LIVE WHERE YOU WORK IN DC

COURSERA CAPSTONE PROJECT PRESENTATION

PROBLEM DEFINITION/BACKGROUND

- Washington DC is a very diverse city and the capital of the United States
- The city is broken up into 8 wards numbered 1-8, and is comprised of 131 neighborhoods
- Competitive job market - opportunities all over the city
- Business professionals often like to live within walking/biking distance of their job
- Washingtonians who accept new job opportunities in other parts of the city, need a way to find housing in a similar neighborhood close to their new work location

Contoso

DATA

- Wikipedia Page - https://en.wikipedia.org/wiki/Neighborhoods_in_Washington,_D.C
  - Lists the 8 wards of DC and the neighborhoods belonging to each ward

- Nominatim – Python library to enrich the neighborhood data with latitude and longitude

- Foursquare – Location-based service to gather data about popular venues in each neighborhood

Contoso

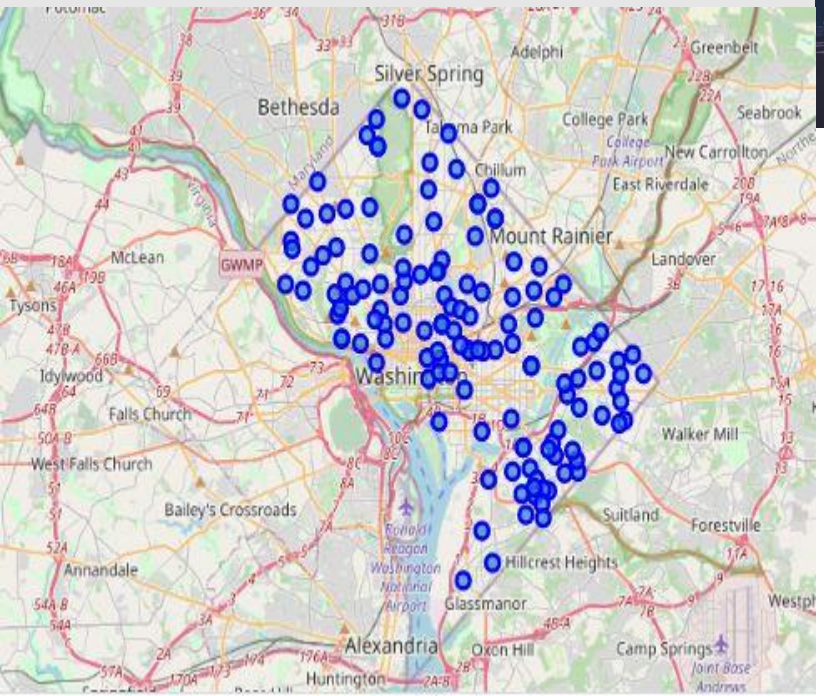
METHODOLOGY

- Obtain the data
- Wrangle the data into appropriate format
  - Drop rows missing key fields
  - Reformat neighborhood names to enable querying for geographic coordinates
  - Convert categorical variables to “dummy” variables
- Exploratory Data Analysis
  - Generate descriptive statistics
  - “Elbow” method to determine the optimal number of clusters for clustering algorithm
- Machine learning
  - K-Means clustering
  - Manually analyze cluster to determine the discriminating venue categories that distinguish each cluster

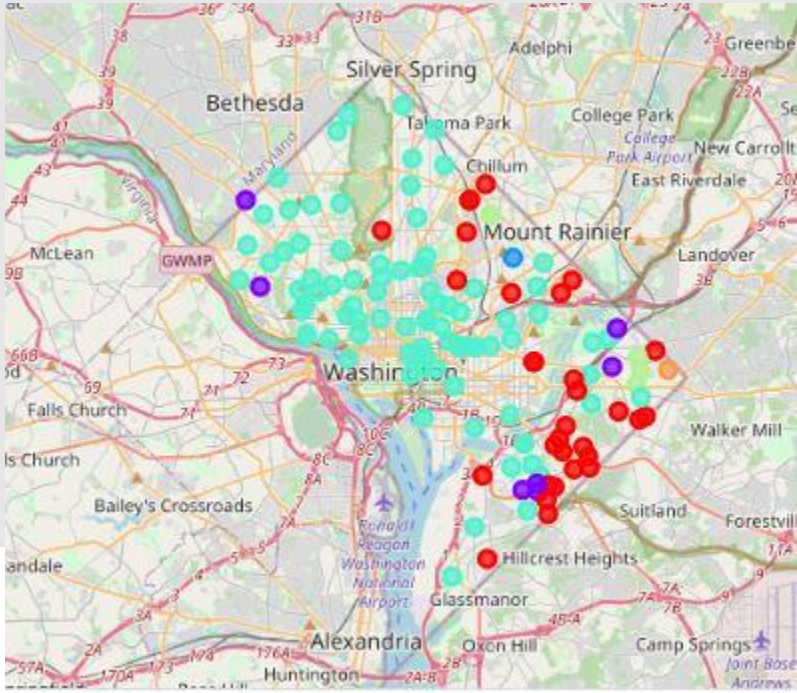
Contoso

RESULTS/DISCUSSION

WASHINGTON, DC



WASHINGTON, DC CLUSTERED



RESULTS/DISCUSSION

MOST NEIGHBORHOODS LIED IN ONE OF THREE CLUSTERS

- A cluster where convenience and liquor stores were the most popular
- A cluster where parks and playgrounds were very popular, and
- A cluster where coffee shops were very prevalent

- If a Washingtonian's current neighborhood was in one of those clusters, they could easily find similar neighborhoods in different parts of the city.

THE THREE CLUSTERS WERE GEOGRAPHICALLY DISPERSED

Contoso

CONCLUSION

- Unsupervised machine learning techniques, specifically clustering, is a viable technique to segment and cluster neighborhoods
  - Enables identifying similar neighborhoods in different locations throughout Washington, D.C.

- This information will be very useful for those business professionals who take new job opportunities across the city, but like to live within walking distance of their employer.

Contoso

Contoso

THANK YOU !

HTTPS://GITHUB.COM/GLENYJ98/COURSERA_CAPSTONE