



Understanding Important Features for Real & Alternative Meat

About Us



Harsha Gurram



Vikramaditya,
Sriramachandra



Dubey, Shriya



Pratham
Khubchandani



Samarth, Verma



Hiral, Yadav



Kathyayani,
Chenimineni

The Problem & Hypothesis

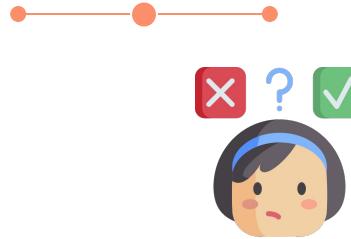
understanding the question

Problem & Hypothesis



The Problem

Real meat category
outperforms plant-based
alternatives



Our Hypothesis

If we create **similar experiences** for consumers when purchasing real and alternative meat, revenue from alternative meats will increase



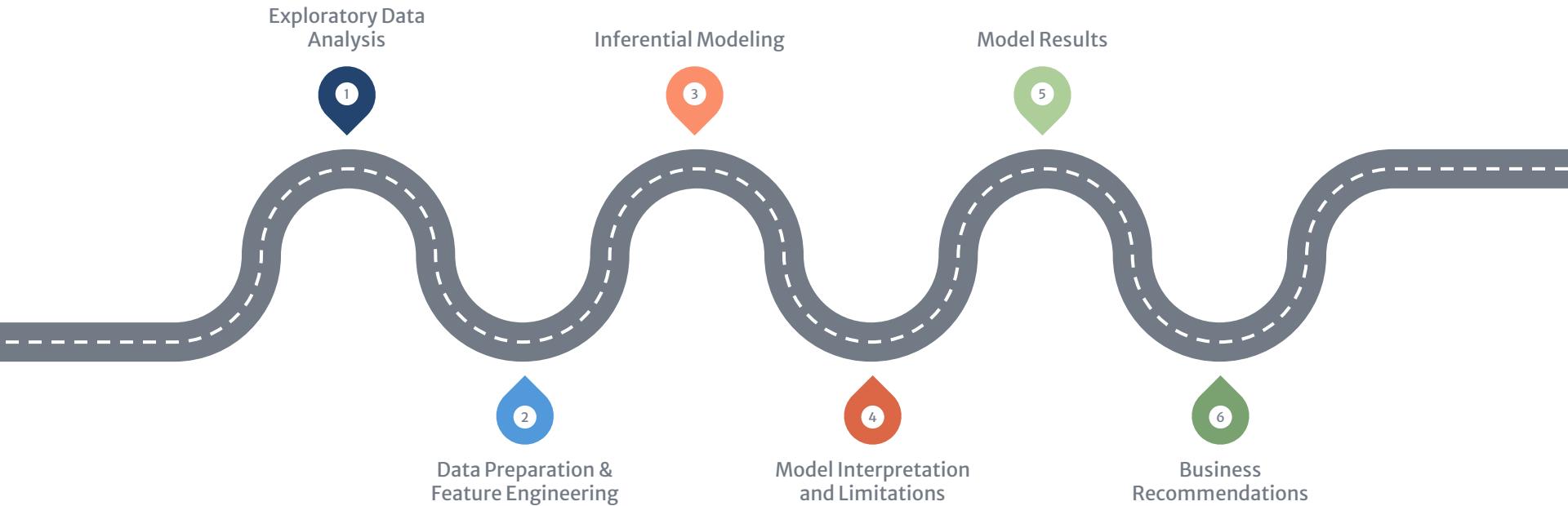
2

Our Approach

our approach to test the hypothesis



Roadmap





3

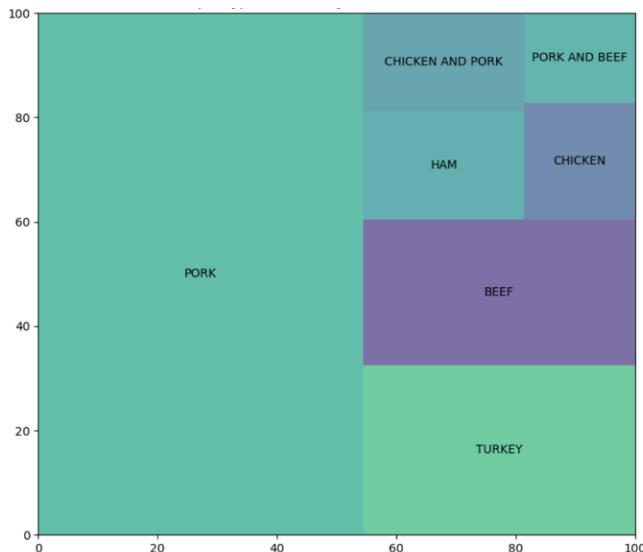
Data Understanding

understanding and preparing our data

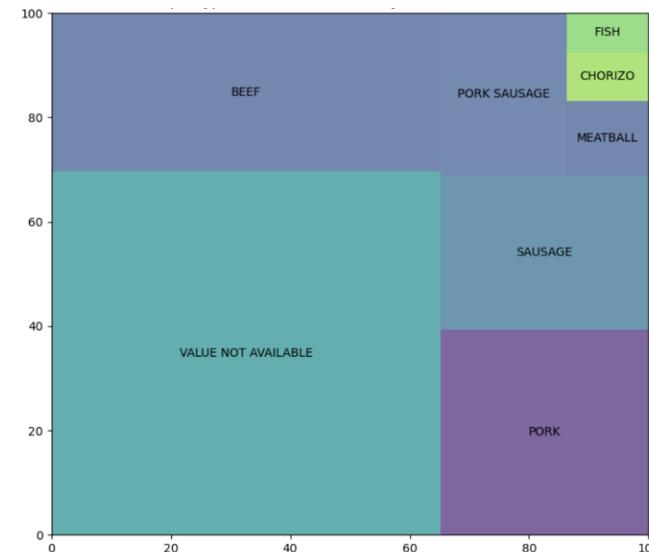


Distribution of Sales for Meat Segments

Source: [supermarketnews.com](https://www.supermarketnews.com/article/2022/07/06/alternative-meat-sales-grow-10-in-q2)

