Analyzing 311 Data - Deliverable 2 - Team 2

Aishwarya Reddy Lachangar(Team Representative)
Juan Guang
Gaoyuan Han
Anoohya Veerapaneni

Project Motivation: City Council Julia wants to gain an understanding of how city is responded to service requests made through the 311 System, and whether or not they are resolving these requests in a fair and unbiased manner.

Goal of our Project:

- First to get the better understanding of past reporting patterns.
- Econometric measures of neighborhood characteristics.
- Analyse Data(past records) and help the Boston residents.

Background to understand:

- What is 311 system.
- Types of cases System receives.
- Basic terminologies.

➔ What have we done so far:

- Data collection and cleaning(as required)
- Data Analysis
- Finding patterns in the past records

➔ Work division:

- All work has been equally divided

➔ Data Used:

- DataVerse 311 Data
- Census data(for demographics)
- Social Vulnerability Index
- Analyze Boston (311 data)

What is the pattern of demographics and service requests: looking at census block group and social vulnerability index?

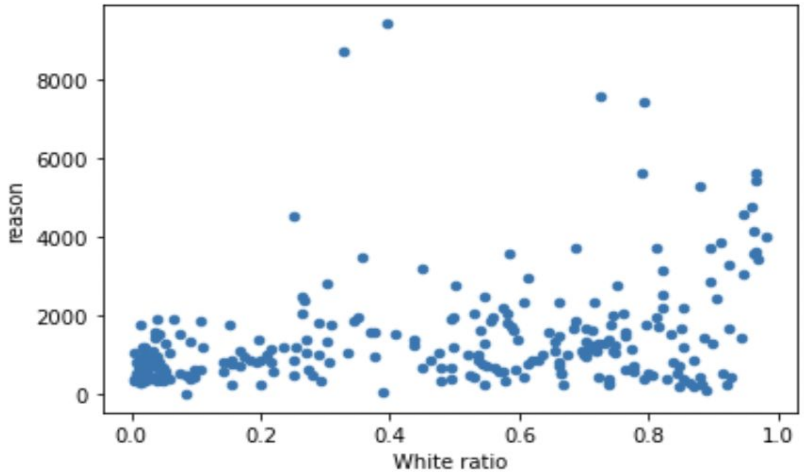
1. We find the top 5 most frequent types/reasons of cases

```
: df_all_311['reason'].value_counts()
```

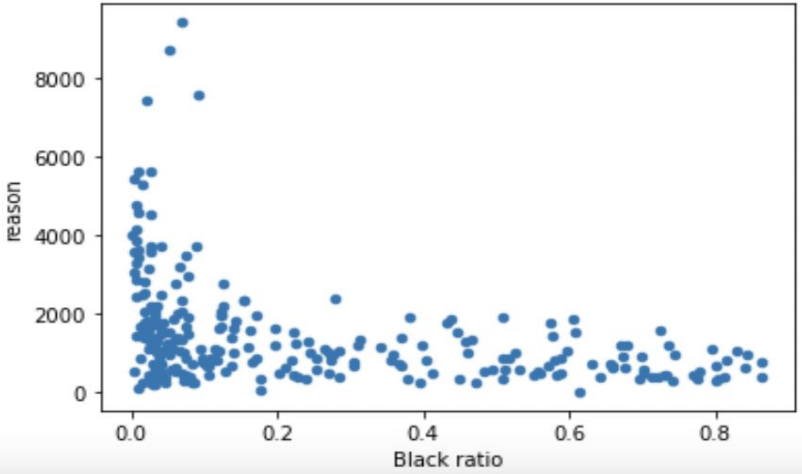
```
: Enforcement & Abandoned Vehicles     364678
  Street Cleaning                      361859
  Sanitation                           348704
  Highway Maintenance                  298163
  Code Enforcement                     189721
```

2. We analyse number of these cases with respect to the census block group and SVI to construct the pattern

Number of cases for Enforcement & Abandoned Vehicles versus White people ratio



Number of cases for Enforcement & Abandoned Vehicles versus Black people ratio



From census: we calculate correlation for each dimension

.

