

Content-based Recommender System for the Clusters of potential interest

Based on OSM data for San Francisco

Mariia Sundeeva. Data Science Capstone. 3/1/23

For whom

People with non-mainstream preferences

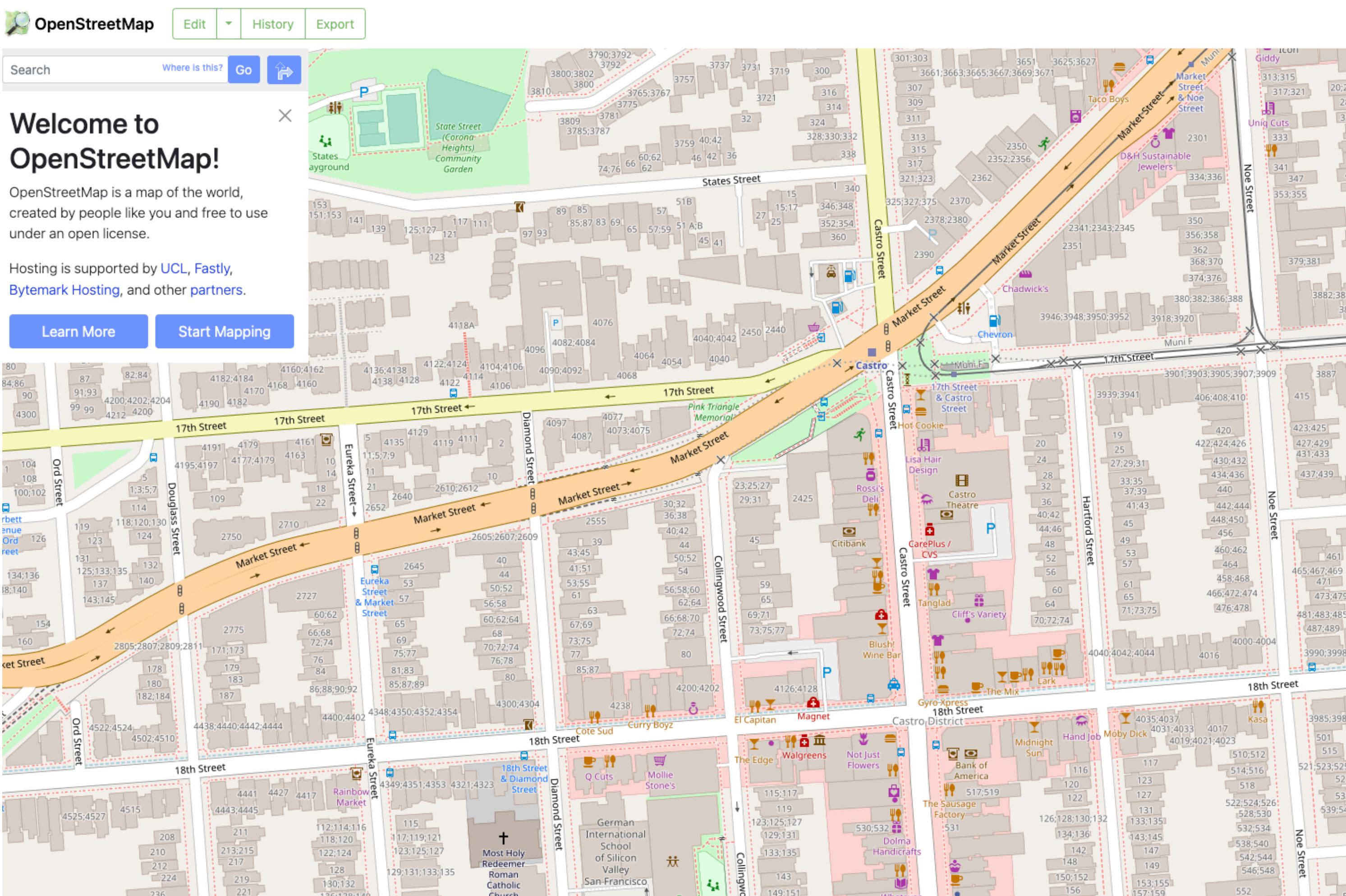
Cluster is something an average person walks around in a comfortable pace

Consists of at least 3 places

Top categories: cafes, restaurants, places of worship, fast food, bars, banks, pubs

Data Collection

Open Street Map via OSMnx



Tag: amenity

Geometry: points
and ways

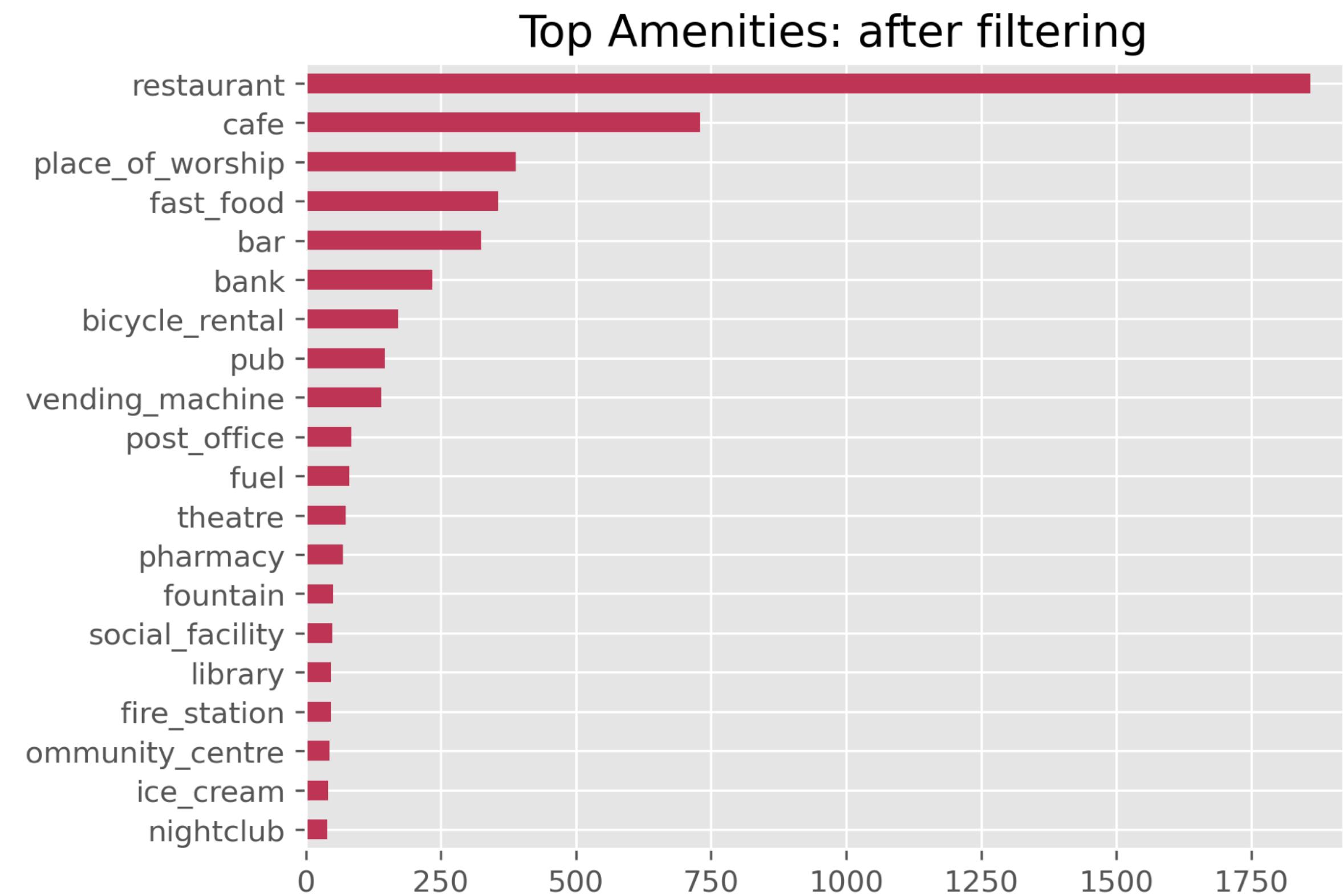
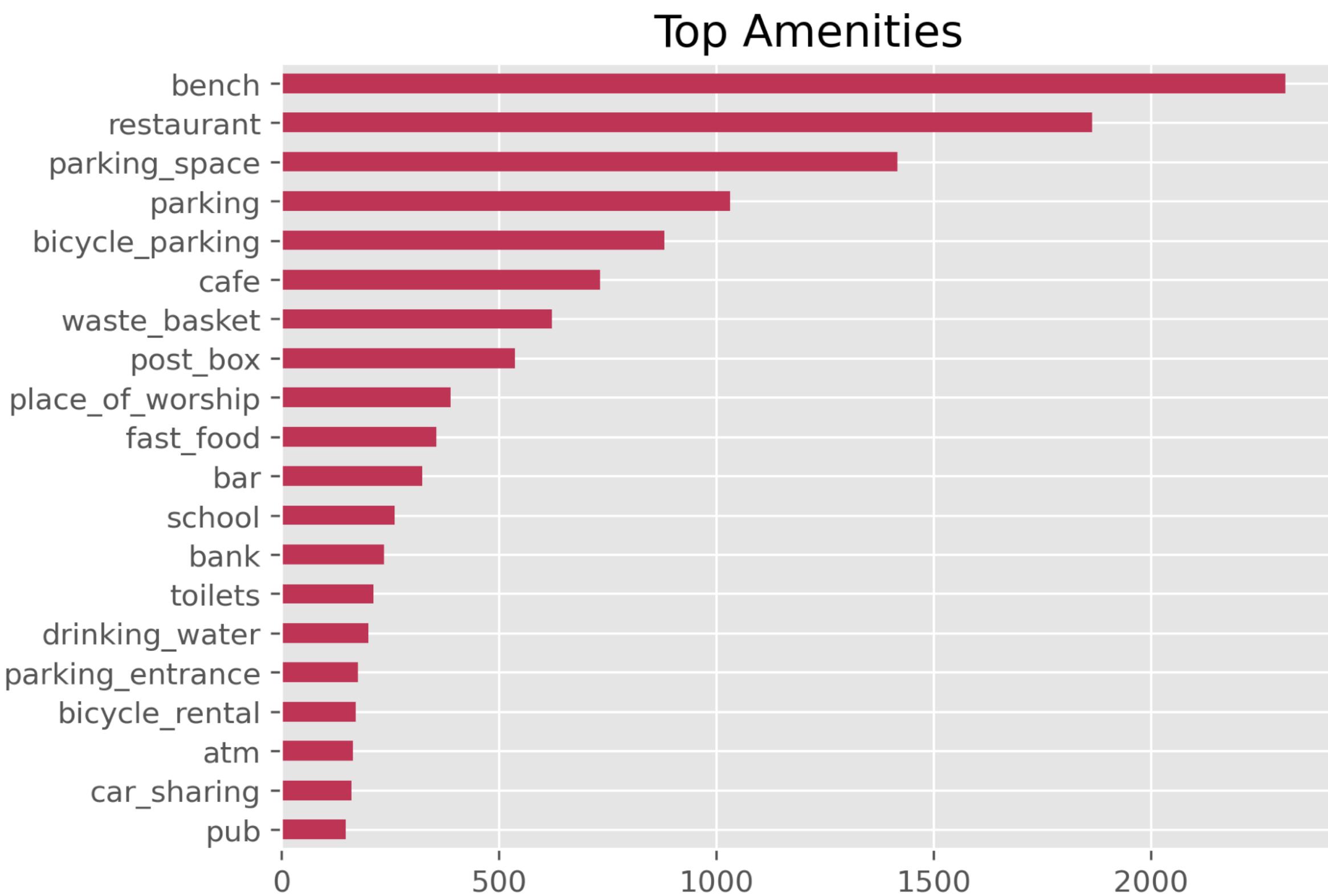
Location: San
Francisco

Observations: 13k

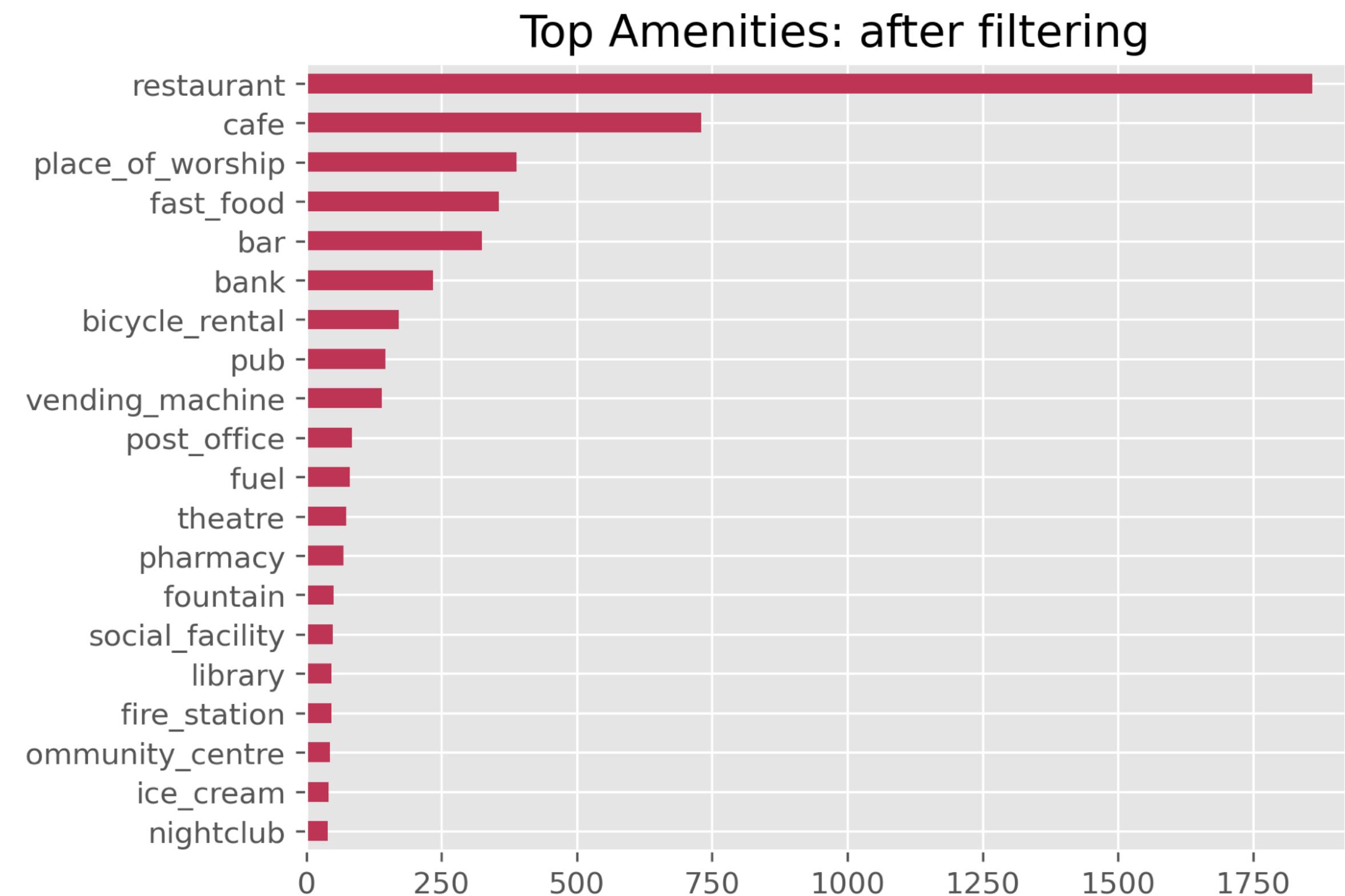
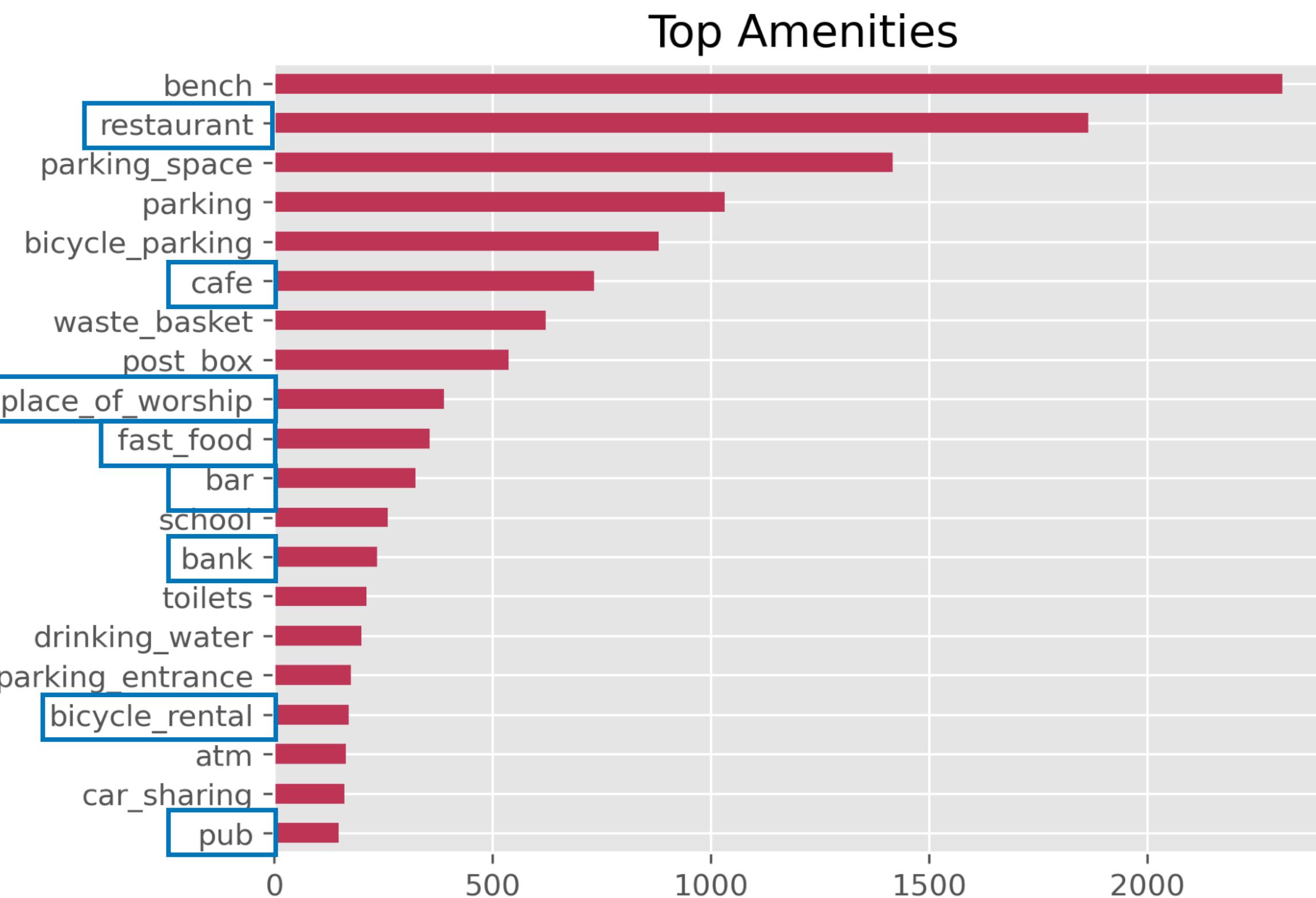
Data Cleaning And Feature Engineering

- Amenities shortlist
 - No Parks' infrastructure
 - No Transit Infrastructure
 - No Bicyclists' infrastructure

Shortlisted Amenities

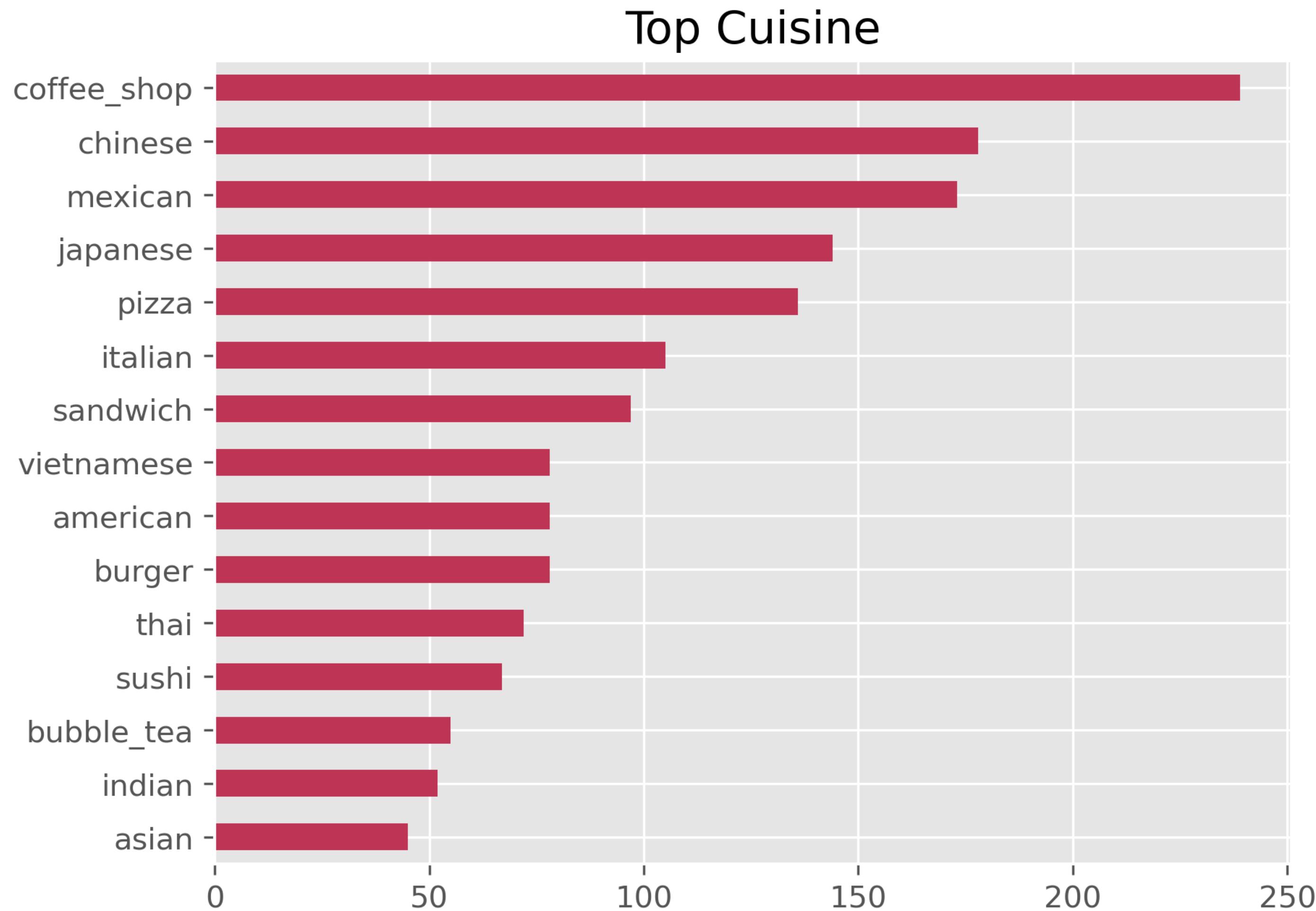


Shortlisted Amenities



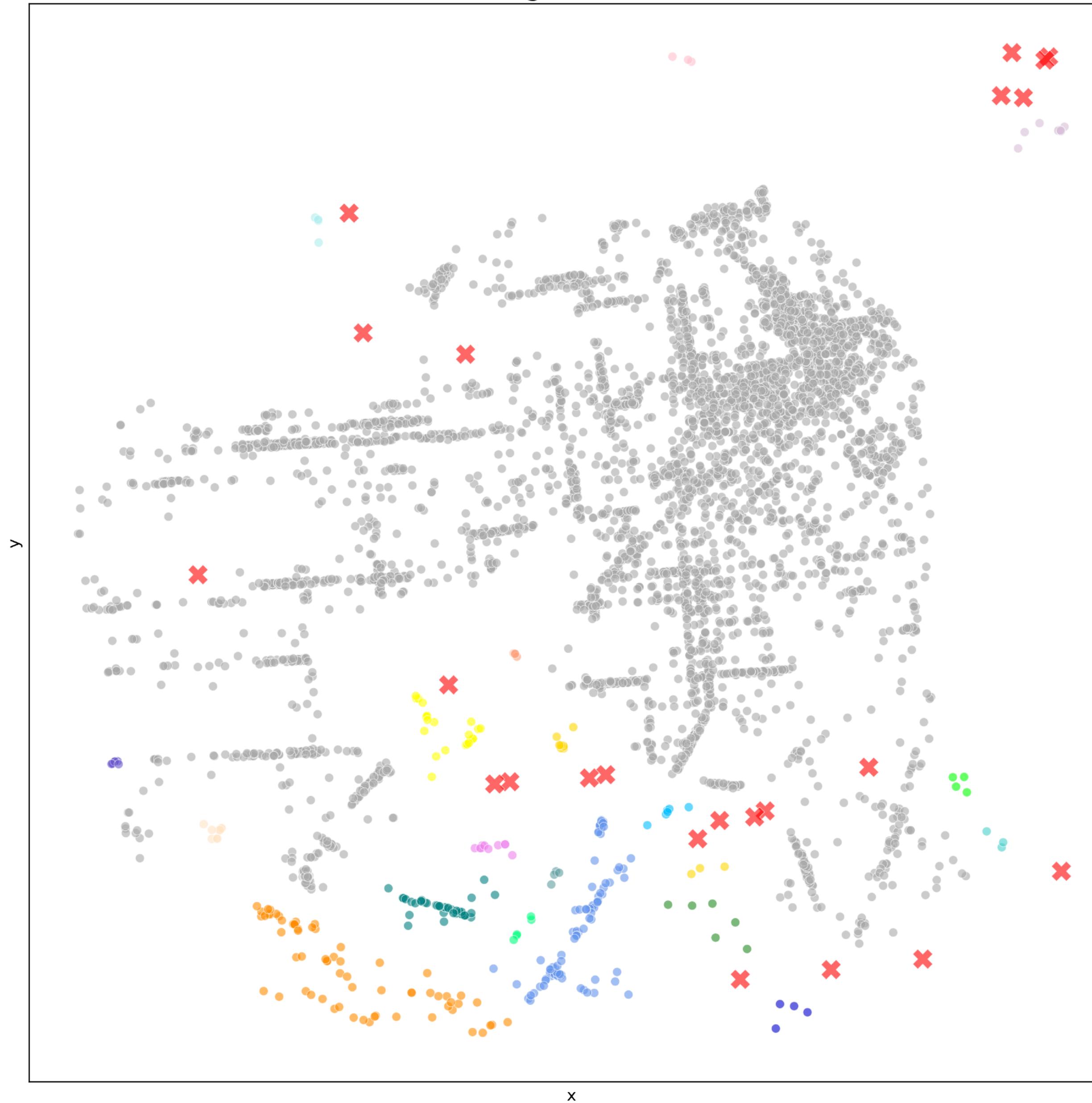
Feature Engineering

Tag: Cuisine

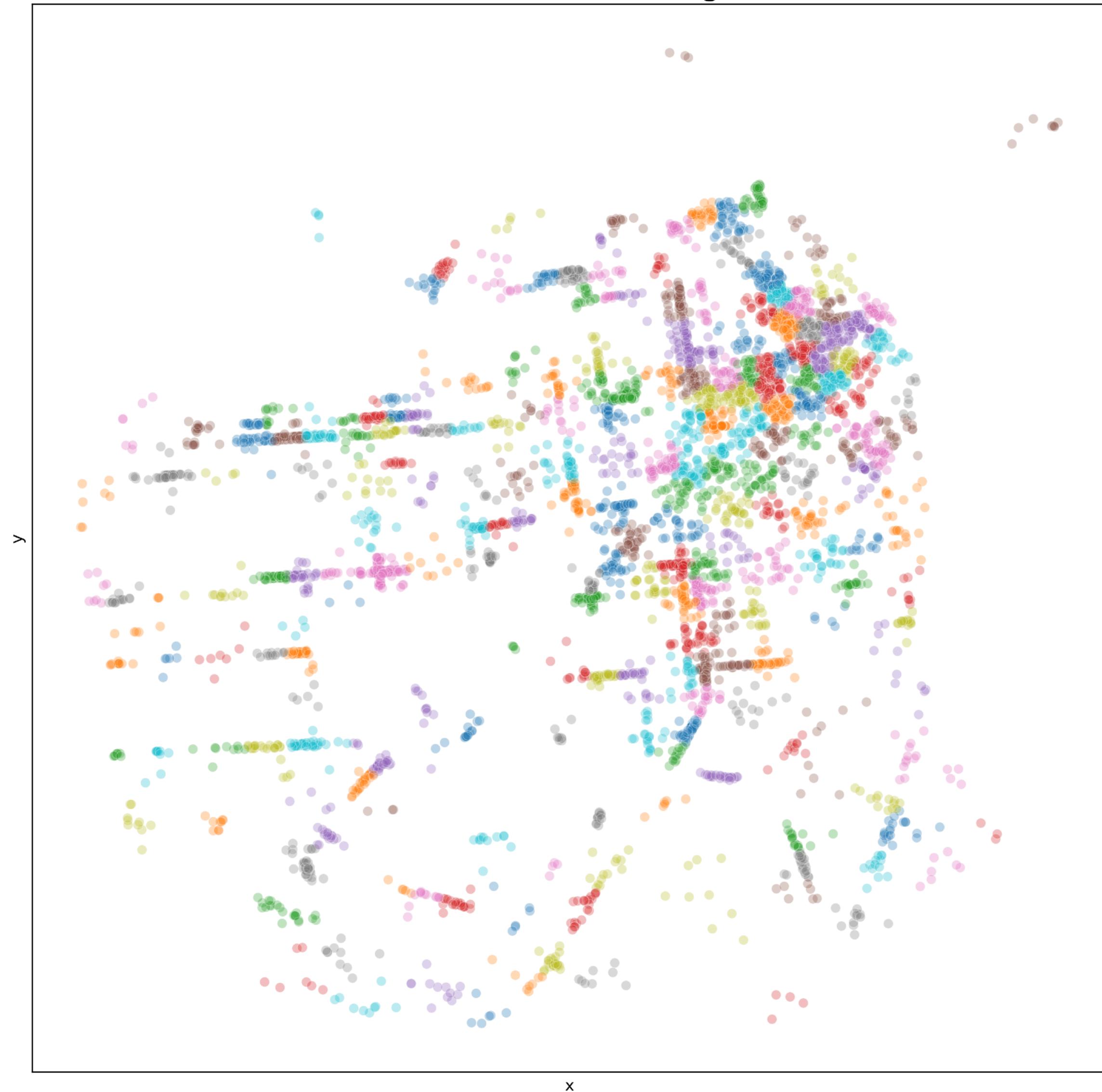


Clustering Using only XY

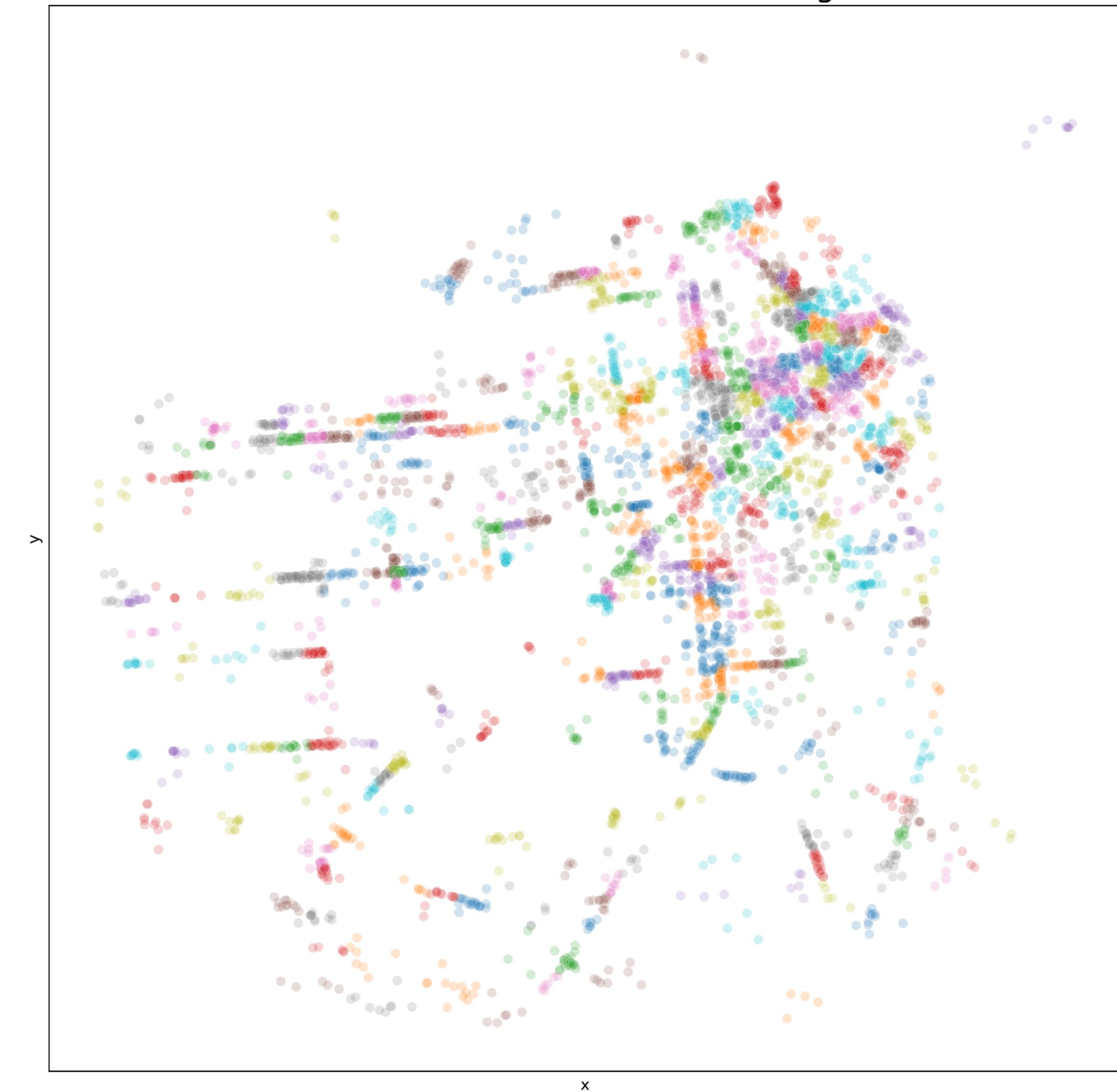
DBSCAN clustering: Clusters and Outliers



K-Means clustering

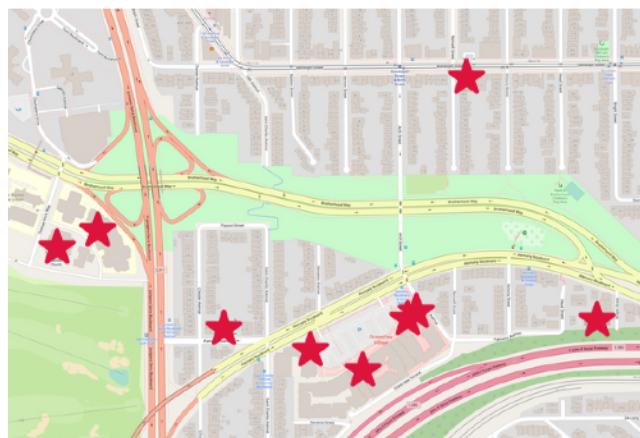


K-Means constrained clustering



Biggest Clusters

Cluster 55



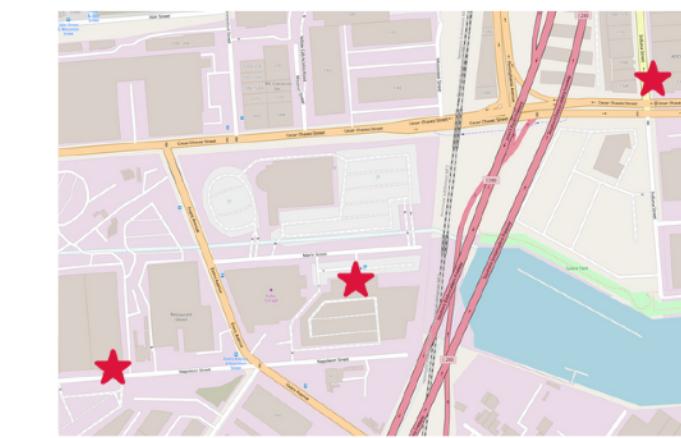
Cluster 274



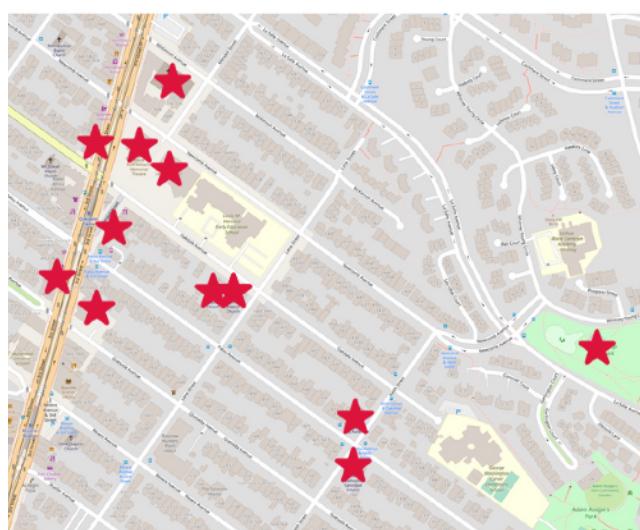
Cluster 154



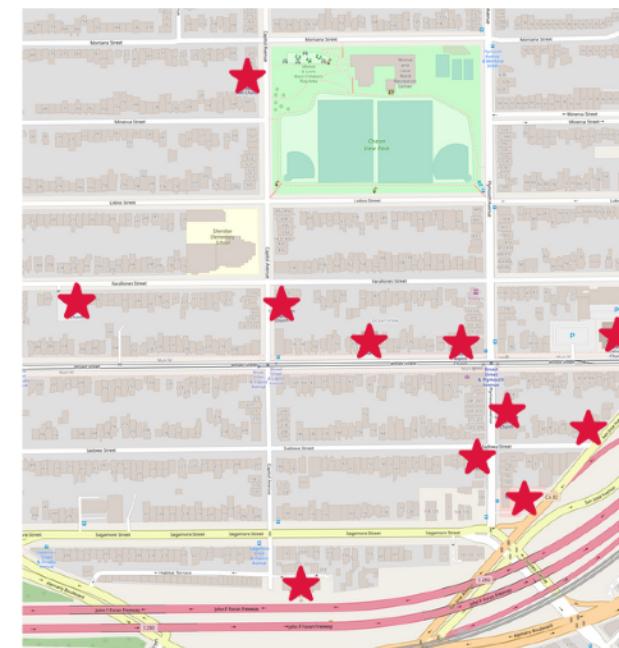
Cluster 280



Cluster 6



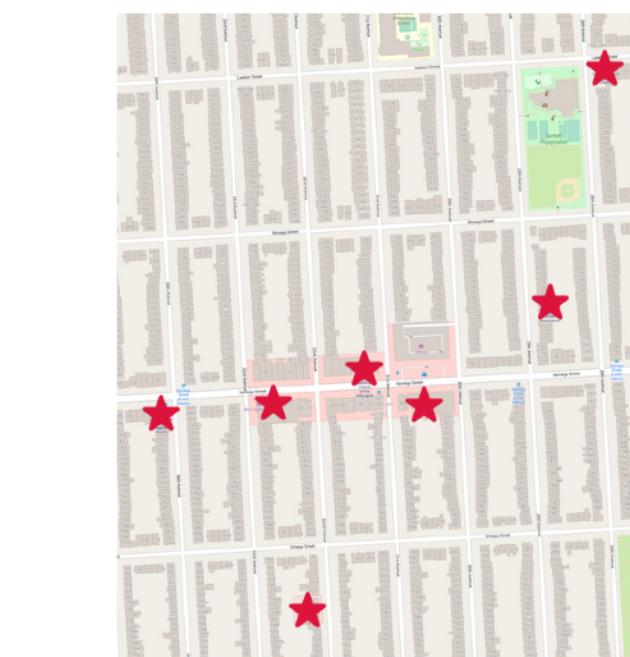
Cluster 106



Cluster 299



Cluster 161



Cluster 181



Cluster 109



Cluster 131

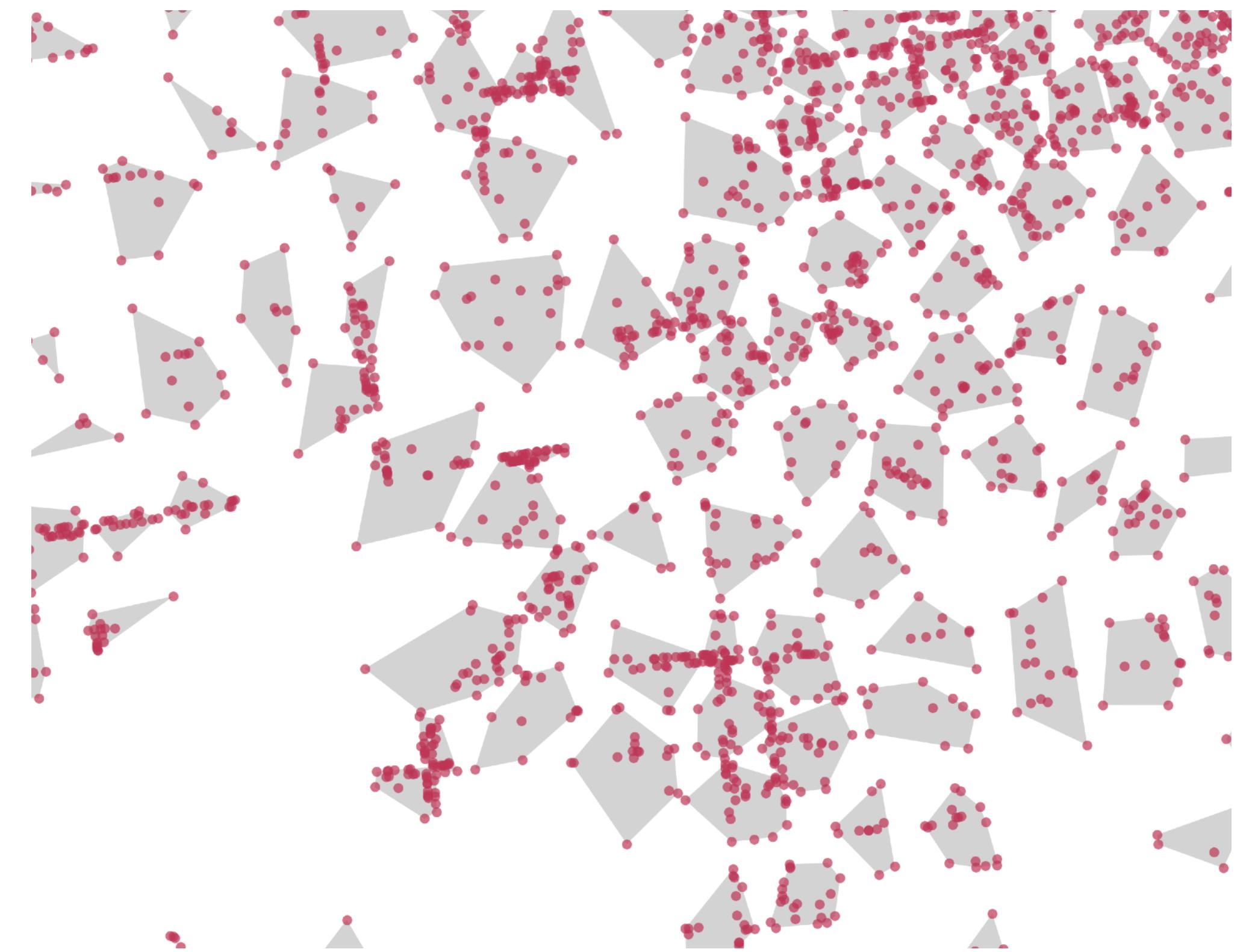


Cluster 102



Cluster to Polygons conversion

GeoPandas



Potential Improvements

- Include more categories
- Collect user feedback
- Add more data