Homework 4
MGSC 410-01
Lily Annen

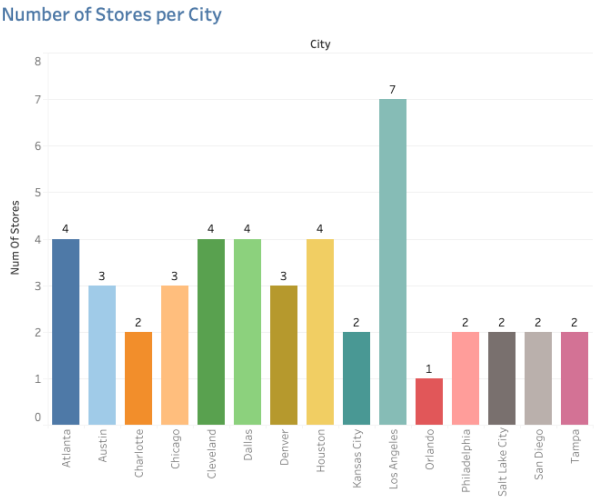
Launch Plan for Retail Startup

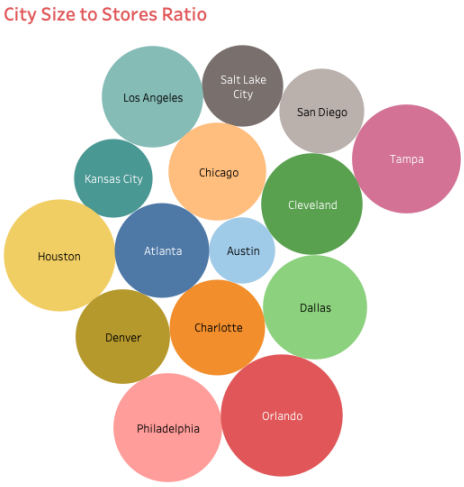
This launch plan recommendation will target the launch of a new retail startup to compete with stores such as Target and Walmart. Using Walmart sales data from 2011-2012, I will detail the best location for a startup to launch a new multi-purpose home and grocery store, ideally for middle to high income households. The datasets provided include geographic, metropolitan, retail, store, and sales data, as well as other features that may contribute to overall store performance. The main focus of this analysis is on the *Designated Market Area (DMA)* data, which includes the major city or metropolitan area where stores are located. The *Metropolitan Statistical (MSA)* data details insights into consumer expenditure in categories such as public transportation and average household expenses, and includes information on population composition, race, median household income, and age per city. The retail dataset is made up of data from 45 competitive stores, and the sales data contains dates, weekly sales per department and whether or not the week is a holiday. The features dataset gives insight to variables that may be related to consumer spending habits, including temperature, fuel cost in the region, promotional markdowns, the Consumer Price Index (CPI), and the unemployment rate at the time. At first glance at the all of the datasets provided, many have overlapping variables that leads to merging information. Another key takeaway that impacts overall consumer behavior is holidays. The holiday variable includes four major holidays: the Super Bowl (February), Labor Day (September), Thanksgiving (November), and Christmas (December). The promotional markdowns often precede prominent holidays, and these weeks are weighted five times higher in the evaluation than non-holiday weeks. This is all important to consider when deciding upon where to launch a new startup to compete with Walmart and Target.

The first step in any analysis is to create a detailed analytical plan. The goal of this plan is to determine the best 10 markets or locations to compete effectively against the established competitors. My action lists of steps includes: (1) open the datasets, (2) explore each dataset and explain all variables and features, (3) clean and preprocess the data for analysis, (4) merge datasets based on overlapping features, (5) break the merged datasets down by DMA (city), (6) export the data to create visualizations, (7) generate a list of questions to prioritize what to show through the analysis, (8) create visualizations to answer these questions and tell a data-driven story, (9) explain and interpret the findings from visualizations, (10) seek out possible other sources relevant to the launch plan, (11) overall analysis and explanation of what markets and locations to choose for the Startup, and (12) write this detailed launch recommendation including all information and visualizations.

Now onto the specifics of the datasets provided. In total, 9 datasets were in the google drive folder given for analysis. These include: *Stores with DMA* dataset (2 different sets), *Sales* dataset, *Features* dataset, *PI-18803 DMAs* dataset (dashboards), and *midwest*, *northeast*, *south*, and *west* datasets. These are comprised of mostly excel (.xlsx) files. All of these datasets are messy and unorganized in terms of *'NA'* values, titles, column names, and duplicate information. My next step was to rename all datasets for simplicity, make column names consistent and explanatory, and to

get rid of data that is not unique. I started with the *Stores with DMA* dataset, which includes 45 rows and 9 columns. The rows detail each store (45 in total), while the columns were readjusted and broken up. I transferred some information to a new file, labeled *Stores with City Ratios*, because these rows had information about the ratio of city size to the number of stores. The city with the most stores was Los Angeles (7 stores) followed by Atlanta, Cleveland, Dallas, and Houston (4 stores), Austin, Chicago, and Denver (3 stores), Charlotte, Kansas City, Philadelphia, Salt Lake City, San Diego, and Tampa (2 stores), and lastly Orlando (1 store).



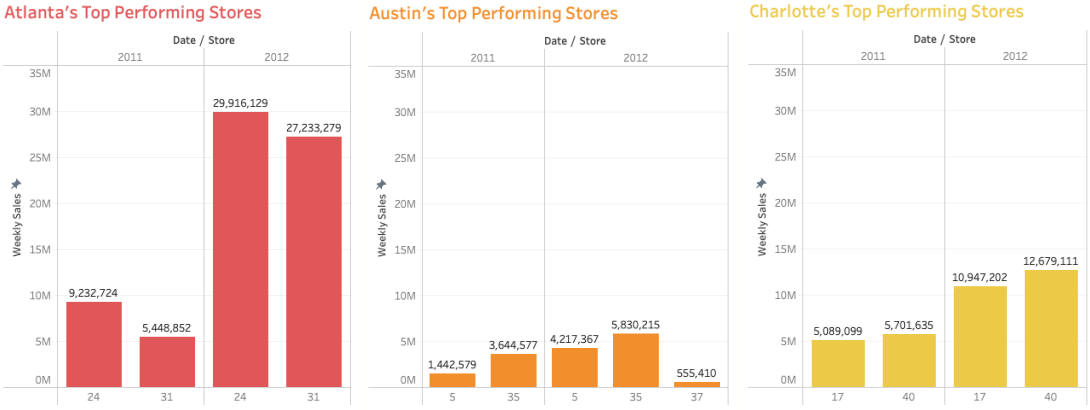


The graph above on the left shows a visual representation of the number of stores per city, while the graph above to the right shows the city size to number of stores ratio. The bigger the city's circle, the larger this ratio is, meaning that the size of the city is larger relative to the number of stores this city has.

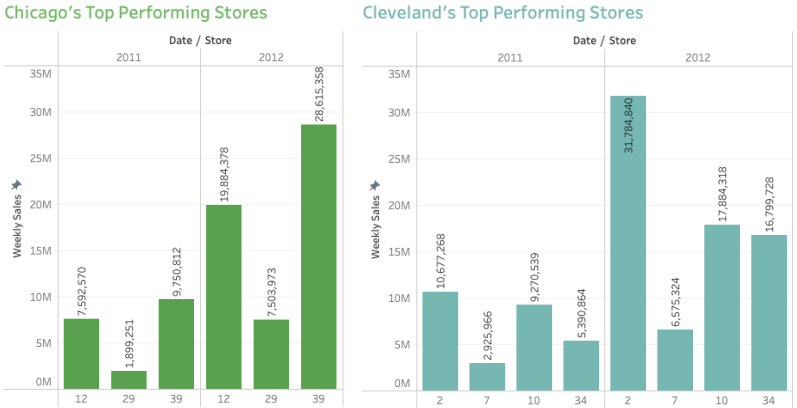
Next, I moved onto the *Sales* dataset, which includes 421,571 rows and 5 columns. I ran code on every dataset used to figure out the number of null values, and what columns the null values belonged to. The features dataset contains 8,191 rows and 12 columns. Compared to the sales dataset, there was overlap in three columns: *Store*, *Date*, and *IsHoliday*. This dataset also included the Markdown variables previously mentioned, but these columns had a large number of *'NA'* values. The *DMA Dashboards* dataset (originally named *PI-18803 DMAs*) includes three excel sheets, two of which display dashboards about vehicle ownership, overall household expenses, and transit expenses - taxi, public transportation, etc. It also includes information about race, age, and the population composition. The regional datasets detailed each region's (*Midwest*, *Northeast*, *South*, *West*) MSA, average annual expenditures, and characteristics from the Consumer Expenditure Survey from 2016-2017. It is important to note that all Stores/Sales data is dated from 2011-2012. After going through each dataset and tidying up the column names and taking account of missing values, I did not realize until later in my analysis that some values were "bad" - what i mean by this is that a substantial amount of rows had values that were not the correct type (ex: *Weekly_Sales* included some values such as *"2011-12-20 2345 00:00:00"* instead of just *"2345."* I again went through each dataset, checked the column type and made sure these columns were overwritten as numeric types (*float64*).

Launch Plan Recommendation

Merging the datasets for ease of analysis was the next step, based on the overlapping features. I did this in two parts, first merging the *Sales* and *Features* datasets based on the variables *"Store," "Date,"* and *"IsHoliday."* After this, I merged the *Sales* and *Features* merged dataset with the *Stores with DMA* dataset, to then use the geographical locations for a further breakdown. After each merge I made sure to recount the number of *'NA'* values, and drop them accordingly. Now to break this merged dataset down by DMA, I grouped the values of each city (15 in total). The following graphs show which stores for each DMA were the highest performing. On the Y-axis is the sum of weekly sales specific for the city, and the X-axis shows the store number and the year the weekly sales were reported. The weekly sales is set at a max of $35,000,000, to show the relative size of the bars compared between each city.

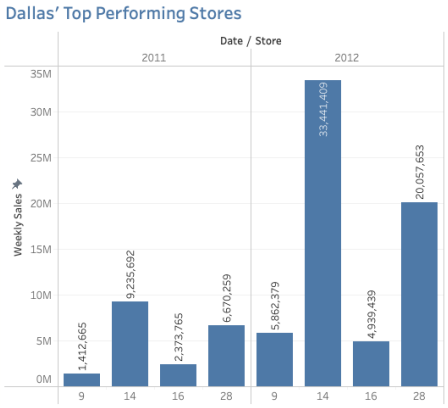


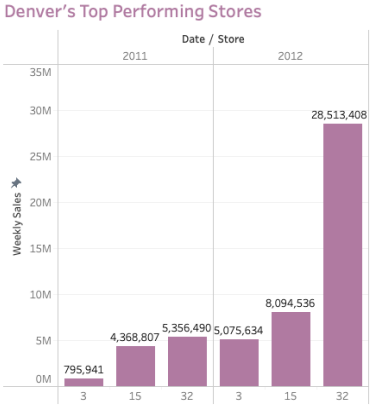
Atlanta's two stores (24, 31) vastly improved in monthly sales from 2011 to 2012. Store 24 had a sum of $9.2 million in weekly sales in 2011, and this increased to $29.9 million in 2012. Store 31 had a sum of $5.4 million in weekly sales in 2011, and this again increased to $27.2 million in 2012. This is a huge improvement compared to Austin, where Store 5 had an increase in weekly sales by roughly $3 million from 2011 to 2012, and Store 35 which increased by roughly $2 million. Additionally, Store 37 looks like it just opened in 2012, and started out very low in terms of weekly sales. Charlotte follows a similar increase in weekly sales with Store 17 and Store 40, but still fell far short of Atlanta.



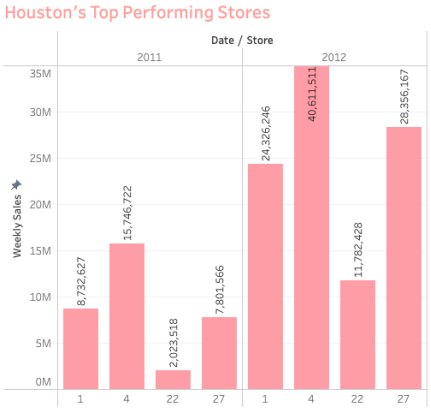
Chicago's weekly sales for each store vastly increased again from 2011 to 2012. Store 12 grew weekly sales from $7.5 million to $19.8 million, and Store 39 grew from $9.7 million to $28.6

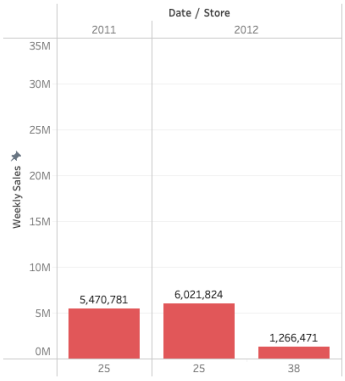
million to be Chicago's top performing stores. Store 29 looks to be much lower in weekly sales compare to the previous. Cleveland, above on the right, has four stores in total. Of these four stores, I would consider only Store 2 to have a great increase in sales, from $10.6 million in 2011 to $31.7 million in 2012.





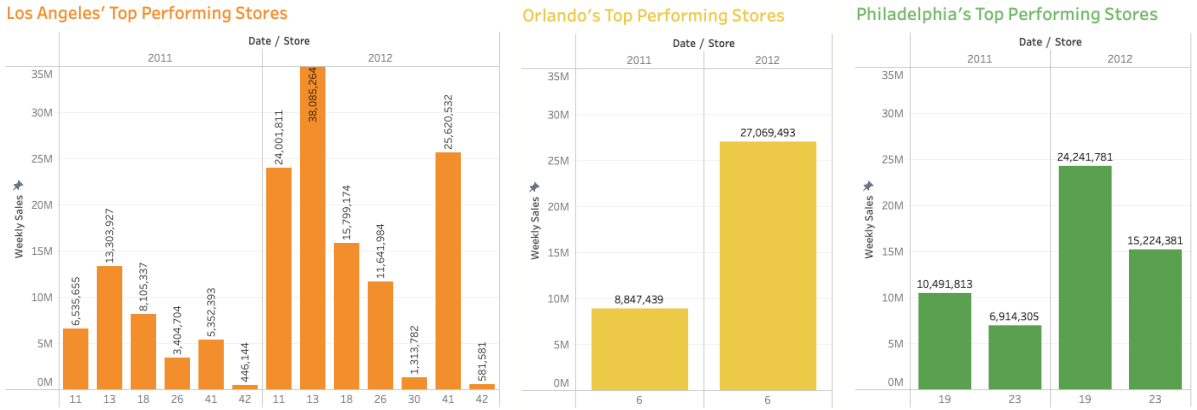
Dallas was able to more than triple it's weekly sales from 2011 to 2012, particularly with Store 14. This store reported sales of $9.2 million in the first year and $33.4 million in the second year. Denver also had takeoff in only one store. Store 32 reported weekly sales of $5.3 million in 2011, and the next year $28.5 million.



