

## **MINERÍA DE DATOS**

## **BASES DE DATOS**

**FLOR KARINA JUÁREZ RODRÍGUEZ**  
**1802920**

**M.C. MAYRA CRISTINA BERRONES REYES**

**SEMESTRE 7**  
**LICENCIADO EN ACTUARÍA**

**SAN NICOLÁS DE LOS GARZA A 14 DE OCTUBRE DE 202**



**Base de datos:** Google Play Store Apps

**Objetivo:** Crear una aplicación para ayudar a los estudiantes en sus actividades escolares.

**Problema planteado:** El uso de herramientas tecnológicas en las escuelas es cada vez más común, pero eso no quiere decir que en el mercado se encuentren las mejores aplicaciones que cumplan con las necesidades del estudiante. Hay ocasiones en las que el estudiante necesita realizar cierta acción y busca una aplicación en Google Play que lo ayude, pero se encuentra con apps de baja calidad, que no funcionan bien, no dan el resultado que se espera, entre otras cosas.

**Solución:** Lo que se puede hacer con ayuda de la base de datos es, primero filtrar por categoría el área que nos interesa (en este caso sería Education) y tomar en cuenta el rango de edad puesto que queremos enfocarnos en estudiantes que tengan un mayor conocimiento del uso del celular (aplicaciones para 12+ aprox.), después de ese filtro lo que seguiría sería buscar las apps peor calificadas y ver en críticas el por qué; una vez que se tenga esa información hacer un análisis de cuál es la que nos conviene tomar como base para crear nuestra propia versión de la app pero mejorada tomando en cuenta los aspectos negativos de las aplicaciones parecidas a la que se quiere crear.



**Base de datos:** Novel Corona Virus 2019 Dataset

**Objetivo:** Saber cuál es el país con más recuperados de COVID-19.

**Problema planteado:** Sabemos que el COVID-19 es algo que ha venido a cambiar la vida a todo el mundo, desde que se comunicó la pandemia hemos tomado muchas medidas de prevención para evitar el contagio, pero desgraciadamente en ciertos casos eso no fue suficiente o también tenemos a la gente que simplemente no se cuidaba y se contagió; pero a pesar de tener este virus que es nuevo para todos hay gente que se logró recuperar y es justo este sector en el cual nos queremos enfocar. Queremos saber cuál fue el país que tuvo más recuperados de COVID-19 y a qué se debió eso.

**Solución:** Tomando la base de datos, seleccionaríamos los datos del apartado de recovered y veríamos cuál país tiene más recuperados de coronavirus hasta cierta fecha, pero como sabemos, cada país tiene diferente número de habitantes se tendría que hacer una proporción para que nuestros datos de comparación tengan un poco más de sentido. Una vez que tengamos nuestro objetivo, podríamos buscar qué es lo que hicieron en ese país y en otros no para que la recuperación fuera la más rápida.



**Base de datos:** Wine Reviews

**Objetivo:** Saber si el catador influye en el precio.

**Problema planteado:** Hay personas que tiene mucha influencia en todo lo que hacemos, muchas veces compramos cierto producto porque cierta persona que nos gusta mucho o cierto experto dijo que era bueno y no nos importa nada más que su opinión e incluso dejamos a lado el detalle del precio. Esto hizo que surgiera la duda de si el precio de los vinos es influenciado por el catador que lo califico.

**Solución:** Dada la base de datos filtramos todos los catadores que se tengan y analizamos los precios que tiene las botellas de vino que calificaron, ya que tengamos un rango de precios por cada catador podemos comparar. Si cierto catador tiene un rango de precios muy alto podemos decir que si influye quién califica el vino a la hora de ponerle precio; además si esto se llega a cumplir podemos hacer una investigación para complementar por qué la opinión de cierto catador pesa más que la de los demás. Otra cosa que podemos analizar si tomamos la columna de country es ver si los catadores califican mejor los vinos de cierto país.



**Base de datos:** Iris Species

**Objetivo:** Comparar los tres tipos de especies en diferentes aspectos

**Problema planteado:** En ocasiones necesitamos saber ciertos datos en específico de cierta cosa, en este caso hablamos de las especies de Iris que nos muestra la base de datos. Los datos que nos brindan nos pueden ayudar a comparar y ver en qué aspecto resalta mas una especie que otra aún y siendo de la misma familia.

**Solución:** Dado que la base nos da los datos de la longitud y anchura de los pétalos y el sépalo de diferentes muestras de flores de la familia Iris, podemos hacer uso de la visualización de datos y mostrar los resultados que nos dan de forma gráfica (buscando que sea el gráfico más apropiado para la comparación), para comparar las diferentes características que tienen las tres especies. Al hacer esto nos podemos dar cuenta si tal especie tiende a tener los pétalos más anchos o largos que las otras dos especies, y así podemos ir comparando todos los datos que nos dan; aunque también podemos obtener que las especies no siguen alguna tendencia en cierta característica.



**Base de datos:** Netflix Movies and TV Shows

**Objetivo:** Identificar si se hacen más películas o series según la clasificación por edades

**Problema planteado:** Si se planea realizar una producción, pero no se sabe si optar por una película o una serie, pero se sabe a qué rango de edad se quiere dirigir se puede tomar esta información como base de apoyo para tomar la decisión más conveniente.

**Solución:** Tomando las columnas de type y rating podemos hacer un algoritmo que nos separe las clasificaciones de edades y una vez hecho esto que nos indique qué es lo que se produce más para cada rango de edad. Si tuviéramos un poco más de información sobre lo que prefiere la audiencia podríamos ampliar el uso de este algoritmo para otros casos, por ejemplo, si se tiene dudas en dónde producir la película, qué directores prefiere la audiencia, la durabilidad de la película o capítulo según sea el caso. Todo esto se pudiese mostrar de manera más fácil haciendo uso de algún gráfico que nos permita comparar todas las edades y sus gustos.