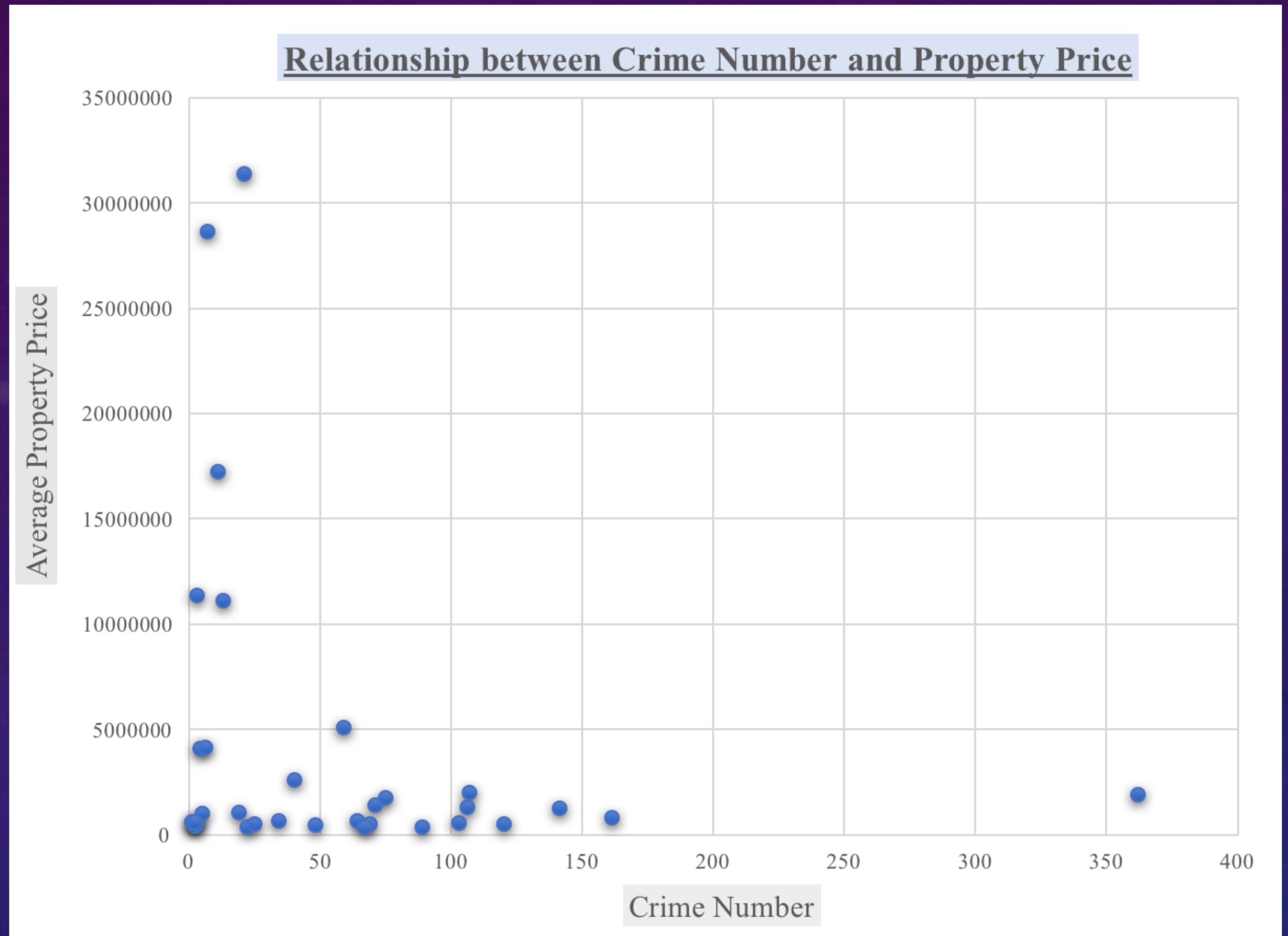
