Deliverable 1

Akshad Ramnath, Neeraja Mehta, Ruiling Zhang, Min Han

Collect and pre-process a preliminary batch of data

Our initial step was to understand the geographic areas of the City of Boston based on race, ethnicity, income, and home ownership/renters. We then used data from Data-neighborhood.csv to perform our analysis. For the first deliverable, we mainly focus on race and ethnicity in the different neighborhoods and how the companies receiving funds are distributed in these neighborhoods.

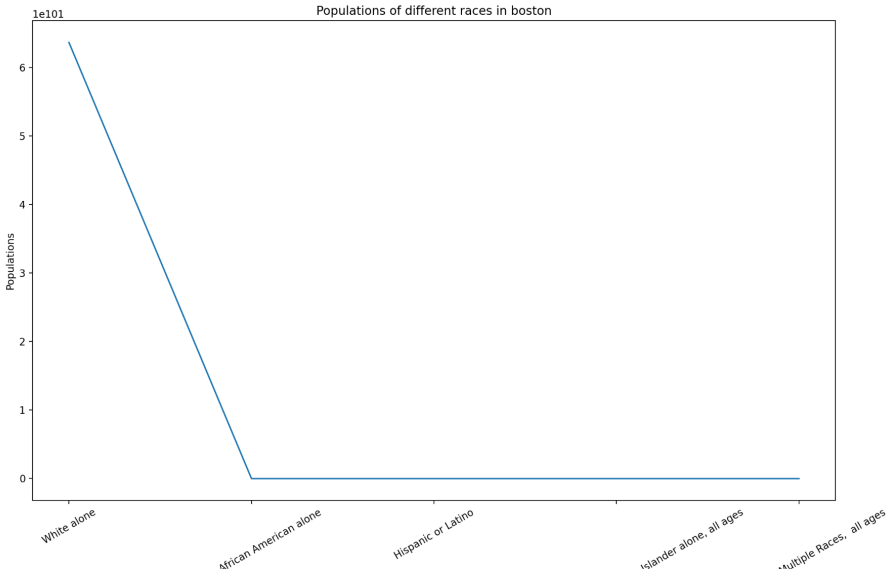
Perform preliminary analysis of the data

1) We use the Boston Neighbourhood Data from https://data.boston.gov/dataset/2020-census-for-boston and filter the data to see how the different ethnic groups, especially Hispanic and Black, are spread among the different neighborhoods of Boston. From our analysis, we found out that Mattapan has the highest black population - 68.29% followed by Hyde Park - 49.96%, Roxbury - 41.52%, and Dorchester - 34.95%. For the LatinX population, we find that East Boston has the highest percentage - 50.38% followed by Roxbury - 30.47% and Roslindale - 25.09%.

2) We have done some analysis for these batches of data. For instance, we collected and pre-processed some data on the relationship between the areas in Boston and the populations of different races, and ages.

For the data of the population, since the value is too big, we decided to use the average values for the comparison.

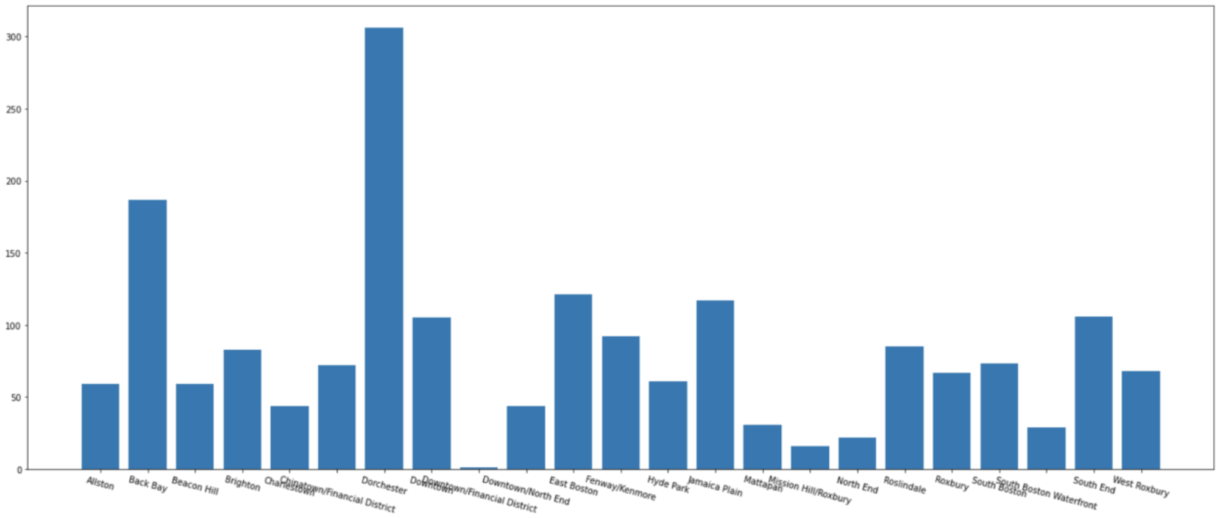
The image below shows that the population of “White alone” is the largest in the Boston area. It is nearly six times the population of the other races. Since the gap between the population of “White alone” and other races is too large, there might be a risk of the distribution of relief funds between different races.



