

# Stoutistical Analysis

Group 4: Andrew Hawthorne / Vicky Lasota / Natalia Lopez / Rajan Patel / Ben Richardson



# Premise of the project

A group of beer-and-coding enthusiasts are “hopping” to bring a fresh brew near you. Using U.S. Census Data and Open Brewery DB, this team is looking to identify the demographics of the area with the most breweries and see what might be ideal conditions for where to open Stoutistical Analysis.

## RESEARCH QUESTIONS

- What areas in the U.S. have the most micro and brewpubs?
- What city has the most micro and brewpubs, and what are the demographic characteristics of the areas within that city with the highest concentration of those breweries?
- What city has the least micro and brewpubs, and how do the demographic characteristics of that city compare with the demographic characteristics of the city with the most micro and brewpubs?

## DATA USED

The analysis was performed with the use of following datasets and APIs:

- [Brewery DB API](#)
- [Geopify API](#)
- Census API - [Parameters Sheet](#) & direct csv [download](#)
- [ZIP/ZTCA crosswalk](#) used in t test Zip-code Analysis

# Our code - download, clean, merge, and chart

```
# URL for GET requests to retrieve brewery data
base_url = 'https://api.openbrewerydb.org/v1/breweries?by_country=United_States'
```

```
#create a new dataframe with only the missing longitude rowsbased on clean_breweries_df
missing_long_df = clean_breweries_df[clean_breweries_df['longitude'].isnull()]
```

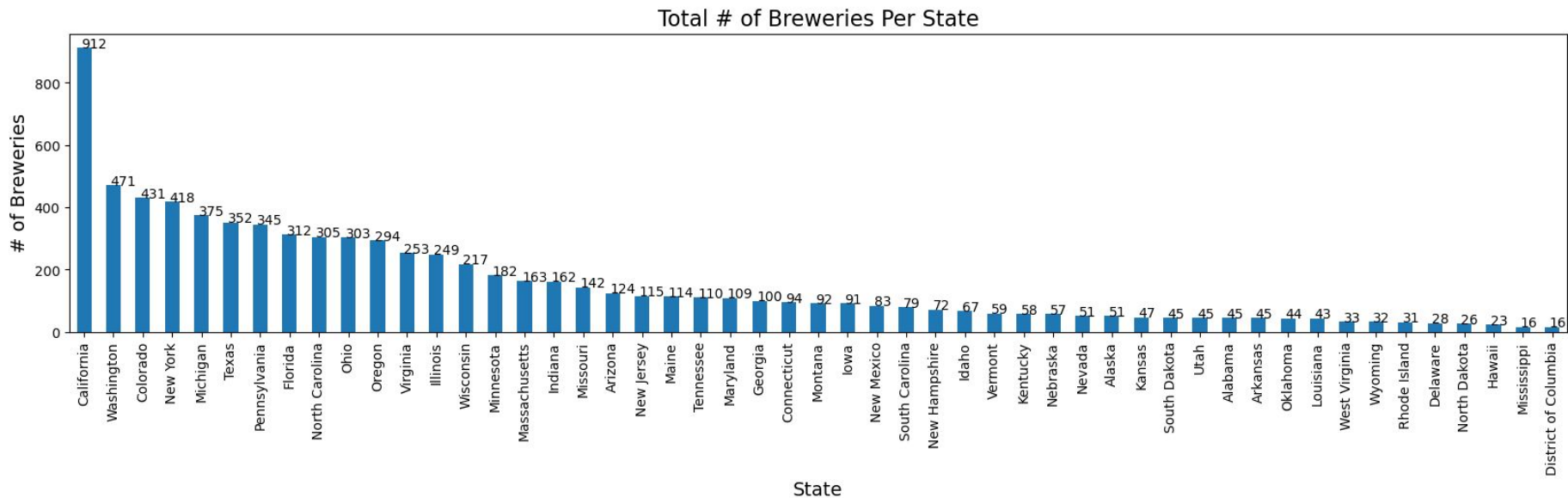
```
# Base URL for census dataset
base_url = 'https://api.census.gov/data/2021/acs/acsse?'
```

```
# Merge two dataframes by City column
merged_df = pd.merge(brewery_last, census_last, on=["City","State"], how="left")
```

```
# Fix misspelling "MIssouri" in state column and remove brewery_id 1986 (displays separately from rest of "Washington" data)
breweries_complete["state"] = np.where(breweries_complete["state"] == "MIssouri", "Missouri", breweries_complete["state"])
breweries_complete = breweries_complete.loc[breweries_complete["brewery_id"] != 1986]
```

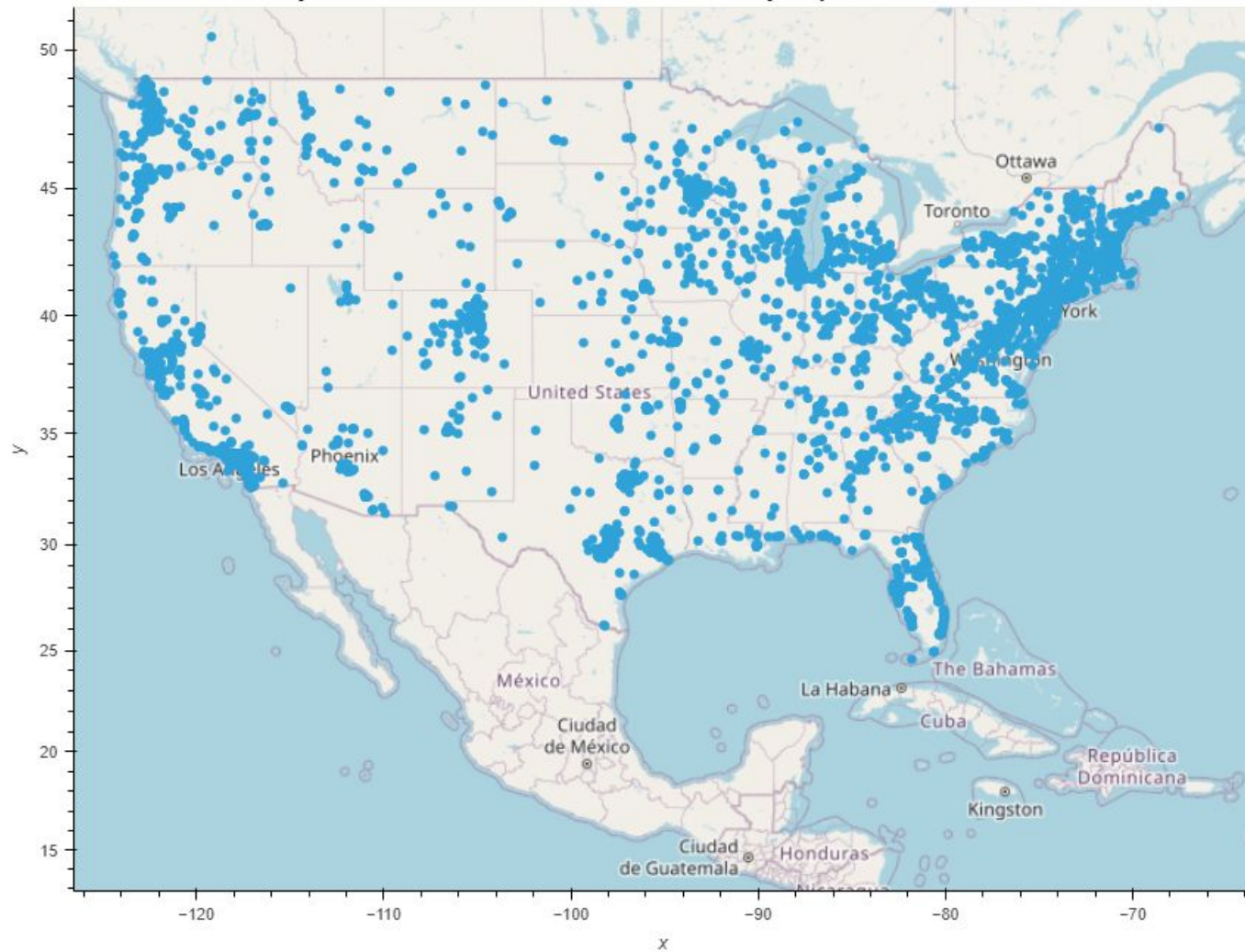
```
# Set dataframe to plot
SummaryCorrelation_df = last_df.groupby(['State', 'City'])[['population_total', 'population_male', 'population_female', 'population_median_age',
    'population_median_age_male', 'population_median_age_female', 'income_household', 'income_household_median', 'income_family_median', 'income_nonfamily_median']].mean()
SummaryCorrelation_df
```

# What areas in the US have the most breweries?



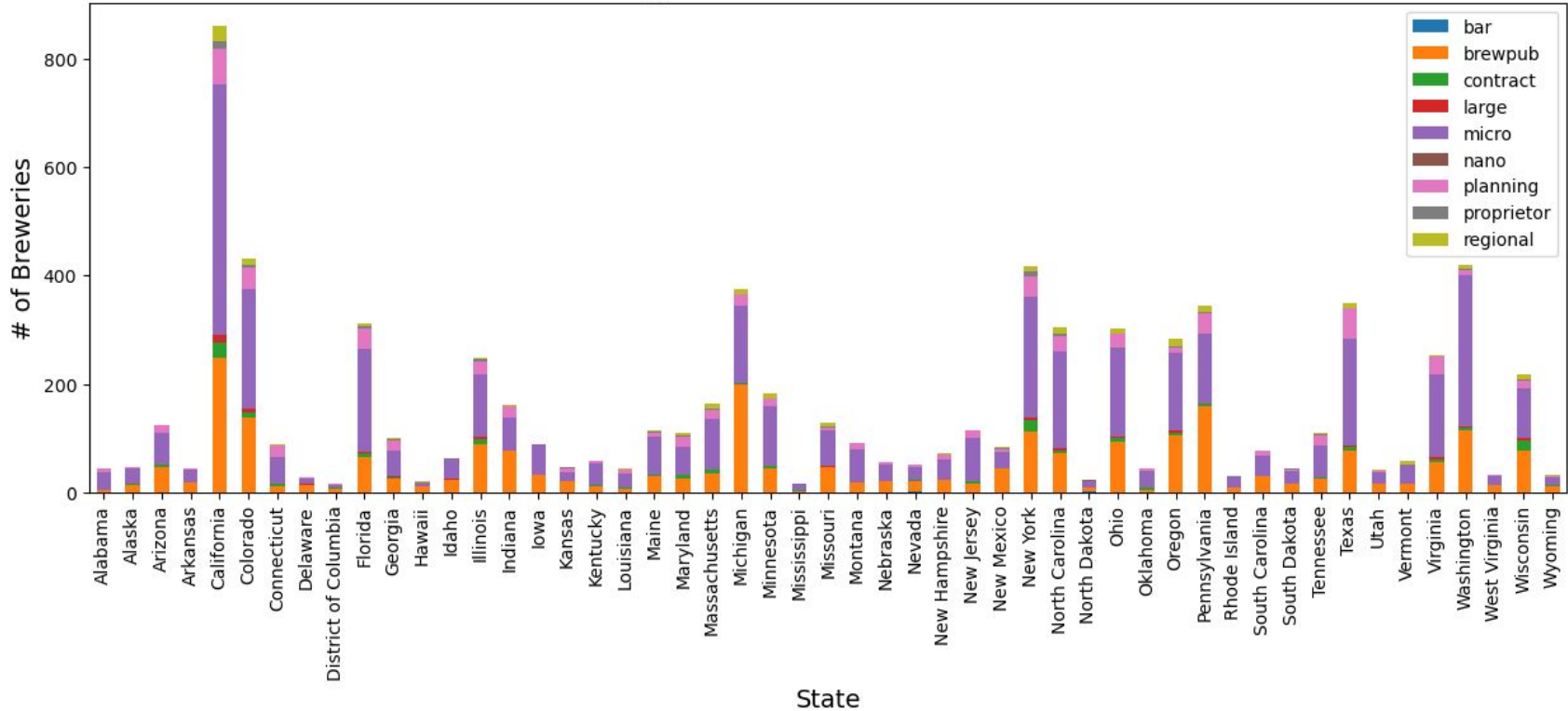
7000+ in Brewery BD dataset

Micro & Pub Brewery Locations in the Continental US - Density Map



## What types of breweries are there?

Types of Breweries Per State

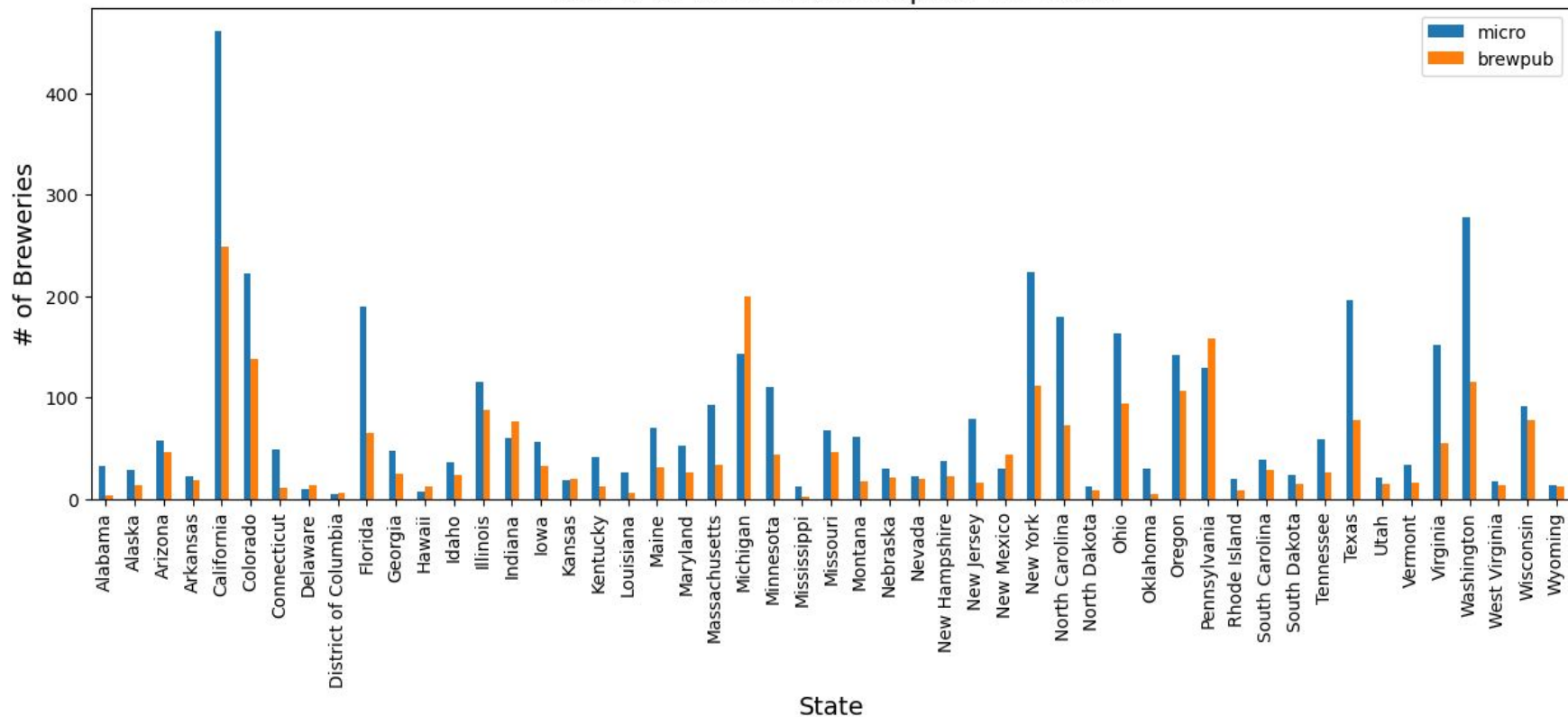


The largest categories were micro and brewpubs (6000+)



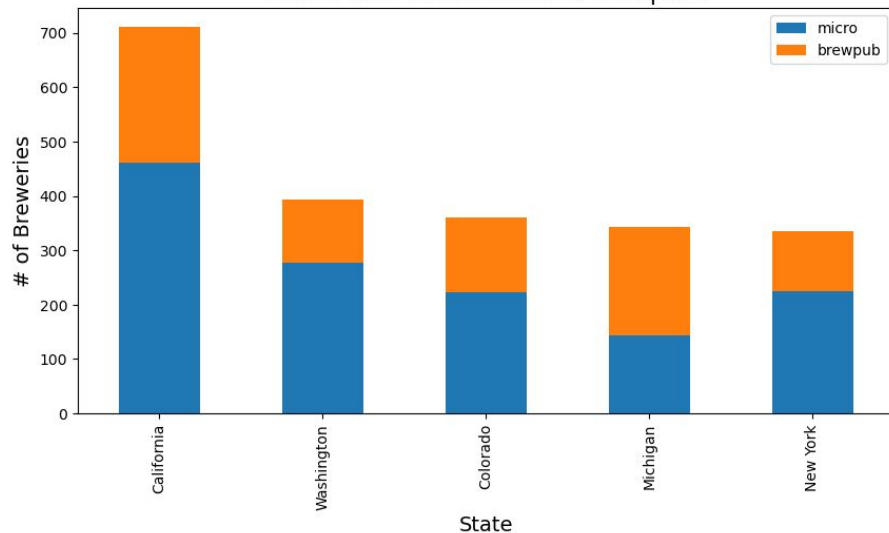
# Total Micro and Brewpubs per State

Total # of Micro and Brewpubs Per State

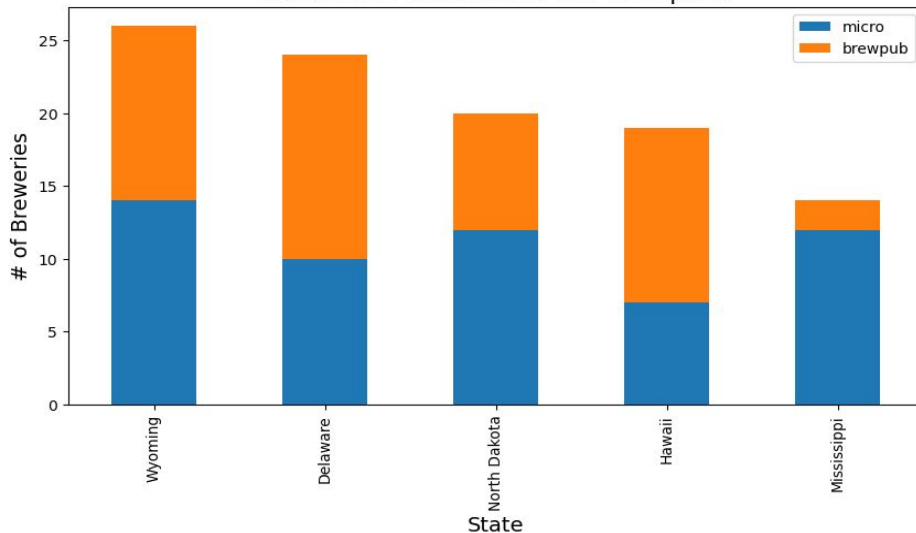


# Micro and Brewpubs in Top / Tail 5 States

States with Most Micro and Brewpubs



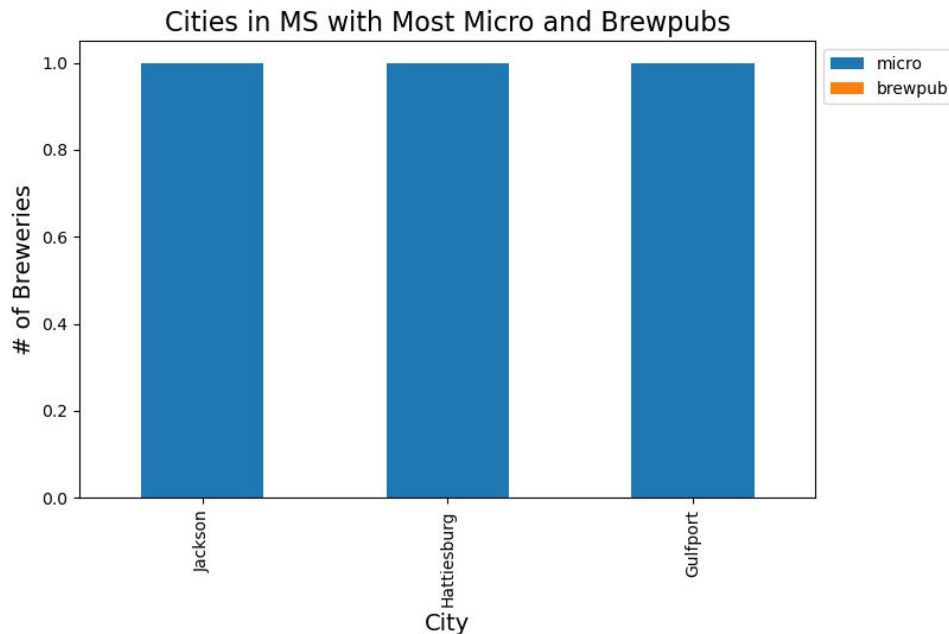
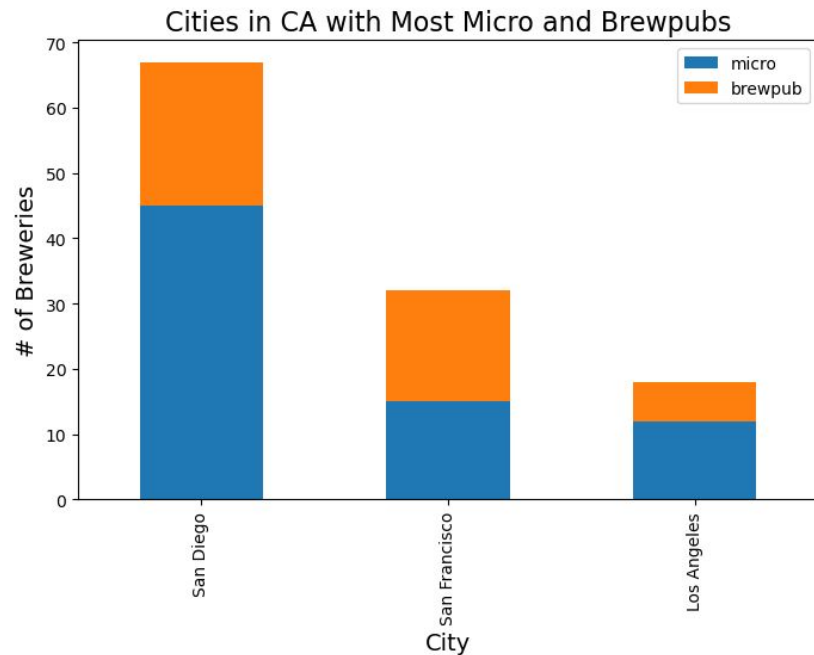
States with Least Micro and Brewpubs



California has 700+ micro/brewpubs and Mississippi has 14 total



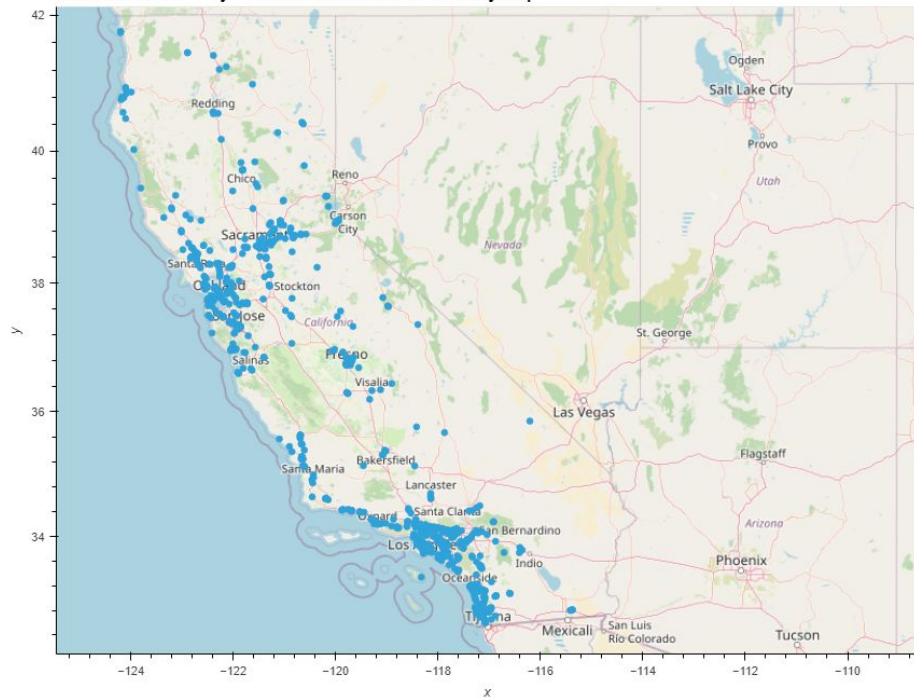
# What cities in the top and tail states had the most micro+brewpubs?



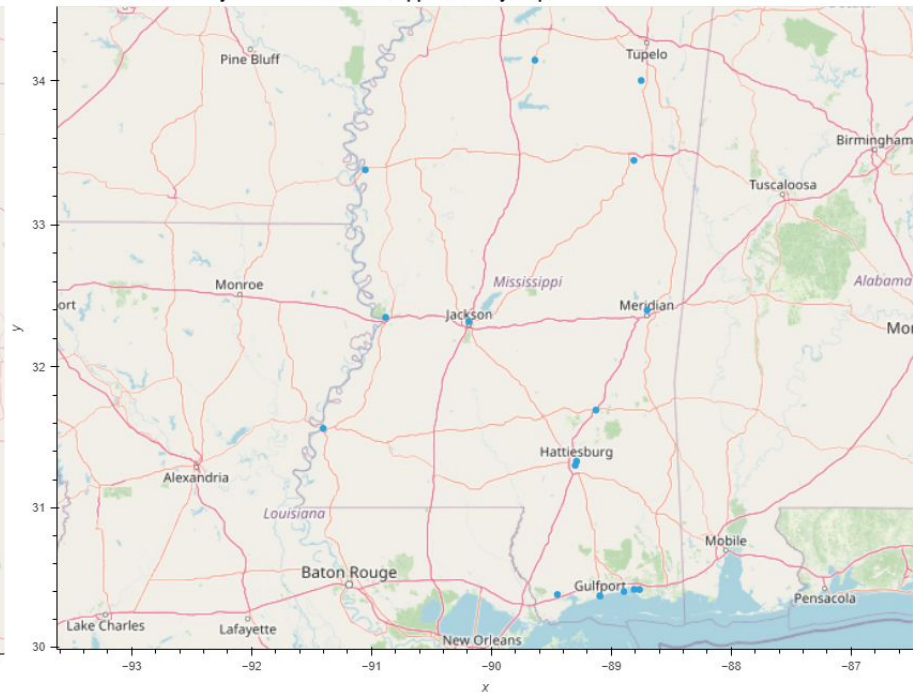
San Diego has 67 micro+brewpubs and the cities on record for Mississippi had about 1 micro or brewpub per city

# Top / Tail States Heat Maps

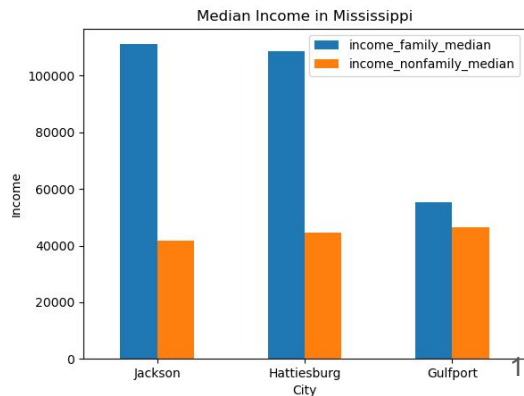
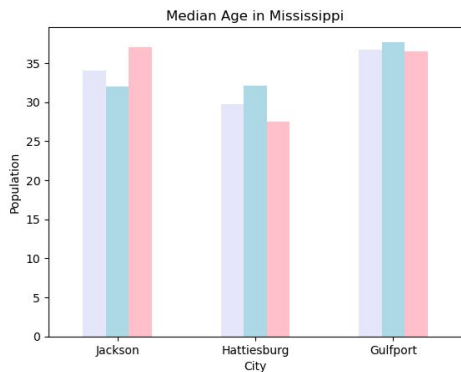
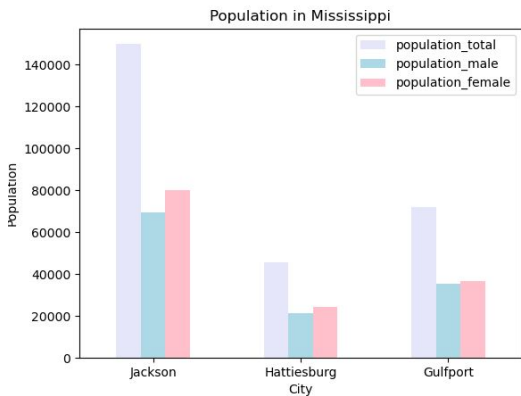
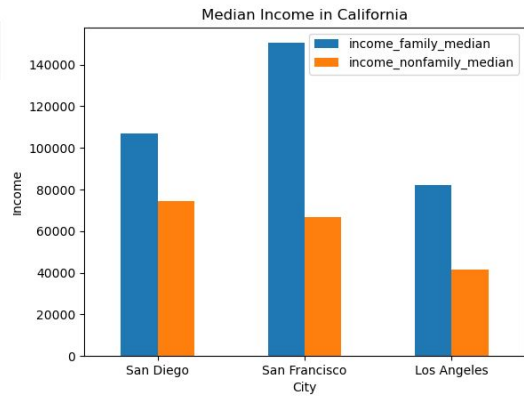
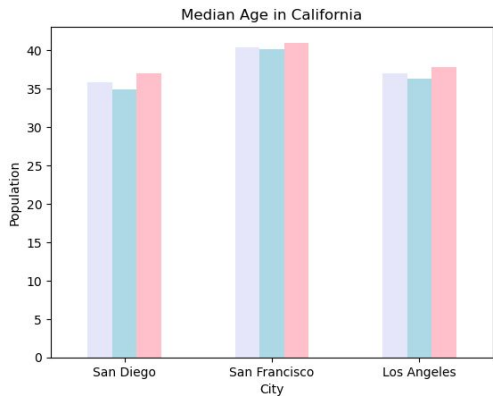
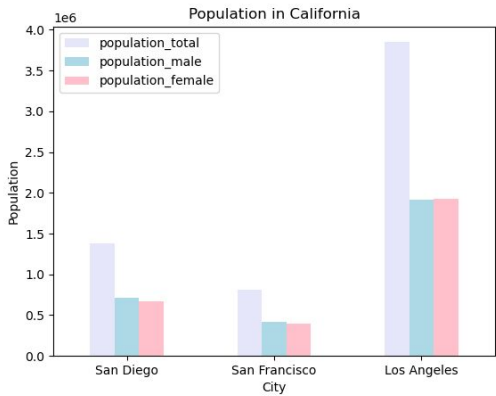
Micro & Pub Brewery Locations in California - Density Map



Micro & Pub Brewery Locations in Mississippi - Density Map



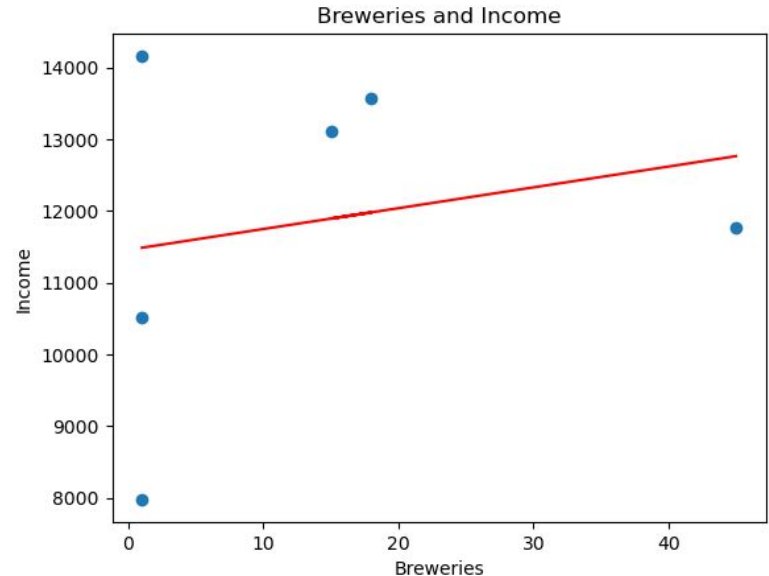
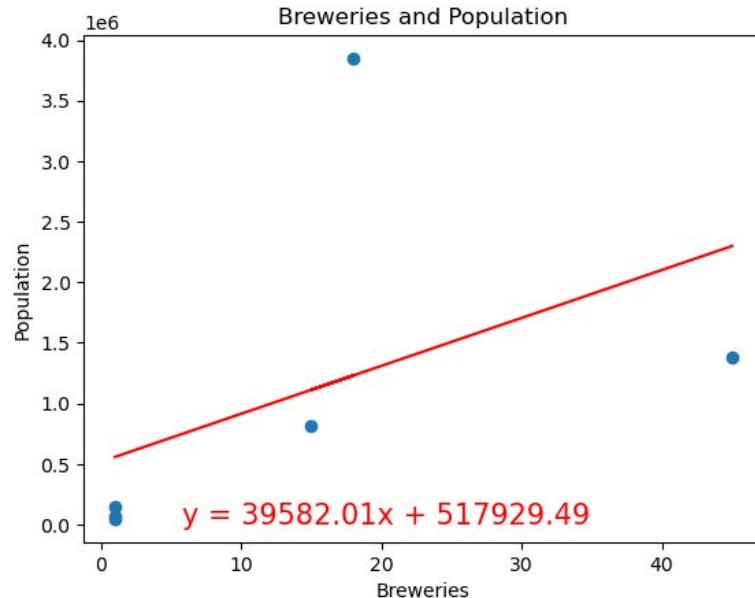
# Demographic characteristics of top cities for micro+brewpub concentration and cities with least amount of micro or brewpubs (but still have them)



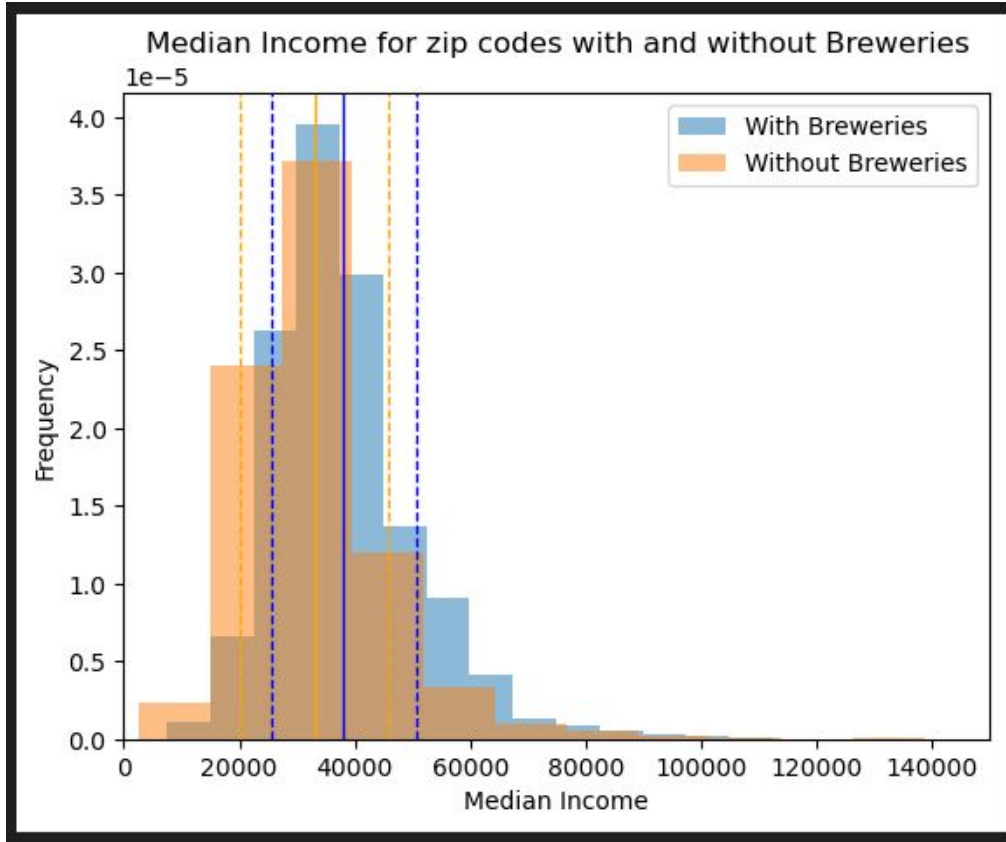
## ***Based on density analysis, what are the characteristics of the ideal location for a new brewery?***

- High population densities
- Age
- High income

R-squared <0.2 (n.s.)



# T-Testing - Median Income by ZIP-CODE



The set is comparing median income for zip codes with no brewery locations against zip codes with brewery locations.

Mean with breweries: \$ 3,8099  
Mean without breweries: \$ 3,3059  
Absolute difference: \$ 5,040.33

`Ttest_indResult(statistic=27.5871559`

`p-value=1.0600259801014812e-160)`

Cohen's d: 0.3978458087054  
-> MEDIUM Effect

# Results:

- Distribution comparison showed high left skew for full zip-code listing while the distribution remained relatively flat for the tail of the US zip code listing further confirming that breweries can be found predominantly in urban areas - aka smaller zip codes with higher populations.
- Urban areas are preferred, however 65% breweries being located in areas with population density less than 400 per square mile - for higher density data brewery distribution mimics zip-code code distribution almost exactly - stick to the suburbs
- Median income plays a role - T-test comparison of median income comparison between zips with and without breweries showed a statistically significant difference with brewery zips showing higher median income - craft beer drinkers are in the 75K annual income bracket
- There is a mild to moderate positive correlation between age bracket and number of breweries - around 60% are in the 25-54 age range - look for affluent millennials

**Conclusion: we advise that busy, wealthy suburban locations with a large population in the 25-54 age range are ideal locations for opening a micro or brewpub in the US.**

# LIMITATIONS

- No cost analysis relating to real estate, taxes, labor, & supplies
- Non-urban tourist locations largely excluded, but based on the analysis worth exploring based on visitor data for example ski resort towns
- No licensing or licensing limitation information
- No sales & profit margin information
- Limited data on closing & planned and no discrimination in type - when totals compared with the Brewers Association summaries
- Quality of beer and types of beer aka product offerings not discussed
- Lack of revenue projection data - requiring further analysis
- Survivorship bias - data on closing was too small to run the comparison

Our data analysis could not focus on tourist locations since our data was residential; we recommend drilling down further by zip code (vs. just city) and use different demographic data to compare city vs. smaller tourist locations. Further market research is required to investigate brewery licensing by state and preferences toward “local” breweries.



# RESOURCES

<https://www.foodandwine.com/beer/craft-beer/cities-most-craft-breweries>

<https://passionpassport.com/american-cities-craft-beer/#:~:text=San%20Diego%2C%20California&text=With%20long%20stretches%20of%20sandy,to%20this%20southern%20California%20oasis>

<https://www.brewersassociation.org/statistics-and-data/national-beer-stats/>

<https://www.brewersassociation.org/resource-hub/business-model/>

<https://www.brewbound.com/news/power-hour-nielsen-shares-2019-craft-beer-consumer-insights/>

<https://thebrewermagazine.com/new-research-reveals-current-beer-consumption-rates-influenced-mostly-by-history-of-personal-choice/>

<https://blacktailnyc.com/who-drinks-craft-beer-demographics/>

<https://www.brewbound.com/news/power-hour-nielsen-shares-2019-craft-beer-consumer-insights/>