



Universidad Autónoma de Nuevo León

Facultad de Ciencias Físico Matemáticas

Minería de Datos

Prof. Mayra Cristina Berrones Reyes

“Análisis de las Bases de Datos”

Sergio Velázquez Rivera

1805244

Gpo: 003

16 de Octubre del 2020, San Nicolas de los Garza



## GOOGLE PLAY STORE APPS – DATA SET

---

**Objetivo:** Identificar una correlación de aplicaciones que sean de paga y las que no, que otorguen la misma función.

**Problema Planteado:** Al obtener esta relación estaríamos identificando que aplicaciones otorgan las mismas funciones que las que son de paga, y poder ofrecerle al usuario una interfaz de calidad con previews que sean agradables a la vista del usuario y que no se pierda en la interfaz entrando en categoría de visualización.

**Solución:** Desarrollar una función que a través de la base de datos clasifique las aplicaciones las aplicaciones por las herramientas que esta pueda proporcionar, revisando también su grado de rating y el número de descargas ya que estos datos nos proporcionarán las aplicaciones más populares para cada categoría de aplicaciones de pago vs gratuitas. Sobreponiendo las aplicaciones gratuitas que cumplan con lo anteriormente mencionado.



## CORONAVIRUS (SARS-CoV-2) – DATA SET

---

**Objetivo:** Intentar predecir la curva de descenso de los casos de contagio causados por el coronavirus (covid-19) en un país en específico.

**Problema Planteado:** Segmentar por agrupaciones a la población contagiada por edades y revisar cuales de estos estratos son los mas vulnerables asi como también los casos de éxito que se recuperaron, entraría en la categoría de predictivo.

**Solución:** Desarrollar un modelo matemático con ayuda de regresiones para el cual podamos revisar por segmentos a la población creando variables por edades y ver cuales son las mas influyentes en los casos con éxito, asi mismo, teniendo muy en cuenta los datos atípicos los cuales serían los asintomáticos para un país en específico, suponiendo que este se comporte como un grupo sin reemplazo, buscando los casos en los que si se pudieron recuperar de la enfermedad y pudieron desarrollar alguna inmunidad al virus.



## WINE REVIEWS – DATA SET

---

**Objetivo:** Identificar y catalogar la ubicación de las mejores cosechas con las que cuentan los viñedos de las zonas más reconocidas.

**Problema Planteado:** Para los consumidores de vinos, estamos hablando de compradores nos resulta un tanto difícil el saber donde conseguirlos revisando sugerencias y cuál es su procedencia.

**Solución:** A través del cluster podemos clasificar cuales son los mejores lugares a partir de su puntuación y calidad de vinos, para un restaurante le podría resultar útil tener esta información para tener un buen lugar de excelente procedencia para poder importar sus listas de vinos y así tener un mejor inventario.



## IRIS SPECIES– DATA SET

---

**Objetivo:** Ayudar a científicos herbolarios o botánicos a clasificar de una manera más rápida el estudio de las plantas y su desarrollo conforme el paso del tiempo.

**Problema Planteado:** Muchos agricultores o herbolarios necesitan saber el paso de maduración sexual de su planta en cuestión, para poder obtener correctamente su cebo o sábila en cuestión que pueda ayudar a las personas

**Solución:** Con ayuda de el promedio de los tallos u hojas de las plantas podemos revisar cuales son las que ya alcanzaron su madurez y están listas para ser utilizadas en la medicina y su distribución, así también revisar algunos datos atípicos que se puedan presentar como lo son hojas o tallos que tengan diferentes mediciones o que simplemente se trata de otra planta ya que con el promedio de las medidas nosotros esperaríamos que se trate de una planta pero puede que estemos hablando de otra.



## NETFLIX MOVIES & TV SHOWS – DATA SET

---

**Objetivo:** Analizar el crecimiento de los programas originales de Netflix y ver el como se esta aceptando por la audiencia y sus usuarios.

**Problema Planteado:** Buscar mejorar la experiencia al usuario al aumentar series o películas originales que sean de interés para la audiencia, entrando en las categorías con mejor rating.

**Solución:** Clasificar los programas y géneros con mejor rating en la base de datos y enfocarnos en las categorías que sean tendencia, ya que las visualizaciones se presentan por momentos, tendríamos que enfocarnos en darle al usuario lo que busca, recomendando los mejores programas del momento y en dado caso de algún dato atípico, que no se salga tanto de la agrupación por elemento.