



UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FCFM

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS



UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

BASES DE DATOS

MINERÍA DE DATOS

GPO: 01

Equipo:

| | |
|---|----------------|
| Leonardo Daniel Castillo Rodríguez | 1844452 |
| Hector Cedillo Charles | 1851642 |
| Gustavo Juan Martínez Ortuño | 1877422 |
| César Alejandro Morales Salazar | 1869864 |
| José Eduardo Navarrete Rosales | 1851701 |

Base de datos: Google Play Store

Hipótesis:

Se puede analizar la información para conocer las preferencias de las personas sobre el tipo de aplicaciones que usan basándonos en la categoría, rating, reviews, etc. Asimismo podemos analizar por qué las personas prefieren descargar aplicaciones gratuitas antes que pagar por una, independientemente de la calidad de dicha app.

Problema Planteado:

Las apps que son de pago representan un porcentaje muy bajo de la población, por lo que uno de los principales problemas es que si desarrollas una app y quieres que sea un éxito, es casi un hecho que debe ser gratuita, de otra forma no tendrá tantas instalaciones. Sin embargo hay personas que consideran que sus apps son tan buenas que no deberían ser gratuitas. Entonces, ¿se está perdiendo dinero al desarrollar una mejor app y que sea de pago antes que bajar su calidad y que sea gratuita?

Objetivo:

Hacer una comparativa de dos cosas: el precio promedio de una app de pago multiplicado por cierta cantidad de instalaciones vs lo que paga Google play store a un desarrollador por una app gratuita multiplicado por el número proporcional de instalaciones. Con dichos resultados se podrán crear gráficas de tendencias para corto, mediano y largo plazo para ver qué opción es más rentable y en qué tiempo.

Base de datos: Coronavirus

Hipótesis:

Se puede conocer la región donde se han confirmado más casos de Covid-19 y de esta forma saber en qué región es más probable que una persona se contagie de Covid-19

Problema Planteado:

Debido a que el virus es altamente contagioso y a la irresponsabilidad de las personas se ha provocado la actual pandemia que nos ha afectado de muchas formas a todas las familias del mundo

Objetivo:

Dar una idea en base a los datos acerca de las zonas del mundo más afectadas por el virus del Covid-19 y por ende darle a conocer a esas zonas las medidas preventivas necesarias y saber que es altamente probable contagiarse en esa zona mencionada

Base de datos: Críticas de vinos

Hipótesis:

Se puede obtener información sobre el vino más popular, vendido o con mayor calificación de una zona en específico.

Problema Planteado:

Identificar el vino adecuado para cada usuario de una región es una labor complicada, ya que hay muchos tipos de vino con diferentes características generales, por lo que muchas veces el vino no es bien recibido por una población, y esto se puede ver fácilmente viendo las calificaciones que le dan a cierto vino, siendo muchas veces bajas, lo que como principal consecuencia afectará a las ventas de los mismos.

Objetivo:

Determinar el mejor vino que hay para cada localidad, de acuerdo con la ubicación, variedad y precio, basado en los países que más compren vino y con ello, poder maximizar las ventas de vino.

Base de datos: Clasificación de plantas

Hipótesis:

Se plantea el realizar un modelo predictivo que dé como resultado el tipo de tulipán con el que se está trabajando, utilizando como datos de entrada los tamaños de cada parte de éste, de tal manera que no sea necesario conocer sobre botánica para nombrar la especie, sino sólo realizar medidas.

Para lograrlo, se pretende establecer intervalos de confianza para las medidas de cada parte del tulipán de acuerdo con los datos observados en la base, y programar el modelo en algún lenguaje de programación que parezca pertinente.

Problema Planteado:

A una persona le podría resultar difícil identificar un tipo de tulipán con sólo verlo, especialmente si no tiene conocimientos suficientes sobre biología o botánica.

Es necesaria la programación de una herramienta que facilite la identificación de cada tipo o especie de tulipán sin que sea necesario tener conocimientos avanzados de botánica.

Objetivo:

Clasificación de especies de tulipanes por tamaño.

Base de datos: Shows de Netflix

Hipótesis:

Se puede determinar el mejor show o película para el usuario, de acuerdo a la categoría, duración y país.

Problema Planteado:

Identificar las películas o shows adecuados para el usuario, para ello, se dispone de datos como las categorías, la duración y el país, dado lo anterior se puede realizar una selección de los posibles gustos hacia la persona, para así poder tener una mayor aceptación y popularidad contra la competencia.

Objetivo:

Realizar un análisis sobre los factores antes mencionados en donde se pueda observar las preferencias de los usuarios de la plataforma. Además, con ayuda de los datos recabados dentro de la plataforma ir actualizando o guardando que es lo que más gusta y que es lo que no gusta a los usuarios, para así poder realizar una mejor selección.