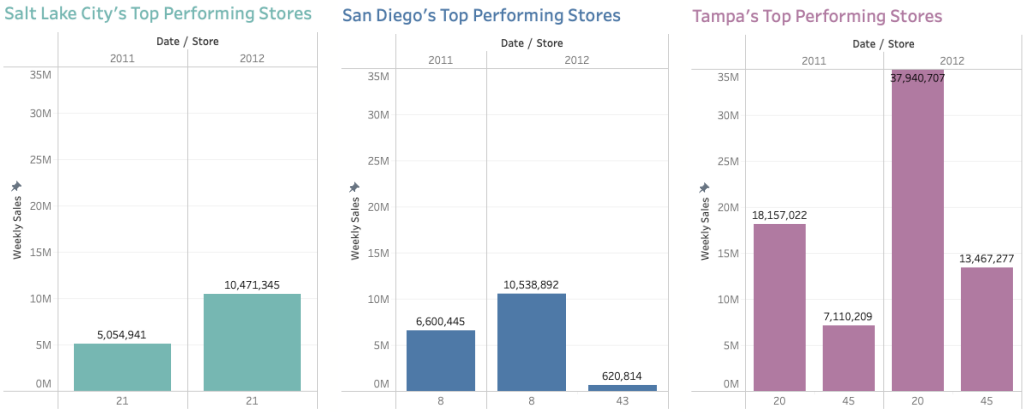


Houston's stores are another top competitor, and all stores at least doubled, if not tripled weekly sales from 2011 to 2012. Store 1 nearly tripled in weekly sales from $8.7 million to $24.3 million, Store 15 more than doubled in sales from $15.7 million to $40.6 million, Store 22 (likely a newer store) grew substantially from $2.0 million to $11.7 million, and lastly, Store 27 nearly quadrupled from $7.8 million to $28.3 million. Kansas City overall was the worst performing city in the data. With one store in 2011, and two stores in 2012, Store 25 and 38 fell short of $6.0 million in weekly sales.

Launch Plan Recommendation



Here we see Los Angeles, with the most amount of stores: seven. Almost all stores, including Store 11, Store 13, Store 18, Store 26, and Store 41, grew weekly sales in the high millions with the exception of Store 42. The specific stores to take note of are Store 13, with an increase from $13.3 million to $38.1 million in sales, and Store 41, with an increase from $5.3 million in sales to $25.6 million. It is likely that store 42 is much newer than the rest of the stores in Los Angeles, and this shows in the sum of weekly sales (under $582,000). Orlando, the only city with one store, Store 6, reported an increase in weekly sales of more than $18 million from 2011 to 2012. Philadelphia, graphed above to the right, had two stores which both doubled in the sum of weekly sales from 2011 to 2012. Although Store 19 and Store 23 vastly grew over this time period, it does not come close to Los Angeles, or even Orlando.

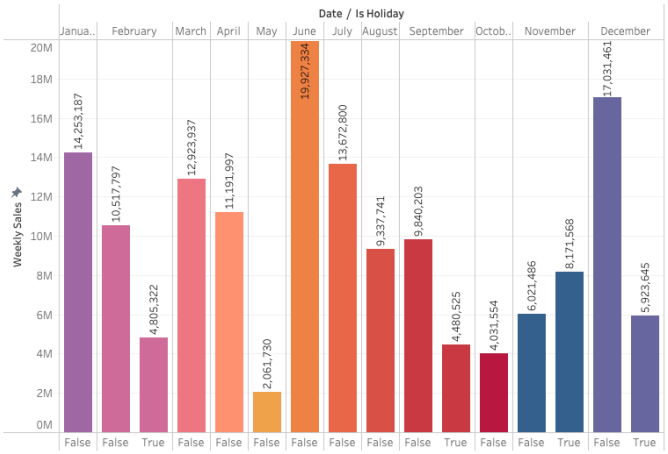


Salt Lake City, with one store, lacked the growth needed to be a substantial city choice for the Startup. In 2011, Store 21 reported $5.0 million in sales, while in 2012, this doubled to $10.4 million. San Diego had one store in 2011 (Store 8), and opened a new store in 2012 (Store 43). It would be interesting to take this analysis further and see when Store 43 opened, because it only reported $620,814 in sales. Store 6, on the other hand, grew from $6.6 million in sales to $10.5 over the yearly period. The last city to take considerable notice of is Tampa. In 2011, Store 20 grew in weekly sales from $18.1 million compared to $37.9 million in 2012. Store 45 lacked in this growth slightly, with an increase from $7.1 million to $13.4 million.

