Date:

Exercise - 6

ANALYZING WEBSITE PERFORMANCE ON SEARCH ENGINE USING SEO AUDIT

AIM:

To identify issues affecting the organic search performance of a website using an SEO Audit.

OBJECTIVE:

Check the status of the following issues on the website:

- Technical SEO issues
- Website structure issues
- On-page SEO issues
- Off-site problems
- User experience issues
- Content gaps and opportunities
- Competitive marketplace insights

REQUIRED TOOLS:

- experte.com
- GTmetrix
- Spotibo
- BrokenLinkCheck
- Google Search Console
- Google Analytics
- Other free tools

PROCEDURE:

1. On-Page Audit

- a. Basics
- Site Indexed or not?

Use the site: keyword in Google Search to check if the site is indexed.

• CTA Included or not?

Ensure Call-To-Actions (CTAs) are available in all portions of the website.

Necessary Pages Exist or not?

Verify the presence of essential pages like About Us, Contact Us, Privacy Policy, Disclaimer, and Terms and Conditions.

• SSL Certificate Used or not?

Check if the lock symbol and HTTPS are present in the website URL.

• SSL Redirect or not?

Ensure HTTPS is available on all web pages.

• Is the Site Mobile Friendly?

Use experte.com to check mobile-friendliness.

Mobile Spend?

Use GTmetrix to check LCP (closer to 2.5 seconds), TBT (less than 300 milliseconds), and CLS score (less than 0.1).

b. Technical

Sitemap Available or not?

Verify if the sitemap.xml file is present.

Robots.txt Available or not?

Check for the presence of the robots.txt file.

How Many Pages Have Thin Content?

Use Spotibo to check for pages with low word content.

How Many Pages Have Duplicate Content?

Use Spotibo to identify duplicate content.

• How Many Pages Have Broken Links?

Use BrokenLinkCheck to find broken links.

• 301 Redirect?

Use Spotibo to check for 301 redirects.

c. Optimization (using Spotibo)

Meta Titles and Meta Descriptions

- URL Structure
- H1 Tags Canonical Tags
- Image Alt Tags

2. Off-Page Audit

• Backlink Profile

Analyze using Ubersuggest or some other free tools available on the web.

• Trust Flow & Citation Flow

Check using tools like Majestic or some other free tools available on the web.

3. Analytics Audit

• Check Traffic, Audience

Use Google Analytics to assess traffic and audience metrics.

• Check Bounce Rate

Analyze bounce rate using Google Analytics.

4. Search Console Audit

• Crawling and Indexing Errors

Use Google Search Console to identify crawling and indexing issues.

Penalties

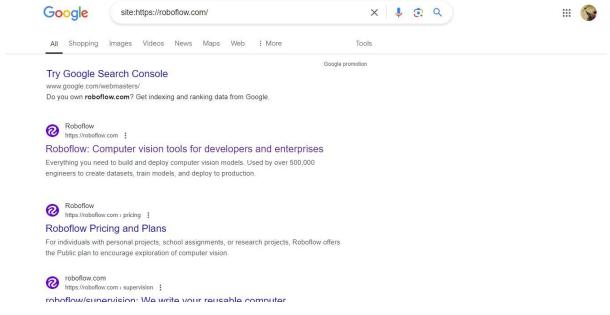
Check for any manual actions or penalties in Google Search Console

OUTPUT:

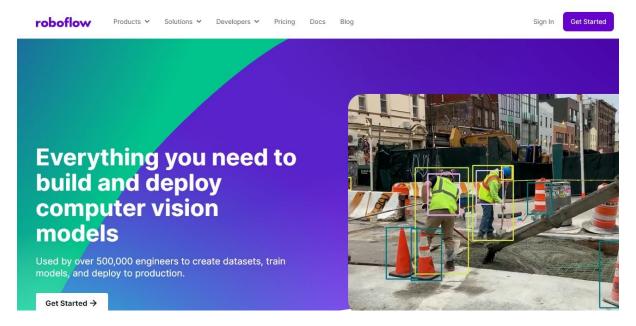
SEO AUDIT for https://roboflow.com/

Basic

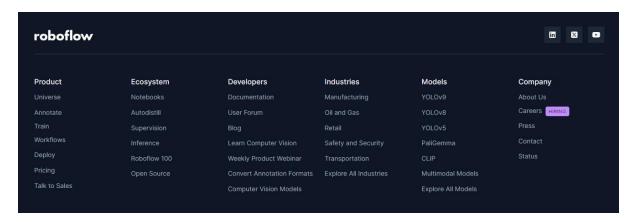
Site indexed



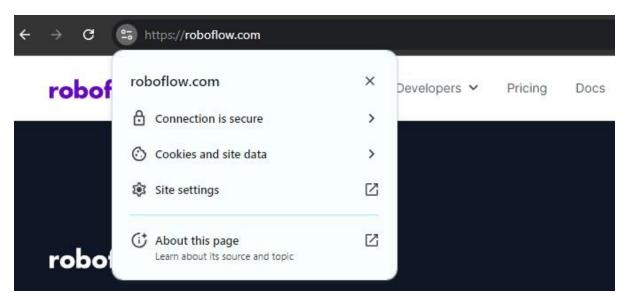
CTA not indexed



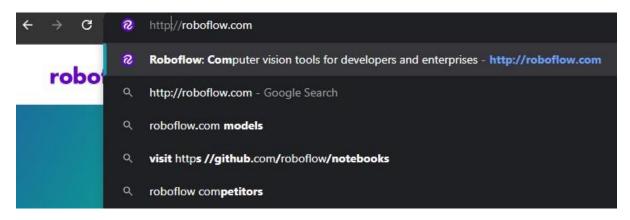
Necessary pages

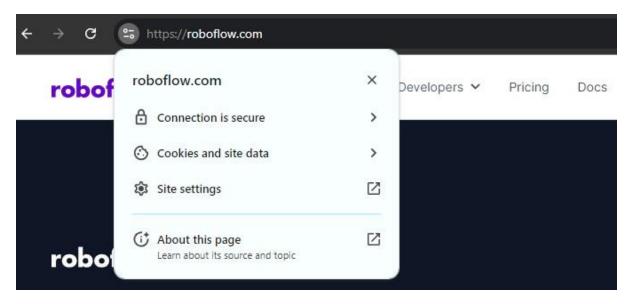


SSL certificate

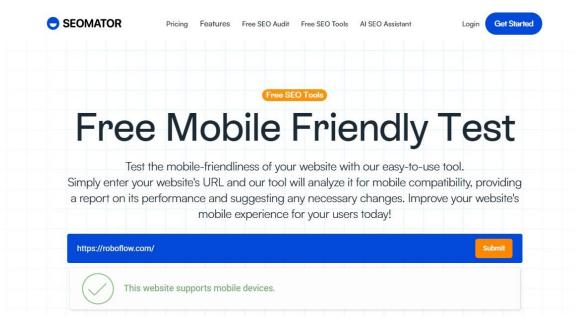


SSL redirect

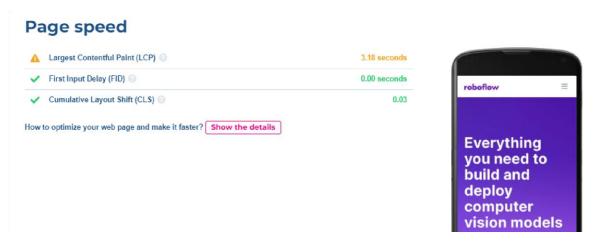




Mobile friendly



Mobile speed



Technical

Sitemap

```
C
               25 roboflow.com/sitemap.xml
<?xml version="1.0" encoding="UTF-8"?>
<urlset xmlns="http://www.w3.org/1999/xhtml">
   <url>
       <loc>
          https://roboflow.com
       </loc>
   </url>
   <url>
       <loc>
          https://roboflow.com/contact
       </loc>
   </url>
   <url>
       <loc>
          https://roboflow.com/maintenance
       </loc>
   </url>
   <url>
          https://roboflow.com/terms-march-10-2023
       </loc>
   </url>
   <url>
       <loc>
          https://roboflow.com/privacy
       </loc>
   </url>
   <url>
       <loc>
          https://roboflow.com/sales
       </loc>
   </url>
   <url>
       <loc>
          https://roboflow.com/features
       </loc>
   </url>
   <url>
       <loc>
          https://roboflow.com/about
       </loc>
```

