



**Universidad Autónoma de Nuevo León**

Facultad de Ciencias Físico Matemáticas

Minería de Datos

## Ejercicios Bases de Datos

---

Técnicas de Minería de Datos

**Helena Patricia Carrillo Soto**

Grupo: 002

Matrícula: 1725370

A 12 de octubre de 2020, San Nicolás de los Garza, Nuevo León

## Google Play

**Nombre de la base de datos:** Google Play Store Apps

### OPCIÓN 1

**Objetivo:** Mejorar la experiencia del usuario de Google Play Store al recomendar aplicaciones acordes a sus gustos.

**Problema planteado:** Google Play Store tiene muchas aplicaciones disponibles para descarga, pero a pesar de ello los usuarios podrían no aprovechar esta variedad debido a que no se le presentan las opciones que son acordes a sus preferencias.

**Solución:** Crear un algoritmo de clasificación o agrupación que nos permita ver las similitudes entre las aplicaciones para hacer recomendaciones al usuario basados en las similitudes de otras aplicaciones con aquellas que ya ha descargado y/o calificado positivamente.

### OPCIÓN 2

**Objetivo:** Dar una vista general a los creadores de aplicaciones de las características que tienen las aplicaciones más descargadas de Google Play Store para que puedan mejorar el desempeño de sus creaciones.

**Problema planteado:** Google Play Store cuenta con una gran variedad de aplicaciones de diversos creadores sin embargo dada la gran cantidad de aplicaciones disponibles algunas podrían quedarse rezagadas respecto al número de descargas lo que podría desanimar a los creadores de contenido.

**Solución:** Crear un algoritmo de clasificación (o agrupación) que nos permita ver las similitudes entre las aplicaciones más descargadas y mejores rankeadas para poder ofrecer a los creadores de contenido una vista general de las características que su aplicación debe tener para ser bien recibida por los usuarios.

## Coronavirus

**Nombre de la base de datos:** Novel Corona Virus 2019 Dataset

**Objetivo:** Identificar las estrategias de salud pública más eficientes para lidiar con la pandemia de COVID-19 evaluando la evolución y propagación de la enfermedad en distintos entornos y en base a los resultados elaborar planes de acción que nos permitan replicar los resultados en nuestro país.

**Problema planteado:** El COVID-19 ha afectado a todo el mundo, pero algunos países han logrado contener la enfermedad mientras que en otros los números de infectados y fallecidos continúan elevándose. Esto provoca naturalmente las preguntas ¿cuáles son las diferentes estrategias que se siguieron en los países que han logrado minimizar el impacto de la pandemia? y ¿Es posible replicarlas en nuestro entorno?

**Solución:** Crear un programa que nos permita visualizar y comparar la evolución de los números de infectados y fallecidos de cada país, para así ser capaces de evaluar las estrategias seguidas por cada uno de ellos y en base a los resultados identificar aquellas estrategias que han tenido más éxito.

## Criticas de vinos

**Nombre de la base de datos:** Wine Reviews

**Objetivo:** Identificar el lugar y condiciones ideales para producir distintos tipos de vino para de esta manera asegurar una correcta o acertada elección al momento de elegir la localización para un nuevo viñedo que busque especializarse en algún tipo de vino en particular.

**Problema planteado:** El sabor y las características del vino se ven fuertemente influenciados por las condiciones en que se cultivaron las uvas. Cada país tiene condiciones distintas por lo que los vinos provenientes de distintos países tienen características distintas. Siendo este un producto cuya mayoría de mercado se compone de expertos y entusiastas de los cuales satisfacer las necesidades y expectativas es considerablemente más difícil que en otros mercados, para los nuevos productores e inversores es de vital importancia asegurarse de que su producto contará con las características necesarias para satisfacer las necesidades del mercado a la vez que maximizan la utilización de sus recursos haciendo inversiones inteligentes basadas en investigaciones congruentes.

**Solución:** Usar la técnica de agrupamiento y visualización que nos permita ver las características que comparten los vinos provenientes de una misma región, así como las que comparten los mejores puntajes de cada una de ellas para así facilitar a los nuevos productores la elección de una localización o bien proporcionar la información necesaria para replicar las condiciones óptimas para conseguir cultivos que produzcan el tipo de vino que se desea. Esto también ayudaría a inversores potenciales del mercado vinícola a saber en qué regiones es más conveniente invertir.

## Clasificación de plantas

**Nombre de la base de datos:** Iris Species

**Objetivo:** Facilitar y agilizar el proceso de clasificación de flores pertenecientes al subgénero Iris basados en las mediciones de sus pétalos y sépalos, especialmente para aquellos que aún no cuentan con experiencia en el ámbito de la floricultura o bien para los entusiastas de la materia.

**Problema planteado:** Iris es un género de plantas rizomatosas que cuenta con más de 300 especies, además de muchos híbridos y cultivares. Iris se usa comúnmente para referirse a todas las especies, así como a otros varios géneros estrechamente emparentados y a una subdivisión dentro del género. Las flores del subgénero Iris mantienen, como ya se ha mencionado, grandes similitudes entre sí por lo que clasificar estas flores es un proceso difícil que requiere de mucha experiencia en materia de floricultura la cual no muchos poseen.

**Solución:** Elaborar un programa de clasificación basado en árboles de decisión que nos permita reconocer la especie de flor basándonos en sus características dadas de una manera más sencilla y rápida.

## Shows de Netflix

**Nombre de la base de datos:** Netflix Movies and TV Shows

**Objetivo:** Visualizar los shows con más reproducciones y mejores rankeados en cada país para con esta información sea posible elegir, acorde al país en que se desee enfocarse, el tipo de show en el que es más conveniente invertir para asegurar un buen recibimiento por parte del público y la crítica y de esta manera garantizar el éxito de la inversión.

**Problema planteado:** Para mantener su competitividad ante el surgimiento de muchas otras plataformas de streaming, Netflix ha optado por invertir en la creación de contenido original. Sin embargo, se debe satisfacer las necesidades de muchos usuarios con diferentes gustos y preferencias y, a pesar del éxito que ha cultivado en los últimos años, Netflix no cuenta con recursos ilimitados por lo que debe usar sus recursos de manera eficiente de tal forma que le permita generar la mayor cantidad de ganancias de cada una de las producciones que decida financiar. Para esto se debe apostar por el tipo de contenido que se sabe es consumido por los mercados a los que se desea enfocar.

**Solución:** Elaborar un programa que nos permita visualizar el tipo de series y películas que más se reproducen y que tienen mejores calificaciones en cada país para de esta forma saber cuál es el tipo de contenido que generaría más ganancias y por ende en el que es mejor invertir.