Team 2 Deliverable4 - Final Report Draftv2

Background and Motivation

The purpose of this project is to understand the current landscape of small businesses in District 4. We want to understand the current landscape of small businesses in District 4 and find features that will attract small businesses to the area while understanding the current issues small businesses face. District 4 is one of the most diverse districts and historically under-resourced in Boston, and current and previous councilors have been working on advocating for this district.

An analysis of the business environment will offer insight into the successes and shortcomings of small businesses in District 4. With this information, District 4 Councillor Brian Worrel can make informed decisions on what aspects to improve upon and focus on to most effectively support small businesses, as well as make District 4 an attractive place to locate to. Allowing small businesses to thrive can improve the community as a whole. With the growing population of chain stores, we are interested in seeing how small businesses are faring in comparison and ways for them to succeed in the community. The data used include the BPDA data, District 4 shapefile, as well as the shape files of the other districts, data from the Certified Business Directory of certified businesses, and Employment Data of Boston neighborhoods.

Based on our preliminary analysis described below, we decided to define a "small business" as businesses with fewer employees and/or less annual revenue than a regular-sized business or corporation and limited this to approximately 50 employees. (need to take a look at the annual revenue/if we can get this data). In addition, we determined that this does not include chain businesses with few employees, as chain businesses have select advantages in that the central organization assumes responsibility for the finances of its units.

Data collection and Exploration

Preliminary Analysis:

In our initial exploration of the businesses in Boston dataset, we noted that Estimated_Employment may be a relevant feature to study. Estimated_Employment is the number of employees a business has the resources for, and this may be indicative of the size of the businesses and its success. We filtered out the businesses for each district and discovered that in District 4, most businesses are in the employment range of anywhere from 2 - 10 employees with some operated by the owner only. However, we found one outlier, Fields Corner Store (located just on the border of D4 and D3) with 937 employees. This may be an inaccuracy in the dataset or this means Fields Corner Store is too large to be considered a "small business".

In comparison with District 6, a neighboring district to some parts of District 4, District 4 lacks the middle ground employment rate that District 6 has. District 6 not only has more

business in the 2-10 range, but also in the 10-30 range as well. The number of businesses in the 2-10 range is essentially doubled from D4 to D6. There is also a larger number of non-small businesses based on the estimated_employment. There may be factors such as district size that contribute to a larger number of businesses in D6 or a larger degree of success, which we will explore in the extension project.

We looked at District 7, which has a similar demographic to District 4, and observed that District 7 also outweighs District 4 in a similar way that District 6 does. There is a healthier middle ground in District 7 of businesses with 10-30 estimated_employment, though not as drastic as District 6. The number of businesses in the 2-10 range is actually very similar to District 4, but differences are apparent/show beyond 10 employees.

The number of businesses in District 4 is __; based on our definition of small businesses, we found that the number of small businesses in District 4 is __. This preliminary study of the estimated_employment of businesses in District 4, which is indicative of the successes and resources available of businesses in District 4, informs our goal of finding reasons for the failures and successes in District 4.

Base Project Questions:
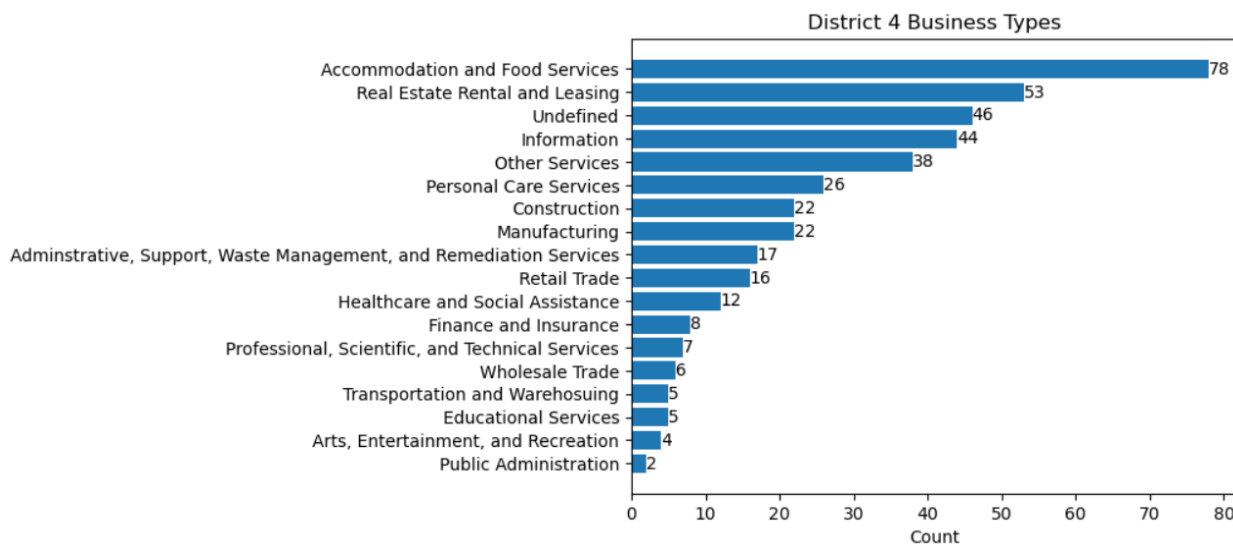<u>What businesses exist?</u>