	Sanitation	Street Cleaning	Enforcement & Abandoned Vehicles	Highway Maintenance	Code Enforcement
White ratio	-0.0246	0.0418	**0.3288**	0.2671	0.1957
Black ratio	0.1845	-0.1125	**-0.4417**	**-0.3091**	**-0.3036**
Asian ratio	-0.2873	-0.0183	0.1907	0.1564	0.1283
Hispanic ratio	-0.1448	0.0274	-0.1895	-0.0486	-0.1284
White ratio (over 18)	-0.1210	0.1395	**0.3806**	**0.3729**	**0.2747**
Black ratio (over 18)	0.1829	-0.0143	**-0.3715**	-0.1926	**-0.2221**
Asian ratio (over 18)	**-0.2851**	0.0863	**0.2589**	**0.2529**	**0.2073**
Hispanic ratio (over 18)	-0.1631	0.1460	-0.0725	0.1506	-0.0120

Correlation coefficient: a number from -1 to 1. -1 suggest a negative correlation, +1 suggest a positive coefficient.

From census: analysis

	Sanitation	Street Cleaning	Enforcement & Abandoned Vehicles	Highway Maintenance	Code Enforcement
White ratio	-0.0246	0.0418	**0.3288**	0.2671	0.1957
Black ratio	0.1845	-0.1125	**-0.4417**	**-0.3091**	**-0.3036**

- Each number(coefficient) represents a story for the communities in a dimension
- For example, **0.3288** shows a positive correlation, it means that if there are more white people in a community, then there tends to be more cases for Enforcement and Abandoned Vehicles reported
- Similarly, **-0.4417** shows a significant negative correlation, meaning that if there are more black people in a community, then there tends to be less cases for Enforcement and Abandoned Vehicles reported

From SVI: we calculate correlation for each dimension

	Sanitation	Street Cleaning	Enforcement & Abandoned Vehicles	Highway Maintenance	Code Enforcement
OlderAdult ratio	**0.2991**	**0.1937**	**0.1179**	**0.2671**	**0.1409**
TotChild ratio	**0.4116**	**0.1648**	0.0174	0.0515	-0.0143
LEP ratio	-0.0929	-0.0514	0.0276	-0.0715	-0.0401
Low_to_No ratio	**-0.1268**	-0.0316	0.0236	-0.0945	0.0026
MedIllnes ratio	0.1105	**0.2088**	**0.2535**	**0.3302**	**0.3035**
Disability ratio	**0.2477**	0.0874	0.0476	0.0601	-0.0187

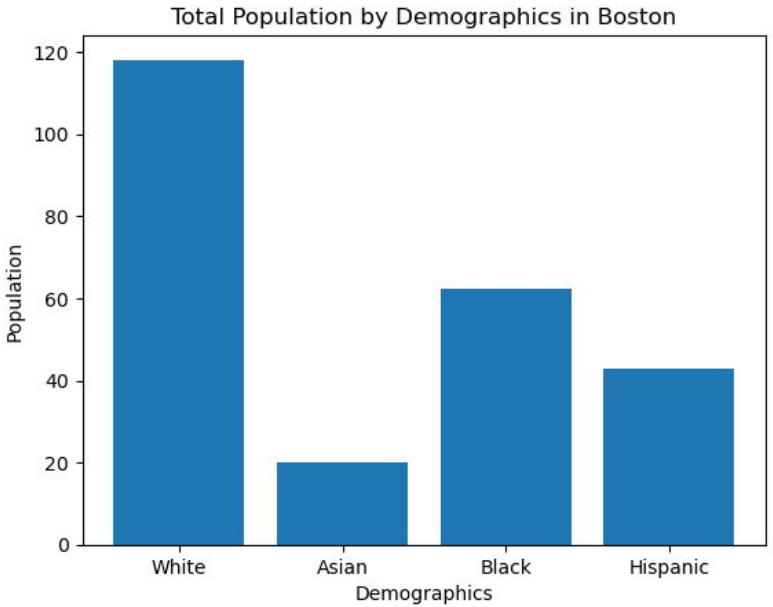
From SVI: analysis

	Sanitation	Street Cleaning
OlderAdult ratio	0.2991	0.1937
TotChild ratio	0.4116	0.1648

- We can see that for Older adult ratio and Children ratio, the coefficients are all significantly positive, this means that if there are lots of children and elder people in a community, then there tends to be more cases about Sanitation and Street Cleaning reported
- Speculation: it is reasonable since if there are lots of children and elder people, then people who take care of them probably want them to live in a clean environment, which leads to more attention about Sanitation and Street Cleaning

Which communities are most empowered based on **ALL** 311 service requests?

Demographics: Overview



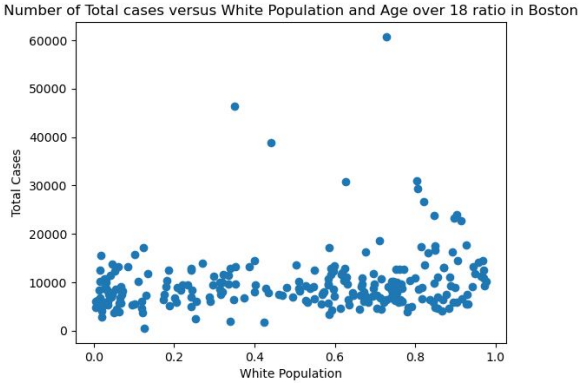
This bar chart shows the proportion of four race with the whole population in Boston city .

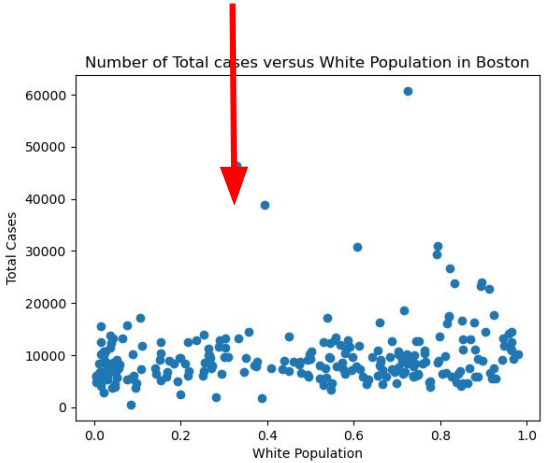
Correlation table for Demographics(Race & Age):

Different races	Correlation ratio between 311 case and different races	Correlation ratio between 311 case and different races over 18 years
White	0.2469	0.2432
Asian	0.0415	0.0411
Black	-0.2938	-0.2923
Hispanic	-0.1335	-0.1299

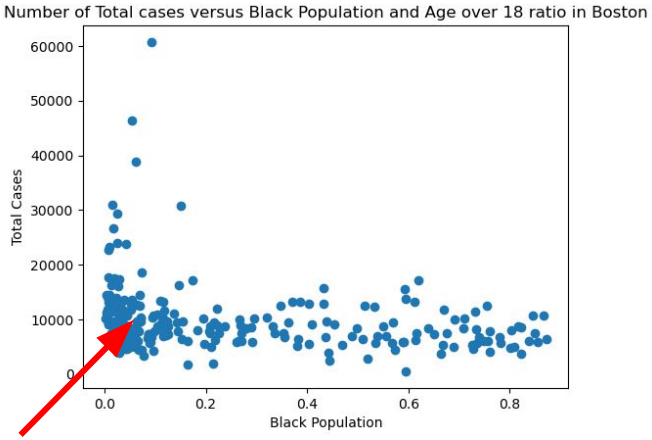
The 0.2469 & 0.2432 shows + correlation that means, as one variable increases the second variable increases too. Similarly, -ve correlation -0.2938 & -0.2923 means, that they are inversely related, one increases and the other decreases.

