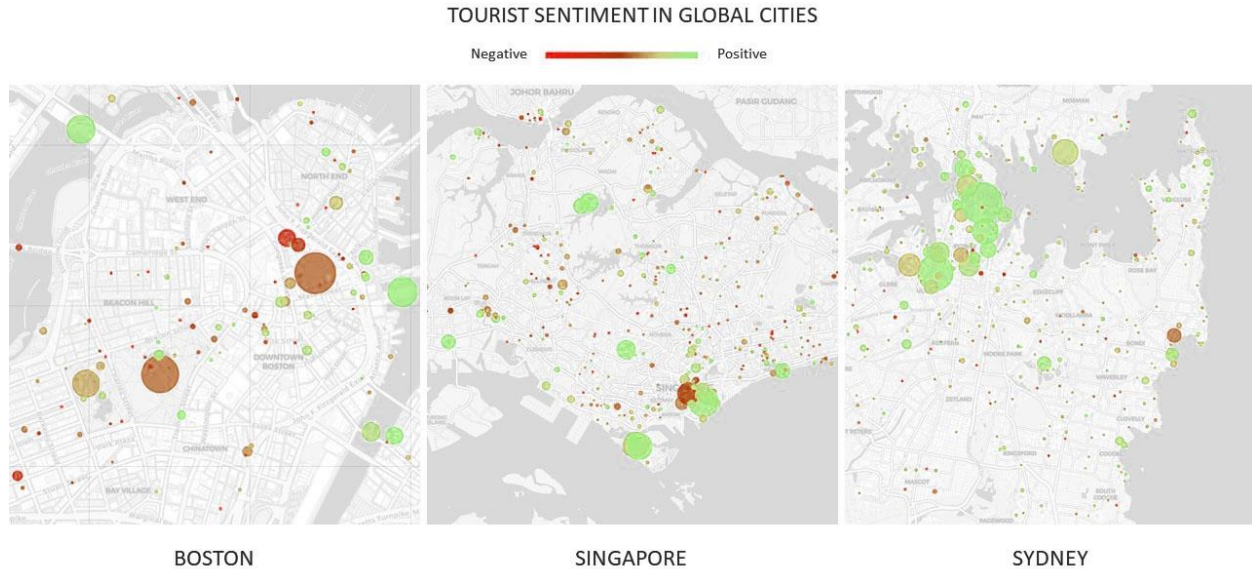


THE TOURIST'S IMAGE OF THE CITY

A data-driven comparative analysis of the features of tourist interest across three global cities

Rohit Priyadarshi Sanatani



Project Brief:

Tourist attractions play major roles in shaping the mental image of cities for residents and visitors alike. While there have been numerous studies over the past decades inquiring into urban tourist sentiment and features of interest, the exponential growth of big-data over the past few years have opened up valuable methodological approaches in this regard (Alaei et al. 2019). Social media data from platforms such as Twitter, Flickr and Tripadvisor has emerged as a popular data-source for such recent research directions (Galí et al. 2015, Valdivia et al. 2017, Cardone et al. 2021). Moreover, geolocated Point of Interest (POI) data available through platforms such as Google Places, and Foursquare provide valuable access to user-reviews, user-photographs, and visitation patterns, at a very high spatio-temporal resolution (Marti et al. 2019).

Drawing upon Point of Interest (POI) data for tourist attractions across three cities across the globe – Boston, Singapore and Sydney - this body of research analyses the predominant themes of discussion and features of interest that contribute towards popular tourist sentiment towards these cities. The study collects over 3500 user reviews and over 6000 photographs across 750 tourist attractions through the Google Places API. This data is then merged with Flickr data streams for additional visual content, and Twitter posts for additional data on topics of interest. Sentiment analysis is then carried out on the textual data using the NLTK Vader analyzer, for the extraction of emotional content. A Latent Dirichlet Allocation (LDA) model is used for topic-modeling for the extraction of high-level concepts and predominant topics of discussion emergent from these cities. A faster-RCNN object detection model trained on the Open-Images V4 dataset is then run on the user photographs for extraction of semantic features contained within them. Unsupervised K-means clustering is carried out on these features, to identify the major themes of tourist interest. Similar clustering is then carried out on the combined dataset, to identify major

themes of discussion and corresponding features of interest that frequently appear together. Finally, a qualitative assessment of these clusters is embarked upon, for a comparative analysis of the major similarities and differences in patterns of tourist interest across the five cities. It is hoped that the methodological framework demonstrated through this work is able to provide valuable insights into the multitude of ways in which different cities across the world are perceived by tourists and visitors.

References:

Alaei, A.R., Becken, S. and Stantic, B., 2019. Sentiment analysis in tourism: capitalizing on big data. *Journal of Travel Research*, 58(2), pp.175-191.

Cardone, B., Di Martino, F. and Sessa, S., 2021. GIS-based fuzzy sentiment analysis framework to classify urban elements according to the orientations of citizens and tourists expressed in social networks. *Evolutionary Intelligence*, pp.1-10.

Galí, N. and Donaire, J.A., 2015. Tourists taking photographs: the long tail in tourists' perceived image of Barcelona. *Current Issues in Tourism*, 18(9), pp.893-902.

Martí, P., Serrano-Estrada, L. and Nolasco-Cirugeda, A., 2019. Social Media data: Challenges, opportunities and limitations in urban studies. *Computers, Environment and Urban Systems*, 74, pp.161-174.

Valdivia, A., Luzón, M.V. and Herrera, F., 2017. Sentiment analysis in tripadvisor. *IEEE Intelligent Systems*, 32(4), pp.72-77.