



UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN



Análisis bases de datos

Fase 2

Docente: Mayra Cristina Berrones Reyes

Keila Rubí Puente González

1807864

14 de octubre de 2020

Google Play Store Apps

Objetivo: Mejorar la calidad de las aplicaciones para que nuestros usuarios estén satisfechos y nos prefieran sobre otras aplicaciones.

Problema Planteado: Muchas de las veces las aplicaciones presentan algunos errores o son poco amigables con el usuario, lo que ocasiona que les den un *Rating* bajo, y que los *Reviews* sean negativos, además, con lo mencionado anteriormente lo que los usuarios hacen es desinstalar la aplicación. Esto ocasiona que el nombre del desarrollador o bien, el de la empresa, queden “marcados” como poco competentes, ya que las personas le comentarán a sus conocidos sobre su mala experiencia y pasará “de boca en boca”, por lo tanto, buscarán mejores opciones.

Solución: Queremos que nuestra aplicación esté de “boca en boca” pero de una manera positiva, para esto primeramente podemos utilizar las *Reglas de Asociación* para analizar que sea correcta la relación que existe entre las columnas de *Rating*, *Reviews*, e *Installs*. Las reglas de asociación aplicarían de la siguiente forma, si tiene un buen *Rating*, entonces tendrá un buen *Review*, y esto implicaría que más usuarios se sentirían seguros de descargarla, ya que las personas la han recomendado. Pero es más seguro analizar más sobre las columnas *Rating* y *Reviews*, ya que el número de instalaciones puede ser un poco engañoso por lo mencionado anteriormente, que puede ser que a simple vista a las personas les agrada la aplicación y tenga muchas instalaciones, pero el problema surge al empezar a utilizarla y es donde se genera la desinstalación de esta por parte del usuario.

Una vez que tengamos detectadas las aplicaciones donde algo anda mal, entonces trabajamos con ellas para la mejora de estas.

Novel Corona Virus 2019 Dataset

Objetivo: Analizar las circunstancias del país que presentó mayor número de personas recuperadas, así como las medidas que adoptaron y con estas conocer el tiempo que les tomó salir de la pandemia.

Problema Planteado: Después de todo la enfermedad resultó diferente a lo que se había planteado inicialmente, es decir, se consideraba que afectaba a adultos mayores sin embargo, afectó incluso a niños y a adolescentes, además, se descubrió que la mortalidad en todas las edades era la misma al menos que alguno contara con alguna enfermedad como diabetes, obesidad, etc., entonces aumentaba la mortalidad. Por lo tanto, a veces no resulta predecible un virus, mucho menos si es nuevo, por esto es necesario conocer cómo trabajó en esto el país que presentó mayor número de personas recuperadas, para

que cuando vuelva a suceder alguna pandemia tengamos una idea de cómo actuar de manera más eficiente.

Solución: Para el análisis las columnas que principalmente nos ayudarán son *Recovered* y *Country Region*, podemos hacer uso de varias técnicas según nuestro enfoque, realmente hay varias cosas que podemos hacer con una base de datos, solo necesitamos estar desde alguna perspectiva, pero para el objetivo que se ha planteado podemos utilizar la *Visualización de datos*, de esta forma será más sencillo observar el comportamiento de las personas recuperadas.

Wine Reviews

Objetivo: Conocer en qué país nos conviene invertir para un viñedo rentable y qué precio ponerle a nuestro vino, de esta manera se podrá lograr una mejor posición en el mercado de los vinos.

Problema Planteado: Supongamos que somos un grupo de personas que quieren iniciar una empresa de vinos y contamos con la base de datos mencionada, esta nos ayudará a estudiar los vinos que tienen mayores puntuaciones, asimismo podemos checar el país de donde proceden, de esta forma estaremos analizando las preferencias de las personas y todo esto nos será útil 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 emplearla para clasificarlos en rangos de puntajes y sea más sencillo analizar las preferencias de las personas, para esto será necesario utilizar la columna *Points*, y una vez que lo hemos elegido podemos 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 nuestras circunstancias como empresa.

Iris Species

Objetivo: Clasificar de manera rápida y eficiente las plantas de Iris que en algún momento tengamos que registrar para algún estudio o algún análisis en particular.

Problema Planteado: Tenemos una base de datos que nos da información de las características que presentan los tres tipos de clasificación de plantas de Iris, ahora

supongamos que tenemos que desarrollar un programa para que cuando nos digan que registremos alguna planta nos diga directamente en qué clasificación entra y asimismo que podamos utilizar los datos almacenados para algún estudio que se tenga que hacer por ejemplo en biología, que tal vez quieren analizar si encuentran más relación entre las clasificaciones de las plantas o si encuentran más propiedades que puedan diferenciarlas aún más.

Solución: Para el problema planteado podemos utilizar los *Árboles de clasificación*, resultará más sencillo por el tipo de datos que tenemos, recordando cómo funcionan estos, consiste en hacer preguntas del tipo $\{x_k \leq c\}$ para las covariables cuantitativas o preguntas del tipo $\{x_k = nivel_j\}$ para las covariables cualitativas, de esta forma el espacio de las covariables es dividido en hiperrectángulos y todas las observaciones que queden dentro de un hiperrectángulo tendrán el mismo valor grupo estimado, aquí es donde entran las características de cada una de las clasificaciones y recordemos que estos árboles de clasificación se hacen por medio de algún programa y no a lápiz y a papel.

Netflix Movies and TV Shows

Objetivo: Analizar los gustos principales del público por país para proporcionar una mayor satisfacción a los usuarios y una mejor experiencia.

Problema Planteado: Queremos obtener mayores ganancias, entonces queremos aprovechar al máximo la base de datos que nos ha sido proporcionada.

El problema puede plantearse en que a veces los usuarios no encuentran algo que los atrape para que se queden mucho tiempo mirando del contenido de entretenimiento con el que se cuenta, entonces queremos saber cómo podemos hacer para aumentar las ganancias por temporada, es decir, según sea Navidad, Halloween, Semana Santa, por mencionar algunas, ya que las personas usualmente están influenciadas por lo que está de moda podemos sacar ventaja de esa forma, de igual manera se puede analizar el director que mas le gusta a la gente.

Solución: Podemos hacer uso de *Patrones Secuenciales* porque como son eventos que se enlazan con el paso del tiempo, analizaremos los datos y se encontraremos subsecuencias que nos pueden resultar muy interesantes dentro de un grupo de secuencias que vendrían siendo los usuarios. De esta forma sabremos qué podemos recomendarles y tenemos más certeza de que les guste y se queden más tiempo mirando series o películas. Para analizar el director que suele gustarle más a las personas por país podemos hacer el uso de la *Visualización de datos*, y además al obtener esta información podemos saber de cuál director comprar los derechos de más películas o series.