What countries to Travel in a Limited Time

Athar Kamal

4th November, 2019

1. Introduction

Background

People all around the world have ideas of travelling to Europe in the summers for a “Euro trip”. Students can take their time because of summer breaks, and enjoy the continent’s 44 countries. However, people who work have limited time and cannot leave their places of work for extended period of time, but the need to enjoy most of Europe is still with them. These travelers want to maximize the number of cities they can travel, and enjoy the different cultures.

Problem & Interest

This project aims to cluster the most populous and famous European cities by estimating the population density as well as the closeness of the airport to the city center so that travelers (who have limited time) can enjoy and maximize different cities and countries.

2. Data Acquisition and Sorting

Data Sources

The initial data about the European countries and cities has been acquired from the Wikipedia page with the list of urban areas. The data was scraped through the *beautifulsoup4* tool and saved into an initial dataframe. Using *GeoPy* , the latitude and longitude of each city were imported to form a dataset. Furthermore, Foursquare API was used to search the cities for airport terminals and the distance of the airport from the city.

Data Cleaning

Scraping through the Wikipedia page gave us these five columns (Rank, City, Country, Population and Density). We got a dataframe of size 86 rows x 5 columns. The data was further enhanced by finding the latitude and longitude of each city. However, 3 of the cities did not have any latitude and longitude and were thus removed. The cities were mapped on Europe through the folium library, which revealed that 4 more cities

had wrong latitude and longitude and these were subsequently removed as well. More cities were dropped using the Foursquare API, to estimate whether an airport near a city existed within a radius of 5 kilometers. Thus, our final dataset contained 42 cities.