PROJECT FINAL REPORT - *GENTRIFICATION IN BOSTON*
Group 3
November 8, 2022

Business Question: How can we prepare for future demand by using existing population data as predictors of neighborhood composition changes? Who will be Interested in our Project?

- Predicting changes in income levels is an important aspect of future demand planning.
- Businesses can predict anticipated changes in income level enabling them to adapt their supply chain to effectively meet changing consumer demand and spend. For example, building equipment lease companies may move stock to warehouses closer to the building sites.
- Businesses can plan contracts and leases in anticipation of the change in Buying Power because of gentrification. For example, entering into new lease agreements for commercial properties at competitive rates.
- This research is applicable to the retail industry, investors and developers who can calibrate their supply chain or begin expansion into a neighborhood in accordance with anticipated changes in the income levels of an area arising from gentrification.

Data: Describe the data sets you used and how they were used to answer your business question(s). You must use at least 2 data sources and at least 2 data sets. Where did you collect the data from? What are the sources of the data? Describe the data collection process and the variables in the data sets. Provide URLs if the data was collected from websites. Provide samples of the data if it is from a private source.

- Our dataset is composed of 134 columns and contains data on hospitals, train stations, population demographic information and niche services such as Starbucks. We have chosen our unit of analysis to be Neighborhoods as it was the common variable between the datasets from where our data was collected and allows us to cover the entirety of the Boston area including the 19 different neighborhoods we have chosen to use.
- By drawing on data from these different aspects of the population we are able to determine the current possible displacement of a population in a neighborhood by a wealthier incoming group. This results in gentrification which leads to more retail locations, restaurants, MBTA lines and more opportunity in the area overall. With the data collected, businesses can closely track gentrification per Boston neighborhood before it occurs, therefore presenting them with the opportunity to plan ahead for an increase in buying power, demand and opportunity.
- Our Unit of Analysis at Neighborhood level is considered Unique.
- We have directly downloaded data from Boston.gov, US Census, Wikipedia as CSV files and read it into Excel or Python.
- We have scraped data from some websites using BeautifulSoup as well as manually searching through Google for determining Starbucks locations around Boston.

Datasets:

Dataset	Description	URL
Train Stations	This dataset was scraped from Wikipedia and gives us the location, Latitude and longitude.	Wikipedia

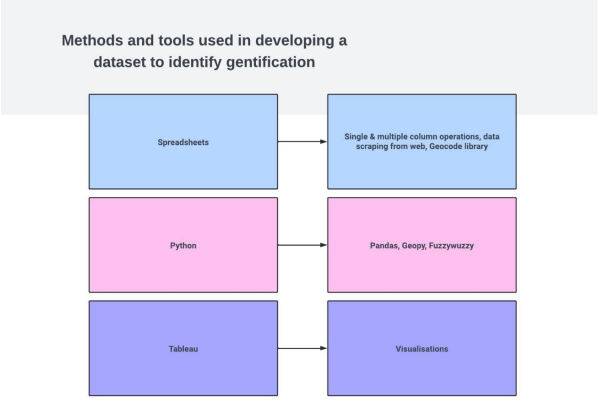
Average price per night of Airbnb listings in selected U.S. cities 2022	This dataset gives us an idea on Rental Prices of houses through Airbnb portal	Manually searched in Airbnb and collected the Average prices in Neighborhood.
Starbucks Data	This gives us an idea of all Starbucks outlets in Boston	Manually searched in Google for the locations of Starbucks in Boston
Boston Parks Data	Dataset on list of Parks in Boston divided by Neighborhoods	https://data.boston.gov/dataset/boston-park-assets
Hospitals Data	Locations of Hospitals in Boston	https://data.boston.gov/dataset/hospital-locations
Restaurant Permits	The number of Permits given to Restaurants segregated by different neighborhoods	https://data.boston.gov/dataset/active-food-establishment-licenses
PROPERTY ASSESSMENT 2017-2022	Gives property assessment values over a period of 2017 till 2022 to know details on old properties, New properties and remodeled houses.	https://data.boston.gov/dataset/property-assessment
Age, Household Type, Race, Nativity, Geographic Mobility, Education, School Enrollment, Means of commuting, Travel time to work, Place of work, Per capita Income, Occupation, Industries, Labour Force, Income, Bedrooms, Vehicles, Poverty Rates	Information on Racial Demographics of population in Boston segregated by Neighborhoods	Source: U.S. Census Bureau, 2015-2019 American Community Survey, BPDA Research Division Analysis
APPROVED BUILDING PERMITS	This dataset includes information about building permits issued by the City of Boston from 2015 to the present. Permits that are being processed or have been denied, deleted, void or revoked are not included in the dataset.	https://data.boston.gov/dataset/approved-building-permits