Figure 1. Business Types in District 4

District 4 has a variety of businesses including but not limited to, Full Service Restaurants, Beauty Salons, Financial Transaction Offices. The most prominent business types are accommodation and food services, real estate rental and leasing, information (wireless telecommunication carriers, newspaper publisher, radio stations) and personal care services (Figure 1). Compared to other districts in Boston, District 4 has an overall lower number of

businesses. From looking at the two most prominent businesses types in District 4 we can see that District 4 has the second lowest number of these business types out of the 9 districts (Figures 2 & 3).
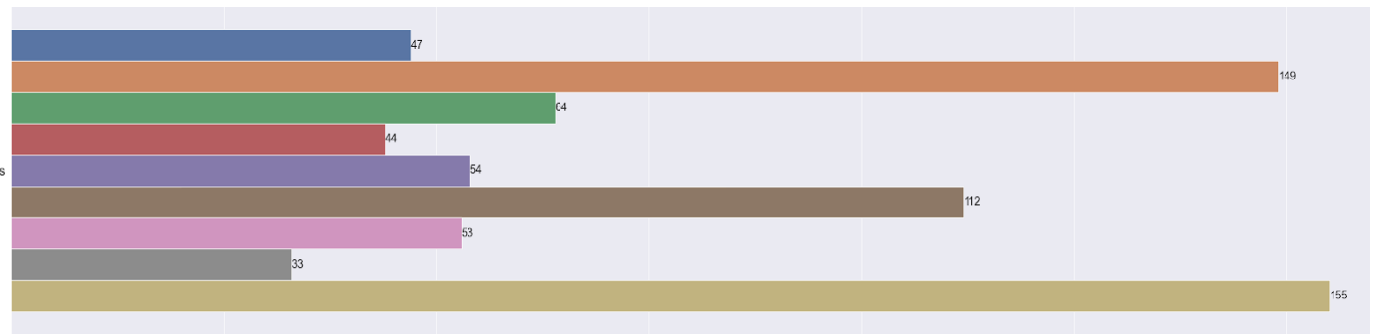

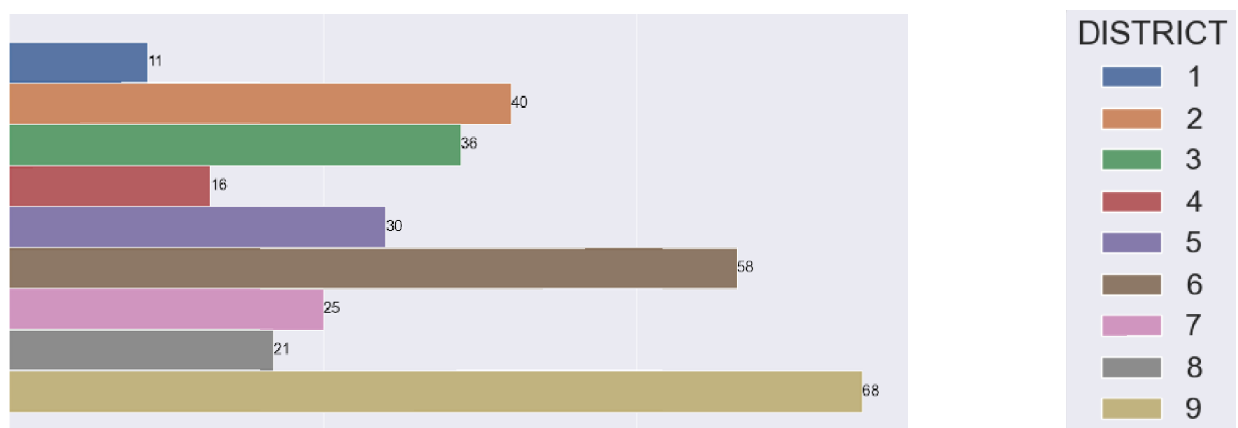Figure 2. Number of Accommodation and Food Service Businesses Across All Districts

Figure 3. Number of Real Estate Rental and Leasing Across All Districts

What businesses are missing?