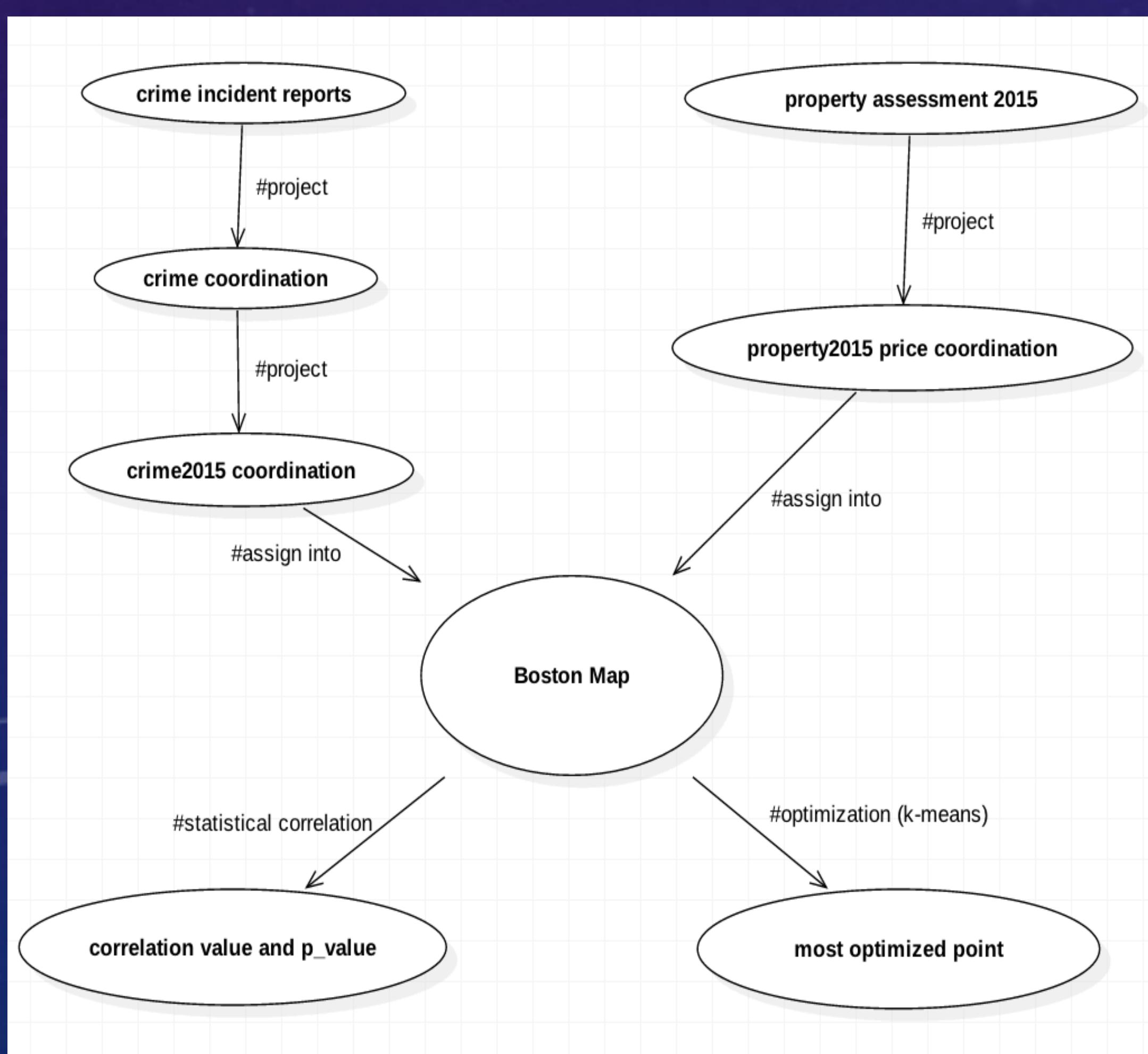


