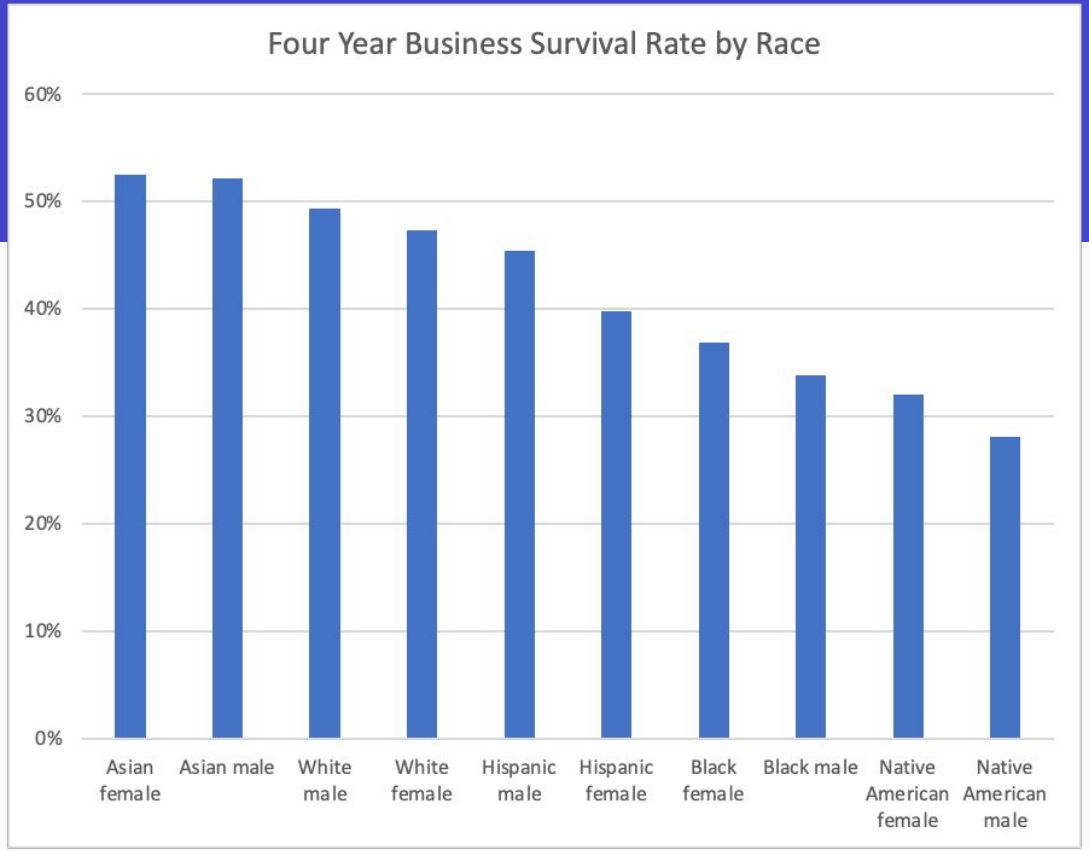
Predicting Minority Status of Boston Women-Owned Businesses

Arlyss, Anushka, Marco, Deekshith
Tufts University, DATA 200
Final Presentation
12/6/2023

Background

Problem

Women and minority owned business have a lower long term survival rate than white, male businesses. This is exacerbated by intersectionalities and spatial barriers.



Robb, Alicia M. “Entrepreneurial Performance by Women and Minorities: The Case of New Firms.” Journal of Developmental Entrepreneurship 7, no. 4 (2002): 383-.

Background

Problem

Women and minority owned business have a lower long term survival rate than white, male businesses. This is exacerbated by intersectionalities and spatial barriers.

Gap

Lack of predictive measures to preemptively provide economic support or financial services

Motivation

Support and promote women-owned businesses economically, and provide extra economic support to women of color

Research question and hypothesis

- **Research question:** How are business types, location, and socioeconomic status of the surrounding area correlated with whether a women-owned business is also minority owned?
- **Hypothesis:** Minority-owned businesses are clustered in areas of relatively lower socioeconomic status.
- **Hypothesis:** Service and retail businesses tend to be owned predominantly by minorities.

Data Overview

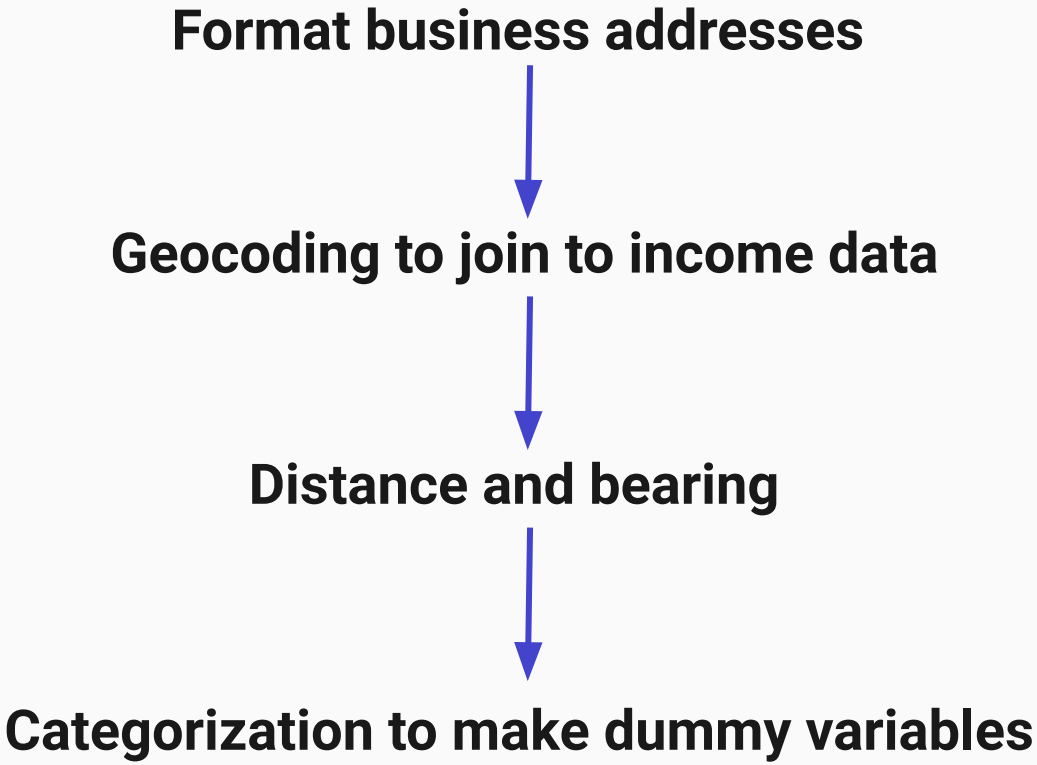
Women-Owned Businesses

- From Boston government
- Information on name, location, contact, business type, demographic information
- 195 non duplicate entries

Median income

- US Census Bureau 2021 5 Year ACS
- Median income by census tract
- Subset to census tracts with women-owned businesses in Greater Boston area

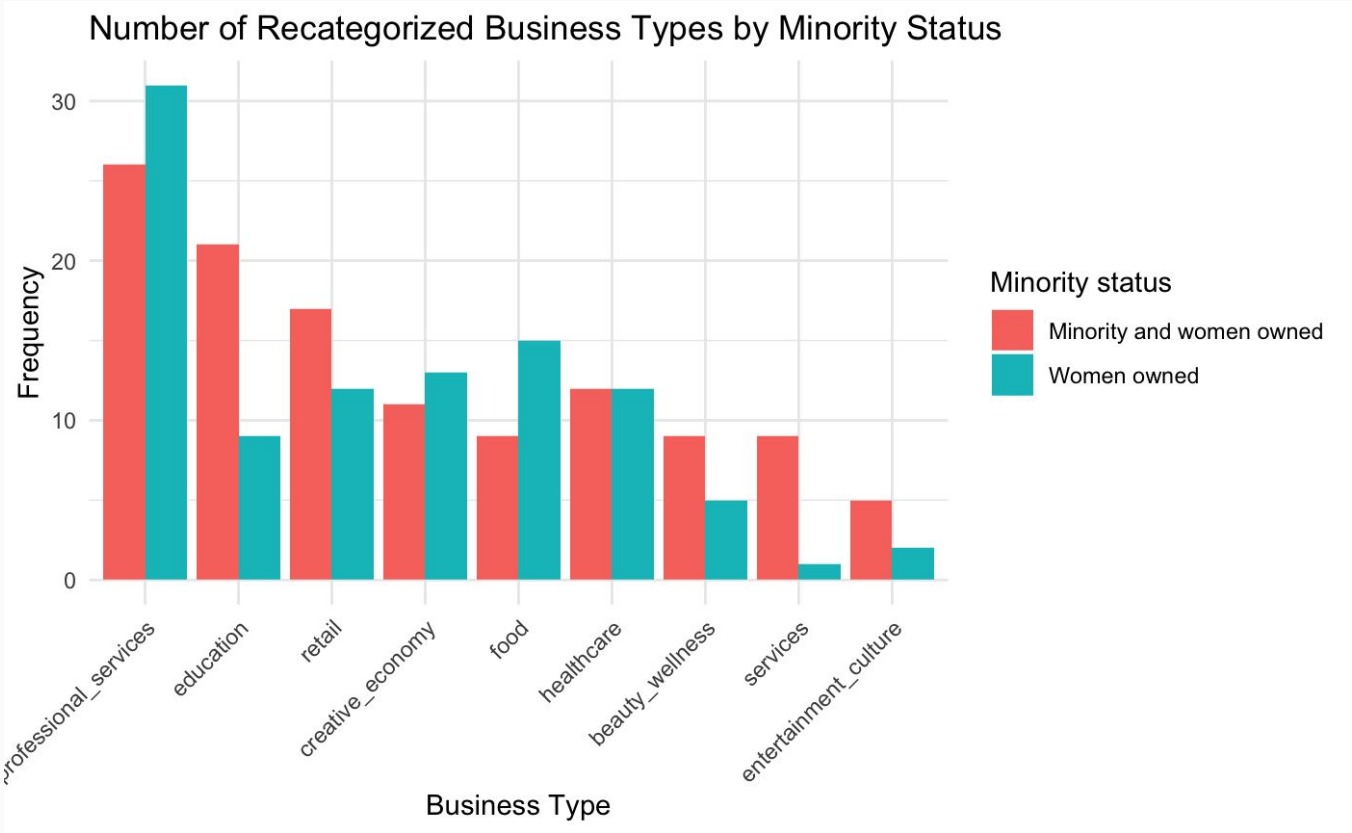
Data processing



Methods

- Exploratory data analysis (EDA)
- Assess EDA takeaway
- Model selection
- Models
- Results and conclusions

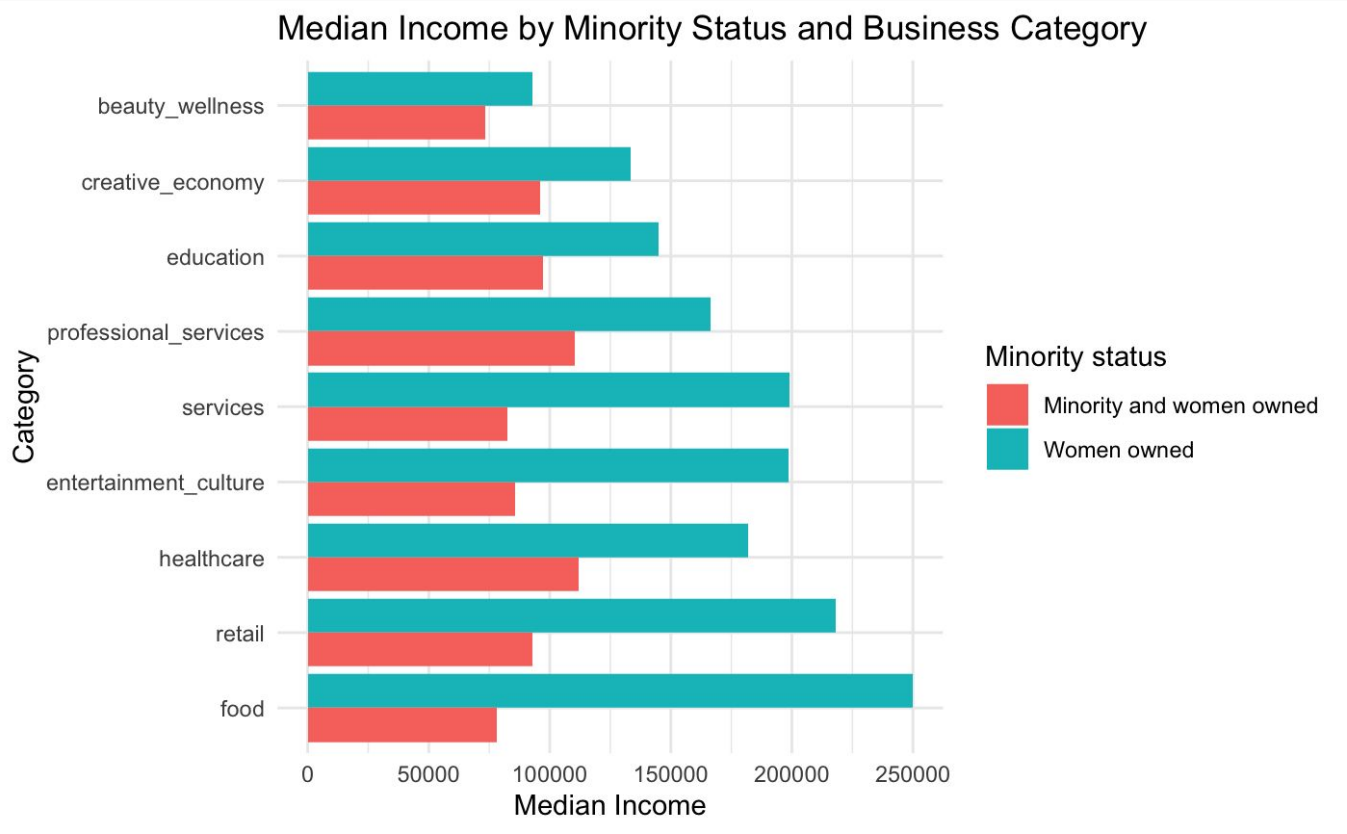
Exploratory Data Analysis



Goal: See how business types are distributed by minority status

Takeaway: Education, retail, beauty and wellness, service, and entertainment businesses are owned substantially more by minorities.

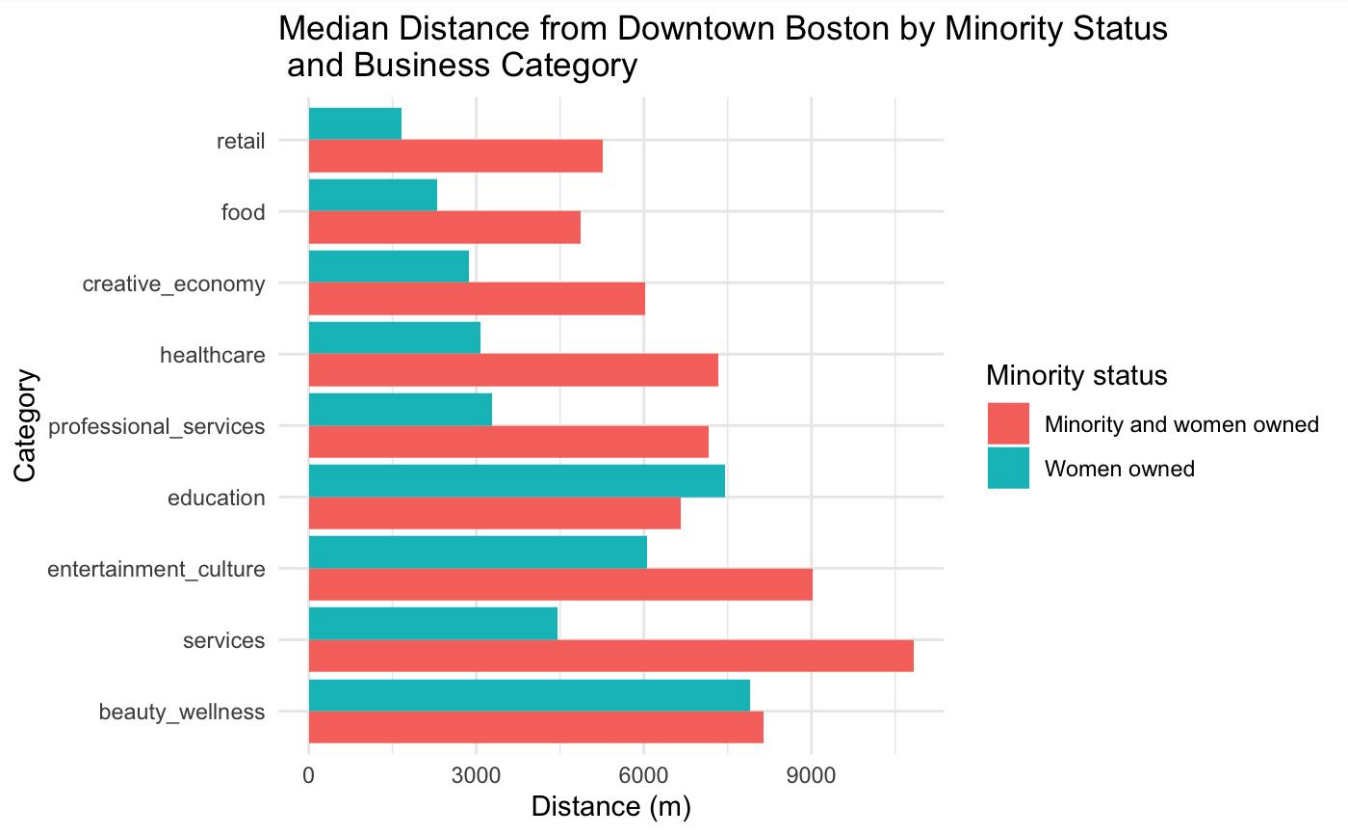
Exploratory Data Analysis



Goal: See how median income varies by business type and minority status

Takeaway: Minority businesses are overwhelmingly located in lower income areas across all business types

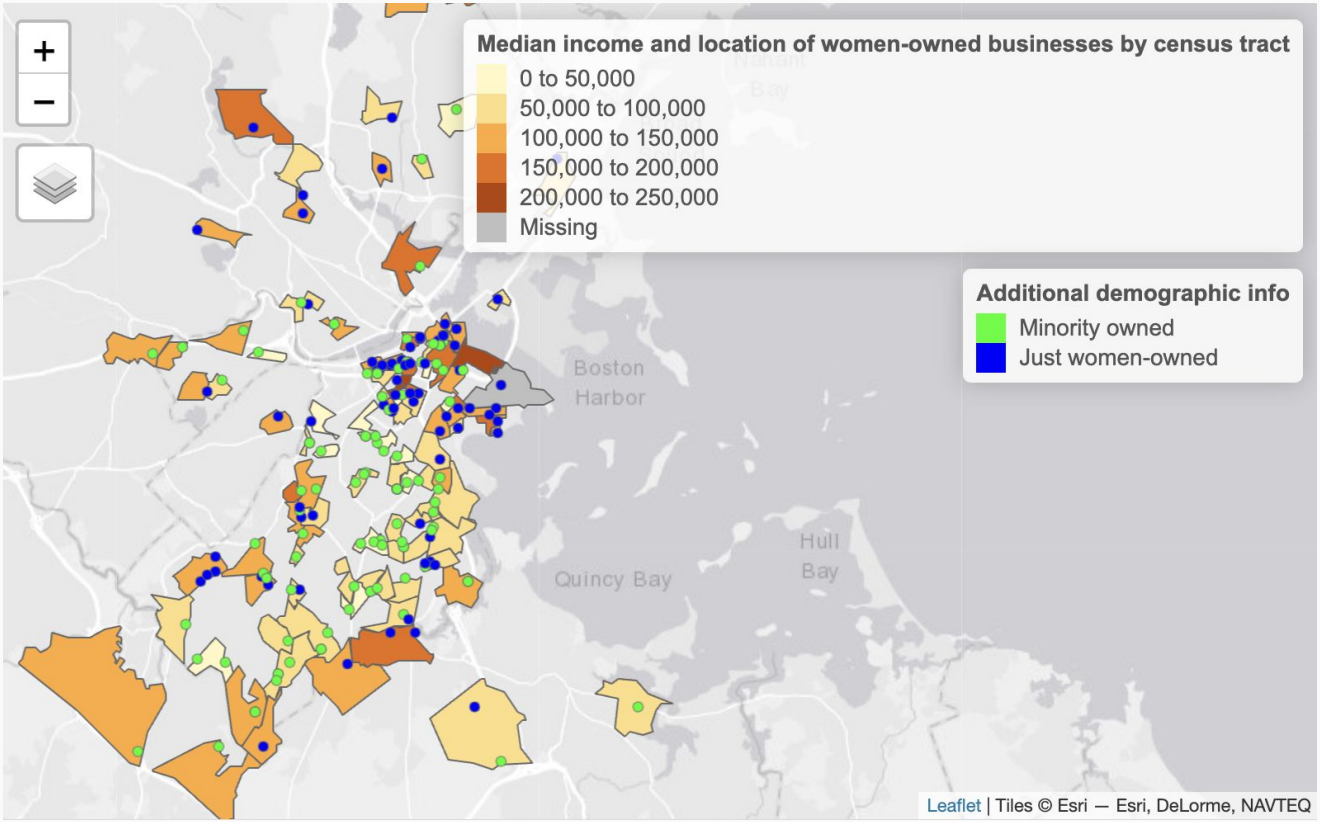
Exploratory Data Analysis



Goal: See how median distance from downtown Boston varies by business type and minority status

Takeaway: Minority businesses, except in education and beauty fields, are overwhelmingly located further from downtown Boston

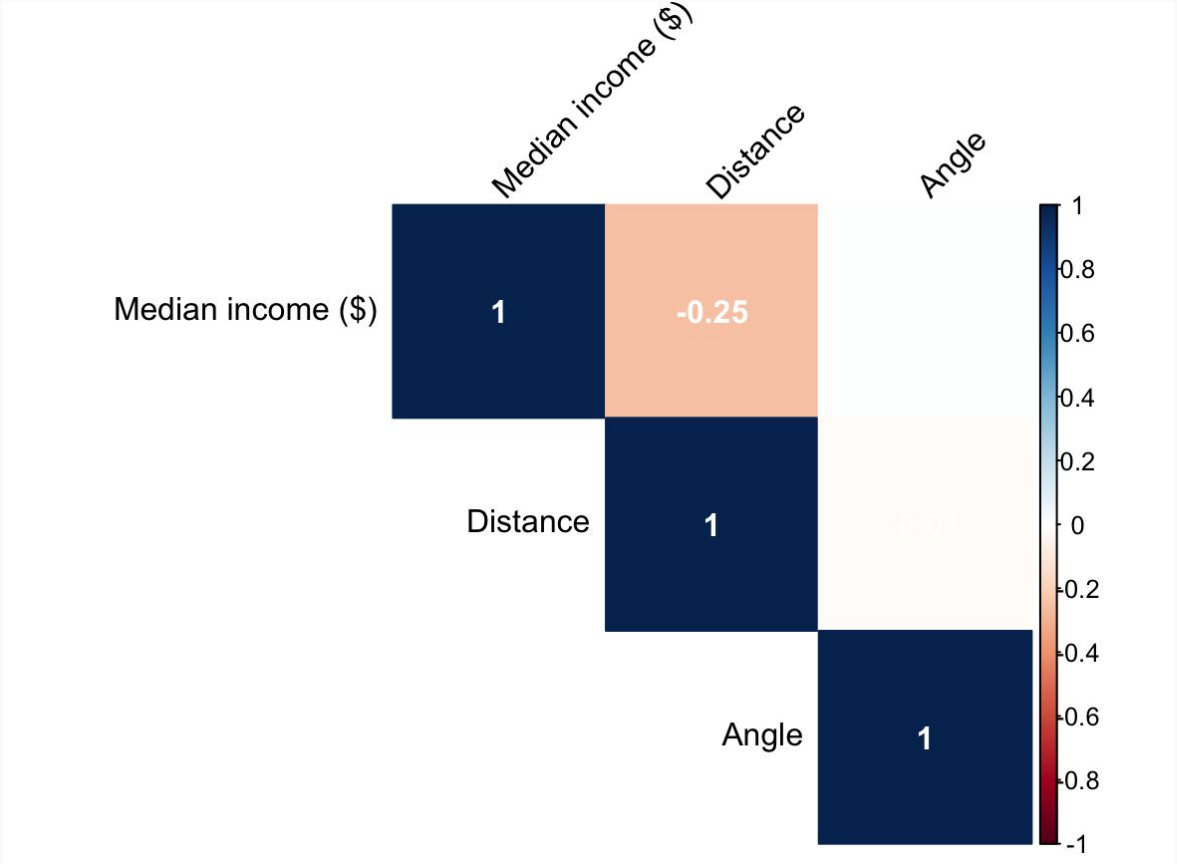
EDA: Correlation of location and income



Goal: Visual distribution of median income of census tracts and minority status of businesses

Takeaway: Non-minority businesses are proportionally more in downtown Boston, in wealthier areas

EDA: Correlation of location and income

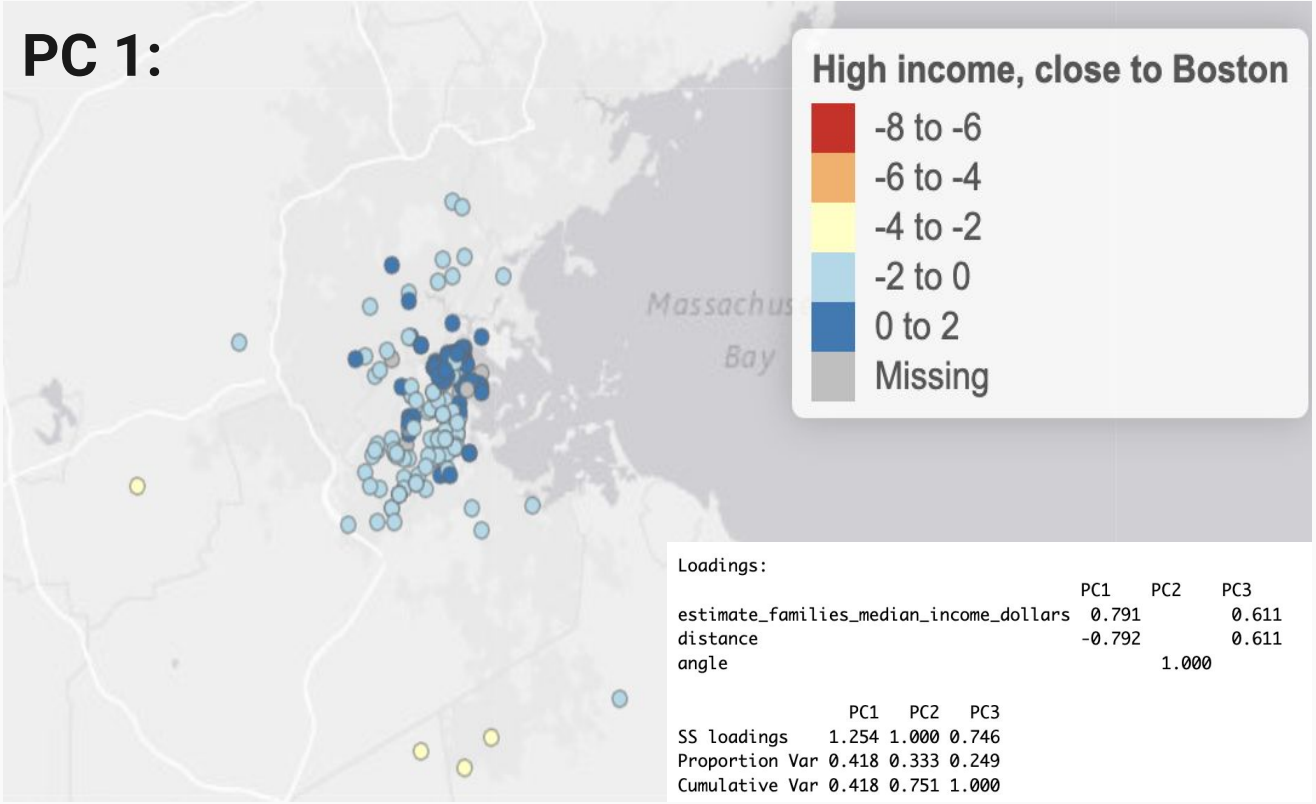


Goal: Correlation between median income, distance, and angle

Takeaway: Only substantial correlation is that distance and median income are inversely proportional

Principal component analysis

PC 1:




```
Loadings:
                                        PC1    PC2    PC3
estimate_families_median_income_dollars  0.791          0.611
distance                                -0.792          0.611
angle                                           1.000

                 PC1   PC2   PC3
SS loadings    1.254 1.000 0.746
Proportion Var 0.418 0.333 0.249
Cumulative Var 0.418 0.751 1.000
```

Goal: Spatial correlation between median income, distance, and angle

Takeaway: Higher income areas are close to Boston, explain the most variance

EDA takeaways

- Minority owned businesses across all business types are located in lower income areas
- Minority owned businesses are predominantly located further from Boston than non-minority owned businesses
- Businesses are more frequently owned by minorities in the fields of education, retail, beauty and wellness, service, and entertainment

Models:

Objective

- Predict whether a women-owned business in Boston is minority-owned

Predictors

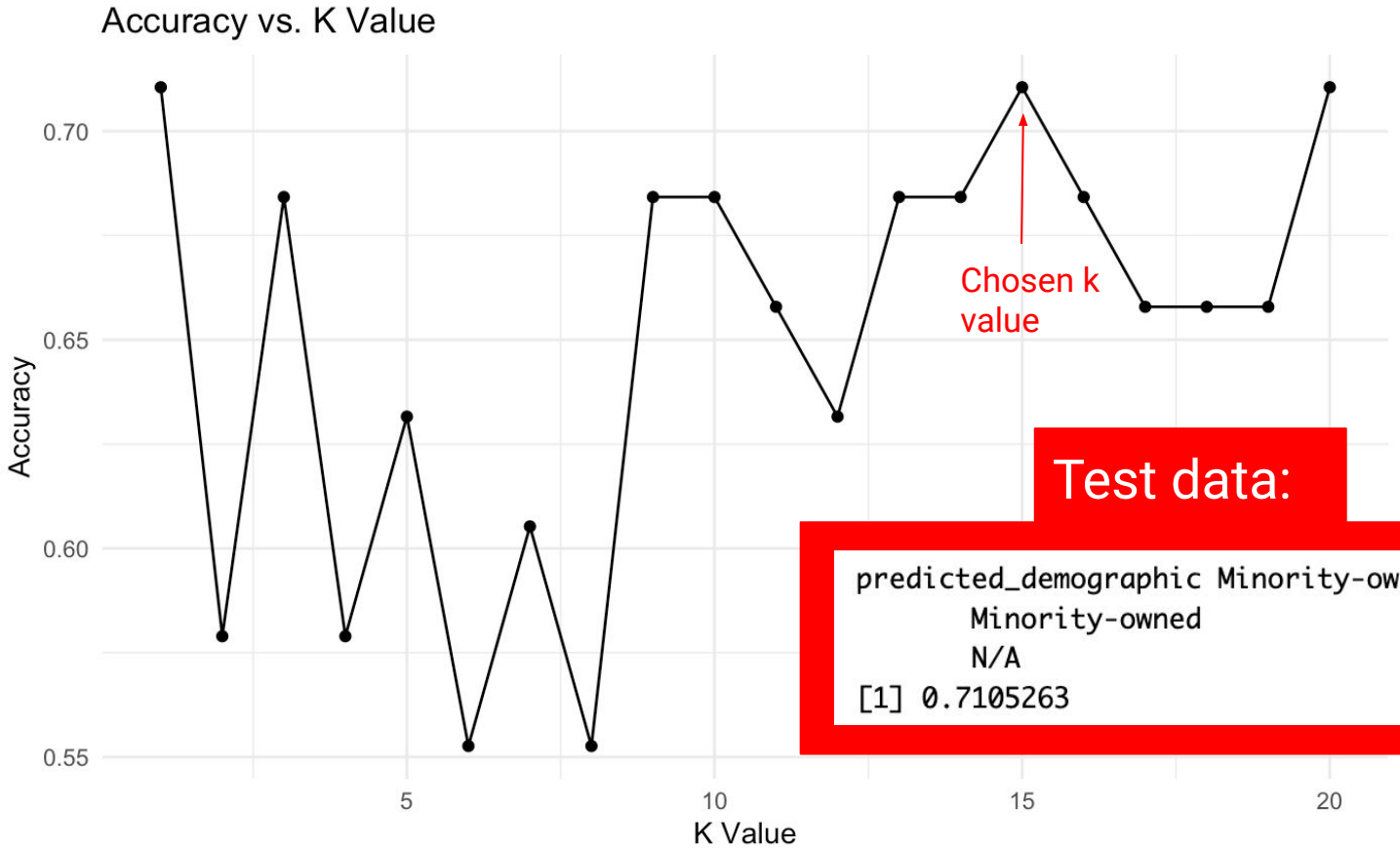
- Median income
- Distance and bearing
- Nine business type dummy variables (not used in kNN)

Models

- kNN
  - Businesses with similar location and income would have similar minority status
- Random forest
  - Binary splitting, can handle dummy categorical and continuous data
- Logistic regression
  - Model binary outcomes and take in dummy variables and continuous data

K-Nearest Neighbors (kNN): 71.1% accuracy

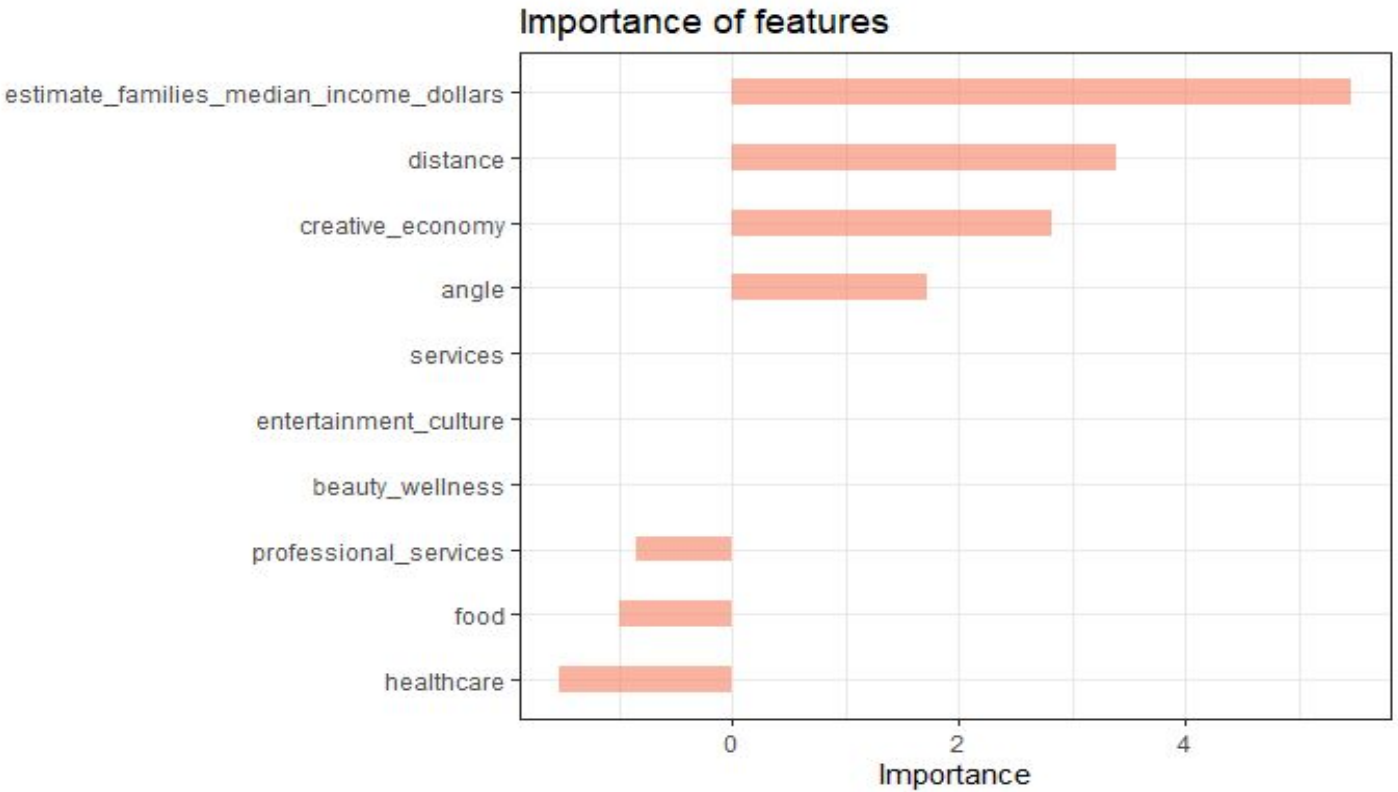
Training data:



Test data:

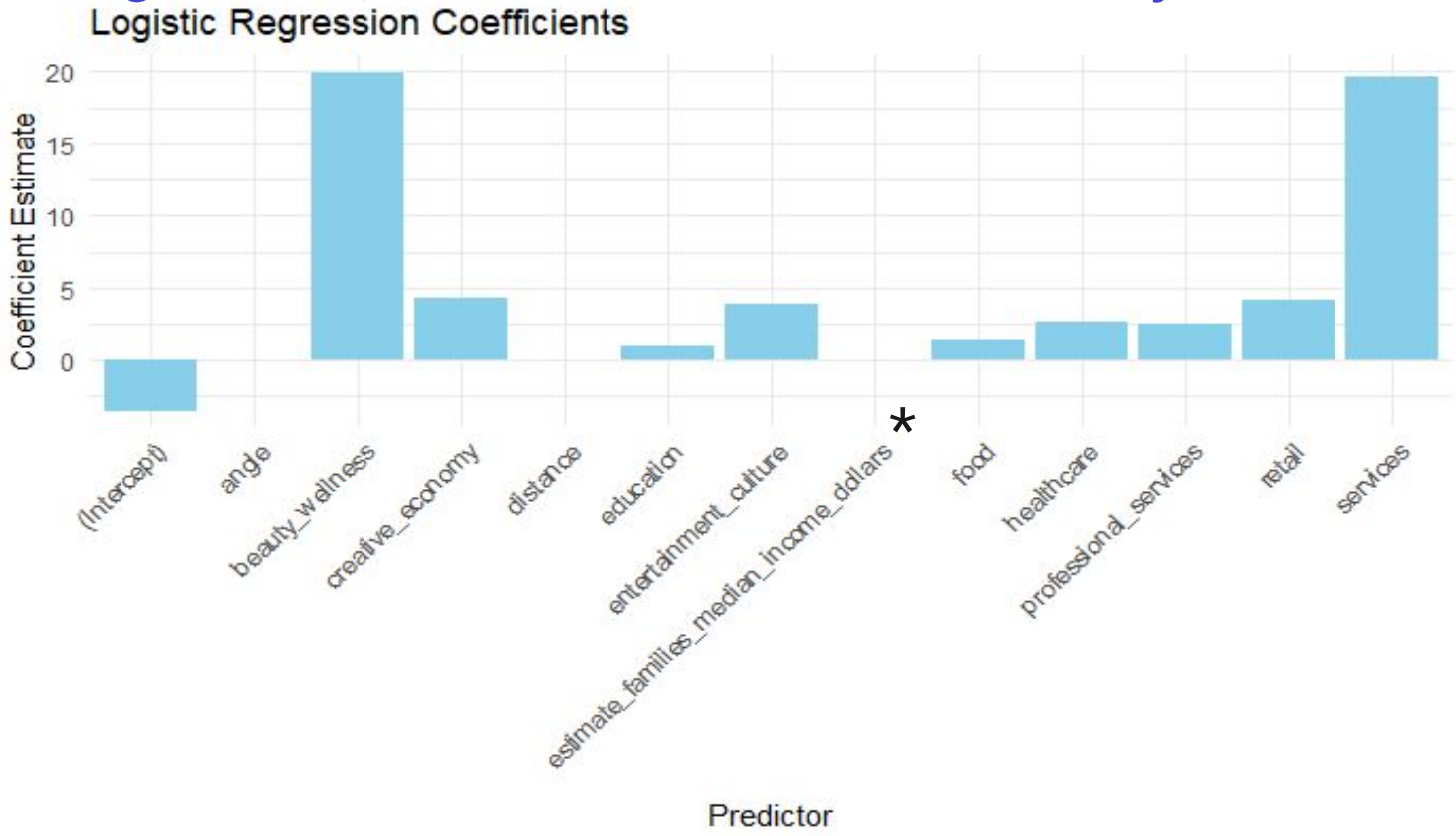
```
predicted_demographic Minority-owned N/A
       Minority-owned                17   4
       N/A                            7  10
[1] 0.7105263
```

Random forest: 60.5% accuracy



Importance as measured by the amount the accuracy would decrease if the variable were to be excluded. Median income and distance are the most important, whereas many business types have a net zero or net negative effect on the model.

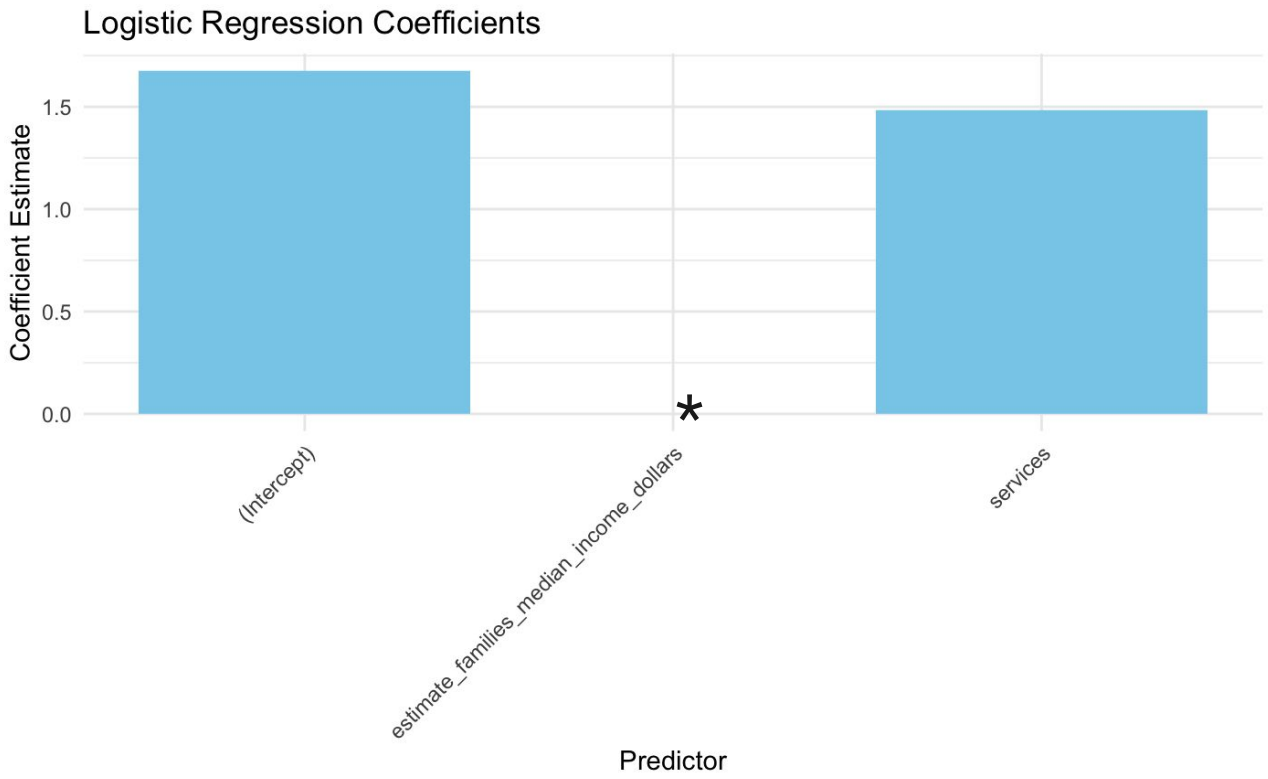
Logistic regression, all variables: 66.5% accuracy



This graph is used to understand which factors are most influential in predicting the outcome and the nature of their impact.

* statistically significant

Logistic regression, forward selection: 65.4% accuracy



*** statistically significant**

Results

kNN

71.1% accuracy

- Most accurate
- Similar location and income good predictors

Random forest

60.5% accuracy

- Least accurate
- Business type not more predictive, most decrease accuracy

Logistic regression, forward selection

65.4% accuracy

- Mid-accuracy
- Only uses income and service work

Conclusion

- Policymakers can use the location and median income to predict if a women-owned business is minority owned

  ➡ Provide additional resources and economic support

- Business types is not the most accurate indicator to predict whether a business is minority-owned
  - Income and distance from center of Boston best predictors

Thank you!

Limitations

Lack of continuous data

- Most data was categorical

Small dataset

- Only 188 businesses

Lack of data about men's businesses

- Compare to women's

No disaggregation by race/ethnicity

- Disaggregated race/ethnicity data may have different financial and business implications

Bias in center of Boston

- Subjective choice on center of Boston for distance

Bias in business categorization

- Different categories may lead to different results

Future work

- Continuously update the model when new data becomes available.

- Compare the model's performance with other similar but advanced models in order to create the best possible model.

- Use new method to handle new dataset, try to implement new method and coding skills to handle data cleaning and missing dataset.

- Look at other gender and race intersections

Sources

Picture:

https://designpickle.com/creative-hub/graphic-design/5-reasons-why-graphic-design-is-important-for-any-business/

Graph:

Robb, Alicia M. “Entrepreneurial Performance by Women and Minorities: The Case of New Firms.” Journal of Developmental Entrepreneurship 7, no. 4 (2002): 383-.