Although District 4 covers a large variety of business types, some businesses have plenty of room to grow within the district. There is a limited number of educational services in District 4 – a total of 8 businesses fall into this category which includes schools, tutoring, Brazilian jiu jitsu, and yoga. District 1 has a total of 5 educational services which is comprised of driving schools, sports, and family tutoring, having the lowest number of educational services across all districts. District 6 has the highest number of educational services among all districts with 24 businesses consisting of early care and education, martial arts, yoga, art academy, dance, and driving school. Seeing that District 6 neighbors some parts of District 4, it may be the case that residents of District 4 travel to District 6 in pursuit of some of these educational services that are not present within District 4. Therefore District 4 could benefit from an increase in these services to increase engagement and maintain more business within the district.

We also noticed a lack of services in arts, entertainment, and recreation – there is one fitness center and one fitness instruction site. However, compared to other districts the gap is not as big as it seems. It is still the case that District 4 can benefit from more services in arts, entertainment, and recreation as it is one of the districts with a lower number of such services.
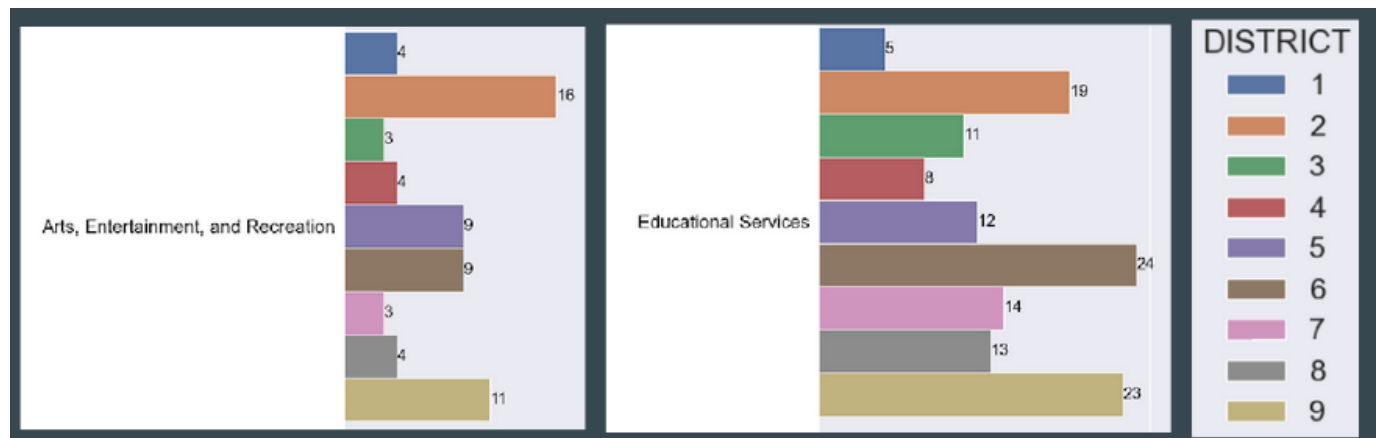

Figure 4. Number of Arts, Entertainment, and Recreation Across Districts

Figure 5. Number of Educational Services Across Districts

How many WBE (women owned) or MBE (minority owned) businesses?

Using the certified businesses directory, we found that only 10 businesses in District 4 have been registered as MBE. Of those 10 businesses, 3 are also woman-owned. The certified business directory dataset contains 680 data entries of businesses from all over the US. To build a more comprehensive picture of the state of businesses in District 4, we plan on exploring the number of businesses in the other districts that have been registered as MBE or women-owned and compare the number between districts. This may reveal information about how accessible it is to be registered, and whether the city council should be advertising these resources to small businesses.

Delete Later/Note: Re: business owner demographics - This may not be comprehensive data, but we can overlay the data with the MBE database to see if any are registered MBEs:
Boston Registry: https://data.boston.gov/dataset/certified-business-directory
State Registry: https://www.sdo.osd.state.ma.us/BusinessDirectory/BusinessDirectory.aspx

How many of the businesses are chains vs. independent operators?

