The Battle of Neigbourhoods

Final Presentation

Karl De Ruyck

June 25, 2021

Overview of Presentation

American Capital Cities

If we describe state capitals by the types of food that are popular in restaurants there, do we see any trends emerge?

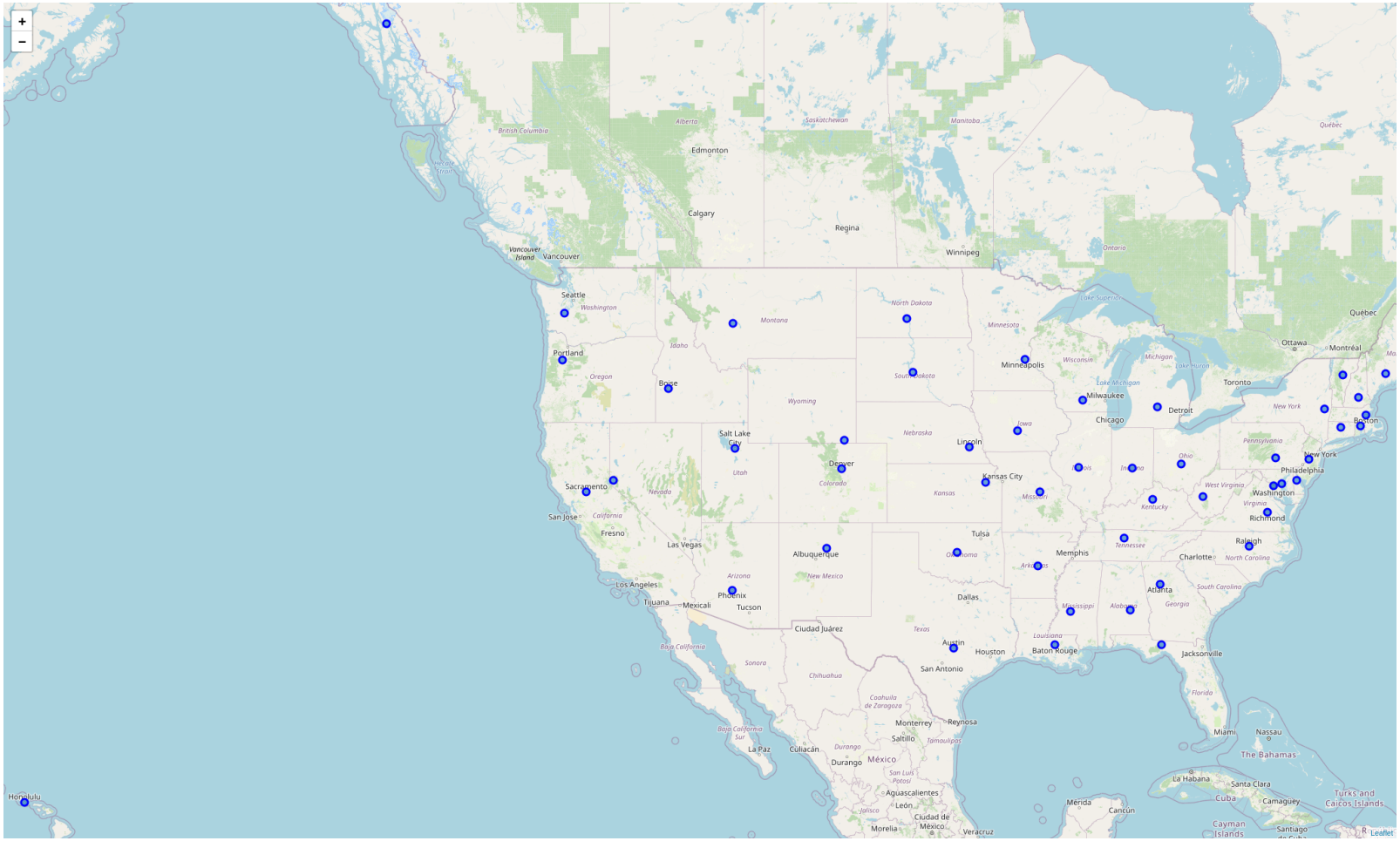

Figure: State capital cities in the United States of America.

The Business Problem

Can each state capital *differentiate themselves* by popular types of restaurant, or do people in state capitals *all prefer the same* types of food?

Perhaps *some* state capitals are similar to each other, while others stand apart?

- **Restaurateurs** may understand where to expand without needing to change their menus to suit the local palate.
- **Tourists** may be enticed to cities for a cuisine they wouldn't easily find elsewhere.
- **Anthropologists** may recognize patterns in human movement that link groups of state capitals with similar tastes.

Data Sources

- A list of American states and respective capital cities was downloaded from Wikipedia
- The latitude and longitude of each city was obtained using the Nominator package
- Popular restaurants in each city were obtained by querying Foursquare
  - Top 10 restaurants from each city
  - Within 2000 m of the co-ordinates returned by Nominator

Exploratory Analysis

Figure: Stacked bar chart showing culinary heterogeneity in each capital city.

By grouping restaurants according to the type of food served, the variety in each city was visualized.

Cities where certain cuisines are predominant can be distinguished from cities where many different cuisines were represented.

By observing long bars of a single colour, cities with predominant popular cuisines are identifiable.

Clustering

Data Preparation

1. List what each city serves
2. Count occurrences of each type
3. Rank cuisines in each city
4. Top 5 ranked cuisines considered representative

k-means Clustering

- Unsupervised learning problem
- Divide 50 cities into k clusters
- Try 3 to 10 clusters
- Optimal clustering at $k = 8$

Clusters Identified

As expected, each of 8 clusters was dominated by particular cuisines.

For example:

- A 3-city cluster prefers Food Courts
- A 6-city cluster prefers Mexican
- A 4-city cluster prefers Bakeries and Seafood

Geolocation of Clusters

If we describe state capitals by the types of food that are popular in restaurants there, do we see any trends emerge?

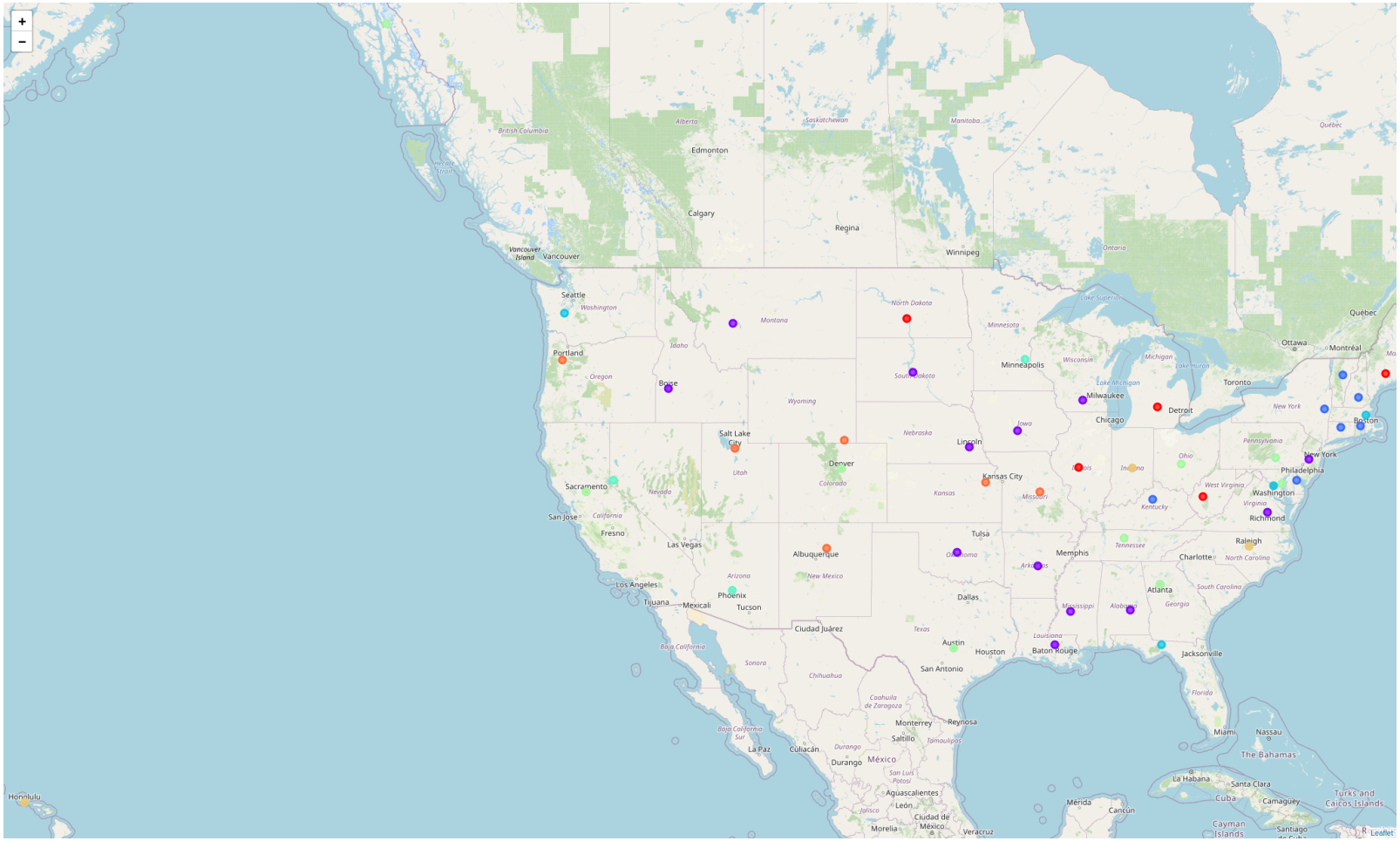

Figure: State capital cities, coloured by clustering according to popular cuisine

Geolocation of Clusters

- An orange cluster includes several central capital cities.
- A purple cluster lies above and below the orange cluster.
- A blue cluster is only the east, particularly the north-east.
- Each state borders at least one other of a different cluster.
- A light blue cluster is widely & sparsely distributed.
- A yellow cluster includes 3 different coastlines.

Assumptions and Sources of Bias

- Restaurants not listed by Foursquare were omitted.
- Patrons not checking-in to Foursquare don't contribute to popularity.
- Foursquare's categorizations were not validated for accuracy.
- Nominator's co-ordinates may not correspond to a city's culinary centre.
- Popular restaurants may lie beyond 2 km from those co-ordinates.

Business Solutions

By examining the clustering of capital cities according to types of food served in their popular restaurants, various business interests may gain valuable insights:

1. Successful restaurants in one city may try expanding to other cities in the same cluster.
2. Culinary tourists need only cross one state border to find a capital city from a different cluster.
3. Some culinary clusters may be linked by other factors besides geolocation.

Conclusions

Some state capitals have similar types of popular restaurant, while other capital cities share other culinary profiles.

Eight clusters were identified by applying the k-means unsupervised clustering machine learning approach. Some clusters included cities with geographic similarities, while others did not appear strongly associated by geolocation.

Future works extending the present body may seek to address and eliminate sources of bias.

Critical input may also be received from interest groups seeking to apply the results of this analysis.