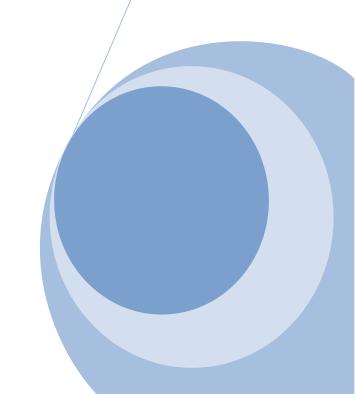# Clustering of the world most visited cities

How to recommend the next best travel destination to clients

The aim of this paper is to help travel agencies in grouping typical destinations based on the most common types of venues found in each location. This will allow a more informed recommendation for their clients seeking advice on where to go next.

Fabrice PETITFRERE
4/20/2020

Contents

PART 1 – INTRODUCTION

Introduction

Do you know what it feels to be thrown in a foreign city and be miles away from home, losing all sense of familiarity? That's what is so particular with our planet is that there is such a variety of cultures and environments. Sometimes however, even far from home, you feel a sense of familiarity with the place you're staying and/or visiting and can't yet tell why this is. There are obviously similarities among all the other differences between towns across all the continents. These similarities are sometimes noticeable enough to make you feel in a familiar place even though environment, culture or language is different. Whether you like a place or not is sometimes linked to these little things, hard to notice. What if we are able to cluster the world cities according to the number and the categories of venues found in each location?

Business problem

Travel agencies role is (more often than we can think) to advise their customers on best holiday locations. Not all people decide of their next holiday plans before they walked into a travel agency. Selecting a destination for their customers required to understand their preferences and presumably know of places they visited and which they liked. Destinations are sometimes quite hard to compare and it would be a good idea to have at least the most visited cities arranged in a few similar groups. It would enable travel agents to say : "Oh you've been to Milan last year, I would suggest you try Lisbon if you haven't been as it has some similarities you may like!" (obviously if we can group Milan and Lisbon together but the analysis will tell us!).

Business audience

This work can be helpful to travel agents in need of information to advise their customers.

PART 2 – DATA

Data sources

To undertake this work, we will first need a pertinent list of destinations to investigate, their accurate locations with GPS coordinates, some information about each location including data about different types of venues that can be found there.

List of destinations

Since we are investigating destinations for tourists, we will use a list of the most visited cities in the world. This list of the top 100 most visited cities is ranked by the number of international visitors, including all international arrivals by land, sea or plane, for tourist of business purposes. More precisely, we will the Euromonitor count which counts a visitor as any person visiting a city another country for at least 24 hours, for a period not exceeding 12 months, and staying in paid or unpaid, collective or private accommodation. Each arrival is counted separately and includes people travelling more than once a year and people visiting several cities during one trip. Useful fields to retrieve contains: city name, country and number of arrivals in 2018.

Source: https://en.wikipedia.org/wiki/List_of_cities_by_international_visitors

GPS coordinates

To find the GPS coordinates of each city location, we will use Geopy, a popular python client for geocoding web services. This will return latitude and longitude for each location.

Source: https://geopy.readthedocs.io/en/stable/#

Venues data

To explore each location, we will then use the Foursquare location data and explore each location for the most popular venue categories. This will be described in more details further on but we will only use the main venue categories which consists of 10 distinct categories.

Source : https://developer.foursquare.com/docs/build-with-foursquare/categories