Introduction

For most people nowadays, the most pressing demand is to improve their living standards. In a crowded city like Boston, one of the main factors that influence people's decision on choosing residential properties is the amount of crime incidents that had happened in their respective area. For this project, we attempt to analyze the relationship between property prices in 2015 and the number of crime incidents that had happened in 2015 in the Boston area.

We use Relational Paradigm to extract certain data from *City of Boston Data Portal*: Property Assessment 2015 and Crime Incident Reports. First, we assume higher property price will lead to lower number of crime incidents. To test our assumption, we use statistical analysis to find if the correlation value between the two is negative, and how strongly they are correlated; we also look at the p-value to check the accuracy of our correlation value. Secondly, we use k-means technique to find the most optimized point where the property is the cheapest yet the safest, which indicates that the frequency of crime incidents happened around that point is the smallest.

Overall, we implement two techniques: statistical analysis and optimization(k-means) to study the relationship between the 2015 property price and the number of crime incidents in Boston area.



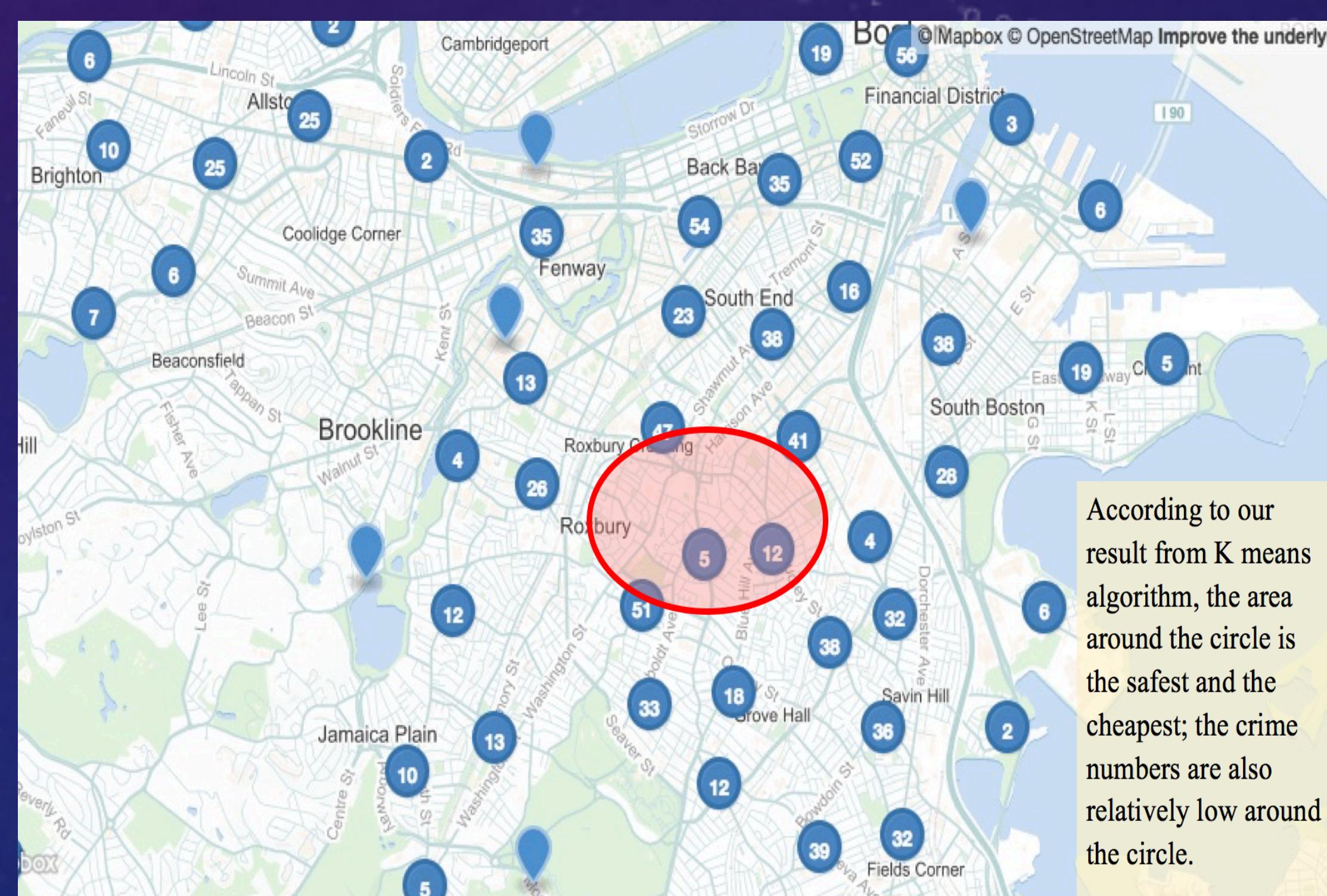
Statistical Analysis

As for the statistical analysis, we want to find the correlation between property price and number of crime incidents; we expected that the higher property price will lead to lower number of crime incidents. We first use the definition of r-tree and polygon to correspond “property15_price_coordination_float” to “zip_to_coor,” compute the number of properties and the average price of properties in each zip code, and the result shows all the zip codes along with its corresponding average property price in Boston area.

Second, we also use the same definition we mentioned earlier to correspond “crime_15coordination” to “zip_to_coor,” compute how many crime incidents happened within each zip code, and the result will show all the zip codes along with its corresponding number of crime incidents. Finally, we find the correlation between the average price of properties and the amount of crime incidents. The number we got is: **-0.246**.

K-Means

For the k-means section, we want to find the most optimized point where the property is the cheapest yet the safest (farthest from crime incidents). To do so, we input “crime_15coordination” to our k-means function and find the most optimized coordinate. Then we correspond the coordinate back to polygon and claim the zip code we get is the area where people can obtain property that is both the cheapest and the safest. The zip code we obtained is “02119”.



Conclusion

The correlation value, **-0.246** shows that property value and crime incident are indeed negatively related (an inverse relationship) which matches our assumption. However, $|(-0.246)|$ also shows that they are weakly related, and the p-value we obtained is very high. Moving forward with this project, we should consider factors such as population density, school districts and police districts more thoroughly since they are also contributors to crime rates. Further, due to time constraint, we were only able to run the small portion of our crime data, which might outcome the unideal correlation and p value. Thus, we assume by running more data in the future, we will be able to obtain a better correlation and p value.