Case Study 5

Sentiment Decider

Due: XXX

**Sentiment Decider**

Welcome to your first day on the job at the fancy tech company, Mamazon. Here at Mamazon, we like to learn all that we can from our customers to make sure that they are happy with our products. In order to do this, we analyze all of the reviews that are posted for every product. Our goal is that eventually we will use the algorithm you have created so that for products with more positive reviews, we can promote those by ranking them higher in the search results. For all products with more negative reviews, we can use your algorithm to punish those by ranking them lower in the search results.

**Problem 1 (60 points) Create Your Sentiment Analyzer**

**Part 1: What's the sentiment of all of the reviews combined?**

To get us started on the right track, you will be creating a function that analyzes the sentiment of all of Mamazon reviews that we will be analyzing. In order to create this function, you will have access to 2 text files: positivewords.txt and negativewords.txt. You will start by creating a dictionary for all of the positive words and a dictionary for all of the negative words. For each word you encounter from all of the amazon reviews, you will keep track of how many positive words you come across and how many negative words you come across.

Your function should take in three parameters: positiveWordsFile, negativeWordsFile, and reviewsFile. These are all strings that represent the name of those files, respectively. Your function will need to:

1. Create the dictionaries to store the positive and negative words

2. Open and read the review file

3. Parse all of the words in the file

4. Count how many positive and negative words you've encountered

5. Print the quanitity of positive and negative words

5. Return a string that lets us know if the majority of the reviews were positive or negative. If there are more positive reviews, return "The Reviews are Mostly Positive". If there are more negative reviews, return "The Reviews are Mostly Negative".

Use the following function header:

**def totalReviewSentiment(***positiveWordsFile, negativeWordsFile, reviewsFile***):**

**Part 2: Make Your Function Cleaner!**

Using functional decomposition, try to break your functions into several smaller sub-functions that are each in charge of one task. Create on “main” function that calls all of your sub-functions in order to get the same results that you got in part 1.

**Part 3: What's the sentiment of a single review?**

For this task, you will be creating a function that analyzes the sentiment of a single Mamazon review. This function will take in a string *review*, and calculate the sentiment for that review. It will also take in *posWordsDict* and *negWordsDict* which both represent dictionaries of each of the positive and negative words.

In order to calculate the sentiment, we are going to first calculate just the positive sentiment. This can be calculated by dividing the number of positive words over the number of total words for a review.

Example #1: If I had 3 positive words and 2 negative words in a sentence, then my \*\*positive sentiment\*\* would be positive words / total words --> 3/5 --> 0.6

Example #2: If I had 12 negative words and 6 positive words in a sentence, then my \*\*positive sentiment\*\* would be positive words / total words --> 6/18 --> 0.3333333333333333

Now that we have our positive sentiment caclulated... how do we calculate the negative sentiment?

Well, it's just the inverse of the positive sentiment!

Example #1: negative words / total words --> 2/5 --> 0.4 == 1 - 0.6(positive sentiment)

Example #2: negative words / total words --> 12/18 --> 0.66666666 == 1 - 0.3333333(positive sentiment)

How can we tell if a review is more positive or more negative?

Well, if we just calculate the positive sentiment - then we can check if that number is above 0.5 or below 0.5.

positiveSentiment > 0.5 == review that is more positive than negative

positiveSentiment < 0.5 == review that is more negative than positive

For your function, you will return a string that states whether this review was more positive or more negative. If it was more positive, print "This review has a positive sentiment". If the review is more negative, print "This review has a negative sentiment".

Use the following function header:

**def reviewSentiment(***review,posWordsDict,negWordsDict***):**

**Problem 2 (20 points) Can You Beat The Hackers? Let's "Game" The Algorithm**

Many hackers and producers of products for Mamazon have figured out how to "game" our algorithm. They have found different ways to promote their products higher up on our recommendation algorithm by getting their products ranked higher on the search page. There are many ways they have done this, and one of them is to write fake reviews that our algorithm thinks are really positive (even if they aren't real).

In order to prep for this kind of algorithmic "gaming" - let's step into the role of the hacker. Try to come up with several examples of reviews that you can pass into your function above that seem "positive" to a human, but come out as "negative" to your algorithm and visa versa. What are some interesting trends you notice? How did you figure out ways to "game" the algorithm?

**Problem 3 (20 points) Put it All Together**

We have written code for you in order for us to loop through all of the Mamazon reviews and come up with the average positive sentiment for the reviews... but there is something wrong with our code! We've encountered 3 hours and we can't figure out how to debug them. The location of the errors is depicted below. Please help us fix our function so we can finish out our sentiment decider! Use the code that is provided in the template jupyter notebook and get the code running so that the test case prints out our final answer!

You will know you have successfully fixed all of the errors if your code prints out the following message:

*The average sentiment of all of the reviews is: 0.5405544234784411*

----------------------------------------------------------------------------------------------------------------------------

Amazing work! On behalf of everyone at Mamazon, we want to thank you for your dedication to our company and to our customers, but especially to those who create the products. We are sure they will be grateful for your sentiment analysis algorithm for their reviews so they can get the rankings they *deserve*!!!