

## **Project Deliverable 3 -- Census Team 2**

Emily Greven, James Heilberg, Yichuan Philip Ma, Qingyang Xu

### **1. Data Collection:**

At this point we have received all relevant data from our client. Since we received aggregate data first, and did not receive household/ other alternative data until last week, we had to focus our analysis on the aggregate data. The data can be found in excel file format on the Google Drive, and the code to break down into smaller subsets of related features can be found on our team github branch ([BU-Spark/ds-boston-bpda-br-census at Team2 \(github.com\)](https://github.com/BU-Spark/ds-boston-bpda-br-census-at-Team2)). We would like to include some analysis on the household/PUMS data, however, as Hugh suggested, it is important that we have a meeting first to discuss complex nuances and meaning of this data before proceeding.

### **2, 3, 4 Refined Preliminary Analysis and Key Questions:**

For this project, our scope was very broad with an overarching question/idea of painting a picture of the Brazilian community in Massachusetts. Our only constraint was that we focus more on making sense of a variety of features and their relations to one another rather than simple one dimensional time series analysis. As such our work flow for completing our clients requests has been to understand the data through metrics and visualizations. Then we worked to find correlations/ relationships between variables that are non-trivial and help to paint a picture. Finally, we used regression ML techniques taught in class to predict how things may change in the future and to find interesting trends.

Specific key questions answered in depth and further analysis can be found in the attached report draft. New visualizations take advantage of ML techniques seen in class and include enhanced predictive analysis.

Answer to overarching project question of showing the story of the Brazilian population can be found in Draft 1 of final report under “**Results**”

### **5. Create a draft of your final report**

See branch “Team 2” -> folder “Final Reports” -> “Draft 1-Team 2”

### **6. Project Scope and Limitations:**

The project scope will be the same as at the time of deliverable 2. We were not able to incorporate the household/PUMS alternative data that was sent by the client yet. That is because this dataset is much more microscopic and nuanced. This will require us to meet with the client about this data specifically which is planned for after the break. At that time we will make the decision with our PM and the client of whether or not we can incorporate it into the scope.