

The Minority Action Project

Steve Kim, Arun Singh, Divya Veerapaneni
Department of Computer Science, Boston University, Boston, MA

The Minority Action Project is a research project designed to identify the resources available for minority-led startups in innovative cities in the United States. The focus of this project is to study specifically women and African-Americans. The objective is to determine what factors contribute to the success of minority-led tech companies in innovative cities. The factors taken into account for this study include gender of entrepreneur, income of entrepreneur, and ethnicity of entrepreneur. To gain a better understanding of how resources are distributed in a single city, we looked specifically at Boston as an example of a typical innovative city in the United States. The specific objectives of this project are the following:

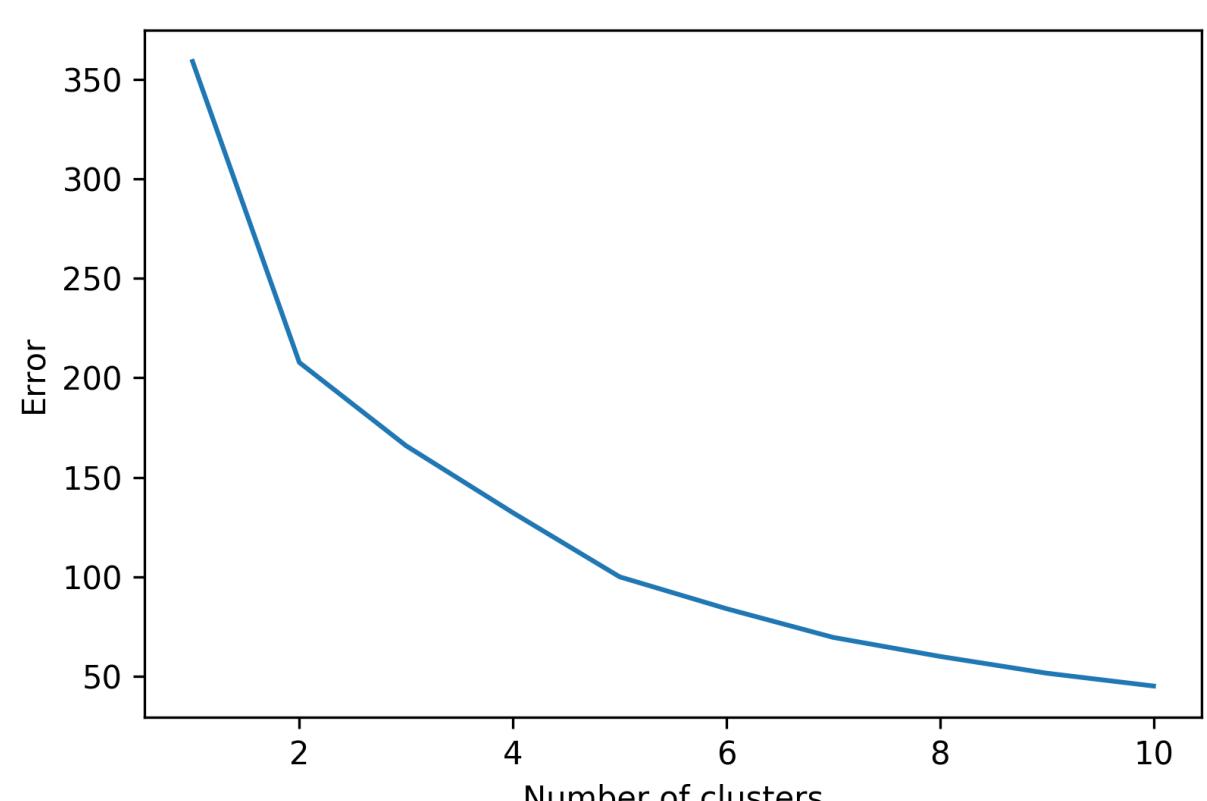
- Determine the distribution of businesses and resources in Boston to gain insight into the distribution of businesses and resources in a typical innovative American city
- Obtain statistics on various demographics such as ethnicity, income, gender, sex, occupation, and education level of entrepreneurs throughout the United States.
- Measure the change in entrepreneurial growth throughout the years in the U.S. Does growth or decline match minority entrepreneurial growth/decline?
- Identify a metric for entrepreneurial success given the datasets provided. What are the variables that contribute to success in a business? Does ethnic/gender data correlate with these definitions?

Methods

To obtain statistics regarding the demographics of American entrepreneurs, we used microdata from the Kauffman Index of Entrepreneurship in America (KIEA), which obtained data on 650,000 Americans through surveys. The Kauffman Index provided data from each year starting from 1996 to 2015. To obtain statistics related to property in Boston, we used a property assessment document from data.boston.gov.

Objective 1

Using property value data from data.boston.gov, we were able to make a plot of businesses in Boston. We used k-means clustering to gain insight into property values throughout Boston. We determined our k by plotting accuracy vs the k value, and chose the k value with the highest accuracy.



Objective 2

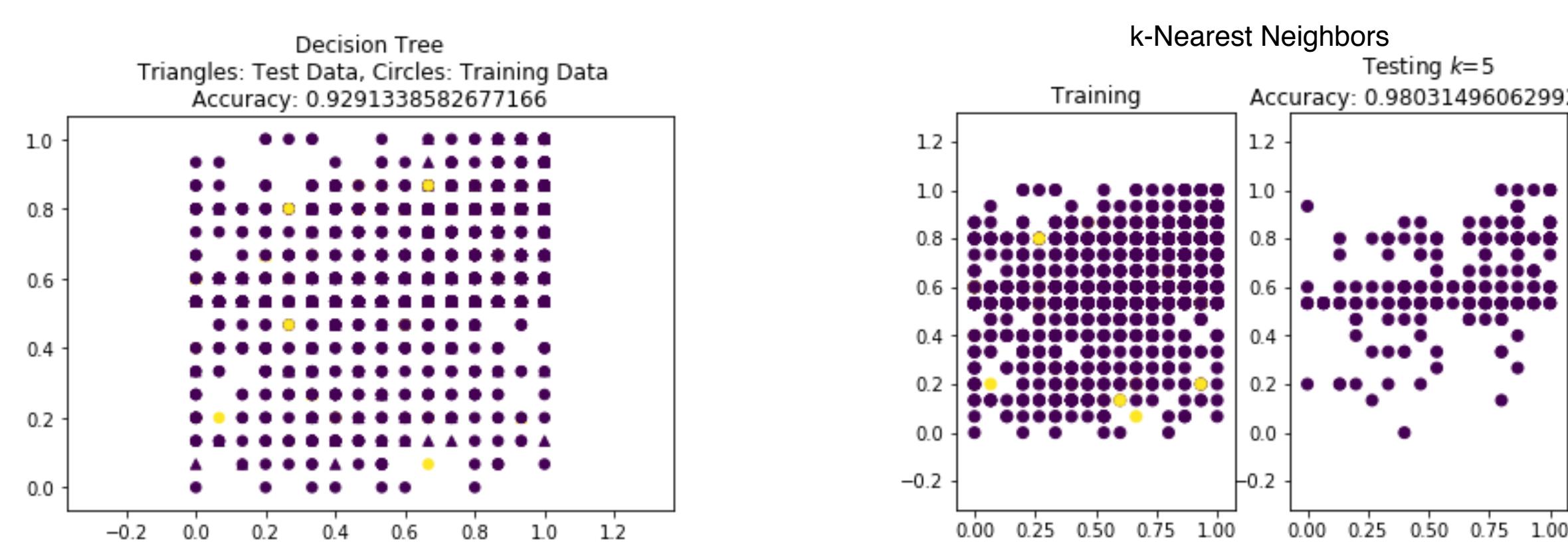
The KIEA dataset gives information on race, family income, education, and gender of a sample of U.S. adults. Using the data, we gave a breakdown of U.S. entrepreneurs based on these demographics for the year 2015.

Objective 3

In order to take objective 2 further, we used the KIEA dataset to track entrepreneurial activity throughout 1996-2015 on a national and minority level. Then, we plotted the change in demographics throughout time to gain insight into minority representation.

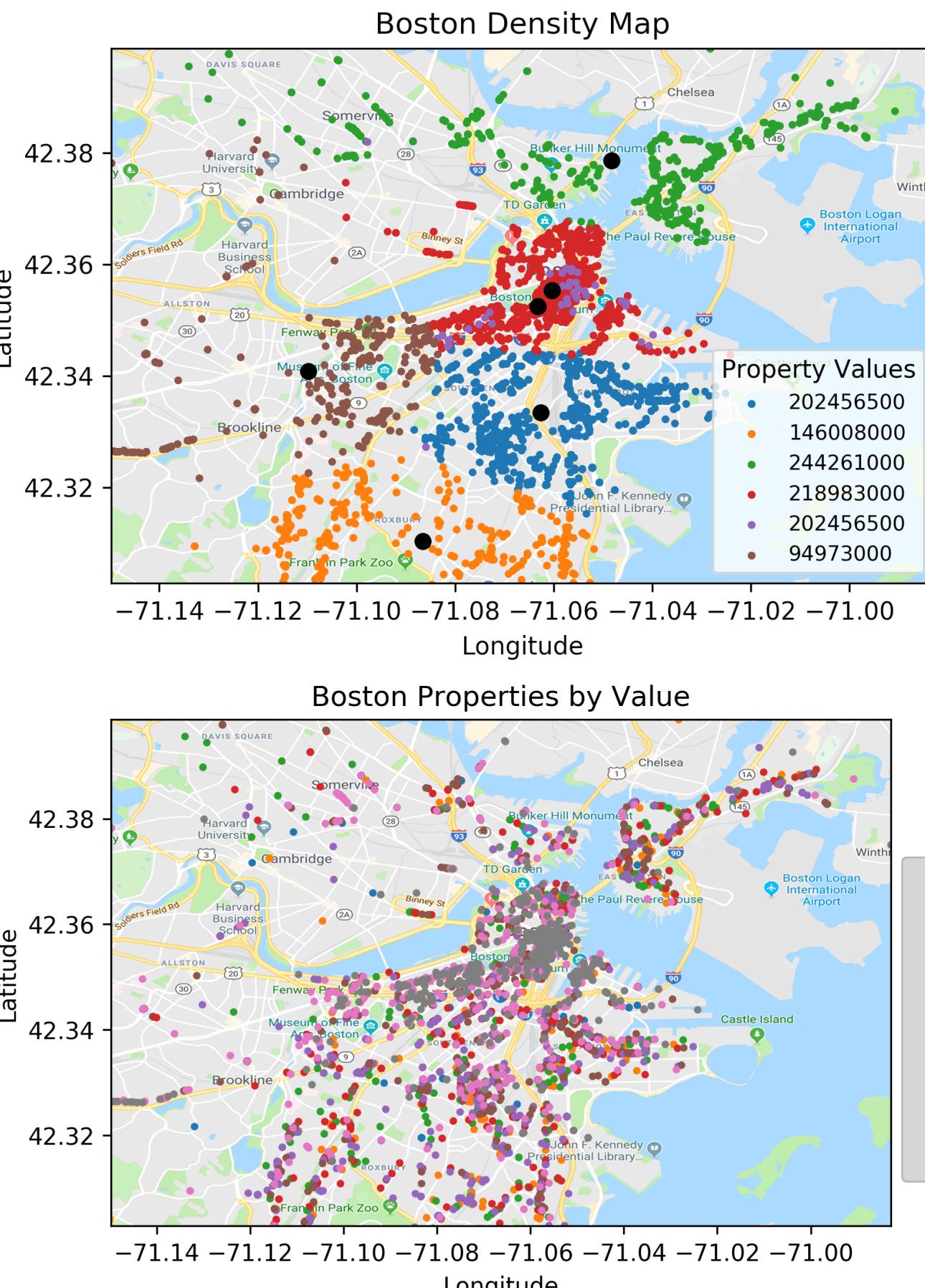
Objective 4

A successful entrepreneur can be defined as one whose business survives a year. We first plotted entrepreneurial success rates for each year. Then, we ran linear regression analyses on each field to determine the most contributing variable to success based on our definition. The most contributing variable was determined by the model with the highest magnitude of coefficient of determination (r^2). Finally, we created linear classification models using the k-Nearest Neighbors algorithm and the Decision Tree algorithm to test whether the given features combined could contribute to success. We determined our k by plotting accuracy vs the k value, and chose the k value with the highest accuracy.

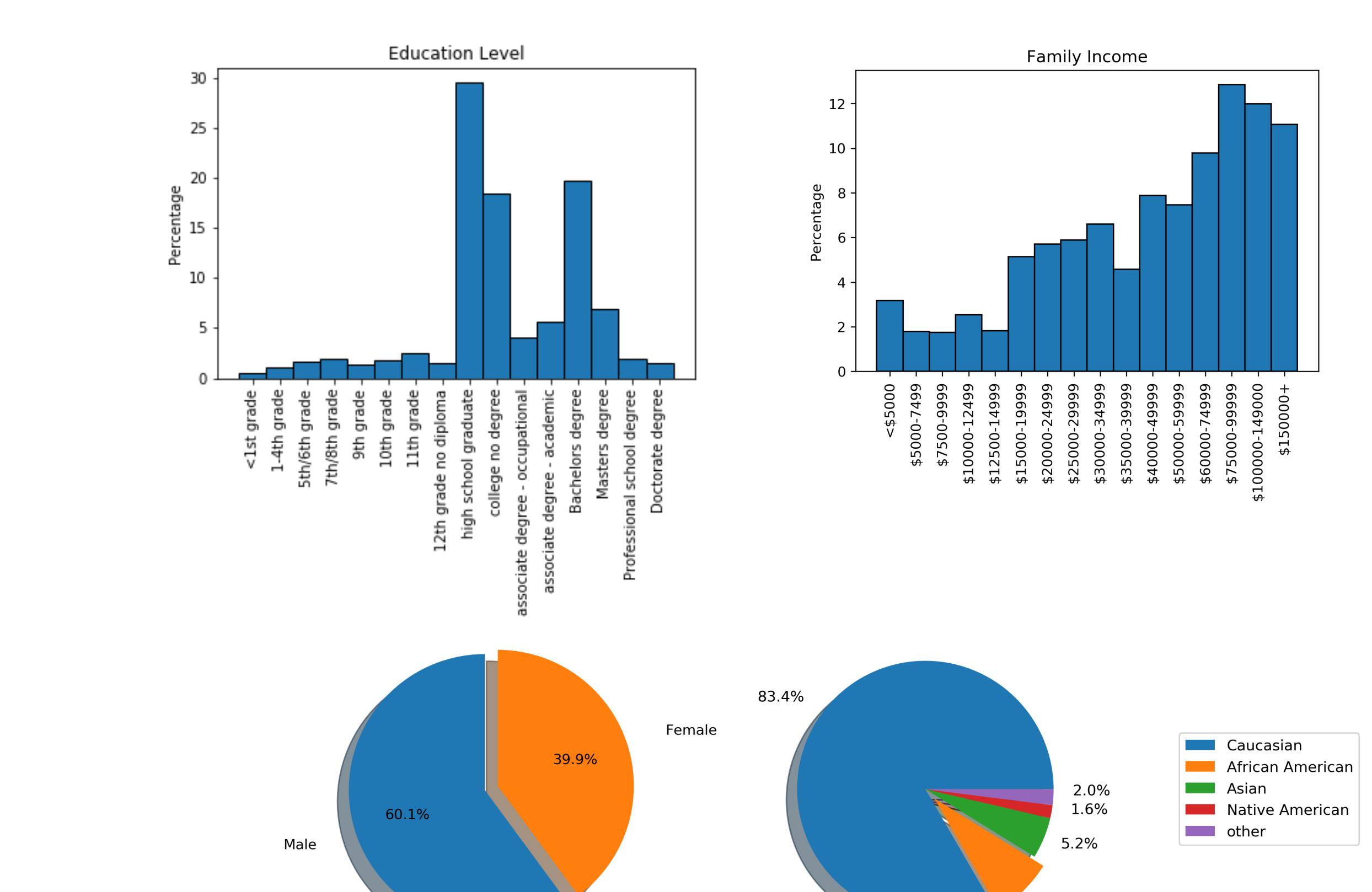


Results & Analysis

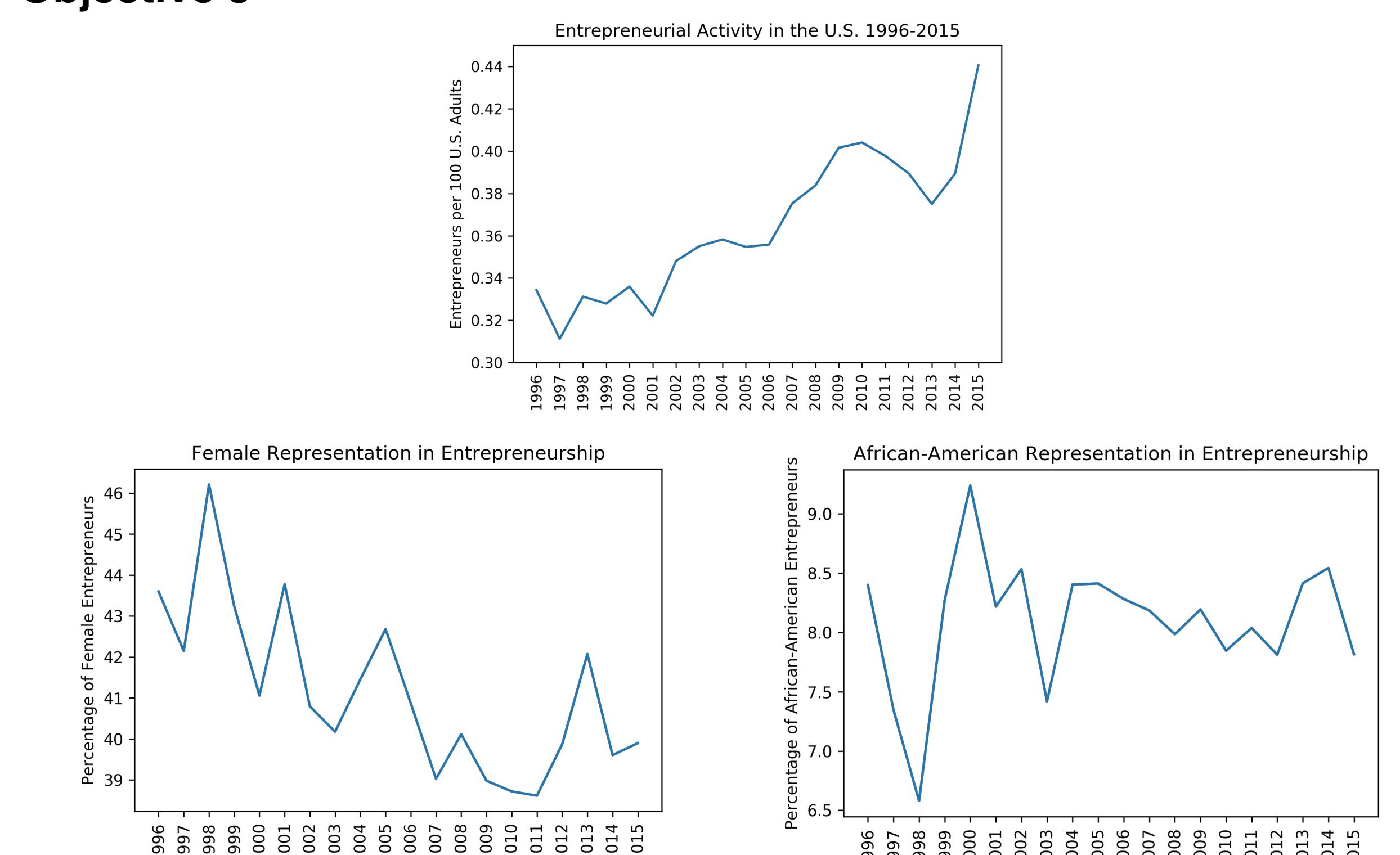
Objective 1



Objective 2



Objective 3



Objective 4



Factors	Coefficient of determination (r^2)
Family Income	0.0269023716
Graduation Level	0.03028840887
State (Connecticut)	0.01126436782
Region (Champaign-Urbana, IL)	0.01149425287

Conclusion

Objective 1

Using k-means as our clustering algorithm, we identified six centroids based on location and property values of businesses in Boston. Results reveal that the regions with the lowest average property are Brookline/Longwood (\$95,000) followed by Roxbury/Dorchester (\$145,000). Both Roxbury and Dorchester have high percentage of African Americans (about 50%) and Hispanics (20%) according to US Census Bureau. The wealthiest regions, in terms of property value, are East Boston, Boston Commons, and Chinatown. The most common property types were retail, parking lots, vacant lots, hotels, and offices.

Objective 2 and 3

Our analyses in objectives 2 and 3 show an overall increase in entrepreneurial activity in the United States. However, it also shows that representation of African Americans in the entrepreneurial community has remained stagnant and that the representation of females has declined. In addition, it shows that for the most recent year (2015), both African Americans and females are underrepresented in the entrepreneurial community when compared to the total U.S. population demographic. According to the U.S. Census Bureau, for the year 2015, 13.2% of the population was African American and 50.8% of the population was female. Using our results and that of the Census Bureau, we see an evident discrepancy in the representation of minorities in the entrepreneurial community. The results of objective 2 also show that the majority of entrepreneurs have a college education of some kind whether incomplete or complete. For 2015, the majority of entrepreneurs are those with a college education, but no degree, and those with a masters degree. Furthermore, we found that more entrepreneurs tend to come from wealthier families.

Objective 4

Based on our definition of success, the rate of success generally increased over time from the years 1996 to 2015. The rate of success was approximately 3%. A sharp decline was observed in the years 2003-2004. The study shows that family income and education level are the two most important factors that contribute to success of entrepreneurs. This is expected since one would expect that being a successful entrepreneur requires a certain amount of capital and professional expertise; however, because the r^2 values for both of these factors are less than 0.1, these factors do not have a significant correlation with success overall. In terms of minorities, both gender and race do not seem to have any direct correlation with success. The classification models we created that used the combined factors of race, gender, family income, region and state to predict success were successful, as the accuracy levels of the decision trees model and k-nearest Neighbors model on the test data were 0.93 and 0.98, respectively.

Acknowledgements

We would like to thank Professor Andrei Lapets and Bassel El Mabsout for their direction in our CS 506 class. In addition, we would like to thank Sarah Greisdorf and Emmanuel Amponash for their support.

References

- [1] Kauffman Index for Entrepreneurship in America Microdata
- [2] data.boston.gov
- [3] U.S. Census Bureau 2015