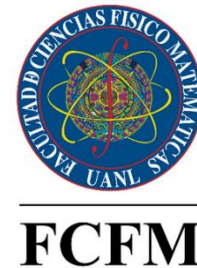




Universidad Autónoma de Nuevo León  
Facultad De Ciencias Físico Matemáticas



## Fase 2 Base de datos 5

### Minería de Datos

#### **Maestra:**

Mayra Cristina Berrones Reyes

#### **Alumnos:**

Francisco García Sánchez Armáss

#### **Matricula:**

#1816358

**Grupo: 003**

**Aula: AV12**

San Nicolás de los Garza, Nuevo León, México a 11 de octubre del 2020



Google Play  
Store

## Aplicaciones de Google Play Store

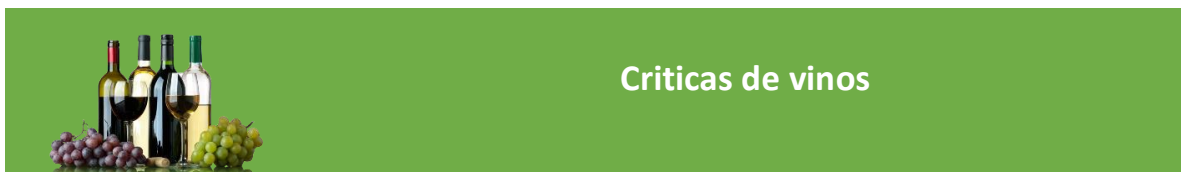
<b>Objetivo:</b>	Diferenciar cuando una aplicación es gratis o de pago, he aquí cuando entra el término “pay to win” el cual entraría como una tercera categoría. Entonces podríamos ver cual de los 3 tipos de categorías tiene más impacto.
<b>Problema planteado:</b>	<p>Lo más rápido que se puede hacer para localizar las distintas aplicaciones que utilizan este término, sería irnos a las aplicaciones gratis, ya que por lo general podemos encontrar diversos títulos como “Free fire”, “PUBG”, “Candy Crush”, etc, que son gratis y dentro de ellas se nos dan opciones para comprar mejoras, avanzar niveles, personalizar el avatar todo esto significa una ganancia monetaria que debemos de hacer nosotros para sentirnos felices.</p> <p>Para las aplicaciones de pago, por lo general ya esta por defecto que vas a tener que comprar distintas mejoras tu experiencia en la aplicación.</p>
<b>Solución:</b>	<ul style="list-style-type: none"><li>• Con la ayuda de la “visualización de datos” podemos mediante distintas graficas poder ver el comportamiento que tienen las aplicaciones de: “pago”, “gratis”, “pay to win”.</li><li>• Si vemos desde el punto de vista estadístico podemos utilizar una “regresión” que en este caso sería una “regresión lineal múltiple” por tener tres betas (<math>B_0</math>, <math>B_1</math>, <math>B_2</math>), esto nos ayudara para predecir el valor de la variable dependiente o para evaluar la influencia que tienen los predictores en ella.</li><li>• En dado caso que queramos saber el comportamiento que tendrán estas 3 categorías podemos usar la “predicción”, ya que nos ayudarán a proyectar los tipos de datos que se verán en el futuro o predecir el resultado de un evento.</li></ul>




## Coronavirus

<b>Objetivo:</b>	Suponiendo que somos una compañía de seguros de vida a nivel mundial, y al tener esta pandemia nos gustaría saber qué países tienen mas casos
------------------	---

	de contagios de COVID-19, para nosotros ver por cuanto podemos asegurar a dichas personas por país.
<b>Problema planteado:</b>	<p>Tendremos que observar el número de contagios que existe por país, en este caso existen mas de 50 países con un alto nivel de contagio, lo que podríamos hacer sería lo siguiente:</p> <ul style="list-style-type: none"> <li>• Apoyarnos en la columna de “resumen” la cual nos va a indicar cuantas personas se han recuperado de covid y cuantas otras han sido contagiadas.</li> <li>• La localización nos servirá para ver las zonas más afectadas de cada país.</li> <li>• El país nos ayudara para ver quienes son más vulnerables que otros.</li> <li>• El genero masculino o femenino, nos ayudara para ver si el hombre es mucho más propenso a contagiarse o la mujer.</li> </ul> <p>Con la ayuda de estas columnas nos ayudara para poder dar un diagnostico de cuanto debemos nosotros cobrar al cliente y a cuanto podemos asegurarlo, para no llegar a perder dinero nosotros.</p>
<b>Solución:</b>	<ul style="list-style-type: none"> <li>• Con la ayuda gráfica, para ser más precisos con el tema “visualización de datos”, podemos realizar dos tipos de graficas. <ol style="list-style-type: none"> <li>1. Seria representar en el eje “y” los países contagiados por covid vs el eje “x” el número de personas contagiadas por día (hombre o mujer), haciéndolo acumulativo día con día.</li> <li>2. Representar en el eje “y” los países contagiados por covid vs el eje “x” el número de personas curadas por día (hombre o mujer), haciéndolo acumulativo día con día.</li> </ol> <p>Estas graficas nos ayudaran para saber que países son los más contagiados, además de indicarnos si es hombre o mujer, el cual nos ayudara con lo siguiente.</p> <p>Después de ver qué países son los más contagiados los catalogamos por las distintas regiones de cada país.</p> </li> <li>• Una vez tengamos lo anterior podemos realizar una “regresión” en este caso una regresión lineal ya que solo nos importa para el hombre(B0) y la mujer(B1). Este dato de predicción nos arrojará que generó es mas propenso a contraer Covid-19 y así nosotros dar por el momento como compañía de seguros de vida una predicción de cuanto podremos cubrir o de cuanto es la suma asegurada por persona.</li> </ul>



