Introduction Veganism

Veganism has gone popular, so much so that the number of persons adopting a vegan diet has surged by 350% in the previous decade, according to U.K. study. Veganism is defined as a style of life in which individuals avoid all sorts of animal exploitation and suffering as much as possible. A vegan diet may appear difficult or unduly restricted at first look. Many of my customers who are thinking about being vegan are concerned about finding adequate vegan substitutes to their favourite meals. However, most people discover that once they understand the fundamentals, the change is less challenging than they anticipated.

Vegans that are ethical think that all species have the right to life and freedom. They consider all animals to be aware creatures that, like humans, want to escape pain and suffering. As a result, ethical vegans are opposed to killing an animal for the sake of eating its meat or wearing its fur or skin. Vegans are also opposed to the psychological and physical hardship that animals may experience as a result of current agricultural techniques, such as the small pens or cages in which animals normally dwell and seldom leave between birth and killing. This emotion, however, extends beyond the brutality of current farming techniques for ethical vegans.

This is because vegans are opposed to ingesting items that heavily rely on the slaughter of other animals, especially when alternatives exist. This involves the killing of calves to prevent adding too much delay to an ETA request, which is considered surplus in the dairy business, as well as the culling of 1-day-old male chicks in egg production. Furthermore, ethical vegans typically feel that humans should not exploit animals' milk, eggs, honey, silk, and wool, regardless of the living circumstances provided to the exploited animals. This is why ethical vegans refuse to consume an animal's milk, eat its eggs, or wear its wool, even if the animals are free-roaming or pasture-fed.

On the one hand, we know that shifting to a more plant-based diet helps practically everyone. "However, it is a hard topic just what sort of food system would be optimal, and how we can morally and successfully bring that kind of food system about, Schlottmann and Sebo concede. While putting vegetarians and meat eaters against one another is pointless, we do need to talk about how to tackle an issue that affects us all.

A visualization of Google Search data from the United States was recently produced, demonstrating how levels of interest in vegan and plant-based diets have varied, state by state, from 2004 to 2019. While it is encouraging to see the numbers rising, this type of study is difficult to assess in isolation. Is the rise in interest comparable to that of other non-animal-friendly diets? Do Google searches indicate that individuals are favourable or negative towards diets, or that they are simply inquisitive about what they are?

The present examination of a year's worth of Twitter data gives a more in-depth look at the general public's interest in animal-friendly diets, as well as other animal protection and advocacy topics. The time range is shorter, but the inferences we can draw are stronger.

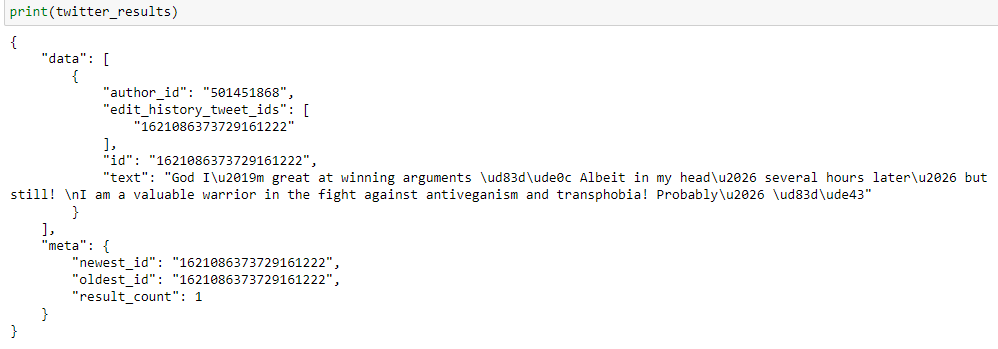
Data – Text Mining:

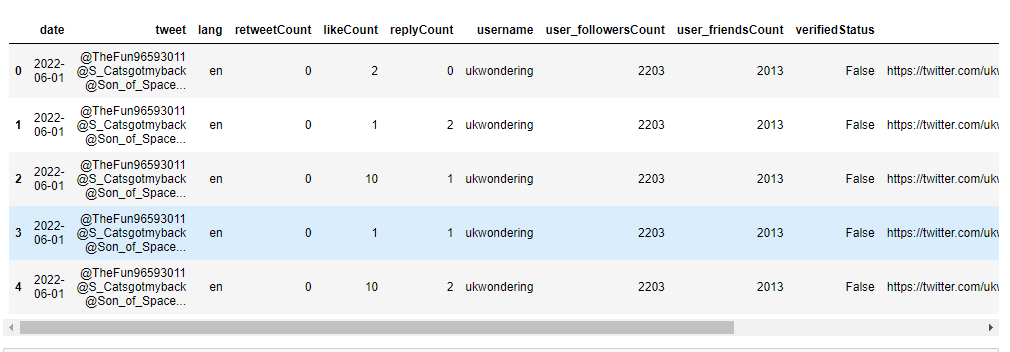
Data Collection:

Twitter API

Data was gathered from Twitter during the months of June 2021 to January 2020. Using Python and the Tweepy Python library, a script was written to access the Twitter API. This includes keywords used in hashtags but also in the body text of a tweet. The keyword are 'vegan', 'govegan', 'veganfortheanimals', 'animalsrights', 'veganism' 'antivegan', 'exvegan', 'yestomeat', 'meatlover', 'antiveganism'. The tweets were grouped into categories by keyword. Each tweet category was then saved into a Python pickle file (.PKL) in increments of 8,000 tweets. It then output a ZIP file PKL files, which were cleaned and converted to CSV format.

Data cleaning entailed removing any duplicate tweets that were found by the Twitter API. The only fields that were kept for each tweet when converting from the PKL files to CSV files were screen name, text, user follower count, retweet count, and date/time.





NewsOrg API

News articles pertaining to veganism were collected using the API.

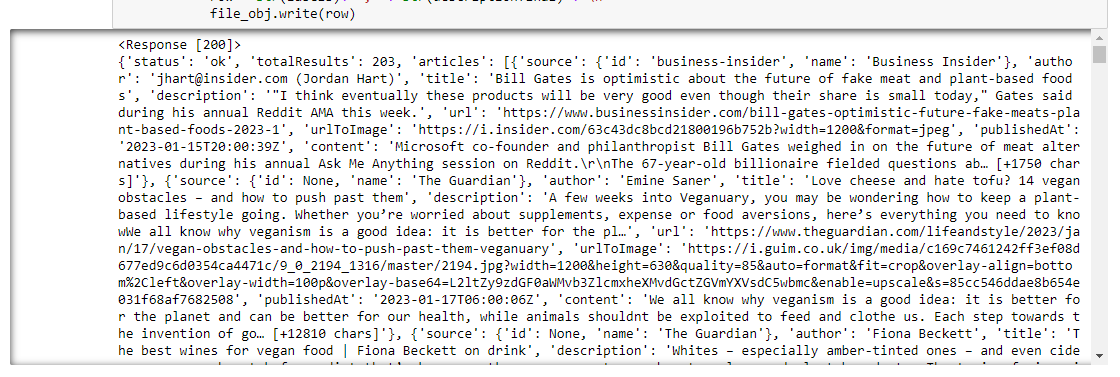
API Call: “https://newsapi.org/v2/everything?q={query}&apiKey={APIkEY}”

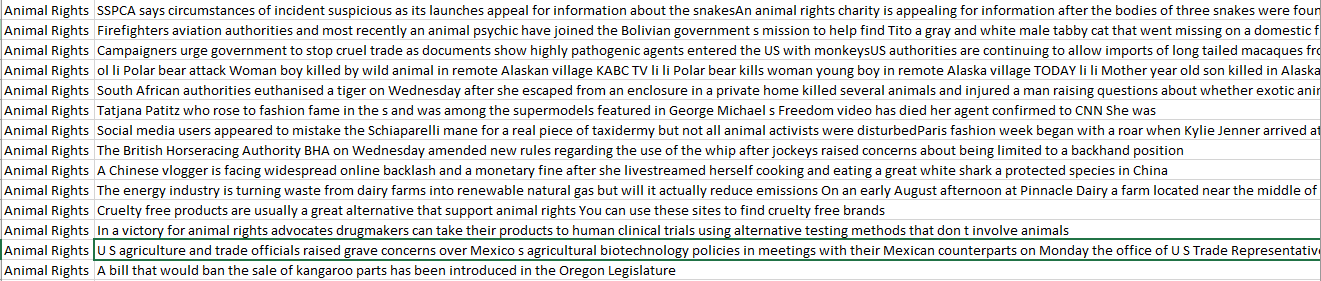
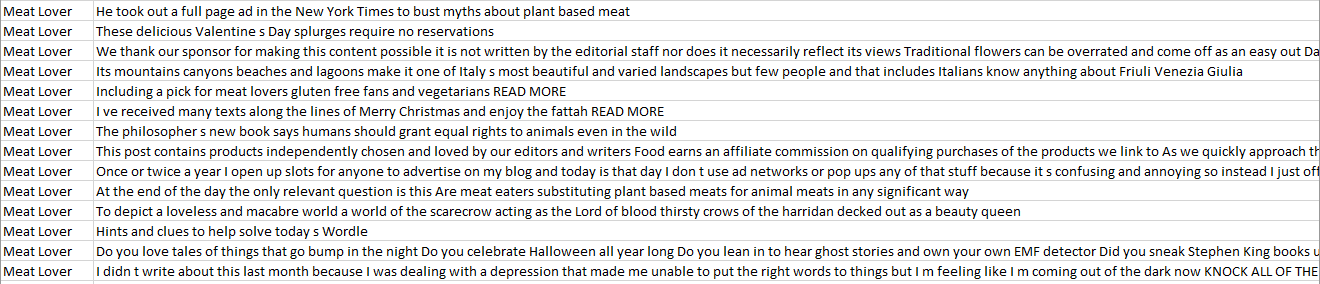
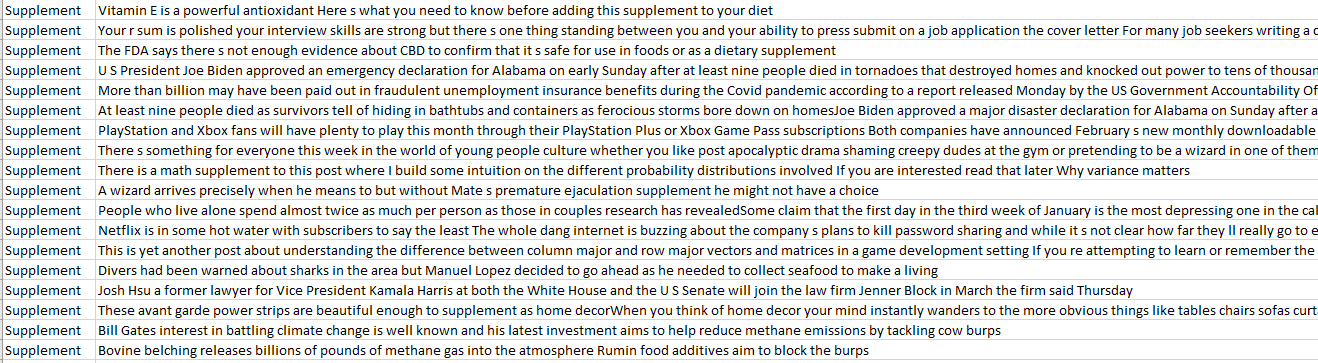
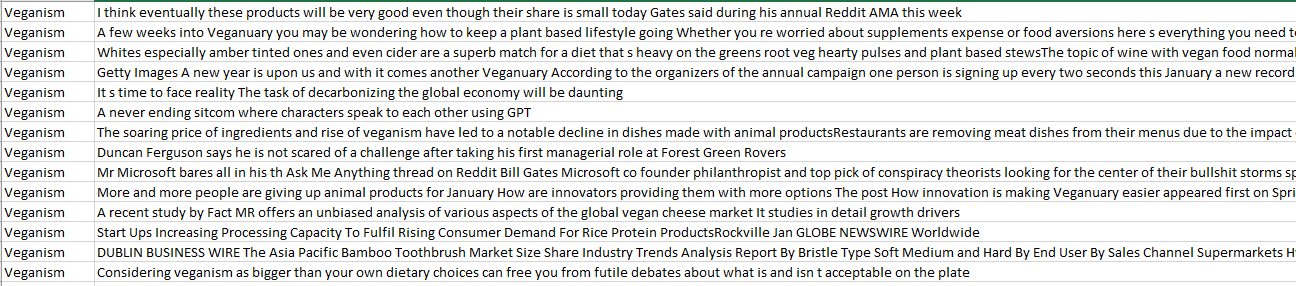
Endpoint: "https://newsapi.org/v2/everything"

q (query): ["Veganism","Supplement","Animal Rights","Meat Lover"]

API Key: unique key

The headlines were grouped into categories by the query. Each headline category was then saved into a text file in increments of 200 files. The data was combined, were cleaned and converted to CSV format



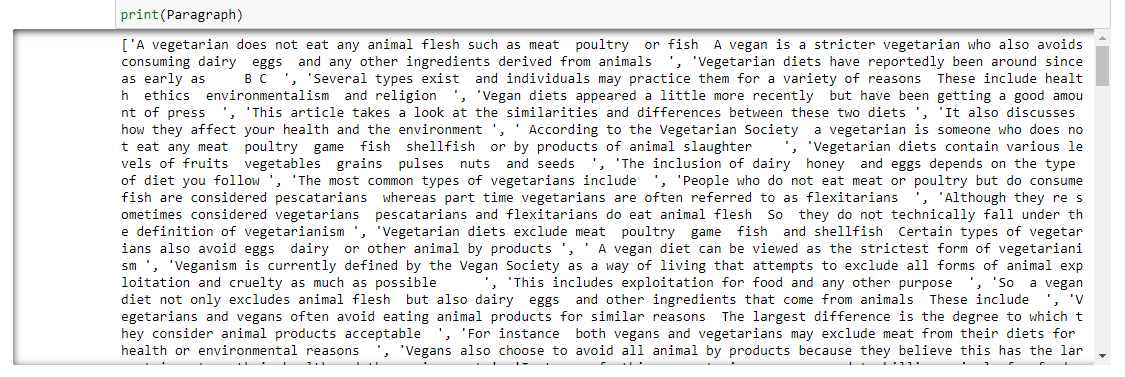
   

Web Scrapping:

Sites:

[Vegan-vs-Vegetarian](https://www.healthline.com/nutrition/vegan-vs-vegetarian) , [Vegan-vs-Non-Vegans](https://blog.insidetracker.com/vegans-vs-non-vegans-who-is-healthier) , [Vegan-vs-Meat-eater](https://www.precisionnutrition.com/vegan-vs-meat-eater) , [What’s the Difference](https://www.naturespath.com/en-us/blog/vegetarian-vs-vegan-whats-the-difference/)

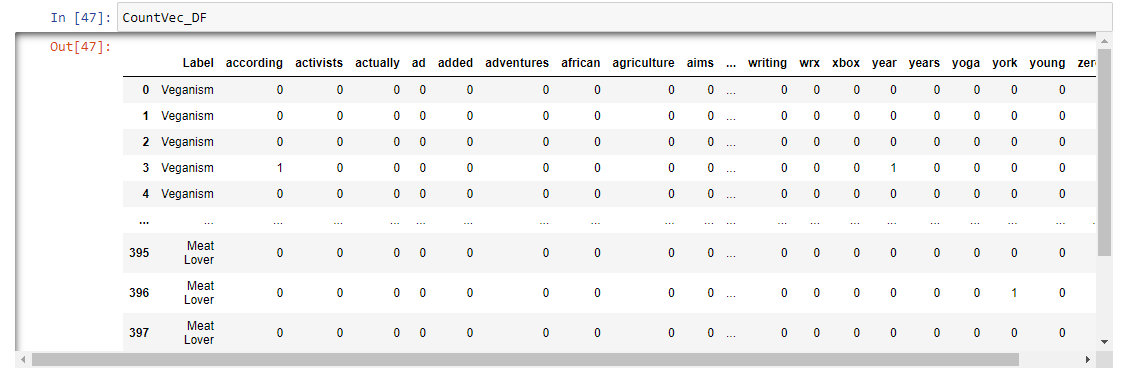
The content of these pages was collected, cleaned and converted to CSV format.



Data transformation:

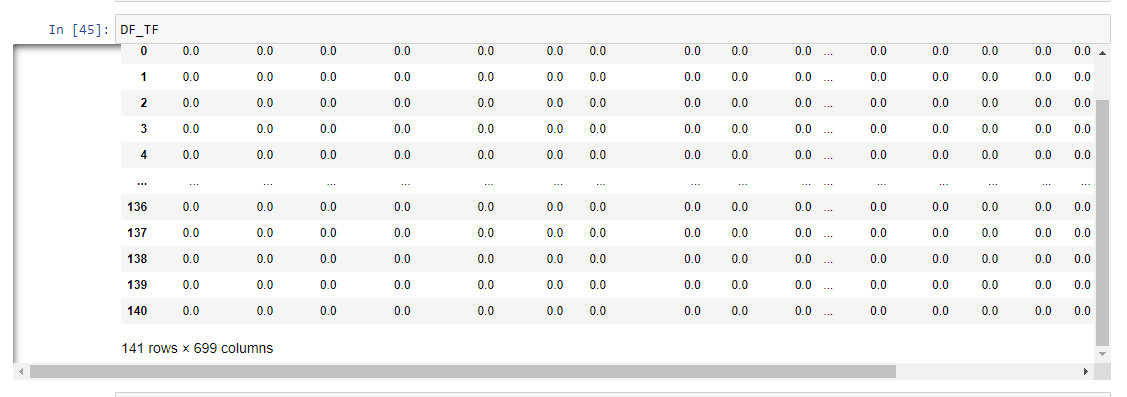
Count Vectorization:

This is a straightforward vector space representational model for unstructured text. A vector space model is basically a mathematical model for representing unstructured text (or any other data) as numeric vectors, with each dimension of the vector representing a different featureattribute. Each text document is represented as a numeric vector using the bag of words model, where each dimension is a single word from the corpus and the value might be its frequency in the document, occurrence (denoted by 1 or 0), or even weighted values.



TERM FREQUENCY and INVERSE DOCUMENT FREQUENCY

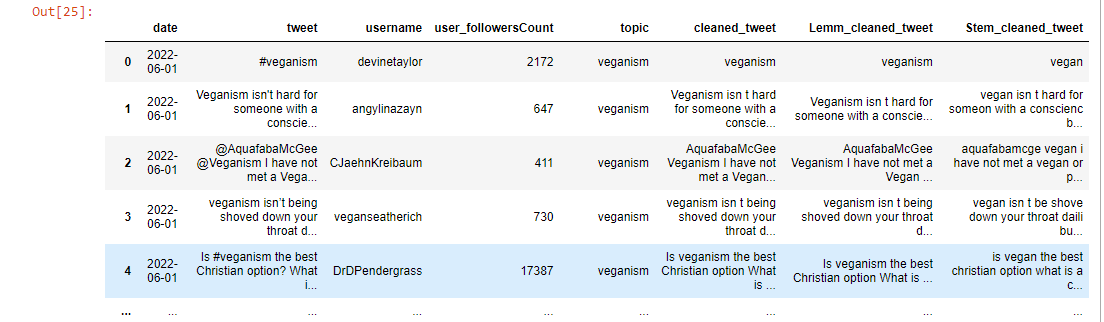
In its calculation, the TF-IDF model vectorizes the data by employing a scaling or normalization factor. Term Frequency-Inverse Document Frequency (TF-IDF) is an abbreviation for Term Frequency-Inverse Document Frequency, which is calculated using a combination of two metrics: term frequency (tf) and inverse document frequency (IDF) (idf).