Due to no existing dataset that contains all chain store names, we have created our own dataset using information from websites like Top 100 Retail Chains 2022 and restaurant chains. This dataset[will be uploaded, file name here] will not be comprehensive at all so we can

perform only basic chain store cleaning to get independently operated businesses. Using our created dataset of chain stores, we filtered through District 4 businesses and discovered that 6 businesses are chain stores. This number may be inaccurate(underestimate) due to the way businesses are named.

So 411 total businesses - 6 chain stores = 405 independently operated businesses(overestimate)

[chain stores in District 4]

Comparison to other Districts

[chain stores in District 2; total businesses in District 2 : 784]

get number of all businesses in each district & the number of chain stores in each district and make a bar graph of #chainstores/all_businesses

Extension Pitch:

Using the employment and spending data, we want to analyze the district's monetary status as it relates to small businesses and compare with other districts to identify areas of improvement for District 4. This extension would analyze the performance of small businesses in District 4. By comparing the performance across businesses, we hope to identify areas of improvement that are likely to aid in the success of a small business. We also hope to identify achievable goals that small business owners entering District 4 can expect to encounter. With

this information, our goal is to implement a foundational report on how District 4 can become more appealing to small business owners and furthermore, identify how they can see success.

We hope to answer the following questions:

<u>Are there any features that can help small businesses?</u>

Types of businesses:

Analyze Spending Data Analysis:

- Spending datasets show Change in In-Person card Spending
    - Keep in mind that it only looks at in-person card spending, so data is limited
    - The data only spans from 1/5/20 - 2/27/22, so we don't have access to spending from more recent months
    - Analyzing the spending dataset for Mattapan Square, we can see that this area appears to be near or at it's pre-Covid % change in consumer spending
    - This is a healthy sign for the district, it shows that, as of February, people were willing to spend roughly just as much money in-person as they were before Covid
    - With most restrictions being lifted since that time, it's possible that % change levels may have even gone up since
    - Unfortunately, the foot traffic dataset end date is 12/26/21 so we can not make any conclusions about how busy the area is overall at the time of the increase in % change in consumer spending
    - However, given the increase in spending data, we assume that foot traffic is also roughly the same as it was pre-Covid
    - Let's look for highest % change pre-Covid and highest % change post-Covid
- This data suggests that small businesses can now experience a fresh period of relief for the first time in nearly 3 years
- We know that these businesses experienced their harshest times during Covid and many had no other option but to close down - after analyzing the data we can conclude that small businesses in District 4 will continue to perform much better than they have in the last 2 and a half years because of the increase in in-person consumer spending
- A return to higher levels of spending and traffic can draw more businesses into the district because of the opportunity that now exists

Mattapan Weekly Spending Data
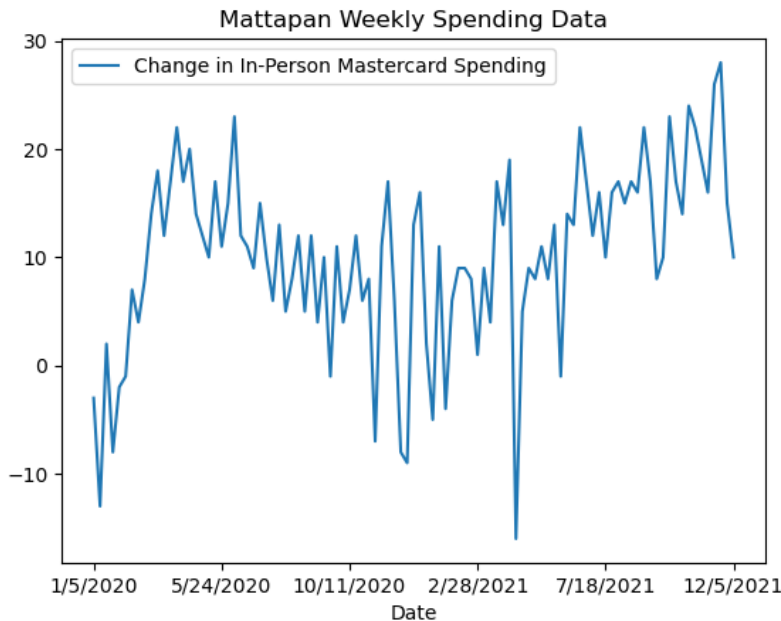
- Graph showing average spending from Jan 5 2020 until March 8 2020 ('pre-covid') and compare to March 15 2020 until November 15 2020 ('during covid') and 9/26/2021 - 12/5/2021 ('post covid')
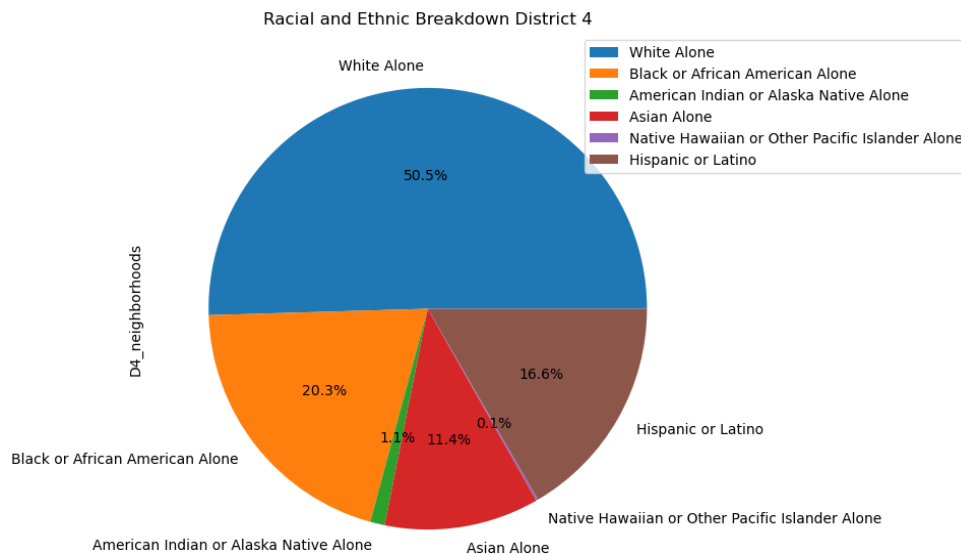
From previous data analysis:
- Food services, Real Estate, Information, and Personal Care Services make up more than half of the number of small businesses in District 4
  - These business types are far more common than the others, which leads us to infer that they should be the most successful of the categories
  - However, compared to all other districts, D4 has nearly the lowest amount of all those business types
  - The Food/Personal Care Services are great candidate categories to attract new small business operators to the area
  - Appealing to new businesses owners of these business types can help push the economy of D4 and further drive development of the smaller and less common business types

Are there any barriers to a business succeeding in District 4?
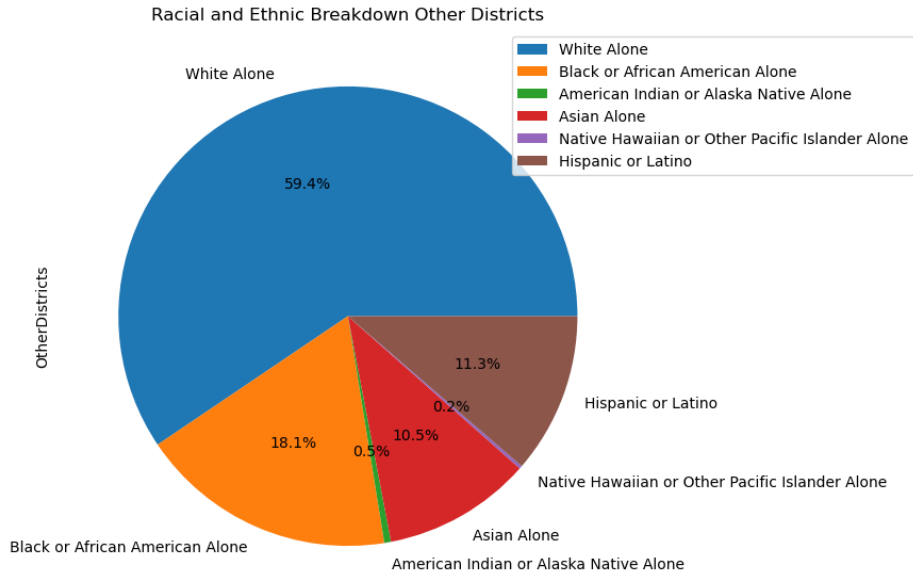- Do these businesses tend to spend a lot of capital? More or less compared to other districts?
  - Result: Unfortunately, we were later informed that the business spending data set is not actually business spending but rather consumer spending at businesses, which is what is described in the extension question above
  - We have slight indirect access to business spending through employment data

from the base project - from that data we can recall that District 4 had significantly lower employment than other districts (at the time of data collection)
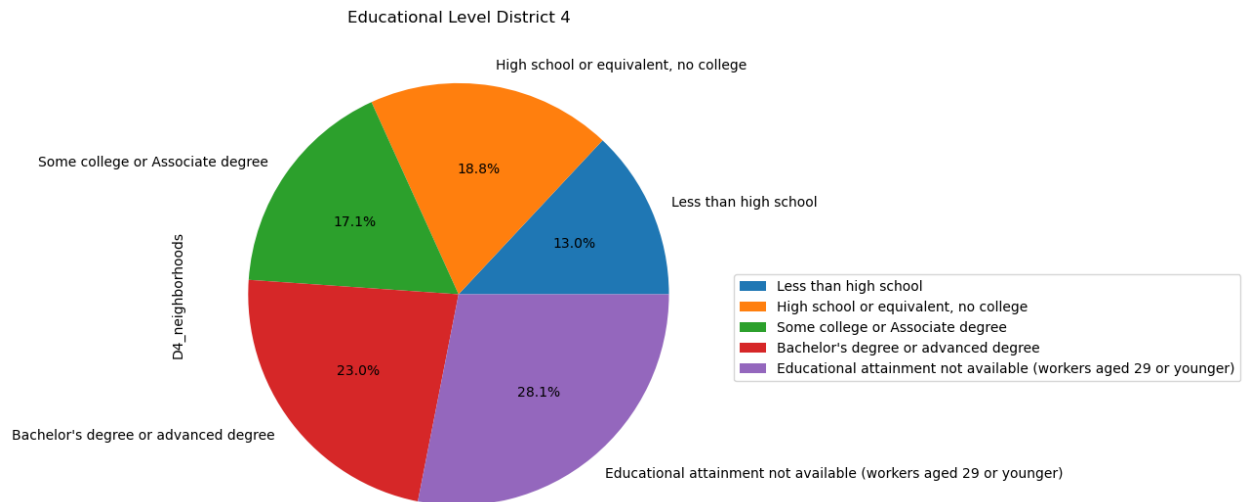
- ○ From this we infer that the spending in D4 is relatively low, mostly due to a decrease in income during the low in-person consumer spending trend
- ○ High cost of operation is not something that operators in District 4 should be concerned about
- Explore if similar business types in other minority districts perform better or worse
    - ○ Findings on District 4's business count by category (shown above) as well as count of employment (from base project) tell us that District 4 lags slightly behind District 7, a district that has similar demographics to District 4, in both aspects
    - ○ While these differences do not directly indicate that District 7 is performing significantly better, it does show that District 7 appears to provide slightly more opportunity
- Analyze the breakdown of employment data in neighborhoods in District 4 using the Employment Data datasets, focusing on the age category, race, education levels, etc of employees (see below for charts)
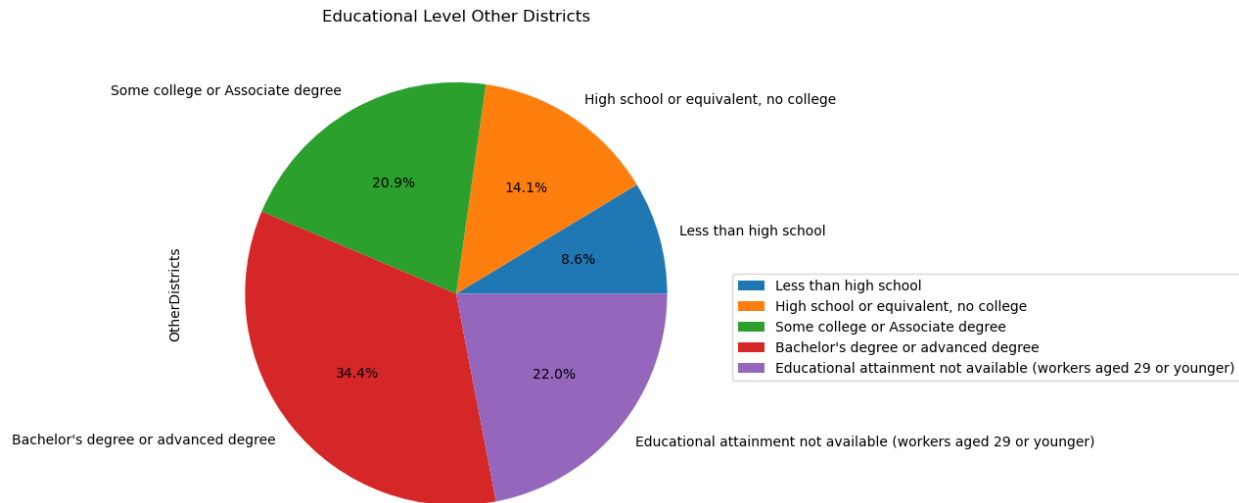
Racial and Ethnic Breakdown District 4

