Tourism Market Segmentation Report – India

---

1. Introduction

This report presents a data-driven segmentation and analysis of **325 top tourist destinations in India**. It leverages key variables such as geographic location, cultural significance, public sentiment (Google reviews), accessibility (airport proximity, camera policies), and operational details (entry fee, best time to visit, and weekly offs).

Advanced techniques like **KMeans clustering** have also been used to identify distinct groups of tourist places based on visitor interest and convenience.

---

2. Dataset Overview

- **Total Entries**: 325 places
- **Geographic Fields**: Zone, State, City
- **Attraction Features**: Type, Significance, Establishment Year
- **Visitor Logistics**:
    - Google Review Rating (1–5 scale)
    - Number of Reviews (in lakhs)
    - Entry Fee in INR
    - Time Needed to Visit (in hours)
    - Weekly Off
    - Best Time to Visit
    - DSLR Allowed or Not
    - Proximity to Airport (within 50km)

This structured dataset supports deep segmentation for travel tech apps, tourism planning, and marketing efforts.

---

3. Data Preprocessing for Clustering

Before performing clustering, the dataset needed to be cleaned and encoded to ensure compatibility with KMeans (which only works with numerical data). The following steps were taken:

1. **Missing Value Handling:**
   Rows with missing values in *State* or *Google review rating* were dropped, as these fields are essential for clustering.
2. **Label Encoding:**
    - *State* and *Zone* columns were converted into numerical format using Label Encoding.
    - This was necessary to allow KMeans to calculate distances between categorical values.
3. **Feature Selection:**
   Only relevant features — *Encoded State* and *Google Review Rating* — were selected for clustering. *Zone* was encoded and preserved for optional future analysis.
4. **Data Normalization (Optional, Not Used):**
   Normalization wasn't applied here since encoded states are already scaled in a limited range, and ratings are on a 1–5 scale.
   If more numerical variables were used, standardization would be advised.

---

4. Market Segmentation using KMeans Clustering

Objective:

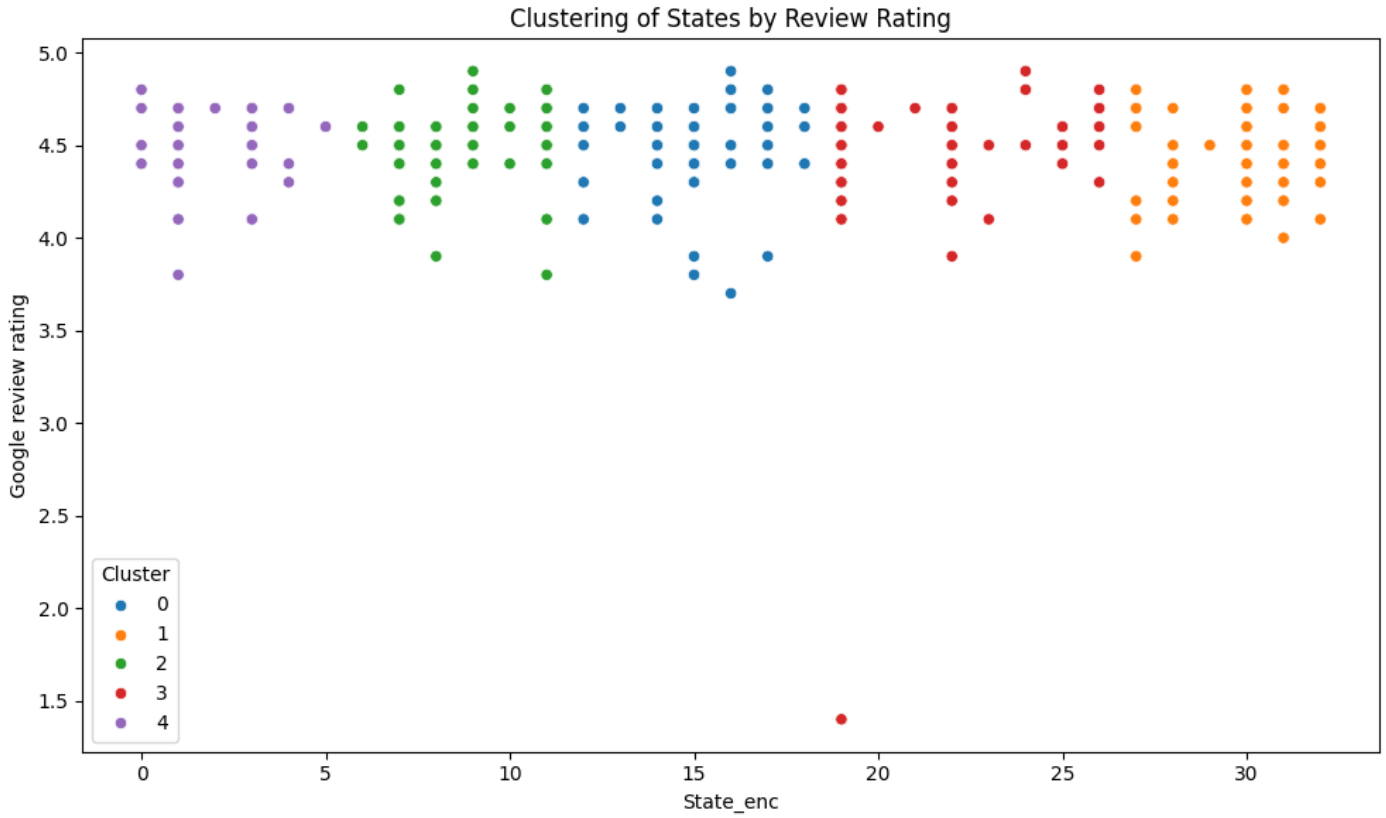
To segment the market based on geographical location (*State*) and customer sentiment (*Google Review Rating*), helping identify target regions for marketing, service improvement, and expansion.