Robots.txt available



Duplicate content

Duplications and canonicals

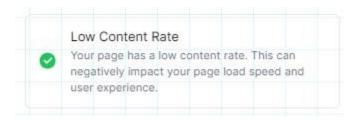


✓ 0 Pages with a 301 redirect ▼ show details → Turn off

Broken link check



Thin link content



Optimization

Meta Titles and Meta Descriptions

10	Title tag is empty or not set ▼	show details Turn off
, ,	Title tag is empty of not set	SHOW details Turn off
10	<u>Duplicate title</u> ▼	show details O
A 6	Title tag is too short ▼	show details Turn off
10	Title tag is too long ▼	show details O Turn off
V 0	There is more than one title tag in <head> ▼</head>	show details Turn off

A 1	Meta description is empty or not set ▼	show details Turn of
V 0	<u>Duplicate meta descriptions</u> ▼	show details Turn of
A 1	Meta description is too short ▼	show details Turn of
<u>A</u> 7	Meta description is too long ▼	show details Turn of
V 0	Pages with more than one meta description in <head> ▼</head>	show details Turn of

URL Structure

H1 Tags



Canonical Tags



Image Alt Tags



Off page audit

Back link profile

Referring domain count

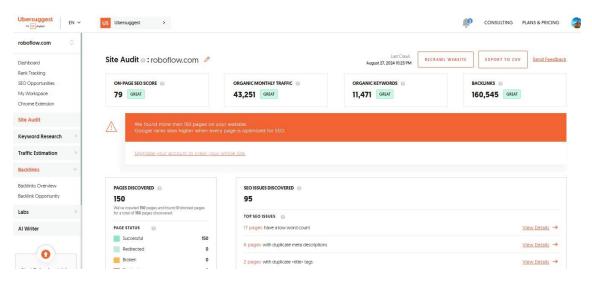
Referring domain authority

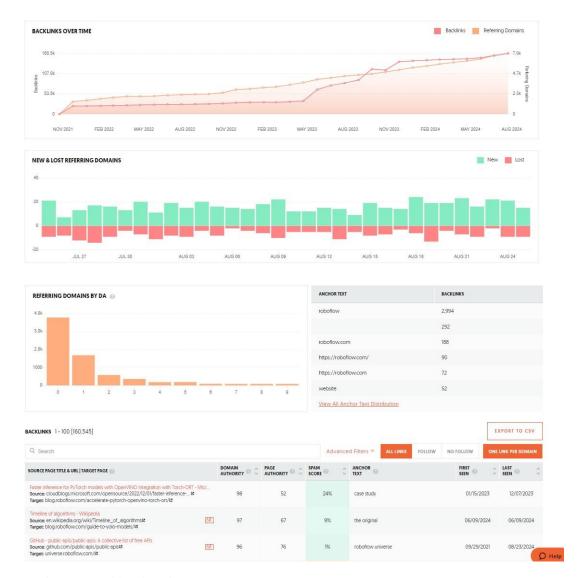
Backlink count

Anchor text spam

Link graph

No follow and Do follow link count

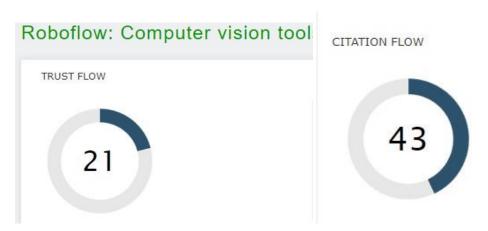




Trust flow and Citation flow

Trust flow is based on quality of backlink - important $% \left(\mathbf{r}_{0}\right) =\mathbf{r}_{0}$

Citation flow is based on number of backlinks

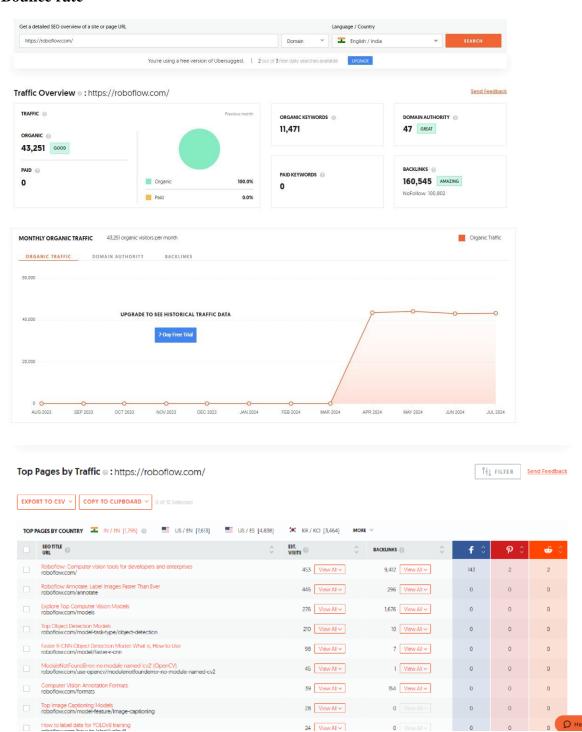


Analytics audit

Traffic

Audience

Bounce rate



Search Console Audit

Crawling and Indexing errors

Indexation and crawling



Penalties

RESULT:

Thus, issues affecting the organic search performance of a website using an SEO Audit have been identified successfully.

	4
11	ota.
v	au.

Exercise – 7

GOOGLE KEYWORD PLANNER - DISCOVER KEYWORDS AND GET SEARCH VOLUME & FORECASTS

AIM:

Explore Google Keyword Planner to discover new keywords and learn how to get search volume and forecasts for better keyword selection in your marketing campaigns

REQUIREMENTS:

Google Ads account. Internet connection. List of potential topics or products.

PROCEDURE:

- 1. Access Google Keyword Planner
 - Go to https://ads.google.com/ and log in.
- 2. Navigate to Keyword Planner
 - Click "Tools & Settings" > "Keyword Planner."
- 3. Discover Keywords Start with Keywords
 - Choose "Discover new keywords."
 - Select "Start with keywords."
 - Enter seed keywords (e.g., "running shoes," "fitness tracker").
 - Refine search options (optional).
 - Click "Get Results."

4. Discover Keywords - Start with a Website

- Choose "Discover new keywords."
- Select "Start with a website."
- Enter your website's URL.
- Refine search options (optional).
- Click "Get Results."

