New Tourist Offices in Rome

Applied Data Science Capstone

September 1, 2020

Introduction

The city of Rome plans to open five new tourist offices.

- They will specialize in long guided walking tours of historic sites.
- Differences in rent and housing prices are insignificant (ample funding).
- Availability of public transport to historic sites from tourist offices is insignificant.

Task

Find five locations, so that each historic site in Rome has low (walking) distance to at least one of the locations.

Data I

Data Sources

- *Foursquare API*: historic sites and their coordinates
- *Wikipedia*: coordinates of Rome and its boroughs

Cleaning

Incorrectly labeled venues were removed.

Feature Selection

- Venue name
- coordinates (latitude and longitude)

Data II

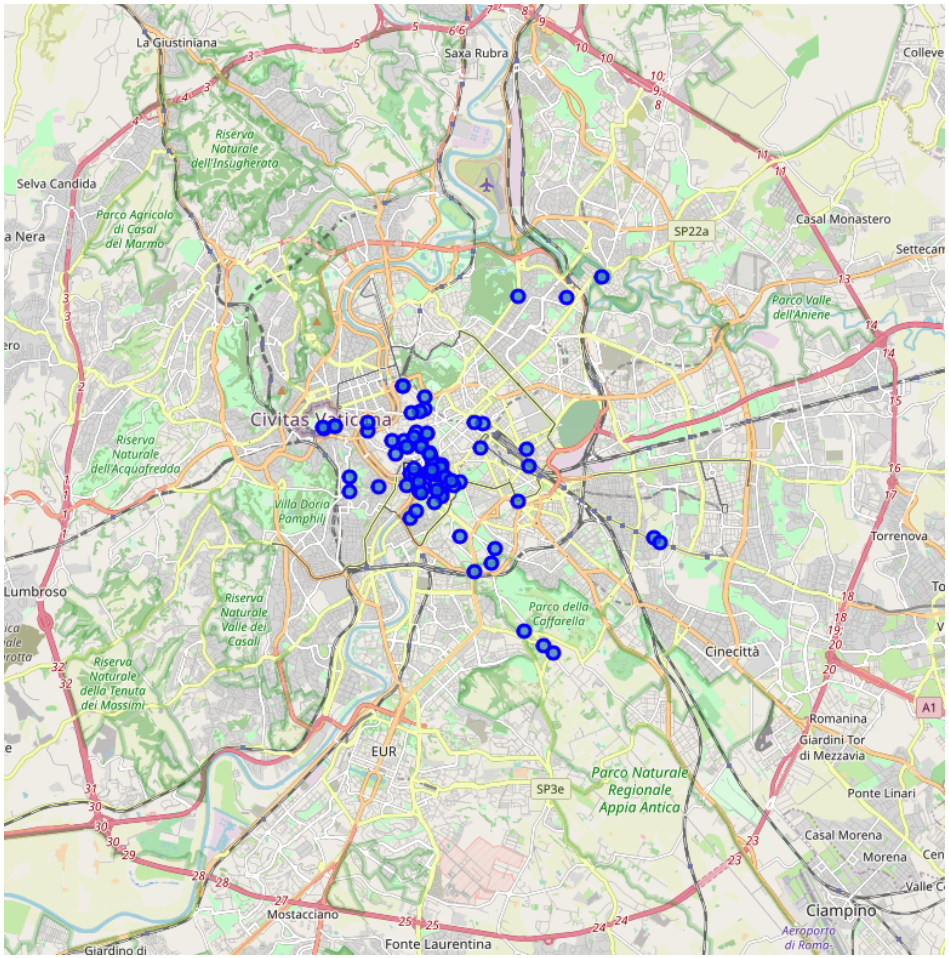

Figure: The most popular historic sites in Rome

Data III

	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: The 10 most popular historic sites in Rome

Methods and Analysis I

Given two points (ϕ latitude, λ longitude)

- (ϕ_1, λ_1),
- (ϕ_2, λ_2),

their distance is:

Haversine Formula

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

where R is the radius of the earth.

Methods and Analysis II

If both points are close to Rome, their distance (up to the constant factor $R > 0$) is roughly

Distance Approximation (up to positive factor)

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

where $\lambda' = \lambda \cdot \cos(41.883333 \cdot \pi/180)$ for all λ, and where $\|\cdot\|$ denotes Euclidean distance.

Methods and Analysis III

Since the K-Means algorithm works with Euclidean distance:

Chosen Method

K-Means clustering algorithm from the *scikit-learn* library

- number of clusters equal to 5,
- had to transform latitude and longitude to modified coordinates.

Results I

- **Every historic site has distance less than** 2.1 km **to one of the locations found by the algorithm,**
- Locations proposed by algorithm:

	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: Coordinates of new tourist offices

Results II

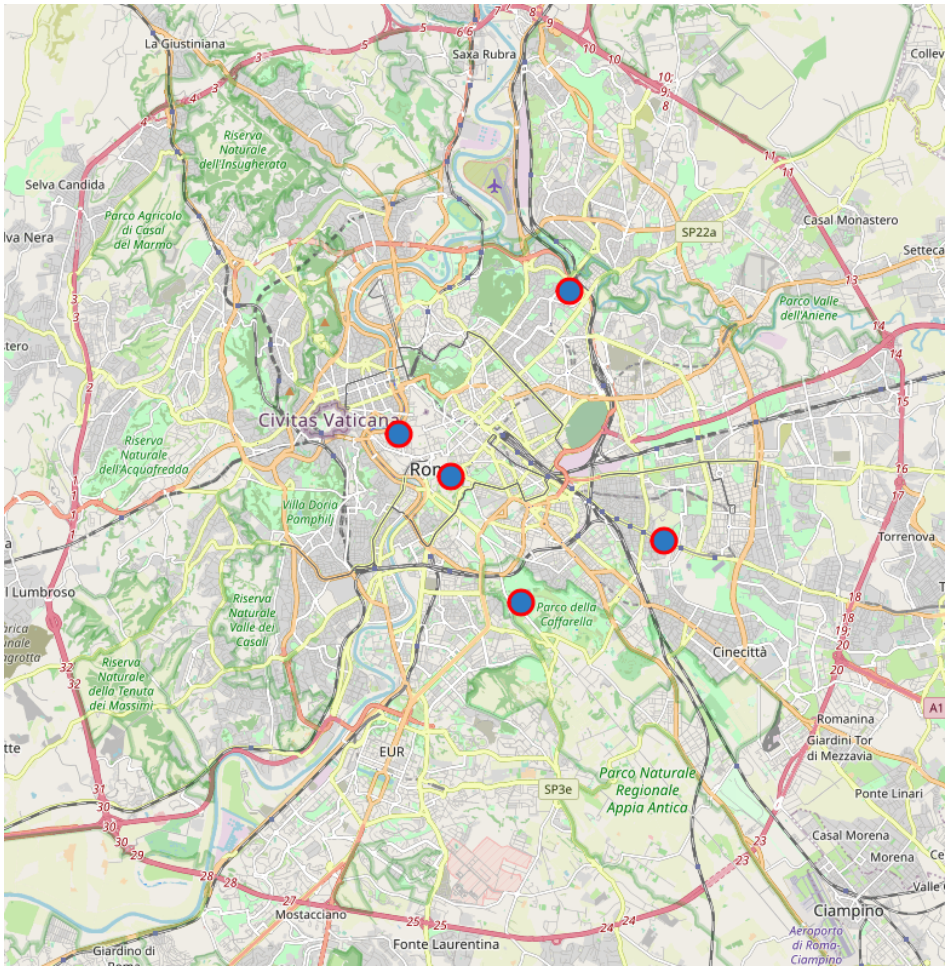

Figure: New tourist offices

Results III

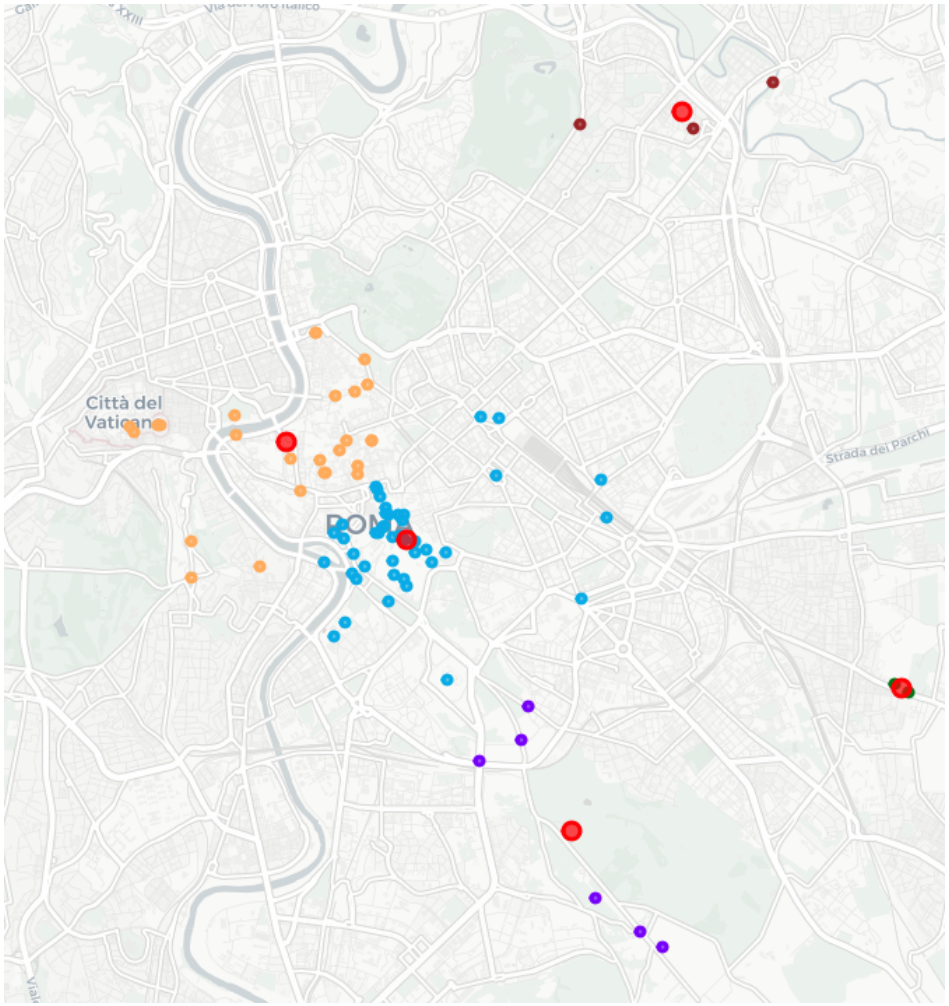

Figure: Historic sites and new tourist offices

Discussion

- Each historic site is within walking distance of one of the tourist offices.
- All five tourist offices can be reached well from the center.
- Two offices in center, accounting for higher density of sites and tourists.

Conclusion

Recommendation

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: Coordinates of new tourist offices

The End