

Untitled

May 1, 2025

```
[4]: import pandas as pd

data = pd.read_csv('C:/Users/brait/OneDrive/Escritorio/Proyecto 2/paulo/df.
↪csv', encoding='latin1')
print(data)
```

| | Nº Contrato | Fecha plantacion | Lat | Lon | \ |
|-----|-------------|------------------|------------|------------|---|
| 0 | 6000000118 | 11/19/2024 | -35.495617 | -71.607243 | |
| 1 | 6000000092 | 10/14/2024 | -33.749477 | -70.705993 | |
| 2 | 6000000230 | 10/30/2024 | -35.811234 | -71.487128 | |
| 3 | 6000000193 | 10/4/2024 | -35.133026 | -71.286488 | |
| 4 | 6000000127 | 10/11/2024 | -34.303551 | -70.764888 | |
| .. | ... | ... | ... | ... | |
| 311 | 6000000025 | 11/20/2024 | -34.053469 | -70.757050 | |
| 312 | 6000000224 | 11/14/2025 | -35.487041 | -71.650732 | |
| 313 | 6000000066 | 10/4/2024 | -33.804619 | -70.892049 | |
| 314 | 6000000180 | 12/7/2024 | -34.270117 | -71.368606 | |
| 315 | 6000000370 | 12/4/2024 | -36.080216 | -71.968250 | |

| | Productor | Zona | \ |
|-----|-------------------------------------|--------|---|
| 0 | AGRICOLA CORDILLERA SPA | Sur | |
| 1 | MAURICIO A. ROJAS GOMEZ AGRI.ARREND | Norte | |
| 2 | CARLOS ZENOBIO VERGARA IBAÑEZ | Sur | |
| 3 | GERALDINE EMA GÓMEZ VALENZUELA | Centro | |
| 4 | AGRICOLA MANUELITA SPA | Norte | |
| .. | ... | ... | |
| 311 | CRISTIAN A. ZAVALLA HURTADO E.I.R.L | Norte | |
| 312 | AGRICOLA BENSANREN SPA | Sur | |
| 313 | SOCIEDAD AGRICOLA BRISAS DEL CAMPO | Norte | |
| 314 | SOCIEDAD AGROTURISTICA CABAÑA BLAN | Centro | |
| 315 | SUGAL CHILE LIMITADA | Sur | |

| | Predio | Tecnico | \ |
|---|------------------------------------|----------------------------|---|
| 0 | PARCELA Nº 15 P.P. PORVENIR LOTE B | Luis Fernando Reyes | |
| 1 | LOTE Nº 1, CASA PATRONAL LINDEROS | RENE ORLANDO TERCILLA PE A | |
| 2 | PREDIO SANTA VICTORIA | Jose Daniel Castillo | |
| 3 | PARCELA Nº13 P.P EL CONDOR | Marcos Daniel Suazo | |
| 4 | PARCELA Nº 16, PP. LOURDES | RENE ORLANDO TERCILLA PE A | |

| | | |
|-----|------------------------------------|---------------------------|
| .. | ... | ... |
| 311 | PARCELA N° 43, PP. LA SOLEDAD | Ana Carolina Rojas |
| 312 | LOTE N°5 PREDIO SANTA INÉS | Braulio Alejandro Ramirez |
| 313 | PREDIO LO CHACON | Ana Carolina Rojas |
| 314 | HIJUELA N° 1, RESTO FUNDO SAN LUIS | Felipe Arturo Farias |
| 315 | PREDIO CUYUMILLACO DE PARRAL | Joao Dos Santos |

| | Tipo riego | Has | Mensuradas | Kilos Netos | % Total defectos | \ |
|-----|------------|------|------------|-------------|------------------|---|
| 0 | ZRGF | 12.3 | 2521450 | 8% | | |
| 1 | ZRGF | 9.2 | 1796820 | 6% | | |
| 2 | ZRGF | 13.6 | 2703260 | 12% | | |
| 3 | ZRGF | 3.6 | 676360 | 8% | | |
| 4 | ZRGF | 22.6 | 4258900 | 9% | | |
| .. | ... | ... | ... | ... | | |
| 311 | ZNOF | 11.3 | 324050 | 20% | | |
| 312 | ZRGF | 9.1 | 124330 | 10% | | |
| 313 | ZRGF | 16.0 | 66900 | 9% | | |
| 314 | ZRGF | 12.5 | 66500 | 20% | | |
| 315 | ZRGF | 49.2 | 221660 | 19% | | |

| | Rendimiento x ha | UNIDAD DE NEGOCIO | Presencia Orobanche | Nivel de Presión | \ |
|-----|------------------|-------------------|---------------------|------------------|---|
| 0 | 188.3 | AGRICULTORES | No | 1.Nula | |
| 1 | 184.4 | AGRICULTORES | No | 1.Nula | |
| 2 | 175.5 | AGRICULTORES | No | 1.Nula | |
| 3 | 172.4 | AGRICULTORES | No | 1.Nula | |
| 4 | 171.7 | AGRICULTORES | No | 1.Nula | |
| .. | ... | ... | ... | ... | |
| 311 | 23.0 | AGRICULTORES | No | 1.Nula | |
| 312 | 18.7 | AGRICULTORES | Si | 2.Baja | |
| 313 | 12.2 | AGRICULTORES | No | 1.Nula | |
| 314 | 4.7 | AGRICULTORES | Si | 4.Alta | |
| 315 | 4.5 | CAMPOS PROPIOS | No | 1.Nula | |

| | Obj rendimiento |
|-----|-----------------|
| 0 | 106 |
| 1 | 106 |
| 2 | 106 |
| 3 | 102 |
| 4 | 106 |
| .. | ... |
| 311 | 90 |
| 312 | 106 |
| 313 | 106 |
| 314 | 102 |
| 315 | 99 |

[316 rows x 17 columns]

```
[7]: valores_perdidos = data.isna()
print(data.isnull().sum())
```

```
Nº Contrato      0
Fecha plantacion  0
Lat              0
Lon              0
Productor        0
Zona             0
Predio           0
Tecnico          0
Tipo riego       0
Has Mensuradas   0
Kilos Netos      0
% Total defectos  0
Rendimiento x ha  0
UNIDAD DE NEGOCIO 0
Presencia Orobanch 0
Nivel de Presión  2
Obj rendimiento   0
dtype: int64
```

```
[9]: data = data.dropna()
print(data)
```

```

Nº Contrato Fecha plantacion Lat Lon \
0 6000000118 11/19/2024 -35.495617 -71.607243
1 6000000092 10/14/2024 -33.749477 -70.705993
2 6000000230 10/30/2024 -35.811234 -71.487128
3 6000000193 10/4/2024 -35.133026 -71.286488
4 6000000127 10/11/2024 -34.303551 -70.764888
.. ..
311 6000000025 11/20/2024 -34.053469 -70.757050
312 6000000224 11/14/2025 -35.487041 -71.650732
313 6000000066 10/4/2024 -33.804619 -70.892049
314 6000000180 12/7/2024 -34.270117 -71.368606
315 6000000370 12/4/2024 -36.080216 -71.968250
```

```

Productor Zona \
0 AGRICOLA CORDILLERA SPA Sur
1 MAURICIO A. ROJAS GOMEZ AGRI.ARREND Norte
2 CARLOS ZENOBIO VERGARA IBAÑEZ Sur
3 GERALDINE EMA GÓMEZ VALENZUELA Centro
4 AGRICOLA MANUELITA SPA Norte
.. ..
311 CRISTIAN A. ZAVALLA HURTADO E.I.R.L Norte
312 AGRICOLA BENSANREN SPA Sur
313 SOCIEDAD AGRICOLA BRISAS DEL CAMPO Norte
```

314 SOCIEDAD AGROTURISTICA CABAÑA BLAN Centro
 315 SUGAL CHILE LIMITADA Sur

| | Predio | Tecnico \ |
|-----|------------------------------------|----------------------------|
| 0 | PARCELA N° 15 P.P. PORVENIR LOTE B | Luis Fernando Reyes |
| 1 | LOTE N° 1, CASA PATRONAL LINDEROS | RENE ORLANDO TERCILLA PE A |
| 2 | PREDIO SANTA VICTORIA | Jose Daniel Castillo |
| 3 | PARCELA N°13 P.P EL CONDOR | Marcos Daniel Suazo |
| 4 | PARCELA N° 16, PP. LOURDES | RENE ORLANDO TERCILLA PE A |
| .. | ... | ... |
| 311 | PARCELA N° 43, PP. LA SOLEDAD | Ana Carolina Rojas |
| 312 | LOTE N°5 PREDIO SANTA INÉS | Braulio Alejandro Ramirez |
| 313 | PREDIO LO CHACON | Ana Carolina Rojas |
| 314 | HIJUELA N° 1, RESTO FUNDO SAN LUIS | Felipe Arturo Farias |
| 315 | PREDIO CUYUMILLACO DE PARRAL | Joao Dos Santos |

| | Tipo riego | Has Mensuradas | Kilos Netos | % Total defectos \ |
|-----|------------|----------------|-------------|--------------------|
| 0 | ZRGF | 12.3 | 2521450 | 8% |
| 1 | ZRGF | 9.2 | 1796820 | 6% |
| 2 | ZRGF | 13.6 | 2703260 | 12% |
| 3 | ZRGF | 3.6 | 676360 | 8% |
| 4 | ZRGF | 22.6 | 4258900 | 9% |
| .. | ... | ... | ... | ... |
| 311 | ZNOF | 11.3 | 324050 | 20% |
| 312 | ZRGF | 9.1 | 124330 | 10% |
| 313 | ZRGF | 16.0 | 66900 | 9% |
| 314 | ZRGF | 12.5 | 66500 | 20% |
| 315 | ZRGF | 49.2 | 221660 | 19% |

| | Rendimiento x ha | UNIDAD DE NEGOCIO | Presencia Orobanche | Nivel de Presión \ |
|-----|------------------|-------------------|---------------------|--------------------|
| 0 | 188.3 | AGRICULTORES | No | 1.Nula |
| 1 | 184.4 | AGRICULTORES | No | 1.Nula |
| 2 | 175.5 | AGRICULTORES | No | 1.Nula |
| 3 | 172.4 | AGRICULTORES | No | 1.Nula |
| 4 | 171.7 | AGRICULTORES | No | 1.Nula |
| .. | ... | ... | ... | ... |
| 311 | 23.0 | AGRICULTORES | No | 1.Nula |
| 312 | 18.7 | AGRICULTORES | Si | 2.Baja |
| 313 | 12.2 | AGRICULTORES | No | 1.Nula |
| 314 | 4.7 | AGRICULTORES | Si | 4.Alta |
| 315 | 4.5 | CAMPOS PROPIOS | No | 1.Nula |

| | Obj rendimiento | ceros_por_fila |
|---|-----------------|----------------|
| 0 | 106 | 0 |
| 1 | 106 | 0 |
| 2 | 106 | 0 |
| 3 | 102 | 0 |
| 4 | 106 | 0 |

```

..          ...
311          90          0
312          106         0
313          106         0
314          102         0
315          99          0

```

[314 rows x 18 columns]

```
[10]: valores_perdidos = data.isna()
      print(data.isnull().sum())
```

```

Nº Contrato          0
Fecha plantacion     0
Lat                  0
Lon                  0
Productor            0
Zona                 0
Predio               0
Tecnico              0
Tipo riego           0
Has Mensuradas       0
Kilos Netos          0
% Total defectos     0
Rendimiento x ha     0
UNIDAD DE NEGOCIO    0
Presencia Orobancha  0
Nivel de Presión     0
Obj rendimiento      0
ceros_por_fila       0
dtype: int64

```

```
[12]: #Como estan distribuidos los rendimientos?
```

```

rendimientos_ha = data['Rendimiento x ha']
rendimientos_ha

```

```

[12]: 0      188.3
      1      184.4
      2      175.5
      3      172.4
      4      171.7
      ...
      311     23.0
      312     18.7
      313     12.2
      314      4.7
      315      4.5

```

Name: Rendimiento x ha, Length: 314, dtype: float64

```
[14]: print(data['Rendimiento x ha'].describe())
```

```
count    314.000000
mean      93.105096
std       31.975025
min        4.500000
25%       72.650000
50%       93.000000
75%      114.425000
max      188.300000
```

Name: Rendimiento x ha, dtype: float64

count: 314 | Hay entonces 314 registros validos (no nulos) en esta columna mean: 93.10 | El rendimiento promedio por hectarea es de 93.10 (qq/ha) std: 31.98 | Los datos varian en promedio +/- 31.98 respecto al promedio. Esto indica una dispersion considerable de los datos. min: 4.50 | El valor mas bajo es muy inferior al promedio lo cual sugiere que puede haber algunos outliers negativos 25%: 72.65 | El 25% de los datos esta por debajo de 72.65 50%: 114.43 | la mediana es decir el valor central es casi igual a la media, lo que sugiere que la distribucion podria ser simetrica 75%: 114.43 | El 75% de los datos están por debajo de este valor mas: 188.30 | el valor mas alto es el doble del promedio, indicando posibles outliers positivos o rendimientos excepcionalmente buenos.

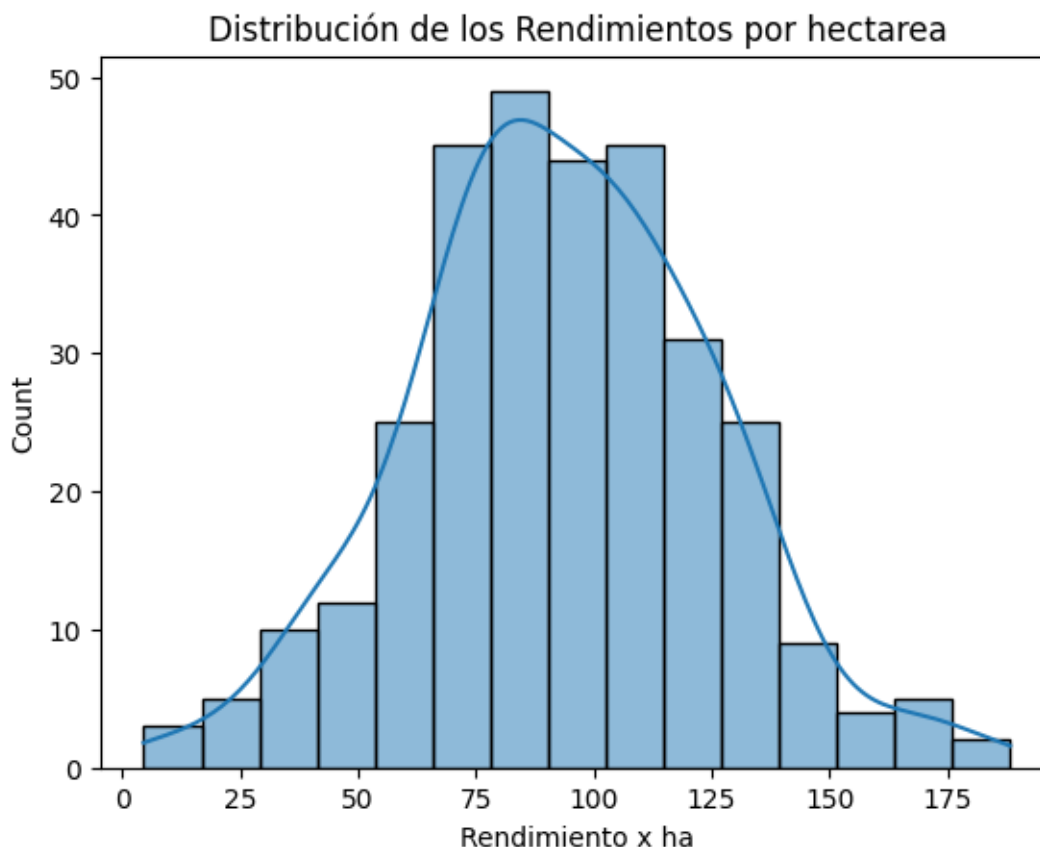
La mayoría de los rendimientos están entre ~73 y ~114, lo que te da un rango típico.

