Nicholas Gingrich
Eren Budur
Daniel Tse
Tiffany Chung

Team 4: City Council - Small Businesses Final Report

Introduction:

The overarching purpose of this project is to analyze various types of data pertaining to the small businesses that exist within District 4 in order to get a better understanding of what the current small business landscape looks like. Especially over the last couple of years with covid-19, the state of small businesses across the globe has changed dramatically and the City Council of District 4 is eager to understand how they can better support the small business community. As such, it is our goal to use the insights from our data analysis to highlight potential areas of reform or potential obstacles that small businesses currently face. Specifically, within our analysis, our project is split up into two major portions, a base project and an extension project. The base project was designed to allow us to perform some preliminary data analysis in order to get a feel for the small business environment that currently exists. Once this was completed, we used our insights from the base project to narrow the scope of our analysis to a particular area within the realm of business that we would like to explore further. In this case we decided to look into the state of various types of business licenses and how those quantities compare to other districts in the Boston area. This project report details our findings from both the base project as well as our extension project, in addition to information on what datasets we used and some of the limitations that we faced along the way.

Datasets Used:

BDPA List of Main Street Businesses: A list of businesses that exist on the “main streets” of the Boston area along with information such as category and location coordinates.
https://bpda.app.box.com/s/683z3xkojdpxrb2subs0ial76d1ye4sp/folder/167044664376

Section 12 Liquor Licenses: Contains a list of businesses that have active registered liquor licenses in the Boston area.
https://data.boston.gov/dataset/liquor-licenses

Cannabis Licenses: Contains a list of businesses that have active registered cannabis licenses in the Boston area.
https://data.boston.gov/dataset/cannabis-registry

Annual Entertainment Licenses: Contains a list of businesses that have active registered annual entertainment licenses in the Boston area.
https://data.boston.gov/dataset/entertainment-licenses-legacy

Food Establishment License: Contains a list of businesses that have active food establishment licenses in the Boston area.
https://data.boston.gov/dataset/active-food-establishment-licenses

Licensing Board Licenses: Contains a list of businesses that have active registered miscellaneous license types (See miscellaneous section for more information on the included types)
https://data.boston.gov/dataset/licensing-board-licenses

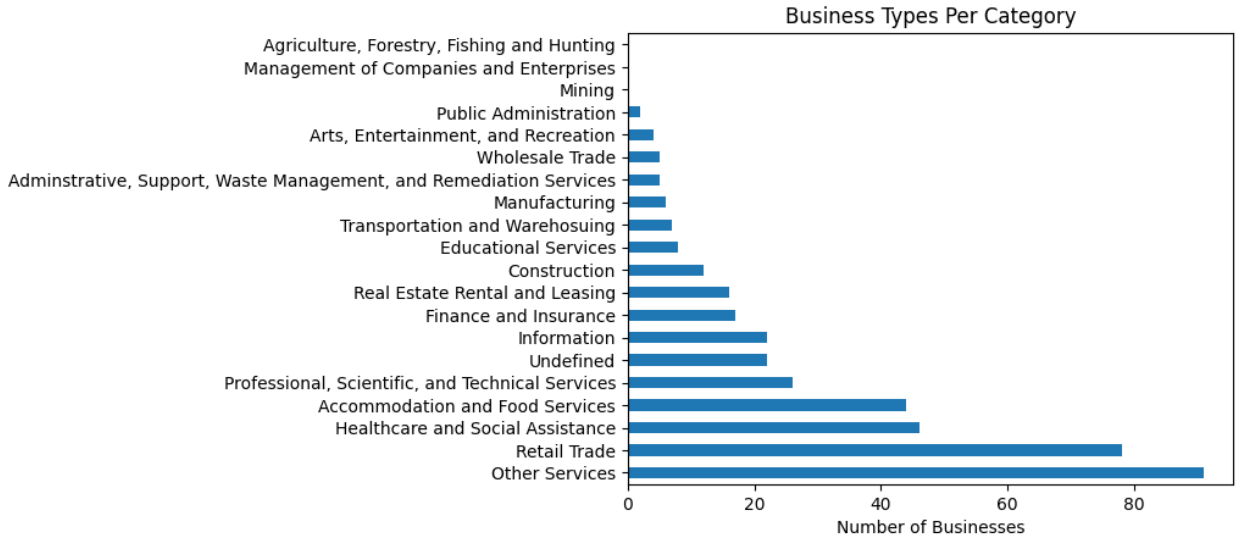
Certified MBE Dataset: A list of businesses registered as one of the following enterprises; women owned, minority owned, small business, small local business, or veteran owned.
https://data.boston.gov/dataset/certified-business-directory

Voting Minutes Documents: A series of PDF’s that contain a record of the city council licensing board hearings. The content of the documents contains information on license violations as well as license applications.
https://www.boston.gov/departments/licensing-board/licensing-board-information-and-members#voting-agendas

Base Project:

What types and how many businesses exist in District 4?

