**Introduction**

I’m an enthusiastic European football fan and also love Chinese cuisine. One of my favorite activities is to enjoy a Chinese meal after watching some live football in a stadium. I’d like to explore that among 30 top European football cities, which ones would be more interesting to invest in opening a Chinese restaurant in close proximity to the local stadium. Also, if it’s possible to break the 30 cities into categories for investment consideration upon different attributes of the cities.

Potential target audience would be those who’re interested in investing in the project.

**Data Description**

Foursquare data will be used to explore the existing Chinese restaurants that are the closest to the local stadiums among the 30 cities.

Population of the 30 cities will be used.

<https://worldpopulationreview.com/continents/cities-in-europe/>

Cost of living, safety index, salary, restaurant index data of the 30 cities will also be used.

<https://www.numbeo.com/cost-of-living/region_rankings_current.jsp?region=150>

<https://www.numbeo.com/cost-of-living/region_prices_by_city?itemId=105&region=150>

<https://www.numbeo.com/property-investment/region_rankings_current.jsp?region=150>

<https://www.numbeo.com/crime/region_rankings_current.jsp?region=150>

An example of use of the data would be categorizing the 30 cities into different classes using K-mean clustering based on a list of attributes such as population, local purchasing power, distance from the stadium to the existing closest restaurant and reciprocal of salary where we would like high values in each attribute. It’s unlikely there is a category or a city that will show highest value in every attribute. Therefore, the classification will help assess investment decisions upon considerations of different attributes.

**Methodology**

Foursquare data will be used to search for the existing nearest Chinese restaurant to the local stadiums within 15 kilometers. The distance, along with latitude and longitude of the restaurants will be used to create a data frame df\_Foursquare. The shortest distance from the stadium to the existing Chinese restaurants from each city will be used because this is an indicator of competition where the higher the distance the more favorable to an investment decision.

Population data are used to create a data frame df\_population where only data from the 30 cities of interest are filtered. City names are replaced with those the same as in df\_Foursquare so the two data frames can be merged to form df\_merge.

All the data from Numbeo are filtered to the 30 cities of interest and merged into one data frame df\_numbeo. City names are replaced with those the same as in df\_merge so the two data frames can be merged to form df\_final.

Since the K-mean clustering method is used to classify the 30 cities some attributes are transformed so they will be clustered in proximity to desire. For example, an investment decision favors high population and low salary therefore, attributes to be clustered would be population and the reciprocal of salary for the city of interest.

The following attributes from df\_final are taken reciprocal: salary, rent index, cost of living index, groceries index, restaurant price index, crime index.

The following are the selected attributes to be analyzed: population, safety index, distance to the stadium of the closest Chinese restaurant, local purchasing power index, gross rental yield city center, gross rental yield outside of center, 1/net salary, 1/rent index, 1/cost of living index, 1/groceries index, 1/restaurant price index, 1/crime index.

The selected attributes are normalized using preprocessing package from sklearn so every attribute weighs the same. 30 cities are classified into 5 clusters using K-mean clustering method because the analysis aims to group data into distinct and well separated from each other in a linear fashion. Once the clustering is complete, each class will be analyzed using scatter plots and bar plots over average values to determine the class and city that is most interesting to an investment decision.

**Results**

K-mean clustering classifies the 30 cities into 5 groups.

Class 0: Milan, Munich, Hamburg, Berlin, Paris, Florence, London, Madrid, Amsterdam, Vienna

Class 1: Saint Petersburg, Kyiv, Moscow,

Class 2: Turin, Rome, Naples, Leipzig, Toulouse, Liverpool, Manchester, Barcelona, Brussels, Athens

