

Universidad Autónoma de Nuevo León  
Facultad de Ciencias Físico Matemáticas

## **MINERÍA DE DATOS**

### Ejercicio Práctico Bases de Datos

7° Semestre

Licenciatura en Actuaría

**Alumno:** Evelyn Lizbeth Trejo Rodríguez

**Matrícula:** 1811917

**Profesor:** Mayra Cristina Berrones Reyes

**Grupo:** 002

## Índice

Google Play Store .....	3
Clasificación de Plantas .....	4
Coronavirus .....	5
Criticas de Vinos .....	7
Shows de Netflix .....	8

## Google Play Store

- **Nombre de la base de datos**

Google Play Store Apps

- **Objetivo**

Poder satisfacer necesidades, o crearlas en el usuario, recomendando aplicaciones relacionadas con las de mayor uso.

- **Problema planteado**

De manera personal, he tenido varias necesidades de las cuales después me entero de que existía una aplicación que me ayudaba a esa necesidad, por lo cual el problema está en que no se recomiendan, o no se sugieren cuando las requerimos, y aun sin tener necesidades, debemos comprender de que los usuarios en ocasiones no saben lo que necesitan, así que otro problema es que no se recomiendan aplicaciones relacionadas con las que el usuario interactúa o le gusta más.

- **Solución**

Desarrollar un algoritmo de agrupación de aplicaciones relacionadas con las mejor calificadas por el usuario o con las que el usuario tenga mayor interactividad para después sugerírselas, además, de acuerdo con lo más buscado en esas aplicaciones o en el buscador, agrupar aplicaciones referentes a esas búsquedas para comenzar a recomendarle las aplicaciones en su inicio.

## Clasificación de Plantas

El conjunto de datos clasifica las plantas de Iris en tres especies, además características de cada una.

- **Nombre de la base de datos**

Iris Species

- **Objetivo**

Proporcionar un mejor análisis e investigación de las plantas Iris, mediante su clasificación de acuerdo con su tipo y sus demás propiedades.

- **Problema planteado**

Hay una gran variedad de tipos de Iris en esa base de datos, con diferentes características, es decir cada una con diferentes tamaños ya sea del pétalo o sépalo, etcétera siendo complicado analizarlas y/o estudiarlas en diferentes ámbitos.

- **Solución**

Una solución sería utilizar el método de árboles de decisión de la técnica de clasificación para organizar o mapear el conjunto de plantas por clase dependiendo de sus características como el ancho del sépalo o pétalo y ancho de los mismos, de manera que nos permita analizarlas más sencillo además de que si lo vemos de una manera comercial, y este tipo de plantas se venden, es bueno tenerlas clasificadas de acuerdo a su tipo y con sus propiedades, ya que un cliente puede pedir cierto tipo de planta, con cierto tamaño de sépalo, etc.

## Coronavirus

La base de datos incluye información a nivel del día sobre los casos afectados por el COVID-19.

- **Nombre de la base de datos**

Novel Corona Virus 2019 Dataset

1)

- **Objetivo**

Poder analizar la información, si de acuerdo con la situación económica de cada país, está relacionada la cantidad de muertos en cada uno por este virus.

- **Problema planteado**

De manera lógica, podríamos pensar que hay regresión entre la economía del país y la cantidad personas recuperadas, debido a que los países con mayor economía cuentan con más capacidad en hospitales y mayor capacidad de atención médica, pero ¿la economía del país influye con la cantidad de muertos? Es un buen interrogante ya que diariamente vemos en las noticias las cifras, siendo Estados Unidos una potencia mundial, se encuentra con la mayor cantidad de muertos, entonces el problema está en que no se sabe exactamente si la economía de un país influye en si las personas mueren o no por este virus.

- **Solución**

Utilizar la técnica de regresión lineal simple para analizar el vínculo entre la variable dependiente que en nuestro caso serán las muertes por coronavirus y una independiente, que sería la economía de cada país, que se puede medir con el PIB, encontrando una relación matemática si es que la hay, así podríamos analizar si realmente estas variables se relacionan, es decir si la economía del país está relacionada con las muertes por covid-19.

2)

- **Objetivo**

Poder determinar cuáles podrían ser las mejores medidas de precaución tomadas por los países con mayor número de personas recuperadas y menor número de contagiados.

- **Problema planteado**

Sabemos que actualmente existen muchos países en donde continua el virus, más sin embargo han sabido controlarlo, por otro lado, se encuentran países, incluido México, en donde continua el virus sin un control, por eso mismo nos interesa analizar las estrategias que han funcionado para equilibrar este virus.

- **Solución**

Mediante el método de visualización darle seguimiento, es decir analizar las medidas de los países con mayor número de personas recuperadas y menor números de infectados por el virus, es decir el país sobresaliente, analizaremos que medidas de precaución, y de acción tomaron.

## Criticas de Vinos

La base de datos contiene reseñas de vinos con variedad, ubicación, bodega, precio y descripción.

- **Nombre de la base de datos**

Wine Reviews

- **Objetivo**

Ser capaces de analizar los precios y costos de los vinos con mejor puntuación por los usuarios, así como también su país de procedencia, para la toma de decisiones como vendedor y productor de vino.

- **Problema planteado**

De manera mercantil, nos encontraríamos en desventaja si no contamos con un análisis referente a una base de datos de vinos, debido a que al comenzar tu viñedo lo estaríamos haciendo sin información completa, pudiendo obtener mejores costos si lo llevamos a cabo en otro país, o de igual forma la calidad y reseña el vino en dicho país, nos es importante como vendedores para tener a consideración nuestras prioridades.

- **Solución**

Haciendo uso de la técnica visualización para proporcionarnos una manera accesible de ver los datos, para analizar de donde procede el vino de mejor calidad, además de encontrar el país donde pudiera reducir nuestros costos, dependiendo del país analizar cuál es el vino preferente para los usuarios, es decir, los vinos con la mejor critica, y al final tomar las decisiones basadas en esos datos nosotros como inversionistas, vendedores y/o productores.

## Shows de Netflix

El conjunto de datos consiste en programas de televisión y películas disponibles en Netflix a partir de 2019.

- **Nombre de la base de datos**

Netflix Movies and TV Shows

- **Objetivo**

Poder obtener un amplio análisis sobre los gustos y preferencias de los usuarios, para mantenerlos activos en la plataforma.

### **Problema planteado**

En estos momentos, estamos de acuerdo que Netflix ha sido una plataforma que hemos frecuentado excesivamente durante esta cuarentena, siendo un problema para el usuario cuando ya no tienen algo nuevo por ver, es decir, han pasado demasiado tiempo que creen que Netflix ya no les ofrece contenido nuevo, siendo perjudicial económicamente para la plataforma debido a que busquen alternativas que llamen más su atención.

- **Solución**

De acuerdo con las películas y series que vean completas o que califiquen mejor, desarrollar un algoritmo de patrones secuenciales para analizar los datos del usuario y encontrar subsecuencias dentro de un grupo de secuencias, a partir de ahí comenzar a recomendar series y películas nuevas, de acuerdo con lo que ya observamos que le agrada al usuario, con el fin de evitar la cancelación de suscripción debido a que no encuentren contenido nuevo de su agrado.