Deliverable 1:

Team mambers :

Weichen Jiang, Class of 2023, weichen@bu.edu.
Cheng-Ping Lin (Team Leader), Class of 2023, momolin@bu.edu.
Kyran Chen, Class of 2023, kyc0510@bu.edu.

1. Collect and pre-process a preliminary batch of data
   Yes, we've gone through and process the data from CLIMATE READY BOSTON SOCIAL VULNERABILITY dataset and 311 requests 8-years around 2020 as the former data is collected in 2020. We did normalization for some of the data to diminish the influence of population.
2. Perform a preliminary analysis of the data
   Yes, we calculated the number of 311 cases in each census group and sorted districts by how many 311 cases they have received. We've also grouped data by case type for further research.
3. Answer 1~2 Key questions
   (1) How do we calculate the social vulnerability index of each census group, which is also: How do we define social vulnerability status of a census group.

   Based on the ranking method of CDC/ATSDR SVI we have found. We first divided each calculated social group by total population. Then we sorted census group by the rate contributed by each social group, for every census group at the first 10% in each sort, wo add one point on their SVI. After going through all social group, we got a rank of social vulnerability status. The SVI rank we've got will be pasted latter in this deliverable.

   (2) How do we estimate 311 request based on the calculated SVI?

   As we've got the SVI of all census group, we analyzed how well the SVI predict the load of 311 request. For every district in each SVI rank, if its' position in 311 cases number list is in the range of it's rank's district, we say that SVI rank did a good predict (For example, let's say district EX has a SVI rank 2 and all districts in 151 to 171 position has a SVI rank 2. If we found that in 311 cases number list the position of district EX is between 151 and 171, we say SVI rank did a good predict).

   (3) Conclusion: After comparing we found that using SVI, 135 out of 183 census blocks received a good predict. We think the bias between the two features

might be that there are spam calls, also the SVI rank only give a approximate predict as it is a rank.