



Desafío PyWeekend: *Reconstruyendo el pasado con Python y AI*

Introducción



Un equipo de investigadores de la ESPOL ha compartido una base de datos del sitio arqueológico **Llano Chico**, que es una parroquia de un barrio histórico y rural del norte de Quito.

La matriz de datos contiene la información morfológica de restos cerámicos. El archivo está en formato Excel, lleno de notas internas, valores nulos, y estructuras por definir. Tu

misión es convertir estos datos en información útil para entender lo que significa cada grupo de restos de cerámica.

Etapas 1: Estructura y visualiza los datos

Tienes en tus manos un archivo real entregado de restos arqueológicos de Llano Chico, donde se han encontrado tumbas y artefactos.

El archivo contiene datos de fragmentos cerámicos recolectados durante excavaciones. Existe la necesidad del equipo de arqueólogos de poder facilitar el pre-procesamiento de estos datos.

¿Qué representa este archivo?



- **Cada fila representa un paquete:** este almacena los restos cerámicos encontrados en un segmento específico de terreno durante la excavación.

Ilustración 1 Identificación del material en el laboratorio de Rumipamba.



- Cada columna representa una característica o parte de los objetos hallados (como tipo de borde, forma, decoración, técnica, color, etc.).
- Los valores numéricos indican cuántos fragmentos con esa característica fueron encontrados en ese paquete.
- Algunas celdas cuentan con **comentarios** que indican características como "CONCAVA", "ENGROSADO", etc., los cuales también aportan información sobre las piezas.

✦ **Nota:** la combinación de los campos "**PROCEDENCIA**" y "**Nro. Paquete**" identifica de forma única cada registro o unidad arqueológica.

Este es un ejemplo de cómo lucen las columnas de los datos:

| Análisis morfológico | | | | | | | | | | | | | | |
|----------------------|------|-------|------|------|------|------|------|------|-------|------|------|------|------|------|
| | | Borde | | | | | | | | | | | | |
| Form | Form | Dire | Dire | Dire | Forn | Forn | Forn | Forn | Forma | Forn | Forn | Forn | Forn | Forn |
| 9 | 10 | 1 | 2 | 3 | 1 | 2 | 3 | 4 | 5 | | | | | |
| | | 3 | | | | | 2 | | 1 | | | | | |

Ilustración 2 Ejemplo del contenido del archivo excel

La imagen anterior (Ilustración 2) muestra una sección del archivo bajo el bloque "**Análisis morfológico**", específicamente enfocado en la categoría "**Borde**" de los fragmentos cerámicos.

Elementos clave:

- Las columnas están organizadas jerárquicamente por tipo de análisis. En este caso, se observa una subcategoría llamada "**Forma**" dentro del análisis del borde.
- Cada columna numerada (como "1", "2", "3", etc.) representa un **tipo específico de forma** de borde. Sin embargo, la descripción de esa forma no está escrita directamente, sino que aparece como un **comentario de la celda**, como el que se ve en (Ilustración 2) con el texto: **HP: CÓNCAVA**. Esto significa que, por ejemplo, la forma número 5 corresponde a un borde de tipo "cóncavo".



- Las celdas en la tabla contienen **valores numéricos** que indican **cuántos fragmentos** con esa forma específica de borde fueron encontrados en un determinado lugar (fila de la matriz).

Tu objetivo

- Leer, interpretar y estructurar esta información de forma automática, creando una estructura organizada (preferentemente en diccionarios anidados) que capture las características clave de las unidades arqueológicas. Por ejemplo:

```
"caracteristicas": {  
    "borde": {  
        "formas": {  
            "cóncava": 1,  
            "recta": 2  
        },  
        "decoracion": {  
            "impresa": 1  
        }  
    }  
}
```

- Por cada fila de la matriz, es importante descartar aquellas columnas que no tengan datos relevantes o estén vacías.
- **Debe escribir el diccionario en un archivo de texto.**
- Analizar las estructuras obtenidas y presentar al menos **3 visualizaciones estadísticas** que considere relevante para descubrir patrones significativos en los datos. Por ejemplo: Frecuencia de aparición de determinadas técnicas o materiales.

⚠ Restricción importante:

No está permitido modificar el archivo Excel original. No se deben agregar columnas, cambiar encabezados ni editar los datos en la fuente.

Toda la limpieza, estructuración y procesamiento debe hacerse desde el código.

✂ Librerías sugeridas

- Lectura y limpieza:
pandas, **openpyxl**
- Estructuración:
json, collections
- Visualización:
matplotlib, seaborn, plotly, etc.