# Battle of cities

EUROPEAN SOFTWARE COMPANY WANTS TO EXPAND INTO THE U.S

## The business problem

The management of a successful European software firm decides that in order to further advance the company's market share, a new R&D office is needed in the United States.

Being European, they have little knowledge about the U.S: they can't decide which city to target. What they do know is that quality coding requires quality coders, and quality coders are only happy if they can work and live in an ideal environment.

What makes an environment ideal? Cities with high quality of life, low crime, good weather, good healthcare, low pollution all comes to mind.

The goal of this research is to use data to find cities within the United States best suited for being the location of the new R&D office, keeping in mind the factors mentioned above. A recommendation of neighborhood within the highest ranked cities should also be part of the final report: the neighborhood chosen should be rich in amenities for the developers to spend their free time and hard-earned pay: restaurants, cinemas, parks all come to mind.

## Data sources

The following data sources should be considered:

- Wikipedia for weather-related data
- Kaggle datasets relevant to the task, e.g. quality-of-life data
- Foursquare API for getting info about the venues in cities

## Data understanding and preparation

The datasets should be first explored via descriptive statistics and data visualization techniques and then cleaned and transformed into a format where machine learning algorithms can take them as input for deriving models that can help quantify differences, show similarities between cities.

## Modelling

K-Nearest Neighbors can be used compare weather aspects of candidate cities to cities the stakeholders know and love in Europe.

Multiple linear regression can yield insights of how different factors work together to establish a certain “quality-of-life” found in one of the input datasets.

K-Means clustering can find different types of neighborhoods within a city.

## Conclusion

The output of the investigation will be a report that communicates the findings in a way that’s interpretable for stakeholders: they should be happy about their choice of city and should be confident moving forward with their plan of expansion into the United States.

## Final thoughts

This research is an iterative process: stakeholders will provide feedback that can help modifying or expanding models, also during model preparation and evaluation, the need for additional data or different format of data can arise: this is normal and each iteration should improve the quality of the findings.