Lab-02-Sentiment Analysis and Data Processing Report

Author: Jace Lopes

Date: 02/24/2025

All the files can be ran in any IDE's

Overview

This report summarizes the tasks completed for analyzing and processing social media data, specifically tweets collected in data.json. The following tasks were performed:

1-Sentiment Analysis of Tweets (tweet_sentiment.py)

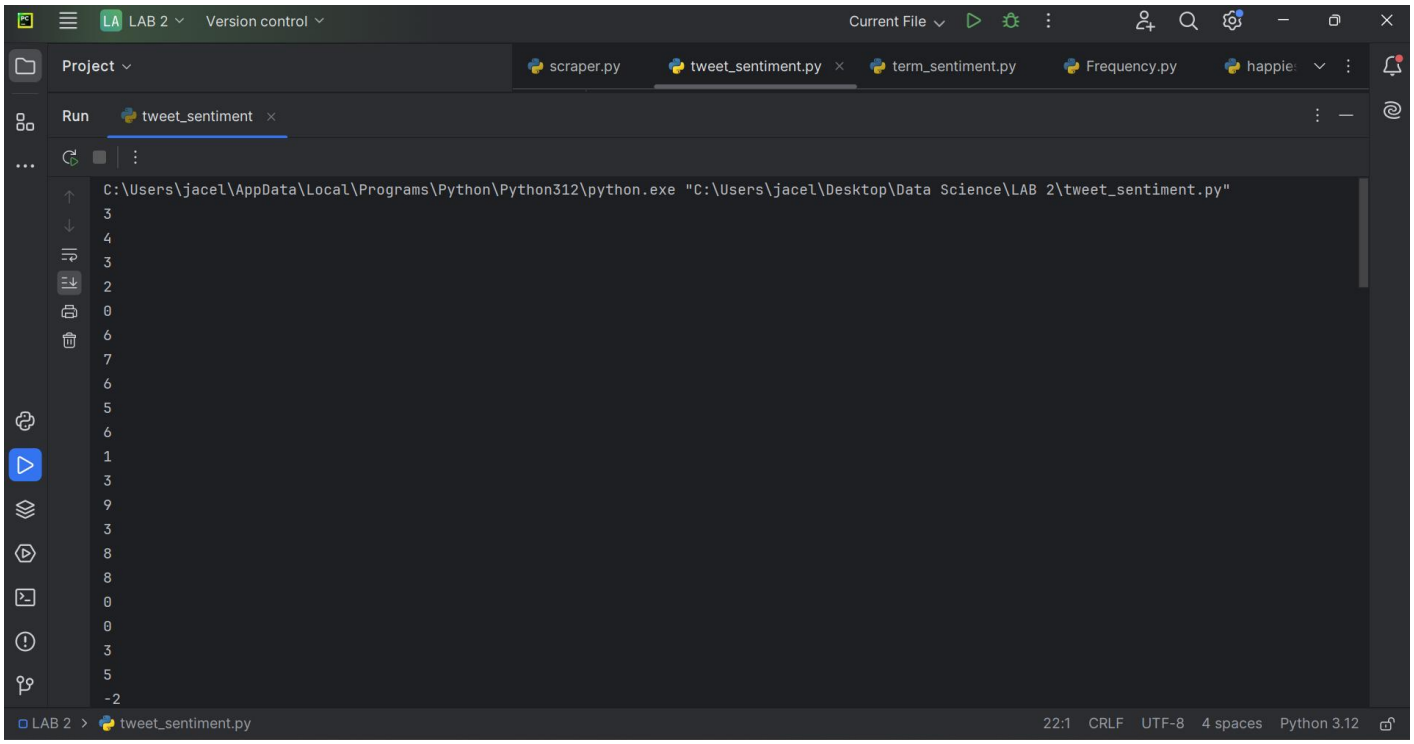
Goal: Compute sentiment scores for each tweet based on AFINN-111.txt.

Approach:

-Loaded AFINN-111.txt into a dictionary.

-Processed each tweet, calculating the sentiment score by summing known word scores.

-Printed sentiment scores for each tweet.



2-Deriving Sentiment for New Terms (term_sentiment.py)

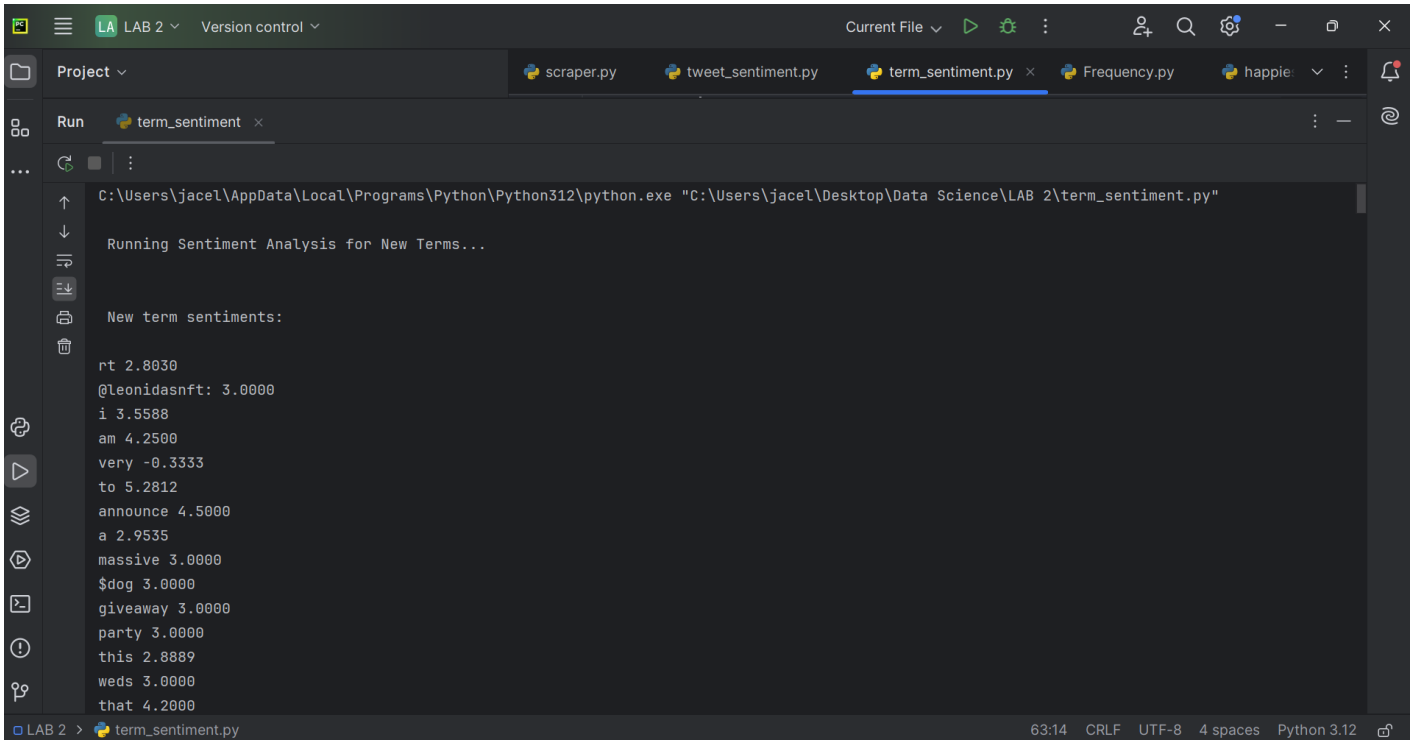
Goal: Assign sentiment values to words not found in AFINN-111.txt.

Approach:

-Identified words in tweets missing from AFINN.

-Assigned them an average sentiment score based on tweets they appeared in.

-Printed <term> <sentiment> format output.



3-Computing Term Frequency (frequency.py)

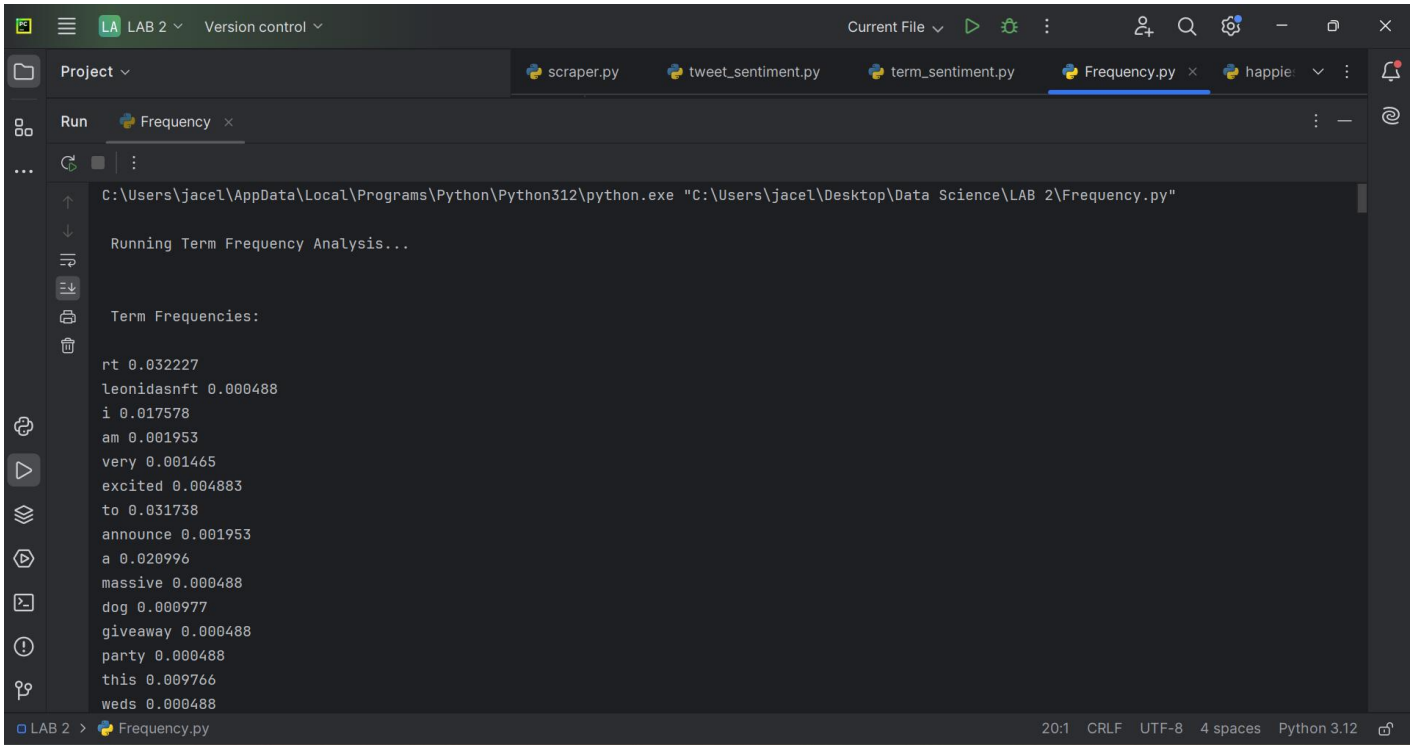
Goal: Calculate and print the frequency of each term in tweets.

Approach:

-Tokenized tweet text, removed punctuation and URLs.

-Counted occurrences of each word and computed their relative frequency.

-Printed <word> <frequency> format output.



4-Finding the Happiest State (happiest_state.py)

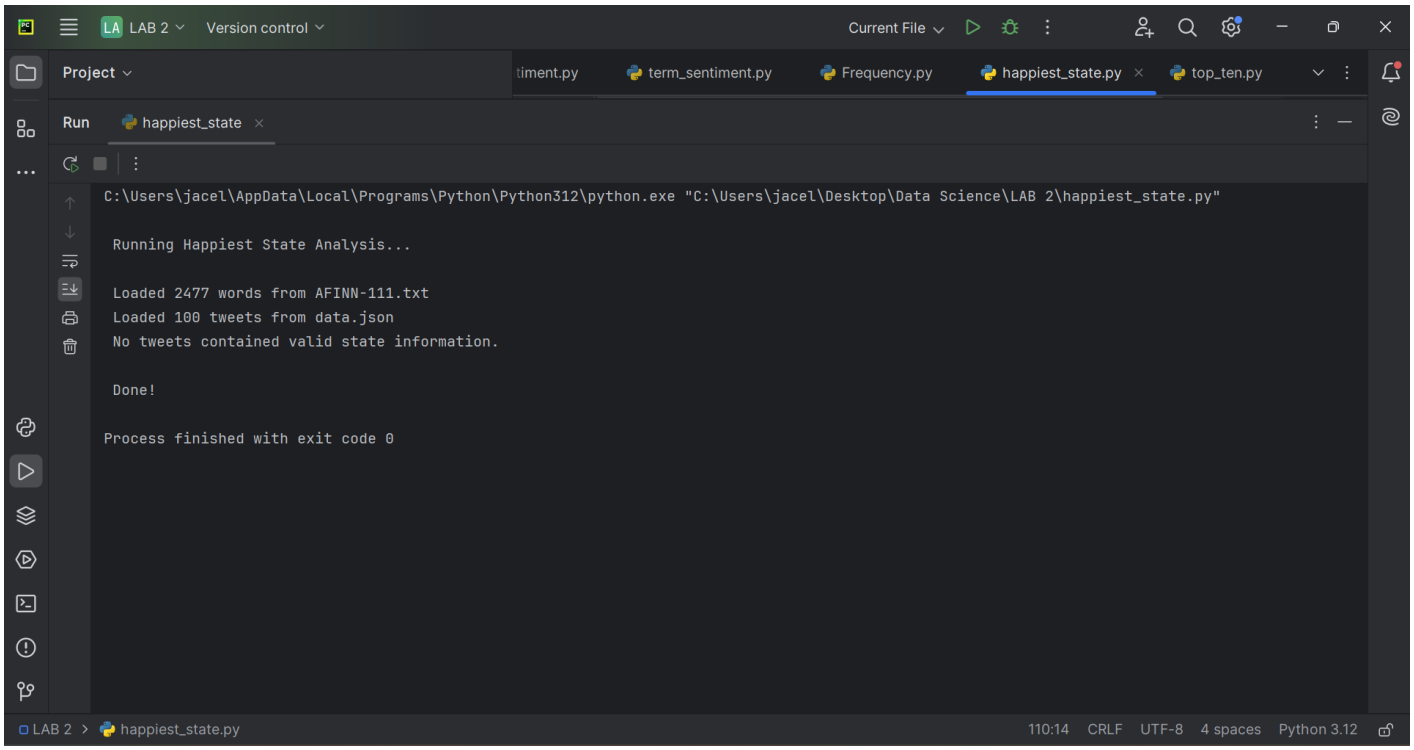
Goal: Identify the U.S. state with the highest average sentiment.

Approach:

-Extracted state names from tweet metadata (place or user.location).

-Computed sentiment scores per state and determined the happiest state.

-Printed the state with the highest average sentiment score.



5-Extracting Top 10 Hashtags (top_ten.py)

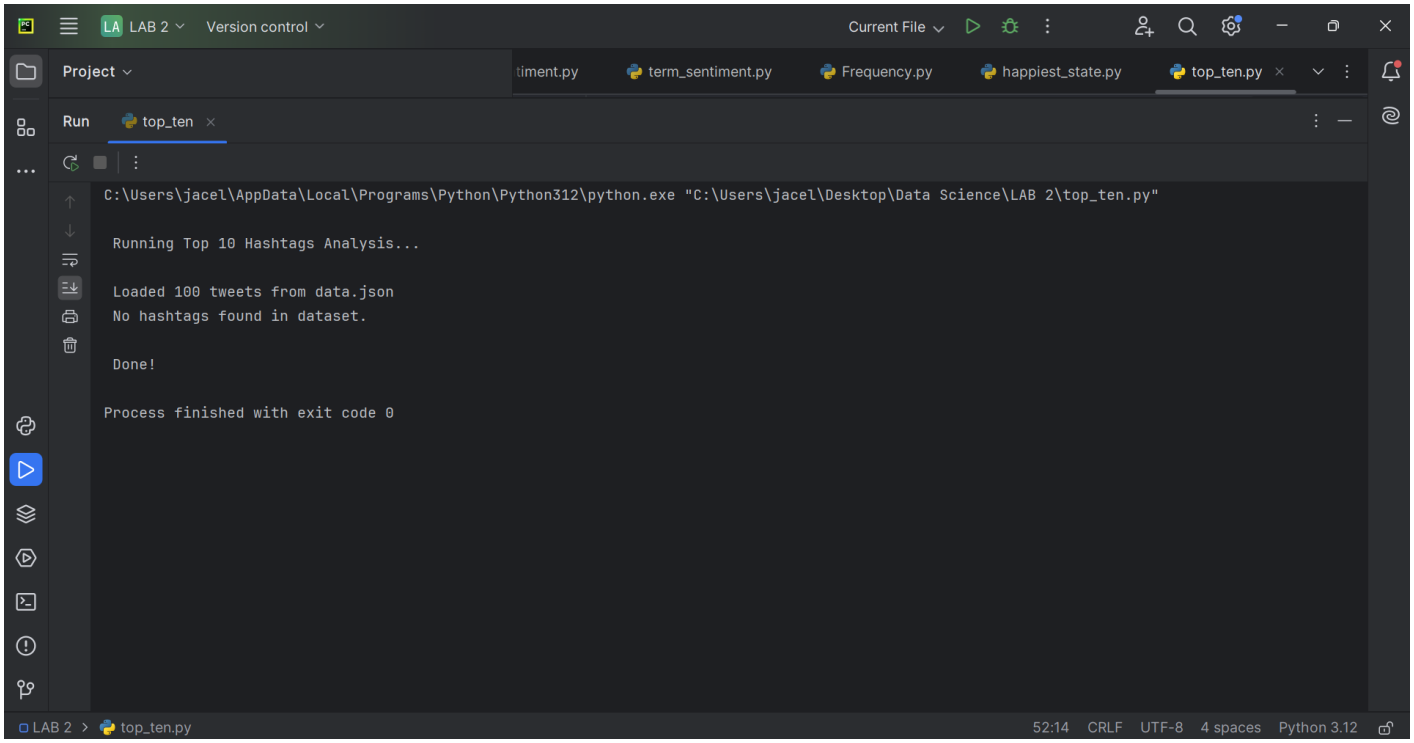
Goal: Identify and print the 10 most frequently used hashtags.

Approach:

-Extracted hashtags from tweet metadata (entities.hashtags).

-Counted occurrences and sorted the top 10 most frequent hashtags.

-Printed <hashtag> <count> format output.


LAB 2
Version control
Current File
Project
timent.py
term_sentiment.py
Frequency.py
happiest_state.py
top_ten.py
Run
top_ten
C:\Users\jacel\AppData\Local\Programs\Python\Python312\python.exe "C:\Users\jacel\Desktop\Data Science\LAB 2\top_ten.py"
Running Top 10 Hashtags Analysis...
Loaded 100 tweets from data.json
No hashtags found in dataset.
Done!
Process finished with exit code 0
LAB 2
top_ten.py
52:14
CRLF
UTF-8
4 spaces
Python 3.12