

Building an Information Retrieval System using University Data

PRI 2023/24

GONÇALO ALMEIDA

ŽAN ŽLENDER

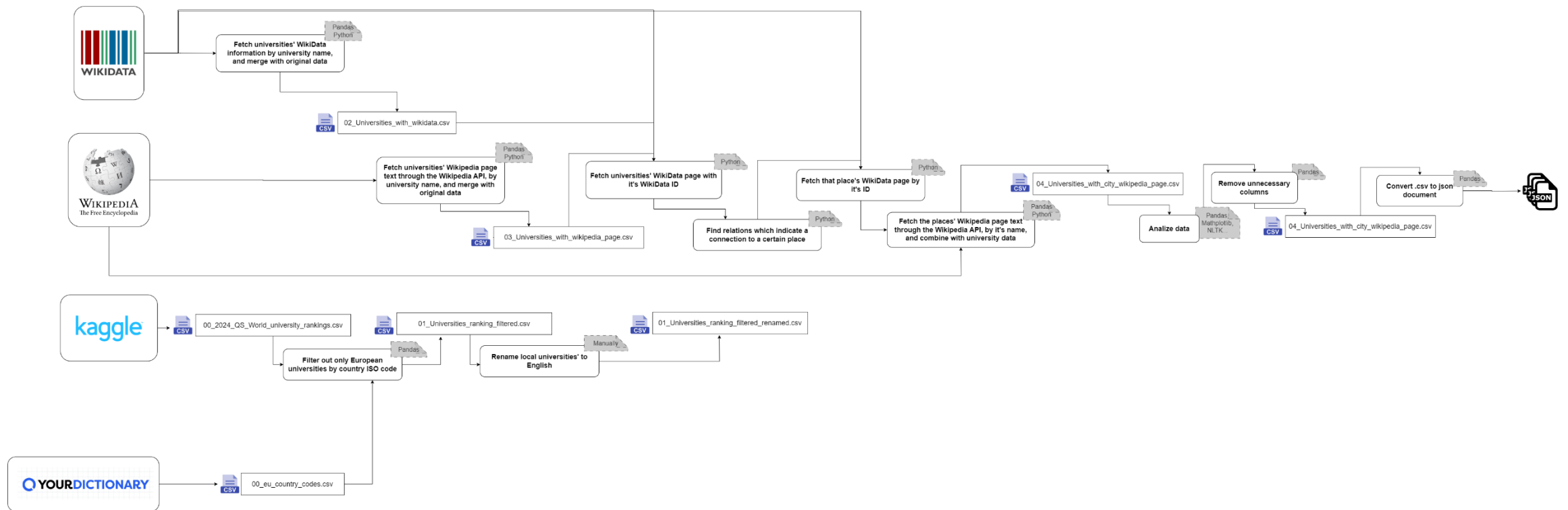
ILINA KIROVSKA

Main idea

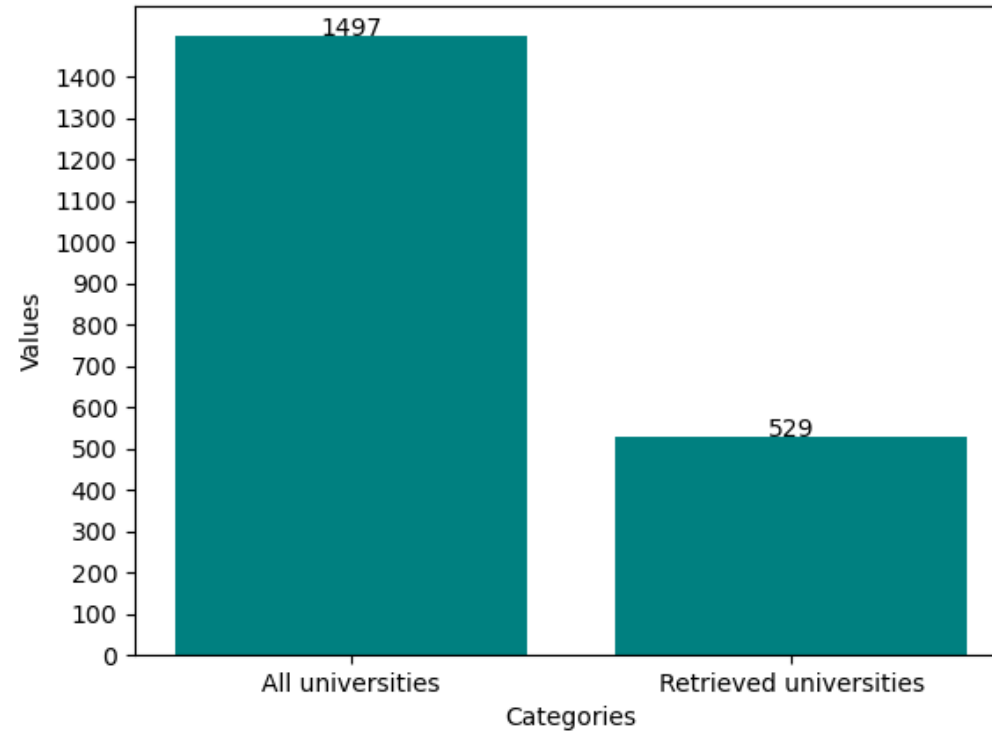
Build a search engine where students can get as much information about universities and various aspects related to them.

