

# Applied Data Science – Final Project:

## A new restaurant in Florence

### 1. Introduction

A famous chain restaurant has decided to open a new diner in Florence. In order to get the maximum income, they decided to involve us (their Data Analysis team) to locate the best area where to open a restaurant. To do so, they located some specific area for each district and gave us specific instructions:

- the main customers are families and people between the age of 12 and 45, so this information must be taken into account when choosing the area;
- it would be better to choose an area where the number of restaurant per person is lower than the others;
- a close bus/train station is valuable.

### 2. Data

In order to execute the task, we are going to exploit the data offered by the Foursquare service; however, other data sources will be involved to fulfil the requests. Particularly, we will use:

- **District location data:** retrieved via Google Maps; those are the information we need to fetch location data through Foursquare api.
- **Population by District and Age data** from the web source [www.opendata.comune.fi.it](http://www.opendata.comune.fi.it); this dataset contains information about the number of people in each district divided by age classes.
- **Family Number by District and N° of Children data** from the web source [www.opendata.comune.fi.it](http://www.opendata.comune.fi.it); this dataset contains information about the number of families in each district divided by number of children, but such information are ignorable for our purpose.

### 3. Methodology

The analysis demanded does not involved any machine learning technique as we are not in the case of a classification or clustering analysis; bosses were actually interested in data and wanted us to depict the actual scenario.

For these purposes, we tried to fulfil their requests and studied different aspects of datasets: first we import and transform data of **Population by District and Age data** and **Family Number by District and N° of Children data**; we had to so as we were not interested in the value itself, but we'd rather gather information related to district characteristics (as, for example, surface). In the end, we get to compute density and percentage of target subpopulation.

Later, as some specific spot where located by bosses and we had geographical information on those, we exploit Foursquare API to find venues nearby given coordinates. In the end we only needed to compute these data to answer questions we were given.

#### 4. Results

The very first evidence we obtained is about families and how they spread through districts. Note that, even if we had many different groups, we split the population in two subsets: families with less than 3 components and families with more (or equal) than 3 components; we made this choice by assuming that the second subset may represent families with at least 1 kid, which are of main interest in our study.

N° componenti	Tot	Under_3_comp	Over_3_comp	Under_3_comp_perc	Over_3_comp_perc	Superficie_km2	Under_3_comp_per_km2	Over_3_comp_per_km2
Quartiere 1	36142	28666	7476	0.793149	0.206851	11396	2.515444	0.656020
Quartiere 2	45861	33168	12693	0.723229	0.276771	23406	1.417073	0.542297
Quartiere 3	19824	13933	5891	0.702835	0.297165	22312	0.624462	0.264028
Quartiere 4	32378	22358	10020	0.690531	0.309469	16991	1.315873	0.589724
Quartiere 5	53783	38901	14882	0.723295	0.276705	28171	1.380888	0.528274

Figure 1. Families distribution data.

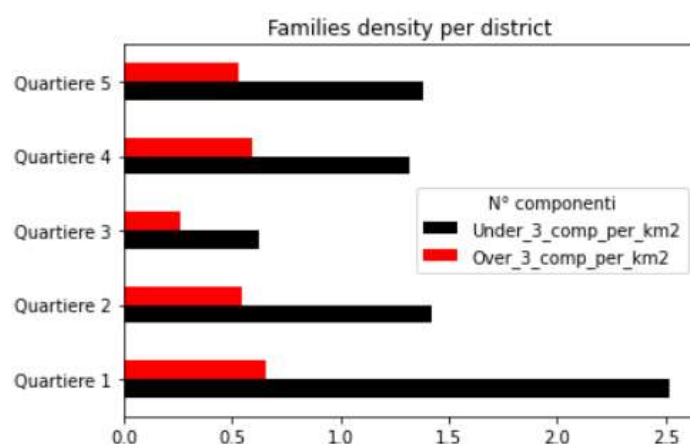


Figure 2. Densities of families under and over 3 components.

As we can see from the graph, in each district the majority of family unites is composed of 1 or at least 2 people, in particular *Quartiere 1* has a huge concentration of this kind of households (with a density of over 2.5). Instead, *Quartiere 3* shows a general low density of families (for all of the categories) due to the great surface of the area in contrast with the relatively small number of people. Other districts have similar distribution.

Concerning the presence of people in target age range the situation depicted is the following:

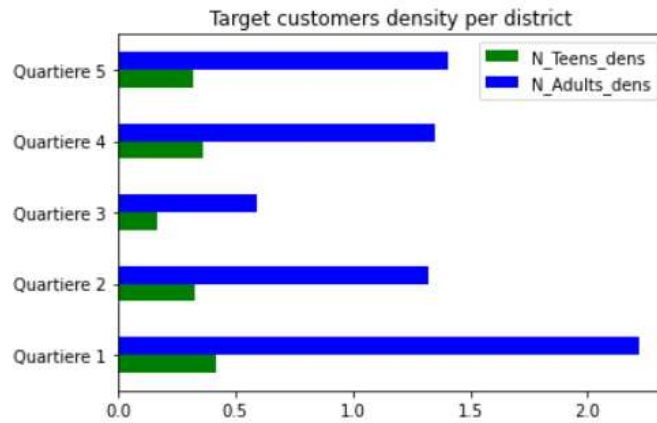


Figure 3. Densities of teenagers and adults per district.

The situation is quite similar to the one we had with families distribution and once again the areas that differ from other are *Quartiere 1* and *Quartiere 3*.

Once we retrieved data through Foursquare API, the last thing we had to examine is the presence of restaurant in each area:

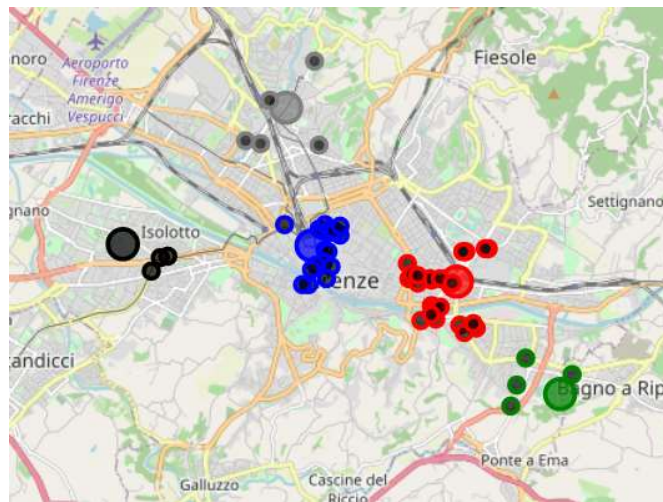


Figure 4. A map of Florence with restaurant marks.

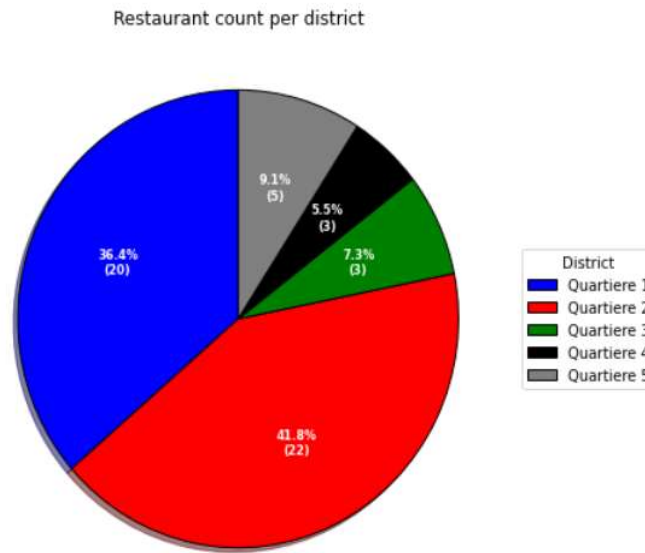


Figure 5. Restaurant count per district.

So, the majority of restaurants (considering coordinates given) are located within *Quartiere 1* and *Quartiere 2*; this of course comes as no surprise, in fact these areas are the most central and are close to main touristic spots.

## 5. Discussions and Conclusions

All in all, if we may suggest what district may be the most suitable for opening a new restaurant we should say *Quartiere 4*. The key points of our choice are the following:

- We immediately discarded *Quartiere 3* as it is too much dispersed;
- *Quartiere 1* seemed a good choice considering the great amount of people in target and families, however the huge presence of other restaurants in this area makes competition a real problem; a similar thought was made for *Quartiere 2*;
- *Quartiere 4* was chosen over *Quartiere 5* because of the slight difference in families and teens density; moreover, there are 2 light rail station close to the spot located (under 1 km), which are valuable.