Approach:

- The dataset was encoded to convert *State* and *Zone* into numerical values for clustering.
- KMeans clustering was applied on *State* and *Google review rating* to group regions with similar satisfaction levels.
- 5 clusters were created to differentiate high and low-performing regions.

Key Insights:

1. **Segmented User Experience:**
   The clusters reveal states with high, moderate, and low customer satisfaction. This helps tailor region-specific strategies.
2. **Data-Driven Expansion Planning:**
   Clusters with low ratings can be flagged for experience improvement, while high-performing clusters can be used as case studies or marketing showcases.
3. **Cluster-wise Strategy Design:**
   Each cluster serves as a distinct user base. TripBuddy can plan targeted promotions, UI tweaks, or support campaigns based on the behaviour of each segment.



5. Geographic Segmentation

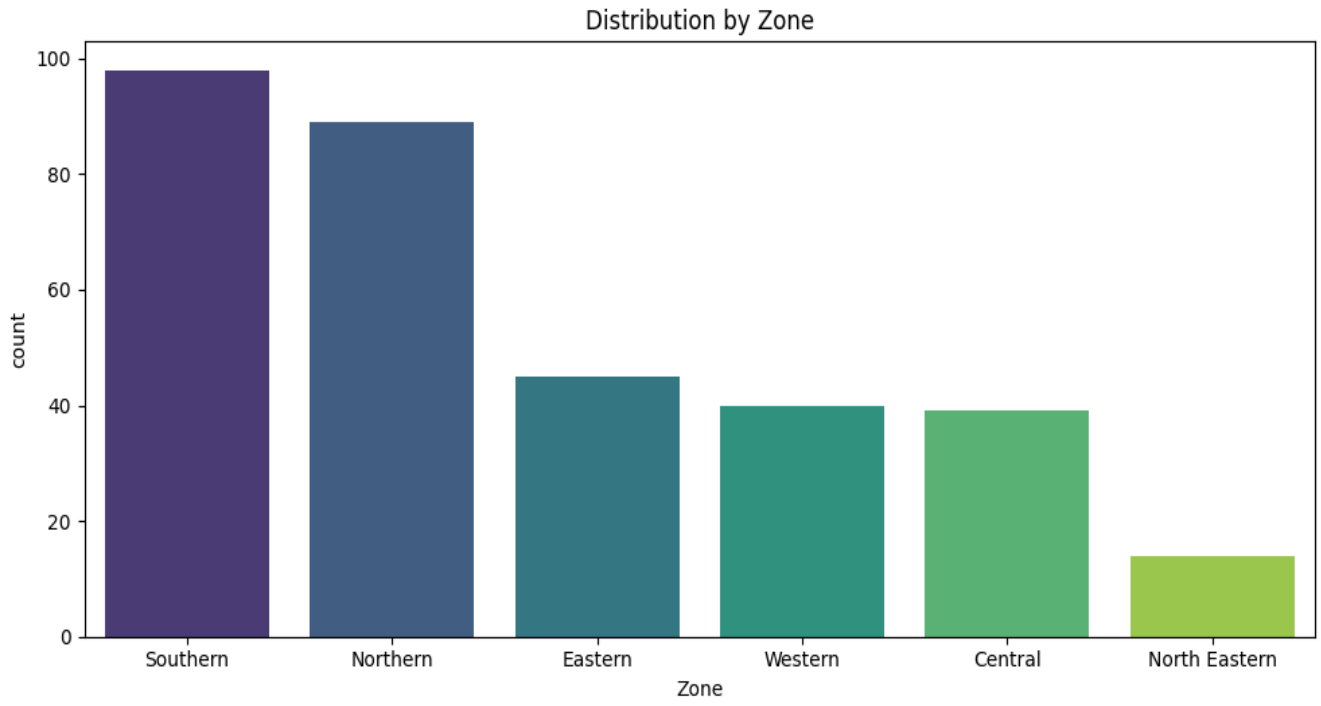
Zone-Wise Distribution:

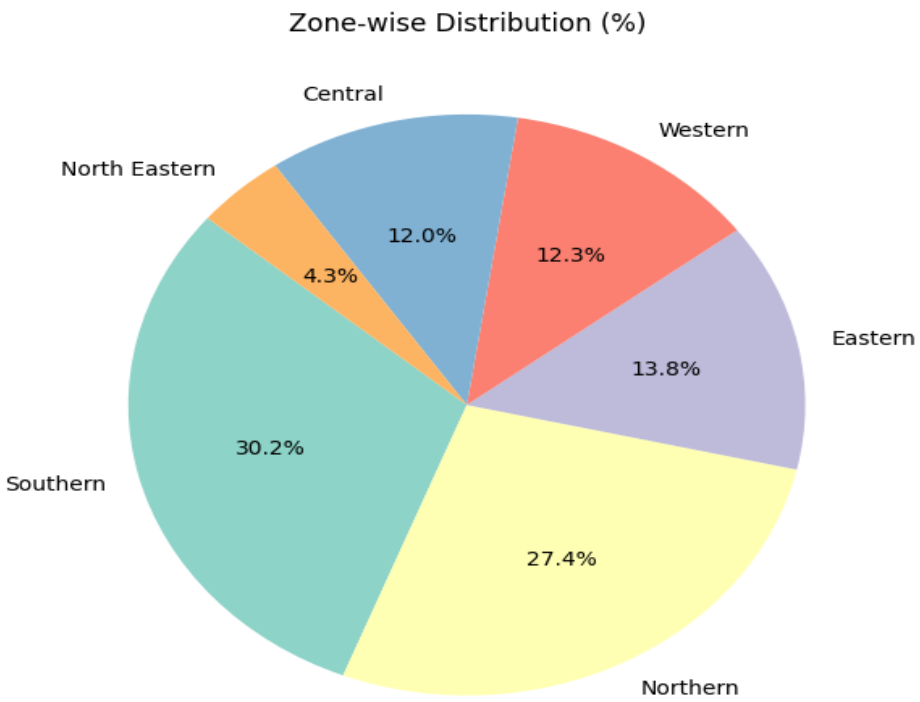
- The **Northern Zone** has the most destinations, driven by heritage-rich states like **Delhi, Uttar Pradesh, and Rajasthan**.
- **Southern and Western Zones** also show strong tourism density, with Tamil Nadu and Maharashtra being notable contributors.

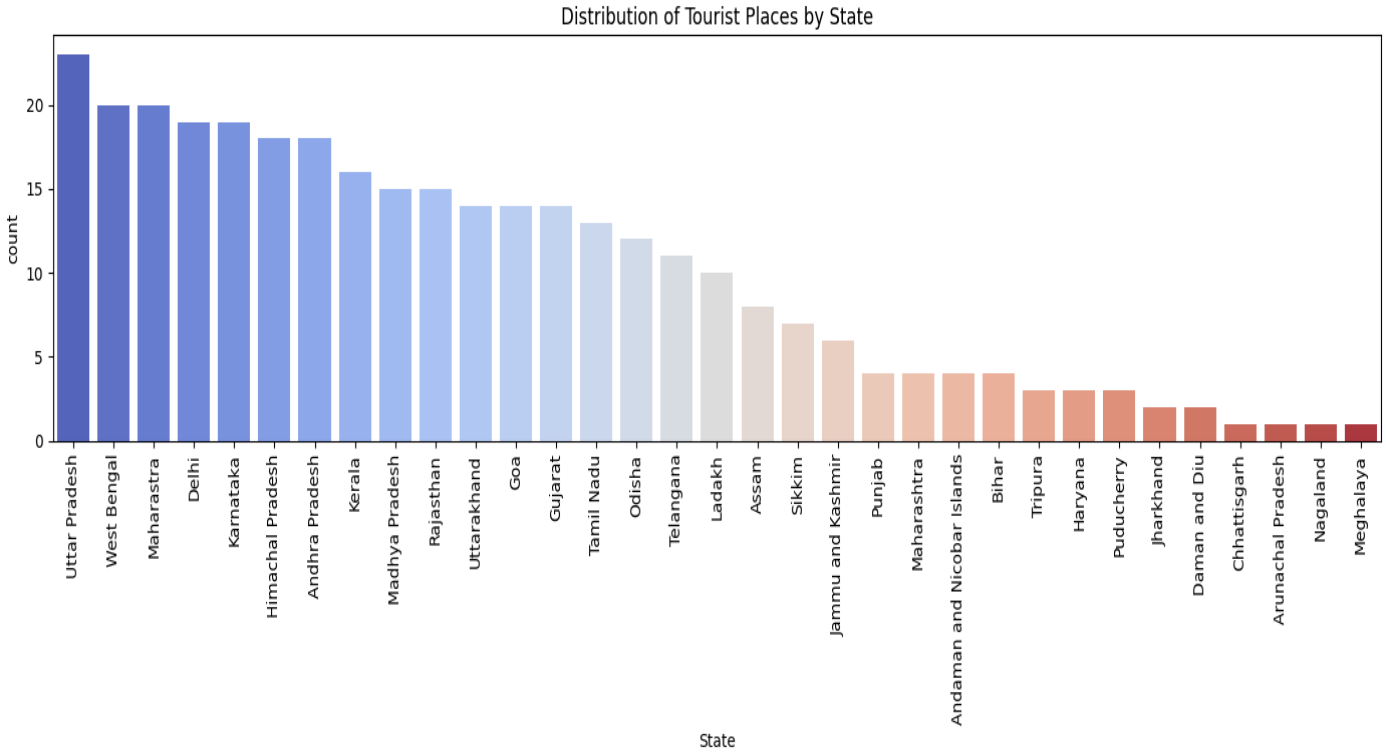
Top 5 States by Number of Places:

1. Delhi
2. Maharashtra
3. Tamil Nadu
4. Uttar Pradesh
5. Rajasthan

These regions collectively shape the core of India's tourism ecosystem, offering a diverse mix of culture, architecture, spirituality, and modernity.







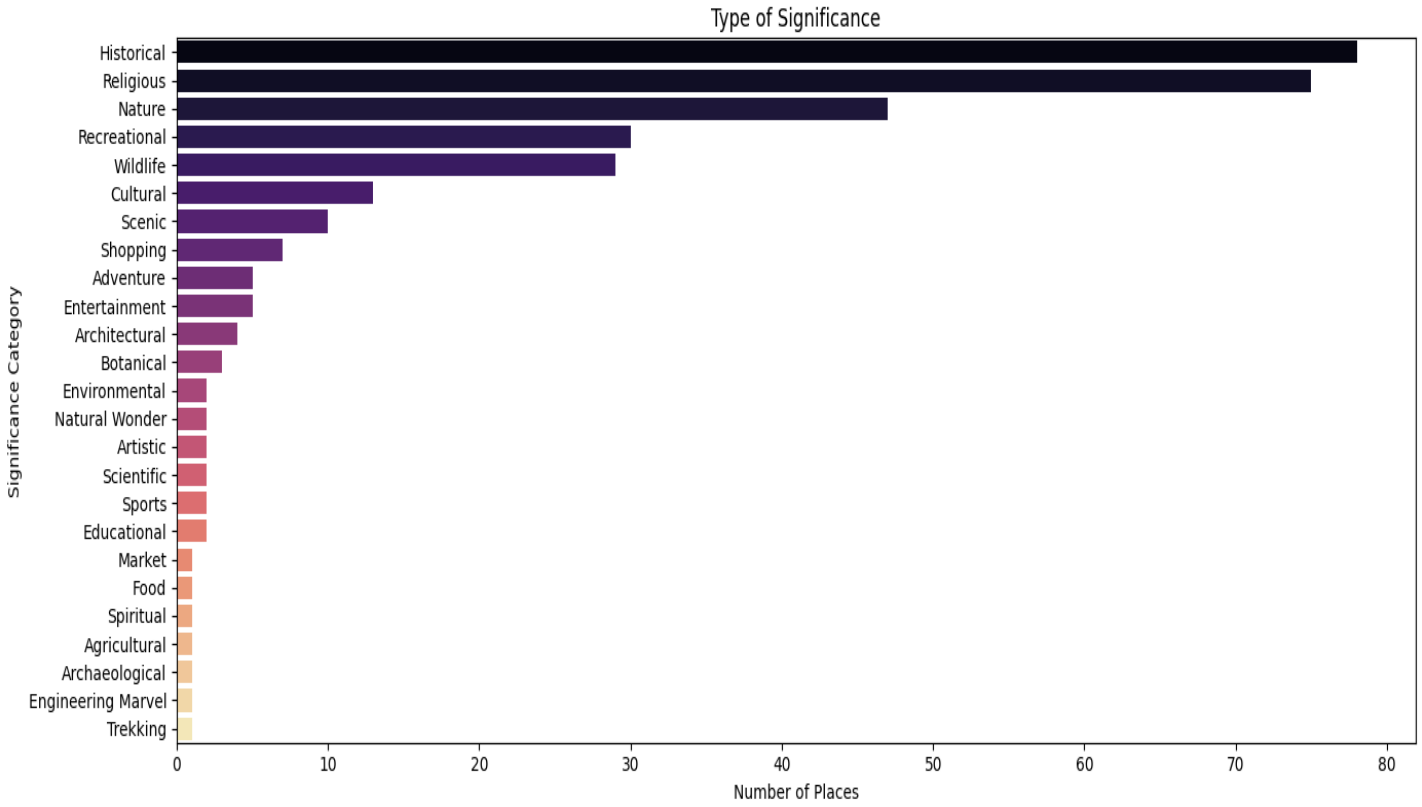
---

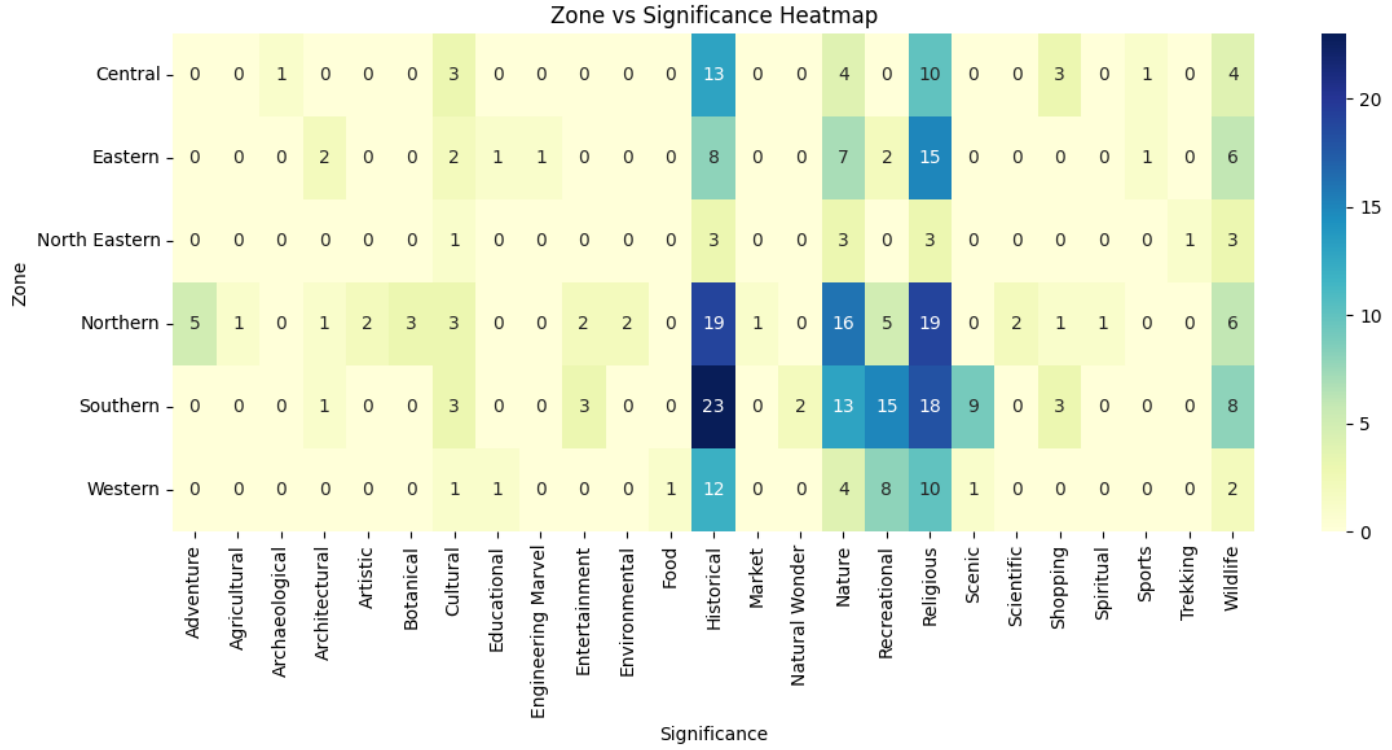
6. Significance-Based Segmentation

Tourist spots are grouped by their dominant **theme or purpose**:

- **Historical** – Forts, palaces, and monuments from India's rich past
- **Religious** – Temples, churches, mosques, and pilgrimage centers
- **Environmental** – Parks, sanctuaries, eco-reserves
- **Scientific** – Observatories and museums
- **Theme Parks** and modern recreational sites

Historical and Religious sites form the majority, showing the strong cultural foundation of tourism in India.





---

7. Review-Based Clustering (KMeans)

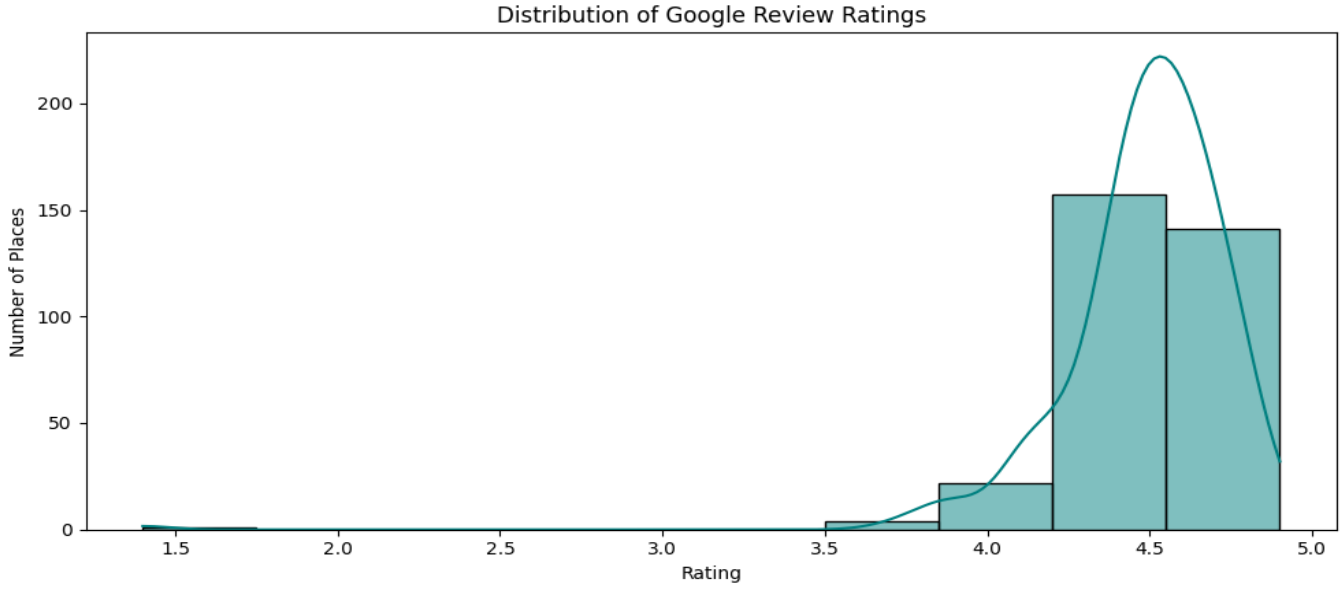
Based on **Google review rating** and **number of reviews**, four distinct segments emerged:

Segment 0: Spiritual Pilgrimage Spots

Key Characteristics:

- 89.6% of places are religious sites
- 83.6% marked as "Anytime" for visiting
- 98.5% have no weekly off

Interpretation:
These are year-round, easily accessible spiritual sites such as temples and shrines. They cater to devotees and spiritual tourists across seasons.



Segment 1: Tourist-Friendly, Well-Connected Attractions

Key Characteristics:

- 99.3% are within 50km of an airport
- 98.6% allow DSLR cameras
- 82.1% have no weekly off

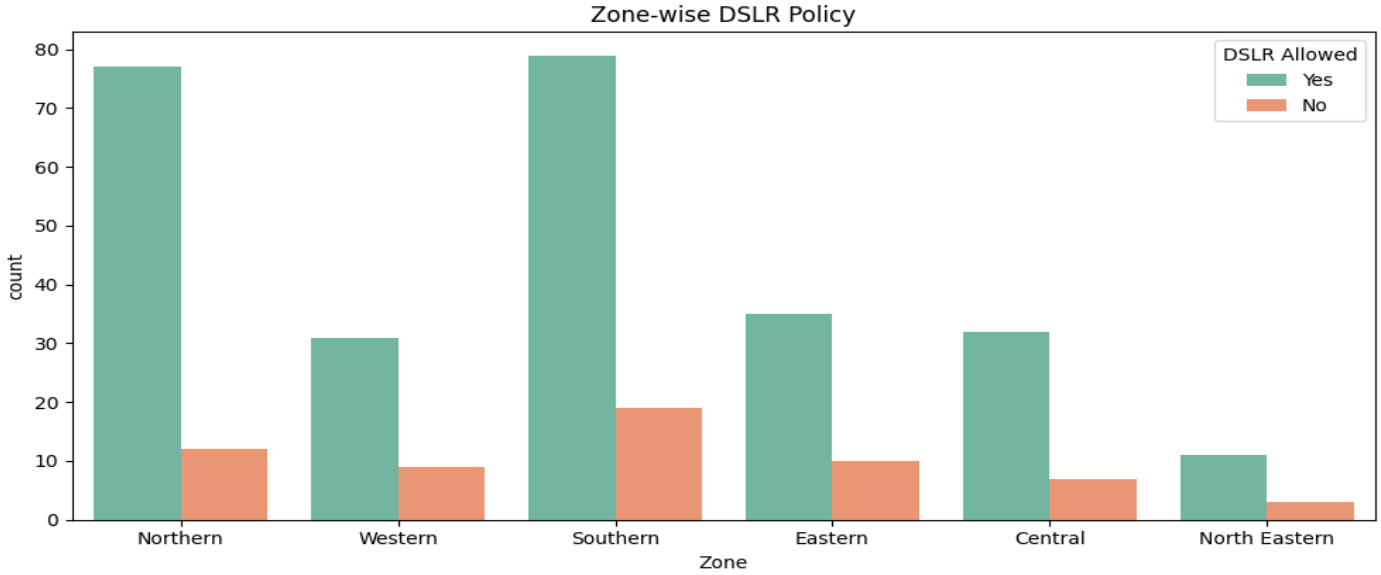
Interpretation:
These are highly accessible, camera-friendly, well-supported tourist sites. Likely located in urban areas or popular cities. Ideal for international tourists and city-based explorers.

