

MBDS

❖ *JOSÉ TRINDADE &
OLGA BELIAI*

DATA SCIENCE
FOUNDATIONS
FINAL PROJECT





Every day people share on Social Media their desire to go on vacation to country x and city y.

This desire is normally expressed in posts, comments or likes, on social media.

- *What if travel agencies could reach out to these accounts and recommend flights and accommodation options? Or even warn them about the travel restrictions?*
- *Wouldn't that increase the likelihood of getting sales?*

Travel agencies are in extinction and need new ways to innovate and diversify their offer and reach.



Our product helps **Outbound Sales teams, in Travel Agencies**, to identify **new customer leads**, and **tailor their offering** by providing flight and accommodation options based on their activity on Social Media.

Road Map



Extract people's activity (posts) on Reddit's Subreddits about Travelling via their official API



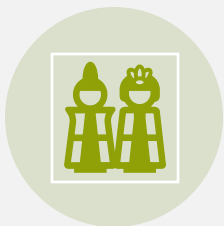
Use a ETL Pipeline to extract, prepare and store the the results accordingly in a sqlite ddbb.



Get the destination for each post using NLP Library spaCy for text processing and entity **extraction**.



Use GoogleMaps API to get geographical data about the destinations people are talking about.



Plot and visualize the top destinations mentioned



Selenium Webscrap Flights and Accomodation options from Google Travel website, to propose to potential client leads.

Files used



extract_ingestion.py

- Extracting activity from Reddit subreddits about traveling
- This Script accepts 4 parameters as inputs: 3 for accessing the Reddit API and 1 as the name of subreddit to retrieve the activity of the users: *python3 extract_ingestion.py CLIENT_ID CLIENT_SECRET USER_AGENT SUBREDDIT_NAME*

process.py

- Interpreting posts from Reddit and sourcing location and travel-restrictions features from GoogleMaps API and other dataset
- This Script accepts 1 parameters which is the key to access the project created on Google Cloud Console, namely the GoogleMaps API: *python3 process.py GOOGLE_PROJECT_KEY*

visualizations.ipnyb

- Visualizing the processed Reddit posts

GenRedditComm.py

- Webscrapping Google Travel and **Generating a Reddit comment**
- This Script accepts 1 parameter which is the Post_ID: *python3 genRedditComm.py POST_ID*

ETL



➤ **raw_reddits table**

raw_reddits(Post_ID , Title , Body , Author , Creation_Date , Score , Num_Comments , URL, Over_18)

➤ **proc_log table**

proc_log(Post_ID, proc_time)

➤ **proc_posts table**

proc_posts(Post_ID, URL , Destinations, Latitude, Longitude, Country_Full_Name, Requirement)

New DatabaseOpen DatabaseWrite ChangesRevert ChangesOpen ProjectSave ProjectAttach DatabaseClose Database

Open an existing database fileDatabase StructureBrowse DataEdit PragmasExecute SQL

