Capstone Project - The Battle of The Cities

Agenda:

A. **Introduction**
B. Data Description
C. Methodology
D. Results
E. Discussion & Conclusion

Introduction

Scenario:

You just started to work as a Junior Data Analyst at a travel agency. Since Italy will be a very sought travel destination after COVID-19, your manager asks you to propose a weekend travel itinerary for busy people.

Ideally, the itinerary should pack different experiences or cities for a weekend trip. Therefore, these cities should be quite different (from a venue's point of view) and quite close from a geographical point of view.

Agenda:

A. Introduction
B. Data Description
C. Methodology
D. Results
E. Discussion & Conclusion

Data Description

We will use the following data sources:

- List of Italian provinces and Regions. The following Wikipedia page was scraped to pull out the necessary information: https://en.wikipedia.org/wiki/List_of_postal_codes_in_Italy. The information obtained i.e. the table of postal codes was transformed into a Pandas DataFrame for further analysis.

- Coordinates data for each capital city in each region. The following csv lists the geographical coordinates of each city: https://simplemaps.com/static/data/country-cities/it/it.csv

Agenda:

A. Introduction
B. Data Description
C. Methodology
D. Results
E. Discussion & Conclusion

Methodology

Steps:

1. Importing data and creating a Pandas DataFrames from the two data sources.

	Province	Code	Region	lat	lng
0	Rome	RM	Lazio	41.900000	12.483333
1	Viterbo	VT	Lazio	42.416667	12.100000
2	Rieti	RI	Lazio	42.400000	12.850000
3	Frosinone	FR	Lazio	41.633333	13.316667
4	Latina	LT	Lazio	41.466667	12.866667

2. Generating a map of Italy and moving towards Veneto

3. Utilizing Foursquare API to explore the Provinces in Veneto

	Province	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	Belluno	Soccer Stadium	Wine Bar	Bar	Fried Chicken Joint	Hotel	Supermarket	Italian Restaurant	Restaurant	Japanese Restaurant	Dessert Shop
1	Padua	Platform	Sushi Restaurant	Supermarket	Italian Restaurant	Hotel	Breakfast Spot	Plaza	Seafood Restaurant	Light Rail Station	Juice Bar
2	Rovigo	Pizza Place	Pub	Italian Restaurant	Park	Shopping Mall	Design Studio	Dessert Shop	Soccer Stadium	Plaza	Diner
3	Treviso	Café	Italian Restaurant	Plaza	Wine Bar	Bar	Pizza Place	Ice Cream Shop	Trattoria/Osteria	Clothing Store	Winery
4	Venice	Italian Restaurant	Hotel	Wine Bar	Plaza	Café	Art Museum	Restaurant	Bar	Bed & Breakfast	Gastropub
5	Verona	Italian Restaurant	Café	Ice Cream Shop	Restaurant	Cheese Shop	Castle	Museum	Campground	Scenic Lookout	Snack Place
6	Vicenza	Café	Italian Restaurant	Plaza	Bar	Art Museum	Ice Cream Shop	Pub	Restaurant	Sandwich Place	Wine Bar

Agenda:

A. Introduction
B. Data Description
C. Methodology
D. Results
E. Discussion & Conclusion

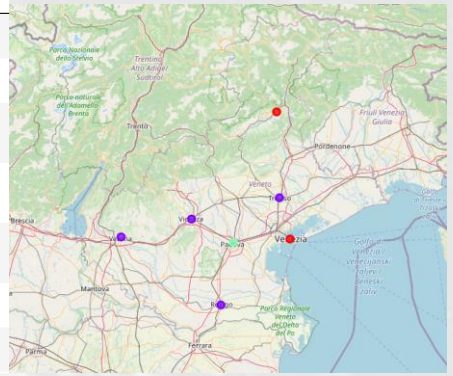
Results

Scenario:

The below table depicts the clustered data along with the top 10 most common venues in that cluster. The 10th column is not visible in the table.

We use the matplotlib and folium packages to visualize the clusters on the Veneto map.

	Province	lat	lng	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
0	Venice	45.438611	12.326667	0	Italian Restaurant	Hotel	Wine Bar	Plaza	Café	Restaurant	Art Museum	Bed & Breakfast	Bar
1	Treviso	45.666667	12.245000	1	Café	Italian Restaurant	Plaza	Wine Bar	Bar	Pizza Place	Ice Cream Shop	Restaurant	Trattoria/Osteria
2	Belluno	46.145000	12.221389	0	Soccer Stadium	Fried Chicken Joint	Bar	Wine Bar	Hotel	Supermarket	Italian Restaurant	Japanese Restaurant	Design Studio
3	Padua	45.416667	11.883333	2	Light Rail Station	Sushi Restaurant	Supermarket	Hotel	Breakfast Spot	Boat or Ferry	Gift Shop	Platform	Plaza
4	Vicenza	45.550000	11.550000	1	Café	Italian Restaurant	Plaza	Bar	Art Museum	Wine Bar	Pub	Restaurant	Sandwich Place
5	Verona	45.450000	11.000000	1	Italian Restaurant	Café	Ice Cream Shop	Restaurant	Soccer Field	Cheese Shop	Castle	Martial Arts Dojo	River
6	Rovigo	45.066667	11.783333	1	Pizza Place	Soccer Stadium	Italian Restaurant	Café	Park	Plaza	Design Studio	Pub	Dessert Shop



Agenda:

A. Introduction
B. Data Description
C. Methodology
D. Results
E. **Discussion & Conclusion**

Discussion & Conclusion

Discussion:

We see that cities/Provinces are clustered as follow:

- Cluster 0: Venice, Belluno
- Cluster 1: Verona, Vicenza, Treviso, Rovigo
- Cluster 2: Padua

1. We want to visit a city for each cluster. Since Cluster 2 has only one city, Padua, this could be a city we want to visit.
2. In cluster 0 there are two cities, Venice and Belluno. However, Venice is much closer to Padua, therefore Venice becomes the second city on our weekend trip.
3. In cluster 1, we have four cities, but Treviso is the nearest to both Padua and Venice, therefore Treviso is going to be the third city on our weekend trip.

Conclusion:

Despite the outcome, using a different k_cluster value can show a slightly different result. As seen in the example above, data was used to cluster cities in Veneto based on the most common venues in those cities. Similarly, it could be interesting to compare clusters in different regions or countries.