

VISVESVARAYA TECHNOLOGICAL UNIVERSITY,

Jnana Sangama, Belagavi-590018.



A Mini-Project Report

On

**ANALYZING INSTAGRAM HASHTAGS: A COMPREHENSIVE TOOL
FOR POPULARITY AND RELATED TAGS**

*Submitted for the partial fulfillment of the requirement of the VIth Semester
DATA SCIENCE AND ITS APPLICATIONS (21AD62)*

Submitted in partial fulfillment of the requirement for the award of the degree of

BACHELOR OF ENGINEERING

In

Artificial Intelligence and Machine Learning

Submitted by

SUMITH SIGTIA, 10X21AI039

Under the Guidance of

Ms. RASHMI PARUTI

Assistant Professor

Department of Artificial Intelligence and Machine Learning

THE OXFORD COLLEGE OF ENGINEERING,

Bommanahalli, Bangalore 560068



Department of Artificial Intelligence and Machine Learning

THE OXFORD COLLEGE OF ENGINEERING

Bommanahalli, Bangalore 560068

2024-2025

ABSTRACT

This project presents an Instagram Hashtag Analysis tool that evaluates hashtag popularity and identifies related hashtags. Utilizing web scraping with BeautifulSoup and requests, and interactive visualization with Streamlit and Plotly, the tool provides insights into hashtag trends. It fetches post counts and related tags, storing data in a JSON file to maintain historical records. The user-friendly interface allows for dynamic analysis and visualization, offering valuable insights for social media strategy and trend monitoring.

CONTENT

Section	Description	Page Number
1. Introduction	Background and Motivation, Objectives	3
2. Literature Review	Instagram and Hashtags, Existing Tools and Methods	3-5
3. Methodology	Data Collection, Data Processing, Analysis	5-6
4. Implementation	Code Structure and Functionality, Complete code	6-8
5. Results	Example Analysis, Visualizations	8-9
6. Discussion	Interpretation of Results, Challenges, Improvements	10
7. Conclusion	Summary of Findings, Future Work	10
8. References	List of References	10

1. Introduction

Background and Motivation

Social media platforms, such as Instagram, have become integral to modern communication and marketing. Hashtags are a key feature on Instagram, enabling users to categorize content and enhance its visibility. As the usage of hashtags grows, so does the need to analyze their effectiveness to understand trends and improve engagement strategies.

Analyzing hashtags can provide insights into:

- **Popularity:** Understanding which hashtags are trending and have high engagement.
- **Related Tags:** Discovering additional hashtags that can help broaden reach and relevance.
- **Trend Monitoring:** Tracking changes in hashtag popularity over time to adapt strategies.

This project aims to address these needs by providing a tool to analyze Instagram hashtags, offering valuable data to users seeking to optimize their social media presence.

Objectives

The primary objectives of this project are:

Fetch and Analyze Hashtag Popularity:

Develop a tool to retrieve the number of posts associated with specific hashtags. Analyze these counts to gauge the popularity and impact of different hashtags.

Identify Related Hashtags:

Provide recommendations for related hashtags based on user input. Enhance the effectiveness of hashtag strategies by suggesting additional relevant tags.

Visualize Data:

Create interactive visualizations to display hashtag popularity and related hashtags. Use tools like Plotly to generate graphs and charts that help users interpret the data easily.

By achieving these objectives, the project aims to empower users with actionable insights into hashtag performance, facilitating more effective social media strategies and trend analysis.

2. Literature Review

Instagram and Hashtags

Instagram Overview: Instagram is a popular photo and video-sharing platform that allows users to post content, follow others, and engage with media through likes, comments, and shares. Launched in 2010, Instagram has evolved into a major social media network with over a billion active users globally.

Role of Hashtags: Hashtags are used on Instagram to categorize content and make it discoverable. Users include hashtags in their posts to:

- **Increase Visibility:** Posts with relevant hashtags appear in search results and hashtag feeds, reaching a broader audience.

- **Categorize Content:** Hashtags help users find content related to specific topics or interests.
- **Engage with Trends:** Trending hashtags can amplify a post's reach and connect it with current trends or events.

