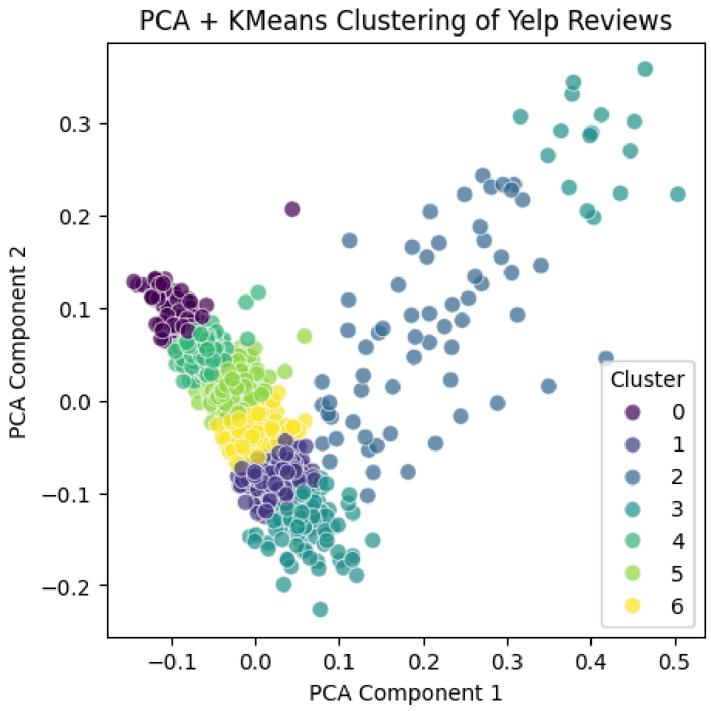

Our hypothesis is that using a baseline BM25 ranking algorithm, we could improve the recommendations that are generated by first grouping documents together using KMeans.

We started with a random sample of 1000 reviews from the Yelp dataset and ran BM25 on both the un-clustered and clustered data. The results that we see are below:

This is our input query:
`Food is great just wish it was bigger and you didnt have to call to make a reservation`

These are recommendations without clustering:

1. works like a strip bar...... we were greeted to sit down on a table of five… Good Luck!
2. I have no idea about the food or service, as the restaurant was closed for lunch on the day I had a reservation for a party of 12. … Do they not know about the longevity of restaurants on Market Street? Better step it up.
3. if i could give this hotel negative stars, i certainly would…. on top of that, it's an outdated hotel... clearly their reservation system reflects that as well.

These are recommendations with KMeans clustering (7 clusters):

1. Great concept! \n\nDefinitely worth the try !
2. Had the 1 meat plate [$10] Brisket was fresh & amazing. … Came right out with a huge serving I took a photo of the nice portion.
3. Easy walk in on a Friday night at 5pm. … They were hiding under the coffee table in the front of the shop. It was a nice quiet atmosphere.