5. Review Keyword Suggestions

- Examine the list of relevant keywords.
- Check average monthly searches and competition.

6. Get Search Volume and Forecasts

- Select desired keywords.
- Click "Add to plan."
- Review selected keywords in "Keywords in your plan" tab.

• Click "Review plan."

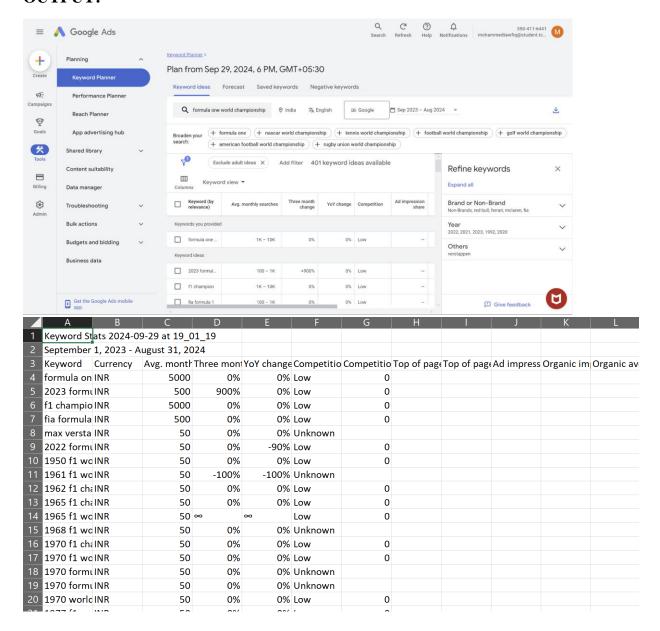
7. Review Search Volume and Forecasts

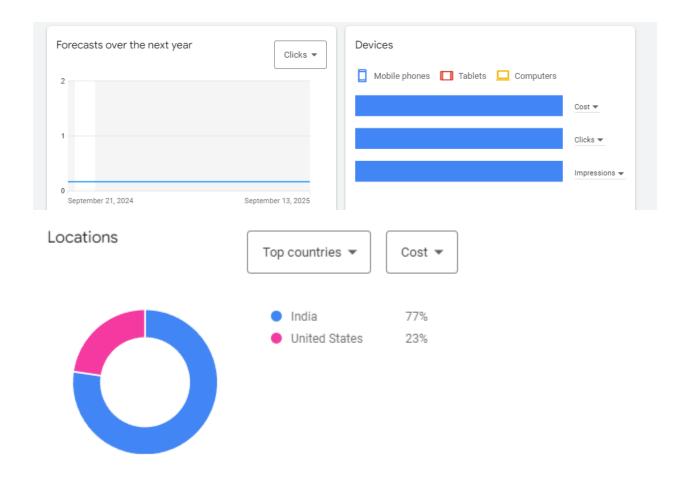
- Check estimated performance metrics (clicks, impressions, costs, etc.).
- Use data to make informed decisions for your campaigns.

8. Export Keywords and Forecasts (Optional)

Click "Download keyword ideas" to export data (CSV or Excel).

OUTPUT:





RESULT:

Thus, Exploring Google Keyword Planner to discover new keywords and learn how to get search volume and forecasts for better keyword selection in your marketing campaigns are implemented successfully.

Date:

Exercise - 8

KEYWORD RESEARCH EXCELLENCE: UNLOCKING LONG-TERM, HIGH-VOLUME, LOW-COMPETITION KEYWORDS FOR SEO

AIM:

Learn how to find high-quality keywords for SEO that have long-term potential, good search volume, and low competition.

REQUIREMENTS:

We will utilize three powerful keyword research tool

- Keywordsheeter
- Google Keyword Planner
- Ahrefs.com.

PROCEDURE:

1. Introduction to Keyword Research

- Understand the importance of keyword research for SEO.
- Recognize the benefits of finding keywords with long-term potential, high search volume, and low competition.

2. Keywordsheeter

- Use Keywordsheeter to generate keyword ideas related to a specific topic or niche.
- Export the results to a CSV file for further analysis.

3. Google Keyword Planner

- Learn how to use Google Keyword Planner to obtain search volume data for the keywords from Keywordsheeter.
- Identify keywords with good search volume that match your website's content.

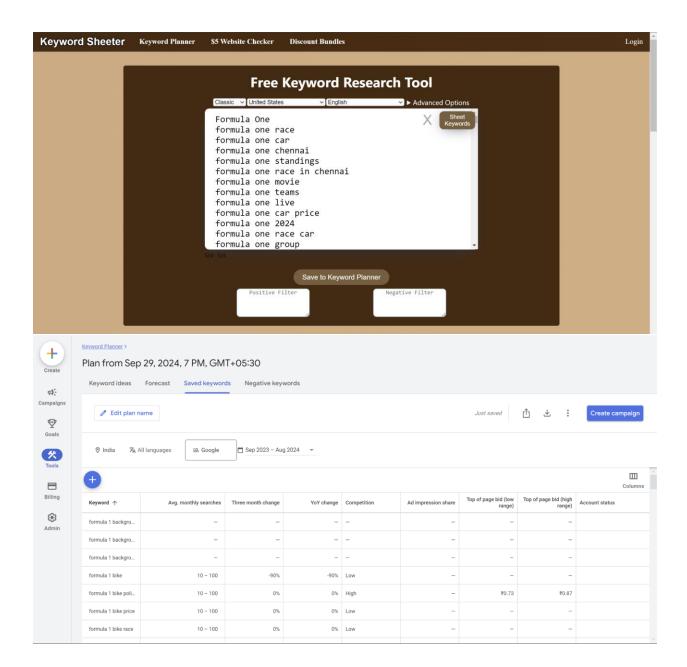
4. Ahrefs.com

- Discover Ahrefs.com as a tool to assess keyword difficulty.
- Evaluate the competition for the selected keywords from Google Keyword Planner.

5. Keyword Selection

 Combine data from Keywordsheeter, Google Keyword Planner, and Ahrefs.com to choose the best keywords with long-term potential, good search volume, and low difficulty.

OUTPUT:



RESULT:

Thus, Learning how to find high-quality keywords for SEO that have long-term potential, good search volume, and low competition is successful.

D 4	
บารา	ω.
Dai	•

Exercise - 9

SOCIAL MEDIA AUDITING USING FANPAGE KARMA WEBSITE

AIM:

To perform a social media audit using the Fanpage Karma website and create various reports using the available options in the categories: Analytics, Engage, Discovery, and More.

TOOLS NEEDED:

- Computer with internet access
- Fanpage Karma account (Sign up if not already done)
- Social media account to audit (e.g., Facebook page, Twitter profile, Instagram account, etc.)

PROCEDURE:

1. Sign in to Fanpage Karma

• Log in to your Fanpage Karma account using your credentials on the website.

2. Select the Social Media Account

• Choose the social media account you want to audit, whether it's your own account or a practice account.

3. Analytics:

- Navigate to the "Analytics" section to gather key insights about your social media account's performance.
- Explore options such as follower growth, engagement metrics, post frequency, and reach.
- Prepare the following Performance Overview Report:
 - Utilize follower growth data, engagement metrics, and reach statistics.
 - Highlight trends and key performance indicators over a specific period.

4. Engage:

- Visit the "Engage" section to assess the engagement level of your posts and interactions with your audience.
- Explore options like top engaging posts, interactions history, and sentiment analysis.
- Prepare the following Engagement Analysis Report:
 - Compile top engaging posts, interactions history, and sentiment analysis.
 - Analyze which content resonates most with your audience and how interactions have evolved.

5. Discovery:

- Explore the "Discovery" section to gain insights into your audience demographics, locations, and interests.
- Utilize options like audience demographics, audience locations, and audience interests.
- Prepare the following Audience Demographics and Interests Report:
 - o Combine audience demographics, locations, and interests data.
 - Create a comprehensive profile of your target audience, aiding content customization.

6. More:

- Navigate to the "More" section to access additional features that provide deeper insights.
- Explore options like "Compare Yourself with Others," "Industry Overview," and "Analyze."
- Prepare the following Competitor Comparison Report:
 - Utilize the "Compare Yourself with Others" feature to assess your account's performance against similar accounts.
 - Identify areas of strength and areas for improvement.
- Prepare the following Industry Insights Report:
 - Utilize the "Industry Overview" options, such as Facebook Catalogue, Social Media Index, and news analysis.
 - Gain insights into your industry's social media landscape and trends.
- Prepare the following Content Analysis Report:
 - Utilize the "Analyze" options, such as Post Tracking and Google Analytics integration.
 - Analyze post performance and gather data from Google Analytics for deeper insights.

OUTPUT:





RESULT:

Thus, a social media audit using the Fanpage Karma website and creating various reports using the available options in the categories: Analytics, Engage, Discovery were implemented successfully.

Data	,
Date	•

Exercise – 10

IMAGE SEO FOR BLOGGER

AIM:

To learn and apply image SEO (Search Engine Optimization) techniques for bloggers, focusing on optimizing images within blog posts to improve search engine visibility and user experience.

TOOLS NEEDED:

- Computer with internet access
- Blogging platform or content management system (e.g., WordPress, Blogger, etc.)
- Images to be used in blog posts

PROCEDURE:

1. Select a Blog Post:

• Choose an existing blog post or create a new one where you want to optimize images for SEO.

2. Image Selection:

• Select relevant and high-quality images that complement your blog content.

3. Image Optimization:

- Optimize each image using image editing software (e.g., Adobe Photoshop) or online tools to adjust dimensions and quality.
- Rename the image file to something descriptive and related to the content.

4. Alt Text:

• Add descriptive and concise alt text to each image. Alt text should convey the image's content and purpose for accessibility and SEO benefits.

5. Image Compression:

• Compress images to reduce their file size without compromising quality. Use tools like TinyPNG or ImageOptim.

6. Image File Format:

• Choose the appropriate image file format. Use JPEG for photographs and images with gradients, and PNG for images with transparency or simpler graphics.

7. Image Dimensions:

• Resize images to match the dimensions needed for your blog layout. Avoid using larger images that slow down page loading.

8. Use Descriptive Filenames:

• Use descriptive filenames that include relevant keywords. For example, instead of "IMG12345.jpg," use "best-chocolate-cake.jpg."

9. Image Captions:

• If relevant, add captions to your images. Captions provide context and engagement.

10. Embed Images in Blog Post:

• Embed optimized images within your blog post where they enhance the content.

11. Image Placement:

• Place images near relevant text to create a cohesive visual experience.

12. Schema Markup:

• If applicable, add schema markup (structured data) to your images to provide search engines with more context about the image.

13. Check Mobile Responsiveness:

• Ensure that your images are mobile-responsive and adapt well to different screen sizes.

14. Image Sitemap:

• If you have multiple images, consider creating an image sitemap and submitting it to search engines using Google Search Console.

15. Preview and Test:

- Preview your blog post to ensure that images are properly displayed and load quickly.
- Test the blog post on different devices and browsers to ensure optimal image display.

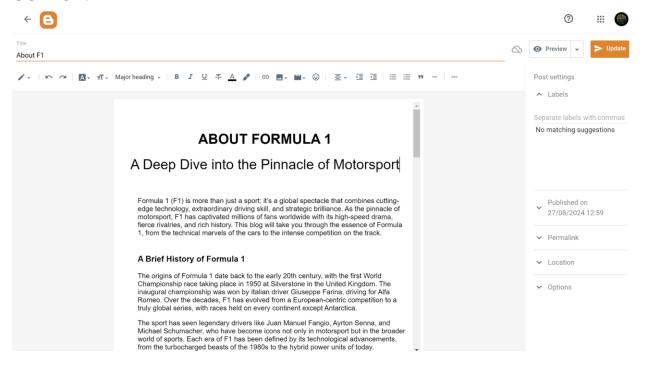
16.SEO Meta Description:

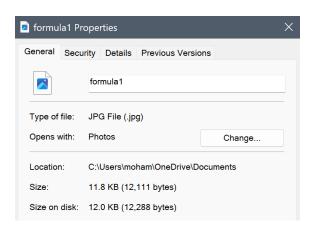
• Craft a compelling meta description for your blog post that includes relevant keywords.

17. Publish and Monitor:

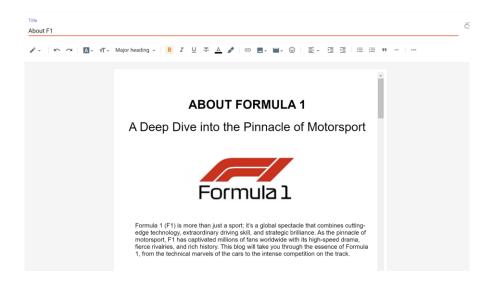
- Publish your optimized blog post.
- Monitor the performance of your blog post using tools like Google Analytics to see how well optimized images contribute to user engagement and traffic.

OUTPUT:









RESULT:

Thus, learning and applying image SEO (Search Engine Optimization) techniques for bloggers, focusing on optimizing images within blog posts to improve search engine visibility and user experience is successfully implemented

	4
11	ota.
v	au.

Exercise - 11

ANALYZING SEASONALITY OF DIFFERENT NICHES USING GOOGLE TRENDS

AIM:

To determine whether the chosen niches are seasonal or evergreen using Google Trends data for specific countries.

NICHES AND COUNTRIES:

- "Usage of Refrigerator" India, UAE, Pakistan
- "Usage of Air Conditioner" India, Sri Lanka, China
- "Usage of Wet Grinder" India, USA, England
- "Usage of Furniture" India, Bangladesh, Australia

PROCEDURE:

1. Access Google Trends

• Open your web browser and go to the Google Trends website.

2. Analyze "Usage of Refrigerator"

- a. In the search bar, type "Refrigerator" and press Enter.
- b. Click on "All categories" and select "Home & Garden."
- c. Choose a time range of 5 years or more.
- d. Select the countries: India, UAE, Pakistan.
- e. Analyze the interest over time, regional interest, and related queries/topics. Note any patterns or trends.

3. Analyze "Usage of Air Conditioner"

- a. Repeat steps a to e, replacing the keyword with "Air Conditioner" and countries with India, Sri Lanka, China.
- b. Analyze the interest over time, regional interest, and related queries/topics.

4. Analyze "Usage of Wet Grinder"

- a. Repeat steps a to e, replacing the keyword with "Wet Grinder" and countries with India, USA, England.
- b. Analyze the interest over time, regional interest, and related queries/topics.

