TOURIST BEHAVIOUR ANALYSIS

Sarvesh Navare - 60004180096
Saurav Tiwari - 60004180097
Shreerang Taparia - 60004180101
Shruti Sawant - 60004180103

Guide Name-Prof. Pankaj Sonawane

Introduction

This project is to analyse tourists behaviour based on the locations they have visited so far, to identify tourist interests and tourism demographics over time variant data. It supports strategic decision-making in tourism destination management. For the sake of this project, examples taken:

- Goa
- Rajasthan
- Himachal Pradesh

Challenges

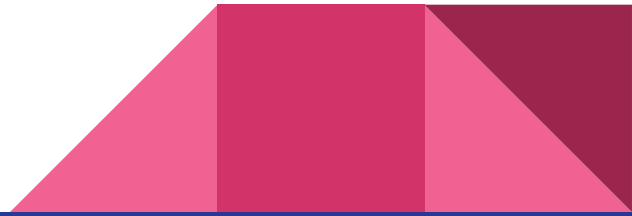
- Scraping of large quantities of data from Instagram without using API.
- Data Cleansing
- Sampling and Mapping of Spatial Coordinates.

Literature Review

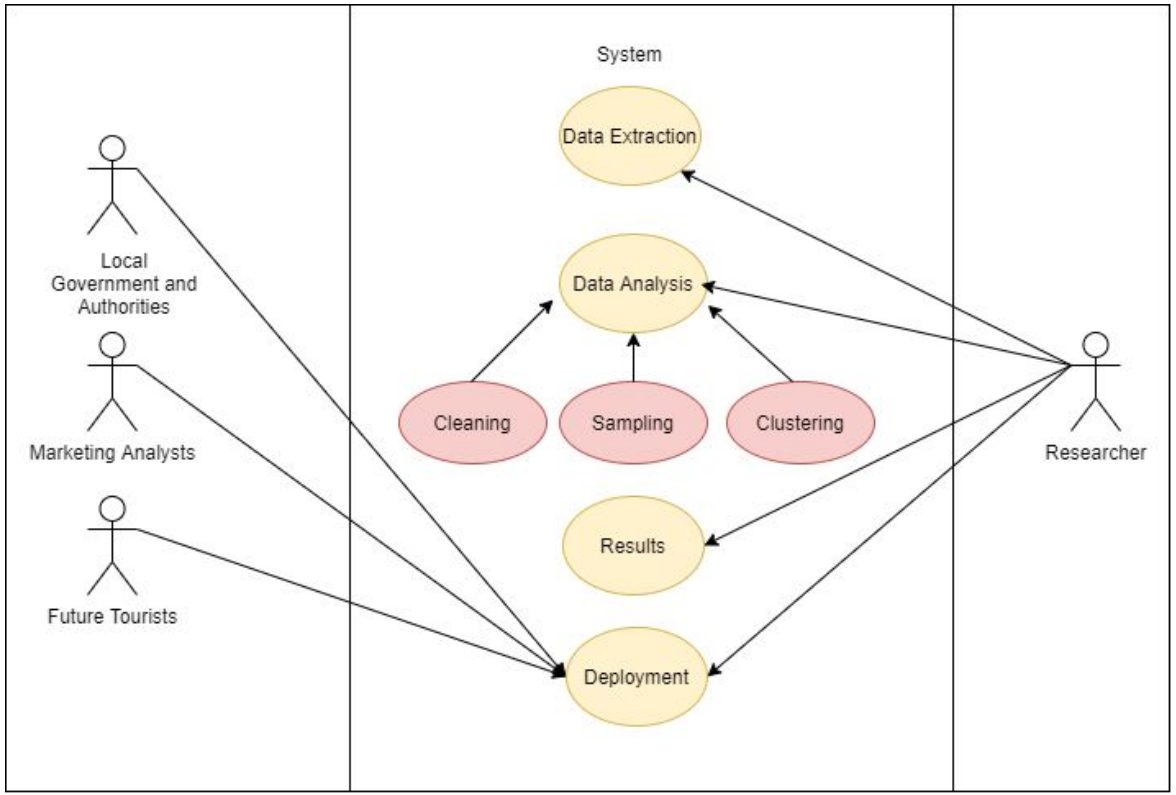
- After studying a few papers, we realised that major research has been done on tourist recommendation systems rather than tourist behaviour analysis.
  10.1109/ICASSP.2010.5495905
  10.1109/PerComW.2013.6529508
- A research done on analyzing spatial data from Twitter faced the hurdle of constant blocking.
  https://doi.org/10.1016/j.trc.2018.09.006

Research Gaps

- Insufficient use of gadgets having sensors while travelling.
- No single source to obtain information in a desired format.
- Presence of online bots hinder the process getting meaningful data.
- No usage of time variant data to compare trends.



