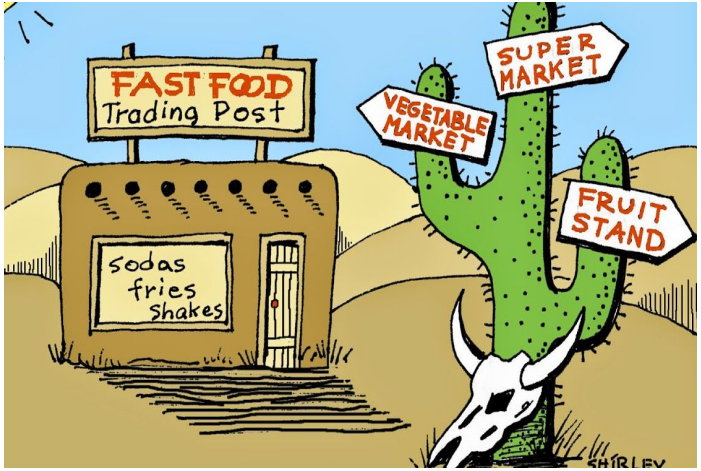
Food Desert Prediction Using Population Health and Social Media Sentiment

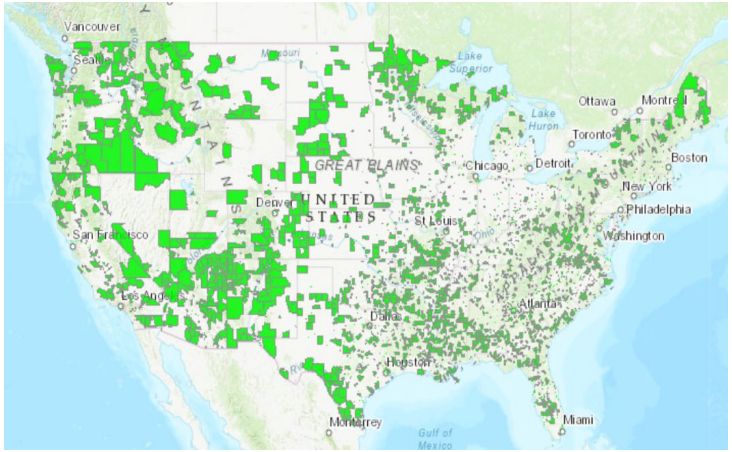
Karina Patel

Github: karinapatel

Project Motivation and Scope

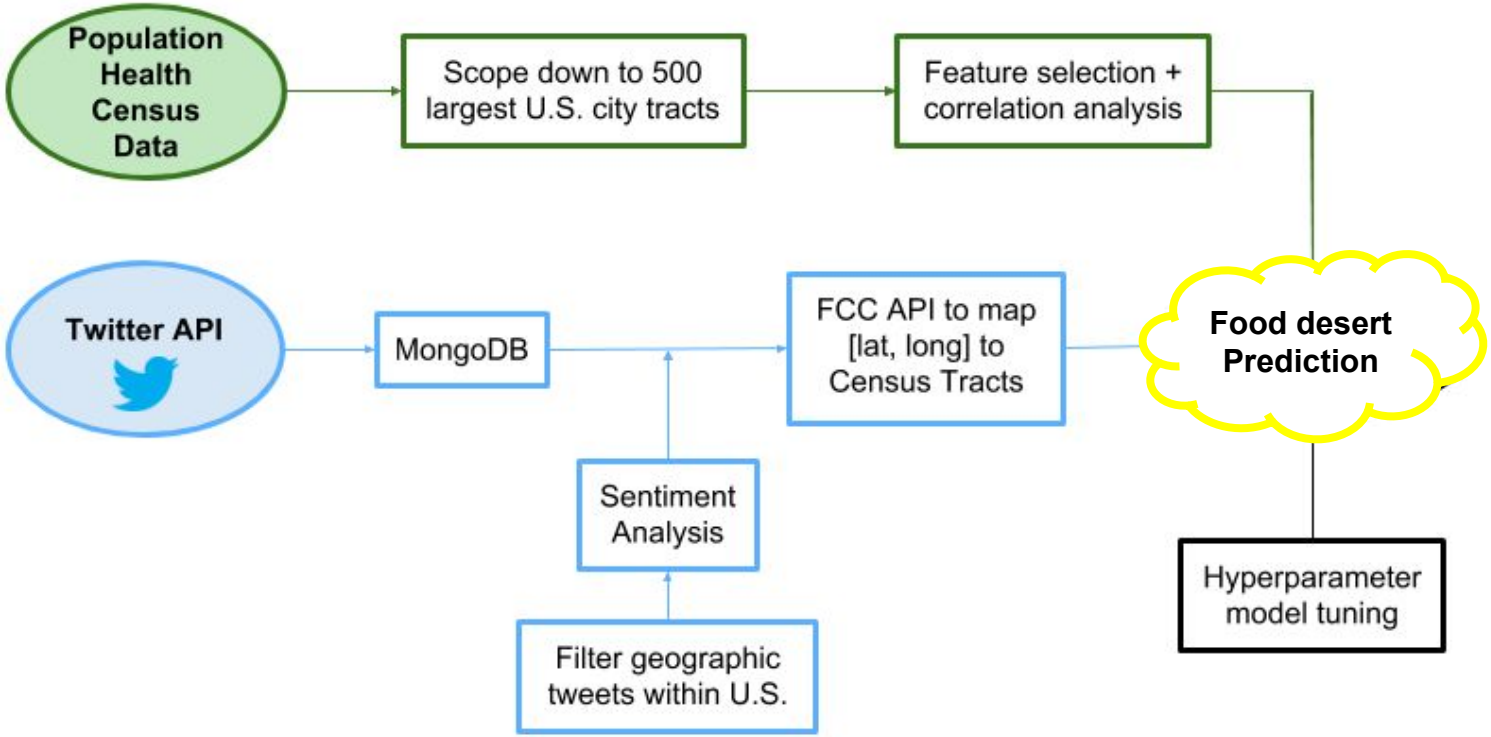
- **Food deserts:** Low income census tracts where over 33% of residents live over a mile from access to fresh groceries


FAST FOOD
Trading Post
sodas
fries
shakes
VEGETABLE MARKET
SUPER MARKET
FRUIT STAND


Vancouver
Seattle
Lake Superior
Ottawa
Montreal
Lake Huron
Toronto
GREAT PLAINS
Chicago
Detroit
Boston
New York
Denver
UNITED STATES
Philadelphia
St Louis
San Francisco
Washington
Los Angeles
Dallas
Atlanta
Houston
Monterrey
Miami
Gulf of Mexico

- **Goal:** Provide up to date predictions to allow for preventative action
  - Grocery store implementation
  - Fresh food initiatives
- Uncover health factors and sentiments correlated with food deserts

Data Collection / Pre-Processing


Population Health Census Data
Scope down to 500 largest U.S. city tracts
Feature selection + correlation analysis
Twitter API
MongoDB
FCC API to map [lat, long] to Census Tracts
Food desert Prediction
Sentiment Analysis
Filter geographic tweets within U.S.
Hyperparameter model tuning

