**Guia de ejercicios: Clase 1**

1. Utilizar [esta notebook](https://github.com/elsonidoq/machine_learning_practico/blob/clase-1/notebooks/clase-1/01_get_the_data.ipynb) para llevar los datos a tu google drive
   1. Entrar a [este link](https://colab.research.google.com/github/elsonidoq/machine_learning_practico/blob/clase-1/notebooks/clase-1/01_get_the_data.ipynb) para abrirlo directamente en colab
2. Explorar los archivos
   1. Leer la documentación en [este link](https://www.imdb.com/interfaces/)
   2. Qué columnas contienen?
   3. Usando la jerga de imdb: ¿Qué *titles* vamos a utilizar para trabajar? Considerar que hay series, cortos, etc. ¿Tienen sentido las duraciones de las preguntas?
   4. Explorar el archivo movie\_gross.csv descargado en el ejercicio 1. ¿Puede ser útil para [nuestro caso de negocio](https://docs.google.com/presentation/d/10G-qAtPrtmdLTrY6qIMdl5R2yPcB3U51wGPA6jotbJM/edit#slide=id.g9c1c2528d1_0_232)?
   5. Explorar cómo combinar los archivos, ¿Todas las películas ocurren en todos los archivos?
   6. Todas las columnas tienen valores útiles?
3. Explorar alternativas de modelado para abordar el caso de negocio
   1. ¿Qué variable podemos predecir?
   2. ¿De qué forma podemos modelarla?
   3. ¿Cómo evaluarías el modelo? Pensar en métricas y sanity checks
   4. ¿Cómo se vería la solución final que utiliza el modelo? ¿Quién la usaría?
   5. Identificar riesgos. ¿Están mitigados con la estrategia de modelado, métricas y sanity checks?
   6. Entrenar un modelo lineal que resuelva el problema delimitado por las respuestas a las preguntas anteriores