

Assignment 4: Retail Location

See code [here](#)

After an analysis of data pertaining to sales metrics across a chain of multipurpose stores as well as DMA and MSA data pertaining to demographic information, I recommend placing new stores for our company in the following 10 locations: Portland (OR), Seattle-Tacoma, Bend, Salt Lake City, Reno, Denver, Cheyenne-Scotts Bluff, Los Angeles, Sacramento-Stockton-Modesto, and San Diego.

Overview/Background

The purpose of this analysis is to create an analytical plan for the launch of retail locations for a company that falls somewhere between a Wal-Mart and Target level multi-purpose home and grocery store. The consumers we are trying to reach are middle to high income shoppers. It's difficult to pin down exactly what constitutes a middle or high income salary, but for our purposes, we'll start with the number \$65,000. According to the U.S. Census Bureau, the median income for the United States was \$65,000 in 2021. Middle class is defined by Pew Research Center "as households that earn between two-thirds and double the median U.S. household income" (CITE). This means that middle income is between \$43,350 and \$130,000.

Analysis

To conclude the aforementioned 10 locations, I began with an exploratory data analysis of the store data. After dropping columns I deemed would be not useful due to multiple missing values and joining the sales, stores, and features tables, I began examining different aspects of and relationships within the data.

The first thing I wanted to know was which stores had the highest and lowest average weekly sales. As shown by the radial bar chart in Figure 1, stores 20, 4, 14, 23, 2, 10, 27, 6, 1, and 39 have the top 10 highest average weekly sales while stores 5, 33, 44, 3, 38, 16, 29, 7, 36, and 30 have the lowest. (See code for simple bar chart of the same information).

From there, I found it important to identify which DMAs these stores were situated in. Looking at Figure 2, stores in Houston and Cleveland dominated half of the highest performing stores with additional stores in Florida and other parts of Texas making up another portion. Only states east of the Mississippi, with the addition of Texas, made the top 10. When looking at a larger sample of the 25 highest selling stores, we see a similar trend in location with a couple more stores ad Texas and other areas already in the top ten with the addition of four stores in Los Angeles and a location in Denver making the list.

Figure 1. Radial bar chart displaying average weekly sales for each store (store 20 has the highest weekly sales and 5 has the lowest)

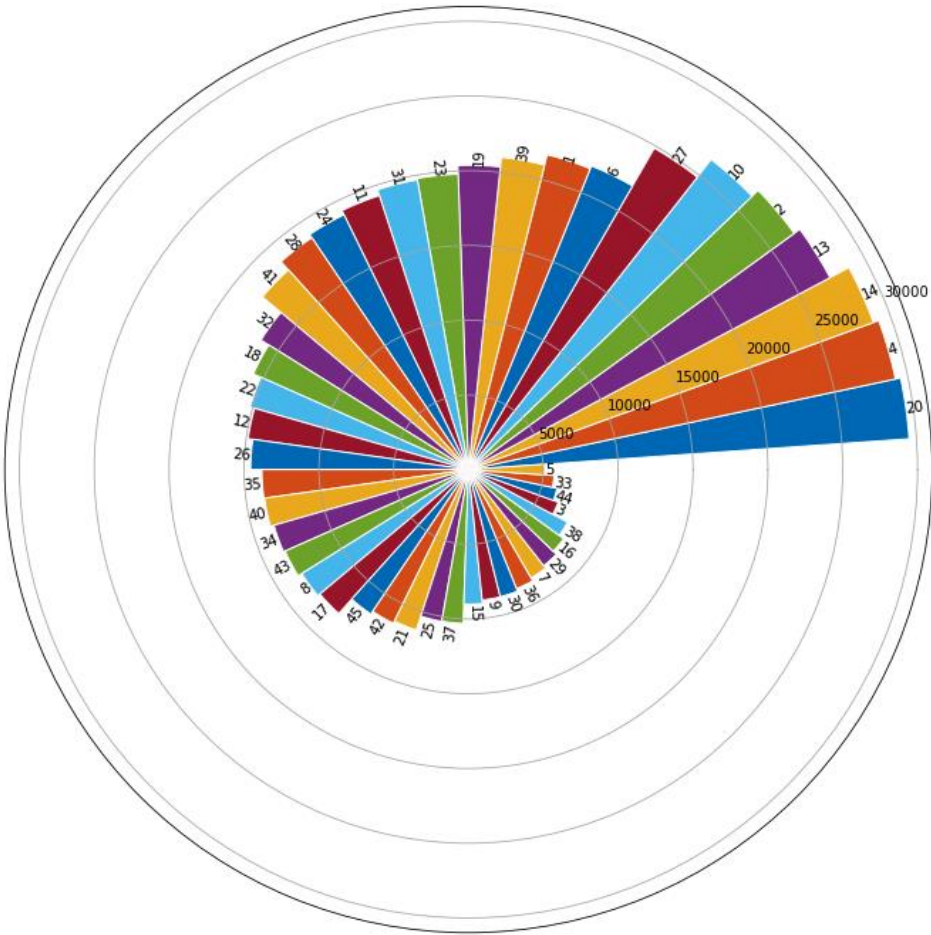


Figure 2. Stores with highest average weekly sales and their DMAs

Store				DMA	
Store	Type	Size		DMA	
3	4	A	205863	Houston	
4	27	A	204184	Houston	
8	20	A	203742	Tampa-St. Pete	
10	6	A	202505	Orlando-Daytona Brach-Melbourne FL	
11	2	A	202307	Cleveland-Akron	
12	14	A	200898	Dallas-Ft.Worth	
14	39	A	184109	Chicago	
19	1	A	151315	Houston	
22	10	B	126512	Cleveland-Akron	
28	23	B	114533	Philadelphia	

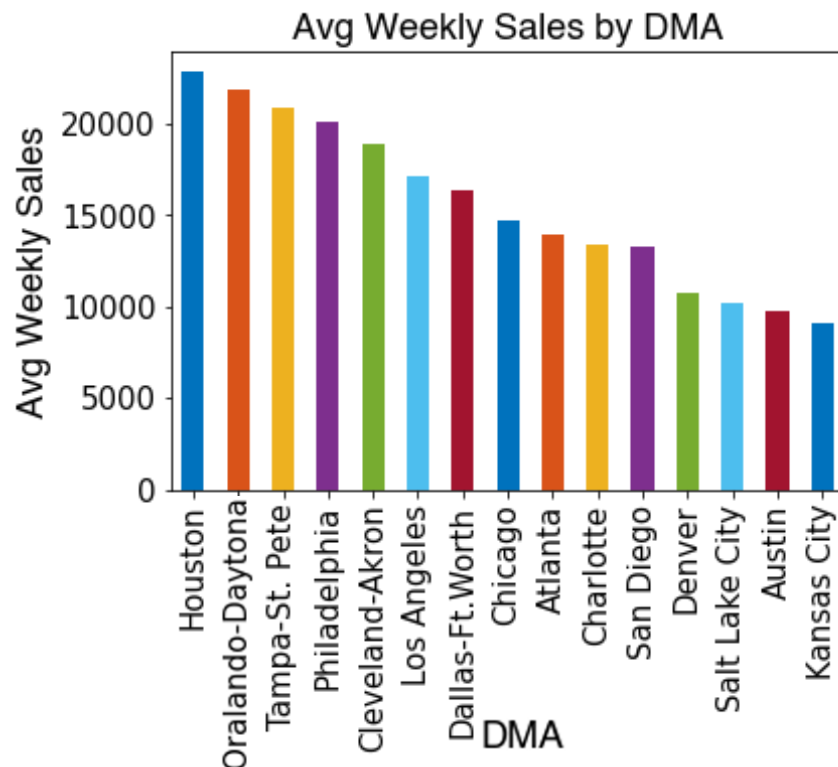
Almost surprisingly, when looking at the stores with the lowest average weekly sales (Figure 3), we see some of the highest selling and some of the lowest selling stores in the same DMAs. This is even more apparent when looking at the 20 highest performing stores.

Figure 3. Stores with lowest average weekly sales and their DMAs

	Store	Type	Size	DMA
31	29	B	93638	Chicago
33	7	B	70713	Cleveland-Akron
34	16	B	57197	Dallas-Ft.Worth
35	30	C	42988	Los Angeles
37	36	A	39910	Salt Lake City
39	44	C	39910	Atlanta
40	33	A	39690	Atlanta
41	38	C	39690	Kansas City
43	3	B	37392	Denver
44	5	B	34875	Austin

Because of the overlap in highest and lowest performing stores being located in the same DMAs, this analysis doesn't provide us much useful information. The next step was to look at the average weekly sales across all stores within a given DMA as shown in Figure 4.

Figure 4. Bar chart illustrating DMAs and average weekly sales



This is a much better indicator of locations that have the highest sales performance. There are only 15 DMAs in the store data set, but the top 10 follow the trend of being all east of the Mississippi with the exception of locations in Texas as well as Los Angeles.

The next step was to begin identifying the cities I wanted to target by income. I started by creating a circle plot to illustrate the different incomes of the different locations in the larger DMA resource we were provided. As shown in Figure 5, San Francisco, Juneau, and Washington DC have the highest income in the country. Despite being ordered and having the highest income locations on the left, this graph is difficult to understand because of the realivly small differences in income. See Figure 6 for the same information (income for each DMA) in a more readable bar chart.

Figure 5. Circle plot of income for each DMA location

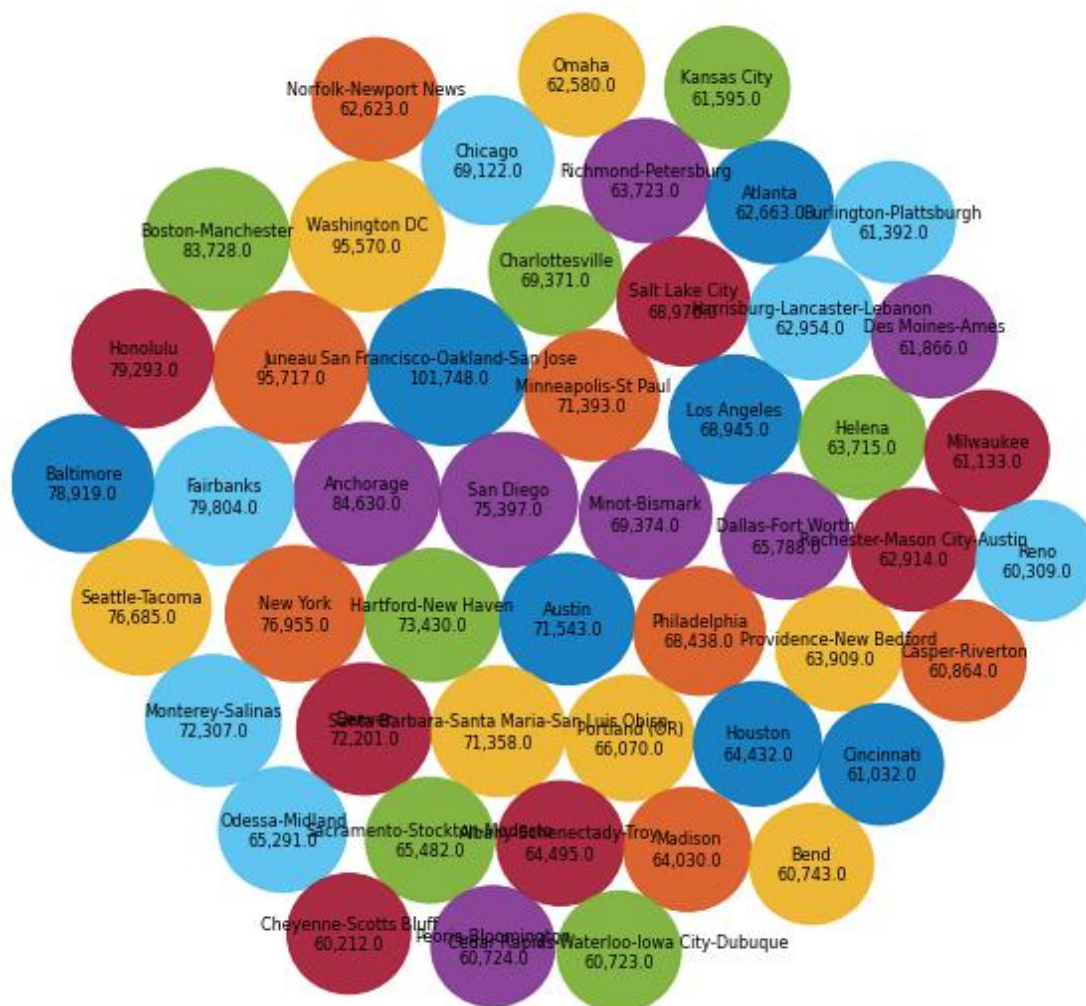
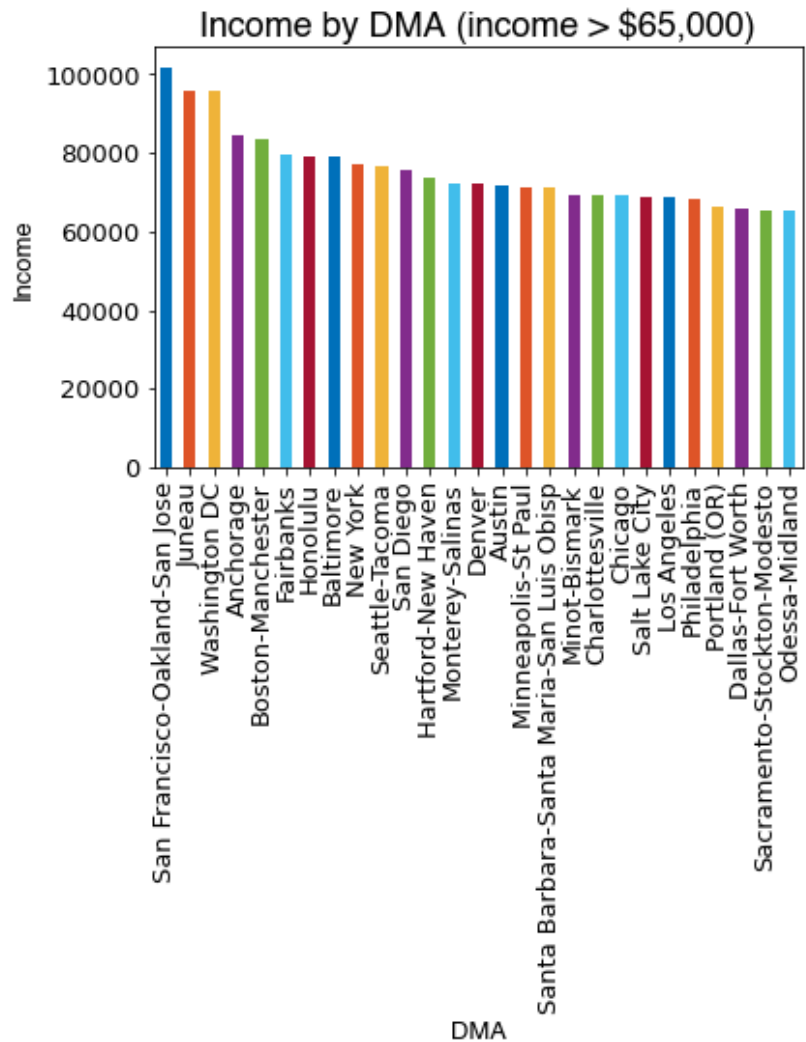


Figure 6. Bar chart of income for each DMA (for DMA's with an income greater than \$65,000)



I started by looking only at locations with an annual income of over \$65,000 as that was the median income for the United States. Despite this, many of our DMAs did not appear in this range. This makes sense when looking at the highest performing DMAs (Figure 7) as half of the incomes fall below the national median income

As the data being analyzed is Wal-Mart data and our retain chain is a little higher class yet still serving a very similar market, I imposed a minimum income of \$60,000 and a maximum income of \$80,000. This left me 45

Figure 7. Income for the top 50% of the DMAs from stores/sales data

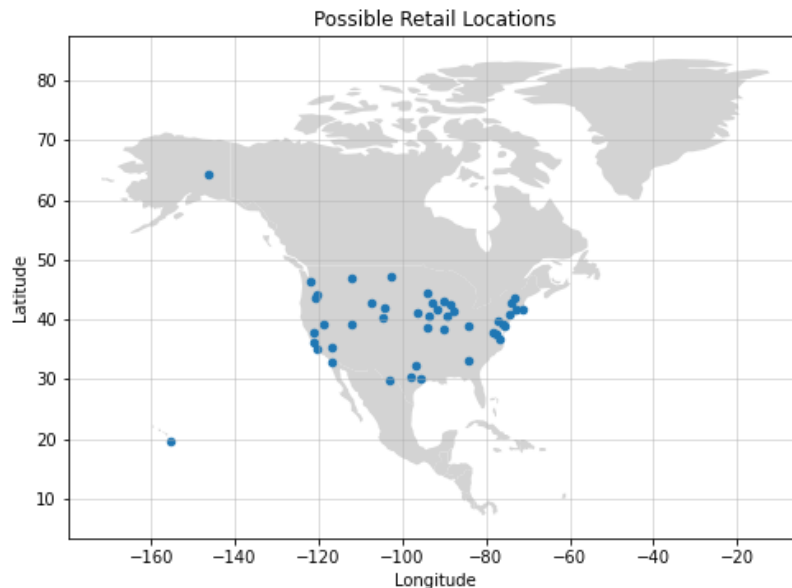
Houston	64432
Orlando-Daytona Beach-Melbourne	52928
Tampa-St Petersburg	52388
Philadelphia	68438
Cleveland	54042
Los Angeles	68945
Dallas-Fort Worth	65788
Chicago	69122

options to pick from. When looking at possible options, there are some being excluded on a location basis. For example, it will be difficult and costly to transport products to locations like Alaska or Hawaii as visualized in Figure 7. I would also like to try to contain the locations within geographical confines to mitigate transportation costs.

Possible retail locations:

'Fairbanks'	'Houston'
'Honolulu'	'Madison'
'Baltimore'	'Providence-New Bedford'
'New York'	'Richmond-Petersburg'
'Seattle-Tacoma'	'Helena'
'San Diego'	'Harrisburg-Lancaster-Lebanon'
'Hartford-New Haven'	'Rochester-Mason City-Austin'
'Monterey-Salinas'	'Atlanta'
'Denver'	'Norfolk-Newport News'
'Austin'	'Omaha'
'Minneapolis-St Paul'	'Des Moines-Ames'
'Santa Barbara-Santa Maria-San Luis Obispo'	'Kansas City'
'Minot-Bismarck'	'Burlington-Plattsburgh'
'Charlottesville'	'Milwaukee'
'Chicago'	'Cincinnati'
'Salt Lake City'	'Casper-Riverton'
'Los Angeles'	'Bend'
'Philadelphia'	'Peoria-Bloomington'
'Portland (OR)'	'Cedar Rapids-Waterloo-Iowa City-Dubuque'
'Dallas-Fort Worth'	'Reno'
'Sacramento-Stockton-Modesto'	'Cheyenne-Scotts Bluff'
'Odessa-Midland'	'St Louis'
'Albany-Schenectady-Troy'	

Figure 7. Possible retail locations plotted



This offers a couple of district groups: East Coast (Baltimore, New York, and Providence), Pacific Northwest/West (California DMAs, Seattle, Salt Lake, and Bend), South-East (Atlanta, Norfolk, and Texas), and Mid-West (Des Moines, Iowa City, St. Louis, Bismarck, and Nebraska).

I have chosen to go with the Pacific Northwest/West as there can be multiple locations in a somewhat reasonable area. In addition, there is less snow and fewer natural disasters than the Mid-West, South-East, and East Coast. Lastly, a couple of the highest performing stores fell within these DMAs. The final 10 locations are:

Portland (OR)	Denver
Seattle-Tacoma	Cheyenne-Scotts Bluff
Bend	Los Angeles
Salt Lake City	Sacramento-Stockton-Modesto
Reno	San Diego

Future Explorations

One of the major areas that was left unexplored with more specific demographic data. An option would have been to identify the customers we want to sell to and move forward from there. That was done a little in regard to income, but could have been more targeted to different genders, races, or ages. This could have utilized clustering as we could have looked at which DMAs have similar demographics of people.

I would have liked to fill out the map visualization in Figure 7. There are ways to make it interactive/easier to read and understand, but this is my first time doing it so I wanted to include it

Lastly, this is a visualization I made of the population in each DMA that I thought was cool (and how I hoped the first income visualization would have turned out).



Additional References

- <https://www.kaggle.com/c/walmart-recruiting-store-sales-forecasting/data>
- <https://www.investopedia.com/financial-edge/0912/which-income-class-are-you.aspx#:~:text=The%20Pew%20Research%20Center%20defines,make%20between%20%2443%2C350%20and%20%24130%2C000.>
- <https://towardsdatascience.com/9-visualizations-that-catch-more-attention-than-a-bar-chart-72d3aeb2e091>
- <https://towardsdatascience.com/the-easiest-way-to-plot-data-from-pandas-on-a-world-map-1a62962a27f3>