

Desafío. Exploración NDVI

Con el fin de analizar el comportamiento del Índice de Vegetación en el área de estudio: “Agrícola Chapi Don Ernesto”, se utilizó “*Jupyter Notebook*” para ejecutar los scripts de Python. En el archivo “Análisis_Edinson_Medina.ipynb”, se puede encontrar la metodología, la cual fue dividida en 6 partes:

1. Importar paquetes.
2. Importar datos.
3. Obtener imágenes satelitales.
4. Extraer los datos estadísticos básicos.
5. Extraer las imágenes NDVI.
6. Análisis exploratorio de los datos.

A continuación, se presenta una breve descripción de cada una:

1. Importar paquetes.

Se importan, en el área de trabajo, los paquetes de Python utilizados:

- *Numpy – Pandas – Geopandas.*
- *Shapely – Descartes.*
- *Matplotlib – Seaborn.*
- *Sentinelhub.*

2. Importar datos.

Se importan los datos suministrados para la resolución del problema en *DataFrames*, para mayor facilidad en el manejo de los datos (es necesario que ambos archivos estén en el mismo directorio donde se encuentre el archivo “.ipynb”):

- “SentinelHubImage-available_dates.xlsx”: Contiene una lista de 89 fechas diferentes para extraer los datos y realizar los cálculos, las cuales oscilan entre el 20-04-2020 y 15-07-2016.
- “farm_map.json”: Contiene información referente a cada uno de los lotes, como por ejemplo coordenadas geográficas de los polígonos, área y rendimiento del cultivo.

Asimismo, se graficaron en conjunto todos los lotes, con el fin de visualizar su distribución espacial (figura 1). En recuadros de color azul se resaltan los polígonos, mientras que, en morado, se observa el recuadro delimitador (*Bounding Box*), que se utilizará para extraer las imágenes satelitales.

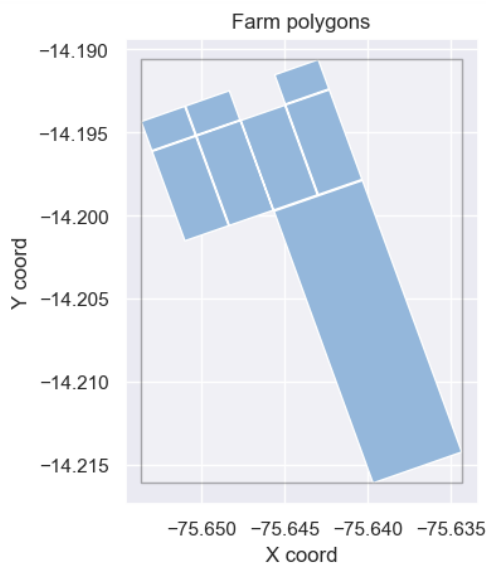


Figura 1. Polígonos de cada lote en el área de estudio.

3. Obtener imágenes satelitales.

Para ello, fue necesario crearse una cuenta gratuita de prueba en *Sentinel Hub*, obteniendo la siguiente credencial:

`INSTANCE_ID = "e7d7de5c-b222-421a-b4fe-6e2bdf2d76a5"`

Para realizar las solicitudes a *Sentinel Hub*, se utilizó la clase "*WcsRequest*", con una resolución para cada imagen de 4m x 4m. Una vez ejecutado el bloque de código, se creará en el directorio la carpeta "*1_Satellite_images*", que tendrá las imágenes satelitales extraídas, clasificadas por fechas.

Para control calidad, en la figura 2, se muestra las imágenes en las fechas 20-04-2020, 15-04-2020 y 11-03-2020.

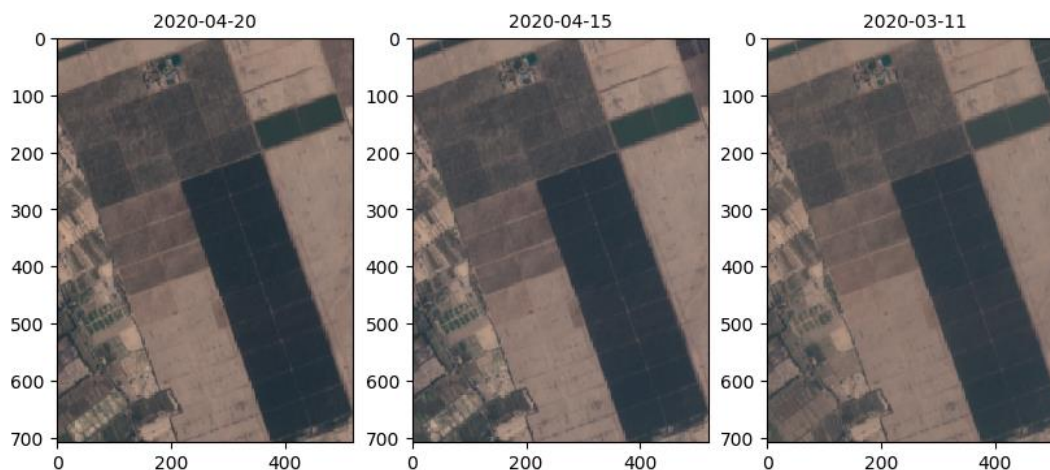


Figura 2. Polígonos de cada lote en el área de estudio.

4. Extraer los datos estadísticos básicos.

Se realizaron nuevas solicitudes a Sentinel Hub, en esta ocasión con la clase “*FisRequest*”, para extraer, para cada lote y para cada fecha, los valores mínimo, máximo, promedio y desviación estándar del Índice de Vegetación (NDVI). Una vez ejecutado el bloque de código, se creará en el directorio la carpeta “2_NDVI_statistics”, donde, adicionalmente, se crearán subcarpetas para cada fecha de interés, con el fin de guardar la información de los lotes.

Con ésta nueva información, se creó un DataFrame, incluyendo las áreas y el rendimiento de cada uno de los cultivos.

5. Extraer las imágenes NDVI.

Para ello, luego de ejecutar los códigos correspondientes, se creará la carpeta “3_NDVI_images”, al igual que los casos anteriores guardará las imágenes de NDVI para cada fecha, en el área de interés.

6. Análisis exploratorio de los datos.

En la figura 3, se muestra el histograma de frecuencia de los valores promedio de NDVI en el área de estudio. Claramente, se observa que tiene un comportamiento bimodal, es decir, existen dos intervalos diferenciados del índice de vegetación que son los más frecuentes.

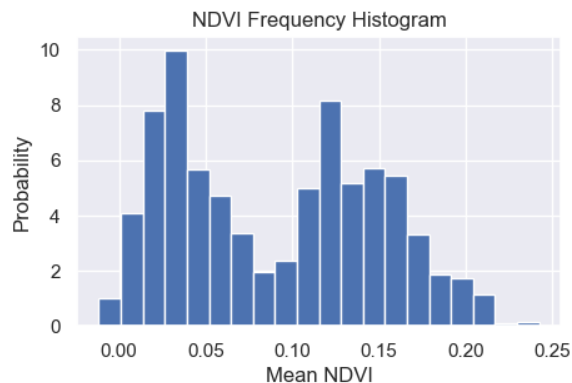


Figura 3. Histograma de los valores promedio de NDVI.

Asimismo, en la figura 4, se muestra la evolución de los valores promedio de NDVI a lo largo del tiempo, donde cada línea representa un lote diferente. Se puede observar con mayor detalle la diferenciación entre dos zonas dentro del área de estudio. Un primer grupo, con cultivos más sanos, con valores de NDVI que oscilan entre 0.10 y 0.25, en contraste con un segundo grupo de lotes, con una evolución más errática en el tiempo (denominados Grupo A y Grupo B respectivamente).

Llama la atención el comportamiento cíclico del Grupo A, donde el índice de vegetación suele mejorar en los últimos meses de cada año. Esto se puede observar desde el segundo semestre del 2016 hasta la presente fecha.

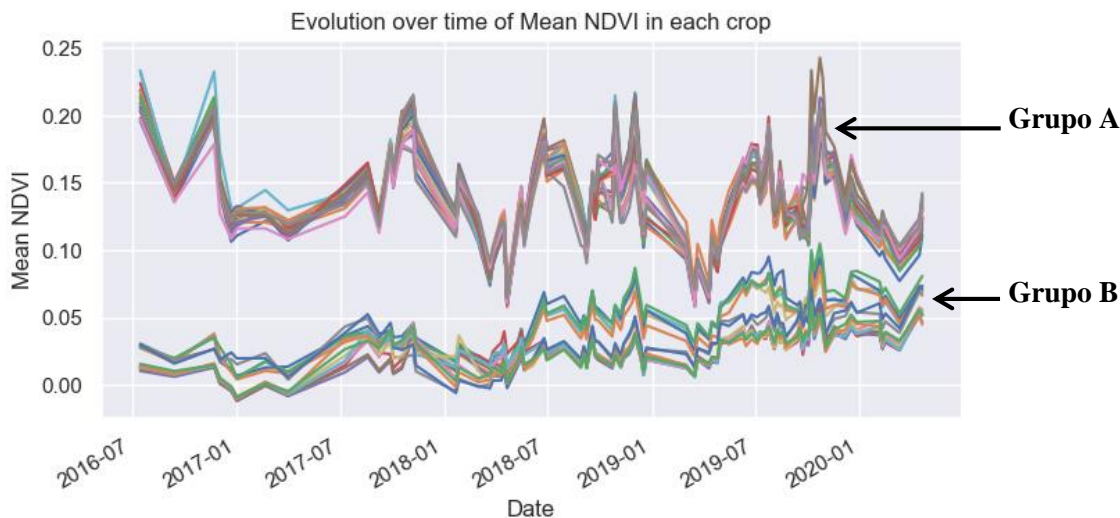


Figura 4. Evolución en el tiempo de los valores promedio de NDVI de cada cultivo.

Adicionalmente, con el fin de analizar la densidad de probabilidad de los datos de NDVI promedio se realizó un “*Violin plot*”. Este tipo de gráficos es especialmente útil, para los casos que, como éste, presentan una distribución multimodal. La figura 5, permite establecer que los lotes enumerados entre 0 - 17 y 18 - 32, se pueden clasificar en dos grupos diferentes.

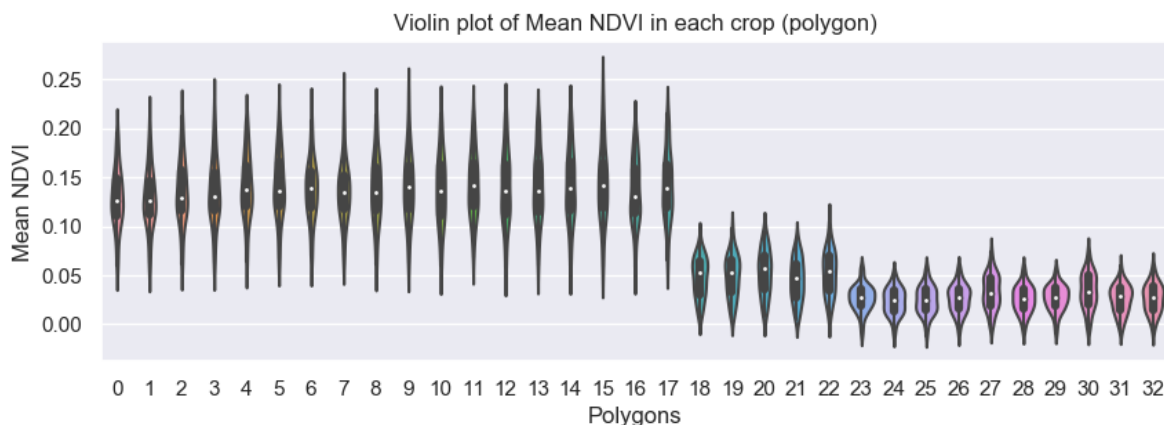


Figura 5. *Violin plot* de los valores promedio de NDVI en cada lote del área de estudio.

Seguidamente, la figura 6 permite conocer la ubicación espacial del Grupo A, en color azul, el cual agrupa a los lotes numerados desde 0 hasta 17, y el Grupo B, en color gris, que concentra el resto de los lotes en el área de estudio. En las imágenes satelitales que se muestran en la figura 2, se puede observar igualmente la diferencia entre cada uno.

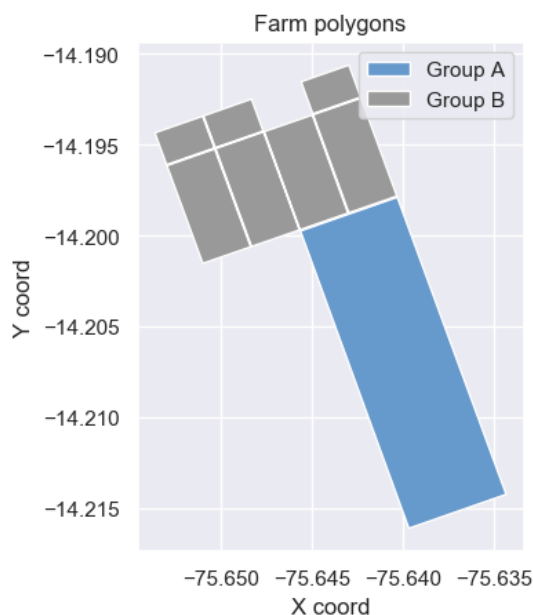


Figura 6. Ubicación espacial de los grupos A y B.