Racial and Ethnic Breakdown Other Districts

Legend:
- White Alone
- Black or African American Alone
- American Indian or Alaska Native Alone
- Asian Alone
- Native Hawaiian or Other Pacific Islander Alone
- Hispanic or Latino

OtherDistricts

- White Alone: 59.4%
- Black or African American Alone: 18.1%
- American Indian or Alaska Native Alone: 0.5%
- Asian Alone: 10.5%
- Native Hawaiian or Other Pacific Islander Alone: 0.2%
- Hispanic or Latino: 11.3%

○ District 4 is far more diverse than the average of all other districts with over 9% more non-White employees, but has a lower educational level overall (below) - more details in the following sections

Educational Level District 4

Legend:
- Less than high school
- High school or equivalent, no college
- Some college or Associate degree
- Bachelor's degree or advanced degree
- Educational attainment not available (workers aged 29 or younger)

D4_neighborhoods

- Less than high school: 13.0%
- High school or equivalent, no college: 18.8%
- Some college or Associate degree: 17.1%
- Bachelor's degree or advanced degree: 23.0%
- Educational attainment not available (workers aged 29 or younger): 28.1%

Educational Level Other Districts

Do businesses in District 4 get visibility online?

- Google maps, Yelp, etc.
  - We plan on merging District 4 businesses with Yelp businesses on coordinates (longitude, latitude) to find if any of the District 4 businesses can be found on Yelp
  - *Result:* Yelp does not allow scraping and instead provides the Yelp Dataset for people interested in Yelp's dataset to use, however, they provide only a subset of all yelp's businesses (around 150,000 businesses), and only 2 businesses in the provided dataset are in Massachusetts and there are no businesses in district 4
  - We've scoured through google maps and found that there is a lot of engagement under frequented small businesses in District 4, and many have great reviews
  - There is a strong online presence of small businesses in District 4 overall which we find to be one of District 4's strong points
  - High engagement shows residents of District 4 are actively looking for the best service options in their area - they desire high quality and they want to continue to spend their money in-person

Does a lack of non small businesses (corporations) negatively impact District 4's landscape?

Income District 4

Income Other Districts
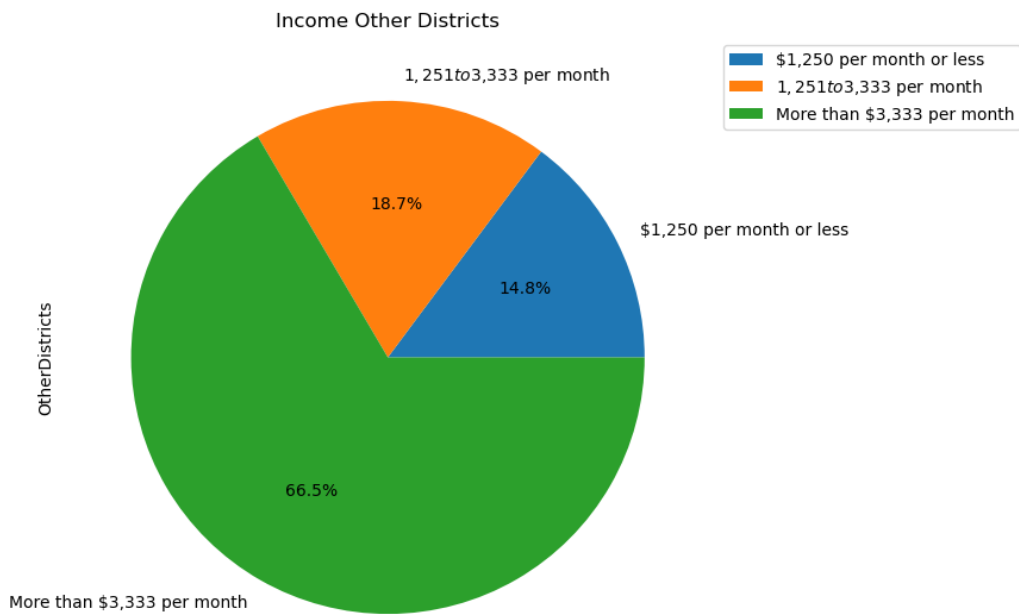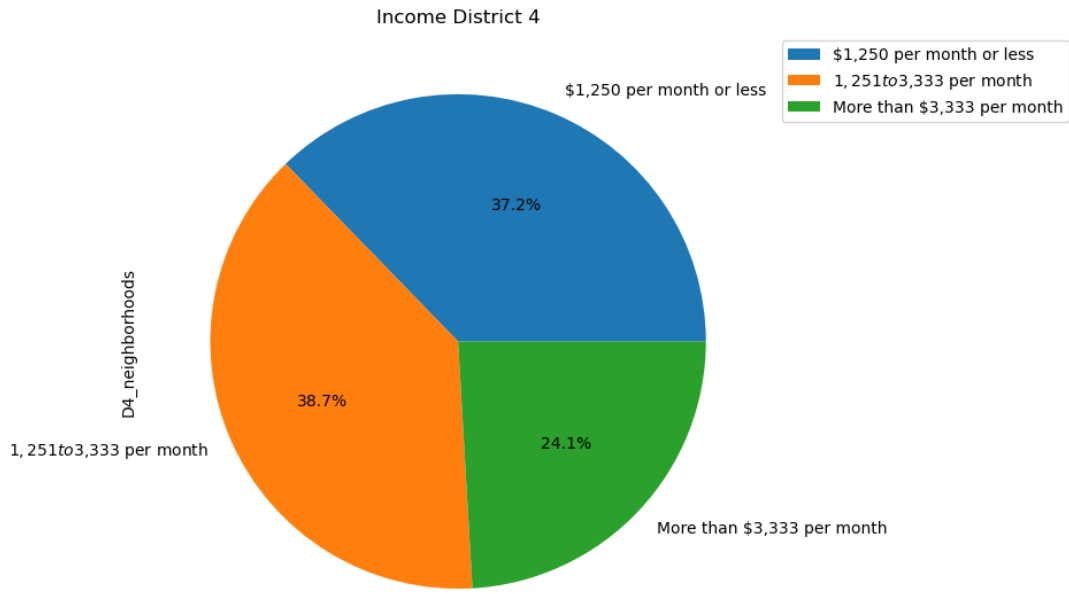
- District 4 has significantly less income than the average of the other districts
- Does this affect the number of customers willing to do business in District 4? (travel to District 4 for food, groceries, etc.)
  - District 4 may experience a lower amount of business overall because there is far less traffic than there would be in a District 2 sort of demographic (higher income, more corporate)
  - For example, residents in District 4 (or other districts) are much more likely to

