Madrid_Pain_Graphs

May 25, 2021

1 Informes de la comunidad de Madrid

Actualizado diariamente, este documento se visualiza mejor aquí.

Datos de la situación de la infección por coronavirus en la Comunidad de Madrid.

Nos descargamos los datos, agrupamos, y calculamos :

- Gráfico de seguimiento.
- Muertes medias diarias, últimos 7 días.
- Muertes medias diarias desde que la comunidad de Madrid publica datos.

[11]:
```
# Miramos si hay nuevos datos a descargar.

!# cd ../data/; FILELIST=" 200509 200508 200507 200506 200505 200504 200503␣
↪200502 200501 200430 200429 200428 200427 200426 200425 200424 200423 200422␣
↪200510 200511 200512 200513 200514 200515 200516 200517 200518 200519 200520␣
↪200521 200522 200523 200524 200525 200526 200527 200528 200529 200530 200609␣
↪200608 200607 200606 200605 200604 200603 200602 200601 200610 200611 200612␣
↪200613 200614 200615 200616 200617 200618 200619 200620 200621 200622 200623␣
↪200624 200625 200626 200627 200628 200629 200630 " ; for fecha in `echo␣
↪$FILELIST`  ; do  FILE=${fecha}_cam_covid19.pdf ;  [ ! -f ../data/${FILE} ]␣
↪&& echo $FILE::::: && wget https://www.comunidad.madrid/sites/default/
↪files/doc/sanidad/$FILE 1>/dev/null  2>/dev/null && ls -altr $FILE ; done

# Miramos solo hoy y los ultimos diez dias
! cd ../data/; FILELIST=`seq -w 0 7 | while read i ; do date +%y%m%d -d "$i day␣
↪ago" ; done` ; for fecha in `echo $FILELIST`  ; do ␣
↪FILE=${fecha}_cam_covid19.pdf ;  [ ! -f ../data/${FILE} ] && echo $FILE:::::␣
↪   && wget https://www.comunidad.madrid/sites/default/files/aud/sanidad/
↪$FILE  1>/dev/null  2>/dev/null && ls -altr $FILE ; done
! cd ../data/; FILELIST=`seq -w 0 7 | while read i ; do date +%y%m%d -d "$i day␣
↪ago" ; done` ; for fecha in `echo $FILELIST`  ; do ␣
↪FILE=${fecha}_cam_covid19.pdf ;  [ ! -f ../data/${FILE} ] && echo $FILE:::::␣
↪   && wget https://www.comunidad.madrid/sites/default/files/doc/sanidad/
↪$FILE  1>/dev/null  2>/dev/null && ls -altr $FILE ; done
```

```
! cd ../data/; FILELIST=`seq -w 0 7 | while read i ; do date +%y%m%d -d "$i day␣
 ↪ago" ; done` ; for fecha in `echo $FILELIST`  ; do  FILE=${fecha}cam_covid19.
 ↪pdf ;  [ ! -f ../data/${FILE} ] && echo $FILE::::    && wget https://www.
 ↪comunidad.madrid/sites/default/files/doc/sanidad/$FILE  1>/dev/null  2>/dev/
 ↪null && ls -altr $FILE ; done
! cd ../data/; FILELIST=`seq -w 0 7 | while read i ; do date +%Y%m%d -d "$i day␣
 ↪ago" ; done` ; for fecha in `echo $FILELIST`  ; do ␣
 ↪FILE=${fecha}_cam_covid19.pdf ;  [ ! -f ../data/${FILE} ] && echo $FILE:::::␣
 ↪   && wget https://www.comunidad.madrid/sites/default/files/doc/sanidad/
 ↪$FILE  1>/dev/null  2>/dev/null && ls -altr $FILE ; done
! cd ../data/; FILELIST=`seq -w 0 7 | while read i ; do date +%y%m%d -d "$i day␣
 ↪ago" ; done` ; for fecha in `echo $FILELIST`  ; do ␣
 ↪FILE=${fecha}_cam_covid19.pdf ;  [ ! -f ../data/${FILE} ] && echo $FILE:::::␣
 ↪   && wget https://www.comunidad.madrid/sites/default/files/$FILE 1>/dev/
 ↪null  1>/dev/null  2>/dev/null && ls -altr $FILE ; done
! cd ../data/; FILELIST=`seq -w 0 7 | while read i ; do date +%-d.%-m.%Y -d "$i␣
 ↪day ago" ; done` ; for fecha in `echo $FILELIST`  ; do  FILE=${fecha}_2.pdf ;
 ↪  [ ! -f ../data/${FILE} ] && echo $FILE::::    && wget https://www.
 ↪comunidad.madrid/sites/default/files/doc/sanidad/$FILE       1>/dev/null ␣
 ↪2>/dev/null && ls -altr $FILE ; done
! cd ../data/; FILELIST=`seq -w 0 7 | while read i ; do date +%-d.%-m.%Y -d "$i␣
 ↪day ago" ; done` ; for fecha in `echo $FILELIST`  ; do  FILE=${fecha}.pdf ; ␣
 ↪[ ! -f ../data/${FILE} ] && echo $FILE::::    && wget https://www.comunidad.
 ↪madrid/sites/default/files/doc/sanidad/$FILE       1>/dev/null  2>/dev/null␣
 ↪&& ls -altr $FILE ; done
! cd ../data/; FILELIST=`seq -w 0 7 | while read i ; do date +%y%m%d -d "$i day␣
 ↪ago" ; done` ; for fecha in `echo $FILELIST`  ; do  FILE=${fecha}cam_covid19.
 ↪pdf ;  [ ! -f ../data/${FILE} ] && echo $FILE::::    && wget https://www.
 ↪comunidad.madrid/sites/default/files/doc/sanidad/$FILE  1>/dev/null  2>/dev/
 ↪null && ls -altr $FILE ; done
```

```
210308_cam_covid19.pdf:::::
210308_cam_covid19.pdf:::::
210313cam_covid19.pdf:::::
210312cam_covid19.pdf:::::
210311cam_covid19.pdf:::::
210310cam_covid19.pdf:::::
210309cam_covid19.pdf:::::
210308cam_covid19.pdf:::::
210307cam_covid19.pdf:::::
210306cam_covid19.pdf:::::
20210313_cam_covid19.pdf:::::
20210312_cam_covid19.pdf:::::
20210311_cam_covid19.pdf:::::
20210310_cam_covid19.pdf:::::
20210309_cam_covid19.pdf:::::
20210308_cam_covid19.pdf:::::
```

```
20210307_cam_covid19.pdf:::::
20210306_cam_covid19.pdf:::::
210308_cam_covid19.pdf:::::
13.3.2021_2.pdf:::::
12.3.2021_2.pdf:::::
11.3.2021_2.pdf:::::
10.3.2021_2.pdf:::::
9.3.2021_2.pdf:::::
8.3.2021_2.pdf:::::
7.3.2021_2.pdf:::::
6.3.2021_2.pdf:::::
13.3.2021.pdf:::::
12.3.2021.pdf:::::
11.3.2021.pdf:::::
10.3.2021.pdf:::::
9.3.2021.pdf:::::
8.3.2021.pdf:::::
7.3.2021.pdf:::::
6.3.2021.pdf:::::
210313cam_covid19.pdf:::::
210312cam_covid19.pdf:::::
210311cam_covid19.pdf:::::
210310cam_covid19.pdf:::::
210309cam_covid19.pdf:::::
210308cam_covid19.pdf:::::
210307cam_covid19.pdf:::::
210306cam_covid19.pdf:::::
```

