**Part 1 - Clean and Analyze the Data**

First Glance: The dataset is a sample of Floridians who currently utilize SNAP benefits. The content is in the form of categorical variables with the exception of age. Qualitative values recorded are ethnicity, gender, county, employment status, primary language, household size, and education level.

I predict that primary language and gender will be important for my analysis. Primary language matter's because a potential language barrier could lead to a systematic underrepresentation. Also, gender may be important because societal expectations for men to provide for the household may lead to them being less likely to ask for federal assistance.

Cleaning the data: I had to clean the data reported for ethnicity. In many cases, there were different categories for the same racial bucket, so I had to edit that value. For example, I changed 'Asian' values to 'asian' and "African American" to "black. Additionally, I had to find a way to evaluate the hispanic demographic. I decided to create the following new buckets: non-hispanic white, total-white, non-hispanic black, total-black, and hispanics. These buckets should allow for a comprehensive analysis.

Visualizations:

Pie Chart: SNAP users by gender
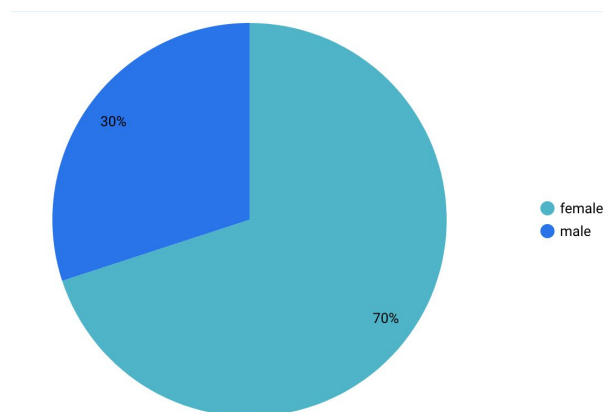
Pivot Table Heat Map: Native language at by ethnicity and gender

ethnicity	gender	english	spanish	other
white	female	6,906	988	255
	male	2,973	418	102
black	female	2,399	322	95
	male	1,012	125	34
Native American and Alask...	female	1,189	168	40
	male	519	79	20
asian	female	670	105	29
	male	307	40	16
Pacific Islander	female	379	57	9
	male	147	21	5
Other	female	323	39	9
	male	168	27	5

Pivot Table Heat Map: Household size by ethnicity and gender

ethnicity	gender	1	4	3	2
white	female	4,062	1,664	820	798	805
	male	1,778	683	346	360	326
black	female	1,383	631	293	251	258
	male	607	214	109	109	132
Native American and Alask...	female	700	278	136	156	127
	male	339	109	58	54	58
asian	female	406	158	95	76	69
	male	187	77	33	39	27
Pacific Islander	female	213	95	49	38	50
	male	80	35	23	18	17
Other	female	205	69	30	34	33
	male	102	39	15	22	22

Insights: Looking at the tables, there are clear disparities in representation of gender and household size in the sample. There may be a potential underrepresentation of non English native speakers.

Gender is distributed in Florida with 51.2% and 48.8% of the population being women and men respectively. The ratio of female and male users is 7:3 which suggests that men are highly underrepresented in SNAP users. As a result, I controlled for gender going forward in my analysis.

Next, householders living alone make up 26.6% of Florida residents; however, when considering white female users who live alone, that demographic makes up 49.85% of the users. A further look into the distribution of single sized households shows it proportionally outweighs all other values regardless of race or gender. It is evident that larger households are not involved in the SNAP program as equally as single households.

Lastly, the native Spanish and English speakers in Florida make up 19.54% and 73.36% of the population respectively. However, these numbers are not representative of the SNAP users. After controlling gender and ethnicity, we can look in the case of white female SNAP users. The native Spanish speakers in this group only make up 12.12% of the total while native English speakers make up 84.75% of the users.

In these cases, the population of recipients is made up of a smaller percentage than their

proportional representation with the Floridian population that qualifies for SNAP. The most evident differences are found in household size and gender.

Part 2 - Synthesis of User Pain Points

Pain Points:

- Unclear marketing
    - Users are unsure what the program is, what it does, and how it can benefit them. There is a lack of familiarity with the program, and people who could benefit from it are unaware of its existence. The branding and advertising aren't direct enough nor do they reach a diverse audience.
    - Consider the quote about the individual who felt as a Libertarian, that he was unable to receive the benefits of the program. SNAP is failing to meet individuals who cultural and/or ideological leanings hinder their desire to enroll. Those populations are still in need of the aid; however, the way it is marketed to them is unconducive of success.
- Unclear communication
    - Program rules and eligibility are not easily understood by users. Overall there seems to be an insufficient amount of information being communicated about the program. Additionally, there is a lack of transparency in the application process. Individuals may be unsure which documents are needed and why some may be denied.
- Barriers of use
    - Many users feel as though the current method is inaccessible. A lengthy application, digital illiteracy, and unsafe physical offices are some of the given examples to why some are not enrolled in the program. It is evident that it is not easy for every individual to apply and access the benefits. There is a need for an improved application process to improve the user experience.
- Unconsidered 'edge cases'
    - The program is not reaching everyone who needs to be supported. In the example of the nurse, we see that not everyone who needs to be supported by the program is currently eligible. Though the nurse's salary makes her unqualified to receive

benefits, her situation still necessitates federal assistance. The selection and approval criteria should be reevaluated to include all edge cases and holistically evaluate needs, especially in the case of a large household size.

Part 3 - Other Data Required

I would like the following data to better identify key leverage points:

- Data on the digital and in person application process
  - I picked this data to run a funnel analysis and determine in which step in the sign up process are people dropping. From that data, one could see where the large reductions in applicants occur and better target improvements. The necessary documents or questions that stump users can be determined.
- Data on denied applicants
  - This information will help determine if there are any patterns of certain groups of people being constantly denied. That insight may indicate an unconsidered "edge case". In that scenario, the insights would suggest underrepresented populations. Additionally, should there be policy changes, it would be a good idea to have this data to reach out to people now eligible.
- Data on the dispersal of information
  - This data would tell how/where people are receiving information about the program. From this, insight could be gained into what is working and what is not to ultimately better focus efforts of communication. A result from this data could lead to better targeted marketing for a more diverse population.

Part 4 - Propose a Hypothesis

Hypothesis: Men are systematically underrepresented in those receiving benefits because of current branding of the SNAP program, so improving the communication to men will help this finding.

Based on the user quotes, unclear communication to potential participants was demonstrated. Since the data suggests men are less likely to receive the food stamps than women, men may be unaware of how SNAP can benefit them. It might take a different approach to reach a male

audience due to a difference in gender roles placed on to men. Since society teaches men to provide and be self reliant, they may find it difficult to accept the information as is presented now. As a result, the intervention for this group is to change the way in which SNAP is communicated to them.

In order to test the key parts of my hypothesis, there are certain metrics to consider. The purpose is to develop target interventions for those who are not utilizing SNAP, and the goal of the hypothesis is to increase usage specifically for men. As such, there are two buckets of metrics to consider; I'd consider how my hypothesis is improving men's participation as well as their overall understanding.

We can measure their participation by looking at engagement with the program. We can define good engagement for this scenario by considering the total number of completed applications in a given week and how frequently each application is approved. From these measurements, we can compare the rate of men joining to prior data and determine whether that change is significant. Long term good engagement means a percentage of male users the same proportion as their representation within the larger population.

In order to track men's understanding of the SNAP program we can directly survey and conduct user research on our target population. The survey data would allow us to get feedback on the communication used directly from the users we are targeting by using a control versus exposed methodology. This process would be surveying a mix of people to determine if we are reaching and having the desired effect on the target audience. Moreover, A/B or multivariate testing can be considered as well. As a result, we could see if perceptions are shifting and understand which attributes of the marketing are most effective.

A counter metric is to monitor that the number of female SNAP users does not decrease. A concern is that in targeting men, women may not receive our communication well, so it is important to ensure that we do not alienate any other subset of users.

Part 5
How long did you spend on the work product assignment?
<2 hours