

# **Educational Migration: Studying the Neighbourhood Amenities of Ljubljana (Slovenia) and Villach (Austria)**

Nic Whitehead

22<sup>nd</sup> August 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.

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>)

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 intended stakeholders for this analysis are Australians who are trying to decide whether to take their families on an educational migration to Europe.

The analysis will 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 purposes of this analysis the assumption is that a family will be looking for an Urban/Suburban lifestyle with certain amenities within a relatively close range of the chosen location.

Some of these amenities are:

1. Gym or fitness centre
2. Cafes
3. Restaurants
4. Supermarkets
5. Parks
6. Cinemas
7. Train Station
8. Theatre
9. Shopping facilities
10. Entertainment Venues

The report will highlight venues of interest in each city, within a 500m radius of each district.

Thus, illustrating 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 was stored in a github repository, then 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 500 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 to their requirements.

## 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.pyplot as plt
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 and creating plots for analysis
- *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 obtaining the required information for the cities of Ljubljana, Villach, Maribor and Graz. The data requires postal codes, districts and latitude/longitude specific to each of the cities.

Sourcing 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.

	City	Postcode	District	Lat	Long
0	Ljubljana	1000	Bežigrad	46.077862	14.517007
1	Ljubljana	1000	Center	46.047856	14.512180
2	Ljubljana	1231	Črnuče	46.103134	14.556109
3	Ljubljana	1000	Dravlje	46.075934	14.455358
4	Ljubljana	1000	Golovec	46.044281	14.556979
5	Ljubljana	1000	Jarše	46.073226	14.553944
6	Ljubljana	1000	Moste	46.054600	14.558398
7	Ljubljana	1260	Polje	46.066593	14.615835
8	Ljubljana	1000	Posavje	46.098312	14.493320
9	Ljubljana	1000	Roznik	46.055225	14.465799
10	Ljubljana	1000	Rudnik	46.009911	14.498683
11	Ljubljana	1261	Sostro	46.046756	14.664018
12	Ljubljana	1210	Šentvid	46.098977	14.455900
13	Ljubljana	1000	Siška	46.071886	14.487714
14	Ljubljana	1211	Šmarna Gora	46.134046	14.466710

\* 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 with Silhouette analysis

For each city location, the same analysis process has been conducted. After all four locations have been completed, the locations can be reviewed as a group to assist stakeholders in determining which will be the most desirable location to live.

### 4.2.1 Ljubljana Analysis

Below is the location map of Ljubljana showing all the locations in the analysis for the city.

Details of the full analysis process also follow.



### 4.2.1.1 Ljubljana Analysis: Clustering of Venues

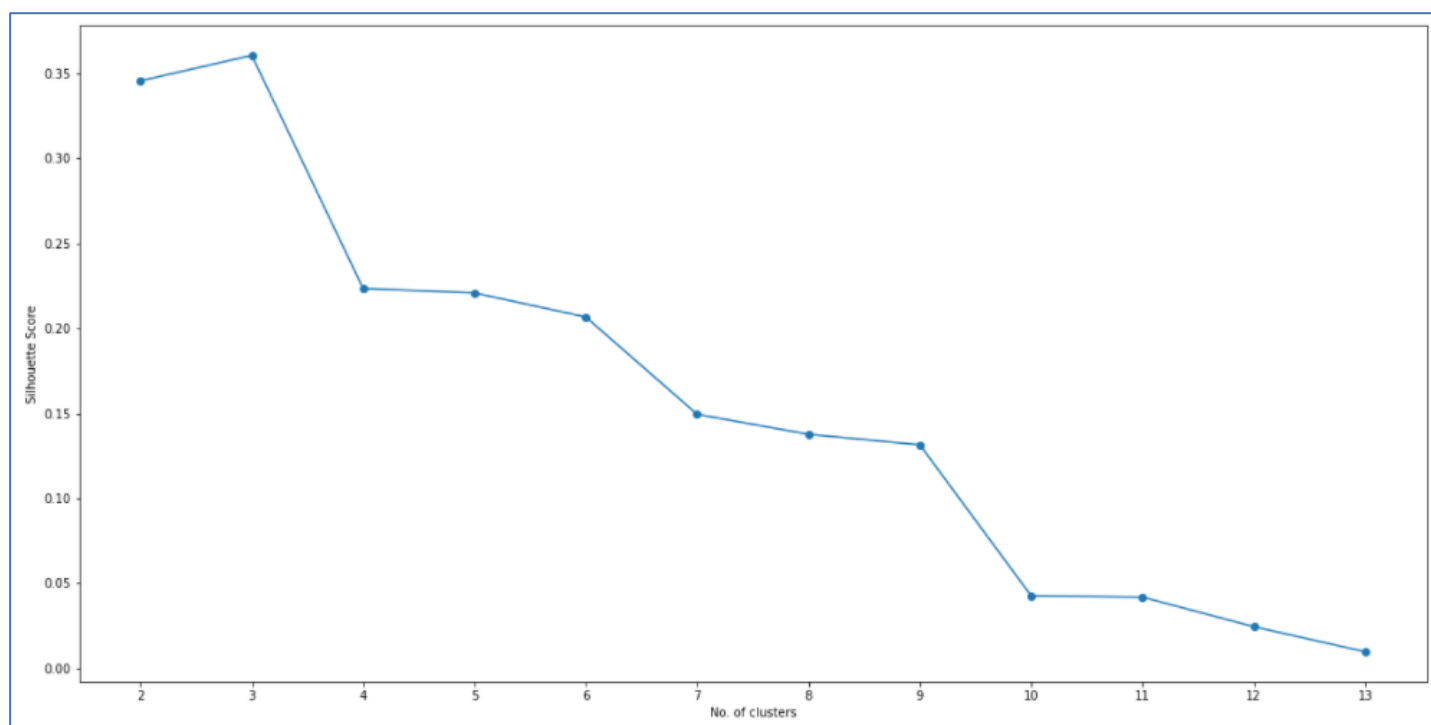
To assist with the decision-making process for potential ex-pats moving to Ljubljana, the districts will be clustered using K-Means clustering. This will group the districts according to the main type of venues found nearby the Lat/Long for the centre of each district. Venues of similar type are clustered together in this un-supervised algorithm.

Once the clusters have been calculated, they are plotted on the city map for easier consumption.

However, K-Means requires a manual entry for the number of clusters. The question is how many clusters will be the optimal number of districts and venues in the sample?

In this case the silhouette method has been chosen to predict the optimal number of clusters for the dataset. The silhouette method allows for a rating of how well separated each cluster is from the other clusters. This is indicated by a value no greater than 1. The closer a value is to 1, the denser and well separated the cluster is from other clusters in the sample. For each city, a simple line plot has been chosen to visually indicate how many clusters will be optimal for the K-Means analysis.

