



# Forecasting County-wise Micro-Business Density (MBD)



# Project Goal

- What is a micro-business?
  - Businesses with 10 or fewer employees
  - Has an online presence
- Current Models:
  - Leverage available internal and census data
  - Use econometric approaches

# Parties of Interest

## Commercial Interests

- Shopify
- Turbo Tax
- Office Depot
- MailChimp
- Home Depot

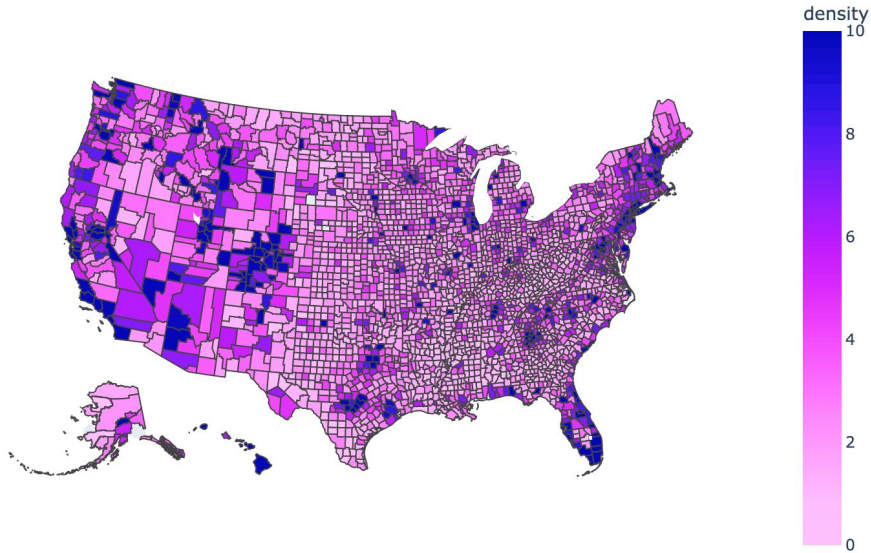
## Academic Interests

- Policymakers
- Economists
- Social Scientists

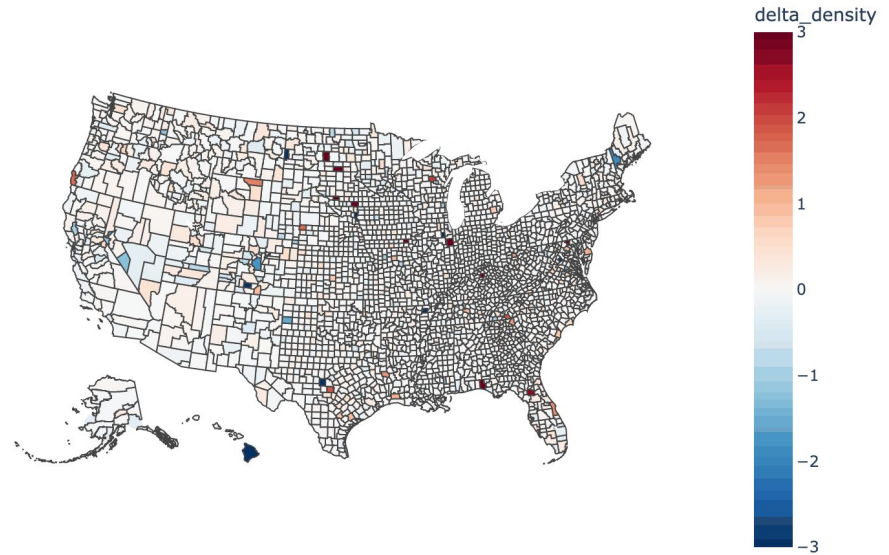
# Visualization of the Target:

## Monthly Micro-Business Density and Density Change

Microbusiness density by county on 2019-08-01

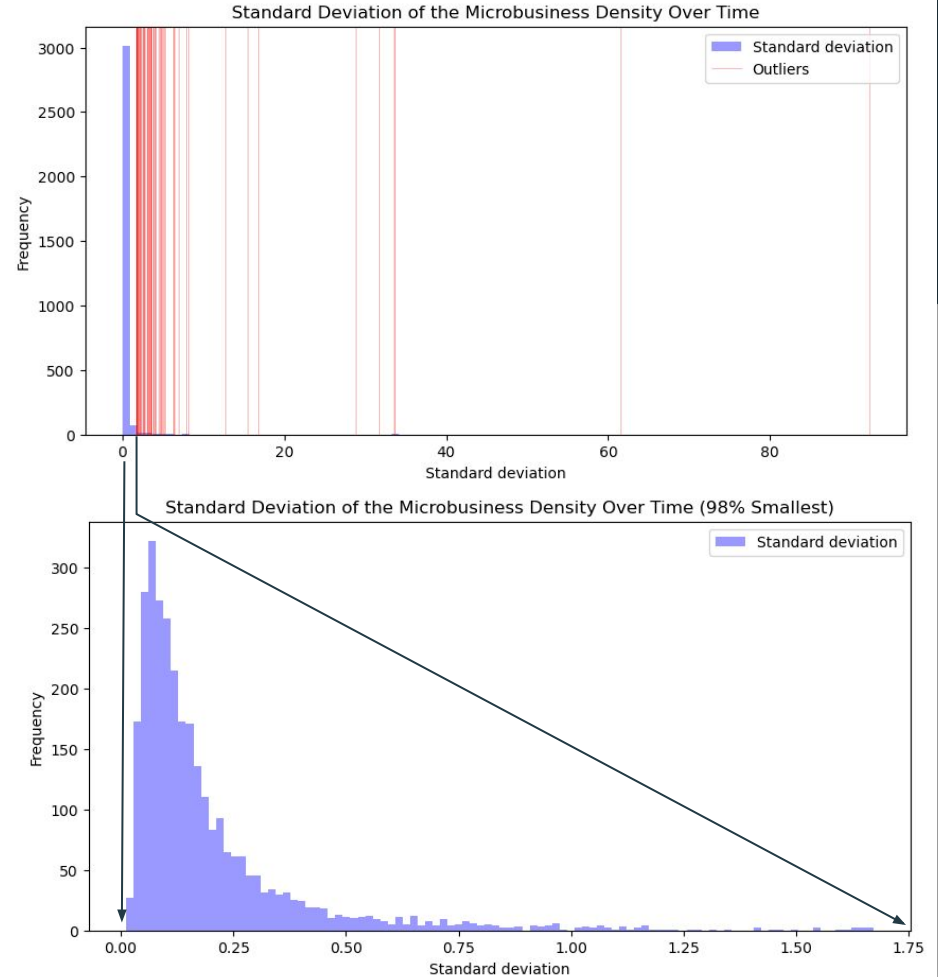
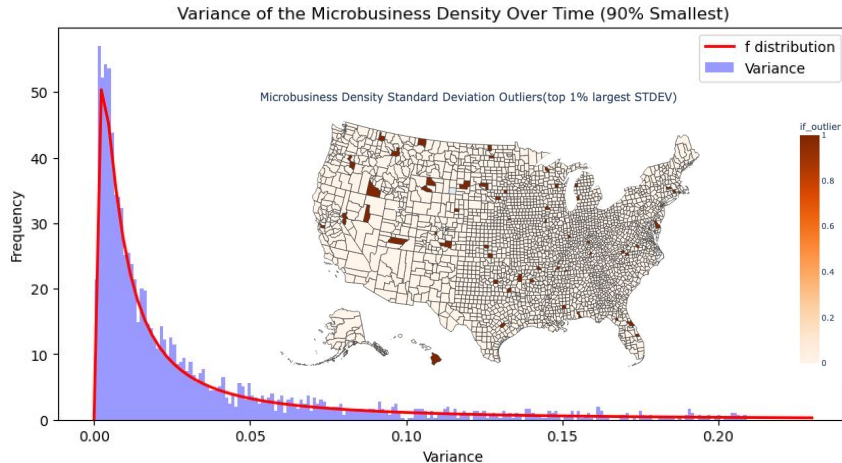


Monthly Change in Microbusiness density by county on 2019-09-01



# Distribution of the Targets

- The MBD's are not nice and normally distributed, it's heavy-tail-ish.
- Outliers, in red, with high leverage exists.
- Candidate Distributions: f, mielke, nct, betaprime, invgamma



# Data Collection & Feature EDA

## Types of Features

- Monthly
- Yearly
- Static

## Important Features

- \*\*
- Business Tax
-

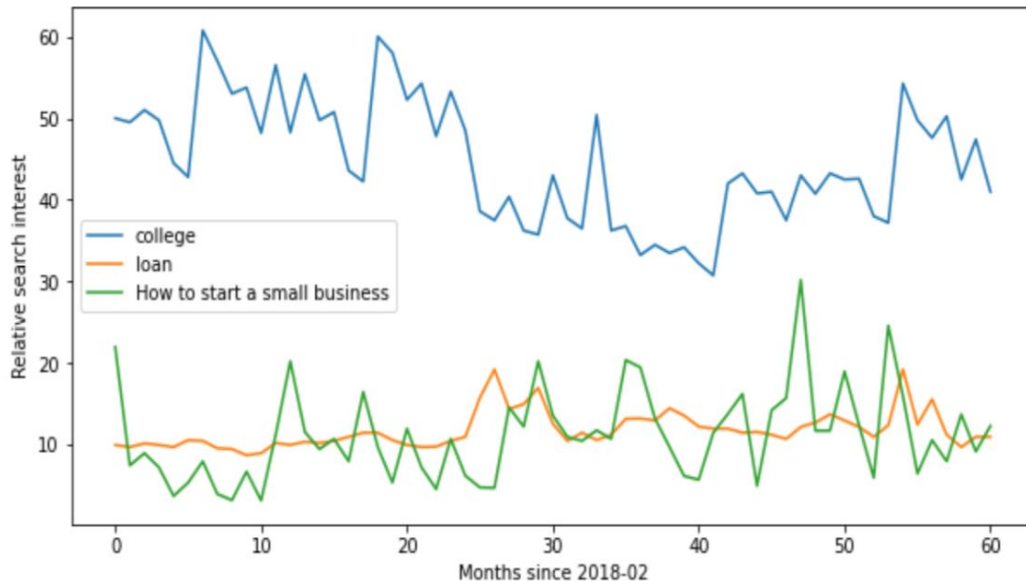
# Google Trends Features

## County-wise queries (7)

- “Alameda County Tax”


## State-wise queries (17)


- “Tax”
- “How to start a small business”
- “Business loan”



# ChatGPT into BERT

## County-wise queries to ChatGPT's API.

 Give me bullet points for why I should or should not start a small business in Autauga County Alabama, 5 pros and 5 cons.

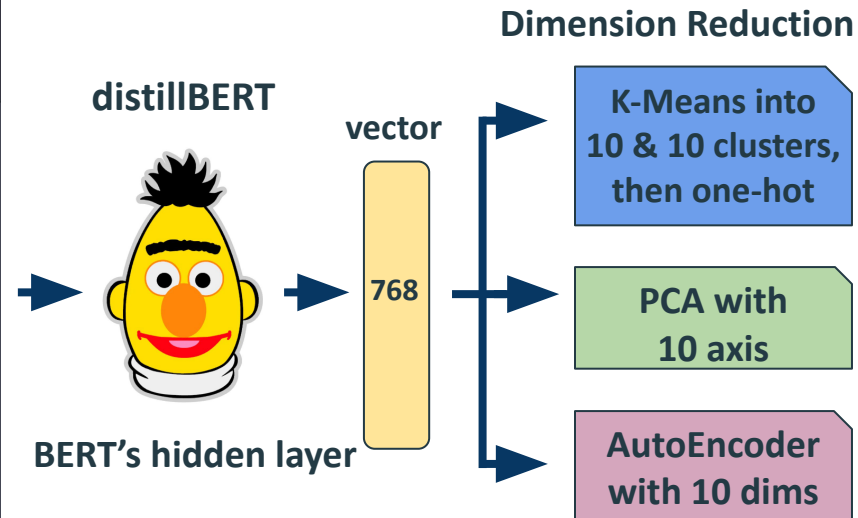
 Pros of starting a small business in Autauga County, Alabama:

- 1. Low cost of living and doing business
- 2. Supportive business community
- 3. Favorable business climate
- 4. Resources available for small businesses, such as the Small Business Administration and local business incubators
- 5. Well-developed transportation infrastructure

Cons of starting a small business in Autauga County, Alabama:

- 1. Limited access to investment capital
- 2. Small market size
- 3. Limited pool of skilled workers
- 4. Economy may be reliant on specific industries
- 5. High competition in some industries.

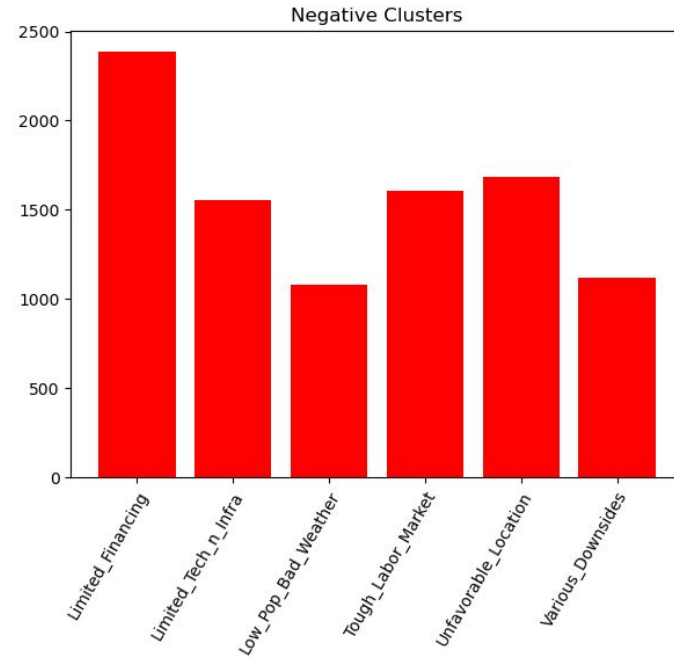
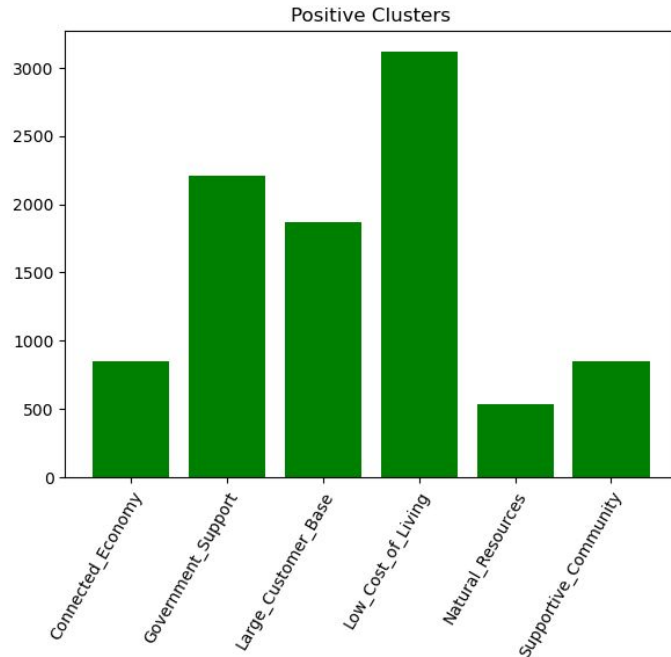
**bullet point or paragraph-wise embedding into tensors**





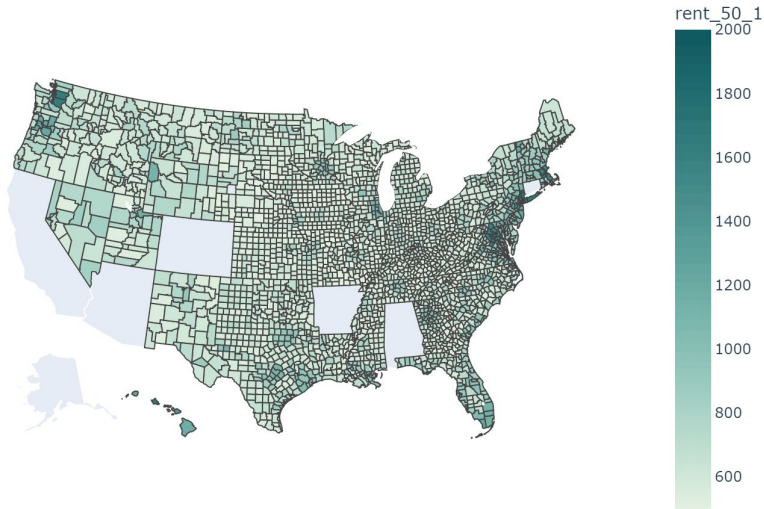
## Cluster 9 - "Government\_Support":

- ❑ There are numerous resources available to help small businesses get started, such as the Small Business Development Center and the Baldwin County Economic Development Alliance.
- ❑ Access to resources such as the Blount County Chamber of Commerce, which provides support and resources to local businesses.
- ❑ Access to resources such as the Small Business Development Center and the Butler County Chamber of Commerce

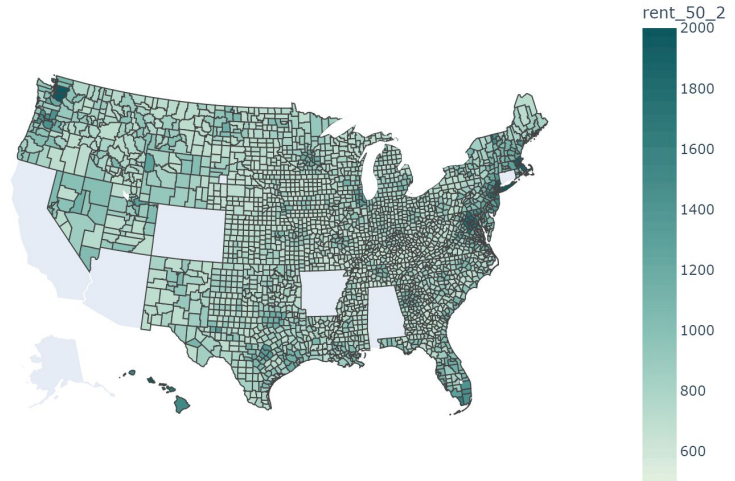


# Yearly Median Rent Features

median rent (1 bed/s) by county for year 2019

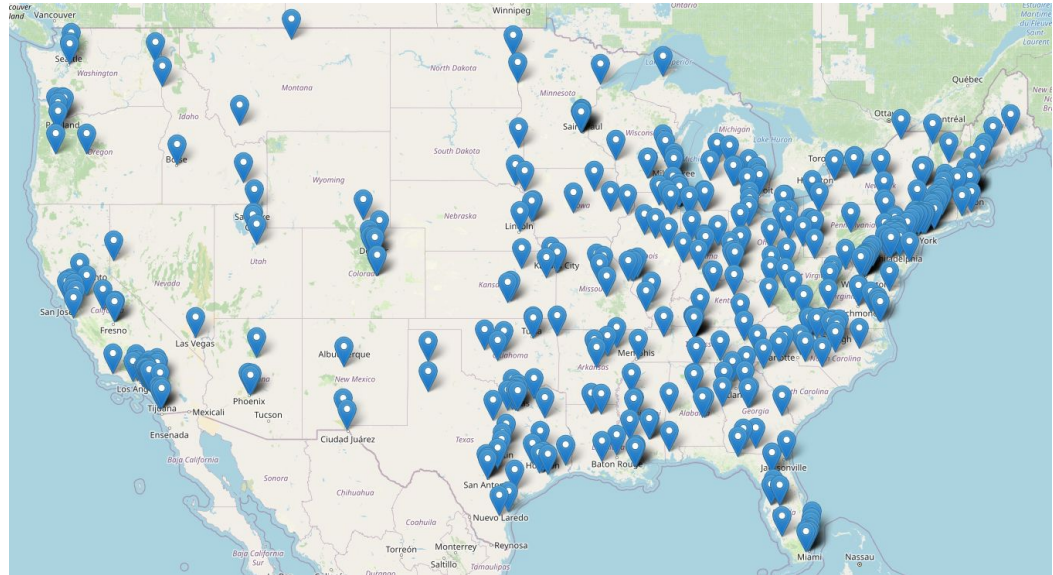


median rent (2 bed/s) by county for year 2019

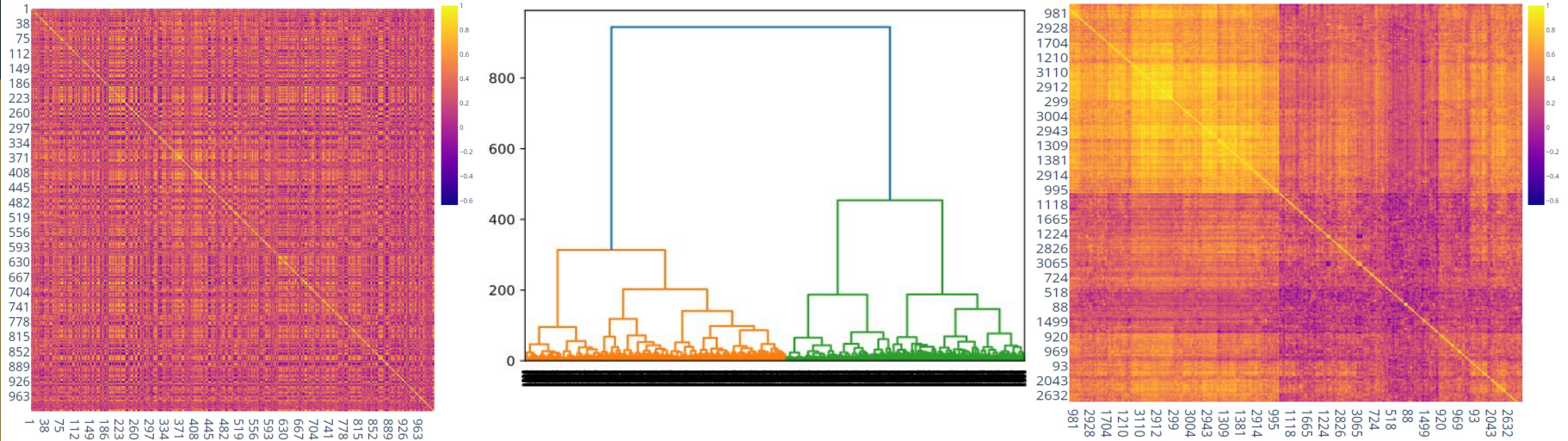


# Nearest National University Features

- Tuition
- Enrollment
- Acceptance Rate
- SAT/ACT Scores
- Graduation Rate
- US News Score
- Average Debt at Graduation



# Clustering Target Density Correlation





# What's next?

- Include additional features (Google Trends, geospatial)
- Drop smallest x% of counties
- Different kinds of clustering
- Other feature engineering