AIST 2120
Extra Credit Program
Blog Analyzer

INTRODUCTION

This programming assignment provides the opportunity to practice using Python's Web Scraping capabilities to analyze content in web sites.

ASSIGNMENT & DISCUSSION

Your task in this programming assignment is to write a script that will scrape blog content from https://grith-llc.com/blog using Python's request module (See Chapter 12, Web Scraping). Your code should determine how many blog articles are present on the page, list them, and provide a count. Then article, your script will download the content for each article using Python's request module. Your code will then count the number and display the links to other Grith articles found in the analyzed article. Your code will also count the number of Amazon Ads and Google Ads found in the article, display them, and provide a count for each variety of add. Contain your scraping to grith-llc.com.

TASKS

1. Include a header/footer
2. Implement this program in a file named **lastN_firstN_bloganalyzer.py**
3. Include a header/footer
4. Request content from web URL; indicate if an issue occurs.
5. List all blog articles found on https://grith-llc.com/blog.
   - List the URL for each article found, do not include other links
   - Provide a count of the URLs.
   - Note that new articles may be added to the site, so make sure your code is flexible enough to address new articles.
6. For each article's URL, scrape it
   - Provide a count of the Blog Articles found in the article. These URLs will contain grith-llc.com in the URL. **DO NOT attempt to scrape a page that does not have grith-llc.com in its URL.**
   - Provide a count of the Amazon Ads found in the article. Do not attempt to scrape-**ONLY COUNT.**
   - Provide a count of the Google Ads found in the article. Do not attempt to scrape-**ONLY COUNT.**
   - If you have questions, ask!
7. Print an analysis of the blog that includes the number and averages of linked articles, amazon ads, and google ads per article. You will need to keep a running tally for each article and divide by the total number of articles.

REQUIREMENTS

- □ **Application**
  - o Implement the program in Python 3.
  - o Ensure you use comments to explain your code
  - o Ensure your file is named properly: **lastN_firstN_bloganalyzer.py**.
  - o If you choose to use a different file editor, note that I will run and evaluate your submissions in IDLE. If you write your code using another IDE, remember to run your finished program in IDLE to ensure no unfortunate surprises. If it doesn't run in IDLE, it doesn't run.

- □ **SUBMISSION**
  - o Submit your source code file via D2L. Neither email nor hardcopy submissions will be accepted.
  - o Ensure you retain a complete copy of your source code files(s).
  - o This extra credit project will be used to replace your lowest programming assignment score.

- □ **DUE DATE: PER D2L INSTRUCTIONS.**

- □ **LATE PENALTY: NO LATE SUBMISSION PERMITTED.**

STYLE GUIDE

- □ **File headers**. Include a header in the below format on all .py source code files.

