Madrid_Pain_Graphs

November 13, 2020

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.

```
[1]: # 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 2 | 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 $FILE ; done
     ! cd ../data/; FILELIST=`seq -w 0 2 | while read i ; do date +%y%m%d -d "$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 2 | 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 $FILE ; done
! cd ../data/; FILELIST=`seq -w 0 2 | while read i ; do date +%Y%m%d -d "$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 $FILE ; done
! cd ../data/; FILELIST=`seq -w 0 2 | while read i ; do date +%y%m%d -d "$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  2>/dev/null && ls -altr $FILE ; done
#200902_cam_covid19.pdf
```

```
201113cam_covid19.pdf:::::
201112cam_covid19.pdf:::::
201111cam_covid19.pdf:::::
20201113_cam_covid19.pdf:::::
20201112_cam_covid19.pdf:::::
20201111_cam_covid19.pdf:::::
```

[2]:
```python
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

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

""" Rellenar dias vacios con interpolacion"""
def interpolate_dataframe(df,freq):
    if freq == 'H':
```

```python
        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(fecha):

    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=(000, 600, 400, 800) , pages='1')

    #print("1 get_daily_date_new_format ",fecha)

    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='1')
        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:
```

```python
            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")

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

    if fecha > '200512' :
        return get_daily_date_new_format(fecha)

    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)
```

```python
    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= sorted(glob.glob('../data/*_covid19.pdf'),
                     key=os.path.getmtime,
                     reverse=True )


    for pdf_file in tqdm(pdf_list,
                         desc="Procesando pdfs diarios"):
        # extract fecha from username , eg : ../data/2200422_cam_covid19.pdf
        fecha = pdf_file.split('/')[2].split('_')[0].replace('cam_','').
replace('_cam_','').replace('cam','')
        if fecha not in BLACKLIST:
            #print("processing", fecha)
            df = get_daily_data(fecha)
        list_df.append(df)

    df = pd.concat(list_df)
```

```python
    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=171.0, style

Got stderr: nov 13, 2020 4:44:14 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:44:16 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:44:17 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:44:19 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:44:21 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:44:23 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

```
Got stderr: nov 13, 2020 4:44:25 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:44:26 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:44:28 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:44:29 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:44:31 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:44:33 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:44:34 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:44:36 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:44:37 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:44:39 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:44:41 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:44:42 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'
```

```
Got stderr: nov 13, 2020 4:44:44 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:44:45 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:44:47 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:44:48 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:44:50 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:44:52 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:44:53 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:44:55 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:44:56 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:44:58 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:44:59 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:45:01 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'
```

```
Got stderr: nov 13, 2020 4:45:03 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:45:04 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:45:06 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:45:07 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:45:09 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:45:10 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:45:12 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:45:13 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:45:14 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:45:16 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:45:17 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:45:19 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'
```

```
Got stderr: nov 13, 2020 4:45:21 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:45:23 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:45:24 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:45:25 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:45:27 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:45:28 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:45:30 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:45:32 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:45:33 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:45:35 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:45:36 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:45:37 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'
```

```
Got stderr: nov 13, 2020 4:45:39 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:45:41 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:45:42 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:45:44 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:45:45 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:45:47 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:45:48 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:45:50 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:45:51 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:45:53 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:45:54 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:45:56 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'
```

```
Got stderr: nov 13, 2020 4:45:58 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:45:59 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:46:01 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:46:03 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:46:04 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:46:06 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:46:08 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:46:09 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'


200519 mal parseada: index 1 is out of bounds for axis 0 with size 1
200518 mal parseada: index 1 is out of bounds for axis 0 with size 1
200517 mal parseada: index 1 is out of bounds for axis 0 with size 1
200516 mal parseada: index 1 is out of bounds for axis 0 with size 1
200515 mal parseada: index 1 is out of bounds for axis 0 with size 1
200514 mal parseada: index 1 is out of bounds for axis 0 with size 1
200513 mal parseada: index 1 is out of bounds for axis 0 with size 1

