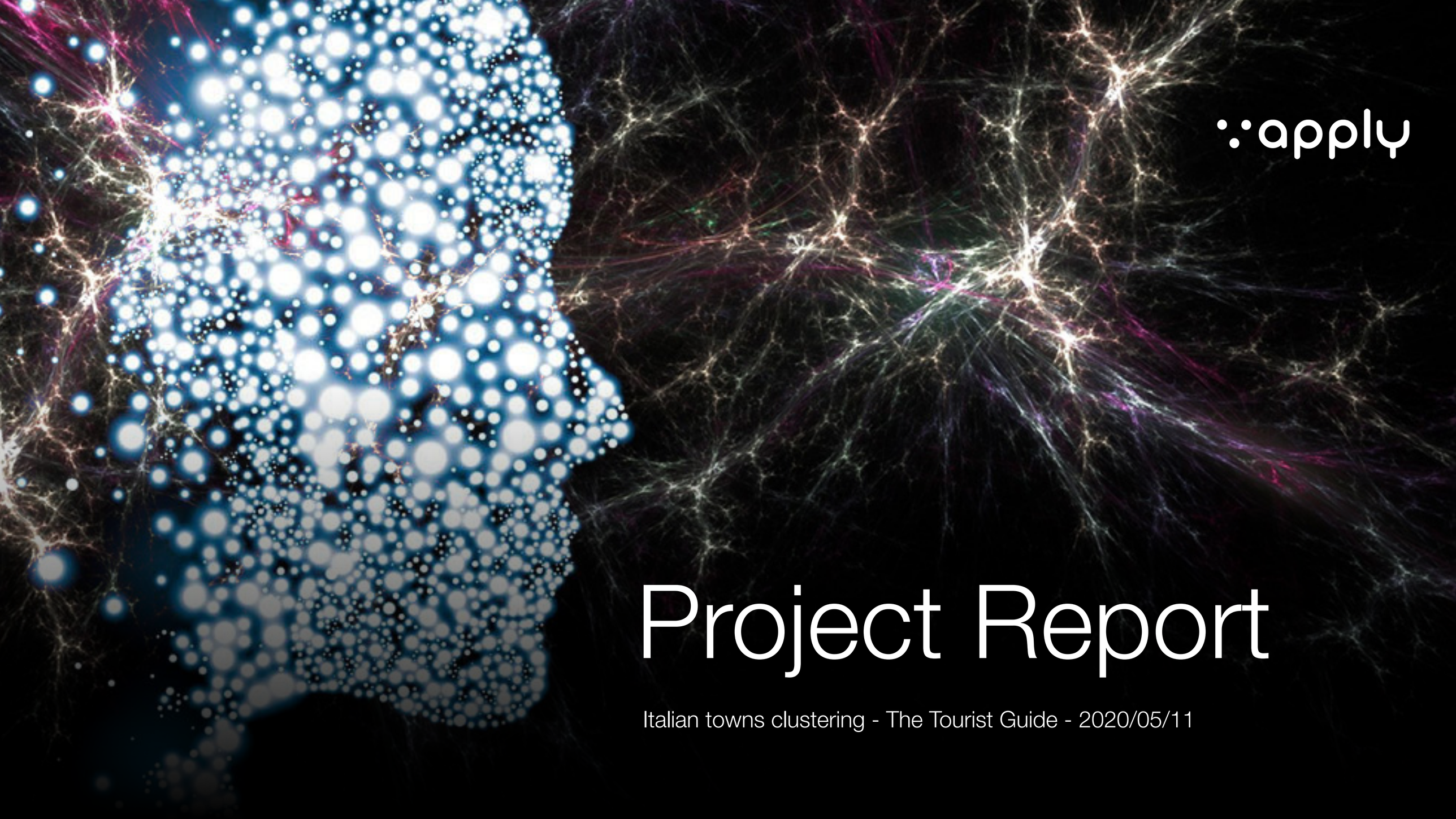
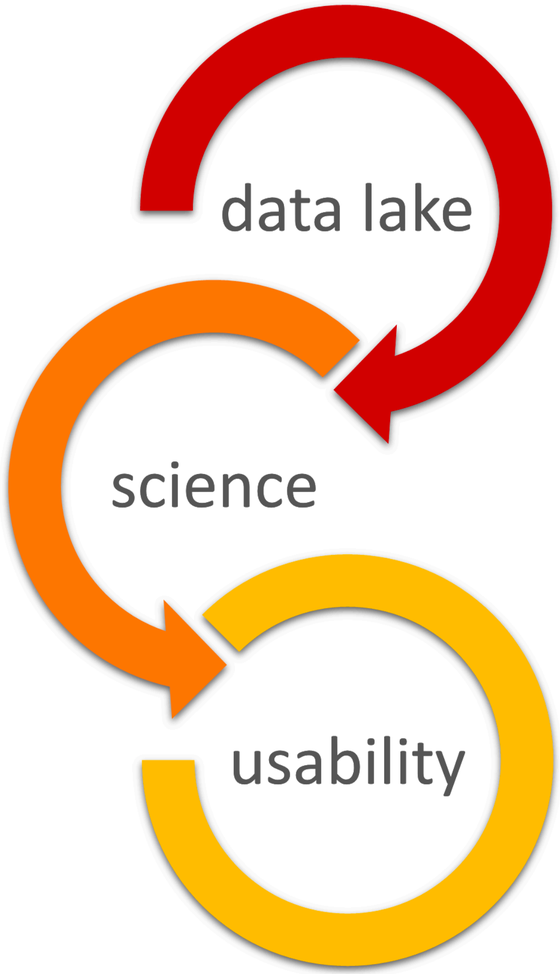


