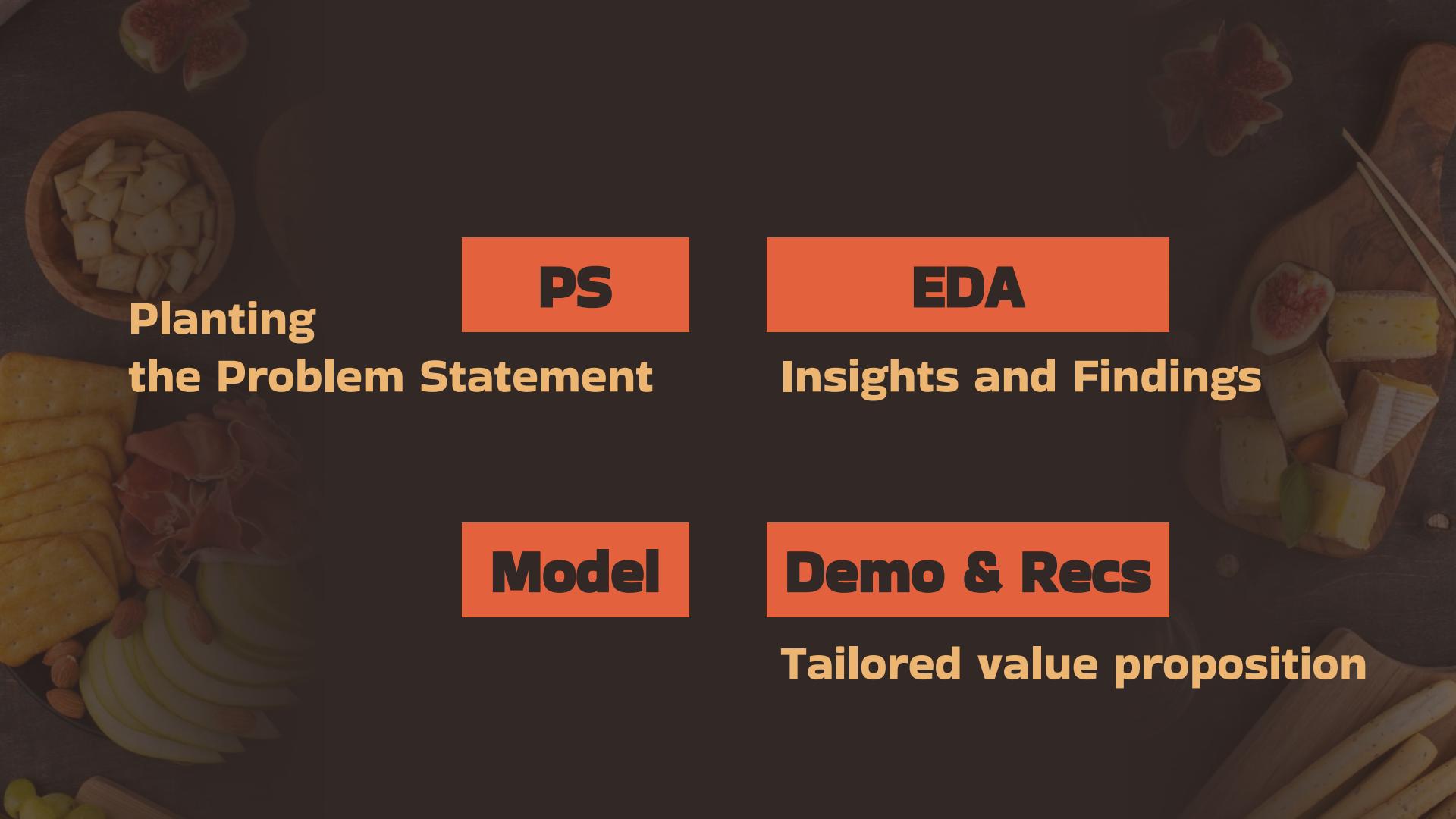


# Project 3

vivalamuffin



**Planting  
the Problem Statement**

**PS**

**EDA**

**Insights and Findings**

**Model**

**Demo & Recs**

**Tailored value proposition**

# F&B in SG: A Fairytale Story?

As Singapore's tourism sector recovers post-pandemic, the food and beverage sector is also experiencing a rebound, buoyed by tourism receipts. F&B sales volume rose 28 percent year-on-year in the third quarter of 2022.

manufacturers, and distributors. The sector is a significant contributor to Singapore's GDP, raking over S\$5 billion in 2022.

Looks like the F&B industry in Singapore is booming...

## KEY INDICATORS OF FOOD & BEVERAGE SERVICES



Total Food & Beverage Sales

Year-on Year  
▲ +6.5%

Month-on-Month  
(Seasonally adjusted)  
▲ +0.2%



Total Food & Beverage Sales

Sales Value  
\$1.0 Billion

Online Sales Proportion  
22.8%

# F&B in SG: Crushing Reality

**1 in 4 F&B businesses in Singapore churn out of business every year.** Many close within the first year of operations. This is the case across segments, from

More often than not, would-be restaurateurs fail to do enough market research before plunging into the business. It is not surprising to hear of young, eager

clients. The saddening statistics is that 40% of F&B businesses don't survive past 5 years. And the 60% that survive, the industry will push the load of

...but almost half of F&B businesses go out of business in the first 5 years!

# Key Problems

Existing businesses	New businesses
Managing operating costs	High setup cost
Training and hiring	Obtaining licenses
Understanding customer base	Research

# Why is Research a problem?

01

Time

02

Cost

03

Effort

# Problem Statement

*“Can a model help F&B business to more effectively understand, segment, and target their customer base?*

*Can we also help prospective F&B businesses with identifying potential locations and supplier networks?”*





**\$5 Billion  
6.5% annual growth**

F&B Market size

# Meat Market - What is at Steak?

**2%**

**Average Annual Growth  
[2021 - 2026]**

**93%**

**SG Population that  
eats Meat**

# Vegan Value - Roots of the Matter

7%

Average Annual Growth  
[2022 - 2027]

7%

Current SG Population  
that identifies with Veg

<https://www.mordorintelligence.com/industry-reports/singapore-plant-based-food-and-beverages-market>

<https://vegeconomist.com/2s-singapore>

<https://www.panesas.com/blog/post/type%20of%20vegetarian%20vs%20vegan%20in%20Singapore#:~:text=Veganism%20or%20Meat%2DFree%20culture,a%20small%20country%20like%20Singapore>

# “Meat” the subreddits

**r/Meat**

2390 posts

Community of meat cooking enthusiasts

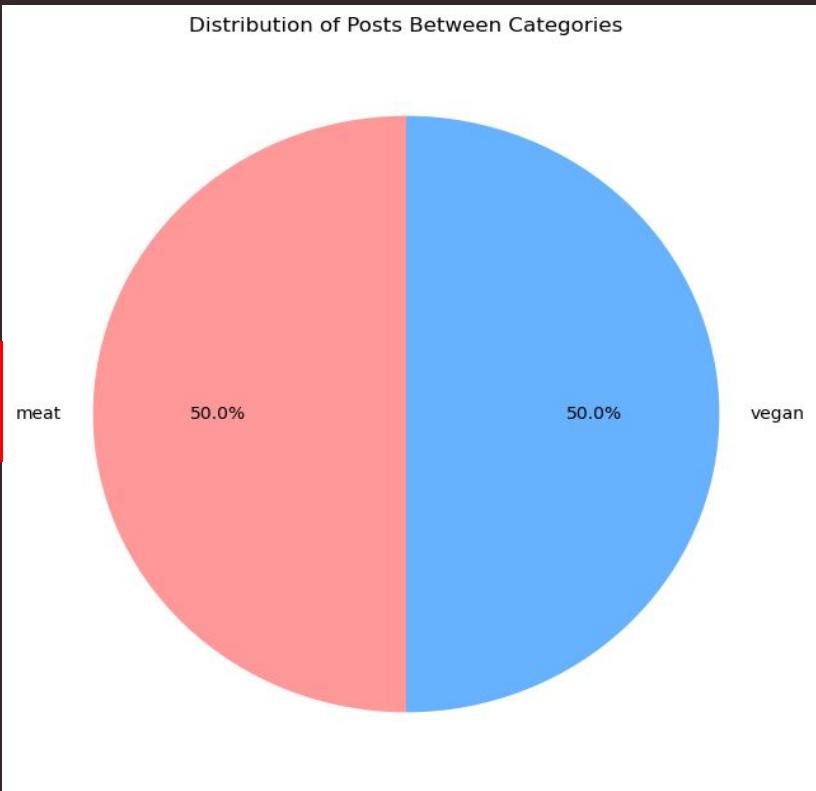


**r/Vegan**

3318 posts

Community of vegans exchanging recipes

# 6000 POSTS



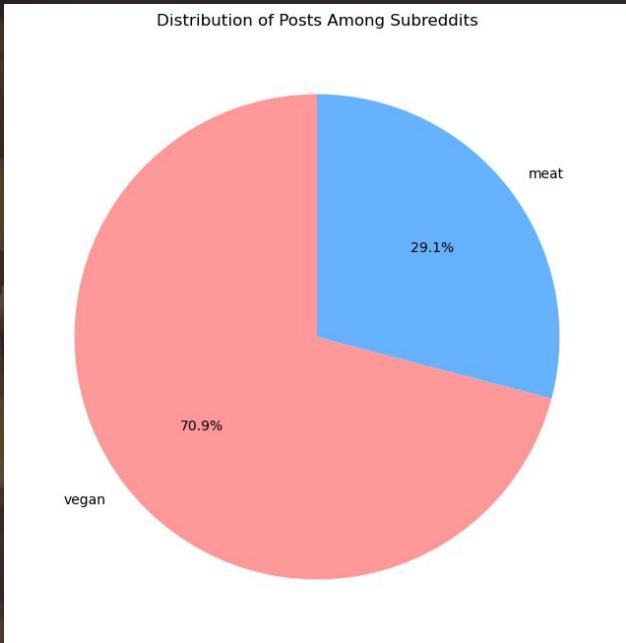
3000

meat

3000

vegan

# Trimming the Fat



**1763**

Total posts left

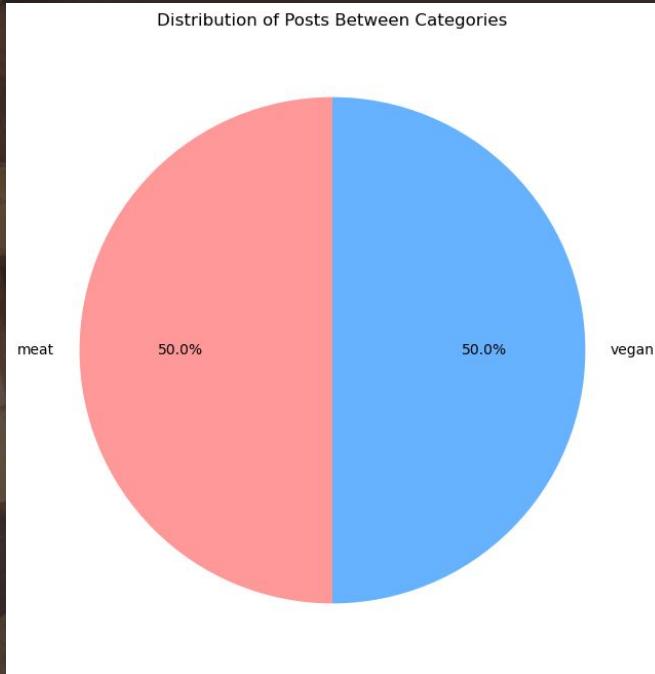
**1250**

Vegan posts

**513**

Meat posts

# SMOTE THIS STEAK!!



**2500**

Total posts after SMOTE

**1250**

Vegan posts

**1250**

Meat posts after SMOTE

## SMOTE

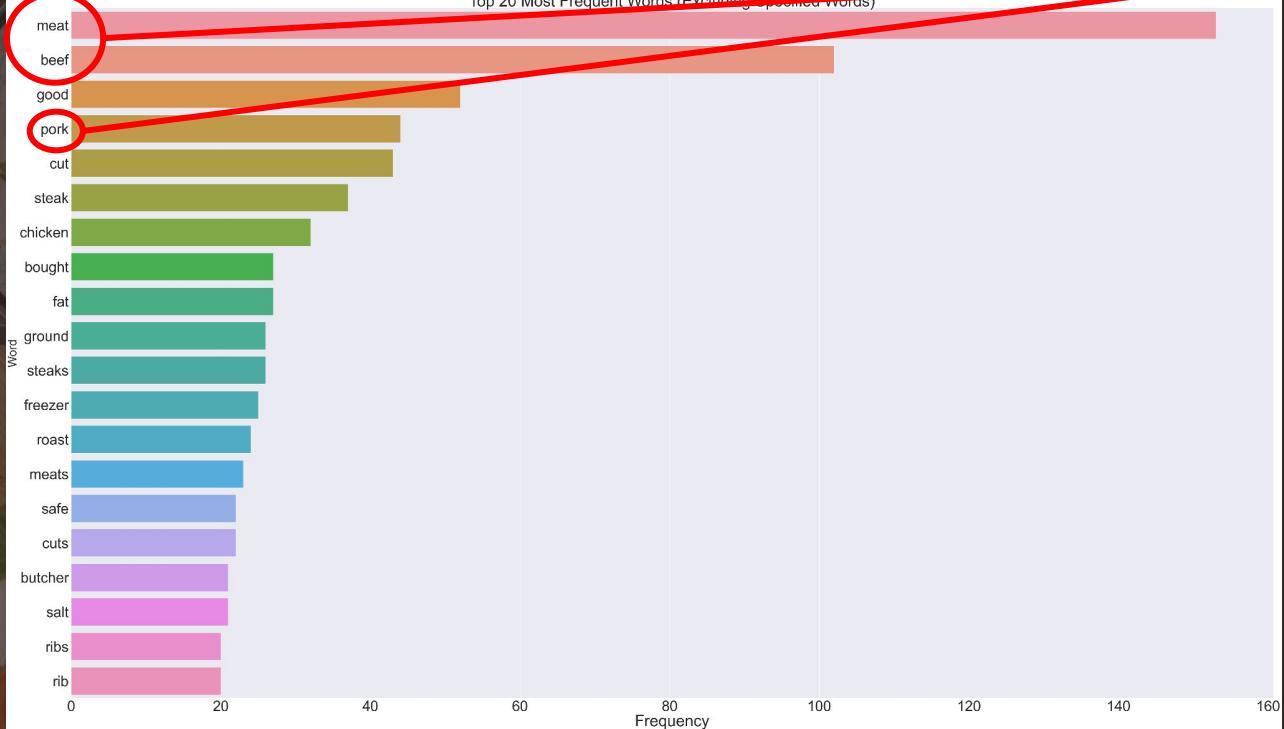
the SMOTE technique is used to address the class imbalance in the dataset.

In this instance, the MEAT subreddit was upscaled to match the vegan's subreddit.

This way, the machine has equal opportunity to train on both subreddits.

**SMOTE IS NOT APPLIED TO TEST DATA**

# Top KeyWords (r/Meat)



## Top Topics

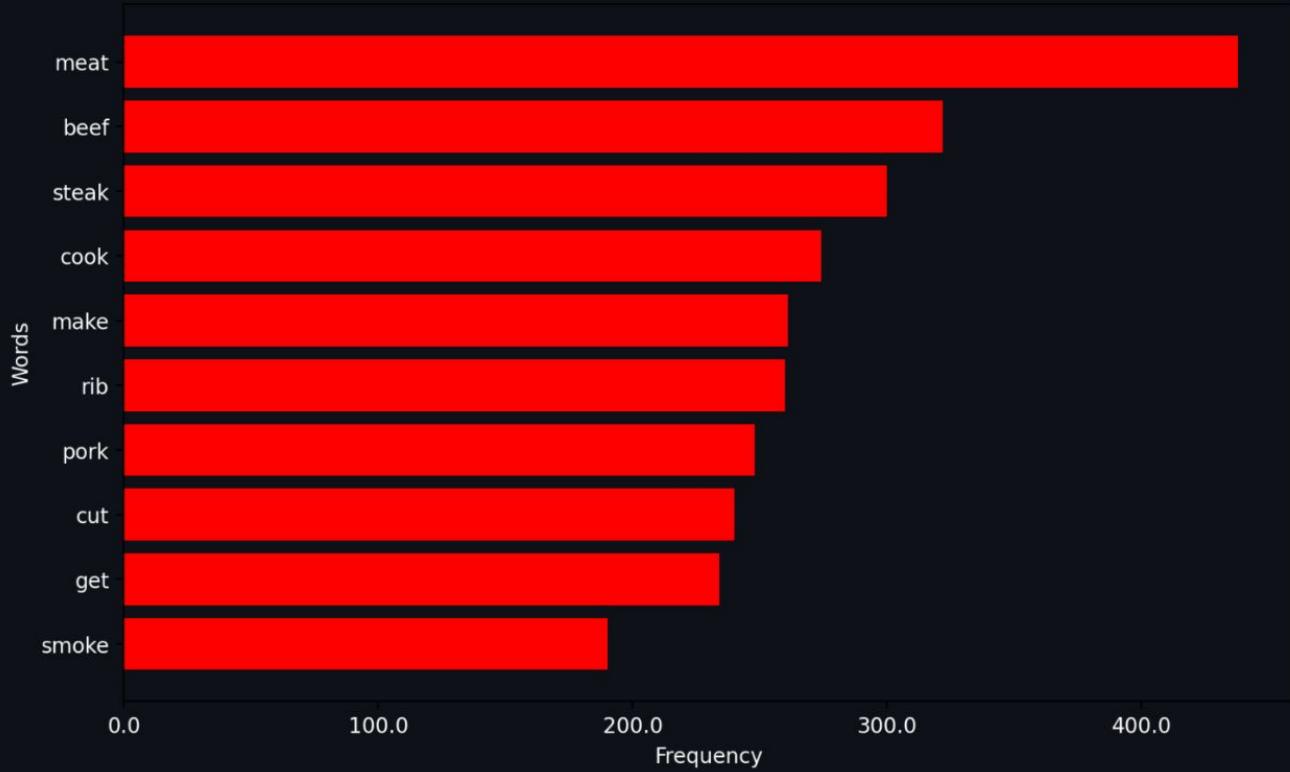
### Meats of Interest

1. Beef
2. Pork
3. Chicken
4. Lamb

### Cooking Equipment

- Oven
- Smoker
- Fridge/Freezer

# Top KeyWords (r/Meat)



## Top Topics

### Meats of Interest

1. Beef
2. Pork
3. Chicken
4. Lamb

### Cooking Equipment

- Oven
- Smoker
- Fridge/Freezer

# Top Themes (r/Meat)

## Top Cuts

Beef

1. Ground Beef
2. Beef Ribs
3. Sirloin Tip
4. Beef Fat

Pork

1. Pork Shoulder
2. Uncured Bacon
3. Pork Chops

## Concerns

### A. Concerns

- Nitrates/Nitrites
- Processed Meats

### B. Refrigeration

### C. Reliability of Store-Bought Meat

## Recipes

### Recipes

1. Sous Vide
2. Steak
3. Brisket

# Top Themes (r/Meat)

## Top Cuts

### Beef

1. Ground Beef
2. Beef Ribs
3. Sirloin Tip
4. Beef Fat

### Pork

1. Pork Shoulder
2. Uncured Bacon
3. Pork Chops

### [FOOD RECALL]

SFA has directed the recall of Bibik's Choice Pepper Chicken Patties due to presence of undeclared egg allergen. The recall is ongoing.

Singapore beef imports have grown steadily in recent years, underpinned by an expanding retail and foodservice sector and its growing popularity among Singaporeans. However, beef

Meat Insider

## Singapore's pork demand not curtailed by high prices

By **Rajeswari Ramanee** - 4 September, 2023, AM

# Top Themes (r/Meat)

## Top Cuts

Beef

1. Ground Beef
2. Beef Ribs
3. Sirloin Tip
4. Beef Fat

Pork

1. Pork Shoulder
2. Uncured Bacon
3. Pork Chops

## Concerns

### A. Concerns

- Nitrates/Nitrites
- Processed Meats

### B. Refrigeration

### C. Reliability of Store-Bought Meat

## Recipes

### Recipes

1. Sous Vide
2. Steak
3. Brisket

# Top Themes (r/Meat)

## Top Cuts

Beef

1. Ground Beef
2. Beef Ribs
3. Sirloin Tip
4. Beef Fat

Pork

1. Pork Shoulder
2. Uncured Bacon
3. Pork Chops

## Concerns

### A. Concerns

- Nitrates/Nitrites
- Processed Meats

### B. Refrigeration

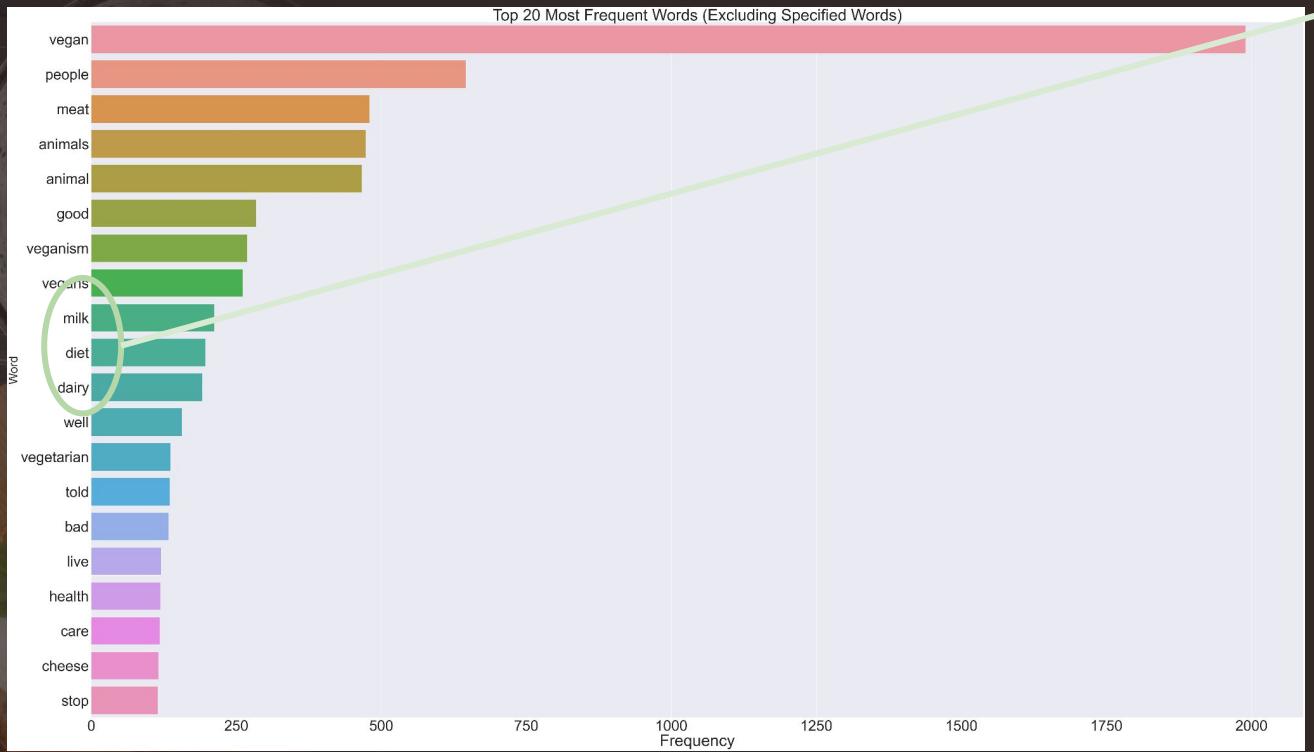
### C. Reliability of Store-Bought Meat Products