Information Quality

- Data Formatting – We formatted different variables across the datasets and standardized the data.
    - For Example: Neighborhood names are standardized
- Multiple Datasets Integration – Merging Datasets using a common Variable
    - We used Neighborhoods as Unit of Analysis to solve the integration issue.

- Different Neighborhoods categorization: Each dataset has categorized Neighborhoods differently. Therefore we had to merge some of the neighborhoods to standardize in all the datasets.
    - For Example: We merged Fenway & Kenmore to Fenway-Kenmore, Allston & Brighton to Allston/ Brighton and Back Bay & Beacon Hill to Back Bay/Beacon Hill in order to have a common variable between all datasets.

Methods and Tools



- Spreadsheets are used for common single and multiple column operations such as Concatenating, Standardizing, Merging etc. We have scraped some data from the Web. We used the Geocode library.
- We used Python for some Data Wrangling in some datasets. We used the libraries Pandas mostly. We used Geopy to convert Latitudes and longitudes to Neighborhoods and used Fuzzywuzzy to give us the best dataset as not all datasets were readily informative.
- We used Tableau software for Data Visualization as it is interactive and gives lots of insights effectively.

<u>Challenges</u>

- Missing Data on some neighborhoods.
    - As the data has been taken from Official sources such as Boston.gov and US Census, we cannot fill the Missing data with our individual perception/analysis. Due to the less row count as neighborhoods are less in Boston, We cannot delete the rows with missing values.
    - No dataset was found on Starbucks, Airbnb locations around Boston and search was conducted manually using each of the 19 selected neighborhoods as filters.

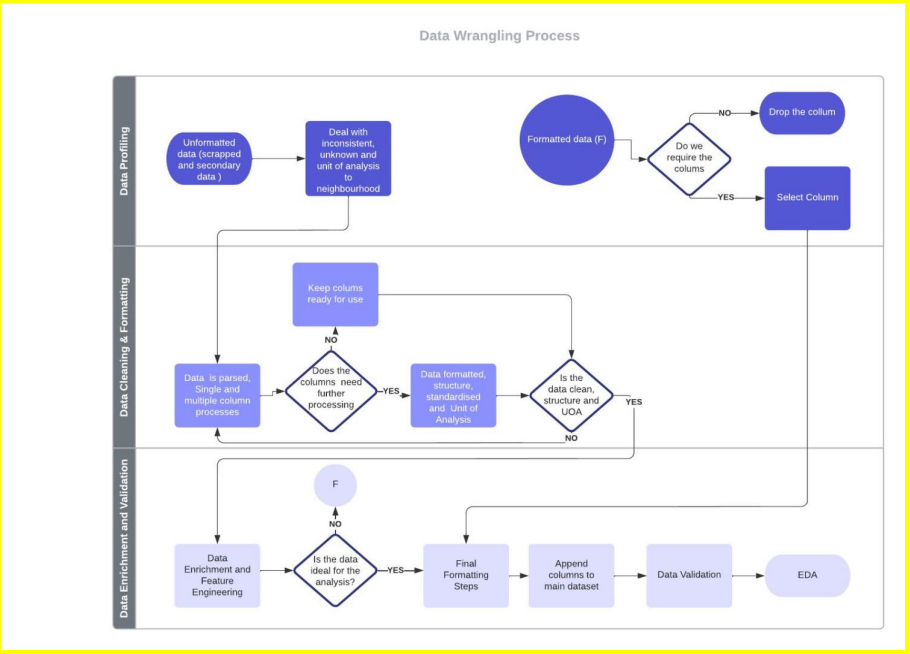
- Non standardization of Neighborhoods. Different sources classified Neighborhoods differently
    - We had to merge some neighborhoods as they were listed differently on different data sheets, for example Fenway & Kenmore was listed as Fenway/Kenmore in some so the areas were standardized, merged and displayed as Fenway/Kenmore in the “master dataset”.

Data Wrangling Process

Construct	Dataset	Process	Rationale
Income	BOSTON NEIGHBORHOOD DEMOGRAPHICS, 2015-2019	Median income Per Capita income Neighborhoods combined to merge with broader neighborhood structure	Lower income areas are expected to be more susceptible to gentrification as demand for properties increase.
Education Levels	BOSTON NEIGHBORHOOD DEMOGRAPHICS, 2015-2019	% of masters degree of better to determine level of academic	Areas with lower education levels indicate susceptibility to gentrification.
Occupation	BOSTON NEIGHBORHOOD DEMOGRAPHICS 2015-2019	% of Finance and insurance, and real estate and rental and leasing occupation to indicate high earning individuals. % Management, business, science, and arts	Concentration of high income
Housing Prices/Values	Property Assessment 2017- current	Concatenated datasets 2017-2022 Created subsets 1. Remodeled homes 2. new construction. Compared value of properties	Property assessed value increased exponentially after remodeling. An increase in remodeled homes may indicate a change in persons occupying property
Niche Services - Starbucks	Manually Obtained Locations through Google Search	Google search was conducted for each of the 19 neighborhoods. Count of Starbucks per neighborhood was generated using the 'count' Excel function.	Starbucks stores locations are located in affluent neighborhoods and those with a majority white consumers as those represent their largest customer segment.
Niche Services - AirBnB	Manually Obtained Locations through Airbnb Search	Airbnb search was conducted for each of the 19 neighborhoods. We took an average of 5 airbnb units for a single day for a family of 4 in each neighborhood.	Does Not provide much insight on Gentrification
Transportation	MBTA Rail Ridership by Time Period, Season, Route/Line, and Stop	This dataset was cleaned and processed using Python. Several unnecessary columns were removed. Using Geopy, the stop name was matched to a Zip code which was then allocated to the corresponding neighborhood. Each available/nearby MBTA line was associated with a specific neighborhood. We also calculated Distance to nearest Train station	By determining what public transportation routes are located in each specific neighborhood we can make inferences on the population of the area and also determine their possible commute time and options.

		but decided not to put it in our final dataset.	
Demographics:Age, Household Type, Race, Nativity, Geographic Mobility, Education, School Enrollment, Means of commuting, Travel time to work, Place of work, Per capita Income, Occupation, Industries, Labour Force, Income, Bedrooms, Vehicles, Poverty Rates	Source: U.S. Census Bureau, 2015-2019 American Community Survey, BPDA Research Division Analysis	The Data wrangling has been done in Excel and Vlookup is then used to associate them with Neighborhoods we finalized. Neighborhoods were divided separately so we had to merge some neighborhoods and standardize the names.	This source has given us a lot of insights on more than 15 factors such as Education, Income, Employment, Poverty, etc. We can enrich this data more by finding out percentages with total neighborhoods population instead of just counting to understand the data better.

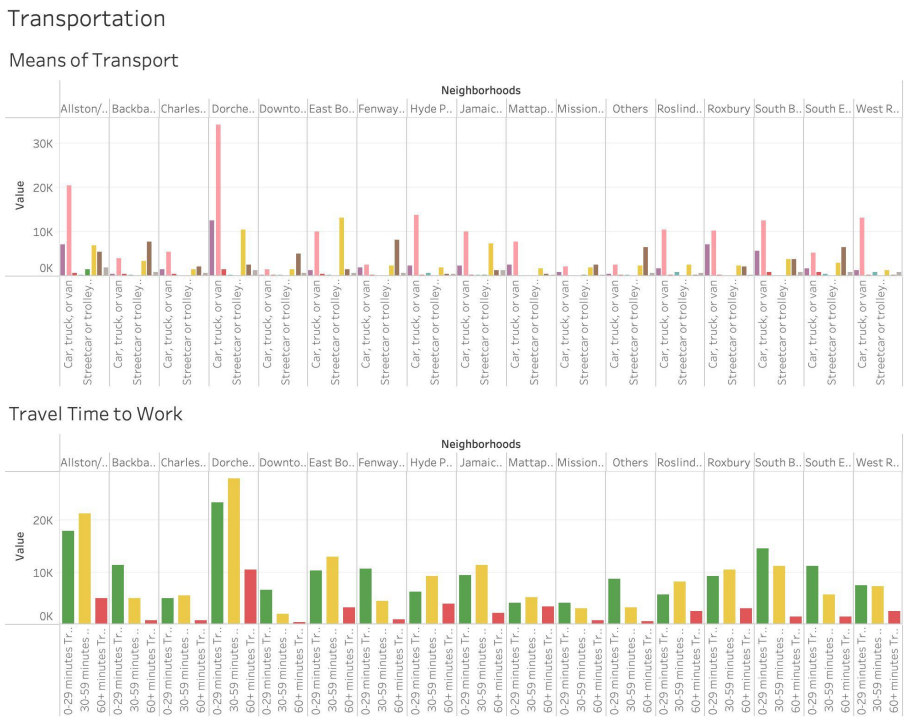
Process Diagram



We first started with Profiling the data. The data was secondary and we scraped some from the internet. Then, we removed some columns which were redundant to our analysis. After that we have then cleaned our data by standardizing the neighborhoods, then converted all the data to represent neighborhoods. We then used some Data Enrichment when we calculated Distance, % of Masters Degree per Neighborhood Population etc. After all the Data wrangling, we then merged the datasets into a single Master Dataset with 19 neighborhoods and 134 Columns each being Core to our Analysis.

Analysis and Results

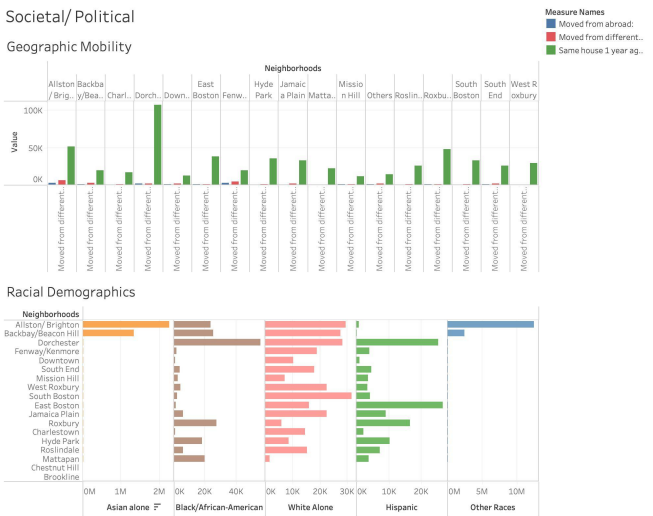
Transportation Dashboard



The transportation dashboard consists of Travel Time to work and Means of Transport.

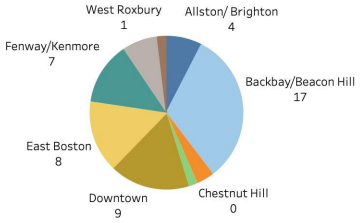
In this visualization we can notice a pattern in which less affluent neighborhoods tend to have longer travel times to and from work than those in affluent neighborhoods. Individuals who live in Dorchester, on average, commute more than 30 minutes to and from work. This can be due to the majority of business being located in other areas such as Downtown or the lack of public transportation accessibility nearby.

Societal/ Political Dashboard



The Societal/ Political Dashboard consists of Geographic Mobility and Racial Demographics. Less affluent Neighborhoods tend to have a high population of minority groups such as African American, Hispanic or Asian. For Example in Dorchester, considered a not so affluent neighborhood, a large minority population is observed mostly composed of Hispanics and African Americans. In comparison, other neighborhoods like Fenway or South Boston tend to be white in majority.

Displacement Effect Dashboard



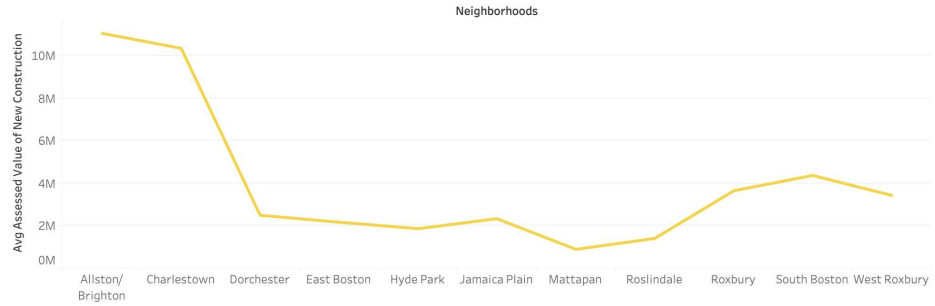
As seen in figure above, we can observe that a large number of restaurants plan to open in the Dorchester, East Boston and Roxbury neighborhoods. This demonstrates a rise in demand for those neighborhoods in the future and thus providing a clear indicator that gentrification may occur and those neighborhoods may become more affluent in the future.

Also as Starbucks being a Niche service, it only establishes its services in highly affluent neighborhoods like Back bay and Downtown. In the pie chart above, the most affluent neighborhood, Back Bay/Beacon Hill has 16 more locations than West Roxbury which can be categorized as a not so affluent neighborhood. It is also interesting to note that Chestnut Hill has no Starbucks locations.

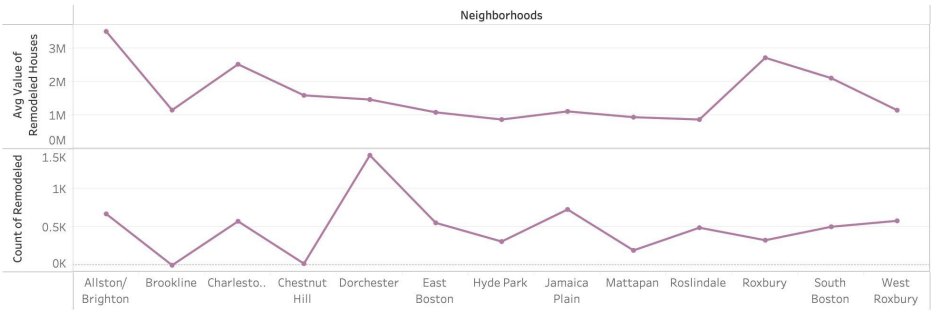
Socio-Economic Dashboard

Socio-Economic

Average Value of New Properties



Average Value of Remodeled Units

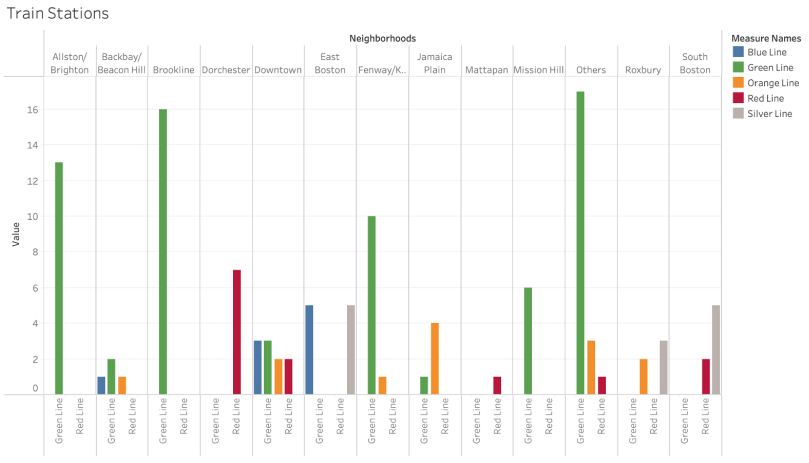


The Socio-Economic dashboard has 2 line graphs which gives insights on average value of renovations or remodeled units and average value of new properties. The value of new properties suggest that the less affluent neighborhoods such as Dorchester, Mattapan etc,. property prices are low when compared to the highly affluent neighborhoods such as Allston/ Brighton.

The average value of remodeled units suggests that the less affluent neighborhoods such as Dorchester, Jamaica Plain, Roxbury have higher counts of remodeling happening. This could be to places such as new restaurants, workplaces etc.

Nearby Train Stations

Train Stations

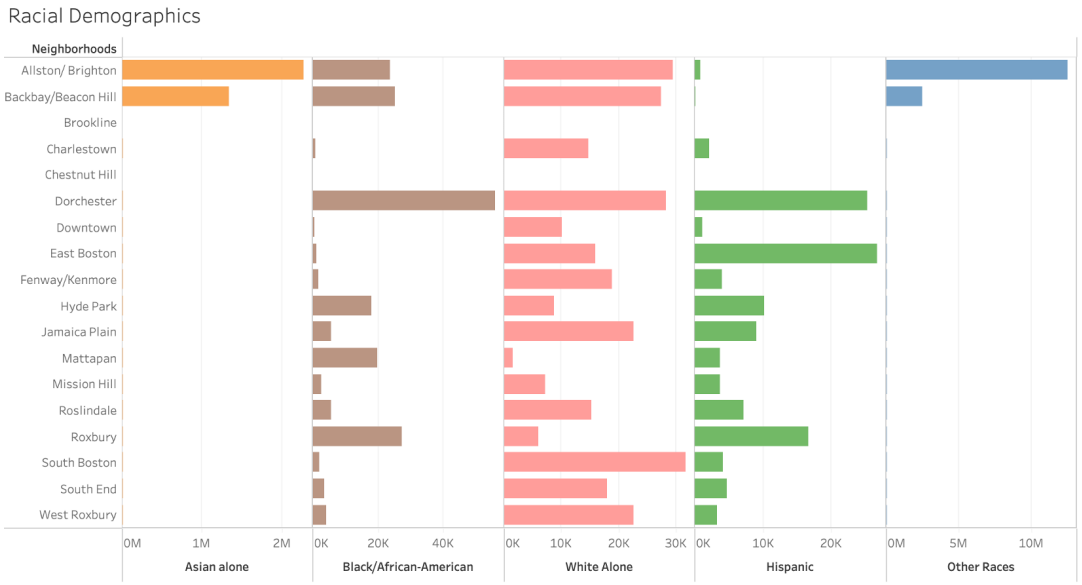


Blue Line, Green Line, Orange Line, Red Line and Silver Line for each Neighborhoods. Color shows details about Blue Line, Green Line, Orange Line, Red Line and Silver Line. The view is filtered on Neighborhoods, which keeps 13 of 19 members.

We can observe an inverse relationship between the affluence of the neighborhood and the number of MBTA lines available nearby. As the neighborhood becomes less affluent, for

example, Dorchester or Roxbury, there are less MBTA lines nearby when compared to a wealthier neighborhood such as Downtown or Back Bay/Beacon Hill.

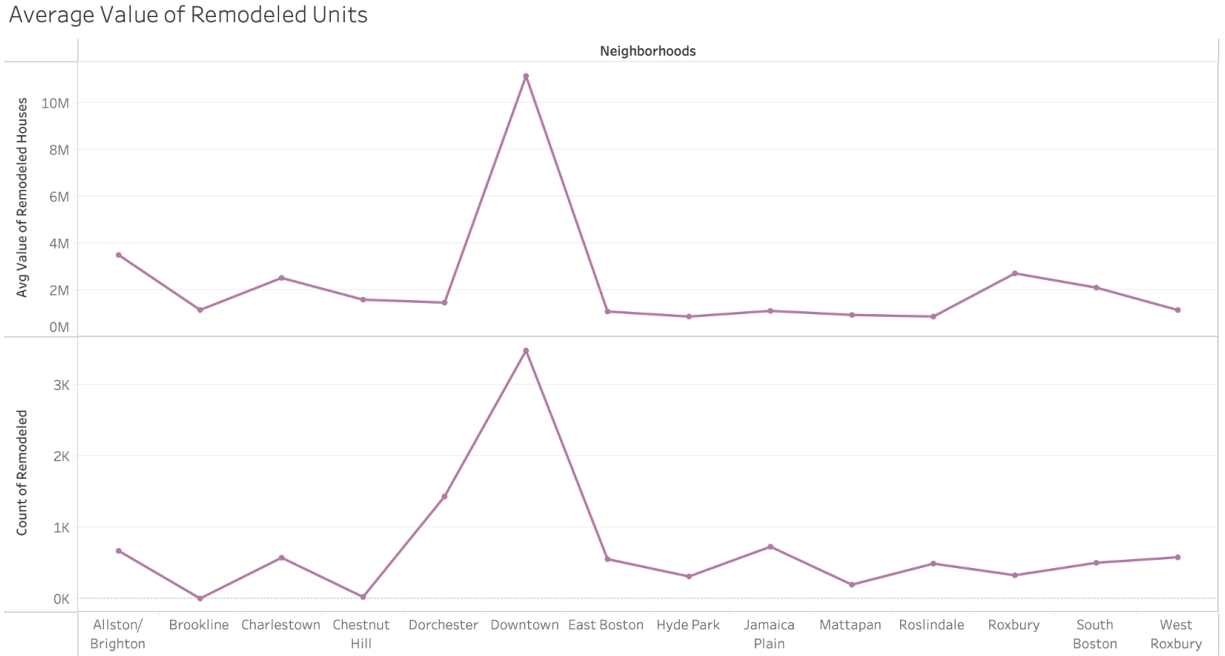
Racial Demographics



Sum of Asian alone, sum of Black/African-American, sum of White Alone, sum of Hispanic and sum of Other Races for each Neighborhoods. The view is filtered on Neighborhoods, which excludes Others.

We can see that less affluent neighborhoods such as Dorchester, East Boston, Roxbury have higher numbers of African Americans and Hispanic who often have lower income.

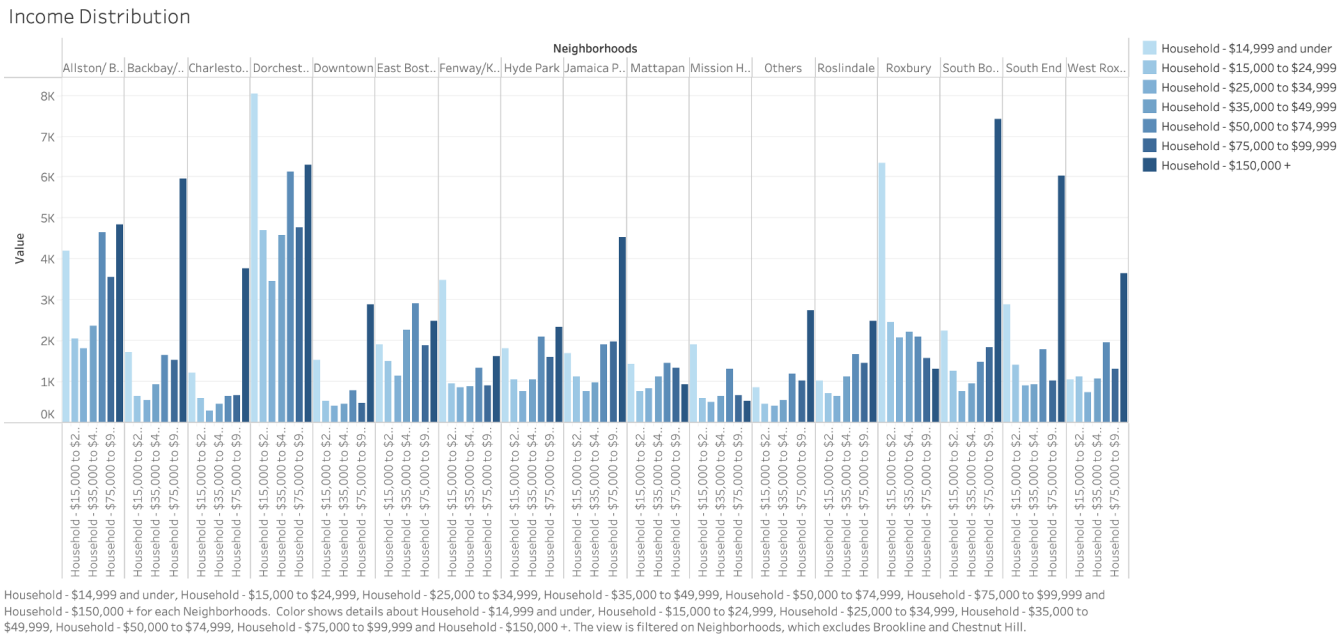
Renovations



The trends of sum of Avg Value of Remodeled Houses and sum of Count of Remodeled for Neighborhoods. The view is filtered on sum of Avg Value of Remodeled Houses, which keeps non-Null values only.

If we remove Downtown from the charts, we can see that the number of renovations are significantly higher in lower affluent areas such as Dorchester, East boston, Jamaica Plain, Roxbury which suggests that those areas can be gentrified in future.

Income



Household - $14,999 and under, Household - $15,000 to $24,999, Household - $25,000 to $34,999, Household - $35,000 to $49,999, Household - $50,000 to $74,999, Household - $75,000 to $99,999 and Household - $150,000 + for each Neighborhoods. Color shows details about Household - $14,999 and under, Household - $15,000 to $24,999, Household - $25,000 to $34,999, Household - $35,000 to $49,999, Household - $50,000 to $74,999, Household - $75,000 to $99,999 and Household - $150,000 +. The view is filtered on Neighborhoods, which excludes Brookline and Chestnut Hill.

We can clearly see that Dorchester, Roxbury have more population whose Annual Income falls under $15,000.

Overall, a total of 14 visualizations have been made on Tableau and have been uploaded into the Powerpoint presentation especially focusing on Visualization

Conclusion

The entirety of the data collected proved to be useful in answering our business question. By merging all of our data, we have generated a “master dataset” that contains information on different important aspects of the Boston population split into neighborhoods. By including data on racial demographics, education levels, proximity to public transportation, counts of niche services, etc. we have identified Boston neighborhoods that have or will possibly become gentrified in the future. In terms of business, our visualizations along with the “master dataset” can allow business, developers and investors to closely track the developments and changes in composition in each neighborhood to anticipate competition and adjust their supply chains, leases and contracts to match the new demand and buying power.

External Materials

Article	**Findings**

Musil et al., 2022	· Research has shown the identification of gentrification based on socio-economic and house price data · Freeman (2005) – city center location, income below median, old housing stock, increase in higher mean educational level, sharp increase in residential prices · Holm & Schulz (2018) household panel data used to estimate the degree of displacement. · Important to measure displacement – Causal link between upgrade and displacement although it is not mandatory – phenomenon is not easily measured. · Suggested – transformation of building stock and shift between housing segments o Privatization of housing stock o Causal link between housing market transformation and exclusionary displacement. o Does the decline of a cheap rental market may be a plausible reason for displacement
Jover & Díaz-Parra, 2020	Introduced touristification – Short term rental market Whereas gentrification refers to a lower-income population being replaced by a higher-income population, touristification refers to an increase in tourist activity, which generally implies a loss of residents.
Curran, 2018	Qualitative research methodology focused on displacement on Chicago neighborhood identified democratic vision as a factor While democratic vision cannot completely forestall gentrification, they can help to draw attention to substantive issues like zoning, affordability, transparency and democratic process that serves to educate long-term residents and gentrifiers alike.
Steinmetz-Wood et al., 2017	Gentrification was positively associated with collective efficacy. Connects to democratic vision Gentrifiers (individuals moving into gentrifying neighborhoods) had higher collective efficacy than individuals that lived in a neighborhood that did not gentrify
(Grube-Cavers & Patterson, 2015; Jones & Ley, 2016)	Link between rail transit and gentrification statistically significant and positive relationships between exposure to urban rail transit stations
(Horn & Merante, 2017; Wachsmuth & Weisler, 2018)	Brought in theme of Airbnb's and impact on rent prices – maybe extended to ownership. Rental Prices are likely to increase because of Airbnb Likely to impact the displacement of individuals

(Washington et al., 2021)	The pace of gentrification in Chicago from 2007 to 2009 was negatively associated with the concentration of blacks and Latinos in neighborhoods that either showed signs of gentrification or were adjacent and still disinvested in 1995. Racial composition has a threshold effect, however, attenuating gentrification when the share of blacks in a neighborhood is greater than 40 percent.

Additional data and analysis

- Generate More columns by Data Enrichment such as Percentages of whole instead of numbers to analyze data more efficiently
- We can run a Machine Learning model and find out which areas are likely to be gentrified in future. We can also identify those variables which can predict the neighborhood which affects the target variable
- Since we have more than 130 columns, we can shorten the dataset and remove unwanted columns which do not add to our analysis of identifying neighborhoods which are likely to be gentrified in future.
- We can create a Database for all the sources and retrieve data more organized.