**Reporte**

**Introducción**

El propósito de este documento es detallar el proceso de revisión y validación de archivos JSON del documento compartido, para garantizar que los datos cumplan con los requisitos y el formato esperado en el proyecto. El archivo JSON contiene información sobre tweets, como texto, usuario, hashtag, fecha y métricas de participación (retweets y favoritos). El objetivo es verificar la integridad de estos datos examinando su tipo y estructura, identificando cualquier inconsistencia que pueda afectar la confiabilidad de análisis posteriores.

**Explicación de la verificación del tipo de datos**

Para garantizar que los datos correspondan al tipo esperado, se implementa un esquema JSON que especifica el tipo de cada campo. Este patrón nos permite comprobar que la estructura de datos de cada tweet es correcta y que cada tipo de datos es el adecuado. Los principales tipos de datos probados son:

**id:** Este campo identifica de forma única cada tweet. (Tipo de dato: string)

**Texto:** Contiene el contenido del texto del tweet. (Tipo de dato: string)

**usuario:** Identifica al usuario que publicó el tweet. (Tipo de dato: string)

**hashtags:** Contiene una lista de #hashtags utilizados en los tweets. (Tipo de dato: array)

**fecha:** Indica la fecha y hora en que se publicó el tweet. (Tipo de dato: string)

**Reenviar:** Indica cuántas veces se ha retuiteado el tweet. (Tipo de dato: number)

**Favoritos:** Muestra el número de tweets que más me gustan. (Tipo de dato: number)

**Instrucciones para verificar la estructura de datos**

Además de verificar el tipo de datos, se verifica la estructura del archivo JSON para garantizar que cumpla con los requisitos esperados. La estructura de datos esperada es una matriz de objetos, donde cada objeto representa un tweet y contiene las claves anteriores.

Para validar la estructura se implementa un proceso de validación, el cual consiste en:

1. Verificar si los campos requeridos están presentes: Cada objeto debe contener todos los campos especificados en el esquema para garantizar la coherencia estructural.
2. Identificar campos vacíos: comprueba si hay campos vacíos en cada objeto (tweet) usando una función que busca valores nulos (Ninguno, '' o []). Esta característica nos permite determinar si un campo existe, pero no tiene datos, lo que puede afectar el análisis.
3. Formato de fecha: proporciona validación para garantizar que las fechas sigan el formato de fecha y hora estándar, lo cual es útil para temporalidades basadas en tweets.

Estos métodos de validación de estructura garantizan que todos los tweets en el archivo JSON mantengan una estructura consistente, lo que hace que los datos sean más fáciles de manipular y analizar de manera confiable.

Conclusión

El proceso de validación de los datos en los tweets y la validación del archivo JSON concluyó que los datos tienen un tipo consistente y una estructura definida en el esquema.La implementación del esquema JSON y la validación de campos vacíos nos permiten identificar y asegurar la calidad y consistencia de los datos antes de utilizarlos en análisis posteriores.

1. Cumplimiento del esquema: El archivo JSON validado cumple con el esquema especificado en cuanto a tipos de datos y estructura.
2. Identificación de campos vacíos: el análisis muestra que algunos campos en algunos tweets pueden estar vacíos, lo que indica la necesidad de preprocesar la interpolación de estos datos o sus datos.
3. Facilidad de integración: la coherencia de la estructura de datos hace que los archivos JSON sean adecuados para la integración en procesos de análisis, visualización o almacenamiento de bases de datos, lo que los hace más fáciles de usar en futuras fases del proyecto.