



UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN  
FACULTAD DE CIENCIAS FÍSICO  
MATEMÁTICAS



# **MINERÍA DE DATOS**

## **Bases de datos**

Maestra: Mayra Berrones

Nombre:

Tania Sarahi Rossel Castillo

Matrícula:

1810461

## Google Play Store Apps

### Objetivos:

Hacer recomendaciones al usuario dependiendo de la aplicación elegida ya sea para una combinación entre aplicaciones para un uso más completo o llevar para a cabo una tarea más completa y/o para recomendar una aplicación similar a la que se busca para tener más variedad de elección

Por ejemplo, en el primer punto de combinación de aplicaciones: Podemos buscar instalar una aplicación para hacer ejercicio y que como recomendación Google Play me presente aplicaciones de dietas para que de esta manera tener mejores resultados al hacer ejercicio.

### Problema planteado:

En ocasiones solo muestran las aplicaciones más populares, pero no las que son más eficientes al momento de realizar alguna tarea que es necesaria, además con ayuda de otras aplicaciones podemos ayudarnos a hacer una tarea más completa y eficiente.

### Solución:

Haciendo uso de las bases de datos de “Categoría” y “Rating” para recomendar la mejor aplicación haciendo un match con el número de Instalaciones para que concuerde que sea lo mejor y así mismo con la tarea a realizar, todo esto haciendo uso de una técnica de reglas de asociación para definir los patrones entre aplicaciones y se dé una tarea más completa, en cuanto al segundo objetivo podría utilizarse la técnica de clasificación para mostrar aplicaciones similares.

## Coronavirus (Novel Corona Virus 2019 Dataset)

### Objetivos:

Analizar si el clima en el que se puede llegar a transmitir el virus de COVID – 19 es un factor importante de contagio de dicha enfermedad tomando en cuenta los ecosistemas de diversos países afectados y de esta manera saber que otras medidas de precaución tomar para reducir el índice de contagio presentado en los países.

### Problema planteado:

Altos contagios de forma exponencial que afecta al sistema de salud por alta demanda de los servicios médicos, además de muerte de personas por insuficiencia de material quirúrgico necesario para su atención médica correspondiente.

### Solución:

Haciendo uso de las bases de datos de “localización” y “país”, podemos usar la técnica de visualización de minería de datos para saber cuáles son los países que tienen un mayor índice de contagios y dependiendo de las medidas establecidas por el gobierno analizar el clima que se presenta con mayor frecuencia de cada país, para poder reestablecer mejores medidas de prevención de contagio del covid-19.

## Netflix (Netflix Movies and TV Shows)

### Objetivo:

Ofrecer a la compañía Netflix recomendaciones sobre qué tipo de entretenimiento preferido por los usuarios para que de esta manera la empresa pueda ofrecer una mejor variedad de películas y/o series que se adapten al perfil del usuario de acuerdo con su preferencia entre estos dos tipos de entretenimiento y con esto mantener a la mayor cantidad de usuarios registrados dentro de la plataforma, para así mantener e incrementar las ganancias como empresa

### Problema planteado:

No tener metas claras sobre el tipo de contenido a ofrecer a los espectadores debido a falta de atención a sus preferencias ya sean películas o series y esto conlleva a un gasto de recursos en entretenimiento que no capta la atención de los televidentes.

### Solución:

Haciendo uso de base de datos de “Tipo” (Película o Serie) usando las técnicas de visualización de minería de datos podemos hacer una representación gráfica sobre las preferencias de los espectadores entre películas o series para que de esta manera la compañía solo se centre en la que tenga mayor preferencia y tener más ingresos como lo mencionamos en el objetivo.

## Clasificación de Plantas (Iris Species)

### 🚦 Objetivos:

Verificar que las plantas y sus características como los son el ancho y largo de sus pétalos y sépalos se encuentren sin anomalías o si se llegan a mostrar anomalías verificar la forma en la que fue plantada o en qué condiciones se llevó a cabo para saber cuál fue la causa de esto.

### 🚦 Problema planteado:

Llevar un listado de posibles plantas que muestren anomalías, ya que puede que no se hayan tomado en cuenta y de esta manera sirvan para posibles investigaciones.

### 🚦 Solución:

Haciendo uso de las bases de datos del tamaño del sépalo con su debido ancho y largo, así mismo de la base de datos del pétalo con su respectivo ancho y largo podemos hacer uso de la técnica de Detección de Outliers por medio de boxplot para verificar cuales son los que se encuentran dentro de los rangos normales que presentan todos los demás sépalos y pétalos, y por consecuencia los que se encuentran fuera de los rangos de los boxplot se presentaran como anómalos. De esta manera se puede llevar un mejor control para saber si se podría llegar a enfrentar a anomalías dentro de las plantas.

## Criticas de vino (Wine Reviews)

### Objetivos:

Viendo desde un punto de vista del dueño de un restaurante donde uno de sus especiales es el vino podríamos presentarle recomendaciones para ofrecer a sus clientes y así mismo invertir de una manera más segura.

### Problema planteado:

El dueño de negocio tiene incertidumbre sobre qué tipo de vinos tiene que comprar y desde que providencia, tomando en cuenta las críticas de los catadores de vinos y también el viñedo dentro de la bodega de donde provienen las uvas que elaboraron el vino, ya que de esta manera se puede respaldar la decisión de inversión.

### Solución:

Haciendo uso de las bases de datos de “designación” donde obtendremos el viñedo donde provienen las uvas, “providencia” será, como lo dice su nombre, el lugar donde proviene el vino y la base de datos de puntos la “calificación” de los catadores de vino, podemos presentar al dueño del restaurante una representación gráfica de las diferentes tipos de bases de datos usando la técnica de visualización para una mejor toma de decisiones y logre apreciar de una manera más digerible el tipo de providencia que puede elegir el vino, el tipo de uva y todo esto basado del puntaje de los catadores.