Hay algunos casos extremos, tanto muy bajos (~4.5) como muy altos (~188), que podrían merecer una revisión más detallada.

La media y la mediana están muy cercanas, lo que sugiere que no hay un sesgo fuerte hacia la izquierda o la derecha, aunque eso deberías confirmarlo con un histograma.

```
[15]: import matplotlib.pyplot as plt
import seaborn as sns

sns.histplot(data['Rendimiento x ha'], kde=True) # Histograma con curva de densidad
plt.title("Distribución de los Rendimientos por hectarea")
plt.show()
```



1. ¿Es simétrica o asimétrica? La distribución es casi simétrica, pero muestra una ligera asimetría hacia la derecha. La curva de densidad y el histograma son visualmente cercanos a una forma simétrica, aunque hay una leve cola más larga del lado derecho.
2. ¿Tiene forma de campana (normal)? Sí, la distribución tiene una forma de campana, lo que sugiere que se aproxima a una distribución normal, aunque no es perfectamente normal (por la ligera cola derecha)
3. ¿Está sesgada a la izquierda o derecha? Está ligeramente sesgada a la derecha (sesgo positivo), ya que:

La media (93.1) es ligeramente mayor que la mediana (93.0).

Hay algunos valores altos (outliers) que alargan la cola hacia la derecha

4. ¿Hay muchos ceros o valores extremos (outliers)? No hay muchos ceros: El valor mínimo es 4.5, así que no hay acumulación en 0.

Sí hay valores extremos, especialmente hacia el extremo superior (máximo de 188.3), lo que contribuye al sesgo a la derecha.

5. ¿Se concentra en ciertos rangos? Sí, la mayoría de los rendimientos se concentran entre 70 y 115 unidades por hectárea. Esto se ve tanto en la gráfica (los bins más altos están en ese rango) como en los percentiles:

25% = 72.65

75% = 114.43

Hacemos lo mismo pero para la columna de los rendimientos esperados.

```
[16]: rendimiento_objetivo = data['Obj rendimiento']
rendimiento_objetivo
```

```
[16]: 0      106
      1      106
      2      106
      3      102
      4      106
      ...
     311      90
     312      106
     313      106
     314      102
     315      99
      Name: Obj rendimiento, Length: 314, dtype: int64
```

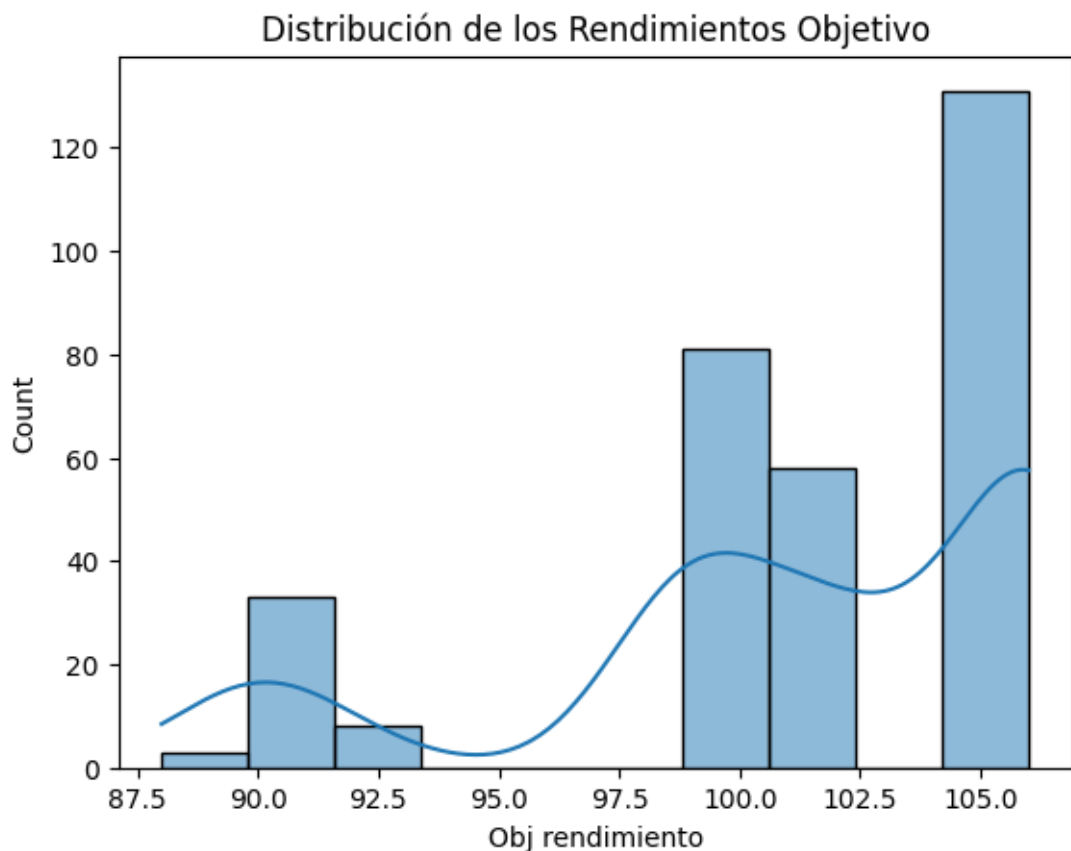
```
[18]: print(data['Obj rendimiento'].describe())
```

```
count      314.000000
mean       101.245223
std         5.298176
min         88.000000
25%         99.000000
50%        102.000000
75%        106.000000
max        106.000000
      Name: Obj rendimiento, dtype: float64
```

count: 314 | Hay 314 observaciones (rendimientos objetivo medidos). mean: 101.25 | El promedio de los rendimientos objetivo es 101.25. std: 5.30 | La desviación estándar es baja, lo que indica poca dispersión. min: 88.00 | El valor más bajo es 88, posible caso atípico, ya que se aleja del resto. 25% (Q1): 99.00 | El 25% de los datos están por debajo de 99. 50% (mediana): 102.00 | La mediana es 102, muy cercana a la media → ligera simetría. 75% (Q3): 106.00 | El 75% de los datos están por debajo de 106. max: 106.00 | El valor máximo también es 106, igual que el tercer cuartil → concentración alta en ese extremo.

```
[19]: import matplotlib.pyplot as plt
import seaborn as sns

sns.histplot(data['Obj rendimiento'], kde=True) # Histograma con curva de
↪densidad
plt.title("Distribución de los Rendimientos Objetivo")
plt.show()
```

1. ¿Es simétrica o asimétrica? Es asimétrica, y muestra una bimodalidad leve (dos picos), lo que rompe la simetría. La mayor concentración está en el extremo derecho.
2. ¿Tiene forma de campana (normal)? No, esta distribución no tiene forma de campana. La densidad muestra varias ondulaciones y una fuerte acumulación hacia la derecha (más de tipo escalón que de campana).
3. ¿Está sesgada a la izquierda o derecha? La distribución está sesgada a la izquierda (sesgo negativo), ya que:

Hay una fuerte concentración de valores altos (105), y una cola hacia valores más bajos (87.5–95).

Esto indica que la mayoría de los datos se acumulan en el lado derecho del eje.

4. ¿Hay muchos ceros o valores extremos (outliers)? No hay ceros.

No se observan outliers extremos visualmente, ya que los datos están agrupados en un rango relativamente estrecho (~88 a 106). Sin embargo, los valores bajos (<90) podrían considerarse moderadamente atípicos, dependiendo del contexto.

5. ¿Se concentra en ciertos rangos? Sí, los valores se concentran principalmente entre 100 y 106. La barra de 105 es la más alta con diferencia, lo que indica que muchos datos están ahí.

Los datos están muy concentrados en un rango pequeño (entre 99 y 106), con una leve cola hacia

valores bajos (88–95), lo cual se refleja también en la gráfica.

La asimetría leve hacia la izquierda es coherente con el valor mínimo (88) siendo mucho menor que la mediana (102) y el máximo (106).

Podría haber un objetivo común en los rendimientos fijado en torno a 105 o 106, lo cual explicaría la acumulación en ese extremo.

Rendimientos por hectárea muestran una distribución más natural y extendida, típica de datos observados en campo, con algunas observaciones altas que influyen en la media.

Rendimientos objetivo, en cambio, parecen definidos por un estándar, con muchos datos agrupados cerca del 105, lo cual sugiere una meta establecida, no necesariamente alcanzada de forma uniforme.

La asimetría opuesta entre ambas (derecha vs. izquierda) indica que muchas observaciones reales no logran alcanzar el rendimiento objetivo, o que existe una variabilidad considerable entre las metas y lo que realmente se obtiene.

```
[ ]: #Cual es el rendimiento promedio?
```

```
Rendimiento promedio real por hectarea: 93.11 aproximadamente
```

```
Rendimiento objetivo promedio: 101.25 aproximadamente
```

```
en comparacion:
```

1. En promedio los rendimientos reales están 8.14 unidades por debajo del objetivo
2. Esto indica una brecha entre el rendimiento esperado y lo alcanzado.

```
[23]: #Cuanto varian los rendimientos entre ellos?
```

```
# Convertir fechas a tipo datetime y asignar semana
```

```
data["Fecha plantacion"] = pd.to_datetime(data["Fecha plantacion"])
```

```
data["semana"] = data["Fecha plantacion"].dt.isocalendar().week
```

```
semanal = data.groupby("semana").agg({  
    "Rendimiento x ha": "mean",  
    "Obj rendimiento": "mean"  
}).reset_index()
```

```
semanal["variacion"] = semanal["rendimiento_objetivo"] -  
    semanal["rendimiento_real"]
```

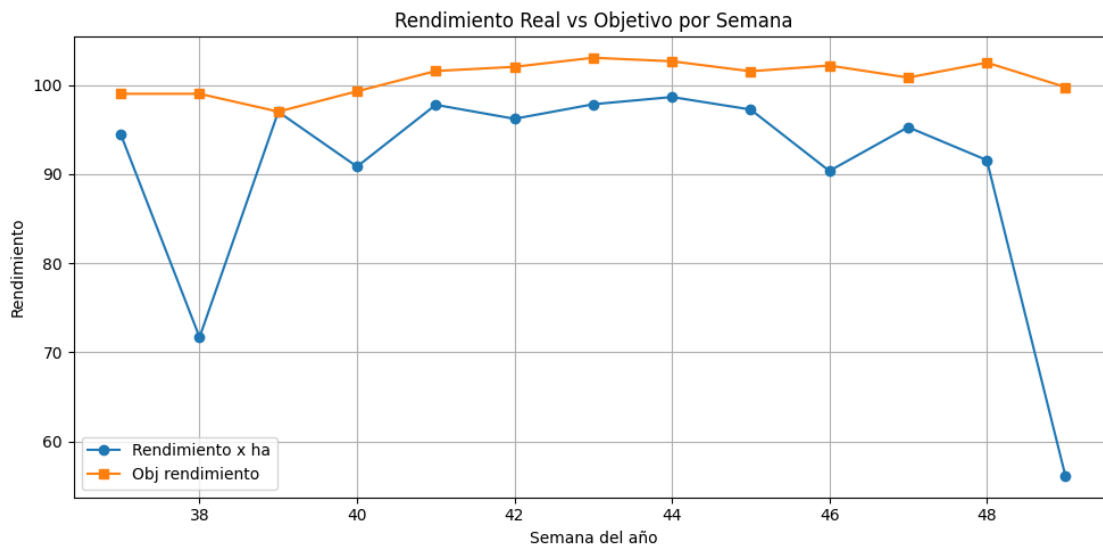
```
[28]: plt.figure(figsize=(10, 5))  
plt.plot(semanal["semana"], semanal["Rendimiento x ha"], marker='o',  
    label="Rendimiento x ha")  
plt.plot(semanal["semana"], semanal["Obj rendimiento"], marker='s', label="Obj  
    rendimiento")  
plt.title("Rendimiento Real vs Objetivo por Semana")  
plt.xlabel("Semana del año")
```

```

plt.ylabel("Rendimiento")
plt.legend()
plt.grid(True)
plt.tight_layout()
plt.show()

# Histograma de la variación semanal
plt.figure(figsize=(8, 5))
plt.hist(semanal["variacion"], bins=10, edgecolor='black')
plt.title("Distribución de la Variación (Objetivo - Real)")
plt.xlabel("Variación")
plt.ylabel("Frecuencia")
plt.grid(True)
plt.tight_layout()
plt.show()

```



```

-----
KeyError                                Traceback (most recent call last)
File ~\AppData\Local\Programs\Python\Python313\Lib\site-packages\pandas\core\indexers\base.
  ~\AppData\Local\Programs\Python\Python313\Lib\site-packages\pandas\core\indexers\base.py:3805, in Index.get_loc(self, key)
    3804 try:
-> 3805     return self._engine.get_loc(casted_key)
    3806 except KeyError as err:

File index.pyx:167, in pandas._libs.index.IndexEngine.get_loc()

File index.pyx:196, in pandas._libs.index.IndexEngine.get_loc()

```

```
File pandas\\_libs\\hashtable_class_helper.pxi:7081, in pandas._libs.hashtable.  
↳PyObjectHashTable.get_item()
```

```
File pandas\\_libs\\hashtable_class_helper.pxi:7089, in pandas._libs.hashtable.  
↳PyObjectHashTable.get_item()
```

KeyError: 'variacion'

The above exception was the direct cause of the following exception:

KeyError Traceback (most recent call last)

Cell In[28], line 14

```
12 # Histograma de la variación semanal  
13 plt.figure(figsize=(8, 5))  
----> 14 plt.hist(semanal["variacion"], bins=10, edgecolor='black')  
15 plt.title("Distribución de la Variación (Objetivo - Real)")  
16 plt.xlabel("Variación")
```

File

```
↳~\AppData\Local\Programs\Python\Python313\Lib\site-packages\pandas\core\frame  
↳py:4102, in DataFrame.__getitem__(self, key)  
4100 if self.columns.nlevels > 1:  
4101     return self._getitem_multilevel(key)  
-> 4102 indexer = self.columns.get_loc(key)  
4103 if is_integer(indexer):  
4104     indexer = [indexer]
```

File

```
↳~\AppData\Local\Programs\Python\Python313\Lib\site-packages\pandas\core\indexes\base.  
↳py:3812, in Index.get_loc(self, key)  
3807 if isinstance(casted_key, slice) or (  
3808     isinstance(casted_key, abc.Iterable)  
3809     and any(isinstance(x, slice) for x in casted_key)  
3810 ):  
3811     raise InvalidIndexError(key)  
-> 3812 raise KeyError(key) from err  
3813 except TypeError:  
3814     # If we have a listlike key, _check_indexing_error will raise  
3815     # InvalidIndexError. Otherwise we fall through and re-raise  
3816     # the TypeError.  
3817     self._check_indexing_error(key)
```

KeyError: 'variacion'

<Figure size 800x500 with 0 Axes>

En general, los rendimientos reales por hectárea se mantuvieron por debajo del rendimiento objetivo a lo largo de las semanas 37 a 49. Las semanas 37 y 49 presentan las mayores desviaciones

negativas, con una diferencia de más de 25 puntos respecto al objetivo. Las semanas más cercanas al cumplimiento del objetivo fueron de la 41 a la 45, donde la diferencia se redujo a menos de 5 puntos. Esto sugiere que hubo factores que afectaron negativamente la producción en semanas específicas, mientras que otras semanas se mantuvo un buen nivel de eficiencia. Sería útil investigar qué ocurrió en las semanas con mayor caída para corregir posibles problemas en el futuro.

#Cual es la desviacionentre el rendimiento esperado y el rendimiento real.

```
[36]: # Diferencia absoluta
data['Diferencia_absoluta'] = abs(data['Rendimiento x ha'] - data['Obj_
rendimiento'])

# Error porcentual absoluto
data['Error_porcentual'] = abs(data['Rendimiento x ha'] - data['Obj_
rendimiento']) / data['Rendimiento x ha'] * 100
print(data['Diferencia_absoluta'])
```

```
0      82.3
1      78.4
2      69.5
3      70.4
4      65.7
...
311    67.0
312    87.3
313    93.8
314    97.3
315    94.5
Name: Diferencia_absoluta, Length: 314, dtype: float64
```

```
[35]: print(data['Error_porcentual'])
```

```
0      43.706851
1      42.516269
2      39.601140
3      40.835267
4      38.264415
...
311    291.304348
312    466.844920
313    768.852459
314   2070.212766
315   2100.000000
Name: Error_porcentual, Length: 314, dtype: float64
```

```
[37]: desviacion_promedio = data['Diferencia_absoluta'].mean()
mape = data['Error_porcentual'].mean()

print(f"Desviación promedio: {desviacion_promedio:.2f}")
```

```
print(f"Error porcentual medio (MAPE): {mape:.2f}%")
```

Desviación promedio: 25.68

Error porcentual medio (MAPE): 52.30%

```
[39]: data['Diferencia'] = data['Rendimiento x ha'] - data['Obj rendimiento']  
std_diferencia = data['Diferencia'].std()  
  
print(f"Desviación estándar de las diferencias: {std_diferencia:.2f}")
```

Desviación estándar de las diferencias: 31.08

```
[ ]:
```

Untitled

May 2, 2025

```
[9]: import pandas as pd
data = pd.read_csv('C:/Users/brait/OneDrive/Escritorio/Proyecto 2/paulo/df.
↳csv', encoding='latin1')
print(data)
```

| | Nº Contrato | Fecha plantacion | Lat | Lon \ |
|-----|-------------|------------------|------------|------------|
| 0 | 6000000118 | 11/19/2024 | -35.495617 | -71.607243 |
| 1 | 6000000092 | 10/14/2024 | -33.749477 | -70.705993 |
| 2 | 6000000230 | 10/30/2024 | -35.811234 | -71.487128 |
| 3 | 6000000193 | 10/4/2024 | -35.133026 | -71.286488 |
| 4 | 6000000127 | 10/11/2024 | -34.303551 | -70.764888 |
| .. | ... | ... | ... | ... |
| 311 | 6000000025 | 11/20/2024 | -34.053469 | -70.757050 |
| 312 | 6000000224 | 11/14/2025 | -35.487041 | -71.650732 |
| 313 | 6000000066 | 10/4/2024 | -33.804619 | -70.892049 |
| 314 | 6000000180 | 12/7/2024 | -34.270117 | -71.368606 |
| 315 | 6000000370 | 12/4/2024 | -36.080216 | -71.968250 |

| | Productor | Zona \ |
|-----|-------------------------------------|--------|
| 0 | AGRICOLA CORDILLERA SPA | Sur |
| 1 | MAURICIO A. ROJAS GOMEZ AGRI.ARREND | Norte |
| 2 | CARLOS ZENOBIO VERGARA IBAÑEZ | Sur |
| 3 | GERALDINE EMA GÓMEZ VALENZUELA | Centro |
| 4 | AGRICOLA MANUELITA SPA | Norte |
| .. | ... | ... |
| 311 | CRISTIAN A. ZAVALLA HURTADO E.I.R.L | Norte |
| 312 | AGRICOLA BENSANREN SPA | Sur |
| 313 | SOCIEDAD AGRICOLA BRISAS DEL CAMPO | Norte |
| 314 | SOCIEDAD AGROTURISTICA CABAÑA BLAN | Centro |
| 315 | SUGAL CHILE LIMITADA | Sur |

| | Predio | Tecnico \ |
|----|------------------------------------|----------------------------|
| 0 | PARCELA Nº 15 P.P. PORVENIR LOTE B | Luis Fernando Reyes |
| 1 | LOTE Nº 1, CASA PATRONAL LINDEROS | RENE ORLANDO TERCILLA PE A |
| 2 | PREDIO SANTA VICTORIA | Jose Daniel Castillo |
| 3 | PARCELA Nº13 P.P EL CONDOR | Marcos Daniel Suazo |
| 4 | PARCELA Nº 16, PP. LOURDES | RENE ORLANDO TERCILLA PE A |
| .. | ... | ... |

| | | |
|-----|------------------------------------|---------------------------|
| 311 | PARCELA N° 43, PP. LA SOLEDAD | Ana Carolina Rojas |
| 312 | LOTE N°5 PREDIO SANTA INÉS | Braulio Alejandro Ramirez |
| 313 | PREDIO LO CHACON | Ana Carolina Rojas |
| 314 | HIJUELA N° 1, RESTO FUNDO SAN LUIS | Felipe Arturo Farias |
| 315 | PREDIO CUYUMILLACO DE PARRAL | Joao Dos Santos |

| | Tipo riego | Has Mensuradas | Kilos Netos | % Total defectos | \ |
|-----|------------|----------------|-------------|------------------|---|
| 0 | ZRGF | 12.3 | 2521450 | 8% | |
| 1 | ZRGF | 9.2 | 1796820 | 6% | |
| 2 | ZRGF | 13.6 | 2703260 | 12% | |
| 3 | ZRGF | 3.6 | 676360 | 8% | |
| 4 | ZRGF | 22.6 | 4258900 | 9% | |
| .. | ... | ... | ... | ... | |
| 311 | ZNOF | 11.3 | 324050 | 20% | |
| 312 | ZRGF | 9.1 | 124330 | 10% | |
| 313 | ZRGF | 16.0 | 66900 | 9% | |
| 314 | ZRGF | 12.5 | 66500 | 20% | |
| 315 | ZRGF | 49.2 | 221660 | 19% | |

| | Rendimiento x ha | UNIDAD DE NEGOCIO | Presencia Orobanche | Nivel de Presión | \ |
|-----|------------------|-------------------|---------------------|------------------|---|
| 0 | 188.3 | AGRICULTORES | No | 1.Nula | |
| 1 | 184.4 | AGRICULTORES | No | 1.Nula | |
| 2 | 175.5 | AGRICULTORES | No | 1.Nula | |
| 3 | 172.4 | AGRICULTORES | No | 1.Nula | |
| 4 | 171.7 | AGRICULTORES | No | 1.Nula | |
| .. | ... | ... | ... | ... | |
| 311 | 23.0 | AGRICULTORES | No | 1.Nula | |
| 312 | 18.7 | AGRICULTORES | Si | 2.Baja | |
| 313 | 12.2 | AGRICULTORES | No | 1.Nula | |
| 314 | 4.7 | AGRICULTORES | Si | 4.Alta | |
| 315 | 4.5 | CAMPOS PROPIOS | No | 1.Nula | |

| | Obj rendimiento |
|-----|-----------------|
| 0 | 106 |
| 1 | 106 |
| 2 | 106 |
| 3 | 102 |
| 4 | 106 |
| .. | ... |
| 311 | 90 |
| 312 | 106 |
| 313 | 106 |
| 314 | 102 |
| 315 | 99 |

[316 rows x 17 columns]


```
[10]: valores_perdidos =data.isna()
print(data.isnull().sum())
```

```
Nº Contrato      0
Fecha plantacion  0
Lat              0
Lon              0
Productor        0
Zona             0
Predio           0
Tecnico          0
Tipo riego       0
Has Mensuradas   0
Kilos Netos      0
% Total defectos  0
Rendimiento x ha  0
UNIDAD DE NEGOCIO 0
Presencia Orobanch 0
Nivel de Presión  2
Obj rendimiento   0
dtype: int64
```

```
[11]: data= data.dropna()
print(data)
```

```

      Nº Contrato Fecha plantacion      Lat      Lon \
0      6000000118      11/19/2024 -35.495617 -71.607243
1      6000000092      10/14/2024 -33.749477 -70.705993
2      6000000230      10/30/2024 -35.811234 -71.487128
3      6000000193       10/4/2024 -35.133026 -71.286488
4      6000000127      10/11/2024 -34.303551 -70.764888
..      ...
311    6000000025      11/20/2024 -34.053469 -70.757050
312    6000000224      11/14/2025 -35.487041 -71.650732
313    6000000066       10/4/2024 -33.804619 -70.892049
314    6000000180      12/7/2024 -34.270117 -71.368606
315    6000000370      12/4/2024 -36.080216 -71.968250
```

```

              Productor      Zona \
0      AGRICOLA CORDILLERA SPA      Sur
1  MAURICIO A. ROJAS GOMEZ AGRI.ARREND  Norte
2      CARLOS ZENOBIO VERGARA IBAÑEZ      Sur
3      GERALDINE EMA GÓMEZ VALENZUELA  Centro
4      AGRICOLA MANUELITA SPA      Norte
..      ...
311  CRISTIAN A. ZAVALLA HURTADO E.I.R.L  Norte
312      AGRICOLA BENSANREN SPA      Sur
313  SOCIEDAD AGRICOLA BRISAS DEL CAMPO  Norte
```

314 SOCIEDAD AGROTURISTICA CABAÑA BLAN Centro
 315 SUGAL CHILE LIMITADA Sur

| | Predio | Tecnico \ |
|-----|------------------------------------|----------------------------|
| 0 | PARCELA N° 15 P.P. PORVENIR LOTE B | Luis Fernando Reyes |
| 1 | LOTE N° 1, CASA PATRONAL LINDEROS | RENE ORLANDO TERCILLA PE A |
| 2 | PREDIO SANTA VICTORIA | Jose Daniel Castillo |
| 3 | PARCELA N°13 P.P EL CONDOR | Marcos Daniel Suazo |
| 4 | PARCELA N° 16, PP. LOURDES | RENE ORLANDO TERCILLA PE A |
| .. | ... | ... |
| 311 | PARCELA N° 43, PP. LA SOLEDAD | Ana Carolina Rojas |
| 312 | LOTE N°5 PREDIO SANTA INÉS | Braulio Alejandro Ramirez |
| 313 | PREDIO LO CHACON | Ana Carolina Rojas |
| 314 | HIJUELA N° 1, RESTO FUNDO SAN LUIS | Felipe Arturo Farias |
| 315 | PREDIO CUYUMILLACO DE PARRAL | Joao Dos Santos |

| | Tipo riego | Has Mensuradas | Kilos Netos | % Total defectos \ |
|-----|------------|----------------|-------------|--------------------|
| 0 | ZRGF | 12.3 | 2521450 | 8% |
| 1 | ZRGF | 9.2 | 1796820 | 6% |
| 2 | ZRGF | 13.6 | 2703260 | 12% |
| 3 | ZRGF | 3.6 | 676360 | 8% |
| 4 | ZRGF | 22.6 | 4258900 | 9% |
| .. | ... | ... | ... | ... |
| 311 | ZNOF | 11.3 | 324050 | 20% |
| 312 | ZRGF | 9.1 | 124330 | 10% |
| 313 | ZRGF | 16.0 | 66900 | 9% |
| 314 | ZRGF | 12.5 | 66500 | 20% |
| 315 | ZRGF | 49.2 | 221660 | 19% |

| | Rendimiento x ha | UNIDAD DE NEGOCIO | Presencia Orobanche | Nivel de Presión \ |
|-----|------------------|-------------------|---------------------|--------------------|
| 0 | 188.3 | AGRICULTORES | No | 1.Nula |
| 1 | 184.4 | AGRICULTORES | No | 1.Nula |
| 2 | 175.5 | AGRICULTORES | No | 1.Nula |
| 3 | 172.4 | AGRICULTORES | No | 1.Nula |
| 4 | 171.7 | AGRICULTORES | No | 1.Nula |
| .. | ... | ... | ... | ... |
| 311 | 23.0 | AGRICULTORES | No | 1.Nula |
| 312 | 18.7 | AGRICULTORES | Si | 2.Baja |
| 313 | 12.2 | AGRICULTORES | No | 1.Nula |
| 314 | 4.7 | AGRICULTORES | Si | 4.Alta |
| 315 | 4.5 | CAMPOS PROPIOS | No | 1.Nula |

| | Obj rendimiento |
|---|-----------------|
| 0 | 106 |
| 1 | 106 |
| 2 | 106 |
| 3 | 102 |
| 4 | 106 |

```

..          ...
311          90
312          106
313          106
314          102
315          99