5. Analyze "Usage of Furniture"

- a. Repeat steps a to e, replacing the keyword with "Furniture" and countries with India, Bangladesh, Australia.
- b. Analyze the interest over time, regional interest, and related queries/topics.

6. Compile Your Findings

- Create a report summarizing your findings for each niche-country combination.
- Include graphs or screenshots of the "Interest over time" graphs and any significant observations you made.

7. Interpret Your Findings

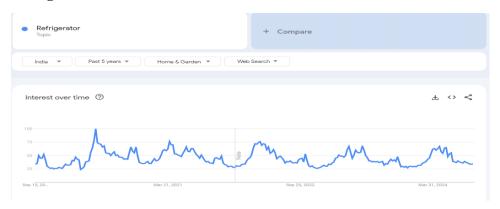
- For each niche, analyze whether the interest is consistent or fluctuates seasonally.
- Compare the interest trends between different countries for each niche.
- Consider the related queries and topics to understand the context and ongoing interest.

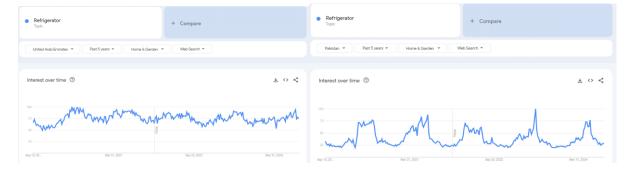
8. Draw Conclusions

- Based on the data, determine whether each niche is seasonal or evergreen in the selected countries.
- Write a brief summary of your conclusions for each niche

OUTPUT:

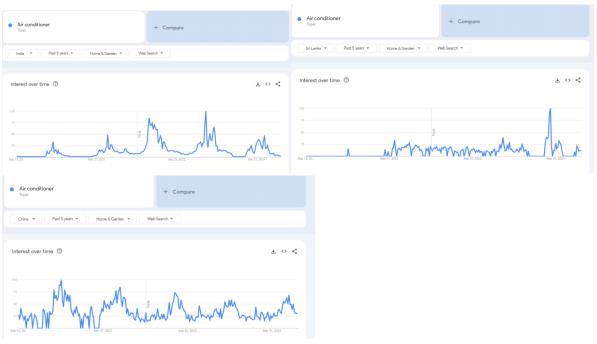
Refrigerator:





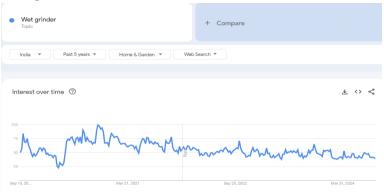
Refrigerator are evergreen

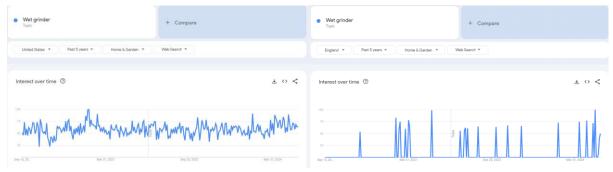
Air conditioner:



Air conditioner are mostly seasonal and evergreen in china

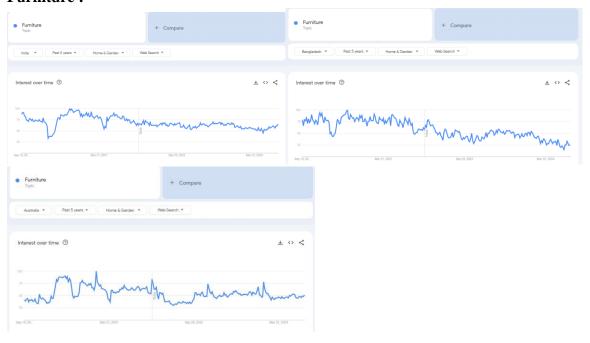
Wet grinder:





Wet grinder are mostly evergreen and seasonal in england

Furniture:



Furnitures are mostly evergreen

RESULT:

Thus, determining whether the chosen niches are seasonal or evergreen for specific countries has been implemented successfully using Google Trends data.

Date:	
Dau.	

Exercise - 12

COMPETITOR ANALYSIS FOR FITLIFE.COM USING UBERSUGGEST

AIM:

To gain practical experience in conducting a comprehensive competitor analysis for a website in the fitness industry using the Ubersuggest platform. By the end of this exercise, participants will be able to utilize various features of Ubersuggest to extract valuable insights about a competitor's online presence, SEO strategy, and overall performance.

PROCEDURE:

1. Set Up Ubersuggest Account

• If you don't have an Ubersuggest account, sign up for one. Once you're logged in, you're ready to start the analysis.

2. Identify Competitors and Gather Information

• For this exercise, our competitor is "FitLife.com." Before you start, gather some basic information about FitLife.com, such as their main products, services, and target audience. This will help you contextualize your findings.

3. Domain Score

- Go to Ubersuggest and enter "fitlife.com" in the search bar.
- On the search results page, you will find the domain score. This score gives you an idea of the overall authority and strength of FitLife.com's domain.

4. Backlinks

- Navigate to the "Backlinks" tab in the left sidebar.
- Explore the backlinks that point to FitLife.com. Look for patterns in the types of websites linking to them and the quality of those links.

5. SEO Keyword Ranking

- Click on the "Organic Keywords" tab to see the list of keywords FitLife.com is ranking for.
- Analyze the keywords to understand the focus of their SEO efforts and the search terms they are targeting.

6. Top SEO Pages

- Head to the "Top Pages" section.
- Identify the pages on FitLife.com that are receiving the most organic traffic. Take note of the content topics and formats that seem to attract the most visitors.

7. Used SEO Keywords

• In the "Organic Keywords" section, analyze the specific keywords that are driving traffic to FitLife.com.

• This will provide insight into their primary keywords and content strategy.

8. Maximum Traffic Pages (Top Pages)

- Review the "Top Pages" list again and pay attention to the pages with the highest organic traffic.
- Take a closer look at these pages to understand what makes them successful.

9. Social Media Presence and Backlinks

- Return to the "Backlinks" section and specifically look for backlinks coming from social media platforms.
- This will show you FitLife.com's social media presence and how it's influencing their backlink profile.

10. Find Search Volume, Position, Estimated Visitors for Keywords

- In the "Organic Keywords" section, explore the keywords FitLife.com is ranking for.
- Look for search volume, position in search results, and estimated monthly visitors for these keywords.

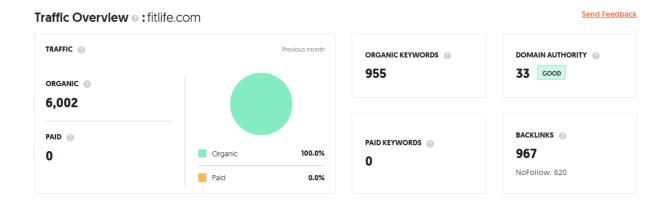
11. In Site Audit

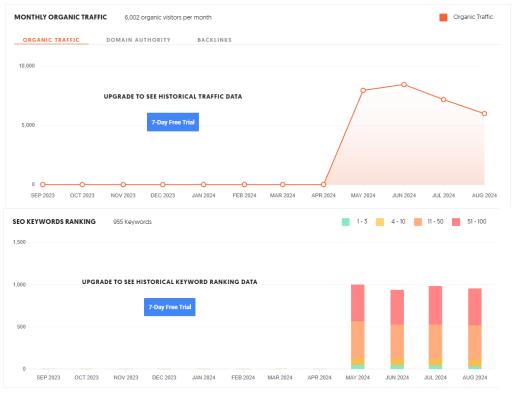
- Go to the "Site Audit" tab.
- Run a site audit for FitLife.com to gather information on their on-page SEO score, organic monthly traffic, organic keywords, broken links, loading time, top SEO issues, and referring domains.

