# How to use Instagram to Travel the World? An Approach to Discovering Relevant Insights from Tourist Media Content

Angel Fiallos

*Universidad Ecotec, Samborondón, Ecuador*

Abstract
This work aims to detect content themes, locations, sentiment, and demographic information on Instagram or similar platforms in a way that supports business decision-making and marketing strategies in the tourism or travel industries. For this purpose, we propose an original combination of NLP methodology and computer vision to be applied to the content of posts associated with a specific hashtag. To demonstrate this, we collected and processed 30,122 images and texts of Instagram posts related to the hashtag `#traveltheworld`, showing the results of the most relevant user interests, places, emotions, and other detected features.

**Keywords**
Natural Language Processing, Data Mining, Computer Vision, Instagram, Tourism

## 1. Introduction

Social media is essential because social networks have made everybody a potential author, so the language is now closer to the user than to any prescribed norms. In this way, share information about events, activities, services, opinions, and experiences on social media channels.

Instagram is a social network that has experienced a rapid increase in users and picture uploads since it was launched in October 2010. However, a few research works have been developed around it in contrast to other social networks like Twitter, where the text is analyzed as the main element in its posts.

Ninety million photos are shared every day through Instagram. Furthermore, users add other features such as hashtags, locations, and text to photos through the platform. These media elements communicate the user's intention behind posting an image but do not necessarily describe the published image [1]. Also, concerning hashtags, several researchers suggest they carry emotional information which is not directly related to the context they appear [2].

Hashtags are also used to create searchable content categories to gain followers by attracting the attention of public users by businesses and are single words or unbroken strings of words preceding the # symbol. Instagram encourages users to make hashtags both specific and relevant,

*ICAIW 2022: Workshops at the 5th International Conference on Applied Informatics 2022, October 27–29, 2022, Arequipa, Peru*
afialloso@ecotec.edu.ec (A. Fiallos)
0000-0002-7828-1207 (A. Fiallos)

rather than tagging generic words, to make photographs stand out and to attract like-minded Instagram users.

Obtaining all possible information from these Instagram posts is essential for gaining user insights, measuring brand reputation, and other important market digital research aspects on several industries, such as tourism, travel, hospitality, and customer services, among others. Also, to evaluate campaigns in business, understand users' social behavior, and avoid costly direct surveys.

The main contribution of this work is proposing a methodology to identify the relevant topics, locations, sentiments, and features from a combination of text and pictures associated with a particular hashtag by combining text mining techniques, sentimental analysis, natural language processing, and computer vision tools. The methodology was applied to a dataset of Instagram photos associated with the hashtag `#traveltheworld`. This popular hashtag refers to more than 15 million posts and is used by travelers to discover new destinations, swap travel tips, and share their experiences.

The rest of this work is structured as follows: Section 2 describes the related work, Section 3 presents the proposed methodology, Section 4 describes the results of the case study analysis, and Section 5 presents the conclusions and future work components, incorporating the applicable criteria that follow.

## 2. Related Word

Few researchers have investigated different ways to detect relevant content topics from Instagram pictures: Hu et al., [3] analyzed free photos of a random sample of users by considering the user's text. Then, the similarity between pictures was calculated in terms of Euclidean distance between their codebook vectors by k-means to obtain clusters of photos. This work shows eight popular picture categories (friends, food, gadget, captioned photos, pets, activities, selfies, fashion) and five distinct types of users in terms of their posted pictures.

Jang et al., [4] performed an analysis of the relationship between LDA-based topics and Likes from the test datasets of 20 million users and their 2 billion LDA-based topics. This work uses a Latent Dirichlet Allocation model over the description text and hashtags written by users. As a result, they identified 20 latent topics prevalent among hashtags added to pictures and presented the top 5 topics.

Amanatidis et al., [5] performed a picture analysis and categorization of the personal experiences of users before, during, and after the covid-19 vaccination process. For this purpose, they used computer vision convolutional neural models and datasets from ImageNet.