Segment 2: Scenic Spots Allowing Photography

Key Characteristics:

- 98.6% allow DSLR usage
- 95.9% are far from airports (more than 50km)

Interpretation:
These are remote yet visually rich attractions — possibly nature trails, offbeat locations, or scenic parks. They appeal to photographers, backpackers, and off-grid travellers.



Segment 3: High-Engagement Tourist Spots
Key Characteristics:

- 90.9% have no weekly off
- 97% allow DSLR cameras
- Average of 2.19 lakh reviews

Interpretation:
These are high-traffic, iconic destinations with major online engagement. Likely well-known landmarks, monuments, or family-friendly attractions that are always open and welcoming to tourists.

Segment 4: Paid Entry Attractions with DSLR Access
Key Characteristics:

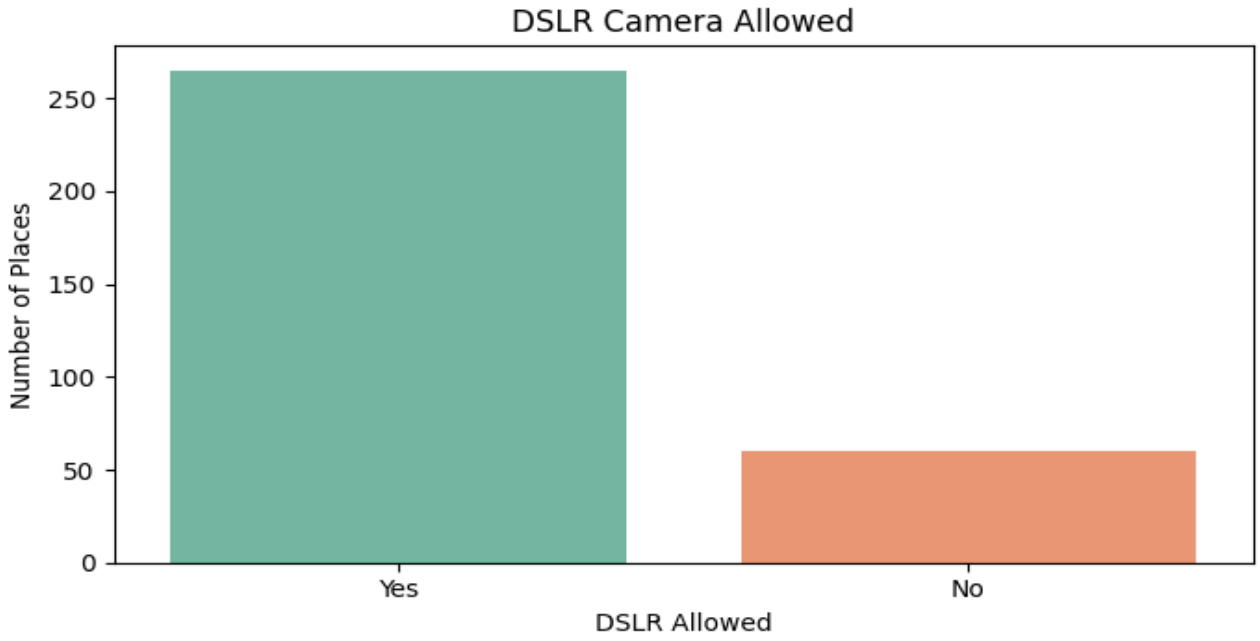
- 100% allow DSLR photography
- Average entrance fee: ₹3.74
- Average visit duration: ~2.5 hours

