



# Leveraging Data Science to Predict New Shopping Mall Location in Rome, Italy

IBM Applied Data Science Capstone  
Kenneth Zong



# Introduction/Business Problem:

- Shopping malls are a place of happiness for the ordinary consumer to actively shop, eat, relax, and spend time with loved ones.
- The mall experience is unique such that there is always something for someone to do. Not only that, but for the average tourist, a shopping mall can be a great way to learn about a country's culture without having to travel many miles.
- Many surrounding clubs, schools, and venues have the flexibility to market themselves more especially at a centralized hub such as a shopping mall. However, what makes the shopping mall experience so great is not just the extracurricular events that are held there, but also the location.
- If the location is within a cluster radius of too many malls, it will be difficult to earn profit as the new mall will be competing with other ones within their vicinity.
- Business Question to be asked: If a contractor is told to suggest a few of the places that a shopping mall would flourish, where would it be located, based on the data of Rome, Italy?

## Data needed:

- This Wikipedia page ([https://en.wikipedia.org/wiki/Category:Subdivisions\\_of\\_Rome](https://en.wikipedia.org/wiki/Category:Subdivisions_of_Rome)) has neighborhood data for list of neighborhoods in Rome , Italy
- Python Geocoder package will be leveraged for latitude and longitude coordinates of the neighborhoods.
- Foursquare API (<https://foursquare.com/>) to get the venue/shopping mall data within the neighborhoods

# Methodology

- Utilize web scraping with Python requests and beautifulsoup packages to extract the list of neighborhoods
- Retrieve the geographical coordinates in the form of latitude and longitude using Python Geocoder package
- Using Foursquare API (<https://foursquare.com/>) to get the venue data
- Analyze each neighborhood by grouping the rows by neighborhood and taking the mean of the frequency of occurrence of each venue category.
- Filter venue category by Shopping Mall
- Perform clustering models on the data by using k-means clustering

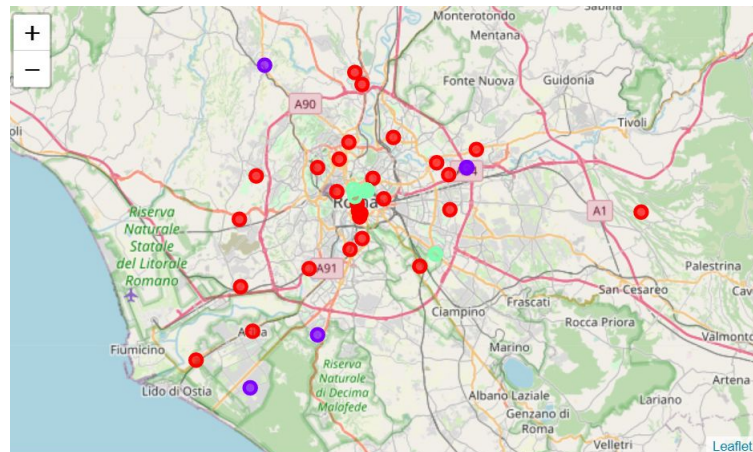
# Results

Categorizing the neighborhoods into 3 various clusters based on number of nearby shopping malls

Cluster 0: neighborhoods with low number to no existence of shopping malls

Cluster 1: neighborhoods with high concentration of shopping

Cluster 2: neighborhoods with moderate number of shopping malls



# Discussion

- It can be seen that there is a great opportunity and high potential areas to open new shopping malls as there is very little to no competition from existing malls within cluster 0.
- Meanwhile, a new shopping mall open in cluster 1 could encounter intense competition due to oversupply and high concentration of shopping malls.
- Therefore, cluster 0 may be the better option if the property developer does not want to encounter any competition.
- However, cluster 2 is possible to be the better option if there is more population within the vicinity and has a niche characteristic such as a unique restaurant, store, etc.

# Conclusion

- The neighborhoods in cluster 0 are the most preferred locations to open a new shopping mall if property stakeholders want to be more conservative and encounter no competition.
- However, cluster 2 with moderate competition may be another option depending on population vicinity and niche market product.
- The findings of this project will help the relevant stakeholders to capitalize on the opportunities in high potential locations while avoiding overcrowded areas in their decisions to open a new shopping mall.

# References

- Category: Subdivisions of Rome in Wikipedia. Retrieved from [https://en.wikipedia.org/wiki/Category:Subdivisions\\_of\\_Rome](https://en.wikipedia.org/wiki/Category:Subdivisions_of_Rome)
- Foursquare Developers Documentation. Foursquare. Retrieved from <https://developer.foursquare.com/docs>



THANK YOU!