Mediterranean Fruit Fly (Ceratitis capitata (Wied))

1. DATOS:

https://www.sag.gob.cl/ambitos-de-accion/mosca-de-la-fruta/publicaciones?field_tema_otros_documentos_target_id=2749&field_tipo_de_publicacion_target_id=244&field_fecha_otros_value=&title=&order=field_fecha_otros&sort=desc&page=0

2. DE PDF A TABLA DE DATOS.

```
!pip install pypdf
!pip install camelot-py
!pip install pyproj
```

```
Collecting pypdf
  Downloading pypdf-5.4.0-py3-none-any.whl.metadata (7.3 kB)
Downloading pypdf-5.4.0-py3-none-any.whl (302 kB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 302.3/302.3 kB 5.0 MB/s eta 0:00:00
Installing collected packages: pypdf
Successfully installed pypdf-5.4.0
Collecting camelot-py
  Downloading camelot_py-1.0.0-py3-none-any.whl.metadata (9.4 kB)
Requirement already satisfied: click>=8.0.1 in /usr/local/lib/python3.11/dist-packages (from camelot-py) (8.1.8)
Requirement already satisfied: chardet>=5.1.0 in /usr/local/lib/python3.11/dist-packages (from camelot-py) (5.2.0)
Requirement already satisfied: numpy>=1.24.4 in /usr/local/lib/python3.11/dist-packages (from camelot-py) (2.0.2)
Requirement already satisfied: openpyxl>=3.1.0 in /usr/local/lib/python3.11/dist-packages (from camelot-py) (3.1.5)
Collecting pdfminer-six>=20240706 (from camelot-py)
  Downloading pdfminer_six-20250416-py3-none-any.whl.metadata (4.1 kB)
Collecting pypdf<4.0,>=3.17 (from camelot-py)
  Downloading pypdf-3.17.4-py3-none-any.whl.metadata (7.5 kB)
Requirement already satisfied: pandas>=2.2.2 in /usr/local/lib/python3.11/dist-packages (from camelot-py) (2.2.2)
Requirement already satisfied: tabulate>=0.9.0 in /usr/local/lib/python3.11/dist-packages (from camelot-py) (0.9.0)
Requirement already satisfied: opencv-python-headless>=4.7.0.68 in /usr/local/lib/python3.11/dist-packages (from camelot-py) (4.11.0.86)
Collecting pypdfium2>=4 (from camelot-py)
  Downloading pypdfium2-4.30.1-py3-none-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (48 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 48.2/48.2 kB 1.4 MB/s eta 0:00:00
Requirement already satisfied: et-xmlfile in /usr/local/lib/python3.11/dist-packages (from openpyxl>=3.1.0->camelot-py) (2.0.0)
Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.11/dist-packages (from pandas>=2.2.2->camelot-py) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.11/dist-packages (from pandas>=2.2.2->camelot-py) (2025.2)
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.11/dist-packages (from pandas>=2.2.2->camelot-py) (2025.2)
Requirement already satisfied: charset-normalizer>=2.0.0 in /usr/local/lib/python3.11/dist-packages (from pdfminer-six>=20240706->camelot-py) (3.4.1)
Requirement already satisfied: cryptography>=36.0.0 in /usr/local/lib/python3.11/dist-packages (from pdfminer-six>=20240706->camelot-py) (43.0.3)
Requirement already satisfied: cffi>=1.12 in /usr/local/lib/python3.11/dist-packages (from cryptography>=36.0.0->pdfminer-six>=20240706->camelot-py) (1.17.1)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.11/dist-packages (from python-dateutil>=2.8.2->pandas>=2.2.2->camelot-py) (1.17.0)
Requirement already satisfied: pycparser in /usr/local/lib/python3.11/dist-packages (from cffi>=1.12->cryptography>=36.0.0->pdfminer-six>=20240706->camelot-py) (2.22)
Downloading camelot_py-1.0.0-py3-none-any.whl (66 kB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 66.6/66.6 kB 2.5 MB/s eta 0:00:00
Downloading pdfminer_six-20250416-py3-none-any.whl (5.6 MB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 5.6/5.6 MB 45.5 MB/s eta 0:00:00
Downloading pypdf-3.17.4-py3-none-any.whl (278 kB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 278.2/278.2 kB 16.7 MB/s eta 0:00:00
Downloading pypdfium2-4.30.1-py3-none-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (2.9 MB)
```

```
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 2.9/2.9 MB 43.5 MB/s eta 0:00:00
Installing collected packages: pypdfium2, pypdf, pdfminer-six, camelot-py
  Attempting uninstall: pypdf
    Found existing installation: pypdf 5.4.0
    Uninstalling pypdf-5.4.0:
      Successfully uninstalled pypdf-5.4.0
Successfully installed camelot-py-1.0.0 pdfminer-six-20250416 pypdf-3.17.4 pypdfium2-4.30.1
Requirement already satisfied: pyproj in /usr/local/lib/python3.11/dist-packages (3.7.1)
Requirement already satisfied: certifi in /usr/local/lib/python3.11/dist-packages (from pyproj) (2025.1.31)
```