12. Compile and Analyze Findings

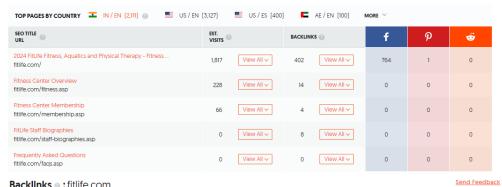
- After completing the above steps, compile your findings into a report.
- Analyze the data to identify FitLife.com's strengths, weaknesses, opportunities, and threats in terms of their online presence and SEO strategy

OUTPUT:



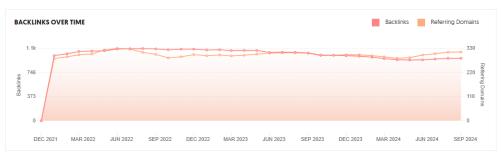


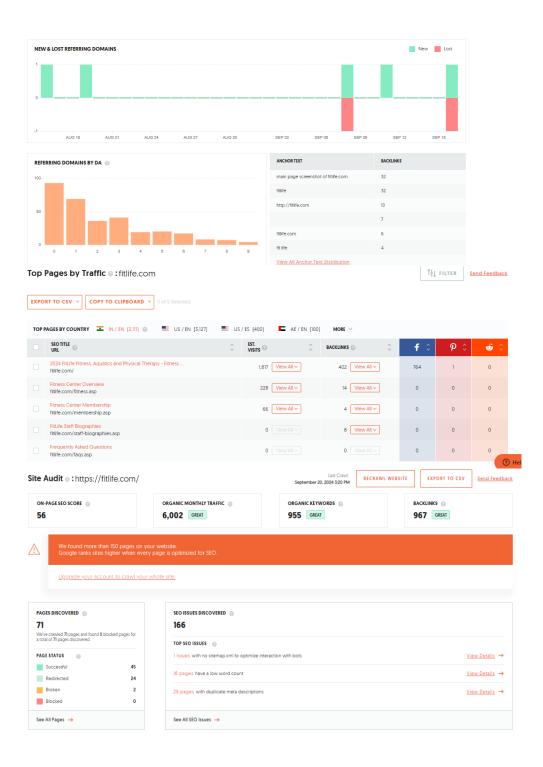
TOP SEO PAGES

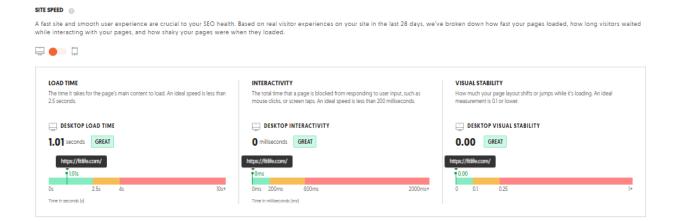


Backlinks : fitlife.com

BACKLINKS 💿 DOMAIN AUTHORITY REFERRING DOMAINS 967 **33** GOOD 314







RESULT:

Thus, competitor analysis for fitlife.com using ubersuggest has been implemented successfully.

Date:

Exercise – 13

ADVANCED GOOGLE SEARCH USING SEARCH OPERATORS

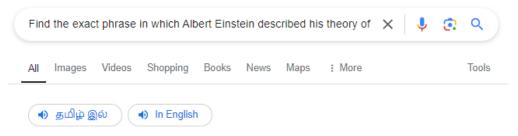
AIM:

To familiarize participants with various advanced Google search commands and demonstrate their practical use in retrieving specific information from the web.

QUESTIONS:

1. "search term"

• Find the exact phrase in which Albert Einstein described his theory of relativity.



He used to say: 'In the [special] relativity theory it is no question of a revolutionary act but of a natural development of lines which have been followed for centuries'. In 1865 or so Maxwell published his equations which describe the behavior of electromagnetic fields. 9 Oct 2016

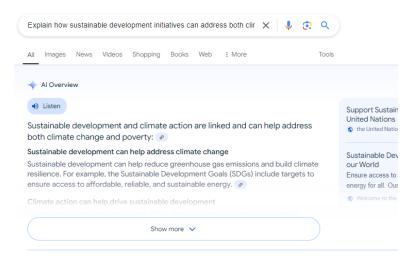
2. OR

• List online courses on artificial intelligence or machine learning available for enrollment.

List online courses on artificial intellige	nce or machine learni	ng availa 🗙 👃 👩 🔾			
All News Images Videos Shoppin	ng Books Web	More Tools			
From sources across t	the web				
Deep learning	·	Generative Al	·	Al strategy	~
Natural language process	· -	Artificial intelligence and	•	Neural network	~
Computer vision	·	Google Al Essentials	V VALUE OF THE PARTY OF THE PAR	Prompt engineering	~
		15 more 🗸			Feedback

3. AND

• Explain how sustainable development initiatives can address both climate change and poverty.



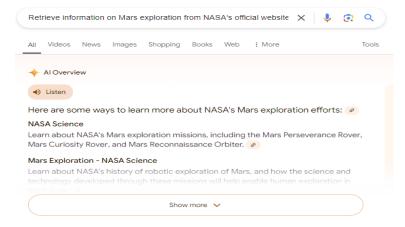
4. -exclude

• Share information about tigers while excluding details about their habitats.



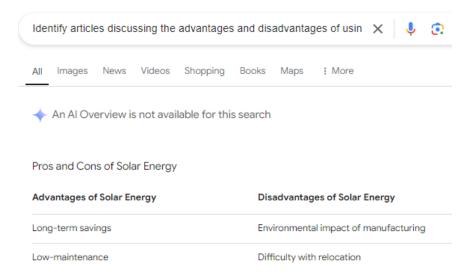
5. site:

Retrieve information on Mars exploration from NASA's official website.



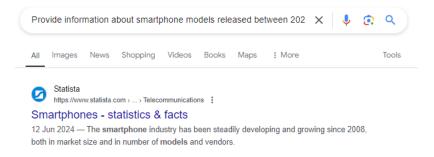
6. ()

• Identify articles discussing the advantages and disadvantages of using solar energy compared to traditional sources.



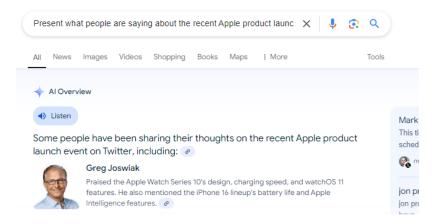
7. #..#

Provide information about smartphone models released between 2020 and 2022.



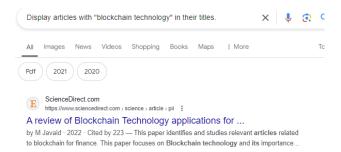
8. @

• Present what people are saying about the recent Apple product launch event on Twitter.



9. intitle:

• Display articles with "blockchain technology" in their titles.



10. allintitle:

• List articles with both "green energy" and "cost-effective solutions" in their titles.



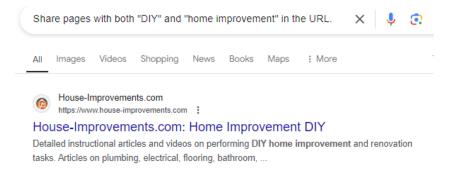
11. inurl:

• Show websites with "recipe" in the URL, particularly those related to Italian cuisine.



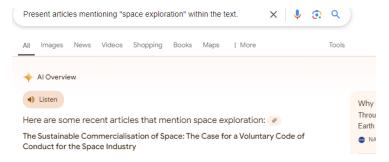
12. allinurl:

Share pages with both "DIY" and "home improvement" in the URL.



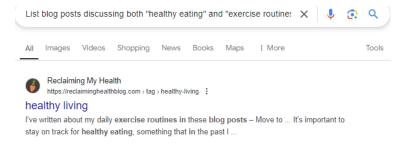
13. intext:

• Present articles mentioning "space exploration" within the text.



14. allintext:

• List blog posts discussing both "healthy eating" and "exercise routines" within the body text.



15. filetype:

• Provide PDF reports on climate change mitigation strategies.



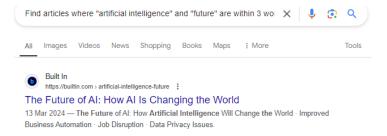
16. related:

• Identify websites related to the World Health Organization's official site.



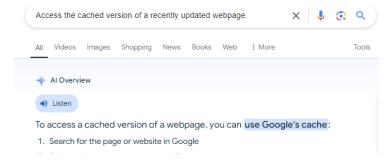
17. AROUND(X)

• Find articles where "artificial intelligence" and "future" are within 3 words of each other.



18. cache:

• Access the cached version of a recently updated webpage.



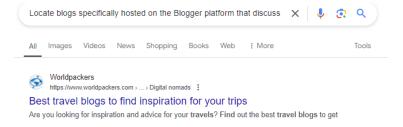
19. source:

• Determine the original source of a news article reporting on recent climate change research.



20. blogurl:

• Locate blogs specifically hosted on the Blogger platform that discuss travel experiences.



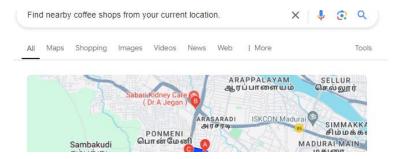
21. loc:placename

• Find results related to tourism in "Kyoto, Japan."



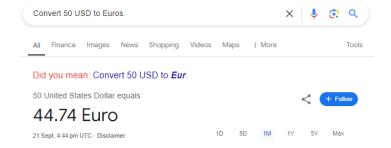
22. location:

• Find nearby coffee shops from your current location.



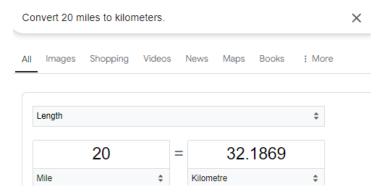
23. \$/€

• Convert 50 USD to Euros.



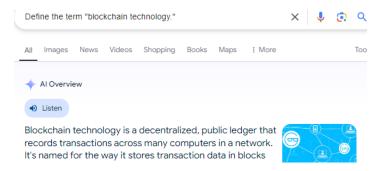
24. **in/to**

Convert 20 miles to kilometers.



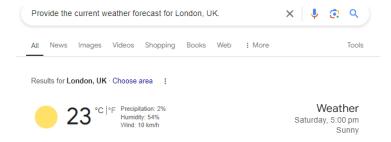
25. define:

• Define the term "blockchain technology."



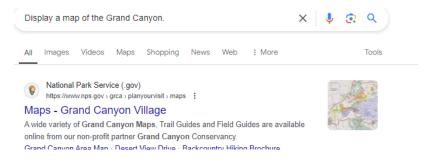
26. weather:

• Provide the current weather forecast for London, UK.



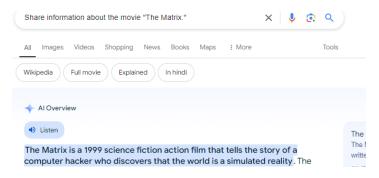
27. map:

• Display a map of the Grand Canyon.



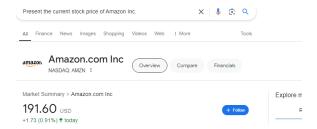
28. movie:

• Share information about the movie "The Matrix."



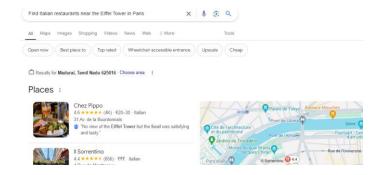
29. stocks:

• Present the current stock price of Amazon Inc.



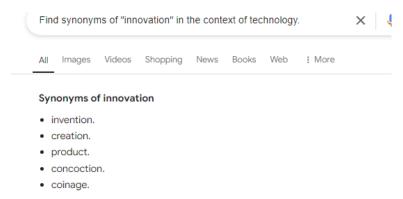
30. near:

• Find Italian restaurants near the Eiffel Tower in Paris.



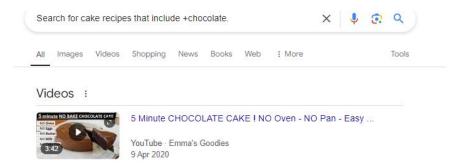
31. ~

• Find synonyms of "innovation" in the context of technology.



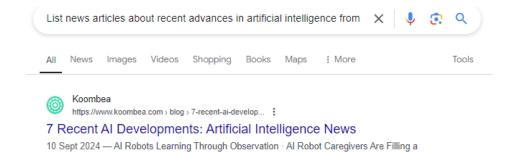
32. +

• Search for cake recipes that include +chocolate.



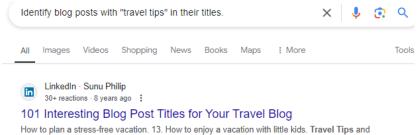
33. daterange:xxxxx-xxxxx

• List news articles about recent advances in artificial intelligence from 2020 to 2022.



34. inposttitle:

• Identify blog posts with "travel tips" in their titles.



35. allinpostauthor:

• List blog posts written by authors who have both "John" and "Smith" in their names.



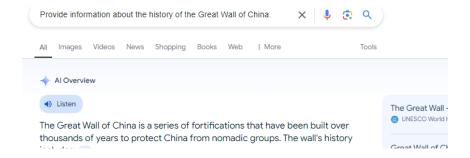
36. inpostauthor:

• Find blog posts specifically authored by "Jane Doe."



37. info:

Provide information about the history of the Great Wall of China.



RESULT:

Thus, familiarizing participants with various advanced Google search commands and demonstrating their practical use in retrieving specific information from the web is successful.

Date:

Exercise - 14

EXTRACTING AND ANALYZING JOURNAL DATA FROM THE GIVEN WEBSITE URL

OBJECTIVE:

Extract journal data from the Journal List webpage of Anna University's website, save the data into an Excel file, and analyze the dataset using Python

MATERIALS REQUIRED:

- Python environment (Google Colab or Jupyter Notebook)
- Internet connection for web scraping

PROCEDURE:

1. Import Required Libraries

- Requests: To fetch the HTML content from the URL.
- BeautifulSoup: To parse the HTML content and locate the table.
- Pandas: To structure the table data into a DataFrame and analyze it.
- Warnings: To suppress unnecessary HTTPS warnings.
- Files (from Colab): To handle file download in Google Colab.

2. Fetch Data from the Website

- Website URL: https://cfr.annauniv.edu/research/academics/journals-list.php
- Use the requests.get() function to retrieve the webpage data.
- Suppress HTTPS Warnings: Use the warnings library to ignore unverified HTTPS warnings.

3. Parse the Webpage Content

- Use BeautifulSoup to parse the HTML and locate the table containing journal information.
- Extract the headers (column names) and rows (journal details) from the table.

4. Structure Data into a DataFrame

- Use pandas to convert the extracted data into a structured DataFrame.
- Ensure the DataFrame has appropriate column names: S.No, Full Journal Title, Print-ISSN, E-ISSN, Publisher, and Country.

5. Save Data to Excel

• Save the DataFrame to an Excel file (e.g., table_data.xlsx) using the to_excel() method.

• Download the file to your local machine.

6. Upload and Read the Excel File

- Upload File: Re-upload the Excel file back into your environment for further analysis.
- Rename the column names manually as S.No, Full Journal Title, Print-ISSN, EISSN, Publisher, Country
- Read Data: Load the data into a pandas DataFrame.

7. Analyze the Data

- Basic Information: Use df.info() to display the structure of the dataset.
- Check for Missing Data: Use df.isnull().sum() to identify any missing values in the dataset.
- Column-wise Unique Values: Use df.nunique() to check how many unique entries exist in each column.

8. Descriptive Statistics

- Statistical Summary: Use df.describe() to generate basic descriptive statistics for numeric columns.
- Top 5 Publishers: Use value_counts() to find the top 5 publishers based on frequency.
- Top 5 Countries: Use value_counts() to list the top 5 countries represented in the journal list.

CODE:

a) Extraction:

```
import requests
from bs4 import BeautifulSoup
import pandas as pd
from google.colab import files
import warnings

warnings.filterwarnings('ignore', message='Unverified HTTPS
request is being made')

def fetch_and_save_table_to_excel(url, output_file):
    response = requests.get(url, verify=False)
    response.raise_for_status()
```

```
soup = BeautifulSoup(response.text, 'html.parser')
  table = soup.find('table')
  headers = [header.text.strip() for header in
table.find all('th')]
  rows = []
  for row in table.find all('tr'):
    columns = [col.text.strip() for col in row.find all('td')]
   if len(columns) > 0:
     rows.append(columns)
  try:
    df = pd.DataFrame(rows, columns=headers)
  except ValueError as e:
   print(f"Error creating DataFrame: {e}")
   max cols = max(len(row) for row in rows)
    df = pd.DataFrame(rows, columns=headers[:max cols])
  df.to excel(output file, index=False)
  files.download(output file)
  print(f"Data has been saved to {output file} and is ready for
download.")
url = 'https://cfr.annauniv.edu/research/academics/journals-
list.php'
output file = 'table data.xlsx'
fetch and save table to excel(url, output file)
```

b) Statistics:

```
import pandas as pd
from google.colab import files
uploaded = files.upload()
for file name in uploaded.keys():
  df = pd.read excel(file name)
print("First few rows of the dataset:")
print(df.head())
df.columns = ['S.No', 'Full Journal Title', 'Print-ISSN', 'E-
ISSN', 'Publisher', 'Country']
print("\nBasic Information about the dataset:")
print("\nSummary of missing values per column:")
print(df.isnull().sum())
print("\nUnique value counts per column:")
print(df.nunique())
print("\nDataset Statistics:")
print(df.describe())
print("\nTop 5 most frequent Publishers:")
print(df['Publisher'].value counts().head(5))
print("\nTop 5 Countries represented in the dataset:")
print(df['Country'].value counts().head(5))
```

OUTPUT:

Error creating DataFrame: 8 columns passed, passed data had 6 columns Data has been saved to table_data.xlsx and is ready for download.

A1	*	(<i>f</i> _∗ Com	mon for all	registered	d Ph.D. scl	holars irrespe	ective of t	he time of	their regi	stration
А	В	С	D	Е	F	G	Н	I I	J	K	L
ars irresp	RTANT GU	Sl.No	l Journal T	Print-ISSN	E-ISSN						
1	ACADEMIC	1069-6563	1553-2712	WILEY	UNITED ST	ATES					
2	ACADEMIC	1040-2446	1938-808X	LIPPINCO	UNITED ST	ATES					
3	ACADEMIC	1076-6332	1878-4046	ELSEVIER S	UNITED ST	ATES					
4	ACADEMY	1941-6520	1941-6067	ACAD MAI	UNITED ST	ATES					
5				ACAD MAI	UNITED ST	ATES					
6		1537-260X			UNITED ST						
7		1558-9080			UNITED ST						
9				ACAD MAI		ATES					
9				PERGAMO							
Saving tab First few Common 0 1	le_data.xl rows of the	sx to tabl e dataset:	e_data (1)	.xlsx	ective of t 1 2 3		neet)-648479 b				
3 4 4 5 FAKE JOURNALS-IMPORTANT GUIDELINE (Click Here) Sl.No \ 0 ACADEMIC EMERGENCY MEDICINE 1069-6563											
1		ACADEII		IC MEDICIN							
2				C RADIOLOG							
3 4				MENT ANNALS ENT JOURNAL							
Full Jou 0	rnal Title 1553-2712			Print-ISS	N I	E-ISSN STATES					
1 2 3	1938-808X 1878-4046 1941-6067		ELSEVIER	S & WILKINS SCIENCE INC MANAGEMEN	C UNITED !	STATES					
4	1948-0989			MANAGEMEN'							
Basic I											
<class td="" <=""><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></class>											
RangeIndex: 12157 entries, 0 to 12156											
Data co		otal 6 (columns)								
# Co.	Lumn		Non-N	lull Cour	nt Dtyp	e					
0 5.1				non-nul					1		1
				non-nul	_		ummary of	missin	-		Tumn:
	int-ISSN			non-nul		-	.No	1 -1.1	0		
	ESSN			non-nul			ull Journa	al litle			
	olisher			non-nu	_	_	rint-ISSN		501		
	ıntry			non-nu	ll obje		-ISSN ublisher		402		
	<pre>dtypes: int64(1), object(5)</pre>								0		
memory usage: 570.0+ KB							ountry		0		
None						d	type: int@	54			

		Datase	t Statistics:							
			S.No							
Unique value counts	per column:	count	12157.000000							
S.No	12157	mean	6079.000000							
Full Journal Title	12157	std	3509.567946							
Print-ISSN	11654	min	1.000000							
E-ISSN	11751	25%	3040.000000							
Publisher	1877	50%	6079.000000							
Country	79	75%	9118.000000							
dtype: int64		max	12157.000000							
Top 5 most frequent Publishers:										
Publisher										
WILEY		1149								
SPRINGER		876								
ELSEVIER		631								
ROUTLEDGE JOURNALS, TAYLOR & FRANCIS LTD 546										
TAYLOR & FRANCIS LTD	434									
Name: count, dtype: int64										
Top 5 Countries represented in the dataset:										
Country										
UNITED STATES 4184										
ENGLAND 2997										
NETHERLANDS 961										
GERMANY 746										
SWITZERLAND 334										
Name: count, dtype: int64										

RESULT:

Thus, Extracting journal data from the Journal List web page of Anna University's website, save the data into an Excel file, and analyzing the dataset using Python is successfully implemented