

Planifiquemos un Viaje Usando Clustering y Google Maps API

Palabras clave: clustering, aprendizaje no supervisado, Google Maps API

Autor: Juan De Dios Santos Rivera

Abstract

En esta charla adoptaremos el papel de planificador de viajes para mostrar un ejemplo de uso de algoritmos de clustering para organizar un viaje a San Juan, Puerto Rico. Usaremos k-medoids, un algoritmo de clustering de aprendizaje no supervisado implementado en la librería `cluster` para agrupar las localizaciones que deseamos visitar, junto a las librerías `httr` para comunicarnos con la API de Google Maps, y `leaflet` para visualizar los clusters y localizaciones en un mapa.

Comenzaremos mostrando como aplicar k-medoids con las métricas de distancia Euclíadiana y Haversine y discutiremos sus limitaciones en escenarios reales de viaje. Luego, integraremos las distancias reales obtenidas con la API de Google Maps para mejorar nuestro modelo de clustering. La presentación cubrirá como preparar y recopilar los datos de lugares, utilizar `httr` para realizar solicitudes a la API de Google Maps y obtener distancias reales de conducción, aplicar k-medoids con las distancias obtenidas, y como crear un mapa interactivo con `leaflet` que muestra las localizaciones agrupadas.

Como ejemplo, suponga que tiene un dataset de coordenadas de varios lugares que visité durante unas vacaciones a San Juan, Puerto Rico. Entrenar un modelo de clustering con este dataset utilizando distancia Euclíadiana (Figura 1) y otro modelo con la distancia de conducción obtenida con Google Maps (Figura 2), produce dos resultados distintos debido a que la distancia Euclíadiana no considera la ruta que hay que seguir para llegar de un punto a otro, ni otros elementos como la curvatura de la Tierra.

Clustered Locations in San Juan, Puerto Rico

Using Euclidean distance as similarity

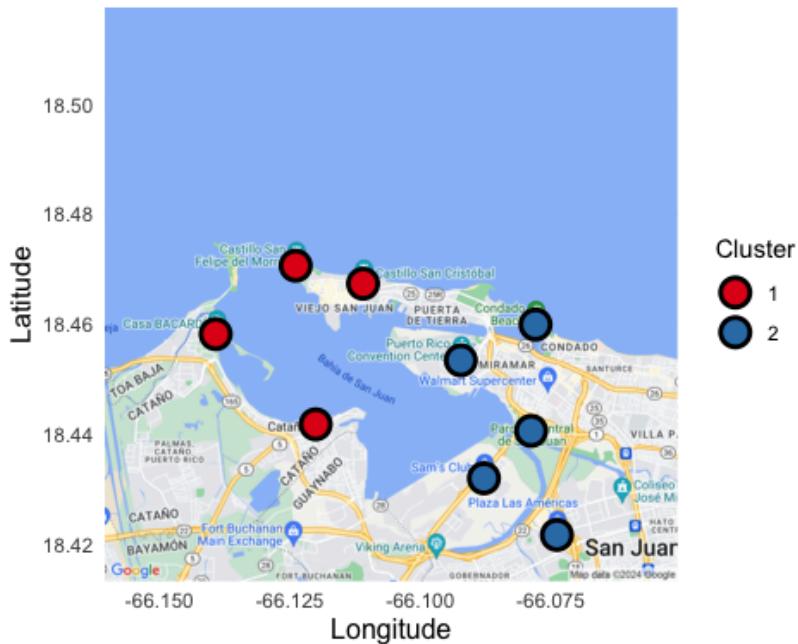


Figura 1

Clustered Locations in San Juan, Puerto Rico

Clusters Based on Driving Distances

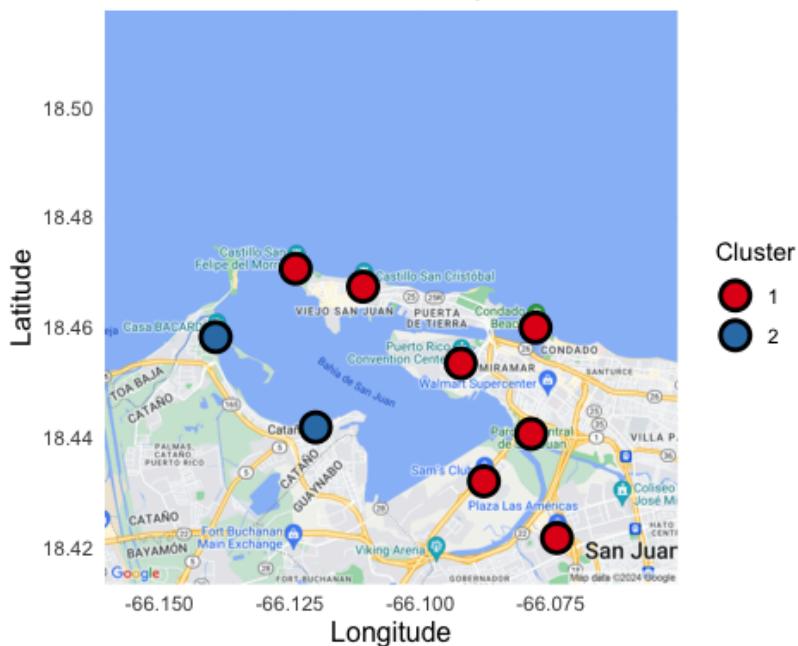


Figura 2