Impact on Social Media Strategy: Effective hashtag usage can significantly enhance engagement and visibility. Businesses, influencers, and content creators leverage hashtags to:

- **Boost Engagement:** Posts with popular hashtags tend to receive higher engagement rates.
- **Track Performance:** Analyzing hashtag performance helps refine social media strategies and identify successful trends.

Existing Tools and Methods

Hashtag Analysis Tools: Several tools are available for analyzing hashtags and their performance. These tools offer various features and capabilities:

1. Hashtagify:

- **Features:** Provides analytics on hashtag popularity, trends, and related hashtags.
- **Limitations:** Some features are only available through paid plans; may not support extensive customization.

2. RiteTag:

- **Features:** Offers real-time hashtag suggestions based on engagement metrics and trends.
- **Limitations:** Primarily focused on real-time suggestions rather than historical analysis.

3. Iconosquare:

- **Features:** Provides detailed analytics on Instagram posts, hashtags, and user engagement.
- **Limitations:** Subscription-based with limited free features; may not offer extensive hashtag exploration.

Challenges and Limitations of Existing Tools:

- **Cost:** Many advanced features require paid subscriptions, which may not be feasible for all users.
- **Limited Customization:** Some tools offer limited flexibility in terms of data analysis and visualization.
- **Data Accuracy:** Tools may face issues with data accuracy or real-time updates, impacting the reliability of insights.

Gap in Existing Solutions: The limitations of existing tools highlight the need for a customizable, open-source solution that offers:

- **Comprehensive Data Analysis:** In-depth analysis of hashtag popularity and trends.
- **Flexibility and Customization:** Options for users to tailor the tool to their specific needs.
- **Cost-Effectiveness:** A free or low-cost alternative to premium tools.

This project aims to address these gaps by developing a tool that provides detailed, customizable insights into Instagram hashtags, helping users optimize their social media strategies effectively.

3. Methodology

Data Collection

Sources:

1. Instagram:

1. **Purpose:** To retrieve the number of posts associated with specific hashtags.
2. **Technique:** Web scraping is used to access Instagram's hashtag pages and extract relevant data.
3. **Tools:** requests library for sending HTTP requests, and BeautifulSoup for parsing the HTML content.

2. Best-Hashtags.com:

1. **Purpose:** To identify related hashtags that are commonly used alongside the specified hashtags.
2. **Technique:** Web scraping is employed to access the page for related hashtags and extract the list.
3. **Tools:** requests library for HTTP requests, and BeautifulSoup for parsing the HTML content.

Challenges:

- **Rate Limiting:** Instagram may impose limits on the frequency of requests, which requires careful handling to avoid being blocked.
- **Data Accuracy:** Ensuring that the data retrieved accurately reflects the current hashtag popularity and related tags.

Data Processing

Parsing and Extraction:

- **Hashtag Data:** Extract relevant details such as the count of posts associated with each hashtag from the scraped data.
- **Related Hashtags:** Extract and compile a list of related hashtags for each specified tag.

Data Storage:

- **File Format:** The data is stored in a JSON file to maintain a persistent record of hashtag statistics and related information.
- **Purpose:** To support historical data retention and facilitate updates without losing previous data.

Data Handling:

- **Updating Records:** New hashtags and their corresponding data are appended to the existing records in the JSON file to preserve historical data.

Challenges:

- **Data Integrity:** Ensuring that new data additions do not overwrite or corrupt existing records, maintaining the accuracy and completeness of the dataset.

Analysis

Data Analysis:

- **Hashtag Popularity:** Analyze the frequency and popularity of hashtags based on the retrieved data.
- **Related Hashtags:** Identify and analyze hashtags that are commonly associated with the specified tags, providing insights into related content.

Tools and Libraries:

- **Streamlit:** Utilized to build an interactive web interface for users to input hashtags and view results.
- **Plotly:** Employed to create interactive visualizations that illustrate hashtag popularity and related tags.
- **pandas:** Used for data manipulation and to prepare data for visualization and analysis.

Challenges:

- **Visualization Accuracy:** Ensuring that the visualizations accurately represent the data and effectively convey insights to the users.
- **User Interaction:** Designing an intuitive and user-friendly interface that allows users to interact with the data and view relevant analyses.

