Determining Retail Categories to Open in the Church Hill Neighborhood of Richmond, VA

Mitch Phillips

Introduction - Background

- **Goal: Determine under-represented retail categories with a high likelihood of success for a growing neighborhood.**
- **For whom: Property managers, entrepreneurs, and existing retail business owners.**
- **Where: Church Hill in Richmond, Virginia, a growing neighborhood.**
- **How: Compare Church Hill retail to similar neighborhoods.**

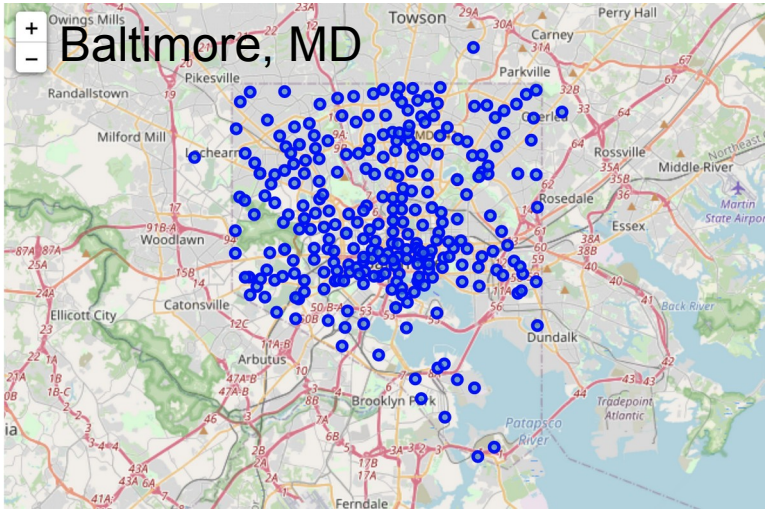
Introduction - Problem

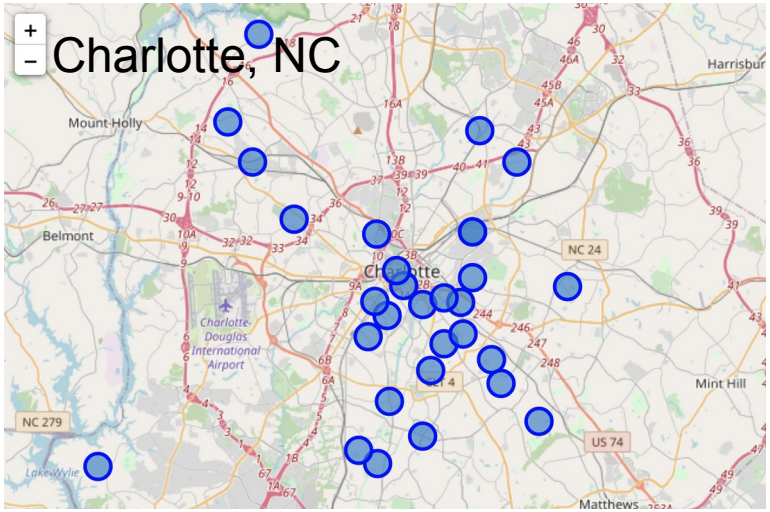
- How can we characterize the retail of Church Hill?
- How can we determine neighborhoods with similar retail to Church Hill?
- How will we compare Church Hill retail to these similar neighborhoods?

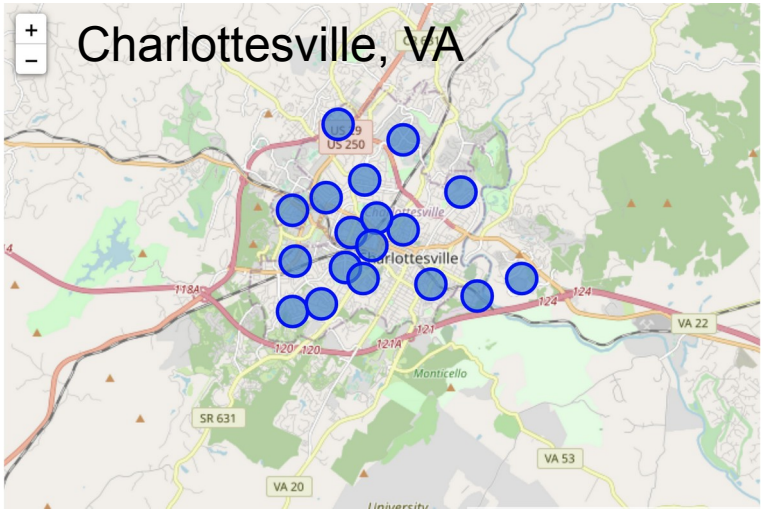
Introduction - Plan

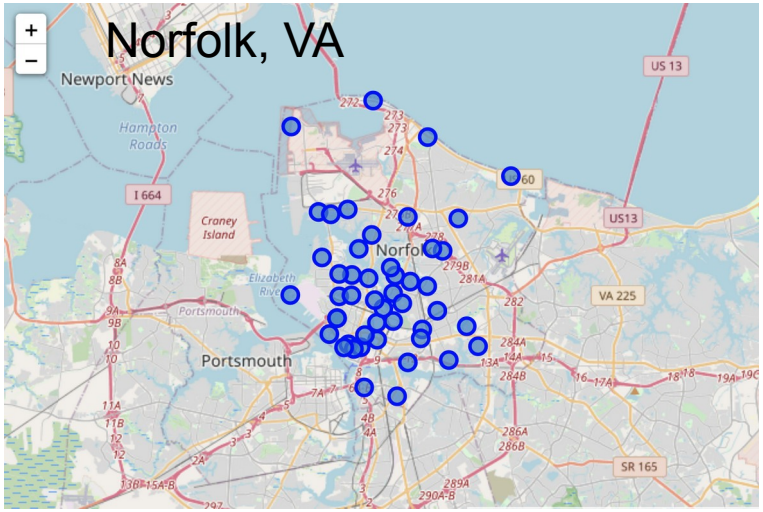
- Compile a list of retail categories present in many neighborhoods.
- Machine Learning clustering to find neighborhoods similar to Church Hill.
- Compare Church Hill retail categories to average frequency of retail categories in the similar cluster.

Neighborhood Data Collection

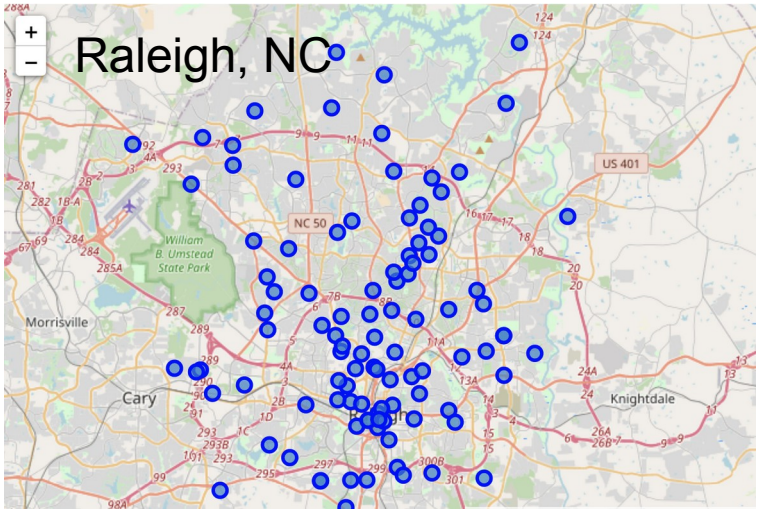


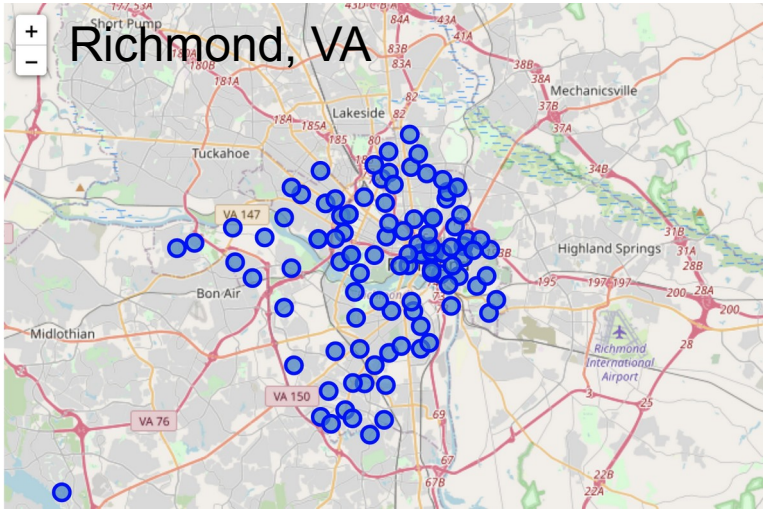


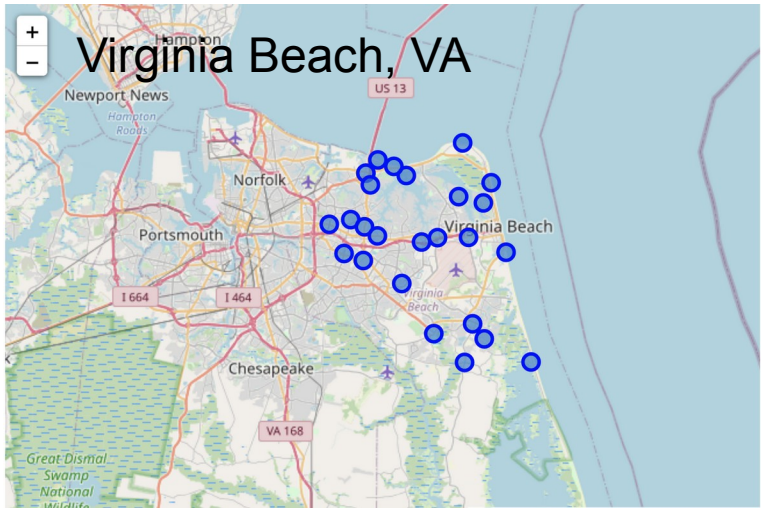


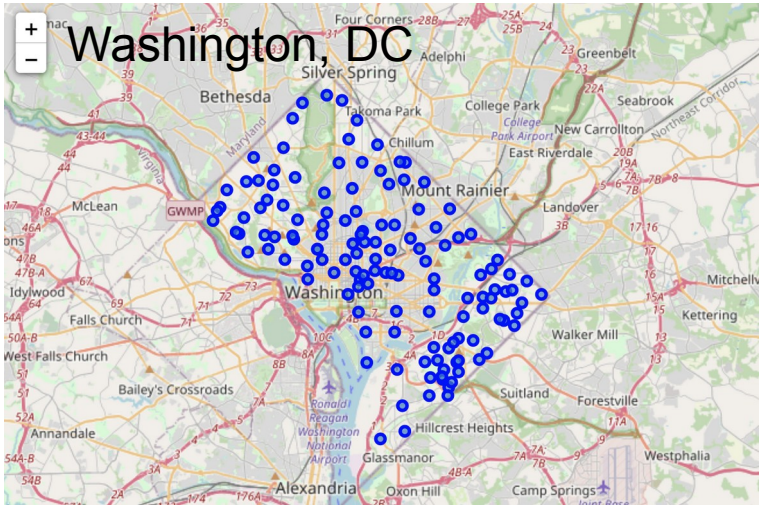


Neighborhood Data Collection









Foursquare Retail Venue Data

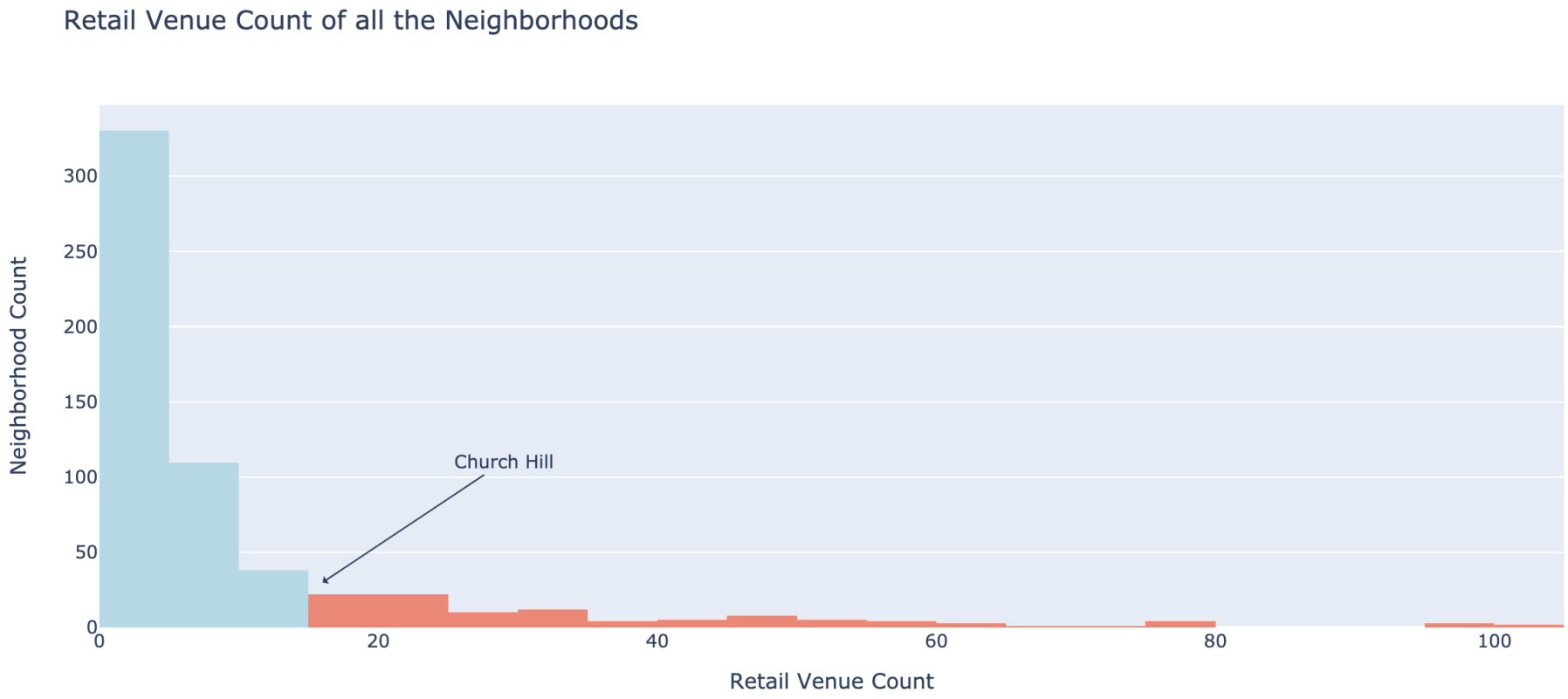
	Neighborhood	City	State	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Arts District	Richmond	VA	37.543453	-77.438963	Pop's Market on Grace	37.542080	-77.438512	Café
1	Arts District	Richmond	VA	37.543453	-77.438963	Secret Sandwich Society	37.541787	-77.438228	Sandwich Place
2	Arts District	Richmond	VA	37.543453	-77.438963	Perly's	37.543848	-77.441436	Deli / Bodega
3	Arts District	Richmond	VA	37.543453	-77.438963	Rappahannock Restaurant	37.542810	-77.439207	Seafood Restaurant
4	Arts District	Richmond	VA	37.543453	-77.438963	Salt & Forge	37.545206	-77.440183	Sandwich Place
...	...	...	...	...	...	...	...	...	...
5735	Starmount	Charlotte	NC	35.141693	-80.868220	Park South Station Pool and Fitness Club	35.145354	-80.864310	Club House
5736	Starmount	Charlotte	NC	35.141693	-80.868220	On The Go Mart	35.143072	-80.875201	Gas Station
5737	Stonehaven	Charlotte	NC	35.155497	-80.763010	Quality Grill Parts, LLC	35.156302	-80.761389	Business Service
5738	Stonehaven	Charlotte	NC	35.155497	-80.763010	Betty Boos Treats	35.155218	-80.756179	Bakery
5739	University City	Charlotte	NC	36.037769	-79.034256	DW EVANS ELECTRIC INC	36.033431	-79.029239	Home Service

5740 rows × 9 columns

We've collected data on a total of 5740 retail venues from the eight cities and 747 neighborhoods.

Data Cleaning – Removing Small Neighborhoods

- **First we get rid of unwanted Foursquare venue entries.**
- **Then we remove entire neighborhoods with less than 15 retail venues.**
- **Now we have 106 neighborhoods to cluster.**



Data Preparation – Category Frequencies

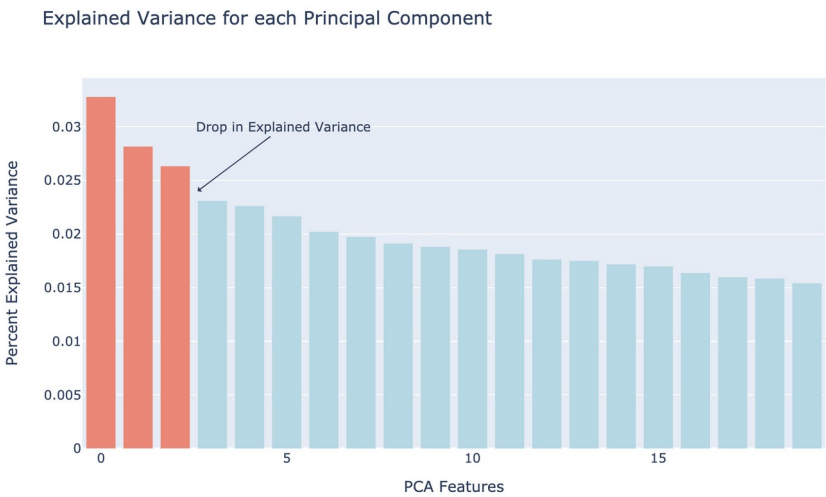
	Neighborhood	City	State	Neighborhood Latitude	Neighborhood Longitude	Neighborhood Venue Count	ATM	Accessories Store	Afghan Restaurant	African Restaurant	...
0	Adams Morgan	Washington	DC	38.921242	-77.043493	52	0.0	0.0	0.019231	0.0	...
1	Aragona Village	Virginia Beach	VA	36.858903	-76.152288	15	0.0	0.0	0.000000	0.0	...
2	Arts District	Richmond	VA	37.543453	-77.438963	23	0.0	0.0	0.000000	0.0	...
3	Bayside	Virginia Beach	VA	36.902925	-76.134380	34	0.0	0.0	0.000000	0.0	...
4	Bellevue	Richmond	VA	37.590832	-77.457554	15	0.0	0.0	0.000000	0.0	...
...	...	...	...	...	...	...	...	...	...	...	...
101	West Raleigh Historic District	Raleigh	NC	35.777619	-78.638765	78	0.0	0.0	0.000000	0.0	...
102	Westhampton	Richmond	VA	37.574101	-77.514645	19	0.0	0.0	0.000000	0.0	...
103	Wilder's Grove	Raleigh	NC	35.798804	-78.564528	19	0.0	0.0	0.000000	0.0	...
104	Willow Lawn	Richmond	VA	37.581870	-77.497587	39	0.0	0.0	0.000000	0.0	...
105	Wyman Park	Baltimore	MD	39.330870	-76.627553	22	0.0	0.0	0.000000	0.0	...

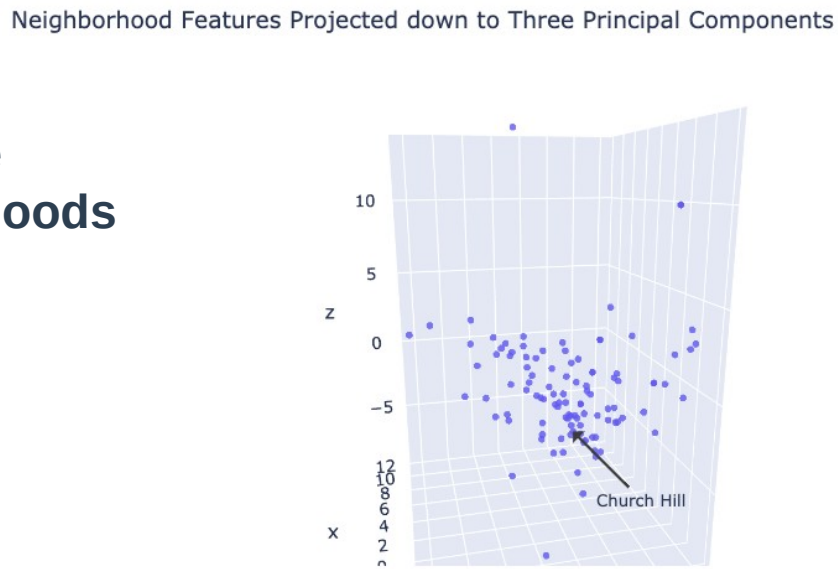
106 rows × 315 columns

- We create a dummy column for each retail category and place a 1 in the corresponding venue row.
- We group each neighborhood together and take the mean of the columns for a category frequency value.
- There are 309 retail categories total.
- We will also include the venue count column so we'll have 310 features for clustering.

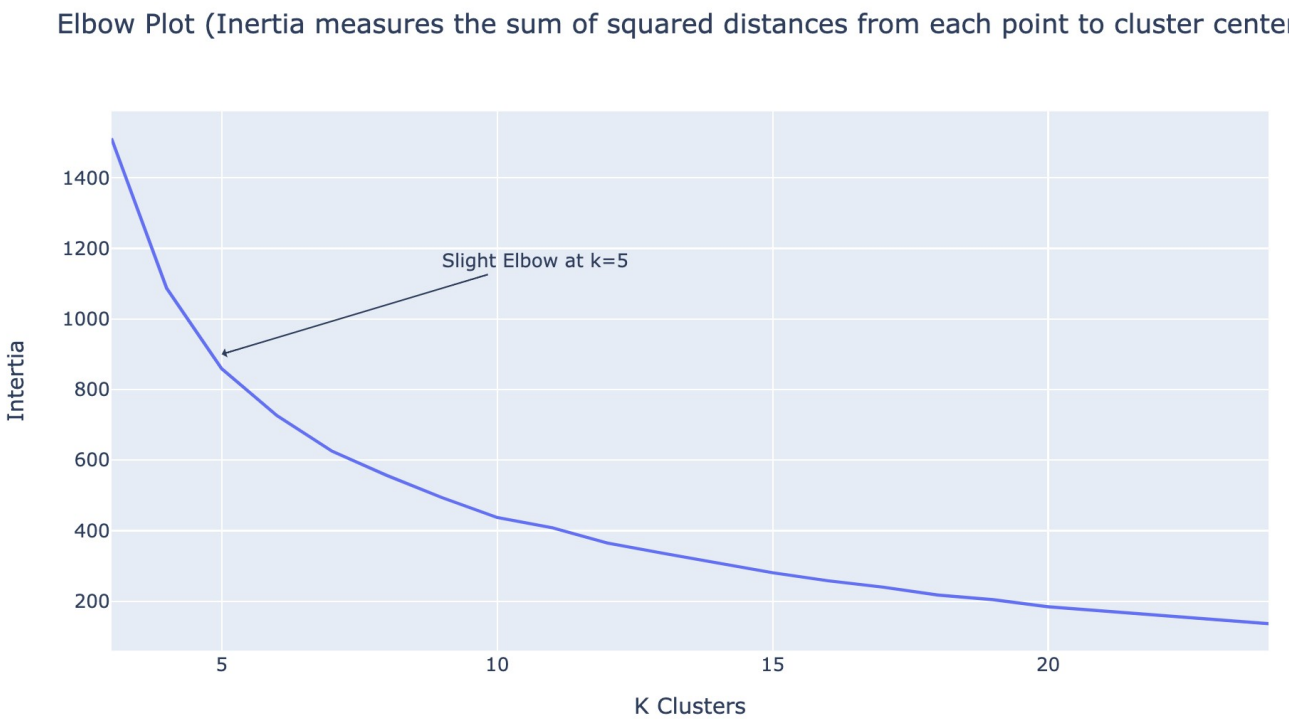
Data Preparation – PCA Dimensionality Reduction

- Address data sparsity with principle component analysis.
- Project the 310-dim data down to lower dimensions.
- Choose 3 dimensions based on explained variance plot.
- Plot the first three principle components to visualize the spread of the 106 neighborhoods before clustering.