## Recipes

### Recipes

1. Sous Vide
2. Steak
3. Brisket

# Top KeyWords (r/Vegan)



## Top Topics

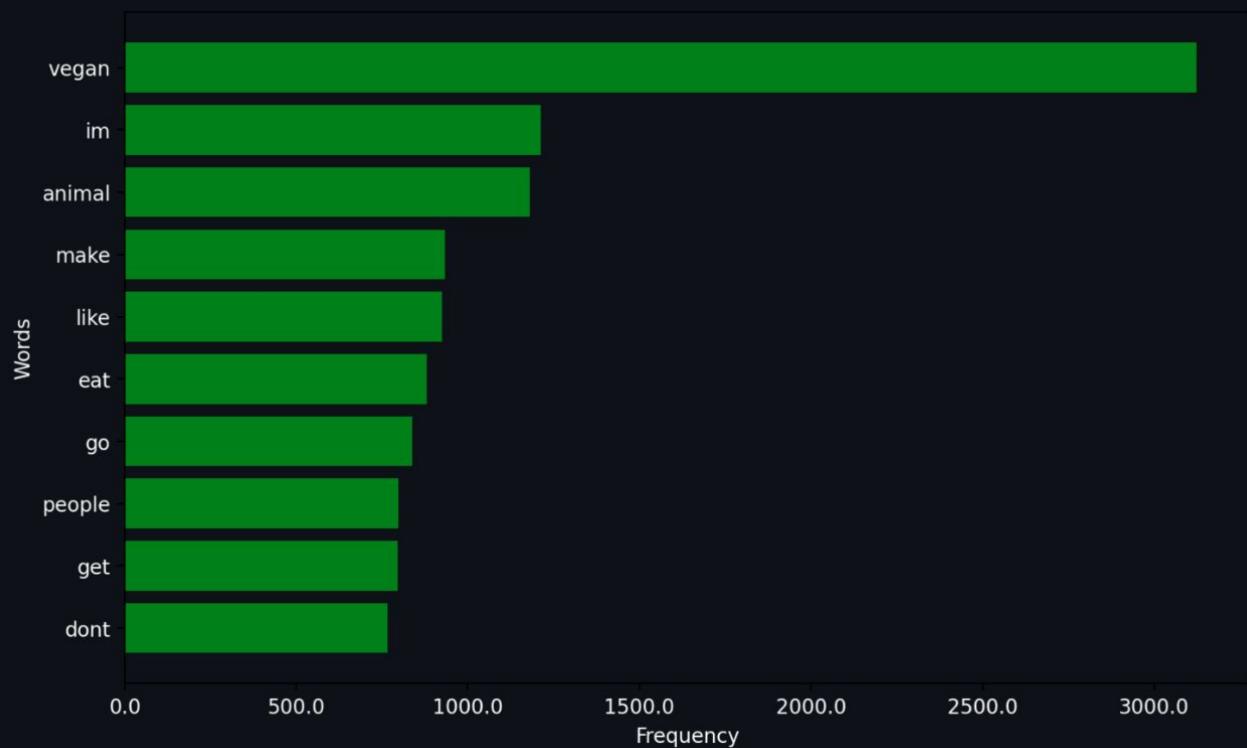
### Nutrition & Options

- Protein
- Vegan Diet
- Vegan

### Animal Welfare

- Animal Rights
- Activism

# Top KeyWords (r/Vegan)



## Top Topics

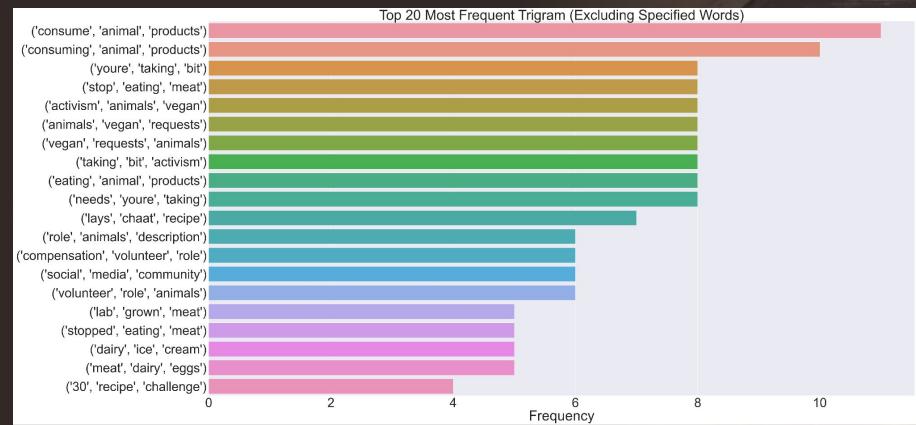
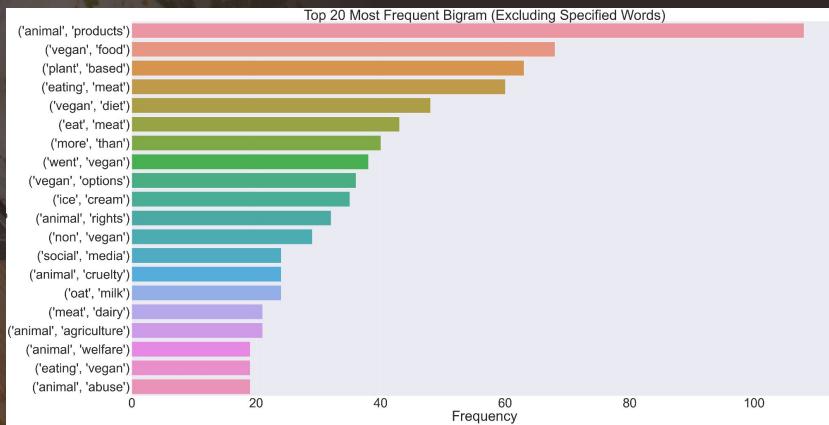
### Nutrition & Options

- Protein
- Vegan Diet
- Vegan

### Animal Welfare

- Animal Rights
- Activism

# Top Phrases (Bigram and Trigrams) (r/Vegan)



# Top Themes (r/Vegan)

## Top Foodstuffs

### Dairy Options

- A. Ice Cream
- B. Oat Milk
- C. Vegan Cheese
- D. Almond Milk
- E. Soy Milk

### Meat Options

- Lab Grown Meat

## Concerns

- 1) Usage of Animal Products
- 2) Climate Change
- 3) Moral Obligations
- 4) Mental Health

## Recipes

### Recipes

- 1. Lays Chaat
- 2. Marble Cake
- 3. Chilli Garlic Noodles
- 4. Vegan Cheese Pizza

# Top Themes (r/Vegan)

## Top Foodstuffs

### Dairy Options

- A. Ice Cream
- B. Oat Milk
- C. Vegan Cheese
- D. Almond Milk
- E. Soy Milk

### Meat Options

- Lab Grown Meat

### Plant-Based Milk Adoption on the Rise in Singapore

In Singapore, the plant-based milk market is rapidly gaining mainstream popularity, with 87% of consumers having tried a plant-based milk beverage. Of those who have tried it, a significant 62% consume it regularly. This growing trend towards alternative milk sources

First Singapore-produced vegan cheese to offer more alternatives to consumers

Why Singapore is the only place in the world selling lab-grown meat

# Top Themes (r/Vegan)

## Top Foodstuffs

### Dairy Options

- A. Ice Cream
- B. Oat Milk
- C. Vegan Cheese
- D. Almond Milk
- E. Soy Milk

### Meat Options

- Lab Grown Meat

## Concerns

- 1) Usage of Animal Products
- 2) Climate Change
- 3) Moral Obligations
- 4) Mental Health

## Recipes

### Recipes

1. Lays Chaat
2. Marble Cake
3. Chilli Garlic Noodles
4. Vegan Cheese Pizza

# Sentiment Analysis

Activist Avo



Chill Carnivore



NEUTRAL MOOD

POSITIVE MOOD

HIGHER BIAS

LOWER BIAS

# Sentiment Analysis

**Activist Avo**



neg	0.078370
neu	0.804984
pos	0.112725
compound	0.045342



**Chill Carnivore**

neg	0.033673
neu	0.874719
pos	0.091607
compound	0.103151

**NEUTRAL** Mood

**POSITIVE** Mood

# Sentiment Analysis (TextBlob)

Veggie Verve



3x Expressive

Average subjectivity coefficient:

0.28



Cold Carnivore

Objective

Average subjectivity coefficient:

0.09

# Persona Classification

## Vegan-Focused Consumer



- 1) More Social Media Savvy
- 2) Interest in trying new food options
- 3) Cares about social causes e.g Climate Change, Mental Health
- 4) Strongly against animal cruelty



## Meat-Focused Consumer

- 1) Aware about dangers of processed, unsafe meats
- 2) Interest in learning to cook meat dishes
- 3) Cares about high quality and freshness of meats
- 4) Against wasting meat foodstuffs

# **Meat the Metrics**

# Metrics Definitions

## Accuracy

Correct Predictions

—

Total Predictions

## Precision

True Positives

—

True Positives + False Positives

## Recall

True Positives

—

True Positives + False Negatives

## F1-Score

$2 \times (\text{Precision} \times \text{Recall})$

—

Precision + Recall

## ROC-AUC

ROC-AUC is a measure that tells us how well a model can separate good things from bad things; higher values means the model is better at this task

# What each model actually mean

## Multinomial Naive Bayes

- Classifier for count data, often used in *text analysis*

## Bernoulli Naive Bayes

- Classifier for binary data, frequently used for text classification with binary features such as *spam filtering*

## Random Forest

- Binary classifier common in finance, healthcare and marketing such as predicting *customer churn from purchase data*.

## XGBoost

- Boosted trees method idea for structured data such as *predicting house prices from attributes*

## Logistic Regression

- Ensemble of decision trees used for classification and regression such as *predicting e-commerce purchase likelihood based on browsing history*

# **Meat vs Vegan**

**Round 1**

# Model Scores

Models	Train Accuracy	Test Accuracy	Precision	Recall	F1-Score	ROC-AUC
Bernoulli Naive Bayes	88.01	84.99	99.02	79.92	88.45	98.81
XGBoost	99.72	94.90	96.46	96.46	96.46	98.64
Random Forest	100	95.75	96.50	97.64	97.06	99.21
Logistic Regression	99.57	96.88	96.91	98.82	97.86	99.22
Multinomial Naive Bayes	98.51	97.17	97.29	98.82	98.05	99.04

# Multinomial Naive Bayes

## Efficiency

Imagine MNB as a robot that quickly reads and sorts comments about meat and vegan subreddits. It handles big lists of words from both groups really fast. (Within seconds)

## Scalability

Even if there are thousands of comments from meat-lovers and plant-lovers, MNB can read them all without getting overwhelmed.

## Handles Discrete Data

MNB counts words like “steak” or “salad” to see how often people talk about them in comments.

## Simplicity in Implementation

MNB’s straightforward approach means fewer chances for errors during setup, ensuring consistent and reliable results for business analysis

# Demonstration

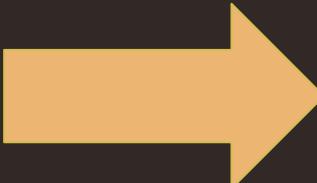


# Bringing it together: Problem & Solution

## Problem

Can a model help F&B business to more effectively understand, segment, and target their customer base?

Can we also help prospective F&B businesses with identifying potential locations and supplier networks?



## Solution

A one-stop shop for all things food research.

- Helps businesses understand their customer base.
- Provides keywords for tailored outreach.
- Helps new businesses scout locations and suppliers, also suggesting menu items.

# Bringing it together: Impact, Impact, Impact

## Cost Reduction

- Reduced research costs.
- Streamlined marketing costs for higher ROI.

## Efficient Market Entry

- Data-driven decisions for optimal location, menu, and suppliers.
- Minimizes risks for new F&B ventures.

## Enhanced Customer Experience

- Enhanced satisfaction through personalized menus.
- Improved customer reviews and ratings.

## Increased Success Rate

- Market insights to help refine offerings and strategies.
- Boosts new business success in a highly competitive market.

# Future work

## Add more diet preferences

Improve model accuracy with data from varied diet-related subreddits (e.g., gluten-free, keto, paleo) for comprehensive user classification.

## Live menu recommendations

Conduct online research to identify trending menu items with high profitability, providing valuable recommendations to businesses.

## Estimate the cost of set up

By analyzing average costs in rental, food, utilities, and manpower, we can estimate the necessary resources to open a restaurant.

## Keywords download

Marketers of existing businesses can effortlessly download their top keywords and seamlessly integrate with Google for online advertisements.

# Thanks!

Do you have any questions?

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



# Model Score (After tuning)

Models	Train Accuracy	Test Accuracy	Precision	Recall	F1-Score	ROC-AUC
Bernoulli Naive Bayes	90.57	91.78	97.47	90.94	94.09	98.27
XGBoost	93.05	94.62	97.19	95.28	94.09	98.78
Random Forest	94.89	96.60	98.02	97.24	97.63	96.60
Logistic Regression	96.17	96.88	96.91	98.82	97.86	99.22
Multinomial Naive Bayes	95.67	97.17	97.29	98.82	98.05	99.06

# Why not use the tuned model?

## Generalization

The goal is to have a model that generalizes well to new, unseen data. The model performs better on the test set. Therefore, it is likely to perform better on new data

## Complexity

A simpler model, often the untuned version is easier to interpret, understand and troubleshoot. If it also performs better or comparably to a more complex tuned model, it's often the better choice

