

Tarea: Sistemas de Recomendación

Prof Luis Eduardo Falcón

- La siguiente Tarea se puede hacer de manera individual o en equipos de máximo 3 integrantes.
 - Esta Tarea requiere descargar la siguiente información:
 - De la siguiente liga de la UCI, debes descargar el archivo RCdata.zip y utilizar los siguientes archivos y variables:
 - **rating_final.csv** : Usar únicamente las variables **userID**, **placeID** y **rating**.
 - **geoplaces2.csv** : Usar las variables **placeID**, **name**, **city**, **latitude** y **longitude**.
 - **chefmozcuisine.csv** : Usar las variables **placeID** y **RCuisine**.
- <https://archive.ics.uci.edu/ml/datasets/Restaurant+%26+consumer+data>
1. De los archivos anteriores generar un solo data frame que contenga todas las variables indicadas al inicio.
 2. Realizar un análisis descriptivo de las variables e indicar si existen datos perdidos en algunas de ellas. De ser así, realiza un análisis para determinar y justificar la decisión que tomes sobre dichos datos perdidos.

Para el caso de los datos perdidos de la variable city, puedes investigar al menos los paquetes “revgeo” o “ggmap” de R, para recuperar el nombre de la ciudad con la latitud y longitud.
 3. ¿De qué ciudades son los restaurantes del estudio?
 4. Si consideramos la popularidad de un lugar como aquellos que tienen la mayor cantidad de evaluaciones por parte de los usuarios (independientemente de si fue positiva o negativa la evaluación), obtener los nombres de los 10 restaurantes más evaluados/populares. ¿Cuántos restaurantes diferentes hay en total?
 5. ¿De qué tipo de comida/cocina son los 10 restaurantes más populares encontrados en el inciso anterior? ¿Cuántos tipos de cocina diferentes hay en total?
 6. Generar la matriz de Utilidad considerando los renglones con la variable userID, las columnas con placeID, y los valores de la matriz con rating. A partir de dicha matriz de utilidad aplicar la factorización SVD (Singular Value Decomposition) para obtener la matriz de variables latentes para los restaurantes.
 7. Si a un usuario le gustó el restaurante “Gorditas Doña Gloria” con placeID: 132834, ¿Qué otros 12 restaurantes (indicar los nombres) le podrías recomendar usando la descomposición SVD con el método de correlación de “pearson” y considerando los 10 valores singulares más grandes de la SVD? Nota: no importa por el momento de qué ciudad sea el restaurante.