

Facultad de Estudios Superiores: Acatlán.

Diplomado en Ciencia de Datos

Proyecto:

*La magia del cine (o de los datos)*

Valencia Martínez Jesús Adolfo

La magia del cine (o de los datos)

Resumen

En el mundo, la información crece de manera constantemente; información que poco a poco se está convirtiendo en activos para su explotación. En el mundo del arte, y más específicamente, en el llamado séptimo arte, este fenómeno no sería la excepción.

Por lo cual, el presente reporte expone la aplicación de diversas metodologías de Ciencia de Datos para obtener resultados de valor haciendo uso de los datos generados por la industria del cine.

La fuente

Los datos de insumo corresponden a una extracción dentro de un sitio web llamado Kaggle. Estos datos se actualizan diariamente y detallan la información de las películas que se encuentran en la página de reseñas de internet llamada TMDb. Esta página que alberga información no solamente de películas, también de series y capítulos de series y, si bien, aunque la misma página ofrece una API para realizar extracción de información, la base que se encuentra en Kaggle contiene información únicamente de películas junto con otra información relevante, por lo cual, se opta por elegir esta última para la extracción de información.

Los datos

Los datos albergan la información de un total de 750,318 películas y 20 características:

|  |  |
| --- | --- |
| Característica | Descripción |
| id | Identificador único en la página de TMDb |
| title | Nombre de la película |
| genres | Géneros relacionados |
| original\_language | Idioma original |
| overview | Resumen o sinapsis |
| popularity | Métrica de TMDb para medir la popularidad dentro del sitio web |
| production\_companies | Compañías productoras |
| release\_date | Fecha de estreno |
| budget | Presupuesto de producción |
| revenue | Ganancia reportada |
| runtime | Duración en minutos |
| status | Estado actual de la película (estrenada o no) |
| tagline | Breve frase alusiva a la trama de la película |
| vote\_average | Calificación promedio que los usuarios le dan a la película |
| vote\_count | Cantidad de calificaciones proporcionadas |
| credits | Nombre de los productores principales |
| keywords | Palabras clave sobre la película |
| poster\_path | URL para obtener el poster de la película |
| backdrop\_path | URL para obtener el backdrop de la película |
| recommendations | Recomendaciones a otras películas relacionadas |

De las características anteriores, hay algunas que, para propósitos de aplicación, no se consideraron para su uso, por lo que se descartaron. Las variables que se descartaron fueron:

|  |  |
| --- | --- |
| 1 | title |
| 2 | production\_companies |
| 3 | credits |
| 4 | recommendations |

Valores ausentes

La información anterior descrita cuenta con valores que no se encuentran registrados o bien, valores ausentes. La presente tabla detalla la cantidad de valores ausentes por característica: