

Data Analytics Research Project

Sahar Mirzabaki

Student ID: 23011185

Lecturer: Dr. Nicolas Garron

MSc. Advanced Computer Science

Liverpool Hope University

February 2024

Abstract:

The purpose of this research is to analyze data regarding the effect of humour, specifically Scouse humour, on the behavior of individuals when confronted with challenges.

1. Main Question:

Does Scouse's humour influence individuals in dealing with difficult experiences across different demographic groups?

2. Methodology

2.1 Descriptive Analysis:

Analyze the distribution of responses to each question. Calculate summary statistics for demographic variables such as age group.

2.2 Thematic Analysis:

Analyze responses to open-ended questions regarding Scouse humor's distinctiveness and role in coping. Identify recurring themes and patterns in participants' perceptions.

2.3 Required tools to aim this analysis:

Handling Missing Data: In this research, pandas is used to identify and handle the missing values. Additionally, Anaconda Navigator, and Visual

Studio Code's development environment facilitate and improve the experience of working with Pandas and NLTK. Other libraries that are used in this research are sklearn and matplotlib.

In this research, textual responses to the questions are categorized into codes using NLTK's tokenization.

Data Review:

Question 1:

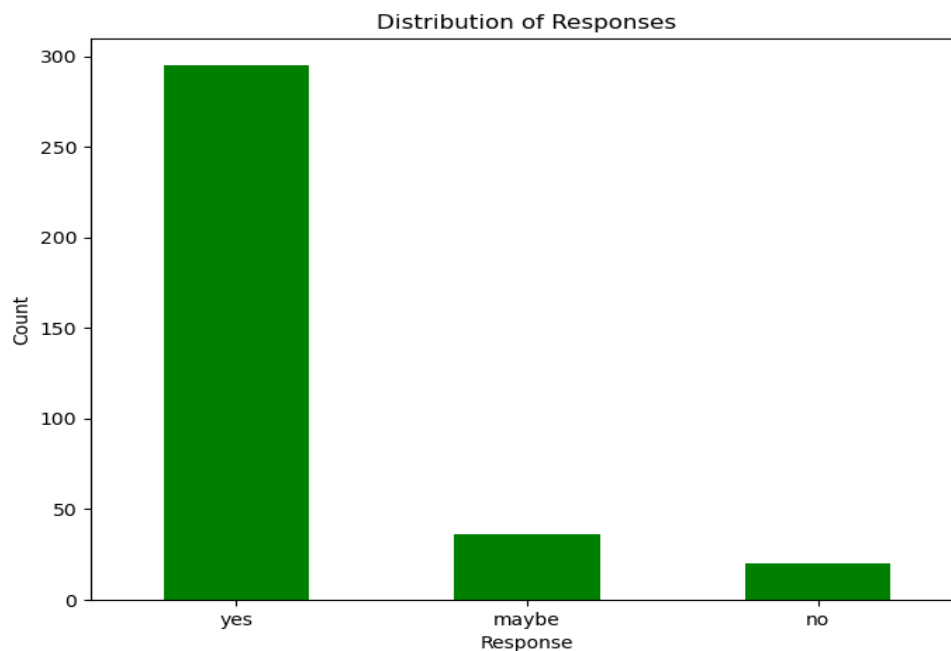
"Do you think there is something distinctive about Scouse humour?"

This dataset comprises 351 responses to this question. The responses are categorized into three choices: 'Yes', 'No', and 'Maybe'.

The number of Yes answers is: 295 (84.04%)

The number of No answers is: 36 (10.25%)

The number of Maybe answers is: 20 (5.698%)



The majority of respondents, around 84%, believe that there is something distinctive about Scouse humor. This suggests that a significant portion of people

recognize unique qualities in Scouse's humor, while a smaller percentage expresses uncertainty or disagreement.

Question 2:

“Do you have any thoughts on what makes it distinctive?”

In this open-ended question, a word cloud is generated to visualize the most common words.



Also, for more analyzing this question, implemented a topic modeling using Latent Dirichlet Allocation (LDA) algorithm and TF-IDF Vectorization to extract the top words. The results are shown below:

Topic 0: people understand language unique observational rude ability laugh mick way

Topic 1: quick humour funny witted just really sharp quite dry time

Topic 2: sarcasm self sarcastic laugh deprecating good dialect local fun close

Topic 3: accent dry wit witty quick don city slang point kind

Topic 4: sense earth joke humour scouser make different dry funnier fast

The words like “funny”, “quick”, “unique”, “slang”, “sharp”, “accent”, and “fast” address the most common reasons for being distinctive.

