

Determinants of the presence, density, and popularity of US food retailers

Syed Fuad

Agricultural & Applied Economics Association Annual Meeting,
Washington DC: July 23-25, 2023

Background

- Food store location shown to have a strong relationship with a variety of demographic and neighborhood attributes.
 - Access to grocery stores is strongly associated with child obesity (Li et al., 2021), mineral deficiency (Cole et al., 2010; Skalicky et al., 2006), tooth decay (Chi et al., 2014), mental health (Melchior et al., 2012; McIntyre et al., 2013) and overall health outcomes (Egeland et al., 2011)
 - Access and presence of food stores are related to food habits (Dubowitz et al., 2015), income (Allcott et al., 2019; Wolfson et al., 2019) economic constraints (Kato and McKinney, 2015), and their interaction (Breyer and Voss-Andreae, 2013)
 - Relation between racial composition, housing segregation, and food store availability (Galvez et al., 2008; Wilcox et al., 2020; Havewala, 2021) and healthy food access (Shannon, 2021)

Motivation

- Comprehensive study on the sociodemographic and relative locational determinants of food store types, and the ability of the same determinants to predict store location
- Recent advancements in geospatial data mining, high resolution demographic data procurement combined with improved computational power and efficient learning models enable us to extract these insights from data

Caveats

- Not a study on:
 - Factors that would affect retailers' or consumers' decisions
 - Understanding consumer or retailer behaviors
- Computational exercise on what predict/determine food store location, density and popularity

Data – Explanatory variables

- US census tract-level explanatory data from the American Community Survey (ACS) 2006-10 5-year estimates
 - Race, unemployment, education, mode of transportation to work, median household income, poverty rate, vehicle ownership, pop. density, etc.
- Other variables at the census tract level that characterize neighborhood amenities, accessibility facilities and relative location efficiency:
 - Walkability index score; transit and auto network density; employment density; food sales tax; crime, etc.
- Data covers 57,463 census tracts ~ 82% of US population and over 70% of land area in the contiguous states

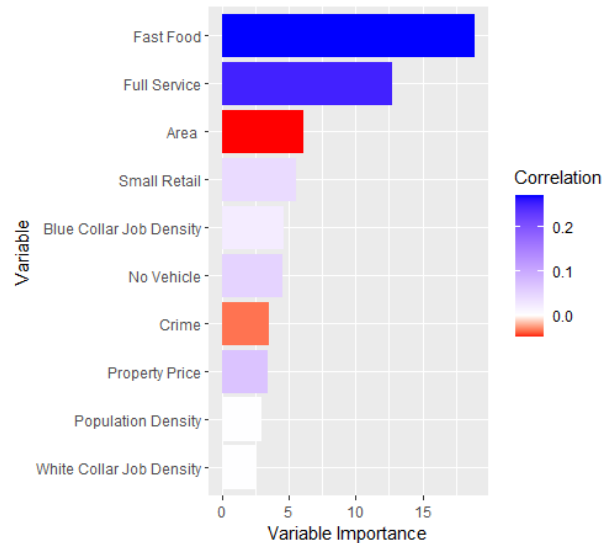
Data – Response variable

- Anonymized mobile phone GPS point of interest (POI) data
 - Visitor aggregations for POIs including aggregated raw counts of and visits to POIs using GPS ‘pings’, the frequency of visits, category of the POI in the North American industry classification system (for example, large retailer, small retailer, fast food or full-service restaurant)
- Three models based on the nature of the response variables:
 - (1) response variable = binary variable indicating whether the particular food retailer type exists in the census tract [gradient boosted logit]
 - (2) response variable = count number of food retailers of a particular type in a census tract [negative binomial]
 - (3) response variable = count number of customers who visited the particular food store type in a census tract [negative binomial]