apply

Project Report

Italian towns clustering - The Tourist Guide - 2020/05/11

Table of Contents



EXECUTIVE SUMMARY

Executive summary

A tour operator would like to have a data driven approach when suggesting destinations in Italy.

Specifically, given the incredible variability of Italy, they asked for some kind of clustering approach, that might be able to suggest similar or different locations.

Hopefully, an added value would be given by some new insights provided by data analysis.

The problem is studied on the basis of data from Foursquare. The approach is based on the assumption that even if Italy is a mix of many traditions and cultures, there exists a common denominator defined by food and wine culture.

The results of this analysis show that it is possible to categorise interesting features of Italian towns on the basis of the food and wine venues that they offer.

INTRODUCTION

Introduction

Italy is one of the main destinations for tourism.

To a tourist, Italy offers many different landscapes, feelings, traditions and cultural backgrounds, because of its geography and because it is a united country only since 1861.

Our **stakeholders** manage an important commercial tourism portal.

The problem they want us to solve is: *find a data driven method to categorise Italian towns, so that it is possible to suggest similar towns or propose different choices to clients*.

In order to solve the problem, we will focus on the analysis of food and wine offerings in different towns, because we know that the approach to food and wine is probably the main feature that describes different locations in Italy.

Introduction

For example, towns where tourism is at maximum levels, use to loose some of their culinary identity, in order to satisfy a wider public. These towns have a stronger *international* taste and feel.

So, the idea is to use food and wine offering to cluster towns.

We will try to use clustering results to identify groups of towns that has a different spirit, so that the tourist can get a suggestion about similar places and types of destinations.

Overall, as clustering is unsupervised, sometimes it can help to find out interesting descriptions and interpretations that highlight specific features. This is a much desired result for this work.

DATA SOURCE

apply

Data Source

Data used in this project are taken from Foursquare.

Basically, for each of the towns involved in the analysis, we will query Foursquare, looking for the first (maximum) 100 venues, with search keywords set to “restaurant” and “wine”.

The search is centred around the centre of the town, basically the city centre or downtown, the core of the ancient town. From this point we open a radius equal to 1000, so that all the very core of the town is covered.

What does a town offers in her “heart”? Traditional food or something more heterogeneous?

Data Source

The table on the right comes with the towns analysed in this project.

They cover Italy north to south.

The numbers in columns “restaurant” and “wine” are the number of such kind of venues found in a circle with radius 1000, centred in the city ancient centre.

We will use the categories of these venues, to cluster the towns explored in the analysis.

town	restaurant	wine
Amalfi	49	11
Bari	70	16
Bologna	100	100
Catania	80	35
Milan	100	100
Naples	100	26
Padua	46	26
Pisa	100	40
Rome	93	13
Trento	99	11
Trieste	100	26
Udine	95	16
Venice	100	100
Verona	100	69
florence	100	100

DATA PREPARATION

apply

Data Preparation

Dummy variables are computed in order to have mean values of occurrences for each category.

	town	American Restaurant	Argentinian Restaurant	Asian Restaurant	BBQ Joint	Bakery	Bar	Bistro	Brazilian Restaurant	Breakfast Spot	Buffet	Burger Joint	Butcher	Cafeteria	Café
0	Amalfi	0.000	0.000	0.000000	0.000000	0.016667	0.000000	0.016667	0.000000	0.000000	0.00000	0.000000	0.000	0.000000	0.100000
1	Bari	0.000	0.000	0.000000	0.000000	0.023256	0.011628	0.023256	0.000000	0.000000	0.00000	0.011628	0.000	0.011628	0.197674
2	Bologna	0.000	0.005	0.000000	0.000000	0.010000	0.005000	0.020000	0.000000	0.000000	0.00000	0.010000	0.005	0.000000	0.100000
3	Catania	0.000	0.000	0.000000	0.008696	0.000000	0.000000	0.017391	0.000000	0.008696	0.00000	0.043478	0.000	0.000000	0.086957
4	Milan	0.005	0.000	0.010000	0.000000	0.030000	0.020000	0.015000	0.000000	0.000000	0.00000	0.005000	0.000	0.000000	0.090000
5	Naples	0.000	0.000	0.007937	0.000000	0.000000	0.007937	0.007937	0.000000	0.000000	0.00000	0.000000	0.000	0.007937	0.079365
6	Padua	0.000	0.000	0.000000	0.000000	0.041667	0.013889	0.000000	0.000000	0.013889	0.00000	0.027778	0.000	0.000000	0.152778
7	Pisa	0.000	0.000	0.000000	0.000000	0.000000	0.007143	0.000000	0.000000	0.000000	0.00000	0.000000	0.000	0.000000	0.128571
8	Rome	0.000	0.000	0.000000	0.000000	0.000000	0.000000	0.009434	0.000000	0.000000	0.00000	0.047170	0.000	0.000000	0.226415
9	Trento	0.000	0.000	0.009091	0.000000	0.027273	0.000000	0.018182	0.000000	0.009091	0.00000	0.009091	0.000	0.000000	0.254545
10	Trieste	0.000	0.000	0.007937	0.000000	0.015873	0.007937	0.023810	0.000000	0.000000	0.02381	0.007937	0.000	0.000000	0.198413
11	Udine	0.000	0.000	0.009009	0.000000	0.036036	0.000000	0.009009	0.000000	0.009009	0.00000	0.009009	0.000	0.000000	0.189189
12	Venice	0.005	0.000	0.000000	0.000000	0.020000	0.020000	0.005000	0.000000	0.000000	0.00000	0.000000	0.000	0.000000	0.065000
13	Verona	0.000	0.000	0.000000	0.000000	0.000000	0.017751	0.005917	0.005917	0.005917	0.00000	0.011834	0.000	0.000000	0.082840
14	florence	0.015	0.000	0.000000	0.000000	0.025000	0.010000	0.015000	0.000000	0.000000	0.00000	0.005000	0.000	0.000000	0.050000

These columns are just a few sample of the many categories we got from Foursquare.

METHODOLOGY

Methodology

We will use K Means clustering algorithm from Sklearn.

The number of clusters is set to three because the analyst thinks that 3 distinct clusters could be interpreted according to a previous hypothesis.

Yet, other numerosities were studied, with 2 or 4 clusters and 3 confirmed to be most reasonable. 2 was not enough and 4 was confirming the results seen with 3, plus adding some noise that was not easily interpretable.

RESULTS

Results

Clustering against 3 clusters has returned an interesting result.

As usual, clusters have to be understood on the basis of domain knowledge, in order to give a significative meaning to each cluster.

These three clusters have a really interesting explanation, as discussed in the following Discussion section.

	town	Cluster Labels
0	Amalfi	1
1	Bari	0
2	Bologna	1
3	Catania	0
4	Milan	1
5	Naples	2
6	Padua	0
7	Pisa	2
8	Rome	0
9	Trento	0
10	Trieste	0
11	Udine	0
12	Venice	1
13	Verona	1
14	florence	1

apply

DISCUSSION

Discussion

Cluster 1 : Very touristic and international

All of the towns in cluster 1 are very common destinations for turists. These towns, due to their turistic relevance, have lost some of their traditional identity, they have acquired many international features. So, even if you can find *also* traditional restaurants and wines, you can easily find food that is not original from that specific place. For an Italian changing food traditions is really a relevant step. So we can claim that these towns, known for tourism, have also acquired many international features, which do not belong to the place itself and this is well seen from restaurant and wine venues.

Discussion

Cluster 2 : Very touristic and traditional

Cluster 2 is made up by Pisa and Napoli, two towns that are very well known by turists. Yet these two towns are very bound to their traditions and thus are less international. In these places you will find many turists, but local traditions are well kept and preserved.

Discussion

Cluster 0 : Beautiful and very traditional

In this cluster all of the towns, with the important exception of Rome, are not very known by turists. These towns are all extremely beautiful and rich with history and art. For these towns, being out of the main turistic trips, is the key to preserve their traditions. Here trditions are **very** well kept and preserved. Rome is an interesting case: Rome is one of the main destinations in the world, but her traditions are so strong that they are almost untouched and surely very well kept and preserved, like in the other towns of this cluster.

CONCLUSION

Conclusion

As it was supposed, it is possible to label Italian towns on the basis of the food venues that they offer.

This is made possible by the Italian culture for food and wine. When an area changes the food traditions, it means that something is changing. Such a change we have seen that, on the basis of domain experience, can be related to the level of globalisation. Globalisation in Italy is often brought by turism, in order to satisfy turistis desires.

For example, the famous lasagne are from Bologna, you cannot expect to eat lasagne in Trento, but you will find lasagne both in Florence and Venice.

But now you know that before eating something in italy you should look for traditional recipes of the specific town and look only for those restaurants that offer only traditional food of the specific place. The other restaurants are for tourists....!