



CENSUS API DATA: ANNUAL BUSINESS SURVEY (2019)

By: Fritz Grunert, David Perhai, Talia Willcott & Isabelle Hyppolite





DATA

The Annual Business Survey provides information on selected economic and demographic characteristics for businesses and business owners by sex, ethnicity, race, and veteran status.

The two main datasets we analyzed are:

1. **Characteristics of Business Owners** - Provides data for owners of respondent employer firms by sector, sex, ethnicity, race, and veteran status for the US, states, and metro areas.
2. **Company Summary** - Provides data for employer businesses by sector, sex, ethnicity, race, veteran status, years in business, receipts size of firm, and employment size of firm for the US, states, and metro areas.

The Characteristics of Business Owners (CBO) dataset is similar to the Company Summary dataset, but with more of a focus on the business owners. The main differences are that it includes more demographic and categorical columns about the owners and that it only has the total number of owners for its data values.





DATA (CONT.)

Possible questions that can be answered using the data:

- Is there a relationship between location and how successful a business is?
- Is there a relationship between race of owner and how successful a business is?
- Is there a relationship between gender of owner and how successful a business is?
- What factors affect average pay in the United States?
- What are the characteristics of business owners based on race?

DATA SOURCE

Bureau, U. S. C. (2022, October 28). *Annual Business Survey (ABS) apis*. Census.gov. Retrieved December 30, 2022, from <https://www.census.gov/data/developers/data-sets/abs.2019.html>



ETL PROCESS

EXTRACTION PROCESS

- Import the correct packages
- Make a variable to choose what columns we needed from the table
- Make a variable to get the key to access the data
- Make a variable to choose between state and US
- Construct a resulting URL from the previous variables we defined
- Create a variable for the pulling of the API
- Pull the text from the request
- Read that text and turning it into a variable
- Turn the first row into the column headers
- Make sure we don't include the column names in the actual data
- Show all the columns so that none get cut off
- Turn the tables we made into CSVs

```

1
2
3 # Import the correct packages
4 import pandas as pd
5 import requests
6
7
8 # 1
9
10 # Choose what columns we needed from the table
11 get = "GEO_ID,NAME,NAICS2017,NAICS2017_LABEL,OWNER_SEX,OWNER_SEX_LABEL,OWNER_ETH,OWNER_ETH_LABEL,OWNER_RACE,OWNER_RACE_L
12
13 # Get key to access the data
14 key = 'fd8c137d32e065e499223dba1594b3c5e7a52b91'
15
16 # Choose between state and US
17 geography = 'us'
18
19 # Resulting url from the previous variables we defined
20 url = (f"https://api.census.gov/data/2018/absco?get={get}" +
21       f"&for={geography}:" +
22       f"&key={key}")
23
24 # Create a variable for the pulling of the API
25 r = requests.get(url)
26
27 # Pulling the text from the request
28 data = r.text
29
30
31 # Reading that text and turning it into a variable
32 owners = pd.read_json(data)
33
34 # Turning the first row into the column headers
35 owners.columns = owners.iloc[0]
36
37 # Making sure we don't include the column names in the actual data
38 owners = owners[1:]
39
40 # Showing all the columns so that none get cut off
41 pd.set_option('display.max_columns', None)
42

```



ETL PROCESS

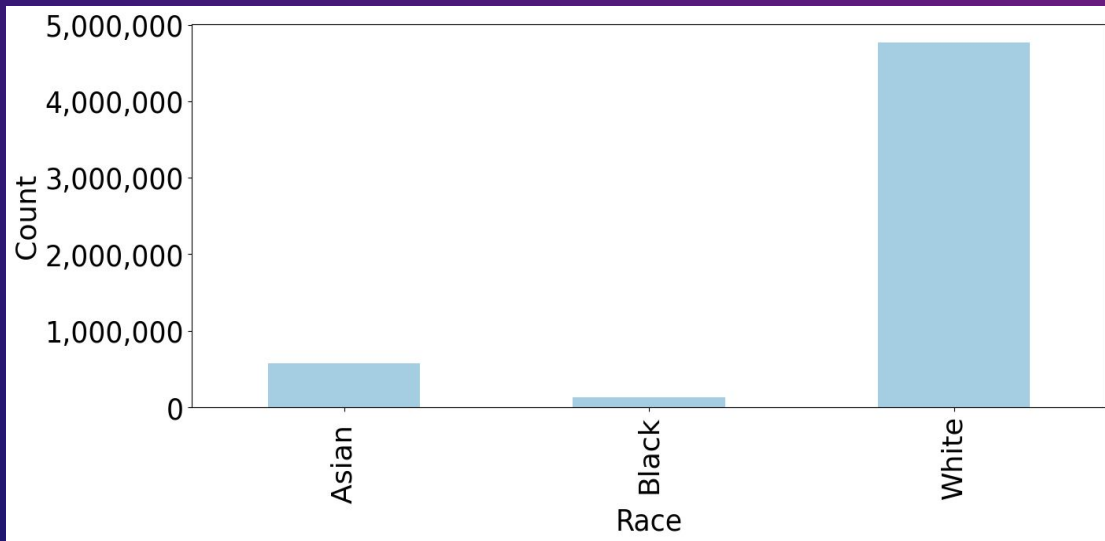
TRANSFORMING PROCESS

- Import the correct packages
- Show all the columns so that none get cut off
- Read in the CSVs we made
- Drop the columns we don't need
- Send the resulting tables to new CSVs

```
96 import pandas as pd
97
98 # Showing all the columns so that none get cut off
99 pd.set_option('display.max_columns', None)
100
101 # Read in the CSVs we made
102 owners = pd.read_csv('owners.csv', low_memory = False)
103 companies = pd.read_csv('companies.csv')
104
105 # Dropping the columns we don't need
106 companies_clean = companies.drop(columns=["Unnamed: 0", "GEO_ID", "NAICS2017", "SEX", "ETH_GROUP", "RACE_GROUP", "VET_GROU
107 owners_clean = owners.drop(columns=["Unnamed: 0", "GEO_ID", "NAICS2017", "OWNER_SEX", "OWNER_ETH", "OWNER_RACE", "OWNER_VE
108
109 # Sending the resulting tables to new CSVs
110 owners_clean.to_csv("owners_clean.csv")
111 companies_clean.to_csv("companies_clean.csv")
112
```



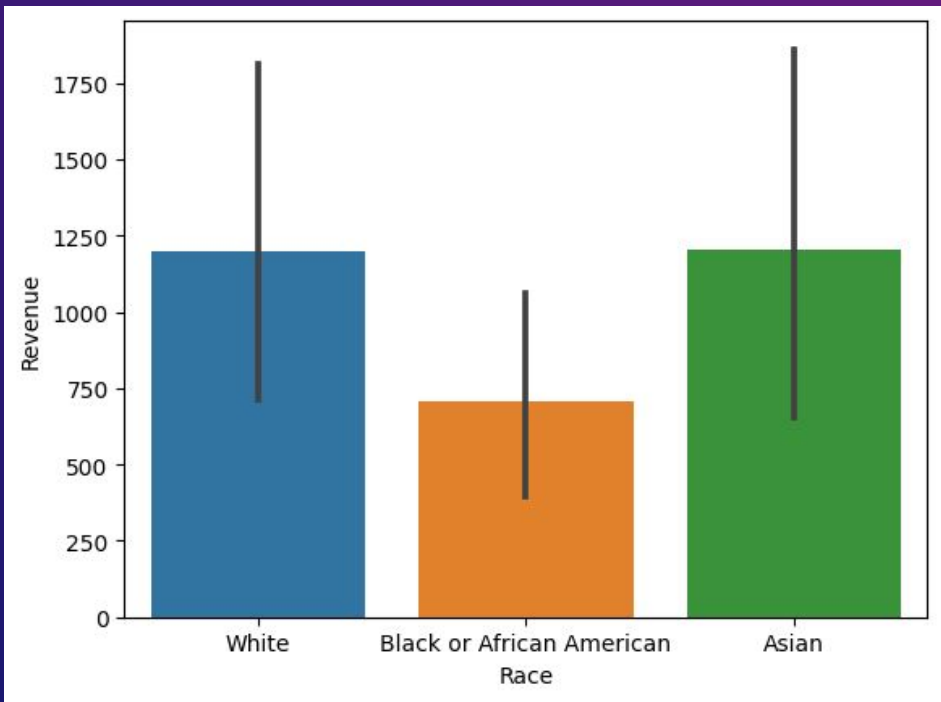
DISTRIBUTION OF RACE



This visual is displaying the distribution of races among business owners. There is a greater number of White business owners than any other race.



REVENUE BY RACE GROUP



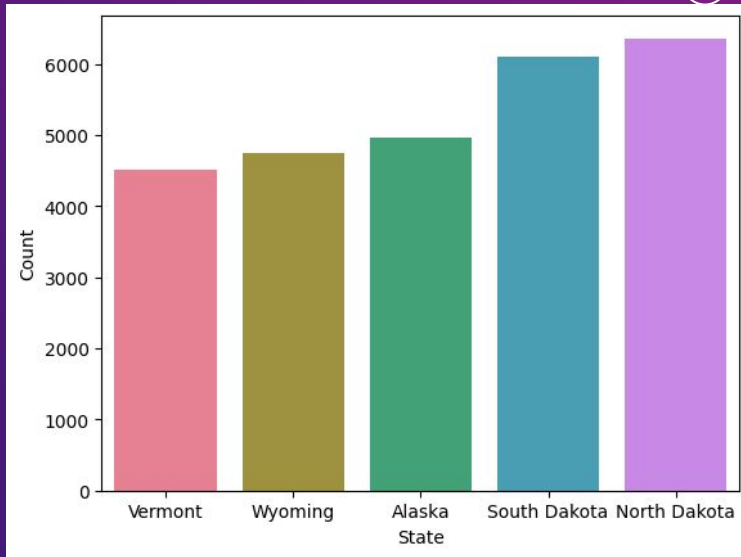
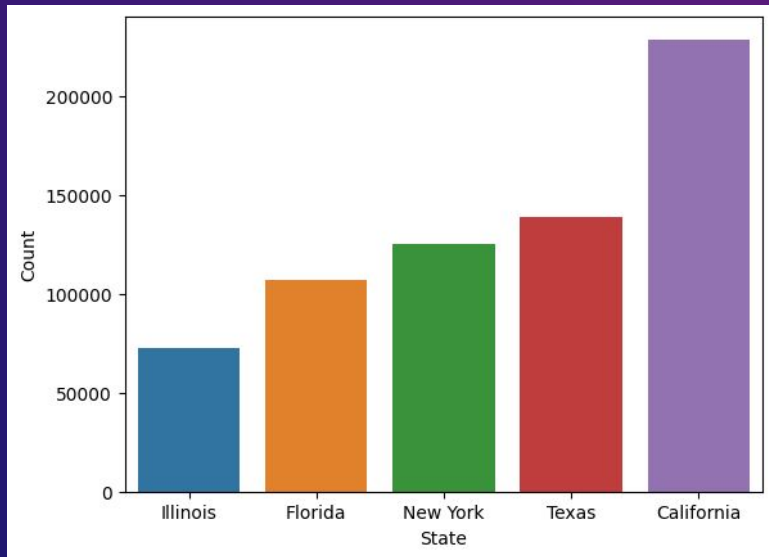
This visual is displaying the average revenue per business owner, measured by sales (RCPPEMP) in thousands of dollars, accrued by each major race group.

Most business owners are white, but they don't necessarily make more revenue based on being the majority. The average Asian business owner is bringing in similar amounts and the average African American business owner is bringing in a bit over half the amount as an average White business owner.





HIGH VALUE SALES



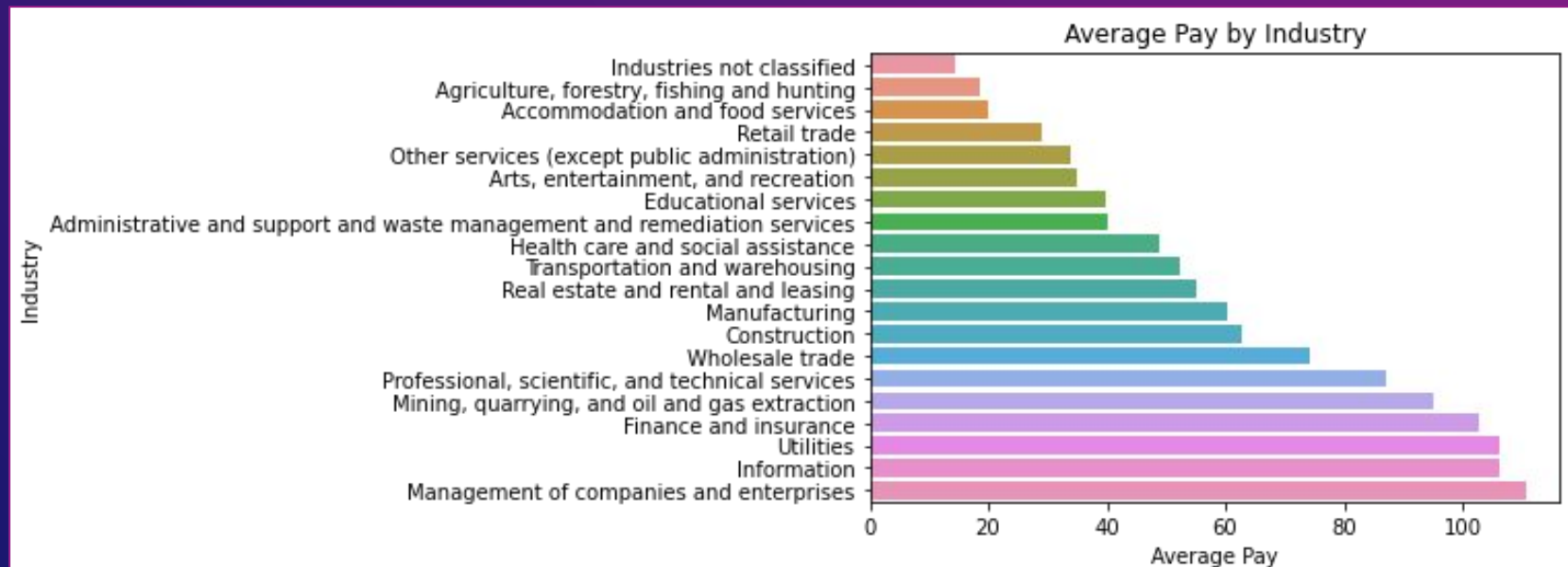
These visuals are displaying the states with the most companies that make over \$1,000,000 in revenue and the states with the least amount of companies that make over \$1,000,000 in revenue.





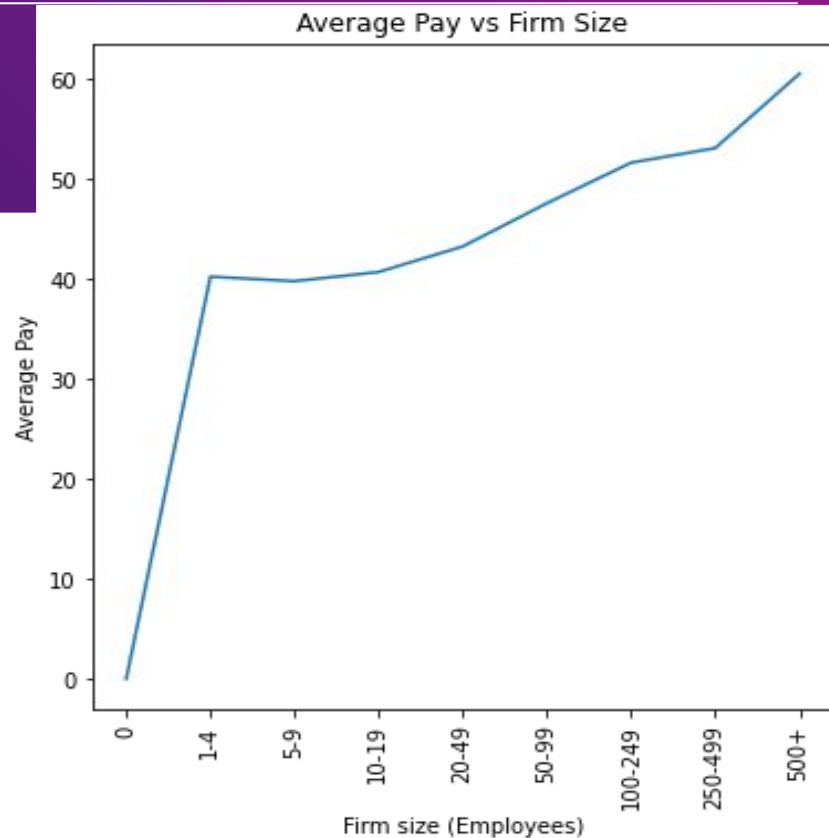
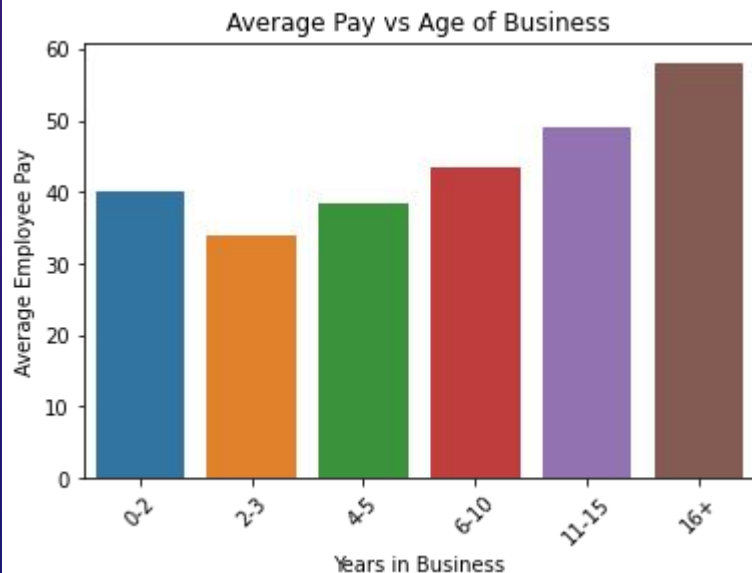
FACTORS THAT AFFECT AVERAGE PAY

BY INDUSTRY





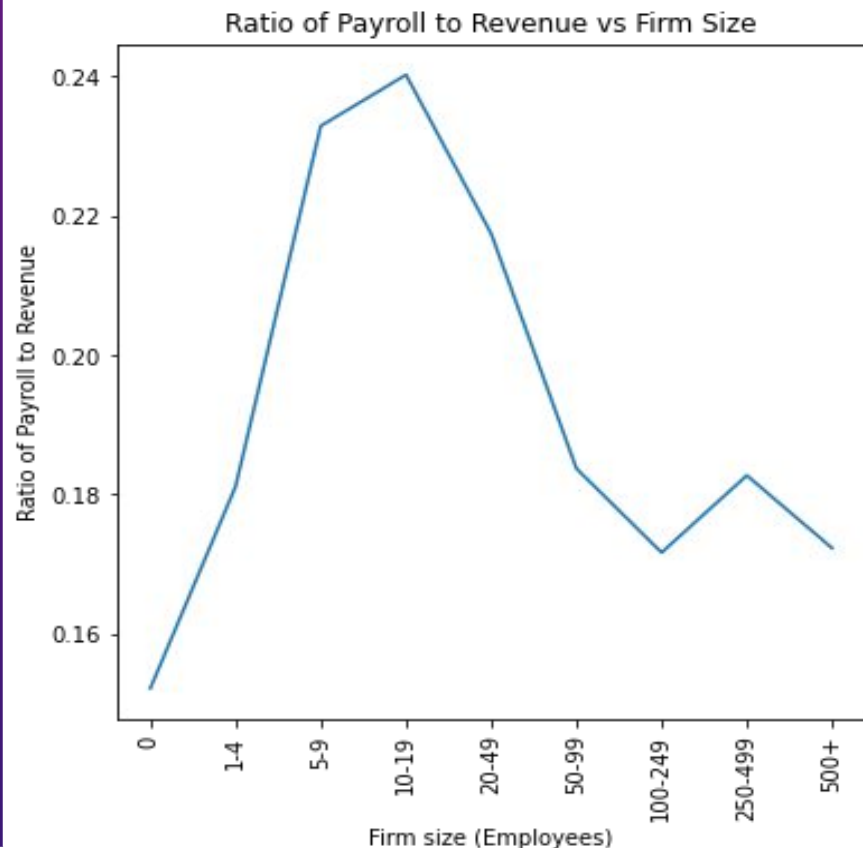
BY AGE AND SIZE



PAYROLL AND REVENUE

REGARDING FIRM SIZE

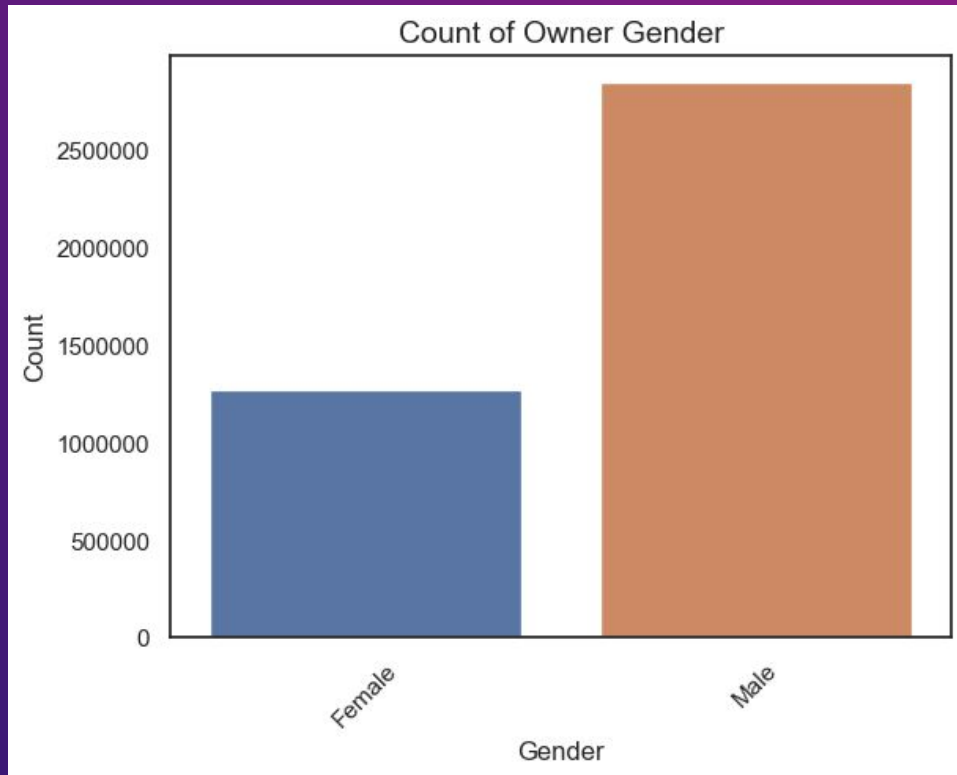
- Shows proportion of revenue spent on payroll
- Inconsistent bucket size from data source
- Firms with zero employees still spend ~15% of revenue on payroll



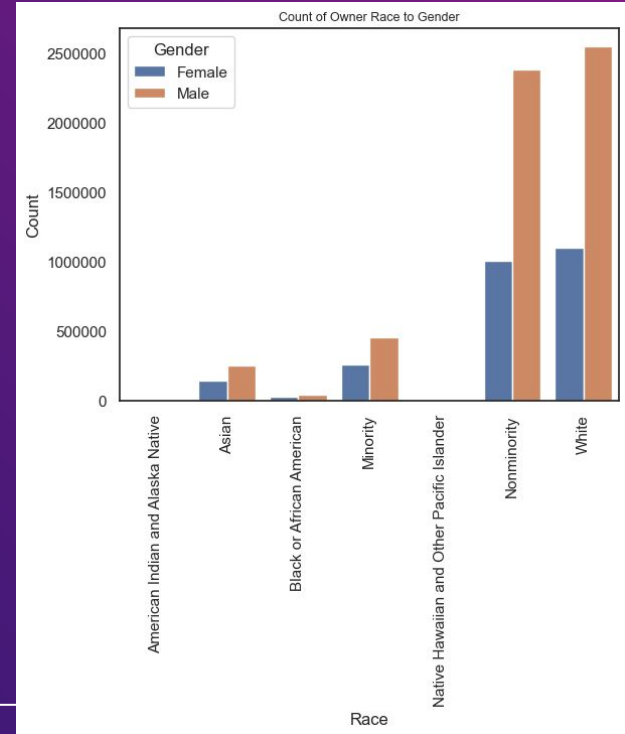
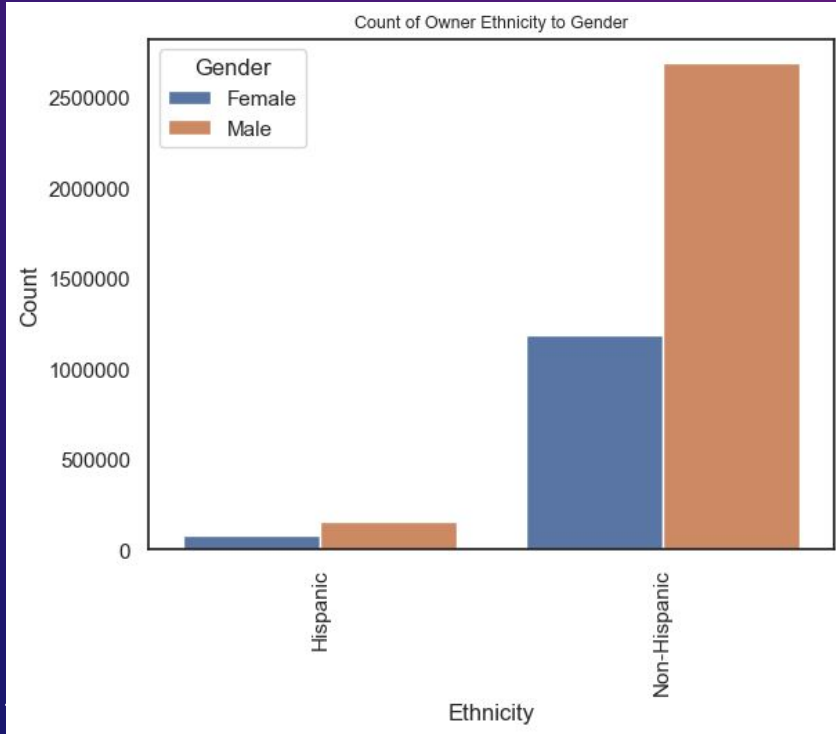


COUNT OF OWNER GENDER

This visualization shows the count of female and male owners as gathered by the survey.

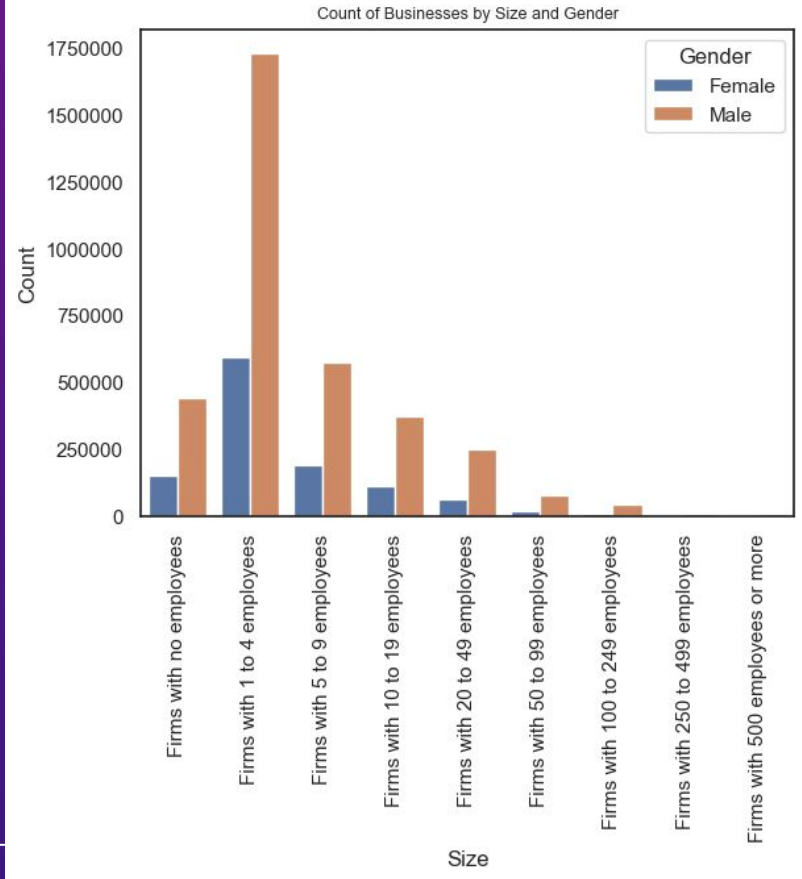


ETHNICITY AND RACE BY GENDER OF OWNERS



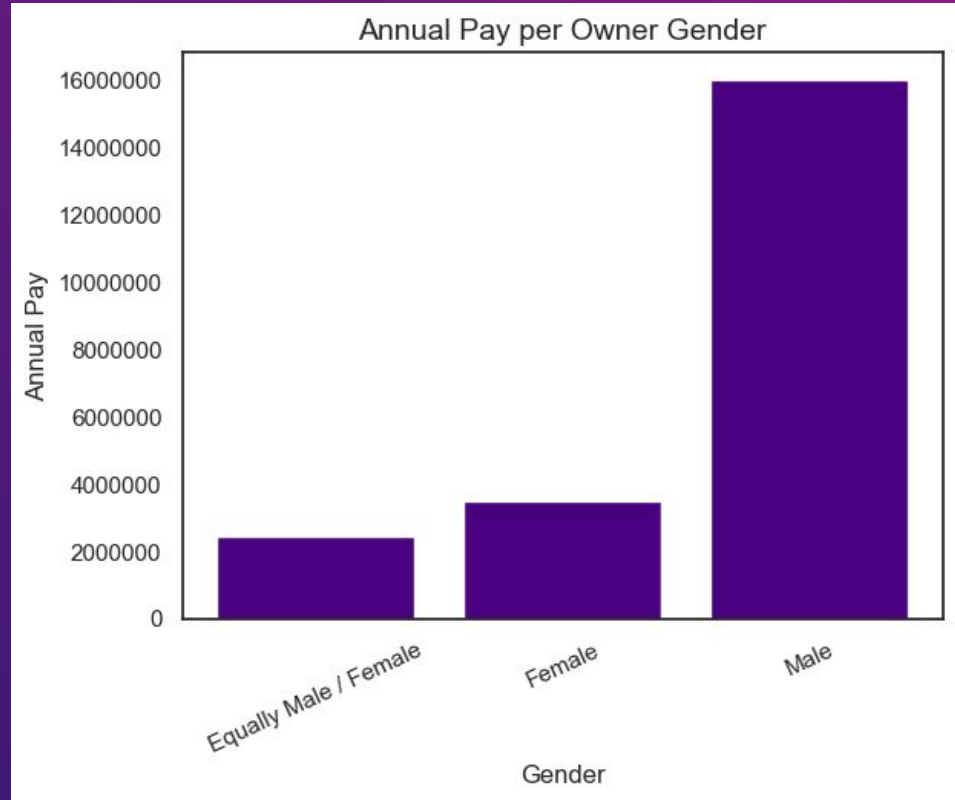
BUSINESS SIZE BY GENDER

This visualization shows the count of employees in businesses by the gender of the owner.



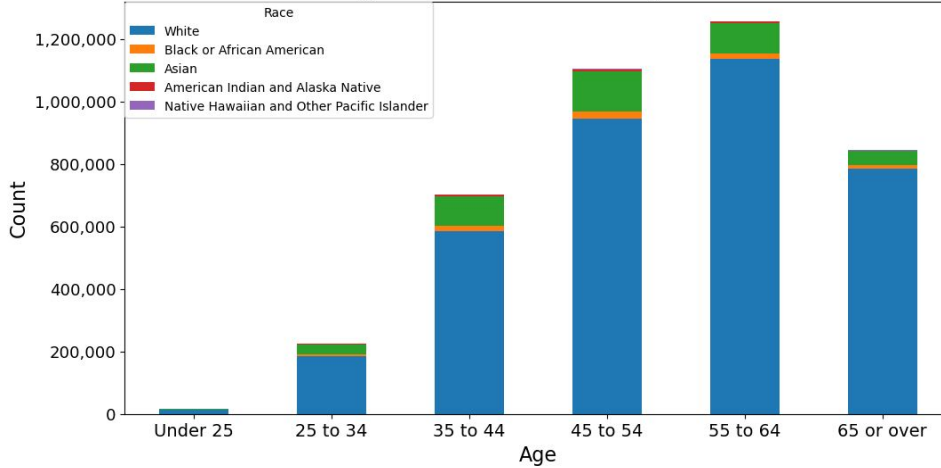
ANNUAL PAY BY GENDER

This visualization shows the average annual pay for employees of businesses that have male owners, female owners or equal male / female owners. Note that the average is the average of the total annual salary, not per employee.



WHAT ARE THE CHARACTERISTICS OF BUSINESS OWNERS BASED ON RACE?

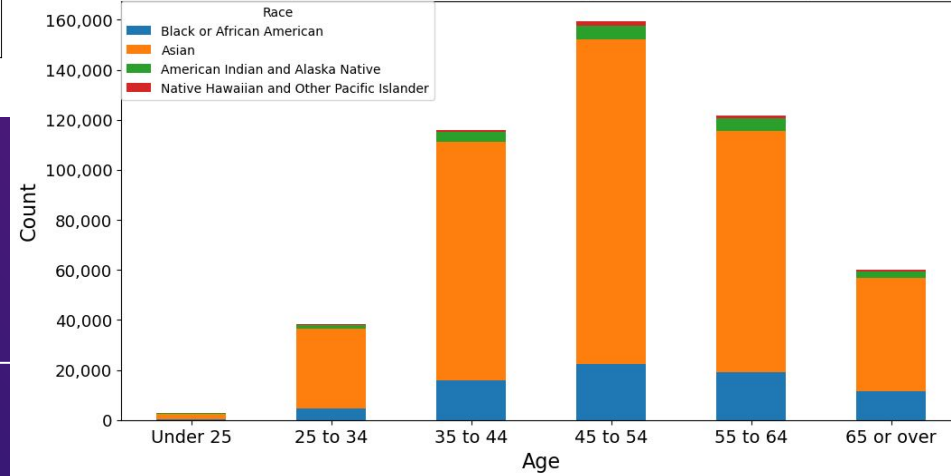
Age of Business Owner by Race



The age of most owners for the most part was over the age of 45, with most business owners being 55-64 years old.

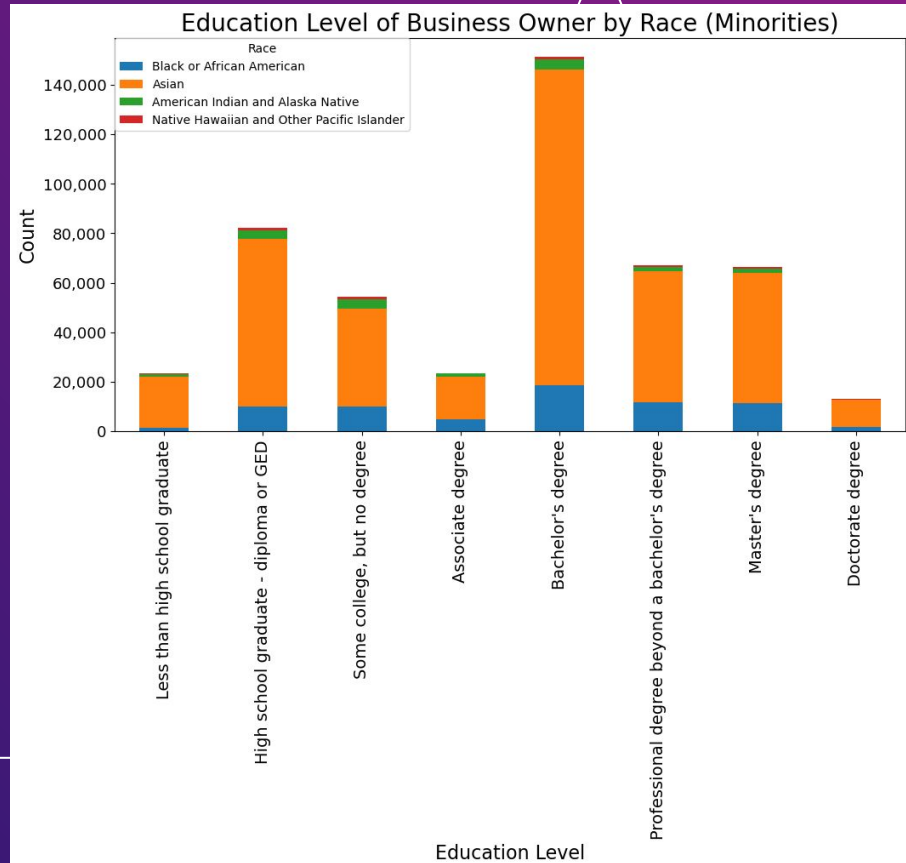
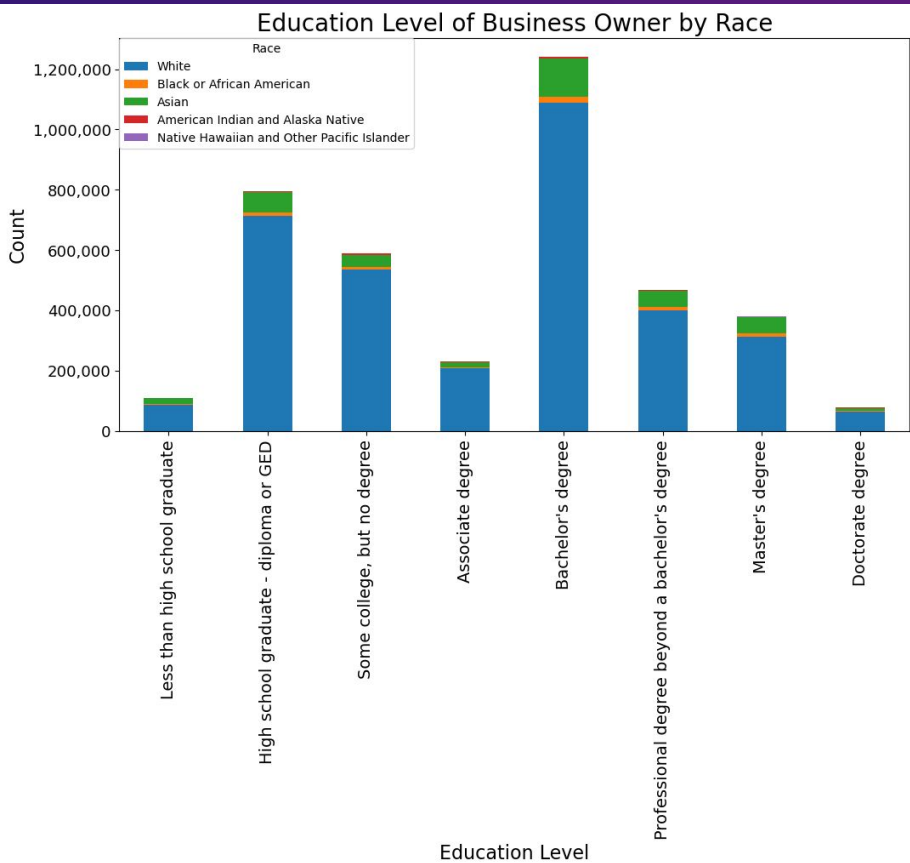
But minority business owners tended to be younger, with most falling between 35-64, and the highest proportion being 45-54 years old.

Age of Business Owner by Race (Minorities)





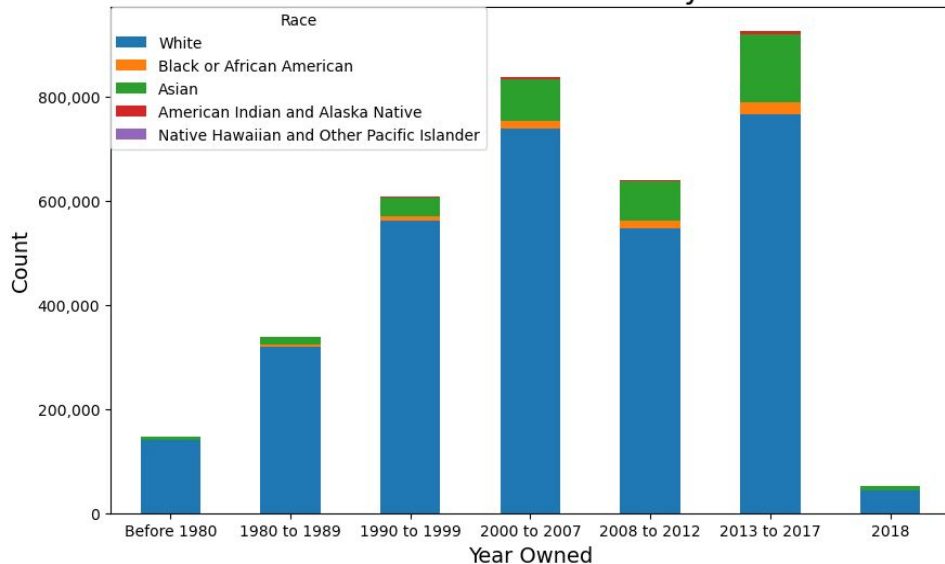
WHAT ARE THE CHARACTERISTICS OF BUSINESS OWNERS BASED ON RACE?





WHAT ARE THE CHARACTERISTICS OF BUSINESS OWNERS BASED ON RACE?

Year Owned the Business by Race



For minorities, there was a slight dip from 2008-2012, but not as steep as the dip from all businesses. Perhaps minority business owners from out of the country saw an opportunity to start a new business when there was less competition.

A large portion of business owners received control of a business between 2000-2007 and 2013-2017. There is a dip in 2008-2012, likely due to the 2008 recession.

Year Owned the Business by Race (Minorities)