Class 3: Seville, Porto, Lisbon, Prague, Warsaw, Zagreb

Class 4: Zurich

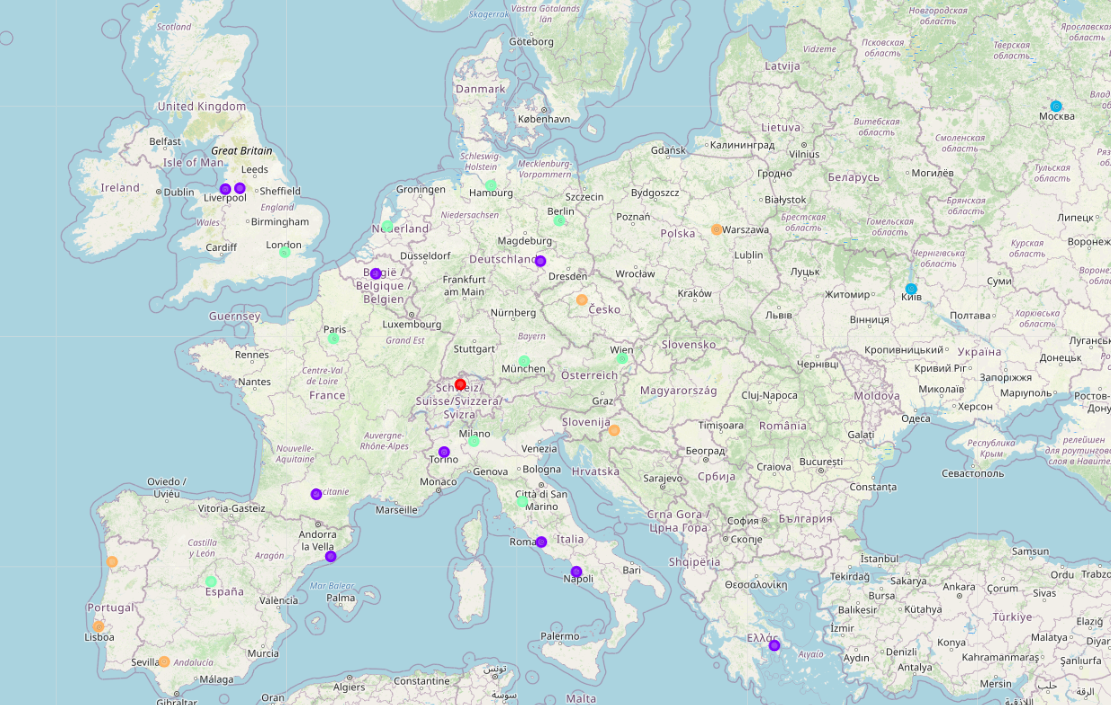


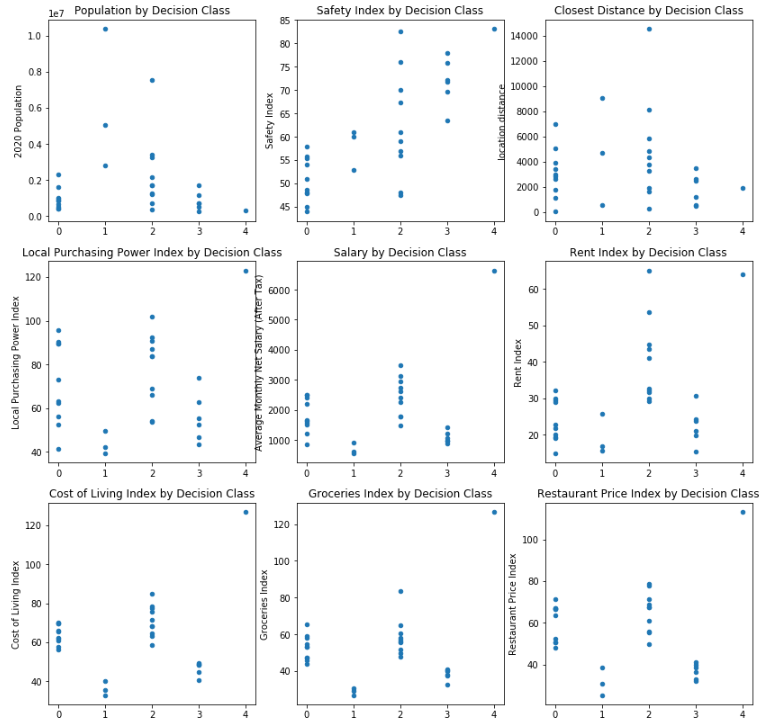
Figure 1. European Map with 30 Cities Classified

Figure 2. Scatter Plots of 9 Attributes by City Classes

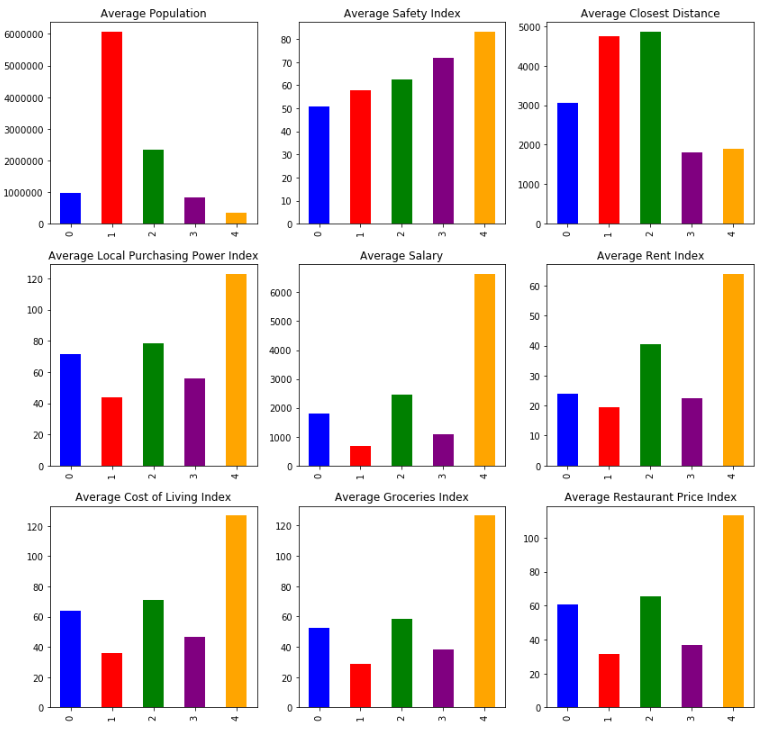


Figure 3. Bar Plot of 9 Attributes in Average by City Classes

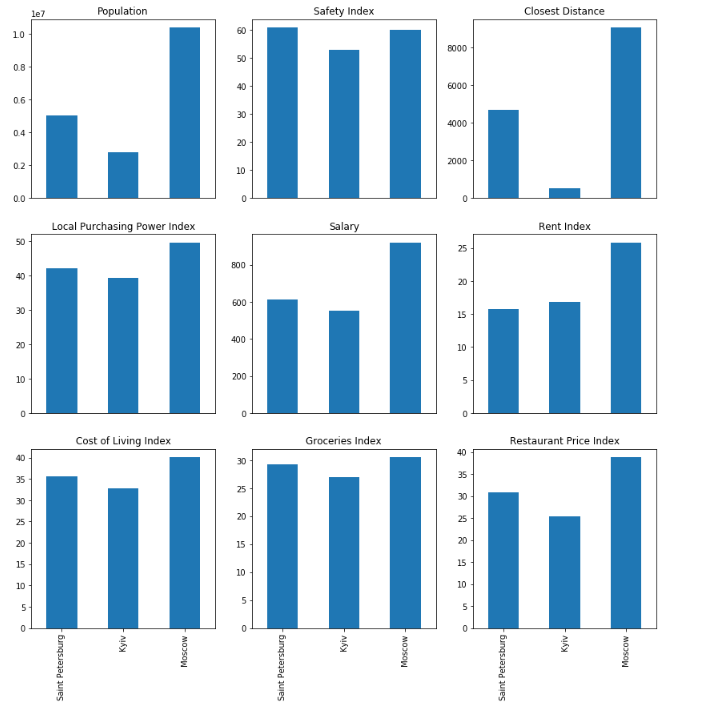


Figure 4. Bar Plot of 9 Attributes of Moscow, Kyiv and Saint Petersburg

**Discussions**

Analyzing Figure 2 and Figure 3, Class 1, where Moscow, Kyiv and Saint Petersburg are classified, has the highest population, distance to the stadium from the closest Chinese restaurants, moderate safety index and the lowest index for all the cost considerations.

Analyzing Figure 4, all three cities have similar attributes while Moscow has the highest population and distance to the stadium from the closest Chinese restaurants. Therefore, the analysis suggests that Moscow is the most favorable city to invest in a Chinese restaurant in close proximity to its local stadium.

The analysis can be significantly improved if the data of the number of annual Chinese visitors to each city and estimated football fans in each city are available.

**Conclusions**

Provided the data available the analysis suggests that Moscow is the most favorable city to invest in a Chinese restaurant in close proximity to its local football stadium.