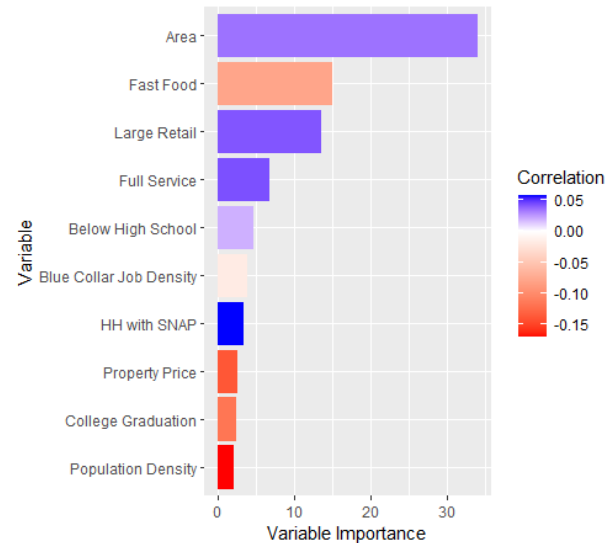
Prediction of food store location

| Dep. variable: store presence | | Large retail | Small retail | Fast food | Full service |
|-------------------------------|--------------|----------------|----------------|----------------|----------------|
| Boosted logit | Accuracy (%) | 72.15 | 92.04 | 77.90 | 74.14 |
| | 95% CI | (71.48, 72.82) | (91.63, 92.44) | (77.27, 78.52) | (73.48, 74.79) |
| | Sensitivity | 75.55 | 48.57 | 65.38 | 77.84 |
| | Specificity | 61.63 | 92.13 | 80.45 | 64.07 |
| | N | 17,400 | 17,400 | 17,400 | 17,400 |

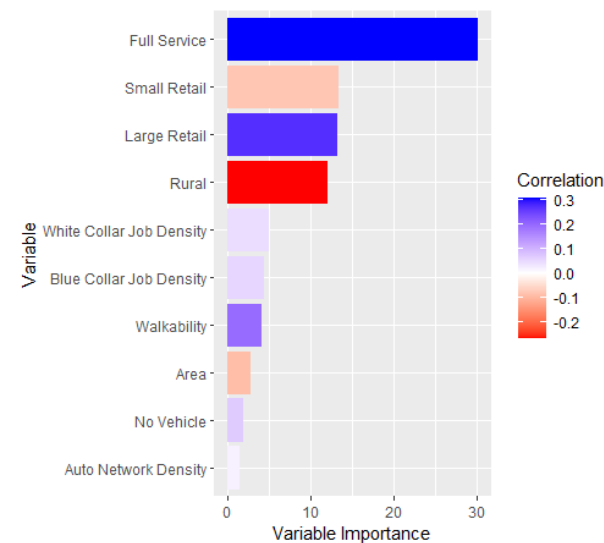
Large retail



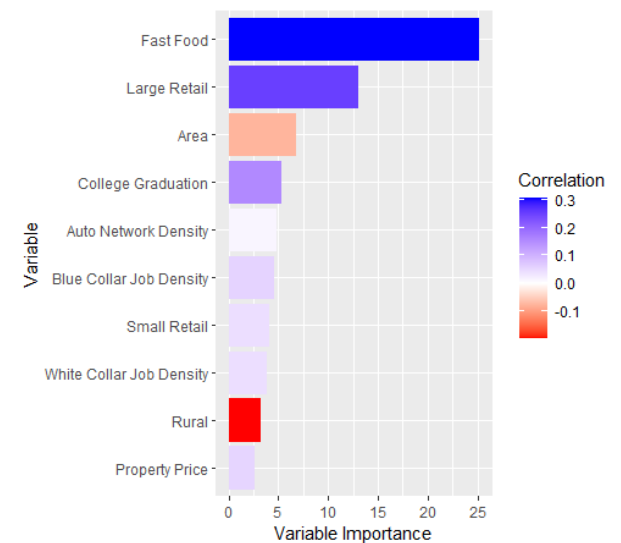
Small retail



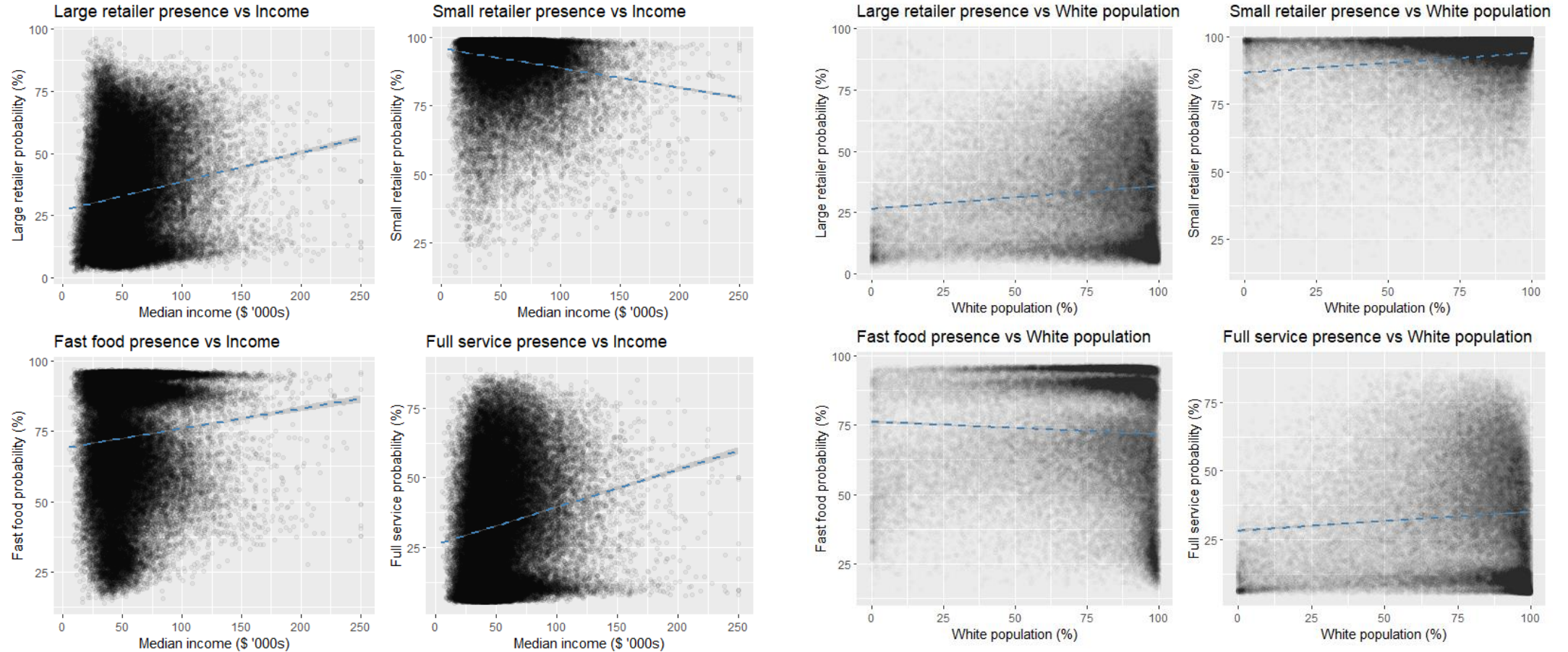
Fast food



Full service



Predicted store probability vs. demographics



Prediction of store count and visits

| | Large retail | | Small retail | | Fast food | | Full service | |
|-----------------|----------------------------|---------------|--------------|---------------|-------------|---------------|--------------|---------------|
| | Panel A: | Panel B: | Panel A: | Panel B: | Panel A: | Panel B: | Panel A: | Panel B: |
| | Store Count | Visitor Count | Store Count | Visitor Count | Store Count | Visitor Count | Store Count | Visitor Count |
| | Food stores | | | | | | | |
| Large Retail | | | + | + | + | + | + | + |
| Small Retail | + | + | | | + | + | + | + |
| Fast Food | + | + | + | + | | | + | + |
| Full Service | - | + | + | + | + | + | | |
| | Income | | | | | | | |
| Poverty Rate | - | - | - | - | + | + | - | - |
| Unemployment | - | - | + | + | + | + | - | - |
| Below HS | - | - | - | - | - | + | - | - |
| Property Value | + | + | - | - | - | - | - | - |
| | Location and Access | | | | | | | |
| Pub. Transport | - | - | - | - | - | - | - | - |
| Area | - | - | - | - | - | - | - | - |
| Rural | - | - | + | + | - | - | - | - |
| Walkability | + | + | + | - | + | + | + | + |
| Transit Density | + | + | + | + | + | + | + | + |
| | Demographics | | | | | | | |
| Black | - | - | + | + | - | - | - | - |
| Hispanic | - | - | + | + | + | + | + | + |
| Asian | - | - | + | + | + | + | + | + |
| | Policy | | | | | | | |
| Sales Tax | - | - | + | + | + | + | + | + |
| BC Job Density | + | - | + | + | + | + | + | + |
| WC Job Density | - | - | - | - | + | + | - | + |
| Crime | - | - | - | + | - | - | + | + |

Summary

- Density and popularity of every store type is consistently and significantly related to the presence, density, and popularity of other store types.
- Low-income neighborhoods have lower presence of and lower foot traffic to large retailers, and full-service restaurants.
 - Residents in these neighborhoods are more likely to visit fast food stores.
- Neighborhoods with higher ethnic diversity are less likely to visit large retailers.
 - Hispanic and Asian neighborhoods have greater presence, counts and visits to small retailers, fast food and full-service restaurants, but have lower counts of and visits to large retailers.
 - Black neighborhoods tend to have lower presence, counts and visits to large retailers, fast food and full-service restaurants, and higher counts of and visits to small retailers.

Addendum - Analysis setup

- Data randomly split into training (70%) and testing (30%) samples
- Ten-fold cross-validation during training: the training set is randomly and iteratively split into ten equal-sized subsamples
 - From these ten subsamples, one subsample is used as a hold-out set and the remaining nine subsamples are used as training sets for model fitting.
- Grid search conducted over a range of values of hyperparameters; shrinkage (λ) ranges from {0.01, 0.05, and 0.10}; split nodes range from {1, 5, and 10}; number of trees range from {100, 300, and 500}.