### 4.2.1.2 Ljubljana Silhouette Score Plot

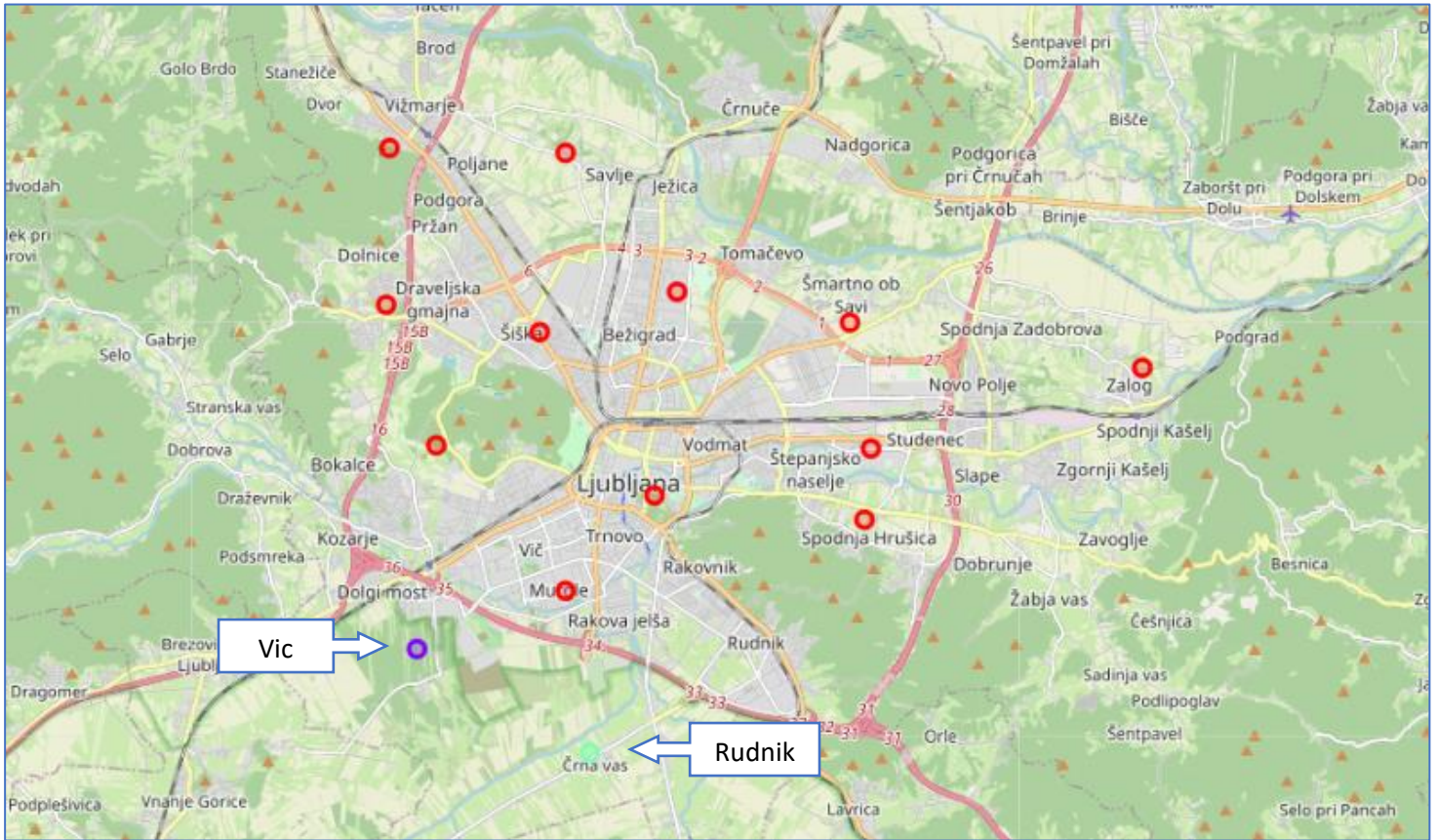


According to the results, the optimal number of clusters for Ljubljana is 3. It is worth noting that at 3 clusters, the silhouette score is only 0.36, which is not a particularly high score by itself. However, when viewed in the context of the other number of clusters, it is proportionally much higher than the majority of clusters tested.

Below is the map plot of Ljubljana with a 3-cluster plot:



#### 4.2.1.3 Ljubljana Clustered Districts (K-Means)



The K-Means algorithm has grouped the majority of districts into one cluster. With two other districts isolated in their own cluster each (Vic – purple/Rudnik – teal).

Once the clusters had been identified and sorted, dataframes were created for each group including their Top 10 venues of interest. These tables can be used by stakeholders to evaluate the suitability of an area for living in.

Each table below represents a colour/group of districts.

## 4.2.1.4 Ljubljana Venues by Clustered Districts

### Cluster 1 – Red Circles

	District	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Bežigrad	0	Pizza Place	Convenience Store	Eastern European Restaurant	Café	Food & Drink Shop	Bus Station	Fast Food Restaurant	Gastropub	Basketball Stadium	Italian Restaurant
1	Center	0	Restaurant	Eastern European Restaurant	Plaza	Coffee Shop	Hostel	Café	Pizza Place	Chinese Restaurant	Vegetarian / Vegan Restaurant	Tea Room
3	Dravlje	0	Bar	Food & Drink Shop	Mobile Phone Shop	Athletics & Sports	Pharmacy	Hardware Store	Historic Site	Hostel	Hotel	Indian Restaurant
4	Golovec	0	Gym	Pub	Bar	Chinese Restaurant	Athletics & Sports	Mexican Restaurant	Paintball Field	Music Venue	Mobile Phone Shop	Liquor Store
5	Jarse	0	Gym / Fitness Center	Restaurant	Athletics & Sports	Pharmacy	Hardware Store	Historic Site	Hostel	Hotel	Indian Restaurant	Italian Restaurant
6	Moste	0	Grocery Store	Café	Supermarket	Electronics Store	Department Store	Pharmacy	Restaurant	Pizza Place	Soccer Field	Music Venue
7	Polje	0	Soccer Field	Clothing Store	Italian Restaurant	Paintball Field	Music Venue	Mobile Phone Shop	Mexican Restaurant	Liquor Store	Athletics & Sports	Pharmacy
8	Posavje	0	Concert Hall	BBQ Joint	Athletics & Sports	Hardware Store	Historic Site	Hostel	Hotel	Indian Restaurant	Italian Restaurant	Liquor Store
9	Roznik	0	BBQ Joint	Trail	Restaurant	Park	Athletics & Sports	Italian Restaurant	Music Venue	Mobile Phone Shop	Mexican Restaurant	Liquor Store
12	Sentvid	0	Hotel	Coffee Shop	Bus Station	Hardware Store	Supermarket	Restaurant	Café	Paintball Field	Music Venue	Mobile Phone Shop
13	Siska	0	Convenience Store	Food & Drink Shop	Music Venue	Gym	Bakery	Bar	Hotel	Bike Rental / Bike Share	Pizza Place	Bus Station
15	Tmovo	0	Athletics & Sports	Shopping Mall	Hotel	Electronics Store	Convenience Store	Restaurant	Bus Station	Pizza Place	Soccer Field	Music Venue

### Cluster 2 – Purple Circle (Vic)

	District	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
16	Vic	1	Paintball Field	Athletics & Sports	Pharmacy	Gym / Fitness Center	Hardware Store	Historic Site	Hostel	Hotel	Indian Restaurant	Italian Restaurant

### Cluster 3 – Teal Circle (Rudnik)

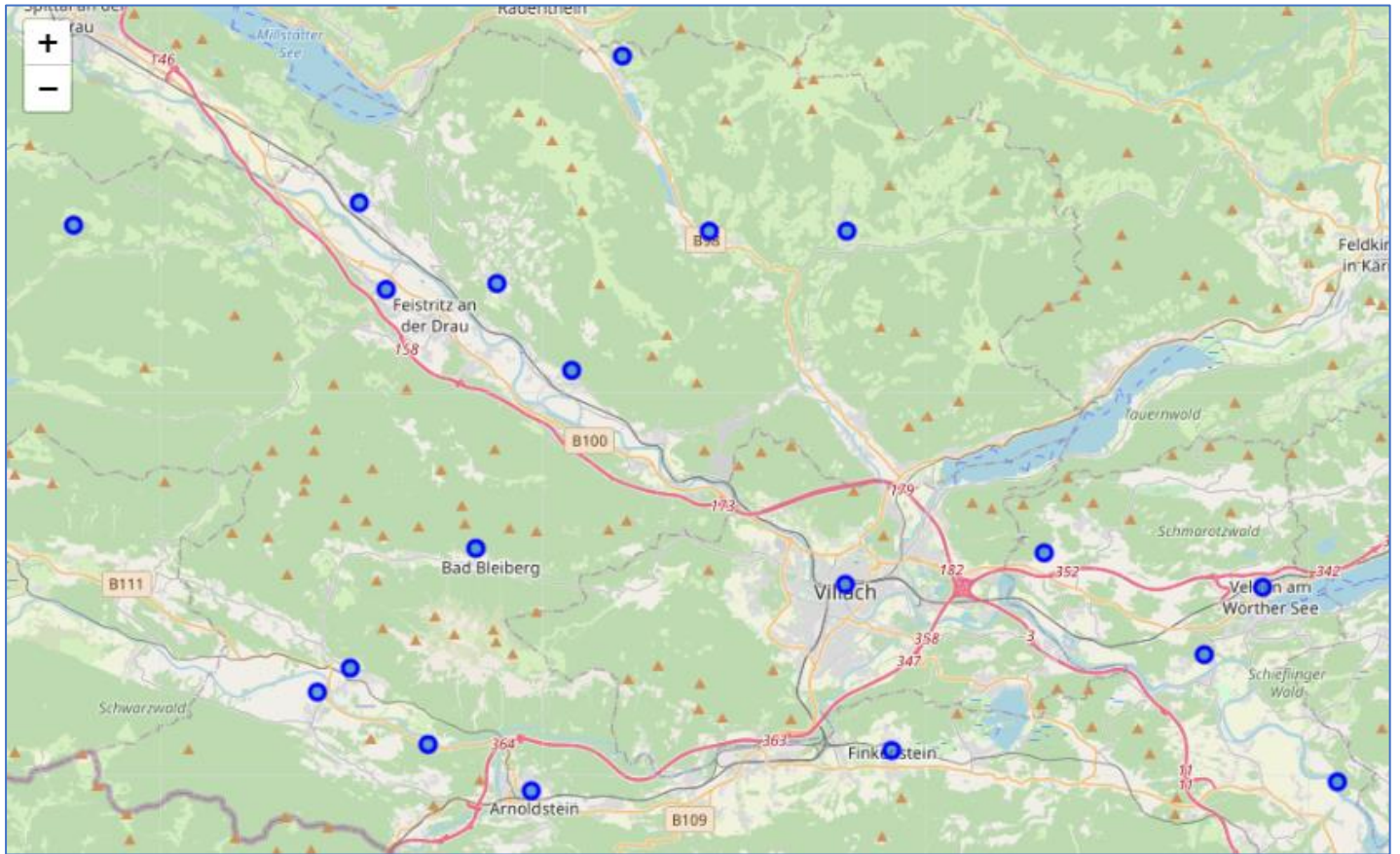
	District	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
10	Rudnik	2	Auto Garage	Grocery Store	Gym	Gym / Fitness Center	Hardware Store	Historic Site	Hostel	Hotel	Indian Restaurant	Italian Restaurant

The above tables will be reviewed later in this report, against the suitability criteria listed in the Business Problem (section 2). All remaining cities analysed in this project are shown below.



