**Task 4: Evaluate information needs for a new project**

[INTRODUCTION](https://ut.daacertificate.com/mc/poa?productID=2654&taskID=3348#introduction)

**Your Task**

Michael Ortiz, the SVP of Alert Analytics, has asked you to evaluate the questions in Julia Chan's ***Generation Luxe Analysis Reques***t and identify any ambiguous questions that require clarification and any questions that a sentiment analysis can't accurately answer.

This task requires you to prepare one deliverable:

***Generation Luxe Analysis Report*** a 3-5 page Word document explaining your analysis of the questions in the ***Generation Luxe Analysis Request*** which includes:

* For each question that you believe can be answered, a brief sketch of the types of features (i.e., data patterns) that you will capture and tabulate with accompanying examples
* A list of questions that can't be reliably answered by using data analytic methods and an explanation of why each question is problematic.
* Any additional information required to clarify ambiguous questions and why that information is needed.
* Optionally, any alternative questions that a sentiment analysis can answer and which will help Julia make the business decisions she has enumerated.

TIP:

In real-world situations, defining the information needs of an analysis requires an exchange with project stakeholders over several cycles. However, if the process exceeds about five revisions, it may be time to revisit what the purpose of the project is supposed to be. It is better to get the confusion out of the way early than realizing the results are not what the stakeholders really wanted.

#### [1. Get Started](https://ut.daacertificate.com/mc/poa?productID=2654&taskID=3348" \l "collapsepoa3298)

1. Download and read the Generation Luxe Analysis Request attached to Michael Ortiz’s email. This document explains the stakeholder's business needs and the questions she proposes that the analysis answer. This task requires that you review and comment on this document.
2. Consider the limitations of data analysis. Read the article [What Data Can't Do](http://www.nytimes.com/2013/02/19/opinion/brooks-what-data-cant-do.html?_r=2&). Do any of the questions the client, Julia Chan, requested seem overly ambitious based on the limitations explained in this article? Keep track of your observations for possible inclusion in your report.

#### [2. Assess the Analyst Request](https://ut.daacertificate.com/mc/poa?productID=2654&taskID=3348" \l "collapsepoa3299)

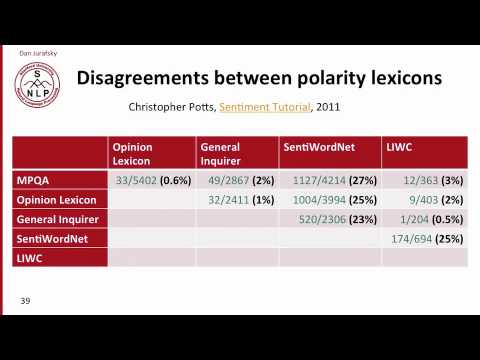
1. **Evaluate the feasibility of answering the questions the client has requested.**Use your hands-on experience in the previous tasks in this course to help you assess if the information requested can be reliably and efficiently inferred from analyzing the text on a web page.
   1. For each of the questions enumerated in the Generation Luxe Analysis Request, think through what features could be captured to determine if the page is about the target topic and secondarily, if so, what features could be captured to classify the sentiment expressed about that topic as positive, negative or neutral.
      * Consider the topic we want to capture the sentiment towards; what word or pattern of words would clearly indicate that the page was about that topic?
      * Consider the attitude expressed on the page; what words or patterns of words that would clearly indicate if the author's attitude towards that topic was positive, negative or neutral.
      * Consider patterns of words that a simple approach might misinterpret.
   2. Based on the experience of trying to extract features in the previous tasks, do you think that there are patterns of words and phrases that will enable reliable answers to each of Julia Chan’s questions?  Provide a few examples for each question.
   3. Keep track of your answers and examples to help you draft your sketch of the types of data patterns that can be feasibly captured to address the questions you believe can be answered and formulate your recommendations regarding questions that you believe can’t be answered.

TIP:

While they are potentially more reliable, the natural language "processing" techniques developed by artificial intelligence researchers are too computationally complex to use effectively across hundreds of millions of web pages. In other words, the techniques can take too long or cost too much to process in the context of millions of webpages.

TIP:

Sentiment analysis experts often leverage existing lexicons and affective word lists to categorize the sentiment of a given text file. These resources are large lists of words and phrases that indicate a type or polarity of sentiment. For example, General Inquirerfrom Harvard University provides lists of words indicating positive outlook, negative outlook, hostility, pleasure, pain and many more. If you want to learn more about sentiment lexicons, watch the video below.



TIP:

Webpages can be segmented by date, language or other meta characteristics by checking for specific information in the meta data or the body of webpages. This processing technique can enable analysis to compare different segments of the data, for example changes in sentiment overtime or differences in sentiment in different geographical regions.

1. **Determine if tabulated sentiment scores will be sufficient to answer each question**. In the previous step, you examined if information can be extracted from an individual page that is relevant to each specific question requested by the client. Now evaluate, for each question, if an enumeration of simple counts of positive, negative and neutral pages about the topic of interest will yield a satisfying answer "in the large." Keep track of your observations to formulate your report.
2. **Identify any questions which need to be better defined to answer reliably.** As you experienced in previous tasks, the program will look for specific words, phrases or patterns of words to identify if the webpage is about the specific topic relevant to the question, such as a specific brand, product, or product feature.The team working on this project will need to enumerate the specific words, phrases and word patterns for each question before the analysis begins.   
   Revisit the questions:
   1. Are all the topics for which sentiment will be captured clearly specified?
   2. Are there any questions for which you were unsure what information to capture?
   3. What clarifications from the client will help specify the topics for which sentiment will be captured?
   4. Note any questions for which the information to be gathered is ambiguous and what clarification is required from the client to eliminate that ambiguity.
3. **Identify any alternative questions.** Revisit the business needs in the document. Brainstorm any other questions which will help Julia make the business decisions she has enumerated. Evaluate the questions you come up with—can they feasibly be answered using sentiment captured on the web? Will tabulated sentiment scores be sufficient to answer each question? Note any questions you want to include in your report and sketch how they will be answered.

#### [3. Write Up and Submit Your Report](https://ut.daacertificate.com/mc/poa?productID=2654&taskID=3348" \l "collapsepoa3300)

1. Draft and your ***Generation Luxe Analysis Report***. Refer to the notes that you made as you worked through the questions in the Generation Luxe Analysis Request and write up a 3-5 page report in a Word document that includes:

* A sketch of the types of features (i.e., data patterns) that can be feasibly captured to accurately address the questions you believe can be answered
* Any information required to clarify ambiguous questions and why that information is needed.
* Each question that can't be accurately answered by a sentiment analysis and your rationale for why that question is problematic.
* Optionally, any other questions that a sentiment analysis can answer which will help Julia make the business decisions she has enumerated, and a sketch of the data patterns you will capture to answer these questions.

1. **Proofread and Submit Your Report** using the Submit Your Work tab.