

Exploración de la Etapa: Proceso

Preguntas Orientadoras:

1. **¿Qué herramientas eliges y por qué? (estamos trabajando en un cuaderno de Jupyter en Visual Studio Code)**
 - La elección de Jupyter en Visual Studio Code es adecuada ya que ofrece un entorno interactivo y flexible para la limpieza y análisis de datos. La combinación de Jupyter y VS Code proporciona un flujo de trabajo eficiente y herramientas poderosas.
2. **¿Ha garantizado la integridad de sus datos?**
 - Se han realizado verificaciones iniciales para identificar valores nulos, duplicados y posibles anomalías. Durante la fase de procesamiento, se implementarán medidas adicionales para garantizar la integridad de los datos.
3. **¿Qué medidas ha tomado para garantizar que sus datos estén limpios?**
 - Se han identificado valores nulos y duplicados durante la carga del DataFrame. Durante el proceso de limpieza, se abordarán estos problemas y se realizarán transformaciones adicionales según sea necesario.
4. **¿Cómo puede verificar que sus datos estén limpios y listos para analizar?**
 - La verificación se realizará mediante la revisión de estadísticas descriptivas, la identificación y manejo de valores atípicos, y la validación de la coherencia de los datos con respecto al contexto del estudio de caso.
5. **¿Ha documentado su proceso de limpieza para poder revisar y compartir esos resultados?**
 - Sí, se documentará cada paso del proceso de limpieza en el cuaderno de Jupyter, asegurando que el proceso sea transparente y comprensible. Esto facilitará la revisión y compartirá los resultados con otros interesados.

Tareas Clave:

- **Verifique los datos en busca de errores:**
 - Se revisarán los datos para identificar y abordar errores como valores nulos, duplicados y posibles anomalías.
- **Elige tus herramientas:**
 - Jupyter en Visual Studio Code será la principal herramienta para el procesamiento y limpieza de datos.
- **Transforme los datos para que pueda trabajar con ellos de manera efectiva:**
 - Se implementarán transformaciones para abordar valores nulos, duplicados y cualquier otro problema identificado durante la verificación de datos.

- **Documente el proceso de limpieza:**
 - Cada paso del proceso de limpieza se documentará en el cuaderno de Jupyter, proporcionando una referencia clara y comprensible para revisar y compartir resultados.

Entregable:

Un cuaderno de Jupyter en Visual Studio Code documentando el proceso de limpieza de datos, con transformaciones y explicaciones detalladas de las decisiones tomadas para garantizar la calidad de los datos.