## 4.2.2 Villach Analysis

Villach location map



Villach shows a significant spread of districts across the countryside. Depending on the nature of the venues within each district and cluster, this may not be the best option for a Suburban/Urban lifestyle.

Villach has 19 districts. The below table shows a sample of the First 5 districts with the Top 10 venues per district:

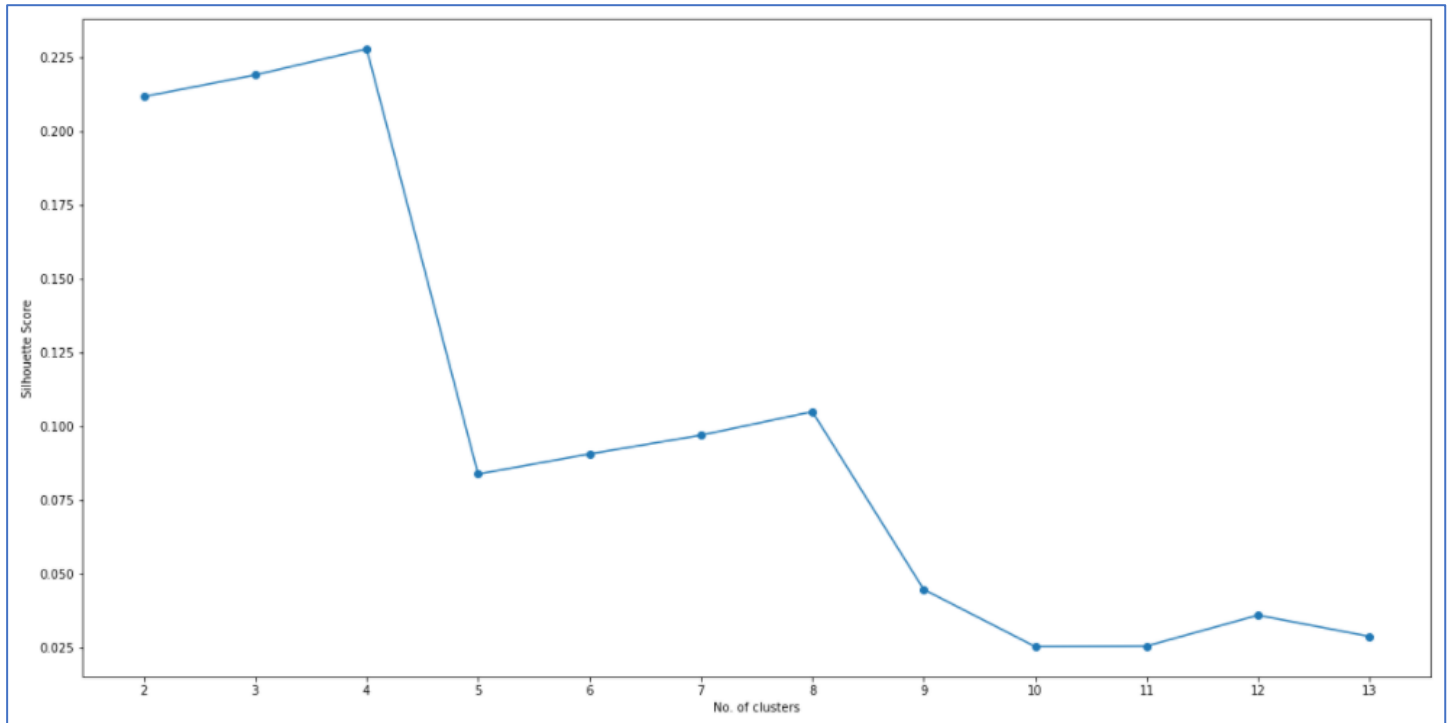
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Afritz am See	Supermarket	Eastern European Restaurant	Construction & Landscaping	Asian Restaurant	Other Great Outdoors	Insurance Office	Light Rail Station	Mediterranean Restaurant	Nightclub	Pizza Place
1	Arnoldstein	Train Station	Home Service	Restaurant	Supermarket	Stables	Ski Area	Shop & Service	Recording Studio	Pizza Place	Other Great Outdoors
2	Arriach	Health & Beauty Service	Austrian Restaurant	Café	Pizza Place	Insurance Office	Light Rail Station	Mediterranean Restaurant	Nightclub	Other Great Outdoors	Restaurant
3	Bad Bleiberg	Athletics & Sports	Cave	Ice Cream Shop	Insurance Office	Light Rail Station	Mediterranean Restaurant	Nightclub	Other Great Outdoors	Asian Restaurant	Hotel
4	Feistritz an der Gail	Supermarket	Department Store	Asian Restaurant	Other Great Outdoors	Ice Cream Shop	Insurance Office	Light Rail Station	Mediterranean Restaurant	Nightclub	Pizza Place

### 4.2.2.1 Villach Analysis: Clustering of Venues

Venues of similar type are clustered together using K-Means. Once the clusters have been calculated, they are plotted on the city map for easier analysis.

Silhouette method has been used to predict the optimal number of clusters for the dataset.

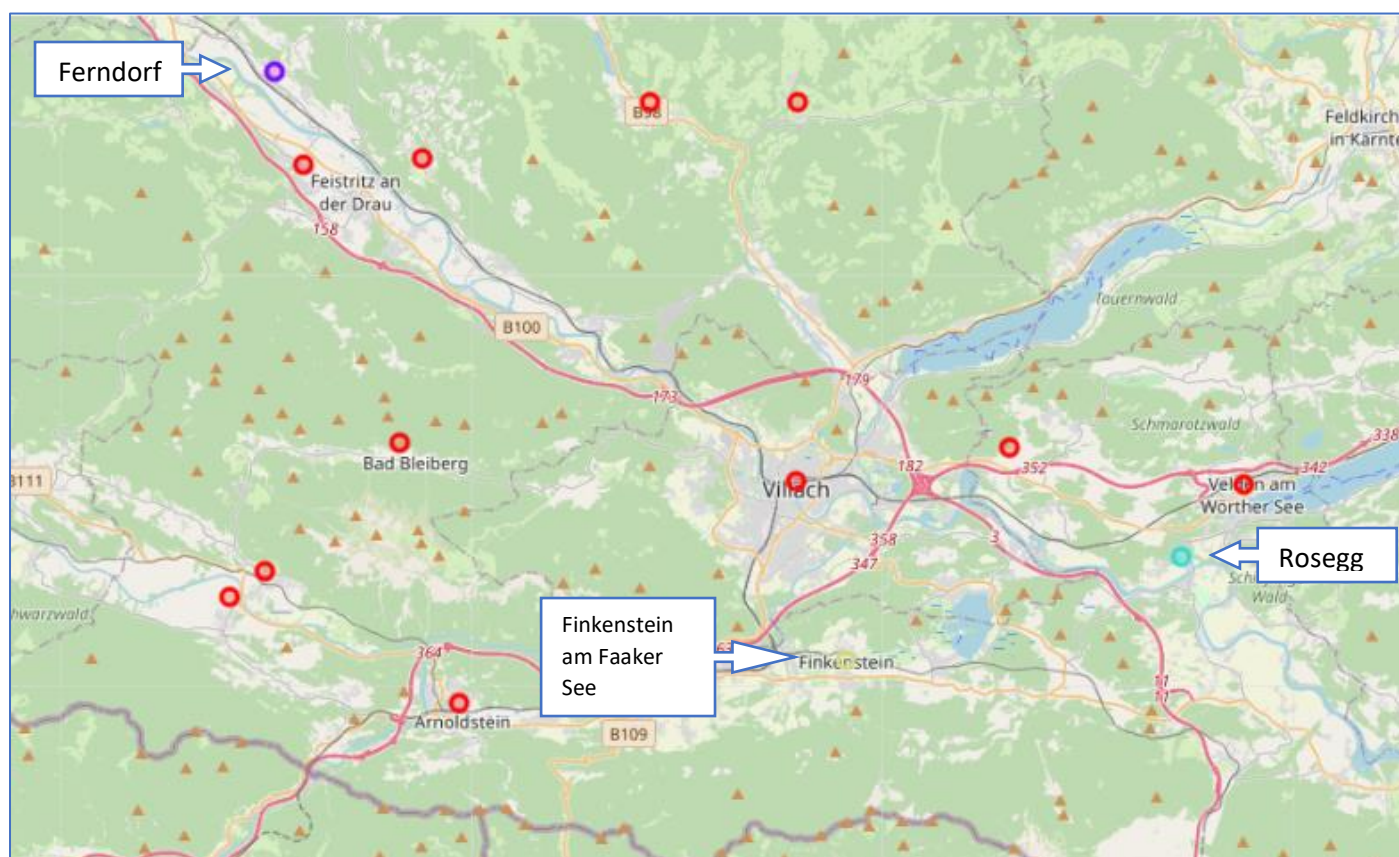
### 4.2.2.2 Villach Silhouette Score Plot



According to the results, the optimal number of clusters for Villach is 4. It is worth noting that at 4 clusters, the silhouette score is only 0.226, which is not a particularly high score by itself. However, when viewed in the context of the other number of clusters, it is proportionally much higher than the majority of clusters tested.