Table: raw_redditsFilter in any column

	Post_ID	Title	Body	Author	Creation_Date	Score	Num_Comments	URL	Over_18
	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter
1	1b8tcb6	What temperature range have you experienced?	My highest was 42C (107F) in South Africa, my lowest was...	Gingerbread_Cat	2024-03-07	58	285	https://www.reddit.com/r/travel/comments/1b8tcb6/...	0
2	1b8hx9x	People who visited Russia as a tourist before the war, wha...	In the past I've had an interest in visiting Russia, but for ...	UltimateLazer	2024-03-07	342	334	https://www.reddit.com/r/travel/comments/1b8hx9x/...	0
3	1b8t3tq	Solo travelers, how often do you meet new people, how ...	Meeting new people is one of the best things. Just came ...	Achilles982	2024-03-07	48	48	https://www.reddit.com/r/travel/comments/1b8t3tq/...	0
4	1b83h16	Bhutan has been nothing short of magical.		sendhelpandthensome	2024-03-06	1058	80	https://www.reddit.com/gallery/1b83h16	0
5	1b8ysk8	Afraid I'll pick the wrong country for my first ever oversea...	25m was hoping this year to take my first ever trip to ...	Frogs_and_Feels	2024-03-07	9	91	https://www.reddit.com/r/travel/comments/1b8ysk8/...	0
6	1b8ft6i	What's the most unique pair of countries you've been to?	My wife was involved in a team building 'odd facts about ...	MonsieurAK	2024-03-07	162	316	https://www.reddit.com/r/travel/comments/1b8ft6i/...	0
7	1b8sfpo	Train Ninja scam - avoid this web, it is scam with train ...	I travel from Seoul to Busan with Bullet train and I review...	kadoslav	2024-03-07	15	15	https://www.reddit.com/r/travel/comments/1b8sfpo/...	0
8	1b8u977	RyanAir customer service	Backstory...	arintonakost2	2024-03-07	10	22	https://www.reddit.com/r/travel/comments/1b8u977/...	0
9	1b8j3nq	Talking to my friend gave me a perspective on how ...	We were just talking about his recent trip to Bangkok and ...	potalliega	2024-03-07	54	16	https://www.reddit.com/r/travel/comments/1b8j3nq/...	0
10	1b8vw9z	There are a lot of air lines that don't participate in the ...	Apparently there's quite the steep commission generated ...	lowspeed	2024-03-07	6	10	https://www.reddit.com/r/travel/comments/1b8vw9z/...	0
11	1b8tg58	Clear+Precheck is significantly slower than Precheck out ...	Every flight out of Newark the last 6 months has been ...	42ATK	2024-03-07	9	14	https://www.reddit.com/r/travel/comments/1b8tg58/...	0
12	1b8v6h4	Visiting Montreal for the first time, requesting an itineary ...	Hello!	homelybliscuit	2024-03-07	5	2	https://www.reddit.com/r/travel/comments/1b8v6h4/...	0
13	1b807t9	What's your biggest travel screw up?	Needing reassurance...	emperorrryping	2024-03-06	460	630	https://www.reddit.com/r/travel/comments/1b807t9/...	0
14	1b8yewl	What's cheaper in LA: Car rental or Uber?	Hello Everyone!	Jiulitajournal	2024-03-07	5	11	https://www.reddit.com/r/travel/comments/1b8yewl/...	0
15	1b8wtmh	How Can I Get My Lost Suitcase from Swissport Stanstea...	Help! #Pegasus Airlines going from Istanbul to Stanstead ...	Affectionate_Menu617	2024-03-07	3	1	https://www.reddit.com/r/travel/comments/1b8wtmh/...	0
16	1b8z2z2	Traveling withn Sweden - suggestions?	Hi folks, I'm looking for any advice you have about our trip...	notsaroundtown	2024-03-07	2	2	https://www.reddit.com/r/travel/comments/1b8z2z2/...	0
17	1b8ysdd	3 month long stay in tokyo?	Hi! my friend and I plan to stay in tokyo for about 3 month...	Juphyias	2024-03-07	2	3	https://www.reddit.com/r/travel/comments/1b8ysdd/...	0
18	1b8osqs	No "liquid rules" at Linate airport?	I'm traveling from Milano Linate back to Berlin this week...	fragrium	2024-03-07	8	6	https://www.reddit.com/r/travel/comments/1b8osqs/...	0
19	1b8xe9i	Bangkok to South Thailand transport suggestions?	Traveling from Bangkok to Krabi, south Thailand (doesn't ...	Delicious_Loquat_893	2024-03-07	2	4	https://www.reddit.com/r/travel/comments/1b8xe9i/...	0
20	1b8e52v	Driving vs flying to Florida from Pennsylvania	My friend and I booked a trip to Miami in July and she is ...	TayZhunna	2024-03-06	34	71	https://www.reddit.com/r/travel/comments/1b8e52v/...	0
21	1b91uwx	6 Day Switzerland Trip	Hi! I'm taking a 6 day trip to Switzerland (May 17-22) to ...	adityaparmar3	2024-03-07	1	1	https://www.reddit.com/r/travel/comments/1b91uwx/...	0
22	1b91ubg	Albania recommendations in May	Albania ...	Equivalent_Length_56	2024-03-07	1	0	https://www.reddit.com/r/travel/comments/1b91ubg/...	0
23	1b91t38	Korean air won't allow seat upgrades?	We are flying with KoreanAir next month and have econo...	Xandar24	2024-03-07	1	0	https://www.reddit.com/r/travel/comments/1b91t38/...	0
24	1b91m7	Afraid I'll pick the wrong country for my first ever oversea...	25m was hoping this year to take my first ever trip to ...	bageiff	2024-03-07	0	0	https://www.reddit.com/r/travel/comments/1b91m7/...	0

1 - 25 of 101Go to: 1

New DatabaseOpen DatabaseWrite ChangesRevert ChangesOpen ProjectSave ProjectAttach DatabaseClose Database

Database StructureBrowse DataEdit PragmasExecute SQL

Table: proc_postsFilter in any column

	Post_ID	URL	Destinations	Latitude	Longitude	Country_Full_Name	Requirement
	Filter	Filter	Filter	Filter	Filter	Filter	Filter
1	1b8tcb6	https://www.reddit.com/r/travel/comments/1b8tcb6/...	south africa	-30.559482	22.937506	South Africa	e-visa, 30 visa-free days, visa required, 90 visa-free days
2	1b8tcb6	https://www.reddit.com/r/travel/comments/1b8tcb6/...	lapland	67.9222304	26.5046438	NULL	NULL
3	1b8hx9x	https://www.reddit.com/r/travel/comments/1b8hx9x/...	russia	61.52401	105.318756	Russia	30 visa-free days, e-visa, 90 visa-free days, visa required, ...
4	1b8hx9x	https://www.reddit.com/r/travel/comments/1b8hx9x/...	ukraine	48.379433	31.16558	Ukraine	30 visa-free days, visa required, 90 visa-free days, visa free...
5	1b8t3tq	https://www.reddit.com/r/travel/comments/1b8t3tq/...	milan	45.4642035	9.189982	Italy	visa required, 90 visa-free days, visa free
6	1b8t3tq	https://www.reddit.com/r/travel/comments/1b8t3tq/...	georgia	42.315407	43.356892	Georgia	90 visa-free days, 30 visa-free days, visa required, 360 vis...
7	1b8t3tq	https://www.reddit.com/r/travel/comments/1b8t3tq/...	austria	47.516231	14.550072	Austria	visa free, visa required, 90 visa-free days
8	1b83h16	https://www.reddit.com/gallery/1b83h16	bhutan	27.514162	90.433601	Bhutan	14 visa-free days, e-visa
9	1b8ft6i	https://www.reddit.com/r/travel/comments/1b8ft6i/...	andorra	42.506285	1.521801	Andorra	visa free, 90 visa-free days, visa required
10	1b8sfpo	https://www.reddit.com/r/travel/comments/1b8sfpo/...	busan	35.2100142	129.0688702	South Korea	30 visa-free days, 180 visa-free days, 60 visa-free days, e-...
11	1b8sfpo	https://www.reddit.com/r/travel/comments/1b8sfpo/...	seoul	37.5518911	126.9917937	South Korea	30 visa-free days, 180 visa-free days, 60 visa-free days, e-...
12	1b8u977	https://www.reddit.com/r/travel/comments/1b8u977/...	london	51.5072178	-0.1275862	United Kingdom	180 visa-free days, visa required, visa free, e-visa
13	1b8j3nq	https://www.reddit.com/r/travel/comments/1b8j3nq/...	jakarta	-6.1944491	106.8229198	Indonesia	e-visa, 30 visa-free days, visa on arrival
14	1b8j3nq	https://www.reddit.com/r/travel/comments/1b8j3nq/...	bangkok	13.7563309	100.5017651	Thailand	visa on arrival, 90 visa-free days, visa required, e-visa, 30 ...
15	1b8tg58	https://www.reddit.com/r/travel/comments/1b8tg58/...	newark	40.735657	-74.1723667	United States	visa required, e-visa, 180 visa-free days, visa free
16	1b8tg58	https://www.reddit.com/r/travel/comments/1b8tg58/...	precheck	NULL	NULL	NULL	NULL
17	1b8v6h4	https://www.reddit.com/r/travel/comments/1b8v6h4/...	fairmont	NULL	NULL	NULL	NULL

