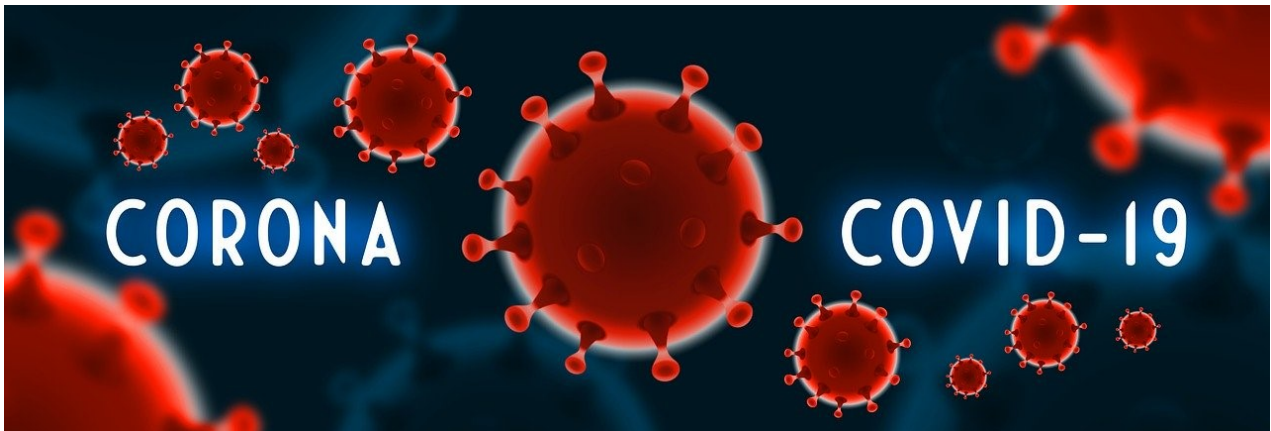


# Práctica 2

## Análisis de datos con Python



### Tema

Evolución de la enfermedad por CORONAVIRUS en España

### Objetivos

Realizar consultas que traten de explicar la evolución de la enfermedad por COVID-19 y su impacto en la población y en los servicios de salud de las Comunidades Autónomas de Castilla y León y Andalucía.

Para alcanzar el objetivo propuesto, se debe realizar un análisis de los datos proporcionados y descritos en la siguiente sección.

### Fuentes de datos

1. Evolución del número de hospitalizados en planta, hospitalizados en UCI, altas y fallecimientos por hospital debidos al coronavirus (COVID-19).

- T\_situacion-de-hospitalizados-por-coronavirus.csv

2. Datos diarios de confirmados, nuevos casos, altas y fallecimientos por provincias debidos al coronavirus (COVID-19) por provincias. Los datos de confirmados se facilitan por la Red de Vigilancia Epidemiológica a las 20:00 horas del día anterior y corresponden a personas a las que se les ha efectuado la prueba de PCR (prueba de detección molecular).

- T\_situacion-epidemiologica-coronavirus-provincias.csv

3. Datos de número de pacientes por provincia, sexo y tramo de edad. El dato de pacientes enfermos incluye todos los diagnósticos compatibles con coronavirus (neumonía por COVID-19 y enfermedad por COVID-19).

- T\_situacion-enfermos-por-coronavirus-por-tramos-sexo.csv

4. Datos del censo de habitantes por provincias de los años 2018 y 2019.

- Poblacion\_municipos\_sexo.csv

5. Datos de las 50 provincias de España (población según padrón municipal de habitantes del INE a 1 de julio de 2019) y de las ciudades autónomas de Ceuta y Melilla.

- Poblacion\_sueficie\_comunidades.csv

## Instrucciones generales

- Las prácticas se realizarán de forma individual
- El nombre del alumno debe figurar en el apartado que se habilita para ello.
- El lenguaje que se debe utilizar para la realización de dicho análisis ha de ser el lenguaje Python. Se penalizarán aquellos códigos que sean poco eficientes o que vayan en contra de la filosofía de cómputo de las librerías de análisis de datos Numpy y Pandas.
- Los ficheros de datos no se pueden manipular manualmente. Se ha de utilizar el lenguaje Python para todas las tareas relacionadas con el análisis; carga de datos, limpieza, procesamiento, transformación, generación de gráficas, etc.
- La visualización de datos nos ayuda a comprender e interpretar los datos. La visualización y generación de gráficas es opcional.
- El estudiante deberá entregar un notebook de Jupyter con los códigos que dan solución a las preguntas que se proponen. Se valorará que dicho documento esté bien organizado y comentado. Recordar que la presentación forma parte de la nota. El estudiante debe comprobar antes de la entrega que todos los códigos funcionan y no saltan errores de ejecución.
- El profesor podrá solicitar la defensa de la práctica por parte del alumno.

