

UNIVERSIDAD AUTÓNOMA DE
NUEVO LEÓN



FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

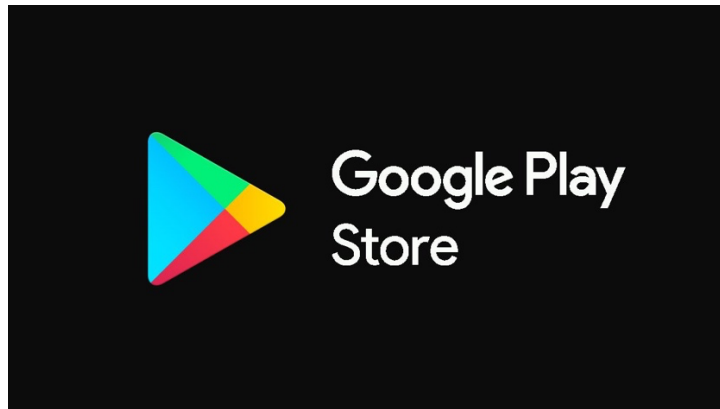
Minería de Datos

Análisis de Base de Datos

Laura Estefany Rodríguez de los Reyes
Matrícula: 1588292

Maestra: Mayra Cristina Berrones Reyes

Google Play Store



<https://www.kaggle.com/lava18/google-play-store-apps/tasks>

Objetivo

Determinar en que categorías se ubican las aplicaciones con mayor numero de descargas, esto es cual de las cateogorías es más popular entre los usuarios.

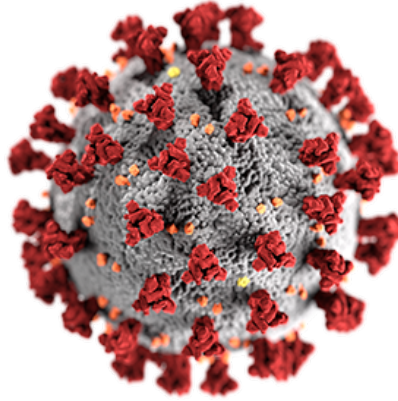
Problema Planteado

Google Play Store cuenta con una gran cantidad de aplicaciones, por lo que a los desarrolladores les gustaría saber cual es la categoría más popular en la que se descargan aplicaciones, esta información puede ayudar a las empresas para analizar su enfoque para ingresar al mercado, ya sea a uno con más, regular o menor competencia.

Solución

Crear un algoritmo que nos permita visualizar los datos para poner observar la distribución del mercado de las aplicaciones en Google Play Store, así como tambien tomar en cuenta la variable de rating, para checar si existe alguna relación entre ambas.

Coronavirus



<https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset>

Objetivo

Obtener el riesgo de contraer la enfermedad dependiendo del país donde te encuentres.

Planteamiento del problema

Actualmente estamos en una pandemia, y dado que los contagios van en aumento cada día pero hay casos en que las personas tienen que viajar por ciertas razones, deseamos obtener la tasa de riesgo de contraer coronavirus dependiendo del país donde se está o se quiere viajar. O simplemente se tiene que la economía no puede parar por lo que las zonas con menor tasa de contagio se pueden abrir.

Solución

Desarrollar un modelo de regresión para poder encontrar esta tasa de contagio, también la podemos ver como la probabilidad de contraerla estando en un lugar. Así como también crear por medio de la técnica de visualización un gráfico con el mapa del mundo donde podamos ver cuáles son los países que presentan mayor número de contagios y consecuentemente la tasa de contagio sería mayor.

Criticas de Vino (Wine Reviews)



<https://www.kaggle.com/zynicide/wine-reviews>

Objetivo

Encontrar un modelo que nos ayude a predecir de donde viene el vino o más bien donde fue hecho, basado en sus características.

Planteamiento del problema

La información que vamos a obtener puede ayudar a las empresas o personas entusiastas por los vinos a saber solo teniendo las características del vino de donde proviene este, además nos ayuda a determinar en que regiones se producen los mejores vinos.

Solución

En la solución de este problema podemos hacer uso de la técnica de agrupación para analizar cada región y por lo tanto las características que presentan los vinos hechos en cada región. Con esto podemos clasificar también a los vinos según sus puntuaciones, ayudándonos a ver en que regiones se producen.

Clasificación de las plantas (Iris Species)



<https://www.kaggle.com/uciml/iris>

Objetivo

Determinar a que tipo de especie de Iris pertenece una planta dadas sus características como longitud, tamaño del sepalo y petalos.

Planteamiento del problema

La información que podemos obtener según nuestro objetivo puede ayudar a las personas que se están adentrando o no tienen mucho conocimiento de como distinguir cada especie de iris a determinarlo con características que tienen a su alcance.

Solución

Para poder resolver este problema podemos desarrollar un modelo donde ingresando los valores de la longitus, tamaño de sepalo y petalos nos de como resultado el tipo de Iris que estamos viendo ya sea setosa, versicolor o virginica. Con esta herramienta podemos ayudar a que las personas novatas puedan posteriormente reconocerlas sin el uso de ayuda extra.

Netflix Movies and TV Shows



<https://www.kaggle.com/shivamb/netflix-shows>

Objetivo

Identificar las preferencias de películas y series en cada país para que así Netflix pueda saber que tipo de contenido habilitar o desarrollar según la zona donde este el usuario.

Planteamiento del problema

Netflix debe conocer las preferencias de los usuarios para así poder retenerlos, ya que si no agrega contenido en la categoría que a la mayoría de los usuarios ven con rapidez estos seguramente se aburrirán y dejarán la suscripción.

Solución

Para la solución del problema crearemos la técnica de visualización para poder ver que categoría es más popular en los diferentes países, para que con esta información Netflix produzca o incluya contenido más frecuente en estas categorías que son más populares. Adicionalmente se puede crear una secuencia personalizada para cada usuario, recomendándole contenido similar al que ha visto.