

• • • •

Analyzing Tourists' Perceptions of Hotel Location: Text Analytics Using Online Reviews



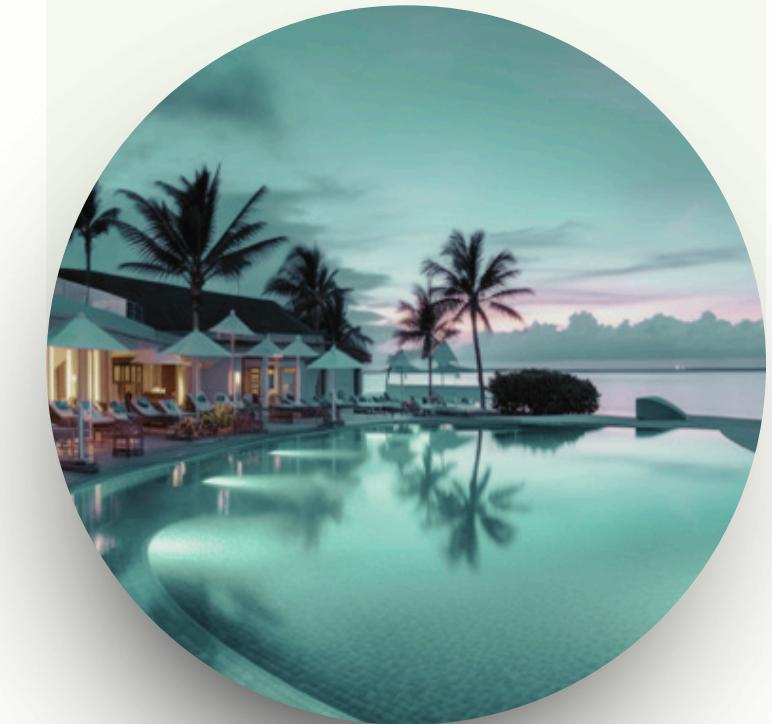
H.P.S.B. Gunarathna
Department of Decision Science

Supervisors :
Dr. Indra Mahakalanda
Dr. Sandun M. Dassanayake



Introduction

- Traditionally, hotel location selection has been approached from the hotelier's perspective, focusing on factors that benefit the business. However, a crucial gap exists in understanding how guests perceive hotel locations based on their travel goals and preferences.
- Travel agencies and hotels can give future guests a more accurate picture of where they'll be staying by understanding how past guests have described the location.



Objective 1

To conduct text analytics on reviews of tourist hotels in Sri Lanka.

Objective 2

To provide recommendations for developing marketing strategies for tourist hotels

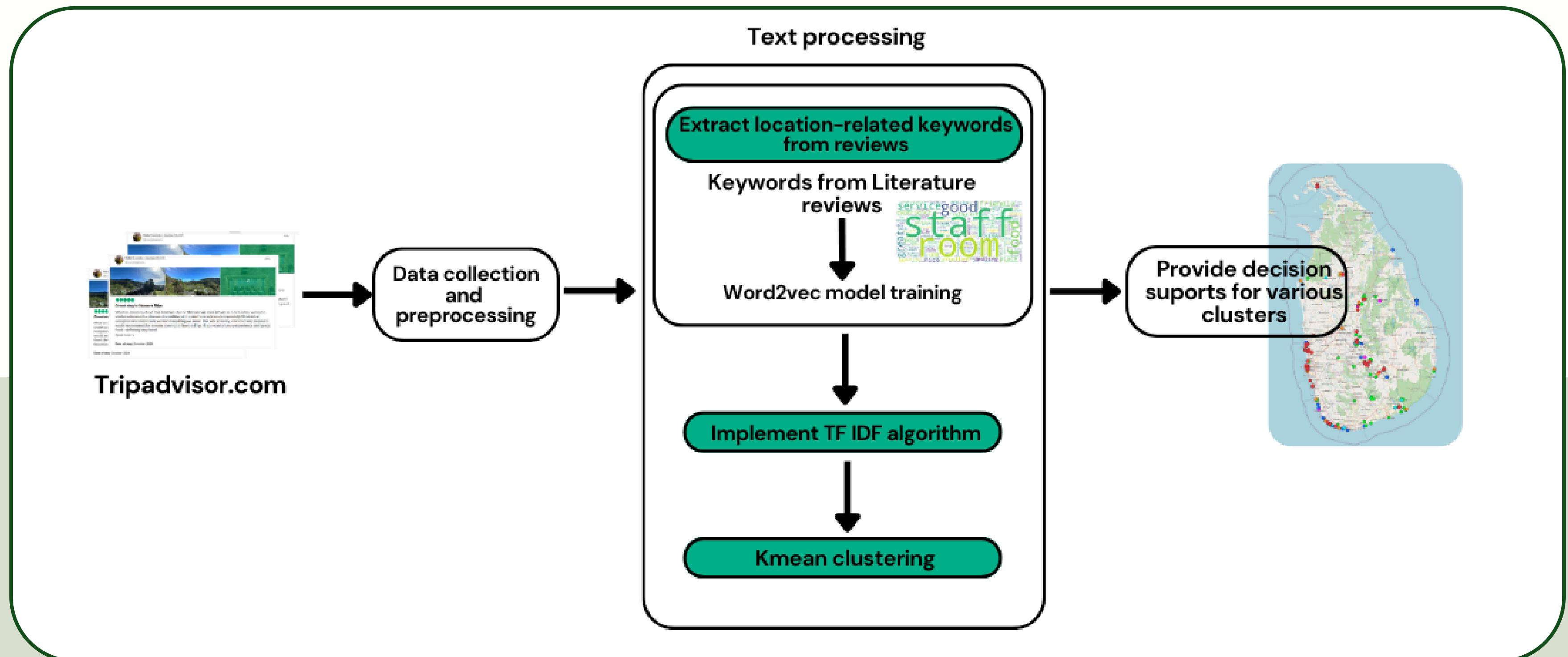


Significance of the study

- The study's findings offer valuable insights for future hoteliers seeking to identify optimal hotel locations aligned with their target markets.
- Understanding the preferences and priorities of their target clientele, hotel operators can utilize the study's findings to develop tailored promotion strategies.
- Online Travel Agencies (OTAs) can utilize this information to personalize recommendation systems, inspiring guests to discover hotels that best suit their needs.

Methodology

Data was collected from TripAdvisor by web scraping the customer reviews from August 2023 to August 2024. A total of 30666 reviews from 266 hotels were used for the analysis.



Results and Discussion



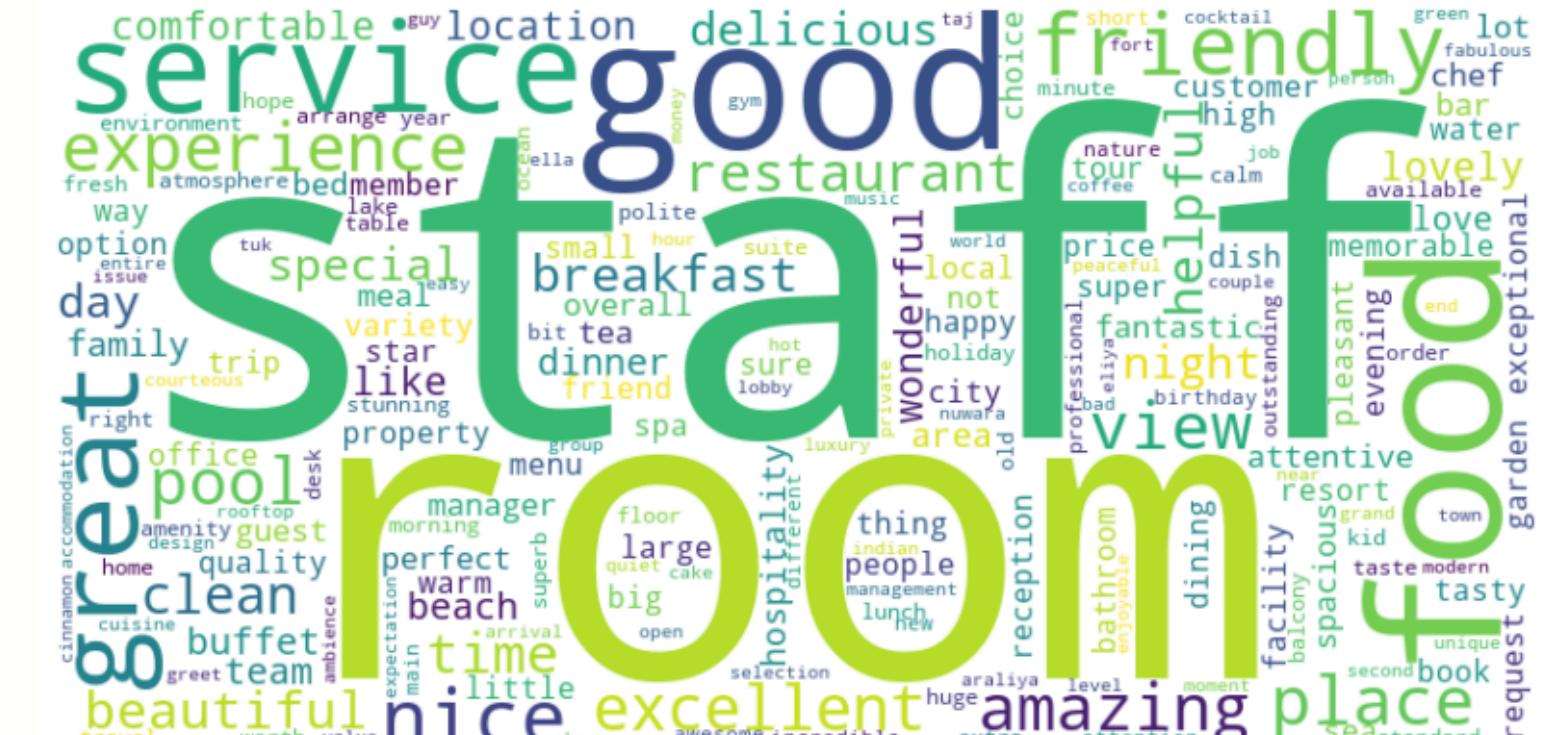
Frequency analysis

A total of 3815 words were extracted from the reviews . This list includes the hotel guest satisfaction dimensions that most affect guest satisfaction such as room, food, service, clean, and location as indicated by Dong et al., 2014

Table 2: Top 80 words in hotel quest reviews

Word	Freq.	Word	Freq.	Word	Freq.	Word	Freq.
staff	24088	night	4843	guest	2767	warm	1724
room	20730	day	4829	spacious	2639	resort	1698
good	19488	special	4744	chef	2637	trip	1695
food	18874	delicious	4703	manager	2538	tasty	1676
service	16789	comfortable	4572	overall	2490	book	1630
great	13350	like	4529	attentive	2439	facility	1624
friendly	10568	location	4433	fantastic	2373	customer	1607
nice	8873	lovely	4429	lot	2249	happy	1604
experience	8851	wonderful	4282	bed	2211	city	1596
place	8725	team	3972	meal	2144	large	1582
view	8462	buffet	3956	high	2016	friend	1578
pool	8149	beach	3920	property	1989	pleasant	1544
amazing	8120	dinner	3891	super	1982	memorable	1541
excellent	7757	family	3555	variety	1921	bathroom	1539
restaurant	7355	love	3343	way	1892	quality	1537
time	7207	hospitality	3239	reception	1880	option	1527
clean	6576	area	3210	exceptional	1874	small	1508
beautiful	6211	tea	2974	big	1859	local	1488
helpful	6143	perfect	2861	sure	1831	evening	1477
breakfast	5543	bar	2773	people	1726	member	1449

Attributes of Hotel Guest Satisfaction





Location-related keyword extraction

Literature reviews



mountain, airport, noise, wildlife, loud, beautiful location, calm, beach, train, waterfall, railway, shopping, city center, historical site, central location, convenient, supermarket, supermarket, landscape, scenery, nature, bus, taxi



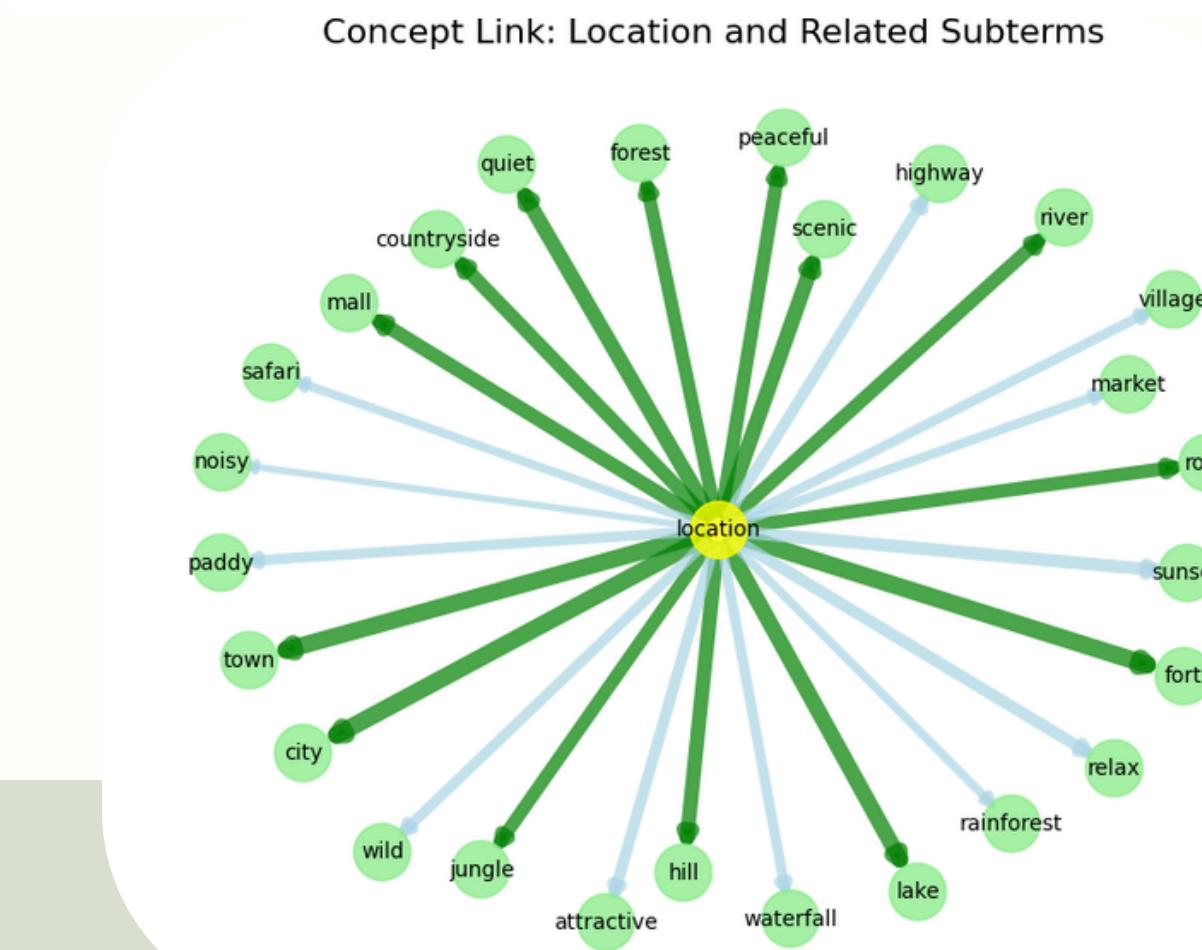
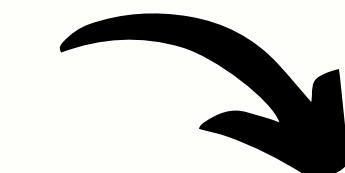
Table 3: Most similar keywords to the predefined list of keywords

Term	Similar words	Similarity scores	Term	Similar words	Similarity scores
airport	safari	0.794564	noise	noisy	0.922960
beach	rock	0.705956	shopping	mall	0.961620
beach	sunset	0.740195	supermarket	market	0.946675
calm	peaceful	0.904825	train	highway	0.686579
calm	quiet	0.824907	wildlife	wild	0.935046
calm	relax	0.798782	wildlife	village	0.927997
calm	attractive	0.739300	wildlife	waterfall	0.887560
convenient	fort	0.796159	wildlife	paddy	0.887436
convenient	city	0.776347	wildlife	jungle	0.886620
convenient	town	0.771431	wildlife	forest	0.886116
landscape	countryside	0.928022	wildlife	rainforest	0.743740
mountain	river	0.925842			
mountain	lake	0.924652			
mountain	hill	0.912254			
mountain	scenic	0.857700			

Word2vec model



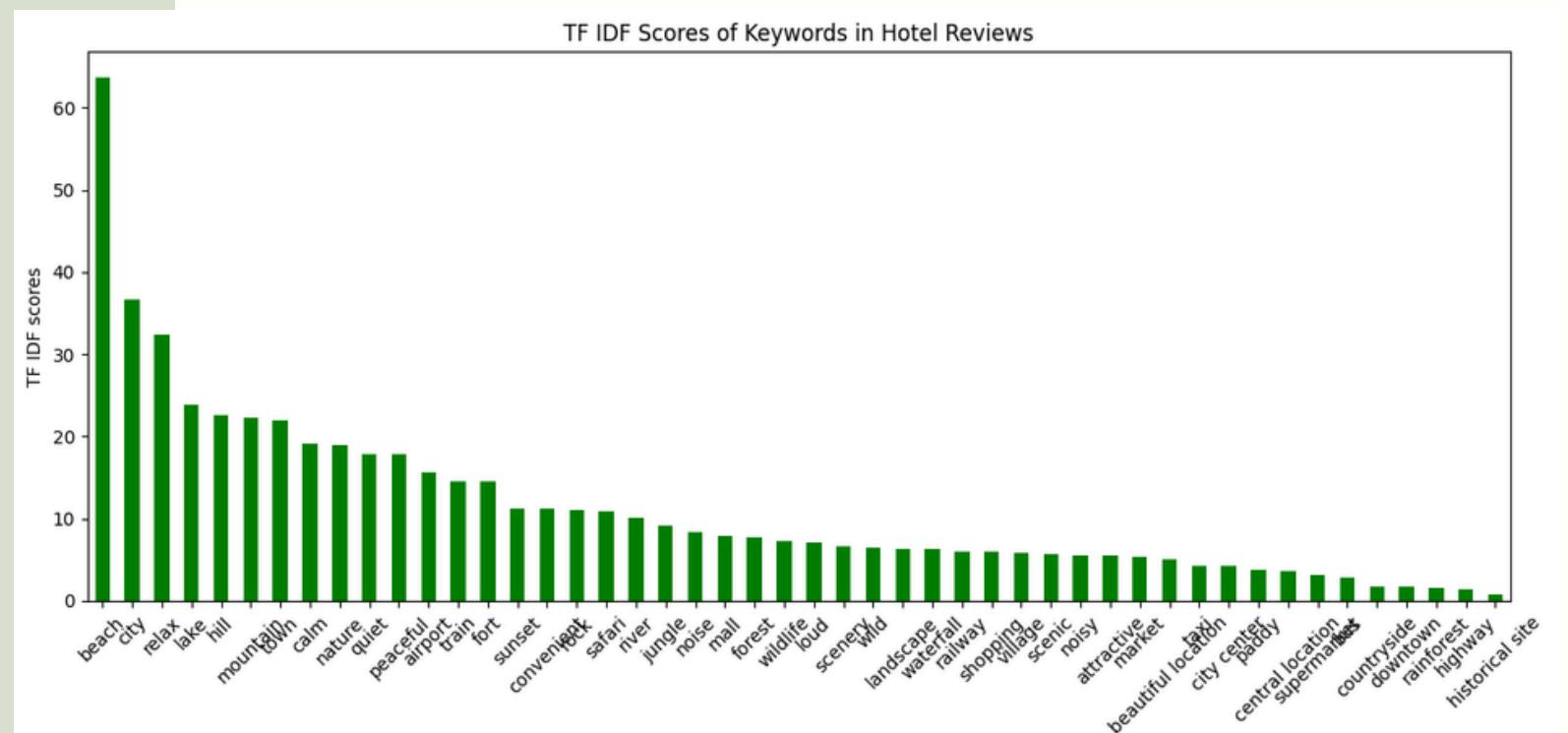
safari, rock, sunset, peaceful, quiet, relax, attractive, fort, city, town, count yside, river, lake, hill, scenic, noisy, mall, market, highway, wild, village, waterfall, paddy, jungle, forest, rainforest





TF IDF Scores

Among all keywords related to hotel locations, keywords such as 'beach', 'relax', 'city', 'lake', and 'clam' have the highest TF IDF scores, and 'downtown', 'historical site' and 'highway' keywords have the smallest TF IDF scores in overall hotel reviews.