Use Case Diagram



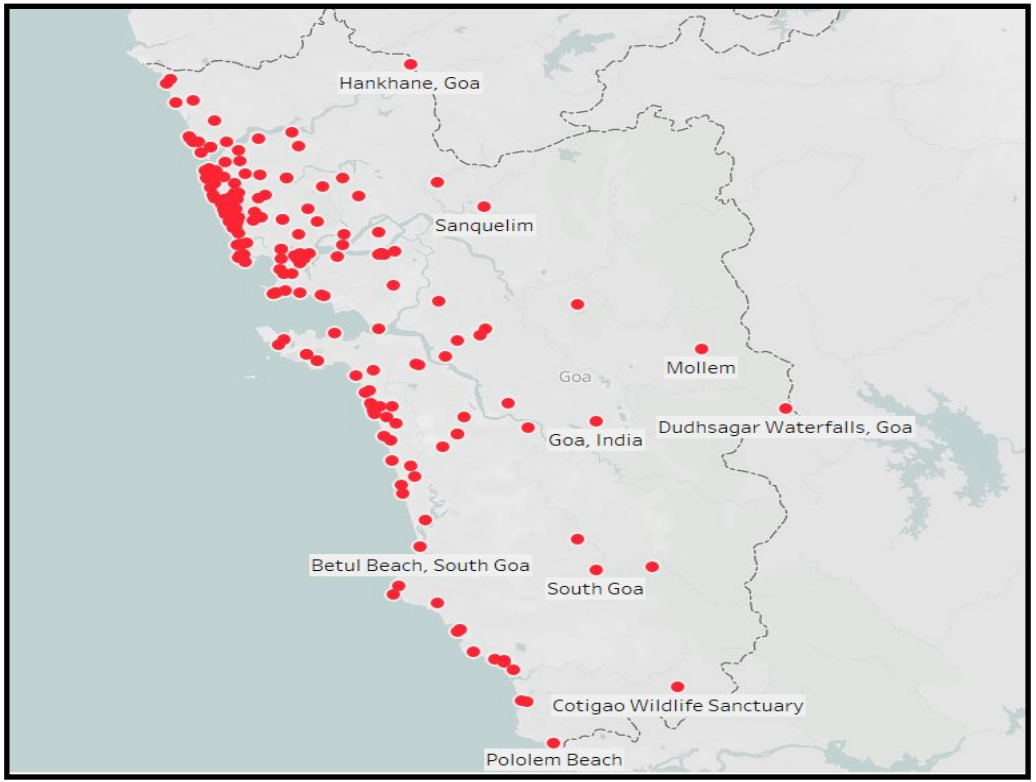
Implementation and Results

- Our objective was to group frequently visited locations according to their latitude and longitude and discover the underlying patterns.
- We have created the dataset from the data extracted by scraping information of users and their geotagged locations from a social media platform - Instagram.
- We have worked on sample tourist data of Goa, Rajasthan and Himachal Pradesh having approximately 1000 records each.
- Several steps of data preprocessing were then carried out on the dataset collected via data extraction.

Implementation and Results

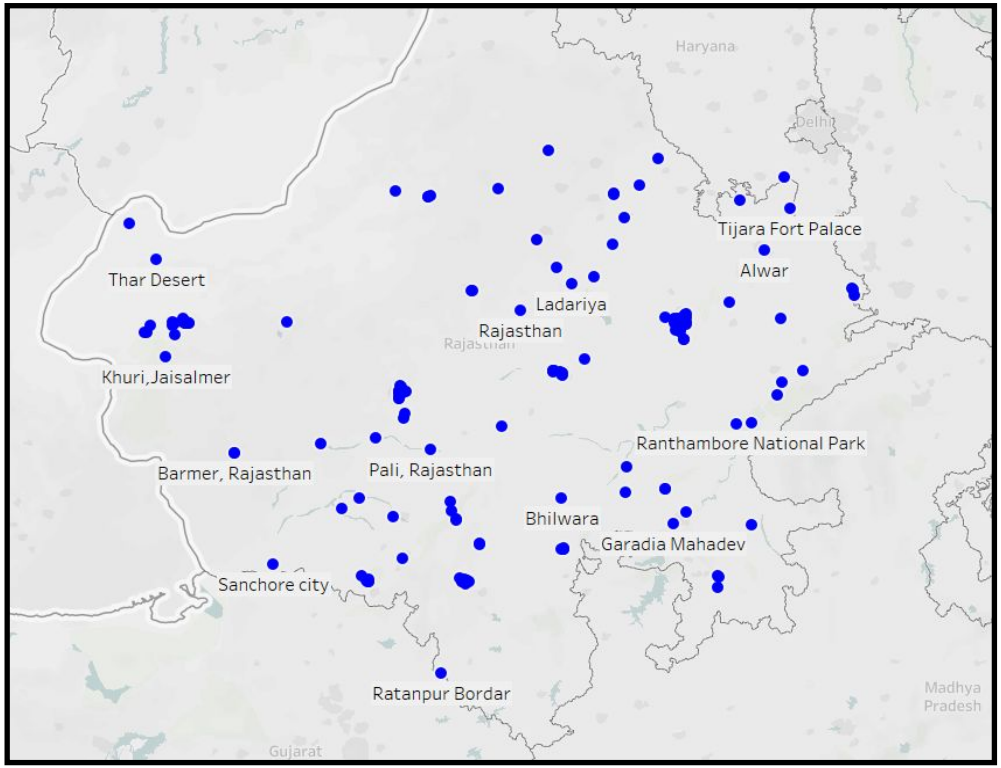
- K-Means clustering algorithm was used to cluster locations and the clusters were mapped.
- We tried to do a quarterwise analysis on the clusters and the number of tourists touring the state from our sample dataset.
- From our analysis, we made an effort to estimate the most popular area of the state.
- Also, the quarter of the year which experienced maximum tourists was identified.

Implementation and Analysis



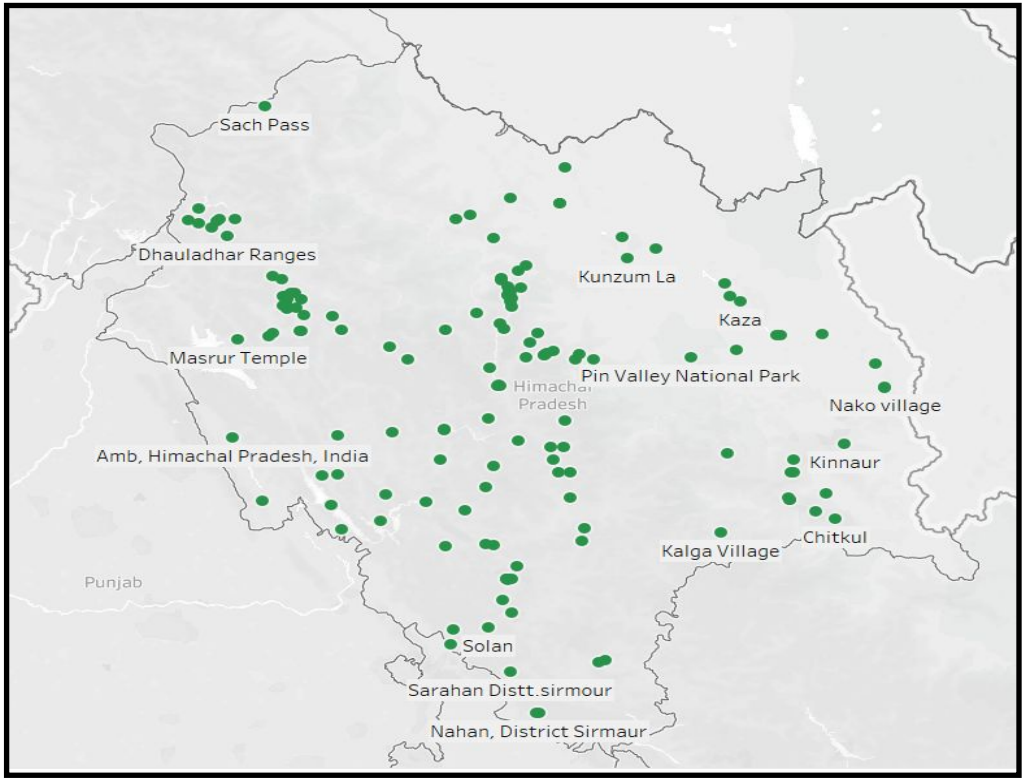
- Unique tourist locations in **Goa** identified from our sample dataset.

