Aviation Project

# Goal

To find a suitable destinations/routes for Airline companies to consider using data-driven approach

# Numbeo.com

Numbeo is website that surveys its users about the living condition in their particular city. Numbeo is extremely useful for us because it provides us with the data we will be using in out formula. The key used is "﻿eojgmx7e3dxnl9", if that doesn’t work then head to their API page (<https://www.numbeo.com/common/api.jsp>) and follow instructions.

# Definitions

|  |  |
| --- | --- |
| Safety index | A measure of how safe a city is |
| Quality of life index | A measure of how good life is in a particular city. It takes into account the following:   * Cost of living * Purchasing power * Levels of pollution * Crime rates * Health system quality * Traffic |
| Safety index | How safe a given city is |
| Cpi index | Consumer Price Index: a measure that indicates the general price of every-day-goods/services, such as a carton of milk or the cost of a taxi ride |
| Groceries index | An indication of the general prices of goods typically found in a groceries store, such as milk, eggs, or meat |
| Rent index | A measure of the amount of rent paid in a particular city relative to the rest of cities |
| Traffic index | An estimation of time spent commuting in a particular city relative to the rest of cities |
| Restaurant prices index | An estimation of the cost of a meal in a particular city relative the rest of cities, it can range from 5-stars restaurants to fast food |

# Methodology

When you make an ﻿indices request to Numbeo, you will get JSON file with multiple attributes, the ones we will focus on are the ones listed in the definitions section (and the amount of the contributions for each one).

First you should start by getting all the cities in their database, and then calculate the distance of each from any given point (Riyadh in this case). Then, you should set a range that you want your targeted destinations to be within and exclude everything else (4000km in this case). Once you have your list, you will then get the indices for each using the ﻿get\_indices\_by\_id function. You can find these steps in In[1] through In[9] in the code.

Once you have your indices, you will then iterate through them and check if:

* They have all the values we need
* They have enough contributors (30 in this case)

If they meet these requirements, we will then apply the formula below In[10]:

Now that you have your scores for each city, you can sort them and write them to file (In[11]/In[12]).

Then, you will need to check the top x cities in your sorted list in TripAdvisor and score them manually (you can try to get API access, we couldn’t do that due to their strict requirements, but it may be loosened in the future). Scores that were picked are as follows:

* Stars
* Reviews / opinions
* Number of phots

Once you got that, it's time to gather some meta-data (average stars, reviews, and photos. In[16] and In[17]).