A	B	C	D	E	F	G	H	I	J	K	L	M
Hotel names	fort	city	town	lake	scenic	quiet	hill	mall	countryside	peaceful	river	rock
98 Acres Resort & Spa	0	0	0	0	0	0	0	0	0.1156	0	0	0.24607
Aarunya Nature Resort - Kandy, Sri Lanka	0	0	0	0	0	0	0	0	0	0	0	0
Aditya Resort	0	0	0	0	0	0	0	0	0	0	0	0
Amaara Forest Hotel Sigiriya	0	0	0	0	0	0	0	0	0	0	0	0.1542 0.15588
Amangalla	0.80135	0	0	0	0	0	0	0	0	0	0	0
Amanwella	0	0	0	0	0	0	0	0	0	0	0	0
Amba Kola	0	0	0	0	0	0	0	0	0	0	0	0
Amethyst Resort Passikudah	0	0	0	0	0	0	0	0	0	0	0.27346	0
Andriyala	0	0	0	0	0	0	0.76451	0	0	0.22186	0	0
Araliya Beach Resort & Spa	0	0	0	0	0	0	0	0	0	0	0	0
Araliya Green City	0	0.96905	0	0	0	0	0	0	0	0	0	0
Araliya Green Hills Hotel	0	0	0	0	0	0	0.94649	0	0	0	0	0
Araliya Red	0	0	0	0	0.11215	0	0	0	0	0	0	0
Ayurvie Weligama	0	0	0	0	0	0.25329	0	0	0	0	0	0
Belmont Boutique Hotel	0	0	0	0	0	0	0	0	0	0	0	0
Berjaya Hotel Colombo	0	0	0	0	0	0	0	0	0	0	0	0
Best Western Elyon Colombo	0	0.79428	0	0	0	0	0	0	0.15057	0	0	0
Blue Wild - Yala	0	0	0	0.7044	0	0	0	0	0	0	0	0
Buckingham Place	0	0	0	0	0	0	0	0	0.04321	0	0	0
Calamansi Cove Villas	0	0	0	0	0	0	0	0	0	0	0	0
Camelot Beach Hotel	0	0	0	0	0	0	0	0	0	0	0	0
Camp Leopard - Yala Safari Glamping	0	0	0	0	0	0	0	0	0	0	0	0
Ceylon Resort	0	0	0	0	0	0	0	0	0	0	0	0
Chaarya Resort & Spa	0	0	0	0	0	0	0	0	0	0	0	0
Chimneys - Nuwara Eliya	0	0	0.4482	0	0	0.37181	0	0	0	0.51844	0	0
Cinnamon Lakeside Colombo	0	0	0	0.93529	0	0	0	0	0.01872	0	0	0
Clove Villa	0	0.59495	0.42714	0	0	0.44292	0	0	0	0.27792	0	0
Coco Royal Beach Resort	0	0	0	0	0	0	0	0	0	0	0	0
Colombo Court Hotel & Spa	0	0.68437	0	0	0	0.21164	0	0	0	0	0	0
Courtyard Colombo	0	0	0	0	0	0	0	0	0.71019	0	0	0



Geospatial analysis

Hotels within the same geographical location can be differentiated according to the perceived value of tourists in that location.