```

[314 rows x 17 columns]

[19]: *#Hay alguna relacion entre los rendimientos y los otros parametros en amarillo?*

```

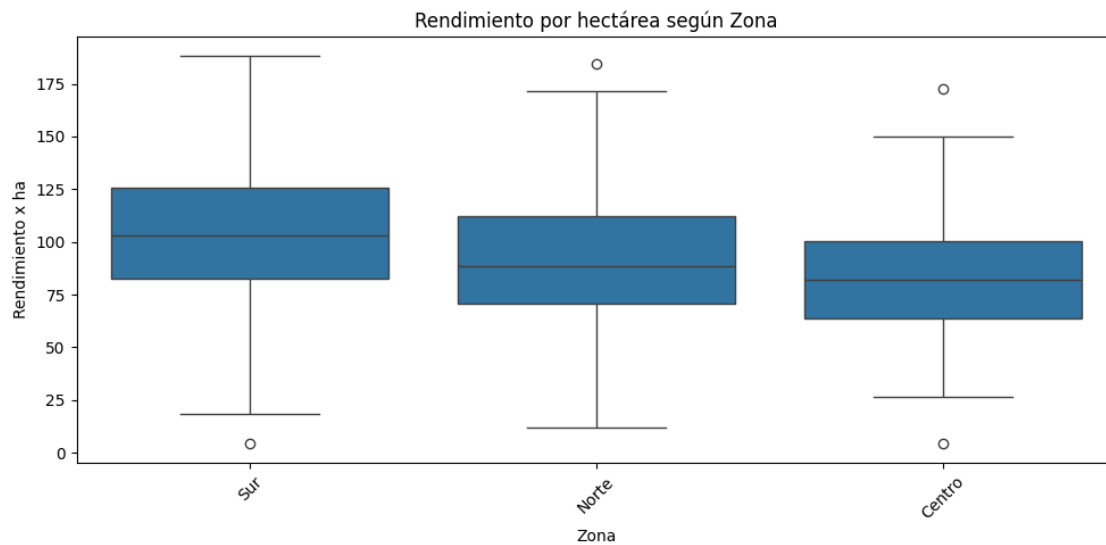
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

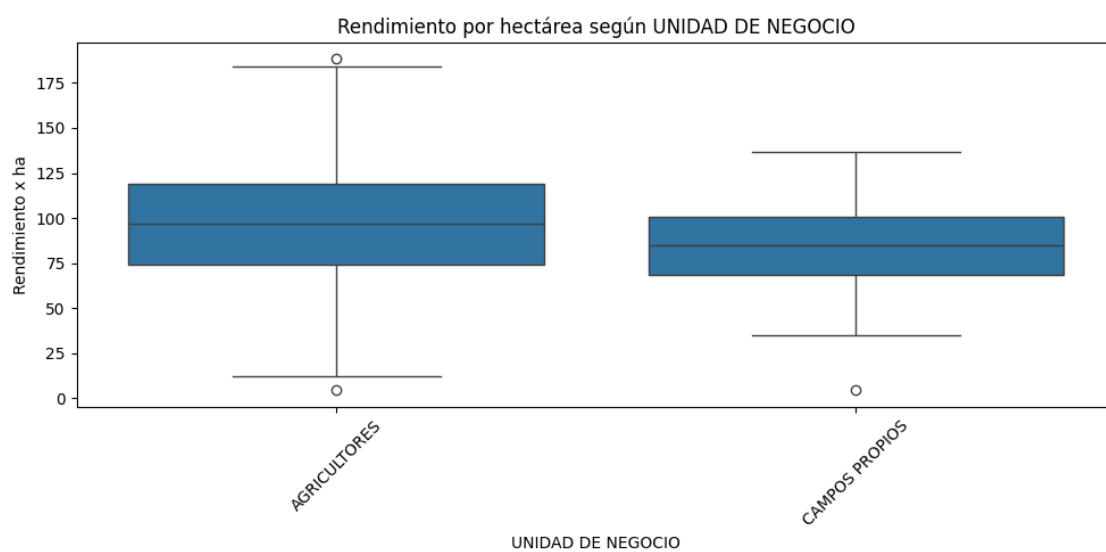
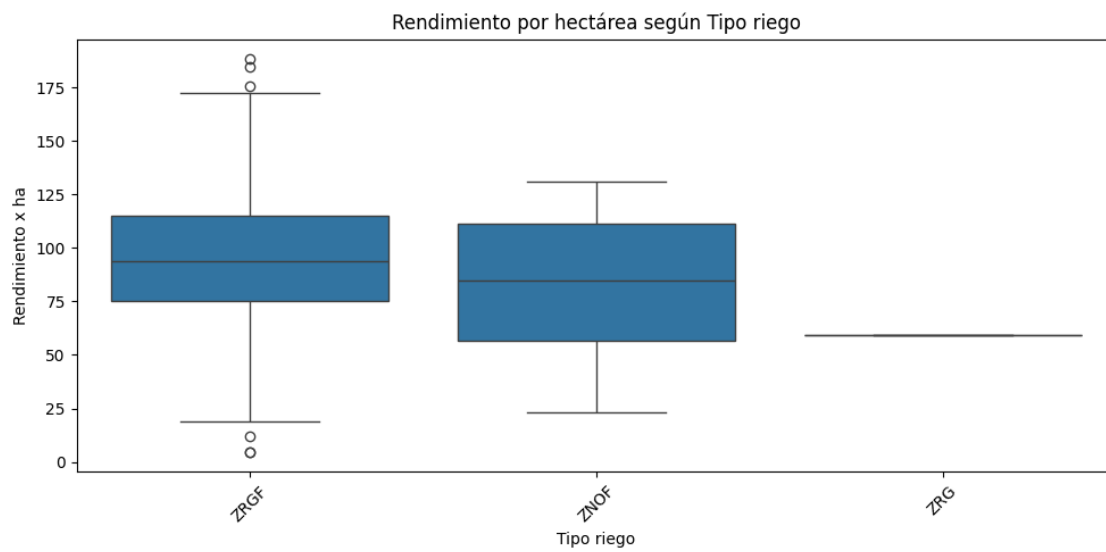
categorical_vars = ['Zona', 'Tipo riego', 'UNIDAD DE NEGOCIO', 'Presencia_
↳Orobanche', 'Nivel de Presión', 'Obj rendimiento']

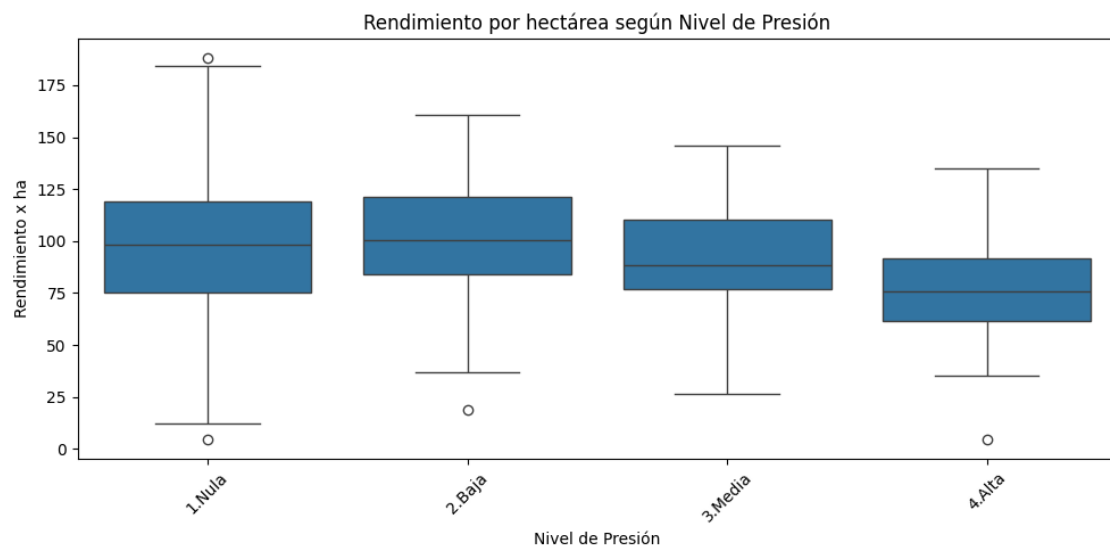
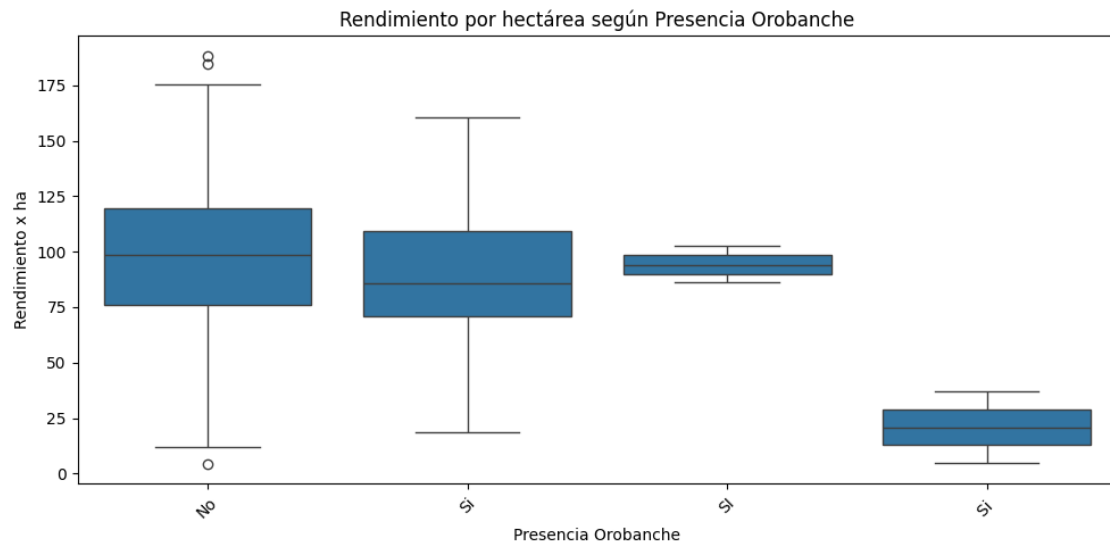
for var in categorical_vars:
    plt.figure(figsize=(10, 5))
    sns.boxplot(x=var, y='Rendimiento x ha', data=data)
    plt.title(f'Rendimiento por hectárea según {var}')
    plt.xticks(rotation=45)
    plt.savefig("rendimiento_zona.png", dpi=300, bbox_inches='tight')
    plt.tight_layout()

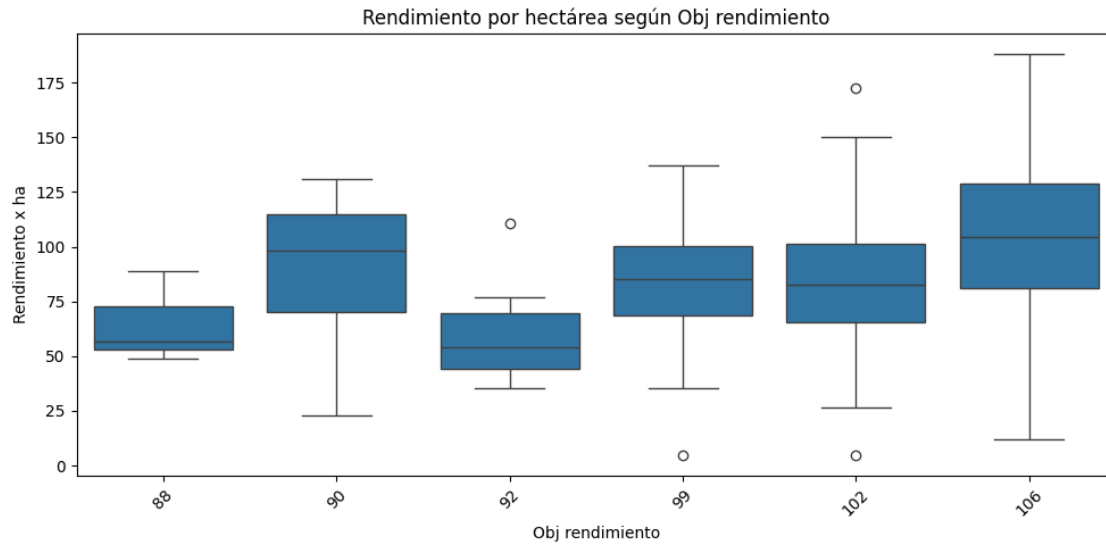
plt.show()

```









#ANALISIS POR ZONA:

Zona Sur Mediana (línea dentro de la caja): Cerca de 100. Rango intercuartílico (IQR) (la caja): Aproximadamente de 75 a 130. Valores extremos (outliers): Un valor muy bajo cerca de 5. Máximo y mínimo (sin contar outliers): Aproximadamente de 20 a 185.

Zona Norte

Mediana: Cerca de 95. IQR: De unos 70 a 120. Outliers: Uno bajo (cerca de 10) y uno alto (alrededor de 180). Máximos y mínimos (sin outliers): Entre 20 y 160 aprox.

Zona Centro

Mediana: Alrededor de 90. IQR: De 65 a 110. Outliers: Uno bajo (cerca de 10) y uno alto (cerca de 170). Rango general: De 25 a 150 aprox.

Comparación general:

Mayor rendimiento: En promedio, la zona Sur tiene la mediana más alta.

Dispersión: La zona Sur también parece tener una mayor dispersión (rango más amplio).

Outliers: Todas las zonas presentan valores atípicos bajos, y Centro y Norte también tienen altos.

Rendimiento más consistente: La zona Centro tiene la caja (IQR) más compacta.

#ANALISIS DE RENDIMIENTO POR HECTAREA SEGUN SU TIPO DE RIEGO:

1. ZRF Mediana: ~100 IQR: entre ~75 y ~120 Rango general: de ~10 a ~185 Outliers: Varios valores bajos (~5-10) y algunos muy altos (~180)

ZNOF Mediana: También alrededor de 100 IQR: más amplio, de ~50 a ~125 Rango general: más estrecho que ZRF, de ~20 a ~130 Outliers: Menos visibles; parece más homogéneo, pero con mayor dispersión hacia el mínimo

3. ZRG Observación clave: Solo hay un valor (línea horizontal sin caja), lo que sugiere que solo hay un dato disponible — no se puede comparar estadísticamente con los otros tipos.

