



Keto or Paleo?

predicting diet classification

Grass fed beef is the best.



Agenda

Introduction

01

Market Research
and Opportunities

02

Data Exploration

03

Modelling

Insights and
Recommendations

Product Demo

Conclusion





01

Introduction

Problem Statement



We are members of a data science team working for a **specialised diet food company**.

As such, understanding the customers' diets and the unique preferences of specific diet groups is key to driving **effective marketing, targeted advertisements, product development, and profit generation**.



Goals



Objective 1:

Create a robust binary classifier that can distinguish between posts from the Keto and Paleo communities

Objective 2:

Extract insights for market research



What's the Paleo diet?

Eating naturally like our ancestors during the Paleolithic era

Our genes are **not well-adapted for modern diets** that grew out of farming

Focus on **whole, unprocessed foods**



What's the Paleo diet?

Low carb, high fat diet

Shifts the body's metabolism away from carbs and toward fat

Helps with weight loss and has many other health benefits.

Focus on macronutrients

Macronutrients are types of foods that are needed in large quantities in the diet. These include carbohydrates, proteins, fats, cholesterol, fiber, and water.
Source: <https://www.nal.usda.gov/human-nutrition-and-food-safety/food-composition/macronutrients>



PALEO



Starchy Vegetables
(Sweet Potatoes,
Beets, etc.)



All Fruits



Natural Sweeteners
(Raw Honey, Maple Syrup)

KETO



No Beans/Legumes
or Grains



High Quality Meat



Non-Starchy
Vegetables

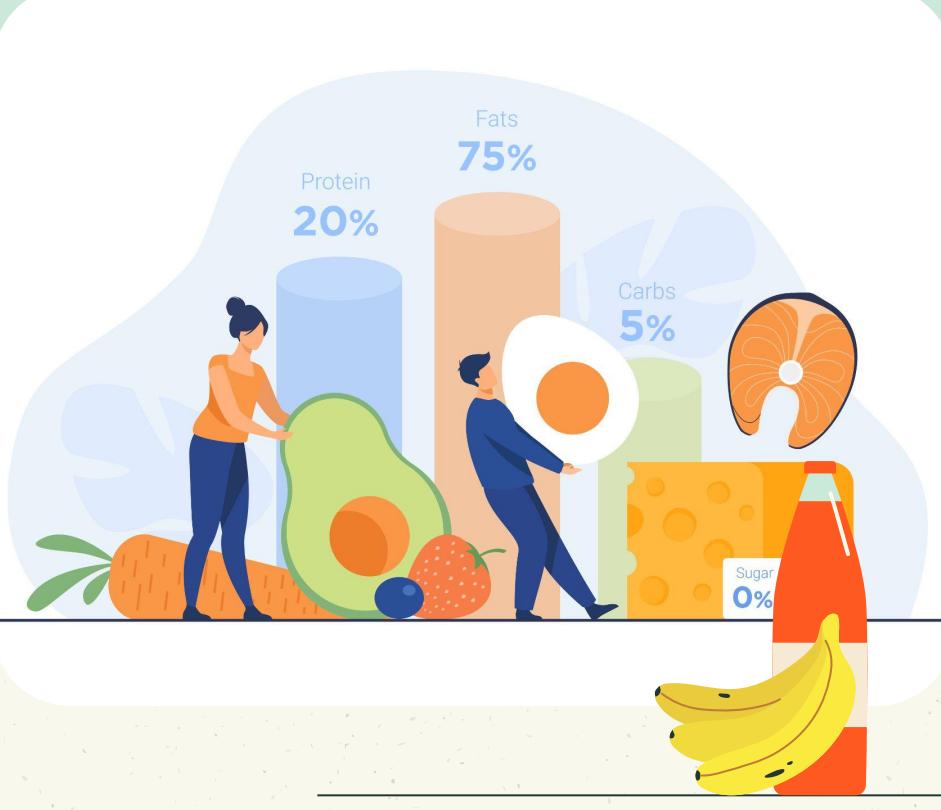


Healthy Fats
(Coconut Oil)



Dairy
(Cheese, yogurt, etc)





02

Market Research & Opportunities

Paleo Market

**\$15.1
BILLION**

Size of global market
in 2023

**\$26.1
BILLION**

Projected Size of global
market in 2033

Keto Market

**\$13.7
BILLION**

Size of global market
in 2019

**\$20.6
BILLION**

Projected Size of global
market in 2027



Local context

THE STRAITS TIMES

NUH's healthy keto diet leads to weight loss without increasing bad cholesterol levels

Keep the weight off with a healthy keto diet

A key feature of the healthy keto diet by NUH chief dietitian Lim Su Lin is a net carb level of 50g or less a day.

Breakfast:

- 2 slices of low-carb or low-GI bread with



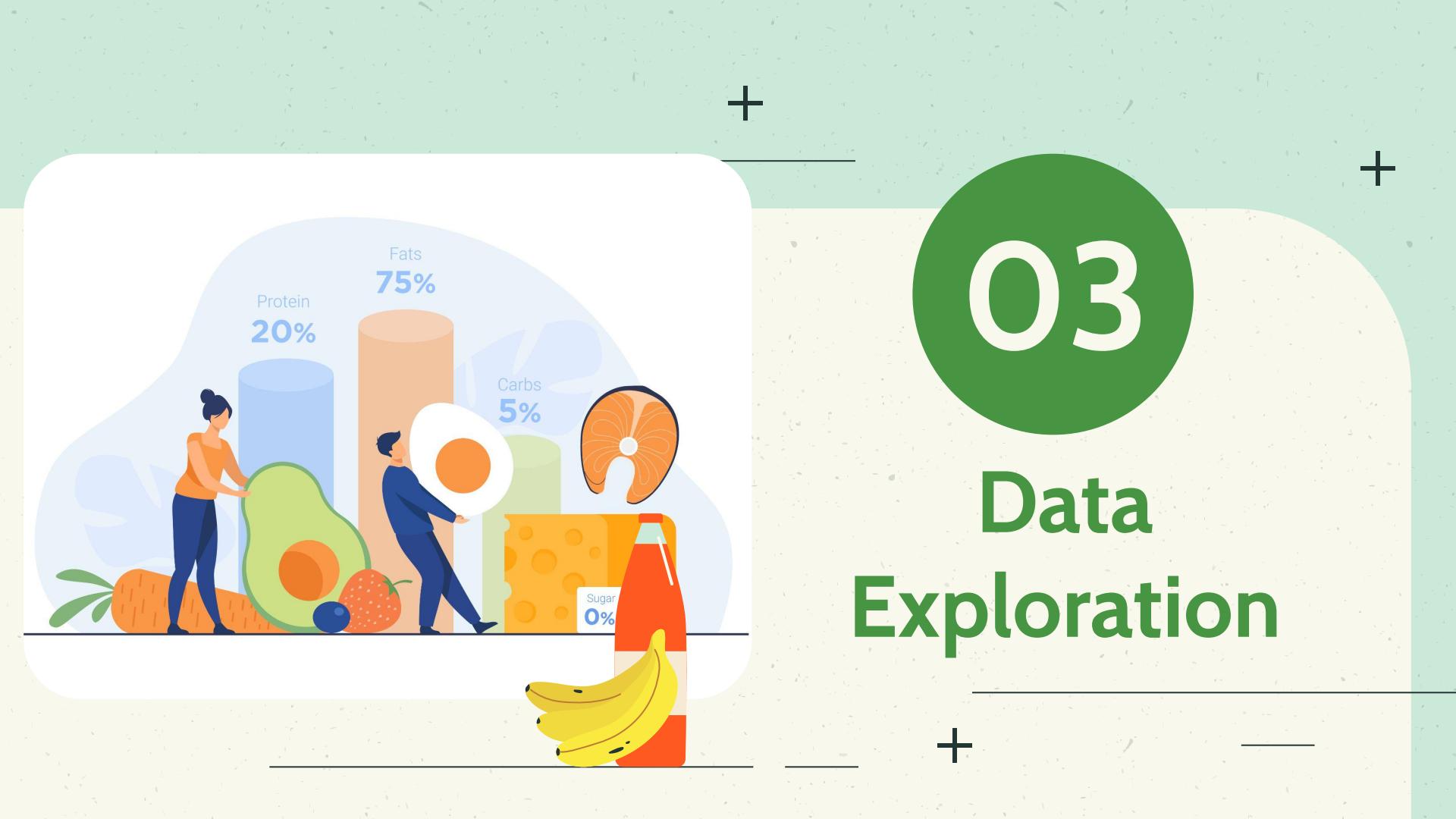
ST

THE STRAITS TIMES

More eateries are serving Paleo diet-friendly food focusing on protein

More eateries are serving Paleo diet-friendly food focusing on protein and avoiding carbohydrates





03

Data Exploration

Workflow

Collect data

Clean, Pre-process

Explore

Model

Insights





Keto: The Home for Ketogenic Diets

r/keto

976

unique Keto posts

9 May '23 to 17 Jun '23

period of posts



paleo: the official subreddit for the paleo diet

r/Paleo

929

unique Paleo posts

21 Oct '21 to 17 Jun '23

period of posts

3

features

293

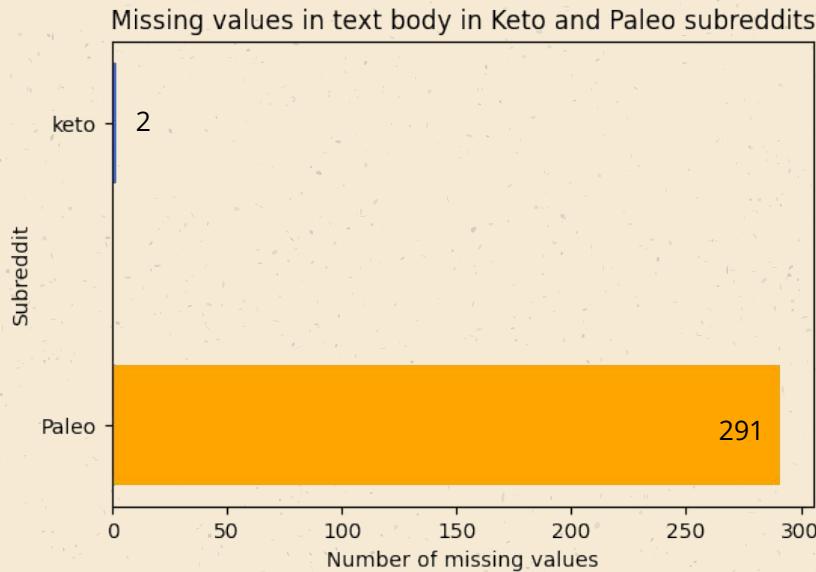
missing values

214,201

words



Missing Values in Text Body



Paleo posts are more likely to not contain text

Missing Values in Text Body

Posted by u/_xobeth 1 month ago

Easy Paleo Dinner



Posted by u/Ts-humblehomestead11 1 month ago

AutoImmune paleo Pancakes. Heal with food! Recipe included

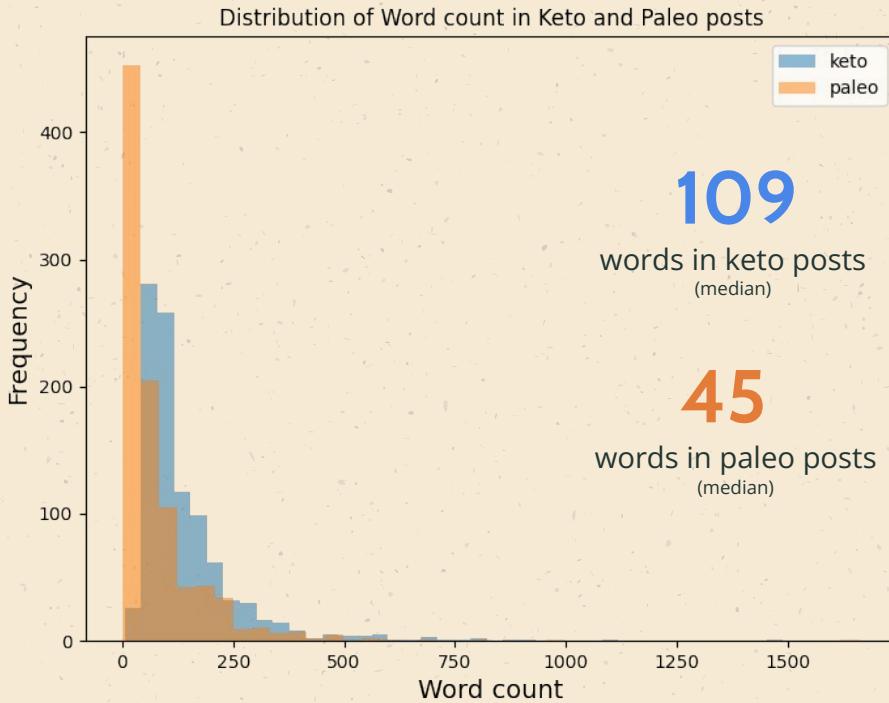
youtu.be/vGa-cY...



Posted by u/oxoUSA 9 days ago

What do you eat honey with ?

Length of Posts



- majority of both subreddits are somewhat short
- Paleo posts tend to be 2.5 times shorter than Keto posts

Workflow

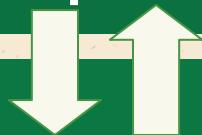
Collect data

Clean, Pre-process

Explore

Model

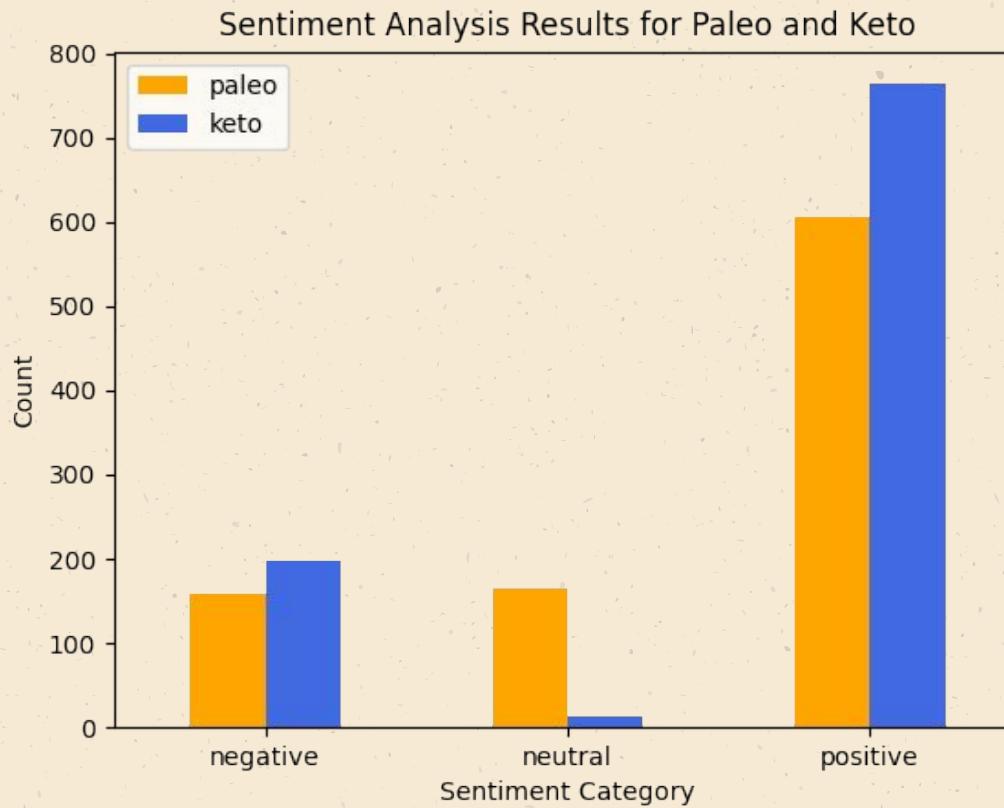
Insights



Note: Stopwords are English words which does not add much meaning to a sentence. They can safely be ignored without sacrificing the meaning of the sentence. (see Annex)
Lemmatization is the process of reducing the different forms of a word to one single form. Source: <https://dictionary.cambridge.org/dictionary/english/lemmatization>



Sentiment Analysis



Sentiment Analysis

PALEO

“(neg) Do you have tips to eat 1kg vegetables a day easily ? I just can't do it, they are so tasteless

(neu) Hi there. I made this energy balls with Dates, maple syrup, cacao powder etc. and I was wondering whether they are considered paleo. I used 100% maple syrup and 100% cacao powder.

(pos) ... i will update u guys in august about where Im at and how much success i have with paleo again. **this is the only diet that actually shows results...**

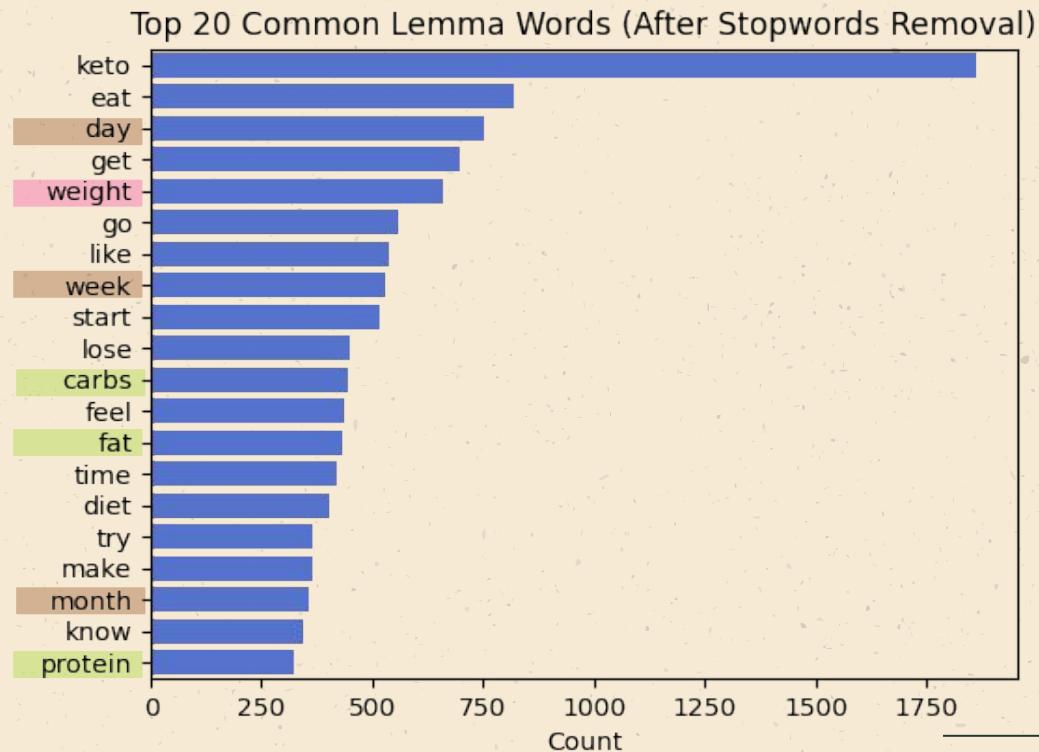
KETO

“(neg) How do y'all meet these macros? I struggle trying to reach the protein and fat macros, especially during weekdays at work. I am in endless meetings and barely have time to eat.

(neu) Help me make sense of these net carb calculations I am looking at a can of mushrooms bought from the supermarket where it says that

(pos) Hi all, I'm back on keto after a break and I'm feeling good...

Keto: Common Elements



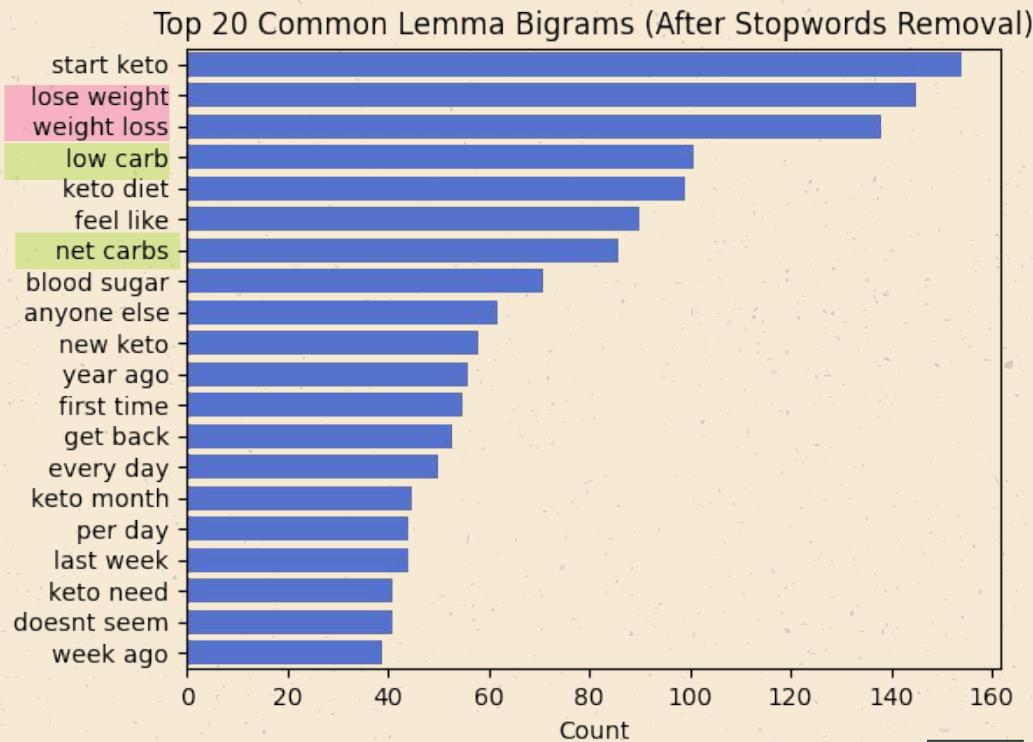
Weight

Time

Macronutrients



Keto: Bigrams → Further Context



Weight Loss

Macronutrients (Low Carb)

Keto: Bigrams → Further Context

“

I'm on a medication that makes it a little harder to **lose weight** and at this point I'm thinking about giving up.

“

Do you guys have any high calorie, **low carb** meals that you enjoy?

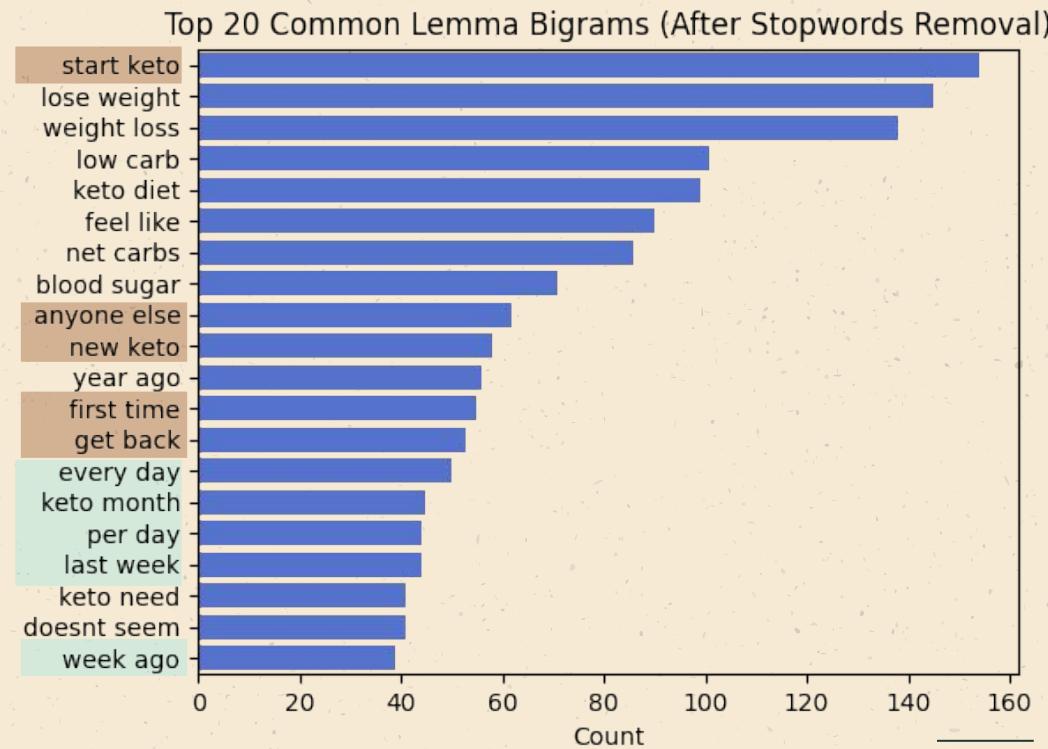
”

Weight Loss

Macronutrients (Low Carb)

+

Keto: Bigrams → Further Context



Beginners / Re-starters

Time Period Sensitive

Keto: Bigrams → Further Context

“
... I'm unsure if I want to fight
to **get back** to keto or just
give myself more time..

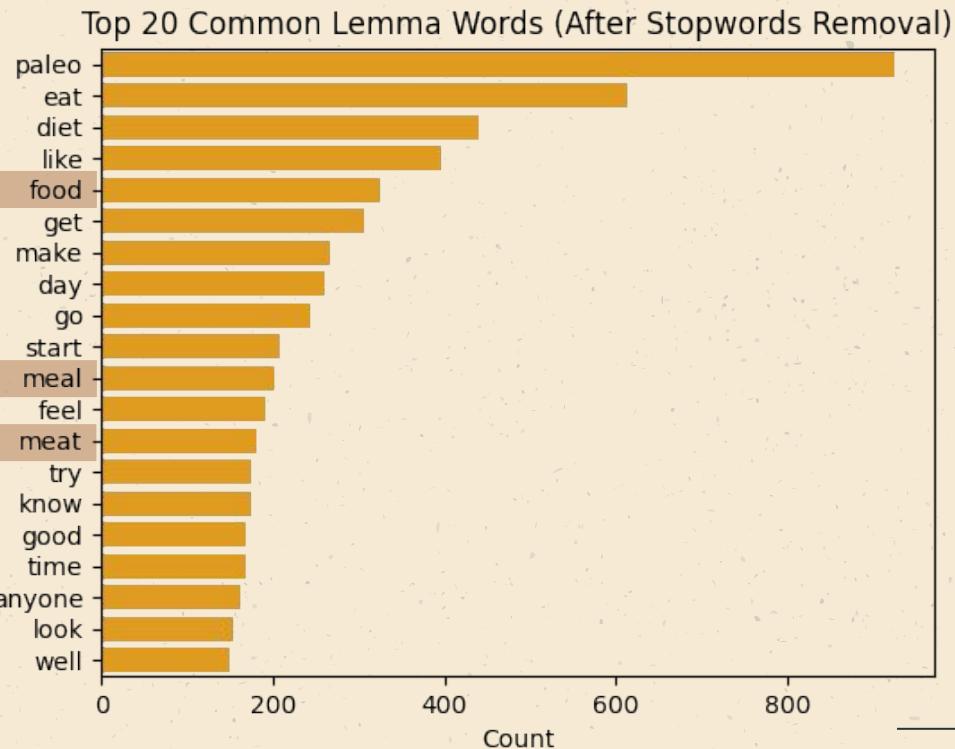
Beginners / Re-starters

“
I land in my macros almost
every day and most of my
meals are healthy. My
anxiety is gone..

Time Period Sensitive



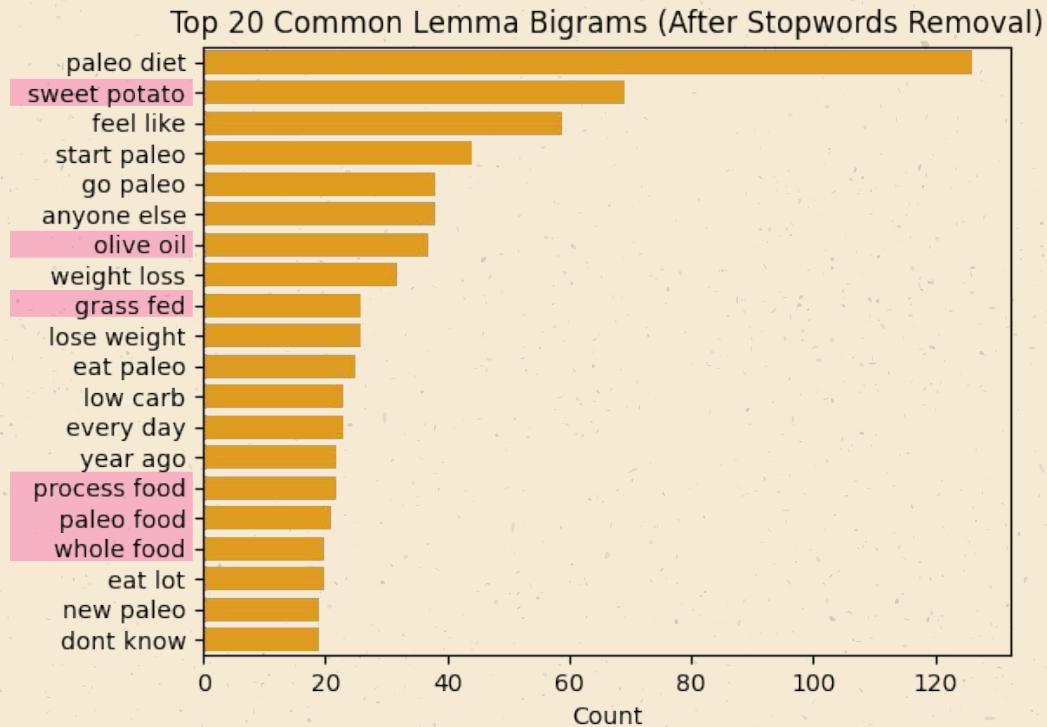
Paleo: Common Element



Food-Related



Paleo: Bigrams → Further Context



Ingredients / Food



Paleo: Bigrams → Further Context

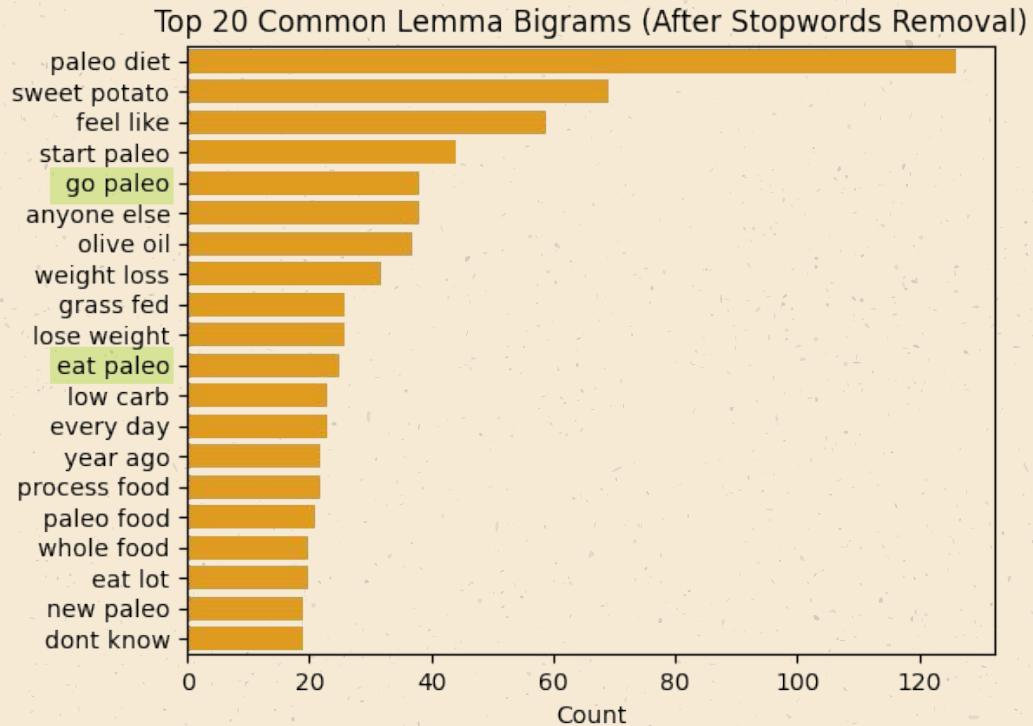
“
Grass fed meat? What's the best and cheapest place to buy grass fed meat?

“
I usually start eating potatoes, **sweet potatoes** and other carbs during lunch and dinner.