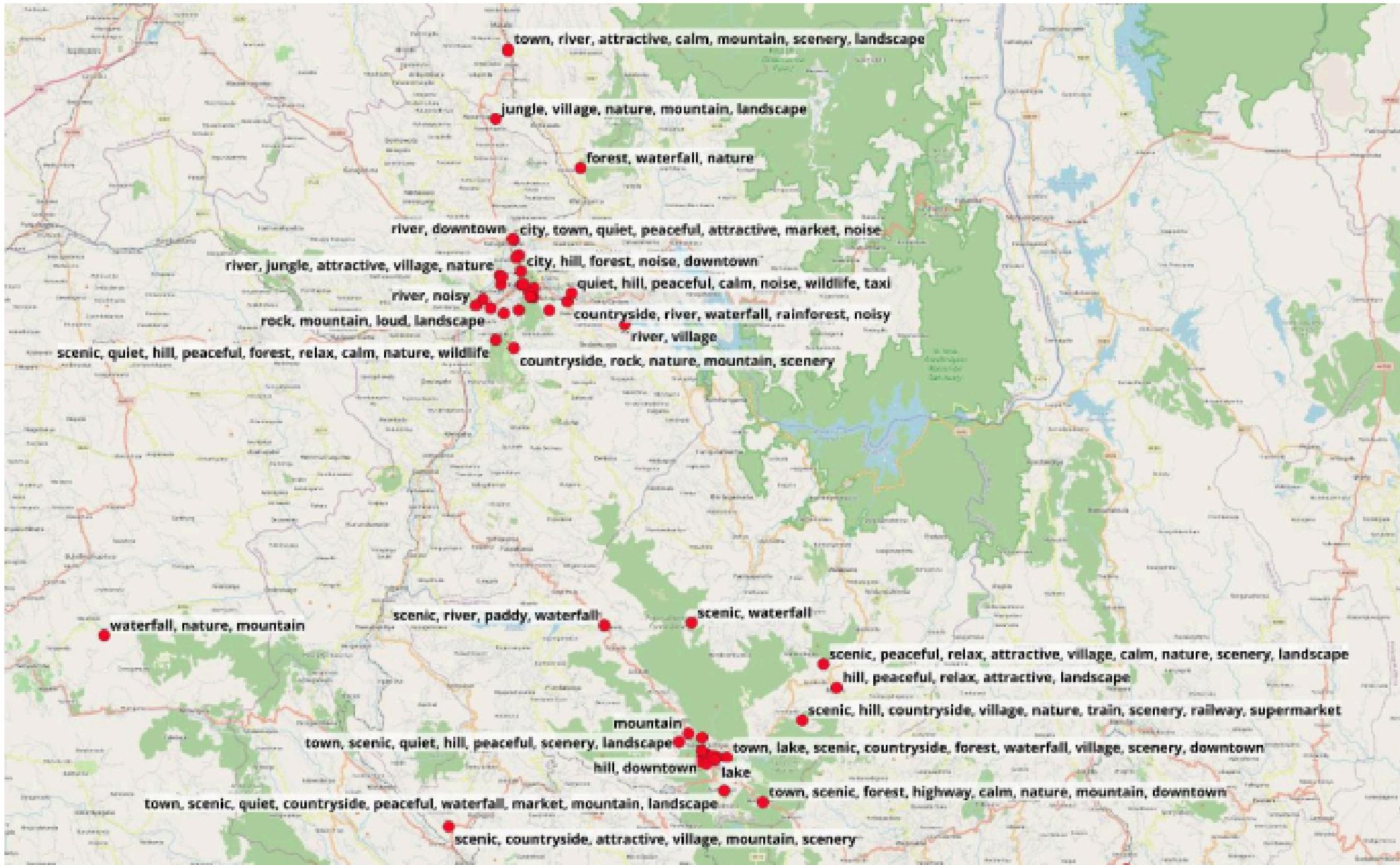


Figure 5:Hotels in Central Province with location related keywords

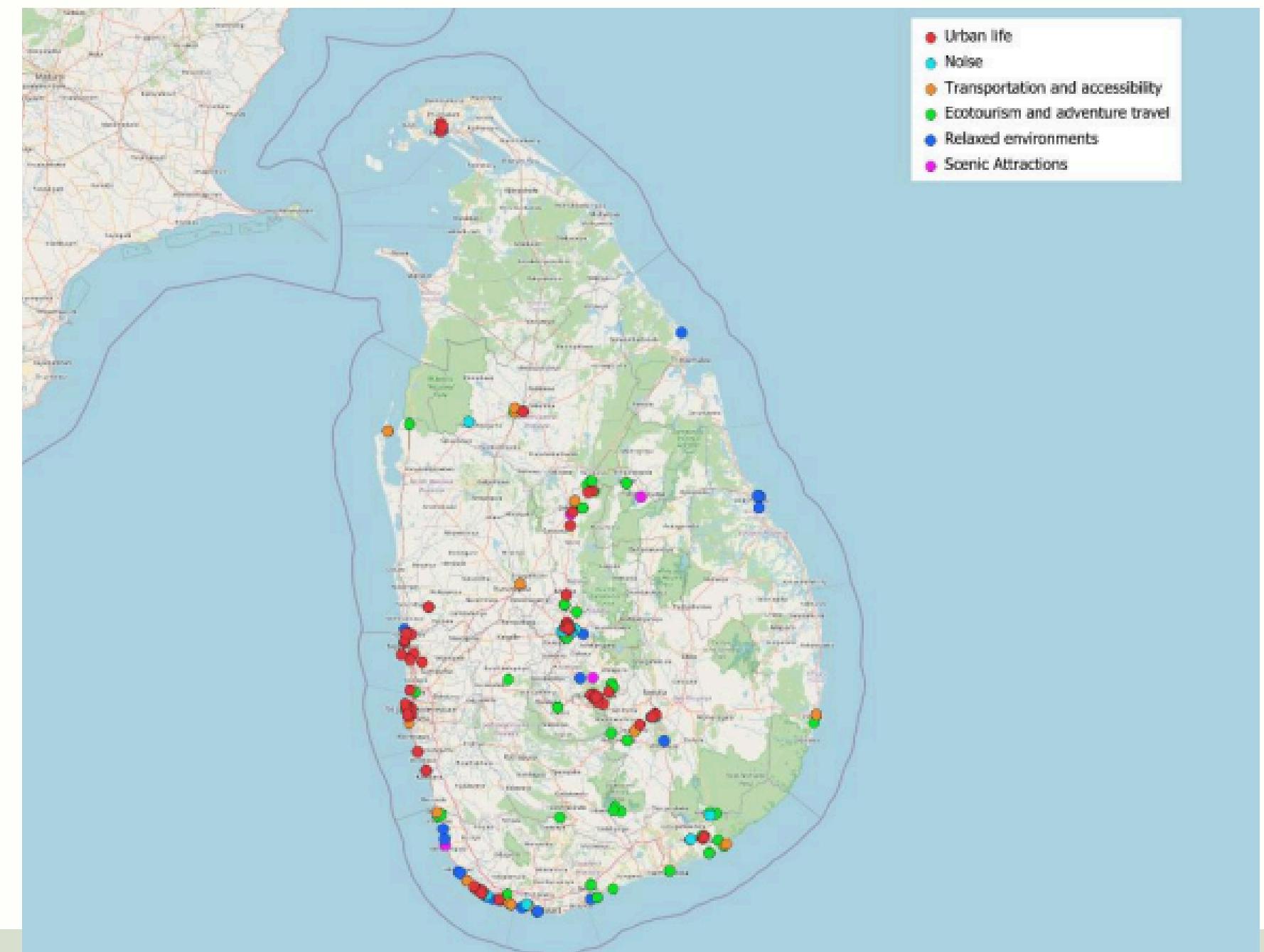


Location related Clusters

The K-means clustering process used in this context effectively organizes hotel location-related words into meaningful groups based on semantic similarity. Ecotourism and adventure travel, Noise, Relaxed environments, Scenic Attractions, Transportation and accessibility, Urban life clusters included the number of hotels as 109,45,123,111,75,117 respectively.

