



Facultad de Ciencias  
Físico Matemáticas



# Base de datos



EQUIPO 9

INTEGRANTES:

KAREN GÓMEZ .  
CECILIA DUARTE.  
EMMANUEL GALVÁN.

GRUPO 003

FEB-22-21

MINERÍA DE DATOS

**Nombre de la base de datos: Aplicaciones de Google Play Store.**○ **Objetivo:**

- Eliminar aplicaciones inservibles.
- Tener más espacio en el servidor para aplicaciones de utilidad.
- Tener un catálogo de calidad.

○ **Problema planteado:**

- Google Play permite que existan aplicaciones dentro de su plataforma que tienen una baja calificación por ser poco útiles o engañosas. Como distribuidor no es correcto que se tengan productos malos dentro de la tienda.
- Debido a las aplicaciones ineficaces dentro de la tienda Play Store, se tiene un mal uso de la ocupación en el espacio del servidor.

○ **Solución:**

Ordenar las aplicaciones por calificación, número de reseñas e instalaciones. De manera que se obtenga aquellas aplicaciones que tengan baja calificación y gran número de reseñas e instalaciones. Esto nos da más certeza para reconocer aquellas aplicaciones que han sido menos útiles para los usuarios y por tanto se pueden eliminar de la tienda.

**Nombre de la base de datos: Nuevo conjunto de datos del virus Corona 2019.**○ **Objetivo:**

- Identificar índices de letalidad por zonas.
- Reducir la cantidad de muertos por el virus.



- **Problema planteado:** A finales de 2019, se dio a conocer la aparición de la enfermedad causada por el virus Corona 2019, que se extendió por todo el mundo. Debido a la letalidad de esta enfermedad, laboratorios de todo el mundo se dispusieron a elaborar vacunas para disminuir la letalidad.

En estos momentos se están distribuyendo las vacunas para el virus y se discute sobre cuál es la manera más eficiente de distribuirlas y repartirlas.

- **Solución:** Los estados en donde se tienen una mayor tasa de letalidad suponen la pérdida masiva de habitantes de la región por lo que se buscará analizar la letalidad que ha tenido el virus en las diferentes regiones, ciudades y países para localizar cuales han tenido una mayor mortalidad por el virus y ser las primeras en tener acceso a la vacuna. Esto a través del número de muertes y el número de recuperados por zonas.

### Nombre de la base de datos: Reseñas de vinos.



- **Objetivo:**
  - Maximizar la calidad de una selección de vinos.
  - Asegurar una buena elección de compra.
  - Reducir tiempos de investigación.
- **Problema planteado:** El vino en los restaurantes suele ser muy caro en comparación a su precio de tienda, la razón de esto es que el restaurante tiene que invertir en comprar vinos, tener una selección variada y no tienen la certeza de que todos se venderán, por lo que el sobreprecio permite cubrir los costos generados por los vinos que no se vendan.
- **Solución:** De cada uno de los vinos obtener su puntuación, variedad y su precio. Con estos datos analizar la mejor selección de modo que se compren los vinos con mejor puntuación de cada variedad sin dejar de lado el costo que representa.

### Nombre de la base de datos: Películas y programas de televisión de Netflix.



- **Objetivo:**
  - Distribuir de la mejor manera la inversión.
  - Ampliar el catálogo de una manera más eficiente.
- **Problema planteado:** Netflix es una plataforma de streaming de películas y series. Su principal objetivo es administrar su dinero para comprar los derechos y licencias del contenido más llamativo y exitoso, así como producir contenido propio. El problema es que el dinero es limitado por lo que dependiendo de las decisiones que se tomen pueden llegar más personas a la plataforma o irse.
- **Solución:** Con la información de la base de datos no es posible determinar cuáles han sido las mejores producciones, sin embargo, podemos descubrir cuáles son las mejores opciones por tipo y género según la popularidad entre los demás productores.  
Esto se puede lograr a través de un análisis de la cantidad de películas y series producidas por género y tipo a lo largo del tiempo.

**Nombre de la base de datos: Especies de iris.****○ Objetivo:**

- Mejorar el tiempo de distribución de la planta Iris.
- Reducir costos de operación.

**○ Problema planteado:**

- El tiempo de identificación de especies de plantas provoca un retraso para los viveros al momento de la distribución.

**○ Solución:**

- Desarrollar el prototipo de un aparato capaz de medir las plantas.
- Crear una herramienta de aprendizaje que sea capaz de identificar las plantas, siendo esta herramienta alimentada de la información en la base de datos.
- Si el prototipo funciona, este puede ser mejorado para identificar más plantas.