Implementation and Analysis



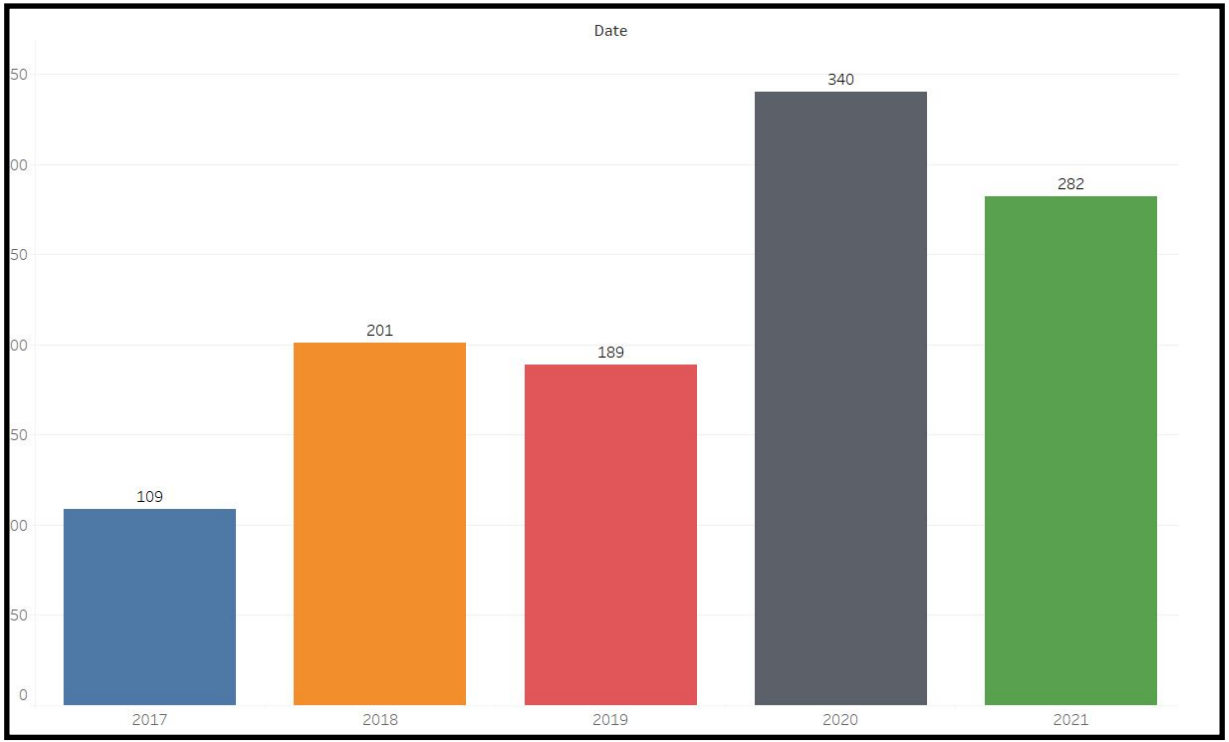
- Unique tourist locations in **Rajasthan** identified from our sample dataset.

Implementation and Analysis



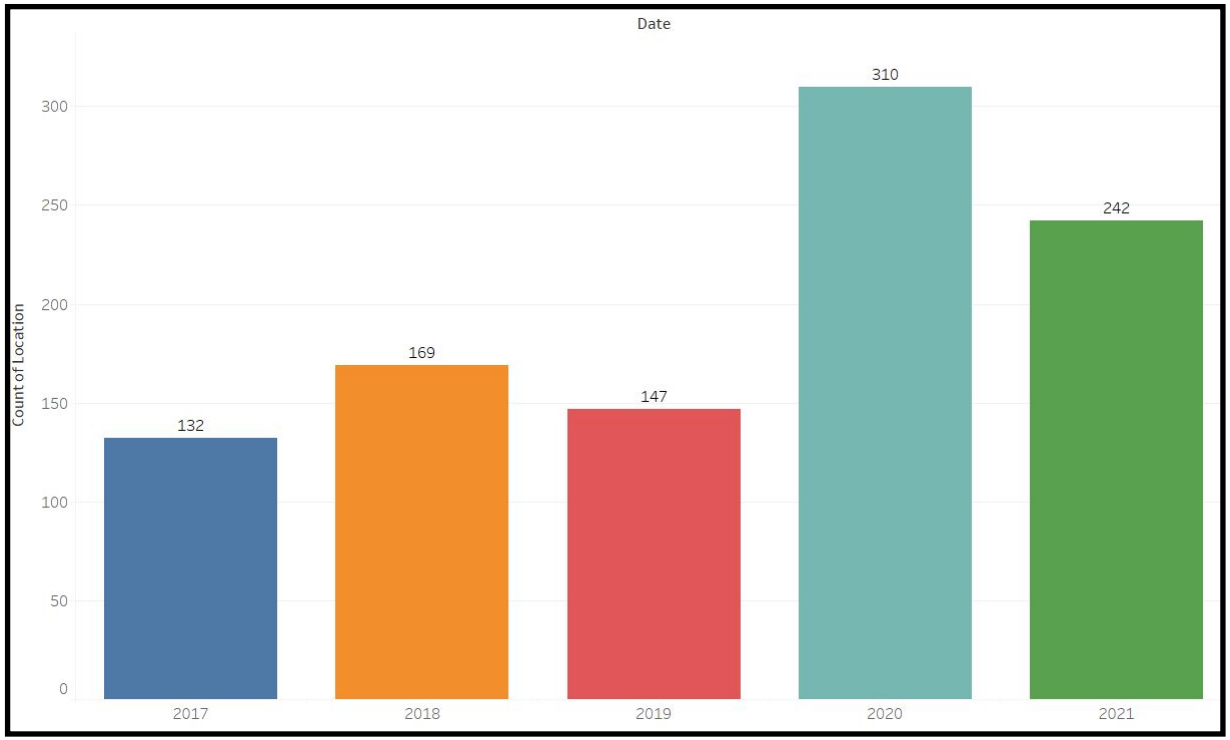
- Unique tourist locations in **Himachal Pradesh** identified from our sample dataset.

