

## EQUIPO 10



## BASES DE DATOS

### TAREA 2

- 1867341 CRUZ LERMA JORGE ANTONIO
- 1864878 CERDA SERRATO ALAN RENÉ
- 1941475 GALINDO VÁZQUEZ JESÚS ALFONSO
- 1630606 VALENCIA GONZÁLEZ JESÚS EDUARDO

## 1.- Base de datos: Google Play Store Apps

**Hipótesis:** La categoría “Game” es la más descargada de todas las que hay en la base de Google Play Store.



**Problema planteado:** La base de datos no tiene ordenada correctamente las categorías esto hace que trabajar con esos datos sea complicado, ya que se reparten en otras categorías. Por ejemplo, el juego “ROBLOX” aparece en repetidas ocasiones y en diferentes categorías.

### **Objetivos:**

- Replantear las categorías para que estén mejor ordenadas.
- Limpiar los datos para que no se repitan en las demás categorías.
- Identificar si vale la pena desarrollar una aplicación de paga.

## 2.- Base de datos: Datos del virus corona 2019

**Hipótesis:** La cantidad de recuperados de COVID va incrementando, ahora se recuperan más, y la de muertos va en decremento, se puede suponer por las vacunas.



**Problema planteado:** Es una cantidad excesiva de datos y falta una columna de vacunados para poder determinar si a raíz de la vacunación el número de muertes disminuye y de personas curadas aumenta.

La vacunación varía en tiempo y cantidad por región.

### **Objetivos:**

- Recatar información para saber cómo está actuando el virus después de las vacunas.
- Identificar dónde está peor la situación del virus.

### 3.- Base de datos: Reseñas de Vinos

**Hipótesis:** La puntuación del vino es proporcional a su precio, en promedio entre más caro tiene mejor puntaje.

**Problema planteado:** No hay una columna de la edad de los vinos en cuestión, por lo que no sabemos qué tan añejado está y eso está relacionado con la calidad y precio del vino.



**Objetivos:**

- Al agregar una columna con el añejamiento del vino podemos asegurar que mientras más tiempo tenga el vino es mejor.
- Identificar si valer la pena comprar vinos los vinos más caros.

### 4.- Base de datos: Clasificación de plantas

**Hipótesis:** Es posible distinguir la especie de Iris (o lirio) conociendo el tamaño de sus sépalos y pétalos.

**Problema planteado:** Los lirios o iris se parecen mucho y tienen usos medicinales. También se usan en perfumes, pero la especie “setosa” puede perjudicar a la salud, por lo que, si no se tiene cuidado, se podría usar la especie equivocada.



Iris Versicolor

Iris Setosa

Iris Virginica

**Objetivo:**

- Identificar un rango de tamaños de sépalos y pétalos que permita identificar las diferentes especies de lirios o iris estudiados.

## **5.- Base de datos: Películas y series de Televisión de Netflix**

**Hipótesis:** Un informe interesante que muestra que el número de programas de televisión en Netflix casi se ha triplicado desde 2010, la cantidad de películas del servicio de streaming ha disminuido en más de 2.000 títulos desde 2010.

**Problema planteado:** Faltan cierto tipos de columnas, por ejemplo, vistas por país, géneros, clasificación, rating, entre otros, por lo que la información aquí es un poco limitada, se podría hacer una base mucho más extensa ya que tiene mucho potencial este tipo de datases bases.



### **Objetivo:**

- Mientras más columnas útiles se agreguen, mejores datos sobre las series/películas podríamos tener.
- Ver si la plataforma va en decadencia al sacar de su catálogo muchas películas y creando sus propias "series originales".