..... ZRF y ZNOF tienen rendimientos similares en la mediana. ZRF muestra más variabilidad (y más outliers). ZRG tiene datos insuficientes para análisis.

#ANALISIS DE RENDIMIENTO POR EL TIPO DE UNIDAD DE NEGOCIO

1. AGRICULTORES:

Mediana: Cerca de 105 IQR: De ~80 a ~130 Rango total: De ~10 a ~185 Outliers: Varios bajos (~5-10) y uno muy alto (~180)

2. CAMPOS PROPIOS:

Mediana: Más baja, ~90 IQR: ~60 a ~110 Rango: más estrecho que Agricultores (~20 a ~150) Outliers: Uno muy bajo (~5), menos extremos

..... AGRICULTORES tienen mayor rendimiento mediano y mayor variabilidad.

CAMPOS PROPIOS presentan menor rendimiento promedio y menos dispersión, aunque también tienen outliers.

Tipo de riego: No hay grandes diferencias entre ZRF y ZNOF, pero ZRF parece tener rendimientos más altos y dispersos.

Unidad de negocio: AGRICULTORES superan a CAMPOS PROPIOS tanto en mediana como en rango superior de rendimientos, aunque también presentan más variabilidad y outliers.

[]: #ANALISIS DE RENDIMIENTO POR HECTARA SEGUN PRESENCIA DE OROBANQUE:

Hay 4 categorías: No y tres Sí (posiblemente un error en los datos, ya que se repite SI varias veces, pero parecen representar distintas intensidades o grupos).

No (sin presencia de Orobanche)

Mediana más alta, superior a 100

IQR: de ~80 a ~130

Rango total: ~5 a ~180

Outliers: varios en ambos extremos

Primer "Sí" (con Orobanche)

Mediana: baja, ~90

IQR: ~65 a ~115

Más dispersión hacia valores bajos

Segundo "Sí"

Caja muy pequeña, mediana ~95, pero con rango estrecho (~85 a ~100)

Baja variabilidad, pocos datos

Tercer "Sí"

Mediana muy baja, ~20-30
IQR: ~10 a ~35
Rendimiento claramente inferior

.....
La presencia de Orobanche impacta negativamente el rendimiento.
Cuanto más severa o avanzada es la infestación (según las categorías duplicadas, de Sí),
más bajo es el rendimiento.
El grupo "No" claramente tiene los rendimientos más altos.

[]: #ANALISIS DE RENDIMIENTO X HECTAREA SEGUN EL NIVEL PRESION.

Niveles: 1. Nula, 2. Baja, 3. Media, 4. Alta

1. Nula y 2. Baja
Medianas: Altas (~100-110)

Rangos amplios: muchos valores altos, incluso outliers por encima de 175

IQR: similares (~80 a ~130)

3. Media
Mediana: Ligeramente menor (~95)

Menor dispersión en general

4. Alta
Mediana más baja, entre ~75 y ~80

IQR más bajo y estrecho (~60 a ~90)

Menos valores extremos altos

Conclusión
A mayor presión, menor el rendimiento por hectárea.
La tendencia es descendente del rendimiento conforme aumenta la presión.
La diferencia entre "Nula/Baja" y "Alta" es significativa visualmente.

La presencia de Orobanche y el aumento en el nivel de presión están asociados con una disminución clara en el rendimiento por hectárea.

Es probable que ambas variables estén relacionadas entre sí (la presión puede deberse en parte a Orobanche u otras plagas).

Los datos también sugieren que hay algunos errores o redundancias en las etiquetas ("Sí", "Sí", "Sí") que convendría revisar y corregir.

#ANÁLISIS DE RENDIMIENTO X HECTÁREA SEGÚN RENDIMIENTO OBJETIVO:

A partir de este análisis visual, parece haber una tendencia general de aumento en el rendimiento por hectárea a medida que el rendimiento objetivo se incrementa, especialmente al comparar los grupos con objetivos más bajos (88 y 92) con los grupos de objetivos más altos (99, 102 y 106).

Sin embargo, la relación no es perfectamente lineal ni directa. Observamos fluctuaciones (como la disminución en el rendimiento para el objetivo de 92) y variaciones en la dispersión de los datos para cada objetivo. La presencia de valores atípicos también indica casos particulares que se desvían de la tendencia general.

```
[33]: import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

data = pd.read_csv('C:/Users/brait/OneDrive/Escritorio/Proyecto 2/paulo/df.
↳csv', encoding='latin1')
print(data)

# Asegúrate de que 'data' está definido antes
# data = pd.read_csv('tu_archivo.csv') # o como lo hayas cargado

categorical_vars = ['Lat', 'Lon', 'Fecha plantacion']

for var in categorical_vars:
    if var in data.columns:
        plt.figure(figsize=(10, 5))
        sns.boxplot(x=var, y='Rendimiento x ha', data=data)
        plt.title(f'Rendimiento por hectárea según {var}')
        plt.xticks(rotation=45)
        plt.tight_layout()
        plt.show()
    else:
        print(f"La columna '{var}' no se encuentra en el DataFrame.")
```

| | Nº Contrato | Fecha plantacion | Lat | Lon \ |
|-----|-------------|------------------|------------|------------|
| 0 | 6000000118 | 11/19/2024 | -35.495617 | -71.607243 |
| 1 | 6000000092 | 10/14/2024 | -33.749477 | -70.705993 |
| 2 | 6000000230 | 10/30/2024 | -35.811234 | -71.487128 |
| 3 | 6000000193 | 10/4/2024 | -35.133026 | -71.286488 |
| 4 | 6000000127 | 10/11/2024 | -34.303551 | -70.764888 |
| .. | ... | ... | ... | ... |
| 311 | 6000000025 | 11/20/2024 | -34.053469 | -70.757050 |
| 312 | 6000000224 | 11/14/2025 | -35.487041 | -71.650732 |
| 313 | 6000000066 | 10/4/2024 | -33.804619 | -70.892049 |
| 314 | 6000000180 | 12/7/2024 | -34.270117 | -71.368606 |
| 315 | 6000000370 | 12/4/2024 | -36.080216 | -71.968250 |

Productor Zona \

| | | |
|-----|-------------------------------------|--------|
| 0 | AGRICOLA CORDILLERA SPA | Sur |
| 1 | MAURICIO A. ROJAS GOMEZ AGRI.ARREND | Norte |
| 2 | CARLOS ZENOBIO VERGARA IBAÑEZ | Sur |
| 3 | GERALDINE EMA GÓMEZ VALENZUELA | Centro |
| 4 | AGRICOLA MANUELITA SPA | Norte |
| .. | ... | ... |
| 311 | CRISTIAN A. ZAVALLA HURTADO E.I.R.L | Norte |
| 312 | AGRICOLA BENSANREN SPA | Sur |
| 313 | SOCIEDAD AGRICOLA BRISAS DEL CAMPO | Norte |
| 314 | SOCIEDAD AGROTURISTICA CABAÑA BLAN | Centro |
| 315 | SUGAL CHILE LIMITADA | Sur |

| | Predio | Tecnico \ |
|-----|------------------------------------|----------------------------|
| 0 | PARCELA N° 15 P.P. PORVENIR LOTE B | Luis Fernando Reyes |
| 1 | LOTE N° 1, CASA PATRONAL LINDEROS | RENE ORLANDO TERCILLA PE A |
| 2 | PREDIO SANTA VICTORIA | Jose Daniel Castillo |
| 3 | PARCELA N°13 P.P EL CONDOR | Marcos Daniel Suazo |
| 4 | PARCELA N° 16, PP. LOURDES | RENE ORLANDO TERCILLA PE A |
| .. | ... | ... |
| 311 | PARCELA N° 43, PP. LA SOLEDAD | Ana Carolina Rojas |
| 312 | LOTE N°5 PREDIO SANTA INÉS | Braulio Alejandro Ramirez |
| 313 | PREDIO LO CHACON | Ana Carolina Rojas |
| 314 | HIJUELA N° 1, RESTO FUNDO SAN LUIS | Felipe Arturo Farias |
| 315 | PREDIO CUYUMILLACO DE PARRAL | Joao Dos Santos |

| | Tipo riego | Has | Mensuradas | Kilos Netos | % Total defectos \ |
|-----|------------|------|------------|-------------|--------------------|
| 0 | ZRGF | 12.3 | 2521450 | 8% | |
| 1 | ZRGF | 9.2 | 1796820 | 6% | |
| 2 | ZRGF | 13.6 | 2703260 | 12% | |
| 3 | ZRGF | 3.6 | 676360 | 8% | |
| 4 | ZRGF | 22.6 | 4258900 | 9% | |
| .. | ... | ... | ... | ... | |
| 311 | ZNOF | 11.3 | 324050 | 20% | |
| 312 | ZRGF | 9.1 | 124330 | 10% | |
| 313 | ZRGF | 16.0 | 66900 | 9% | |
| 314 | ZRGF | 12.5 | 66500 | 20% | |
| 315 | ZRGF | 49.2 | 221660 | 19% | |

| | Rendimiento x ha | UNIDAD DE NEGOCIO | Presencia Orobanche | Nivel de Presión \ |
|-----|------------------|-------------------|---------------------|--------------------|
| 0 | 188.3 | AGRICULTORES | No | 1.Nula |
| 1 | 184.4 | AGRICULTORES | No | 1.Nula |
| 2 | 175.5 | AGRICULTORES | No | 1.Nula |
| 3 | 172.4 | AGRICULTORES | No | 1.Nula |
| 4 | 171.7 | AGRICULTORES | No | 1.Nula |
| .. | ... | ... | ... | ... |
| 311 | 23.0 | AGRICULTORES | No | 1.Nula |
| 312 | 18.7 | AGRICULTORES | Si | 2.Baja |
| 313 | 12.2 | AGRICULTORES | No | 1.Nula |

| | | | | |
|-----|-----|----------------|----|--------|
| 314 | 4.7 | AGRICULTORES | Si | 4.Alta |
| 315 | 4.5 | CAMPOS PROPIOS | No | 1.Nula |

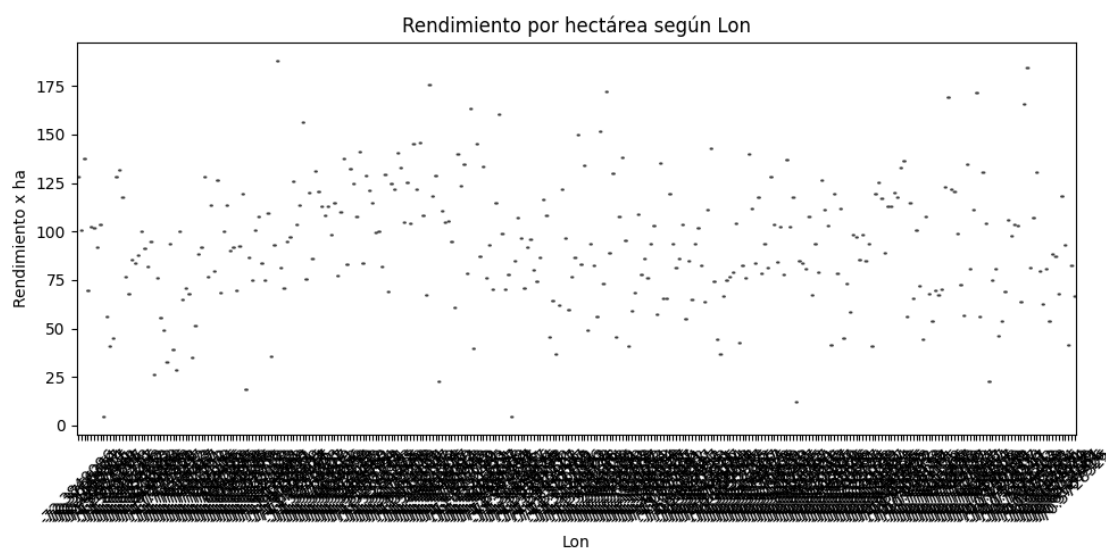
```

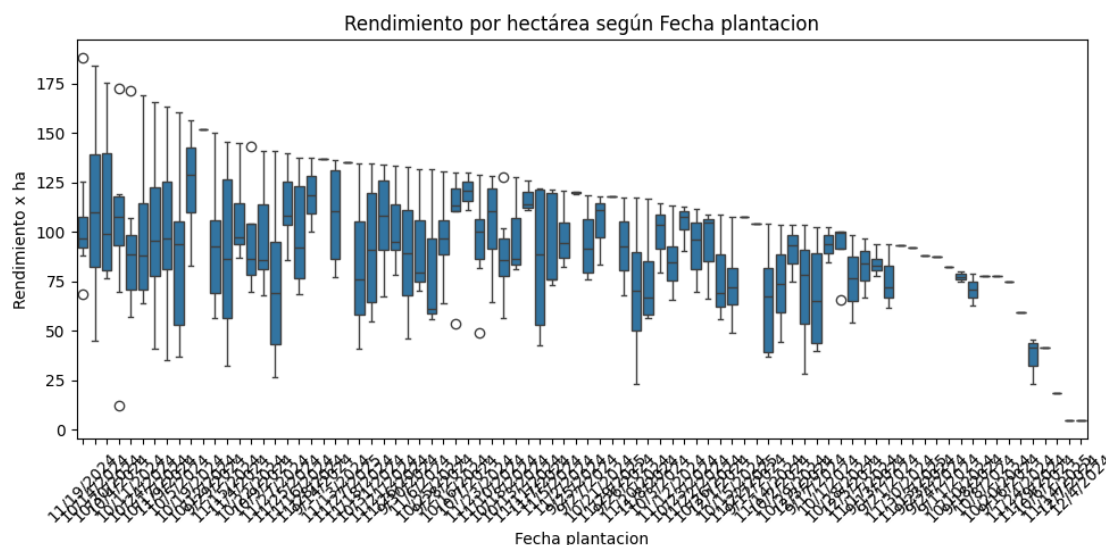
Obj rendimiento
0          106
1          106
2          106
3          102
4          106
..         ...
311         90
312        106
313        106
314        102
315         99

```

[316 rows x 17 columns]

La columna 'Lat' no se encuentra en el DataFrame.





Al observar el gráfico de izquierda a derecha, que presumiblemente representa fechas de plantación más recientes a más antiguas (dada la disposición), podemos notar una tendencia general descendente en el rendimiento por hectárea a medida que las fechas de plantación se hacen más recientes.

Fechas de Plantación más Antiguas (lado izquierdo del gráfico): Las cajas tienden a estar ubicadas en niveles más altos del eje vertical, lo que sugiere que las plantaciones realizadas en estas fechas históricamente han tenido rendimientos por hectárea más altos en promedio. También se observa una mayor variabilidad en algunos de estos grupos, con cajas más grandes y bigotes más largos. Fechas de Plantación más Recientes (lado derecho del gráfico): Las cajas se desplazan hacia niveles más bajos del eje vertical, indicando que las plantaciones más recientes tienden a tener rendimientos por hectárea más bajos. La dispersión de los datos también parece disminuir en algunas de las fechas de plantación más recientes, con cajas y bigotes más cortos. Puntos Específicos a Considerar:

Medianas: La línea dentro de cada caja (la mediana) muestra una clara disminución general a medida que avanzamos de fechas de plantación más antiguas a más recientes. Esto refuerza la tendencia descendente observada. Dispersión (Tamaño de las Cajas y Bigotes): La variabilidad en el rendimiento por hectárea parece ser mayor en algunas de las fechas de plantación más antiguas, lo que podría indicar una mayor influencia de otros factores (variedad de la planta, condiciones climáticas de diferentes años, prácticas de manejo, etc.) en esas plantaciones. En las fechas más recientes, la variabilidad parece reducirse en algunos casos. Valores Atípicos: Se observan valores atípicos tanto por encima como por debajo de las cajas en varias fechas de plantación. Estos representan casos particulares con rendimientos significativamente diferentes al resto de las plantaciones realizadas en esas fechas. Es importante investigar estos valores atípicos para entender sus posibles causas. Brechas: Hay algunas fechas de plantación para las que la caja parece estar muy comprimida o los bigotes son muy cortos, lo que sugiere una menor variabilidad en el rendimiento para esas plantaciones específicas.

La fecha de plantación influye significativamente en el rendimiento, y las plantaciones más tempranas suelen rendir más. Esto puede deberse a mejores condiciones iniciales del cultivo, mejor aprovechamiento del ciclo climático, o menos presión de plagas/enfermedades.

#Hay alguna relacion entre el nivel de presio y el rendimiento?

Sí, hay una relación inversa. Del gráfico, se observa que: A medida que el nivel de presión aumenta (de 1: Nula a 4: Alta), el rendimiento tiende a disminuir. La mediana del rendimiento baja progresivamente con cada aumento en el nivel de presión. También parece haber una mayor variabilidad en niveles bajos de presión y una menor variabilidad (rendimientos más consistentemente bajos) en niveles de presión alta.

..... A mayor nivel de presión, menor rendimiento por hectárea. Esto sugiere que el nivel de presión (posiblemente de plaga, enfermedad o maleza) tiene un efecto negativo sobre el rendimiento del cultivo.

#HaY alguna relacion entre la fecha de plantacion y el rendimiento?

Tendencia descendente: A medida que las fechas de plantación se hacen más recientes (de izquierda a derecha en el gráfico), el rendimiento por hectárea tiende a disminuir.

Medianas en descenso: Las medianas dentro de cada grupo de fecha de plantación se desplazan hacia valores más bajos a medida que la fecha de plantación se hace más reciente.

Disminución de la dispersión: En las fechas de plantación más recientes, se observa menos variabilidad en el rendimiento (cajas más pequeñas y bigotes más cortos).

Valores atípicos: Hay valores atípicos que pueden ser interesantes para investigar, ya que algunos puntos tienen rendimientos significativamente diferentes de los demás

..... Las plantaciones más antiguas tienen rendimientos más altos, y a medida que las plantaciones se hacen más recientes, el rendimiento tiende a ser más bajo y con menos variabilidad.

Untitled1

May 3, 2025

```
[1]: import pandas as pd
data = pd.read_csv('C:/Users/brait/OneDrive/Escritorio/Proyecto 2/paulo/df.
↳csv', encoding='latin1')
print(data)
```

| | Nº Contrato | Fecha plantacion | Lat | Lon \ |
|-----|-------------|------------------|------------|------------|
| 0 | 6000000118 | 11/19/2024 | -35.495617 | -71.607243 |
| 1 | 6000000092 | 10/14/2024 | -33.749477 | -70.705993 |
| 2 | 6000000230 | 10/30/2024 | -35.811234 | -71.487128 |
| 3 | 6000000193 | 10/4/2024 | -35.133026 | -71.286488 |
| 4 | 6000000127 | 10/11/2024 | -34.303551 | -70.764888 |
| .. | ... | ... | ... | ... |
| 311 | 6000000025 | 11/20/2024 | -34.053469 | -70.757050 |
| 312 | 6000000224 | 11/14/2025 | -35.487041 | -71.650732 |
| 313 | 6000000066 | 10/4/2024 | -33.804619 | -70.892049 |
| 314 | 6000000180 | 12/7/2024 | -34.270117 | -71.368606 |
| 315 | 6000000370 | 12/4/2024 | -36.080216 | -71.968250 |

| | Productor | Zona \ |
|-----|-------------------------------------|--------|
| 0 | AGRICOLA CORDILLERA SPA | Sur |
| 1 | MAURICIO A. ROJAS GOMEZ AGRI.ARREND | Norte |
| 2 | CARLOS ZENOBIO VERGARA IBAÑEZ | Sur |
| 3 | GERALDINE EMA GÓMEZ VALENZUELA | Centro |
| 4 | AGRICOLA MANUELITA SPA | Norte |
| .. | ... | ... |
| 311 | CRISTIAN A. ZAVALLA HURTADO E.I.R.L | Norte |
| 312 | AGRICOLA BENSANREN SPA | Sur |
| 313 | SOCIEDAD AGRICOLA BRISAS DEL CAMPO | Norte |
| 314 | SOCIEDAD AGROTURISTICA CABAÑA BLAN | Centro |
| 315 | SUGAL CHILE LIMITADA | Sur |

| | Predio | Tecnico \ |
|----|------------------------------------|----------------------------|
| 0 | PARCELA Nº 15 P.P. PORVENIR LOTE B | Luis Fernando Reyes |
| 1 | LOTE Nº 1, CASA PATRONAL LINDEROS | RENE ORLANDO TERCILLA PE A |
| 2 | PREDIO SANTA VICTORIA | Jose Daniel Castillo |
| 3 | PARCELA Nº13 P.P EL CONDOR | Marcos Daniel Suazo |
| 4 | PARCELA Nº 16, PP. LOURDES | RENE ORLANDO TERCILLA PE A |
| .. | ... | ... |

| | | |
|-----|------------------------------------|---------------------------|
| 311 | PARCELA N° 43, PP. LA SOLEDAD | Ana Carolina Rojas |
| 312 | LOTE N°5 PREDIO SANTA INÉS | Braulio Alejandro Ramirez |
| 313 | PREDIO LO CHACON | Ana Carolina Rojas |
| 314 | HIJUELA N° 1, RESTO FUNDO SAN LUIS | Felipe Arturo Farias |
| 315 | PREDIO CUYUMILLACO DE PARRAL | Joao Dos Santos |

| | Tipo riego | Has Mensuradas | Kilos Netos | % Total defectos | \ |
|-----|------------|----------------|-------------|------------------|---|
| 0 | ZRGF | 12.3 | 2521450 | 8% | |
| 1 | ZRGF | 9.2 | 1796820 | 6% | |
| 2 | ZRGF | 13.6 | 2703260 | 12% | |
| 3 | ZRGF | 3.6 | 676360 | 8% | |
| 4 | ZRGF | 22.6 | 4258900 | 9% | |
| .. | ... | ... | ... | ... | |
| 311 | ZNOF | 11.3 | 324050 | 20% | |
| 312 | ZRGF | 9.1 | 124330 | 10% | |
| 313 | ZRGF | 16.0 | 66900 | 9% | |
| 314 | ZRGF | 12.5 | 66500 | 20% | |
| 315 | ZRGF | 49.2 | 221660 | 19% | |

| | Rendimiento x ha | UNIDAD DE NEGOCIO | Presencia Orobanche | Nivel de Presión | \ |
|-----|------------------|-------------------|---------------------|------------------|---|
| 0 | 188.3 | AGRICULTORES | No | 1.Nula | |
| 1 | 184.4 | AGRICULTORES | No | 1.Nula | |
| 2 | 175.5 | AGRICULTORES | No | 1.Nula | |
| 3 | 172.4 | AGRICULTORES | No | 1.Nula | |
| 4 | 171.7 | AGRICULTORES | No | 1.Nula | |
| .. | ... | ... | ... | ... | |
| 311 | 23.0 | AGRICULTORES | No | 1.Nula | |
| 312 | 18.7 | AGRICULTORES | Si | 2.Baja | |
| 313 | 12.2 | AGRICULTORES | No | 1.Nula | |
| 314 | 4.7 | AGRICULTORES | Si | 4.Alta | |
| 315 | 4.5 | CAMPOS PROPIOS | No | 1.Nula | |

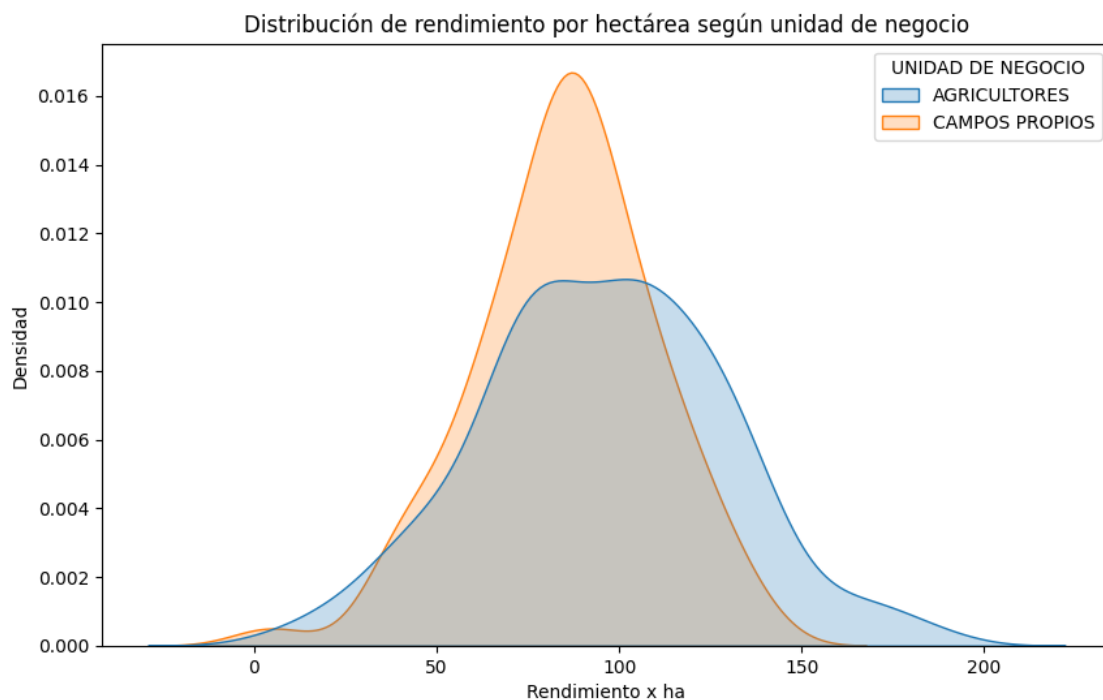
| | Obj rendimiento |
|-----|-----------------|
| 0 | 106 |
| 1 | 106 |
| 2 | 106 |
| 3 | 102 |
| 4 | 106 |
| .. | ... |
| 311 | 90 |
| 312 | 106 |
| 313 | 106 |
| 314 | 102 |
| 315 | 99 |

[316 rows x 17 columns]

```
[12]: #Como es la distribucion del rendimiento entre agricultores y campos propios?
```

```
import seaborn as sns
import matplotlib.pyplot as plt

plt.figure(figsize=(10,6))
sns.kdeplot(data=data, x='Rendimiento x ha', hue='UNIDAD DE NEGOCIO',
            common_norm=False, fill=True)
plt.title('Distribución de rendimiento por hectárea según unidad de negocio')
plt.xlabel('Rendimiento x ha')
plt.ylabel('Densidad')
plt.show()
```



1. ¿Tiene forma de campana? La distribución del rendimiento por hectárea en la unidad de negocio “AGRICULTORES” se asemeja a una curva de campana, aunque presenta una forma ligeramente aplanada. Por otro lado, “CAMPOS PROPIOS” también muestra una distribución unimodal, pero con una forma más estrecha y puntiaguda, menos simétrica que la anterior.
2. ¿Está sesgada? En “AGRICULTORES”, la distribución presenta un leve sesgo positivo (hacia la derecha), evidenciado por una cola más extendida en ese extremo. En contraste, la distribución de “CAMPOS PROPIOS” muestra un sesgo negativo (hacia la izquierda), con una cola más prolongada hacia valores más bajos de rendimiento.
3. ¿Dónde se concentran los datos? Para “AGRICULTORES”, la mayor densidad de observaciones se encuentra entre 80 y 120 unidades de rendimiento por hectárea, con un pico central bien definido. En cambio, “CAMPOS PROPIOS” concentra sus datos en un rango más bajo,

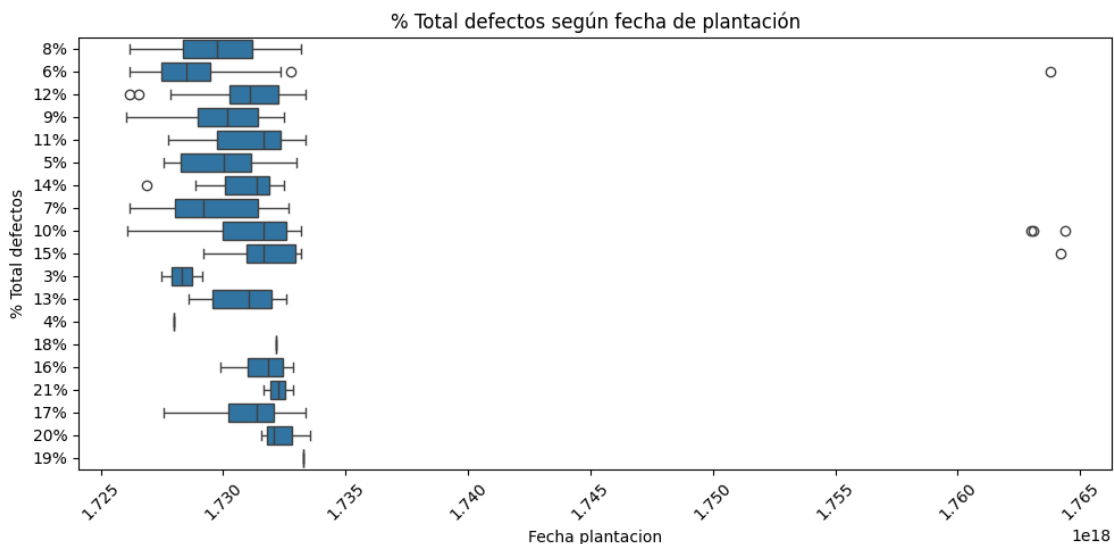
aproximadamente entre 60 y 100 unidades, con un pico más agudo alrededor de los 80.

4. ¿Se distribuyen diferente entre unidades? Sí, existe una diferencia clara entre las distribuciones. “CAMPOS PROPIOS” tiende a concentrar sus rendimientos en valores ligeramente más bajos y en un rango más estrecho, mientras que “AGRICULTORES” muestra una distribución más dispersa, con tendencia a rendimientos ligeramente superiores en promedio. Estas diferencias se reflejan tanto en la forma (más achatada vs. más concentrada) como en el sesgo de cada curva, lo que sugiere comportamientos distintos en el rendimiento según la unidad de negocio.

[14]: *#Hay alguna relacion entre el porcentaje defecto y fecha de plantacion?*

```
import seaborn as sns
import matplotlib.pyplot as plt

plt.figure(figsize=(10, 5))
sns.boxplot(x='Fecha plantacion', y='% Total defectos', data=data)
plt.title('% Total defectos según fecha de plantación')
plt.xticks(rotation=45)
plt.tight_layout()
plt.savefig("defectos_por_fecha.png", dpi=300, bbox_inches='tight')
plt.show()
```

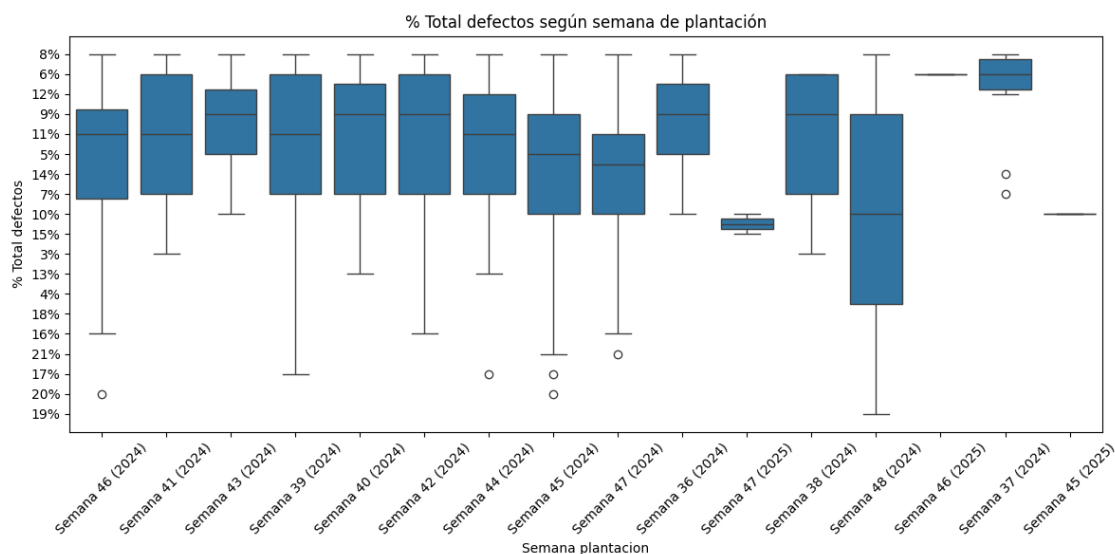


```
[15]: import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# Asegúrate de convertir correctamente la columna a datetime
data['Fecha plantacion'] = pd.to_datetime(data['Fecha plantacion'],
errors='coerce')
```

```
# Extraer la semana del año como nueva columna (por año para mayor claridad)
data['Semana plantacion'] = data['Fecha plantacion'].dt.strftime('Semana %U_
↳ (%Y)')

# Ahora graficamos usando esa columna
plt.figure(figsize=(12, 6))
sns.boxplot(x='Semana plantacion', y='% Total defectos', data=data)
plt.title('% Total defectos según semana de plantación')
plt.xticks(rotation=45)
plt.tight_layout()
plt.savefig("defectos_por_semana.png", dpi=300, bbox_inches='tight')
plt.show()
```



- Resumen del gráfico: Eje X (Semana de plantación): Agrupadas por semana y año. Algunas semanas se repiten en distintos años (por ejemplo, Semana 46 en 2024 y 2025). Eje Y (% Total de defectos): Va aproximadamente de 0% a 20%. Cajas (boxplots): Representan la variabilidad del porcentaje de defectos para cada semana; la línea dentro es la mediana. Puntos fuera de los bigotes: Son valores atípicos.
- Observaciones clave:
 - Variabilidad por semana: La dispersión del % de defectos varía considerablemente entre semanas. Algunas presentan una concentración alta (cajas más pequeñas), mientras que otras tienen mayor variabilidad (cajas más amplias).
 - Comparación de la misma semana en distintos años: Al observar semanas específicas repetidas en ambos años:

Semana 46: En 2024 la mediana se sitúa en torno al 9–10%. En 2025 sube considerablemente, cerca del 16%.

Semana 47: Ambas presentan medianas altas, pero 2025 mantiene el nivel elevado de la semana anterior.

Semana 36: En 2025 se observa un nivel más bajo de defectos en comparación con 2024.

Semana 37: Muestra un comportamiento inverso: mayor mediana en 2024 que en 2025.

- Tendencia temporal dentro del mismo año: No se aprecia una tendencia clara de aumento o disminución del % de defectos a medida que avanzan las semanas dentro de un año. Por ejemplo, en 2024, semanas consecutivas pueden tener niveles de defectos tanto altos como bajos.
- 3. Posible patrón observado: El año de plantación parece influir más en el % total de defectos que la semana específica dentro del año.
- En 2025, las semanas 46 y 47 presentan consistentemente porcentajes de defectos más altos que en 2024.
- Sin embargo, para otras semanas (como la 36 y 37), la tendencia no es consistente, lo que sugiere que otros factores podrían estar interviniendo.
- No hay una tendencia clara por semana, pero sí diferencias importantes entre años, lo que sugiere que factores asociados al año (clima, prácticas agrícolas, calidad de insumos, etc.) podrían estar afectando el nivel de defectos.

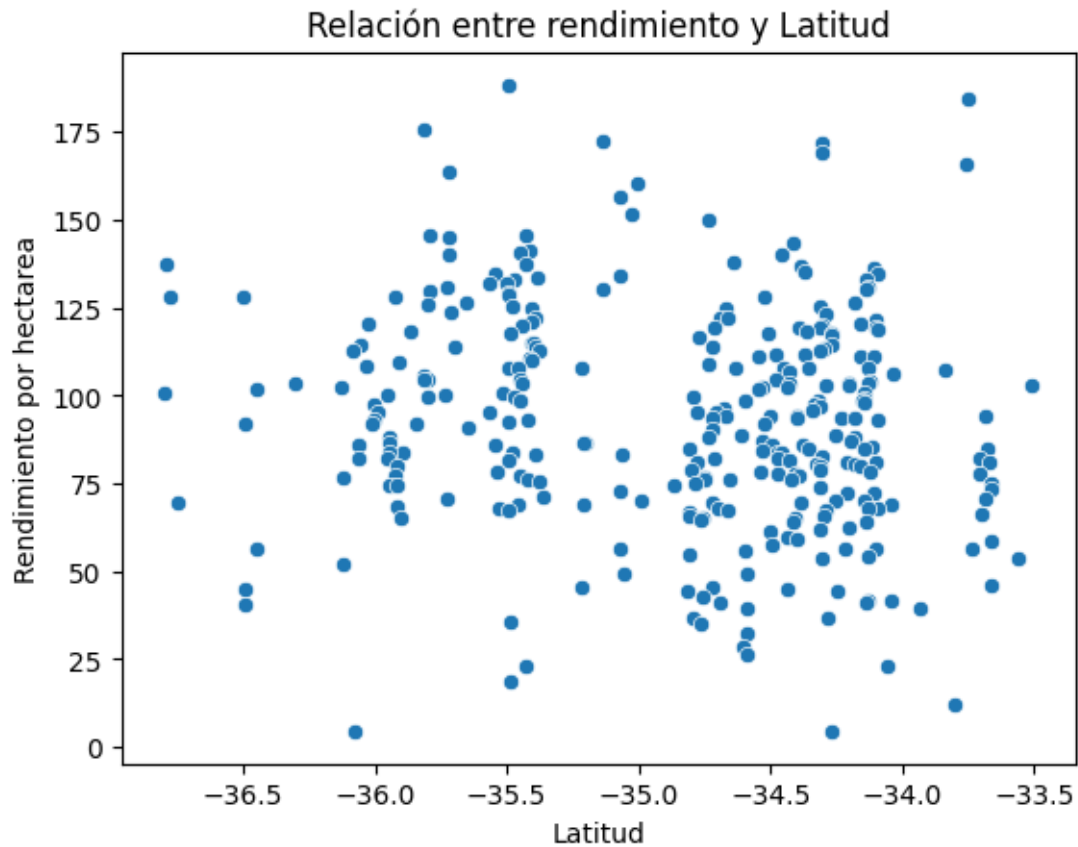
[28]: #9. Hay alguna relacion entre Lat y Lon y los rendimientos?

```
import seaborn as sns
import pandas as pd
import matplotlib.pyplot as plt

# Crear un DataFrame de ejemplo
df = data

# Crear el scatter plot
sns.scatterplot(data=data, x='Lat ', y='Rendimiento x ha')

# Título y etiquetas
plt.title('Relación entre rendimiento y Latitud')
plt.xlabel('Latitud')
plt.ylabel('Rendimiento por hectarea')
plt.savefig("rendimiento_ha_latitud.png", dpi=300, bbox_inches='tight')
# Mostrar la gráfica
plt.show()
```

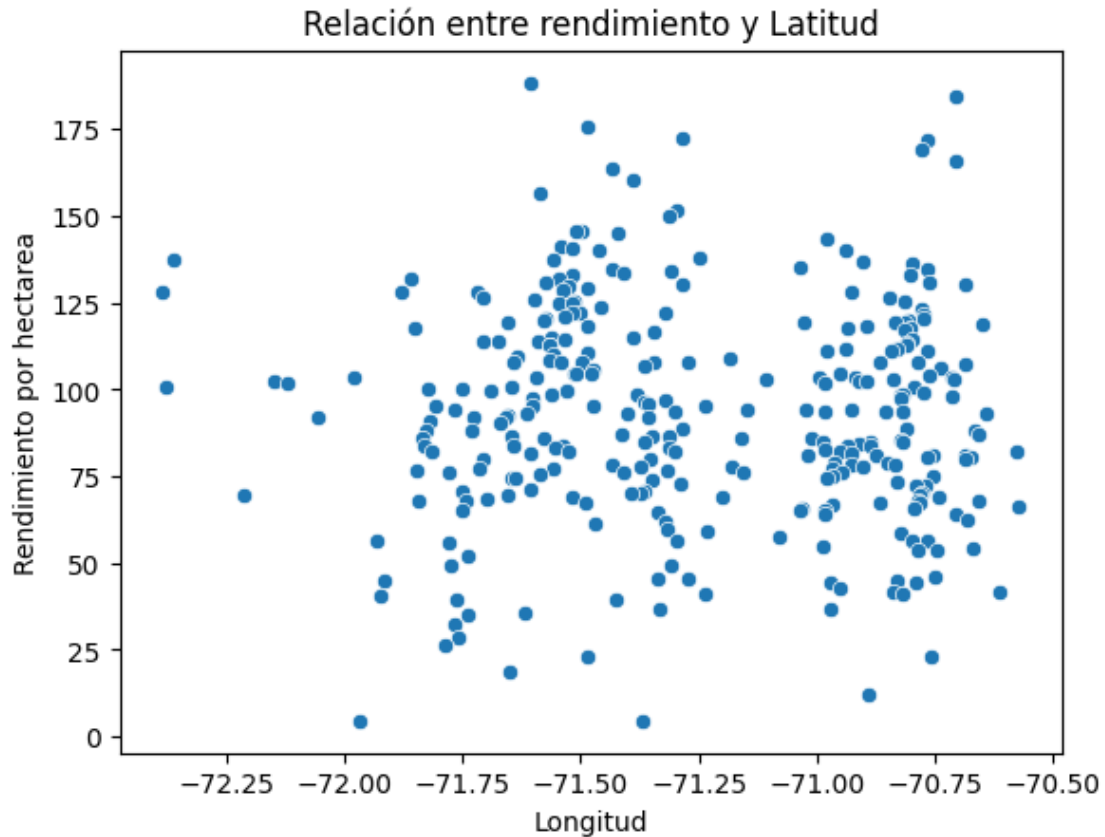


```
[31]: import seaborn as sns
import pandas as pd
import matplotlib.pyplot as plt

# Crear un DataFrame de ejemplo
df = data

# Crear el scatter plot
sns.scatterplot(data=data, x='Lon', y='Rendimiento x ha')

# Título y etiquetas
plt.title('Relación entre rendimiento y Latitud')
plt.xlabel('Longitud')
plt.ylabel('Rendimiento por hectarea')
plt.savefig("rendimiento_ha_Longitud.png", dpi=300, bbox_inches='tight')
# Mostrar la gráfica
plt.show()
```



Observaciones visuales: 1. Rendimiento vs Latitud: No hay una tendencia clara y continua (ni ascendente ni descendente). El rendimiento parece estar agrupado en zonas específicas de latitud, con varios “clusters”. Eso sugiere que hay zonas geográficas donde los rendimientos tienden a ser similares, pero no es una relación lineal clara.

2. Rendimiento vs Longitud: Similar al gráfico anterior: se observan agrupaciones, pero sin una tendencia lineal fuerte. Hay dispersión dentro de cada grupo, lo que sugiere que hay otros factores además de la ubicación que influyen en el rendimiento.

```
[37]: # Correlación de Pearson
correlacion_pearson = data[['Lat ', 'Lon', 'Rendimiento x ha']].
    ↪corr(method='pearson')
print("Correlación de Pearson:")
print(correlacion_pearson)

# Correlación de Spearman
correlacion_spearman = data[['Lat ', 'Lon', 'Rendimiento x ha']].
    ↪corr(method='spearman')
print("\nCorrelación de Spearman:")
print(correlacion_spearman)
```

Correlación de Pearson:

| | Lat | Lon | Rendimiento x ha |
|------------------|-----------|-----------|------------------|
| Lat | 1.000000 | 0.833599 | -0.147486 |
| Lon | 0.833599 | 1.000000 | -0.025964 |
| Rendimiento x ha | -0.147486 | -0.025964 | 1.000000 |

Correlación de Spearman:

| | Lat | Lon | Rendimiento x ha |
|------------------|-----------|-----------|------------------|
| Lat | 1.000000 | 0.844097 | -0.150588 |
| Lon | 0.844097 | 1.000000 | -0.038950 |
| Rendimiento x ha | -0.150588 | -0.038950 | 1.000000 |

Conclusión: No hay una relación fuerte entre latitud o longitud con el rendimiento por hectárea. Si hay alguna relación, la latitud parece tener un ligerísimo efecto negativo: al subir la latitud, podría bajar un poco el rendimiento. Pero estas correlaciones son tan bajas que no permiten hacer predicciones confiables sin considerar otras variables (tipo de suelo, clima, prácticas agrícolas, etc.).