[1]:
```
from tabula import read_pdf
from IPython.display import display, HTML
import os
import pandas as pd
import glob
import re
from tqdm.notebook import tqdm
import warnings
import os.path
#import datetime
warnings.filterwarnings('ignore')


os.environ["JAVA_HOME"] = "/usr/lib/jvm/java-1.8.0-openjdk-1.8.0.141-1.b16.
 ↪el7_3.x86_64/jre"

# Auxiliary functions
from datetime import datetime, date, time, timedelta
```

```
df_cache = pd.read_csv("/root/kaggle/covid19-madrid/madrid_results.csv")

def query_cache(fecha):
    """ Query cache file to avoid parse pdf
    return empty dataframe is not found"""
    try :
        if '.' in fecha:
            date_regexp='%d.%m.%Y'
        else:
            date_regexp='%y%m%d'
        date_formatted = datetime.strptime(fecha,date_regexp ).
 ↪strftime('%Y-%m-%d')
        df = df_cache.query( 'Fecha==@date_formatted')
    except:
        print("Cache miss:" , fecha)
        return pd.DataFrame()
    #print(f"fecha {fecha},{date_formatted}")

    try:
        df['Fecha'] =  pd.to_datetime(date_formatted, format='%y-%m-%d')
    except :
        df['Fecha'] =  pd.to_datetime(date_formatted, format='%Y-%m-%d')
    df.set_index('Fecha', inplace=True, drop=True)
    return df

""" Rellenar dias vacios con interpolacion"""
def interpolate_dataframe(df,freq):
    if freq == 'H':
        rng = pd.date_range(df.index.min(), df.index.max() + pd.Timedelta(23,␣
 ↪'H'), freq='H')
    elif freq == 'D' :
        rng = pd.date_range(
                        datetime.strptime(str(df.index.min())[:10]+' 00:00:00',␣
 ↪"%Y-%m-%d %H:%M:%S") ,
                        datetime.strptime(str(df.index.max())[:10]+' 00:00:00',␣
 ↪"%Y-%m-%d %H:%M:%S"),
                        freq='D')
        df.index = pd.to_datetime(df.index)
    df2 = df.reindex(rng)
    df = df2
    for column in df.columns :
        s = pd.Series(df[column])
        s.interpolate(method="quadratic", inplace =True)
        df[column] = pd.DataFrame([s]).T
    return df

def get_daily_date_new_format_vacunas(fecha,filename):
```

```
    """ Se añadio una página de vacunas que obviamos"""
    print(f"get_daily_date_new_format_vacunas, {fecha}, {filename}")
    file_path = '../data/'+fecha+'_cam_covid19.pdf'
    if not os.path.isfile(file_path):
        file_path = '../data/'+fecha+'cam_covid19.pdf'
    if not os.path.isfile(file_path):
        file_path = filename
    df_pdf =  read_pdf(file_path,area=(150, 625, 400, 900)  , pages=2)[0]

    df = df_pdf['Datos sanidad mortuoria.'].astype(str).str.replace(r".", '').
↪replace("(", ' ').replace(r"*","")
    df = pd.DataFrame(df)
    dict = {}

    dict['HOSPITALES'] = df[df['Datos sanidad mortuoria.'].str.
↪contains('Hospitales')].iloc[0]['Datos sanidad mortuoria.'].split(' ')[0]
    dict['DOMICILIOS'] = df[df['Datos sanidad mortuoria.'].str.
↪contains('Domicilios')].iloc[0]['Datos sanidad mortuoria.'].split(' ')[0]
    dict['CENTROS SOCIOSANITARIOS'] = df[df['Datos sanidad mortuoria.'].str.
↪contains('Centros')].iloc[0]['Datos sanidad mortuoria.'].split(' ')[0]
    dict['OTROS LUGARES'] = df[df['Datos sanidad mortuoria.'].str.
↪contains('otros')].iloc[0]['Datos sanidad mortuoria.'].split(' ')[0]
    cadena_a_parsear = df[df['Datos sanidad mortuoria.'].str.contains('otal')].
↪iloc[0]['Datos sanidad mortuoria.']
    dict['FALLECIDOS TOTALES'] = re.search(r'(\d+)', cadena_a_parsear)[0]

    try:
        df2_pdf =  read_pdf(file_path,area=(300, 100, 800, 400) , pages=2)
        dict['PACIENTES UCI DIA']       = df2_pdf[0].loc[3:3].values[0][1].
↪replace(".",'')
        dict['PACIENTES UCI ACUMULADOS']= df2_pdf[0].loc[6:6].values[0][1].
↪replace(".",'')
    except Exception as e:
        print(f"{fecha} mal parseada: {e}")

    df = pd.DataFrame.from_dict(dict, orient='index').T
    #print("4.5 get_daily_date_new_format")

    if '.' in fecha :
        try:
            df['Fecha'] = pd.to_datetime(fecha, format='%d.%m.%Y')
        except :
            df['Fecha'] =  pd.to_datetime(fecha, format='%d.%m.%y')
    else:
        try:
            df['Fecha'] =  pd.to_datetime(fecha, format='%y%m%d')
```

```
        except :
            df['Fecha'] =  pd.to_datetime(fecha, format='%Y%m%d')

    df.set_index('Fecha', inplace=True, drop=True)
    return df



def get_daily_date_new_format(fecha,filename):
    print(f"get_daily_date_new_format({fecha},{filename})")
    PAGINA_DE_DATOS=1

    file_path = '../data/'+fecha+'_cam_covid19.pdf'
    if not os.path.isfile(file_path):
        file_path = '../data/'+fecha+'cam_covid19.pdf'
    if not os.path.isfile(file_path):
        file_path = filename

    #print("Analizando:" + file_path)
    df_pdf =  read_pdf(file_path,area=(000, 600, 400, 800) ,␣
↪pages=PAGINA_DE_DATOS)

    # Parche, para los saltos de linea en el pdf
    if 'Unnamed: 0' not in df_pdf[0].columns :
        return pd.DataFrame()


    df = df_pdf[0]
    df = df['Unnamed: 0'].astype(str).str.replace(r".", '').replace("(", ' ')
    df = df.T
    df.columns = df.iloc[0]
    df = df.iloc[1:]

    #print("2 get_daily_date_new_format")

    df = pd.DataFrame(data=df)
    df

    dict = {}
    try:
        df2_pdf =  read_pdf(file_path,area=(300, 100, 800, 400) ,␣
↪pages=PAGINA_DE_DATOS)
        dict['PACIENTES UCI DIA']       = df2_pdf[0].loc[3:3].values[0][1].
↪replace(".",'')
        dict['PACIENTES UCI ACUMULADOS']= df2_pdf[0].loc[6:6].values[0][1].
↪replace(".",'')
    except Exception as e:
```

```
        print(f"{fecha} mal parseada: {e}")

    dict['HOSPITALES'] = df[df['Unnamed: 0'].str.contains('Hospitales')].
↪iloc[0]['Unnamed: 0'].split(' ')[0]
    dict['DOMICILIOS'] = df[df['Unnamed: 0'].str.contains('Domicilios')].
↪iloc[0]['Unnamed: 0'].split(' ')[0]
    dict['CENTROS SOCIOSANITARIOS'] = df[df['Unnamed: 0'].str.
↪contains('Centros')].iloc[0]['Unnamed: 0'].split(' ')[0]
    dict['OTROS LUGARES'] = df[df['Unnamed: 0'].str.contains('otros')].
↪iloc[0]['Unnamed: 0'].split(' ')[0]
    #print("3 get_daily_date_new_format")

    cadena_a_parsear = df[df['Unnamed: 0'].str.contains('otal')].
↪iloc[0]['Unnamed: 0']

    dict['FALLECIDOS TOTALES'] = re.search(r'(\d+)', cadena_a_parsear)[0]

    #print("4 get_daily_date_new_format")
    df = pd.DataFrame.from_dict(dict, orient='index').T
    #print("4.5 get_daily_date_new_format")

    if '.' in fecha :
        try:
            df['Fecha'] = pd.to_datetime(fecha, format='%d.%m.%Y')
        except :
            df['Fecha'] =  pd.to_datetime(fecha, format='%d.%m.%y')
    else:
        try:
            df['Fecha'] =  pd.to_datetime(fecha, format='%y%m%d')
        except :
            df['Fecha'] =  pd.to_datetime(fecha, format='%Y%m%d')

    #print("5 get_daily_date_new_format")

    df.set_index('Fecha', inplace=True, drop=True)
    #print(df)
    return df


def get_daily_data(fecha,filename):
    #print(f"""get_daily_data: {fecha}""")
    #print(f"""../data/{fecha}_cam_covid19.pdf""")

    if  fecha > "210228" :
        #print(f"Detected vacunas format {fecha}")
        return get_daily_date_new_format_vacunas(fecha,filename)
    if fecha > '200512' :
```

```
        return get_daily_date_new_format(fecha,filename)

    col2str = {'dtype': str}
    kwargs = {'output_format': 'dataframe',
              'pandas_options': col2str,
              'stream': True}

    df_pdf =  read_pdf('../data/'+fecha+'_cam_covid19.
 ↪pdf',pages='1',multiple_tables = True,**kwargs)


    df = df_pdf[0]

    df = df[df['Unnamed: 0'].notna()]
    df = df[(df['Unnamed: 0']=='HOSPITALES') | (df['Unnamed: 0'] ==␣
 ↪'DOMICILIOS') | (df['Unnamed: 0'] == 'CENTROS SOCIOSANITARIOS') |␣
 ↪(df['Unnamed: 0'] == 'OTROS LUGARES') | (df['Unnamed: 0'] == 'FALLECIDOS␣
 ↪TOTALES')]
    df = df[['Unnamed: 0','Unnamed: 2']]
    df['Unnamed: 2'] = df['Unnamed: 2'].astype(str).str.replace(r".", '')
    df = df.T
    df.columns = df.iloc[0]
    df = df.iloc[1:]

    df['Fecha'] =  pd.to_datetime(fecha, format='%y%m%d')
    df = df.rename_axis(None)

    df.set_index('Fecha', inplace=True, drop=True)
    df.index
    df.dropna()
    #df = df.T
    return df

def get_all_data( ):
    #BLACKLIST = ["200429","200422"]
    #BLACKLIST = ["200514",]
    BLACKLIST = []
    df = pd.DataFrame()
    list_df = []

    #pdf_list= (glob.glob('../data/*_covid19.pdf'),
    #                key=os.path.getmtime,
    #                 reverse=True )
    pdf_list= set(glob.glob('../data/*202*pdf') + glob.glob('../data/
 ↪*cam_covid19.pdf'))
```

```
    for pdf_file in tqdm(pdf_list,
                         desc="Procesando pdfs diarios"):
        # extract fecha from username , eg : ../data/2200422_cam_covid19.pdf

        format_point_occurences = pdf_file.split('/')[2].split('_')[0].count(".
↪")
        # Hack to fix filename inconsistences on remote server
        if format_point_occurences > 2 :
            day = pdf_file.split('/')[2].split('_')[0].split('.')[0].zfill(2)
            month = pdf_file.split('/')[2].split('_')[0].split('.')[1].zfill(2)
            year = pdf_file.split('/')[2].split('_')[0].split('.')[2][-2:]
            fecha = year+month+day
            fecha=fecha.replace('.pdf','')
        else :
            fecha = pdf_file.split('/')[2].split('_')[0].replace('cam_','').
↪replace('_cam_','').replace('cam','')
        if fecha not in BLACKLIST:
            # query cache, otherwise parse pdf
            df = query_cache(fecha)
            if df.empty:
                df = get_daily_data(fecha,pdf_file)
        list_df.append(df)

    df = pd.concat(list_df)
    df = df.fillna(0)
    df = df.astype(int)
    df = df.drop_duplicates()

    df = df.sort_values(by=['Fecha'], ascending=True)

    df['HOSPITALES hoy'] = df['HOSPITALES'] - df['HOSPITALES'].shift(1)
    df['CENTROS SOCIOSANITARIOS hoy'] = df['CENTROS SOCIOSANITARIOS'] -␣
↪df['CENTROS SOCIOSANITARIOS'].shift(1)
    df['FALLECIDOS TOTALES hoy'] = df['FALLECIDOS TOTALES'] - df['FALLECIDOS␣
↪TOTALES'].shift(1)

    df = df.sort_values(by=['Fecha'], ascending=False)

    return df

total =  get_all_data()

total.to_csv('/root/kaggle/covid19-madrid/madrid_results.csv')
total
```

```
HBox(children=(FloatProgress(value=0.0, description='Procesando pdfs diarios', max=314.0, style
```

```
get_daily_date_new_format(200516,../data/200516_cam_covid19.pdf)
200516 mal parseada: index 1 is out of bounds for axis 0 with size 1
Cache miss: 20200814
get_daily_date_new_format(20200814,../data/20200814_cam_covid19.pdf)
get_daily_date_new_format(201219,../data/201219_cam_covid19.pdf)

Got stderr: abr 10, 2021 7:51:42 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: abr 10, 2021 7:51:43 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'


get_daily_date_new_format(201210,../data/201210_cam_covid19.pdf)

Got stderr: abr 10, 2021 7:51:45 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'
```