Got stderr: nov 13, 2020 4:52:11 PM
org.apache.pdfbox.pdmodel.font.PDCIDFontType2 <init>
INFORMACIÓN: OpenType Layout tables used in font CIDFont+F1 are not implemented
in PDFBox and will be ignored
nov 13, 2020 4:52:11 PM org.apache.pdfbox.pdmodel.font.PDCIDFontType2 <init>
INFORMACIÓN: OpenType Layout tables used in font CIDFont+F2 are not implemented
in PDFBox and will be ignored
nov 13, 2020 4:52:11 PM org.apache.pdfbox.pdmodel.font.PDCIDFontType2 <init>
```

```
INFORMACIÓN: OpenType Layout tables used in font CIDFont+F3 are not implemented
in PDFBox and will be ignored
nov 13, 2020 4:52:12 PM org.apache.pdfbox.pdmodel.font.PDCIDFontType2 <init>
INFORMACIÓN: OpenType Layout tables used in font CIDFont+F1 are not implemented
in PDFBox and will be ignored
nov 13, 2020 4:52:12 PM org.apache.pdfbox.pdmodel.font.PDCIDFontType2 <init>
INFORMACIÓN: OpenType Layout tables used in font CIDFont+F2 are not implemented
in PDFBox and will be ignored
nov 13, 2020 4:52:12 PM org.apache.pdfbox.pdmodel.font.PDCIDFontType2 <init>
INFORMACIÓN: OpenType Layout tables used in font CIDFont+F3 are not implemented
in PDFBox and will be ignored
nov 13, 2020 4:52:12 PM org.apache.pdfbox.pdmodel.font.PDCIDFontType2 <init>
INFORMACIÓN: OpenType Layout tables used in font CIDFont+F1 are not implemented
in PDFBox and will be ignored
nov 13, 2020 4:52:12 PM org.apache.pdfbox.pdmodel.font.PDCIDFontType2 <init>
INFORMACIÓN: OpenType Layout tables used in font CIDFont+F2 are not implemented
in PDFBox and will be ignored
nov 13, 2020 4:52:12 PM org.apache.pdfbox.pdmodel.font.PDCIDFontType2 <init>
INFORMACIÓN: OpenType Layout tables used in font CIDFont+F3 are not implemented
in PDFBox and will be ignored
```

[2]:

Fecha	CENTROS SOCIOSANITARIOS	DOMICILIOS	FALLECIDOS TOTALES \
2020-11-13	4935	1125	18438
2020-11-12	4932	1121	18369
2020-11-11	4932	1121	18369
2020-11-10	4932	1121	18369
2020-11-09	4929	1116	18199
…	…	…	…
2020-04-26	4236	798	12855
2020-04-25	4170	788	12612
2020-04-24	4068	775	12252
2020-04-23	3996	769	12056
2020-04-22	3932	761	11852

Fecha	HOSPITALES	OTROS LUGARES	PACIENTES UCI ACUMULADOS \
2020-11-13	12348	30	6230
2020-11-12	12286	30	6205
2020-11-11	12286	30	6186
2020-11-10	12286	30	6161
2020-11-09	12124	30	6143
…	…	…	…
2020-04-26	7800	21	0
2020-04-25	7633	21	0

```
2020-04-24       7388          21                      0
2020-04-23       7271          20                      0
2020-04-22       7144          15                      0


            PACIENTES UCI DIA  HOSPITALES hoy  CENTROS SOCIOSANITARIOS hoy  \
Fecha
2020-11-13                428            62.0                          3.0
2020-11-12                428             0.0                          0.0
2020-11-11                437             0.0                          0.0
2020-11-10                447           162.0                          3.0
2020-11-09                451             0.0                          0.0
...                       ...             ...                          ...
2020-04-26                  0           167.0                         66.0
2020-04-25                  0           245.0                        102.0
2020-04-24                  0           117.0                         72.0
2020-04-23                  0           127.0                         64.0
2020-04-22                  0             NaN                          NaN


            FALLECIDOS TOTALES hoy
Fecha
2020-11-13                    69.0
2020-11-12                     0.0
2020-11-11                     0.0
2020-11-10                   170.0
2020-11-09                     0.0
...                            ...
2020-04-26                   243.0
2020-04-25                   360.0
2020-04-24                   196.0
2020-04-23                   204.0
2020-04-22                     NaN

[158 rows x 10 columns]
```

[3]: `interpolate_dataframe(total,'D')`

```
[3]:           CENTROS SOCIOSANITARIOS  DOMICILIOS  FALLECIDOS TOTALES  \
    2020-04-22                   3932.0       761.0             11852.0
    2020-04-23                   3996.0       769.0             12056.0
    2020-04-24                   4068.0       775.0             12252.0
    2020-04-25                   4170.0       788.0             12612.0
    2020-04-26                   4236.0       798.0             12855.0
    ...                             ...         ...                 ...
    2020-11-09                   4929.0      1116.0             18199.0
    2020-11-10                   4932.0      1121.0             18369.0
    2020-11-11                   4932.0      1121.0             18369.0
    2020-11-12                   4932.0      1121.0             18369.0
```

```
2020-11-13                       4935.0        1125.0                 18438.0

             HOSPITALES  OTROS LUGARES  PACIENTES UCI ACUMULADOS  \
2020-04-22      7144.0           15.0                       0.0
2020-04-23      7271.0           20.0                       0.0
2020-04-24      7388.0           21.0                       0.0
2020-04-25      7633.0           21.0                       0.0
2020-04-26      7800.0           21.0                       0.0
…                   …              …                         …
2020-11-09     12124.0           30.0                    6143.0
2020-11-10     12286.0           30.0                    6161.0
2020-11-11     12286.0           30.0                    6186.0
2020-11-12     12286.0           30.0                    6205.0
2020-11-13     12348.0           30.0                    6230.0

             PACIENTES UCI DIA  HOSPITALES hoy  CENTROS SOCIOSANITARIOS hoy  \
2020-04-22                0.0             NaN                          NaN
2020-04-23                0.0           127.0                         64.0
2020-04-24                0.0           117.0                         72.0
2020-04-25                0.0           245.0                        102.0
2020-04-26                0.0           167.0                         66.0
…                           …              …                            …
2020-11-09              451.0             0.0                          0.0
2020-11-10              447.0           162.0                          3.0
2020-11-11              437.0             0.0                          0.0
2020-11-12              428.0             0.0                          0.0
2020-11-13              428.0            62.0                          3.0

             FALLECIDOS TOTALES hoy
2020-04-22                     NaN
2020-04-23                   204.0
2020-04-24                   196.0
2020-04-25                   360.0
2020-04-26                   243.0
…                              …
2020-11-09                     0.0
2020-11-10                   170.0
2020-11-11                     0.0
2020-11-12                     0.0
2020-11-13                    69.0

[206 rows x 10 columns]
```

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

```
df
```

[4]:

```
            CENTROS SOCIOSANITARIOS  DOMICILIOS  FALLECIDOS TOTALES  \
Fecha
2020-11-13                    4935        1125               18438
2020-11-12                    4932        1121               18369
2020-11-11                    4932        1121               18369
2020-11-10                    4932        1121               18369
2020-11-09                    4929        1116               18199
...                            ...         ...                 ...
2020-04-26                    4236         798               12855
2020-04-25                    4170         788               12612
2020-04-24                    4068         775               12252
2020-04-23                    3996         769               12056
2020-04-22                    3932         761               11852


            HOSPITALES  OTROS LUGARES  PACIENTES UCI ACUMULADOS  \
Fecha
2020-11-13       12348             30                      6230
2020-11-12       12286             30                      6205
2020-11-11       12286             30                      6186
2020-11-10       12286             30                      6161
2020-11-09       12124             30                      6143
...                ...            ...                       ...
2020-04-26        7800             21                         0
2020-04-25        7633             21                         0
2020-04-24        7388             21                         0
2020-04-23        7271             20                         0
2020-04-22        7144             15                         0


            PACIENTES UCI DIA  HOSPITALES hoy  CENTROS SOCIOSANITARIOS hoy  \
Fecha
2020-11-13                428              62                            3
2020-11-12                428               0                            0
2020-11-11                437               0                            0
2020-11-10                447             162                            3
2020-11-09                451               0                            0
...                       ...             ...                          ...
2020-04-26                  0             167                           66
2020-04-25                  0             245                          102
2020-04-24                  0             117                           72
2020-04-23                  0             127                           64
2020-04-22                  0               0                            0


            FALLECIDOS TOTALES hoy
Fecha
2020-11-13                      69
```

16

```
2020-11-12                        0
2020-11-11                        0
2020-11-10                      170
2020-11-09                        0
…                                 …
2020-04-26                      243
2020-04-25                      360
2020-04-24                      196
2020-04-23                      204
2020-04-22                        0

[158 rows x 10 columns]
```

[5]:
```python
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()
```

[6]:
```python
total.head()
```

[6]:
```
             CENTROS SOCIOSANITARIOS  DOMICILIOS  FALLECIDOS TOTALES  \
Fecha
2020-11-13                      4935        1125                18438
2020-11-12                      4932        1121                18369
2020-11-11                      4932        1121                18369
2020-11-10                      4932        1121                18369
2020-11-09                      4929        1116                18199


             HOSPITALES  OTROS LUGARES  PACIENTES UCI ACUMULADOS  \
Fecha
```

```
2020-11-13        12348              30                        6230
2020-11-12        12286              30                        6205
2020-11-11        12286              30                        6186
2020-11-10        12286              30                        6161
2020-11-09        12124              30                        6143
```

```
            PACIENTES UCI DIA   HOSPITALES hoy   CENTROS SOCIOSANITARIOS hoy  \
Fecha
2020-11-13                428             62.0                           3.0
2020-11-12                428              0.0                           0.0
2020-11-11                437              0.0                           0.0
2020-11-10                447            162.0                           3.0
2020-11-09                451              0.0                           0.0
```

```
            FALLECIDOS TOTALES hoy
Fecha
2020-11-13                    69.0
2020-11-12                     0.0
2020-11-11                     0.0
2020-11-10                   170.0
2020-11-09                     0.0
```

[7]:
```python
# 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}""")
```

```
239.0 300.0
R0_estimada = 0.7966666666666666
casos_españa_estimados = 71700.0
```

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.

```
[8]: 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', )
```
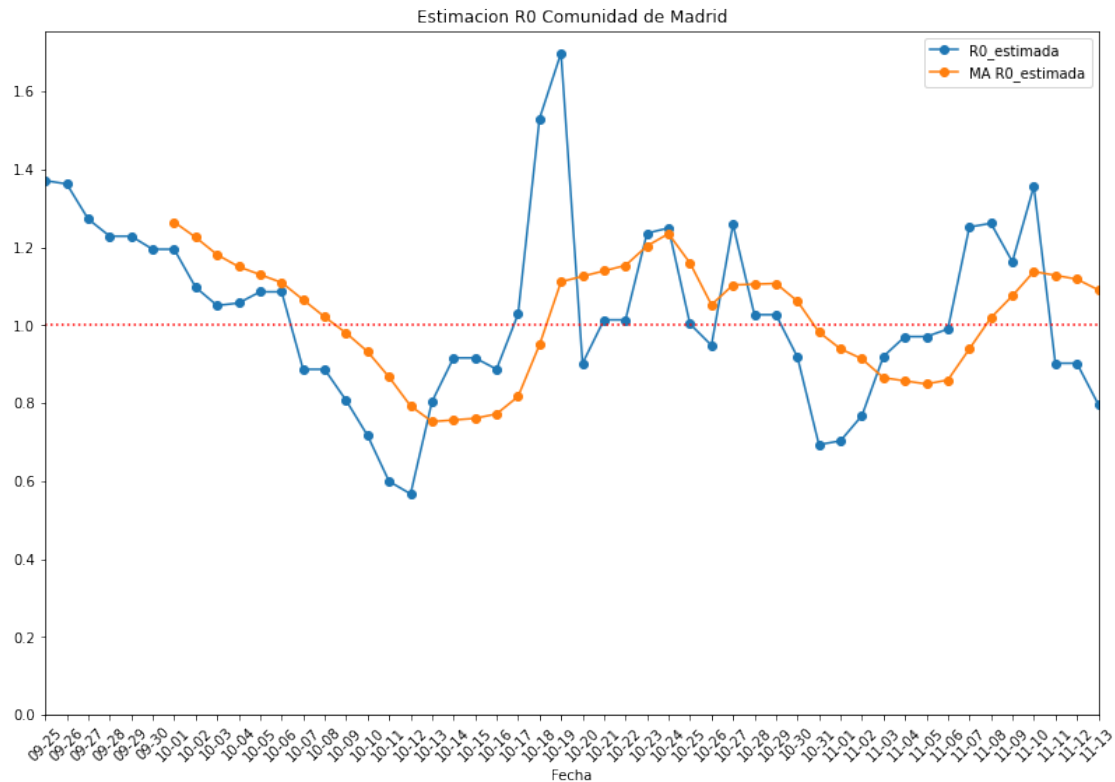

[8]: `<pandas.io.formats.style.Styler at 0x7f4520de6748>`

[9]: ```
R0_estimada * 1.2
```

[9]: `0.956`

[10]: ```
HTML("<h2>Gráfico muertes diarias en Madrid, según Comunidad de Madrid </h2>")
```

[10]: `<IPython.core.display.HTML object>`

[11]: ```
import pandas as pd
import io
```

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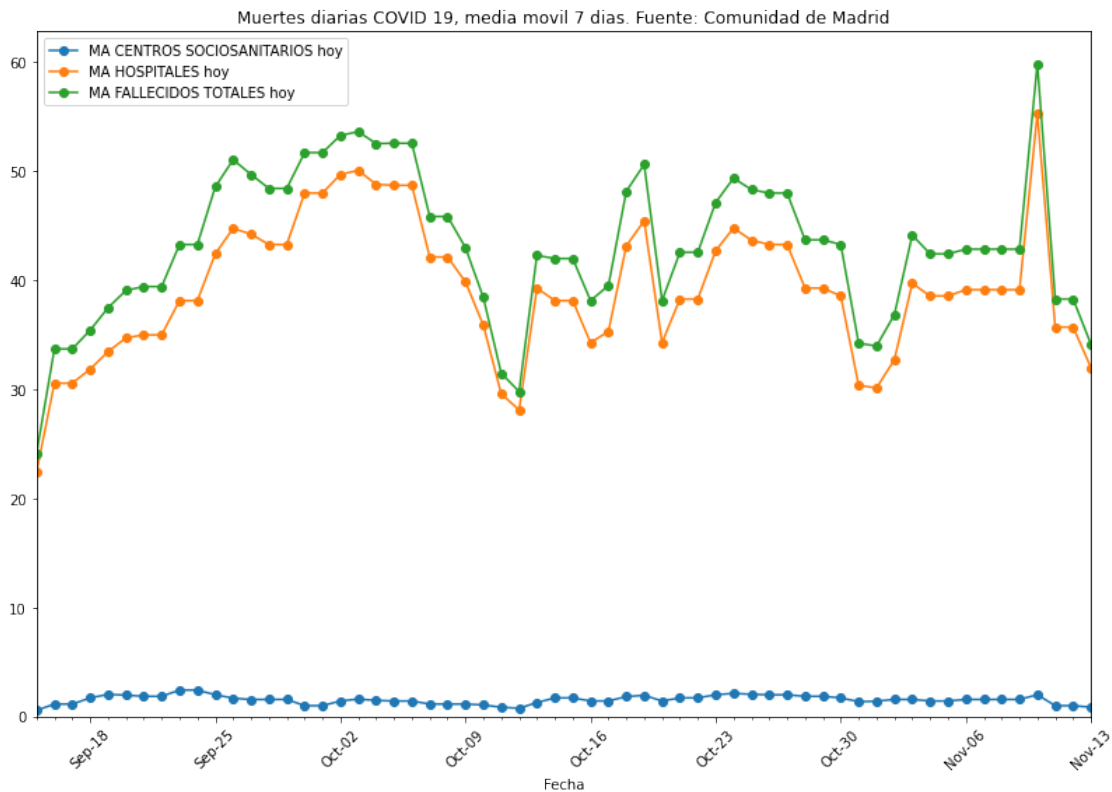
ax.set_title("Muertes diarias COVID 19, media movil␣
 ↪"+str(VENTANA_MEDIA_MOVIL)+" dias. Fuente: Comunidad de Madrid")
ax.set_ylim(ymin=0)

plt.show()
```

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

```
[12]: <IPython.core.display.HTML object>
```

```
[13]: 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:] )
```

```
[13]: <pandas.io.formats.style.Styler at 0x7f451f332cf8>
```

```
[14]: 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:] )
```

```
[14]: <pandas.io.formats.style.Styler at 0x7f451f23f6a0>
```

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

```
[15]: <IPython.core.display.HTML object>
```

```
[16]: 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'  )
```

[16]: <pandas.io.formats.style.Styler at 0x7f451f2ae6d8>

[17]: ```
HTML("<h2>Muertes medias diarias desde que la comunidad de Madrid publica
 ↪datos</h2>")
```

[17]: <IPython.core.display.HTML object>

[18]: ```
# 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')
```

[18]: <pandas.io.formats.style.Styler at 0x7f451f261dd8>

[ ]:

[19]: ```
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)
```

[20]: ```
df_pdf = read_pdf(file_path,area=(300, 100, 800, 400) , pages='1')
df_pdf
```

Got stderr: nov 13, 2020 4:52:19 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

```
[20]: [                           Unnamed: 0         256.667
    0                                  NaN     (acumulados)
    1                           Pacie ntes    Pacientes en
    2                        Hospitalizados            UCI
    3                                3.047            501
    4                        (ingresados en   (ingresados en
    5                              el dia)         el día)
    6                               59.757          5.234
    7                          (acumulados)            NaN
    8                                  NaN    (acumulados)
    9                                  NaN            NaN
    10  e: Dirección General de Salud Pública, Servici…            NaN]
```

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

      pd.DataFrame(df_pdf)
```

```
0 ::                              Unnamed: 0         256.667
0                                      NaN     (acumulados)
1                               Pacie ntes    Pacientes en
2                            Hospitalizados            UCI
3                                    3.047            501
4                            (ingresados en   (ingresados en
5                                  el dia)         el día)
6                                   59.757          5.234
7                              (acumulados)            NaN
8                                      NaN    (acumulados)
9                                      NaN            NaN
10  e: Dirección General de Salud Pública, Servici…            NaN
```

```
[21]:                                        0
     0                                    Unn…
```

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

```
[22]: list
```

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

```
[23]: pandas.core.frame.DataFrame
```

```python
[24]: total
```

```
[24]:            CENTROS SOCIOSANITARIOS  DOMICILIOS  FALLECIDOS TOTALES  \
     Fecha
     2020-11-13                     4935        1125              18438
```

2020-11-12	4932	1121	18369
2020-11-11	4932	1121	18369
2020-11-10	4932	1121	18369
2020-11-09	4929	1116	18199
...	...	...	...
2020-04-26	4236	798	12855
2020-04-25	4170	788	12612
2020-04-24	4068	775	12252
2020-04-23	3996	769	12056
2020-04-22	3932	761	11852

	HOSPITALES	OTROS LUGARES	PACIENTES UCI ACUMULADOS \
Fecha			
2020-11-13	12348	30	6230
2020-11-12	12286	30	6205
2020-11-11	12286	30	6186
2020-11-10	12286	30	6161
2020-11-09	12124	30	6143
...	...	...	...
2020-04-26	7800	21	0
2020-04-25	7633	21	0
2020-04-24	7388	21	0
2020-04-23	7271	20	0
2020-04-22	7144	15	0

	PACIENTES UCI DIA	HOSPITALES hoy	CENTROS SOCIOSANITARIOS hoy \
Fecha			
2020-11-13	428	62.0	3.0
2020-11-12	428	0.0	0.0
2020-11-11	437	0.0	0.0
2020-11-10	447	162.0	3.0
2020-11-09	451	0.0	0.0
...	...	...	...
2020-04-26	0	167.0	66.0
2020-04-25	0	245.0	102.0
2020-04-24	0	117.0	72.0
2020-04-23	0	127.0	64.0
2020-04-22	0	NaN	NaN

	FALLECIDOS TOTALES hoy
Fecha	
2020-11-13	69.0
2020-11-12	0.0
2020-11-11	0.0
2020-11-10	170.0
2020-11-09	0.0
...	...

```
2020-04-26                          243.0
2020-04-25                          360.0
2020-04-24                          196.0
2020-04-23                          204.0
2020-04-22                            NaN

[158 rows x 10 columns]
```

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

```
Got stderr: nov 13, 2020 4:52:20 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'

Got stderr: nov 13, 2020 4:52:22 PM
org.apache.pdfbox.pdmodel.font.PDTrueTypeFont <init>
ADVERTENCIA: Using fallback font 'LiberationSans' for 'Arial,Bold'
```

[25]:

Fecha	PACIENTES UCI DIA	PACIENTES UCI ACUMULADOS	HOSPITALES	DOMICILIOS
2020-10-05	501	5234	10803	1035

Fecha	CENTROS SOCIOSANITARIOS	OTROS LUGARES FALLECIDOS TOTALES
2020-10-05	4874	30 16742

[26]: `total`

[26]:

Fecha	CENTROS SOCIOSANITARIOS	DOMICILIOS	FALLECIDOS TOTALES
2020-11-13	4935	1125	18438
2020-11-12	4932	1121	18369
2020-11-11	4932	1121	18369
2020-11-10	4932	1121	18369
2020-11-09	4929	1116	18199
…	…	…	…
2020-04-26	4236	798	12855
2020-04-25	4170	788	12612
2020-04-24	4068	775	12252
2020-04-23	3996	769	12056
2020-04-22	3932	761	11852

Fecha	HOSPITALES	OTROS LUGARES	PACIENTES UCI ACUMULADOS
2020-11-13	12348	30	6230
2020-11-12	12286	30	6205

```
2020-11-11        12286              30                          6186
2020-11-10        12286              30                          6161
2020-11-09        12124              30                          6143
...                 ...             ...                          ...
2020-04-26         7800              21                             0
2020-04-25         7633              21                             0
2020-04-24         7388              21                             0
2020-04-23         7271              20                             0
2020-04-22         7144              15                             0

                PACIENTES UCI DIA  HOSPITALES hoy  CENTROS SOCIOSANITARIOS hoy  \
Fecha
2020-11-13                    428            62.0                          3.0
2020-11-12                    428             0.0                          0.0
2020-11-11                    437             0.0                          0.0
2020-11-10                    447           162.0                          3.0
2020-11-09                    451             0.0                          0.0
...                           ...             ...                          ...
2020-04-26                      0           167.0                         66.0
2020-04-25                      0           245.0                        102.0
2020-04-24                      0           117.0                         72.0
2020-04-23                      0           127.0                         64.0
2020-04-22                      0             NaN                          NaN

                FALLECIDOS TOTALES hoy
Fecha
2020-11-13                        69.0
2020-11-12                         0.0
2020-11-11                         0.0
2020-11-10                       170.0
2020-11-09                         0.0
...                                ...
2020-04-26                       243.0
2020-04-25                       360.0
2020-04-24                       196.0
2020-04-23                       204.0
2020-04-22                         NaN

[158 rows x 10 columns]
```

[ ]: