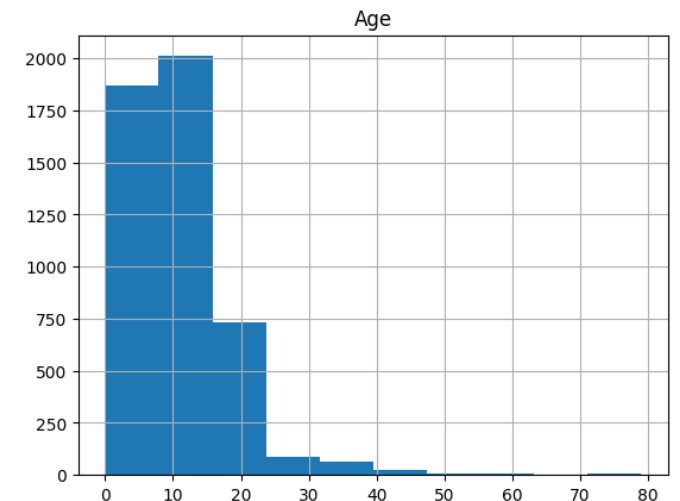


Demographics Analysis Report

Resident population

- 55% of our Resident population are female and 45% of them are male
- 29% of our Resident population does not have children in the house
- Resident's age range from 0 to 79 years (M=10.6, STD =7.8). The age distribution is skewed right, here are other insights:
 1. There is a 2:1 Ratio of 0-12 year olds to all other age combined
 2. 58.8% of the newborns (age 0-1) are female
 3. We see the most number of 7-year-old kids in the population, and the 2nd highest number be 5-year-old kids.



Monthly Income distribution

- In total , female earns \$289,141.77 and male earns \$277,133,22
- On average, a woman makes \$3,755.1 and a man makes \$4,398.9 each month

Household with no children makes \$3,092.7 monthly income on average and those with children makes \$4,466.9 monthly income on average

Total income with respect to Age:

12-year-old residents have the highest monthly income, roughly 100K. The next highest monthly income is reported by 8-year-olds (77K) and then 5-year-olds (53K) .

| Age | |
|-----|-----------|
| 3 | 10130.36 |
| 4 | 35956.39 |
| 5 | 53525.00 |
| 6 | 19764.66 |
| 7 | 36782.39 |
| 8 | 77551.00 |
| 9 | 25324.36 |
| 10 | 26983.08 |
| 11 | 14623.36 |
| 12 | 105892.00 |
| 13 | 26757.39 |
| 14 | 26126.00 |
| 15 | 20800.00 |
| 16 | 8500.00 |
| 17 | 801.00 |
| 18 | 18320.00 |
| 19 | 4166.00 |
| 20 | 25580.00 |
| 21 | 20558.00 |
| 27 | 2160.00 |
| 73 | 974.00 |
| 79 | 5000.00 |

Name: Monthly Income, dtype: float64

Demographic distribution – by Market

- The majority of our resident population are spread between the East and Southwest market, approximately 1120 in the East and 1190 in Southwest. In the East, the majority of the population are spread between North Carolina and Georgia. As for the Southwest market, Arizona has the highest population.
- As for other markets (Bay Area, Central, SoCal and Washington) , we see an average of 540 residents in each area. The Sacramento market has the least number of people in our resident population.
- Of all states, Oregon is least popular for our residents.

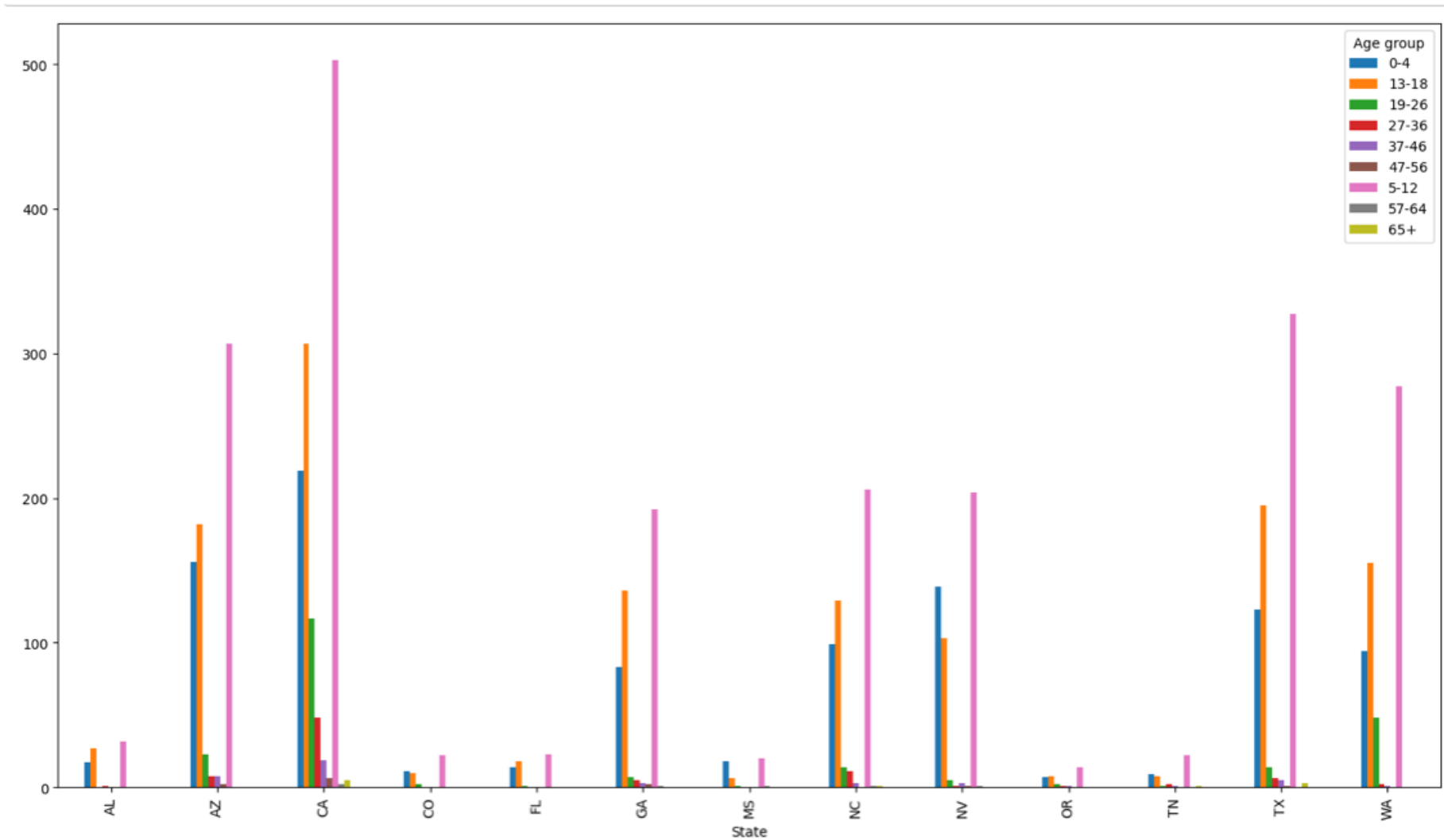
| Market | State | |
|------------|-------|-----|
| Bay Area | CA | 482 |
| Central | TX | 674 |
| East | AL | 77 |
| | FL | 56 |
| | GA | 429 |
| | MS | 46 |
| | NC | 464 |
| | TN | 44 |
| Sacramento | CA | 340 |
| SoCal | CA | 404 |
| Southwest | AZ | 686 |
| | CO | 45 |
| | NV | 457 |
| Washington | OR | 33 |
| | WA | 577 |

Demographic distribution – by Age

- Most states have the highest number of residents with ages 5-12, followed by those with ages 13-18, and then those who ages 0-4.
- However, in Nevada, residents who ages 0-4years outnumbered those between 13-18 years.
- People who are 65+ reside in the Bay Area, Central, East and SoCal Market only. The oldest resident (79 year-old) resides in SoCal.
- Mississippi has a median age of 6, the youngest across all states. Alabama has a median age of 11, which is the oldest across all states.
- California has more than 500 residents in age 5-12yrs. Recall from earlier that we found the most number of residents are 7-year-old. It turns out that the a quarter of them are located in CA.

The graph on the next page shows distribution by age.

Demographic distribution – by Age



Tackling missing Gender information

Gender is not complete in this dataset, we can only rely on first name to estimate gender. By observing that female and male names have distinctive characteristics, there are multiple ways we can approach this:

1) Observe the last character of the name. (This approach is being used for our analysis)

Names ending in a, e, i, h, and l i.e. Carla, Jessie, Sarah, and Carli, are likely to be female, while names ending in k, o, r, s and t i.e. Mark, Carlos, and Sergio are likely to be male.

2) Count the number of syllables in the first name.

Market research and linguistic studies have shown that 1) of all female names from 1880 to 2017, 29% have three-syllable names compared to 11% for male names, and 2) 21% of male names have only one syllable, compared to 6% of female names [<https://journals.linguisticsociety.org/proceedings/index.php/PLSA/article/download/4741/4354>]

We can set further criteria in our gender identifier to detect female and male names based on the number of syllables, and in term of code writing, we will count the number of vowels in the name and use the result to classify a name

3) Make use of NLTK Data in Python [`\nltk_data\corpora\names`]

The python NLTK has existing data for male and female names in two exhaustive text files. We can map the first names in our demographic csv to the two text files and a match between the two files will give us the right classification. In cases when there is no match, we can add a combination of the two methods discussed above to create our gender classifier.

Recommendations for marketing campaign for pet insurance

Percentage of resident population with pets: 23.5%

- Target audience: Households with little girl aged 7-13 years old [Fig.1]
note: Households with children are almost twice as likely to have pets
- Location : 1) CA (50% Bay Area and 50% SoCal) [Fig.2][Table1]
2) Arizona (as the most popular state in the East market for pets) [Table1,2][Fig.3]
3) North Carolina (as the most popular state in the Southwest market for pets) [Table1,2][Fig.3]

Other recommendations :

- Average cost of pet insurance? \$45/ month
- Relationship between cost and audience group's monthly income. Consider our data:
 - 1) Monthly income is not complete
 - 2) Given our young resident population, the monthly income reported likely comes from the parent rather than the person whose name listed in the data table him/herself

Due to the two reasons above, monthly income is not an informative and inclusive attribute for us to study and find the persons. However, based on our existing data, it is beneficial for us to set up different tiers of the pet insurance to better tailor our audience group in different regions:

- a) Accident & Illness Coverage \$45/month
- b) Accident-Only Coverage \$15/month

This way we offer more flexibility for our audience in regards to their different financial capabilities.

Fig.1

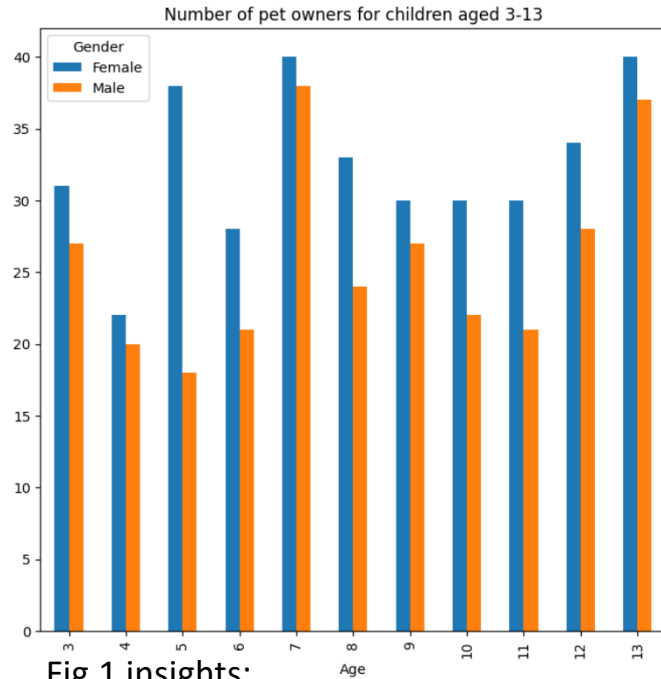


Fig.1 insights:

Female outnumber male in all age from 3 to 13. In addition, we see the most pet owners for people with age 7, 5, and 13.

Table 2

Top counts by age in AZ & NC

| AZ | | NC | |
|-----|----|-----|----|
| Age | | Age | |
| 7 | 15 | 7 | 14 |
| 8 | 13 | 14 | 12 |
| 16 | 11 | 9 | 11 |

Fig.2

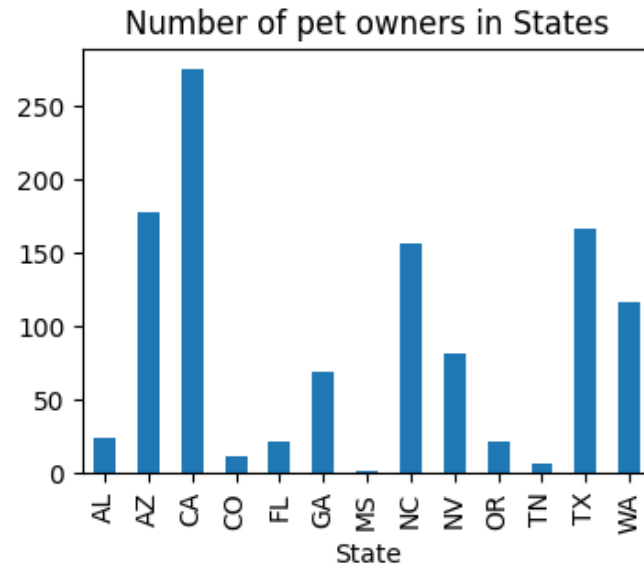


Table 1

| Market | State | Count |
|------------|-------|-------|
| Bay Area | CA | 96 |
| | TX | 167 |
| | AL | 24 |
| Central | FL | 22 |
| | GA | 69 |
| | MS | 2 |
| East | NC | 156 |
| | TN | 6 |
| Sacramento | CA | 83 |
| | CA | 96 |
| SoCal | AZ | 178 |
| | CO | 11 |
| Southwest | NV | 81 |
| | OR | 21 |
| Washington | WA | 116 |

Insights from Table 1:

Bay Area, Sacramento and SoCal all ties back to CA. Thus CA is the state with most pet owners. We are highly likely to find our target audience here.

Fig.3

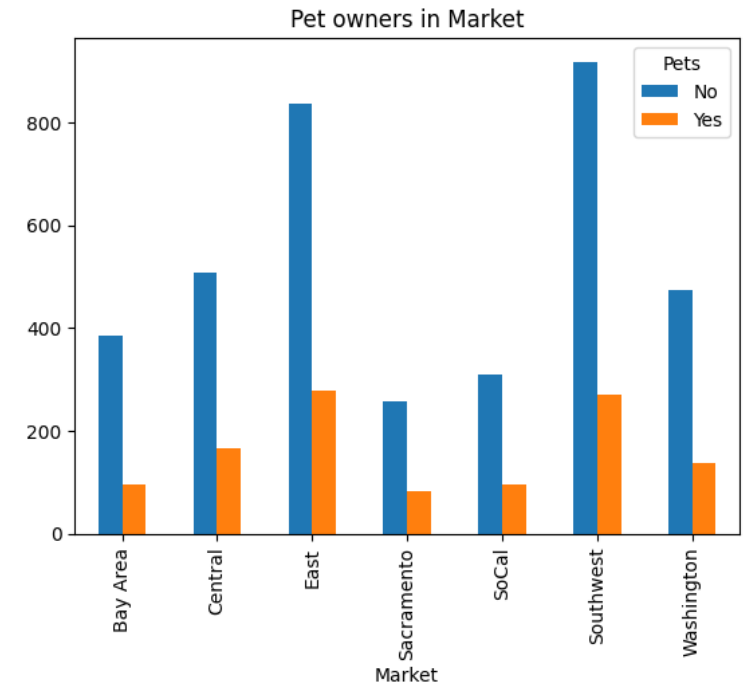


Fig.3 insights:

The East and Southwest market has most pet owners.