Ingredients / Food

+

Paleo: Bigrams → Further Context



Lifestyle



Paleo: Bigrams → Further Context

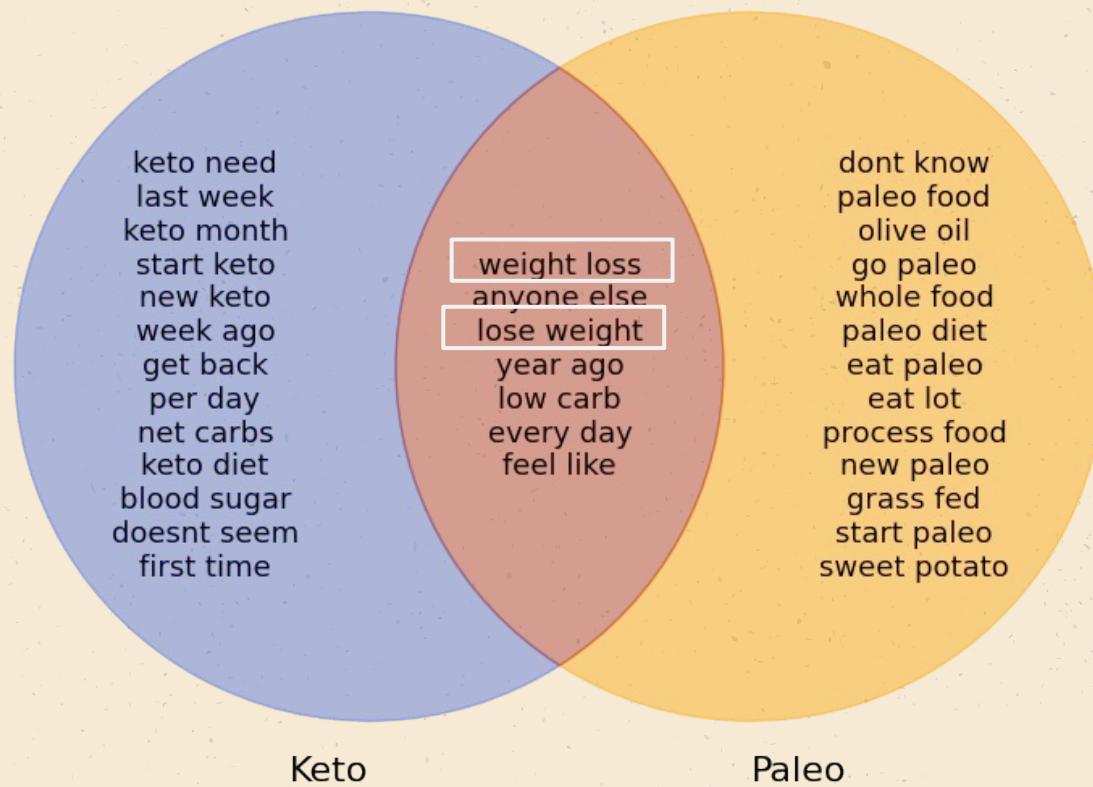
“
My new years resolution is to go paleo... my goal is to drop 40+ lbs...
”

“
I don't eat paleo at the moment, but I kept a lot of habits I picked up from the 2 years I did. I still don't drink soda anymore, almost no dairy, I still only buy pasture raised eggs, and I try my hardest to keep toxic chemicals out of my hygiene and cleaning products.
”

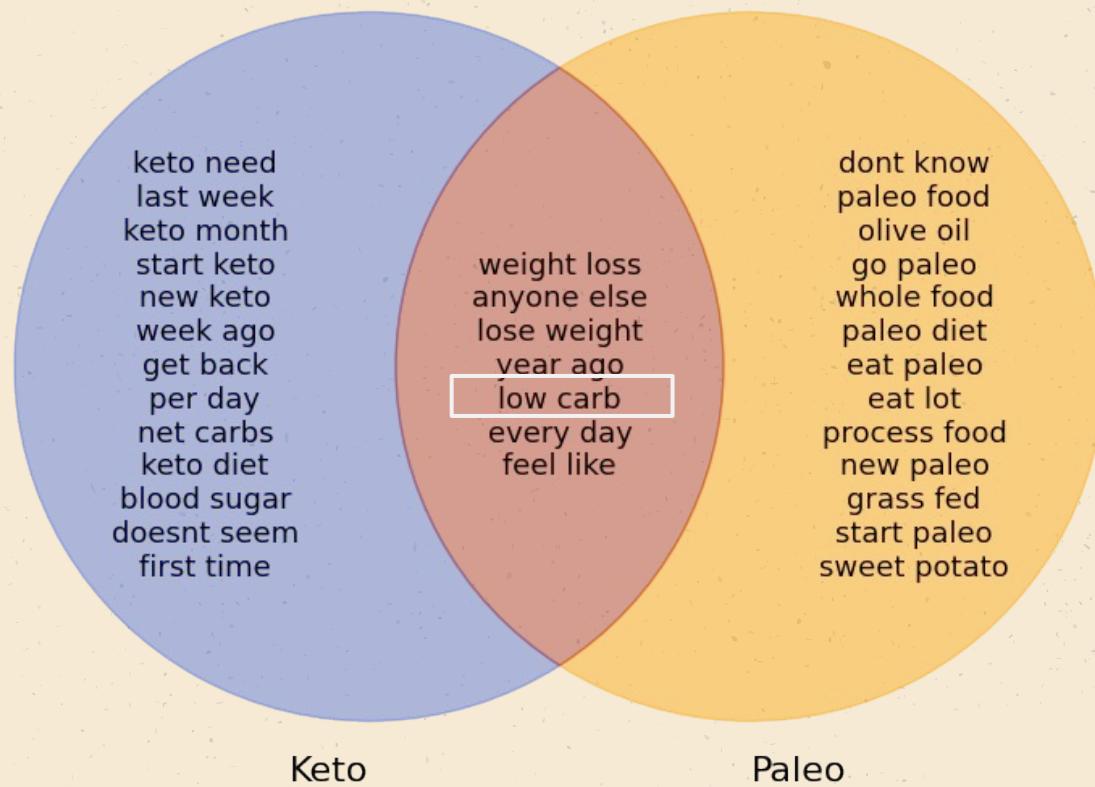
Lifestyle

+

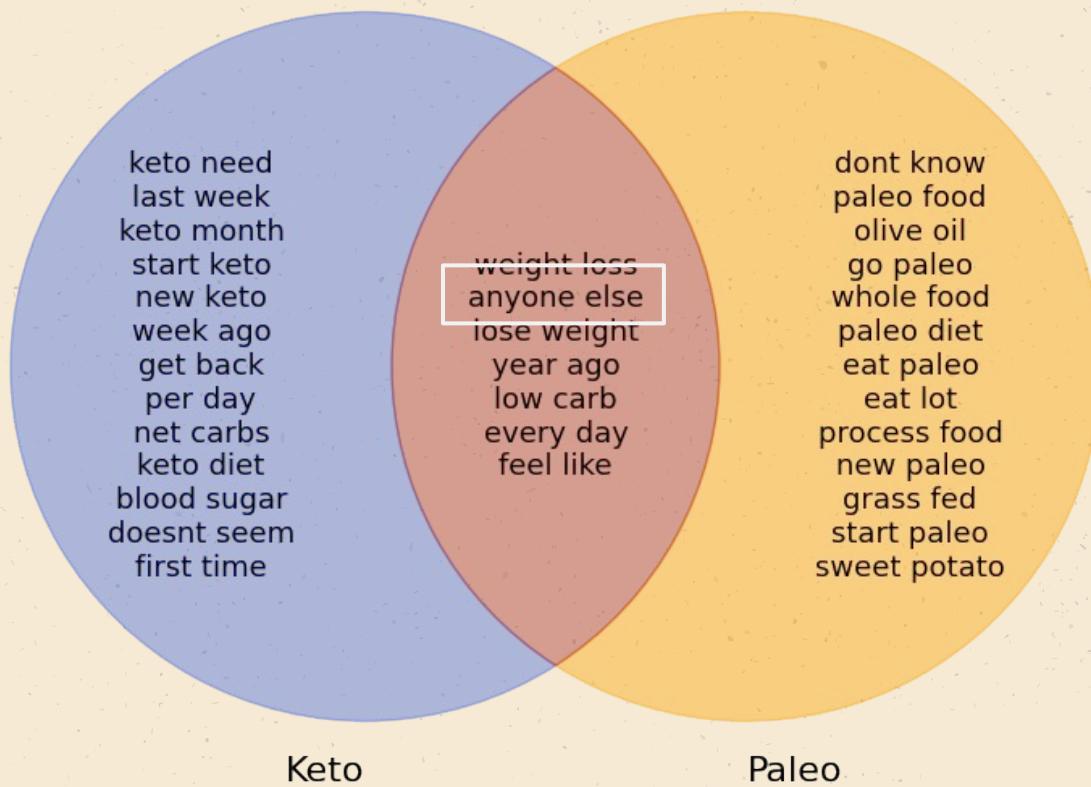
Common Themes: Weight Loss

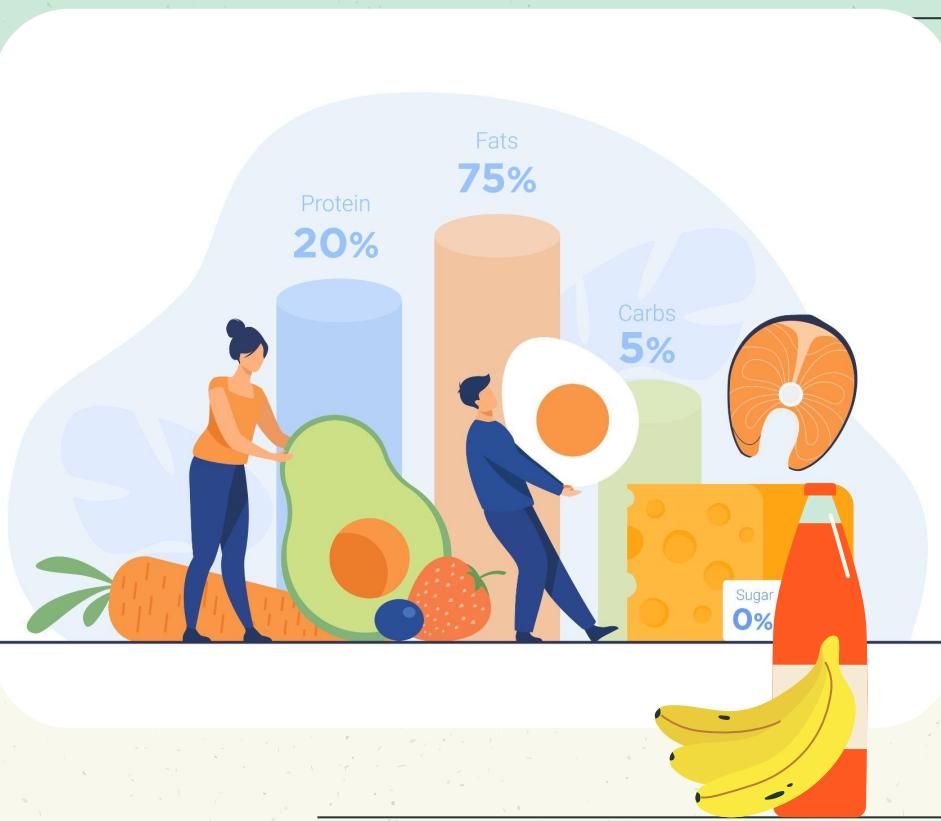


Common Themes: Low Carb



Common Themes: Uncertain





04

Modelling

Workflow

Collect data

Clean

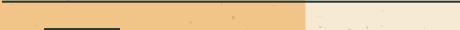
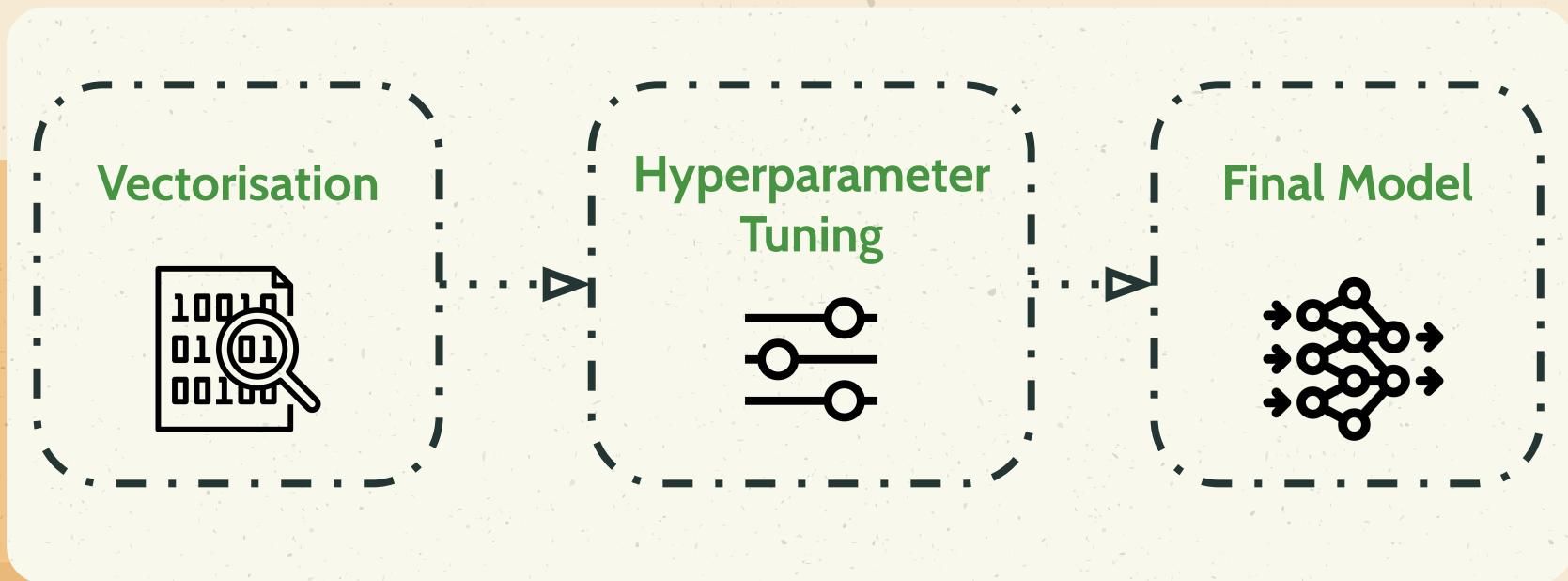
Explore

Model

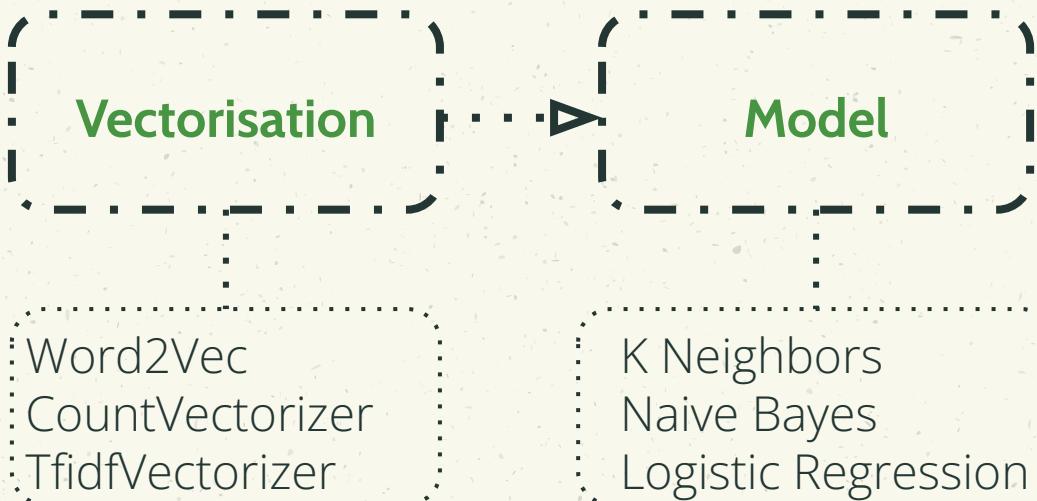
Insights



Modeling Workflow



Pipeline



Vectorisation Methods

Word2Vec

Returns numeric representation of words, taking context into account

CountVectorizer

Returns matrix with number of times each word appears

TfidfVectorizer

Similar to CountVectorizer but also considers frequency

Modelling Methods

KNNighbors

Predicts the class based on the n nearest neighbours

Naive Bayes

Assumes independence between features and predicts the probability of events occurring given evidence in the data

