Choosing a suburb city when you work in Dallas

Capstone Project

Jan 22, 2021

Background & Goal

- Dallas is a major city concentrating many businesses and corporate headquarters
- Hundreds of thousands of jobs and people looking for housing in the suburbs - better cost-benefit ratio
- Goal: Segment / cluster the suburb cities in the Dallas-Fort Worth Metroplex area, within 35 miles distance from downtown Dallas, to group different profiles in terms of most common services available

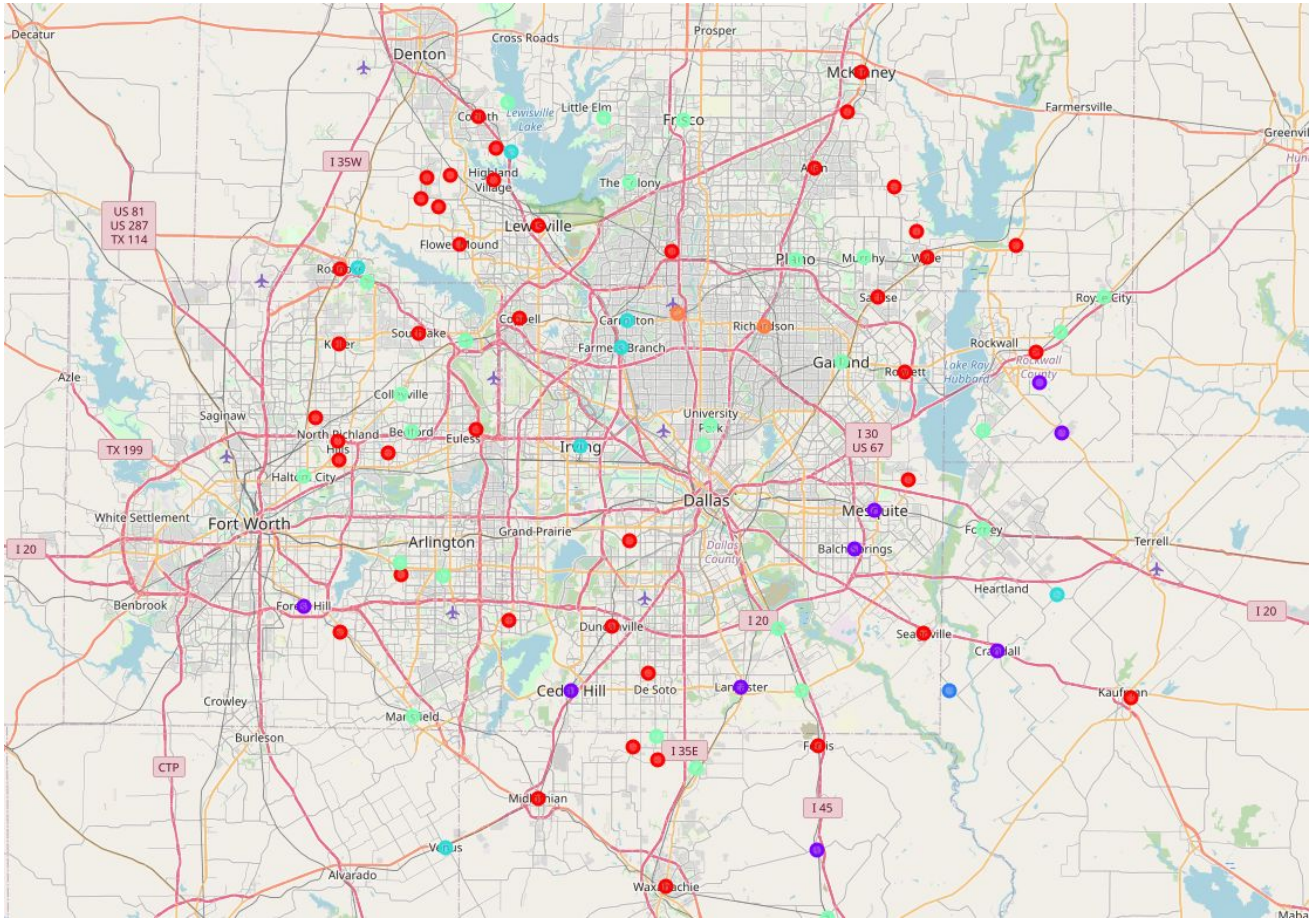
Data Acquisition and Cleaning

1. Produce a CSV file with the list of all the cities in the DFW area
2. Filter out cities farer than 35 miles from downtown Dallas
3. Obtain the geolocation (latitude and longitude) of each city
4. Obtain the nearby POIs from FourSquare - within 6 KM of each city's geolocation
5. Pivot the POIs to count the number of schools, hospitals, groceries / malls, parks, theaters, restaurants and hotels per city
6. Rank the venue categories for each city, from the 1st to the 7th most common

Methodology

- KMeans clustering model to segment the DFW cities by most common category of venue (K = 7)
- Cities' geolocations (latitude and longitude) obtained with geopy
- Venues within 6 KM of each geolocation obtained with Foursquare API
- Venues pivoted and grouped to count each category and rank the most common categories
- Folium map displaying the different clusters by colors

Results - Map of Clusters



Results - List of Clusters

- Cluster 0 : **Predominance of Shopping services, with Theaters more numerous than Hotels** - 45 cities
- Cluster 1 : **Predominance of Shopping services, with Hotels more numerous than Theaters** - 9 cities
- Cluster 2 : **Restaurants first, Theaters second** - 1 city
- Cluster 3 : **Restaurants first, Shopping second** - 7 cities
- Cluster 4 : **Restaurants first, Shopping second - similar to Cluster 3** - 27 cities
- Cluster 5 : **Higher proportion of Hotels** - 1 city
- Cluster 6 : **Restaurants first, Shopping second - similar to Cluster 3** - 2 cities

Main Findings

- K=7 produces similar result than K=4 because the rankings of the categories selected for clustering are very similar across the suburb cities
- Food and Shopping services are almost always the most frequent, with outdoor activities usually in third place and schools (education) predominantly in the last place
- Forcing a clear difference among the clusters:
    - In 3 clusters the shopping activity prevails
    - In 4 clusters the restaurants (food) are more common

Conclusion

- DFW cities have close profile in terms of most common categories of service
- Differentiating the cities as per the frequency of categories of services is not very helpful
- Quantitative economic attributes like house prices and household income or, qualitative ones like school rates likely produce a better differentiation than venue categories