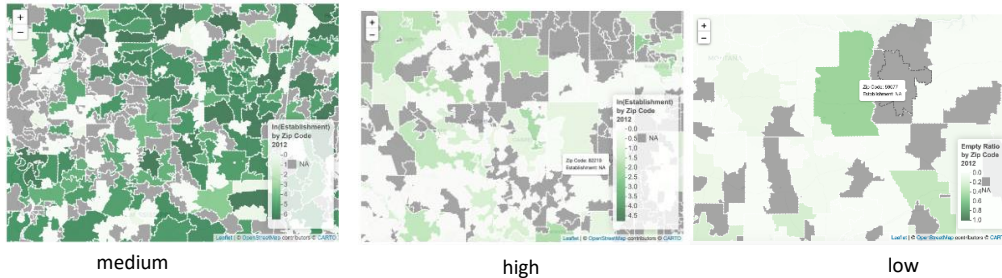


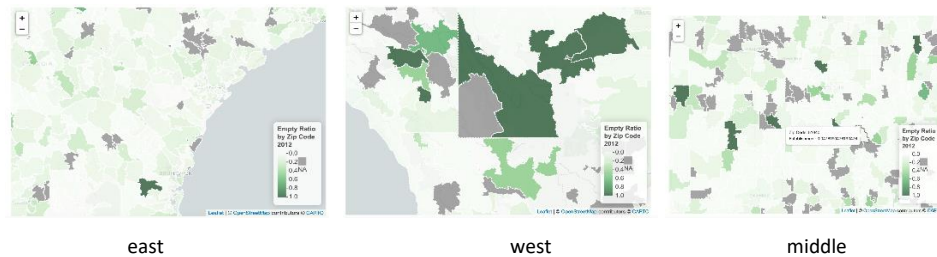
**Group#25 Illini Datathon2020**  
**Jasmine(Fangyun) Yi, Francis Fang, Eva(Yifan) Hu, Yuqing Xu**

Approaching to the data visualization model, we utilized maps, bar plot and tree plot to study the relationship between revenue and areas, revenue and industries and industries and areas. We hope our findings can offer help to the marketing strategy.

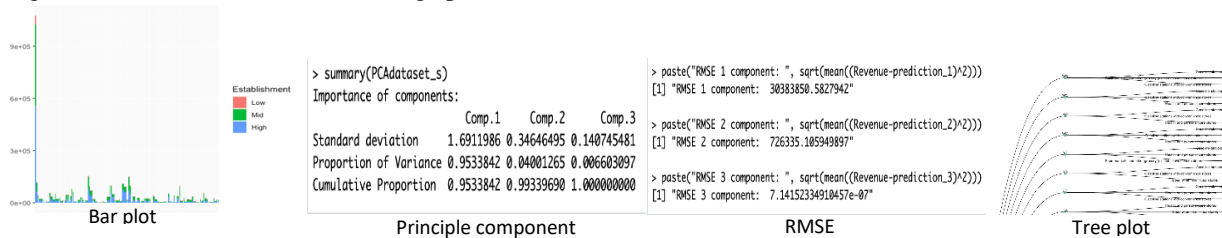
As for the revenue and areas, we divided all the establishments to three categories – low revenue establishments with a number of establishments less than \$100,000, medium revenue establishments with a number of establishments between \$100,000 and \$1,000,000 and high revenue establishments with a number of establishments over \$1,000,000. We turned numbers into logarithms to get a concise dataset. After that, we generated three dynamic maps to indicate the nation-wide distributions of these three categories of cities.



We also defined the establishment empty rate as the ratio between number of establishments that can't make profit for an entire year over the total number of establishments.



Across the country, it is easy to find that most cities contribute a medium to high revenue to our business. A small number of establishments offer less revenue than others and they come with smaller population. For establishment empty rate, we noticed that the rate for eastern coast is lower than western coast and central area. The reason can be that more populations brings more business transactions. We can clearly see the relationship between revenue and industry from the following bar plot. Food and beverage stores provide most revenue for our business, followed by Clothing and clothing accessories stores, Motor vehicle and parts dealers, Gasoline stations and Miscellaneous store retailers. Mostly, medium and high revenue establishments take a large portion of revenue.



We also formed a regression model between the annual revenue and different size of establishments. The results from principle components and RMSE between test and train data both show that medium size establishments are the main source of revenue. Based on previous analysis, we raised an interest in finding the relationship between areas and industries. From the tree plot, we displayed the top 3 industries that provide most revenue in each area.

Overall, we understand that our business mainly targeting on medium and large size of establishments and perform more trade in Claymont, New York, Boston and Atlanta. For the future business, we recommend that the marketing plan should maintain our strength in east coast and develop more small and medium size business in Denver, Modesto and Beaver. After that, we may also do research on our competitors and form a differentiate strategy to gain more revenues. We can investigate more in each city to figure out our customers' behaviors as well as the unique business pattern that dominates the area.