Another method that has been done to analyze this question is the implementation of the Key-Means algorithm on data. In this algorithm, the data is divided into several clusters according to their similarities.

In this case we have 10 clusters and each cluster represents some similar word for example cluster 1 is:

Cluster 1:

it is witty and thought provoking, quick and witty, witty along with accent, dry and witty, witty character, offensive humour, famous across the country to be witty, we are so quick with a witty reply

As evident from the analysis, responses containing the term "witty" have been clustered together. These clusters provide a clearer perspective on the responses as a whole.

Question 3:

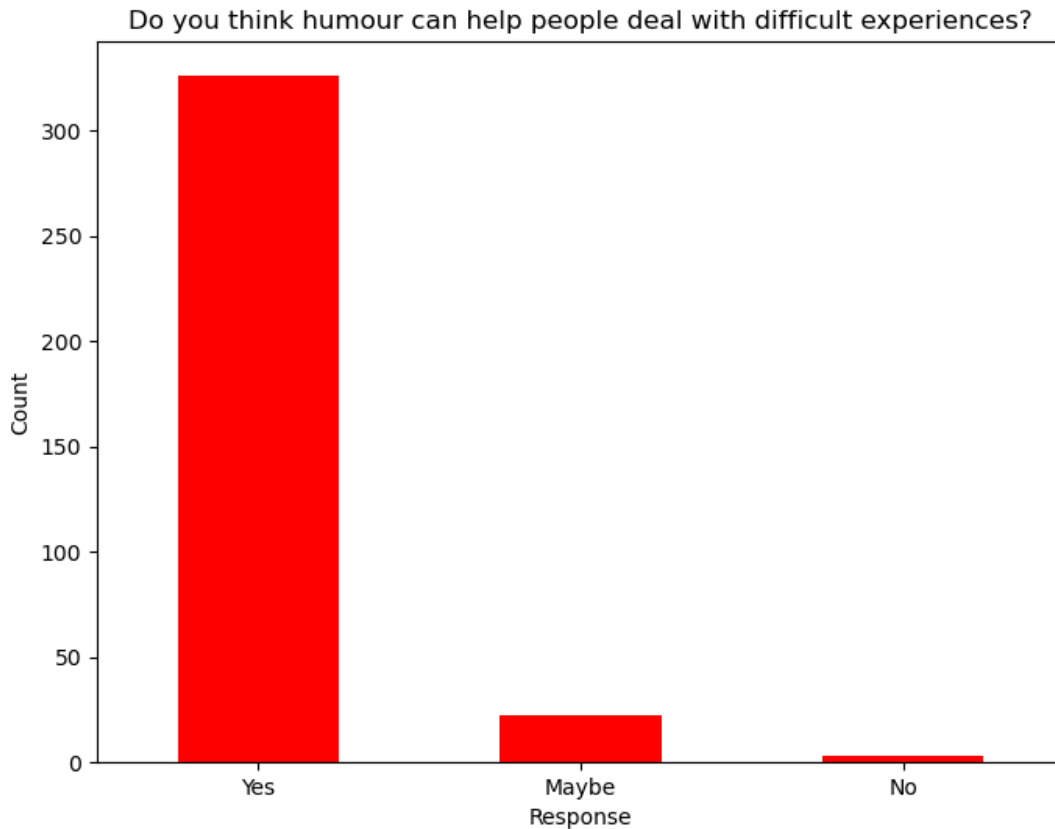
“Do you think humour can help people deal with difficult experiences?”

The responses are categorized into three choices: 'Yes', 'No', and 'Maybe'.

The number of Yes answers is: 326 (92.87%)

The number of No answers is: 3 (0.85%)

The number of Maybe answers is: 22 (6.26%)



The vast majority of respondents, almost 93%, believe that humor can help people deal with tough times. This shows that most people recognize the importance of humor in helping us cope with difficult experiences.

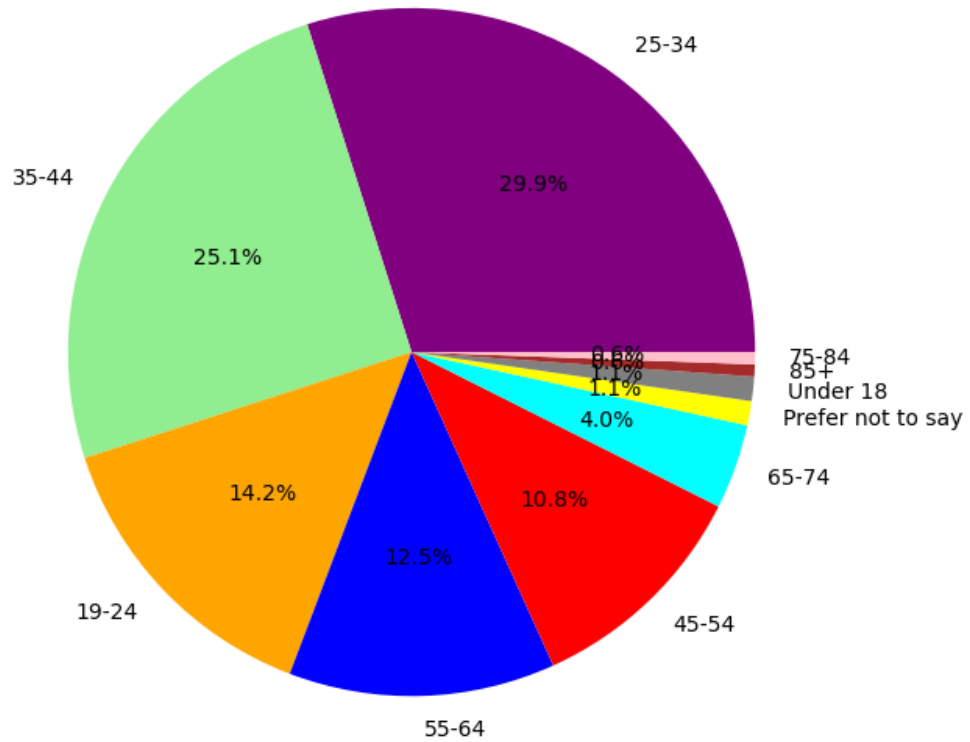
Question 4:

“Would you mind saying what age group you belong to?”

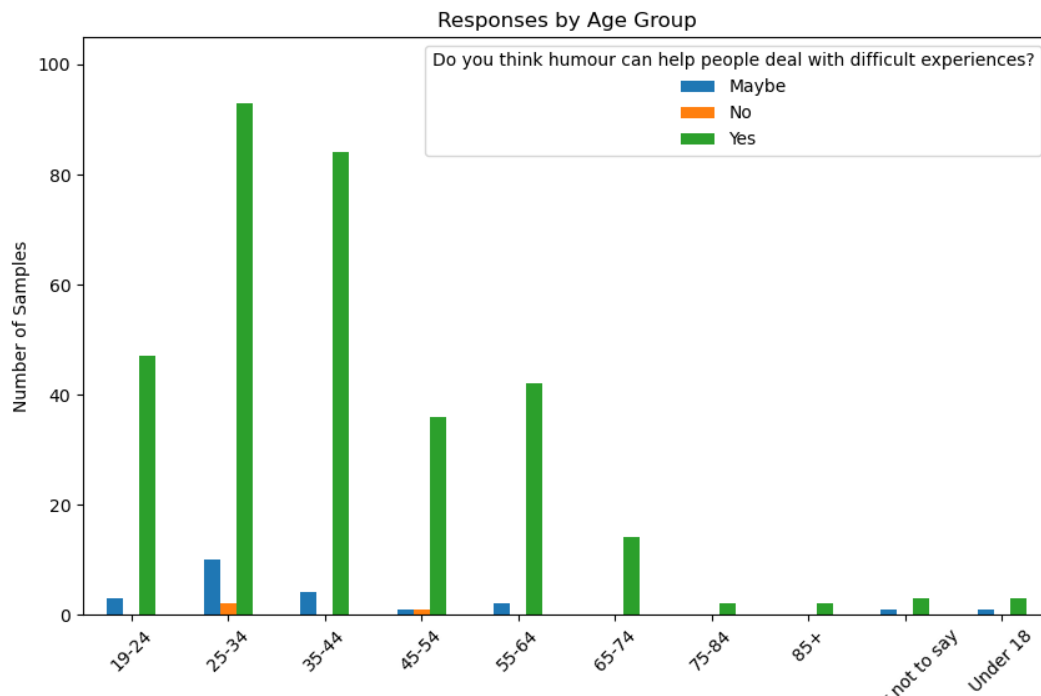
The following pie chart represents the age range for this survey.

This pie chart highlights that the age groups 25-34 and 35-44 are the most prevalent among respondents, comprising the majority of participants.

Distribution of Age Groups



Also, the below chart illustrates the distribution of responses ('Yes', 'No', and 'Maybe') to the question: 'Do you think humor can help people deal with difficult experiences?' across different age ranges.