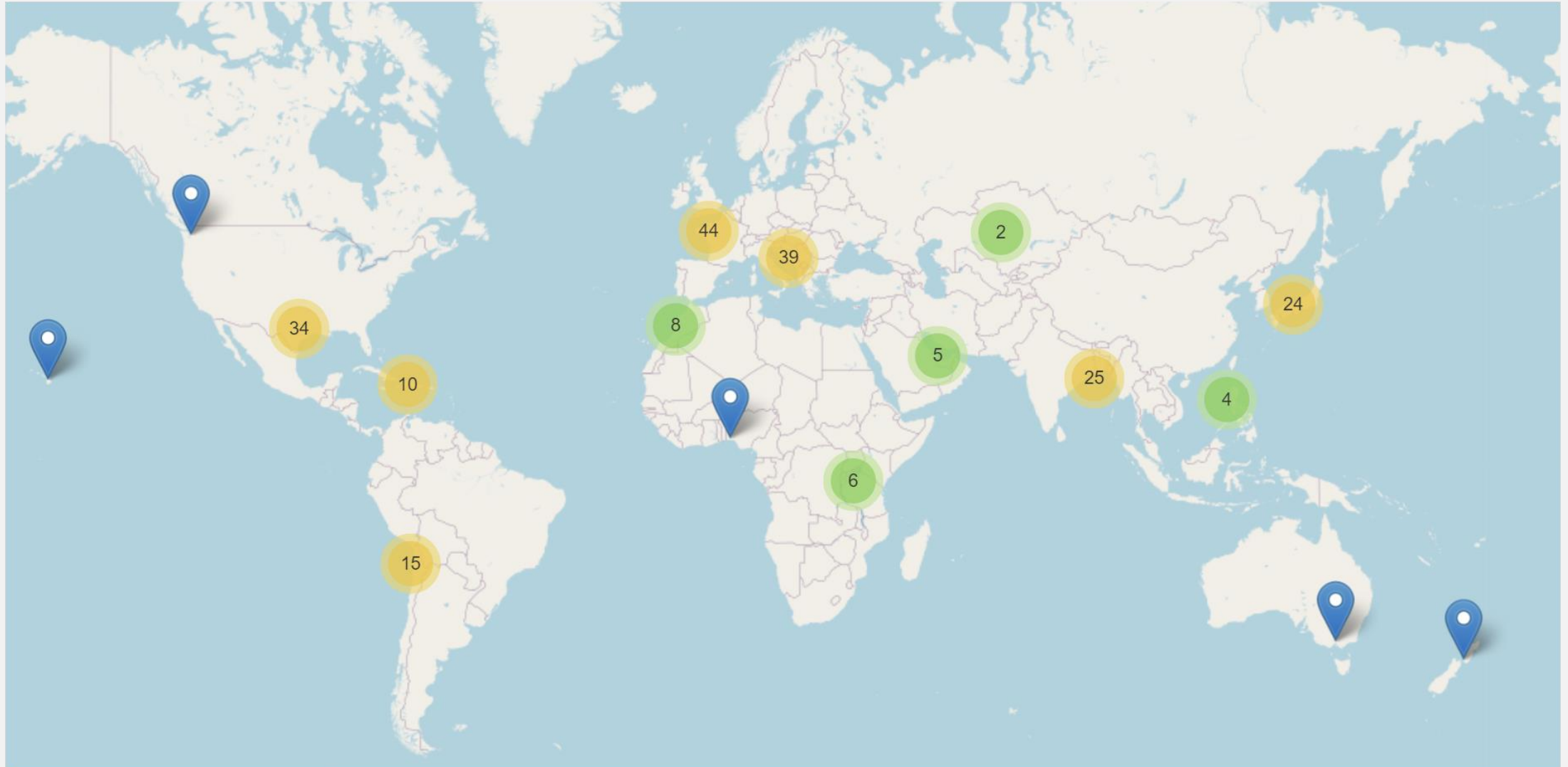
New DatabaseOpen DatabaseWrite Changes

Table: proc_logFilter in any column

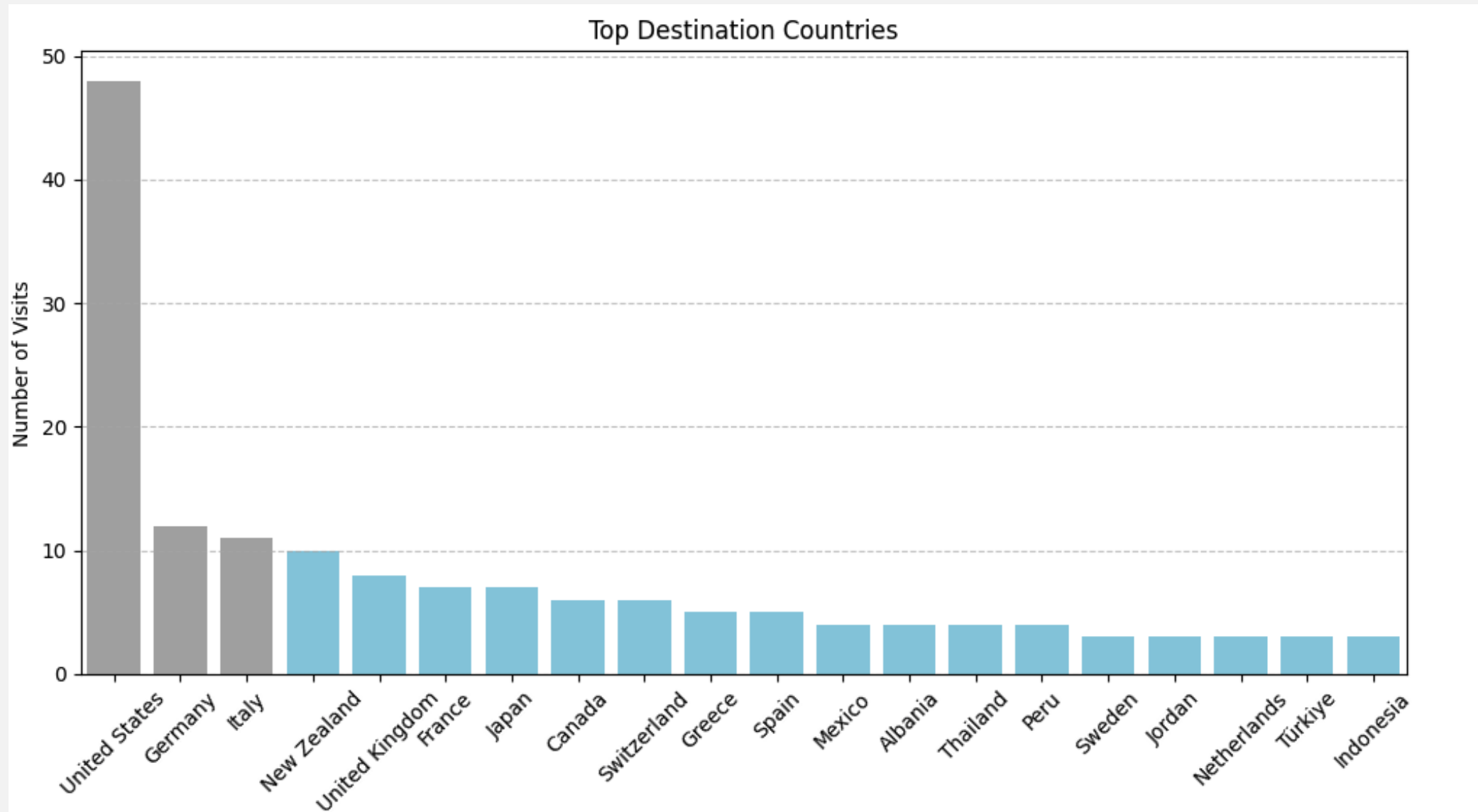
	Post_ID	proc_time
	Filter	Filter
1	1b8tcb6	07/03/2024 21:45:54
2	1b8hx9x	07/03/2024 21:45:54
3	1b8t3tq	07/03/2024 21:45:54
4	1b83h16	07/03/2024 21:45:54
5	1b8ysk8	07/03/2024 21:45:54
6	1b8ft6i	07/03/2024 21:45:54
7	1b8sfpo	07/03/2024 21:45:54
8	1b8u977	07/03/2024 21:45:54
9	1b8j3nq	07/03/2024 21:45:54
10	1b8vw9z	07/03/2024 21:45:54
11	1b8tg58	07/03/2024 21:45:54
12	1b8v6h4	07/03/2024 21:45:54
13	1b807t9	07/03/2024 21:45:54
14	1b8yewl	07/03/2024 21:45:54
15	1b8wtmh	07/03/2024 21:45:54
16	1b8zhz2	07/03/2024 21:45:54
17	1b8ysdd	07/03/2024 21:45:54
18	1b8osqs	07/03/2024 21:45:54
19	1b8xe9i	07/03/2024 21:45:54
20	1b8e52v	07/03/2024 21:45:54
21	1b91uwx	07/03/2024 21:45:54
22	1b91ubg	07/03/2024 21:45:54
23	1b91t38	07/03/2024 21:45:54
24	1b91m7	07/03/2024 21:45:54

1 - 25 of 101

Visualization – Destinations mentioned in the Reddit posts



Top 20 Countries mentioned



Retrieved visa requirements & restrictions for destinations with 2024 Passport Index Dataset

- Number of visa-free days, where available
- Visa free
- Visa on arrival
- E-visa
- No admission
- Covid ban



Use-Case example

CLI: python3 genRedditComm.py 1b8sfpo



Query for 1b8sfpo



WebScape Flight_Price



WebScape Hotel_Price and Hotel_Name



Generate .txt file with post info and text content to publish as a comment on Reddit

https://www.reddit.com/r/travel/comments/1b8sfpo/train_ninja_scam_avoid_this_web_it_is_scam_with/

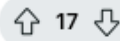


r/travel • 17 hr. ago
kadoslav



Train Ninja scam - avoid this web, it is scam with train tickets.

I travel from Seoul to Busan with Bullet train and I reviewed some google search, ok, found one for around 100.000 krw, expensive but I have no clue how much it should be, ok. Now we are about to go and just quickly checked the ticket before departure and there is price 59.000 on ticket itself. So ticket was more than 41000 krw overpriced! They paid a lot of google ads to block any normal results and make sure to spam your google search with that website. Be more careful than I was.



17



17



Share



RedditComment_1b8sfpo_seoul.txt

```
URL of the reddit post: https://www.reddit.com/r/travel/comments/1b8sfpo/
train_ninja_scam_avoid_this_web_it_is_scam_with/
```

```
**Hi!! Do you want to travel to seoul?**
Fly there for 803 euros
Stay at Chris Hostel for 32.0 euros
The **travel restrictions** can be the following:
- 30 visa-free days
- 180 visa-free days
- 60 visa-free days
- e-visa
- visa required
- 90 visa-free days
```

```
Please check the exact restrictions applied to your country before planning your trip.
```

```
**TRAVEL HERE: [FlyLeads.com] (https://www.flyleads.com/seoul)**
```

From Reddit comments we've retrieved the posts authors:

With the information about comments authors and the destinations mentioned by them we can target the users with ads showing the best weekend flight & hotel combo deals.

The lowest prices have been webscraped from Google Travel search for flights and accomodations with a timeframe +60 days from the closest Friday.

Note: as we were unable to retrieve the author's country a randon departure airport was chosen (Frankfurt)



Future Work

Here are some points for improving our project:

Implement Advanced NLP: Utilize state-of-the-art Natural Language Processing (NLP) techniques with context-aware models to enhance the interpretation of social media posts and accurately extract travel intentions.

Scalable Database Migration: Transition to a more robust and scalable database system, such as PostgreSQL or MongoDB, to improve data management efficiency and query response times.

Real-Time Analytics: Develop a system for real-time analytics and processing to capture and respond to travel-related social media activity promptly.

Geolocation Accuracy: Integrate more sophisticated geolocation methods to accurately determine user locations and provide more relevant travel recommendations.

Personalized Travel Packages: Employ machine learning algorithms to offer personalized travel packages based on user interests, budget, and behavior analysis.

CRM Integration and Lead Management: Facilitate seamless integration with CRM systems to enhance lead tracking, customer interactions, and conversion rates.

Resources:

- Reddit API - <https://old.reddit.com/prefs/apps/>
- GoogleMaps API - <https://console.cloud.google.com/>
- 2024 Passport Index Dataset: <https://github.com/ilyankou/passport-index-dataset>
- Google Travel Flights & Hotels: <https://www.google.com/travel>

