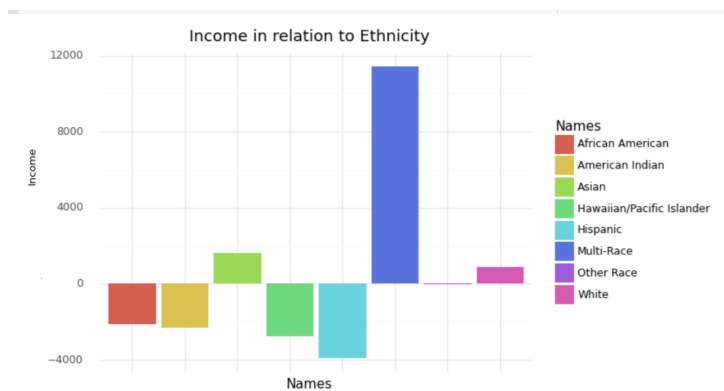


Analytical Plan and Recommendation:

For this assignment, the tactic that was used was going through all the data individually, looking for what to combine and analyze and then cross checking the results to find similarities.

The data used were the DMA data sets (PI-18803 DMA.xlsx), the MSA files where all the MSA.xlsx files were combined into one sheet, the stores data-set DMA.xlsx, the sales data-set.csv, and then a combined version of the sales data-set.csv and features.csv. The first dataset that was looked at was the DMA PI-18803.

In order to be able to compete with competitors, placing one store in a rich DMA/MSA area with high incomes could be beneficial because people will buy more, and the store will not go out of business. Below is a linear regression model along with a coefficient bar graph which was created in order to understand the relationship between income and ethnicity. The most positive correlation based on the statistics was Multi-Race.



More analysis was done on income and household count because these were crucial factors against competitors in order to have sales in the market. This was the DMA rankings for largest household counts with 2+ vehicles in relation to the largest incomes:

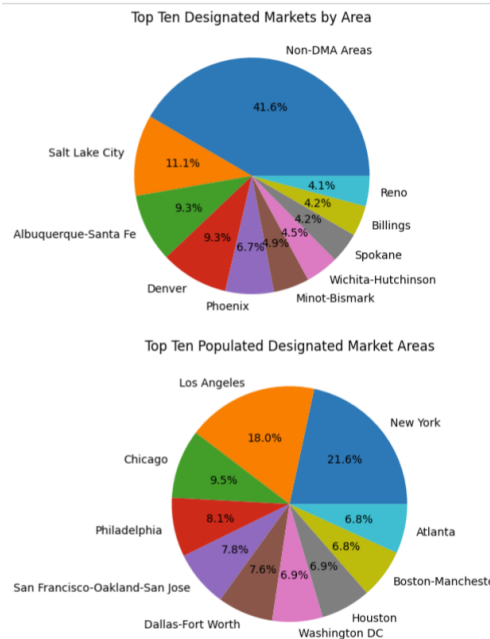
- The top 10 places with the highest income, with 2+ vehicles, and highest household count are:

	index	DMA	Name
196	196	807	San Francisco-Oakland-San Jose
168	168	747	Juneau
12	12	511	Washington DC
164	164	743	Anchorage
7	7	506	Boston-Manchester
166	166	745	Fairbanks
165	165	744	Honolulu
13	13	512	Baltimore
2	2	501	New York
200	200	819	Seattle-Tacoma

When taking Multi-Race ethnicities into the account, these were the top DMAs:

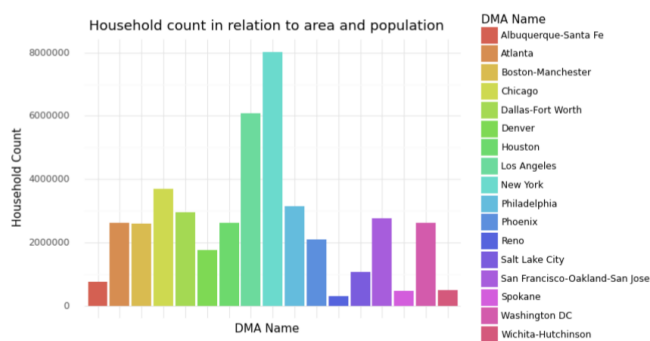
	index	DMA	Name
194	194	803	Los Angeles
2	2	501	New York
165	165	744	Honolulu
196	196	807	San Francisco-Oakland-San Jose
200	200	819	Seattle-Tacoma
12	12	511	Washington DC
207	207	862	Sacramento-Stockton-Modesto
172	172	753	Phoenix
86	86	602	Chicago
5	5	504	Philadelphia

Area and population were factors that were analyzed next because the more area a place has, the more room to build more markets. As well as this, the more populated a place, the more people over 18+ that will be willing to come to a certain market. Two pie charts were created to analyze this data in relation to the DMA name. Although I did not narrow down to just these places, I kept these twenty in mind.



After this, the values that were taken into account were household count in relation to area and population in different DMAs since the more people there are in a household, the more they will spend at a store. Then, the median household count overall in this dataset was found. The median was 301,762, so anything greater than this number when analyzing the data would be ideal. The population over 18+ values and the area of the top 20 DMAs found above were measured in relation to household count if the household count was above 301,762. After analyzing this data and creating a bar graph for the 20 places listed above in relation to area and population as well as household count being greater than the median, the list of prominent DMAs from this file was narrowed from 20 to 17.

Number of observations in DMA1x2: 17



The next priority in this dataset was looking at the modes of transportation that were overall the most prominent to see how people would prefer to get to a store if a new market opened. In order to do this, a new column was created in order to sum up all of the different transportation types and see which one is the most prominent overall. Overall, the most prominent was people taking 2 or more vehicles. I will take this information into account after looking at more analysis from different data sets.

```

HHlds No Vehicles          5082993.0
HHlds 1-2 Vehicles        14146151.0
HHlds 2+ Vehicles         24845314.0
HHld Exp - Public Transport  15283.0
HHld Exp - Intercity Bus Fare  449.0
HHld Exp - Mass Transit     2629.0
HHld Exp - Taxi            586.0
HHld Exp - Other Public Transportation 1775.0
dtype: float64
The most prominent transportation mode in the dataset is: HHlds 2+ Vehicles

```

The next dataset that was analyzed was the MSA dataset in order to see which areas were most prominent and analyze the items that affected the average annual expenditure the most. Before I did this, I looked at the areas in the MSA file in comparison to my top 17 areas from the DMA files. The similar ones included New York, Philadelphia, Boston in the northeast, Chicago in the Midwest, LA, San Francisco, and Denver in the West, and Washington DC, Atlanta, Dallas, and Houston in the South. In regards to looking at characteristics of an area of what would make a good store, averages of the average number in a consumer unit, adults 65 and older, vehicles, and earners overall were taken in order to get a better idea of which demographics were going to the stores. The averages were:

Average number in consumer unit: 2.475 (Highest in the west)

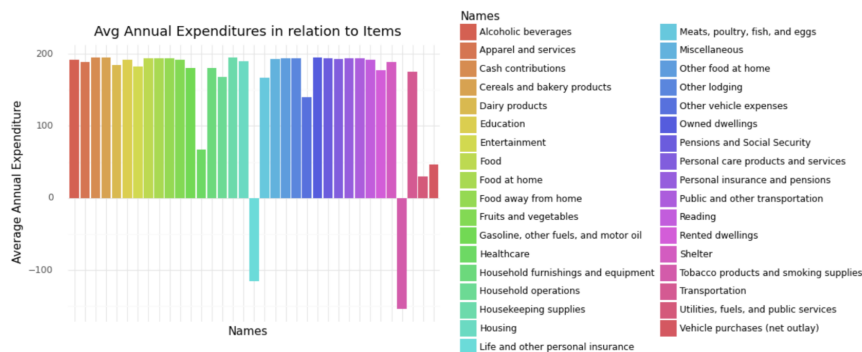
Adults 65 and older: 0.4 (Same across)

Vehicles: 1.9(Highest in Midwest and west)

Earners: 1.3(Highest in West)

In order to see the influence on expenditure, a linear regression model was created which can be seen below. Items that had negative effects were tobacco products, alcohol, and meats, poultry, and fish, so I knew I would not want to place markets where this was prominent. Cereals, bakery products, food in general, food at home, fruits and vegetables had positive effects.

Then I looked at the DMA stores dataset just to see which stores here had the biggest sizes, and then I wanted to compare them to the stores in the sales data set. The most significant stores are shown below.



index	Store	Type	Size	DMA
0	0	13	A 219622	Los Angeles
1	1	11	A 207499	Los Angeles
2	2	28	A 206302	Dallas-Ft. Worth
3	3	4	A 205863	Houston
4	4	27	A 204184	Houston
5	5	19	A 203819	Philadelphia
6	6	24	A 203819	Atlanta
7	7	31	A 203750	Atlanta
8	8	20	A 203742	Tampa-St. Pete
9	9	32	A 203007	Denver

A bar graph was then created with the sales-data set below to see the comparison between weekly sales and stores to see which stores are most profitable and compare them to DMA stores. In this bar graph, I found that some of the stores with the highest sales were 2, 4, 10, 13, 14, 20, 27, 6, 1, and 39 similar to what I found in the DMA

dataset. These stores are located in Cleveland-Akron, Houston, LA, Dallas Fort Worth, Tampa St.Pete, Orlando-Daytona Beach-Melbourne, and Chicago.



Then, the files for the sales dataset and the features dataset were combined and a weighted average was taken to see what values would be the most significant temperature wise. This factor is important because people tend to settle down and start families in places with good temperatures. Results shown below:

#Weighted Temperature: 68.06682396

To summarize, overall what I found after all this analysis was:

- **The 17 prominent DMAs from the DMA data set in regards to household count, area, and population: Albuquerque-Santa Fe, Atlanta, Boston-Manchester, Chicago, Dallas-Fort Worth, Denver, Houston, Los Angeles, New York, Philadelphia, Phoenix, Reno, Salt Lake City, San Francisco-Oakland-San Jose, Spokane, Washington DC, Wichita-Hutchinson. (DMA/MSA results)**
- **All of the prominent stores after combining the results from the DMA stores dataset and sales dataset were located in Los Angeles, Dallas-Fort Worth, Houston, Philadelphia, Atlanta, Tampa-St. Pete, Denver, Cleveland-Arkon, Orlando-Daytona, and Chicago. (DMA stores/sales results)**
- **Characteristics a market area would need to compete against established competitors according to my analysis:**
 - **Areas where having cars is prominent, especially either one or two pls vehicles would be beneficial.**
 - **Average number in consumer unit: 2.475**
 - **Adults 65 and older: 0.4**
 - **Vehicles: 1.9**
 - **Earners: 1.3**
 - **Cereals, bakery products, food, fruits and vegetables are services that have positive effects over alcohol. Meats, poultry, fish and tobacco products had not so positive to negative correlations with average annual expenditure.**
 - **This shows that most of these characteristics seem to describe families since the average people in a house is above 2, there are more than 2 cars, and more than**

one earner. Furthermore, alcohol is not very positive but cereals and healthy options are showing the residents maybe of older age or in a family once again. Although the areas with older residents above 65 were only .4, so I gave priority to younger adults.

- Overall, based on the weighted averages, most value was given to a temperature of about 68.
- After analysis, comparing the DMA/MSA results with the DMA store/sales results, the stores that were the same in both were Los Angeles, Dallas-Fort Worth, Houston, Philadelphia, Atlanta, Denver, and Chicago.
 - Overall, the average temperature was 68 while fuel price was 3149 and CPI of 1966203417 when it was not a holiday.
 - Other characteristics looked for:
 - Highest vehicle(HHId) rate: LA, Philadelphia, Chicago(DMA dataset)
 - Lowest to Highest meat consumptions
 - Lowest to Highest alcohol/tobacco
 - Highest to Lowest family food consumption of food/food at home/fruits and vegetables
- A sum of all the bad items in each state(tobacco, meats/poultry/fish, and alcohol) and a sum of all the good items(food, cereal/bakery products/fruits and vegetables) in each state found from the MSA linear regression graph was taken. I wanted to rank the states by overall sentiment, so the difference between the positive sum and negative sum. For example, I added up the meat consumption value as well as tobacco value and alcohol value together which are the negatives and subtracted from the cereal and other food values which are the positives for each state to see if the difference was negative or positive. I found the values, for example the number value associated with tobacco for a certain state from the MSA file. The more positive or higher the number, the better the sentiment.
 - Los Angeles: 8,386
 - Dallas-Fort Worth: 5,977
 - Houston: 8,629
 - Philadelphia: 8,042
 - Atlanta: 6,301
 - Denver: 8,356
 - Chicago: 7,941
- I picked the top five which were Houston, Denver, Los Angeles, Philadelphia, and Chicago, and Los Angeles, Philadelphia and Chicago also were highest for vehicles.
 - Then I checked the weighted averages of each individual state by picking a store they did well in from the graphs above to make sure that they matched closely with the overall weighted average above.
 - Houston(store 4)
 - Weighted temperature: 67.898
 - Denver(store 32)
 - Weighted temperature: 68.2214
 - Los Angeles(store 13)
 - Weighted temperature: 68.09268
 - Philadelphia(store 19)
 - Weighted temperature: 68.10733
 - Chicago(store 39)
 - Weighted temperature: 68.37089

Overall, each state is pretty consistent in being close to the value of 68 for weighted temperature showing they have ideal temperatures for stores. The temperature is overall very stable and consistent throughout the data given. Therefore, I do so far recommend Los Angeles, Philadelphia, Chicago, Houston, and Denver.

- The other prominent places I found after analyzing the datasets which can be seen in the graphs above were Albuquerque-Santa Fe, Boston-Manchester, Phoenix, Reno, Salt lake City, San Francisco-Oakland-San Jose, Spokane, Washington DC, Wichita-Hitchinson, Tampa-St. Pete, Cleveland-Arkon, New York, and Orlando-Daytona.
- The top 10 places with the highest income, with 2+ vehicles, and highest household count(this was important because it would mean there could be multiple earners above 1.3 in the household as well as the average consumer unit being above 2.5) are San Francisco-Oakland-San Jose, Juneaeu, Washington DC, Anchorage, Boston Manchester, Fairbanks, Honolulu, Baltimore, New York, Seattle-Tacoma from the earlier analysis.
 - When comparing these to the ones in the bullet above(other prominent places), the same ones are San Francisco-Oakland-San Jose, Washington DC, Boston Manchester, and New York.
- When Multi-Race ethnicities were taken into account in order to have a store in a rich market, the top were Los Angeles, New York, Honolulu, San Francisco-Oakland-San Jose, Seattle-Tacoma, Washington DC, Sacramento-Stockton-Modesto, Phoenix, Chicago, Philadelphia.
 - When comparing these to the ones in the bullet above(other prominent places), the same ones are San Francisco-Oakland-San Jose, Washington DC, Phoenix, and New York.
- Therefore, I chose to recommend the most common places that were repeated in each result with high household count, income, and other important factors stated above. These places are San Francisco-Oakland-San Jose, Boston-Manchester, Phoenix, New York, and Washington DC.

Overall, the top best market places I would recommend are Houston, Denver, Los Angeles, Philadelphia, Chicago, San Francisco-Oakland-San Jose, Boston Manchester, Phoenix, New York, and Washington DC.