



Fundamentos de Programación

Ciclo 1 - Reto 5

Descripción del problema:

Una consultora financiera a estado almacenando información relacionada con la puntuación dada por sus usuarios ha diferentes negocios. Este ejercicio lo ha venido realizando con el fin de construir un sistema de recomendación para sus clientes.

Ellos han estado consultando que podrían hacer con la información disponible, en su búsqueda encontraron que pueden tener un sistema de recomendación tipo filtro colaborativo-contenido en el que pueden usar las puntuaciones de los usuarios respecto a los negocios que frecuentan y las categorías a los que esos negocios pertenecen.

Usted ha sido contratado para desarrollar este sistema. Con las categorías de los negocios deberá construir una matriz de peso, a partir de esta determinará el perfil del usuario con el que podrá evaluar una matriz de negocios candidatos y obtener una ponderación desde la que determinar que negocios recomendar.

Los pasos a seguir son:

1. Construir una matriz de categorías. Esta matriz tiene como filas las categorías comunes a los negocios frecuentados, y como columnas el identificador del negocio (secuencia alfanumérica), el interior de la matriz esta compuesto de ceros, se tendrá el valor 1 para las celdas que correspondan a la categoría a la que pertenece el negocio, ver Tabla 1.
2. Multiplicar cada una de las columnas de la matriz de categorías por la puntuación dada por el usuario para los negocios, esta es la matriz de pesos, ver Tabla 2.
3. Determinar el perfil de usuario. Se obtiene para cada uno de los géneros un perfil sumando cada una de las filas y dividiendo ese valor entre la puntuación total de los negocios (la cual es la suma de todos los valores de la matriz de pesos), ver Tabla 2.
4. Construir la matriz candidata, igual que la matriz de categorías pero para los negocios sin visitar,



ver Tabla 3.

5. Cada una de las filas de la matriz candidata se multiplica por su respectivo perfil obtenido en el paso 3 (si no hay perfil para la categoría se multiplica por 0), ver Tabla 4.
6. Se obtiene la ponderación para cada uno de los negocios, sumando los valores de sus respectivas columnas, ver Tabla 4.

La ponderación hace referencia al interés del usuario de visitar un nuevo negocio. Para el ejemplo se obtiene que al cliente se le pueden recomendar en orden de mayor interés a menor interés los negocios: 1qCuOcks5HRv67OHovAVpg (0.53846), MKYk4F4HSzHF8v-4cYe3Ww (0.47692), McikHxxEqZ2X0joaRNKlaw (0.33077)... Tener en cuenta que en el *json* de salida se redondean las ponderaciones, ver Tabla 7.

La matriz de categorías puede tener una dimensión diferente a la de la matriz candidata. Debe retornar un json indentado a cuatro espacios, con todos los negocios de la matriz candidata y sus respectivas ponderaciones, ordenados por ponderación de mayor a menor y excluyendo los que tengan ponderación 0.

Para cierto usuario se tienen las siguientes puntuaciones:

yXuao0pFz1AxB21vJjDf5w 5
fNGIbpazjTRdXgwRY_NIXA8
b9WZJp5L1RZr4F1nxclOoQ8
zaXDakTd3RXyOa7sMrUE1g6
zgy27FSnvwdINfk5cXBIyQ4
WETE_LykpcnrC1sFcQ5EGg8
6o3RK6rTcN3nw - j - r2nQmA6



Esqueleto: Úselo como base para el desarrollo de su solución. Recuerde que es obligación usar la función *recomiendaNegocio* con sus respectivos parámetros. Las otras funciones son una recomendación.

```
import json
import pandas as pd
import numpy as np

def preProcesado(DF):
    # Funcion que prepara los generos en un DF para ser codificadas en la matriz
    # Devuelve tambien los generos en una lista
    categoriasDF = DF['categories'].apply(pd.Series)
    categoriasDF['business_id'] = DF['business_id']
    categoriasDF = categoriasDF.drop_duplicates(['business_id'])
    categoriasDF.set_index('business_id', inplace=True)

    categories = [categoriasDF[categorie].unique() for categorie in categoriasDF.
                   columns]
    categories = [categorie for lista in categories for categorie in lista if all
                  ([pd.isnull(categorie) == False, categorie != ' ', categorie != '', len(
                    str(categorie)) > 1])]
    categories = list(set(categories))

    return categoriasDF, categories

def codificaMatriz(DF, genres: list, producto: list):
    # Funcion que inserta unos en la matriz
    # Su codigo aqui
    return #Debe retornar un DF como el de la Tabla 1.

def recomiendaNegocio(url_puntuacion:str, url_perfil:str, url_recomendacion:str)
->json:
    # Su codigo aqui
```

Nota: Debe cargar solo las *funciones* y los *import*, con el mismo nombre dado, recibiendo los mismos parámetros y respetando la estructura de salida. No debe redondear valores de cálculos intermedios, ni usar *input()* y *print()*, todos los datos se debe pasar como argumentos a las funciones. La separación para el csv es ';'. Usar *round* a 5 cifras para las ponderaciones de salida.