Table 4: Location-related clusters and associated keywords

Location-related clusters	Keywords
Scenic Attractions	['lake', 'scenic', 'river', 'sunset', 'waterfall', 'landscape', 'attractive']
Urban life	['fort', 'city', 'town', 'mall', 'airport', 'shopping', 'supermarket', 'market', 'downtown', 'central location', 'city center']
Noise	['noisy', 'noise', 'loud']
Transportation and accessibility	['train', 'railway', 'bus', 'taxi', 'convenient', 'highway']
Relaxed environments	['quiet', 'peaceful', 'relax', 'calm', 'beautiful location', 'scenery', 'beach', 'paddy', 'village', 'historical site']
Ecotourism and adventure travel	['safari', 'hill', 'rock', 'forest', 'jungle', 'wild', 'nature', 'mountain', 'rainforest', 'wildlife']

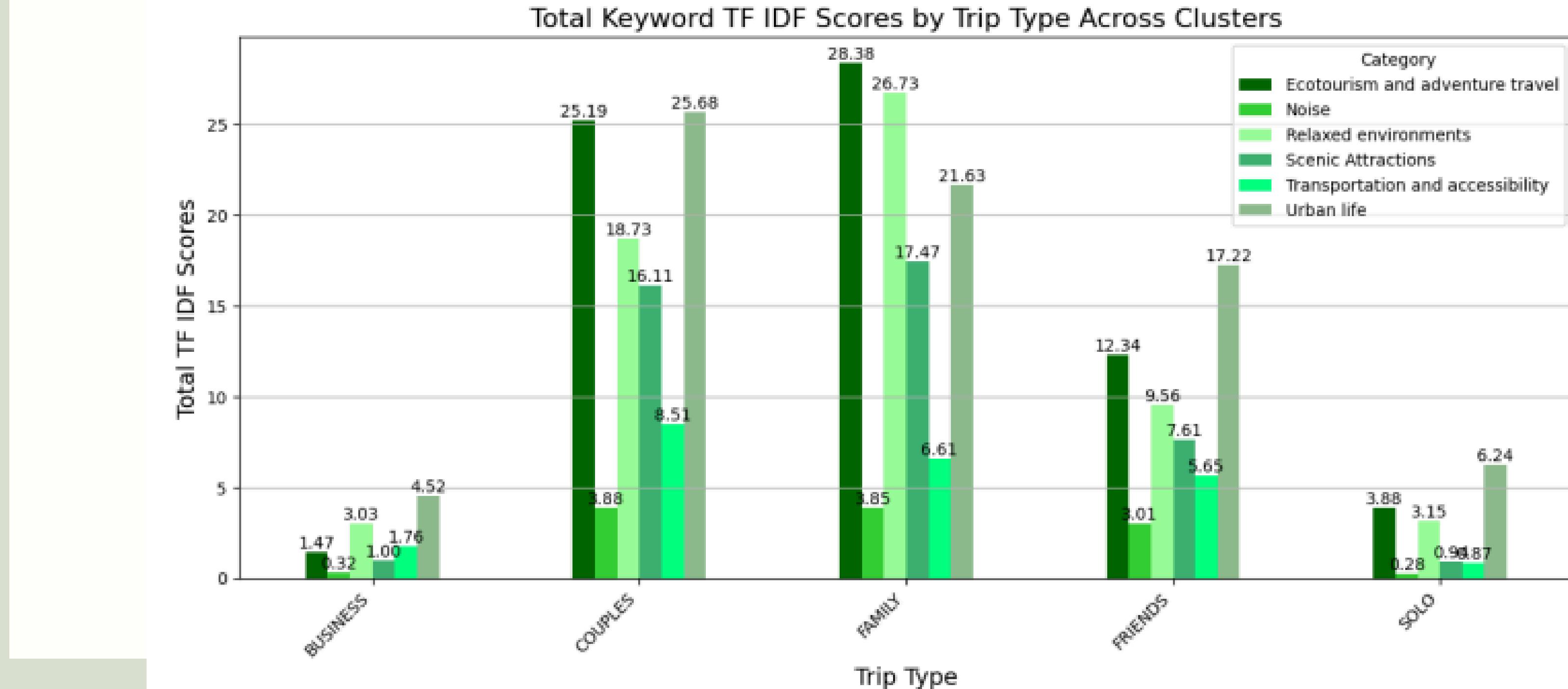


Geographical distribution of Hotels in Sri Lanka with location-related clusters



TF IDF Scores by Trip Type across Clusters

Each travel group has unique destination preferences that reflect their motivations and travel styles



Conclusion

- Online travel agencies can use these findings to improve the customer's booking experience. For example, although the TripAdvisor website suggests amenities, there are no options to easily identify the location of hotels. When travelers find hotels based on their preferences, they can suggest these keywords. Furthermore, hotels can use these findings for target market strategies.
- Such location-related insights help hotels refine their offerings and develop targeted marketing strategies. For instance, Keywords contained in the Ecotourism and Adventure Travel cluster were mostly mentioned by couples and families. Hotels in this cluster can highlight outdoor activities, local wildlife experiences, and eco-friendly practices in their promotional materials. This targeted approach allows hotels to resonate more deeply with their desired clientele, enhancing guest satisfaction and loyalty.

References

- Abrahams, A. S., Fan, W., Wang, G. A., Zhang, Z., & Jiao, J. (2015). An integrated text analytic framework for product defect discovery. *Production and Operations Management*, 24(6), 975–990. <https://doi.org/10.1111/poms.12303>
- Athula C. Gnanapala. (2014a). Factors Affecting Customer Satisfaction Related to the Tourist Hotel Industry in Sri Lanka. *J. of Tourism and Hospitality Management*, 2(7). <https://doi.org/10.17265/2328-2169/2014.07.001>
- Ban, H. J., Choi, H., Choi, E. K., Lee, S., & Kim, H. S. (2019). Investigating key attributes in experience and satisfaction of hotel customer using online review data. *Sustainability (Switzerland)*, 11(23). <https://doi.org/10.3390/su11236570>
- Chaves, M. S., Gomes, R., & Pedron, C. (2012). Analysing reviews in the Web 2.0: Small and medium hotels in Portugal. *Tourism Management*, 33(5), 1286–1287. <https://doi.org/10.1016/j.tourman.2011.11.007>
- Dickinger, A., & Mazanec, J. A. (2015). Significant word items in hotel guest reviews: A feature extraction approach. In *Tourism Recreation Research* (Vol. 40, Issue 3, pp. 353–363). Taylor and Francis Ltd. <https://doi.org/10.1080/02508281.2015.1079964>
- Dong, J., Li, H., & Zhang, X. (2014). Classification of Customer Satisfaction Attributes: An Application of Online Hotel Review Analysis. 238–250. https://doi.org/10.1007/978-3-662-45526-5_23
- Gretzel, U., & Yoo, K. H. (2008). Use and Impact of Online Travel Reviews. In *Information and Communication Technologies in Tourism 2008* (pp. 35–46). Springer Vienna. https://doi.org/10.1007/978-3-211-77280-5_4
- Guenther, M., & Sucholutsky, N. (2017). Text mining using n-gram variables. In *The Stata Journal* (Vol. 17, Issue 4). 17 <http://www.schonlau.net/publication/17ngram.pdf> Electronic copy available at: <https://ssrn.com/abstract=2759033>

- Jatnika, D., Bijaksana, M. A., & Suryani, A. A. (2019). Word2vec model analysis for semantic similarities in English words. *Procedia Computer Science*, 157, 160– 167. <https://doi.org/10.1016/j.procs.2019.08.153>
- Ko, C.-H. (2018). Exploring Big Data Applied in the Hotel Guest Experience. *OALib*, 05(10), 1–17. <https://doi.org/10.4236/oalib.1104877>
- Lee, K. W., Kim, H. B., Kim, H. S., & Lee, D. S. (2010). The determinants of factors in FIT guests' perception of hotel location. *Journal of Hospitality and Tourism Management*, 17(1), 167–174. <https://doi.org/10.1375/jhtm.17.1.167>
- Nicholas, C. K. W., & Lee, A. S. H. (2017). Voice of customers: Text analysis of hotel customer reviews (Cleanliness, overall environment and value for money). *ACM International Conference Proceeding Series*, Part F132530, 104–111. <https://doi.org/10.1145/3152723.3152739>
- Parzych, K., & Brkić-Vejmelka, J. (2020). Guests' assessment of hotel facilities and services: Zadar case study. *European Journal of Tourism, Hospitality and Recreation*, 10(3), 241–250. <https://doi.org/10.2478/ejthr-2020-0021>
- Phucharoen, C., & Sangkaew, N. (2020). Does firm efficiency matter in the hospitality industry?: An empirical examination of foreign demand for accommodation and hotel efficiency in Thailand. *Journal of Tourism Analysis*, 27(1), 62–80. <https://doi.org/10.1108/JTA-08-2019-0034>
- Rahman, H., Kazi, M. M., Islam, M. S., Maoya, J., Arif, M. M., Tagwar, M., Hasan, M., Ali, M. S., Jabid, T., & Islam, M. (2023). An Analysis of Bangla Tweets on Social Media Platform for Polarity Detection Using Machine Learning Algorithms. *2023 4th International Conference on Big Data Analytics and Practices*, IBDAP 2023. <https://doi.org/10.1109/IBDAP58581.2023.10271974>
- Wang, J., Zhao, Z., Liu, Y., & Guo, Y. (2021). Research on the role of influencing factors on hotel customer satisfaction based on bp neural network and text mining. *Information (Switzerland)*, 12(3), 1–19. <https://doi.org/10.3390/info12030099>
- Xiang, Z., & Krawczyk, M. (2016). What Does Hotel Location Mean for the Online Consumer? Text Analytics Using Online Reviews. In *Information and 18 Communication Technologies in Tourism 2016* (pp. 383–395). Springer International Publishing. https://doi.org/10.1007/978-3-319-28231-2_28

- Xiang, Z., Schwartz, Z., Gerdes, J. H., & Uysal, M. (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction? *International Journal of Hospitality Management*, 44, 120–130. <https://doi.org/10.1016/j.ijhm.2014.10.013>
- Yang, Y., Luo, H., & Law, R. (2014). Theoretical, empirical, and operational models in hotel location research. *International Journal of Hospitality Management*, 36, 209–220. <https://doi.org/10.1016/j.ijhm.2013.09.004>
- Yang, Y., Mao, Z., & Tang, J. (2018). Understanding Guest Satisfaction with Urban Hotel Location. *Journal of Travel Research*, 57(2), 243–259. <https://doi.org/10.1177/0047287517691153>
- Yang, Y., Wong, K. K. F., & Wang, T. (2012). How do hotels choose their location? Evidence from hotels in Beijing. *International Journal of Hospitality Management*, 31(3), 675–685. <https://doi.org/10.1016/j.ijhm.2011.09.003>
- Zainol, Z., Lee, A. S., Nohuddin, P. N., Farizah Ibrahim, N., & Hanafi Ahmad Hijazi, M. (2022). Examining the Relationship of Keyword Analysis using Online Traveller Hotel Reviews. In *International Journal on Perceptive and Cognitive Computing (IJPCC)* (Vol. 8).
- Zhou, L., Ye, S., Pearce, P. L., & Wu, M. Y. (2014). Refreshing hotel satisfaction studies by reconfiguring customer review data. *International Journal of Hospitality Management*, 38, 1–10. <https://doi.org/10.1016/j.ijhm.2013.12.004>