

Análisis de Caso

DATA WRANGLING 

Análisis de Caso

Nombre del tema: Data Wrangling con Pandas

Situación inicial

Una empresa de tecnología financiera recibe datos de múltiples fuentes, pero estos datos suelen estar desorganizados, contener valores nulos y registros duplicados, lo que afecta la precisión de los informes financieros. Actualmente, los analistas de datos enfrentan dificultades para consolidar la información y preparar reportes confiables.

El equipo ha decidido implementar Data Wrangling con Pandas para limpiar, transformar y optimizar los datos antes de analizarlos. Sin embargo, necesitan definir un proceso eficiente para la organización de los datos, la eliminación de inconsistencias y la integración de múltiples fuentes en un único DataFrame. Tu misión será diseñar una solución utilizando Pandas para asegurar la calidad y estructura de los datos.

Descripción del Caso

En este caso, asumirás el rol de un **científico de datos** encargado de optimizar la manipulación de datos en la empresa. Para ello, deberás aplicar técnicas avanzadas de **Data Wrangling** para mejorar la calidad y eficiencia del análisis.

Las tareas incluirán:

- **Detección y eliminación de valores nulos y duplicados.**
- **Transformación de datos categóricos y numéricos** para su correcta interpretación.
- **Uso de funciones de agregación y mapeo** para estructurar los datos.
- **Reorganización y optimización de DataFrames** mediante ordenamiento y filtrado.

Deberás utilizar **Pandas** para preparar un conjunto de datos que pueda ser utilizado en modelos analíticos y reportes estratégicos de la empresa.

Instrucciones

Para completar este caso, realiza las siguientes tareas:


1. **Carga y exploración de datos:**

- Importa un dataset en formato **CSV** en un DataFrame.
 - Inspecciona los datos con `.head()`, `.info()` y `.describe()`.
 - Identifica valores nulos y duplicados.
2. **Limpieza y transformación de datos:**
- Imputa valores nulos utilizando estrategias adecuadas (media, mediana o moda).
 - Elimina registros duplicados.
 - Convierte columnas categóricas en variables numéricas si es necesario.
3. **Optimización y estructuración de datos:**
- Aplica funciones de **groupby** y agregación.
 - Filtra los datos para obtener subconjuntos de interés.
 - Renombra y reorganiza columnas para mejorar la interpretación.
4. **Exportación de datos:**
- Guarda el DataFrame procesado en un **archivo CSV** sin incluir el índice.
 - Exporta los datos limpios a **Excel** para su visualización y reporte.

Entregables

Los participantes deberán entregar un informe con:

1. **Código fuente en Python** con la implementación de las tareas.
2. **Explicación detallada** de cada paso realizado y su justificación técnica.
3. **Ejemplo de los datos antes y después de la transformación.**
4. **Conclusiones** sobre la importancia de **Data Wrangling** en la calidad de los datos.

Este análisis de caso te permitirá aplicar **Pandas** para estructurar y optimizar datos en entornos de negocio y análisis predictivo. 

¡Muchas gracias!

Nos vemos en la próxima lección