Tabla 1. Matriz de Categorías

	yXua0pFz1Ax2B1vJjDf5w	fNGHlpsazjTRdXgwyRY_NIXA	l6WZJp5L1RZz4F1nxcI0oQ	zaXDakT3RXY0a7sMtUE1g	zgy27FSuwwdINRk5cXBlyQ	WETE_LykpencrC1sFcQ5EGg	6o3RK6rTcN3aw-j-r2uQmA
Lounges	0	1	0	0	0	0	0
American (Traditional)	0	1	0	0	0	0	0
Cafes	0	0	0	1	0	0	0
Irish	0	0	0	0	0	1	0
Bars	0	1	0	0	0	1	0
Pubs	0	0	0	0	0	1	0
Grocery	1	0	0	0	0	0	0
Restaurants	0	1	1	1	0	1	0
Event Planning & Services	0	0	0	0	1	0	0
Breakfast & Brunch	0	0	1	0	0	0	0
Nightlife	0	1	0	0	0	1	0
Hotels	0	0	0	0	1	0	0
Food	1	0	0	0	0	0	0
Hotels & Travel	0	0	0	0	1	0	0
Puntuacion	5	8	8	6	4	8	6

Tabla 2. Matriz de Pesos y Perfil de Usuario

	yXua0pFz1Ax2B1vJjDf5w	fNGHlpsazjTRdXgwyRY_NIXA	l6WZJp5L1RZz4F1nxcI0oQ	zaXDakT3RXY0a7sMtUE1g	zgy27FSuwwdINRk5cXBlyQ	WETE_LykpencrC1sFcQ5EGg	6o3RK6rTcN3aw-j-r2uQmA	Perfil de Usuario
Lounges	0	8	0	0	0	0	0	0.061538
American (Traditional)	0	8	0	0	0	0	0	0.061538
Cafes	0	0	0	6	0	0	0	0.046154
Irish	0	0	0	0	0	8	0	0.061538
Bars	0	8	0	0	0	8	0	0.123077
Pubs	0	0	0	0	0	8	0	0.061538
Grocery	5	0	0	0	0	0	0	0.038462
Restaurants	0	8	8	6	0	8	0	0.230769
Event Planning & Services	0	0	0	0	4	0	0	0.030769
Breakfast & Brunch	0	0	8	0	0	0	0	0.061538
Nightlife	0	8	0	0	0	8	0	0.123077
Hotels	0	0	0	0	4	0	0	0.030769
Food	5	0	0	0	0	0	0	0.038462
Hotels & Travel	0	0	0	0	4	0	0	0.030769

Tabla 3. Matriz Candidata

	5TH9W0RazJz4kBuHlpsaA	MRKt4F0R5dF9w-4YcFWw	0aB2wQZ2VZvZWYywxXA	gC8SpwclagZBRCabzDly	lgC5a5b5BHCv0H8wAVpg	z0W8qJp4EHBZB8S-2Sew	Md4HxclqZ2XjpaRNK3aw	v6b54aaA1YDbaYygb6g	e7C3XGz4dJLbaa0Vfow	HweEhZKLN0Uc9pwekgQ	Perfil de Usuario
Restaurants	0	1	0	0	1	1	1	0	0	0	0.230769
Home & Garden	0	0	0	1	0	0	0	0	0	0	0.000000
Body Shape	0	0	1	0	0	0	0	0	0	0	0.000000
Department Stores	1	0	0	0	0	0	0	0	0	0	0.000000
Arts& Crafts	0	0	0	0	0	1	0	1	0	0	0.000000
Beveries	0	0	0	0	0	0	1	0	0	0	0.000000
Numerics & Computing	0	0	0	1	0	0	0	0	0	0	0.000000
Books, Maps, Music & Video	0	0	0	0	0	0	0	0	1	1	0.000000
Fashion	1	0	0	0	0	0	0	0	0	0	0.000000
Comedy Clubs	0	0	0	0	0	0	0	1	0	0	0.000000
Food	0	0	0	0	0	0	1	0	1	0	0.038462
Automotive	0	0	1	0	0	0	0	0	0	0	0.000000
Pubs	0	0	0	0	1	0	0	0	0	0	0.061538
Bars	0	1	0	0	1	0	0	0	0	0	0.123077
Hardware Stores	0	0	0	0	1	0	0	0	0	0	0.000000
Americas (New)	0	0	0	0	1	0	0	0	0	0	0.000000
Bookstores	0	0	0	0	0	0	0	0	0	1	0.000000
Building Supplies	0	0	0	1	0	0	0	0	0	0	0.000000
Shipping	1	0	0	1	0	0	0	0	1	1	0.000000
Grocery	0	0	0	0	0	0	0	0	1	0	0.038462
Video & Video Game Rental	0	0	0	0	0	0	0	0	1	0	0.000000
Nightlife	0	1	0	0	1	0	0	1	0	0	0.123077
Home Services	0	0	0	1	0	0	0	0	0	0	0.000000
Arts & Entertainment	0	0	0	0	0	1	0	0	0	0	0.000000
American (Traditional)	0	0	0	0	0	1	1	0	0	0	0.061538