Launch Plan Recommendation

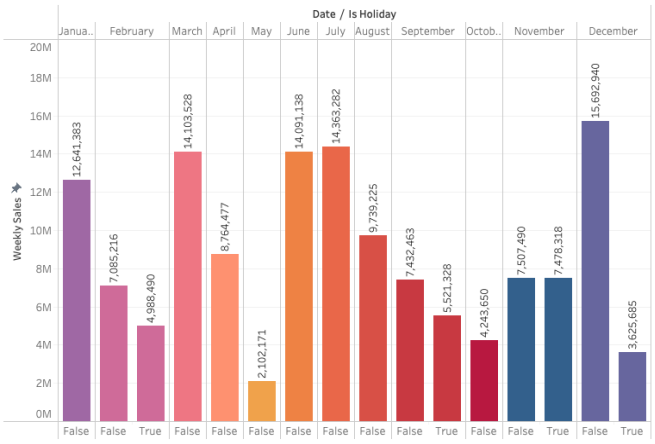
There are many takeaways from looking at each city's weekly sales graphs above. It is assumed that stores with extremely low values are newer, while the time period ranging from 2011 to 2012 can give further insights of any new stores in a city. After analyzing these graphs and looking at the changes in weekly sales from 2011 to 2012 for each store, the top three performing stores based on sales are in the cities of Los Angeles, Houston, and Tampa. The next seven top performing markets are Atlanta, Chicago, Cleveland, Dallas, Denver, Orlando, and Philadelphia. These are the market contenders for a startup to be located that will compete with Walmart and Target, but for the rest of the analysis I will mainly focus on the weekly sales growth in Los Angeles, Houston, and Tampa.

Los Angeles' Sales per Month/Holiday



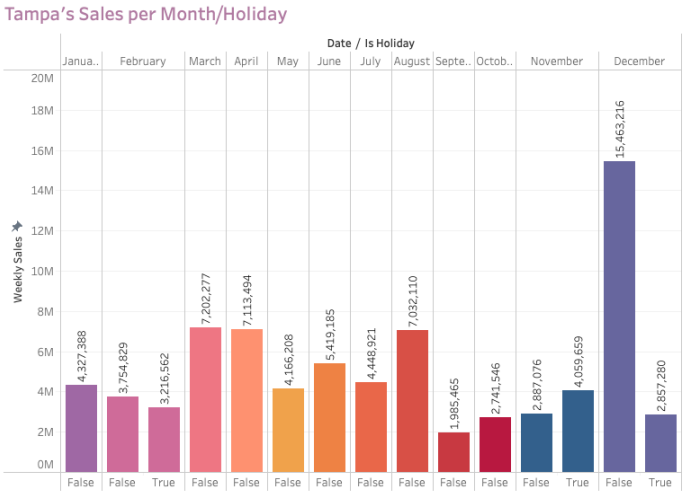
In the graph above, we can see the trends as well as whether the week was marked as a holiday over the course of a year period. The top months in terms of sales were June and December. In June, the Los Angeles stores reported a sum of $19.9 million in sales, with no holidays in this period. On the other hand, the other top month was December, right before Christmas, with a high $17.0 million in sales. At the end of December this dropped off. It is likely that from this graph, the increase in sales in June could be due to tourism, while in December, many promotional markdowns increased Christmas shopping among residents. Overall, we can see a consistent dropoff after the holidays in February (the Superbowl), September (Labor Day), and December (Christmas).

Houston's Sales per Month/Holiday



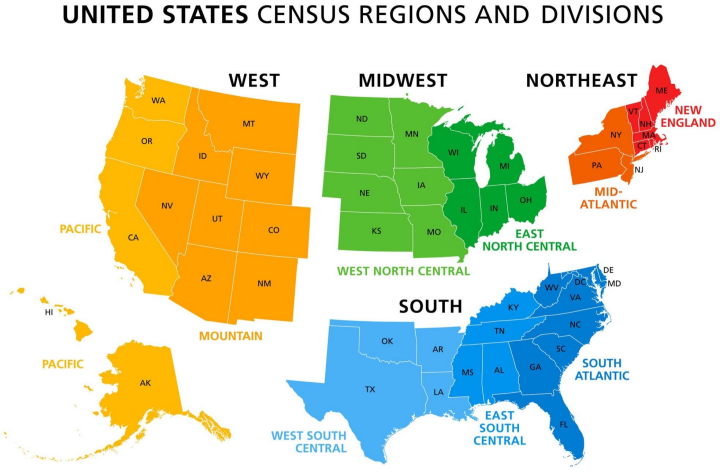
Launch Plan Recommendation

Los Angeles seems to be unique in the jump in sales in the summer months, because Houston follows only the Christmas season sales trends. We see again the dropoff post holidays in February, September, and December in Houston. The primary holiday to take in consideration for promotional materials and shopping at Walmart and Target is Christmas, and most markdown prices should be around this time of year.



Tampa looks to fall short in the monthly sales trends compared to Los Angeles and Houston. The weekly sales for Tampa over the year are stagnant at under $7.2 million, and then this nearly doubles in the Christmas season to $15.4 million in weekly sales. This is the greatest increase in a holiday season seen thus far. It is important to keep notice of the fact that Tampa only has two stores, while Houston has four stores, and Los Angeles has seven stores.

With also factoring in the regional data from the Consumer Expenditure Survey, consumer spending patterns in Los Angeles, Houston, and Tampa are revealed. Below is a breakdown of the regions divided for the US Census, which include West, Midwest, Northeast, and South.



Los Angeles is apart of the West region and both Houston and Tampa are apart of the South.

Average Annual Expenditures and Characteristics (Consumer Expenditure Survey, 2016-2017)

	Los Angeles	Houston	Tampa
Income before taxes	$76,471	$80,250	$60,340
Avg annual expenditures	$66,971	$67,304	$48,654
Income / Expenditures	1.14185	1.19235	1.24019

Starting with the West regional data, in Los Angeles the consumer unit characteristics from 2016-2017 give income before taxes as $76,471, and average annual expenditure as $66,971. In the South regional data provided, the consumer unit characteristics in Houston give income before taxes as slightly higher at $80,250, and average annual expenditure as $67,304. While in Tampa, the reported income before taxes is a much lower $60,340 and average annual expenditure is $48,654. This tells us a few key things: first, that although residents in Los Angeles on average have a lower income before taxes then residents in Houston, they spend a similar amount in annual expenditure. Also, the income before taxes is much lower in Tampa, and this trend follows in annual expenditures being lower subsequently. Additionally, looking at the ratio of income to expenditures, Los Angeles has the lowest at 1.14, followed by Houston at 1.19, and Tampa at 1.24. Los Angeles residents spend the majority of their income, compared to Houston and Tampa.

Concluding this recommendation plan, the launch of a Startup to compete against the competitors Walmart and Target should be primarly located in Los Angeles, Houston, or Tampa, with additional locations to consider including Atlanta, Chicago, Cleveland, Dallas, Denver, Orlando, or Philadelphia, in no particular order. Los Angeles has the most amount of stores and residents spend a large portion of their income. Taking into account the fact that this was the only city where weekly sales peaked in the summer months as well as the Christmas season, this is an ideal location. However, the number of stores already (seven) in this area could be a downside. Houston slightly combats this issue, because there are only four top performing stores, and Tampa gives way for another top market choice with only two top performing stores. Ideally, the retail Startup should launch at the beginning of the year, to give time for marketing and resident realization well advance to the Christmas season. Additionally, promotional efforts and markdowns should be a top priority in the month of December. Taking into consideration the *DMA dashboards* that detail consumer vehicle ownership, consumption habits, population composition, and expenditures on public transportation, the store should be in an area that is easily accessible, with public transit (busses, taxis, etc.) close by. In whole, the West and South regions of the United States are the most profitable for a retail store. Any new store will be slow in weekly sales growth at first, but with mass marketing and consumer outreach, this will hopefully increase sales at least for the following year. With more time, I would have liked to find data relating to how long the competitive stores in the datasets provided were open for, and figure out the best month in the year to have a grand opening for a retail startup store.