[1]:

Fecha	CENTROS SOCIOSANITARIOS	CENTROS SOCIOSANITARIOS hoy	DOMICILIOS	\
2021-04-10	5063	0.0	1329	
2021-04-09	5063	0.0	1329	
2021-04-08	5063	0.0	1327	
2021-04-07	5063	0.0	1327	
2021-04-06	5063	0.0	1327	
…	…	…	…	
2020-04-26	4236	66.0	798	
2020-04-25	4170	102.0	788	
2020-04-24	4068	72.0	775	
2020-04-23	3996	64.0	769	
2020-04-22	3932	NaN	761	

Fecha	FALLECIDOS TOTALES	FALLECIDOS TOTALES hoy	HOSPITALES	\
2021-04-10	23355	0.0	16933	
2021-04-09	23355	45.0	16933	
2021-04-08	23310	0.0	16890	
2021-04-07	23310	26.0	16890	
2021-04-06	23284	0.0	16864	
…	…	…	…	
2020-04-26	12855	243.0	7800	
2020-04-25	12612	360.0	7633	
2020-04-24	12252	196.0	7388	

```
2020-04-23               12056                   204.0        7271
2020-04-22               11852                     NaN        7144

            HOSPITALES hoy  OTROS LUGARES  PACIENTES UCI ACUMULADOS  \
Fecha
2021-04-10             0.0             30                     10280
2021-04-09            43.0             30                     10251
2021-04-08             0.0             30                     10225
2021-04-07            26.0             30                     10186
2021-04-06             0.0             30                     10152
…                      …               …                      …
2020-04-26           167.0             21                         0
2020-04-25           245.0             21                         0
2020-04-24           117.0             21                         0
2020-04-23           127.0             20                         0
2020-04-22             NaN             15                         0

            PACIENTES UCI DIA
Fecha
2021-04-10                485
2021-04-09                477
2021-04-08                474
2021-04-07                457
2021-04-06                447
…                         …
2020-04-26                  0
2020-04-25                  0
2020-04-24                  0
2020-04-23                  0
2020-04-22                  0

[308 rows x 10 columns]
```

[2]:
```
total
df = total
df = df.fillna(0)
df = df.astype(int)
df
```

[2]:
```
            CENTROS SOCIOSANITARIOS  CENTROS SOCIOSANITARIOS hoy  DOMICILIOS  \
Fecha
2021-04-10                     5063                            0        1329
2021-04-09                     5063                            0        1329
2021-04-08                     5063                            0        1327
2021-04-07                     5063                            0        1327
2021-04-06                     5063                            0        1327
…                              …                               …        …
```

2020-04-26	4236	66	798
2020-04-25	4170	102	788
2020-04-24	4068	72	775
2020-04-23	3996	64	769
2020-04-22	3932	0	761

Fecha	FALLECIDOS TOTALES	FALLECIDOS TOTALES hoy	HOSPITALES	\
2021-04-10	23355	0	16933	
2021-04-09	23355	45	16933	
2021-04-08	23310	0	16890	
2021-04-07	23310	26	16890	
2021-04-06	23284	0	16864	
…	…	…	…	
2020-04-26	12855	243	7800	
2020-04-25	12612	360	7633	
2020-04-24	12252	196	7388	
2020-04-23	12056	204	7271	
2020-04-22	11852	0	7144	

Fecha	HOSPITALES hoy	OTROS LUGARES	PACIENTES UCI ACUMULADOS	\
2021-04-10	0	30	10280	
2021-04-09	43	30	10251	
2021-04-08	0	30	10225	
2021-04-07	26	30	10186	
2021-04-06	0	30	10152	
…	…	…	…	
2020-04-26	167	21	0	
2020-04-25	245	21	0	
2020-04-24	117	21	0	
2020-04-23	127	20	0	
2020-04-22	0	15	0	

