IBM

Data Science project week 4.

Introduction:

For the IBM Coursera capstone project I want to create a concern about an unequal distribution of recourses in Russia. Because the companies are not distributed equally I want to demonstrate the inequality and give a visual justification for the problem.

The problem

The concern is, of course, that the corporations are not distributed equally because of the initial problem with infrastructure. However for industries are forced into the further regions because of their specialty: oil, forestry, etc.

By using data analysis I will demonstrate this problem and, hopefully provide an insight for the companies to set their headquarters elsewhere. This also will partly solve the infrastructure problem, due to the increased investments into the regions as a whole.

Target audience:

Corporations and the potential branch managers.

Data:

We need:

* The current location of the company venues
* List of the location cities
* The industry list

Extraction of the data:

* Wikipedia table of companies, their locations via cities
* Geocoder package
* Beautiful soup package
* Foursquare API