

Tareas calificadas por los compañeros: Proyecto Capstone - La Batalla de los Vecindarios (Semana 2)

Introducción/Problema

El actual trabajo correspondiente al entrenamiento de ciencia de datos, consistirá en realizar un ejercicio geoespacial para ubicar las mejores zonas geográficas de un restaurante, o café de comida Americana en el país de la India. Puntualmente este proyecto está orientado a responder a las personas interesadas en invertir y desarrollar un modelo de negocio de mayor anuencia de público, basado en sus preferencias. Dada la complejidad del caso de estudio, buscaremos la forma de agrupar zonas que contengan la mayor cantidad de negocios del rubro de comidas u ocio (según sea el caso). Una de las principales prioridades es ubicar este tipo de comercios lo más cerca del centro de la ciudad. Buscaremos en lo posible presentar las mejores zonas de comidas las cuales servirán como propuesta para un grupo exclusivo de inversores. (Actualizado).

Selección de Datos

Para el desarrollo del presente proyecto, emplearemos: Pandas, Numpy y Geopy, este último con la cual obtendremos las coordenadas de diferentes ubicaciones. Folium para representar gráficamente la zona de vamos a segmentar. Y la API de Foursquare: para obtener la lista vanues y sus detalles alrededor de la mejor zona de ubicación geoespacial.

Metodología

Obtener las coordenadas de la ciudad objetivo.

Obtener la lista de barrios y sus coordenadas.

Exploración de los lugares más visitados en las localidades de destino.

Agrupación de las localidades.

Analizando los clusters formados.

Exploración

Exploramos con la creando de un mapa agregando marcadores para localidades, establecemos un marcador para Bangalore y marcadores de localidades. aplicamos el uso de la API de Foursquare para explorar las localidades y conseguir los 100 mejores recintos de cada localidad, con la conversión de la lista de lugares en un marco de datos, podemos identificar el número de sedes por cada Localidad que ofrecen comida americana.

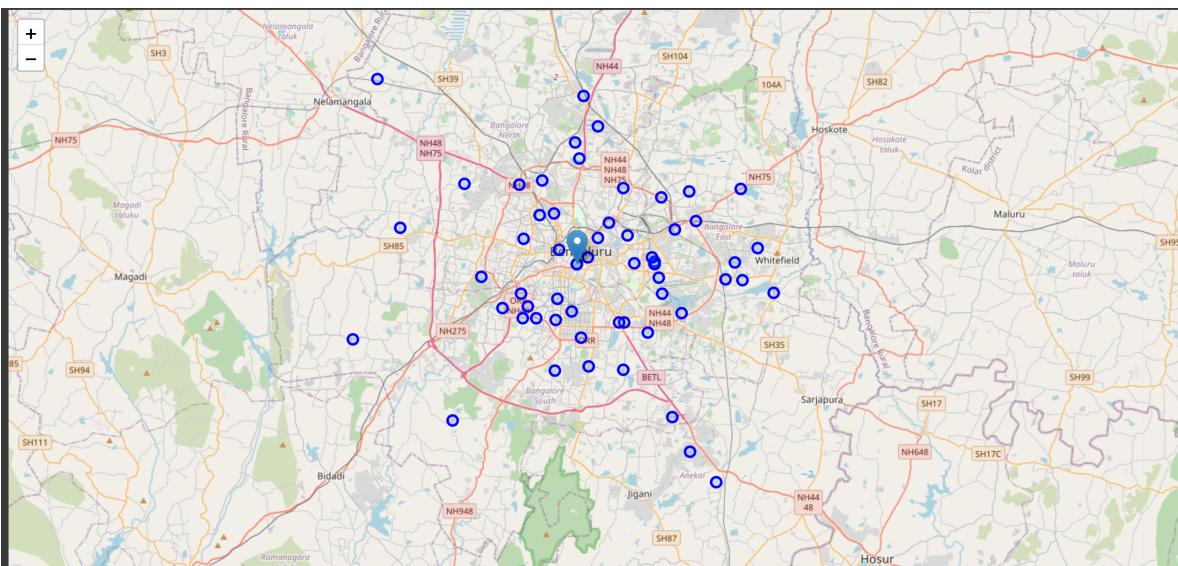
Identificamos El número total de categorías únicas de 206, y el número de localidades que tienen restaurantes americanos (12), identificando las zonas candidatas para elaborar la recomendación de inversión.

```
[21] 1 #Number of venues for each Locality  
2 venues_df.groupby(['Locality']).count()
```

	Latitude	Longitude	Venue name	Venue Lat	Venue Lng	Venue Category	
Locality							
Agara, Bangalore	1	1	1	1	1	1	
Arekere	81	81	81	81	81	81	
Banashankari	100	100	100	100	100	100	
Banaswadi	67	67	67	67	67	67	
Basavanagudi	100	100	100	100	100	100	
Basaveshwaranagar	57	57	57	57	57	57	
Begur, Bangalore	12	12	12	12	12	12	
Bellandur	78	78	78	78	78	78	
Bengaluru Pete	100	100	100	100	100	100	
Bidadi	100	100	100	100	100	100	
Bommasandra	8	8	8	8	8	8	
Chandapura	4	4	4	4	4	4	

▶ 1 neig df.head()

	Locality	Latitudes	Longitudes
0	Agara, Bangalore	12.842920	77.485820
1	Arekere	12.885670	77.596730
2	Banashankari	12.922310	77.569880
3	Banaswadi	13.019265	77.655603
4	Basavanagudi	12.938980	77.571370

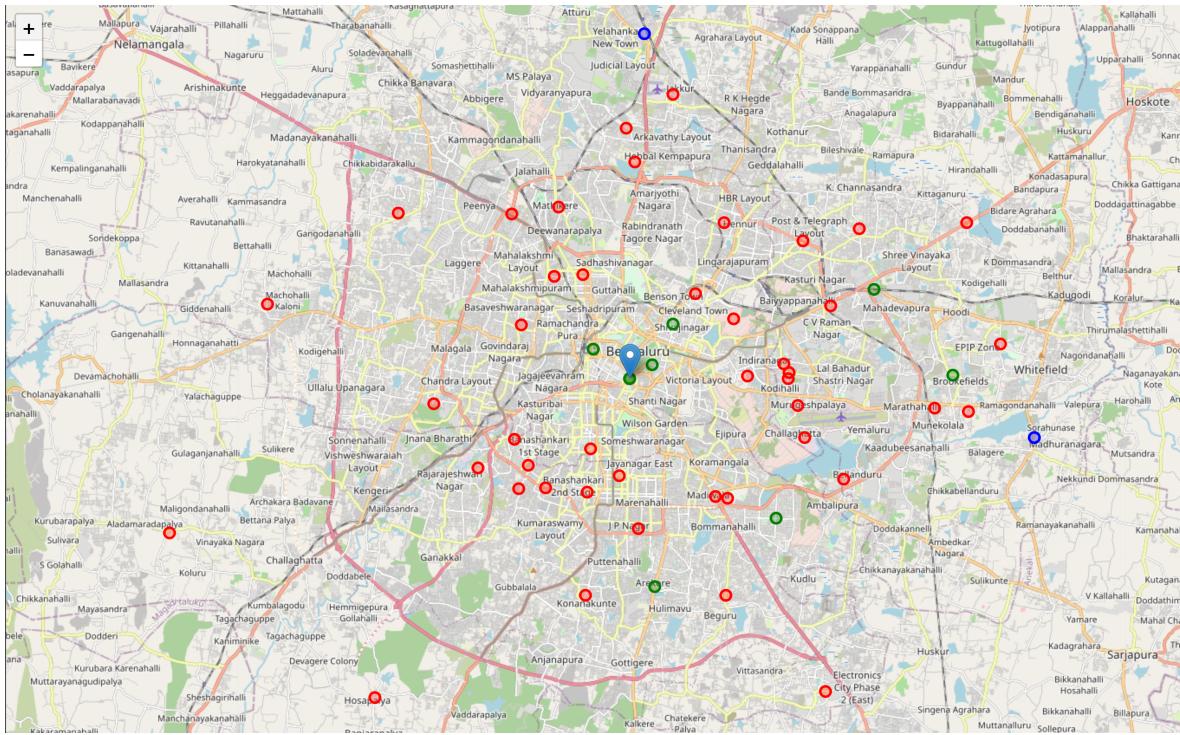


Modelo KNN

Ejecutamos el algoritmo Agrupamiento de K-medias sobre una marco de datos para el agrupamiento de los datos. Al ejecutar el K-means obtenemos las etiquetas para las primeras 10 localidades. Aquí es donde creamos un marco de datos de `blr_american`, procedemos a agregar etiqueta a `blr_labels` y logramos una fusión de los marcos de datos

blr_labels y neig_df para obtener la latitud y la longitud de cada una de las localidades. Agrupando las localidades de acuerdo a sus Etiquetas de cada Segmentación/Clúster, procedemos a Trazar el clúster en el mapa, establecemos el marcador para Bangalore y Obtener los colores para los grupos con sus respectivos marcadores de localidades. El algoritmo KNN, es un clasificador de aprendizaje supervisado, y se basa en establecer la cercanía de las clasificaciones o predicciones sobre la agrupación de un punto en el plano.

	Locality	American Restaurant	Cluster Label
0	Agara, Bangalore	0.000000	0
1	Arekere	0.012346	1
2	Banashankari	0.000000	0
3	Banaswadi	0.000000	0
4	Basavanagudi	0.000000	0



Conclusiones

Del análisis anterior podemos inferir que el grupo 1 (mostrado en color rojo) casi no tiene restaurantes americanos con los números más altos del mismo en el grupo 2 (mostrado en color verde) y un número moderado de restaurantes americanos están presentes en el grupo 3 (mostrado con color azul) ubicado en la parte central de la ciudad.

Este análisis presenta una gran oportunidad para que los empresarios aprovechen el potencial no utilizado de las zonas exteriores de la ciudad de Bangalore mediante la apertura de restaurantes americanos.

También es evidente que el grupo 2 (alrededor de la parte central de la ciudad) está sufriendo de alta competencia y exceso de oferta, por lo que los desarrolladores deben evitar invertir mucho en esta área.

Los desarrolladores con propuestas de venta únicas que pueden destacarse de la competencia moderada en el clúster 3 pueden asumir un riesgo moderado y atraer a los clientes que ya visitan la localidad de este clúster debido al restaurante americano existente.