

## **Minería de datos**

**Mayra Cristina Berrones Reyes**

<b>Nombre:</b>	<b>Matrícula:</b>
Ávila Rodríguez Delfino Azael	1861240
Gutiérrez Herrera Luisa Fernanda	1794593
Montemayor Guerra Diego Marcelo	1847156
Ovalle Salas Britany	1852645
Sánchez Bracho Joss Alberth	1796551

San Nicolás de los Garza, Nuevo León a 16 de agosto de 2021.

---

### Google Play Store

---

- **Hipótesis:** Se puede determinar cuáles son de las mejores apps para cada categoría, esto tomando en cuenta la relación de la cantidad de descargas junto con la calificación y las reviews, también tomando en cuenta la naturaleza de estas últimas.  
Para llevar a cabo esta tarea podemos ir filtrando por categoría, luego ordenar por las que tengan una mayor calificación y finalmente poner un segundo orden la cantidad de descargas. Esto porque es importante ver si el puntaje se debe a que tiene pocas calificaciones buenas o si tiene muchos usuarios que realmente están felices con la app.
- **Problema planteado:** Se puede tomar algo de tiempo para escoger la mejor aplicación para un uso en específico, esto por nuestros estándares o necesidades. Con esto se planea tomar la mejor app para cada género, ya sea filtros y edición de fotos y videos, y también se puede optar por si es viable pagar por una app según nuestro análisis.
- **Objetivo:** Tomar una relación la cual nos diga cual app es mejor para el tipo de uso que le daremos.

---

### Coronavirus

---

- **Hipótesis:** Es posible determinar cuál fue un mejor enfoque para evitar la propagación de la pandemia, por un lado, los países que pusieron medidas más restrictivas a la población, como China y Japón, y por el otro, países que no quisieron vulnerar las libertades de la población como India y Estados Unidos. El principal beneficio que este análisis podría generar es crear una mejor estrategia para evitar la propagación de un virus entre la población.  
Para llevar a cabo esta tarea, es posible filtrar la base de datos por la columna Country/Region con los países que queramos, y revisar el número de casos confirmados en la columna Confirmed.
- **Problema planteado:** Se ha reconocido que el imponer medidas más relajadas para el control de la propagación del virus COVID 19, como el que no sea obligatorio el uso de cubrebocas, un libre tránsito en el país, el no cerrar fronteras, entre otras medidas, ha hecho que el impacto de la pandemia no haya disminuido.  
Fuente: [https://www.lavozdegalicia.es/noticia/especiales/aniversario-pandemia/2021/03/07/china-reivindica-exito-control-pandemia/0003\\_202103G7E48991.htm](https://www.lavozdegalicia.es/noticia/especiales/aniversario-pandemia/2021/03/07/china-reivindica-exito-control-pandemia/0003_202103G7E48991.htm)

- **Objetivo:** Identificar qué países fueron los que tuvieron un menor número de casos positivos de COVID 19 a lo largo de la pandemia. Si fueron Japón y China, o Estados Unidos e India.

---

### *Críticas de vinos*

---

- **Hipótesis:** Es posible determinar aquellos viñedos que puedan producir uvas de alta calidad y, entre los diversos efectos positivos que se puedan generar, el principal beneficio es el siguiente: una mejor administración de los recursos como los terrenos, vides, materiales requeridos para la cosecha, entre otros. Para llevar a cabo dicha tarea, es posible usar información de la base de datos como la siguiente: # de puntos asignados por *WineEnthusiast*, descripción de las personas que probaron el vino o el precio que los consumidores están dispuestos a pagar.
- **Problema planteado:** Se ha reconocido que las uvas de baja calidad provocan que el vino producido tenga defectos y los viñedos son los responsables de la cosecha de dichas uvas.  
Fuente: <https://hogardevinos.com/defectos-del-vino-y-otros-problemas/>
- **Objetivo:** Identificar los 10 mejores viñedos que cosechan uvas de buena calidad.

---

### *Clasificación de plantas*

---

- **Hipótesis:** Se puede reducir el tiempo que tarda una persona en identificar con cuál especie de Iris (Setosa, Versicolor o Virginica) es con la que cuenta. Para llevar a cabo dicha tarea, se pueden considerar las columnas "Petal.Length" y "Petal.Width" que nos indican el largo y ancho de los pétalos por especie, ya que cada una de estas cuenta con distintas medidas que podrían facilitar su clasificación.
- **Problema planteado:** A primera vista, las distintas especies de Iris parecen ser la misma gracias a las similitudes que estas presentan y resulta difícil el poder diferenciarlas rápidamente.
- **Objetivo:** Identificar de manera más rápida con qué especie de Iris se cuenta dependiendo de sus características, por ejemplo, la medida del largo o ancho del pétalo que tienen.

---

### *Shows de Netflix*

---

- **Hipótesis:** Se pueden identificar a los directores que más títulos tienen en Netflix por país, lo cuál podría ser útil al momento de querer tener un listado en caso de tener nuevo proyecto.
- **Problema planteado:** Se ha estudiado que la mayoría de los consumidores de Netflix son personas entre 18 y 54 años por lo que invertir en una categoría que no se adapte a este público podría resultar desfavorable al igual que trabajar con directores que sólo han realizado títulos infantiles u otros que no se adapten a este rango de edad.  
*Fuente:* <https://www.marketinginsiderreview.com/estrategia-marketing-netflix/>
- **Objetivo:** Obtener el listado por país de los 10 directores con más títulos en Netflix con su respectivo título realizado y categoría a la que pertenece.