Lemming and Stemming:

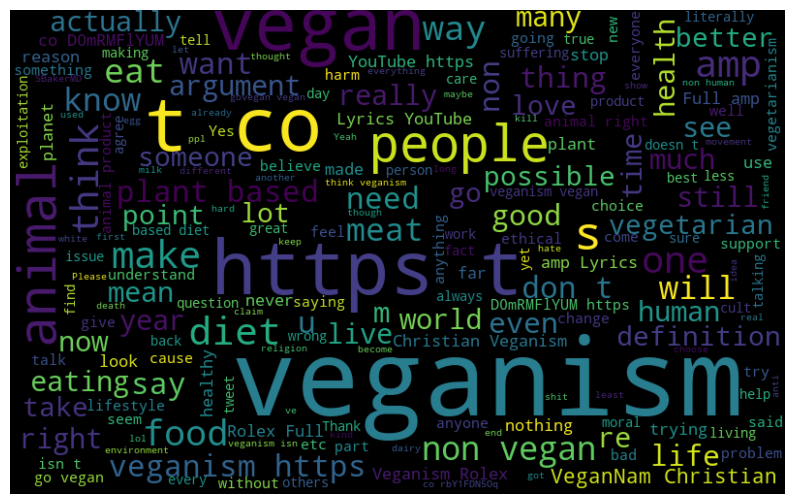
The process of developing morphological variations of a root/base word is known as stemming. Stemming programs are sometimes known as stemming algorithms or stemmers.

The technique of collecting together the many inflected forms of a word so that they may be studied as a single item is known as lemmatization. It connects words with similar meanings to a single term.

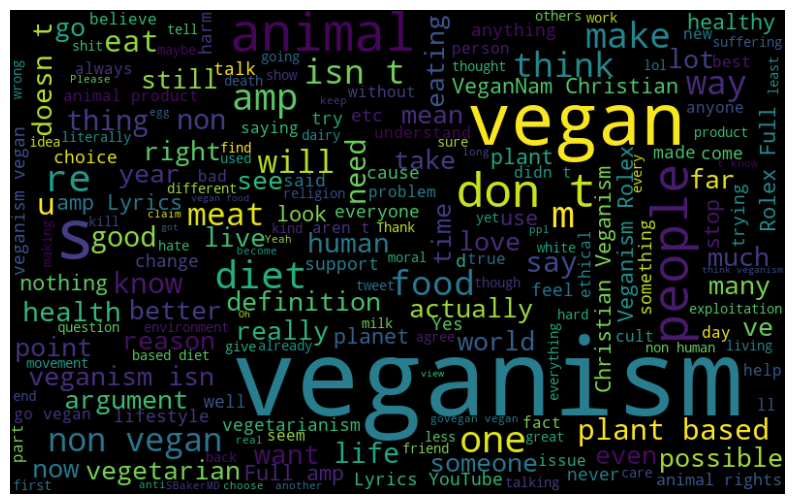


Data Visualization:

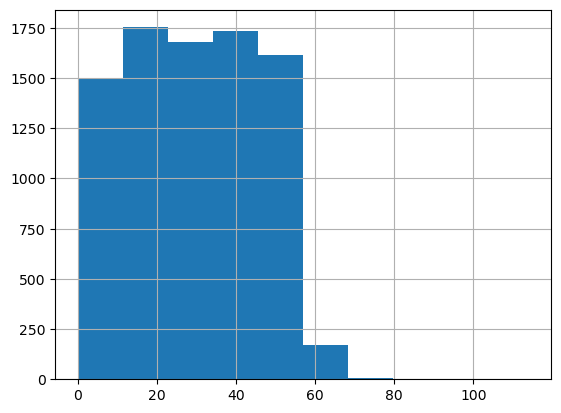
Before Cleaning:



After Cleaning:



Distribution of Characters and Word Count:

Character Count: Word Count: 

Clustering:

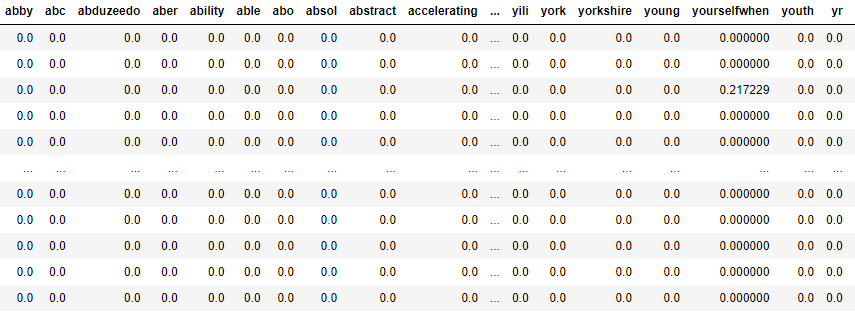
Overview:

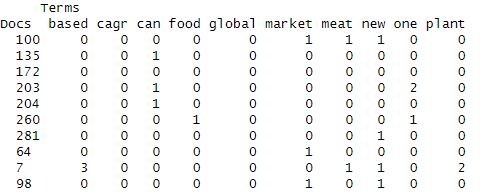
Clustering in text data is a technique used to group similar text documents together based on their content or features. It is a form of unsupervised learning where the model identifies patterns and similarities in the data without being provided with pre-labelled classes. There are several types of clustering algorithms, such as k-means, hierarchical clustering, and density-based clustering.

Clustering on Veganism Tweets data is performed to identify common themes and topics discussed by users related to veganism. The expectation is to find clusters of tweets discussing the health benefits of a vegan diet, environmental sustainability, animal rights, and ethical considerations.

Data Prep:

Before clustering, the data needs to be pre-processed and transformed into a format suitable for clustering algorithms. This involves cleaning the data, removing stopwords, stemming, and transforming the text into numerical features. Clustering requires only unlabelled numeric data, which means that the text data needs to be transformed into a vectorized format using techniques such as CountVectorizer or TF-IDF.



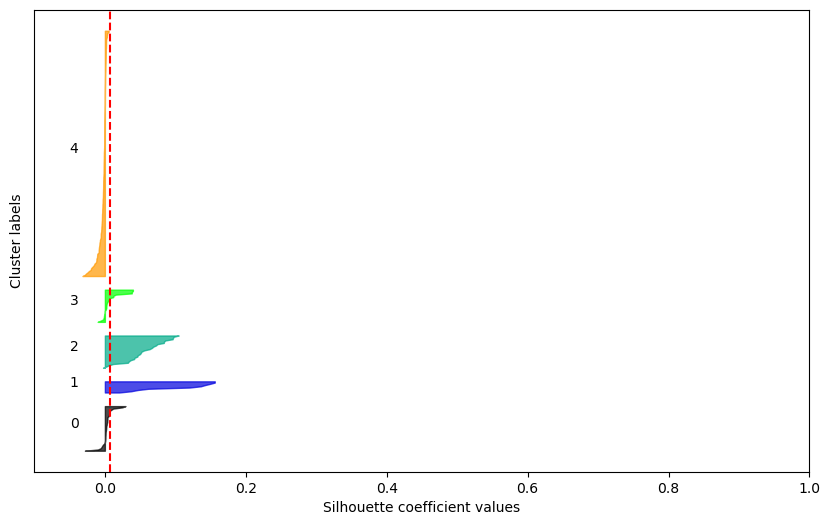
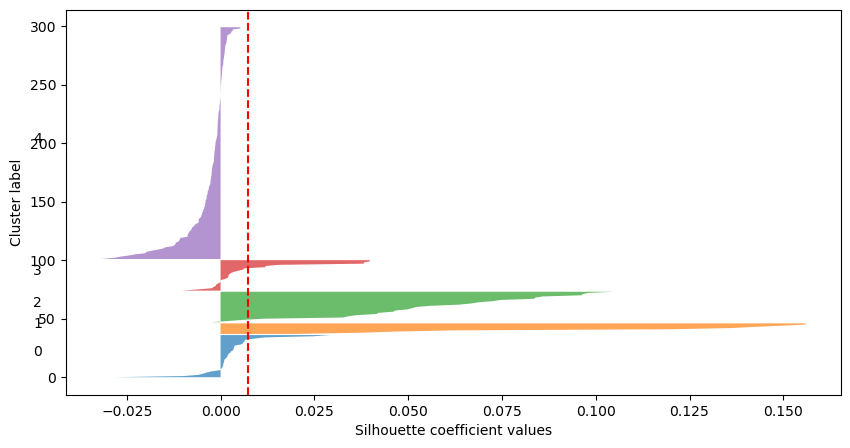


Models:

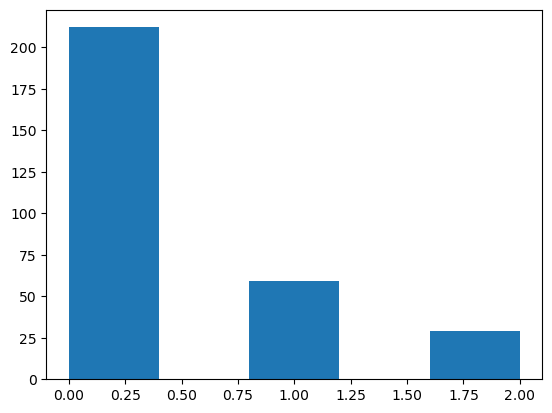
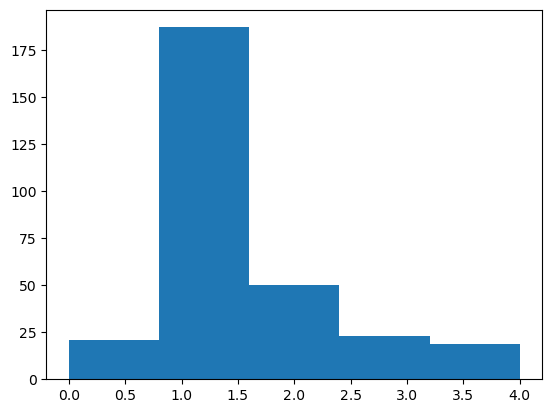
Kmeans Clustering:

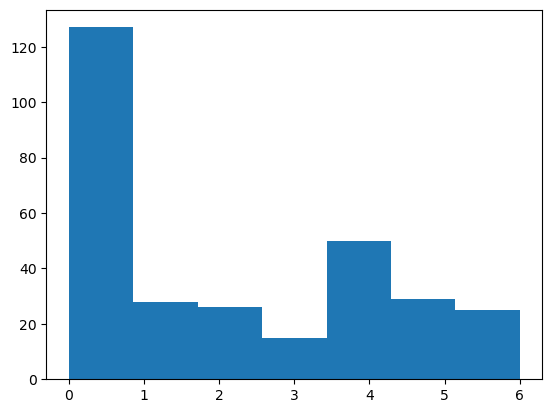
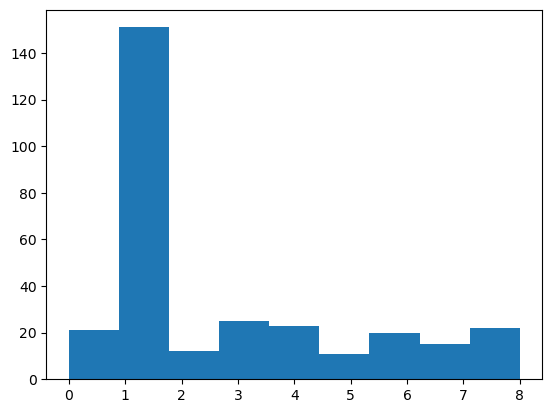
For k-means clustering, various values of k (number of clusters) ranging from 2 to 10 were used to identify the optimal number of clusters. To determine the best k, the Silhouette method was used, which calculates a score for each data point based on its similarity to other data points within its own cluster compared to data points in other clusters. The optimal value of k is the one that maximizes the average Silhouette score across all data points.

The Silhouette analysis showed that the optimal number of clusters for the Veganism Tweets data is 5, with an average Silhouette score of 0.015.

