# **Teacher Notes**

# **Sentiment Analysis**

## **Motivation and Essential Understandings**

Data science is a field that crosses multiple disciplines. Data scientists utilize many different methods and statistical techniques to gain knowledge and insights about data, including using text data for natural language processing. One such method is called Sentiment analysis. The goal of this lesson is for students to gain exposure to the kinds of data that can be used with sentiment analysis and to think critically about how the data can be analyzed using a lexicon-based approach to this method.

* What is data science?
* What techniques do data scientists use to analyze data?
* How can Sentiment/Lexicon Analysis be used to analyze customer perceptions?

## **Context and Dataset**

Many customers have differing perceptions of flying during the COVID 19 pandemic. The question is how to collect data regarding customer response to COVID 19 as well as what to do with data once it is collected. Lexicon-based Sentiment Analysis is a method that applies a number value to words/phrases (lexicons) based on the positive or negative nature of the response. You will be guiding your students to understand and evaluate the use of Sentiment/Lexicon Analysis to analyze customer response data.

## **Learning Objective**

## This lesson will introduce students to the application of sentiment analysis to customer perceptions or reviews. This lesson will also teach students to open RStudio Cloud, modify a file, and generate their own results.

## **Learning Outcomes**

Students will be able to:

* Describe the field of data science.
* Describe what natural language processing is and how it might be used
* Understand the use of Lexicon-based Sentiment Analysis to analyzing a structured dataset
* Apply Lexicon-based Sentiment Analysis to a twitter dataset containing 100,000 tweets
* Identify relationships between Lexicon-based Sentiment Analysis and the broader fields of AI / ML
* Modify function parameters in RStudio Cloud
* Run sentiment analysis code in R and understand outputs

## **Analytical Concepts and Skills**

1. Formulating hypotheses and Asking questions
2. Analyzing and interpreting data
3. Using computational thinking
4. Constructing explanations based on the analysis of data

## **Target Audience**

This lesson is geared towards undergraduate Students whose interest is in business, marketing, and mass communications. This lesson assumes students understand how to open and navigate RStudio Cloud and have a basic understanding of an Rmd file. This lesson also assumes students have experience with R code and are comfortable with basic data manipulation.

## **Lesson Outline/Narrative**

**Time allotment**: This lesson can be taught in one class period, with the option of another class period being used if the assignment is to be done in-class (as opposed to take-home). If students are not familiar with RStudio Cloud, another class period should be used for students to create free accounts and walk through the site.

**Teacher Overview Materials**

Please review the following documents prior to delivering this lesson:

* Sentiment\_analysis.pptx

**Technical Tools:** This lesson will be delivered via zoom.

|  |  |
| --- | --- |
| **Part 1: Introduction to Data Science/The Problem/Sentiment Analysis** | |
| **Narrative:**  Introduce the field of data science. Introduce the data science pipeline. Very briefly how ML fits in. Discuss where this lesson’s components fall into the data science pipeline  **The problem:**   * How do customers feel about airlines as they respond to covid-19? * Which airlines are customers more positively to, and which are they responding more negatively to? * Have customer feelings about airlines been relatively consistent since covid began, or have their feelings shifted since March? If so, how?   **Business understanding:**   * What might we do with the answers to these questions? How could we use these answers?   **Discussion: Analysis techniques used in this lesson**   * **One approach to answering these questions is by using a machine learning technique called sentiment analysis.** * **What is sentiment analysis?** * **When might you use sentiment analysis?** * **Where does sentiment analysis fit into the data science pipeline? (modeling)**   **Assessments:**  Pre-assessment | |
| **File/Material** | **Description/Notes** |
| sentiment\_analysis.pptx | Power Point presentation to walk through concepts |
| Q&A | Questions to check students’ understanding as they follow along with the Power Point |
| **Part 2: Data Understanding/Data Analysis** | |
| **Narrative:**  **Discuss Data understanding:**  Walk through the data set of tweets from customers to 6 different airlines since Covid began. Explain what exploratory data analysis (EDA) is. Show a word cloud of tweets (this is EDA, part of data understanding).  Explain that we iteratively gain more understanding through EDA and will likely modify the questions we're asking as we gain new insights.  **Discuss Data Preparation:**  Explain that we often have to do some data cleaning before we begin - explain what this is without going into detail about how it was used in this particular case, e.g. explain why stop words ("the", "a", etc.) were removed before we made word clouds in the last step, because those words don't tell us interesting things.   * **Note** that cleaning is often 80% of the work in data science, and there are some really complex methods for this part of the pipeline such as data imputation   **Discuss Data Modeling:**  Explain how sentiment analysis works at a concept-level and how to interpret it. Emphasize the type of data needed for sentiment analysis to be successful (Quiz on this at some point using multiple choice examples). Show graphs from results comparing between several , walk through some of the insights.   * We previously discussed what you might do with the results from these questions. Now we have the answers! How is this useful to airlines? What does this really tell us? Get students to discuss this in groups.   **Discuss Data evaluation and deployment:**  Briefly explain the importance of initially and iteratively checking the validity of our models before deploying them, or building the infrastructure needed so we can keep running this model continuously rather than having to redo this analysis over and over.   * How / When might the methods used in this lesson be applicable to the students’ domain? (e.g. provide some example applications of sentiment analysis in Mass Communication) * Ask students in groups what other scenarios they could use sentiment analysis in and exactly what data set they might be able to use. How might they get that data? Is it actually accessible? What questions could they answer?   **As an ending / recap slide: Tie back into the pipeline and relation to AI / ML. Review**  **Assessments:**   * Assignment to be distributed as take-home or in-class work (Mod\_Questions.docx) which can be guided by *Getting Started in RStudio Cloud.docx*   + Note that this assignment is broken into 4 parts. Parts can be assigned partially or over time, though they should be covered in sequential order. For example, parts 1 and 2 should be assigned and completed before or alongside part 3. * End of lesson Survey | |
| **File/Material** | **Description/Notes** |
| sentiment\_analysis.pptx | Power Point presentation to walk through concepts |
| Q&A | Questions to check students’ understanding as they follow along with the Power Point |
| Create\_Mass\_Communications\_Lesson\_Overview.docx | Overview of instructional design components of lesson |
| Create\_Questions.docx | Assignment to be distributed as take-home or in-class work |
| Getting Started in RStudio Cloud.docx | Instructions to access the online files needed in RStudio Cloud and guided setup instructions |