Implementation and Analysis



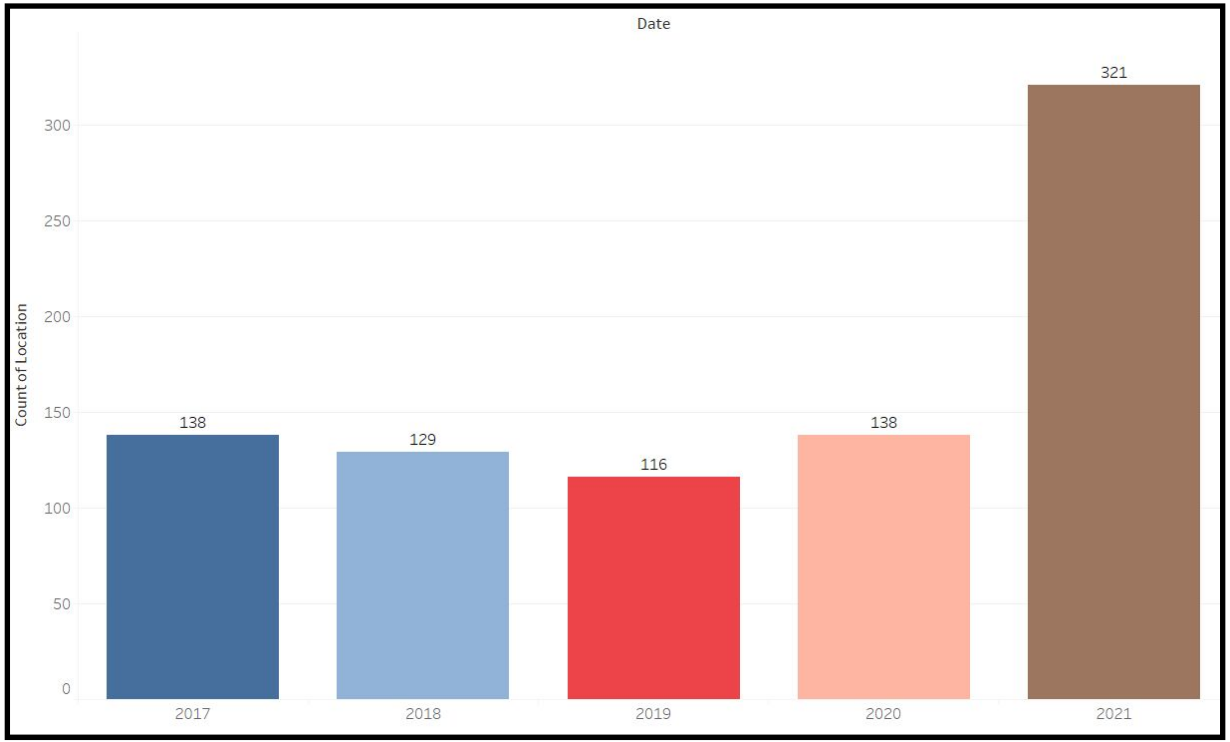
- Count of tourists visiting **Goa** per year

Implementation and Analysis



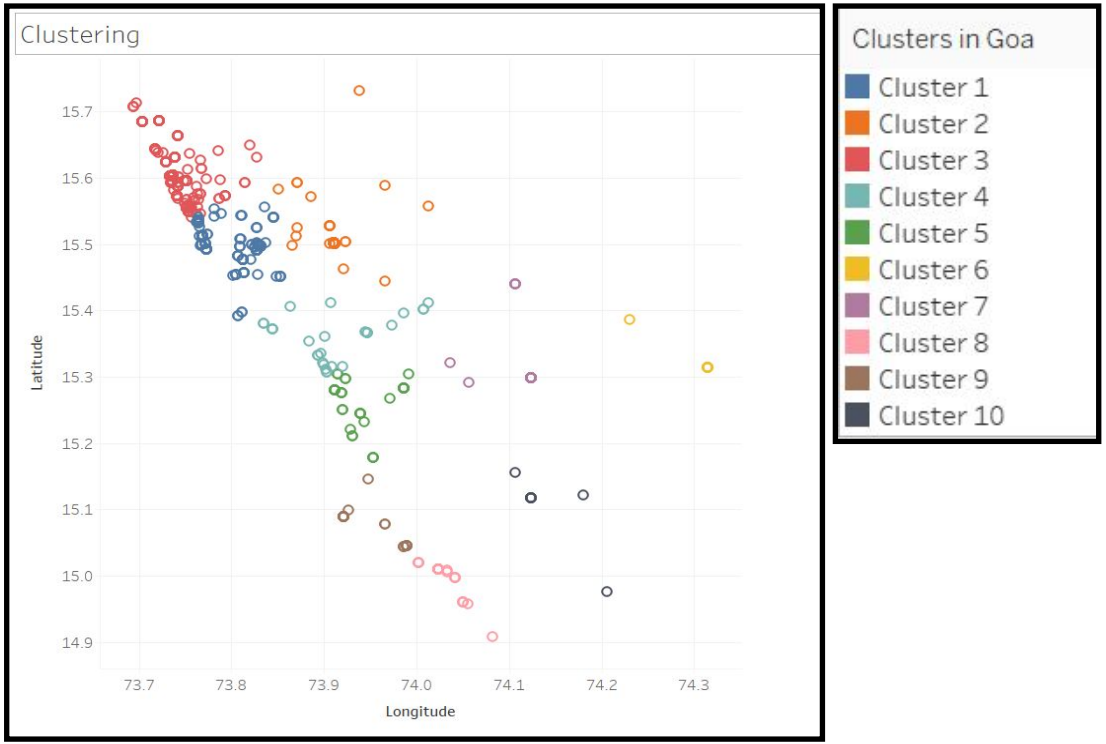
- Count of tourists visiting **Rajasthan** per year

