this project has 3 main function
the first function gets tries to find the subdomains or a certain website t tries multiple possibilities from a list of possibilities stored in a file

```
with open('subdomains_dictionary.bat', 'r') as file:
   for line in file:

      line = line.strip()

      if line.endswith('.'):
         url0 = 'https://www.' + line[:-1] + '.' +after_www
      else:
         url0 = 'https://www.' + line + '.' + after_www
      try:
         response = requests.get(url0)
         if response.status_code == 200:
            print('Request was successful.')
            print(url0)
            with open("subdomains_output.bat", "a") as file:
               # Write the string to the file
               file.write(url0+"\n")
      except requests.exceptions.ConnectionError:
            # handle the connection error
            print('no')
```

this piece of code loops thghrough every line of possible subdomain and cocatinates them to the url given by the user using the right format then sends a request and get either a response 200 or an error if the response is 200 it stores the url with the subdomain in another file using this piece of code :

```
with open("subdomains_output.bat", "a") as file:
               # Write the string to the file
               file.write(url0+"\n")
```

if not it just prints no

the second part of this codes does the same thing as the first one but instead of looking for subdomains it looks for directories using this piece of code :

```
 with open('dirs_dictionary.bat', 'r') as file:
   for line in file:

      line = line.strip()
      if line.startswith('.'):
         line=line[1:]
```

```
    url1='https://' + url.strip() + '/'+ line
    try:
       response = requests.get(url1)
       if response.status_code == 200:
          print('Request was successful.')
          print(url1)
          with open("directories_output.bat", "a") as file:
             # Write the string to the file
             file.write(url1+"\n")
    except requests.exceptions.ConnectionError:
          # handle the connection error
           print('no')
          # None
```

the last part of this code takes the html of the main url and the html of the directories found and extracts all hrefs from these pages using thi code :

```
url2='https://'+url
response = requests.get(url2)
html_content = response.text

# Define a regular expression to match all href links
href_regex = r'href=[\'"]?([^\'" >]+)'

# Use regex to find all href links in the HTML content
href_links = re.findall(href_regex, html_content)

# Print the href links found
for link in href_links:
   with open("files_output.bat", "a") as file:
    file.write(link+"\n")


# this part of the code finds all html pages of the found directories and gets the href
with open('directories_output.bat', 'r') as file:
   for line in file:
      response = requests.get(line)
      html_content = response.text

      # Define a regular expression to match all href links
      href_regex = r'href=[\'"]?([^\'" >]+)'

      # Use regex to find all href links in the HTML content
```

```
    href_links = re.findall(href_regex, html_content)

    # Print the href links found
    for link in href_links:
       with open("files_output.bat", "a") as file:
       file.write(link+"\n")
```

THE DIFFICULTIES ENCOUNTERED:

- first the subdomain names were not all given in the sam syntax some of them ended with a dot that neede to be removed which i did
-the scond problem encountered was to import some libraries needed and download them
-i had to add try catch in rder for the code not to stop runnning
-writing to files was deleting previous informations which i fixed and only deleted all content in the beggining of the script