

TALLER 01 - ORANGE DATA MINING

Sergio Abraham Botina Josa

Brayan Ricardo Pisso Ramírez

Alejandro Vargas Orozco

Angel Rafael Alfonso Arias

Universidad Nacional de Colombia Sede Manizales

Marzo 2024

COFFE DISTRIBUTION ACROSS 94 COUNTRIES

La presente base de datos consiste en el dataset generado por USDA (United States Department of Agriculture), con datos de café durante una ventana de tiempo que abarcan desde los años de 1960 hasta 2023, con datos de 94 países, la base de datos se encuentra en el siguiente link: [Coffee Distribution Across 94 Counties \(kaggle.com\)](https://www.kaggle.com/datasets/usda/coffee-distribution-across-94-countries)

DESCRIPCIÓN GENERAL DE LA BASE DE DATOS

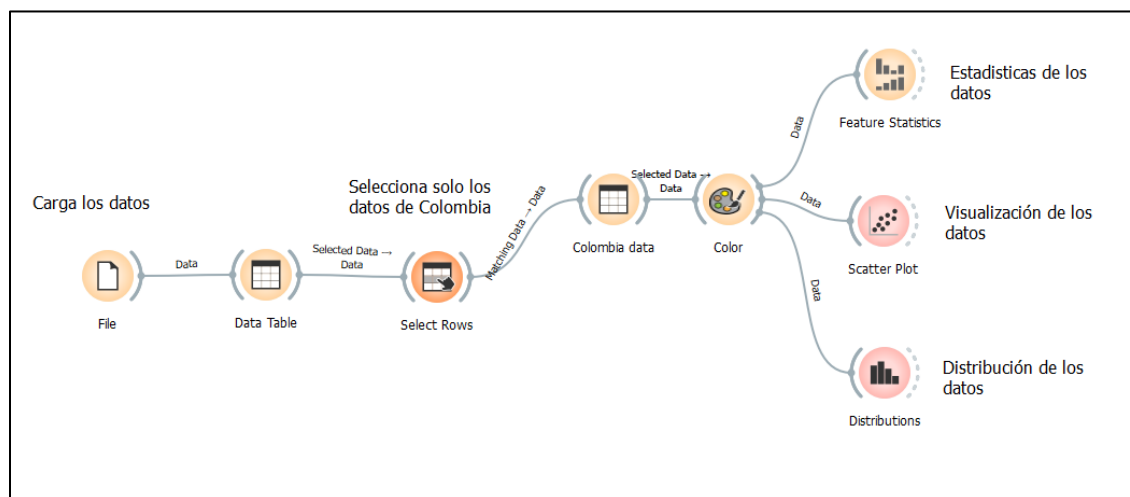


Fig. 1 Flujo de trabajo en Orange Data Mining para dataset

La figura anterior muestra el flujo de trabajo elegido para el análisis de esta base de datos, consiste en una primera parte de carga de datos y una posterior visualización de los datos a través del bloque Data Table, para este análisis solo nos enfocamos en el país de Colombia por lo tanto seleccionamos las filas correspondientes a la descripción de "Colombia", después de estos realizamos una inspección visual de los datos que se discutirá a continuación.

Información base

Name	Psd_coffe
Size	6016 rows, 21 columns
Features	1 categorical, 20 numeric

Tabla 1 Información básica de los datos

Datos de Colombia

Para el análisis se selecciono solo los datos del país de Colombia, con este primer filtrado de datos tenemos 64 filas, una por cada año desde 1960 hasta el 2023.