- ☐ Possible search tasks:
 - ☐ Search for universities by country or by city
 - ☐ Search for universities by rank, size, age
 - ☐ Search for the best university in a given country/city
 - ☐ Search for information about a city by university
 - ☐ Search for universities/cities by using the most common words(keywords) from the Wikipedia text

Data pipeline

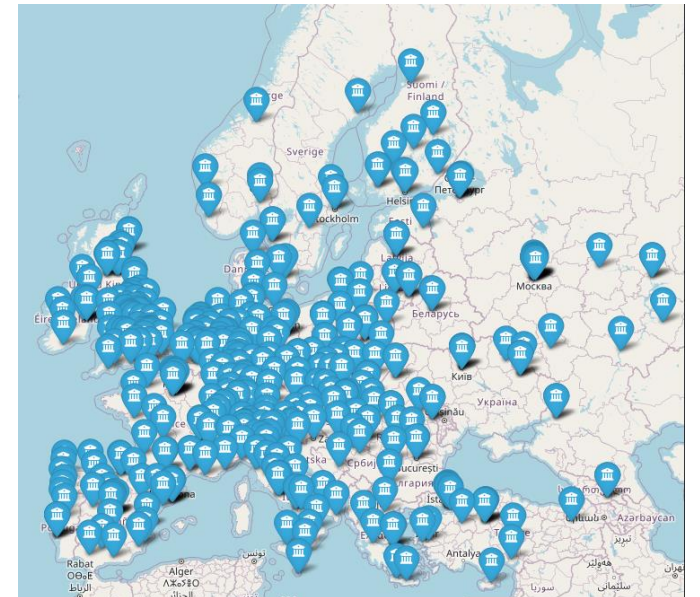
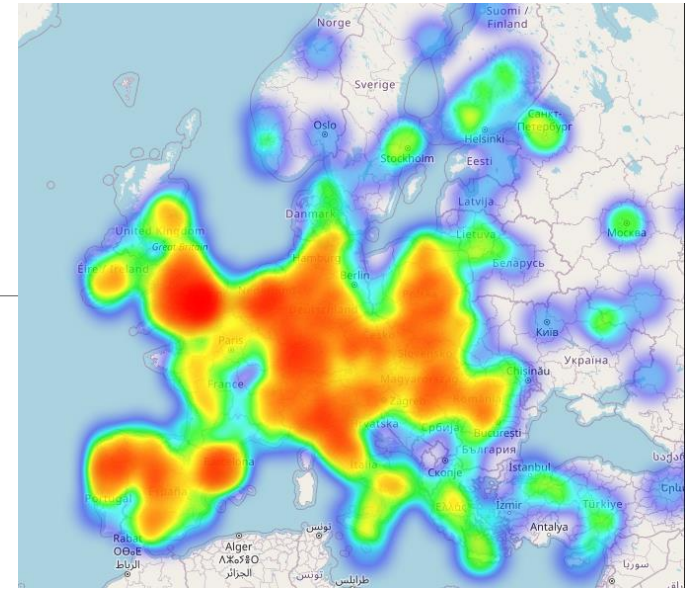
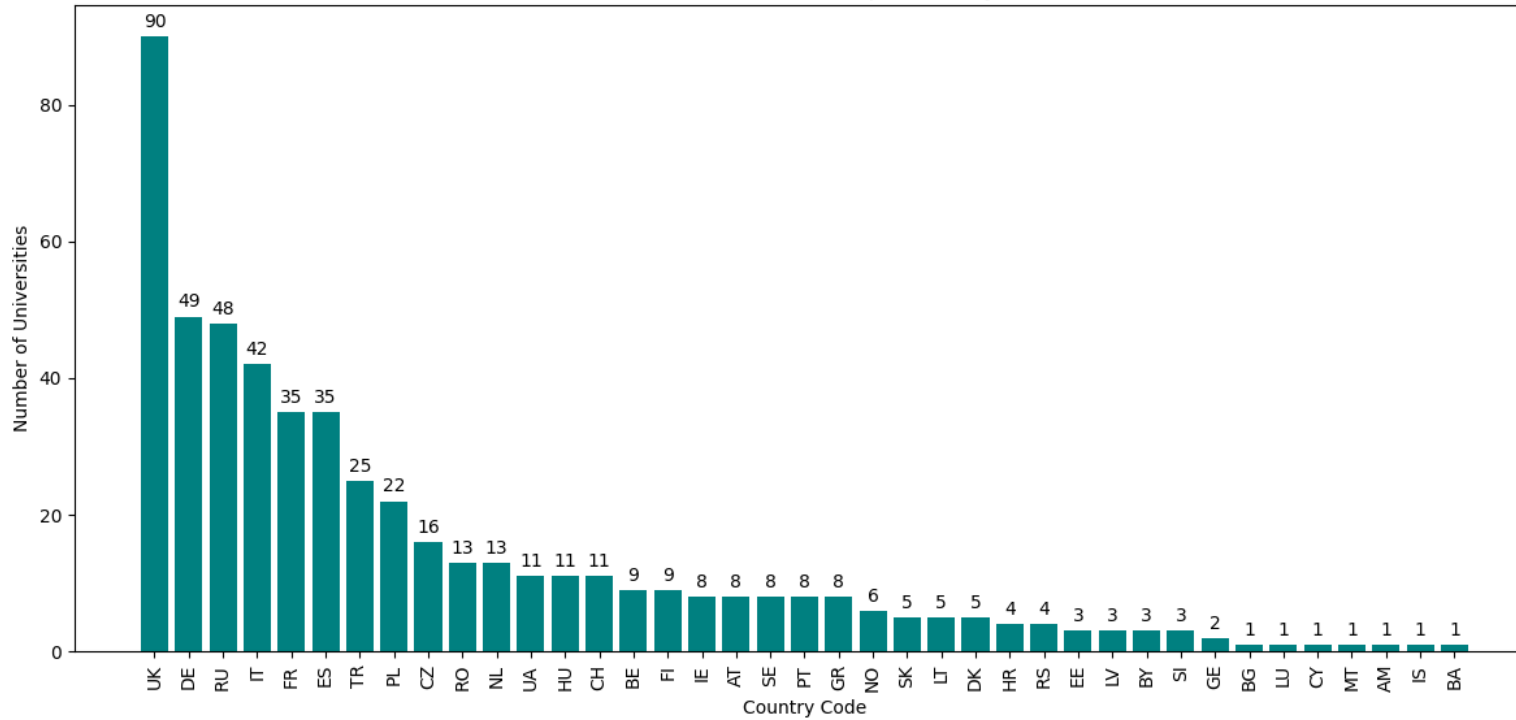