Implementation and Analysis



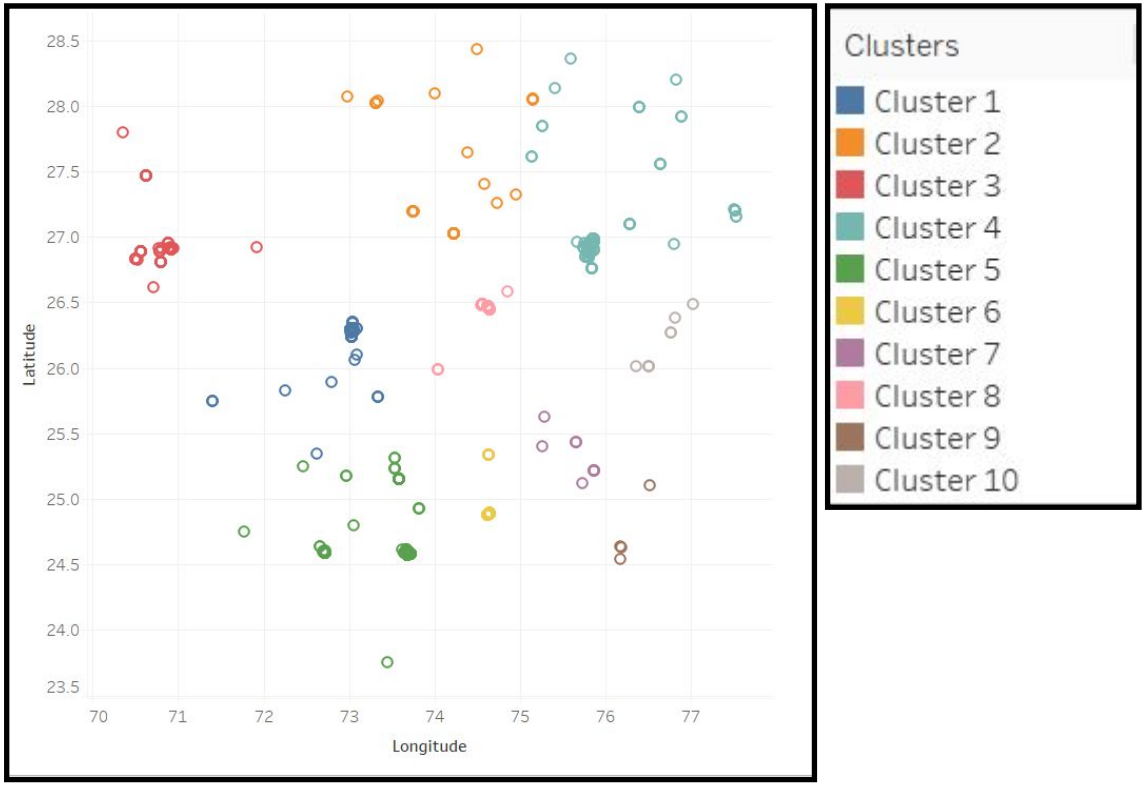
- Count of tourists visiting **Himachal Pradesh** per year

Implementation and Analysis



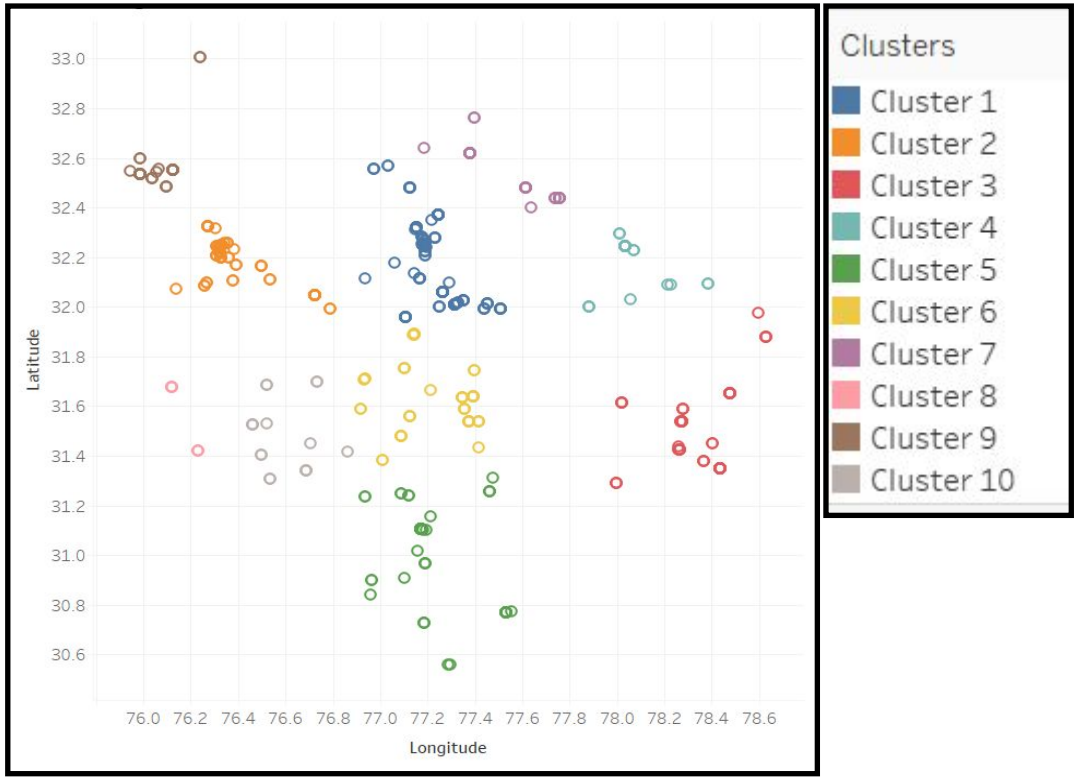
- Clustering of locations in **Goa** based on latitude and longitude

Implementation and Analysis



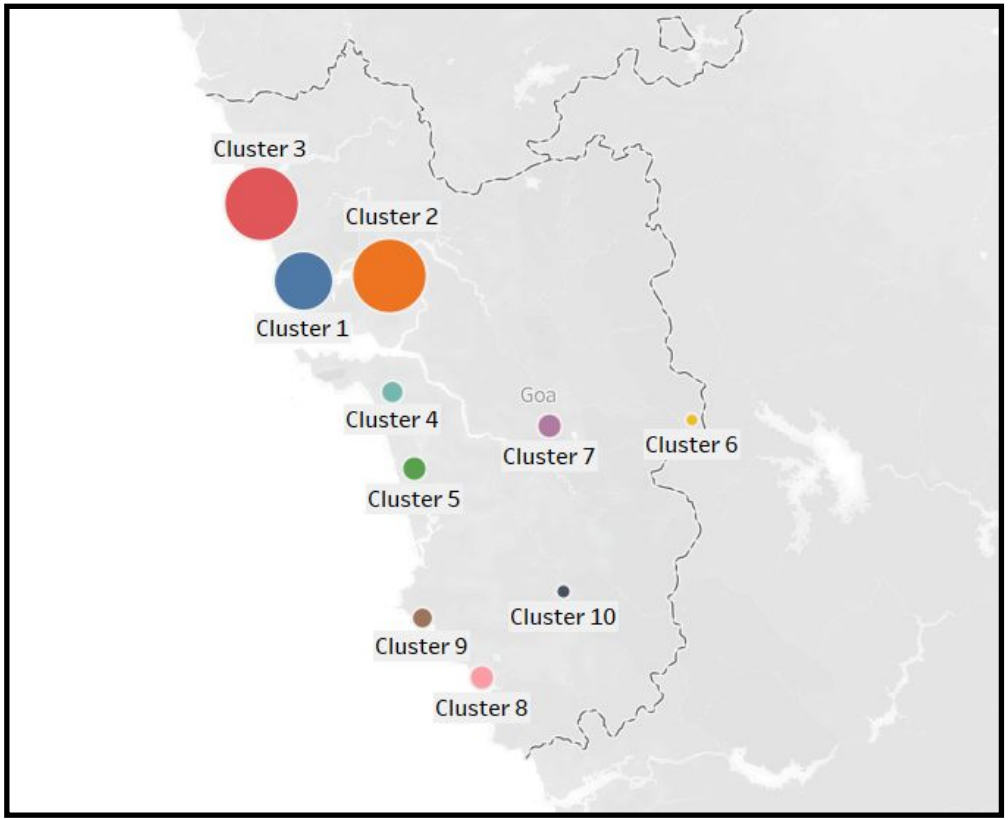
- Clustering of locations in **Rajasthan** based on latitude and longitude

Implementation and Analysis



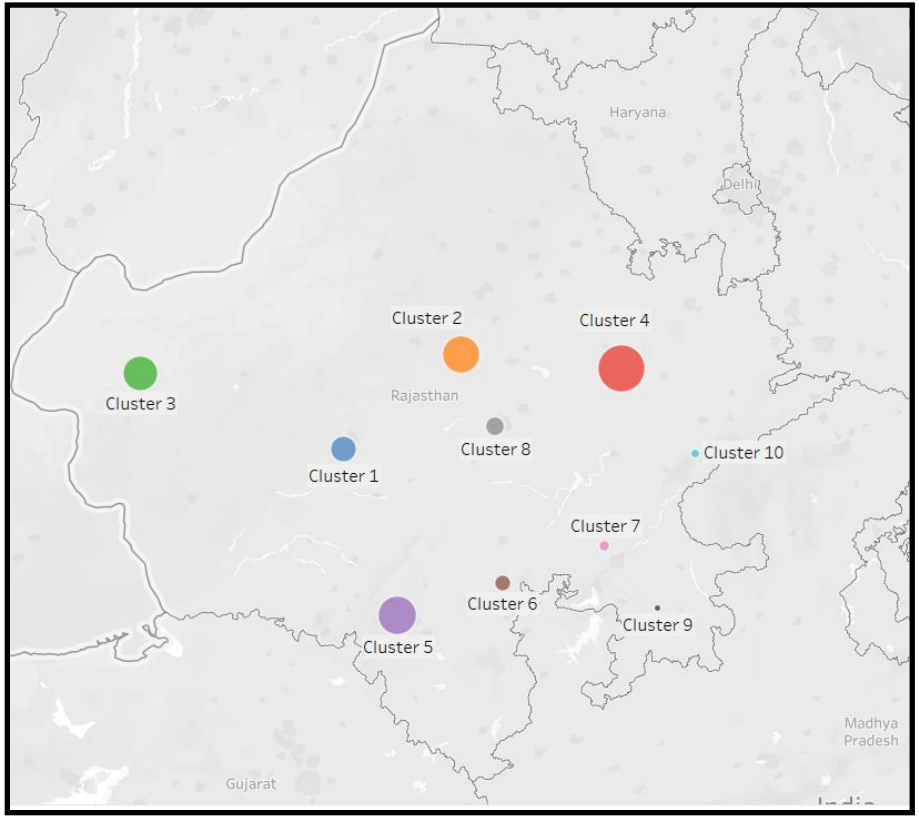
- Clustering of locations in **Himachal Pradesh** based on latitude and longitude