Fecha	PACIENTES UCI DIA
2021-04-10	485
2021-04-09	477
2021-04-08	474
2021-04-07	457
2021-04-06	447
…	…
2020-04-26	0
2020-04-25	0
2020-04-24	0
2020-04-23	0
2020-04-22	0

```
[308 rows x 10 columns]
```

[75]:
```
import re
import PyPDF2




def get_daily_date_edades(fecha,filename,page):
    """ Se añadio una página de vacunas que obviamos"""

    print(f"get_daily_date_new_format_vacunas, {fecha}, {filename}")
    file_path = '../data/'+fecha+'_cam_covid19.pdf'
    if not os.path.isfile(file_path):
        file_path = '../data/'+fecha+'cam_covid19.pdf'
    if not os.path.isfile(file_path):
        file_path = filename

    #pages =count_pdf_pages(file_path)
    print(f"get_daily_date_new_format_vacunas, {fecha}, {filename}")



    col2str = {'dtype': str}
    kwargs = {'output_format': 'dataframe',
              'pandas_options': col2str,
              'multiple_tables' :True,
              'stream': True,
             'area' : (150, 00, 400, 400)}

    df_pdf =  read_pdf('../data/'+fecha+'_cam_covid19.pdf',pages=page,**kwargs)
    df = df_pdf[0]
    for column in ['Unnamed: 2','Unnamed: 3','Unnamed: 6','Unnamed: 7']:
        print(column,df.columns.values )
        if column in df.columns.values :
            df.drop( column , inplace=True,axis=1)
    df = df.dropna()
    df = df.rename(columns = {'Unnamed: 0': 'Ages' , 'Mujeres': 'Female', ␣
↪'Hombres': 'Male'}, inplace = False)
    #df['Date'] = datetime.strptime(fecha,date_regexp ).strftime('%Y-%m-%d')
    if '.' in fecha:
        date_regexp='%d.%m.%Y'
    else:
        date_regexp='%y%m%d'
    date_formatted = datetime.strptime(fecha,date_regexp ).strftime('%Y-%m-%d')
```

```
    df['Fecha'] =  pd.to_datetime(date_formatted, format='%Y-%m-%d')


    return df

fecha='210404'
filename='../data/210404_cam_covid19.pdf'


df_ages= get_daily_date_edades(fecha,filename,11)
df_ages
```

```
get_daily_date_new_format_vacunas, 210404, ../data/210404_cam_covid19.pdf
get_daily_date_new_format_vacunas, 210404, ../data/210404_cam_covid19.pdf
Unnamed: 2 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 2' 'Unnamed: 3' 'Hombres'
'Total'
 'Unnamed: 6' 'Unnamed: 7']
Unnamed: 3 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 3' 'Hombres' 'Total' 'Unnamed:
6'
 'Unnamed: 7']
Unnamed: 6 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 6' 'Unnamed:
7']
Unnamed: 7 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 7']
```

[75]:

	Ages	Edad Mujeres	Male	Total	Fecha
0	0-9	2	8	10	2021-04-04
1	10-19	3	1	4	2021-04-04
2	20-29	6	10	16	2021-04-04
3	30-39	23	26	49	2021-04-04
4	40-49	61	132	193	2021-04-04
5	50-59	221	502	723	2021-04-04
6	60-69	525	1.305	1.830	2021-04-04
8	70-79	1.525	2.975	4.500	2021-04-04
9	80-89	4.103	4.878	8.981	2021-04-04
10	90>	4.073	2.535	6.608	2021-04-04
11	Edad no confirmada*	134	152	286	2021-04-04
12	Total general	10.676	12.524	23.200	2021-04-04