## Trabajo previo

Previamente es indispensable que hagáis un estudio de los conjuntos de datos para comprender su significado. Si entendemos mal los datos, podremos llegar a resultados incoherentes, como por ejemplo que se producen más fallecimientos que población total censada. Es recomendable que dediques un apartado para comentar el significado de los datos con los que vas a trabajar.

## Detección de copias

- Se realizará control de copias de todas las prácticas comparando las entregas. Se consideran infracciones graves la copia, la reproducción total o parcial del código de otros alumnos o cualquier código extraído de Internet o de cualquier otra fuente.
- En caso de cometer alguna infracción se aplicará el Reglamento de Régimen Académico de CUNEF.

In [204]:

```
#####
## Nombre y apellidos:
#####
```

## Carga de datos (importante)

A partir de los ficheros proporcionados en el proyecto (**deberás cargar correctamente los datos de fechas**), deberás crear los dataframes necesarios para realizar las cuestiones asociadas a cada una de las secciones que se describen a continuación.

## Sección 1.

1. Mostrar un listado de los hospitales de los cuales se conoce la incidencia del COVID-19. ¿Qué hospital ha sufrido el mayor número de fallecimientos?. ¿En qué provincia se encuentra dicho hospital?
2. Mostrar un listado de las provincias que han recogido datos de evolución del Covid-19. Para cada provincia se desea conocer el total de altas producidas. Calcular el número total de altas por Comunidad Autónoma.
3. Mostrar un listado de las comunidades autónomas de España

## Sección 2.

1. Seleccionar los datos de evolución de Covid-19 registrados en el hospital Complejo Asistencial de Soria.
2. Seleccionar los datos de evolución de Covid-19 registrados en la provincia de Soria.
3. Seleccionar los datos de evolución de Covid-19 registrados el día 2020-03-22.
4. Se desea estudiar la incidencia que ha tenido el Covid-19 dependiendo de los grupos de edad. Seleccionar los datos de evolución de Covid-19 asociados al grupo de edad 70-79.

## Sección 3.

1. Se desea estudiar la incidencia que ha tenido el Covid-19 dependiendo de los grupos de edad. Seleccionar los datos de evolución de Covid-19 asociados al grupo de edad 70-79 en la provincia de siglas SO.
2. Se desea estudiar la evolución que ha tenido el Covid-19 dependiendo de los grupos de edad. Seleccionar los datos de evolución de Covid-19 asociados a la provincia de siglas SA y al grupo de edad 60-69.
3. Se desea estudiar la evolución de los nuevos fallecimientos producidos por Covid-19 en las distintas provincias. Seleccionar los datos de evolución de Covid-19 asociados a la provincia de Valladolid el día '2020-05-10'.
4. Se desea estudiar la evolución de los nuevos fallecimientos producidos por Covid-19 en las distintas provincias. Seleccionar los datos de evolución de Covid-19 recogidos en el mes de Abril.

## Sección 4.

1. En el fichero `T_situacion-epidemiologica-coronavirus-provincias.csv` se encuentra la evolución diaria de enfermos por Covid-19 en varias de las provincias Españolas. Calcula el número total de casos confirmados en la provincia de Palencia.
2. Calcula el número total de fallecimientos en la provincia de Zamora. Usa el fichero `T_situacion-epidemiologica-coronavirus-provincias.csv` donde se encuentra la evolución diaria de enfermos por Covid-19 en varias de las provincias Españolas.
3. Calcula el número total de altas en la provincia de Valladolid.
4. Calcula el número total de enfermos de sexo varón mayores de 80 años producidos en la provincia de Zamora (ZA)

## Sección 5.

1. El fichero `T_situacion-enfermos-por-coronavirus-por-tramos-sexo.csv` recoge los datos de enfermos por sexo y edad. Muestra los datos de la provincia de León (LE) ordenados por fecha y grupo de edad en orden ascendente.
2. Muestra los datos asociados al grupo de edad '40-49' ordenados por fecha y provincia en orden descendente. El fichero `T_situacion-enfermos-por-coronavirus-por-tramos-sexo.csv` recoge los datos de enfermos por sexo y edad.
3. Muestra los datos de evolución de enfermos por Covid asociados a la provincia de Soria. Los datos han de estar ordenados por fecha y nombre de hospital de forma decreciente de fecha y ascendente de hospital.
4. Muestra los datos asociados a las provincias de la comunidad autónoma de Andalucía. Muestra los datos ordenados de forma creciente de Superficie y decreciente de población
5. Muestra los datos de evolución de casos confirmados de Covid asociados a la provincia de Ávila. Los datos han de estar ordenados por fecha de forma decreciente.

## Sección 6.

1. Añade una nueva columna llamada `ratio_hombres` al dataframe asociado al fichero `T_situacion-enfermos-por-coronavirus-por-tramos-sexo.csv`. El valor de dicha columna representa el porcentaje de hombres enfermos con respecto al total de enfermos
2. Añade una nueva columna llamada `diferencias` al dataframe asociado al fichero `T_situacion-enfermos-por-coronavirus-por-tramos-sexo.csv`. El valor de dicha columna representa la diferencia entre los hombres enfermos y mujeres enfermas
3. Añade una nueva columna llamada `nuevos_hospitalizados` al dataframe asociado al fichero `T_situacion-de-hospitalizados-por-coronavirus.csv`. El valor de dicha columna representa la suma de los nuevos hospitalizados en planta y los nuevos hospitalizados en uci.

## Sección 7

1. Se desea conocer la media de **nuevos hospitalizados en planta** en cada uno de los hospitales.
2. Se desea conocer el número máximo de **nuevos casos positivos** registrado en cada uno de las provincias.
3. Se desea conocer el número máximo de **nuevos casos positivos** registrado en cada una de las fechas.

## Sección 8.

1. ¿Cuántos valores NaN aparecen en la columna 'Posición' del dataframe obtenido de la lectura del fichero `T_situacion-epidemiologica-coronavirus-provincias.csv`?
2. Sustituye los valores NaN de la columna 'Posicion' del dataframe obtenido de la lectura del fichero `T_situacion-epidemiologica-coronavirus-provincias.csv` por el valor '0,0'.
3. Elimina las filas que contengan algún valor NaN del dataframe obtenido de la lectura del fichero `T_situacion-epidemiologica-coronavirus-provincias.csv`.
4. Seleccionar las columnas de fecha, hospital, altas y fallecimientos del dataframe asociado al fichero `T_situacion-de-hospitalizados-por-coronavirus.csv` y guardar el resultado en un fichero con nombre `nuevo_f1.csv`.

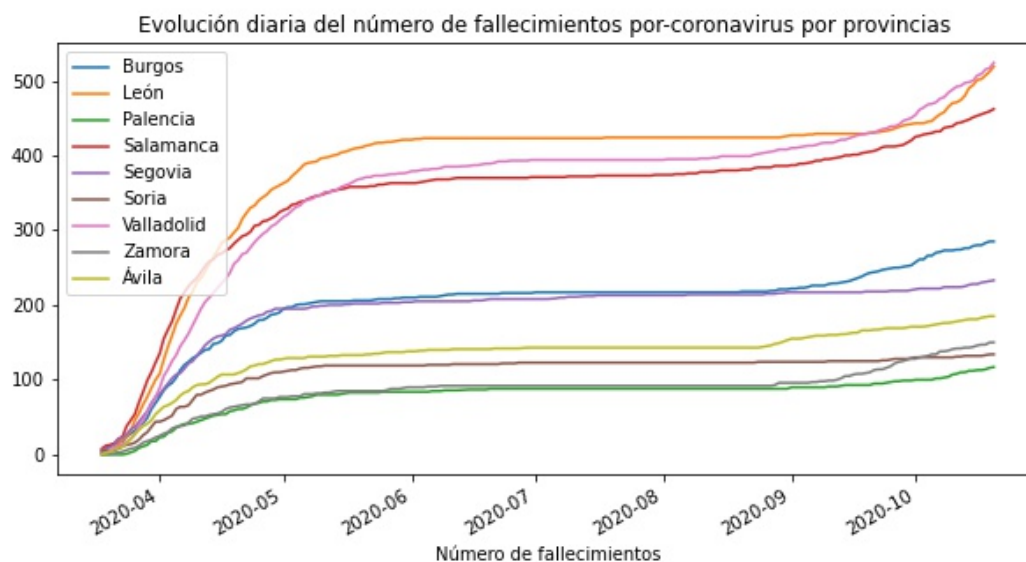
## Sección 9.

