



Universidad Autónoma de Nuevo León

Facultad de Ciencias Físico Matemáticas



FCFM

Semestre: 7

Análisis bases de datos

Materia:

Minería de Datos Gpo 002

Maestro:

Mayra Cristina Berrones Reyes

Alumno:

Jesús Ramón Castro Hernández

1887860

Contenido

Google Play Store Apps	3
Objetivos.....	3
Problema Planteado	3
Solución	3
Novel Corona Virus 2019 Data set.....	3
Objetivos.....	3
Problema Planteado	3
Solución	3
Wine Reviews	4
Objetivos.....	4
Problema Planteado	4
Solución	4
Iris Species	4
Objetivos.....	4
Problema Planteado	4
Solución	5
Netflix Movies and TV Shows	5
Objetivo	5
Problema Planteado	5
Solución	5

Google Play Store Apps

Objetivos

Una empresa desarrolladora de videojuegos en para móviles desea saber cuál es el precio ideal para lanzar su aplicación al mercado. El objetivo del análisis es dar una recomendación a la empresa en materia de pricing.

Problema Planteado

El problema para esta investigación es generar una relación entre el precio de la aplicación, relacionada a su tamaño (en MB), haciendo enfoque en las aplicaciones que cumplan con características de categoría de aplicación y su formato, de costo.

Solución

Para la solución de este problema se propone generar una ecuación a través del método de Regresión, donde limpiaremos la base de datos filtrando por categoría dejando solo GAMES, y filtrando por su tipo de aplicación PAID. Tras eso, utilizaremos las variables de SIZE expresado en megabytes y la variable PRICE expresada en dólares. Con esto podremos establecer una ecuación lineal que relacione las variables, para así otorgar un precio justo.

Novel Corona Virus 2019 Data set

Objetivos

El objetivo es dar el conocimiento del riesgo de fallecimiento relacionado a la edad del paciente.

Problema Planteado

Un país desea conocer la relación de los fallecidos de COVID con sus edades, para enfocar sus estrategias de prevención a los sectores más vulnerables, haciendo este análisis para cada sexo.

Solución

Para la solución se recomienda aplicar los métodos de clasificación, con la clasificación bayesiana, donde podemos estimar las probabilidades a través del teorema de Bayes. Esto

nos dará una mejor perspectiva de los grupos con mayor riesgo, considerando las probabilidades de muerte, en base a las edades y el sexo.

Wine Reviews

Objetivos

El objetivo es encontrar a que catador conviene invitar a catar nuestro vino dependiendo de la variedad de las uvas utilizadas en la fabricación del mismo.

Problema Planteado

Un viñedo busca conocer cuál es el catador de vino que le conviene a la empresa para que califique su vino, en base al tipo de uvas que utilizan para fabricarlo. En base a esto, organizar sus estrategias de marketing y maximizar sus buenas notas.

Solución

Para la solución de este problema, se propone utilizar los métodos de la regla de asociación, de sea A dado B, donde podemos separar las calificaciones por rangos específicos, y utilizar las variables variedad y de catador, para así obtener la mayor cantidad de combinaciones, asociando con que probabilidad cierto catador da una alta calificación en base a la variedad de las uvas utilizadas. Con este proceso, podremos dar una recomendación al viñedo para mejorar sus estrategias.

Iris Species

Objetivos

El objetivo es organizar las plantas, en base a sus características, de forma que podamos identificar datos atípicos, es decir, ver si alguna planta de nuestro invernadero presenta características inusuales y estudiarla más de cerca.

Problema Planteado

Un invernadero recibió un cargamento de 150 plantas, separadas en distintas variantes de la especie IRIS, la intención del dueño del invernadero es analizar cuales no cumplen con los estándares de cada especie para solicitar un cambio con su proveedor.

Solución

Para la solución, se recomienda aplicar el algoritmo de Centroid Based Clustering, para obtener los clústeres, y a través de las agrupaciones, analizar los datos que quedaron fuera de los grupos de su especie, y proceder a solicitar su cambio para mantener el estándar de la especie.

Netflix Movies and TV Shows

Objetivo

Usando esta base de datos, el objetivo es visualizar las características de las películas y series, para darles una expectativa a los productores.

Problema Planteado

Una casa productora de películas y series desea conocer la cantidad de temporadas de una serie o la duración de la película dependiendo de su género, para así, presupuestar en base al promedio cuanto tendrán que invertir en un proyecto de dicho género.

Solución

A través de la visualización de datos, podemos obtener los promedios de duración para cada uno de los géneros, filtrando entre películas y series. Una forma cómoda es mostrar los datos en gráficos para ver donde se concentra la duración en las películas con un histograma o una gráfica de barras para la cantidad de temporadas por serie.