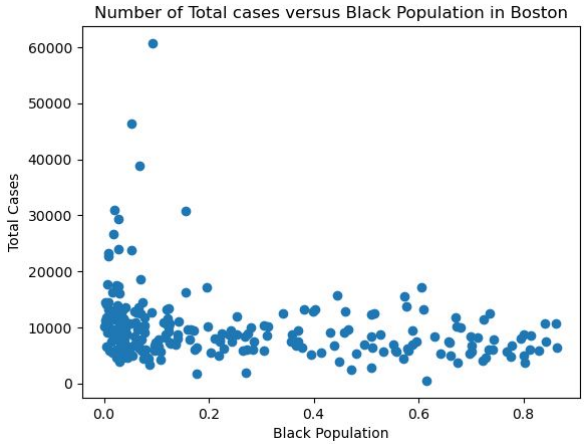
Scatter plot for census + 311 data(race & age):



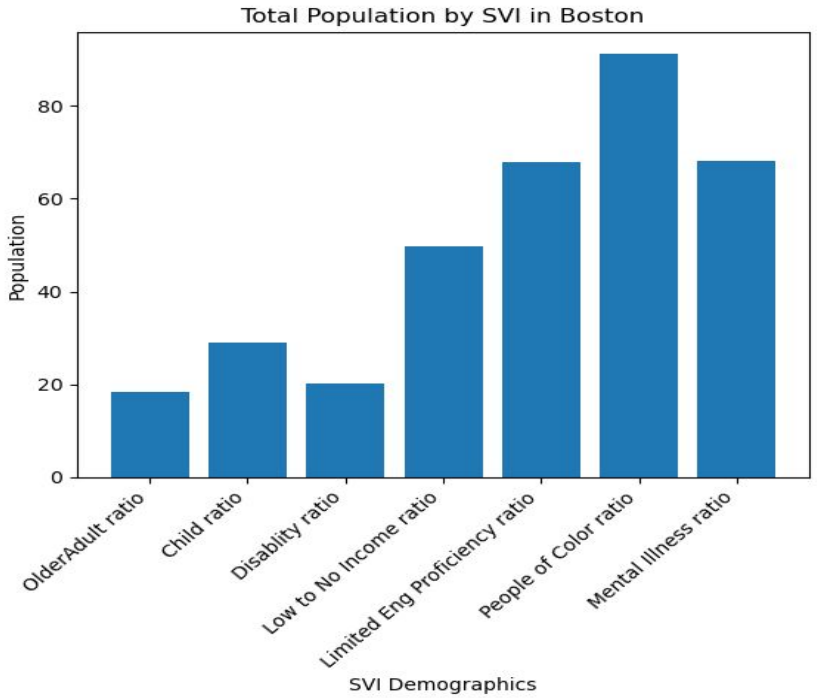


The four graphs shows the visualization of the correlation that we saw in previous slide.