Manikonda [6] concluded that on Twitter, you could locate informational content, while on Instagram, the content is more personal and social. To reach this conclusion, the researchers performed a textual and visual analysis of the media content posted on these two platforms from the same set of users. Our paper differs from those mentioned because it uses a multidisciplinary approach of techniques (computer vision and natural language processing) and validates which one provides better results depending on the objectives to be achieved, in this case, focused on tourist experiences and user demographics.

## 3. Methodology

First, the topics, locations, sentiments, and demographic information were detected following the steps shown in Figure 1.

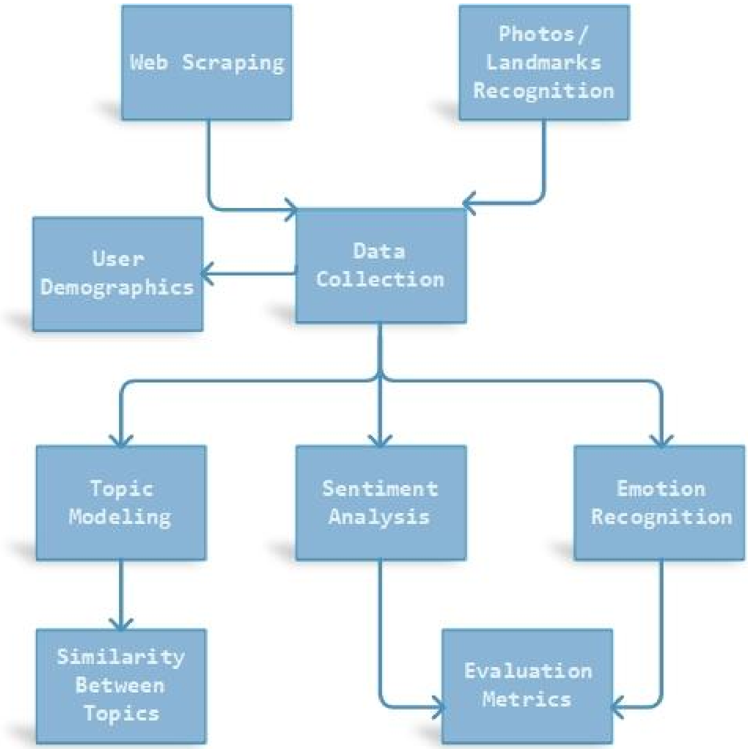


**Figure 1:** Proposed methodology.

### 3.1. Data Collection

A scraping process developed with Python, using BeautifulSoup, and Selenium libraries, was applied to collect a dataset of 50.510 publications related to the hashtag #traveltheworld from the Instagram Platform. These data include the following features per publication: image file, post id, user id, hashtags, upload date, post text, locations, and likes count.

Then, a sample of 30,122 photos was selected from user accounts with an average of at least 150 likes and 100 followers to avoid downloading photos that belong to fake accounts. The hashtags and the text were taken as post descriptions for this work. Figure 2 shows an Instagram post sample.

Once the photo collection was obtained, an image recognition process was applied to the digital files to retrieve the visual description. Using Microsoft Cognitive Services API[1], multiple executions were run to obtain the visual description of each picture in JSON Format. The API has

[1]Microsoft Cognitive Services https://azure.microsoft.com/es-es/services/cognitive-services/


**Figure 2:** Instagram post sample.

a collection of SDK applications and machine-learning services developed for the Bing Oxford Project and Microsoft Research. Figure 3 shows how computer vision API returns information about the visual content of an image.

The Microsoft Cognitive Services API also recognizes natural and manmade landmarks worldwide by comparing them to a library of known places. Figure 4 shows an example of a response once the recognition process is applied over a landmark photo.