Implementation and Analysis



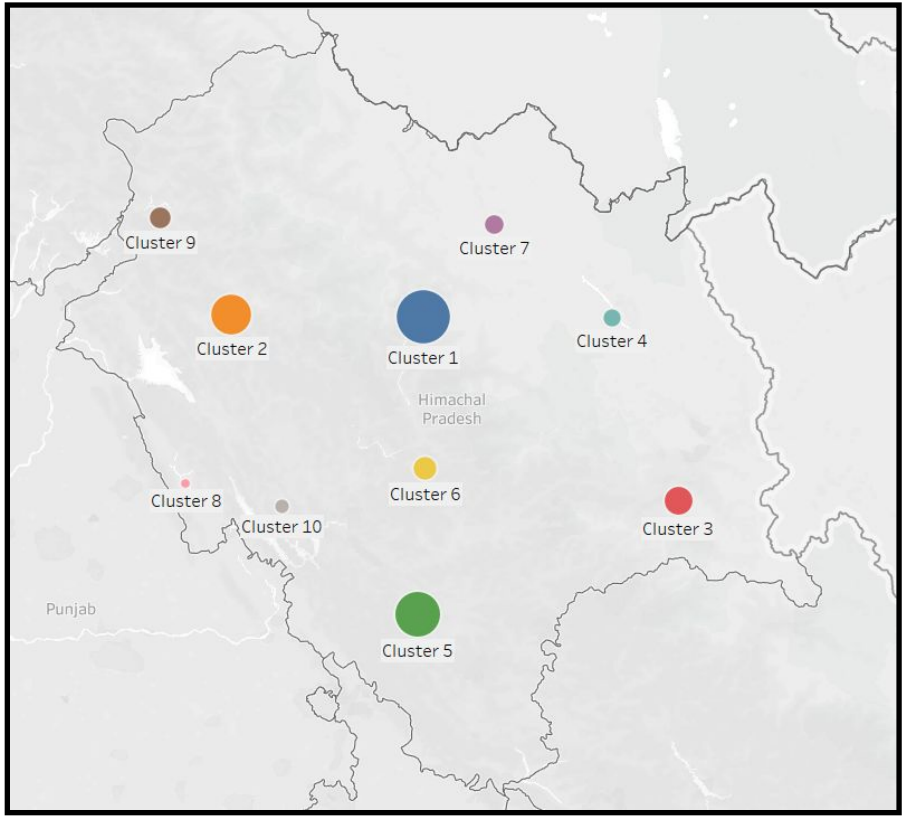
Clusters in Goa	Latitude	Longitude	Strength	No of Locations
Cluster 1	15.5	73.8	226.0	50.0
Cluster 2	15.5	73.9	359.0	19.0
Cluster 3	15.6	73.7	362.0	61.0
Cluster 4	15.4	73.9	28.0	22.0
Cluster 5	15.3	73.9	35.0	14.0
Cluster 6	15.3	74.3	8.0	2.0
Cluster 7	15.3	74.1	35.0	4.0
Cluster 8	15.0	74.0	34.0	8.0
Cluster 9	15.1	74.0	24.0	7.0
Cluster 10	15.1	74.1	10.0	4.0

Implementation and Analysis



Clusters in Rajasthan	Latitude	Longitude	Strength	Locations
Cluster 1	26.2	73.0	83.0	21.0
Cluster 2	27.1	74.2	181.0	15.0
Cluster 3	26.9	70.8	156.0	26.0
Cluster 4	27.0	75.9	292.0	50.0
Cluster 5	24.7	73.5	192.0	34.0
Cluster 6	25.0	74.6	30.0	5.0
Cluster 7	25.3	75.7	12.0	6.0
Cluster 8	26.4	74.6	42.0	11.0
Cluster 9	24.7	76.3	4.0	4.0
Cluster 10	26.2	76.7	8.0	5.0

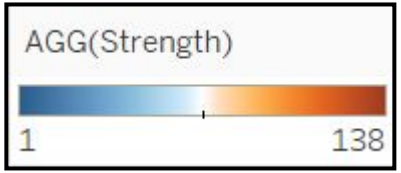
Implementation and Analysis



Clusters of Himachal	Latitude	Longitude	Strength	Locations count
Cluster 1	32.2	77.2	264.0	35.0
Cluster 2	32.2	76.4	148.0	27.0
Cluster 3	31.5	78.4	72.0	15.0
Cluster 4	32.2	78.1	27.0	8.0
Cluster 5	31.1	77.2	187.0	22.0
Cluster 6	31.6	77.2	48.0	18.0
Cluster 7	32.5	77.5	32.0	7.0
Cluster 8	31.6	76.2	7.0	2.0
Cluster 9	32.6	76.1	40.0	9.0
Cluster 10	31.5	76.6	17.0	9.0

Results and Discussions

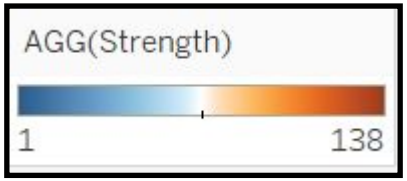
Clusters in Goa	Quarters Quarter1	Quarter2	Quarter3	Quarter4
Cluster 1	77 21.27%	39 17.73%	28 20.74%	82 20.30%
Cluster 2	121 33.43%	64 29.09%	52 38.52%	122 30.20%
Cluster 3	114 31.49%	78 35.45%	32 23.70%	138 34.16%
Cluster 4	8 2.21%	8 3.64%	2 1.48%	10 2.48%
Cluster 5	14 3.87%	10 4.55%	4 2.96%	7 1.73%
Cluster 6	2 0.55%	2 0.91%	1 0.74%	3 0.74%
Cluster 7	6 1.66%	6 2.73%	7 5.19%	16 3.96%
Cluster 8	8 2.21%	5 2.27%	7 5.19%	14 3.47%
Cluster 9	8 2.21%	7 3.18%	1 0.74%	8 1.98%
Cluster 10	4 1.10%	1 0.45%	1 0.74%	4 0.99%
Grand Total	362 100.00%	220 100.00%	135 100.00%	404 100.00%



- Clusterwise analysis of tourists visiting **Goa** in each quarter

Results and Discussions

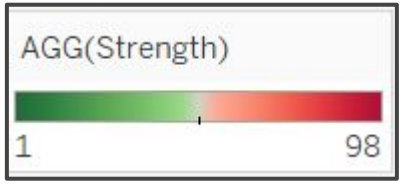
Clusters in Goa	Quarters Quarter1	Quarter2	Quarter3	Quarter4	Grand Total
Cluster 1	77 34.07%	39 17.26%	28 12.39%	82 36.28%	226 100.00%
Cluster 2	121 33.70%	64 17.83%	52 14.48%	122 33.98%	359 100.00%
Cluster 3	114 31.49%	78 21.55%	32 8.84%	138 38.12%	362 100.00%
Cluster 4	8 28.57%	8 28.57%	2 7.14%	10 35.71%	28 100.00%
Cluster 5	14 40.00%	10 28.57%	4 11.43%	7 20.00%	35 100.00%
Cluster 6	2 25.00%	2 25.00%	1 12.50%	3 37.50%	8 100.00%
Cluster 7	6 17.14%	6 17.14%	7 20.00%	16 45.71%	35 100.00%
Cluster 8	8 23.53%	5 14.71%	7 20.59%	14 41.18%	34 100.00%
Cluster 9	8 33.33%	7 29.17%	1 4.17%	8 33.33%	24 100.00%
Cluster 10	4 40.00%	1 10.00%	1 10.00%	4 40.00%	10 100.00%



- Quarter wise analysis of tourists visiting locations belonging to respective clusters in **Goa**

Results and Discussions

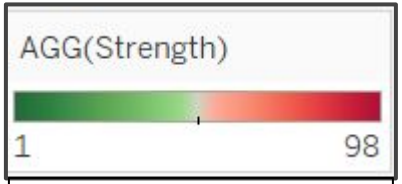
Clusters in Rajasthan	Quarters Quarter1	Quarter2	Quarter3	Quarter4
Cluster 1	7.37% 25	8.41% 26	9.02% 12	9.13% 20
Cluster 2	18.29% 62	23.95% 74	17.29% 23	10.05% 22
Cluster 3	16.81% 57	11.65% 36	11.28% 15	21.92% 48
Cluster 4	28.91% 98	27.83% 86	36.09% 48	27.40% 60
Cluster 5	18.29% 62	21.68% 67	18.05% 24	17.81% 39
Cluster 6	2.06% 7	2.27% 7	4.51% 6	4.57% 10
Cluster 7	1.77% 6	1.29% 4	1.50% 2	
Cluster 8	5.01% 17	1.62% 5	1.50% 2	8.22% 18
Cluster 9	0.59% 2	0.65% 2		
Cluster 10	0.88% 3	0.65% 2	0.75% 1	0.91% 2
Grand Total	100.00% 339	100.00% 309	100.00% 133	100.00% 219



- Clusterwise analysis of tourists visiting **Rajasthan** in each quarter

Results and Discussions

Clusters in Rajasthan	Quarters Quarter1	Quarter2	Quarter3	Quarter4	Grand Total
Cluster 1	30.12% 25	31.33% 26	14.46% 12	24.10% 20	100.00% 83
Cluster 2	34.25% 62	40.88% 74	12.71% 23	12.15% 22	100.00% 181
Cluster 3	36.54% 57	23.08% 36	9.62% 15	30.77% 48	100.00% 156
Cluster 4	33.56% 98	29.45% 86	16.44% 48	20.55% 60	100.00% 292
Cluster 5	32.29% 62	34.90% 67	12.50% 24	20.31% 39	100.00% 192
Cluster 6	23.33% 7	23.33% 7	20.00% 6	33.33% 10	100.00% 30
Cluster 7	50.00% 6	33.33% 4	16.67% 2		100.00% 12
Cluster 8	40.48% 17	11.90% 5	4.76% 2	42.86% 18	100.00% 42
Cluster 9	50.00% 2	50.00% 2			100.00% 4
Cluster 10	37.50% 3	25.00% 2	12.50% 1	25.00% 2	100.00% 8



- Quarter wise analysis of tourists visiting locations belonging to respective clusters in **Rajasthan**

Results and Discussions

Cluste	Quarters Quarter1	Quarter2	Quarter3	Quarter4
Cluster 1	33.25% 126	35.43% 79	26.05% 31	23.14% 28
Cluster 2	12.93% 49	16.14% 36	10.92% 13	41.32% 50
Cluster 3	5.54% 21	12.56% 28	8.40% 10	10.74% 13
Cluster 4	4.49% 17	2.69% 6	1.68% 2	1.65% 2
Cluster 5	24.54% 93	16.14% 36	36.13% 43	12.40% 15
Cluster 6	7.12% 27	3.14% 7	5.88% 7	5.79% 7
Cluster 7	3.69% 14	5.38% 12	4.20% 5	0.83% 1
Cluster 8	1.06% 4	0.90% 2		0.83% 1
Cluster 9	6.33% 24	3.14% 7	6.72% 8	0.83% 1
Cluster 10	1.06% 4	4.48% 10		2.48% 3
Grand Total	100.00% 379	100.00% 223	100.00% 119	100.00% 121

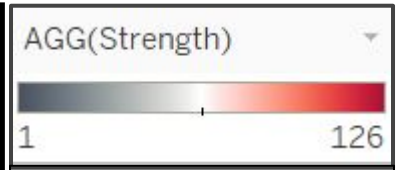
AGG(Strength)

1 126

- Clusterwise analysis of tourists visiting **Himachal Pradesh** in each quarter

Results and Discussions

Clusters ..	Quarters Quarter1	Quarter2	Quarter3	Quarter4	Grand To..
Cluster 1	47.73% 126	29.92% 79	11.74% 31	10.61% 28	100.00% 264
Cluster 2	33.11% 49	24.32% 36	8.78% 13	33.78% 50	100.00% 148
Cluster 3	29.17% 21	38.89% 28	13.89% 10	18.06% 13	100.00% 72
Cluster 4	62.96% 17	22.22% 6	7.41% 2	7.41% 2	100.00% 27
Cluster 5	49.73% 93	19.25% 36	22.99% 43	8.02% 15	100.00% 187
Cluster 6	56.25% 27	14.58% 7	14.58% 7	14.58% 7	100.00% 48
Cluster 7	43.75% 14	37.50% 12	15.63% 5	3.13% 1	100.00% 32
Cluster 8	57.14% 4	28.57% 2		14.29% 1	100.00% 7
Cluster 9	60.00% 24	17.50% 7	20.00% 8	2.50% 1	100.00% 40
Cluster 10	23.53% 4	58.82% 10		17.65% 3	100.00% 17



- Quarter wise analysis of tourists visiting locations belonging to respective clusters in **Himachal Pradesh**

Conclusion

- Our aim to help the tourism industry revive, which suffered huge losses in the pandemic last year, has been fulfilled.
- The analysis presented is constructed from every minute data available of the previously visited tourists.

  Our analysis presents that:

State	Popular areas	Best Quarters
Goa	North Goa	Quarter 1 Quarter 4
Rajasthan	Central and Southern Rajasthan	Quarter 1 Quarter 2
HImachal Pradesh	Central and Southern Himachal Pradesh	Quarter 1 Quarter 2

Future Scope & Expansion

- Comprehensive review system based on live experiences extracted from social media which will help authorities and businessmen in the area to understand the needs, discomforts of the people and take necessary actions.

- Web Application that works on automated data extraction of the user entered tourist attraction from various sources and generates trends and suggestions regarding the same.

THANK YOU