Community Planning in Washington D.C

Outline

- Background
- Community Needs
- Data Sources
- Characterization of D.C.
- Analysis by Zip Code
- Venue Categorization
- Correlation Analysis
- Venue Clustering
- Conclusion

Background

- Washington D.C. has a racially and economically diverse population. At 68 square miles with a population of 705,749 people, D.C. is made up of more than two dozen different neighborhoods with various cultures and personalities.
- To determine whether the needs of the communities are being met by the businesses and services within those communities, one needs to look at the racial and economic distribution within each community.
- For instance, do communities with higher median household incomes have better resources?
- Is there racial bias when allocating resources?

Community Needs

- **Access to fresh, healthy, and affordable food:**
  - Access to grocery stores and farmer's markets increase the likelihood that members of the community will be able to maintain a balanced and nutritious diet. In contrast, areas with high concentrations of fast food restaurants, which typically serve food high in fat, sodium, and calories, are more likely to inhibit community members' ability to maintain that balanced diet.
- **Access to recreational or fitness activities:**
  - Nearby fitness facilities, parks, or other outdoor areas increase the likelihood that members of the community will be able to lead an active lifestyle
- **Access to educational resources:**
  - D.C. is known for having a plethora of museums with free admission as well as bookstores which provide additional educational opportunities to members of the community outside of traditional schooling.
- **Access to fast food options:**
  - If a community has more fast food options than grocery stores, members of the community are more likely to choose these options rather than healthy, nutritious meals.
- **Diversity of restaurant choices:**
  - The selection of restaurants within a community may impact people's decisions when it comes to selecting nutritious and delicious foods.

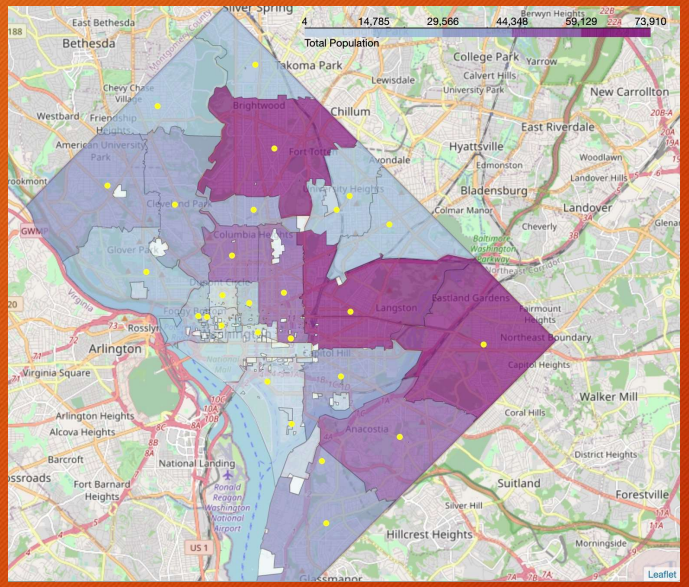
Data Sources

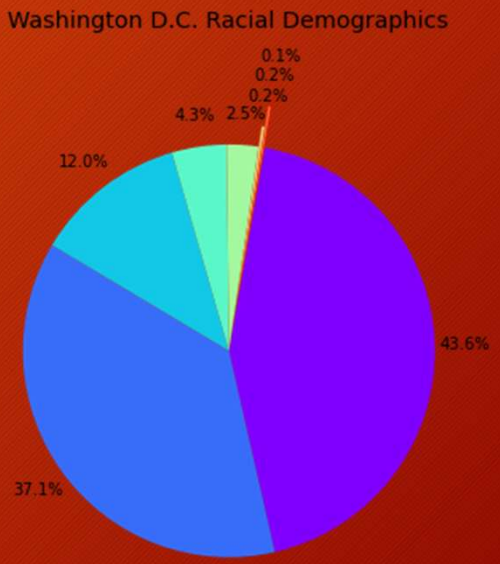
- D.C. Health Matters[1]
  - Racial Demographic Data
    - White, Black, American Indian/Alaska Native, Asian, Native Hawaiian/Pacific Islander, Mixed Race, Hispanic/Latino, and Other Races
  - Median Household Income
  - 29 Zip Codes around D.C.
- FourSquare API[2]
  - Venue Data (Venue location, name, category)
- GeoJSON data[3]
  - Longitude/Latitude information for every D.C. zip code

[1]https://www.dchealthmatters.org/
[2]https://api.foursquare.com/v2/venues/
[3]https://raw.githubusercontent.com/benbalter/dc-maps/master/maps/zip-codes.geojson

Characterization of D.C.