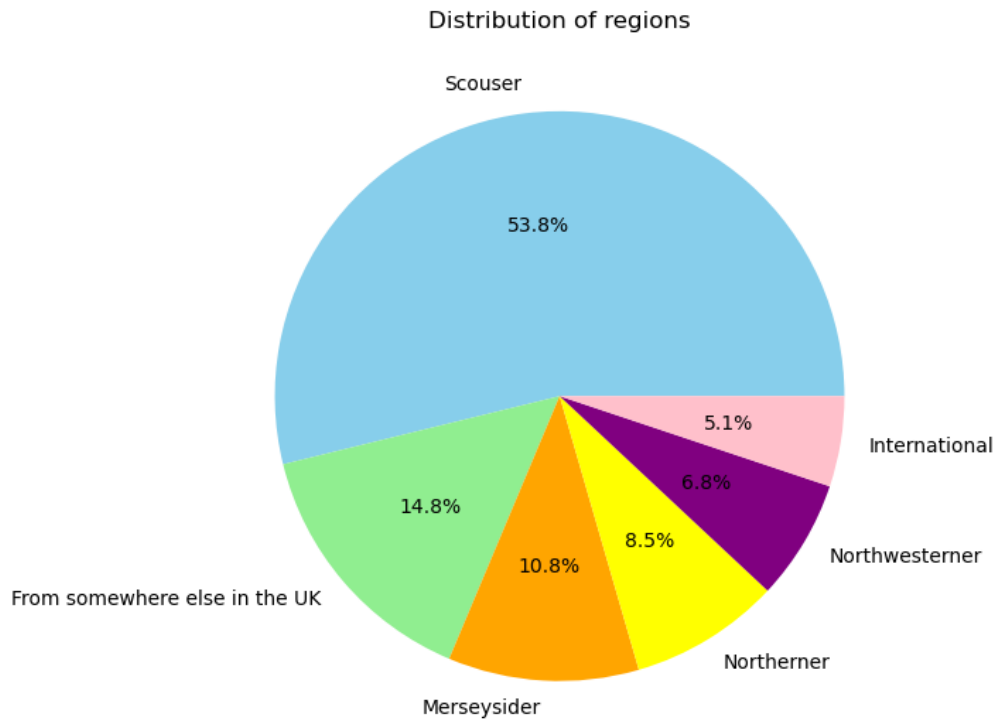
This chart shows that the highest proportion of 'No' answers is associated with the 25-35 age range. However, it's essential to note that approximately 30% of participants fall within the 25-34 age bracket alone, which complicates direct interpretation. Additionally, the 45-54 age group also shows a significant percentage of 'No' responses compared to other age groups.

Question 5:

“How would you identify yourself from the following:”

The answers are: Scouser, From somewhere else in the UK, Merseysider, Northerner, Northwesterner, International

The below chart shows the percentage of each region:



Critical Thinking:

While our analysis helps us understand common themes and patterns in responses, we need to be careful when interpreting the results. Clustering responses based on keywords might make things seem simpler than they actually are. We should keep in mind that people's attitudes and beliefs are complex and influenced by many factors.

Moreover, while the utilization of advanced analytical methods like topic modeling and clustering enhances our understanding of the data, but these methods have limits. The results can change based on how we set things up, so we cannot always trust them completely.

Conclusion:

Our findings revealed that a significant majority of respondents perceive Scouse's humor to be effective in facing difficult experiences.

Key characteristics such as quick wit, unique language and dialect, and a strong sense of humor rooted in local culture, are influential in this choice.

References:

<https://github.com/yarikoptic/scikit-learn>

<https://scikit-learn.org/stable/modules/clustering.html#k-means>

https://scikitlearn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html

[Scikit-learn: Machine Learning in Python](#), Pedregosa *et al.*, JMLR 12, pp. 2825-2830, 2011.

J. D. Hunter, "Matplotlib: A 2D Graphics Environment," in *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90-95, May-June 2007, doi: 10.1109/MCSE.2007.55.

McKinney, W. & others, 2010. Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference*. pp. 51–56.

Bird, S., Klein, E. & Loper, E., 2009. *Natural language processing with Python: analyzing text with the natural language toolkit*, "O'Reilly Media, Inc."

Van Rossum, G. & Drake Jr, F.L., 1995. *Python reference manual*, Centrum voor Wiskunde en Informatica Amsterdam.

Anon, 2020. *Anaconda Software Distribution*, Anaconda Inc. Available at: <https://docs.anaconda.com/>.

Oesper, L. et al., 2011. WordCloud: a Cytoscape plugin to create a visual semantic summary of networks. *Source code for biology and medicine*, 6(1), p.7.