2.1 Datos en pdf.

Leemos, sacamos datos del pdf, corregimos coordenadas, seleccionamos datos de interés.

Graficamos en mapa

```
from pypdf import PdfReader
import camelot
import pandas as pd
from IPython.display import display
from pypdf import PdfWriter
from google.colab import files

# Crear el objeto para combinar PDFs
pdf_writer = PdfWriter()

# Subir archivos PDF manualmente
pdf_files = files.upload()  # Esto devuelve un diccionario con los nombres de los archivos


# Agregar cada PDF al nuevo documento
for pdf in pdf_files.keys():
    pdf_writer.append(pdf)

# Guardar el archivo final
pdf_writer.write("PDF_unido.pdf")
pdf_writer.close()

# Cargar el PDF
pdf_path = "PDF_unido.pdf"
reader = PdfReader(pdf_path)

# Obtener el número de páginas
num_paginas = len(reader.pages)
print(f"El PDF tiene {num_paginas} páginas.")

# Extraer todas las tablas del PDF
tables = camelot.read_pdf(pdf_path, pages="all")

# Contar cuántas tablas se detectaron
num_tablas = len(tables)
print(f"Se encontraron {num_tablas} tablas en el PDF.")

# Crear un DataFrame con todas las tablas
df_total = pd.concat([table.df for table in tables], ignore_index=True)
```

```
#llenan Nan con NaN
df_total.fillna("NaN", inplace=True)

df_limpio = df_total.dropna(subset=[12]) # seleccionar sin NaN columna latitud.

# Supongamos que tu DataFrame se llama df_total y la columna se llama 'mi_columna'
df_numeros = df_limpio[pd.to_numeric(df_limpio[12], errors='coerce').notna()]

display(df_numeros)
```

Mostrar salida oculta

```
from google.colab import drive
drive.mount('/content/drive')
```

2.2 Transformación de datos gps para ser leido por folium o kepler

```
import pandas as pd
import folium
from pyproj import Transformer

# Definir el sistema de referencia correcto (ajusta según la zona UTM de Chile)
transformer = Transformer.from_crs("EPSG:5361", "EPSG:4326", always_xy=True)  # UTM 19S a WGS84

# Suponiendo que df_numeros tiene las coordenadas en las columnas 12 y 13
df_numeros[['lon', 'lat']] = df_numeros.apply(lambda row: pd.Series(transformer.transform(row.iloc[13], row.iloc[12])), axis=1)

display(df_numeros)
```

```
<ipython-input-15-5b2cca56d263>:9: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df_numeros[['lon', 'lat']] = df_numeros.apply(lambda row: pd.Series(transformer.transform(row.iloc[13], row.iloc[12])), axis=1)
```

