NEW TOURIST OFFICES IN ROME

Abstract. In this report I discuss the optimal placement of five new tourist offices in Rome. These tourist offices will specialize in long guided walking tours of historic sites. Ample funding is available, making differences in rent and housing prices insignificant.

1. introduction

1.1. **Background.** Rome, the capital city of Italy, is one of the oldest cities in Europe, with a history spanning more than 2 millennia. It is one of the most visited cities in the world, with millions of tourists coming every year. As capital of the ancient Roman empire, Rome is home to a large number of historic sites.

1.2. **Problem.** The city of Rome is planning to open five new tourist offices in Rome, which specialize in *long* guided walking tours of historic sites. The city is interested in the optimal placement of these five tourist offices.

Ample funding is available, making differences in rent and housing prices insignificant for our purposes. Since these tours are *walking* tours, availability (and quality) of public transport connecting different sights and tourist offices is insignificant for our purposes as well.

The problem then becomes to find five locations, so that each historic site in Rome has low (walking) distance to at least one of the locations.

2. Data

2.1. **Sources.** I used the *Foursquare API* in order to obtain a list of the various historic sites of Rome, including their respective geographic coordinates (latitude and longitude). Since the Foursquare API never returns more than 50 venues per call, I needed to make several API calls to cover all of Rome.

Finally, I used *Wikipedia* to get the coordinates for Rome itself, as well as the coordinates for several of its boroughs.

2.2. **Cleaning and Feature Selection.** Some venues were incorrectly labeled as historic sites; I removed them from the data.

Since the explore command of the Foursquare API returns venues ordered by popularity, the only features significant for our purposes were name and geographic coordinates, as is obvious from the above problem statement.

This is the author's final project for the "Applied Data Science Capstone" course on Coursera, offered by IBM.

3. Methodology

3.1. **Data Visualization.** First I displayed a map of Rome and its boroughs (Figure 1). This visualization (and all remaining visualizations) were created using the *Folium* library. After carrying out several GET requests to the Foursquare API, I combined the significant

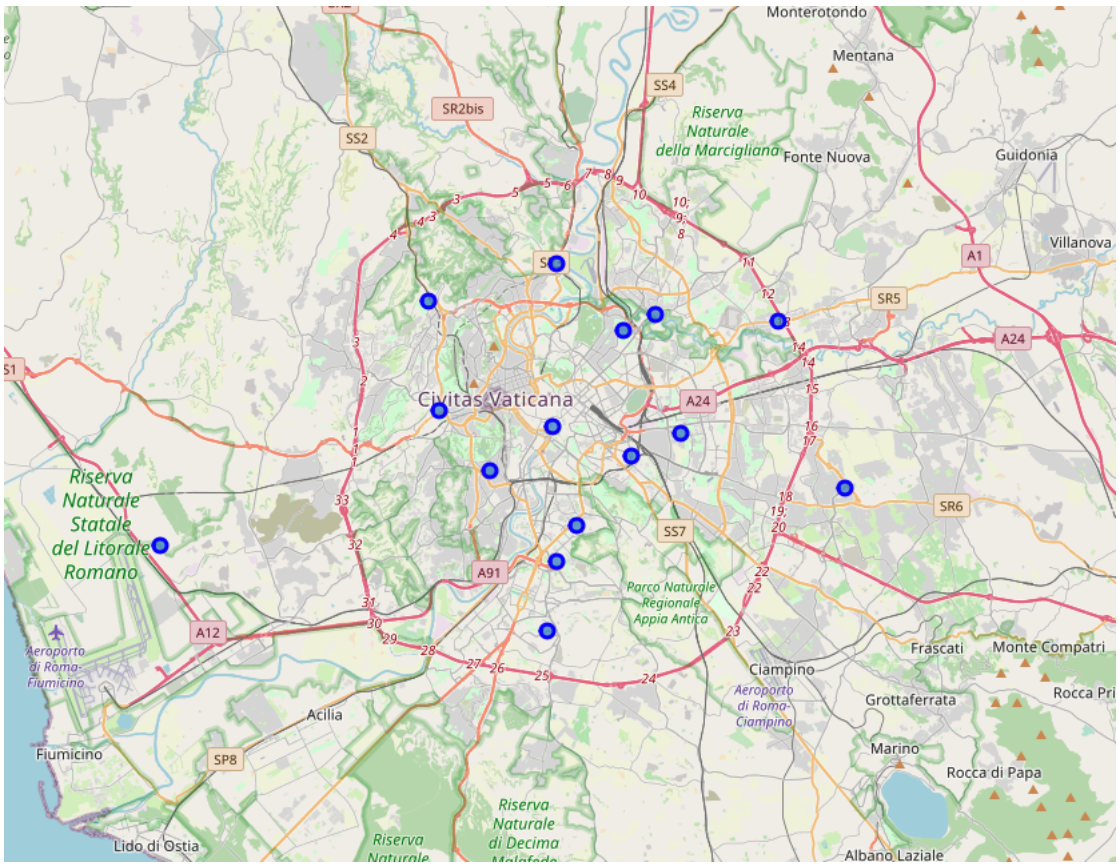

Figure 1. Rome and its boroughs

features into a dataframe (head only, Figure 2), which I then displayed on a map (Figure 3).

3.2. **Methods and Analysis.** In order to compute distances between locations, one needs to obtain a formula in terms of the latitude and longitude. Using (a slight variation of) spherical coordinates, one easily arrives at the haversine formula: given two points (ϕ_1, λ_1) and (ϕ_2, λ_2), where ϕ_j denotes latitude and λ_j denotes longitude, the distance of these two points computes to

$$2R \arcsin\left(\sqrt{\sin^2\left(\frac{\phi_2 - \phi_1}{2}\right) + \cos(\phi_1)\cos(\phi_2)\sin^2\left(\frac{\lambda_2 - \lambda_1}{2}\right)}\right),$$

where R is the radius of the earth. When restricting attention to points close to Rome, which has coordinates $(\phi_0, \lambda_0) = (41.883333 \cdot \pi/180, 12.5 \cdot \pi/180)$ (in radians), a simple power series

	Historic Site	Latitude	Longitude
0	Colosseo	41.890633	12.492378
1	Foro Romano	41.892030	12.487037
2	Foro di Cesare	41.894128	12.485232
3	Palatino	41.888234	12.487209
4	Stadio palatino (Stadio palatino | Stadio di D...	41.887582	12.487496
5	Foro di Traiano	41.894729	12.484871
6	Terme di Caracalla	41.878990	12.492443
7	Portico d'Ottavia	41.892382	12.478500
8	Scalinata di Trinità dei Monti	41.905974	12.482647
9	Teatro di Marcello	41.891931	12.479798

FIGURE 2. The 10 most popular historic sites in Rome

expansion gives the following approximation to the above expression:

$$\begin{aligned} &2R\sqrt{\left(\frac{\phi_2-\phi_1}{2}\right)^2+\cos^2(\phi_0)\cdot\left(\frac{\lambda_2-\lambda_1}{2}\right)^2} \\ =&R\cdot\|(\phi_1,\lambda_1\cos(\phi_0))-(\phi_2,\lambda_2\cos(\phi_0))\|, \end{aligned}$$

where $\|\cdot\|$ denotes the Euclidean norm (2-norm) on $\mathbb{R}^2$. Defining $\lambda' = \lambda\cdot\cos(\phi_0)$ for all λ, the latter expression can be rewritten as

$$R\cdot\|(\phi_1,\lambda_1')-(\phi_2,\lambda_2')\|.$$

Since R is just a constant (positive) factor, and one is only interested in *comparing* distances, one can drop that factor and consider the following expression instead:

$$\|(\phi_1,\lambda_1')-(\phi_2,\lambda_2')\|.$$

But this is just the Euclidean distance (in the new coordinates), which means that one can use the K-Means clustering algorithm (which uses Euclidean distance).

After establishing this elementary fact, I transformed the coordinates (first to radians, and then to modified coordinates, see Figure 4, head only). Then I used the **K-Means clustering algorithm** from the *scikit-learn* library with number of clusters equal to 5, i.e., to the number of new tourist offices.

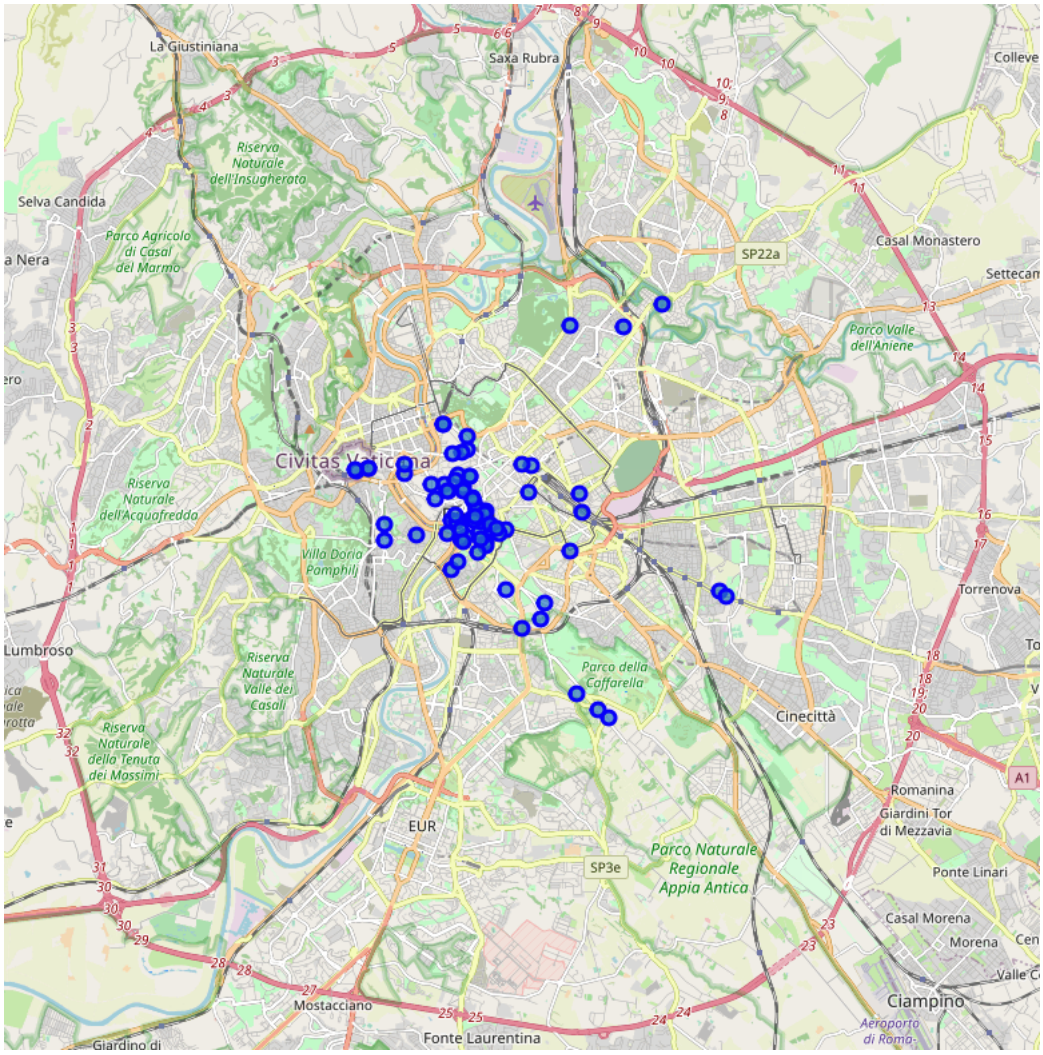

FIGURE 3. Most popular historic sites in Rome

	Historic Site	Latitude (rad)	modified Longitude
0	Colosseo	0.731129	0.162327
1	Foro Romano	0.731154	0.162258
2	Foro di Cesare	0.731190	0.162234
3	Palatino	0.731088	0.162260
4	Stadio palatino (Stadio palatino | Stadio di D...	0.731076	0.162263

FIGURE 4. Historic sites with modified coordinates

4. Results

The coordinates for the five locations in Rome found by the algorithm are the following:

	Latitude	Longitude
0	41.891833	12.487710
1	41.900684	12.472790
2	41.878225	12.548237
3	41.930894	12.521352
4	41.865195	12.507724

Figure 5. Coordinates of new tourist offices

Every historic site has distance *less than* 2.1 km to one of the proposed locations. I displayed these locations on a map of Rome, together with the historic sites:

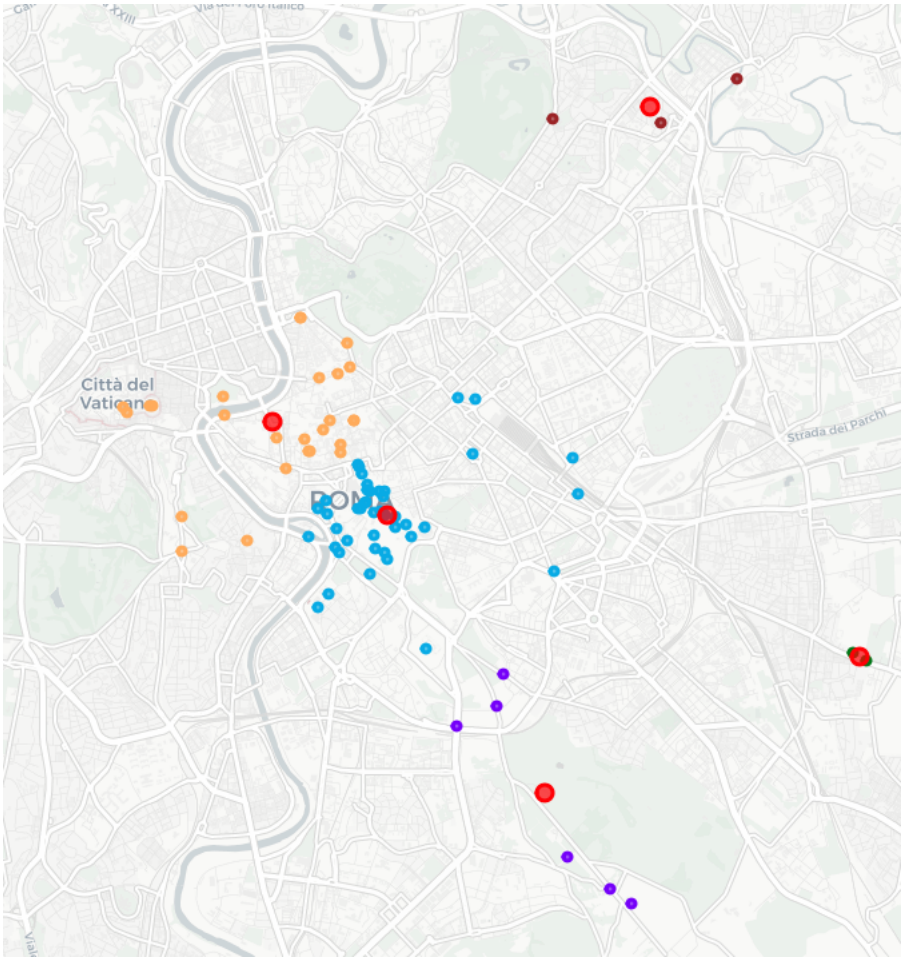

Figure 6. Tourist offices (red) and historic sites

5. Discussion

Each of the historic sites has distance less than 2.1 km to at least one of the tourist offices. While, of course, the path one has to take is usually longer than the geometric distance (due to the layout and terrain of the city), the corresponding factor rarely exceeds 1.5, giving us an estimated longest path of roughly 3 km. Hence, **each historic site is within walking distance of at least one of the tourist offices**, as desired.

One readily verifies that all five potential new offices are conveniently located by manually checking public transport schedules. Furthermore, the fact that two of the new offices are quite central (and hence close to each other) is advantageous, since the density of historic sites and tourists is higher in the centrum.

6. Conclusion

In conclusion, I recommend that the city of Rome establishes the five new tourist offices in the following locations:

	Latitude	Longitude
0	41.891833	12.487710
1	41.900684	12.472790
2	41.878225	12.548237
3	41.930894	12.521352
4	41.865195	12.507724

Figure 7. Coordinates of new tourist offices

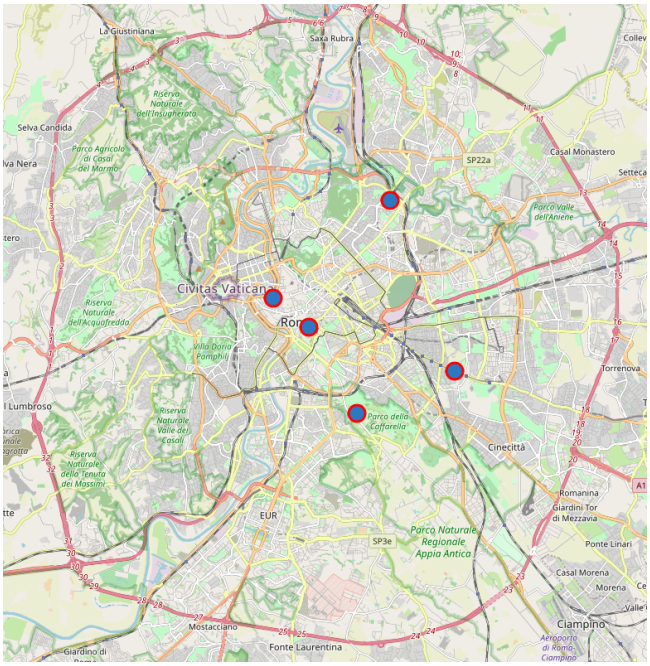

Figure 8. Location of new tourist offices