

# Applied Data Science Capstone

Looking for some  
inspiration about where to go?



# Problem

## Introduction: The business problem and the stakeholders

In economics, Okun's law is an empirically observed relationship relating unemployment to GDP. It states that.. “for every 1% increase in the unemployment rate, a country's GDP will be an additional roughly 2% lower than its potential GDP”.

In our “increasingly mobile society” and globalized economy, working abroad is no longer a novelty for the fortunate few – it's an ordinary fact.

There are many exciting gains to be made by taking the plunge and settling overseas. It increases the possibility to find your “dream” job and it is also a great way to actually save money. Moving to another country offers plenty of opportunities for boosting your career, helping you to acquire new, varied skills and experience, and to establish an international network which may well pay dividends in the future.

So.. Looking for some inspiration about where to go? Do you want to see the world, travel, explore, meet people, discover new cultures, and escape the 9–5 with its commutes, endless meetings and oppressive cubicle walls. Maybe you've just graduated from college or high school, but you just don't feel ready to settle down – after all, there's a whole world out there to see?

Sounds too good to be true? Curious about this idea, but confused where to start? Well, think again! Because, that is exactly what tens of thousands of people are doing right now, all across the world.



# Data Analysis

- I. The wikipedia page of List of countries by GDP (nominal) and unemployment rate ([https://en.wikipedia.org/wiki/List\\_of\\_countries\\_by\\_GDP\\_\(nominal\)\)](https://en.wikipedia.org/wiki/List_of_countries_by_GDP_(nominal)))-([https://en.wikipedia.org/wiki/List\\_of\\_countries\\_by\\_unemployment\\_rate](https://en.wikipedia.org/wiki/List_of_countries_by_unemployment_rate)) contains all relevant details in order to find the 10 countries with the highest GDP and lowest unemployment rate. I have used Beautifulsoup and pandas library to create the initial data-frame. For a clean and understandable data-frame some of the rows are renamed or dropped.
- II. The wikipedia page of Special wards of Tokyo ([https://en.wikipedia.org/wiki/Special\\_wards\\_of\\_Tokyo](https://en.wikipedia.org/wiki/Special_wards_of_Tokyo)). I have used Beautifulsoup and pandas library to create the initial data-frame. For a clean and understandable data-frame some of the wards are renamed. After this initial preparation, I moved on to the next step to obtain coordinates using Geopy library and Folium to visualize geographic details.
- III. The average rent-price data for each ward of Tokyo was obtained from Tokyo Rental Apartment average value page (<https://utinokati.com/en/details/apartment-rent-market/area/Tokyo/>).
- IV. Finally, I make use of Foursquare API to obtain the most common venues based on my preferences within 1 kilometer of each major district. Concluding our project the K-means algorithm was used to cluster the boroughs. K-Means algorithm is one of the most common cluster method of unsupervised learning. Using this method we will be able to cluster our results per area and examine their properties accordingly.

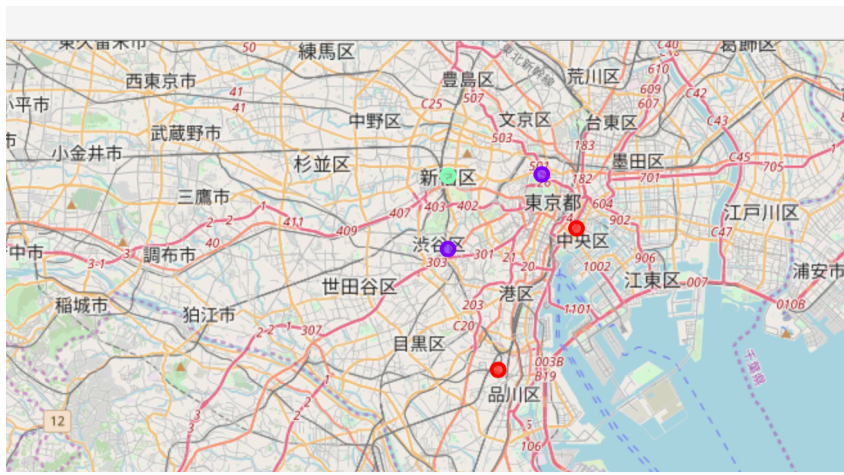
# World Map with ratio data



World map.

# Results

**Tokyo東京! Just Perfect. There's no lovelier sight than the sakura trees on a sunny spring morning - symbolising new beginnings, they put us in a good mood for the rest of the day.**





# Conclusion

The average rent prices are overall the same, with in the range 3.7 - 4.3.

Since I'm in love with Japanese cuisine I will go with Chuo and Shinagawa ie cluster 0 because the Japanese Sushi Ramen Restaurants are the most common venue.

Some points that are to be followed in the future for more accurate and better results are the following:

- Built useful models to predict whether and how much a country's economy will improve
- Accuracy of the models has room for improvement.
- Capture more data from the economy in order to have a better fit on the model and improve its accuracy in the predictions.
- Ideas include:
  - More macroeconomic variables
  - Financial data