Section 3: Week 5: Mining Theories Through Data Problem

Nate Bachmeier

TIM-8130: Data Mining

February 9th, 2020

North Central University

# Imagine Boston 2030

As the city of Boston continues to grow, the Mayor has recognized that not all citizens are equally gaining those rewards. The dichotomy of the metropolitan area has led to distinct clusters, disproportionally filled with wealth or crime. Nearly one in five residents live in poverty and do not have a path to improve their situation. Instead, Walsh asks the community to imagine a different Boston, one that within a generation creates access to affordable housing, reduce crime, and improve wage inequality (Kovatch et al., 2017). This transformational process not only raises the standard of life for Bostonians, but it also encourages investment through tourism and economic expansion. Meeting their ambitious goals requires a strategy to track progress and ensure the progressive agenda is unfolding correctly. The Boston Planning and Development Agency (2017) publishes several Key Performance Indicators (KPI) around each of the major pillars of the plan, in addition to estimating January 2000 to present. This approach produces a highly efficient hierarchical reporting structure that aligns the mission with specific questions and supporting evidence. For instance, the objective *Reduce Housing Cost Burden for Bostonians* qualifies success as low-income residents are not living paycheck to paycheck, and quantifies it through the ratio of wages to housing costs. According to their chart, the ratio has held steady for low-income residents between 45-50% since 2000. When a population does not have disposable income, they are at high risk that minor incidents can become life-impacting. Consider the cascade of events from a 150$ auto-part failure, a broken leg, or a sick child—each scenario forcing the person to take on debt, further compounding their problems, and getting further from the finish line. These human problems need solutions, but what moves the needle the most in the least amount of time? Merely writing a blanket check or applying broad policy can be ineffective without addressing the crux of the issue.

## Improving Housing Affordability

The improvement plan suggests that the price of housing is too high, and this limits access to a broad audience. While their evidence supports this observation, does it tell the entire story? Wikipedia (2020) states that Boston is 89.63 square miles, of which land encompasses 48.42, and this needs to hold approximately 700,000 residents. According to the US Census, the population is consistently increasing at 1.25% annually (ZipAtlas, 2020). With the supply of land being finite and the continuous increase in demand, one needs to reassess the initial statement. That reassessment requires an agreement whether housing prices (a) are too high or (b) are fair. Redfin (2020) specializes in listing housing quotes and supports exporting query results into Comma Separated Value (CSV) files. An analysis of 351 homes in the Boston area shows that the median listing time is five days, and the 95th percentile is sixteen days. If houses are selling within one to two weeks, that suggests that prices are *fair* and a different knob needs to turn (e.g., wages or commutability). Reviewing descriptive statistics shows the housing quartiles as (600k, 900k, 1.3M, 7.3M) and highlights the diversity price ranges available. These ranges naturally form in alignment with the demographic distribution of the city. For instance, in the financial district (02110), housing is at the 75th percentile with median income at 152,000$. Roxbury (02119) has a median income of 27,000$ and housing in the 25th percentile. These vast community differences provide additional evidence that the government needs neighborhood-specific policies. These policies need to classify each community based on a series of economic attainability features, instead of focusing solely on race. That would represent a stark pivot from the current administration’s approach that treats the entire city as a single cluster. Using distinct clusters with specific classifications makes logical sense, as a median person in Roxbury cannot afford the median mortgage, irrespective of their physical attributes.

## Improving Wages

If housing prices are fair but unattainable, then another lever such as wages needs to be explored. The median income in Boston is 72,000/year in part due to a booming professional and business service industry, and unemployment is near 2.0% (BLS, 2020). However, a more in-depth analysis shows that that the top 1/3rd of zip codes account for 56% of the wealth. Also, by ignoring the top 3 districts, the city’s median income reduces to 39,000$. This price point makes attainable housing for the actual middle class at roughly 1137.50$/month. For members of the lower-class, such as food services and health care support, they earn between 12.50 to 16.99$ per hour (13 to 18k per year) (DataUSA, 2020). This group would need median rent of around 400$/month, assuming a 35% wage to rent ratio. Even with the cities commitment to building 69,000 additional low-income housing by 2030 (Boston.gov, 2019), the does not meet the current needs of the bottom eight neighborhoods. In both scenarios, the manageable housing price is below the fair market price and requires either a wealth redistribution or population migration. A wealth redistribution could be either implicit (e.g., higher minimum wage) or explicit (e.g., taxation). Alternatively, substitutes can improve the commuter experience and enable people to trade extended travel to work for lower housing prices. Evaluating the political or sociological impact of these trade-offs is outside the scope of this research.

## Reducing the Wealth Inequality

An ability to accumulate housing equity has considerable influence over the net wealth of a family. For most families, it is the most expensive purchase they will ever make and their only leveraged investment. Munoz et al. (2015) report that the lack of homeownership causes many non-white Bostonians to have a median net worth of less than 10,000$. While some of their numbers are misleading, such as segmenting specific groups arbitrarily, causing an equal emphasis on say Native Hawaiians (0.04%) versus the African Americans (25.26%). That is not to say one group is less important than another, but group populations of 247 versus 175k people will likely have different error rates, making direct comparisons challenging. Similar to other literature, Munoz et al. use a single partition per race for analysis. This limited approach might distort the reality compared with say, slicing the data by industry and age group. Any attempt to partition solely on race will miss that the distribution is 50% white, 25% black, and 25% other—with 50% of the wealth controlled by 7.5% of the population. Since this subset has the same influence as the other 92.5%, the make-up of that group has tremendous potential to skew derived wealth statistics.

## Reducing Crime

Boston claims to be the fourth safest major city in the country, with 2,852 severe crimes per 100k residents. Image Boston (2020) defines severe crimes to include homicide, rape, robbery, aggravated assault, burglary, larceny, vehicle theft, and arson. When these types of crime occur, it impacts the local community by discouraging tourism and consumer spending at local establishments. The Boston dashboard reports that the trend is decreasing; however, this is partially due to the accounting system. Analyzing the Crime Incident Reports (2020) provides a different perspective, such as the number of shootings and sexual assault cases increasing each year. There are also nearly 2,000 larceny reports each year, so while the percentages are decreasing due to population growth, the count is not. To the defense of Boston, 63% of all police reports are for property damage or theft. However, the distribution of crime is not uniform across the city, with Roxbury (B2) the highest followed by Dorchester (C11), Suffolk (D4), and then Downtown (A1). The crime wave also fluctuates based on the area and day of the week (see Figure 1). For instance, Tuesday is the safest period for women due to the extreme outlier for sex and alcohol-related reports. Downtown (A1) experiences 25% more complaints between the start and end of the week. Meanwhile, Roxbury (B2) and Monument Square (A15) decline on Saturday.

Figure 1: Reports Per Day of Week

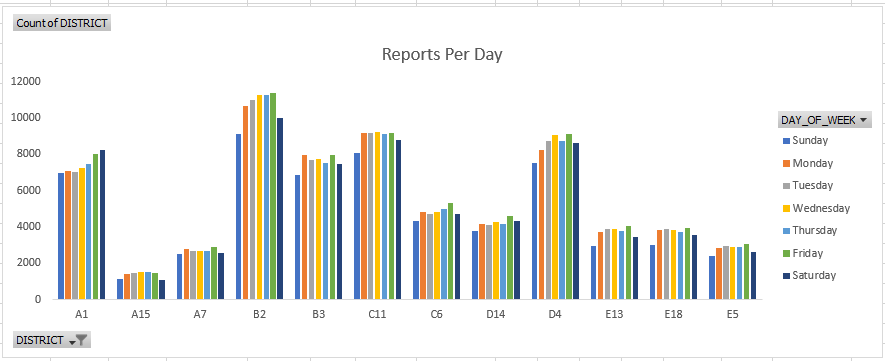


Figure 2: Reports Per Hour of Day