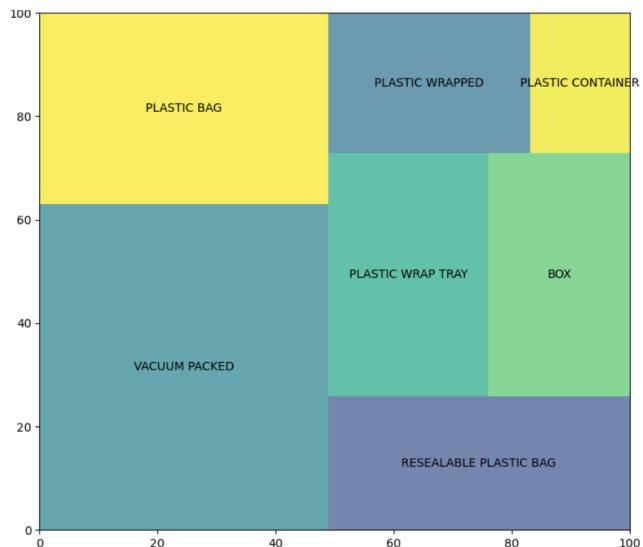


Real Meat

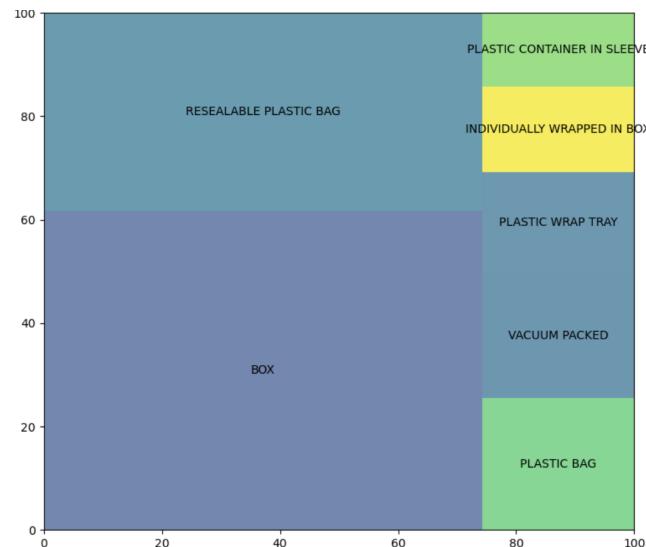


Alternative Meat

Distribution of Packaging for Meat Segments



Real Meat

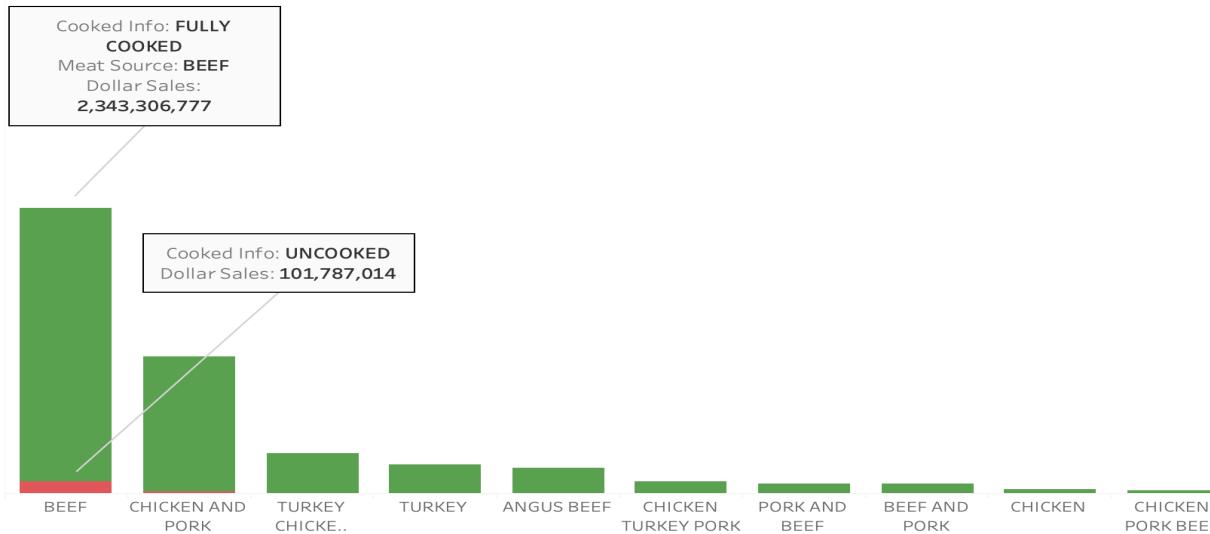


Alternative Meat

Product Sales by Cooking Info & Product Category



FULLY COOKED
UNCOOKED





3

Data Preparation

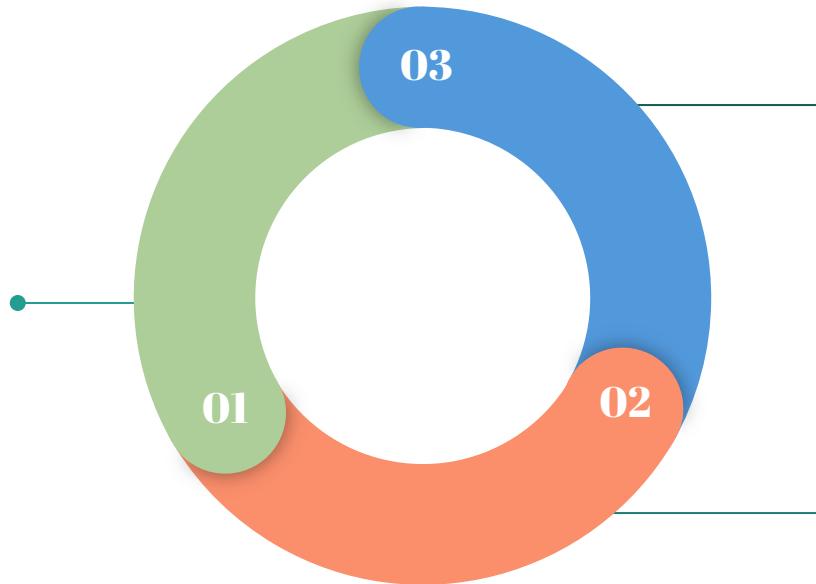
preparing our data for modeling



Our process is easy

Group Data by Product UPC

We grouped data by product UPC so that we can understand the average motivation across each product. Because of grouping data by product UPC, we engineered new features to capture effect of geography



Identifying Multicollinearity

Removed quantitative independent variables with high correlation and qualitative variables that failed the Chi-Squared Test for independence. (Heatmaps in appendix)

Data Standardization & Encoding

We encoded the qualitative variables. Afterwards, we split the data into a train–test split. We standardized the quantitative data in training data set and fit the test data set with the same scale.



4

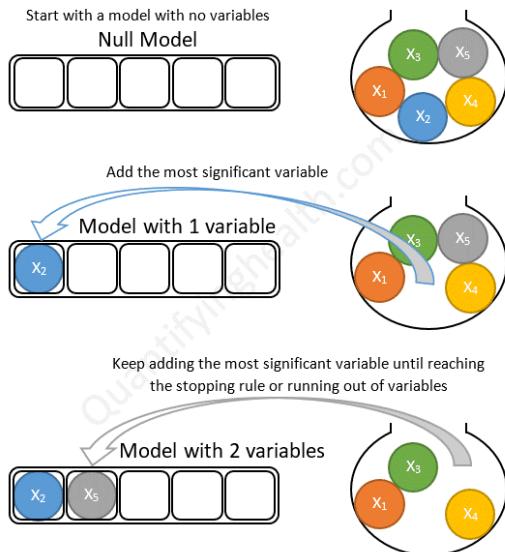
Data Modeling & Inference

building and understanding the models



Our Forward Selection Model

Forward stepwise selection example with 5 variables:



Infographic from quantifyinghealth.com



Grouping

In our forward selection model, we consider encoded variables from one category as one variable (one circle in the box)

Thresholds

Since our goal is to rank the features by importance, we set various thresholds to create rankings for each of the variables

Important Variables Identified From FS

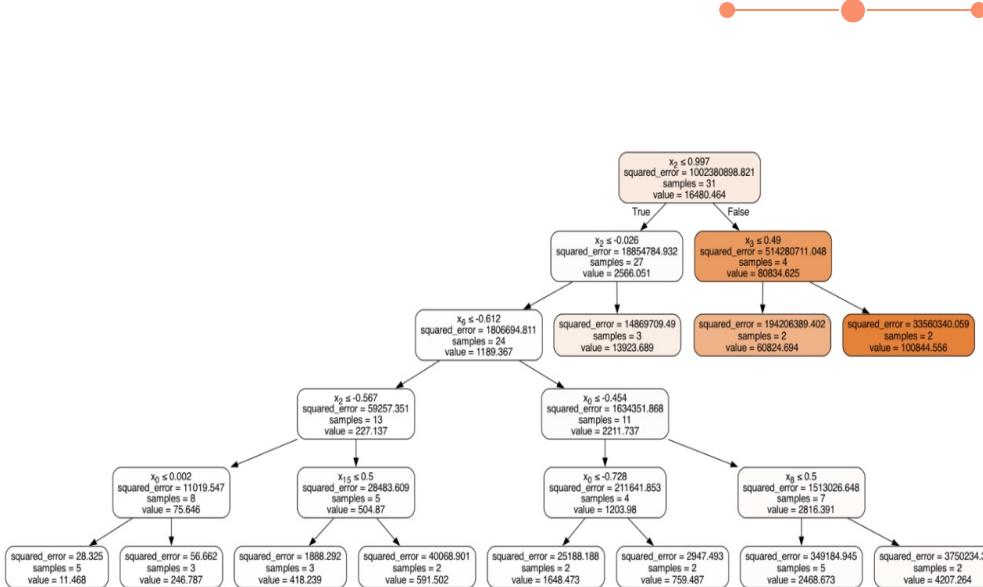
Threshold for Adj R^2:	0.05	0.01	0.005	0.001	0.0005	0.0001
Column Name	ACV Weighted Distribution	Category Name	Cooked Info	Total Ounces	Meat Source	Form
Column Name				Flavor/Scent	Price Per Unit	Package
Column Name					Brand	
					Franchise	
					Name	

Real Meat

Threshold for Adj R^2:	0.05	0.01	0.005	0.001	0.0005	0.0001
Column Name	ACV Weighted Distribution	Price Per Unit	Total Ounces	Package	Type of Substitute	
Column Name	Cooked Info	Popularity			Flavor/Scent	
Column Name						

Alternative Meat

Random Forest Model (with Hyperparameter Tuning)



Tuning For:

- `n_estimators`
- `max_depth`
- `min_samples_split`
- `min_samples_leaf`

Performance:

$$R^2 \approx 0.81$$

Important Variables From Random Forests



Column Name	Importance	Column Name	Importance
ACV Weighted Distribution	0.715	ACV Weighted Distribution	0.9836
Popularity	0.0778	Popularity	0.0154
Price per Unit	0.0704	Price per Unit	0.001
Cooked Info_UNCOOKED	0.0342	Price per Volume	0
Package_VACUUM PACKED	0.0194	Availability Zone Count	0

Real Meat

Alternative
Meat

Recommendations & Limitations

our conclusions and a brief discussion of model
limitations

Our Recommendations

Create alternative meat products that are cooked **like** their real meat counterpart

1

Create more products that have a **similar** quantity to real meat products

2

Increase the total number of regions a product is available

4

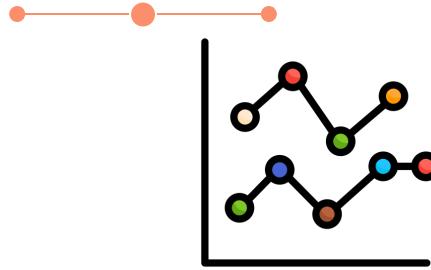
Create packaging that is **like** real meat while emphasizing the health benefits associated

Model Limitations



Lower # of Observations Products in Alternative Pork Model

- Model will have higher variance and less reliability since n is low



Correlation Is Not Causation

- Identified trends and established recommendations based on observed patterns
- Important to establish causality through further testing