Twitter Data Term Prevalence

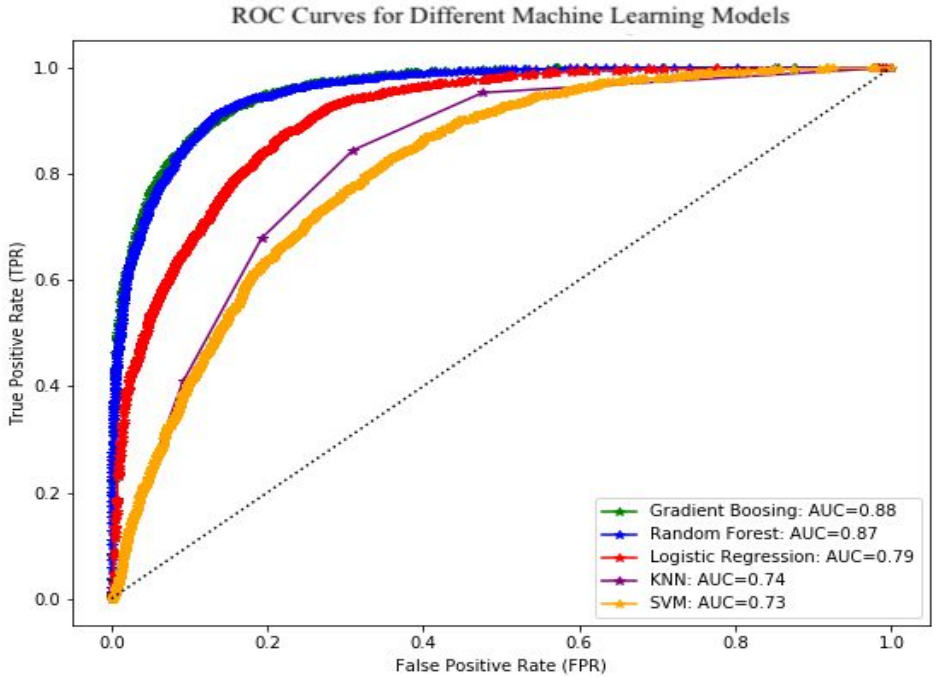

HEALTHYFOODS
brusselssprouts
banana
vegetable
avocado
beef
zucchini
kebab
fruit
cauliflower
mango
peanutbutter
apple
omelet
oatmeal
fish
celery
lettuce
grapes
chicken
milk
yogurt
bellpepper
tofu
quinoa
oats
almond
lentils
lamb
kale
carrot
melon
mushroom
asparagus
orange
pork
meat
spinach
corn
cherry
broccoli
cucumber
salami
blueberry
rice
turkey
eggplant
raspberry
bean
hummus
egg


FASTFOODSTORES
jimmyjohn's
wendy's
jambajuice
steaknshake
domino's
culver's
jackintheboxjason'sdeli
pizzahut
littlecaesars
mcdonald's
fiveguys
jerseymike's
checkers
papajohn's
in-n-out
kfc
elpolloloco
chipotle
subway
mcalister'sdeli
tacobell
starbucks
dunkindonuts
auntieanne's
firehousesubs
sonic
pandaexpress
timhortons
panerabread
bostonmarket
arby's
popeyeslouisianakitchen
carl'sjr
whitecastle
moe'ssouthwestgrill
zaxby's
whataburger
baskin-robbins
bojangles
deltaco
qdobamexican
papamurphy's
dairyqueen
wingstop
burger

Modeling Results

Gradient Boosting produced the strongest results:

Metric	Score
ROC-AUC Score	0.88
Accuracy	0.89
F1 Score	0.84
Recall	0.83
Precision	0.86


ROC Curves for Different Machine Learning Models
True Positive Rate (TPR)
False Positive Rate (FPR)
0.0
0.2
0.4
0.6
0.8
1.0
Gradient Boosing: AUC=0.88
Random Forest: AUC=0.87
Logistic Regression: AUC=0.79
KNN: AUC=0.74
SVM: AUC=0.73

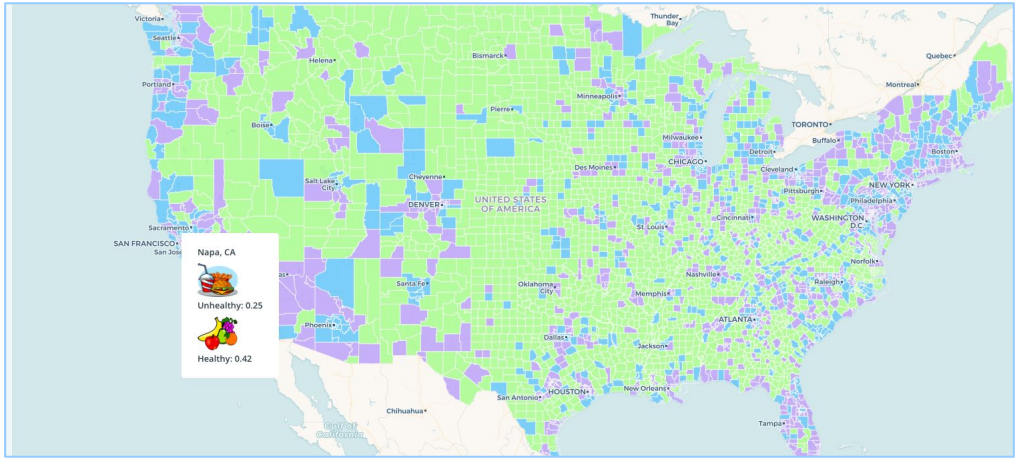
The strongest contributors to my model were health checkups, obesity, tract composition, and healthy food sentiment

