

Crime, Cannabis, and Connectivity

DS3000 / DS5110 (Rachlin)



In this assignment you are a Data Scientist investigating crime trends over the past 4-5 years. You are particularly interested in exploring whether the Covid-19 pandemic impacted patterns of crime in the City of Boston.

1. Use Panda's DataFrames to analyze and visualize crime data from 2018, 2020, and 2022.
2. You are free to use any advanced visualization tools you like. You may find the Seaborn library particularly handy, but it is not a requirement. (seaborn.pydata.org). It is required that your visuals be captioned and publication ready. At least one of your visualizations should include a geospatial map of crime like how you mapped out the trees of Boston in Homework 1. You should apply some method of interpolation or convolution to spatially smooth your data over the area of Boston and use array-based operations to measure and visualizes *changes* over time.
3. Summarize your analysis with a 2-page written research report that includes an Introduction, Methods, Analysis, Conclusions, and References.

Extra credit – both sections (10 points): I have also provided you with locations of approved and proposed Cannabis dispensaries. Some research has suggested that the presence of Cannabis dispensaries may not significantly contribute to an increase in crime rates. Investigate and draw your own conclusions with a captioned visualization for full credit.

Extra credit – both sections (10 points): Access to WiFi can be considered an issue of social justice. In today's increasingly digital world, internet access has become a vital tool for individuals to access information, education, job opportunities, government services, healthcare resources, and social connections. However, not everyone has equal access to the internet, which can exacerbate existing social and economic inequalities. I have provided you with a list of approved wireless antenna installation requests in the City of Boston. Create a map of Boston that identifies regions of lesser or greater WiFi access based on proximity to WiFi antennas. You can carry out this analysis by measuring at each location the distance to the nearest antenna, or you could try some form of interpolation or convolution that counts the mean number of antennas within a given square area. A captioned visualization is sufficient for full credit.

Submit your Research Paper (.pdf), code (.py or .ipynb), and other visualizations to GradeScope.