



UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FCFM



FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS



Ejercicio práctico de Bases de Datos

Materia: Minería de Datos

Maestra: Mayra Cristina Berrones Reyes

Alumno: Francisco Javier Hernandez Ortiz

Matricula: 1599327

Nombre de la base de datos: Google Play Store Apps

- **Objetivo:** Por medio de una campaña publicitaria, generar clientes potenciales, que deseen adquirir una aplicación digital, para su negocio o empresa, ofreciendo la mejor alternativa en base al comportamiento de aplicaciones similares, a la deseada por el usuario.
- **Problema planteado:** El tener una gran diversidad de categorías en las que se puede realizar una aplicación, haría que nuestra estrategia publicitaria sea demasiado extensa, aumentando los costos de publicidad a gran escala.
- **Solución:** Se realizara un análisis previo a la base de datos, con la finalidad de encontrar las categorías, donde se encuentra concentrado la mayor parte del mercado y respecto a ellas, ver el comportamiento de una aplicación gratuita contra una aplicación con costo, sin pasar por alto el impacto que tiene esto en el rating de la aplicación.

De esta manera resolveremos 2 de nuestros problemas principales, en primera instancia, obtendremos las categorías donde centraremos la primera etapa de nuestra estrategia publicitaria y como extra tenemos un reporte a detalle del comportamiento de las aplicaciones en el mercado, lo que nos dará herramientas para ofrecer alternativas a cada cliente, respecto a su petición.

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

- **Objetivo:** Dirigir en mayor porcentaje la atención médica, a las personas mas vulnerables, detectando patrones que indiquen mayor probabilidad de mortandad.
- **Problema planteado:** La diversidad de pacientes que se tienen contagiados, no se restringe a una edad en especifico ni a un sexo en particular, por lo que, para cada clínica del estado, se tienen diferentes estadísticas que no nos permiten definir la población vulnerable.
- **Solución:** En base al comportamiento nacional, se buscarán patrones que relacionen la edad, el sexo y las condiciones en las que se encuentra cada paciente contagiado. Una vez teniendo el análisis, se creará un modelo que nos permitirá rankear a nuestros pacientes y poner mayor énfasis en aquellos con mayor tasa de mortandad.

Nombre de la base de datos: Wine Reviews

- **Objetivo:** Nuestro viñedo se encuentra en un proceso de renovación constante, nuestra marca propia ha alcanzado su punto más alto, por lo que deseamos ampliar nuestra gama de productos, en función de la demanda del mercado, conservando la calidad de los productos de nuestra empresa y un precio razonable que se encuentre al alcance de la mayor parte de los demandantes de nuestros productos actuales.
- **Problema planteado:** Se decidió certificar a nuestros empleados en técnicas de viñedos de otros países. Entre la diversidad de opciones que se tiene, empatar las 3 necesidades principales que se desean obtener, que son calidad, precio y satisfacción del cliente. La dificultad de encontrar la mejor opción es una decisión que se tiene que analizar a gran escala.
- **Solución:** Se analizará a detalle la base de datos, encontrando patrones que engloben las necesidades de nuestra empresa, teniendo así definidas principalmente, aquellas técnicas que reúnan las 3 características deseadas, y proponerlas al consejo, quien seleccionará la mejor o las mejores, de acuerdo a nuestras necesidades.

Nombre de la base de datos: Iris Species

- **Objetivo:** Por medio de ingeniería genética, se desea crear una nueva especie de iris, que reúna cualidades de los 3 tipos que se tienen actualmente. Se desea encontrar la cualidad más notoria en cada tipo, e implementarla en la nueva especie, teniendo lo más sobresaliente de cada una.
- **Problema planteado:** A pesar de las diferencias entre los 3 tipos, es difícil clasificarlos entre sí, ya que encontrar patrones promedios se vuelve un tema muy complicado, por las propias diferencias entre especies del mismo tipo. Enfrentándonos inicialmente a encontrar esa cualidad sobresaliente que deseamos.
- **Solución:** Analizar a gran escala diferentes muestras de nuestras flores, nos ayudarán a encontrar clasificaciones mas notorias, teniendo así, de cada tipo, la característica con mayor notoriedad para nuestro nuevo prototipo de iris.

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

- **Objetivo:** La innovación para nuestra empresa es vital para permanecer competentes en el mercado, por lo que nuestros contenidos originales nos posicionan en números importantes de rating. Se desea generar un nuevo contenido, totalmente diferente, que presente en una misma producción a los actores que generan mayor rating a sus proyectos, sin importar el género en el que se encuentren y definir la mejor opción para este proyecto, ya sea serie o película.
- **Problema planteado:** Ante la gran diversidad de contenidos, nos encontraremos con infinitas combinaciones, que nos dan escenarios completamente distintos, en los que analizar uno en uno, sería imposible para un ser humano.
- **Solución:** Desarrollar un programa que encuentre las combinaciones óptimas que respeten el principal concepto del proyecto, que sería tener un espectáculo completamente diversificado. El sistema nos encontrará la mejor combinación, que contemplara la aceptación entre la diversidad de espectadores, países, actores, categorías. Nos dará ese punto donde la mayor parte de los espectadores intersecan entre una gran diversidad.