

Tarea 1: Utilizando una de las bases de datos que vimos en clase:

Proyecto Gutenberg: <https://www.gutenberg.org/>

Wikipedia

Kaggle: <https://www.kaggle.com/datasets>

Utilicen la extracción de palabras clave, stopwords, lematización, y stemming para hacer n preprocesamiento de la información. Utilizar gráficos adecuados para representar la información, y describir de manera general cual puede ser el propósito de la base de datos. (Conclusiones de lo que se puede hacer con los datos recopilados).

Ejemplo en mi GitHub:

https://github.com/mayraberrones94/FCFM/blob/master/Semana_1_Pre_procesamiento_de_datos.ipynb

El trabajo será entregado de la siguiente manera:

- Archivo en formato PDF en la plataforma de Teams.
- Archivo PDF, documento de código e imágenes adjuntas, en su repositorio personal de GitHub.

Recuerden que en el archivo PDF deben de agregar en las referencias el URL de su GitHub personal y el URL de la página de donde sacaron su base de datos.

Se evaluará:

- Que se tengan los puntos acordados en clase.
- Utilizar correctamente los ejemplos que se dieron.
- Que todas las modificaciones que se vayan haciendo a su base de datos inicial estén explicadas correctamente en el documento.
- Pueden agregar código a su trabajo, pero es importante que solo pongan las líneas necesarias y que estén explicadas correctamente.
- Todas las graficas que se agreguen deben estar etiquetadas correctamente y referenciadas en el texto (asegúrense de que no solo muestran los resultados, sino también los explican).
- Es importante agregar la discusión de resultados y referencias.
- Recuerden que pueden usar los elementos de los ejemplos de manera personalizada. Si encuentran resultados en los que es necesario que ustedes tengan que modificar manualmente las librerías, discutan que tuvieron que hacer y cual fue el resultado.

Puntuación:

Las tareas individuales pueden llegar a tener hasta 6 puntos.

6 puntos - Excelente: Sobrepasa la expectativa del trabajo. Se agregaron elementos de interés a la tarea, como orientar el trabajo con elementos que sirvan para su proyecto de tesina.

5 – 4 puntos – Bueno: Cumple con lo que se espera de la tarea. Baja un poco si se tiene faltas de ortografía o inconsistencias en la discusión de resultados.

3 – 2 puntos – Regular: Faltan puntos que se comentaron para la tarea. Tiene algo que puede ser rescatable, pero falta mucho contenido.

1 punto – Malo: Se intento hacer algo, pero al final no tiene mucho sentido.

0 puntos: No entrego la tarea.

NOTA: Recuerden que la puntuación se evalúa a la mitad de los puntos que pudieron haber obtenido si lo entregan a destiempo por una semana. Después de la primera semana todos los trabajos se evalúan a 0.