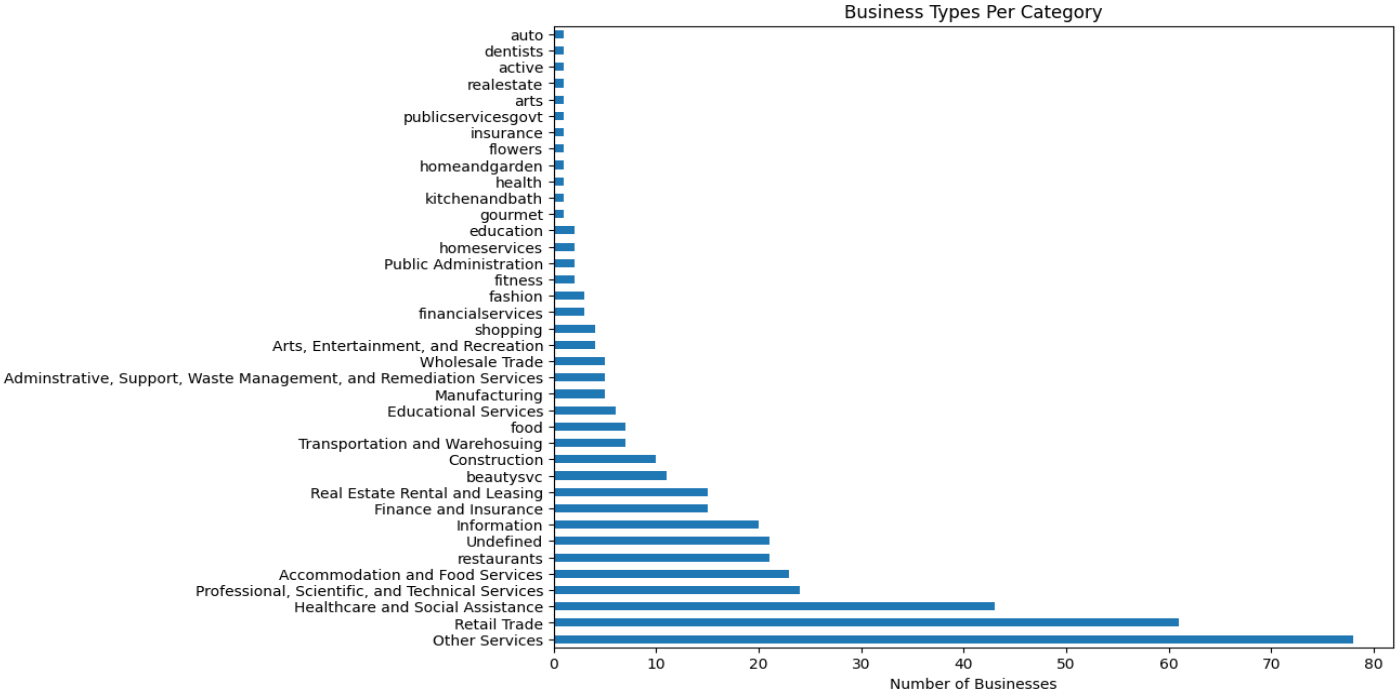
From the preliminary data analysis on the types of businesses that exist, we can see that District 4 contains a variety of different small business types. The bar graph visual below shows the distribution of the quantity of each business type as defined by the filtered BPDA data set. If we move past the “Other services" category and look at “Retail Trade”, we observed that this category is by far the most dominant business type within District 4. Furthermore, while not as dominant, we also observed the “Health Care and Social Assistance " and the “Accommodation and Food Services” categories represent a proportionally large share of the total number of small businesses. The remaining businesses consist of varying categories that share a proportionally small, yet somewhat equal share, which suggests that District 4 has a diverse range of more niche small businesses in addition to the dominant business types mentioned above.



While we initially used the categories provided by BPDA dataset, we found a couple of issues with these provided labels. First, we felt as though the labels themselves were a bit vague and in some cases arbitrary, and secondly, there was a large proportion of businesses that are labeled as “other services” or “undefined”. For these reasons we decided to try to improve the business type labels by utilizing the Yelp API in order to import new labels for all of the BPDA businesses that were listed on yelp. It certainly was not the case that all of the small businesses were in the Yelp database but we were able to cut down on the number of “other services” and “undefined” businesses as well as provide some more descriptive labels for the District 4 BPDA businesses. See the

chart on the next page for an updated list of categories along with a map with all of the BDPA businesses plotted.

Chart of Businesses by Yelp Category in District 4



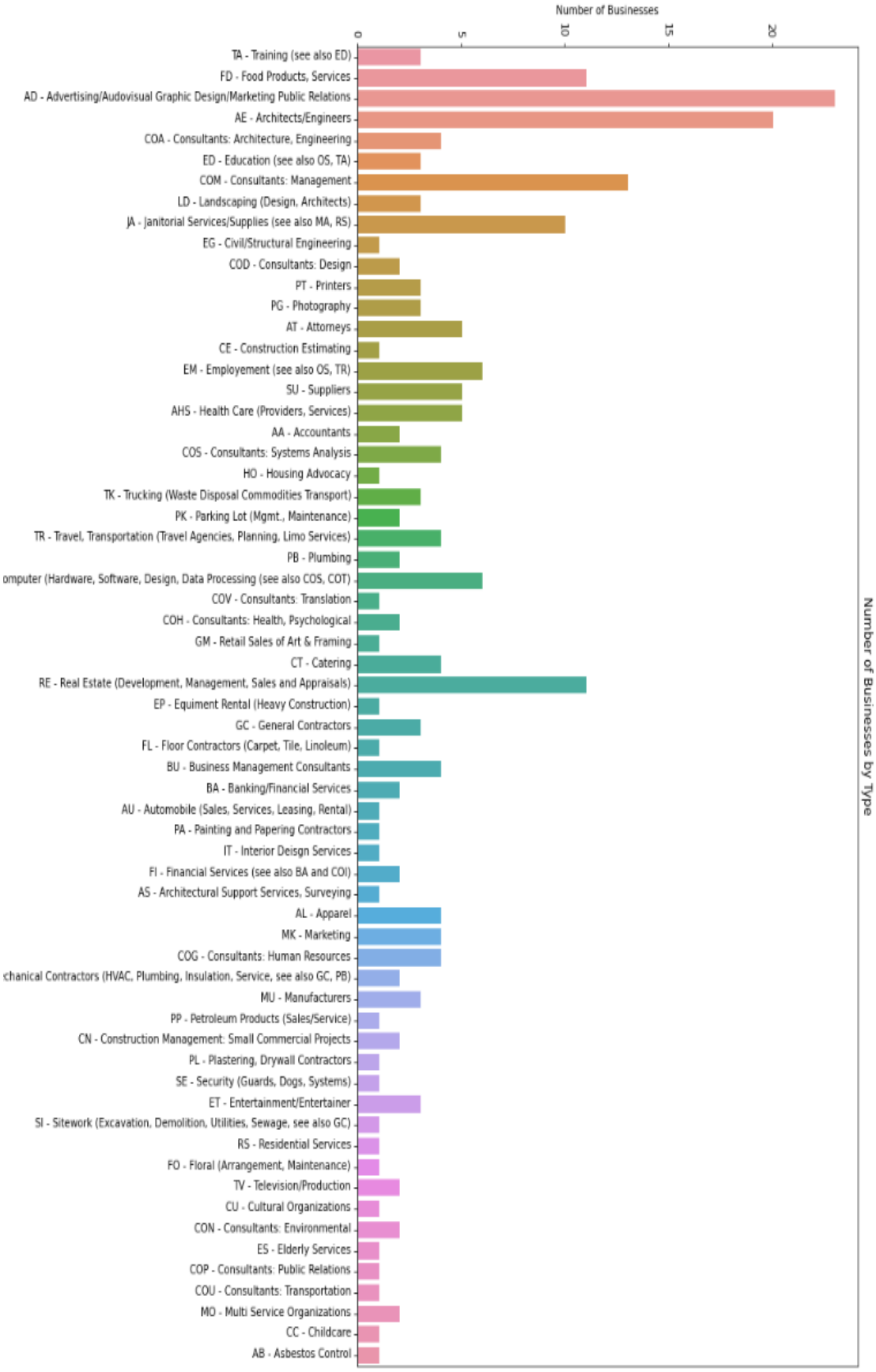
Map of BDPA Main Street Businesses

How many woman owned or MBE businesses exist within District 4?