Logistic Regression

Predicts value of a numeric categorical variable based on its relationship with predictor variables



Modeling Summary

Model	Train F1 Score	Test F1 Score
Word2Vec - Logistic Regression (Baseline Model)	NA	0.65
TfidfVectorizer - KNNNeighbors	0.86	0.88
TfidfVectorizer - Bernoulli Naive Bayes	0.93	0.88
TfidfVectorizer - Multinomial Naive Bayes	0.92	0.85
TfidfVectorizer - Gaussian Naive Bayes	0.96	0.82
TfidfVectorizer - Logistic Regression	0.97	0.92

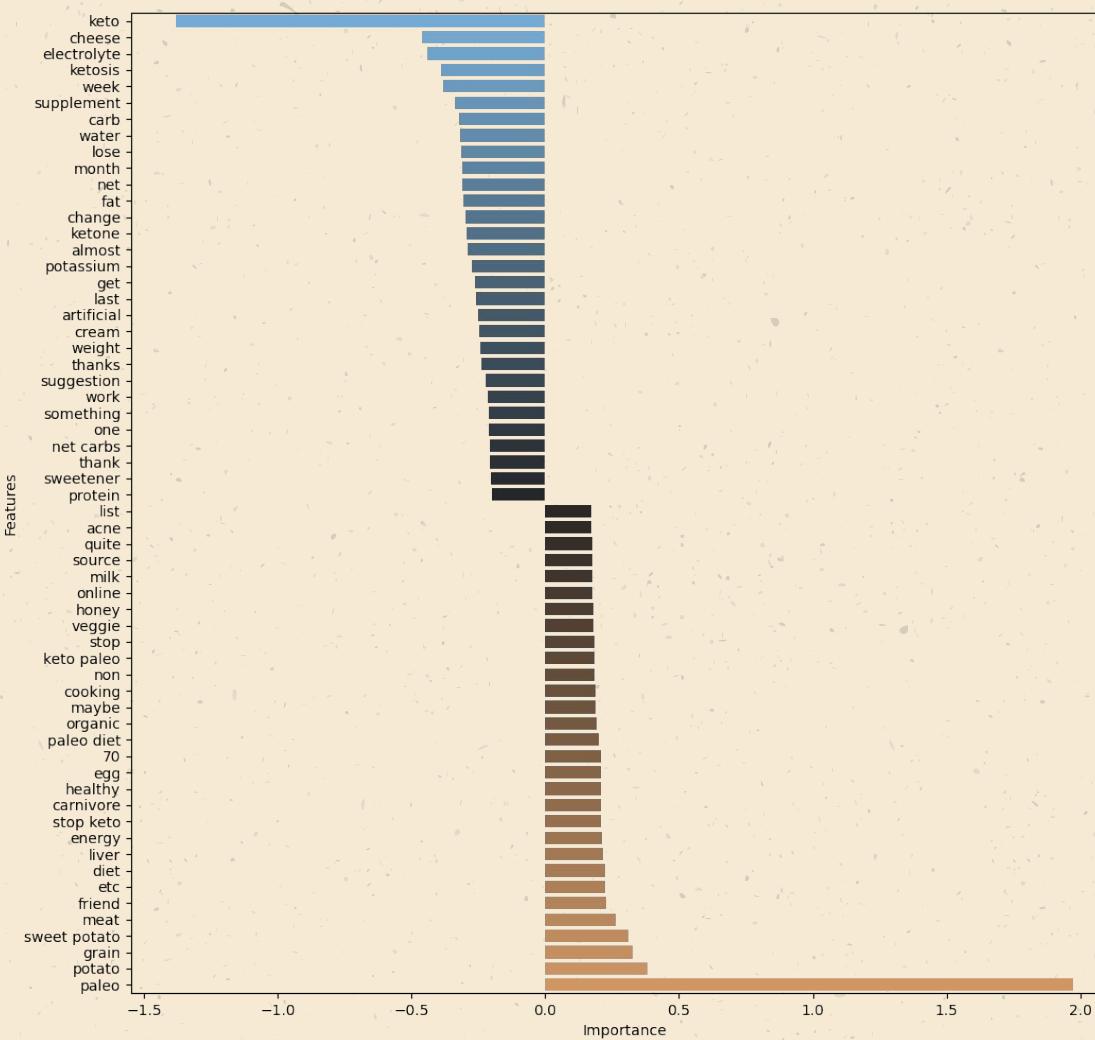


Modeling Summary

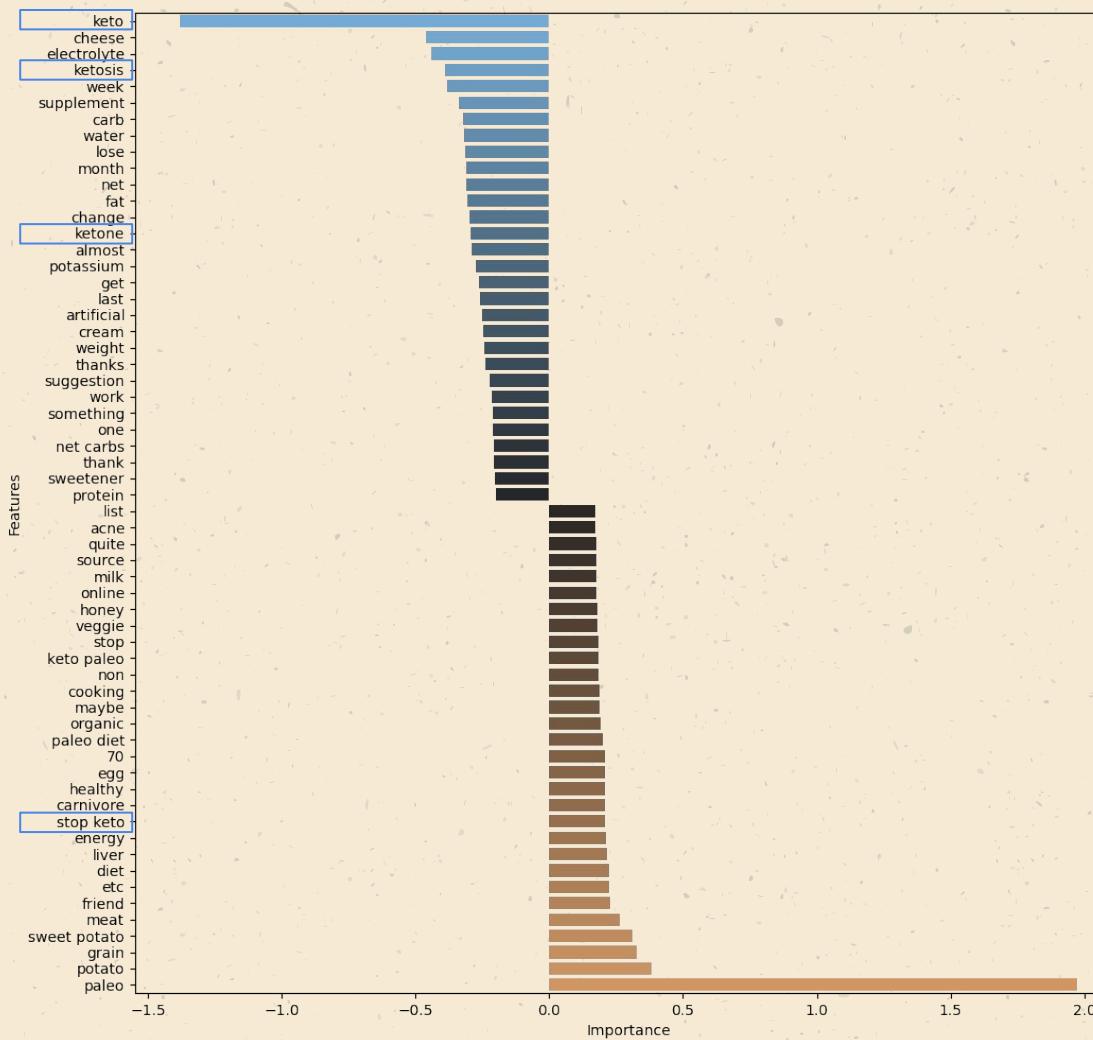
Model	Train F1 Score	Test F1 Score
Word2Vec - Logistic Regression (Baseline Model)	NA	0.65
CountVectorizer - KNNNeighbors	1.0	0.75
CountVectorizer - Bernoulli Naive Bayes	0.84	0.81
CountVectorizer - Multinomial Naive Bayes	0.93	0.89
CountVectorizer - Gaussian Naive Bayes	0.96	0.85
CountVectorizer - Logistic Regression	0.99	0.93