4. Implementation

Code Structure and Functionality

Overview: The project is implemented using Python, Streamlit, and several libraries to fetch, process, and visualize Instagram hashtag data. The code is organized into different functions and components to handle various aspects of the application.

Key Components:

1. Data Fetching:

1. Functions to scrape Instagram and Best-Hashtags.com for hashtag data and related hashtags.
2. Error handling and logging are included to manage issues during data retrieval.

2. Data Processing:

1. Parsing functions to extract relevant information from the scraped HTML content.
2. Storage functions to save and update the JSON database with the latest data.

3. User Interface:

1. Streamlit is used to create an interactive web interface where users can input hashtags and view results.
2. Includes components like text inputs, number inputs, and buttons to manage user interactions.

4. Data Visualization:

1. Plotly is used to generate interactive charts and graphs to visualize hashtag popularity and related tags.
2. The visualizations help users interpret data and understand trends.

Functionality:

- **Fetching Hashtag Count:** Retrieves the number of posts associated with a hashtag.
- **Best Hashtags:** Provides a list of related hashtags to help users expand their hashtag strategy.
- **Interactive Dashboard:** Allows users to input hashtags, view results, and explore visualizations of the data.

Complete code:

```
import streamlit as st
import requests
from bs4 import BeautifulSoup
import json
import pandas as pd
import plotly.express as px
import logging
logging.basicConfig(level=logging.INFO)
st.title("Instagram Hashtag Analysis")

def get_count(tag):
    try:
        url = f"https://www.instagram.com/explore/tags/{tag}/"
        headers = {'User-Agent': 'Mozilla/5.0'}
        s = requests.get(url, headers=headers)
        soup = BeautifulSoup(s.content, 'html.parser')

        meta_tag = soup.find('meta', {'property': 'og:description'})
        if meta_tag and 'content' in meta_tag.attrs:
            content = meta_tag['content']
            match = re.search(r'([\d.]+)([KMB]?)', content)
            if match:
                number = match.group(1).replace(".", "")
                suffix = match.group(2)
                if suffix == 'K':
                    count = int(float(number) * 1_000)
                elif suffix == 'M':
                    count = int(float(number) * 1_000_000)
                elif suffix == 'B':
                    count = int(float(number) * 1_000_000_000)
                else:
                    count = int(number)
            logging.info(f"Fetch count for #{tag}: {count}")
            return count
        logging.warning(f"Count not found for #{tag}")
        return None
    except Exception as e:
        logging.error(f"Error fetching data for tag '{tag}': {e}")
        return None

def get_best(tag, topn):
    try:
        url = f"https://best-hashtags.com/hashtag/{tag}/"
        s = requests.get(url)
        soup = BeautifulSoup(s.content, 'html.parser')

        tag_box = soup.find("div", {"class": "tag-box tag-box-v3 margin-bottom-40"})
        if tag_box:
            tags = tag_box.text.split()[:topn]
            return tags
        return []
    except Exception as e:
        logging.error(f"Error fetching best hashtags for '{tag}': {e}")
        return []

@st.cache_data
def load_data():
    # Try to load from session state if it exists if 'data' not in
```



```

st.session_state:
    try:
        with open("database.json", "r") as f: st.session_state.data
            =json.load(f)
        except FileNotFoundError: st.session_state.data =
            {"hashtag_data": {}}
    return st.session_state.data

def save_data(data):
    with open("database.json", "w") as f: json.dump(data, f,
        indent=4)

data = load_data()

num_tags = st.sidebar.number_input("Select number of tags", 1, 30) tags = []
sizes = [] st.sidebar.header("Tags")
col1, col2 = st.sidebar.columns(2)

for i in range(num_tags):
    tag = col1.text_input(f"Tag {i}", key=f'tag_{i}')
    size = col2.number_input(f"Top-N {i}", 1, 10, key=f'size_{i}') tags.append(tag)
    sizes.append(size)

if st.sidebar.button("Create Hashtags"): tab_names = ["all"]
    tab_names = tab_names + [tags[i] for i in range(num_tags)] tag_tabs =
    st.tabs(tab_names)
    all_hashtags = []

    hashtag_data = []
    for i in range(num_tags):
        hashtags = get_best(tags[i], sizes[i]) for hashtag in
        hashtags:
            if hashtag in data["hashtag_data"]:
                hashtag_count = data["hashtag_data"][hashtag] else:
                hashtag_count = get_count(hashtag.replace("#", "")) if hashtag_count is
                None:
                    hashtag_count = 0 # Default to 0 if count fetching fails data["hashtag_data"][hashtag] = hashtag_count
                hashtag_data.append((f'{hashtag}<br>{hashtag_count:,}', hashtag_count))

        tag_tabs[i+1].text_area(f"Tags for {tags[i]}", " ".join(hashtags)) all_hashtags =
        all_hashtags + hashtags

    tag_tabs[0].text_area("All Hashtags", " ".join(all_hashtags)) st.header("Hashtag

Count Data")
df = pd.DataFrame(hashtag_data, columns=["hashtag", "count"])
df = df.sort_values("count")

# Save back to session state and to the file
st.session_state.data = data save_data(data)

fig = px.bar(df, x='hashtag', y='count') st.plotly_chart(fig,
    use_container_width=True)

```

5. Results

Analysis Overview: The analysis focuses on understanding the popularity of various Instagram hashtags and identifying related hashtags that can enhance engagement strategies. The tool allows users to input hashtags, retrieve their popularity counts, and explore related hashtags.

Example Hashtag Analysis:

Select number of tags

2 - +

Tags

Tag 0 Top-N 0

travel 3 - +

Tag 1 Top-N 1

foodie 2 - +

Create Hashtags

Instagram Hashtag Analysis

all travel foodie

All Hashtags

#travel #nature #travelphotography #foodie #food

Hashtag Count Data



```
database.json X app.py 5
database.json > ...
1 {
2   "hashtag_data": {
3     "#bangalore": 19000000,
4     "#mumbai": 78000000,
5     "#karnataka": 15000000,
6     "#india": 276000000,
7     "#culture": 44000000,
8     "#art": 100000000,
9     "#photography": 100000000,
10    "#travel": 74800000,
11    "#nature": 82200000,
12    "#travelphotography": 24300000,
13    "#foodie": 25700000,
14    "#food": 52900000
15  }
16 }
```

6. Discussion

Interpretation of Results: The analysis reveals that hashtags with high post counts, such as #Travel, have broad engagement potential, while related hashtags offer additional targeting opportunities. Visualizations provide clear insights into hashtag popularity and relationships, enabling users to optimize their social media strategies effectively.

Challenges:

- **Data Accuracy:** Ensuring reliable data extraction from Instagram due to potential rate limiting and changes in website structure.
- **Scalability:** Handling a large number of hashtags and maintaining performance in real-time data retrieval.
- **User Interface:** Designing an intuitive interface that balances functionality with ease of use.

Improvements:

- **Enhanced Data Accuracy:** Implementing more robust error handling and updating data retrieval methods to adapt to changes in Instagram's structure.
- **Scalability Solutions:** Optimizing data storage and retrieval processes to manage larger datasets efficiently.
- **Interface Enhancements:** Refining the user interface based on user feedback to improve usability and accessibility.

7. Conclusion

Summary of Findings: The project successfully developed a tool for analyzing Instagram hashtags, offering insights into popularity and related tags. The visualizations and interactive features provide users with valuable data to enhance their social media strategies. The tool effectively identifies popular hashtags and suggests related ones, aiding in content optimization.

Future Work:

- **Extended Features:** Integrate more advanced analytics, such as trend prediction and sentiment analysis.
- **Broader Integration:** Expand the tool to support other social media platforms beyond Instagram.
- **User Feedback:** Continuously improve the tool based on user feedback and evolving social media trends.

8. References

1. Instagram Official Website: <https://www.instagram.com>
2. Best-Hashtags.com: <https://best-hashtags.com>
3. Plotly Documentation: <https://plotly.com/python/>
4. BeautifulSoup Documentation: <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>
5. Requests Documentation: <https://docs.python-requests.org/en/latest/>