1. Crea una nueva columna llamada `size` en el dataframe asociado al fichero `Poblacion_sueficie_comunidades.csv`. El valor de dicha columna será 'P' para aquellas provincias cuya Población sea inferior a los 100.000 habitantes, 'M' para aquellas provincias cuya población sea inferior a 1.000.000 de habitantes y 'G' para aquellas provincias cuyo número de habitantes sea igual o superior al millón.
2. Crea una nueva columna llamada `alarma` en el dataframe asociado al fichero `T_situacion-de-hospitalizados-por-coronavirus.csv`. El valor de dicha columna será verde si el número de nuevos fallecimientos es inferior a 3, amarillo si el número de nuevos fallecimientos es inferior a 10 y rojo si si el número de nuevos fallecimientos es superior o igual a 10.
3. Crea una nueva columna llamada 'iniciales' en el dataframe asociado al fichero `T_situacion-de-hospitalizados-por-coronavirus.csv`. El valor de dicha columna es las iniciales de cada una de las provincias. Así, la inicial de Burgos es 'B'.
4. Crea una nueva columna llamada 'Código numérico' en el dataframe asociado al fichero `Poblacion_sueficie_comunidades.csv`. El valor de dicha columna es el código numérico de la

Comunidad Autónoma. Por ejemplo, para la Comunidad Autónoma de Cataluña con código 'C09', su código numérico será el 9.

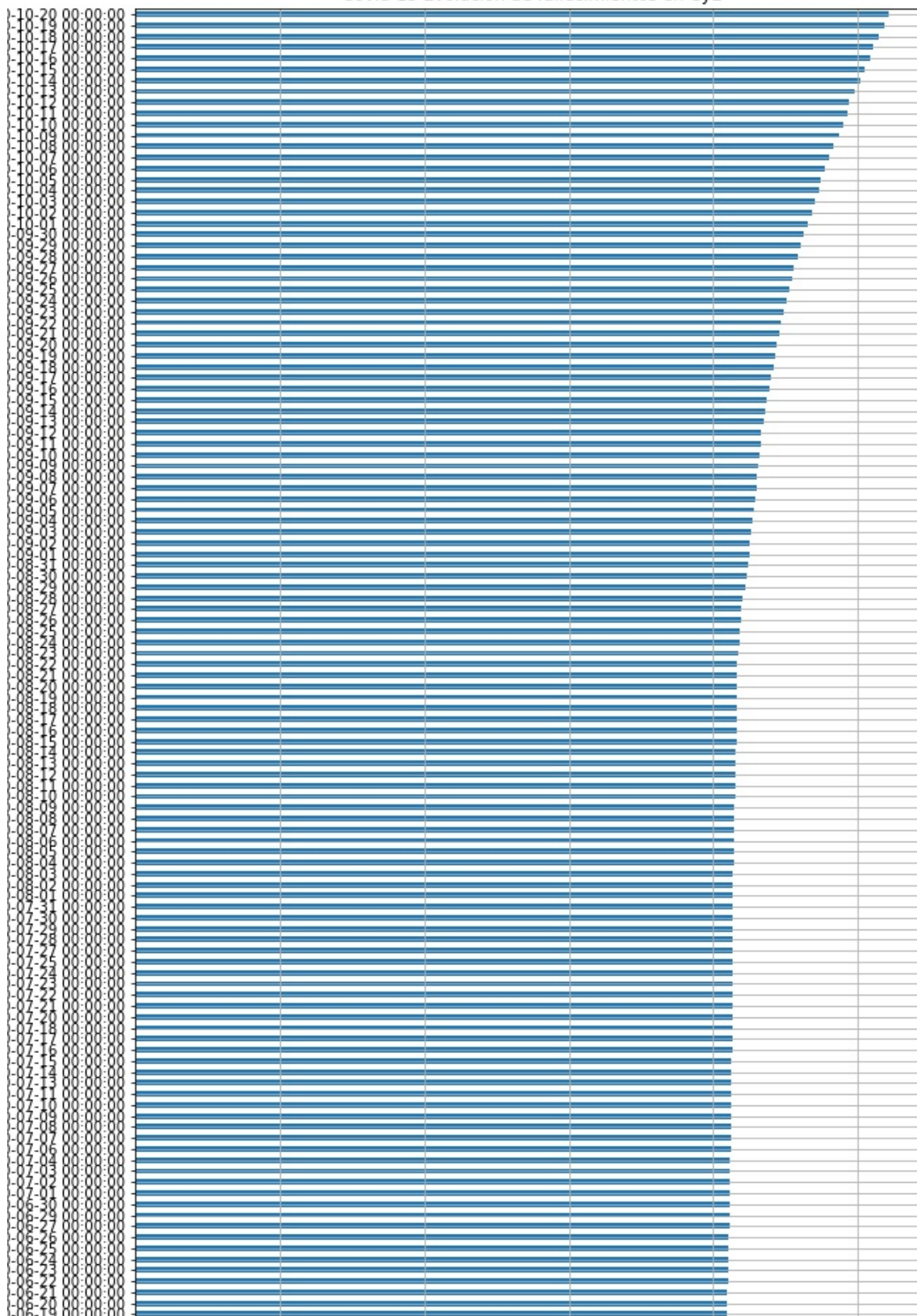
## Sección 10 (Opcional)

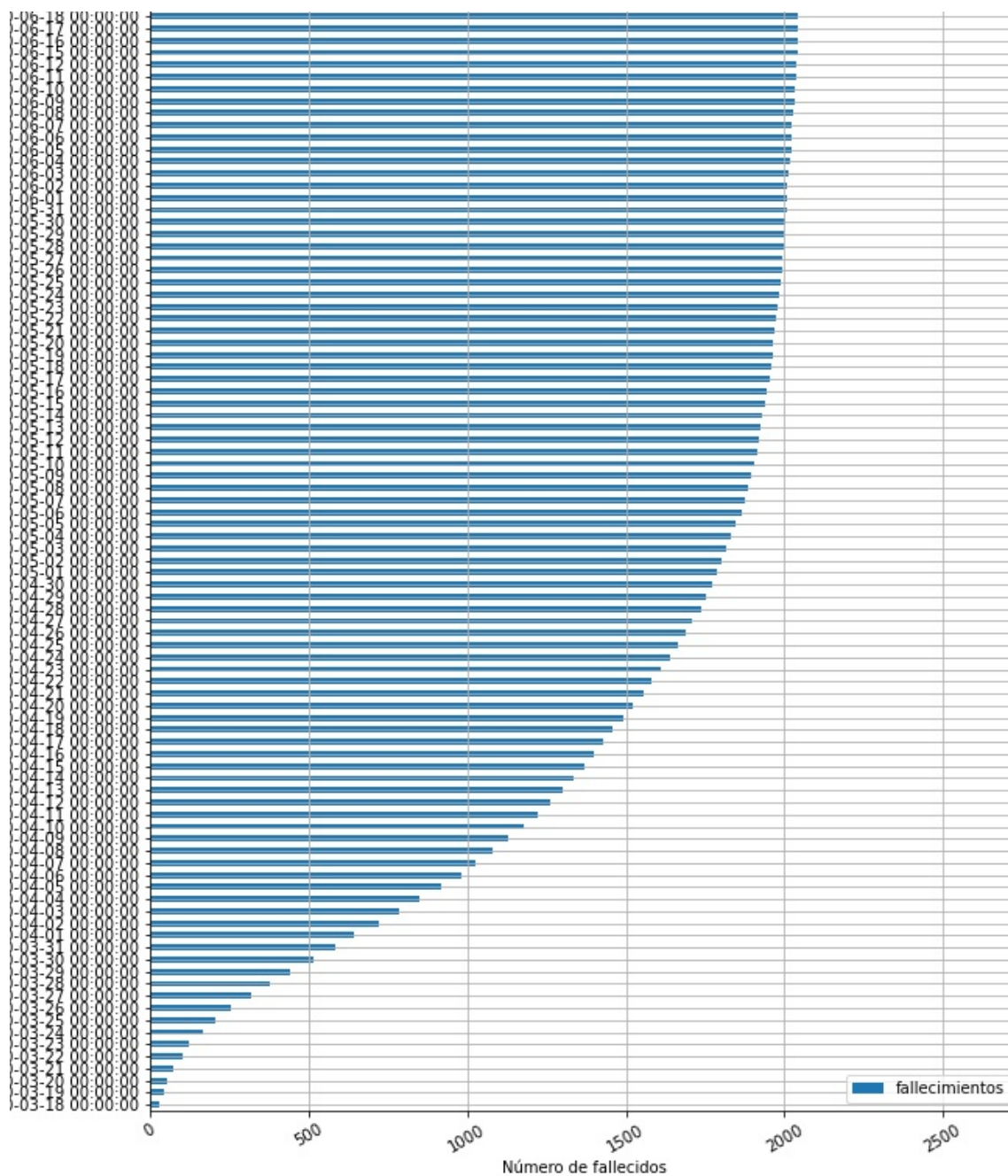
1. Representa la evolución diaria del Número de fallecimientos por-coronavirus por provincias de Castilla y León



2. Evolución diaria de fallecimientos en la C.A de Castilla y León.

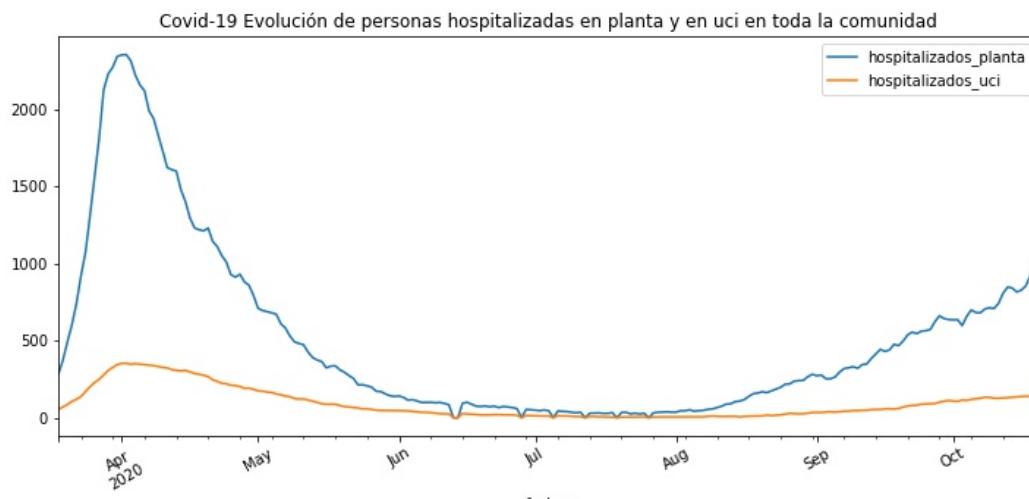
Covid-19 Evolución de fallecimientos en CyL



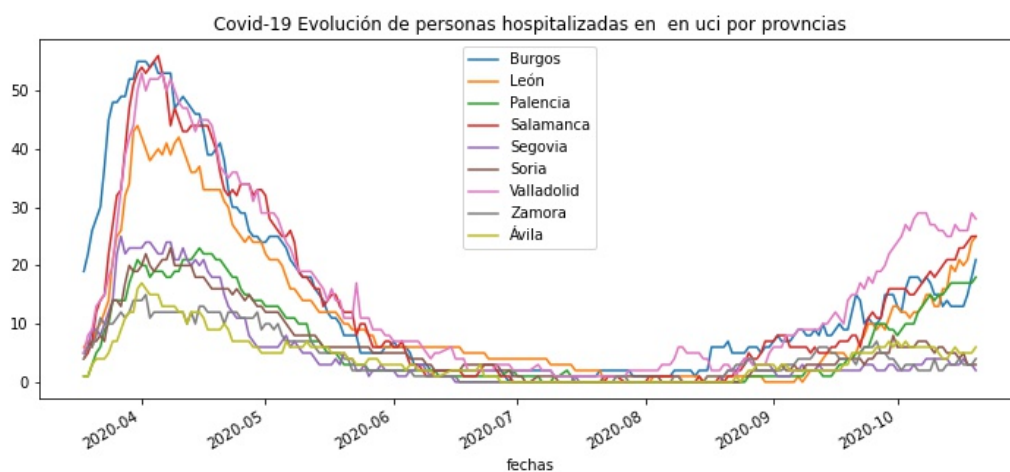


### 3. Evolución diaria de casos hospitalizados en planta y en unidades críticas en Castilla y León

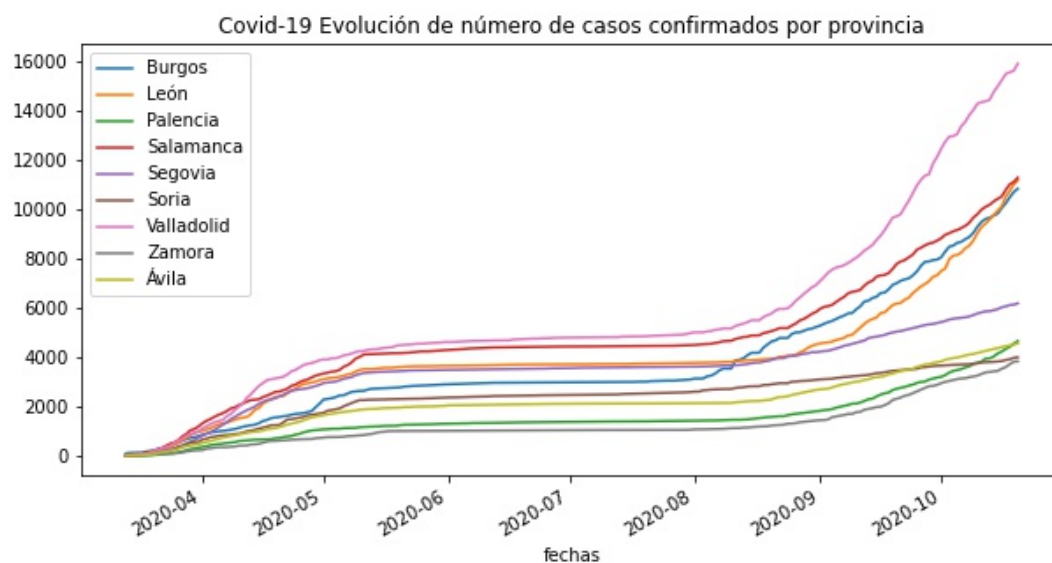




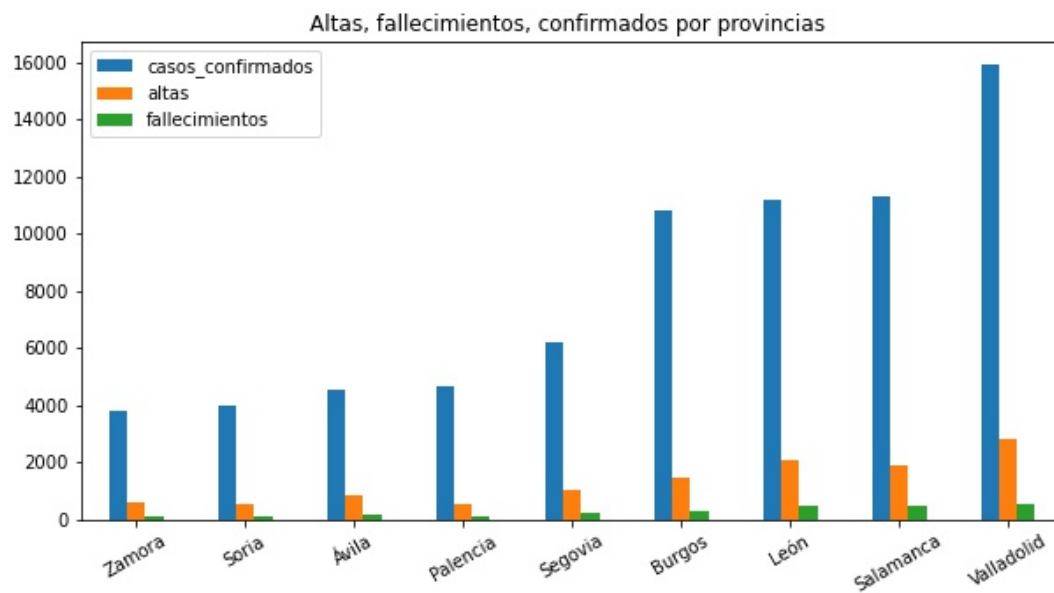
#### 4. Evolución diaria de casos hospitalizados en unidades críticas (por provincias)



#### 5. Evolución de número de casos confirmados por provincia



#### 6. Altas, fallecimientos, confirmados por provincias de Castilla y León.



In [ ]:



(<http://creativecommons.org/licenses/by-nc-sa/4.0/>)