CS – Finding Prevalent Jeopardy Words

DS 4002 – Fall 2023

Submission format: GitHub repository link

Individual Assignment

General Description: Submit to canvas a link to your GitHub repository

Preparatory Assignment: None

Why am I doing this? This is a chance to use the lessons learned in the course to employ skills at creating and optimizing models. Additionally, it will challenge you to analyze the results with respect to real world trends.

- Course Learning Objective: Analyzing text data
- Course Learning Objective: Prepare findings for written communication

What am I going to do? You will read the hook document to understand the prompt. The document will outline the tasks you would need to accomplish for this assignment. You will create word clouds for text analysis. Create as many word clouds as you find necessary to sufficiently analyze the data. You will then analyze the data to find trends within the data and compare this to the real world.

Tips for success: Talk to the professor and TA and be sure to ask any questions

How will I know I have Succeeded: You will meet expectations on CS - Finding Prevalent Jeopardy Words when you follow the criteria in the rubric below.

Spec Category	Spec Details
Formatting	The final public GitHub repository should contain all of the following components: - README.md - License - Deliverable Model: Word Cloud - Figures Folder - Model Analysis document
README.md	Goal: Provide a brief overview of the case study and analysis outcomes Includes the following sections: - Title, Date, Author - Research Question - Brief Context - Word Cloud code - Conclusions analysis (brief assessment of words clouds and what trends you found) - References used (in IEEE format)
License	Goal: Explain to repository viewers the terms of using and/or sharing code found in this repository - MIT license is generally the default
Word Cloud Model	Goal: Create a file with Word Cloud models for determining word prevalence - Should have enough models to conduct analyses on the most prevalent words in Jeopardy - Models can vary in terms of size, types of words included/excluded, year range, etc. - Models should contain enough information to show trends within the show and trends within the real world at the time
Figures	Goal: This folder contains all the figures generated by your project - Include with every figure with relevant notes about the figure
Model Analysis	Goal: Create a brief document outlining the different word clouds created and what takeaways they provide - Should include information about WHAT specifications the word cloud model has and WHY you chose these specifications - Discuss the takeaways from the word cloud in terms of what types of questions Jeopardy asks and how they have changed throughout the years - Provide insight to how these changes may correlate to changes within Jeopardy and the real world - Explain how the information from the word clouds could potentially help someone who would hope to be on jeopardy someday - Should be no longer than 1 page

Finding Frequently used Words in Jeopardy through Word Cloud Analyses

A UVA Data Science Case Study by David Bergman, 2023

Prompt: The television industry is a worldwide market and there exist certain genres within television that have extensive viewership. Gameshows, specifically, gain widespread attention with their ability to attract audiences with the appeals of wealth and trivia. Jeopardy is one of the most popular game shows in the United States as it is considered a "cultural staple and consistently rates as one of the most popular shows in the country." [1] Additionally, Jeopardy has educational applications as demonstrated by Mark Simkin's article on playing Jeopardy in the classroom: Jeopardy can be used as a learning and teaching tool, something that proves to be highly valuable in classroom settings [2]. Knowing that Jeopardy has such benefits and applications to Americans, it is important for us to understand what comprises Jeopardy questions and how they are produced. A greater understanding of questions and their qualities will help us better understand the educational practicality of Jeopardy.

Word Cloud models, while seemingly elementary, are useful for finding trends within text data. Word cloud models create visuals that show the prevalence of a word based on its size within an image. The customizability of word clouds makes them useful for this situation, as a user can create many word clouds with a variety of boundaries with the aim of finding key information within the dataset. Your goal in this project will be to develop word clouds to find key trends within Jeopardy that can be used to draw broad conclusions about the educational properties of Jeopardy.

Task: You will be provided with a dataset of all the questions in Jeopardy since 1984. Using this dataset, you are tasked with creating word clouds to determine trends within the game of Jeopardy. You may create however many word clouds using whatever specifications you'd like, as long as you are able to acquire enough information to perform analysis and find trends within the game of Jeopardy.

Deliverables: Successful completion of this case study will involve the creation of a GitHub repository consisting of the following items:

1) Word Cloud models created using the dataset
2) Explanation for each word cloud, explaining what the specifics of a given word cloud was and why these specifics were chosen
3) Brief write-up explaining what trends can be extrapolated from the models you created. The trends will likely be in relation to the changes found within the types of questions asked over time and how certain real-world events have caused certain topics to be more relevant.

[1] "Jeopardy (American television game show)." Britannica, https://www.britannica.com/topic/Jeopardy-American-television-game-show. Accessed: September 13, 2023.
[2]. Simkin, Mark G. "Playing Jeopardy in the Classroom: An Empirical Study." Department of Information Systems College of Business Administration University of Nevada, vol. 24, no. 3, pp. 203-210, Fall 2013. http://jise.org/volume24/n3/JISEv24n3p203.pdf