

# Capstone Project - The Battle of the Neighborhoods (Week 2)

Applied Data Science Capstone

World Heritage Tourism

# Leisure Tourism vs. Adventure

- World Heritage sites serve as point of reference for individuals who **love to travel and are history, architecture enthusiast**
- These sites are situated in variety of terrains and have different levels of accessibility and availability of amenities. For a recreational tourist accessibility and comfort may be preferred while tourists with adventurous tendency may like to use this opportunity explore more.
- Individuals planning a vacation have no easy way to classify these sites to suit their style.
- Tour operators are interested in running a guided tours for tourists of recreational inclination to remote sites, their selling point will be arranging comfort and amenities for this cohort to encourage them to enjoy the sites which they would not have seen otherwise.

# A Data Based Approach for Grouping



**Gather Locational data about the sites from public websites such as**

-

UNESCO  
(<https://whc.unesco.org/en/list/>)

World Heritage Site by Wikipedia  
([https://en.wikipedia.org/wiki/World\\_Heritage\\_Site](https://en.wikipedia.org/wiki/World_Heritage_Site))



**Following data points will be collected manually -**

Name of the site  
Latitude  
Longitude  
Country  
Built-in



**Enhance data using Foursquare API.**



**collect following information within 20 kms radius of the site –**

Venue Name  
Venue Latitude  
Venue Longitude  
Venue Category



**Create a model using unsupervised machine learning algorithm, K-Means Clustering.**

# Comfortable Tourism



**When basic amenities are easily available –**

- Food
- Accommodation
- Transport



**More return of investment when**

- More point of interests close by
- Additional features such as beaches around

# Adventurous Tourism



ISOLATED SITES



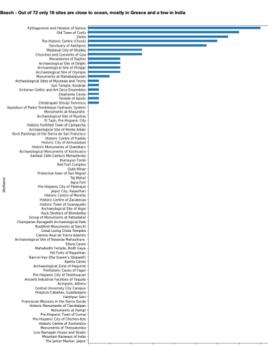
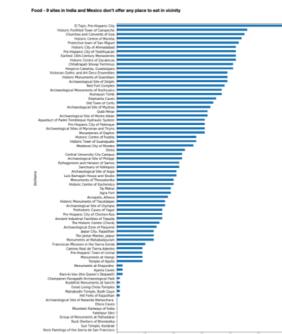
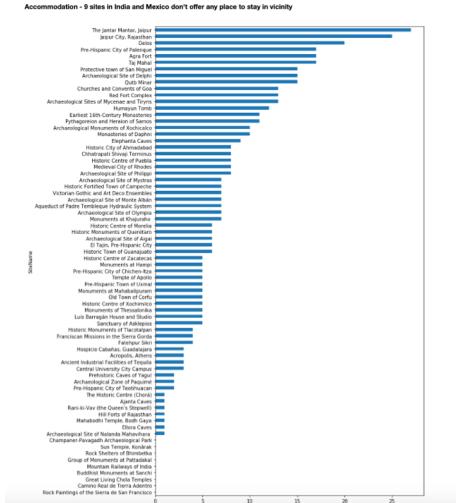
NATURAL HABITATS  
AROUND



NO FORMAL  
ACCOMMODATION

# What Data Tells

- 209 unique categories
- 72 sites to develop model
- 9 out of 72 sites don't have formal accommodation in 20km range
- 9 out of 72 sites don't have a place to eat in 20km range.
- 18 out of 72 sites offer beaches in 20km range



# Model Development

- K-Means Clustering Algorithm
- Top 5 features in use
- Total 6 clusters expected

	SiteName	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Acropolis, Athens	Café	Food	Drink	Historic Site	Museum
1	Agra Fort	Food	Accommodation	Historic Site	Café	Transport
2	Ajanta Caves	ATM	Food	Accommodation	Mobile Phone Shop	Historic Site
3	Ancient Industrial Facilities of Tequila	Food	Drink	Museum	Accommodation	Plaza
4	Aqueduct of Padre Tembleque Hydraulic System	Food	Accommodation	Plaza	Convenience Store	Historic Site
5	Archaeological Monuments of Xochicalco	Food	Accommodation	Garden	Sport Center	Dessert Shop
6	Archaeological Site of Aigai	Food	Café	Accommodation	Drink	Cupcake Shop
7	Archaeological Site of Delphi	Food	Accommodation	Café	Drink	Historic Site
8	Archaeological Site of Monte Albán	Food	Drink	Café	Accommodation	Dessert Shop
9	Archaeological Site of Mystras	Food	Café	Accommodation	Drink	Plaza
10	Archaeological Site of Nalanda Mahavihara	Transport	Accommodation	Historic Site	Mountain	Zoo
11	Archaeological Site of Olympia	Food	Café	Accommodation	Historic Site	Drink
12	Archaeological Site of Philippi	Food	Café	Accommodation	Drink	Supermarket
13	Archaeological Sites of Mycenae and Tiryns	Food	Accommodation	Café	Dessert Shop	Historic Site
14	Archaeological Zone of Paquimé	Food	Accommodation	Department Store	Convenience Store	Museum
15	Buddhist Monuments at Sanchi	Transport	Food	Dessert Shop	Historic Site	Bank
16	Camino Real de Tierra Adentro	Food	Farm	Mountain	Business Service	Racetrack
17	Central University City Campus	Food	Dessert Shop	Park	Café	Theme Park Ride / Attraction
18	Champaner-Pavagadh Archaeological Park	Food	Transport	Temple	Ski Area	Historic Site
19	Chhatrapati Shivaji Terminus	Food	Café	Drink	Accommodation	Dessert Shop
20	Churches and Convents of Goa	Food	Accommodation	Drink	Café	Beach
21	Delos	Food	Accommodation	Drink	Beach	Café
22	Earliest 16th-Century Monasteries	Food	Accommodation	Drink	Historic Site	Dessert Shop
23	El Tajín, Pre-Hispanic City	Food	Accommodation	Dessert Shop	Historic Site	Drink
24	Elephanta Caves	Food	Café	Accommodation	Drink	Dessert Shop
25	Ellora Caves	Historic Site	Accommodation	Castle	Zoo	Electronics Store
26	Fatehpur Sikri	Accommodation	Historic Site	Sculpture Garden	Border Crossing	Park
27	Franciscan Missions in the Sierra Gorda	Food	Accommodation	Scenic Lookout	Mountain	Plaza
28	Great Living Chola Temples	Historic Site	Transport	Food	Clothing Store	Museum
29	Group of Monuments at Pattadakal	Sculpture Garden	Historic Site	Scenic Lookout	Drink	Museum
30	Hill Forts of Rajasthan	Transport	Accommodation	Food	Trail	Zoo
31	Historic Centre of Morelia	Food	Dessert Shop	Accommodation	Drink	Café

# Results

- Following clusters were determined by algorithm
  - Urban Setup (Comfortable for Tourists)
  - City Setup (Easily Doable)
  - Urban Setup with more to do for tourists (Economical)
  - Isolated Site (Adventurous)
  - Remote but Connected
  - Remote Site and not much to do

# Conclusion



Categorization model will help tourists to plan their upcoming trip in a more personalized way by knowing ahead of time, what they are signing up.



Using basic categorization users of this report can further fine tune their trip based on the additional research done about the country and the area they are planning to visit.



Stakeholders can also use this report to plan their offering.



I will be working further to enhance dataset using more popularity and social media data to fine tune this model. It was evident while analyzing the data some of the remote sites may not be having enough data as not many people may have reported it to Foursquare.