



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

Análisis Bases de Datos

Arturo Del Ángel De La Cruz 1809895

14 de Octubre de 2020

Nombre de la base de datos: Google Play Store Apps

En esta base de datos encontraremos diversas cosas sobre las aplicaciones que integran Google Play

Objetivo: Lograr que el interfaz de Google Play Store sea más atractivo para descargar más aplicaciones, poniendo primero las aplicaciones que tengan una relación con juegos o aplicaciones que sean parecidas o complementarias a las que ya han sido instalados con anterioridad por el usuario.

Problema Planteado: Las aplicaciones que se muestran al abrir la Google Play Store son principalmente las que están de moda en ese momento, al igual en las categorías en los rating tenemos las más populares, por lo que al usuario en ambas partes se le pueden ofrecer las mismas aplicaciones en estos 2 apartados y así tener menos oportunidad de que el usuario conozca diferentes que puedan no ser nuevas pero si de gusto para el usuario.

Solución: Relacionar las aplicaciones que se encuentran actualmente en el software del usuario ya instaladas y actualizadas para relacionar que aplicaciones pueden ser de su agrado. Al igual se puede utilizar la base de datos de reseñas y verificar si el usuario ha dado alguna buena reseña a algunas de las aplicaciones para relacionar esos tipos de aplicaciones con él, también verificar el tamaño de las aplicaciones en su respectiva base de datos para comparar con su almacenamiento disponible y que al usuario no se le recomienden aplicaciones para las cuales no tiene espacio para descargar.

Nombre de la base de datos: Coronavirus

En esta base de datos contiene la información diaria sobre el coronavirus (casos, muertes, recuperados, etc.).

Objetivo: Poder encontrar una relación entre los países, el género, los contagiados y los recuperados para conectar los casos que se han proporcionado y poder crear estrategias para conseguir que tanto el número de recuperados como el de muertes pueda reducir significativamente y aumente el número de recuperados para quienes ya contrajeron la enfermedad.

Problema Planteado: Se sabe que existen casos asintomáticos por lo que puede haber mucha gente que no se haya hecho la prueba y aun así llevar el virus con ellos, por lo que no se tiene una estadística del todo completa y una serie de datos que nos diga el número exacto de personas contagiadas.

Solución: Con la base de datos poder tomar los valores de los porcentajes entre los contagiados de los países, del sexo que está contagiado y de los casos que son asintomáticos

y con estos datos poder crear una regresión para poder proyectar un aproximado de los contagiados realmente de Coronavirus y con esto tomar las medidas de seguridad que sean relacionadas con este número y no se tomen en cuenta solo los que se han reportado por hacerse la prueba.

Nombre de la base de datos: Crítica de vinos.

En esta base de datos tenemos las críticas que le hacen a los vinos los clientes y las razones por las cuales prefieren un vino de otro.

Objetivo: Conseguir vinos que sean atractivos para los clientes y pueda aumentar su consumo dentro de los restaurantes, bares, etc. y no se queden mucho tiempo sin consumir y generando espacio a las bodegas de estos lugares de consumo y al mismo tiempo generar más ingresos.

Problema Planteado: Que los clientes de estos lugares de consumo no adquieran vino suficiente de algunos tipos y este se quede en el lugar por mucho tiempo y no sea factible seguir adquiriendo de este tipo de vino y en cambio conseguir vinos que sean de mejor agrado para los clientes.

Solución: Realizando una herramienta podemos conseguir que estos lugares vean de donde los clientes prefieren que el vino sea, como se calificó este mismo en WineEnthusiast, cuál es su costo y de aquí ellos puedan obtener un presupuesto y el área del cultivo del vino. Teniendo estos datos el establecimiento podrá conseguir los mejores vinos para sus clientes y ellos puedan consumirlos con mayor frecuencia.

Nombre de la base de datos: Clasificación de Plantas.

Esta base de datos contiene las 3 especies de Iris con 50 muestras de cada una de ellas, que conociendo estas flores puede ser algo complicado saber cuál es la especie la cual esta es de Iris.

Objetivo: Obtener la especie de la Iris a la cual pertenece una planta teniendo en cuenta diferentes factores ya que esta puede ser Isis-Setosa, Iris-Versicolor o Iris-Virginica y estas a simple vista parecen ser idénticas por lo que buscamos optimizar esta clasificación que tenga pruebas de que cada planta pertenece a cierta especie.

Problema Planteado: La clasificación de estas plantas puede depender de pequeños cambios a lo que es la longitud o el ancho del pétalo o del sépalo que a la vista del ojo humano puede llegar a ser hasta imposible percibir.

Solución: Con la base de datos podemos crear una herramienta que tenga ya sea un Árbol aleatorio o un Bosque aleatorio para ser más precisos donde se puedan verificar los distintos datos que tiene esta base de datos desde la longitud del pétalo o sépalo hasta la longitud de estas y de ahí ir separando en cada nodo que vaya avanzando y al final poder concluir a que especie de las Iris pertenece esa flor.

Nombre de la base de datos: Shows de Netflix.

Dentro de este conjunto de datos podemos ver los programas de televisión y películas que podremos encontrar en Netflix a partir del 2019.

Objetivo: Que la clientela de Netflix siga interesada en los programas o películas que Netflix saque para poder reproducir y de aquí Netflix pueda decidir que nuevos estreno traer, al igual se busca traer nuevos usuarios a la plataforma y que puedan seguir con su membresía un largo tiempo debido al interés que tiene por el catálogo.

Problema Planteado: Los estrenos que traiga Netflix pueden no ser atractivos para ciertos usuarios por lo que lleguen a cancelar la membresía y dar mala publicidad de la plataforma debido a su experiencia y esto evitar que Netflix consiga más usuarios los cuales pudieran si estar conformes con el catálogo.

Solución: Crear una herramienta donde Netflix vea de donde prefiere que sean las películas, si prefiere que se estrenen más películas o programas de televisión, también tiene sus estrenos originales de Netflix, con estos datos verificar el éxito que estos están teniendo en la plataforma y darle al usuario una buena variedad de donde pueda encontrar algo que sea de su agrado.