Data characterization

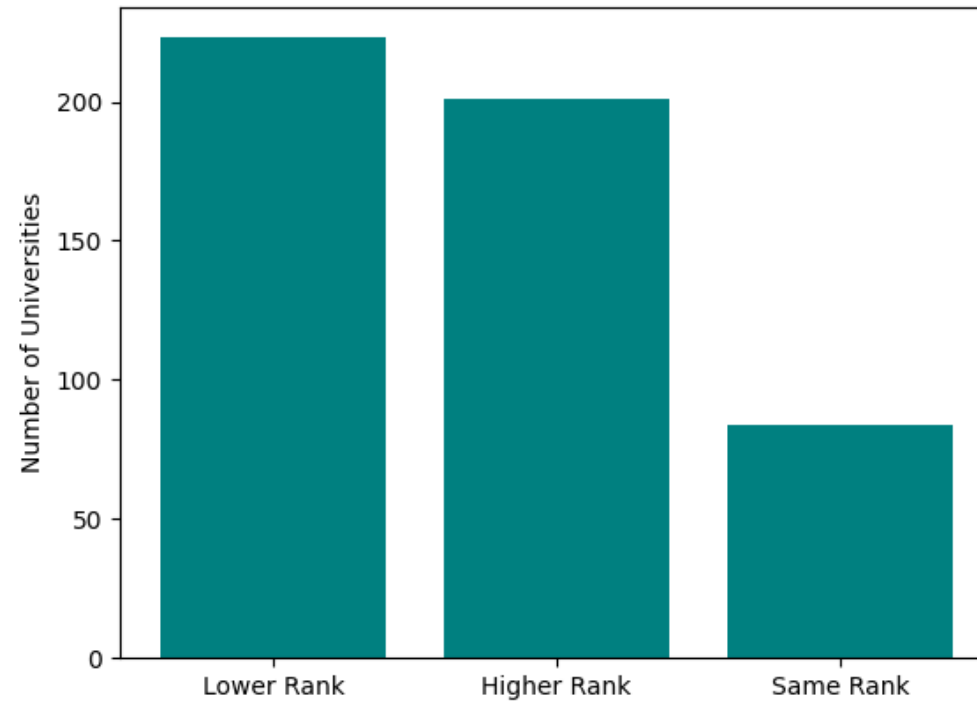


Location analysis

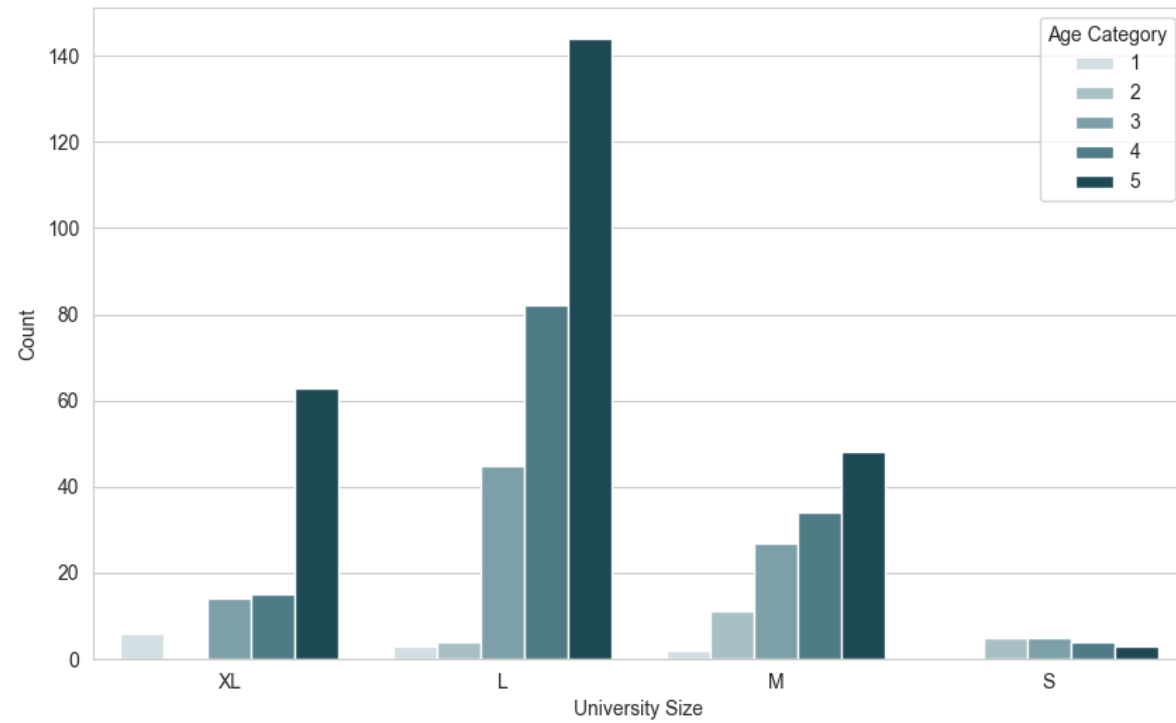
Number of Universities per Country



Rank analysis



Size and age analysis



Text analysis



Future work

After gathering, cleaning, and storing the data, as well as performing an in-depth analysis to identify its key features we have created a reliable dataset that will serve as the foundation for milestone 2, where we will implement an information retrieval tool on the project datasets and explore it using free-text queries.