

# Examen de Procesamiento Mod2

Certificación en procesamiento con Scala

México, Junio 2022



## Instrucciones

De acuerdo a los temas vistos en el Módulo 2 del Curso de Certificación en Procesamiento con Scala, desarrolla la solución para el ejercicio que se solicita a continuación:

Como parte del Festival de Gastronomía de la CDMX, hemos realizado una encuesta a cierto número de personas, de las cuales vamos a elegir un subgrupo que recibirá un tour completamente gratis para conocer las comidas más bizarras de toda la República Mexicana.





La **rama** a utilizar para este examen se llama: **insumosExamen**

Las tablas proporcionadas son las siguientes:

- **t\_kdit\_food**: Contiene la lista de algunos platillos de la República Mexicana que son considerados raros.
- **t\_kdit\_foodie**: Contiene la lista de las personas a las que se les hizo la encuesta sobre sus preferencias gastronómicas.

Los schemas de las tablas son:

- **t\_kdit\_food.input.schema**: Contiene el schema de la tabla t\_kdit\_food.
- **t\_kdit\_foodie.input.schema**: Contiene el schema de la tabla t\_kdit\_foodie.

Para elegir a los ganadores tenemos las siguientes reglas que deberán implementarse sobre los datasets, tú puedes crear los métodos que creas conveniente pero al menos **estos 4 los debes respetar con la firma** que se incluye a continuación:

1. De la tabla `t_kdit_food` quédate únicamente con los platillos que **NO** tengan los siguientes niveles de picor: **1,2,3,5,6,7,8,9**. Utiliza una secuencia para almacenar estos valores y filtra con ella la tabla. Este dataset se utiliza en el **paso 3**.

```
spicyFilter(food: DataFrame): DataFrame
```

2. Del dataset de Foodie, utilizando **Window Functions** devuelve dos registros por nacionalidad que contengan las mujeres que consumen más picante, para esto auxiliate del campo **ROW\_NUMBER**, el cual debes conservar. Este dataset es el que utilizaremos en el **paso 4**.

```
justWomen(foodie: DataFrame): DataFrame
```

3. Regresa una cadena que contenga los códigos y los nombres (**CODE: NAME**) de los registros de la tabla Food que filtramos en el **paso 1** separados por comas. Cuando el platillo no contenga productos de origen animal (**CONTAINS\_ANIMAL\_PRODUCTS = N**) deberá aparecer un asterisco entre paréntesis: (\*).

Ej: "01: torta de gelatina(\*), 10: tascalate"

```
listOfFood(spicyFilter: DataFrame): String
```

4. Genera el campo "FOOD\_LIST" que almacenará la cadena resultante en el **paso 3**, este campo debe ser agregado al dataset de Foodie que resultó del **paso 2**.

```
concatFood(foodie: DataFrame, food: DataFrame): DataFrame
```

5. El resultado final debe cuadrar con el esquema de salida: `t_kdit_report1.output.schema`.

## NOTAS IMPORTANTES:

1. Los archivos los encontrarás en la carpeta `src/test/resources/`
  - a. Los insumos en `input/`
  - b. Los schemas en `schemas/`
2. El archivo a usar para parametrizar el proceso debes crearlo y colocarlo en `src/main/resources/config` y `src/test/resources/config`, respectivamente.
3. La partición de salida deberá ser por el campo "ROW\_NUMBER".

## PRUEBAS UNITARIAS

Elige 3 métodos y crea su correspondiente prueba unitaria del caso de ÉXITO únicamente, recuerda que la prueba unitaria debe comprobar el correcto funcionamiento del método,

NO VALIDES ÚNICAMENTE EL NÚMERO DE REGISTROS, corrobora que se cumple la regla de negocio aplicada o no será válida tu prueba.

## PRUEBAS DE ACEPTACIÓN

Realiza las comprobaciones que creas pertinentes para que el negocio apruebe tu desarrollo.

Toma en consideración que el resultado esperado debe ser el siguiente:

SPICY_TOLERANCE	NAME	PREFERENCES	NATIONALITY	GENRE	ROW_NUMBER	FOOD_LIST	
8	GERTRUDIS	Healthy Food	MOROCCAN	FEMALE	1	18: pipian, 26: pollo al cacahuate, 28: chilaquites más picantes del mundo(*)	
7	ANITA	Healthy Food	GERMAN	FEMALE	1	18: pipian, 26: pollo al cacahuate, 28: chilaquites más picantes del mundo(*)	
10	GABRIELA	Street Food	SPANISH	FEMALE	1	18: pipian, 26: pollo al cacahuate, 28: chilaquites más picantes del mundo(*)	
5	LETICIA	Raw Vegan Food	SPANISH	FEMALE	2	18: pipian, 26: pollo al cacahuate, 28: chilaquites más picantes del mundo(*)	
5	KIM	Healthy Food	AMERICAN	FEMALE	1	18: pipian, 26: pollo al cacahuate, 28: chilaquites más picantes del mundo(*)	
5	NAOMI	Vegan Food	RUSSIAN	FEMALE	1	18: pipian, 26: pollo al cacahuate, 28: chilaquites más picantes del mundo(*)	
4	AMARAL	Healthy Food	ARGENTINIAN	FEMALE	1	18: pipian, 26: pollo al cacahuate, 28: chilaquites más picantes del mundo(*)	
7	NANCY	Healthy Food	MEXICAN	FEMALE	1	18: pipian, 26: pollo al cacahuate, 28: chilaquites más picantes del mundo(*)	
7	MARIA	Healthy Food	MEXICAN	FEMALE	2	18: pipian, 26: pollo al cacahuate, 28: chilaquites más picantes del mundo(*)	
7	RAMONA	Healthy Food	NIGERIAN	FEMALE	1	18: pipian, 26: pollo al cacahuate, 28: chilaquites más picantes del mundo(*)	
10	AISHA	Healthy Food	INDIAN	FEMALE	1	18: pipian, 26: pollo al cacahuate, 28: chilaquites más picantes del mundo(*)	
8	ELBA	Healthy Food	INDIAN	FEMALE	2	18: pipian, 26: pollo al cacahuate, 28: chilaquites más picantes del mundo(*)	
7	AMANDA	Healthy Food	FRENCH	FEMALE	1	18: pipian, 26: pollo al cacahuate, 28: chilaquites más picantes del mundo(*)	

## EVIDENCIAS:

En la carpeta de evidencias coloca una imagen con el resultado de los 4 métodos que se te piden con el show del dataframe final o la cadena que retorne. Las imágenes deben mostrar el dataframe completo y los campos deben visualizarse sin cortes.

*Recuerda que la carpeta de evidencias la puedes encontrar en el **classroom** dónde encuentraste este examen y ahí deberás crear una carpeta con tu **nombre completo** para subir las **evidencias** solicitadas.*



¡MUCHA SUERTE!