Social Vulnerability Index: Overview

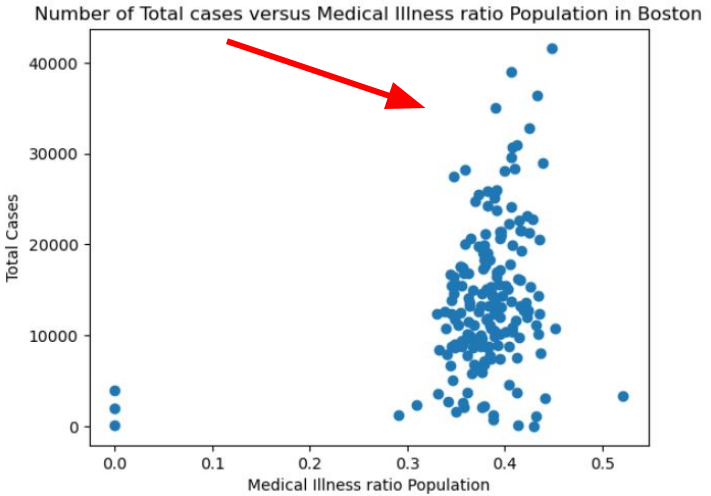


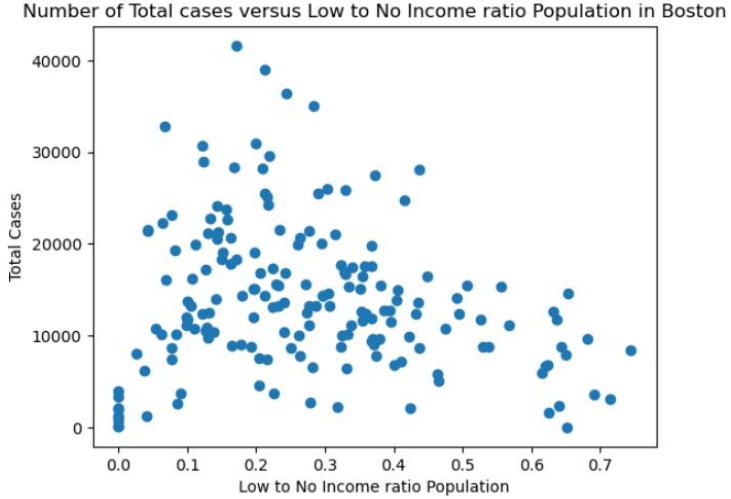
This bar chart shows the proportion of different communities compared to whole communities population in Boston City.

Correlation Table for SVI:

Communities	Correlation ration between 311 cases and
Older adult	0.1465
Disability	-0.0462
People of colour	-0.1179
Limited Eng proficiency	-0.0902
Medical illness	0.2921
Low to no income	-0.1494
Child	0.0679

Medical Illness community has a +ve correlation of 0.2921, whereas Low to No Income community has a -ve correlation of -0.1494.





Exploration Analysis: Street Cleaning & Sanitation

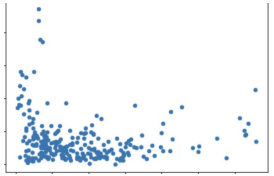
Rationale	*Why is this extension important? What information would this reveal? Why is it interesting to you and your team?* We have already analyzed the 311 data with respect to the composition of communities, including the racial and age composition, and the dimensions included in SVI. From our analysis, especially with the 311 data, I think we have some significant and helpful findings that tell us about communities and their needs in Boston. Then it is very natural for us to think about the government's awareness and response to the 311 service and the composition of communities. We want to explore that in the next step. Our exploration work hopefully can show the government's awareness and response to the 311 service and the composition of communities or how good is the government doing?

Exploration Analysis: Street Cleaning & Sanitation

Questions for Analysis	*What information do you hope to find based on looking at this additional data? What are you curious about? Do you have any hypotheses?*
	We want to construct analysis with respect to the most frequent types of cases. We notice that Sanitation is a very frequent case from 311 data, so we want to see the performance of corresponding service. For Sanitation, we have data about the current trash collecting schedule. In the next step, we are going to discover the relationship or correlation between the unclosed sanitation cases and the frequency of trash collecting in a location. Our hypothesis is that for places where there are lots of unclosed Sanitation cases, the trash collecting service is not frequent enough. In general, we are curious about how the work from the government corresponds to the actual situation of the communities and try to discover some incoordination between the actual need of people and government's planning.

Exploration Analysis: Street Cleaning & Sanitation

Data Sets & Sources	*Include the dataset(s) you will be using and any additional information that will be used directly in the extension* https://data.boston.gov/dataset/trash-schedules-by-address/
Data Visualizations	*Include a description of your proposed charts/ graphs for the final deliverable and axis for each* The graph will be like this: our y axis will be the number of trash collecting service per week for a location denoted by a zip code; our x axis will be the ratio of unclosed Sanitation cases in a location denoted by a zip code.



Thank you