

Primera entrega proyecto INF280

1 Introduction

El objetivo de esta entrega es definir las variables relevantes de estudio y la pregunta a resolver. Esto mediante una exploración del dataset utilizando gráficos e índices estadísticos.

2 Requerimientos

Se deberá cumplir como mínimo con lo siguiente:

1. Introducir el estudio del dataset. Esto consiste en dar un preámbulo de por qué es relevante su investigación y en qué está enfocada (*se recomienda realizarla luego de completar las otras secciones*). Es importante mencionar que debe tener un **”valor agregado”** añadido por el grupo y **focalizar** tanto la contextualización como la relevancia según la pregunta de estudio definida. Además deberá buscarse una imagen que tenga relación con el tema e insertarla al inicio del google colab/ jupyter notebook.

Para un mayor entendimiento consultar la referencia 1.

2. Definir una hipótesis o una pregunta en la cual se basarán los análisis del proyecto.
3. Definir 2 variables principales de estudio y 2 variables secundarias. Estas variables deben tener una ”potencial relación” entre ellas que nos ayuden en el estudio. Deben ser **2 variables principales** pero en cuanto a las secundarias es a libertad de cada grupo cuántas incorporar. Para seleccionar las variables de estudio, deben generar gráficos y conclusiones de los mismos que **apoyen las decisiones tomadas**.

Hint: Si tienes muchas variables numéricas quizás sea bueno trabajar con mapas de calor para tener una idea preliminar

Hint: Si tienes variables categóricas, lo mejor será graficar el comportamiento de una variable en función de la separación de las categorías y analizar

4. Hacer ”limpieza del dataset” esto es eliminar valores atípicos (outliers) y valores ”NaN”. Además, de reducir las columnas presentes y dejar a las variables de interés.

5. Presentar la información del dataset para tener una visión general de como se distribuyen los datos con gráficos.

3 Modalidad

Cada tutor asignado debe fijar una reunión preliminar. En esta se darán a conocer un par de preguntas de ejemplo para que cada grupo comience a familiarizarse con la generación de gráficos y tengan una referencia del tipo de preguntas que se pueden hacer para argumentar la elección de variables (punto 3) para cada tema.

Recordamos que se evalúa la dimensión de **profundidad** por lo que, si se utilizan más tipos de gráficos (histogramas por categoría, heatmap, boxplot, etc..) con un correcto análisis, este item tendrá el puntaje máximo. No es necesario excederse, deben evaluar qué tipos de gráficos pueden aportarles información valiosa y utilizarlos en base a su criterio.

4 Entrega

El archivo a entregar debe ser un google colab o jupyter notebook con los requerimientos mínimos desarrollados y presentes en el documento, es decir, Introducción, gráficos con sus respectivas conclusiones y mencionar a modo de conclusión las variables que van a estudiar.

Para entregar el archivo se abrirá un recurso en aula que estará disponible hasta el **Domingo 5 de Mayo** hasta las 23:59hrs, posteriormente dentro de la semana del 6 de Mayo se darán comienzo a las presentaciones en donde deberán pasar el archivo ".ipynb" a un formato de exposición (ppt, prezi, etc) para luego presentar la información del mismo. Pronto se darán más detalles de la rúbrica de la presentación y los aspectos más formales de la misma.

5 Tutores Asignados

Tema	Tutor
1	Nicolás Rodríguez
2	Nicolás Rodríguez
3	Eduardo Flores
4	Lilly Guerra
5	Eduardo Flores
6	Esteban Gárate
7	Nicolás Horta
8	Esteban Gárate
9	Carlos Kuhn
10	Carlos Kuhn
11	Carlos Kuhn
12	Lilly Guevara
13	Nicolás Rodríguez
14	Nicolás Horta
15	Nicolás Horta
16	Lilly Guevara
17	Eduardo Flores



Contexto

MyAnimeList, también conocido como MAL, es una aplicación web que se dedica a la información sobre anime, manga y novelas ligeras. Esta aplicación es muy popular entre los fans de anime y manga, ya que les permite buscar, leer reseñas y calificar las series y películas de anime, así como crear listas de seguimiento de sus títulos favoritos.

Debido a la gran cantidad de datos que MyAnimeList posee sobre sus usuarios y animes, se le ha solicitado a su grupo que realice una serie de tareas para ayudar a la plataforma a analizar estos datos. Uno de los objetivos principales es identificar el público que más visualiza animes en la plataforma, enfocándonos en su género, edad y cantidad de horas viendo anime.

Además, también se les ha pedido realizar un análisis de datos de animes. Esta tarea puede ayudar a identificar los animes más vistos, mejor calificados, más nuevos, entre otros aspectos. Estos datos pueden ser utilizados para crear secciones en la página web de MyAnimeList que muestren esta información, permitiendo a los usuarios conocer los animes más populares y recientes en el sitio.

Figure 1: Referencia de Introducción Tarea 1 INF280 s2023-01