[76]:
```
pd.set_option('display.max_rows', 500)

def get_all_data( ):
    #BLACKLIST = ["200429","200422"]
    #BLACKLIST = ["200514",]
    BLACKLIST = []
    df = pd.DataFrame()
    list_df = []
```

```
    #pdf_list= (glob.glob('../data/*_covid19.pdf'),
    #                key=os.path.getmtime,
    #                 reverse=True )
    pdf_list= set( glob.glob('../data/2104*cam_covid19.pdf'))


    for pdf_file in tqdm(pdf_list,
                         desc="Procesando pdfs diarios"):
        # extract fecha from username , eg : ../data/2200422_cam_covid19.pdf

        format_point_occurences = pdf_file.split('/')[2].split('_')[0].count(".
↪")
        # Hack to fix filename inconsistences on remote server
        if format_point_occurences > 2 :
            day = pdf_file.split('/')[2].split('_')[0].split('.')[0].zfill(2)
            month = pdf_file.split('/')[2].split('_')[0].split('.')[1].zfill(2)
            year = pdf_file.split('/')[2].split('_')[0].split('.')[2][-2:]
            fecha = year+month+day
            fecha=fecha.replace('.pdf','')
        else :
            fecha = pdf_file.split('/')[2].split('_')[0].replace('cam_','').
↪replace('_cam_','').replace('cam','')

        for page in range(9, 13):
            try:
                df= get_daily_date_edades(fecha,filename,11)
                list_df.append(df)
            except:
                pass

    return list_df
list_df =      get_all_data( )
df_ages = pd.concat(list_df)
```

```
HBox(children=(FloatProgress(value=0.0, description='Procesando pdfs diarios', max=10.0, style=
```

```
get_daily_date_new_format_vacunas, 210401, ../data/210404_cam_covid19.pdf
get_daily_date_new_format_vacunas, 210401, ../data/210404_cam_covid19.pdf
Unnamed: 2 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 2' 'Unnamed: 3' 'Hombres'
'Total'
 'Unnamed: 6' 'Unnamed: 7']
Unnamed: 3 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 3' 'Hombres' 'Total' 'Unnamed:
6'
 'Unnamed: 7']
Unnamed: 6 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 6' 'Unnamed:
7']
Unnamed: 7 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 7']
```

```
get_daily_date_new_format_vacunas, 210401, ../data/210404_cam_covid19.pdf
get_daily_date_new_format_vacunas, 210401, ../data/210404_cam_covid19.pdf
Unnamed: 2 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 2' 'Unnamed: 3' 'Hombres'
'Total'
 'Unnamed: 6' 'Unnamed: 7']
Unnamed: 3 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 3' 'Hombres' 'Total' 'Unnamed:
6'
 'Unnamed: 7']
Unnamed: 6 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 6' 'Unnamed:
7']
Unnamed: 7 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 7']
get_daily_date_new_format_vacunas, 210401, ../data/210404_cam_covid19.pdf
get_daily_date_new_format_vacunas, 210401, ../data/210404_cam_covid19.pdf
Unnamed: 2 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 2' 'Unnamed: 3' 'Hombres'
'Total'
 'Unnamed: 6' 'Unnamed: 7']
Unnamed: 3 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 3' 'Hombres' 'Total' 'Unnamed:
6'
 'Unnamed: 7']
Unnamed: 6 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 6' 'Unnamed:
7']
Unnamed: 7 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 7']
get_daily_date_new_format_vacunas, 210401, ../data/210404_cam_covid19.pdf
get_daily_date_new_format_vacunas, 210401, ../data/210404_cam_covid19.pdf
Unnamed: 2 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 2' 'Unnamed: 3' 'Hombres'
'Total'
 'Unnamed: 6' 'Unnamed: 7']
Unnamed: 3 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 3' 'Hombres' 'Total' 'Unnamed:
6'
 'Unnamed: 7']
Unnamed: 6 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 6' 'Unnamed:
7']
Unnamed: 7 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 7']
get_daily_date_new_format_vacunas, 210406, ../data/210404_cam_covid19.pdf
get_daily_date_new_format_vacunas, 210406, ../data/210404_cam_covid19.pdf
Unnamed: 2 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 2' 'Unnamed: 3' 'Hombres'
'Total'
 'Unnamed: 6' 'Unnamed: 7']
Unnamed: 3 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 3' 'Hombres' 'Total' 'Unnamed:
6'
 'Unnamed: 7']
Unnamed: 6 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 6' 'Unnamed:
7']
Unnamed: 7 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 7']
get_daily_date_new_format_vacunas, 210406, ../data/210404_cam_covid19.pdf
get_daily_date_new_format_vacunas, 210406, ../data/210404_cam_covid19.pdf
Unnamed: 2 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 2' 'Unnamed: 3' 'Hombres'
'Total'
```

```
 'Unnamed: 6' 'Unnamed: 7']
Unnamed: 3 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 3' 'Hombres' 'Total' 'Unnamed:
6'
 'Unnamed: 7']
Unnamed: 6 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 6' 'Unnamed:
7']
Unnamed: 7 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 7']
get_daily_date_new_format_vacunas, 210406, ../data/210404_cam_covid19.pdf
get_daily_date_new_format_vacunas, 210406, ../data/210404_cam_covid19.pdf
Unnamed: 2 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 2' 'Unnamed: 3' 'Hombres'
'Total'
 'Unnamed: 6' 'Unnamed: 7']
Unnamed: 3 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 3' 'Hombres' 'Total' 'Unnamed:
6'
 'Unnamed: 7']
Unnamed: 6 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 6' 'Unnamed:
7']
Unnamed: 7 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 7']
get_daily_date_new_format_vacunas, 210406, ../data/210404_cam_covid19.pdf
get_daily_date_new_format_vacunas, 210406, ../data/210404_cam_covid19.pdf
Unnamed: 2 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 2' 'Unnamed: 3' 'Hombres'
'Total'
 'Unnamed: 6' 'Unnamed: 7']
Unnamed: 3 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 3' 'Hombres' 'Total' 'Unnamed:
6'
 'Unnamed: 7']
Unnamed: 6 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 6' 'Unnamed:
7']
Unnamed: 7 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 7']
get_daily_date_new_format_vacunas, 210402, ../data/210404_cam_covid19.pdf
get_daily_date_new_format_vacunas, 210402, ../data/210404_cam_covid19.pdf
Unnamed: 2 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 2' 'Unnamed: 3' 'Hombres'
'Total'
 'Unnamed: 6' 'Unnamed: 7']
Unnamed: 3 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 3' 'Hombres' 'Total' 'Unnamed:
6'
 'Unnamed: 7']
Unnamed: 6 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 6' 'Unnamed:
7']
Unnamed: 7 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 7']
get_daily_date_new_format_vacunas, 210402, ../data/210404_cam_covid19.pdf
get_daily_date_new_format_vacunas, 210402, ../data/210404_cam_covid19.pdf
Unnamed: 2 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 2' 'Unnamed: 3' 'Hombres'
'Total'
 'Unnamed: 6' 'Unnamed: 7']
Unnamed: 3 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 3' 'Hombres' 'Total' 'Unnamed:
6'
 'Unnamed: 7']
```

```
Unnamed: 6 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 6' 'Unnamed:
7']
Unnamed: 7 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 7']
get_daily_date_new_format_vacunas, 210402, ../data/210404_cam_covid19.pdf
get_daily_date_new_format_vacunas, 210402, ../data/210404_cam_covid19.pdf
Unnamed: 2 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 2' 'Unnamed: 3' 'Hombres'
'Total'
 'Unnamed: 6' 'Unnamed: 7']
Unnamed: 3 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 3' 'Hombres' 'Total' 'Unnamed:
6'
 'Unnamed: 7']
Unnamed: 6 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 6' 'Unnamed:
7']
Unnamed: 7 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 7']
get_daily_date_new_format_vacunas, 210402, ../data/210404_cam_covid19.pdf
get_daily_date_new_format_vacunas, 210402, ../data/210404_cam_covid19.pdf
Unnamed: 2 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 2' 'Unnamed: 3' 'Hombres'
'Total'
 'Unnamed: 6' 'Unnamed: 7']
Unnamed: 3 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 3' 'Hombres' 'Total' 'Unnamed:
6'
 'Unnamed: 7']
Unnamed: 6 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 6' 'Unnamed:
7']
Unnamed: 7 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 7']
get_daily_date_new_format_vacunas, 210408, ../data/210404_cam_covid19.pdf
get_daily_date_new_format_vacunas, 210408, ../data/210404_cam_covid19.pdf
Unnamed: 2 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 2' 'Unnamed: 3' 'Hombres'
'Total'
 'Unnamed: 6' 'Unnamed: 7']
Unnamed: 3 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 3' 'Hombres' 'Total' 'Unnamed:
6'
 'Unnamed: 7']
Unnamed: 6 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 6' 'Unnamed:
7']
Unnamed: 7 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 7']
get_daily_date_new_format_vacunas, 210408, ../data/210404_cam_covid19.pdf
get_daily_date_new_format_vacunas, 210408, ../data/210404_cam_covid19.pdf
Unnamed: 2 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 2' 'Unnamed: 3' 'Hombres'
'Total'
 'Unnamed: 6' 'Unnamed: 7']
Unnamed: 3 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 3' 'Hombres' 'Total' 'Unnamed:
6'
 'Unnamed: 7']
Unnamed: 6 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 6' 'Unnamed:
7']
Unnamed: 7 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 7']
get_daily_date_new_format_vacunas, 210408, ../data/210404_cam_covid19.pdf
```

```
get_daily_date_new_format_vacunas, 210408, ../data/210404_cam_covid19.pdf
Unnamed: 2 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 2' 'Unnamed: 3' 'Hombres'
'Total'
 'Unnamed: 6' 'Unnamed: 7']
Unnamed: 3 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 3' 'Hombres' 'Total' 'Unnamed:
6'
 'Unnamed: 7']
Unnamed: 6 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 6' 'Unnamed:
7']
Unnamed: 7 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 7']
get_daily_date_new_format_vacunas, 210408, ../data/210404_cam_covid19.pdf
get_daily_date_new_format_vacunas, 210408, ../data/210404_cam_covid19.pdf
Unnamed: 2 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 2' 'Unnamed: 3' 'Hombres'
'Total'
 'Unnamed: 6' 'Unnamed: 7']
Unnamed: 3 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 3' 'Hombres' 'Total' 'Unnamed:
6'
 'Unnamed: 7']
Unnamed: 6 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 6' 'Unnamed:
7']
Unnamed: 7 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 7']
get_daily_date_new_format_vacunas, 210407, ../data/210404_cam_covid19.pdf
get_daily_date_new_format_vacunas, 210407, ../data/210404_cam_covid19.pdf
Unnamed: 2 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 2' 'Unnamed: 3' 'Hombres'
'Total'
 'Unnamed: 6' 'Unnamed: 7']
Unnamed: 3 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 3' 'Hombres' 'Total' 'Unnamed:
6'
 'Unnamed: 7']
Unnamed: 6 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 6' 'Unnamed:
7']
Unnamed: 7 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 7']
get_daily_date_new_format_vacunas, 210407, ../data/210404_cam_covid19.pdf
get_daily_date_new_format_vacunas, 210407, ../data/210404_cam_covid19.pdf
Unnamed: 2 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 2' 'Unnamed: 3' 'Hombres'
'Total'
 'Unnamed: 6' 'Unnamed: 7']
Unnamed: 3 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 3' 'Hombres' 'Total' 'Unnamed:
6'
 'Unnamed: 7']
Unnamed: 6 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 6' 'Unnamed:
7']
Unnamed: 7 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 7']
get_daily_date_new_format_vacunas, 210407, ../data/210404_cam_covid19.pdf
get_daily_date_new_format_vacunas, 210407, ../data/210404_cam_covid19.pdf
Unnamed: 2 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 2' 'Unnamed: 3' 'Hombres'
'Total'
 'Unnamed: 6' 'Unnamed: 7']
```

```
