Introduction to Python

Claudia Carroll, TRIADS

**Homework Exercises: Introduction to Text Data**

1. Write the code to print the contents of a text file titled “text\_file.txt”

file = open(“text\_file.txt”, mode=“r”, encoding= “utf-8”)

text = file.read()

print(text)

OR:

text = open(“text\_file.txt”, mode=“r”, encoding= “utf-8”).read()

print(text)

1. Write the code to add a new line of text to “text\_file.txt”

file = open(“text\_file.txt”, mode=“a”, encoding= “utf-8”)

file.write(“new line of text”)

1. Write the following loop in list comprehension format in one line of code:

filtered\_tokens = []

for token in tokens:

if token not in punctuation:

filtered\_tokens.append(token.lower())

filtered\_tokens = [token.lower() for token in tokens if token not in punctuation]

1. If you have a list of tokens called token\_list, what are two ways you can remove the punctuation from the token list?

* Load the punctuation list from string, spacy, nltk or another NLP library and remove the tokens in that list from your own list of tokens:

import string

punctuation = list(string.punctuation)

tokens = [token for token in tokens if token not in punctuation]

* Iterate over your tokens list and keep only those tokens that are alphanumeric:

tokens = [token for token in tokens if token.isalpha()]