# Acme Problem Instructions

**Problem 3 20 points**

**Introduction**

The hypothetical business scenario related to Problems 3 through 6 pertains to a customer feedback web application for the fictious Acme company. Acme customers can enter product reviews on the company’s web site. For Problems 3 through 6, image that a process runs periodically to get the customers’ product reviews from the company’s web application. This process puts the reviews into a file that you’ll use as input for this problem.

Your job in Problems 3 through 6 is to write R scripts to do various kinds of processing on the customer review data.

Hint: it might be helpful for you to open the **customer\_product\_reviews.tsv** file in a text editor to review it, but don’t make any changes to the data. There are a small number of records in the file so that it will be easier for you to understand the data and then to write your code.

The code you write for this Problem 3 will examine each product review record in the input file and then determine if the customer’s sentiment about the product is positive, negative or unknown.

**Note #1:** To keep this problem simple, the information below describes how you can determine if a sentiment is positive, negative or unknown.

Positive Sentiment

Positive sentiment words: **good, great, happy, super, like**.

Also, any word that starts with the letters “nic” must be regarded as a positive sentiment word.

Negative Sentiment

Negative sentiment words: **bad, terrible, stinks, awful, cheap**.

Also, any word that starts with the letters “dis” must be regarded as a negative sentiment word.

Unknown Sentiment

If a review has no positive sentiment words and no negative sentiment words then the sentiment for that review is unknown.

You can assume that no review will have both positive and negative sentiment words, i.e. the sentiment for each review will be clearly positive, negative or unknown.

**Note #2:** You can assume that all of the words that indicate a sentiment will be in all lowercase.

**Instructions**

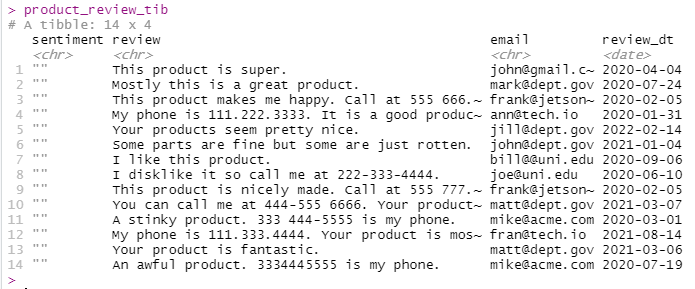
**Part 1**

Read the input file named **customer\_product\_reviews.tsv** to create a tibble named **product\_review\_tib**. You should have previously downloaded this file from Blackboard.

**Part 2**

Write code in this part to add a variable named “sentiment” to **product\_review\_tib**.

Print **product\_review\_tib** to the console. You must see the following:



Note: the data in the **email** and **review\_dt** variables are not used in this Problem 3. You’ll use the **review\_dt** in a later problem. You won’t use the **email** variable in this problem set assignment, but it will “ride along” in these problems. When you’re doing professional data analysis work, you’ll often encounter data in a dataset that is not used directly in your processing. Sometimes you’ll omit the unneeded data, but other times you may decide to retain the data for possible future use.

**Part 3**

Step 1

This step is already coded in the starter script. Don’t add or change any code to this step.

The code in this step creates a character vector that contains a regular expression that can be used for matching on positive sentiment words.

Step 2

Create character vector named **negative\_word\_regex\_pattern** that contains a regular expression that can be used to match on negative sentiment words. Refer to the Introduction section of this problem for which words are regarded as negative sentiment words.

Hint: if you are unsure how to code this step, look again at Step 1.

**Part 4**

Write code that uses the regular expressions from Part 3 to evaluate each product review to determine if the sentiment for that review is positive, negative or unknown.

Your code must populate the **sentiment** variable in the **product\_review\_tib** object as described below:

1. If the review contains a positive word then put **positive** into the sentiment variable for that observation.
2. If the review contains a negative word then put **negative** into the sentiment variable for that observation.
3. If the review does not contain a positive word or a negative word then put **unknown** into the sentiment variable for that observation.

Hint: you could consider using the str\_detect function

**Part 5**

Write code that writes the data now in **product\_review\_tib** to an output file named **customer\_sentiments.tsv**. There must be a column header record in the output file.

You may visually compare your output file to the file named **customer\_sentiments.tsv** to verify your script creates the output file as expected. Viewing the **instructor\_customer\_sentiments.tsv** file will let you see what needs to be in the column header record, i.e. the first record in the file.

Note: When you run your code for this problem, you might see RStudio warning messages such as what’s shown below, you can ignore these warning messages.