Unnamed: 3 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 3' 'Hombres' 'Total' 'Unnamed:
6'
 'Unnamed: 7']
Unnamed: 6 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 6' 'Unnamed:
7']
Unnamed: 7 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 7']
get_daily_date_new_format_vacunas, 210407, ../data/210404_cam_covid19.pdf
get_daily_date_new_format_vacunas, 210407, ../data/210404_cam_covid19.pdf
Unnamed: 2 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 2' 'Unnamed: 3' 'Hombres'
'Total'
 'Unnamed: 6' 'Unnamed: 7']
Unnamed: 3 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 3' 'Hombres' 'Total' 'Unnamed:
6'
 'Unnamed: 7']
Unnamed: 6 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 6' 'Unnamed:
7']
Unnamed: 7 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 7']
get_daily_date_new_format_vacunas, 210410, ../data/210404_cam_covid19.pdf
get_daily_date_new_format_vacunas, 210410, ../data/210404_cam_covid19.pdf
Unnamed: 2 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 2' 'Unnamed: 3' 'Hombres'
'Total'
 'Unnamed: 6' 'Unnamed: 7']
Unnamed: 3 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 3' 'Hombres' 'Total' 'Unnamed:
6'
 'Unnamed: 7']
Unnamed: 6 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 6' 'Unnamed:
7']
Unnamed: 7 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 7']
get_daily_date_new_format_vacunas, 210410, ../data/210404_cam_covid19.pdf
get_daily_date_new_format_vacunas, 210410, ../data/210404_cam_covid19.pdf
Unnamed: 2 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 2' 'Unnamed: 3' 'Hombres'
'Total'
 'Unnamed: 6' 'Unnamed: 7']
Unnamed: 3 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 3' 'Hombres' 'Total' 'Unnamed:
6'
 'Unnamed: 7']
Unnamed: 6 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 6' 'Unnamed:
7']
Unnamed: 7 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 7']
get_daily_date_new_format_vacunas, 210410, ../data/210404_cam_covid19.pdf
get_daily_date_new_format_vacunas, 210410, ../data/210404_cam_covid19.pdf
Unnamed: 2 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 2' 'Unnamed: 3' 'Hombres'
'Total'
 'Unnamed: 6' 'Unnamed: 7']
Unnamed: 3 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 3' 'Hombres' 'Total' 'Unnamed:
6'
 'Unnamed: 7']
Unnamed: 6 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 6' 'Unnamed:
```

```
7']
Unnamed: 7 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 7']
get_daily_date_new_format_vacunas, 210410, ../data/210404_cam_covid19.pdf
get_daily_date_new_format_vacunas, 210410, ../data/210404_cam_covid19.pdf
Unnamed: 2 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 2' 'Unnamed: 3' 'Hombres'
'Total'
 'Unnamed: 6' 'Unnamed: 7']
Unnamed: 3 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 3' 'Hombres' 'Total' 'Unnamed:
6'
 'Unnamed: 7']
Unnamed: 6 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 6' 'Unnamed:
7']
Unnamed: 7 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 7']
get_daily_date_new_format_vacunas, 210403, ../data/210404_cam_covid19.pdf
get_daily_date_new_format_vacunas, 210403, ../data/210404_cam_covid19.pdf
Unnamed: 2 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 2' 'Unnamed: 3' 'Hombres'
'Total'
 'Unnamed: 6' 'Unnamed: 7']
Unnamed: 3 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 3' 'Hombres' 'Total' 'Unnamed:
6'
 'Unnamed: 7']
Unnamed: 6 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 6' 'Unnamed:
7']
Unnamed: 7 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 7']
get_daily_date_new_format_vacunas, 210403, ../data/210404_cam_covid19.pdf
get_daily_date_new_format_vacunas, 210403, ../data/210404_cam_covid19.pdf
Unnamed: 2 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 2' 'Unnamed: 3' 'Hombres'
'Total'
 'Unnamed: 6' 'Unnamed: 7']
Unnamed: 3 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 3' 'Hombres' 'Total' 'Unnamed:
6'
 'Unnamed: 7']
Unnamed: 6 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 6' 'Unnamed:
7']
Unnamed: 7 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 7']
get_daily_date_new_format_vacunas, 210403, ../data/210404_cam_covid19.pdf
get_daily_date_new_format_vacunas, 210403, ../data/210404_cam_covid19.pdf
Unnamed: 2 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 2' 'Unnamed: 3' 'Hombres'
'Total'
 'Unnamed: 6' 'Unnamed: 7']
Unnamed: 3 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 3' 'Hombres' 'Total' 'Unnamed:
6'
 'Unnamed: 7']
Unnamed: 6 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 6' 'Unnamed:
7']
Unnamed: 7 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 7']
get_daily_date_new_format_vacunas, 210403, ../data/210404_cam_covid19.pdf
get_daily_date_new_format_vacunas, 210403, ../data/210404_cam_covid19.pdf
```

```
Unnamed: 2 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 2' 'Unnamed: 3' 'Hombres'
'Total'
 'Unnamed: 6' 'Unnamed: 7']
Unnamed: 3 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 3' 'Hombres' 'Total' 'Unnamed:
6'
 'Unnamed: 7']
Unnamed: 6 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 6' 'Unnamed:
7']
Unnamed: 7 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 7']
get_daily_date_new_format_vacunas, 210405, ../data/210404_cam_covid19.pdf
get_daily_date_new_format_vacunas, 210405, ../data/210404_cam_covid19.pdf
Unnamed: 2 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 2' 'Unnamed: 3' 'Hombres'
'Total'
 'Unnamed: 6' 'Unnamed: 7']
Unnamed: 3 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 3' 'Hombres' 'Total' 'Unnamed:
6'
 'Unnamed: 7']
Unnamed: 6 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 6' 'Unnamed:
7']
Unnamed: 7 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 7']
get_daily_date_new_format_vacunas, 210405, ../data/210404_cam_covid19.pdf
get_daily_date_new_format_vacunas, 210405, ../data/210404_cam_covid19.pdf
Unnamed: 2 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 2' 'Unnamed: 3' 'Hombres'
'Total'
 'Unnamed: 6' 'Unnamed: 7']
Unnamed: 3 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 3' 'Hombres' 'Total' 'Unnamed:
6'
 'Unnamed: 7']
Unnamed: 6 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 6' 'Unnamed:
7']
Unnamed: 7 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 7']
get_daily_date_new_format_vacunas, 210405, ../data/210404_cam_covid19.pdf
get_daily_date_new_format_vacunas, 210405, ../data/210404_cam_covid19.pdf
Unnamed: 2 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 2' 'Unnamed: 3' 'Hombres'
'Total'
 'Unnamed: 6' 'Unnamed: 7']
Unnamed: 3 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 3' 'Hombres' 'Total' 'Unnamed:
6'
 'Unnamed: 7']
Unnamed: 6 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 6' 'Unnamed:
7']
Unnamed: 7 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 7']
get_daily_date_new_format_vacunas, 210405, ../data/210404_cam_covid19.pdf
get_daily_date_new_format_vacunas, 210405, ../data/210404_cam_covid19.pdf
Unnamed: 2 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 2' 'Unnamed: 3' 'Hombres'
'Total'
 'Unnamed: 6' 'Unnamed: 7']
Unnamed: 3 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 3' 'Hombres' 'Total' 'Unnamed:
```

```
6'
 'Unnamed: 7']
Unnamed: 6 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 6' 'Unnamed:
7']
Unnamed: 7 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 7']
get_daily_date_new_format_vacunas, 210409, ../data/210404_cam_covid19.pdf
get_daily_date_new_format_vacunas, 210409, ../data/210404_cam_covid19.pdf
Unnamed: 2 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 2' 'Unnamed: 3' 'Hombres'
'Total'
 'Unnamed: 6' 'Unnamed: 7']
Unnamed: 3 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 3' 'Hombres' 'Total' 'Unnamed:
6'
 'Unnamed: 7']
Unnamed: 6 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 6' 'Unnamed:
7']
Unnamed: 7 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 7']
get_daily_date_new_format_vacunas, 210409, ../data/210404_cam_covid19.pdf
get_daily_date_new_format_vacunas, 210409, ../data/210404_cam_covid19.pdf
Unnamed: 2 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 2' 'Unnamed: 3' 'Hombres'
'Total'
 'Unnamed: 6' 'Unnamed: 7']
Unnamed: 3 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 3' 'Hombres' 'Total' 'Unnamed:
6'
 'Unnamed: 7']
Unnamed: 6 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 6' 'Unnamed:
7']
Unnamed: 7 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 7']
get_daily_date_new_format_vacunas, 210409, ../data/210404_cam_covid19.pdf
get_daily_date_new_format_vacunas, 210409, ../data/210404_cam_covid19.pdf
Unnamed: 2 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 2' 'Unnamed: 3' 'Hombres'
'Total'
 'Unnamed: 6' 'Unnamed: 7']
Unnamed: 3 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 3' 'Hombres' 'Total' 'Unnamed:
6'
 'Unnamed: 7']
Unnamed: 6 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 6' 'Unnamed:
7']
Unnamed: 7 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 7']
get_daily_date_new_format_vacunas, 210409, ../data/210404_cam_covid19.pdf
get_daily_date_new_format_vacunas, 210409, ../data/210404_cam_covid19.pdf
Unnamed: 2 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 2' 'Unnamed: 3' 'Hombres'
'Total'
 'Unnamed: 6' 'Unnamed: 7']
Unnamed: 3 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 3' 'Hombres' 'Total' 'Unnamed:
6'
 'Unnamed: 7']
Unnamed: 6 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 6' 'Unnamed:
7']
```

```
Unnamed: 7 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 7']
get_daily_date_new_format_vacunas, 210404, ../data/210404_cam_covid19.pdf
get_daily_date_new_format_vacunas, 210404, ../data/210404_cam_covid19.pdf
Unnamed: 2 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 2' 'Unnamed: 3' 'Hombres'
'Total'
 'Unnamed: 6' 'Unnamed: 7']
Unnamed: 3 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 3' 'Hombres' 'Total' 'Unnamed:
6'
 'Unnamed: 7']
Unnamed: 6 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 6' 'Unnamed:
7']
Unnamed: 7 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 7']
get_daily_date_new_format_vacunas, 210404, ../data/210404_cam_covid19.pdf
get_daily_date_new_format_vacunas, 210404, ../data/210404_cam_covid19.pdf
Unnamed: 2 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 2' 'Unnamed: 3' 'Hombres'
'Total'
 'Unnamed: 6' 'Unnamed: 7']
Unnamed: 3 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 3' 'Hombres' 'Total' 'Unnamed:
6'
 'Unnamed: 7']
Unnamed: 6 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 6' 'Unnamed:
7']
Unnamed: 7 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 7']
get_daily_date_new_format_vacunas, 210404, ../data/210404_cam_covid19.pdf
get_daily_date_new_format_vacunas, 210404, ../data/210404_cam_covid19.pdf
Unnamed: 2 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 2' 'Unnamed: 3' 'Hombres'
'Total'
 'Unnamed: 6' 'Unnamed: 7']
Unnamed: 3 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 3' 'Hombres' 'Total' 'Unnamed:
6'
 'Unnamed: 7']
Unnamed: 6 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 6' 'Unnamed:
7']
Unnamed: 7 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 7']
get_daily_date_new_format_vacunas, 210404, ../data/210404_cam_covid19.pdf
get_daily_date_new_format_vacunas, 210404, ../data/210404_cam_covid19.pdf
Unnamed: 2 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 2' 'Unnamed: 3' 'Hombres'
'Total'
 'Unnamed: 6' 'Unnamed: 7']
Unnamed: 3 ['Unnamed: 0' 'Edad Mujeres' 'Unnamed: 3' 'Hombres' 'Total' 'Unnamed:
6'
 'Unnamed: 7']
Unnamed: 6 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 6' 'Unnamed:
7']
Unnamed: 7 ['Unnamed: 0' 'Edad Mujeres' 'Hombres' 'Total' 'Unnamed: 7']
```

[79]:
```
df_ages.query( 'Ages=="80-89"')
```

[79]:

	Ages	Edad Mujeres	Male	Total	Fecha
9	80-89	4.102	4.874	8.976	2021-04-01
9	80-89	4.102	4.874	8.976	2021-04-01
9	80-89	4.102	4.874	8.976	2021-04-01
9	80-89	4.102	4.874	8.976	2021-04-01
9	80-89	4.118	4.889	9.007	2021-04-06
9	80-89	4.118	4.889	9.007	2021-04-06
9	80-89	4.118	4.889	9.007	2021-04-06
9	80-89	4.118	4.889	9.007	2021-04-06
9	80-89	4.103	4.878	8.981	2021-04-02
9	80-89	4.103	4.878	8.981	2021-04-02
9	80-89	4.103	4.878	8.981	2021-04-02
9	80-89	4.103	4.878	8.981	2021-04-02
9	80-89	4.121	4.893	9.014	2021-04-08
9	80-89	4.121	4.893	9.014	2021-04-08
9	80-89	4.121	4.893	9.014	2021-04-08
9	80-89	4.121	4.893	9.014	2021-04-08
9	80-89	4.121	4.893	9.014	2021-04-07
9	80-89	4.121	4.893	9.014	2021-04-07
9	80-89	4.121	4.893	9.014	2021-04-07
9	80-89	4.121	4.893	9.014	2021-04-07
9	80-89	4.129	4.903	9.032	2021-04-10
9	80-89	4.129	4.903	9.032	2021-04-10
9	80-89	4.129	4.903	9.032	2021-04-10
9	80-89	4.129	4.903	9.032	2021-04-10
9	80-89	4.103	4.878	8.981	2021-04-03
9	80-89	4.103	4.878	8.981	2021-04-03
9	80-89	4.103	4.878	8.981	2021-04-03
9	80-89	4.103	4.878	8.981	2021-04-03
9	80-89	4.118	4.889	9.007	2021-04-05
9	80-89	4.118	4.889	9.007	2021-04-05
9	80-89	4.118	4.889	9.007	2021-04-05
9	80-89	4.118	4.889	9.007	2021-04-05
9	80-89	4.129	4.903	9.032	2021-04-09
9	80-89	4.129	4.903	9.032	2021-04-09
9	80-89	4.129	4.903	9.032	2021-04-09
9	80-89	4.129	4.903	9.032	2021-04-09
9	80-89	4.103	4.878	8.981	2021-04-04
9	80-89	4.103	4.878	8.981	2021-04-04
9	80-89	4.103	4.878	8.981	2021-04-04
9	80-89	4.103	4.878	8.981	2021-04-04

[ ]:
```
total
VENTANA_MEDIA_MOVIL=7
df = interpolate_dataframe(total,'D')
```

```
df.index.name = 'Fecha'
df = df.sort_values(by=['Fecha'], ascending=True)
df['HOSPITALES hoy'] = df['HOSPITALES'] - df['HOSPITALES'].shift(1)
df['CENTROS SOCIOSANITARIOS hoy'] = df['CENTROS SOCIOSANITARIOS'] - df['CENTROS␣
↪SOCIOSANITARIOS'].shift(1)
df['FALLECIDOS TOTALES hoy'] = df['FALLECIDOS TOTALES'] - df['FALLECIDOS␣
↪TOTALES'].shift(1)


df['MA CENTROS SOCIOSANITARIOS hoy'] = df['CENTROS SOCIOSANITARIOS hoy'].
↪rolling(window=VENTANA_MEDIA_MOVIL).mean()
df['MA HOSPITALES hoy'] = df['HOSPITALES hoy'].
↪rolling(window=VENTANA_MEDIA_MOVIL).mean()
df['MA FALLECIDOS TOTALES hoy'] = df['FALLECIDOS TOTALES hoy'].
↪rolling(window=VENTANA_MEDIA_MOVIL).mean()

df = df.sort_index(ascending=False)
df_master = df.copy()
```

[ ]:
```
total.head()
```

[ ]:
```
# Hacemos lo contrario
# En lugar de sacar el nº de muertos dado el nº de infectados, como lo primero␣
↪lo sabemos (en madrid), sacamos lo segundo y extrapolamos al conjunto de␣
↪españa
df = df_master

R0_estimada = df['FALLECIDOS TOTALES hoy'].values[0:7].sum()  / df['FALLECIDOS␣
↪TOTALES hoy'].values[7:14].sum()
print(df['FALLECIDOS TOTALES hoy'].values[0:7].sum(), df['FALLECIDOS TOTALES␣
↪hoy'].values[7:14].sum() )
print(f"""R0_estimada = {R0_estimada}""")
PROPORCION_ENFERMOS_MUERTOS=750000/15000 # Esta es la proporcion enfermos␣
↪muertos (15.000 muertos para 750.000 afectados)
RATIO_NO_HEMOS_COLAPSADO=2 # La mitad de los muertos se ha calculado del␣
↪colapso. Como ahora no hemos colapsado
PESO_MADRID_MUERTES_TOTALES=1/3
casos_españa_estimados =  df['FALLECIDOS TOTALES hoy'].values[0:5].sum() *␣
↪PROPORCION_ENFERMOS_MUERTOS * RATIO_NO_HEMOS_COLAPSADO /␣
↪PESO_MADRID_MUERTES_TOTALES
print(f"""casos_españa_estimados = {casos_españa_estimados}""")
```

1.1 Gráfico estimacion R0

Considerando solo los datos de Madrid, estimamos el R0 a partir del nº de muertos (considerando que el nº de muertos es una combinacion lineal del nº de enfermos), por lo que es posible calcular

el ratio igual.