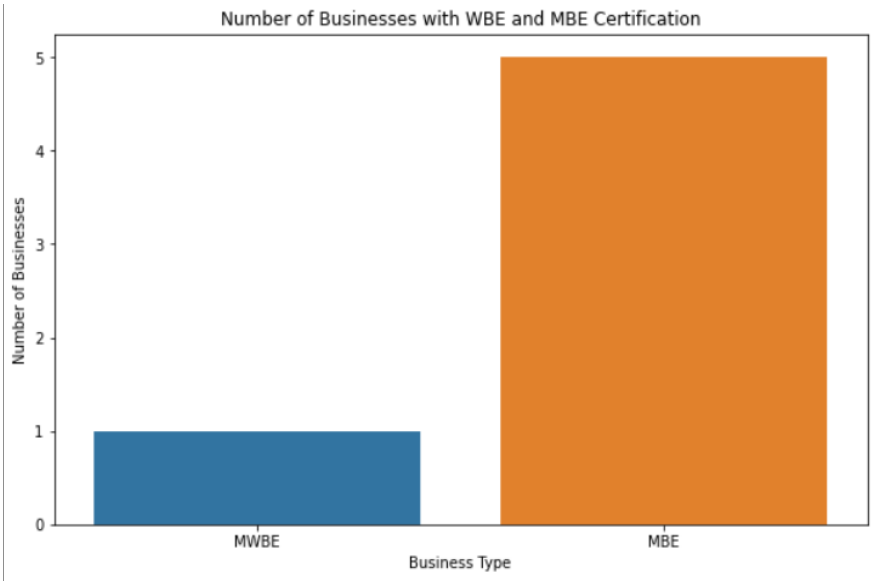
When it came to analyzing the women and minority owned businesses within District 4, our first step was to look through all businesses in Boston that fit these criteria, which is what we see in the first two graphs above. As shown in the first graph, there is an abundance of advertising/marketing businesses, closely followed by architectural/engineering firms. Then, when we looked at the number of businesses by certification type, we found that 80 were minority owned businesses, almost 60 were women owned businesses, and their intersection of minority women owned businesses was about 75.

Our dataset for women and minority owned businesses was limited by only having an address present to determine whether the business was in District 4. We tried different APIs to extract the location data but none of them were accurate enough. So, we decided to merge the dataset with the main road businesses dataset to find any overlapping addresses. This was not the most accurate approach as we were only limited to businesses located on the main roads but it did give us a small yet informative set of businesses to look into. The few businesses that did pass the cross-reference were a range of different types, from auto shops to restaurants. The data that we did pull from this method is too limited to make any solid, overarching conclusions but we can see from the last two graphs that the number of women owned or MBE businesses within District 4 is very low.

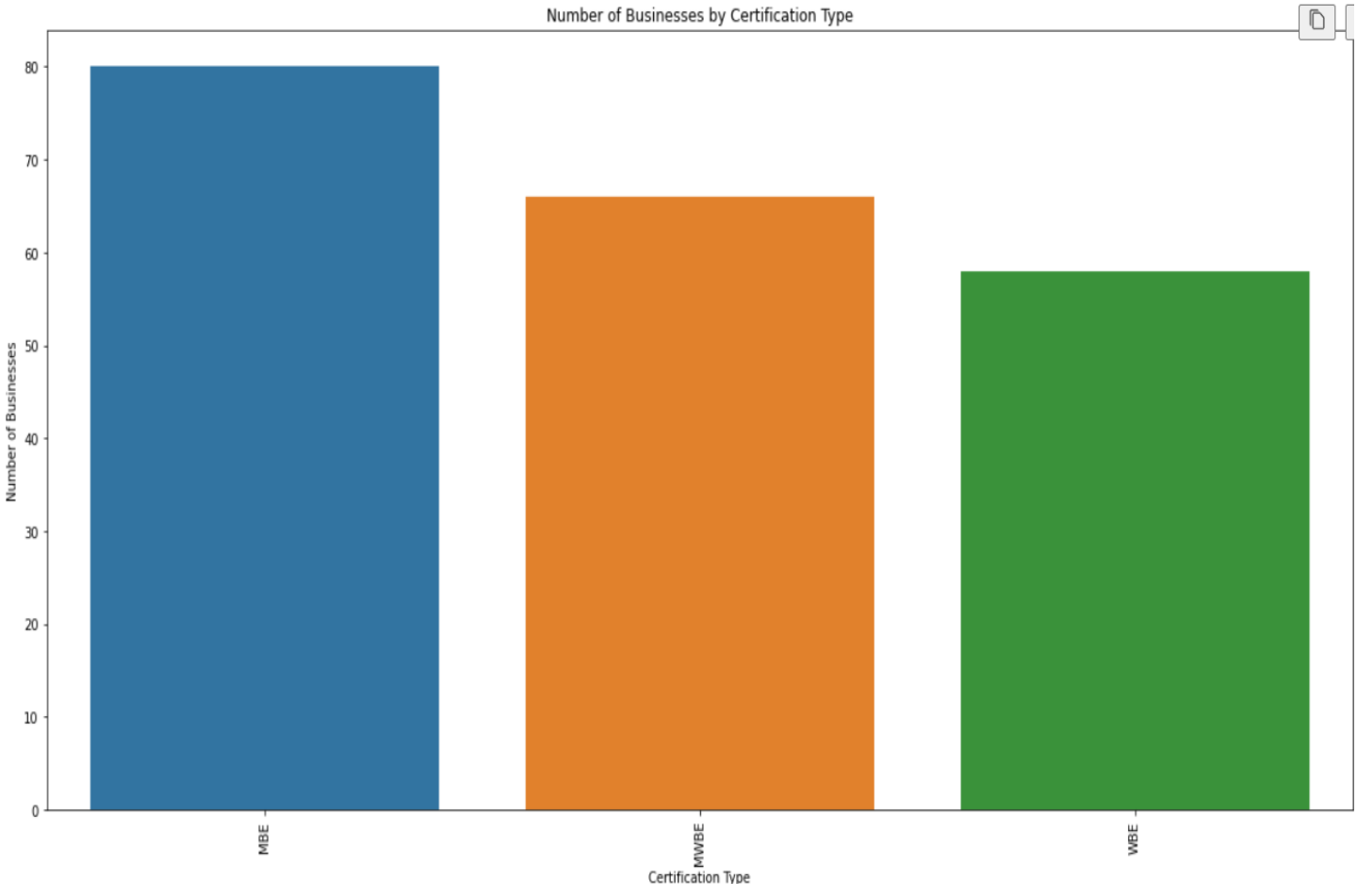
Number of Certified Businesses by Business Type in Boston Area



Number of Businesses with MBE and WMBE Certification in District 4



Number of Businesses by Certification Type



Extension Project:

Background:

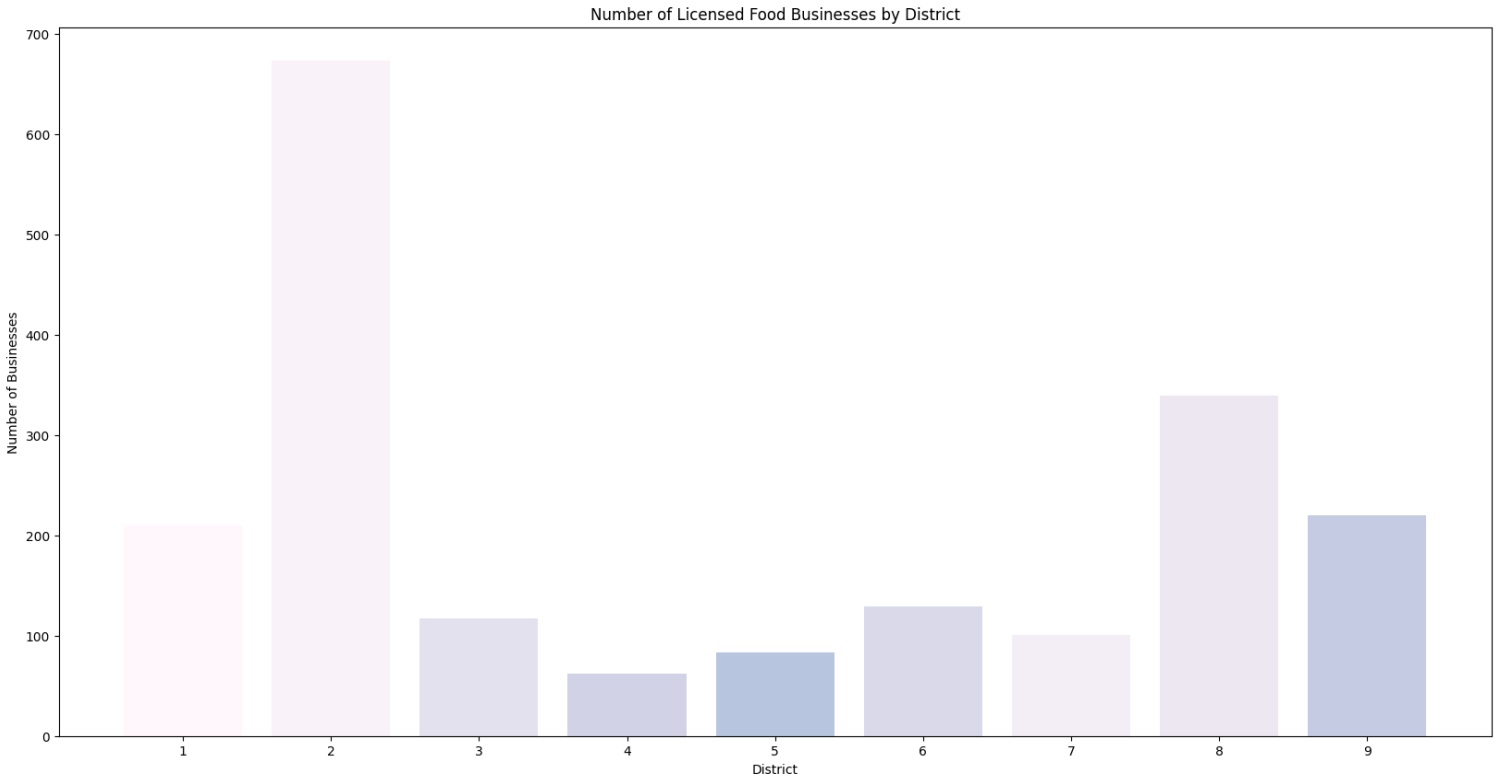
For our extension project we decided to take a comprehensive look at the state of business licenses in District 4 as it compares to other districts. Specifically, within the realm of licenses we decided to look at liquor licenses, food establishment licenses, annual entertainment licenses, cannabis licenses, and miscellaneous business licenses. We feel as though this extension project is important as it ties back into the original goal of the project, being to get an idea of what the current business landscape looks like and to identify potential obstacles that small businesses currently face. By analyzing the business license landscape we aim to be able to identify shortcomings in the quantity of certain license types as it compares to other districts and to further investigate as to what the reasoning is behind the lack of licenses that we observed in our base project analysis. We are hoping that by highlighting this information for the Boston City Council, they will be able to use this information to better understand how they can assist the small business community.

Food Establishment License Insights:

Our initial data analysis on the number of businesses in District 4 from our base project yielded a number that was much lower than we expected, and after seeing the results, we decided to see if a more popular license — like food — would show a more even distribution between all the districts. For this section of the extension project, we analyzed the data of currently active food establishments that carried food licenses (and later used this data to compare with other analyses).

The data analysis for this section was fairly straightforward and is shown below. As seen in the bar graph, District 4 still stands as the district with the lowest number of food licensed businesses (roughly 55) which is in line with what we originally found with the low quantity of business in distinct 4. Districts 5 and 7 are not far ahead of District 4, standing at around 80 and 95 businesses respectively, but District 4 has proportionally less licensed food establishments relative to any other district in Boston.

Number of Food Licenses Per District

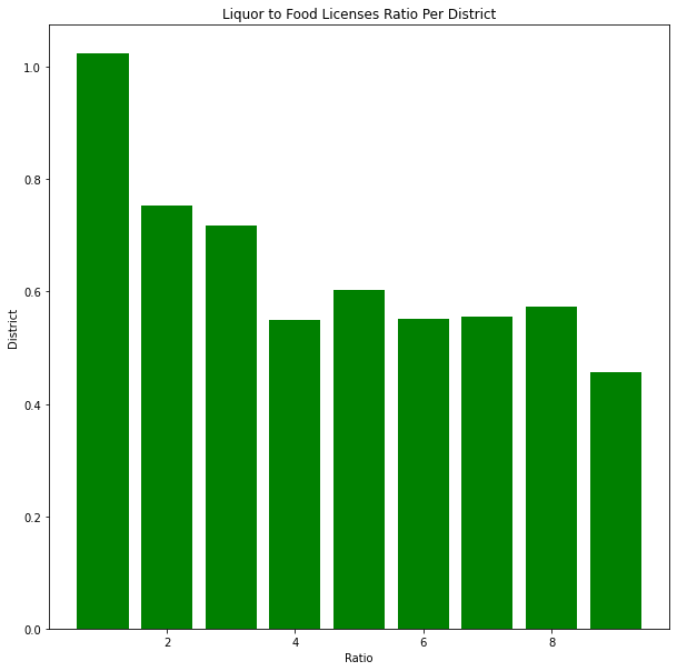


Liquor Licenses vs Food Licenses Insights:

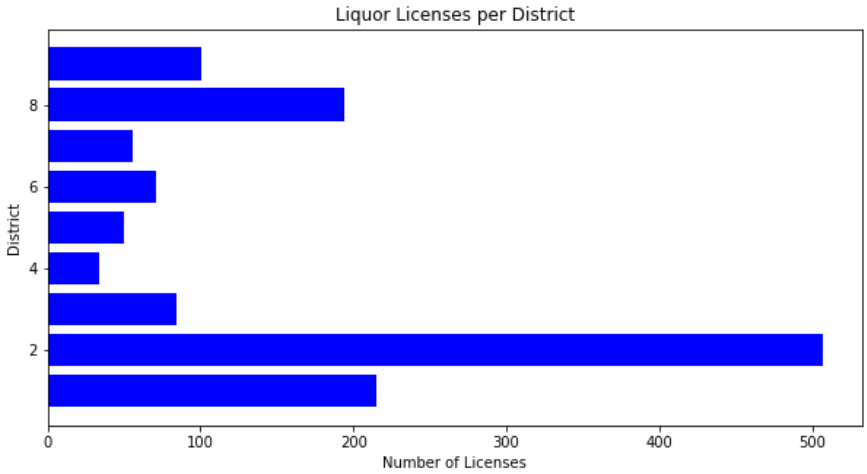
As we realized that the number of businesses was much lower in District 4, we decided to explore the relations of the different licenses. One of the license types that we analyzed was liquor licenses. When we take a look at the liquor licenses we can see that District 4 has the least amount of licenses as compared to the other districts. We could argue from this that District 4 has room to build more businesses in this domain however further analysis is needed. Further we could also observe that there must be some underlying reasons that could explain the low license count. However, when the data is presented side by side with the number of food licenses, we see that District 4 is not lagging behind but rather the district has a low number of food establishments as well.

As we can see in the ratio graph below, most of the districts have similar proportions. When we compare the food and liquor licenses in districts 4-9, they only vary slightly by a factor of .1 or so and then it can be observed that districts 1-3 trend higher. This type of ratio analysis is helpful for understanding the context of two connected domains; food and alcohol, and how they are related. In addition to this sort of analysis, it would be beneficial to explore why there are so few licenses or businesses in general. From just this data analysis alone, we might speculate that it's possible that the district is populated with a high level of residential buildings or perhaps there are a high number of commercial vacancies. While we were not able to investigate the reasoning behind the low number of businesses, we did decide to pursue an explanation as to why district 4 has a lower number of liquor licenses, the results of which can be found at the end of the report.

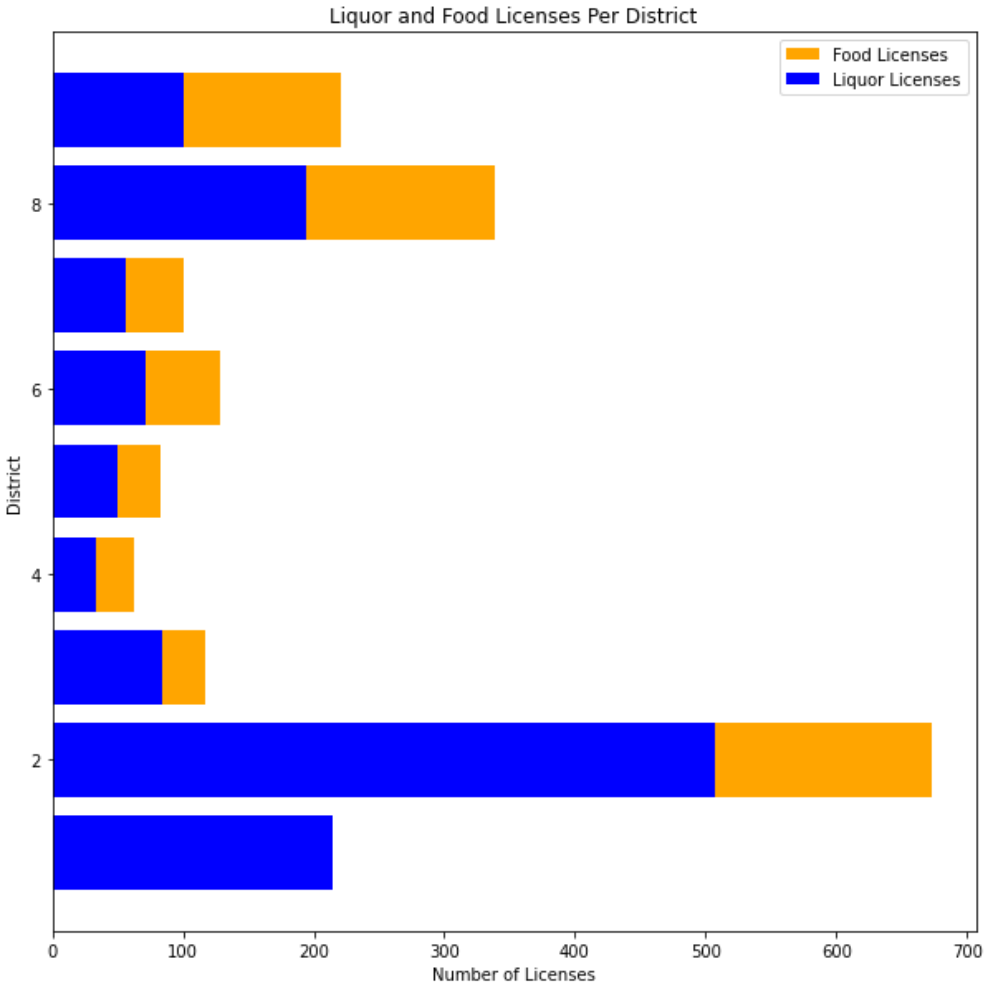
Liquor to Food License Ratio Per District



Quantity of Liquor Licenses Per District



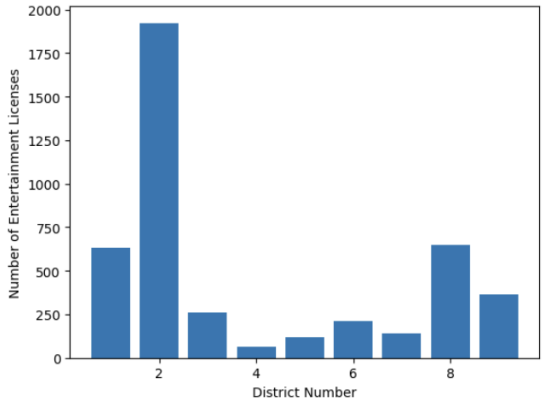
Aggregate Liquor and Food License Quantities Per District



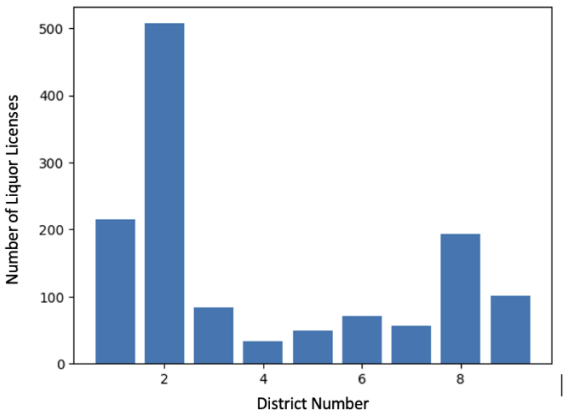
Annual Entertainment License Insights:

One of the other areas that we focused on within the realm of licenses was annual entertainment licenses. This is defined by the Boston City Council to include "concert, dance exhibition, automatic amusement devices (pin ball or video games), cabaret, and public shows consisting of dancing by patrons, performers, recorded or live music, use of amplification, theatrical play, exhibition, film, floor show, or light show, and any kind of dynamic audio or visual show." From our analysis, there are a couple key insights that we thought were particularly interesting. Firstly, we found that the number of annual entertainment licenses was proportionally lower than other Boston districts. If we look at the chart below on the right, we can see that District 4 has the fewest number of annual entertainment licenses out of all of the districts with a total of 65 currently active licenses. However this alone is not necessarily surprising considering that our BPDA analysis indicates that "Arts, Entertainment, and Recreation" business type is one of the most infrequent businesses types that exist within District 4. That being said, it should follow that they have fewer annual entertainment licenses than other districts which is in fact the case. Additionally, we decomposed the 65 annual entertainment licenses figure into the number of different businesses that have an active license and we reported that out of the 65 licenses, they are split up amongst only 14 different businesses. This indicates that to begin with, besides the fact that the majority of businesses in District 4 do not have entertainment licenses, the businesses that do have annual entertainment licenses tend to have on average 4.6 different licenses. Lastly, we noticed an interesting correlation between the number of annual entertainment licenses and the number of liquor licenses per district. If we look at the chart below to right, which plots the number of liquor licenses per district, and compare the trends to the chart on the left which consists of annual entertainment licenses, there exists a very similar trend across all of the districts which suggests that there is a larger scale factor that could affecting the number of licenses per district such as the total number of businesses within each district.

<u>Entertainment and Liquor Licenses Per District</u>



Entertainment Licenses Per District



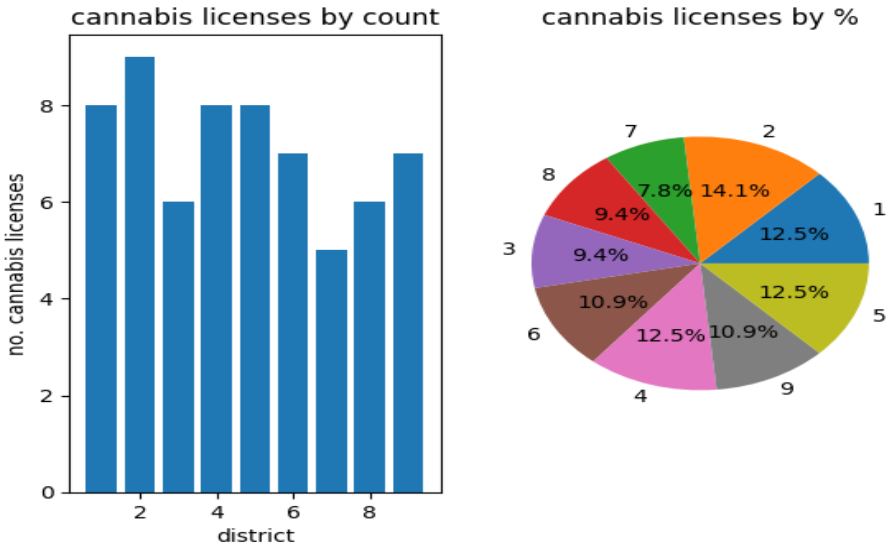
Liquor Licenses Per District

Cannabis License Insights:

Another type of license we analyzed was marijuana licenses. We looked at data for all establishments in Boston with cannabis-licensed applicants, specifically looking into if, like food and liquor, District 4 was relatively lower in marijuana licenses as well. Based on some simple analysis, we found that District 4 is not particularly low in cannabis licenses when compared to all other districts:

As shown in the pie chart, District 4 holds 12.5% of the cannabis licenses in Boston, which is tied for the second highest percentage among all districts. While there are a total of about 1,450 liquor licenses in all of Boston, there are only 64 establishments in Boston with approved cannabis licenses, which may account for the relatively even distribution of cannabis licenses across districts. If there were more total cannabis licenses, perhaps we would see a similar lack of licenses in district 4 in comparison to the others.

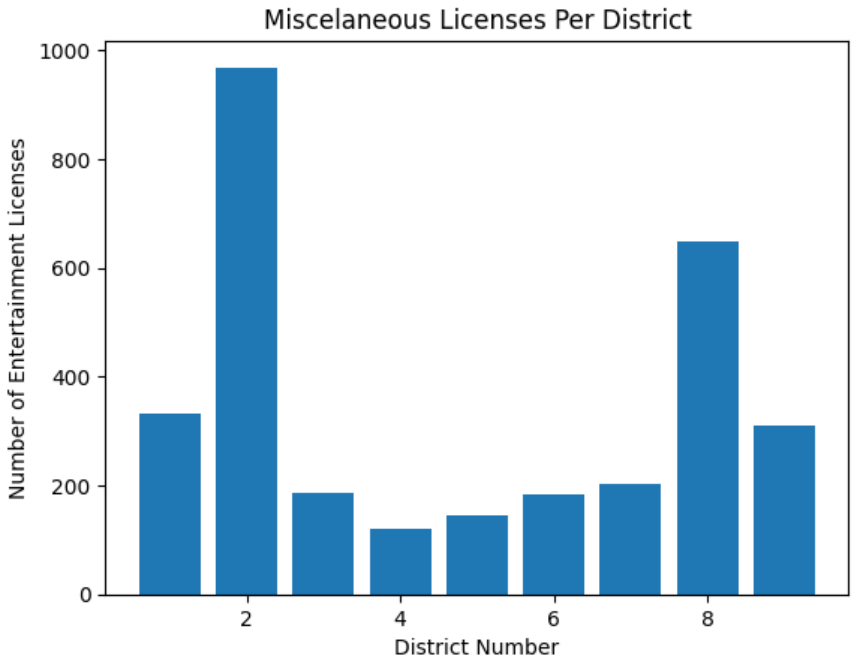
<u>Cannabis Licenses by Count and Percentage of Total Cannabis License by District</u>



Miscellaneous License Insights:

In addition to various other license data discussed so far, we wanted to perform a truly comprehensive analysis on the state of licenses in District 4, so we decided to include in our analysis the miscellaneous licenses dataset in addition to the definite license categories that we have already discussed. Even though this category by definition may not provide a lot of insight on any one specific license group as a whole, we still felt that it was important to consider as the individual types of licenses included in this miscellaneous category may provide useful insight. The Boston City Council defines the miscellaneous license group to include “Billiards, Bowling, Clubs/Veterans' Groups, Common Victualler (Food Service), Alcohol Beverage, Innholder, Dormitories/Lodging houses, and Retail Package Stores”. From our analysis we have determined that the distribution of miscellaneous licenses as displayed on the chart below is very similar to what we have seen for other license data sets, where District 4 has the fewest number of miscellaneous licenses. However in this case, the difference between the District 4 levels and the levels present in other districts is not as drastic as it is for other types of license data such as the annual entertainment data. In this situation, it seems as though because the miscellaneous category covers a variety of different license types, the fact that District 4 has the fewest number of licenses could suggest that District 4 as a whole contains fewer businesses in total then other districts.

Miscellaneous License Quantities Per District



Voting Minutes Analysis:

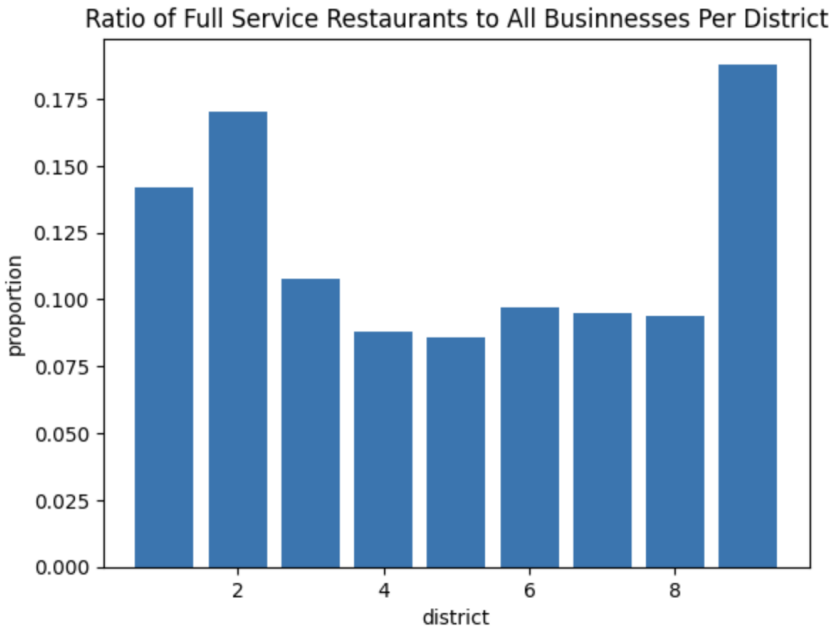
In an effort to understand why license counts were proportionally lower than other districts, we were interested to see if a potential cause of this was a high rejection rate of license applications in district 4. So, we decided to parse the voting minutes documents for the city council licensing department, which outline, among other things, the status of whether or not certain license applications are granted or not. After parsing all of the 2022 and 2021 voting minute files, we discovered that for the most part it was somewhat abnormal for a license application to be rejected by the Licensing Board, at least at this step in the process. It is possible that there are other obstacles that applicants face prior to this point in the process which may manifest itself in some form of rejection, however with the available data we are unable to draw a definite answer on that end. Regardless, from this portion of analysis we reasoned that the low license counts were not the product of a high number of application rejections. This conclusion indicates that the low counts are a product of one of two scenarios: the first option is that there are a number of businesses such as restaurants or entertainment facilities that could benefit from such licenses but for some reason they are not applying; and the second option is that there are simply less of these types of businesses in district 4 and therefore by default there are less licenses. In the next section we attempt to answer this question.

Full Service Restaurants with Liquor Licenses:

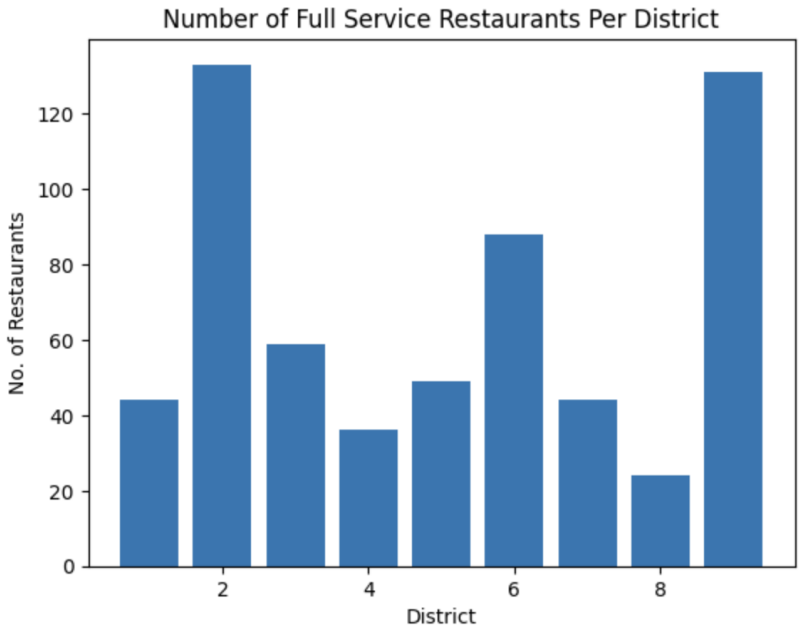
In order to reason about whether the low license counts were resultant of an overall low number of business that require licenses, or if the counts are a result of a lack of applications from businesses that could use a license, we decide to focus on a specific area within licensing that the city council expressed interest in which was liquor licensing. So, we performed an analysis that compared the number of restaurants and liquor stores across the districts, and then further we attempted to identify the proportion of full service restaurants that have an active liquor license across all of the districts. An important note about this portion of the analysis is that we had to resort to using the BDPA dataset of businesses which only contains a portion of the businesses in the Boston area, which are on the main streets of Boston. So while we believe that this subset of businesses is sufficient to characterize the larger scale trends, it is important to understand the context of the analysis.

The graph below shows the number of full service restaurants per district as well as the proportion of full service restaurants to the total number of businesses in the BPDA dataset. In this case it can be observed that while district 4 does not contain the fewest number of full service restaurants, it is certainly among the lower counts out of all of the districts. Next looking at the proportion of full service restaurants to the total number of businesses we can see that district 4 is not too far from the average ratio which suggests that because the average proportion is rather low, around .95, perhaps the low liquor license count could in part be due to the fact that district 4 as a whole doesn't contain as many businesses in general and thus by extension doesn't contain as many restaurants.

Proportion of FSR's to All Businesses In BPDA Dataset By District

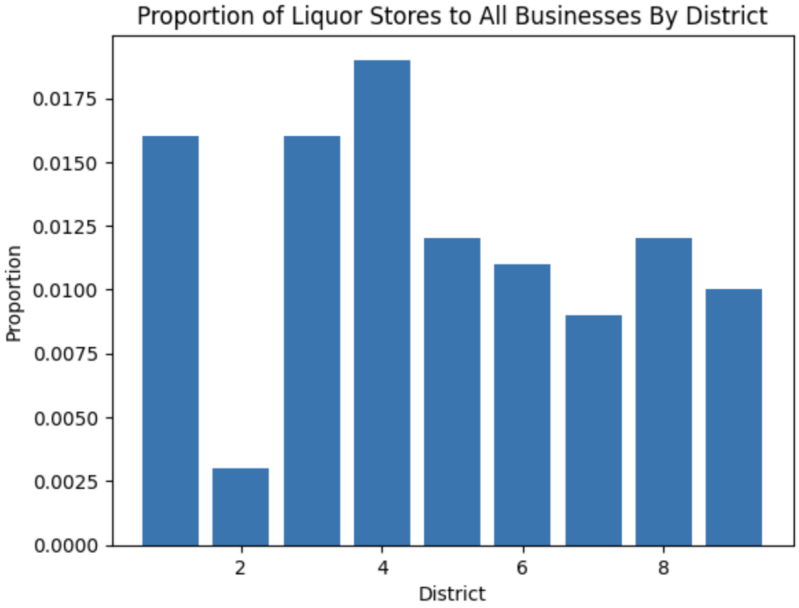


Quantity of FSR's Per District



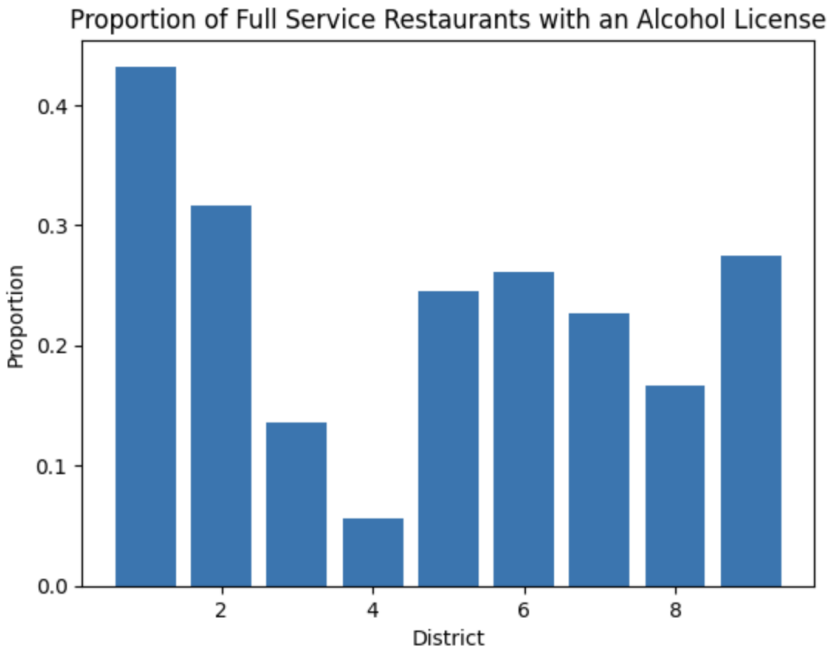
Next, we looked at the number of businesses labeled as “beer, wine, and liquor stores” in the BPDA dataset for every district, and further we looked at the proportion of liquor stores to the total number of businesses. We were surprised to find that for all of the districts, the liquor store count was lower than expected as depicted in the chart below. This could be in part due to the fact that perhaps liquor stores are not as common on the main business streets and thus they would not be included as readily in the BPDA dataset, but the consistency across all of the districts indicates that as a whole the Boston area is not heavily populated with liquor stores. Alternatively, this could suggest that Boston as a whole lacks in the area of liquor stores when considering the number of total businesses in each district.

Proportion of Liquor Stores to All Business By District



Lastly, we decided to calculate the proportion of full service restaurants that were listed in the active section 12 alcohol license dataset. By doing this we aimed to establish if it was the case that there are a lot of restaurants without alcohol licenses in district 4, which would indicate to us whether this was the reasoning for the low liquor license count. Below are the resulting proportions. There was some difficulty identifying common businesses across the BDPA full service restaurants and the active liquor licenses as the data was not consistent across the two datasets in terms of name and exact address, however we can gather that it most likely not the case that the majority of full service restaurants have an alcohol license. Furthermore, the results indicate that there is potential room for expanding the number of liquor licenses by increasing the number of full service restaurants that have liquor licenses. This could either be accomplished through an increase in liquor licenses distributed to already existing full service restaurants or on the other hand through an increase in new full service restaurants that obtain a liquor license as we know that district 4 does not contain a high number of these businesses from our earlier analysis.

Proportion of FSR's in Liquor LIcense Dataset



Project Limitations:

Although we have been able to work around the obstacles that we have faced so far, we think it is important to highlight these limitations in order to fully understand the context of our analysis. Firstly we would like to discuss the BPDA list of businesses dataset that we have been provided with, which contains a list of all of the businesses that exist on “main streets” in and around the Boston area. In this case, while there is no official definition for what constitutes a “Main Street”, from our observations it seems as though this indicates a highly dense area of businesses that exist along a single street. While this is most likely a good indicator of the majority of the businesses in the area, it is certainly not a complete list of the businesses that currently exist in District 4 which may slightly impact the integrity of our analysis. We attempted to find a more comprehensive list of businesses and further tried to use the Google Places API to construct our own list of businesses but we were unsuccessful. In addition to the BPDA dataset, the other major limitation that we encountered was with the MBE (minority business enterprises) dataset that we used. While it did contain a comprehensive list of MBE’s in the area, it did not provide any sort of coordinate data which made it very difficult to identify which MBE’s were actually located within district 4. For the time being, in order to get around this we cross referenced the MBE dataset with our filtered BPDA dataset to determine which businesses existed in both datasets and we were able to identify at least the MBE’s that exist on the main streets of District 4. Moving forward, if the city council is interested in pursuing this domain of analysis in a future spark project, a more comprehensive list of businesses in District 4 and the greater boston area would severely help to increase the potential for more analysis as well as the accuracy.