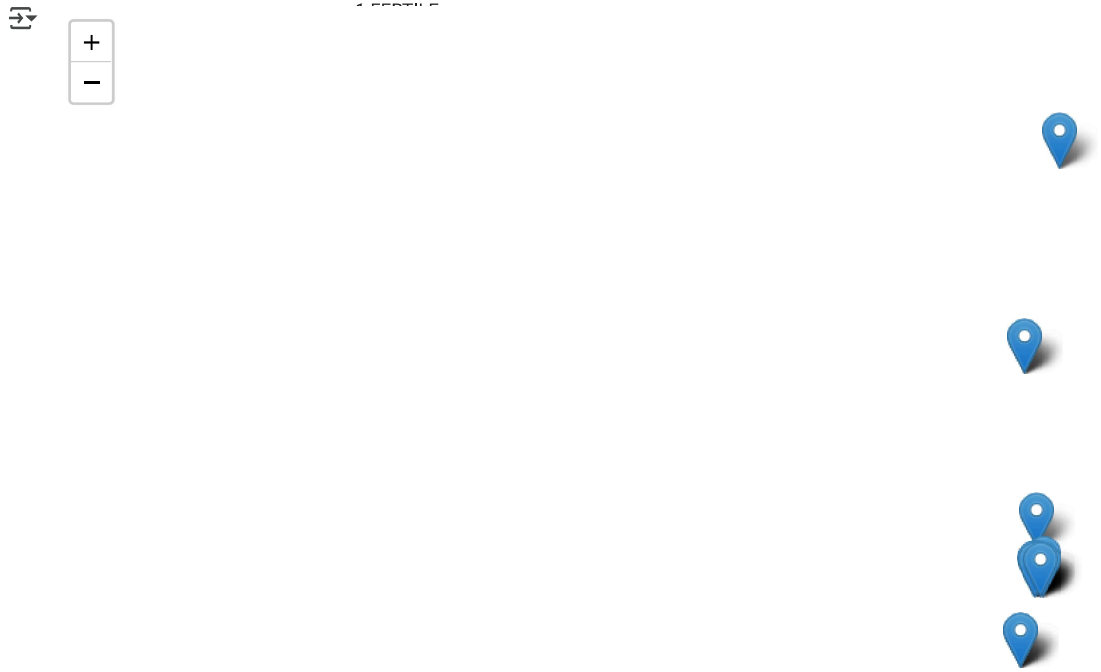
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	lat	lon
4	23-Jan-2025	1		1	1	MATURE NO \nINSEMINATED FEMALE	COPIAPÓ-ATACAMA	VICENTE MERINO \nJARPA #6870	JACKSON/TRIMEDLURE	PEAR TREE - Pyrus \ncommunis	JOSÉ HERRERA - 23-01-2025	ON PROCESS	6967201	374312	NaN	-27.412790	-70.271437
5	23-Jan-2025	2	1		1	MATURE MALE	COPIAPÓ-ATACAMA	COLLIPULLI #1601	JACKSON/TRIMEDLURE	ORNAMENTAL TREE	JOSÉ HERRERA - 23-01-2025	ON PROCESS	6967857	374409	NaN	-27.406878	-70.270388
11	27-Feb-25	1		1	1	MATURE INSEMINATED \nFEMALE	PEÑALOLEN	QUEBRADA DE AROMA \n#1984	JACKSON/ \nTRIMEDLURE	FIG TREE -Ficus carica	NICOLE SILVA - 27-02-2025		6294631	357917	NaN	-33.478386	-70.529208
17	04-03-25	1	0	1	1	1 \nINSEMINATED \nMATURE \nFEMALE	LAS CONDES - \nMETROPOLITANA	LUXEMBURGO \n#9937	MCPHAIL/hydrolyze\nd protein	Peach tree - \nPrunus \npersica	ESTEFANIA \nLEAL - 04-03-\n2025		6304740	357164	NaN	-33.387138	-70.535703
21	10-03-25	1		1	1	FERTILE YOUNG \nFEMALE NOT \nINSEMINATED	MALLOA - O'HIGGINS	Callejóm los \nCopihues s/n \nPelequén Viejo	JACKSON/TRIMEDLURE	FIG TREE - (Ficus \ncarica)	OSCAR ROJAS - \n10.03.2025	BBBB	6187036	325724	1	-34.443498	-70.896959
22	14-03-25	2	1		1	FERTILE YOUNG \nMALE	MALLOA - O'HIGGINS	El Rosedal s/n	JACKSON/TRIMEDLURE	ORNAMENTAL \nTREE	OSCAR ROJAS - 14-03-\n2025	PENDIENTE	6186327	325779	2	-34.449898	-70.896505
23	14-03-25	3		1	1	FERTILE YOUNG \nFEMALE NOT \nINSEMINATED	MALLOA - O'HIGGINS	Las Hortensias	MC PHAIL/PROTEINA	ORNAMENTAL \nTREE	OSCAR ROJAS - 14-03-\n2025	PENDIENTE	6186328	325716	3	-34.449878	-70.897191
24	14-03-25	4		1	1	Inseminated \nmature female	MALLOA - O'HIGGINS	Los Copihues s/n	MC PHAIL/PROTEINA	MEMBRILLERO - \n(Cydonia \noblonga)	OSCAR ROJAS - 14-03-\n2025	PENDIENTE	6187030	325733	4	-34.443553	-70.896863
30	24-01-25	1	0	1	1	1 MATURE VIRGIN FERTILE FEMALE	PEDRO AGUIRRE \nCERDA -	MARILUAN \nFRENTE # 2268	JACKSON/TRIMEDLU\nRE	Plum tree \n(Prunus	NICOLE SILVA-24-\n01-2025	HE 2653/2025 \nBBBB Yellow	6294470	345225	NaN	-33.478077	-70.665789
31	28-01-24	2	4	1	5	1 MATURE VIRGIN FERTILE FEMALE \nAND 4 MATURE ...	SANTIAGO- \nMETROPOLITANA	ARAUCO # 848	JACKSON/TRIMEDLU\nRE	Prunus spp .	NICOLE SILVA-29-\n01-2025	HE 2906/2025 \nAABB Green	6295184	347183	NaN	-33.471921	-70.644601
32	29-01-24	3	5	0	5	4 YOUNG FERTILE MALES AND 1 \nMATURE FERTILE MALE	SANTIAGO- \nMETROPOLITANA	ARAUCO # 856 \nC	JACKSON/TRIMEDLU\nRE	Orange tree \n(Citrus \nsinesis)	ARACELI \nAGUIRRE-29-01-\n2025		6295168	347176	NaN	-33.472065	-70.644679

```
# Crear mapa centrado en Chile
Map = folium.Map(location=[-33.45, -70.65], zoom_start=6)  # Coordenadas aproximadas de Santiago

# Agregar puntos al mapa
```

```
for _, row in df_numeros.iterrows():
    folium.Marker([row['lat'], row['lon']], popup=f"Coordenada: ({row['lat']}, {row['lon']})").add_to(Map)

# Mostrar mapa
Map
```