Answer one key question

Question 1: Where did business assistance go during the pandemic? What were the demographic profiles of the communities where the businesses were located?

According to the analysis of the Small Business Relief Fund 2.0, we plot the following histogram. It can be seen that the number of companies in Dorchester is the highest in Boston, which is around 300, followed by Back Bay. Companies in the remaining neighborhood are less than 150. Most of the business assistance went to Dorchester and Back Bay during the pandemic.



The demographic profiles of the communities are shown below.

	Total:	White alone	Black or African American alone	Hispanic or Latino	Asian, Native Hawaiian and Pacific Islander alone, all ages
Allston	28621	14634	1451	3657	7173
Back Bay	19588	14056	718	1326	2604
Beacon Hill	9336	7521	252	537	630
Brighton	48330	30596	2289	4978	7801
Charlestown	19120	13626	990	2075	1650
Chinatown	7143	1898	297	477	4281
Dorchester	122191	27411	42714	25285	13360
Downtown	13451	9174	537	961	2286
East Boston	43066	15760	1403	21700	1932
Fenway	37733	20456	2396	3643	9218
Hyde Park	33009	7449	15171	7901	677
Jamaica Plain	41012	22032	4686	8921	2985
Longwood	4096	2573	334	381	657
Mattapan	23834	1489	16277	4079	490
Mission Hill	17886	6950	2469	3397	4228
North End	10805	9306	141	528	445
Roslindale	29386	13428	6045	7373	1018
Roxbury	54905	7182	22796	16728	3277
South Boston	37917	29139	1529	3887	2077
South Boston Waterfront	5579	4315	160	316	482
South End	29373	16618	2959	3783	4677
West End	7705	4933	338	613	1440
West Roxbury	31561	20918	3312	3567	2451

Over 50% of the population in Dorchester is either LatinX or Black. However, BackBay's total Black and LatinX population combined is below 11%.

Refine project scope and list of limitations with data and potential risks of achieving the project goal

One potential limitation is that the neighborhood data set as well as the company dataset have the area names but not the longitude and latitude which might make it difficult to plot on a map and may make the data hard to read.

Submit a PR with the above report and modifications to the original proposal

1. **What is the project focus/overall goal?**
    a. The city allocates money to various funds designed to offer relief and support to benefit residents and businesses in Boston. Councilor Mejia wants to understand whether these benefits are being distributed equitably and if the money the city said they were going to spend has reached and impacted the people they said it would serve. The goal is to assess whether the money is going and if it is being equitably distributed.
2. **What type of data will you collect or be analyzing?**
    a. Relief funds, capital investments, and Census Data: Demographic information
3. **What are the potential limitations of the project?**
    a. There might be other variables not mentioned in the data collected that effect which areas get funding. For example, natural disasters, power outages, etc., are external factors that can affect the funding distribution
4. **Why is this project important or Why did you choose this project?**
    a. This project is important because through the data analysis we can see how the relief funds are being distributed to the residents of Boston and if the residents that actually need the money are getting access to the funds so they can make use of them.
5. **What are your team's next steps? (include action items/tasks)**

a. Scheduling weekly meetings, further analysis of the cleaned data, plotting the various data sets into the various census block groups