**Educational Migration:**

**Studying the Neighbourhood Amenities of Ljubljana (Slovenia) and Villach (Austria)**

Nic Whitehead  
  
26th July 2021

**1. Introduction**

Tertiary education in Australia can be an expensive and sometimes cost-prohibitive exercise for many residents. Ex-pats (non-Australian permanent residents) living in Australia can be subject to international student rates which are even more highly priced. With no access to government assisted loans, many ex-pats and children of ex-pats are excluded from the quality tertiary education programs available in Australia. č

So, what possibilities do these ex-pats have to educate either themselves or their children without expending vast sums of money or taking on massive personal debt?

Well, for those are lucky enough to have ties to the European Union, or can only afford lower-cost tuition fees, then Slovenian universities may be the answer. Slovenia provides free and affordable courses to both EU citizens and international students. Many of the courses are in English and can be purchased at a fraction of the cost in Australia.

Some parents may want to consider this as an option for their families where finances allow. Then at the same time, take the opportunity to live and work in Europe while their children study at an international university.

The Universities of Ljubljana and Maribor are the 2 top-ranked universities in Slovenia. Ljubljana is the capital of Slovenia and is the site of the main university campus. Villach is a sizeable community just across the Austrian border and is a short 90-minute train ride away.

Maribor is the largest city in Eastern Slovenia. It is the economic, administrative, educational, and cultural centre of the area. Graz is the second largest city in Austria (*pop. 633,168 @ 2015*) and is less than 90-minutes away from Maribor by train. [*(https://en.wikipedia.org/wiki/Graz)*](https://en.wikipedia.org/wiki/Graz)

Both pairs of cities offer many amenities, employment and lifestyle opportunities. However, which city would be best to live in while the children participate in further education?

**2. Business Problem**

The aim of analysis is to help Australians who decide to take their families on an educational migration to Europe. It is intended to assist in choosing which areas in each city will provide the most relevant amenities and thus provide guidance on where to live for a period of time. There are 4 cities to analyse, all of which provide a variety of lifestyle options.

However, depending on where the children are studying (Ljubljana or Maribor), the challenge is deciding which city of each pair will be most appropriate to meet the needs of the family lifestyle.

For the purpose of this analysis the assumption is that a family will be looking for specified amenities within a relatively close range of the chosen location. These amenities are:

* Supermarket
* Shopping
* Cafés
* Restaurants
* Gym/Fitness Centre
* Movie Theatre
* Library
* Medical Centre
* Parks

The report will illustrate which areas in each city has most of the venues of interest within a 1,000m radius. Thus, being the most suitable location to consider staying while a member/s of the family are studying in Slovenia.

**3. Data Description**

Geographical location data is required for Ljubljana, Villach, Maribor and Graz. Postal codes in each city serve as a starting point. Postal codes will be used to find the neighbourhoods, districts, venues and their most popular venue categories.

**3.1 Postcode & GPS Data**

To derive the solution, postcode and GPS data is obtained from a variety of web sources and compiled as a CSV. This is due to scarcity and inconsistent formatting of the information for each location. Each CSV contains the following:

1. *district*: Name of the district within the city location
2. *post\_code*: Postal codes for each district
3. *latitude*: Latitude for district
4. *longitude*: Longitude for district

Ljubljana, Villach, Maribor and Graz have few distinct districts, so compiling this list manually is relatively straight forward.

The CSV will be imported into a pandas dataframe to be used as a starting point for the analysis.

**3.2 Foursquare API Data**

This analysis will require data about different venues in different districts of each location. In order to gain that information, "Foursquare" locational information will be used. Foursquare is a location data provider with information about all manner of venues and events within an area of interest. Such information includes venue names, locations, menus and photos. As such, the foursquare location platform will be used as the sole data source since all the stated required information can be obtained through the API.

After finding the list of districts, a connection to the Foursquare API is established to gather information about venues inside each district, for each city. For each district, we have chosen the radius to be 1,000 meters.

