Where to Retire in San Diego

A Machine Learning Approach

IBM Data Science Capstone Project

Dapeng Wang

February, 2020

Introduction

- San Diego is a well know place to retire.
- There are over 100 neighborhoods in the city. Which area to choose as the place to live is a question for many retiree-to-be.
- In this study we scraped and analyzed data from all these neighborhoods using machine learning method.
- The model can give recommendations based on people’s preference

Data

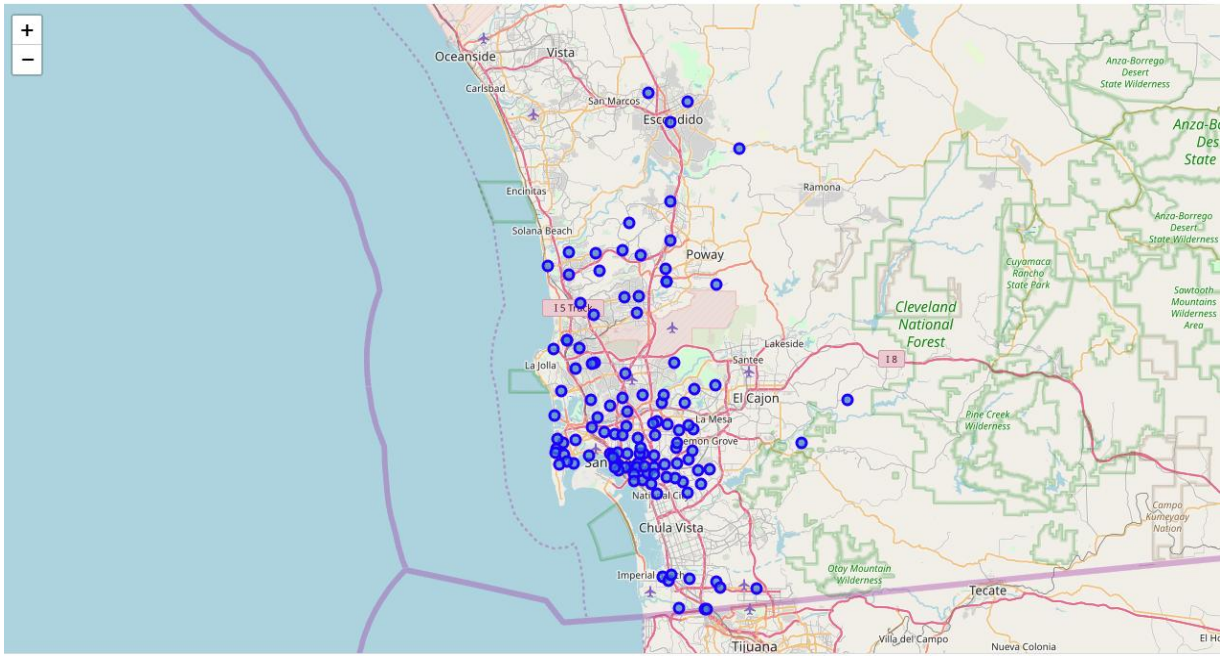
- I included the following data in this study:
  - A complete list of neighborhoods in San Diego from Wikipedia page. https://en.wikipedia.org/wiki/List_of_communities_and_neighborhoods_of_San_Diego
  - Latitude and Longitude data from geocoder package
  - Venue data in each neighborhood from Foursquare API

Methodology

- Neighborhood names were scraped with beautifulsoup library
- Latitude and Longitude data were retrieved with geocoder library
- Overlay of the neighborhood over the map of San Diego were created with Folium library
- Kmeans clustering were used to divide the neighborhood is groups
- Beighborhood ranking were calculated from the user preference and the neighborhood venue dataframe

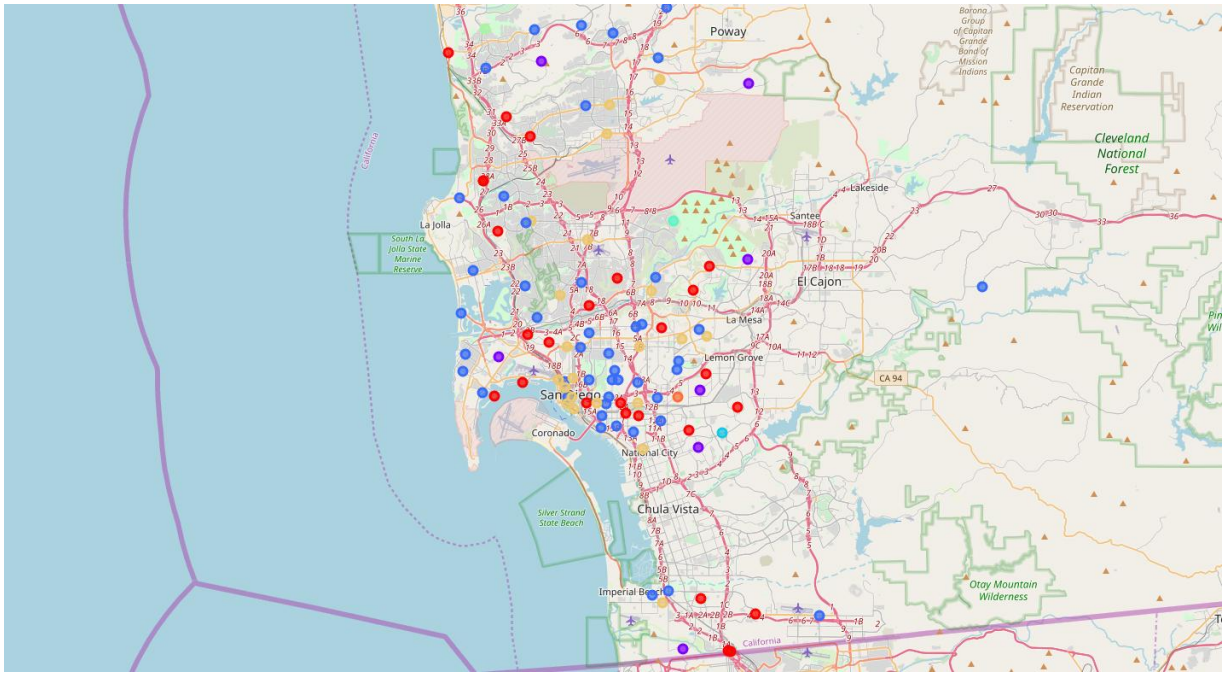
Results

- Overly of the neighborhood over map of San Diego



Results

- Overlay of the neighborhood clustering over map of San Diego

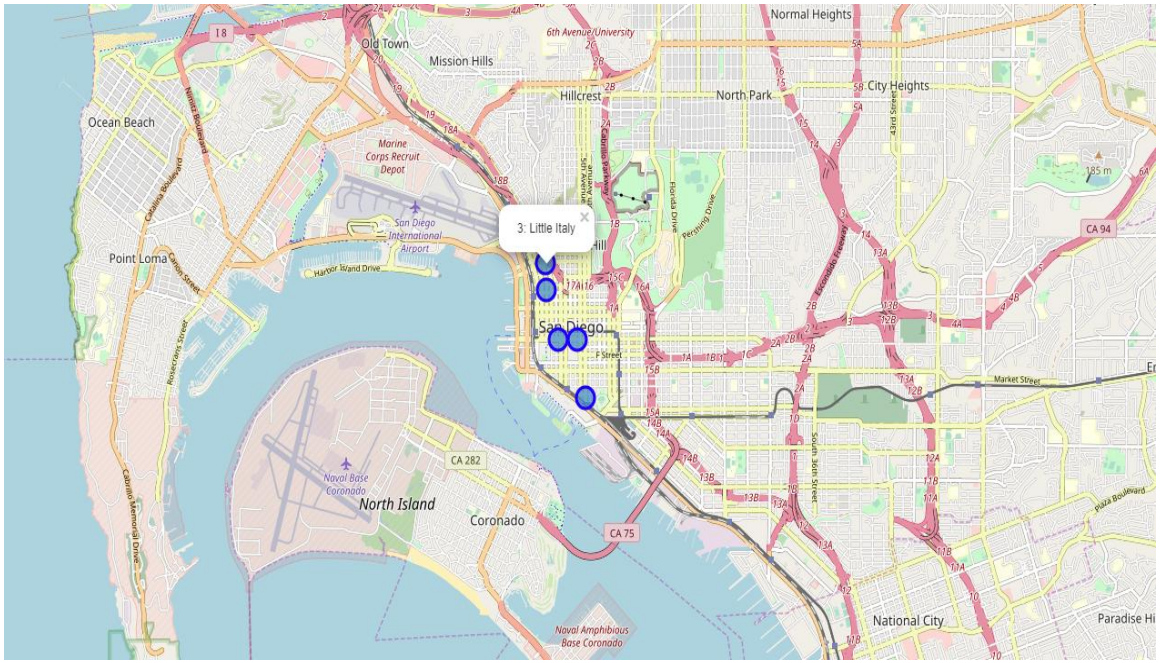


Recommendation

- Here is one user preference:
  - Likes ParkRec and Theater very much (scale 10)
  - Likes Fast Food and Public Transit (scale 5)
  - Does not like Gas Station (scale 0)
- The model gives these top 5 recommendation:
  1. Gaslamp Quarter
  2. Midtown
  3. Little Italy
  4. Columbia
  5. Horton Plaza

Recommendation

- Overlay of the recommended neighborhood over map of San Diego



Conclusion

- In this study, we successfully answered a question “Where to retire in San Diego?” by collecting and analyze the neighborhood geospatial and venue data.
- A machine learning model was build to cluster the neighborhoods and also give personalized recommendations based on user preference