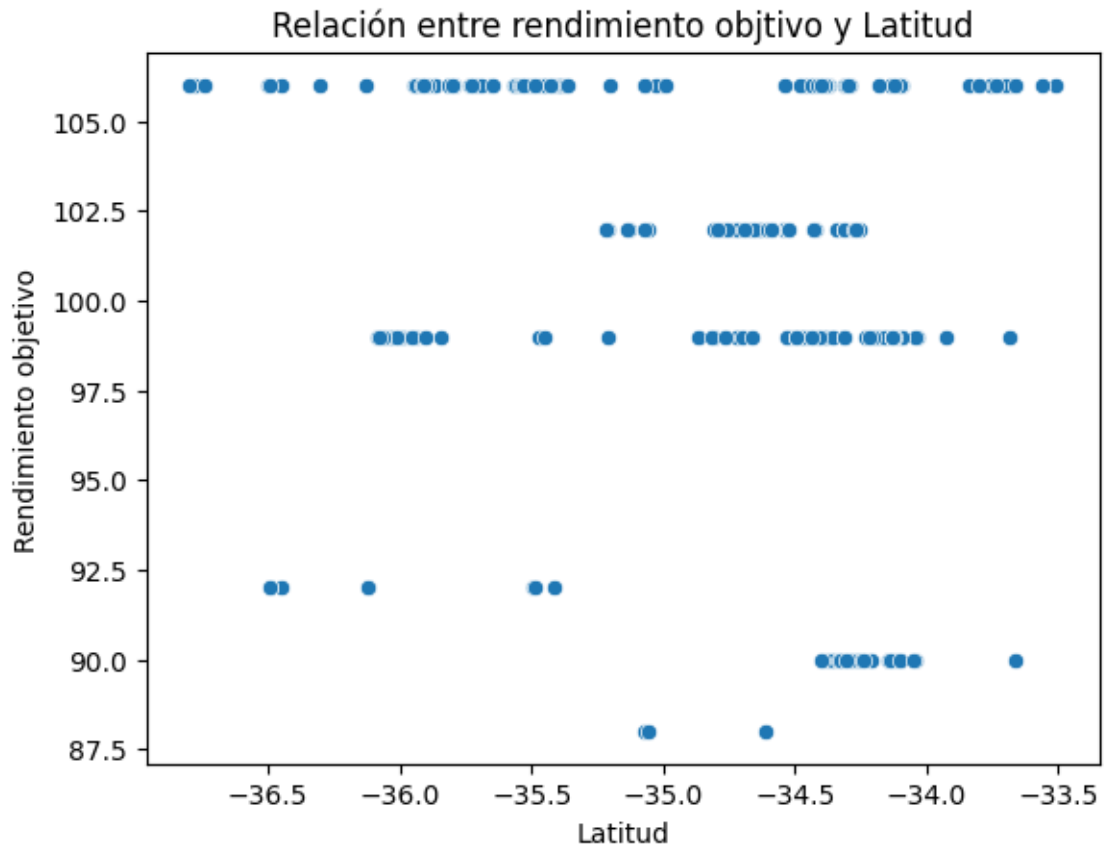
```
[29]: #rendimientos objetivo y latitud y longitud:

import seaborn as sns
import pandas as pd
import matplotlib.pyplot as plt

# Crear un DataFrame de ejemplo
df = data

# Crear el scatter plot
sns.scatterplot(data=data, x='Lat ', y='Obj rendimiento')

# Título y etiquetas
plt.title('Relación entre rendimiento objetivo y Latitud')
plt.xlabel('Latitud')
plt.ylabel('Rendimiento objetivo')
plt.savefig("rendimiento_latitud.png", dpi=300, bbox_inches='tight')
# Mostrar la gráfica
plt.show()
```

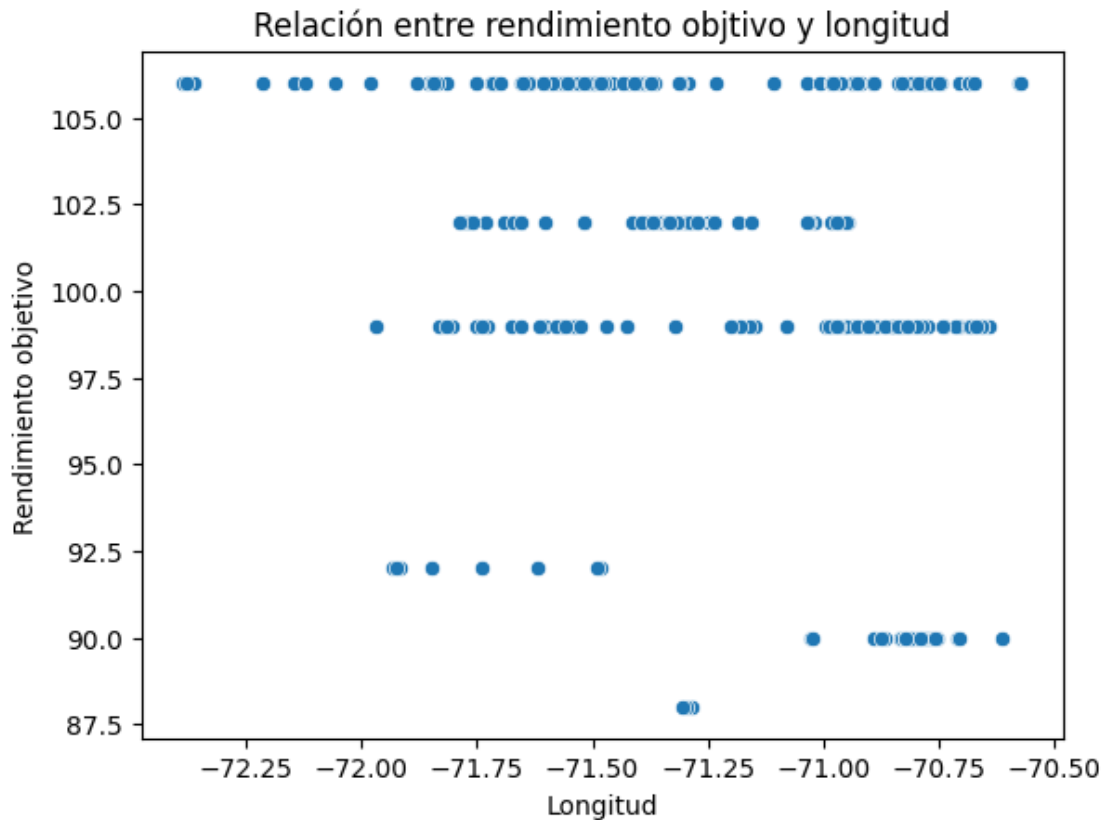


```
[32]: import seaborn as sns
import pandas as pd
import matplotlib.pyplot as plt

# Crear un DataFrame de ejemplo
df = data

# Crear el scatter plot
sns.scatterplot(data=data, x='Lon', y='Obj rendimiento')

# Título y etiquetas
plt.title('Relación entre rendimiento objetivo y longitud')
plt.xlabel('Longitud')
plt.ylabel('Rendimiento objetivo')
plt.savefig("rendimiento_longitud.png", dpi=300, bbox_inches='tight')
# Mostrar la gráfica
plt.show()
```



1. Gráfico 1: Rendimiento objetivo vs Longitud Eje X: Longitud (valores entre -72.5 y -70.5 aproximadamente). Eje Y: Rendimiento objetivo (entre ~88 y 106).

2. Observaciones:

No se aprecia una tendencia clara (ni creciente ni decreciente). Los puntos están bastante dispersos a lo largo del eje X. El rendimiento objetivo parece mantenerse dentro de un rango similar para distintas longitudes. Posible agrupamiento de datos alrededor de ciertos valores de longitud, lo cual podría sugerir zonas geográficas específicas, pero no implica causalidad directa.

[]: 2. Gráfico 2: Rendimiento objetivo vs Latitud
Eje X: Latitud (entre -36.5 y -33.5 aprox).
Eje Y: Igual al gráfico anterior.

2. Observaciones:

Al igual que el anterior, los datos están dispersos sin una tendencia clara. Se observan concentraciones en ciertas latitudes, pero el rendimiento varía dentro de esas concentraciones.

No hay un patrón obvio que indique que a mayor o menor latitud haya un rendimiento significativamente distinto.

No se observa una relación clara entre: Latitud y rendimiento objetivo Longitud y rendimiento objetivo

¿Por qué? Los puntos están distribuidos de forma dispersa sin una tendencia evidente (ni creciente ni decreciente). No se forma un patrón lineal, curvo ni concentraciones que sugieran correlación fuerte. La variabilidad del rendimiento objetivo parece similar a lo largo de todas las latitudes y longitudes.

```
[34]: #aunque visualmente no hay una relacion clara vamos a calcular el coeficiente_
      ↪ de Pearson y Spearman
```

```
# Pearson: para relaciones lineales
pearson_corr = data[['Lat ', 'Lon', 'Obj rendimiento']].corr(method='pearson')
print("Correlación de Pearson:")
print(pearson_corr)
```

Correlación de Pearson:

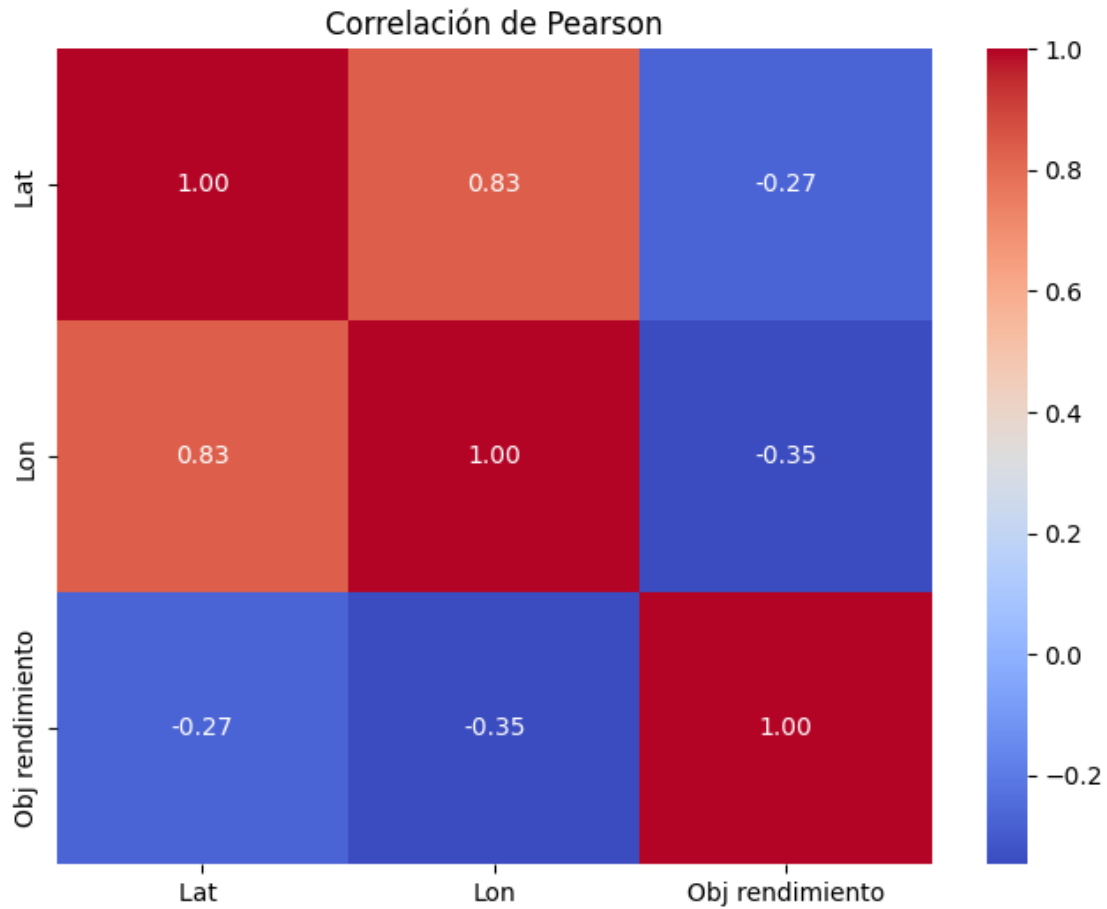
| | Lat | Lon | Obj rendimiento |
|-----------------|-----------|-----------|-----------------|
| Lat | 1.000000 | 0.833599 | -0.272624 |
| Lon | 0.833599 | 1.000000 | -0.347207 |
| Obj rendimiento | -0.272624 | -0.347207 | 1.000000 |

```
[35]: # Spearman: útil para relaciones monótonas (no necesariamente lineales)
spearman_corr = data[['Lat ', 'Lon', 'Obj rendimiento']].corr(method='spearman')
print("\nCorrelación de Spearman:")
print(spearman_corr)
```

Correlación de Spearman:

| | Lat | Lon | Obj rendimiento |
|-----------------|-----------|-----------|-----------------|
| Lat | 1.000000 | 0.844097 | -0.302803 |
| Lon | 0.844097 | 1.000000 | -0.332676 |
| Obj rendimiento | -0.302803 | -0.332676 | 1.000000 |

```
[36]: plt.figure(figsize=(8,6))
sns.heatmap(pearson_corr, annot=True, cmap="coolwarm", fmt=".2f")
plt.title("Correlación de Pearson")
plt.show()
```



1. ¿Hay relación? Sí, pero es débil y negativa.

- Los valores están entre -0.27 y -0.35, lo cual indica:
- Una ligera tendencia a que el rendimiento disminuya cuando la latitud o longitud aumentan.
- Pero no es una relación fuerte (como lo sería un valor cercano a -0.7 o menor).
- La correlación Spearman reafirma esto: hay una pequeña tendencia monótona negativa, pero sigue siendo débil.