## Overfitting

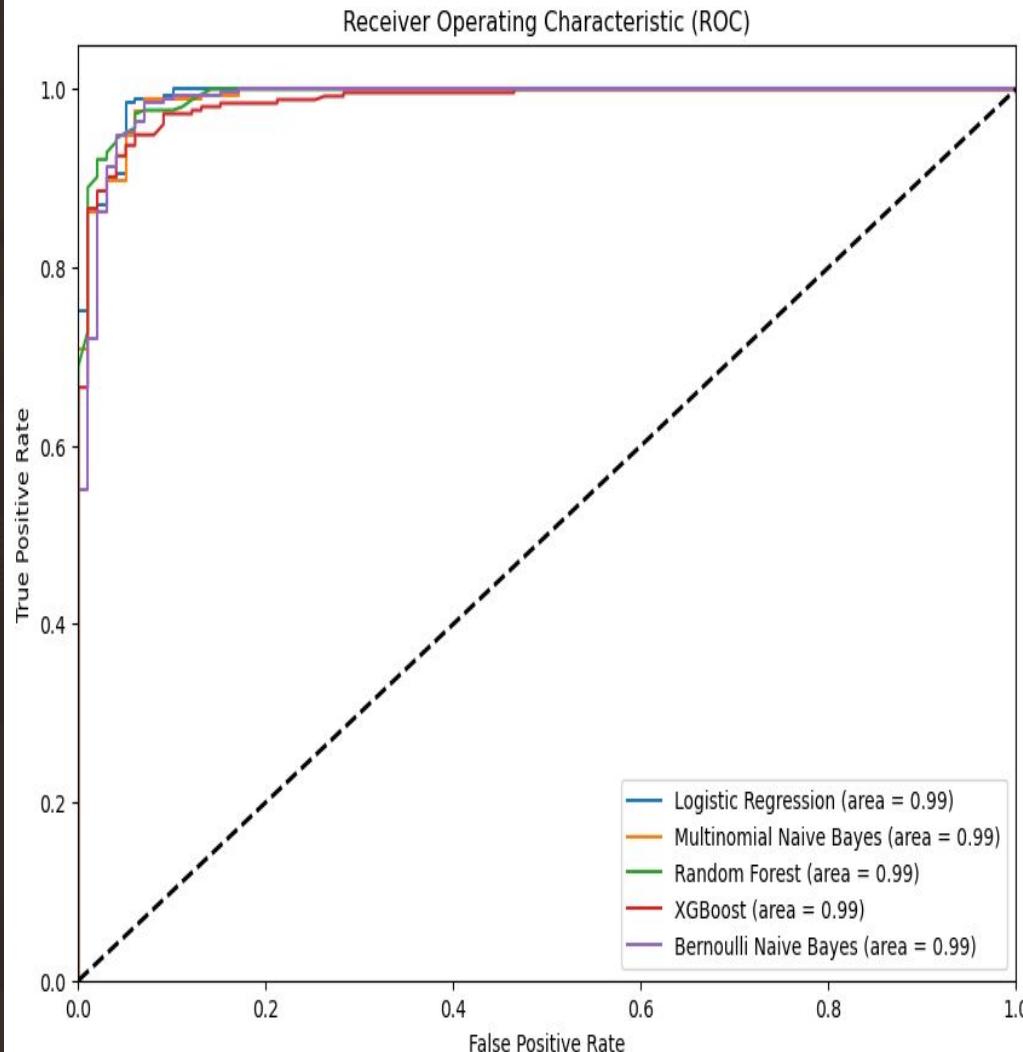
Tuning can lead to overfitting, where the model becomes too specialized to the training data and performs poorly on new data

## Efficiency

Untuned models might be more computationally efficient in terms of prediction time especially tuning involves adding complexity

# Question: What is ROC-AUC?

- ROC-AUC stands for Receiver Operating Characteristics - Area Under the Curve
- True positives vs False positives: The ROC curve plots the true positive rate for different threshold values
- Performance Indicator: AUC measures the entire two-dimensional area underneath the ROC curve, providing a single value summary of model performance
- In summary : ROC-AUC measures a model's ability to distinguish between classes. With 1.0 being perfect and 0.5 being no better than random guessing



# Metrics Definitions

## Precision

Of all the instances predicted as positive, how many were actually positive? It's about the exactness or quality of the prediction

## Recall (Sensitivity)

Of all the actual positive instances, how many were correctly predicted? It focuses on capturing as many positives as possible

## Accuracy

Measures the proportion of all predictions that are correct. It's the overall rate of correct predictions among all cases

## F1-Score

A balance between Precision and Recall. It's the Harmonic mean of the two, ensuring neither metric is favored too heavily

## ROC-AUC

Represents the model's ability to distinguish between positive and negative classes. The AUC quantifies the overall performance of the classifier with 1 being perfect and 0.5 no better than random guessing

# Types of Data

## Count Data

Whole number values counting occurrences like the number of meat based or vegan based words in a post/comment

## Boosted trees method

Using multiple trees in a sequence, where each new tree tries to correct the mistakes of the previous one.

## Continuous Data

Values that can take any number within a range like height or weight

## Binary Data

Data with only 2 possible outcomes like Yes/No or 1/0

## Ensemble

Combining multiple models' prediction for a better overall guess

# **SMOTE OR SMOKED**

Synthetic Minority Over-sampling Technique aka SMOTE, is a method to create synthetic samples in datasets to balance class distribution. (Meat vs Vegan)

**It's used in events when there is imbalanced data**

# Average Subjectivity Coefficient

In the context of text analysis, it could potentially refer to a metric that measures the average degree to which a set of texts expresses personal opinions or feelings rather than objective information.

## Average

Refers to the mean or central value of a set of numbers

## Subjectivity

This often relates to how someone's judgement is shaped by personal opinions and feelings instead of outside influences

## Coefficient

A number showing how much personal opinion or feeling is in a statement, with a higher number meaning more opinion