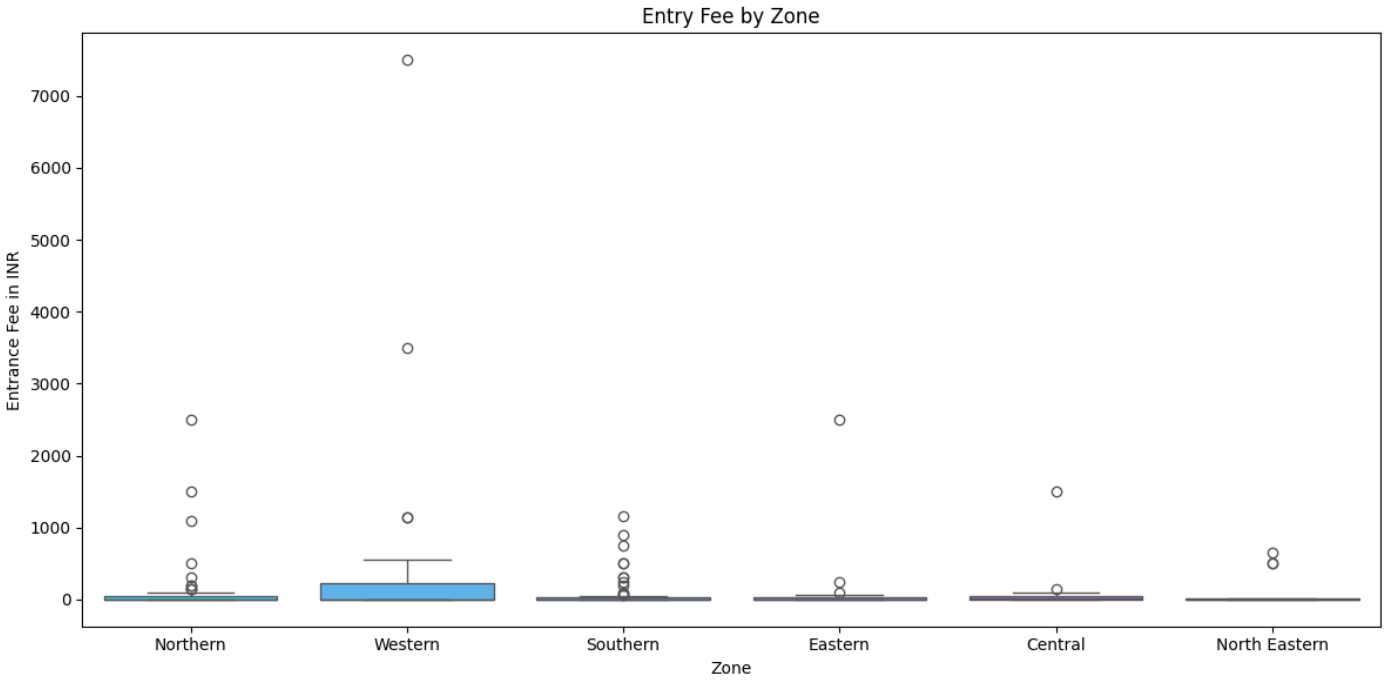
Interpretation:
These are structured, content-rich sites like museums, guided tours, art galleries, or historical palaces. Often visited by students, international tourists, and knowledge seekers.

---

8. Accessibility & Amenities Overview

- **DSLR Allowed**: 265 places (Yes), 60 places (No)
- **Airport Proximity**: Most tourist places are within 50 km of an airport
- **Entrance Fee**: Varies by site — many are free; some have minor fees
- **Time to Visit**: Ranges from quick (30 mins) to long (up to 6+ hrs)
- **Best Time to Visit**: Evenly split across Morning, Afternoon, Evening, and "Anytime"


DSLR Camera Allowed
Number of Places
250
200
150
100
50
0
Yes
No
DSLR Allowed


Entry Fee by Zone
Entrance Fee in INR
7000
6000
5000
4000
3000
2000
1000
0
Northern
Western
Southern
Eastern
Central
North Eastern
Zone

9. Conclusion & Strategic Suggestions

This report delivers a structured segmentation of India's tourism landscape using clustering analysis and key visitor-centric variables. Through the application of KMeans clustering on geo-sentiment data (State and Google Review Rating), five distinct segments of tourist destinations were identified, each showcasing unique behavioral and accessibility patterns.
Key findings indicate that:

- **Religious and Historical attractions** dominate India's travel ecosystem, reflecting both cultural heritage and visitor interest.
- **Public engagement**, as captured by Google review ratings and counts, serves as a strong indicator of destination popularity and traveler satisfaction.
- **Accessibility parameters** such as airport proximity and DSLR policy play a critical role in shaping a tourist's experience and decision-making.
- **Clustering analysis** helps reveal patterns in traveler preference, making it easier to distinguish between over-visited hotspots, hidden gems, and niche attractions.

These insights can guide decision-making for tourism operators, government planners, and AI-based platforms like TripBuddy to improve service delivery, marketing, and infrastructure investments.

Strategic Recommendations

1. **Promote Underrated High-Rated Spots (Segment 1):**
   Focus marketing and outreach on tourist-friendly but less-popular attractions with high ratings. These locations offer strong potential to offload crowd pressure from iconic destinations.
2. **Enhance Infrastructure for Scenic Remote Sites (Segment 2):**
   Improve road access, offer shuttle services, or integrate virtual previews and 360° online walkthroughs to attract travellers hesitant to visit remote scenic locations.
3. **Integrate Segments into Recommendation Systems:**
   Use cluster insights to personalize destination suggestions in travel platforms like TripBuddy AI, offering smarter, user-centric itinerary planning.
4. **Upgrade Amenities in Cultural and Educational Sites (Segment 4):**
   Invest in modern visitor infrastructure—digital guides, e-ticketing, multilingual support—especially for museums, galleries, and historical sites that cater to students and international travellers.
5. **Boost Local Partnerships for Sustainable Tourism:**
   Collaborate with local communities, guides, and artisans to create immersive travel experiences and ensure that tourism benefits are distributed at the grassroots level.
6. **Monitor and Re-Segment Periodically Using Fresh Data:**
   Establish a framework for continuous feedback analysis and re-clustering to adapt to changing travel trends, ensuring agility and relevance in campaign planning.