



UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN®



FCFM

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

# Análisis BD

---

Nombre	Matricula
Miguel Alejandro Noriega Ortega	1798528

**Maestra:** Mayra Cristina Berrones Reyes

**Materia:** Minería de datos

**Fecha:** 13/10/2020

**Nombre de la base de datos:** Google Play Store Apps

**Objetivo:** Facilitar el proceso de análisis de los factores que influyen en el que una aplicación sea exitosa en la tienda

**Problema planteado:** La cantidad de aplicaciones que están actualmente en la tienda son demasiadas lo cual hace complicado el identificar que factores son los que hacen que una aplicación tenga éxito, ya sea su precio, el nombre, su categoría, su tamaño, publico al que va dirigido entre otros, por lo cual me gustaría hacer el proceso mas fácil para así identificar más rápido estos factores

**Solución:** Por medio de clustering y reglas de asociación creo que se podría llegar a encontrar los factores que pudieran ser clave para que una de estas aplicaciones tenga éxito en la tienda, ya sea su tema, su precio , su clasificación o si es una combinación de todos estos y cuanto influye cada uno en el éxito de la misma , de esta manera se podría identificar fácilmente cuales son los aspectos mas importantes a considerar y asegurarse de ser excelentes en ellos para así mejorar las probabilidades de tener éxito en la tienda , habría que analizar las aplicaciones que tienen el mejor rating , las mas descargadas y el mayor numero de reviews para poder analizar que es lo que tienen en común, creo que la cantidad de reviews y la cantidad de descargas reflejan el que tan buenas podrían ser mas allá del rating ya que el rating puede ser de 5 pero si solo la calificaron 10 personas pues no es muy significativo en cambio una calificación de 4 o 4.5 calificada por 2k personas o con muchas descargas podría indicar que de verdad es buena por algo

**Nombre de la base de datos:** Novel Corona Virus 2019 Dataset

**Objetivo:** El poner en números claros que tan grave es la enfermedad basándonos en su duración, los síntomas, cuando fue al hospital y las fechas en las que inicio y termino con los síntomas y el resultado final del paciente para ver si se curo o falleció ya que al tener los datos se puede ver que porcentaje de los pacientes que se enferman logran curarse, cuanto duran con síntomas y aproximadamente cuanto tardan en ir al hospital desde que presentan los síntomas, de esta manera se puede tener en revisión a una persona con síntomas si esta ya se encuentra en los días en los que la enfermedad suele empeorar o complicarse para darle una atención mas seria, además con los datos de los pacientes fallecidos se puede encontrar patrones que pudieron hacer que la enfermedad se complicara, ya sea la edad, el no acudir a tiempo al hospital, entre otros.

**Problema planteado:** La enfermedad del covid-19 a ocasionado muchos problemas a través de estos meses, con esta base de datos podemos encontrar los patrones de comportamiento de la enfermedad para así poder contrarrestarla de manera mas eficaz, cuanto tarda en manifestarse, cuando es que se complica, que factores la pueden complicar y que se puede hacer para mejorar la posibilidad de recuperación.

**Solución:** Por medio de reglas de asociación y patrones secuenciales se podría analizar a los pacientes curados y fallecidos para poder averiguar que se hizo bien o mal para llegar al resultado final de la enfermedad, si se recupero o falleció para así poder terminar con una serie de datos concretos sobre la enfermedad y el como tratarla de manera eficaz, salvando así muchas vidas en un futuro

**Nombre de la base de datos:** Wine Reviews

**Objetivo:** Encontrar la mejor variedad de vinos según el sabor o su tipo para después encontrar que es lo que realmente los vuelve buenos ya que por lo general al momento de adquirir un producto nos influenciamos por el hecho de que el ser mas caro significa que es mejor sin embargo viendo la base de datos podemos ver que el mejor calificado no es el mas costoso por lo tanto seria muy útil identificar que factores son los importantes al momento de escoger un vino para así como consumidor poder hacer mejores decisiones y como dueño de un restaurante talvez dejar de pagar tanto por los vinos caros sabiendo que el precio no siempre es el mejor parámetro para evaluar la calidad

**Problema planteado:** El escoger un vino puede ser una tarea complicada si no se conoce del tema, con tantos tipos y variantes que existen muchas veces caemos en dejarnos guiar por el precio sin embargo ese no es necesariamente el mejor indicador por lo cual creo que seria útil el analizar la base de datos para poder verificar cuales son los factores que realmente hacen que el vino sea bueno , ya sea la provincia, la región, el viñedo, el sabor , el país, etc

**Solución:** Por medio de la clasificación y las reglas de asociación creo que seria posible talvez dividir en categorías a los vinos como talvez por país o por sabor y después por medio de reglas de asociación identificar que es lo que hace bueno a los mejores vinos de cada categoría para así poder tomar mejores decisiones a la hora de escoger un vino y no solo dejarnos guiar por el precio el cual por lo general podría tener mas que ver por una marca o por temas de publicidad mas que por la calidad del producto en si

**Nombre de la base de datos:** Iris Species

**Objetivo:** Poder identificar las plantas que tuvieron un crecimiento fuera de lo común ya sea muy grandes o pequeñas en comparación con las demás de su especie para así poder estudiar a fondo su proceso de cuidado y plantado a través del tiempo de manera que podamos mejorar la forma en la que lo hacemos para así tener un proceso mas eficiente y si se quiere poder generar como resultado plantas mas grandes o más pequeñas según se prefiera

**Problema planteado:** Revisando la base de datos podemos notar que en algunos casos el tamaño de los sépalos y los pétalos es anormalmente grande o pequeño comparado con las demás plantas por lo cual seria interesante identificar estos casos y hacer un mejor estudio de que se les hizo o cual fue la causa de este crecimiento anormal

**Solución:** Por medio de clasificación podemos dividir a las plantas según su especie y después por medio de la Detección de outliers podemos identificar esos casos en los cuales el crecimiento es anormal comparado con las demás de dicha especie para investigar a fondo que fue lo que causo esta anomalía en la planta y como se puede corregir

**Nombre de la base de datos:** Netflix Movies and TV Shows

**Objetivo:** Poder encontrar los factores que ocasionan que una película o serie tenga una buena calificación por los críticos, por lo general una serie o película es juzgada desde antes de ser vista por las personas esto en base a los actores o el director que participan en ellas , sin embargo esto también puede ser contraproducente ya que al crear una imagen y una expectativa de la misma, al momento en que se ve el trabajo se puede terminar decepcionado al no ser lo que se esperaba , lo cual puede terminar en malas críticas en algunos casos infundadas.

**Problema planteado:** Identificar la mejor formula para poder tener un éxito en Netflix y en el cine analizando el peso de los actores, el director, la duración de la película entre otros factores para determinar cuál es la característica más importante para considerar al momento de crear una película para de esa manera poder invertir mas en los aspectos que mas le importan a la gente.

**Solución:** Para llegar a este objetivo se tendría que usar una base de datos secundaria la cual incluya calificación por parte de alguna pagina como lo puede ser IMDB o Rotten Tomatoes y después ya combinando ambas por medio de clasificación dividir las series de las películas y posteriormente dividir las películas y series por género , con clustering encontrar los grupos que presentan características similares y por ultimo con reglas de asociación definir que es lo que tienen en común estas películas y series que las llevaron ya sea a fracasar o a tener éxito.