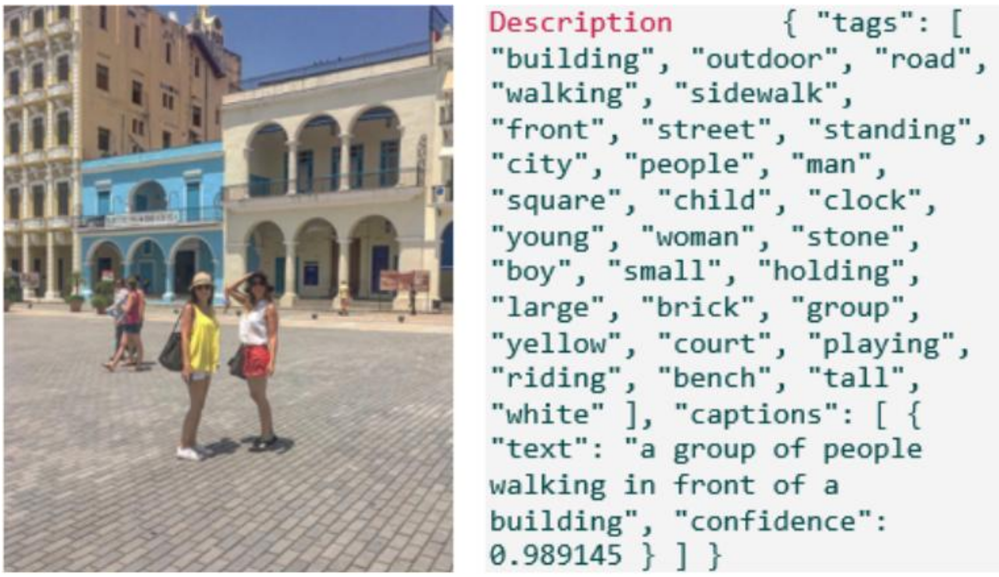


**Figure 3:** Visual content response from an IG photo.

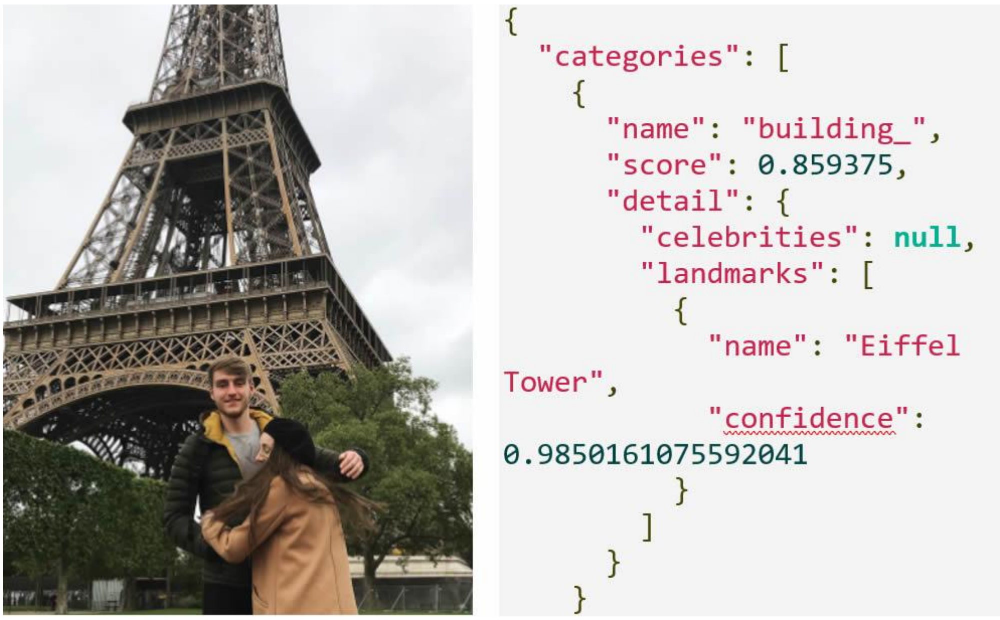


**Figure 4:** Visual content response from a landmark IG photo.

### 3.2. Terms Detection

Some text-mining processes were applied to documents to determine the most relevant topics. First, a data preprocessing [7] was executed separately for posts descriptions and visual description files with the following steps:

- Each document was transformed into words (lexical analysis).
- Empty words (articles, prepositions, marks, conjunctions, numbers, punctuation, and other words that did not semantically describe the content) were deleted.
- Stemming process was executed where non-essential parts of terms, such as suffixes and prefixes, were eliminated to keep their essential part (lemma) of the terms.

