



UNIVERSIDAD AUTONOMA DE NUEVO LEON



FACULTAD DE CIENCIAS FISICO MATEMATICAS

Ejercicios de Bases de Datos **Técnicas de Minería de Datos**

MINERIA DE DATOS

ALUMNA = MAGALY RIVERA VALDEZ

MATRÍCULA = 1823340

MAESTRO: MAYRA CRISTINA BERRONES REYES

a) Google Play Store Apps



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

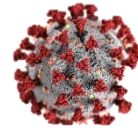
Problema planteado ->

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.

b) Coronavirus (Novel Corona Virus 2019 Dataset)



Objetivo ->

Tomar la mejor decisión de si mantener el cierre de fronteras o abrirlas.

Problema planteado ->

Uno de los países en el que más medidas preventivas para el Covid-19 han tenido es Tailandia pues decidió aislarse de todo el mundo cerrando sus fronteras tanto para su gente como para los extranjeros, nadie puede salir ni entrar aún a la fecha, esto ha traído cosas positivas y al mismo tiempo negativas, pues a pesar de que ha sido un país al que menos le ha afectado la pandemia en sus ciudadanos también su PIB ha bajado y expertos dicen que si Tailandia no abre sus puertas al mundo pronto, es muy probable que Tailandia tenga la peor situación económica de su historia reciente, lo que queremos saber, es si el país esta preparado para abrir sus puertas, que tan vulnerable puede ser su población y si saldrá contraproducente en el ámbito económico o al contrario ayude a recuperar la economía

Solución ->

Desarrollar una herramienta que nos permita identificar las poblaciones más susceptibles y cuáles son las características que comparten entre ellas, con la finalidad de que Tailandia pueda comparar las características de su población con la de los grupos más susceptibles y darse cuenta de si su población pudiera entrar en ese

grupo, al tener este panorama, Tailandia podría estudiar si pueden hacerle frente en capacidad hospitalaria, así como el impacto económico que pudiera tener para tomar la mejor decisión.

c) Criticas de Vinos (Wine Reviews)



Objetivo ->

Conocer en qué país nos conviene invertir para tener un buen viñedo rentable y en qué precio poner a la venta nuestro vino, de esta manera se podrá lograr una mejor posición en el mercado de los vinos.

Problema planteado ->

Vamos a suponer que somos un grupo de personas que quieren iniciar una empresa de vinos y que contamos con la base de datos antes mencionada, esto nos va a ayudar a estudiar los vinos que tienen las mayores puntuaciones, asimismo podemos checar el país del cuál proceden y de esta forma estar analizando las preferencias de los clientes, y todo esto nos va a ayudar para elegir nuestro tipo de vino y el precio más rentable

Solución ->

Para el problema planteado podemos utilizar la Agrupación como herramienta, con ella estaremos analizando cada región y sus características, para escoger el lugar donde nos conviene invertir para el viñedo, también podemos emplearlo para clasificarlos en rangos de puntajes y sea más sencillo analizar las preferencias de las personas, para esto va a ser necesario utilizar la columna Points, y una vez que lo hemos elegido podremos analizar las reseñas que se han hecho respecto a ese vino de mayor puntaje y asimismo observaremos las características que presenta y podemos observar la columna Price de la región que escogimos como guía, con esto solucionaremos el problema del precio ya que tendremos de dónde basarnos para elegir el nuestro de acuerdo con las circunstancias de la empresa.

d) Clasificación de Plantas (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, tienen diferentes características, es decir, que cada una con diferentes tamaños ya sea del pétalo o sépalo, etcétera siendo complicado analizarlas y/o estudiarlas en diferentes ámbitos.

Solución ->

Podríamos utilizar el método de árboles de decisión de la técnica de la 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 sencillas además de que si lo vemos de una manera comercial, y este tipo de plantas se venden, es buen 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.

e) Shows de Netflix (Netflix Movies and TV Shows)The Netflix logo, consisting of the word "NETFLIX" in red capital letters on a black rectangular background.***Objetivo ->***

Ofrecer a la compañía de Netflix recomendaciones sobre qué tipo de entretenimiento preferido por los usuarios para que de esta manera la empresa pueda ofrecer una 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 ridículas 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 podremos hacer una representación gráfica sobre las preferencias de los espectadores entre películas o series para que de esta manera la compañía sólo se centre en la que tenga mayor preferencia y tener más ingresos como lo mencionamos en el objetivo.