

Battle of cities

EUROPEAN SOFTWARE COMPANY WANTS TO EXPAND INTO THE U.S.

Introduction/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

The following data sources should be considered:

Foursquare API¹

We'll use Foursquare for getting info about the venues in cities such as trending venues and venue exploration, where we'll input city neighborhood coordinates and receive a list of venues nearby.

Wikipedia

We'll use Wikipedia as our source for weather-related data, such as average temperature, average rainfall and so on. An example of the data we'll be working with:

¹ Endpoint documentation at: <https://developer.foursquare.com/docs/api/endpoints>

Saint Pierre and Miquelon	Saint-Pierre	-2.6 (27.3)	-3.2 (26.2)	-1.4 (29.5)	2.0 (35.6)	5.6 (42.1)	9.6 (49.3)	14.1 (57.4)	16.2 (61.2)	13.5 (56.3)	8.9 (48.0)	4.5 (40.1)	0.4 (32.7)	5.7 (42.3)
United States	Albuquerque	2.4 (36.4)	5.2 (41.4)	8.9 (48.1)	13.3 (56.0)	18.7 (65.6)	23.8 (74.9)	25.7 (78.3)	24.6 (76.2)	20.7 (69.3)	14.2 (57.5)	7.2 (44.9)	2.4 (36.3)	14.0 (57.2)
United States	Anchorage	-8.3 (17.1)	-6.6 (20.2)	-3.0 (26.6)	2.7 (36.8)	8.8 (47.8)	12.9 (55.2)	14.9 (58.8)	13.7 (56.7)	9.2 (48.6)	1.6 (34.8)	-5.4 (22.2)	-7.2 (19.0)	2.8 (37.1)
United States	Atlanta	6.3 (43.3)	8.4 (47.1)	12.3 (54.2)	16.4 (61.5)	21.1 (69.9)	25.2 (77.4)	26.9 (80.5)	26.3 (79.4)	22.9 (73.2)	17.1 (62.8)	12.0 (53.6)	7.3 (45.1)	16.8 (62.3)
United States	Austin	10.8 (51.5)	12.8 (55.0)	16.5 (61.7)	20.7 (69.2)	24.8 (76.6)	27.9 (82.2)	29.4 (85.0)	29.9 (85.8)	26.7 (80.0)	21.8 (71.2)	16.1 (61.0)	11.4 (52.5)	20.7 (69.3)
United States	Baltimore	0.8 (33.5)	2.4 (36.4)	6.8 (44.2)	12.4 (54.3)	17.6 (63.6)	22.8 (73.0)	25.3 (77.6)	24.3 (75.7)	20.2 (68.4)	13.7 (56.7)	8.3 (47.0)	2.9 (37.3)	13.1 (55.6)
United States	Boise	-0.4 (31.3)	2.5 (36.5)	6.9 (44.5)	10.4 (50.8)	15.1 (59.1)	19.7 (67.5)	24.3 (75.8)	23.7 (74.7)	18.3 (64.9)	11.6 (52.8)	4.4 (40.0)	-0.7 (30.7)	11.4 (52.5)
United States	Boston	-1.5 (29.3)	0.0 (32.0)	3.7 (38.6)	9.1 (48.4)	14.6 (58.2)	20.0 (68.0)	23.2 (73.7)	22.4 (72.4)	18.4 (65.2)	12.4 (54.3)	7.2 (45.0)	1.7 (35.0)	10.9 (51.7)
United States	Charlotte	5.1 (41.2)	7.2 (45.0)	11.3 (52.3)	15.8 (60.5)	20.3 (68.5)	24.7 (76.4)	26.4 (79.6)	25.8 (78.4)	22.2 (71.9)	16.3 (61.3)	11.1 (51.9)	6.3 (43.4)	16.1 (60.9)
United States	Chicago	-5.6 (21.9)	-2.9 (26.7)	5.4 (41.7)	12.6 (54.6)	16.5 (61.7)	21.8 (71.2)	25.3 (77.6)	24.8 (76.7)	18.4 (65.1)	13.3 (56.0)	5.3 (41.5)	-5.3 (22.5)	10.8 (51.4)

Figure 1 Average temperature data ²

Kaggle

The Movehub City Rankings³ will be one input. This dataset has valuable features such as Health Care index, Pollution index and a more general and ambitious-sounding “Quality of Life” index for major cities around the world. We’ll just concentrate on U.S. cities.

The “Hurricanes and Typhoons, 1851-2004” dataset⁴ will be important as well, Europeans are afraid of such extreme weather phenomena, and high frequency of these can be a reason for vetoing a candidate city.

All datasets shall 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.

² Source: https://en.wikipedia.org/wiki/List_of_cities_by_average_temperature

³ <https://www.kaggle.com/blitzr/movehub-city-rankings>

⁴ <https://www.kaggle.com/noaa/hurricane-database>