



UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FCFM



FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN  
FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

MINERÍA DE DATOS

EJERCICIO PRÁCTICO  
DE BASES DE DATOS

WENDY OLIVIA BAZÚA CORRALES  
1887913

**Objetivo:**

Suponiendo que somos parte de Google Play Store, encontrar una mejor manera organizar las aplicaciones cuando los usuarios las busquen por categoría para que muestre aplicaciones que tienen mejor aceptación.

**Problema planteado:**

Hay muchas aplicaciones que tienen muchas descargas pero al ver su rating y el número de reviews, podemos observar que tienen muy poca participación o muy baja calificación.

**Solución:**

Primero usar una técnica de clasificación para separar las aplicaciones de la base de datos por categoría.

Después encontrar como se relacionan el tamaño de la aplicación con el número de reviews y la calificación que tiene para descartar aplicaciones que tienen muchas descargas pero muy baja participación porque eso nos da un indicio de que la aplicación es llamativa pero al final no es lo que los usuarios deseaban y por lo tanto se refleja en estos dos datos.

Y así poder mostrar primero al usuario aplicaciones que tengan una mejor relación de estos tres factores porque basados en usuarios pasados es más probable que esas aplicaciones satisfagan sus necesidades.

**Objetivo:**

Crear una predicción que nos diga como se comportaría una segunda ola de contagios en los países basado en la información que se cuenta de la primera ola y así poder mejorar protocolos en distintos países.

**Problema planteado:**

Ante la incertidumbre de la posibilidad de que haya una segunda ola de contagios de COVID-19, es muy probable que si no se previenen las consecuencias utilizando el ejemplo pasado de la primera ola, vuelva a haber resultados catastróficos en muchos lugares al no mejorar protocolos.

**Solución:**

Se pueden clasificar los datos por países, inclusive se pueden clasificar también por la categoría del país en cuanto a si son del primer mundo, segundo mundo o tercer mundo y así utilizar técnicas de predicción para ver como podría afectar una segunda ola si se maneja la situación de igual manera que en la primera ola y si datos de los países de la misma categoría se comportan similarmente, para así ver que tanto hay que reforzar las medidas de sanidad y las restricciones que debe haber para evitar una segunda ola peor a la primera.

## ≡ CRITICAS DE VINOS ≡

### **Objetivo:**

Evitar el sesgo que puede haber dependiendo el catador basandose en el tipo de resultados que da normalmente.

### **Problema planteado:**

Al ordenar los datos de calificación ascendentemente, podemos observar que los nombres de los catadores que dan malas calificaciones son generalmente los mismos y pasa de igual manera al ordenarlos descendientemente. Esto nos dice que dependiendo el catador hay un sesgo en cuanto a la calificación que dan.

### **Solución:**

Para tener calificaciones más objetivas, sugiero primero clasificar los datos por catador. Despues, partir las calificaciones en 3 grupos: calificación baja, calificación media y calificación alta. A partir de estas divisiones, utilizar un método de visualización para ver como se distribuyen las calificaciones de los catadores. Asi podremos identificar a los catadores que dan muchas malas calificaciones, a los que dan muchas buenas calificaciones y los que tienen más distribuidas sus calificaciones. Entonces por ejemplo, cuando un catador que normalmente califica mal, vuelva a calificar mal proceder a que un catador neutro lo califique para tener una calificación libre de sesgo y más acertada.

## ≡ CLASIFICACION DE PLANTAS ≡

### **Objetivo:**

Crear un instrumento que nos ayude a ver como se relacionan la medida del sépalo y el pétalo en las tres especies de Iris: Iris-Setosa, Iris-Virginica e Iris-Versicolor esto para fines de estudios en biología por ejemplo.

### **Problema planteado:**

Un grupo de biólogos de cierta universidad que se encuentran estudiando este tipo de flor en tres de sus especies necesitan saber como se relacionan las medidas de sépalo y pétalo de las respectivas especies.

### **Solución:**

Iniciando con clasificar los tres tipos de especies que son Iris-Setosa, Iris-Virginica e Iris-Versicolor podemos proceder a utilizar herramientas de visualización como gráficos de dispersión y de barras para ver la relación entre las medidas del sépalo y de los pétalos de las flores.

Al hacer esto les puede servir inclusive para comparar los tipos de especie entre si haciendo diagramas de boxplot para compararlos y de esta manera ver en una sola gráfica las tres especies y sus respectivas medidas, esto puede ser servicial para entender el comportamiento de cada tipo de especie y comprender en general a esta flor a rasgos generales porque se trata de la misma clase.

**Objetivo:**

Mejorar el sistema de sugerencias que te dan al terminar de ver una película o serie para que el usuario pueda encontrar contenido que le interese de manera más sencilla y eficiente.

**Problema planteado:**

Como la base de datos no nos da información sobre la calificación que le da el usuario a la película o serie y tampoco nos dice de que géneros se trata, se necesita encontrar una relación entre los datos que nos permita dar una sugerencia al usuario acercata a las cosas que normalmente ve.

**Solución:**

Relacionar los integrantes del cast de la película con los directores, otros miembros del cast e inclusive con el rating (que en esta base de datos es la clasificación que tienen) para que cuando una persona termine de ver una película o serie, le sugiera cosas lo más similares posibles a lo que acaba de ver.

Y de igual manera, tomar en cuenta cuando un usuario deja una serie o una película sin terminar, mostrarle menos contenido relacionado con lo que dejó a medias porque nos indica que no le gustó o no le interesó lo suficiente para terminarlo, por lo tanto contenido similar puede no interesarle.