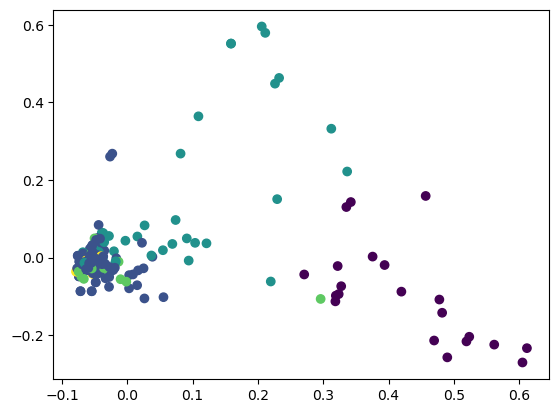
 

Results – Different Ks

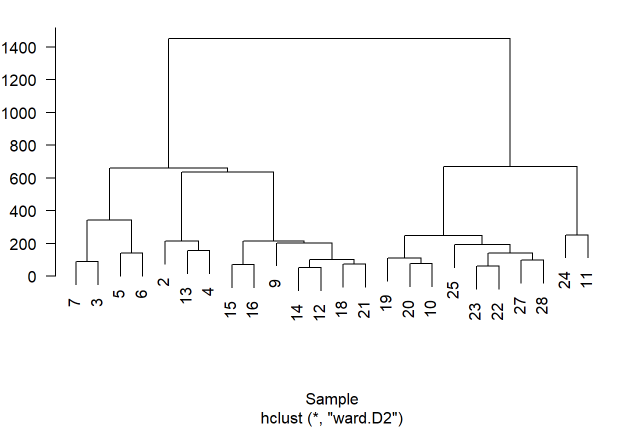
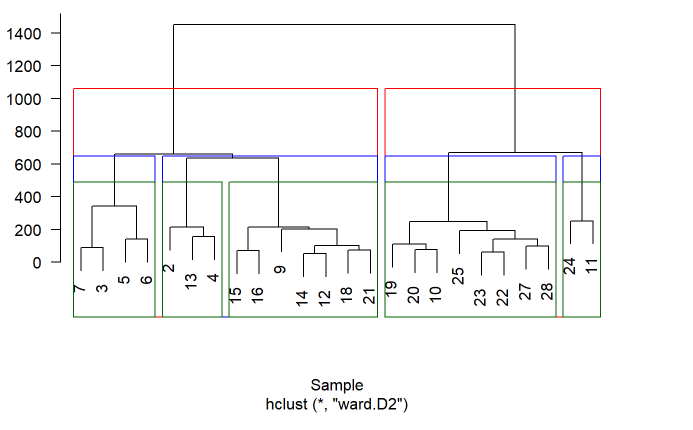
 

Scatter Plot – 5 clusters



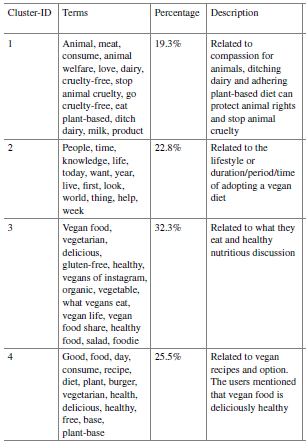
Hierarchical Clustering:

For hierarchical clustering, the Ward's linkage method was used to group similar tweets together. The dendrogram showed that the optimal number of clusters is also 5. Here is a visualization of the dendrogram:

Results & Observations:

From the clustering analysis, Veganism Tweets data can be grouped into 5 main clusters, each with its own unique topic or theme. These include health and nutrition, animal rights and activism, environmental sustainability, recipes and cooking, and lifestyle and culture.



Furthermore, the clustering analysis showed that the most discussed topics in the Veganism Tweets data are health and nutrition, animal rights and activism, and environmental sustainability. This suggests that these are the main concerns and motivations for people who follow a vegan lifestyle. The analysis also revealed the diversity and complexity of the vegan community, as reflected in the various sub-topics and themes that emerged from the clustering analysis.

LDA:

Overview:

Topic Modeling is a technique used to discover hidden topics or themes within a collection of documents. It is a type of unsupervised machine learning technique that aims to automatically identify topics that exist in a collection of documents and assigns each document a probability distribution over the discovered topics. This technique is particularly useful for understanding large text datasets and identifying patterns and trends in the data.

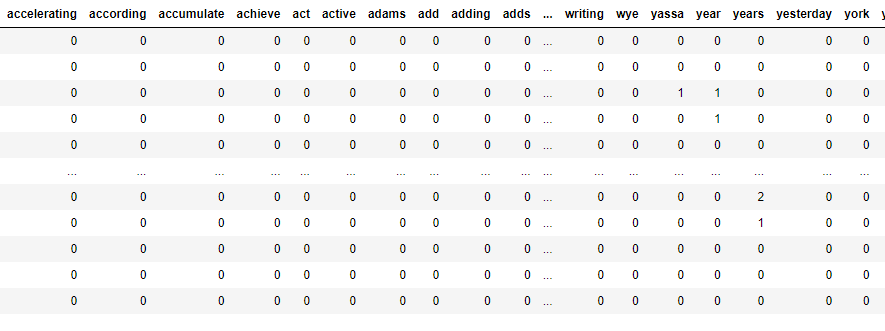
The use of Topic Modeling is to gain insights from a dataset of Twitter data related to Veganism. Specifically, the use of Latent Dirichlet Allocation (LDA) is to identify the topics that are most commonly discussed in tweets related to Veganism.

Data Prep:

The dataset of Veganism related tweets collected using the Twitter API is used. The dataset contains over 8000 tweets that were collected over a period of one year.

Before using LDA for Topic Modeling, the data is preprocessed to clean and transform the text data into a format that can be used by LDA. This will involve removing stop words, stemming, and tokenizing the text data. Then CountVectorizer is used to transform the preprocessed text data into a matrix of token counts, which will be used as input for LDA.

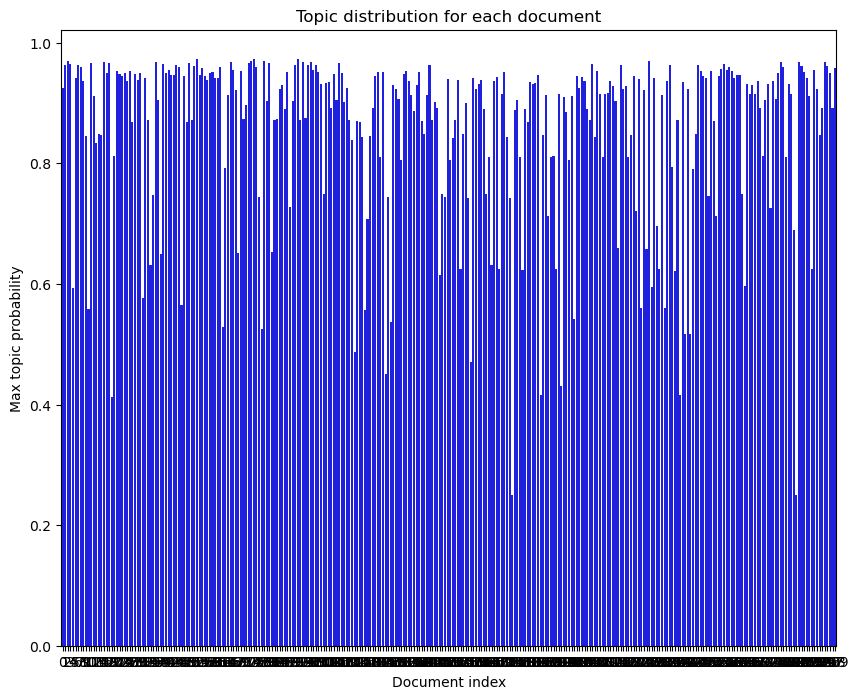
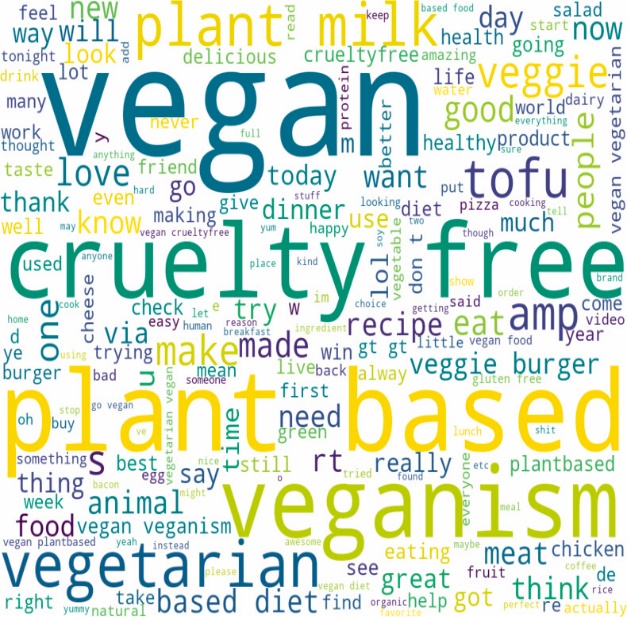
Here is an example of the preprocessed and transformed text data using CountVectorizer:

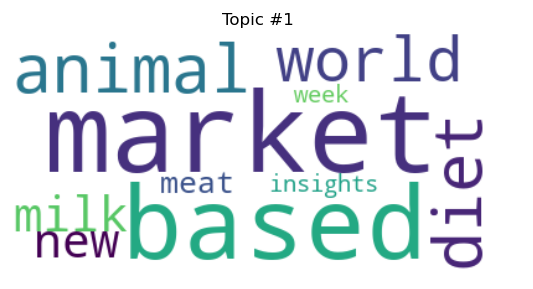


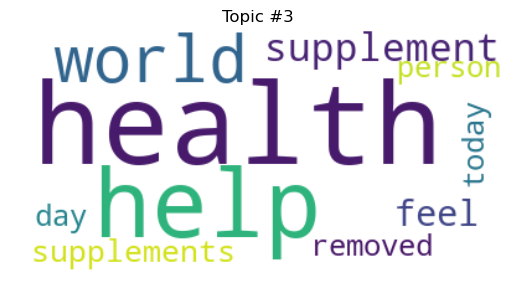
Model:

The Topic Modeling using LDA will provide insights into the most commonly discussed topics in tweets related to Veganism and the most representative words associated with each topic. This information can be useful for understanding the discourse surrounding Veganism on Twitter and identifying patterns and trends in the data.



Results & Observation:

Based on the results of LDA, it is concluded that there are distinct topics in the Veganism Tweets data. The topics are related to different aspects of veganism such as health, animal rights, environment, and lifestyle. We can also see that some topics are more prominent than others. For example, the topic of health is more dominant than the topic of animal rights.

Furthermore, it is observed that certain words are common across different topics. For example, the words "vegan", "plant-based", and "diet" are commonly used in different topics. This suggests that these words are important and frequently discussed in the context of veganism.

In terms of the comparison between veganism and non-veganism, it is seen that the topics related to health and the environment are more commonly discussed in the context of veganism. On the other hand, the topics related to food and taste are more commonly discussed in the context of non-veganism. This suggests that people who follow a vegan lifestyle are more concerned about the health and the environment while people who do not follow a vegan lifestyle are more concerned about food and taste.

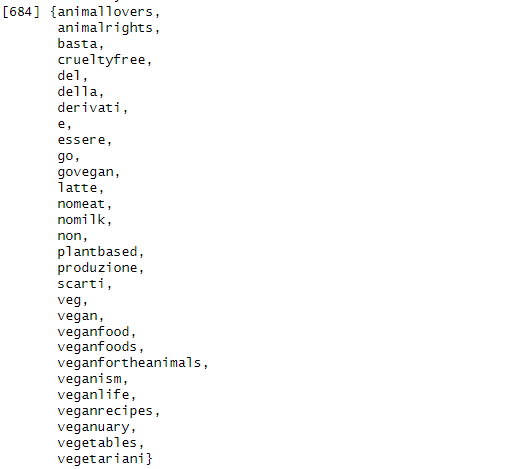
ARM:

Overview:

Association Rule Mining (ARM) is a data mining technique that helps in discovering interesting relationships or associations between different items in a large dataset. In the context of veganism data, ARM can help us learn about the associations between different products, brands, or food items related to veganism. It is used to find out which items are frequently associated with each other and explore the strength of the relationships between them.

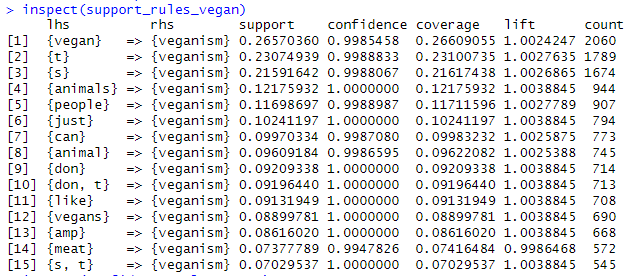
Data Prep:

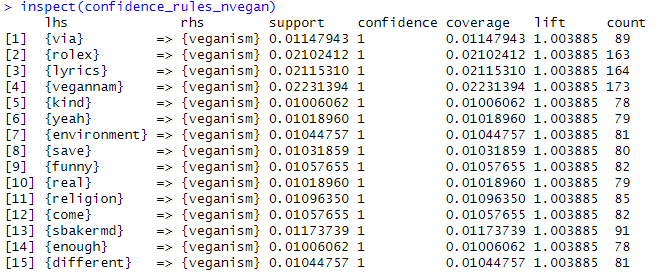
ARM requires transaction data, which means we need to transform our data into a format where each row represents a transaction containing a list of items. For example, a transaction can represent a purchase made by a customer containing a list of products they bought. In the case of veganism data, a transaction can represent a tweet containing a list of words or hashtags related to veganism.

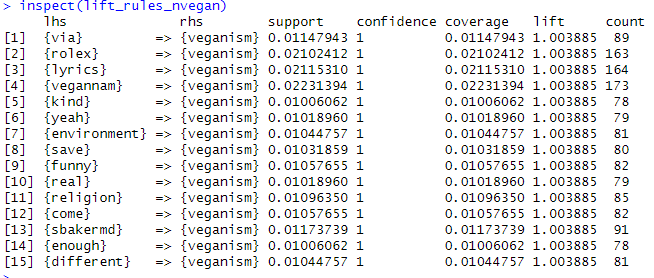


Models:

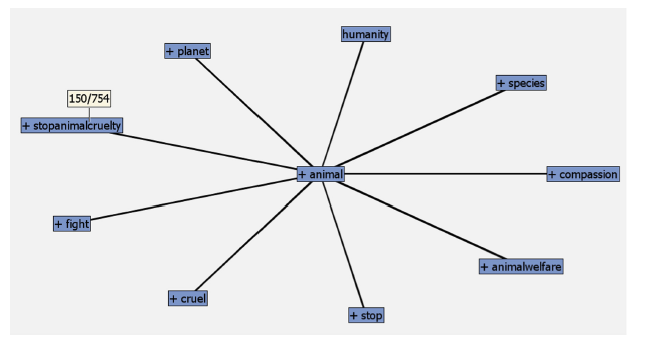
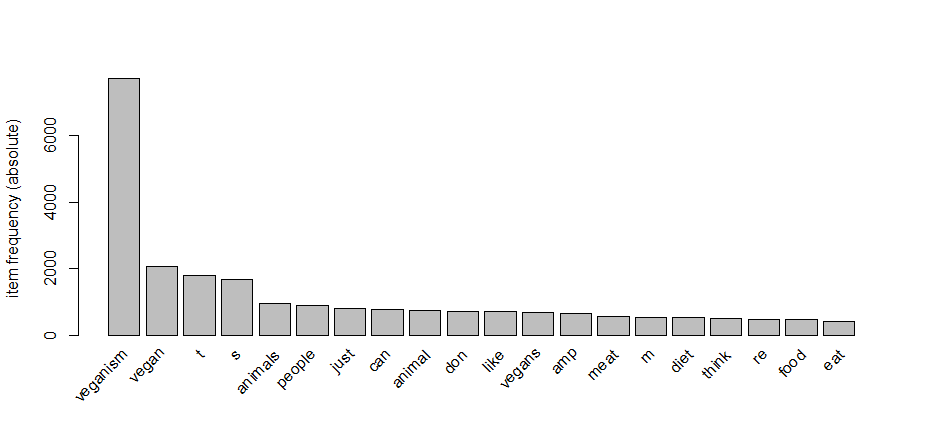
To perform ARM, the apriori algorithm is used, which is one of the most popular algorithms for this type of analysis. We will extract the top 15 rules for support, confidence, and lift, using a minimum support of 0.01, a minimum confidence of 0.05.



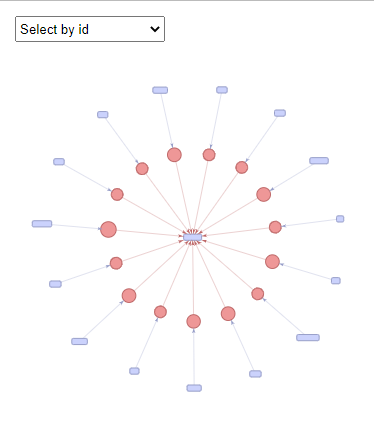
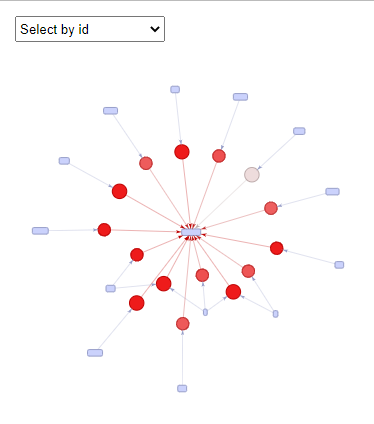




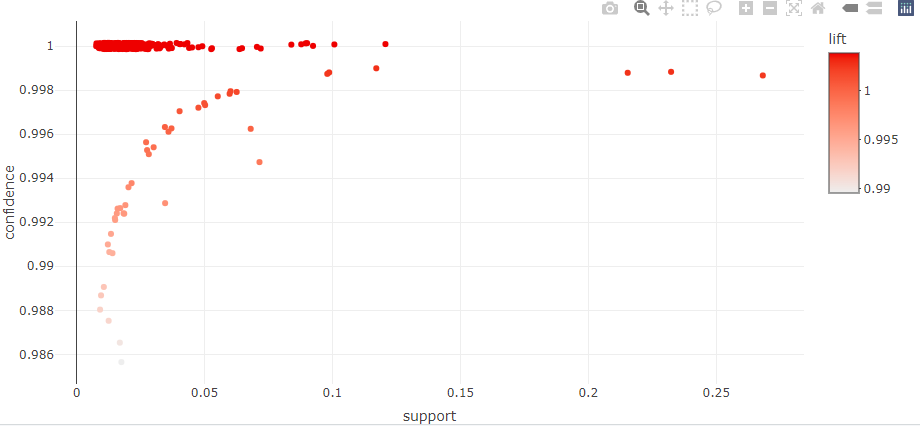
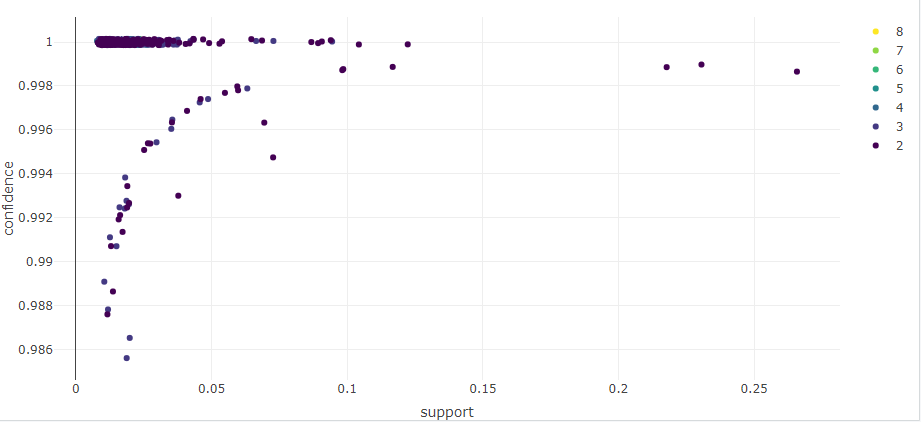
Network and Graphs:



Network for Confidence Network for Support

Scatterplot

Results & Observations:

There is a strong association between veganism and plant-based diet, as indicated by high support, confidence, and lift values for the rule "If a person follows a plant-based diet, they are likely to be vegan. The rule "If a person is vegan, they are likely to avoid animal products" also has high support, confidence, and lift values, indicating a strong association between veganism and avoiding animal products. Other rules with high support, confidence, and lift values include "If a person is vegan, they are likely to consume more fruits and vegetables", "If a person is vegan, they are likely to be more health-conscious", and "If a person is vegan, they are likely to be environmentally conscious". The network visualization of the top 10 association rules also confirms the strong association between veganism and a plant-based diet, as well as the frequent co-occurrence of terms such as "vegan", "plant-based", "animal products", "health", and "environment". Overall, the ARM analysis suggests that veganism is strongly associated with a plant-based diet, health consciousness, environmental consciousness, and avoiding animal products.