<b>Objetivo:</b>	La cantidad de puntos que WineEnthusiast califica a los distintos tipos de vinos en una escala del 1 al 100, se puede decir que entre más alta calificación tenga, mejor es el vino.
<b>Problema planteado:</b>	Visualizar con la ayuda de la columna, del país de donde es el vino, ya que nos ayudara a tener una idea de que país es el mas demandante en la fabricación de vinos, una vez tengamos el o los países donde se fabrica más vino, nos apoyamos con la calificación que de WineEnthusiast que les otorga por los distintos vinos que fabrica cada región y así poder ver en un top 10, cuáles son los mejores.
<b>Solución:</b>	<ul style="list-style-type: none"> <li>• Con la ayuda de la “visualización de datos”, podemos ver el comportamiento que toma cada país en la fabricación de vinos, también al saber que países son los que fabrican mas vino podemos ver por región.</li> <li>• Al tener los mejores países seleccionamos una gran cantidad de vinos que fabrique cada país para irlos catalogando como en “clustering”.</li> <li>• Una vez tengamos los vinos en distintos grupos, queda el tema de “clasificación”, he aquí donde entra WineEnthusiast el cual nos calificará los mejores vinos y nos dará un top 10 de mejores vinos.</li> </ul>

 <h3>Clasificación de plantas de las 3 especies de Iris</h3>	
<b>Objetivo:</b>	Obtener la mejor planta, de los tres tipos de especies de iris, dado la predicción de la longitud del pétalo y anchura del pétalo, al final solo sumariamos las predicciones para determinar la mejor especie de iris.
<b>Problema planteado:</b>	Identificar en la base de datos de iris, la longitud del pétalo y anchura (mediante una regresión lineal), una vez teniendo esto, sumamos cada una de ellas y la que resulte con mayor calificación esa sería la mejor especie de iris.
<b>Solución:</b>	<ul style="list-style-type: none"> <li>• Al tener todos nuestros datos sin un orden lo mejor es ordenarlos en las 3 distintas especies de iris, es por ello, que se utiliza el “clustering”.</li> <li>• Una vez agrupados en las distintas especies, se puede utilizar la “regresión” en este caso sería la regresión lineal (B0 para longitud del pétalo, B1 para la anchura del pétalo) ya que nos interesa saber cuánto es la predicción.</li> </ul>

	<ul style="list-style-type: none"> <li>Finalmente, con la “visualización de datos” podemos sumar el total de las predicciones de la regresión lineal para cada especie y a manera de 3 barras podemos observar cual es la mejor especie de iris.</li> </ul>
--	---

<b>NETFLIX</b>	<b>Series y películas de Netflix</b>
<b>Objetivo:</b>	Mejorar la búsqueda, relacionada con el director de la película, una vez que seleccionemos a nuestro director se nos mostraran todas las películas o series que realizo.
<b>Problema planteado:</b>	Mediante la base de datos de Netflix, podemos utilizar la columna de “Identificador: una película o un programa de televisión”, el cual nos ayudara mostrando todas las películas o series que existen en el catalogo de Netflix, una vez teniendo esto, las relacionamos con el director el cual se guardaran en una matriz, sin embargo para que sea mucho mas atractivo para el publico una vez teniendo las películas o series del director las clasificamos en los distintos géneros que existen y entonces ya terminaríamos
<b>Solución:</b>	<ul style="list-style-type: none"> <li>Con la “visualización de datos” mediante diversas graficas (para el eje y: actor, para el eje x: el nombre de la película o serie) podemos ir guardando las películas o series que realizaron diferentes directores.</li> <li>Al tener las películas y series de cada director, podemos realizar un “clustering” que nos ayudara a calificar por genero la película o serie que dirigió.</li> </ul>