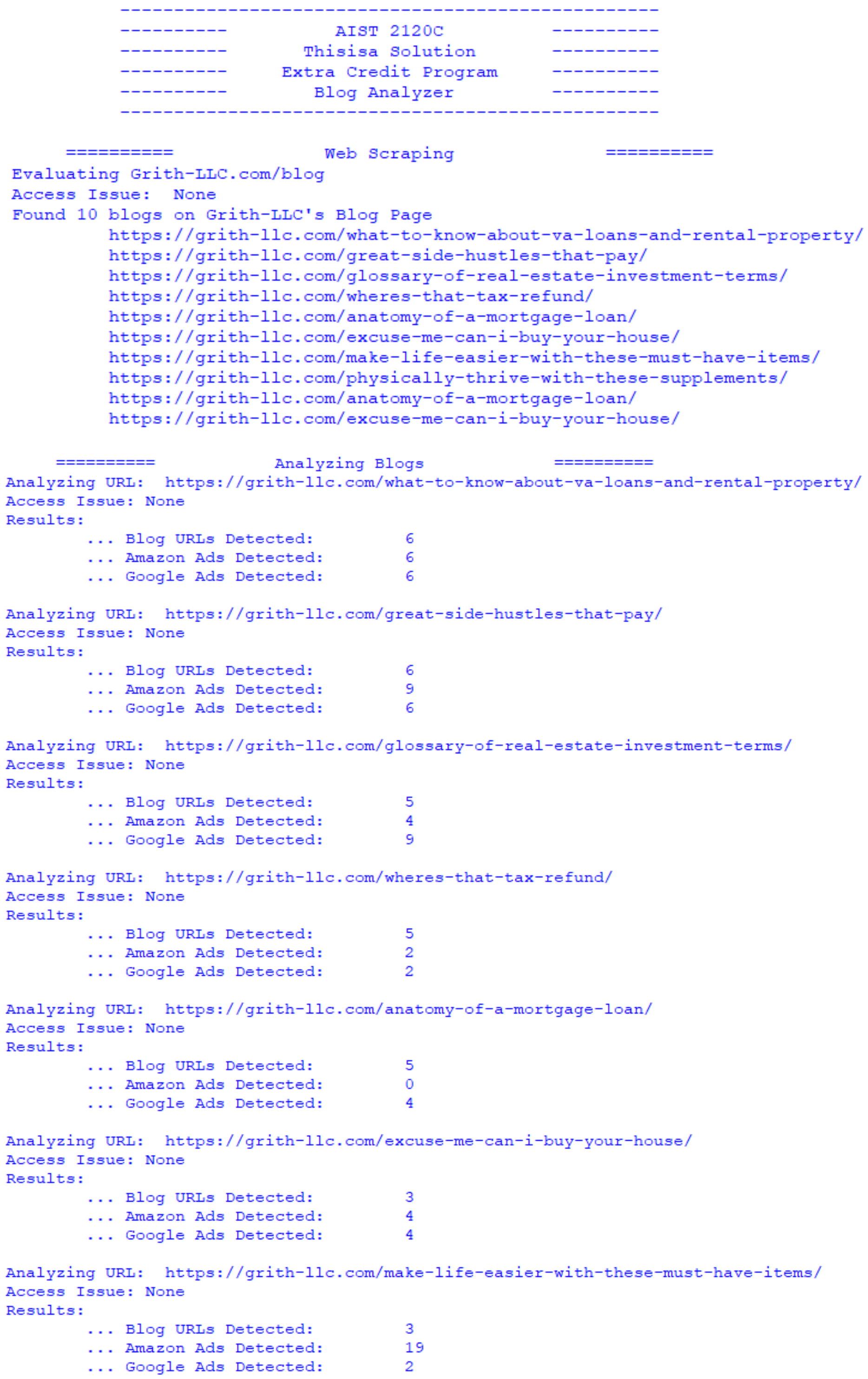
```
# Your Name
# AIST 2120,  Extra Credit
# Submission Date
# lastN_firstN_bloganalyzer.py
```

- □ **Example Screen Output**. Your screen output should look exactly like the example on the next page with minor differences for names and possibly values.
  - ▪ Keep in mind that the output is to inform the user and should be both clear and pleasing to those reading it.

Hints:

- The only imports you need for this program are requests and re; however, you may use other libraries.
- Use your web browsers 'View Page Source' and 'Inspect' to find strings of text that show up for every blog, amazon ads, and google ads.
- Videos for web scraping are already in you content folder.

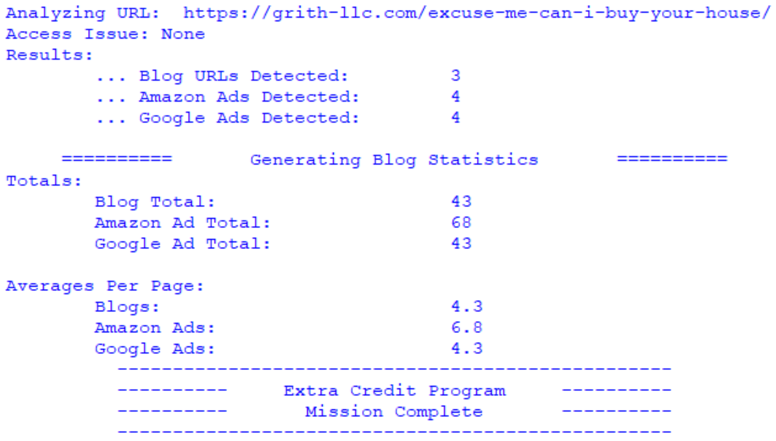
Output: additional articles may be added, so numbers and articles in this example may change.



```
          --------------------------------------------------
          ----------         AIST 2120C         ----------
          ----------      Thisisa Solution      ----------
          ----------    Extra Credit Program    ----------
          ----------       Blog Analyzer        ----------
          --------------------------------------------------

     ==========              Web Scraping              ==========
Evaluating Grith-LLC.com/blog
Access Issue:  None
Found 10 blogs on Grith-LLC's Blog Page
         https://grith-llc.com/what-to-know-about-va-loans-and-rental-property/
         https://grith-llc.com/great-side-hustles-that-pay/
         https://grith-llc.com/glossary-of-real-estate-investment-terms/
         https://grith-llc.com/wheres-that-tax-refund/
         https://grith-llc.com/anatomy-of-a-mortgage-loan/
         https://grith-llc.com/excuse-me-can-i-buy-your-house/
         https://grith-llc.com/make-life-easier-with-these-must-have-items/
         https://grith-llc.com/physically-thrive-with-these-supplements/
         https://grith-llc.com/anatomy-of-a-mortgage-loan/
         https://grith-llc.com/excuse-me-can-i-buy-your-house/

     ==========          Analyzing Blogs          ==========
Analyzing URL:  https://grith-llc.com/what-to-know-about-va-loans-and-rental-property/
Access Issue: None
Results:
        ... Blog URLs Detected:         6
        ... Amazon Ads Detected:        6
        ... Google Ads Detected:        6

Analyzing URL:  https://grith-llc.com/great-side-hustles-that-pay/
Access Issue: None
Results:
        ... Blog URLs Detected:         6
        ... Amazon Ads Detected:        9
        ... Google Ads Detected:        6

Analyzing URL:  https://grith-llc.com/glossary-of-real-estate-investment-terms/
Access Issue: None
Results:
        ... Blog URLs Detected:         5
        ... Amazon Ads Detected:        4
        ... Google Ads Detected:        9

Analyzing URL:  https://grith-llc.com/wheres-that-tax-refund/
Access Issue: None
Results:
        ... Blog URLs Detected:         5
        ... Amazon Ads Detected:        2
        ... Google Ads Detected:        2

Analyzing URL:  https://grith-llc.com/anatomy-of-a-mortgage-loan/
Access Issue: None
Results:
        ... Blog URLs Detected:         5
        ... Amazon Ads Detected:        0
        ... Google Ads Detected:        4

Analyzing URL:  https://grith-llc.com/excuse-me-can-i-buy-your-house/
Access Issue: None
Results:
        ... Blog URLs Detected:         3
        ... Amazon Ads Detected:        4
        ... Google Ads Detected:        4

Analyzing URL:  https://grith-llc.com/make-life-easier-with-these-must-have-items/
Access Issue: None
Results:
        ... Blog URLs Detected:         3
        ... Amazon Ads Detected:        19
        ... Google Ads Detected:        2
```

Some URLs not shown. Include them in your output using this format.

(continued on next page)



```
Analyzing URL:  https://grith-llc.com/excuse-me-can-i-buy-your-house/
Access Issue: None
Results:
        ... Blog URLs Detected:         3
        ... Amazon Ads Detected:        4
        ... Google Ads Detected:        4

     ==========       Generating Blog Statistics       ==========
Totals:
        Blog Total:                     43
        Amazon Ad Total:                68
        Google Ad Total:                43

Averages Per Page:
        Blogs:                          4.3
        Amazon Ads:                     6.8
        Google Ads:                     4.3
          --------------------------------------------------
          ----------     Extra Credit Program     ----------
          ----------       Mission Complete       ----------
          --------------------------------------------------
```