CONTINUACION CODIGO:

4.- Mostrar en mapa

```
!pip install matplotlib seaborn geopandas folium
```

```
Requirement already satisfied: matplotlib in /usr/local/lib/python3.11/dist-packages (3.10.0)
Requirement already satisfied: seaborn in /usr/local/lib/python3.11/dist-packages (0.13.2)
```

```
Requirement already satisfied: geopandas in /usr/local/lib/python3.11/dist-packages (1.0.1)
Requirement already satisfied: folium in /usr/local/lib/python3.11/dist-packages (0.19.5)
Requirement already satisfied: contourpy>=1.0.1 in /usr/local/lib/python3.11/dist-packages (from matplotlib) (1.3.2)
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.11/dist-packages (from matplotlib) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in /usr/local/lib/python3.11/dist-packages (from matplotlib) (4.57.0)
Requirement already satisfied: kiwisolver>=1.3.1 in /usr/local/lib/python3.11/dist-packages (from matplotlib) (1.4.8)
Requirement already satisfied: numpy>=1.23 in /usr/local/lib/python3.11/dist-packages (from matplotlib) (2.0.2)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.11/dist-packages (from matplotlib) (24.2)
Requirement already satisfied: pillow>=8 in /usr/local/lib/python3.11/dist-packages (from matplotlib) (11.1.0)
Requirement already satisfied: pyparsing>=2.3.1 in /usr/local/lib/python3.11/dist-packages (from matplotlib) (3.2.3)
Requirement already satisfied: python-dateutil>=2.7 in /usr/local/lib/python3.11/dist-packages (from matplotlib) (2.8.2)
Requirement already satisfied: pandas>=1.2 in /usr/local/lib/python3.11/dist-packages (from seaborn) (2.2.2)
Requirement already satisfied: pyogrio>=0.7.2 in /usr/local/lib/python3.11/dist-packages (from geopandas) (0.10.0)
Requirement already satisfied: pyproj>=3.3.0 in /usr/local/lib/python3.11/dist-packages (from geopandas) (3.7.1)
Requirement already satisfied: shapely>=2.0.0 in /usr/local/lib/python3.11/dist-packages (from geopandas) (2.1.0)
Requirement already satisfied: branca>=0.6.0 in /usr/local/lib/python3.11/dist-packages (from folium) (0.8.1)
Requirement already satisfied: jinja2>=2.9 in /usr/local/lib/python3.11/dist-packages (from folium) (3.1.6)
Requirement already satisfied: requests in /usr/local/lib/python3.11/dist-packages (from folium) (2.32.3)
Requirement already satisfied: xyzservices in /usr/local/lib/python3.11/dist-packages (from folium) (2025.1.0)
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.11/dist-packages (from jinja2>=2.9->folium) (3.0.2)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.11/dist-packages (from pandas>=1.2->seaborn) (2025.2)
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.11/dist-packages (from pandas>=1.2->seaborn) (2025.2)
Requirement already satisfied: certifi in /usr/local/lib/python3.11/dist-packages (from pyogrio>=0.7.2->geopandas) (2025.1.31)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.11/dist-packages (from python-dateutil>=2.7->matplotlib) (1.17.0)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.11/dist-packages (from requests->folium) (3.4.1)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.11/dist-packages (from requests->folium) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.11/dist-packages (from requests->folium) (2.3.0)
```

```
import geopandas as gpd
import folium
import matplotlib.pyplot as plt
from shapely.geometry import Point

# Crear GeoDataFrame con las coordenadas
geometry = [Point(lon, lat) for lon, lat in zip(df_numeros["lon"], df_numeros["lat"])]
gdf = gpd.GeoDataFrame(df_numeros, geometry=geometry)

# Crear mapa de Chile con Folium
Map = folium.Map(location=[-33.45, -70.65], zoom_start=6,  tiles="CartoDB dark_matter")  # Estilo más limpio

# Agregar puntos como marcadores
for _, row in df_numeros.iterrows():
    folium.CircleMarker(location=[row["lat"], row["lon"]], radius=5, color="blue", fill=True).add_to(Map)

#

# Agregar capa WMS de áreas verdes (ajusta la URL según el dataset)
folium.raster_layers.WmsTileLayer(
    url="https://ide.minvu.cl/arcgis/services/MapaBase/MapServer/WMSServer",
    layers="0",  # Ajusta el número de capa según el dataset
    name="Áreas Verdes",
    format="image/png",
    transparent=True
).add_to(Map)

#

# Mostrar mapa en Google Colab
display(Map)
```

https://ide.minvu.cl/

https://ide.minvu.cl/maps/864ec380054143228818c6a4ccbbf570

https://ide.minvu.cl/datasets/864ec380054143228818c6a4ccbbf570_0/explore?location=-33.482166%2C-70.637530%2C12.06

```
# PAPELERA

import geopandas as gpd
```

```
import matplotlib.pyplot as plt
from shapely.geometry import Point
import numpy as np

# Crear GeoDataFrame con las coordenadas del DataFrame existente
geometry = [Point(lon, lat) for lon, lat in zip(df_numeros["lon"], df_numeros["lat"])]
gdf = gpd.GeoDataFrame(df_numeros, geometry=geometry)

# Crear un hexbin map con Matplotlib
fig, ax = plt.subplots(figsize=(8, 6))
hb = ax.hexbin(gdf.geometry.x, gdf.geometry.y, gridsize=50, cmap="coolwarm", mincnt=1)  # Ajusta gridsize para cambiar tamaño de hexágonos
```