



UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN



FCFM

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

EJERCICIO PRÁCTICO BASES DE DATOS

5 BASES DE DATOS

LESLYE MARISOL HERNÁNDEZ BOLAÑOS

1819111

GPO3

PROF. MAYRA CRISTINA BERRONES REYES

MINERÍA DE DATOS

NUEVO LEÓN, MX, SEPTIEMBRE 2020

Google Play Store Apps

Objetivo: Apps en la tienda de Google Play y cómo se desenvuelven las apps en ella.

Problema planteado: Determinar si una aplicación tiene un buen alcance para el público.

Solución: Si hablamos de una nueva aplicación que se lanzará al mercado, podremos analizar la columna del tipo de aplicación y cuál tiene más descargas y a su vez mejor rating, por otro lado, también podremos analizar, en caso de ser de paga, entre que rangos de precios están las aplicaciones más vendidas, esto podemos analizarlo con la técnica de reglas de asociación.

Novel Corona Virus 2019 Dataset

Objetivo: Mostrar información relevante del número de casos observados, confirmados y el número de muertos, por provincia, edad de las personas con coronavirus.

Problema planteado: Predecir el número de muertos en una ciudad, según las variables más importantes en una persona.

Solución: Podremos analizar la información recabada de esta base de datos y con la técnica de regresión múltiple podremos ajustar una ecuación donde la variable de respuesta sea el número de muertos de un país y las variables regresoras sean la edad, genero, región etc. Podremos analizar que variables son las que mejor describen al modelo, así descartando las que no se adecuan al modelo, lo que a su vez nos podrá ayudar para que el gobierno tome la decisión en que ámbito deberá tomar mejor medida, por ejemplo, si un regresor que mayor influye es la edad, poder centrarse en ese rango de edad y tomar medidas en torno a la población

Wine Reviews

Objetivo: Determinar la calidad de un buen vino.

Problema planteado: Al determinar la calidad de un buen vino en base a las características más importantes.

Solución: Emplear un modelo de regresión, donde podremos predecir que variable es la que predice mejor el modelo. De igual forma esta información nos podrá ser útil, en caso de que alguna empresa decida irse mejor por obtener ganancias del vino a ser reconocido por la calidad de este mismo, pues de la misma manera podremos determinar que variables son más importantes al momento de que un cliente elija un vino, esto se puede determinar con una técnica de asociación o de patrones secuenciales, es decir, si el vino viene de x región o es vendido en x lugar con un determinado rango de precios podrá tener mayor probabilidad de ser consumido.

Iris Species

Objetivo: Detectar anomalías en el crecimiento de una planta

Problema planteado: Puede resultar difícil determinar cuál es el patrón de crecimiento que una planta sigue, al igual que cuales son los elementos que le impiden desarrollarse correctamente.

Solución: A través de técnicas como la detección de outliers podremos analizar que planta es la que mostró un crecimiento diferente al resto, de esta forma como se tiene una serie de tiempo de la planta a lo largo de su vida de crecimiento, se podrá realizar una regresión para determinar con que tendencia tiende a desarrollarse y a su vez determinar cuales son los factores que están influyendo para que se desarrolle correctamente

Netflix Movies and TV Shows

Objetivo: Información relevante de las películas y programas de televisión que están disponibles en la plataforma, la cual, pueden ser de utilidad para directores/creadores de contenido al momento de plantearse la idea de lanzar un nuevo proyecto en la plataforma.

Problema planteado: Un director que decida que su película o serie de televisión de x género sea reproducida en la plataforma de Netflix.

Solución: Al momento de presentar un proyecto para una película independiente, la plataforma de Netflix necesita aprobar o no la propuesta, por lo que la información de la base de datos puede resultar de gran utilidad para el equipo de trabajo de la película o hasta para el mismo equipo de Netflix, pues podrán determinar por medio de un clustering separado por región o género del filme, cuáles son los que cuentan con mayor aceptación o los que más se adaptan a su plataforma, de este modo pueden seguir teniendo el mismo crecimiento que hasta ahora tienen y a su vez, si comúnmente tienen más películas de x género, al aceptar una con características similares, el televidente tiene más probabilidad de llegar a ver esa misma película, lo cual le conviene tanto a la plataforma como al equipo de trabajo de la película para saber con anticipación si tiene probabilidad de ser aceptada o no.