Training the K-Means Model



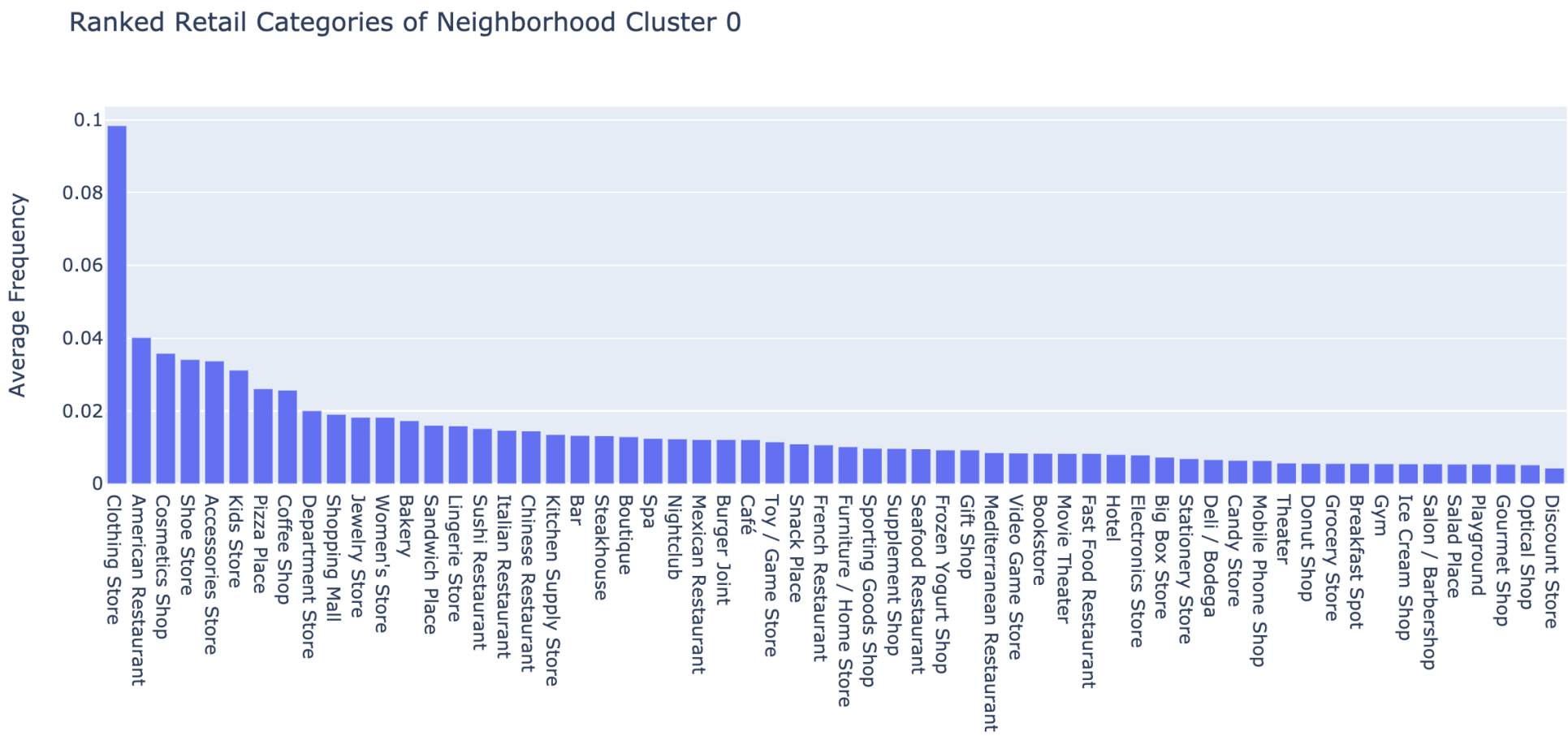
- **Choose K-Means clustering for efficiency and control over number of clusters.**
- **Choose k=5 clusters based on inertia measurements in elbow plot.**

K-Means Clustering Results

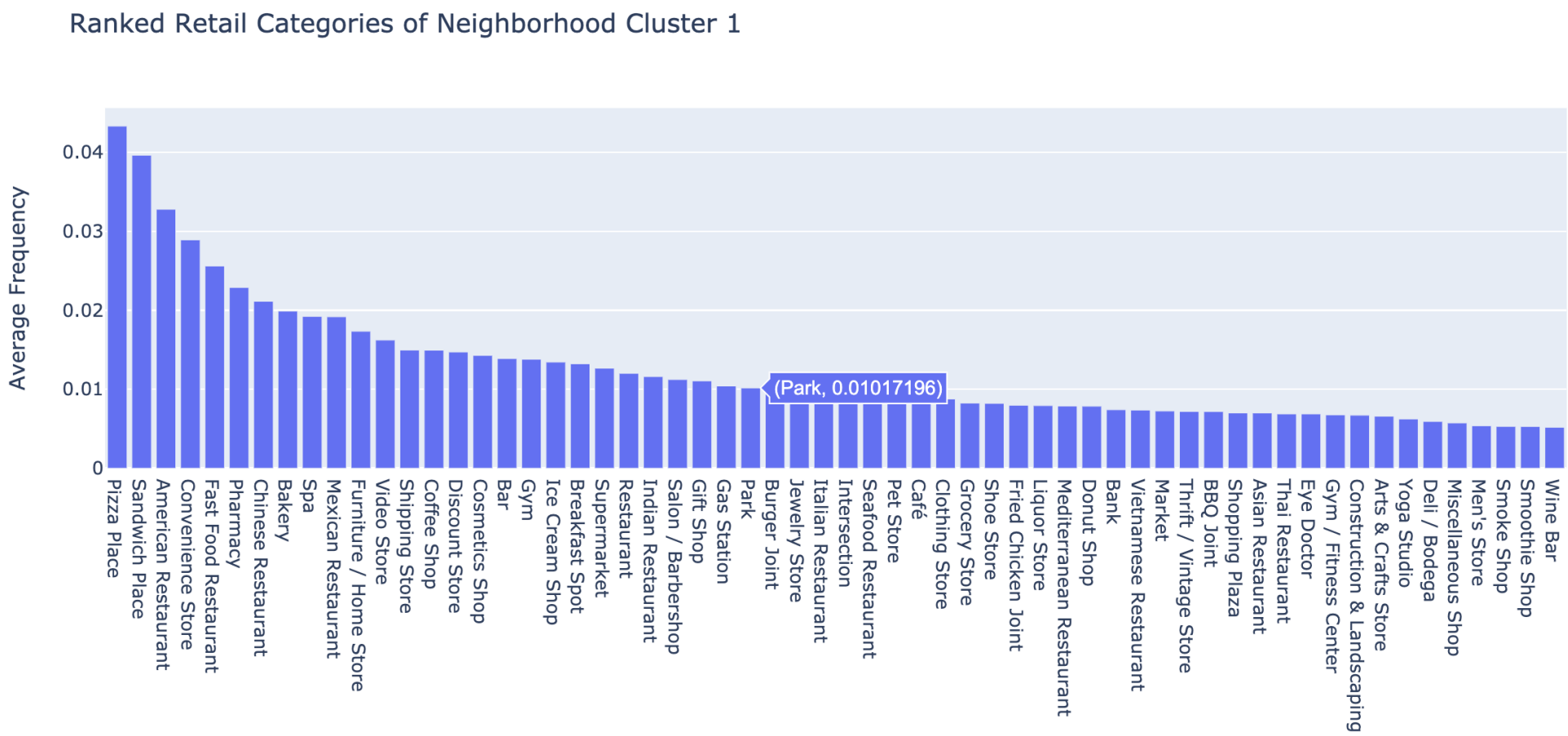
- **Church Hill gets assigned to cluster 3 along with 50 other neighborhoods.**
- **It is the largest and most central cluster in the 3D PCA plot.**
-

Cluster Label	Neighborhood
0	8
1	28
2	17
3	50
4	3

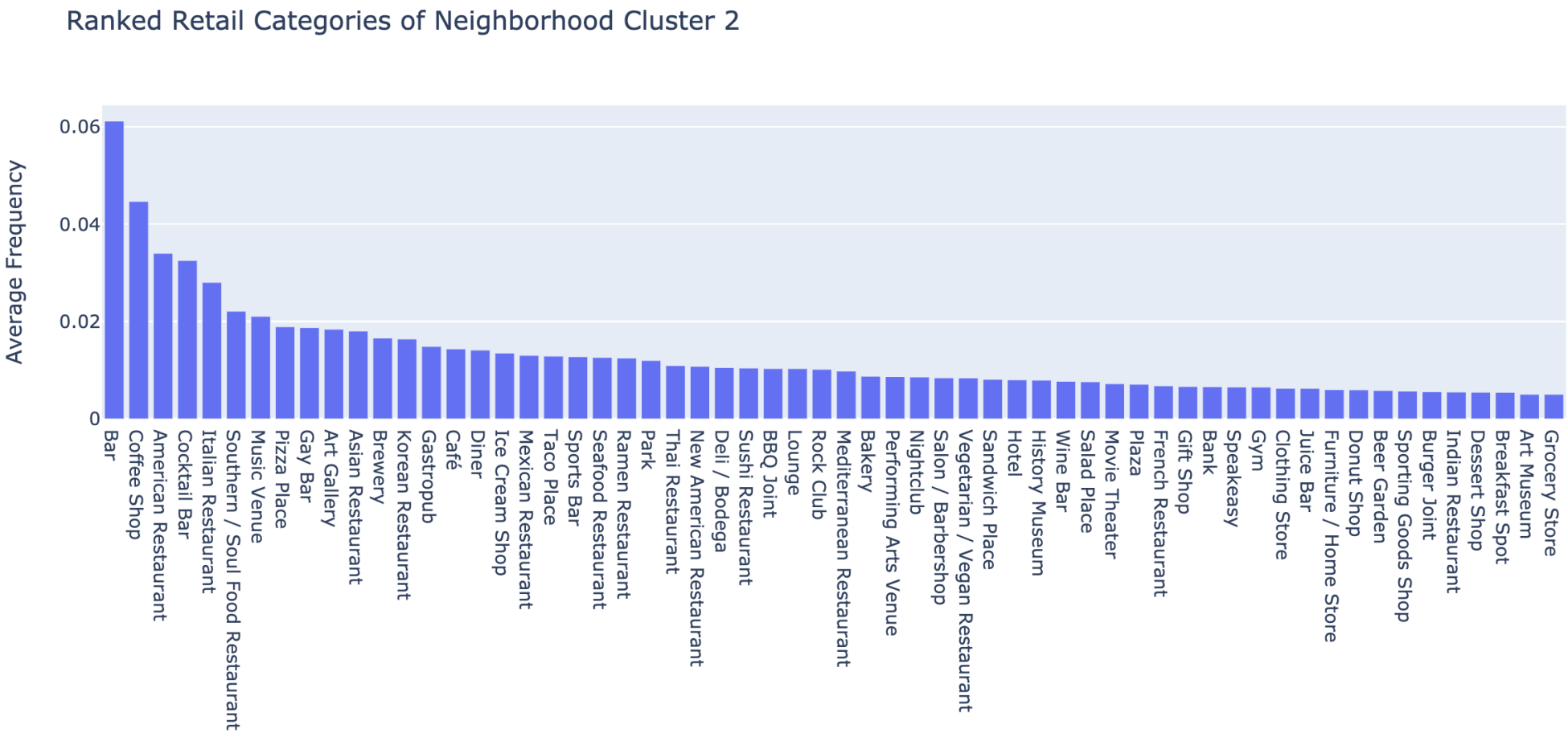
Analysis - Ranked Retail Categories of Cluster 0



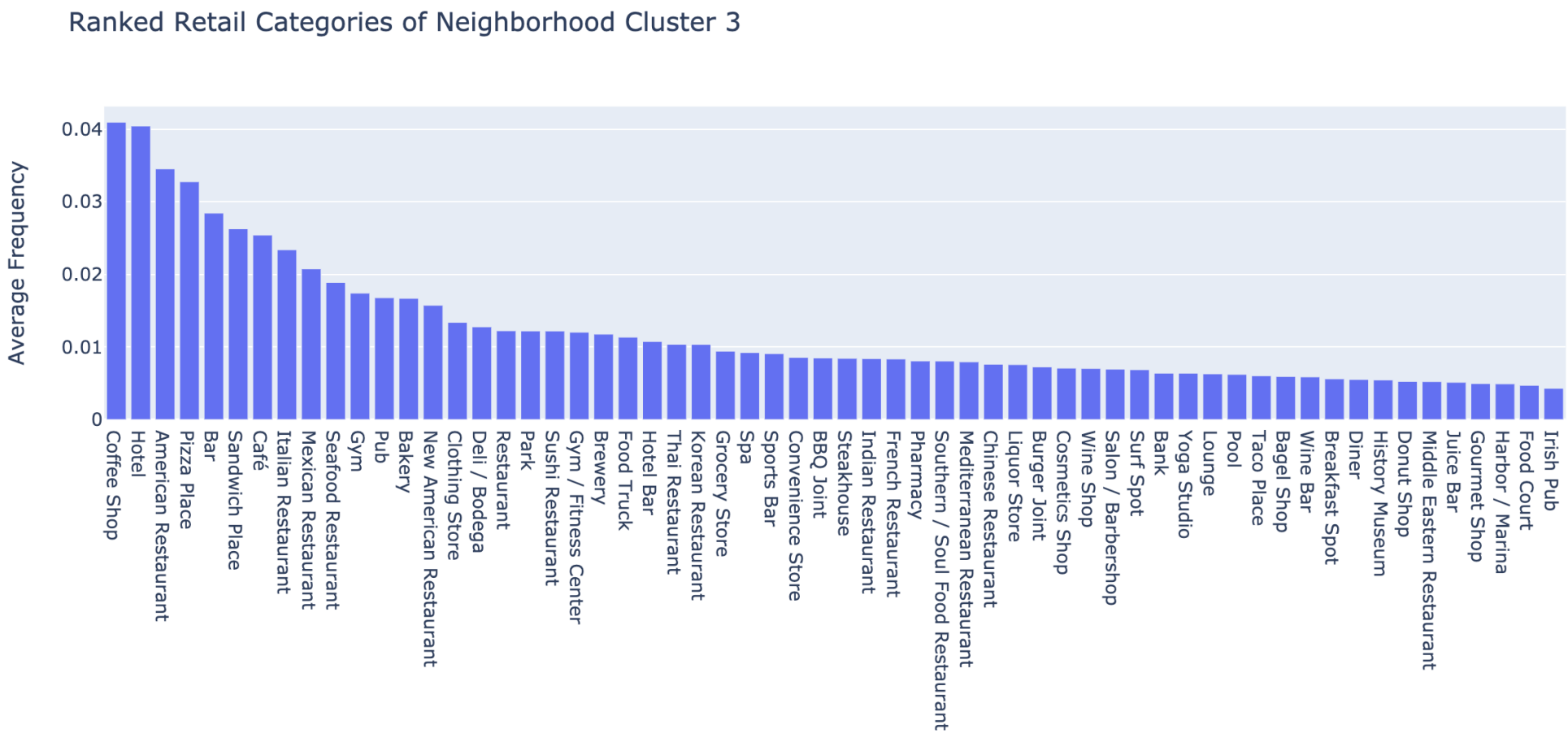
Analysis - Ranked Retail Categories of Cluster 1



Analysis - Ranked Retail Categories of Cluster 2

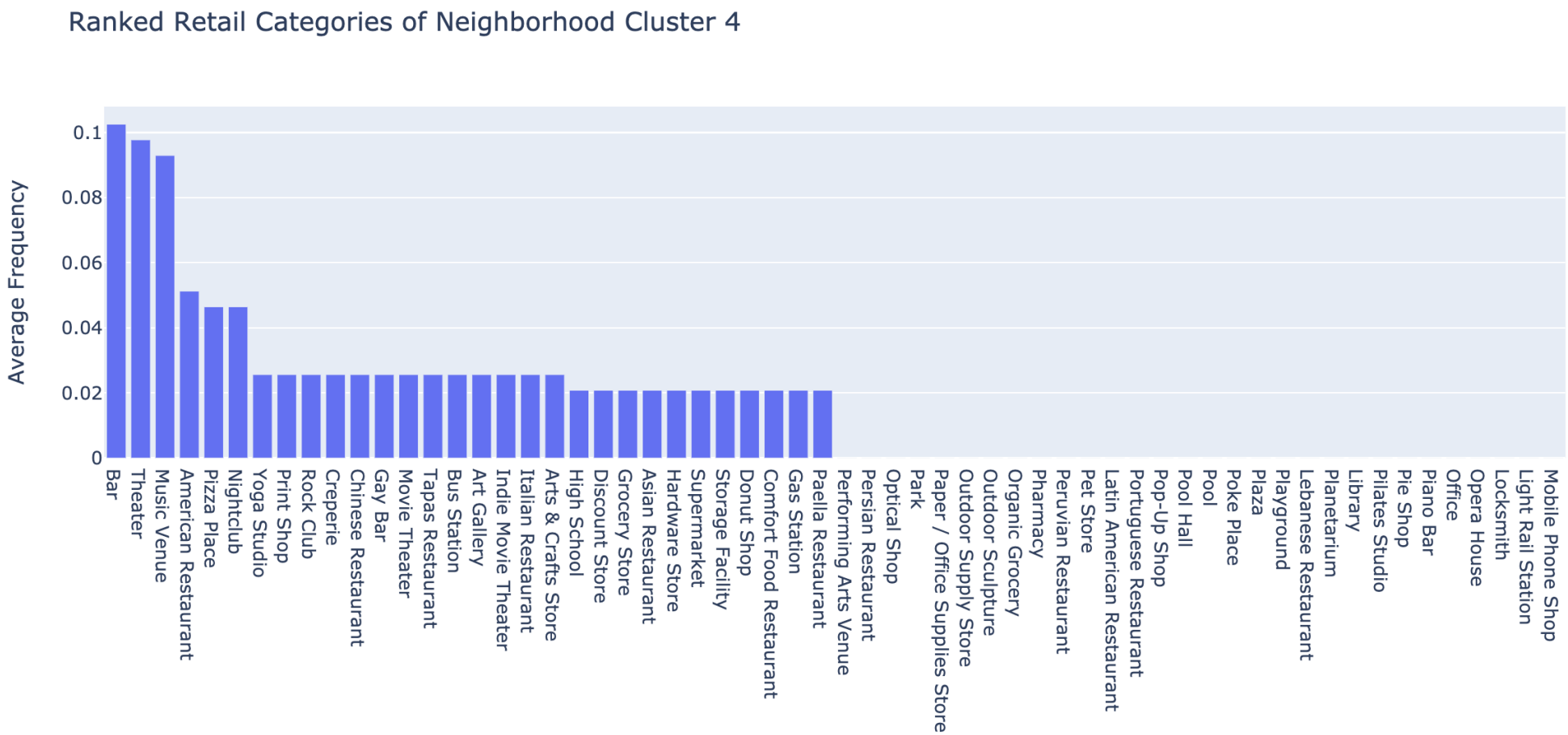


Analysis - Ranked Retail Categories of Cluster 3

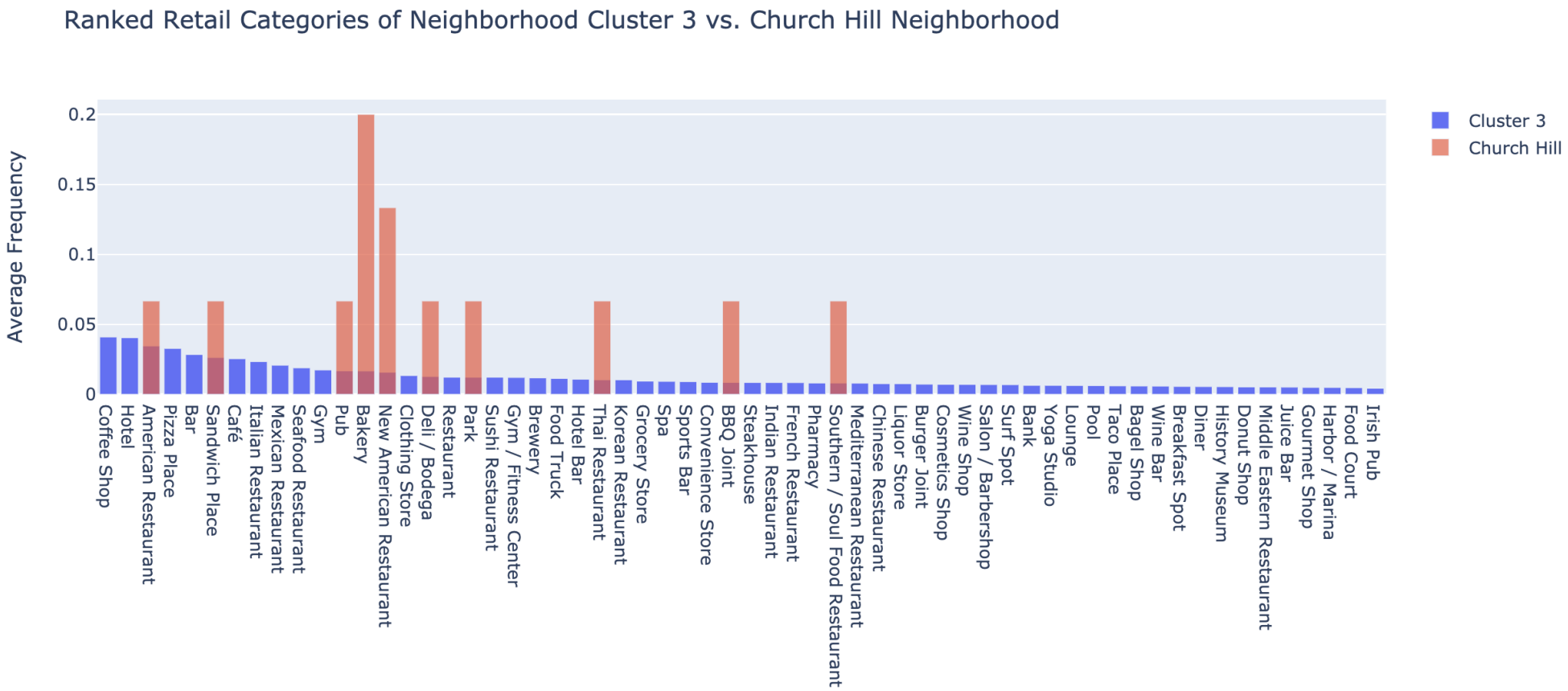


Church Hill belongs to this cluster so we will come back to this plot.

Analysis - Ranked Retail Categories of Cluster 4



Analysis- Church Hill Compared to its Cluster



Categories in blue without red bars are not represented in Church Hill.

Conclusion - Results for Church Hill

Top Retail Choices:

1) **Coffee shop**
2) **Hotel**
3) **Pizza place**
4) **Bar**
5) **Cafe**
6) **Italian Restaurant**
7) **Seafood Restaurant**
8) **Gym**
9) **Clothing store**
10) **Sushi restaurant**

Conclusion – Validity of Results

- **Some retail venues seem to be missing from Foursquare.**
- **Data sparsity made quality clustering difficult.**
- **From personal experience with Church Hill I can say these results makes sense.**

Conclusion – Further Work

- Try other clustering algorithms such as DBSCAN or entropy-weighted K-Means.

- Reduce data sparsity by:
- … adding more neighborhoods,
- … adding more data about the neighborhoods,
- … adding more finely tuned retail subcategory information