The data retrieved from Foursquare contains information of venues within a specified distance of the longitude and latitude of the postcodes. The information obtained per venue as follows:

1. *Neighbourhood*: Name of the District
2. *Neighbourhood Latitude*: Latitude of the District
3. *Neighbourhood Longitude*: Longitude of the District
4. *Venue*: Name of the Venue
5. *Venue Latitude*: Latitude of Venue
6. *Venue Longitude*: Longitude of Venue
7. *Venue Category*: Category of Venue

Based on all the information collected for Ljubljana, Villach, Maribor and Graz, there will be sufficient data to build the model. Districts will be clustered together based on similar venue categories. Using this data, stakeholders can then compare city areas against the required venue categories and decide which location is most suited.

**4. Methodology**

The model will be created with Python so the following packages will need to be imported to process the data.:

**import** **pandas** **as** **pd**

**import** **requests**

**import** **numpy** **as** **np**

**import** **matplotlib.cm** **as** **cm**

**import** **matplotlib.colors** **as** **colors**

**import** **folium**

**from** **sklearn.cluster** **import** Kmeans

**import** **json**

**from** **geopy.geocoders** **import** Nominatim

**from** **pandas.io.json** **import** json\_normalize

Package breakdown:

* *Pandas*: To collect and manipulate data in CSV and conduct data analysis
* *requests*: Handle http requests
* *matplotlib*: Detailing the generated maps
* *folium*: Generating maps of Ljubljana, Villach, Maribor & Graz
* *sklearn*: To import Kmeans which is the machine learning model that we are using.
* *Json*: library to handle JSON files
* *geopy.geocoders/Nominatim*: convert an address into latitude and longitude values
* *pandas.io.json/json\_normalize:* transform JSON file into pandas dataframe

The approach taken here is to explore each of the cities individually, plot the map to show the neighbourhoods being considered and then build the model by clustering all the similar neighbourhoods together. Finally, a new map is plotted with the clustered neighbourhoods. Then insights and observations can be made from the resulting visuals.

**4.1 Data Collection**

In the data collection stage, we begin with collecting the required data for the cities of Ljubljana, Villach, Maribor and Graz. The data requires postal codes, districts and latitude/longitude specific to each of the cities.

Obtaining this information directly and cleanly from web locations was challenging. Most pages relating to each of these cities had partial data and were very inconsistent in their formatting. This made web-scraping difficult.

Fortunately, there are few districts in each area, compared to other larger European or American locations. Therefore, the decision was taken to research and manually collect the relevant data.

A CSV was created and loaded into github, then read into pandas to begin the location analysis.

Table

Description automatically generated

The table above contains the postcodes, districts and Lat/Long for all four cities.

**4.2 City by City Analysis**

At this point, analysis using the following processes can begin:

* Map visuals
* FourSquare
* K-Means Clustering

For each city location, the same analysis process will be conducted. Then the locations can be reviewed as a group to assist stakeholders in determining which will be the most desirable location to live.

**4.3 Analysis Process**

* Firstly, a map will be created to view each location and its respective districts, using folium and geopy
* Establish a connection to FourSquare using a previously created API
* Return the top 100 venues within 1,000m of each District’s Longitude and Latitude
* Clean and structure the resulting Json file ready for additional analysis
* Group the districts and calculate the mean of the frequency of occurrence of each category
* Limit the venues to the 9 priority categories identified by stakeholders
* Return each District with the frequency of occurrences of the venues of interest
* Stakeholders can then view the data as a pandas table to decide which area in the city might be best based on the criteria.
* The venues will be clustered into 5 groups using K-Means and a plot created using folium and matplotlib
* Each of the 5 cluster will then be returned as a pandas table for final review by stakeholders

To complete the analysis and report, this process will be conducted for all 4 locations (Ljubljana, Villach, Maribor and Graz).

These visuals and tables can be viewed collectively by stakeholders to determine which of these locations would be the most suitable to look for a permanent residence while they or someone from their family studies in either Ljubljana or Maribor.