43.1% improvement in test score from baseline

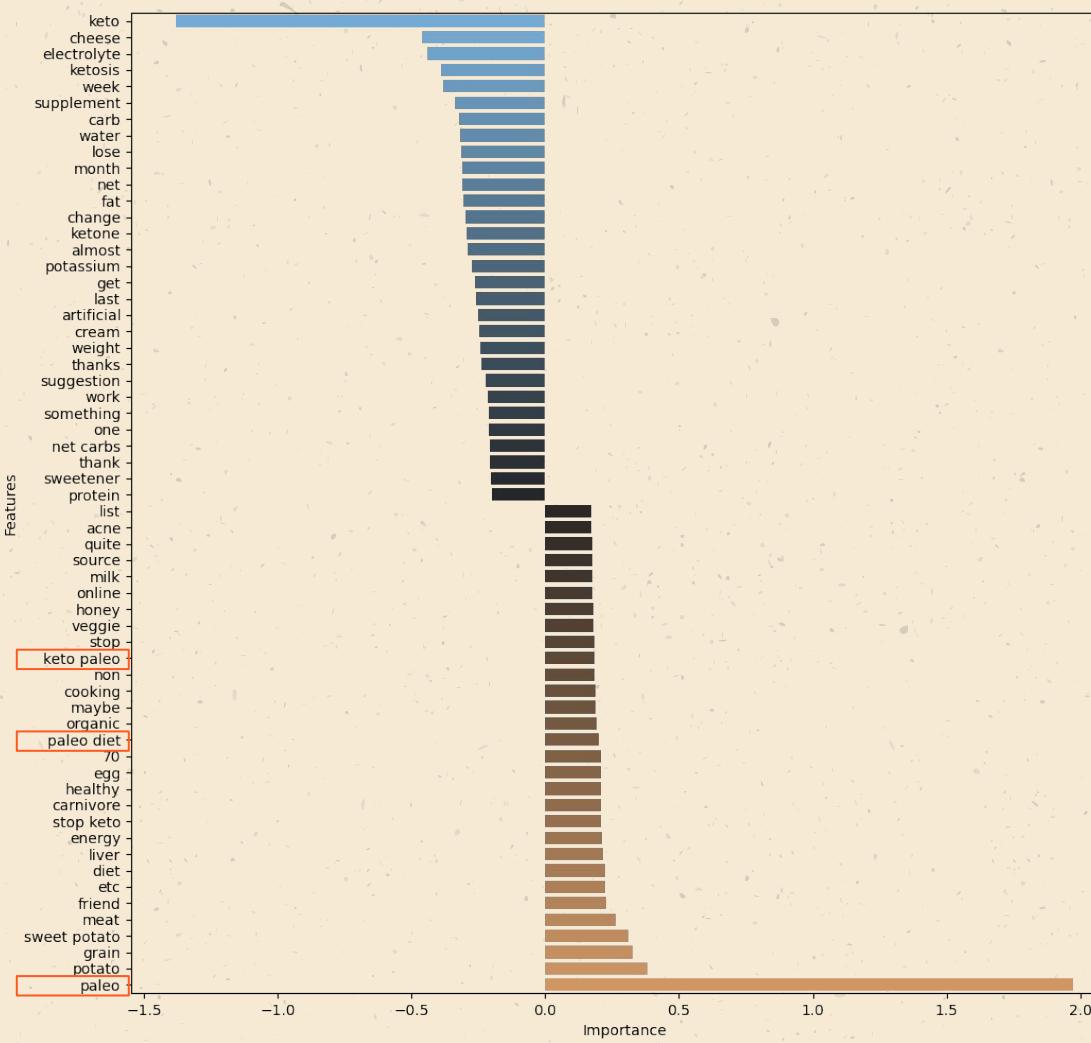
Feature Importance



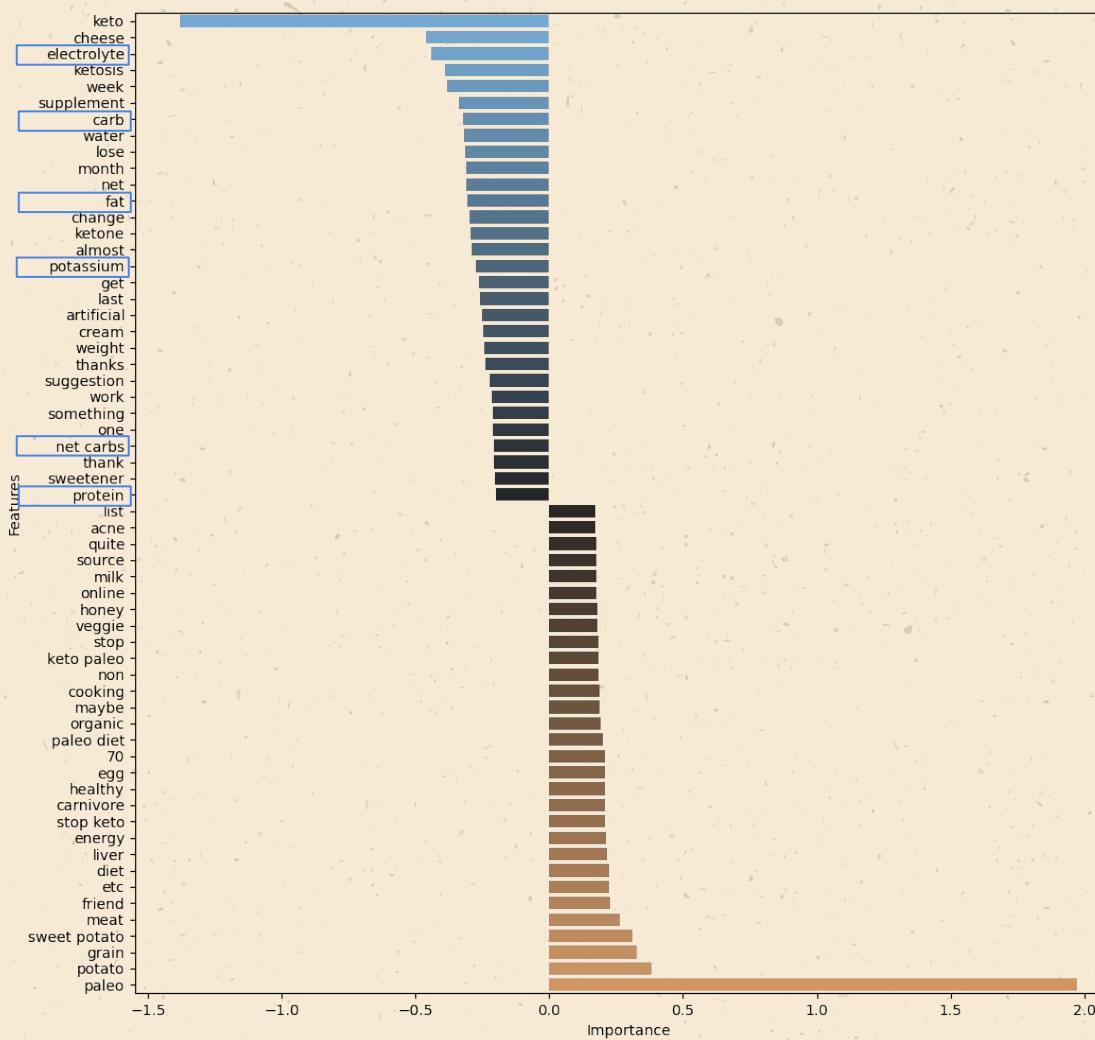
Keto



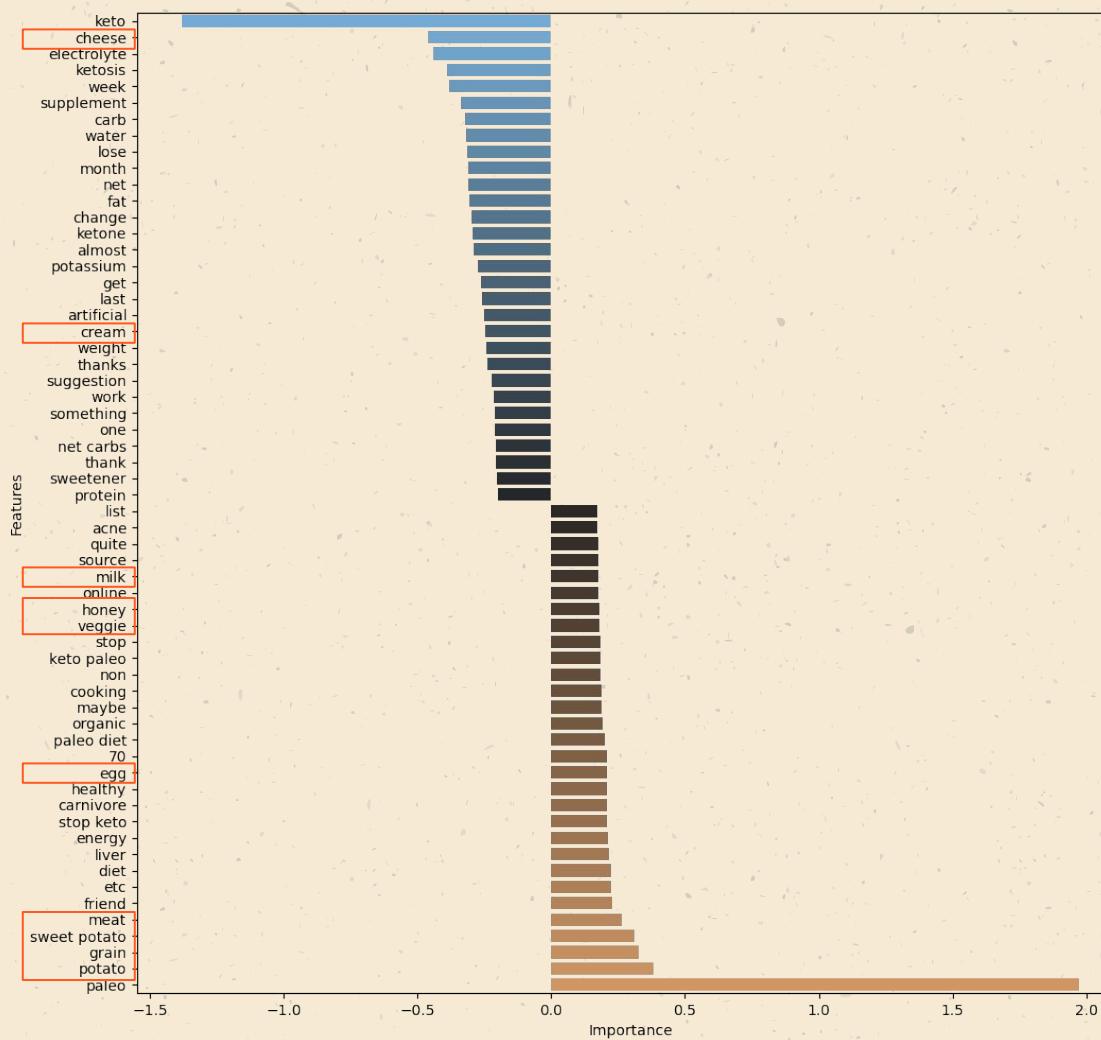
Paleo

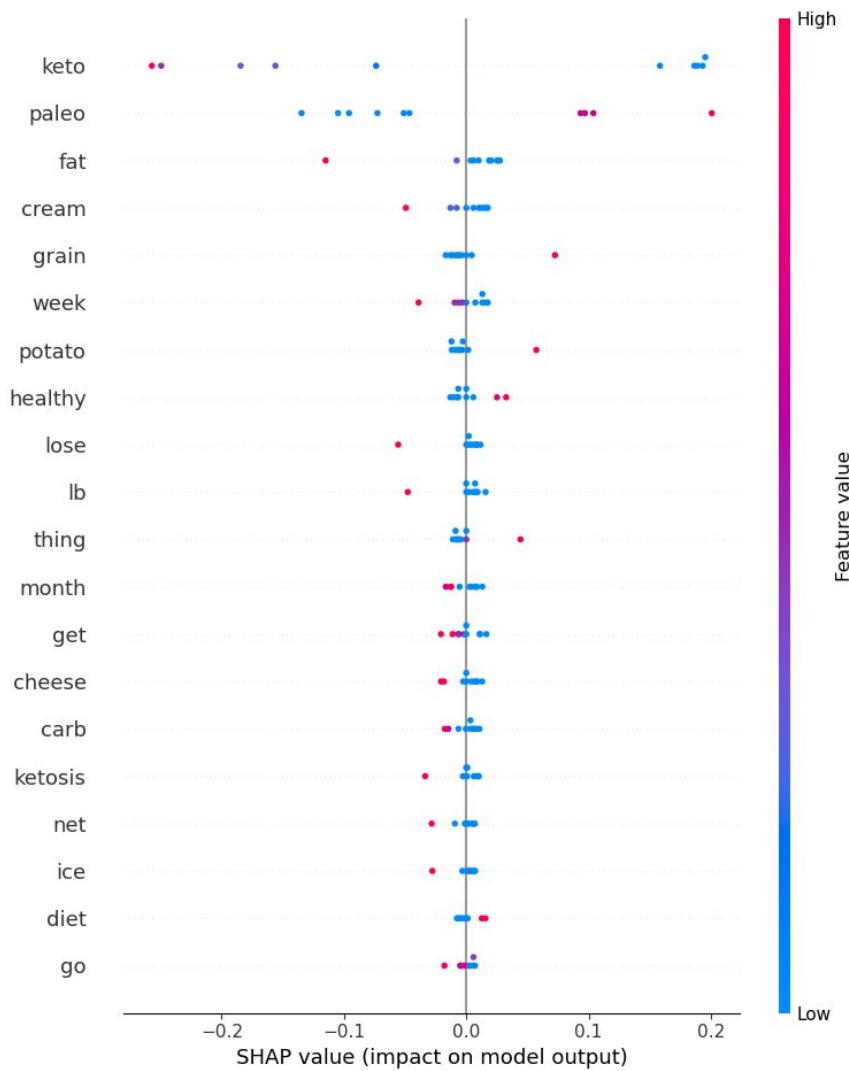


Macro & Micro Nutrients

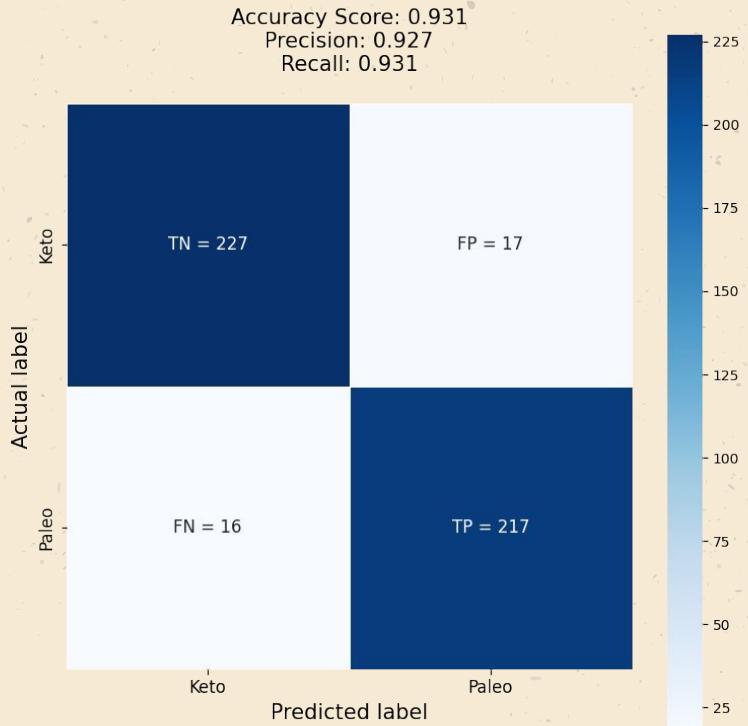


Ingredients



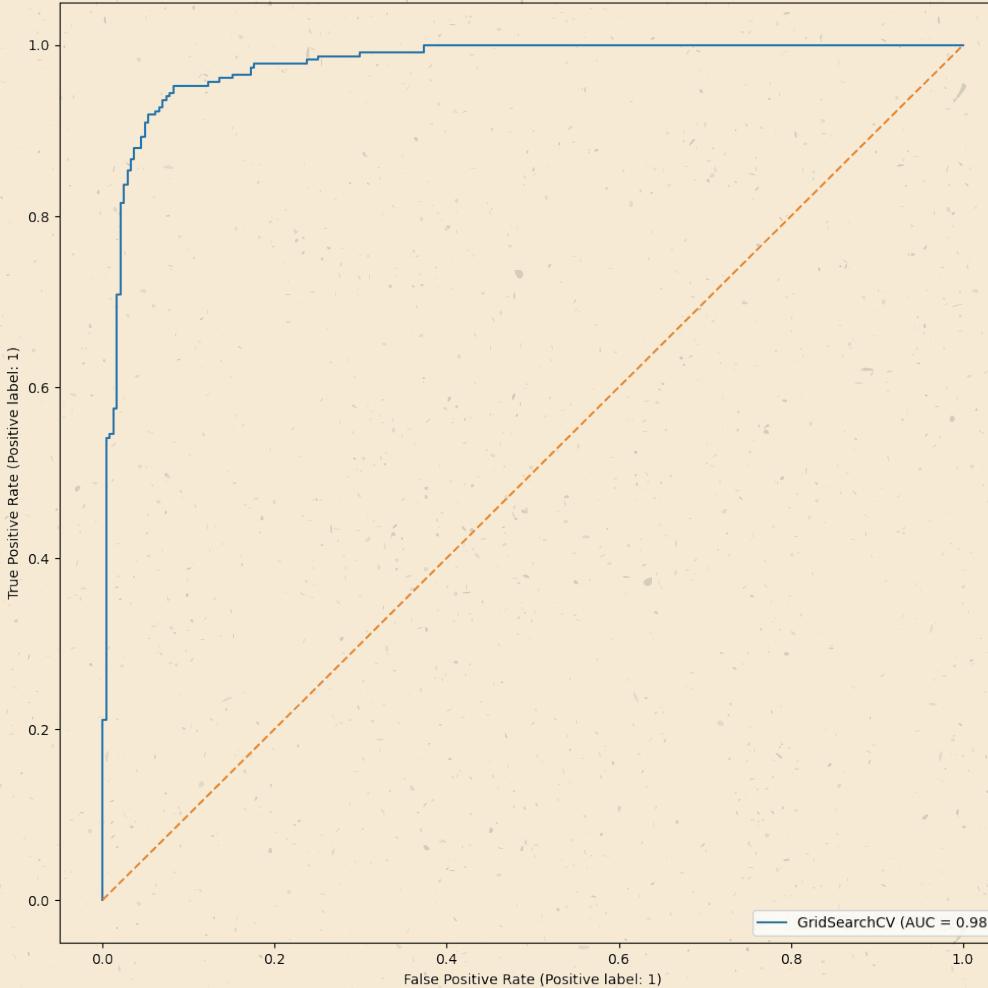


Confusion Matrix



ROC

AUC =
0.98





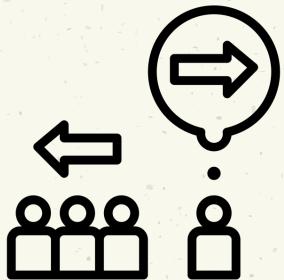
05

Insights & Recommendations

Insights



Differing Focus



Shared Themes



Navigating Uncertainty



Micronutrients are vitamins and minerals needed by the body in small amounts.

Differing Focus



Shared Themes

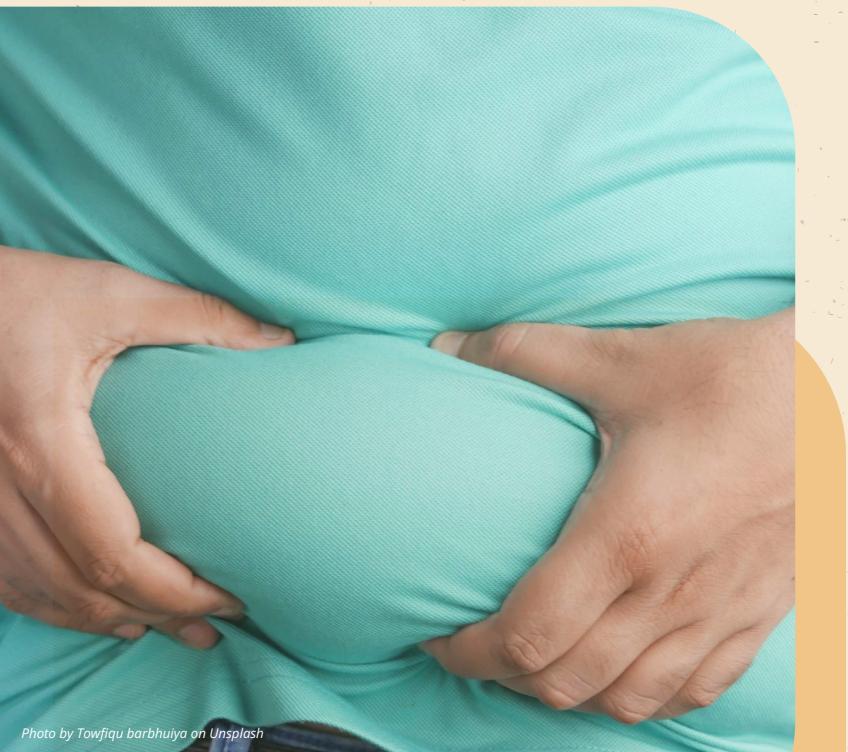


Photo by Towfigh barbhuuya on Unsplash



<https://www.wholesomeyum.com/low-carb-keto-food-list/>

Navigating Uncertainty

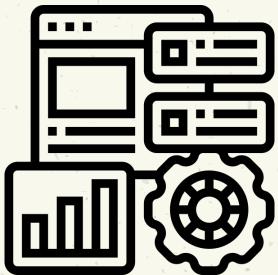


Photo by Siora Photography on Unsplash

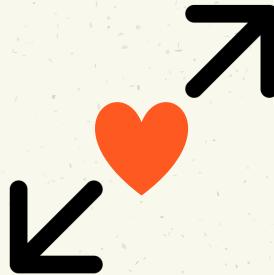
Recommendations



Dynamic Tracking



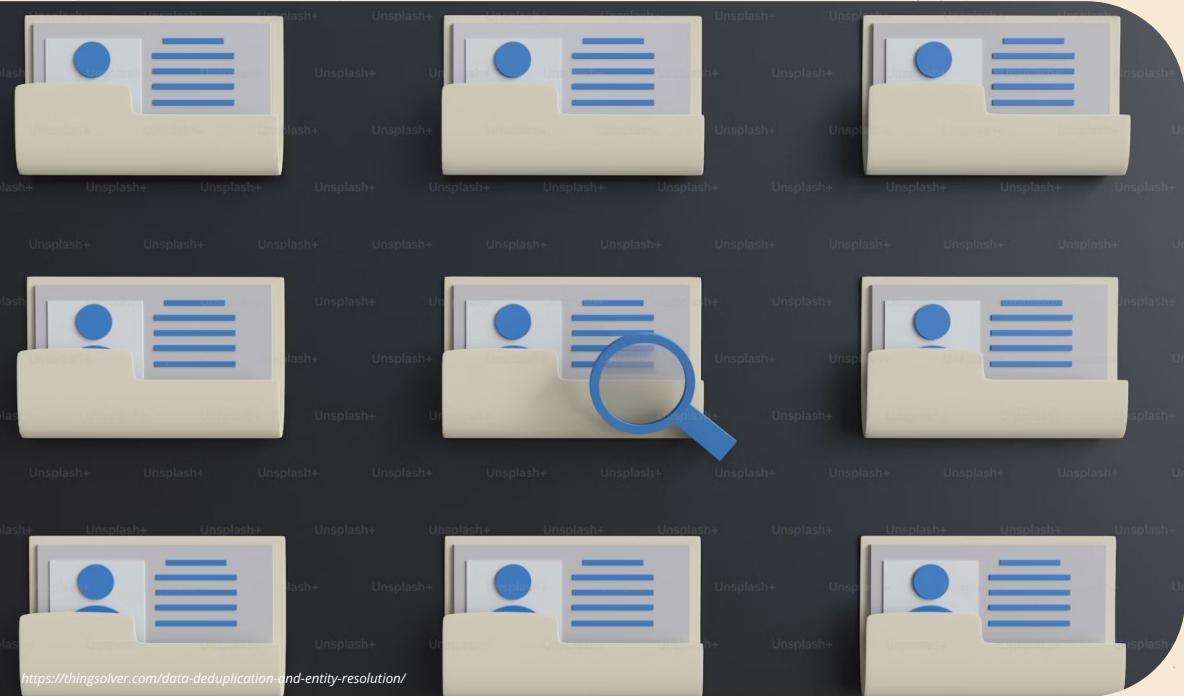
Broaden Appeal



Educational Content

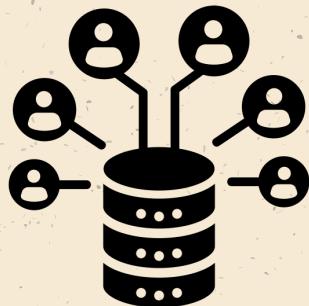


Dynamic Tracking



- + Dynamic User Analysis
- + Personalized Recommendations

Dynamic Tracking

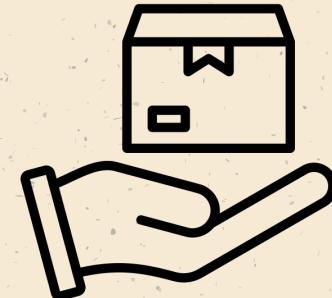


Collect Data

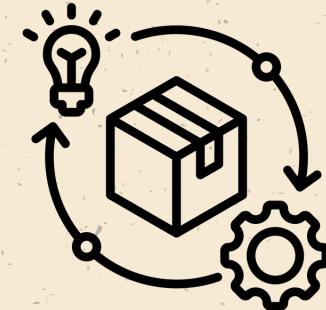
Upload Data



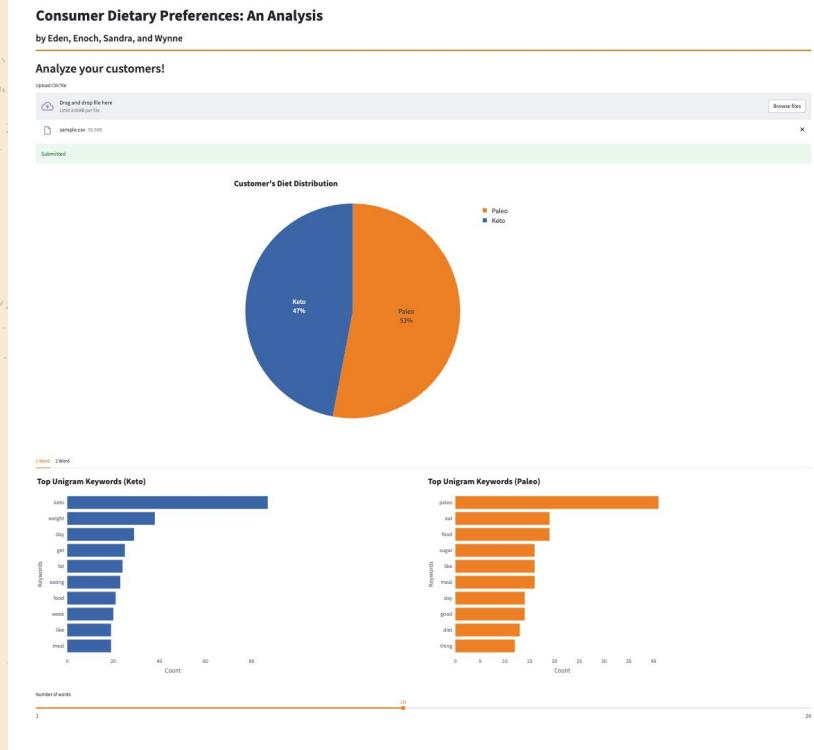
Product
Recommendation



Product
Development



Dynamic Tracking



Insights Businesses Consumers

Consumer Dietary Preferences: An Analysis
by Eden, Enoch, Sandra, and Wynne

Consumer View

Ask us anything!

Enter here
Except from fishes, from what hunter gatherer were finding their omega 3 ?
Submit

Successfully submitted!

Paleo Puffs
PALEO PUFFS TWO CHEESE + CHEESELESS

Paleo Chips
GRAIN FREE ARTISAN TROPIC PALEO PLANTAIN STRIPS CRISP

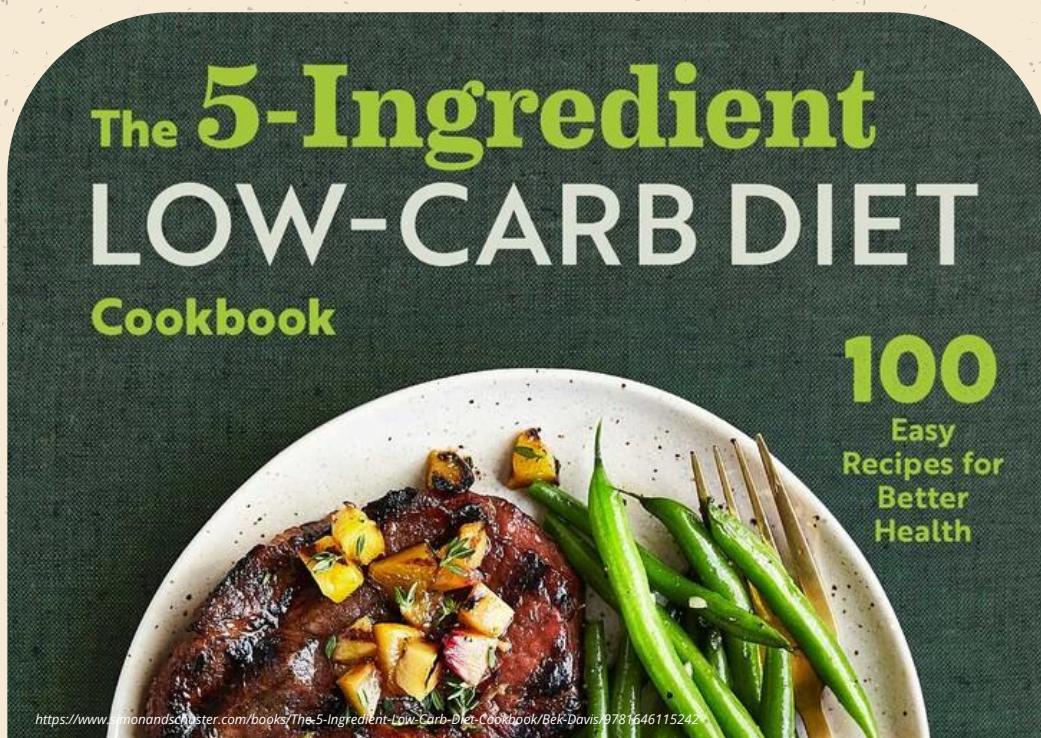
Paleo Muesli
mymuesli PALEO MUESLI

Broaden Appeal

- + Weight Loss
- + Low Carb
- + Calorie Labeling
- + Trending Topics



Educational Content



<https://wwwimonandschuster.com/books/The-5-Ingredient-Low-Carb-Diet-Cookbook/9781646115242>

- + Blog Articles
- + Instructive Videos
- + Curated Recipes

Future Enhancement

Multiclass Classification

- + Enhanced Recommendations
- + Expand Reach
- + Diversify Customer Base



<https://www.flexsuccess.com.au/blog/diet-types-and-why-they-all-work>

Multiclass Classification is where each data sample is assigned one and only one label from more than two classes.

Source: <https://vitalflux.com/difference-binary-multi-class-multi-label-classification/#:~:text=Binary%20Classification%20is%20where%20each,from%20more%20than%20two%20classes.>



06

Product Demo



07

Conclusion

Conclusion



Objective 1

- + F1-Score: 0.93
- + Data-driven
- + Personalized
- + Dynamic



Objective 2

- + Analyse Preferences
- + Insights
- + Strategies





Thank You

Do you have any questions?

CREDITS: This presentation template was created by [Slidesgo](#), including icons by [Flaticon](#), and infographics & images by [Freepik](#)



Annex



Customised Stopwords

i	they	do
me	them	does
my	their	did
myself	theirs	doing
we	themselves	a
our	what	an
ours	which	the
ourselves	who	and
you	whom	but
your	this	if
yours	that	or
yourself	these	because
yourselves	those	as
he	am	until... etc



pinned	community
top	want
subreddit	would
ask	think
question	really
ha	nutshell
beginner	utshell
hello	support
talk	thread
wa	ive
also	im