- D.C. Population not evenly distributed throughout zip codes
- Dark purple is more highly populated
- Light blue is less populated

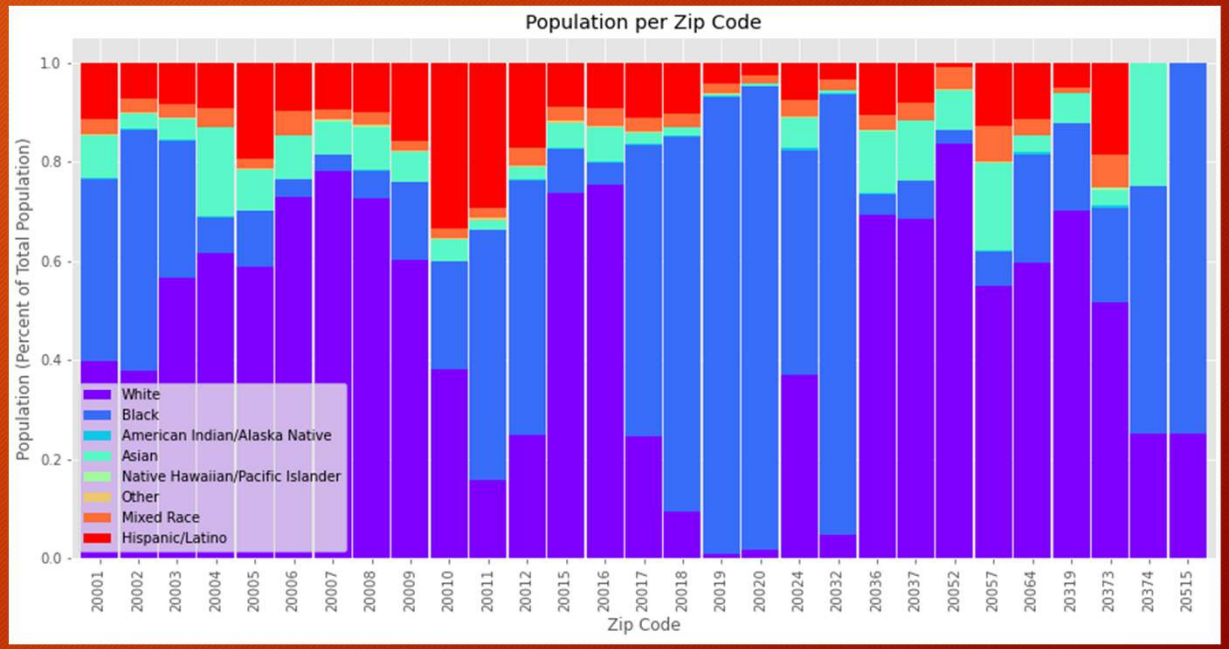




- Racial distribution of D.C.
- Mostly Black and White populations

Population by Zip Code

- Several zip codes are either primarily White or primarily Black
- Races with lower populations also seem clustered in certain zip codes rather than evenly distributed



Venue Categorization

- Identified keywords to group venues into one of 5 “Community Needs”

Community Need	*Keywords*
Athletic/Recreational Facility	'Bike | Tennis | Recreation | Yoga | Field | Skate | Pool | Park | Martial Arts | Gym | Dance'
Restaurants	'Restaurant | Steakhouse | Noodle | Diner'
Fast Food Places	'Fast Food | Food Truck | Joint | Place'
Grocery Stores	'Grocery | Supermarket |Fish Market | Farmers Market | Bodega'
Educational Resources	'College | Student | Bookstore | Museum'

- Search Venue Category for keywords
- If found, Venue is placed in Community Need group
- Restaurants were most prevalent type of Venue (245)
- Other Community Need counts were:
  - Fast Food Places (131)
  - Athletic/Recreational Facilities (77)
  - Educational Resources (42)
  - Grocery Stores (42)

Correlation Analysis

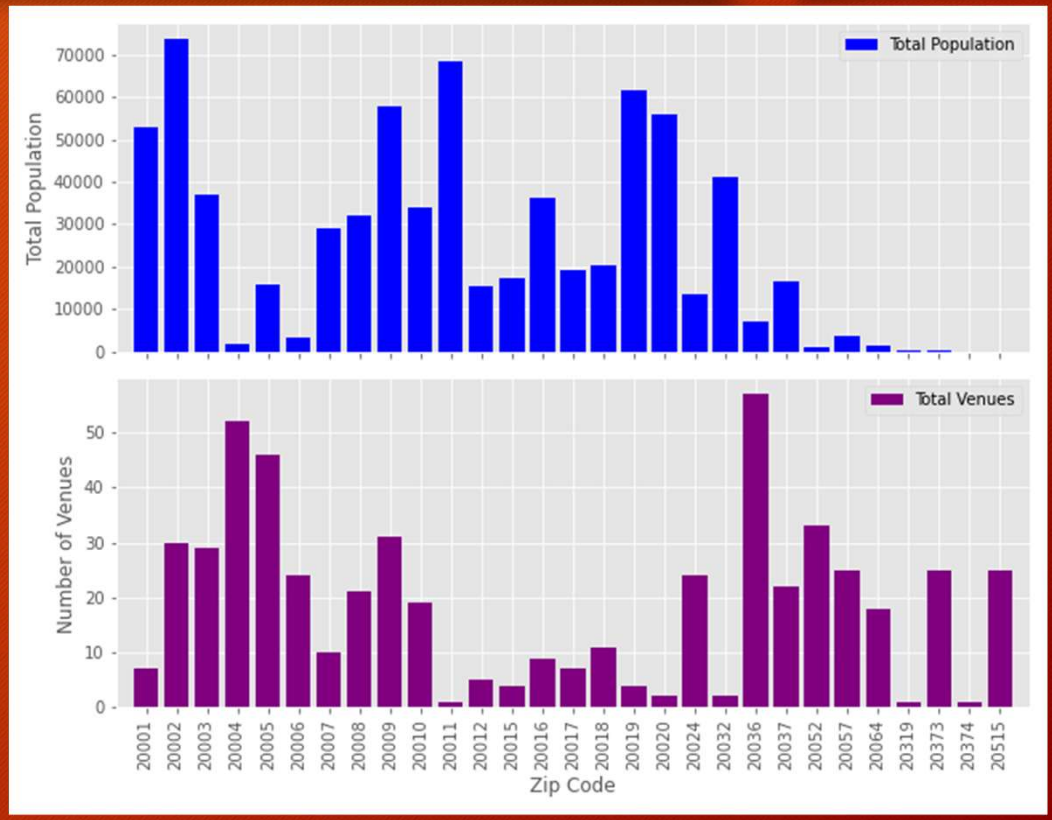
- Pearson Coefficient (Correlation coefficient)
  - Range from -1 to 1
  - Coefficient = 1; Total Positive Correlation
  - Coefficient = 0; No Correlation
  - Coefficient = -1; Total Negative Correlation
- P-Value (Represents confidence level that correlation is significant)
  - P- value << 0.001: 99.9% confidence; strong evidence that the correlation is significant
  - P-value << 0.05: 95% confidence; moderate evidence that the correlation is significant
  - P-value << 0.1: 90% confidence; weak evidence that the correlation is significant
  - P-value >> 0.1: Less than 90% confidence; no evidence that the correlation is significant

Correlation Analysis (Population vs. Venues)

- Number of venues per zip code does not align with total population per zip code (high populated zip codes have low numbers of venues)
- Is there a correlation?

- Pearson Coefficient = -0.25
- P-Value = 0.19

- No evidence of significance of correlation



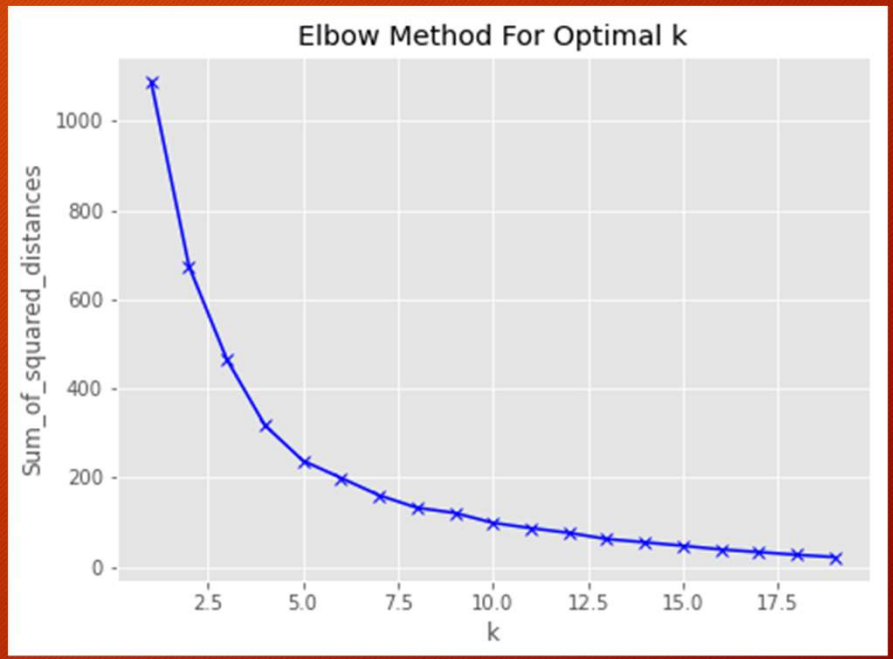
Correlation Analysis (Other Variables)

- Only positive correlations are with White and Asian populations
- Pattern of negative correlations with Black communities
  - Median Household Income
  - Grocery Stores
  - Athletics/Recreational Facilities
- Is there racial bias when choosing where to develop venues?

Evidence	Variable 1	Variable 2	Pearson Coeff	P-Value
Moderate	Asian	Median Household Income	0.391	0.0359
Moderate	Black	Total Venues	-0.376	0.0444
Moderate	Black	Median Household Income	-0.415	0.0252
Weak	Black	Athletics	-0.348	0.0643
Weak	Black	Grocery Stores	-0.354	0.0596
Weak	Black	Fast Food	-0.337	0.0737
Weak	Other	Fast Food	-0.35	0.0629
Weak	Total Population	Fast Food	-0.342	0.0444
Moderate	White	Median Household Income	0.424	0.0218
Weak	White	Grocery Stores	0.331	0.0643

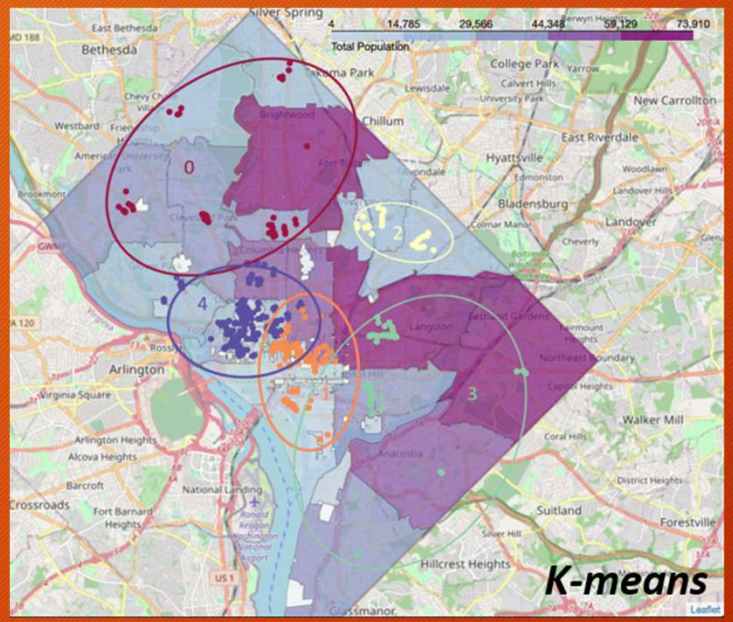
Venue Clustering

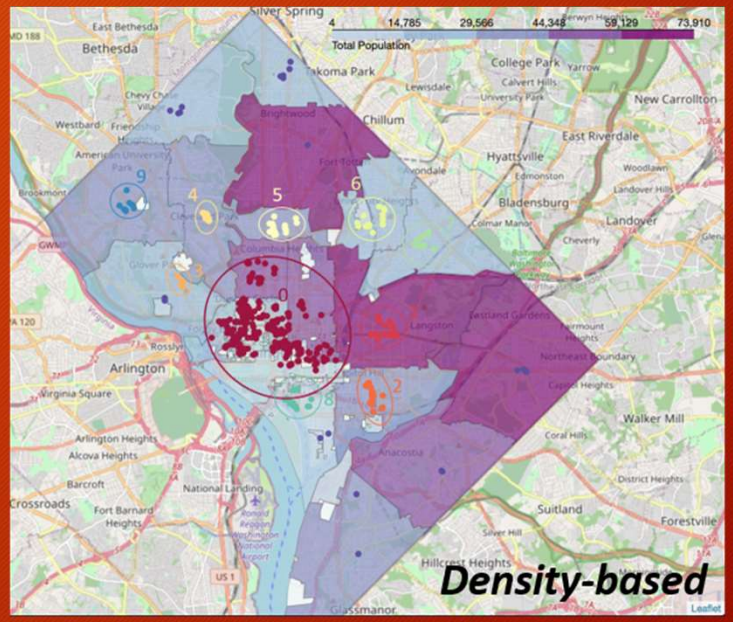
- Two methods of clustering tested:
  - K-means clustering
  - Density-based clustering
- K-means
  - Used elbow method to find optimal value of K (K=5)
- Density-based
  - Finds number of clusters based on density (n clusters = 10)



Venue Clustering

- Density-based clustering did better job of finding approximate clusters
- Majority of venues are in central D.C. (not highly populated regions of D.C.)





Conclusions

- Data presented shows potential racial bias in development of venues in D.C.
- Venue development seems to target central D.C. over more highly populated areas
- Restaurants are the most prevalent type of venue in the dataset

- Is venue development driven by tourism rather than residence?

Future Research

- More research needs to be conducted to determine the driving factors of venue development in D.C.
- Could development be driven by racial bias or the tourism industry?
- Additional data needs:
  - Information on K-12 education to better understand this community need
  - Information tying regions to the tourism industry (which zip codes in D.C. have more tourist attractions)