travel to District 2 for either work or recreation than residents in District 2 are to travel to District 4 for the same reasons (at no fault to District 4, but this helps us understand why District 4 lags slightly behind it's neighbors)

- District 4's customers will primarily come from only District 4 - an important thing to note
  - This does not eliminate the possibility of prosperity for small businesses in District 4 to succeed
- Note what District 4 is lacking in comparison to other districts — health services and educational services that could be focused on to develop district 4
  - District 4 can benefit from more services in arts, entertainment, and recreation as it is one of the districts with a lower number of such services. This can keep District 4 residents within the district for these services rather than having them travel to neighboring districts like District 6/District 7.

maybe we can also look at size of the district by finding area of the shape file and compare the # of businesses in District 4 to the # of businesses in other districts/area of district

Our conclusion

A reasonable benchmark for small businesses in District 4

- Essentially - what is an achievable goal for small businesses?
  - In demand and attractive business types
  - Strong online presence to succeed locally
  - Expect neighborhood traffic - not large scale store-front sales
  - Modest yet profitable business with minimal cost of operation with little to no barriers of entry

Challenges faced

Challenges faced during data collection, exploration, and analysis mainly involved unsubstantial data or inconsistent information across multiple datasets.

- Cell phone traffic data does not provide a point of reference previous to 2019 and doesn't span a substantial enough timeline or geographical area
- It was difficult at first to filter out which data could be useful/meaningful to us
- Some inconsistencies
  - within a dataset some aspects are not categorized or are categorized inaccurately between multiple datasets
  - some business names not correct on dataset – name may have changed
  - data may be outdated
- Finding Minority owned businesses after seeing such low number MBE certified

businesses in the MBE certifications dataset
- ○ Certified Business Directory dataset is not comprehensive
- ○ While businesses can put MWBE status on their Yelp pages, the Yelp dataset does not include this information
- Foot traffic and spending data are for a street/neighborhood and not business
  - ○ Can skew information due to District divisions (reword this but essentially a neighborhood can be in more than one district so the information gleaned from the datasets that are only provided via neighborhoods are not exactly accurate to the district)

Suggestions for the future of the project
We are using the shapefile from before the redistricting; we believe this is important in helping understand the business environment of the original District 4.

More up to date data to help with finding more conclusive and relevant results
- use updated shape file