Applied Data Science Capstone Project - Katarzyna Hejczyk-Mika

"Business integration trips in the Silesian region".

Data

A description of the data and how it will be used to solve the problem

Geometries of cities in the Silesian region were extracted from json file taken from the NYU Spatial Data Repository website[1]. Those data represent the second-level administrative divisions of Poland and contain all regions (voivodeships) and their corresponding cities in Poland. The geometries are represented as multipolygons. To find the centroid location coordinates values for each city, the library Shapely[2] was used. Geometries of cities belonging to the Silesian region were used to look for venues in those regions and to visualise results on maps.

Venues of interest in the Silesian region were found using the Forsquare API[3]. Multiple searches were performed to look for all venue categories, search for selected venues and explore popular spots. The foursquare website also contains information about all venue's categories and subcategories. Initially, the following categories were chosen to find the most recommended venues in all cities belonging to the Silesian Voivodeship: Cave, Forest, Hill, Lake, Mountain, National Park, Nature Preserve, River, Bowling Alley, Go-Kart Track, Golf Course, Golf Driving Range, Paintball Field, Reservoir, Scenic Lookout, Sculpture Garden, State / Provincial Park, Windmill and Vineyard. In the next step, the location of interest was further investigated using categories such as Recreation Center, Convention Center, Resort and Hotel. Finally, the location around the hotel was further explored.

All venues data were analysed, clustered, and the folium library used to visualise data.

---

[1] https://geo.nyu.edu/catalog/stanford-xh662zc5620
[2] https://pypi.org/project/Shapely/
[3] https://developer.foursquare.com/docs/