Battle of the neighborhoods: Finding US cities similar to Phoenix

Author: Bhaumik Choksi
Capstone Project for Coursera

Business Problem

- How do we identify similarities between different cities based on the available venues in each city?

- What cities in the US are similar to Phoenix in terms of venues?

Data Sources

- Foursquare:
  - Web API that allows users to create a free account
  - Return nearby venue data for a given latitude and longitude
  - Dynamic in nature

- US Cities Data:
  - Obtained from opendatasoft.com
  - Contains names and coordinates of the 1000 most populous cities in the US
  - Downloadable as a CSV file
  - Static in nature

Methodology

- Preprocessing
  - Data is parsed, required columns are extracted
  - Venue data is obtained from the API
  - Venue categories are one-hot encoded and normalized using mean
- Clustering
  - K-means algorithm
  - One-hot encoded data is clustered
  - Similar cities tend to lie in the same cluster
- Principal Component Analysis (PCA)
  - Reducing dimensionality of the data
  - In our case, data is represented using only 2 dimensions
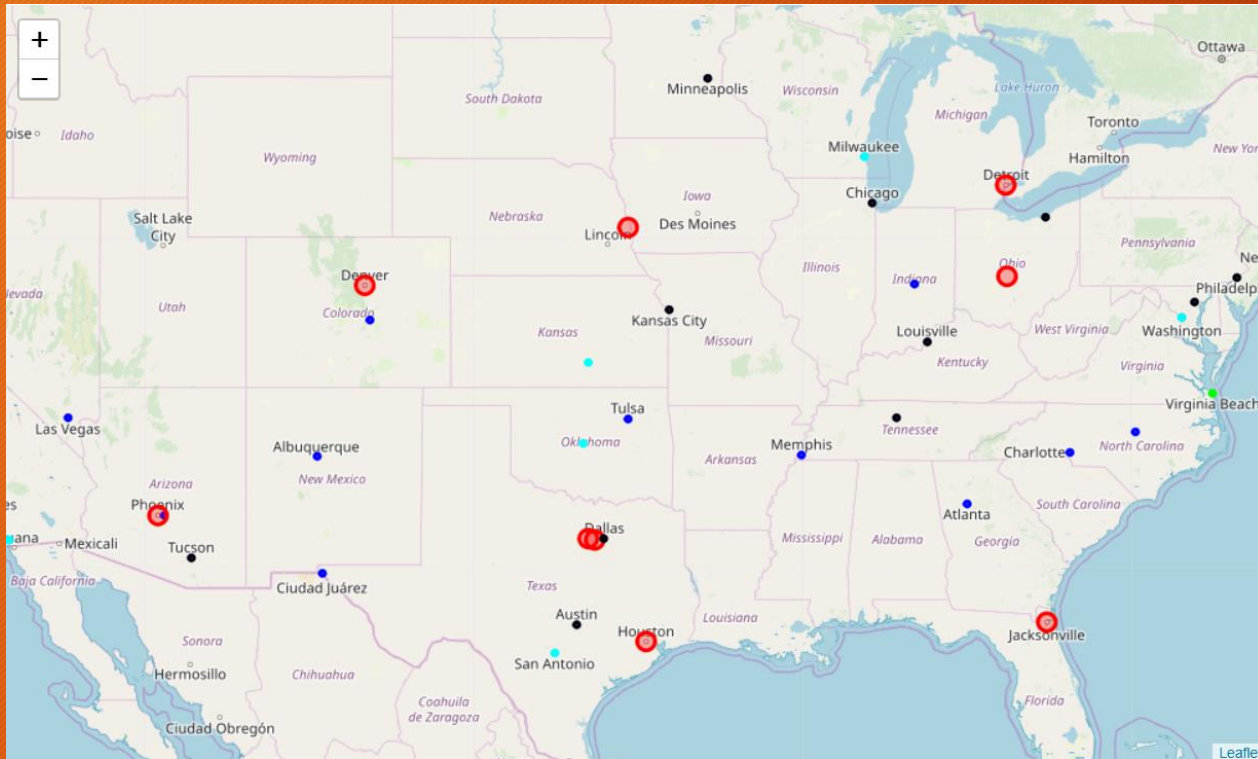 - Helps generate plot and visualize clusters

Results

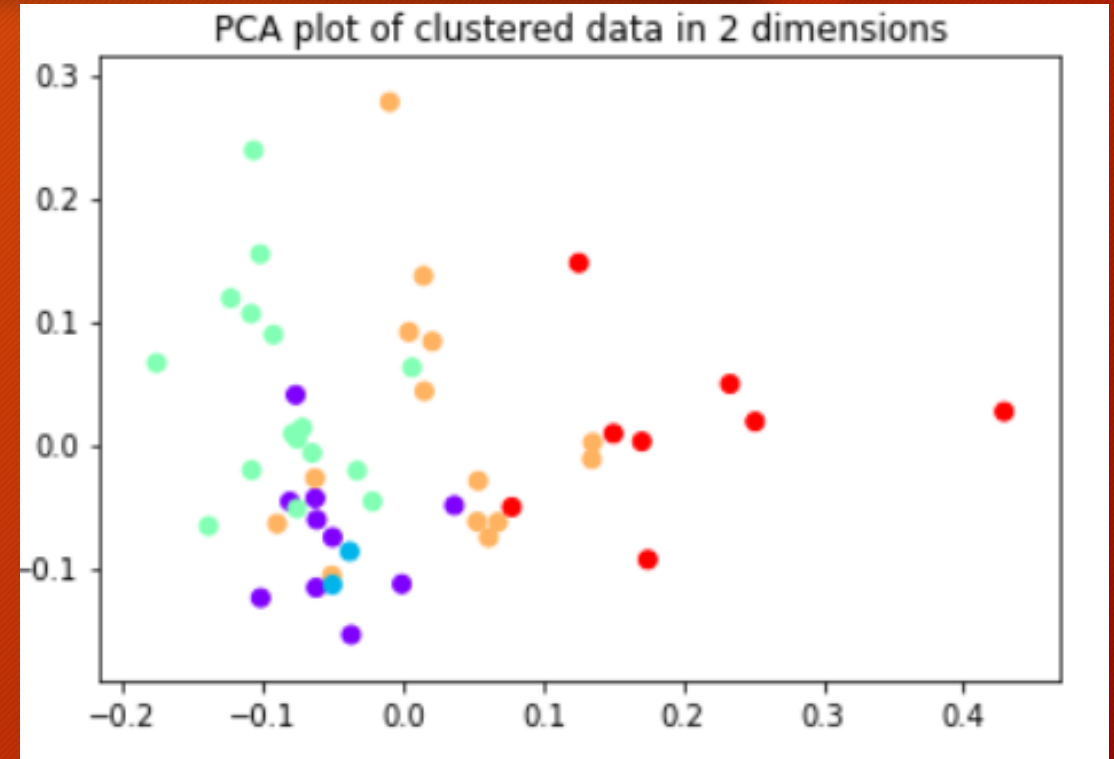Fig 1. Folium generated map of city clusters (Phoenix and similar cities are in red)

Fig 2. PCA plot with colors corresponding to cluster labels

Conclusion

- Similar cities can easily be identified using location and venue data
- PCA reveals the extent of similarity
- Interesting patterns can be discovered by analyzing more cities
- Results depend on the amount of venues processed