Por otra parte, con el fin de validar la hipótesis de la posible relación entre el NDVI promedio con el rendimiento (Kg/Ha) del cultivo, se realizaron tres gráficos cruzados que se encuentran en las figuras 7, 8 y 9.

Con éstos, se puede establecer una relación entre ambas variables. Es decir, los lotes que tienen menor rendimiento son los que igualmente tienen un índice de vegetación menor (Grupo B en óvalo gris). En contraste con el Grupo A (óvalo color azul), que tienen mejor rendimiento y con NDVI en promedio mayor.

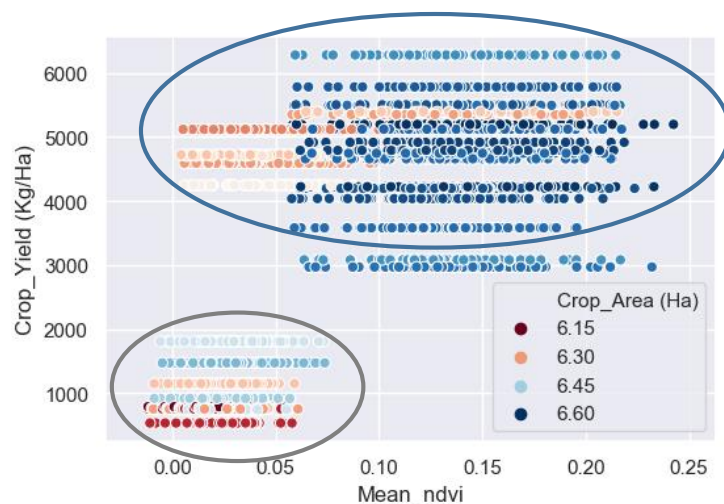


Figura 7. Gráfico cruzado: Rendimiento vs. NDVI promedio. Legenda: Área de los lotes.

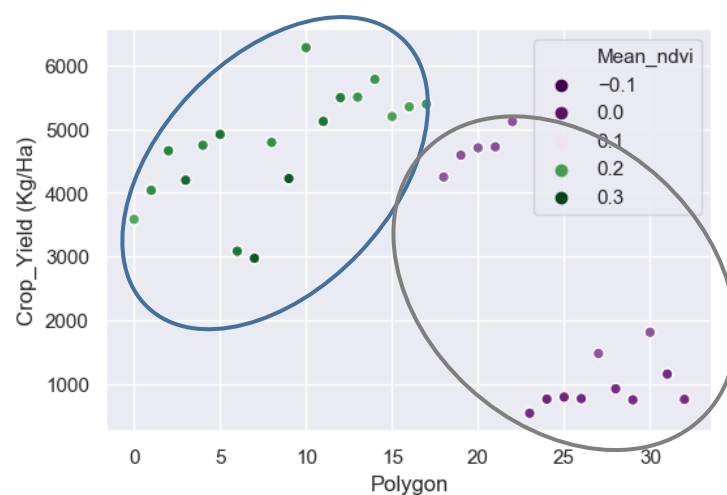


Figura 8. Gráfico cruzado: Rendimiento vs. Lotes. Legenda: NDVI promedio.

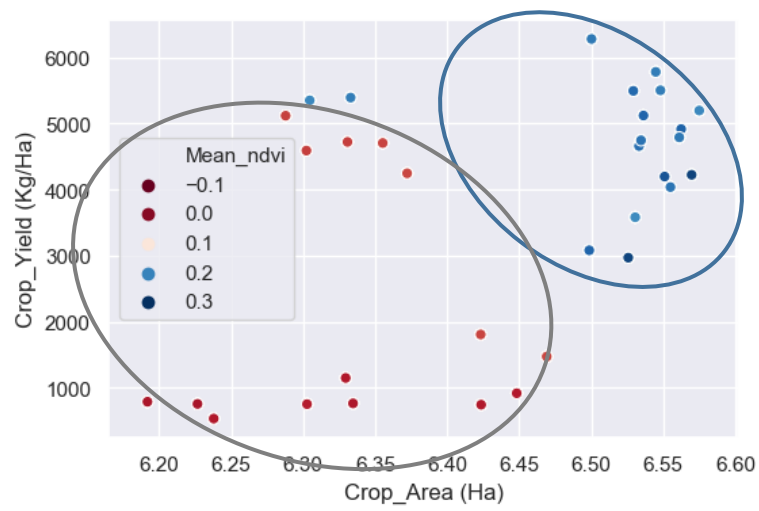


Figura 9. Gráfico cruzado: Rendimiento vs. Área de los lotes. Legenda: NDVI promedio.