	Pizza Place	Coffee Shop	Shopping Mall	Italian Restaurant	Plaza
20131031	Café	Hotel	Park	Fast Food Restaurant	Restaurant	Pizza Place	Coffee Shop	Shopping Mall	Italian Restaurant	Plaza
20140731	Café	Hotel	Park	Fast Food Restaurant	Restaurant	Pizza Place	Coffee Shop	Shopping Mall	Italian Restaurant	Plaza
20141031	Café	Hotel	Park	Fast Food Restaurant	Restaurant	Pizza Place	Coffee Shop	Shopping Mall	Italian Restaurant	Plaza
20150731	Café	Hotel	Park	Fast Food Restaurant	Restaurant	Pizza Place	Coffee Shop	Shopping Mall	Italian Restaurant	Plaza
20151031	Café	Hotel	Park	Fast Food Restaurant	Restaurant	Pizza Place	Coffee Shop	Shopping Mall	Italian Restaurant	Plaza
20160731	Café	Hotel	Park	Fast Food Restaurant	Restaurant	Pizza Place	Coffee Shop	Shopping Mall	Italian Restaurant	Plaza
20161031	Café	Hotel	Park	Fast Food Restaurant	Restaurant	Pizza Place	Coffee Shop	Shopping Mall	Italian Restaurant	Plaza
20170731	Café	Hotel	Park	Fast Food Restaurant	Restaurant	Pizza Place	Coffee Shop	Shopping Mall	Italian Restaurant	Plaza
20171031	Café	Hotel	Park	Fast Food Restaurant	Restaurant	Pizza Place	Coffee Shop	Shopping Mall	Italian Restaurant	Plaza
20180731	Café	Hotel	Park	Fast Food Restaurant	Restaurant	Pizza Place	Coffee Shop	Shopping Mall	Italian Restaurant	Plaza
20181031	Café	Hotel	Park	Fast Food Restaurant	Restaurant	Pizza Place	Coffee Shop	Shopping Mall	Italian Restaurant	Plaza
20190731	Café	Hotel	Park	Fast Food Restaurant	Restaurant	Pizza Place	Coffee Shop	Shopping Mall	Italian Restaurant	Plaza
20191031	Café	Hotel	Park	Fast Food Restaurant	Restaurant	Pizza Place	Coffee Shop	Shopping Mall	Italian Restaurant	Plaza

From table depicted on the previous slide, we can see that for each date there are exactly the same top 10 venues. That can indicate that foursquare.com database does not consist of data for each time period but just the latest one.

To plot data easier, numbers from 1 to 10 were assigned to the venues from above table, based on its popularity:



From above graph it can be noticed that the most popular venue is of type "Cafe"

Based on above data it looks like there is no change in "popularity" over time among Venue Types. Looks like "foursquare" database does not consist of the data from various periods of time for analysed region.

3.2 Finding the best location for the "most popular" Venue

Now when we know what is the most popular Venue let's find its average distance from the city centre.
Table below contains few examples of the city centre and cafe coordinates:

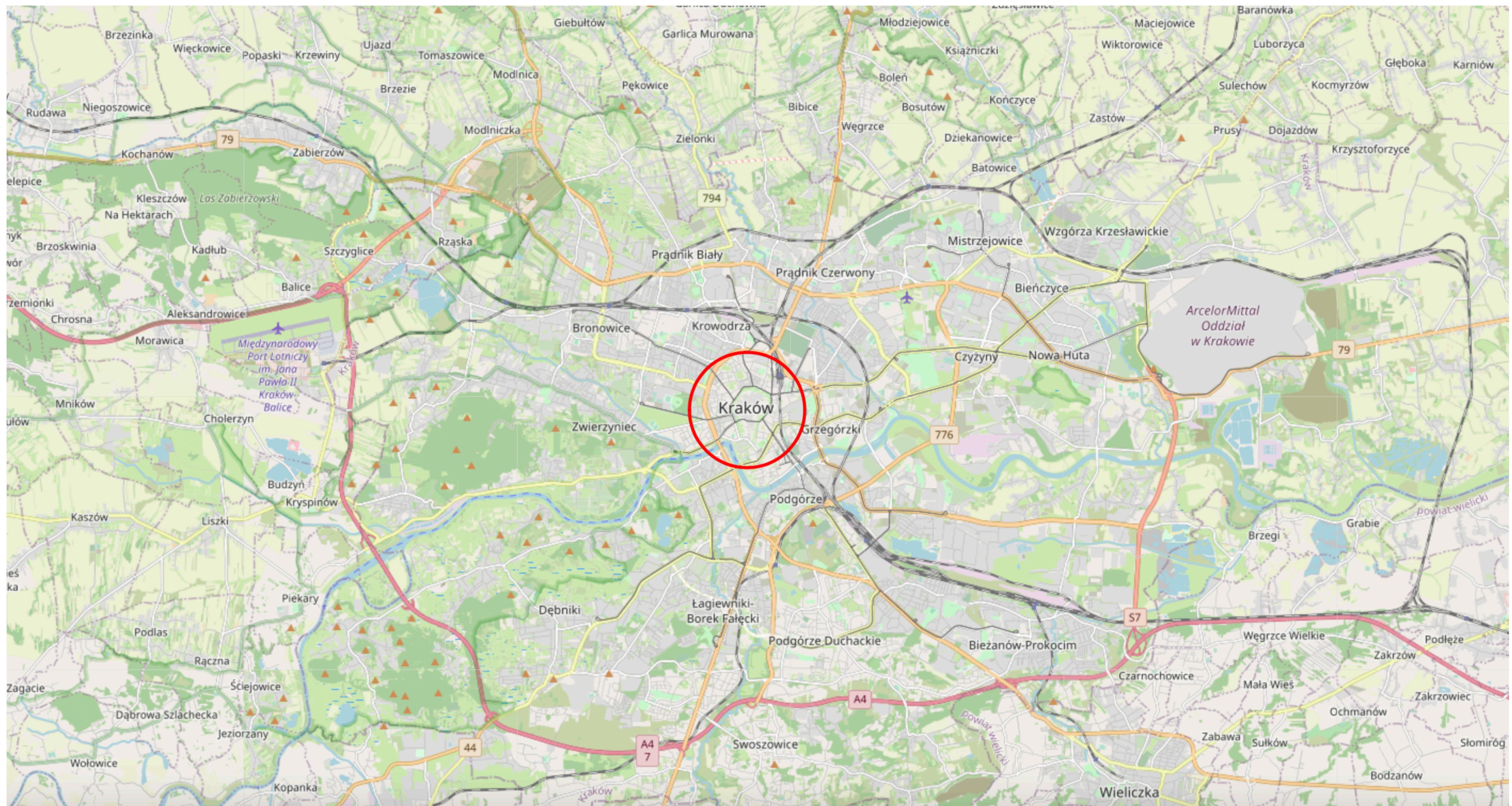
Venue Category	City Latitude	City Longitude	Venue Latitude	Venue Longitude
Café	52.2167	21.0333	52.225629	21.019006
Café	52.2167	21.0333	52.236939	21.024575
Café	52.2167	21.0333	52.233204	21.010875
Café	52.2167	21.0333	52.239527	21.022618
Café	52.2167	21.0333	52.222851	20.988527

In order to calculate the distance between city centre and "Venue Category" (Cafe) the "Haversine" formula will be used. Next based on all distances median and mean values can be calculated:

Median of the distance: 1434.04

Mean of the distance: 1618.35

Printing the area of interest over the city of interest (Krakow)



4. Conclusions



Based on the above analysis, it can be concluded, that the venue which we should invest in, is of type "Cafe". Its ideal location should be close to the city center.

Please note that due to limited number od data included in foursquare database for the analysed cities, it wasn't viable to check the trend in popularity over time, of different venues. Thus no regression was used in order to predict the best investment for the future!!!.