Bienvenidos!
Semana 3

Datos No Estructurados y Semiestructurados

Especialización en Economía, opción Ciencia de Datos

FCS - Udelar

Guillermo Lezama

La clase de hoy

- HTML
- Scraping and APIs
- Una intro a texto
- Actividad grupal y Proyecto Final

- ¿Dudas del cuestionario o alguna de las tareas?

- Comentarios sobre Ejemplos

¿Qué son los datos de texto?

- Datos no estructurados compuestos por caracteres, palabras y oraciones.
- Ejemplos: noticias, tuits, libros, respuestas de encuestas.

Texto: Flujo de trabajo básico

Preprocesamiento

- Normalizar texto (convertir a minúsculas, ej.: "Hola" → "hola").
- Remover palabras comunes (stop words: "de", "y", "el", etc.).
- Limpiar puntuación, símbolos y números.

Tokenización

- Dividir texto en unidades mínimas significativas (palabras o términos).

Transformación y extracción de información

- Contar frecuencia de palabras clave.
- Identificar frases relevantes o expresiones comunes.
- Clasificar textos por sentimiento o temática.
- Crear visualizaciones gráficas (ej.: nube de palabras, gráficos de barras).
- Interpretar resultados y extraer conclusiones.

El lunes

- 1 o 2 slides
- Pregunta que quisieran responder.
- ¿Qué tipo de datos estructurados precisarían para responder eso?
- ¿Qué fuente de datos no estructurados van a usar
- Es un cuarto de la nota del trabajo final.
- ¿Cómo evalúo el trabajo final? Complejidad de la tarea, entender por qué importa lo que hacen…

Proyecto Final

- Similar a actividad grupal.
- La diferencia: si van a trabajar con los datos.
- Objetivo: Transformar un conjunto de datos desestructurados en algo estructurado + contar algo de esos datos + articularlo con una historia a contar.
- 9 de Junio: Contar qué piensan hacer
- Entrega: 31 de Julio
- 3 opciones: Texto, imagen, Sonido.
- Grupos de a 4.
- Fin de semana **(ANTES DEL LUNES DE NOCHE)**: formulario sobre qué formato quieren trabajar
- Agregar si hay algún tipo de texto, imágenes o sonidos que quieran trabajar

Para próxima clase

- Cuestionario
- Llenar formulario
- Tareas del notebook
- Tarea adicional en mi notebook Guillermo.ipynb