Second, the TF-IDF (Term Frequency-Inverse document frequency) model [8] was applied to evaluate the key terms in the documents. TF-IDF measures the weight of a term based on the term frequency (TF) and inverse document frequency (IDF). Then, a document-term matrix was created with the TF-IDF, and the dispersed terms were deleted to conserve the most relevant terms.

### 3.3. Topic Modeling

Topic modeling is a text mining technique that employs unsupervised and supervised statistical machine-learning techniques to identify patterns in a corpus or a large amount of unstructured text. It can take a vast collection of documents and group the words into clusters of words, identifying topics using the process of similarity.

We applied Non-Negative Matrix Factorization to determine relevant topics in both documental corpus. NMF [9] is a linear-algebra optimization algorithm to extract meaningful information about topics from decomposing the document-term matrix A. in two k-dimensional factors W (document-topic matrix) and H (topic-term matrix).

### 3.4. Sentimental Analysis

Sentimental analysis is a technique that uses natural language processing to identify, extract, quantify, and explore affective states and subjective information from text. Generally, the sentimental analysis used a text classification approach based on machine learning.

The text classification assumes that each sample is assigned to one and only one label. On the other hand, multi-label classification assigns to each sample a set of target labels that are not mutually exclusive. However, many of text multi-label classification methods ignore the word order, opting to use word bag models or TF-IDF weighting to create document vectors.

Convolutional neural networks (CNN) utilize layers with convolving filters that are applied to local features [10]. Initially invented for computer vision, CNN models are adequate for NLP and have achieved excellent results in semantic parsing. Kim and Berger [11, 12] demonstrated that CNN models using semantic word embeddings such as Word2Vec [13] significantly outperform the Binary Relevance method with bag-of-words features on a large-scale multi-label.

We designed a simple CNN network composed of an input layer with five different n-grams window sizes and one convolution layer on top of word vectors obtained from the Word2Vec unsupervised neural language model. These vector representations essentially feature extractors that encode words' semantic features in their dimensions. To conduct the experiment, first, we trained a dataset provided by FigureEigth[2], which contains approximately 19.000 tweets that have been labeled in neutral, positive, and negative sentimental categories.

### 3.5. Emotion Recognition

Face API allows the detection of human faces together with facial attributes that contain predictions of facial features based on automatic learning. The characteristics of available facial attributes are age, emotion, gender, and posture, among others. The API also integrates recognition of emotions and returns the degree of confidence of a set of emotions for each face detected. The process is applied to a set of Instagram photos that, during the process of image recognition, refer to some of the values: "man", "men", "woman," or "women". For each photo, the emotional response with the highest score is compared to the emotion classified manually by observers (ground truth). Figure 5 shows a response from Face API.

## 4. Results

### 4.1. Relevant Terms and Topics

Computer Vision API was applied over 30.122 images and detected 1.816 unique terms related to images' visual contents. After the preprocessing routines, 1.801 terms (99.12%) were conserved

[2]FigureEight https://www.figure-eight.com/wp-content/uploads/2016/07/text_emotion.csv

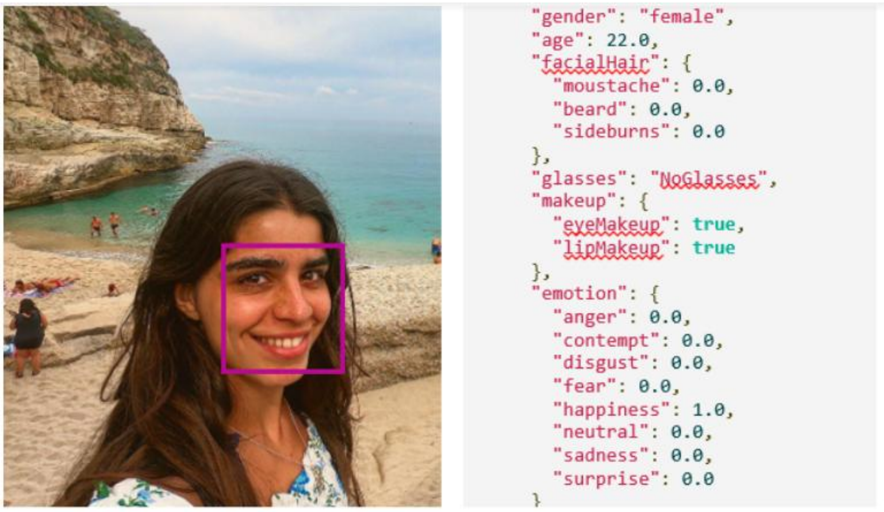


**Figure 5:** Face API Response.

for the following analysis. Figure 6 shows a word cloud with the most relevant terms related to the visual content of images. Figure 7 shows the most frequent terms with higher TF-IDF weights. They are building, groups, people, water, person, mountain, woman, cities, beaches, and streets, among others. These terms have a TF-IDF weight greater than 12.500 and suggest that most of the pictures are related to building structures, people, urban cities, sports activities, and natural tourism attractions.


**Figure 6:** Wordcloud of relevant terms.

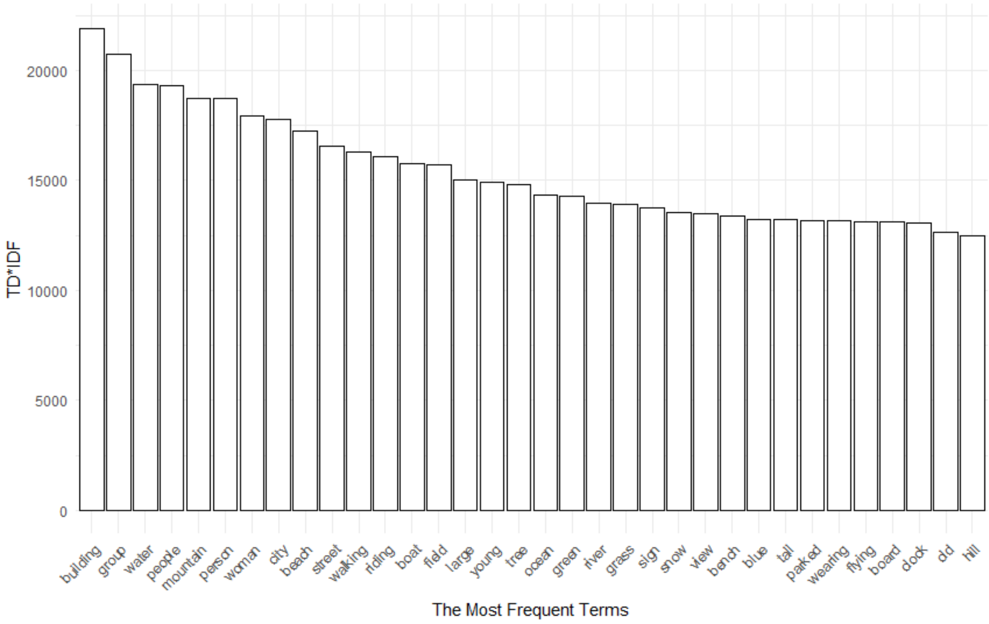


**Figure 7:** Most Frequent terms in corpus.

Table 1 presents the six terms more associated with the key terms “mountain,” “woman”, “water”, “building”, “people,” and “city”; for example, the term “mountain” is related to hills, nature, background, view, and field. Key terms were set considering the most frequent terms illustrated in Figure 7. Associated terms have a correlation, a quantitative measure between 0

**Table 1**
Correlation Terms detected in datasets.

| **Terms** | **Correlation Terms** | | | |
|---|---|---|---|---|
| Mountain | Hill<br>0.58 | Naure<br>0.53 | Back<br>0.44 | Rocky<br>0.38 |
| Woman | Young<br>0.68 | Person<br>0.66 | Girl<br>0.55 | Wearing<br>0.42 |
| Water | Boat<br>0.62 | Ocean<br>0.57 | River<br>0.53 | Lake<br>0.52 |
| Building | Old<br>0.46 | Clock<br>0.45 | Street<br>0.43 | tower<br>0.38 |
| People | Group<br>0.70 | Walking<br>0.30 | Crowd<br>0.26 | Man<br>0.24 |
| City | Street<br>0.55 | Traffic<br>0.50 | Tall<br>0.41 | Clock<br>0.34 |

and 1 of the occurrences of words in several documents. In this respect, whether two terms always appear together, then the calculated correlation is 1.

Using NMF, we detected the most relevant topics of visual descriptions. They are shown in Table 2. and refer to natural landscapes, people's actions, cities and buildings, sea and related activities, food, and other outdoor photos.

**Table 2**
Topics of visual descriptions from IG Photos.

| Topic | Terms |
|---|---|
| 1 | city, building, street, front, clock, tower, tall, old, large, sign. |
| 2 | body, water, boat, ocean, beach, lake, doc, river, large, sunset. |
| 3 | person, woman, young, hold, wear, pose, man, front, girl, standing. |
| 4 | mountain, field, hill, grass, green, tree. |
| 5 | table, sit, food, plate, room, close, wooden, white, indoor, cake. |

Next, a corpus of 24.719 documents and 21.972 terms were created with Instagram posts. After preprocessing, 18.810 terms (85.61%) were conserved for the topic modeling. The relevant topics results of user descriptions are shown in Figure 8. These topics refer to events, exclamations of admiration, visits to specific tourist sites, emotions, and engagements. In order to ensure that content is coherent and to eliminate redundancy in topic terms, we reduce the hashtags related to travel. Figure 8 shows the most frequent terms with higher TF-IDF weights greater than 800.

On the other hand, Table 3 shows the topics of the text content written by users. These

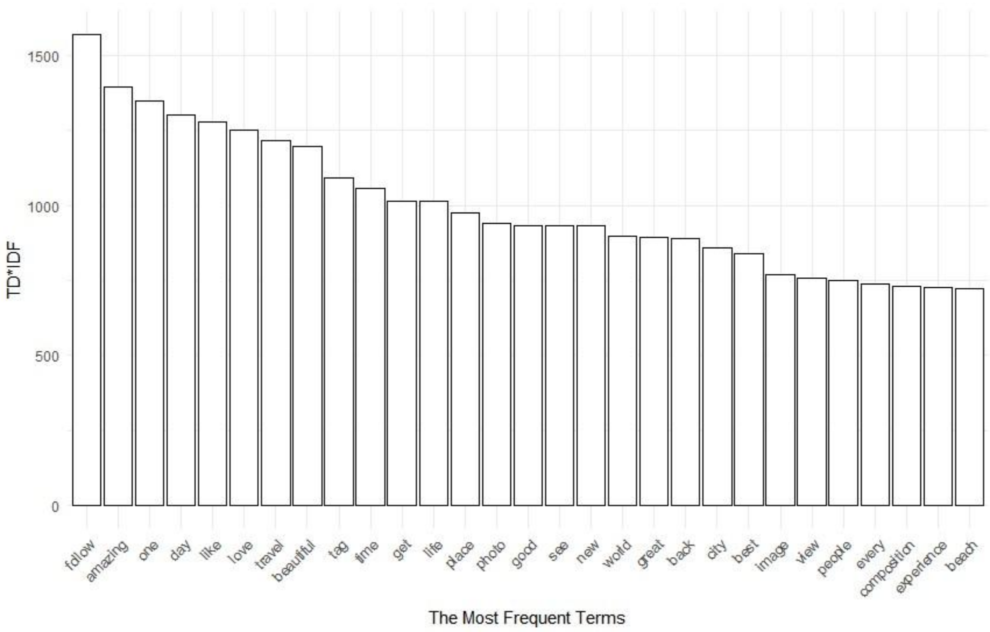


**Figure 8:** Most frequent terms of visual descriptions.

**Table 3**
Topics of text content written by users from IG Posts.

| Topic | Terms |
|---|---|
| 1 | view, top, stunning, enjoy, city, nice, room, beautiful, sea, hotel, climb. |
| 2 | tag, follow, friend, like, comment, someone, photo, credit, share, picture. |
| 3 | travel, world, happy, destination, capture, blog, escape, explore, live, inspiration. |
| 4 | day, beautiful, love, place, time, good, life, see, back, world, sunset. |
| 5 | love, fall, madly, city, hubby, guess, place, live want, someone, people. |

topics refer to travelers' stories, expressions of admiration, and social media engagements. The average cosine distance between the topics mined from the users' descriptions and the visual description was 0.290, which means there is a low similarity between both documents.

Then, the user descriptions do not allow us to identify the features and elements of the images in a specific way because they refer to narrations of events, situations, or opinions of events related to the photos.

## 4.2. Locations and Landmarks

The geolocations were added by users in 19.782 (65.69%) Instagram posts, so the locations for the remaining photos were detected using Computer Vision API landmark properties. A total of 2.26% of pictures were retrieved by this method. Table 4 shows the places identified in Instagram photos which are more highest count rate.

The identified locations include famous monuments and buildings, such as the Eiffel Tower,

**Table 4**
Popular Places Ranking from IG posts linked to hashtag #traveltheworld.

| Location | Count | Percentage |
|---|---|---|
| Paris, France | 214 | 1.08% |
| New York, New York | 172 | 0.87% |
| London, United Kingdom | 122 | 0.61% |
| Rome, Italy | 111 | 0.56% |
| Barcelona, Spain | 110 | 0.55% |
| Bali, Indonesia | 109 | 0.55% |
| Prague, Czech Republic | 96 | 0.48% |
| Amsterdam, Netherlands | 95 | 0.48% |
| Iceland | 69 | 0.35% |
| Venice, Italy | 68 | 0.34% |
| Los Angeles, California | 66 | 0.33% |
| Lisbon, Portugal | 62 | 0.31% |
| Istanbul, Turkey | 57 | 0.29% |
| San Francisco, California | 54 | 0.27% |
| Berlin, Germany | 51 | 0.26% |
| Total | | 7.33% |

Sagrada Familia, Pantheon Rome, Grand Central Terminal, Brooklyn Bridge, and Trevi Fountain, among others, those that were positioned to your specific city or country through the GeoPy tool[3]. These values can be contrasted with TripAdvisor info, the largest travel website in the world, where Paris, New York, London, Rome, Barcelona, Bali, and Prague, among others, are mentioned as the most popular locations in the World in TripAdvisor. Therefore, the results presented in Table 4 could be a good reference for worldwide tourism stats.

### 4.3. Users Demographics

We used a scraping process to retrieve a total of 17.752 unique photos of user profiles from the Instagram posts. The Face API process was applied to the profile's photo collection to recognize facial properties. Once the process was finished, we selected the photos with an exposure value greater than 0.5, and the genre and age properties could be detected. In total 5.560 (31.32 %).

The rest of the photos of user profiles, among other reasons, did not show the face of the user, belonged to business profiles, or had low quality and did not allow identification of gender and age properties. Table 5 shows the percentages belonging to the user genre groups, and table 6 shows the percentages belonging to the user groups by age range:

**Table 5**
Genre Percentages.

| Genre | Count | Percentage |
|---|---|---|
| Female | 3575 | 64.30% |
| Male | 1985 | 35.70% |

### 4.4. Emotion Recognition and Text Sentiment Analysis

An ideal visual experience on Instagram social network happens when the sentiment and emotions transmitted from text and photo(s) or video(s) are similar. Classifying emotions in publications requires a lot of effort and manual work from experienced teams. Therefore, emotion recognition and text sentiment analysis can help predict the emotions of a social media post.

A sample of 114 photos was taken that referred to a person with a visible face. It was automatically classified using Face API, the feelings expressed in the images for each of the following categories: anger, disgust, fear, joy, sadness, and surprise. In addition, we use our Word2Vec model to classify the sentiment found in the text of the user's IG publications. Figure 9 shows the sentiment and emotion percentages, where joy is the most frequent emotion available in people's photos, and neutral is the most regular sentiment available in text content.

