A Study of New Jersey Congressional District Five

Karel Hoppe

September 1, 2019

Introduction

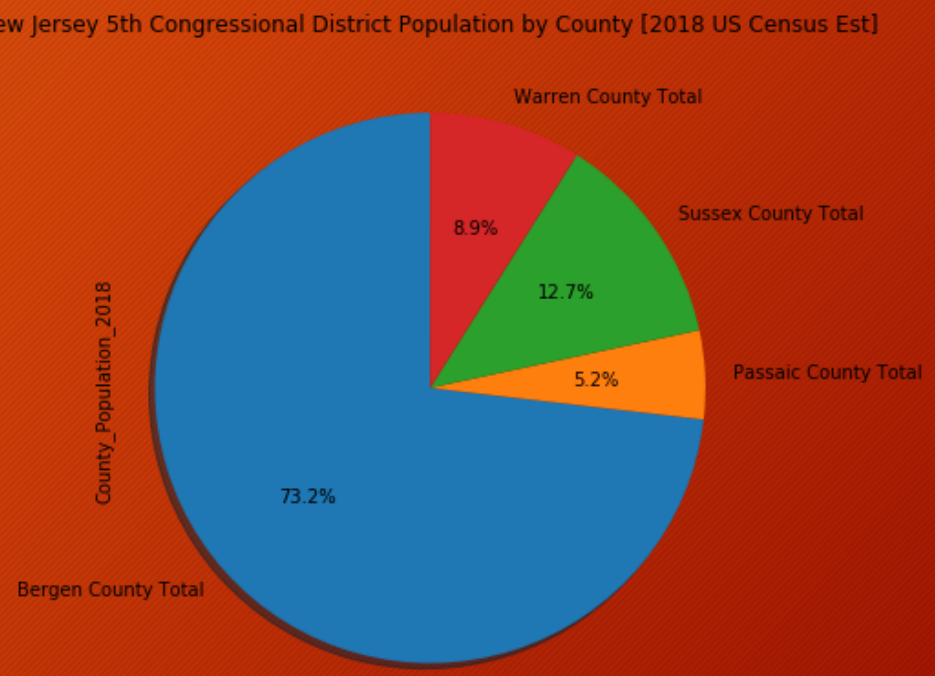
- There are twelve Congressional districts in New Jersey
- Can we use data science to help us understand one?
  - Yes we can
    - Use publicly aviliable data
    - Import the data into Pandas
    - Review the data
    - Consider user K Means Clustering

The Data

- Wikipedia
  - A great place to get the list of communities in NJ5
  - Used as a check against other data sources
- US Census
  - A great place for demographic data such as population size or ethnic groups
  - Not perfect as towns under 5,000 have less detailed information
- State of New Jersey Division of Elections
  - Good source of information on election results for NJ5
- Fourquare
  - Location data which tells us about a community
    - Restaurants, commuter hubs, parks, ect

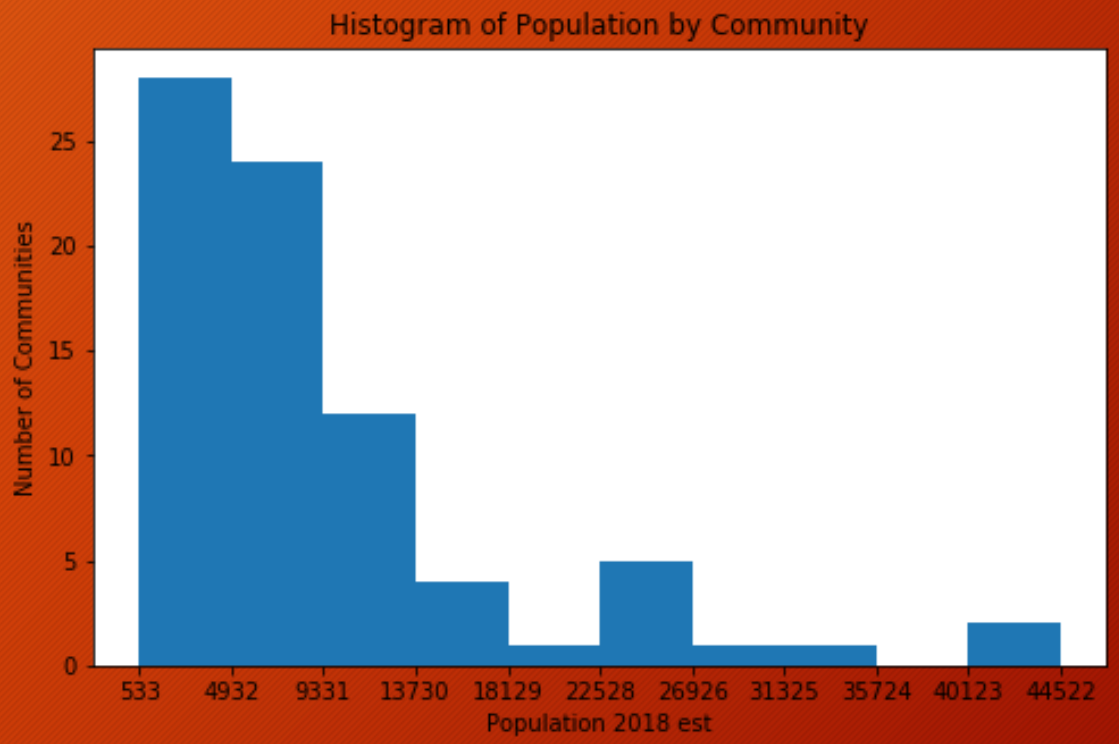
Insights

- Here we can see Bergen County is the biggest county in the District
- The data showed the Democrats are strongest in Bergen County



Further Insights

- Here we can see there are no big cities in the district
- Communities less than 5,000 in population are the largest in number



K Means

- K Means was used to cluster the communities based on US Census data
- Voter data was also used for K Means clustering
- Finally K Means was used to cluster based on Foursquare data
- For each case we decided on using four clusters

Conclusion

- The data provided a lot of clarity on NJ5. Using Data Science tools to graphicly display the data and Pandas to work with the data provided good insights
- At the same time K Means did not give us a lot just yet. It seems the US Census data had perhaps too many points so we could try to reduce them and run K Means many more times. We could also try something different such as regression analysis.
- For the Foursquare data K Means again was imperfect but insightful. For example it did appear the more affluent communities were grouped away from the less affluent.