

13 DE OCTUBRE DE 2020



---

UANL®

# ÁNALIS BASES DE DATOS

JOSÉ IGNACIO CONTRERAS RUIZ

1887848

Facultad de Ciencias Físico Matemáticas

## Contenido

TV Shows and Movies listed on Netflix .....	2
Objetivo: .....	2
Problema: .....	2
Solución: .....	2
Google Play Store APPS.....	2
Objetivo: .....	2
Problema: .....	2
Solución: .....	2
Novel Corona Virus 2019 Dataset .....	3
Objetivo: .....	3
Problema: .....	3
Solución: .....	3
Wine Reviews .....	3
Objetivo: .....	3
Problema: .....	3
Solución: .....	3
Iris Species.....	4
Objetivo: .....	4
Problema: .....	4
Solución: .....	4

## TV Shows and Movies listed on Netflix

### Objetivo:

Encontrar como se encuentra distribuida la producción de los diferentes géneros de manera global en Netflix

### Problema:

La información se encuentra por ID del contenido en Netflix

### Solución:

Filtrar la información creando diferentes listados por países, una vez teniendo estos listados generar un porcentaje según el género al que pertenezca el contenido en Netflix, ya sea Show de Tv o película.

Con esta información podremos analizar y crear visualizaciones de la manera que crea cada país su contenido, o buscando ser más objetivo, cual es el porcentaje de su contenido más comercial respecto a género.

Shows de Netflix (<https://www.kaggle.com/shivamb/netflix-shows>)

## Google Play Store APPS

### Objetivo:

Determinar el tipo de contenido que más descargan los usuarios con dispositivos móviles con sistema operativo Android.

### Problema:

Poder analizar y limpiar la base de datos de acuerdo a lo que más nos interese saber para crear una correcta evaluación.

### Solución:

Analizando los ratings otorgados por el público comprobar si realmente las aplicaciones mejor rankeadas con las que mayor porcentaje de instalación tienen, una vez hecho eso lo que será necesario hacer es comprobar el género o género a los que pertenecen estas aplicaciones y determinar si hay alguna similitud entre ellas y poder crear agrupaciones.

Ya con las correctas agrupaciones únicamente será necesario contabilizar el número de observaciones en cada uno para evaluar qué tipo de aplicación o aplicaciones son las más descargadas por los usuarios.

Google Play Store (<https://www.kaggle.com/lava18/google-play-store-apps>)

## Novel Corona Virus 2019 Dataset

### Objetivo:

Conocer en qué momento la pandemia salió de control para en un futuro poner atención a ese punto

### Problema:

Encontrar repunte en la serie de tiempo y evaluarla de acuerdo a las procedencias de los pacientes.

### Solución:

Comparar las fechas de reporte junto con el análisis de si provenía de Wuhan, estaba en Wuhan o no tuvo contacto directo con la ciudad para conocer el plazo aproximado de colchón que se tiene para poder generar un bloqueo completo en las fronteras de ciudades infectadas para evitar la propagación global.

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

## Wine Reviews

### Objetivo:

Conocer si el precio de la botella tiene relación con el rating recibido.

### Problema:

Los catadores conocen el precio y no se tiene conocimiento del rating que hubieran otorgado de no conocerlo.

### Solución:

Conociendo la base de datos es posible crear un análisis que agrupe el precio con el rating recibido e identificar si están directamente relacionados con una regresión sencilla, para poder conocer aún más la opinión del catador o aquel que realizó el review se pueden llevar a cabo diferentes funciones de sentiment analysis para asociar las palabras con un valor número y crear un promedio de si verdaderamente el sentimiento es positivo o negativo según las palabras utilizadas.

Criticas de vinos (<https://www.kaggle.com/zynicide/wine-reviews>)

## Iris Species.

### Objetivo:

Catalogar las diferentes especies de Iris y crear un modelo predictivo.

### Problema:

Encontrar una relación entre las diferentes variables dentro de la base de datos que nos ayuden a predecir a que especie pertenece una Iris ingresando las diferentes medidas de las variables que conocemos.

### Solución:

Crear un árbol de clasificación CHAD con alguna paquetería que nos otorgue la probabilidad de que una planta pertenezca a alguna de las tres especies si conocemos el tamaño de su tallo, de su sépalo y sus medidas completas.

Clasificación de plantas (<https://www.kaggle.com/uciml/iris>)