**How similar are Indonesian Embassies based on their location**

boy.setiawan

April 4th, 2021

1. **Introduction**
   1. **Problem Description**

**An Embassy** is a representative of a country in other countries, their existence helps to indicate a relation among countries and serve as a way to communicate or strengthen the ties. Their location follows a strict and complex requirements both from the country it comes from and the country it resides, usually in a special diplomatic compound or district. Despite all of the careful planning and requirements, the decision to establish an embassy could have come from other necessities such as certain neighborhood/district/area, near to and close from certain amneties, places that could support the embassy mission etc.

The knowledge of how a certain embassies are similar or different could **help give a bigger view to Indonesian Foreign Affairs Officials** to understand the general environment their embassies are located. Should a certain embassies need to be treated differently, do embassies with certain criteria experience the same or different stress level of working for their staffs, do certain embassies experience certain disturbance etc.

The task to find and group embassies into their own similar group could be a daunting task. From forming the basic data, to gathering and analyzing the data must be done without bias and subjectivity. The end result must be able to give an insight of what the embassies really are so they can be interpreted and understand as a whole new information or to strengthen the already known facts.

* 1. **Data Description**

In order to accomplish the analysis, we will need data on Indonesian Embassies abroad and their latitude and longitude so we can combine it with data about venues surrounding the embassies which we intend to explore from FourSquare API to give a better understanding about the embassy's neighborhood, venues and places.

* 1. **Solution**

Our solution is to cluster the embassies and see the differences and similarities between them based on the data gathered about venues surrounding the embassies.

* 1. **Methodology**

1. **Data Gathering**
2. List of Indonesian Embassies Abroad

Apparently to get data about the list of Indonesian Embassies abroad is going to be a bit difficult, this is because the Ministry of Foreign Affairs website doesn't display the data to be easily analyzable. But with the help of google and wikipedia we find a page that we could extract the data that we need as a starting point.

<https://id.wikipedia.org/wiki/Kedutaan_besar_Republik_Indonesia>

after scraping the page with BeautifulSoup we manage to import the data in a Data Frame as so.

Table

Description automatically generated

There are 98 Indonesian Embassies according to our data. As we can see the data needs to be cleaned since the scrapping process leave us with some characters we need to eliminate.

Table

Description automatically generated with low confidence

After we clean the Data Frame we need to change some values and drop some rows from the Data Frame. This is because some of the data used pronunciation of capitals in Indonesian language, others not so common name and data that does not interest us such as Indonesian representatives for the United Nations.

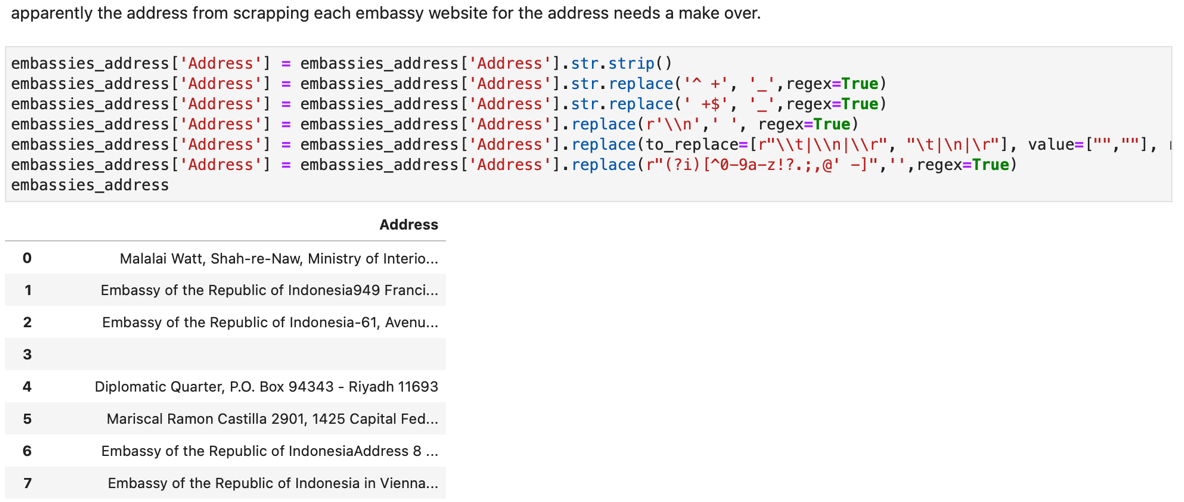
The final tally is 93 Indonesian Embassies abroad.

1. The Embassies Address

It would be nice if everything was served on a plate but in this task, we need to have some wits up our sleeves. What we need are longitudes and latitudes of the embassies, as with the problem above it is not readily available to us. But there is hope since the address is available for us to extract from the Ministry of Foreign Affairs website for they have a common URL that we can explore.

<https://kemlu.go.id/CAPITAL>

with this information we need to iterate this URL with our Data Frame and do another scrapping to get the all the addresses.



After looking at the result we need to take a decision to save the data frame and edit the address manually since the format is not unison. Not a very satisfying task but something we need to take since creating regex rules does not look like a viable solution for this problem. After that we integrate it with our Data Frame and voila a list of Indonesian Embassies abroad with their address.

Table

Description automatically generated

1. The Embassies Geo Location

As our initial data requirement is to have a list of embassies with their geolocation, we need to translate the address into latitude and longitude coordinate. To do that we use a common library in python called geocoder and loop through the address.

Graphical user interface, text, application, email

Description automatically generated

And then format the result in a Data Frame and combine with our embassies Data Frame.

Table

Description automatically generated

Unfortunately, there are some embassies without geolocation information.

Table

Description automatically generated

In this case, 40 embassies are still without geolocation information.

To overcome this obstacle and after googling about free forward geocoding, we stumble open positionstack.com which offer a diverse and rich API in their service. After following the instruction and documentation, we translate the remaining address to get their latitude and longitude.

Graphical user interface, text, application, email

Description automatically generated

And do the process all over again and see if there is any embassy without geolocation information, which there are and decide to do it manually since the number is quite small.

Table

Description automatically generated

Let’s see if we get our geolocation data right with displaying a sample.

Map

Description automatically generated

1. The Embassies Venues

Now let’s get venues data surrounding the embassies from FourSquare.

Graphical user interface, text, application

Description automatically generated

Graphical user interface, table

Description automatically generated

Which resulted in an astonishing 2.232 venues which we can work on.

1. **Data Exploration**
   1. **Discussion**
   2. **Conclusion**