Below is the map plot of Villach with 4 clusters:

### 4.2.2.3 Villach Clustered Districts (K-Means)



The K-Means algorithm has grouped the majority of districts into one cluster. With three other districts isolated in their own cluster each (Ferndorf – purple/Finkenstein am Faaker See – yellow/Rosegg - teal).

Once the clusters had been identified and sorted, dataframes were created for each group including their Top 10 venues of interest. These tables can be used by stakeholders to evaluate the suitability of an area for living in.

Each table below represents a colour/group of districts.

### 4.2.2.4 Villach Venues by Clustered Districts

#### Cluster 1 – Red Circles

	District	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
17	Arnoldstein	0	Train Station	Home Service	Restaurant	Supermarket	Stables	Ski Area	Shop & Service	Recording Studio	Pizza Place	Other Great Outdoors
18	Bad Bleiberg	0	Athletics & Sports	Cave	Ice Cream Shop	Insurance Office	Light Rail Station	Mediterranean Restaurant	Nightclub	Other Great Outdoors	Asian Restaurant	Hotel
20	Notsch im Gailltal	0	Castle	Shop & Service	Restaurant	Pizza Place	Other Great Outdoors	Ice Cream Shop	Insurance Office	Light Rail Station	Mediterranean Restaurant	Nightclub
21	Paternion	0	Train Station	Insurance Office	Ski Area	Construction & Landscaping	Other Great Outdoors	Ice Cream Shop	Light Rail Station	Mediterranean Restaurant	Nightclub	Pizza Place
24	Meeting	0	Café	Hotel	Train Station	Greek Restaurant	Gastropub	Mediterranean Restaurant	Other Great Outdoors	Ice Cream Shop	Restaurant	Supermarket
25	Velden am Wörther See	0	Nightclub	Bar	Bistro	Café	Asian Restaurant	Casino	Restaurant	Pizza Place	Light Rail Station	Mediterranean Restaurant
26	Afritz am See	0	Supermarket	Eastern European Restaurant	Construction & Landscaping	Asian Restaurant	Other Great Outdoors	Insurance Office	Light Rail Station	Mediterranean Restaurant	Nightclub	Pizza Place
27	Arriach	0	Health & Beauty Service	Austrian Restaurant	Café	Pizza Place	Insurance Office	Light Rail Station	Mediterranean Restaurant	Nightclub	Other Great Outdoors	Restaurant
28	Feistritz an der Gail	0	Supermarket	Department Store	Asian Restaurant	Other Great Outdoors	Ice Cream Shop	Insurance Office	Light Rail Station	Mediterranean Restaurant	Nightclub	Pizza Place
31	Fresach	0	Restaurant	Asian Restaurant	Home Service	Supermarket	Stables	Ski Area	Shop & Service	Recording Studio	Pizza Place	Other Great Outdoors
35	Wernberg	0	Stables	Bed & Breakfast	Asian Restaurant	Other Great Outdoors	Ice Cream Shop	Insurance Office	Light Rail Station	Mediterranean Restaurant	Nightclub	Recording Studio

Once again, most venues are concentrated in the largest cluster. These venues will need to be reviewed against the key criteria of the stakeholders.

**Cluster 2 – Purple Circle (Ferndorf)**

	District	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
30	Ferndorf	1	Light Rail Station	Asian Restaurant	Home Service	Supermarket	Stables	Ski Area	Shop & Service	Restaurant	Recording Studio	Pizza Place

**Cluster 3 – Teal Circle (Rosegg)**

	District	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
22	Rosegg	2	Recording Studio	Asian Restaurant	Home Service	Supermarket	Stables	Ski Area	Shop & Service	Restaurant	Pizza Place	Other Great Outdoors

**Cluster 4 – Yellow Circle (Finkenstein am Faaker See)**

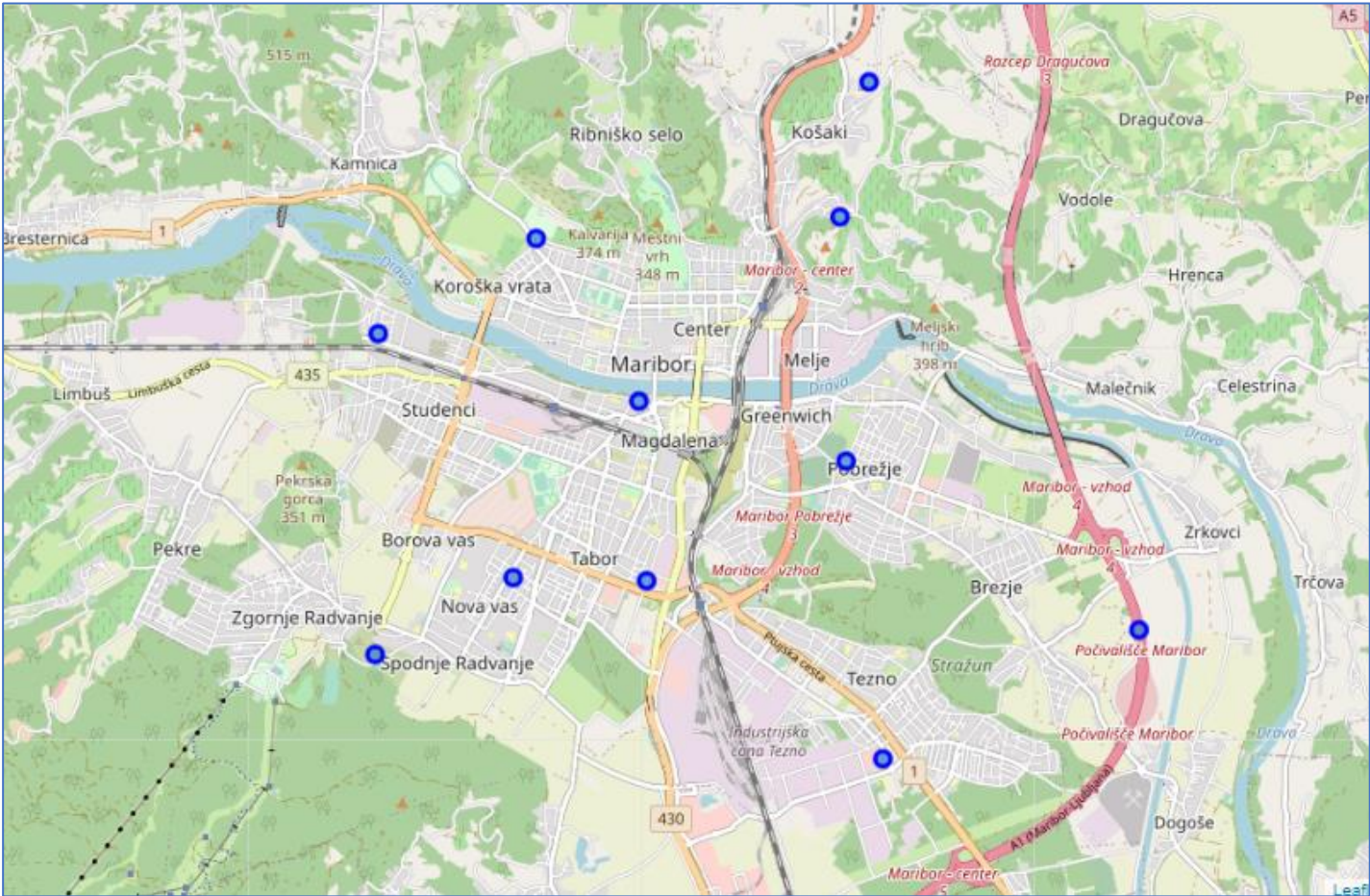
	District	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
19	Finkenstein am Faaker See	3	German Restaurant	Asian Restaurant	Other Great Outdoors	Ice Cream Shop	Insurance Office	Light Rail Station	Mediterranean Restaurant	Nightclub	Pizza Place	Home Service

The above tables will be reviewed later in this report, against the suitability criteria listed in the Business Problem (section 2).



4.2.3 Maribor Analysis

Maribor location map



The districts of Maribor are not too far from the city centre. It is likely that some of these areas will have many venues that are preferred by an expat family from Australia. It could also be true that no matter where a family chooses to live, they will not be far from a wide assortment of venues.

Maribor has 11 districts. The below table shows a sample of the First 5 districts with the Top 10 venues per district:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Brezje Dogose Zrkovci	Fast Food Restaurant	Asian Restaurant	Italian Restaurant	Mediterranean Restaurant	Mexican Restaurant	Museum	Nightclub	Paper / Office Supplies Store	Pizza Place	Pub
1	Center	Indian Restaurant	Bar	Jazz Club	Mediterranean Restaurant	Mexican Restaurant	Museum	Nightclub	Paper / Office Supplies Store	Pizza Place	Pub
2	Koroska Vrata	Café	Trail	Gym	Asian Restaurant	Restaurant	Mediterranean Restaurant	Mexican Restaurant	Museum	Nightclub	Paper / Office Supplies Store
3	Magdalena	Bar	Restaurant	Café	Cocktail Bar	Eastern European Restaurant	Pub	Pizza Place	Nightclub	Museum	Grocery Store
4	Nova Vas	Bar	Hotel	Paper / Office Supplies Store	Department Store	Scenic Lookout	Mediterranean Restaurant	Mexican Restaurant	Museum	Nightclub	Pizza Place

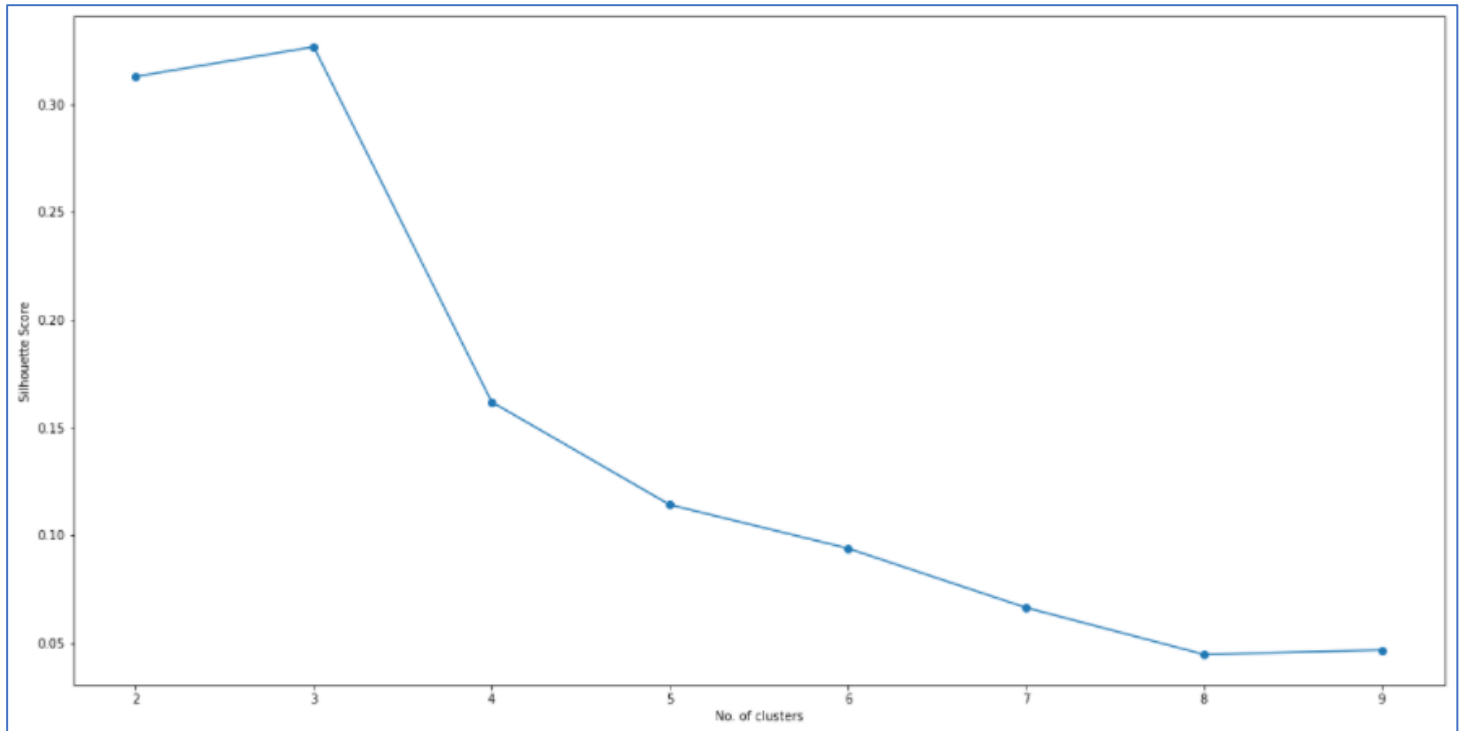


### 4.2.3.1 Maribor Analysis: Clustering of Venues

Venues of similar type are clustered together using K-Means. Once the clusters have been calculated, they are plotted on the city map for easier analysis.

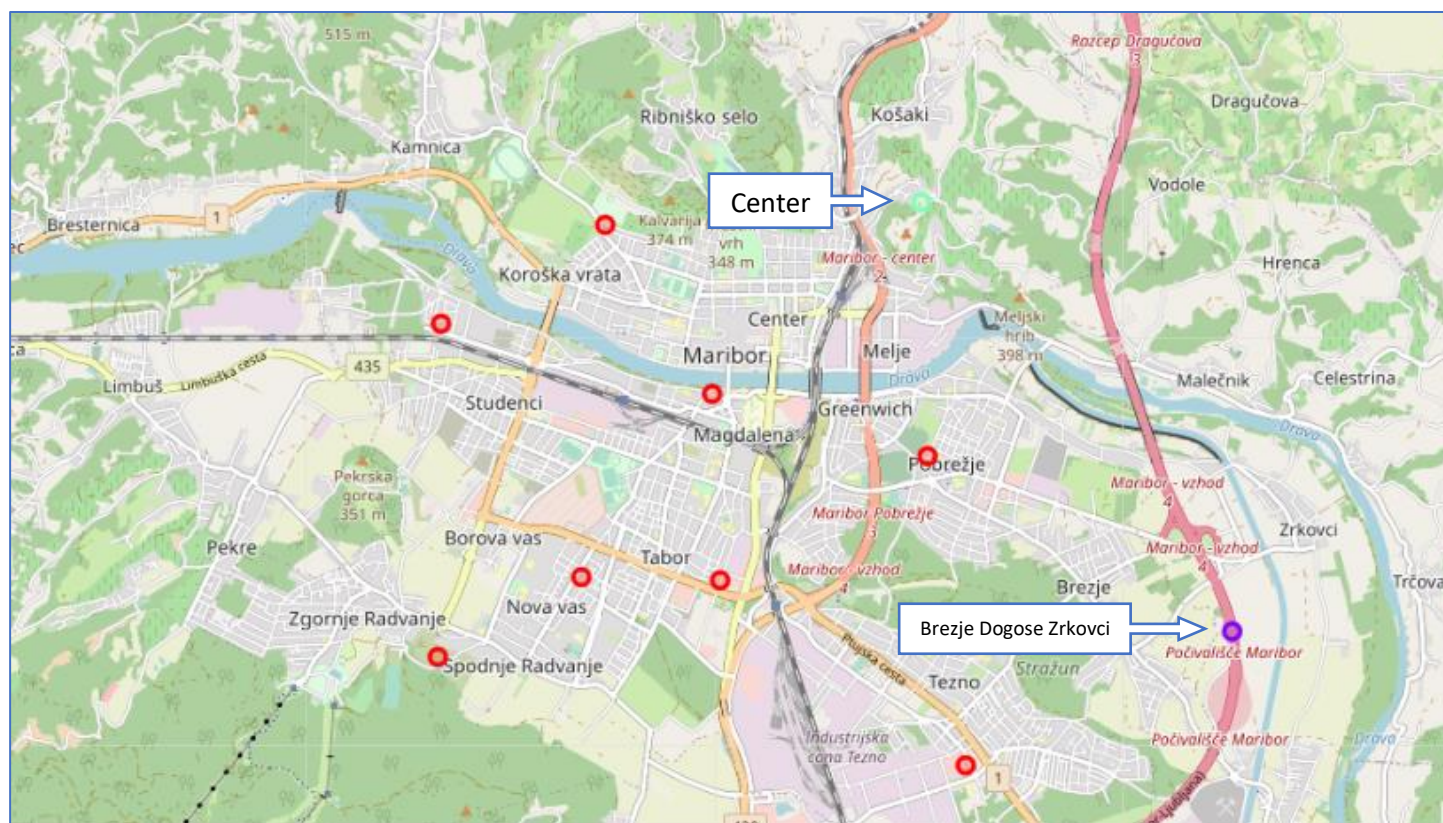
Silhouette method has been used to predict the optimal number of clusters for the dataset.

### 4.2.3.2 Maribor Silhouette Score Plot



According to the results, the optimal number of clusters for Maribor is 3. It is worth noting that at 3 clusters, the silhouette score is only 0.325, which is not a particularly high score by itself. However, when viewed in the context of the other number of clusters, it is proportionally much higher than the majority of clusters tested.

Below is the map plot of Maribor with 3 clusters:



As in previous examples, the K-Means algorithm has grouped the majority of districts into one cluster. With two other districts isolated in their own cluster each (Brezje Dogose Zrkovci – purple/Center - teal).

Once the clusters had been identified and sorted, dataframes were created for each group including their Top 10 venues of interest. These tables can be used by stakeholders to evaluate the suitability of an area for living in.

Each table below represents a colour/group of districts.

#### 4.2.3.4 Maribor Venues by Clustered Districts

### Cluster 1 – Red Circles

District	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Koroška Vrata	0	Café	Trail	Gym	Asian Restaurant	Restaurant	Mediterranean Restaurant	Mexican Restaurant	Museum	Nightclub	Paper / Office Supplies Store
Magdalena	0	Bar	Restaurant	Café	Cocktail Bar	Eastern European Restaurant	Pub	Pizza Place	Nightclub	Museum	Grocery Store
Nova Vas	0	Bar	Hotel	Paper / Office Supplies Store	Department Store	Scenic Lookout	Mediterranean Restaurant	Mexican Restaurant	Museum	Nightclub	Pizza Place
Pobrezje	0	Café	Gym	Nightclub	Pub	Restaurant	Asian Restaurant	Scenic Lookout	Mediterranean Restaurant	Mexican Restaurant	Museum
Radvanje	0	Café	Fried Chicken Joint	Asian Restaurant	Scenic Lookout	Mediterranean Restaurant	Mexican Restaurant	Museum	Nightclub	Paper / Office Supplies Store	Pizza Place
Studenci	0	Asian Restaurant	Theme Park	Hotel Pool	Vietnamese Restaurant	Pub	Jazz Club	Mediterranean Restaurant	Mexican Restaurant	Museum	Nightclub
Tabor	0	Shopping Mall	Bar	Sporting Goods Shop	Electronics Store	Eastern European Restaurant	Shoe Store	Fast Food Restaurant	Italian Restaurant	Chinese Restaurant	Café
Tezno	0	Asian Restaurant	Restaurant	Supermarket	Food & Drink Shop	Mediterranean Restaurant	Mexican Restaurant	Museum	Nightclub	Paper / Office Supplies Store	Pizza Place



### Cluster 2 – Purple Circle (Brezje Dogose Zrkovci)

District	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Brezje Dogose Zrkovci	1	Fast Food Restaurant	Asian Restaurant	Italian Restaurant	Mediterranean Restaurant	Mexican Restaurant	Museum	Nightclub	Paper / Office Supplies Store	Pizza Place	Pub

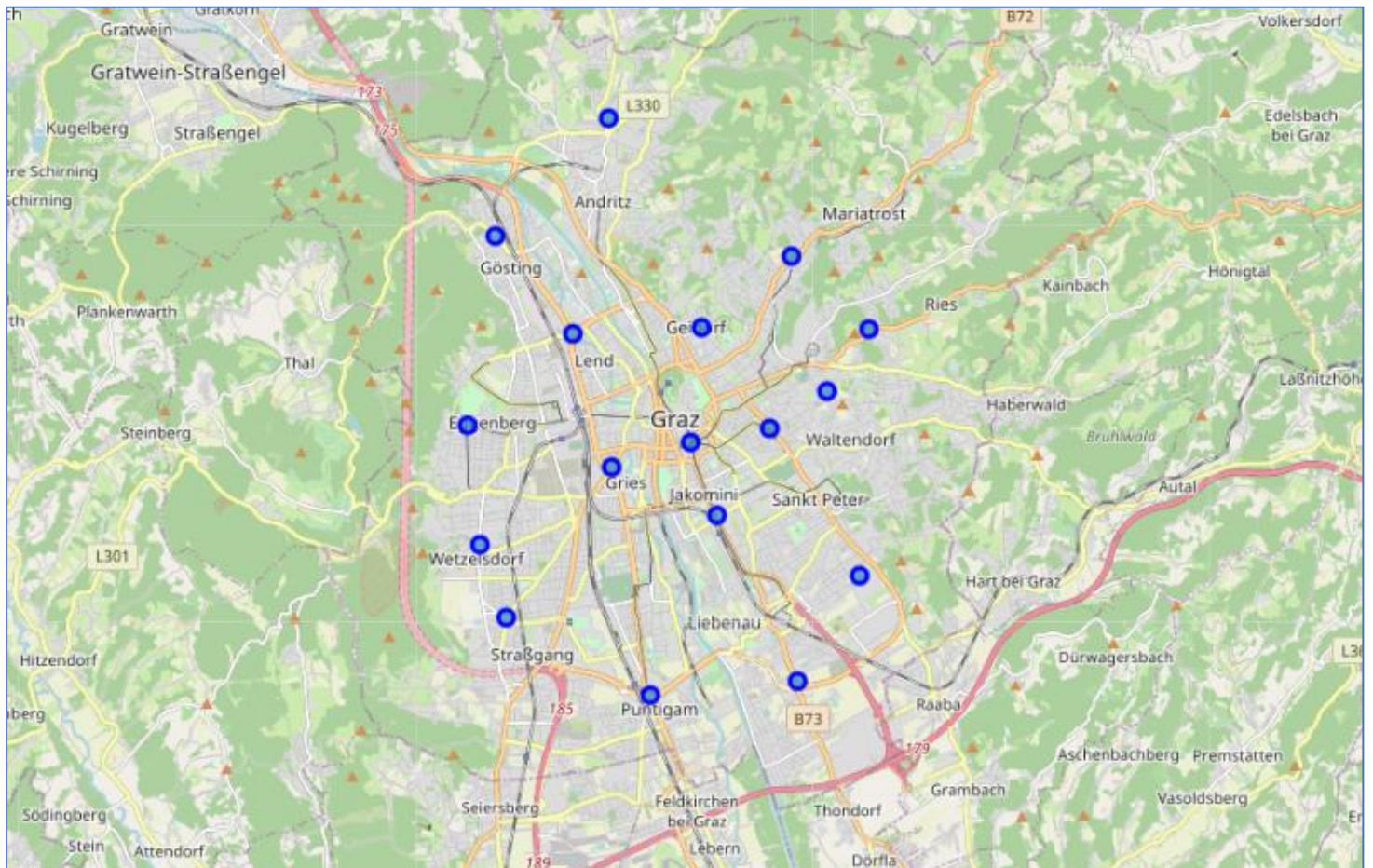
### Cluster 3 – Teal Circle (Center)

District	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Center	2	Indian Restaurant	Bar	Jazz Club	Mediterranean Restaurant	Mexican Restaurant	Museum	Nightclub	Paper / Office Supplies Store	Pizza Place	Pub

The above tables will be reviewed later in this report, against the suitability criteria listed in the Business Problem (section 2).

## 4.2.4 Graz Analysis

Graz location map



The districts of Graz are quite centralised around the city centre, with few districts spread into outlying areas. It is likely that these areas will have many venues that are preferred by an expat family from Australia. It could also be true that no matter where a family chooses to live, they will not be far from a wide assortment of venues.

Graz has 17 districts. The below table shows a sample of the First 5 districts with the Top 10 venues per district:

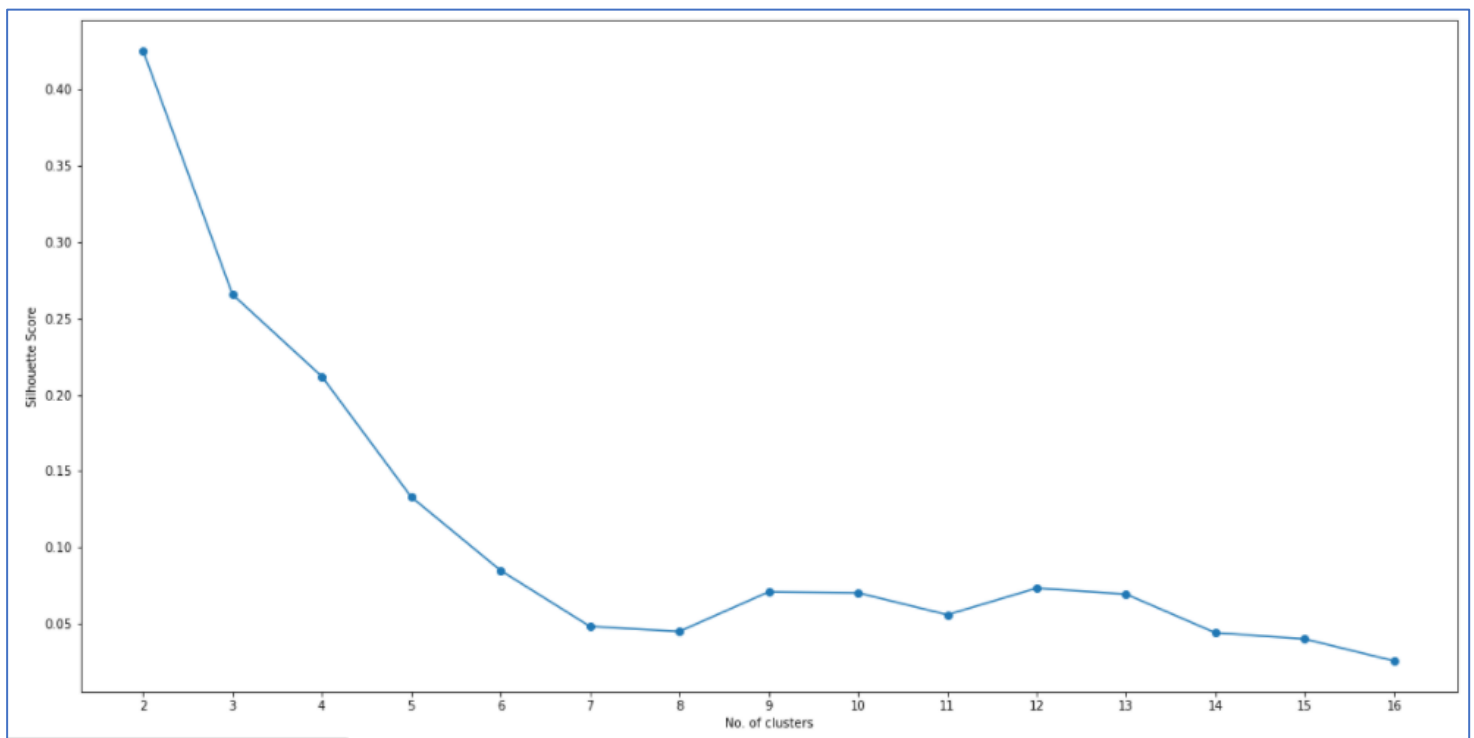
Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Andritz	Italian Restaurant	Café	Supermarket	Gastropub	Park	Pool Hall	Plaza	Pizza Place	Pharmacy	Peruvian Restaurant
Eggenberg	Park	History Museum	Café	Restaurant	Airport Service	Pool Hall	Plaza	Pizza Place	Pharmacy	Peruvian Restaurant
Geidorf	Italian Restaurant	Seafood Restaurant	Café	Supermarket	Park	Opera House	Plaza	Pizza Place	Pharmacy	Peruvian Restaurant
Gosting	Chinese Restaurant	Café	Supermarket	Airport Service	Park	Pool Hall	Plaza	Pizza Place	Pharmacy	Peruvian Restaurant
Gries	Café	Pizza Place	Clothing Store	Shoe Store	Climbing Gym	Mobile Phone Shop	Italian Restaurant	Art Gallery	Hotel	Hot Dog Joint

#### 4.2.4.1 Graz Analysis: Clustering of Venues

Venues of similar type are clustered together using K-Means. Once the clusters have been calculated, they are plotted on the city map for easier analysis.

Silhouette method has been used to predict the optimal number of clusters for the dataset.

#### 4.2.4.2 Graz Silhouette Score Plot



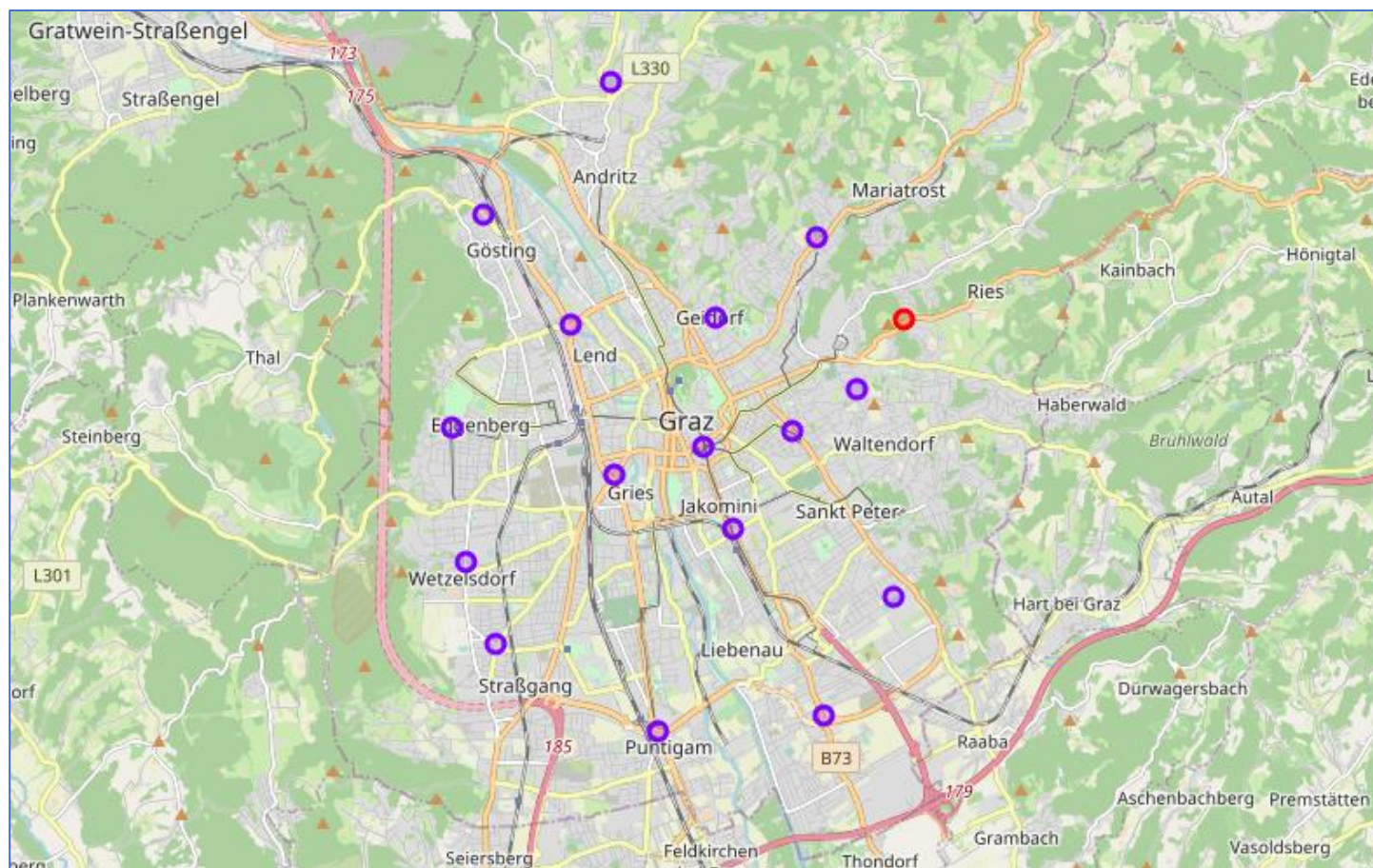
According to the results, the optimal number of clusters for Graz is 2, with a silhouette score is only 0.425, which is not a particularly high score by itself. However, when viewed in the context of the other number of clusters, it is proportionally much higher than the majority of clusters tested.

What is interesting, is that this dataset appears to contain venues that have greater similarity, and the clusters are better separated than all the other cities viewed so far.



For the Graz clustering exercise, silhouette scores of 3 or 4 could be used as these are similar scores to the other cities tested.

Below is the map plot of Graz with 2 clusters:



As in previous examples, the K-Means algorithm has grouped the majority of districts into one cluster. With one other district isolated in its own cluster (Ries – red).

This example clearly shows how tightly grouped these districts are within the highly developed areas of the city. Graz could be a good option for an urban lifestyle with easy access to a wide variety of attractions and venues.

Once the clusters had been identified and sorted, dataframes were created for each group including their Top 10 venues of interest. These tables can be used by stakeholders to evaluate the suitability of an area for living in.

Each table below represents a colour/group of districts.



## 4.2.4.4 Graz Venues by Clustered Districts

### Cluster 1 – Red Circle (Ries)

District	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Ries	0	Grocery Store	Airport Service	Opera House	Pool Hall	Plaza	Pizza Place	Pharmacy	Peruvian Restaurant	Park	Nightclub

### Cluster 2 – Purple Circles

District	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Innere Stadt	1	Austrian Restaurant	Plaza	Bar	Café	Greek Restaurant	Vegetarian / Vegan Restaurant	Pizza Place	Bakery	Steakhouse	Rock Club
St. Leonhard	1	Café	Light Rail Station	Supermarket	Plaza	Bakery	Gastropub	Basketball Court	Park	Airport Service	Pool Hall
Geidorf	1	Italian Restaurant	Seafood Restaurant	Café	Supermarket	Park	Opera House	Plaza	Pizza Place	Pharmacy	Peruvian Restaurant
Lend	1	Pizza Place	Supermarket	Airport Service	Opera House	Pool Hall	Plaza	Pharmacy	Peruvian Restaurant	Park	Nightclub
Gries	1	Café	Pizza Place	Clothing Store	Shoe Store	Climbing Gym	Mobile Phone Shop	Italian Restaurant	Art Gallery	Hotel	Hot Dog Joint
Jakomini	1	Hotel	Tram Station	Supermarket	Fast Food Restaurant	Event Space	Airport Service	Park	Plaza	Pizza Place	Pharmacy
Liebenau	1	Grocery Store	Café	Falafel Restaurant	Italian Restaurant	Movie Theater	Multiplex	Music Venue	Nightclub	Opera House	Pub

St. Peter	1	Soccer Field	Supermarket	Hotel	Italian Restaurant	Café	Airport Service	Plaza	Pizza Place	Pharmacy	Peruvian Restaurant
Waltendorf	1	Austrian Restaurant	Trattoria/Osteria	Pharmacy	Gastropub	Airport Service	Park	Pool Hall	Plaza	Pizza Place	Peruvian Restaurant
Mariatrost	1	Bus Station	Austrian Restaurant	Supermarket	Breakfast Spot	Airport Service	Pool Hall	Plaza	Pizza Place	Pharmacy	Peruvian Restaurant
Andritz	1	Italian Restaurant	Café	Supermarket	Gastropub	Park	Pool Hall	Plaza	Pizza Place	Pharmacy	Peruvian Restaurant
Gosting	1	Chinese Restaurant	Café	Supermarket	Airport Service	Park	Pool Hall	Plaza	Pizza Place	Pharmacy	Peruvian Restaurant
Eggenberg	1	Park	History Museum	Café	Restaurant	Airport Service	Pool Hall	Plaza	Pizza Place	Pharmacy	Peruvian Restaurant
Wetzelsdorf	1	Asian Restaurant	Market	Supermarket	Restaurant	Airport Service	Park	Pool Hall	Plaza	Pizza Place	Pharmacy
Strasgang	1	Airport Service	Austrian Restaurant	Bar	Park	Pub	Pool Hall	Plaza	Pizza Place	Pharmacy	Peruvian Restaurant
Puntigam	1	Bar	Brewery	Ice Cream Shop	Café	Supermarket	Sandwich Place	Rock Club	Restaurant	Pool Hall	Multiplex

## 5. Results of Slovenia & Austria City Analysis

With the analysis of all four cities complete, it is time to make comparisons against the key criteria of stakeholders.

However, the analysis criteria suggests that the choice of city is primarily going to be based on which location is chosen for study. If the preferred university is in Ljubljana, then the choice will be between Ljubljana and Villach. However, if Maribor University is preferred, then the cities to compare will be Maribor and Graz.

The best way to view the results and determine which will be the best location, is to review one pair of cities at a time. Each cluster within each city will be compared against the primary venue types preferred by the stakeholders.

In each city, a primary cluster will be identified as the optimal location for ex-pats to reside in. Then this will be compared against the top contender for the corresponding city. Each cluster will be ranked by the number of key venues identified in the group.

The key venues of interest are:

1. Gym or fitness centre
2. Cafes
3. Restaurants
4. Supermarkets
5. Parks
6. Cinemas
7. Train Station
8. Theatre
9. Shopping facilities
10. Entertainment Venues

A visual analysis of each cluster was conducted using the high-level categories shown above. Note that the categories above are not specific venue categories found in FourSquare. Rather they are broad groupings of amenities that the stakeholders are seeking in their preferred area or district.

The results are listed below in each table.

### 5.1.1 Ljubljana Clusters

CLUSTER	TYPE/DISTRICT	MATCHING VENUES
1	Red Circles	87
2	Purple Circle (Vic)	6
3	Teal Circle (Rudnik)	6

It is clear that **cluster 1** would be the preferred area to settle in from a lifestyle point of view. A new resident seeking an area to live in, who is also looking for the venue types listed above, should look within these districts of Ljubljana.

### 5.1.2 Villach Clusters

CLUSTER	TYPE/DISTRICT	MATCHING VENUES
1	Red Circles	68
2	Purple Circle (Ferndorf)	6
3	Teal Circle (Rosegg)	6
4	Teal Circle (Rudnik)	8

It is clear that **cluster 1** would be the preferred area to settle in from a lifestyle point of view. A new resident seeking an area to live in, who is also looking for the venue types listed above, should look within these districts of Villach.

### 5.1.3 Ljubljana Primary Cluster vs Villach Primary Cluster

On the basis of the above data, it would appear that **cluster 1 in Ljubljana** offers significantly more options that match with the key criteria of our stakeholders.

Therefore, should the University of Ljubljana be chosen as the preferred university it would be better to reside within Ljubljana to gain access to a wider variety of the key amenities.

### 5.2.1 Maribor Clusters

CLUSTER	TYPE/DISTRICT	MATCHING VENUES
1	Red Circles	62
2	Purple Circle (Brezje Dogose Zrkovci)	9
3	Teal Circle (Center)	8

It is clear that **cluster 1** would be the preferred area to settle in from a lifestyle point of view. A new resident seeking an area to live in, who is also looking for the venue types listed above, should look within these districts of Maribor.

### 5.2.2 Graz Clusters

CLUSTER	TYPE/DISTRICT	MATCHING VENUES
1	Red Circle (Ries)	9
2	Purple Circles	136

It is clear that **cluster 2** would be the preferred area to settle in from a lifestyle point of view. A new resident seeking an area to live in, who is also looking for the venue types listed above, should look within these districts of Graz.

### 5.2.3 Maribor Primary Cluster vs Graz Primary Cluster

On the basis of the above data, it would appear that **cluster 2 in Graz** offers significantly more options that match with the key criteria of our stakeholders.

Therefore, should the University of Maribor be chosen as the preferred university it would be better to reside within Graz to gain access to a wider variety of the key amenities.

## 6. Discussion

The preceding analysis is a fairly simplistic method for selecting a location to settle in. While I appreciate that there are many other factors to choosing the best possible district in these four cities, it does provide some guidance to those who may want to engage in a migration for educational reasons.

Without a method of targeting the districts, moving to these cities would be a daunting prospect for Australian residents with little experience of Slovenia or Austria. That is where this analysis can be useful to direct the focus on a reduced number of locations that provide the lifestyle and local amenities that the stakeholders prefer.

It was interesting to note that the primary K-Means groupings were quite large, with a few outliers in each example. What this implies is that the most common venues were largely restricted to centralised areas. Which is what I did expect to some degree.

This does prove that the best place to access most of the required amenities is closer to the centre of each city.

Interestingly, the best location for families who have students in Ljubljana was to stay in Ljubljana. While the optimal location for those that have students in Maribor is Graz, Austria. Also noteworthy is that Graz provides the widest range out of all the cities of appropriate amenities for our stakeholders. This is likely due to Graz being the city with the largest population by a significant margin.

In future, this analysis would benefit from some additional refinements.

First, it would be useful to overlay the average price of rental/sale properties in each district as well. This would further target the districts for stakeholders. They would then be able to choose specific districts that are within their price range and that are nearby their preferred venue types.

Second, refining the FourSquare venues down to the types of amenities listed in the stakeholder's primary list would also assist in providing a better result. These venue types could be developed into a table that groups the venue types from FourSquare into the broad preferred categories. A table like this can be merged with the venue data and used to limit the results specifically to stakeholder-required amenities. The resulting table could be used in the K-Means dataframe instead and may provide quite different clustering results.

## **7. Conclusion**

With the ever-increasing cost of tertiary education in Australia, families seeking to find other means of educating either children (or themselves), may find this kind of analysis useful to assist in their relocation.

So, for Australians able to take advantage of cost-effective education in Slovenia, then choosing to live in either Ljubljana or Graz would appear to provide the best lifestyle options.

Although there are many other factors that may influence exactly where they would want to live, this kind of data can provide a good platform for reducing any potential for error in choosing an optimal location to stay for a considerable period of time.