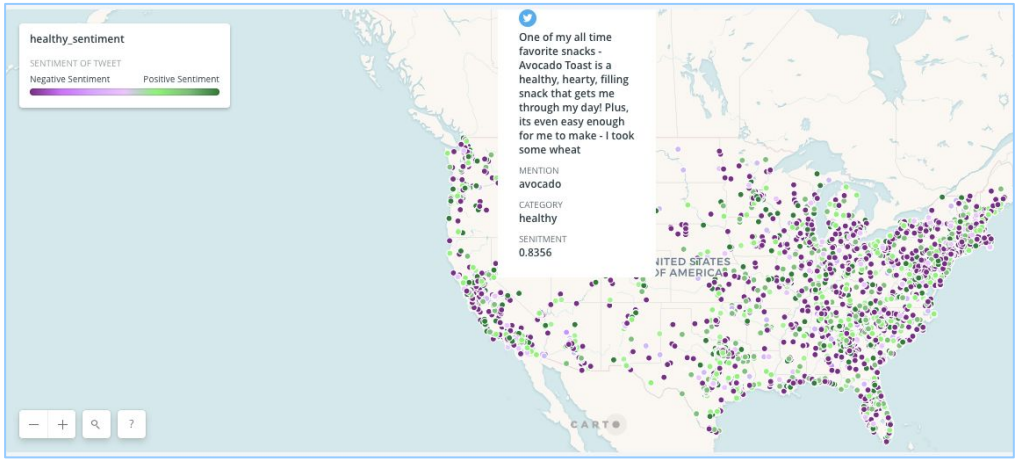
Flask Interactive Web App

[Link to webapp]

Visualized my Twitter geo data using Carto and added the following maps to my web app:

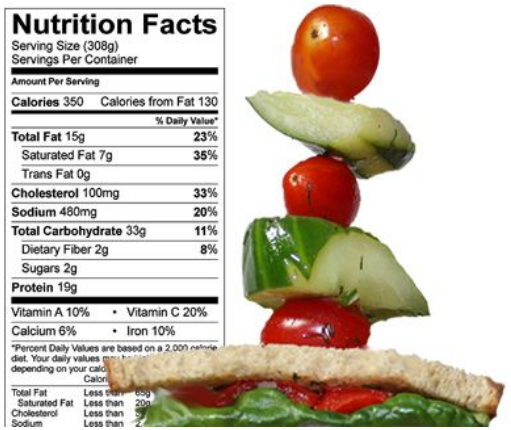
- Tweet map
- County level sentiment map
- State level county map


Napa, CA
Unhealthy: 0.25
Healthy: 0.42
UNITED STATES OF AMERICA
DENVER
CHICAGO
TORONTO
NEW YORK
WASHINGTON D.C.
ATLANTA
HOUSTON
SAN FRANCISCO


healthy_sentiment
SENTIMENT OF TWEET
Negative Sentiment
Positive Sentiment
One of my all time favorite snacks - Avocado Toast is a healthy, hearty, filling snack that gets me through my day! Plus, its even easy enough for me to make - I took some wheat
MENTION
avocado
CATEGORY
healthy
SENTIMENT
0.8356
CARTO

Limitations / Next Steps

1. Estimate nutrition of tweets using the Food Composition API
    - Break down each tweet into nutritional components: fat, carbs, protein


Nutrition Facts
Serving Size (308g)
Servings Per Container
Amount Per Serving
Calories 350 Calories from Fat 130
% Daily Value*
Total Fat 15g 23%
Saturated Fat 7g 35%
Trans Fat 0g
Cholesterol 100mg 33%
Sodium 480mg 20%
Total Carbohydrate 33g 11%
Dietary Fiber 2g 8%
Sugars 2g
Protein 19g
Vitamin A 10% • Vitamin C 20%
Calcium 6% • Iron 10%

2. Update web app maps automatically when new data is pulled


mongoDB
CARTO

Thank you!

Questions?