[3]GeoPyt https://geopy.readthedocs.io/en/latest/

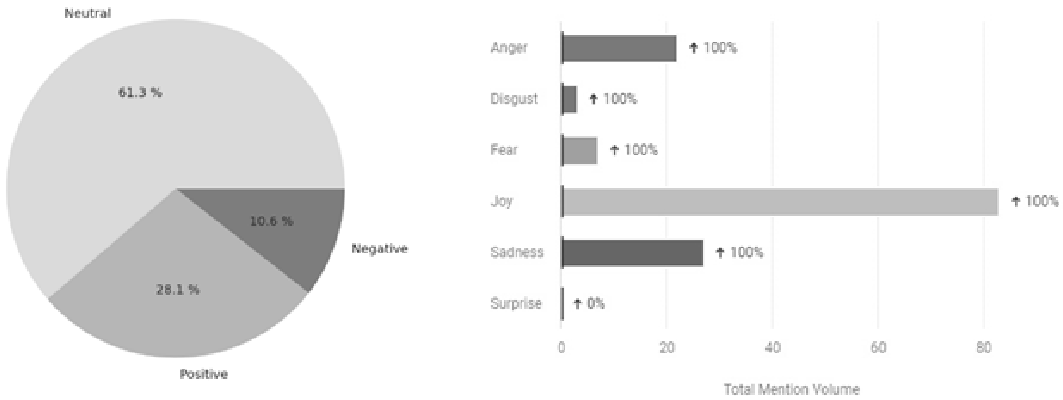


**Figure 9:** Sentiment and Emotion Percentages.

# 5. Conclusions

The proposed methodology allows obtaining more useful inferred information from any collection of publications associated with a particular hashtag on Instagram or other social networks at a low cost and effort.

The low similarity between the topics is mined from the content written by users, tourists usually, and the visual descriptions from photos because users generally refer to situations or opinions regarding the photos. In contrast, the visual analysis produces tags more related to the actual content of the images. We can also determine that the emotions transmitted in Instagram posts are better predicted using photos instead of text written by users, but only when a quality image containing a face with high confidence is available.

The results of the most frequent worldwide photo locations are similar to the most popular places on TripAdvisor. For this reason, the methodology of this work can be helpful in areas such as digital marketing, market research, opinion polls, social studies, and other fields. Also, the findings can be valuable for decision-making, creating new marketing strategies, and other studies such as consumer profile analysis, as well as being complementary to textual content from social network reports and third-party social listening platforms.

In future work, we will consider exploring stories and reels' visual content and text comments on user descriptions to evaluate if they improve prediction values using the text.

# References

[1] S. M. Mohammad, S. Kiritchenko, Using hashtags to capture fine emotion categories from tweets, Computational Intelligence 31 (2015) 301–326.

[2] F. Kunneman, C. Liebrecht, A. van den Bosch, The (un) predictability of emotional hashtags in twitter, 2014.

[3] Y. Hu, L. Manikonda, S. Kambhampati, What we instagram: A first analysis of instagram photo content and user types, 2014.
[4] J. Y. Jang, K. Han, D. Lee, No reciprocity in" liking" photos: analyzing like activities in instagram, in: Proceedings of the 26th ACM conference on hypertext & social media, 2015, pp. 273–282.
[5] D. Amanatidis, I. Mylona, I. Kamenidou, S. Mamalis, A. Stavrianea, Mining textual and imagery instagram data during the covid-19 pandemic, Applied Sciences 11 (2021) 4281.
[6] L. Manikonda, V. V. Meduri, S. Kambhampati, Tweeting the mind and instagramming the heart: Exploring differentiated content sharing on social media, 2016.
[7] K. J. Cios, W. Pedrycz, R. W. Swiniarski, Data mining and knowledge discovery, in: Data mining methods for knowledge discovery, Springer, 1998, pp. 1–26.
[8] D. Jurafsky, J. H. Martin, Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition, 2000.
[9] P. O. Hoyer, Non-negative matrix factorization with sparseness constraints., Journal of machine learning research 5 (2004).
[10] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proceedings of the IEEE 86 (1998) 2278–2324.
[11] Y. Kim, Convolutional neural networks for sentence classification, 2014.
[12] M. J. Berger, Large scale multi-label text classification with semantic word vectors, Technical report, Stanford University (2015).
[13] N. Azam, J. Yao, Comparison of term frequency and document frequency based feature selection metrics in text categorization, Expert Systems with Applications 39 (2012) 4760–4768.