Tabla 4. Matriz Candidata Perfilada y Ponderaciones

	5TH9W0RazJz4kBuHlpsaA	MRKt4F0R5dF9w-4YcFWw	0aB2wQZ2VZvZWYywxXA	gC8SpwclagZBRCabzDly	lgC5a5b5BHCv0H8wAVpg	z0W8qJp4EHBZB8S-2Sew	Md4HxclqZ2XjpaRNK3aw	v6b54aaA1YDbaYygb6g	e7C3XGz4dJLbaa0Vfow	HweEhZKLN0Uc9pwekgQ	Perfil de Usuario
Restaurants	0.000000	0.230769	0.000000	0.000000	0.230769	0.230769	0.000000	0.000000	0.000000	0.000000	0.230769
Home & Garden	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Body Shape	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Department Stores	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Arts& Crafts	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Beveries	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Numerics & Computing	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Books, Maps, Music & Video	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Fashion	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Comedy Clubs	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Food	0.000000	0.000000	0.000000	0.000000	0.000000	0.038462	0.000000	0.038462	0.000000	0.000000	0.038462
Automotive	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Pubs	0.000000	0.000000	0.000000	0.000000	0.061538	0.000000	0.000000	0.000000	0.000000	0.000000	0.061538
Bars	0.000000	0.123077	0.000000	0.000000	0.123077	0.000000	0.000000	0.000000	0.000000	0.000000	0.123077
Hardware Stores	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Americas (New)	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Bookstores	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Building Supplies	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Shipping	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Grocery	0.000000	0.000000	0.000000	0.000000	0.000000	0.038462	0.000000	0.038462	0.000000	0.000000	0.038462
Video & Video Game Rental	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Nightlife	0.000000	0.123077	0.000000	0.000000	0.123077	0.000000	0.000000	0.000000	0.000000	0.000000	0.123077
Home Services	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Arts & Entertainment	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
American (Traditional)	0.000000	0.000000	0.000000	0.000000	0.061538	0.061538	0.000000	0.000000	0.000000	0.000000	0.061538
Puntuacion	0.0000	0.4702	0.0000	0.0000	0.3846	0.2921	0.3307	0.1238	0.0702	0.0000	



Tabla 5. Valores de entrada

Entrada	Tipo	Descripción
url_puntuacion	str	url al archivo csv que contienen las puntuaciones dadas por el usuario
url_perfil	str	url al archivo json con los negocios puntuados y sus respectivas categorías
url_recomendacion	str	url al archivo json con los negocios candidatos y sus respectivas categorías

Tabla 6. Tabla de salida (retorno)

Tipo	Descripción
json	Documento json con el id de los negocios como llave y ponderación como valor. Los valores se encuentran ordenados por ponderación, de mayor a menor. Se omiten los negocios con ponderación 0

Tabla 7. Entradas y Salidas

Entradas			Salida
url_puntuacion	url_perfil	url_recomendacion	return

https://git.io/JZKue	https://git.io/JZK0H	https://git.io/JZKE4	<pre> { "1qCuOcks5HRv67OHovAVpg": 0.53846, "MKyk4F4HSzHF8v-4cYe3Ww": 0.47692, "McikHxxEqZ2X0joaRNKlaw": 0.33077, "sbW8qHJgzEI42B0S-3New": 0.29231, "tv8cS4aaA1VDaInYgggb6g": 0.12308, "eT5Ck7Gg1dBJobca9VFovw": 0.07692 } { "wqu7ILomIOPSduRwoWp4AQ": 0.38, "MKyk4F4HSzHF8v-4cYe3Ww": 0.29, "1qCuOcks5HRv67OHovAVpg": 0.29, "PK6aSizckHFWk8i0xt5DA": 0.23, "6ilJq_05xRgek_8qUp36-g": 0.23, "SQ0j7bgSTazkVQlF5AnqyQ": 0.15, "P1fJb2WQ1mXoiudj8UE44w": 0.15, "sRqB6flj3GtTZIZJQxf.oA": 0.15, "1u2NauOy8xDj7aFQET680A": 0.07, "0hrB2iwQZ52VZcZWYywxrA": 0.06, "3gmBc0qN.LtGbZAJtHWZg": 0.05, "gClSqcwuLsgyZRCzhJc7Dg": 0.02 } </pre>
https://git.io/JZKuk	https://git.io/JZK0A	https://git.io/JZKEg	