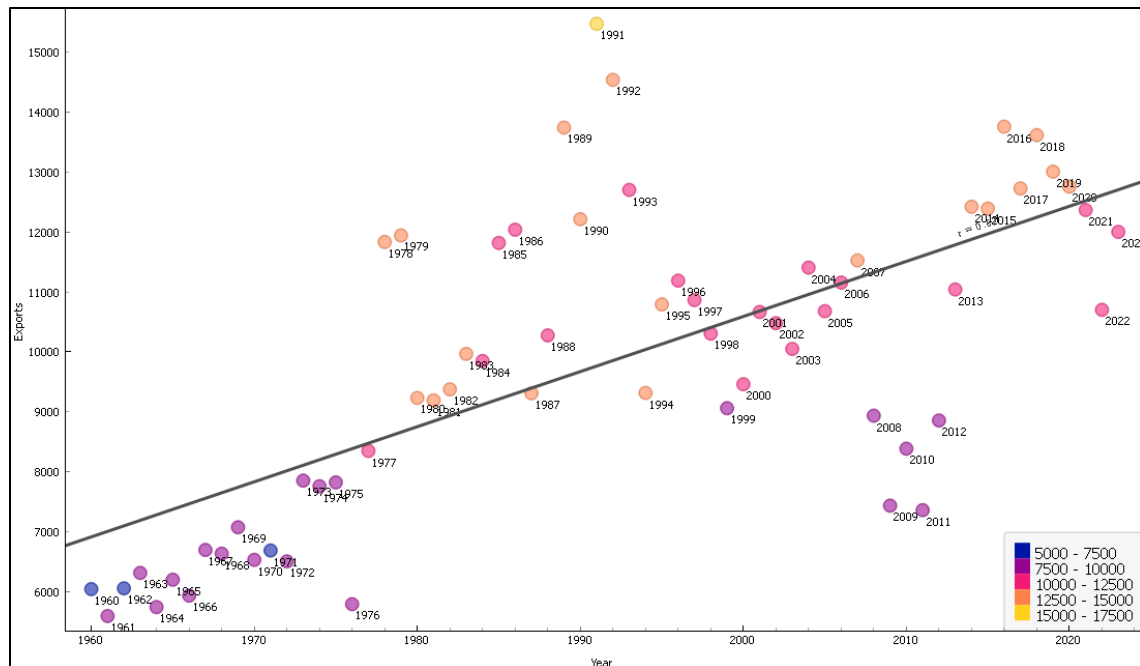


Fig. 2 años vs exportaciones de café y producción de arábico

La figura anterior refleja una tendencia positiva en relación con los años y las exportaciones de café en Colombia. Esto indica que a medida que transcurren los años, las cantidades de exportación han experimentado un aumento. Además, la escala de color representa la producción de café arábica. Con esta dimensión adicional, se aprecia una relación positiva entre las cantidades de exportación y la producción de este tipo de café.

Podríamos concluir que existe una correlación directa entre el aumento de la producción de café arábica y las exportaciones de café en Colombia, independientemente del año. Por lo tanto, si un caficultor colombiano tiene interés en exportar su producto, sería una buena idea considerar el cultivo de este tipo de café.

NINTENDO GAMES DATA SET

Esta base de datos tiene datos de videojuegos para las plataformas de Nintendo, de diversos géneros, fechas de lanzamientos, popularidad, puntaje de usuarios, puntaje meta y otras features.

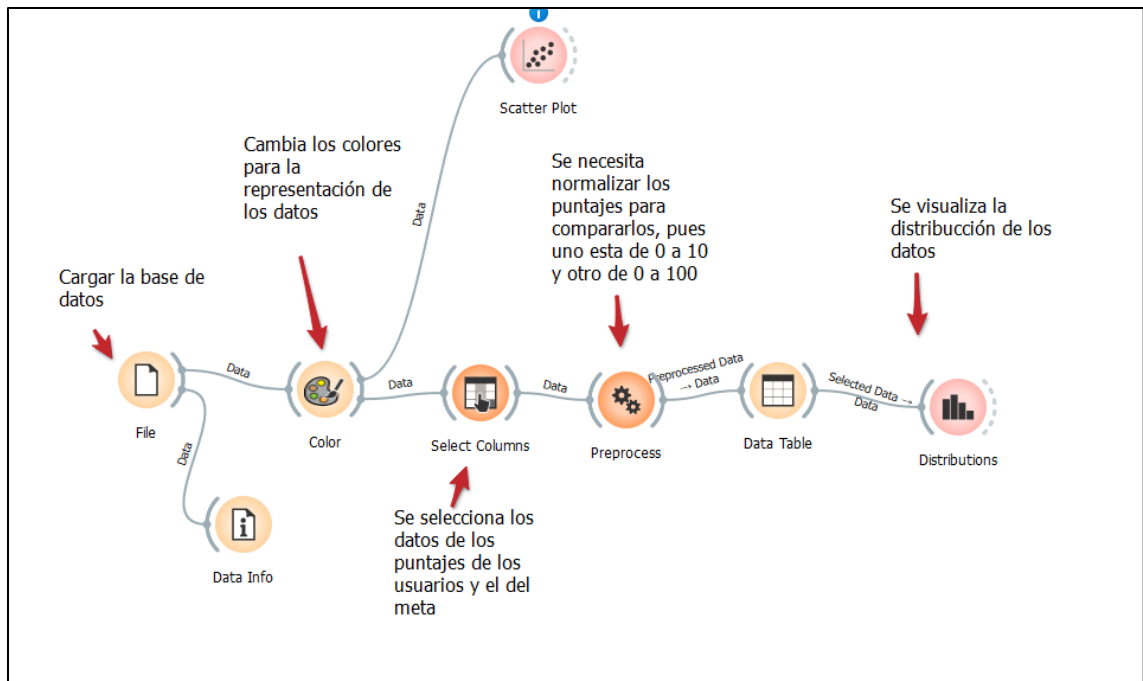


Fig. 3 Flujo de trabajo en Orange Data Mining para la base de datos de Nintendo Games.

La anterior figura muestra el flujo de trabajo para la base de datos de videojuegos de Nintendo, la primera parte es la carga de los datos, agregamos un "Data Info" para el resumen de la información de los datos, seleccionamos los datos de los puntajes de los usuarios el puntaje meta, después normalizamos estos datos, pues unos están en el rango de 0 a 10 y otro de 0 a 100, si realizamos una comparación de los datos sin realizar la normalización los resultados no serían los correctos.

DESCRIPCIÓN DE LA BASE DE DATOS

Name	NintendoGames
Size	1094 rows, 9 columns
Features	2 categorical, 2 numeric
Metas	5 text

Tabla 2 Información básica base de datos

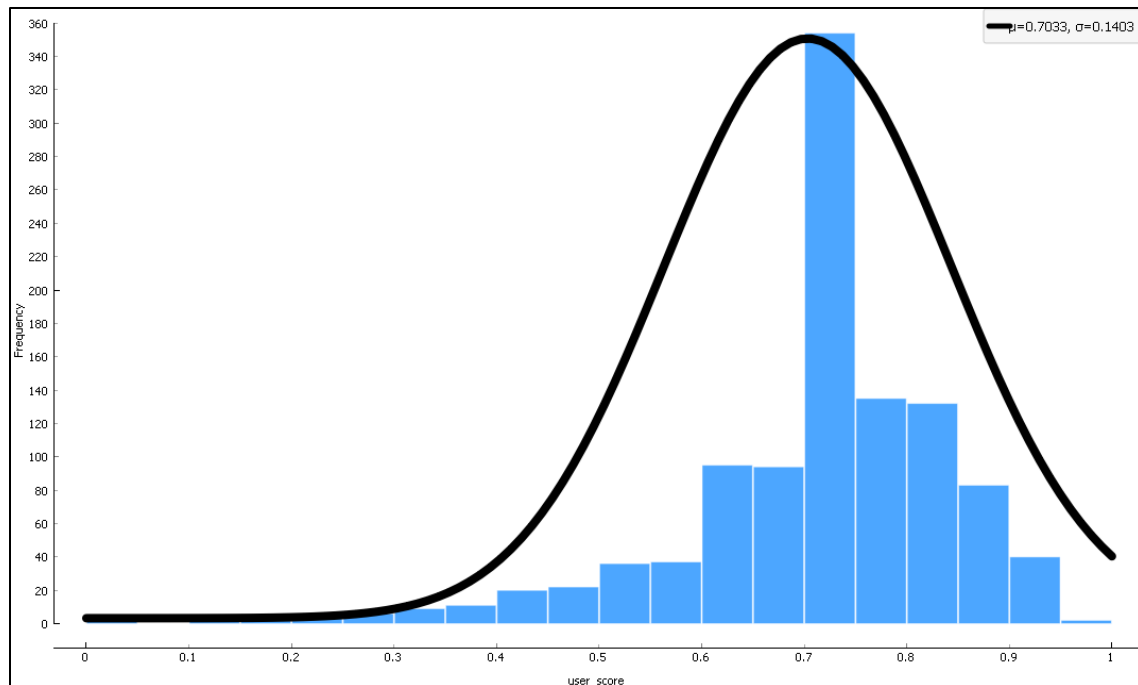


Fig. 4 Distribución de datos para puntuación de usuarios

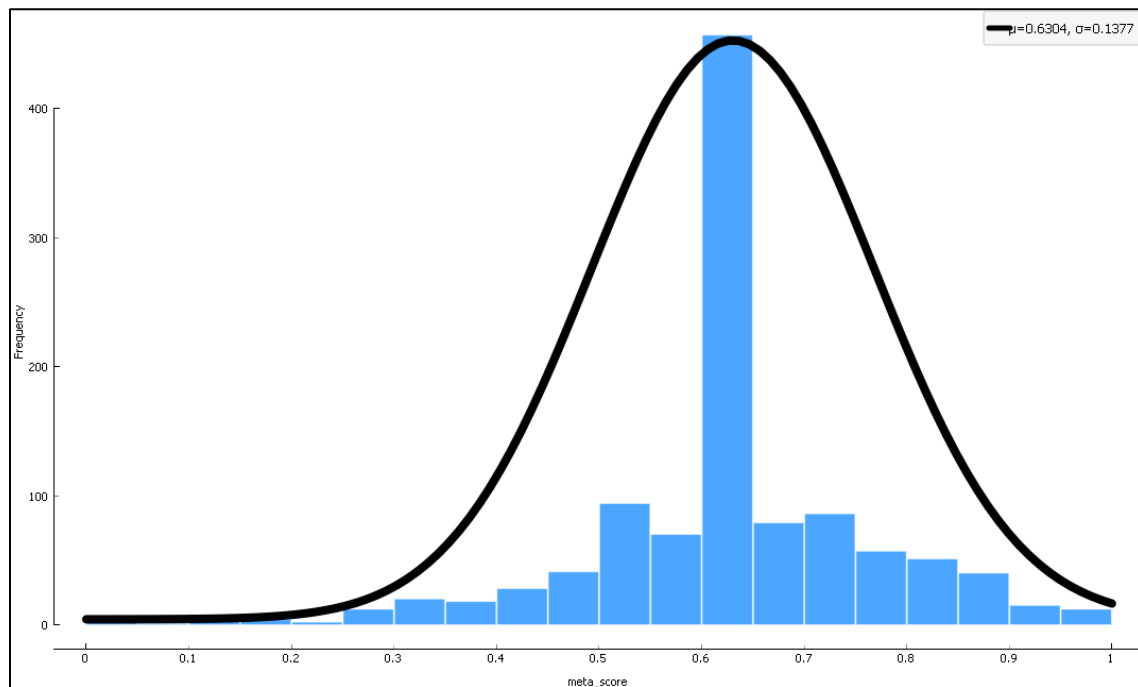


Fig. 5 Distribución de datos para puntuación de meta

Las figuras anteriores muestran las distribuciones de los datos para los puntajes de los usuarios el puntaje general que obtienen los videojuegos llamados “meta score”, los puntajes tiene una distribución normal entre los valores de 0.6 a 0.8, pero hay un comportamiento que se hace evidente, los usuarios tienden a calificar los videojuego con un mayor puntaje, a comparación del meta-score que suma los puntajes hechos por críticos o empresas dedicadas a esto.

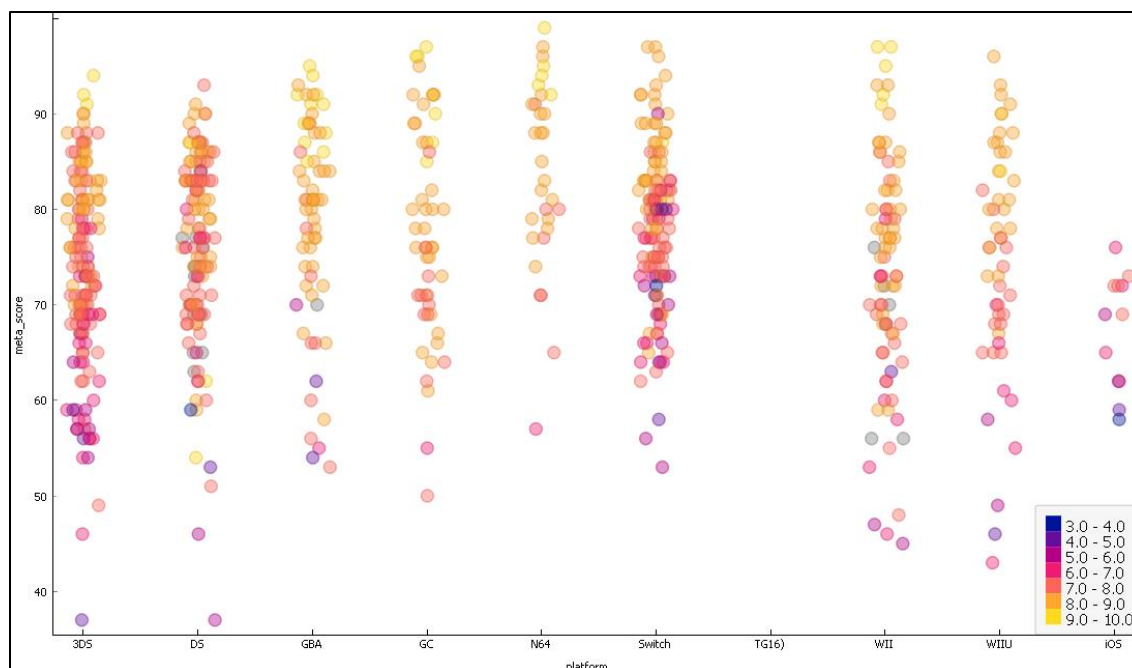


Fig. 6 plataformas vs puntaje meta más puntaje de los usuarios

La gráfica anterior presenta los datos por plataforma y puntaje meta, con los colores que representan los valores de puntaje otorgados por los usuarios. De estas observaciones, se puede concluir en primer lugar que los videojuegos de la plataforma de Nintendo 64 (N64) son los mejor evaluados, tanto en el puntaje meta como en el puntaje proporcionado por los usuarios. Se destaca que hay una plataforma, TG16, para la cual no se disponen de datos, sugiriendo su posible eliminación en la creación de modelos futuros.