Para calcular el R0, sacamos la suma de muertos de la última semana, entre la suma de muertos de la semana anterior.

[ ]:
```
from datetime import datetime, timedelta
import seaborn as sns
from matplotlib import pyplot as plt
import matplotlib.dates as mdates

df = df_master

def calcular_estimaciones_R0(df):
    def calcular_R0_dia(dia,df):
        dia_semana_anterior = dia - timedelta(days=7)
        return dia,df.loc[dia:dia - timedelta(days=6)]['FALLECIDOS TOTALES␣
 ↪hoy'].sum() / df.loc[dia- timedelta(days=7):dia -␣
 ↪timedelta(days=13)]['FALLECIDOS TOTALES hoy'].sum()

    VENTANA_MEDIA_MOVIL=7

    df_R0_estimada = pd.DataFrame([calcular_R0_dia(dia,df) for dia in df.
 ↪index[0:50]],columns=['Fecha','R0_estimada'])

    df_R0_estimada = df_R0_estimada.sort_values(by=['Fecha'], ascending=True)
    df_R0_estimada['MA R0_estimada'] = df_R0_estimada['R0_estimada'].
 ↪rolling(window=VENTANA_MEDIA_MOVIL).mean()
    df_R0_estimada = df_R0_estimada.sort_values(by=['Fecha'], ascending=False)
    df_R0_estimada.set_index('Fecha', inplace=True, drop=True)
    return df_R0_estimada

df= calcular_estimaciones_R0(df_master)
#df=df[['R0_estimada']]
df

chart_df=df[df.columns[-3:]]
chart_df.plot(legend=True,figsize=(13.5,9), marker='o')

plt.gca().xaxis.set_major_formatter(mdates.DateFormatter('%m-%d'))
plt.gca().xaxis.set_major_locator(mdates.DayLocator(interval=1))
plt.xticks(rotation=45)

ax = plt.gca()
ax.axhline(1, color='r',linestyle =  ':' )


ax.set_title("Estimacion R0 Comunidad de Madrid")
ax.set_ylim(ymin=0)
```

```
plt.show()

df.style.format ({ c : "{:20,.3f}" for c in df.columns }).
 ↪background_gradient(cmap='Wistia', )
```

[ ]:
```
RO_estimada * 1.2
```

[ ]:
```
HTML("<h2>Gráfico muertes diarias en Madrid, según Comunidad de Madrid </h2>")
```

[ ]:
```
import pandas as pd
import io
import matplotlib.dates as mdates
from matplotlib import pyplot as plt

df = df_master
chart_df=df[df.columns[-3:]].head(60)
chart_df.plot(legend=True,figsize=(13.5,9), marker='o')


plt.gca().xaxis.set_major_formatter(mdates.DateFormatter('%b-%d'))
plt.gca().xaxis.set_major_locator(mdates.DayLocator(interval=7))
plt.xticks(rotation=45)

ax = plt.gca()
plt.setp(ax.get_xminorticklabels(), visible=False)

ax.set_title("Muertes diarias COVID 19, media movil␣
 ↪"+str(VENTANA_MEDIA_MOVIL)+" dias. Fuente: Comunidad de Madrid")
ax.set_ylim(ymin=0)

plt.show()
```

[ ]:
```
 from IPython.display import display, HTML
HTML("<h2>Comparamos los datos de hoy, de hace una semana y de un mes </h2>")
```

[ ]:
```
from matplotlib import colors

def background_gradient(s, m, M, cmap='PuBu', low=0, high=0):
    rng = M - m
    norm = colors.Normalize(m - (rng * low),
                            M + (rng * high))
    normed = norm(s.values)
    c = [colors.rgb2hex(x) for x in plt.cm.get_cmap(cmap)(normed)]
    return ['background-color: %s' % color for color in c]

df = df_master
```

```
df.style.format ({ c : "{:20,.0f}" for c in df.columns }).
 ↪background_gradient(cmap='Wistia', subset= df.columns[-3:] )
```

[ ]:
```
df = df_master
pd.concat([df.head(1).tail(1) , df.head(8).tail(1) , df.head(30).tail(1)]).
 ↪astype(int)[['MA HOSPITALES hoy','MA CENTROS SOCIOSANITARIOS hoy','MA␣
 ↪FALLECIDOS TOTALES hoy']].style.format ({ c : "{:20,.0f}" for c in df.
 ↪columns }).background_gradient(cmap='Wistia', subset= df.columns[-3:] )
```

[ ]:
```
from IPython.display import display, HTML
HTML("<h2>Muertes medias diarias, últimos 7 días, con datos</h2>")
```

[ ]:
```
from datetime import date

df = df_master
inicio_crisis = df.head(7).index[6]
df=df.head(7)
dia_mas_reciente = df.index[0]
dias_transcurridos_inicio_crisis = dia_mas_reciente - inicio_crisis
df = pd.DataFrame((df.head(1).max(axis=0) - df.tail(1).max(axis=0) ) / ␣
 ↪dias_transcurridos_inicio_crisis.days  ).
 ↪T[['HOSPITALES','DOMICILIOS','CENTROS SOCIOSANITARIOS','OTROS␣
 ↪LUGARES','FALLECIDOS TOTALES']]
df.style.format ({ c : "{:20,.0f}" for c in df.columns }).
 ↪background_gradient(cmap='Wistia'  )
```

[ ]:
```
HTML("<h2>Muertes medias diarias desde que la comunidad de Madrid publica␣
 ↪datos</h2>")
```

[ ]:
```
# Calculamos los incrementos medios, desde que tenemos fechas
df = df_master
df = pd.DataFrame((df.head(1).max(axis=0) - df.tail(1).max(axis=0) ) / df.
 ↪shape[0] ).T[['HOSPITALES','DOMICILIOS','CENTROS SOCIOSANITARIOS','OTROS␣
 ↪LUGARES','FALLECIDOS TOTALES']]
df.style.format ({ c : "{:20,.0f}" for c in df.columns }).
 ↪background_gradient(cmap='Wistia'  )
```

[ ]:
```

```

[ ]:
```
from tabula import read_pdf
from IPython.display import display, HTML
import os
import pandas as pd
import glob
import re
```

```
from tqdm.notebook import tqdm
import warnings
import os.path
fecha="201005"
import os
file_path = '../data/'+fecha+'_cam_covid19.pdf'
if not os.path.isfile(file_path):
    file_path = '../data/'+fecha+'cam_covid19.pdf'
#print("Analizando:" + file_path)
```

[ ]:
```
df_pdf =  read_pdf(file_path,area=(300, 100, 800, 400) , pages='1')
df_pdf
```

[ ]:
```
for x,y in enumerate(df_pdf):
    print(x,"::",y)

pd.DataFrame(df_pdf)
```

[ ]:
```
type(df_pdf)
```

[ ]:
```
type(df_pdf[0])
```

[ ]:
```
total
```

[ ]:
```
get_daily_date_new_format("201005")
```

[ ]:
```
total
```

[91]:
```
#get_daily_date_new_format_vacunas( "210313", "../data/210313_cam_covid19.pdf")
```

[13]:
```
fecha = '210307'
file_path = f"../data/{fecha}_cam_covid19.pdf"

dict={}
PAGINA_DE_DATOS=2
df2_pdf =  read_pdf(file_path,area=(300, 100, 800, 400) , pages=PAGINA_DE_DATOS)
dict['PACIENTES UCI DIA']       = df2_pdf[0].loc[3:3].values[0][1].replace(".
 ↪",'')
dict['PACIENTES UCI ACUMULADOS']= df2_pdf[0].loc[6:6].values[0][1].replace(".
 ↪",'')
dict
```

```
Got stderr: mar 13, 2021 7:52:53 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'
```

[13]: {'PACIENTES UCI DIA': '511', 'PACIENTES UCI ACUMULADOS': '9447'}

[ ]: