# Introduction & Business Problem

With international traveling becoming more accessible throughout the world, deciding on where to go for your next trip is increasingly that of an arduous task. Where each convenience for which the intention is to give travellers a better understanding of potential destinations is followed by confusion if presented with too much information and too many options. Though detailed prospecting of destinations and potential planning cannot be avoided, alleviating the process with simple but relevant clues in areas such as relative density of landmark sites, child friendliness and cuisine could facilitate the final decision. Descriptive analyses could certainly be applied in order to help reach decisions on where to go and would most likely be the sole option for first time travellers as well as for those with very specific tastes. To alleviate the process, segmenting cities according to their venues characteristics around the city cores we can provide generalised profiles for prospective travellers to make use of when deciding on their travel destinations.

Once the travel profiles have been created, further application refinement could be developed by adding further data layers from either other FourSquare services internally or by connecting data from external API services such as TripAdvisor to our own as to add detailed information on for example identified venues, which in turn could lead to further refinements in our clustering of travel destinations.

## Main Audience

The main audience for the app will be **prospective travellers** **of all kinds** who will be given a choice of a set of segments of cities for their next travel destinations, characterised by venue and amenity availability rather than for example geography and popularity.

## Secondary Uses

Other use cases would be for **establishment owners** to use the differences between the clusters to identify business opportunities. Cities with a low hotel density could be interesting for **developers**. Up and coming expat cities will need amenities to cater for the influx of the new category of residents for which **city planners** can use data to pro-actively cater for relevant entrepreneurs to set up shop. In the same way, cities with ambitions to attract tourists may want to understand what world tourist cities make available for their guests.

# Data Sources

Find below the main data sources to be used in this exercise:

**Data Source 1 -** <https://www.worldatlas.com/citypops.htm>

The data list is a gross list of the largest cities in the world. Other criteria may be used such as most visited destinations to further find already, in a sense, empirically proven concepts. Doing so would however potentially filter away alternatives that may not have been “discovered” yet. The main parameters that we will be using are the city names for which we can use to extract the location co-ordinates for insertion into the Foursquare API calls. However, for the sake of completeness, we will also extract the population should we want for example to create per capita measures later on.

|  |  |  |
| --- | --- | --- |
| **Data Source** | **Feature** | **Description** |
| World Atlas Web Page | City | Name of cities to be used to extract the geolocations of the same. |
| World Atlas Web Page | Population | Population of cities in absolute count. Will be used as reference only and only included should there be measures that can be transformed using the values. |

**Data Source 2 – Foursquare Venues API – Venue Details**

Once the cities have been extracted from Data Source – 1 Foursquare Venues API to extract venues and their categories within the **maximum radius of 100,000 meters from the city core**. The features that will be extracted will be only 2 from the Venues API:

|  |  |  |
| --- | --- | --- |
| **Data Source** | **Feature** | **Description** |
| Foursquare Venues API | Name | Name of venue. It will not be used in the clustering but still good to extract to have as reference. |
| Foursquare Venues API | Venue Category | Level 1 venue category to be extracted. Level 2 category is a bit too detailed and will as such be left out of scope. |

**Source:** <https://developer.foursquare.com/docs/api/venues/details>

**Merged data set**

After extracting all relevant features we will be merging them into the below data frame for further analysis & clustering.

|  |  |  |
| --- | --- | --- |
| **Data Source** | **Feature** | **Description** |
| World Atlas Web Page | City | Name of cities to be used to extract the geolocations of the same. |
| World Atlas Web Page | Population | Population of cities in absolute count. Will be used as reference only and only included should there be measures that can be transformed using the values. |
| Foursquare Venues API | Venue Name | Name of venue. It will not be used in the clustering but still good to extract to have as reference. |
| Foursquare Venues API | Venue Category | Level 1 venue category to be extracted. Level 2 category is a bit too detailed and will as such be left out of scope. |