



MÁSTER EN DATA SCIENCE Y ANÁLISIS DE DATOS

ESTUDIO DE LA INFLUENCIA ENTRE LA TEMPERATURA MEDIA DE UN LUGAR Y LA INCIDENCIA DEL CORONAVIRUS SARS-CoV2

AUTOR: Beatriz Cubillas García

DIRECTOR: Rafael Marín Sastre

FECHA: Septiembre 2021

ENTIDAD COLABORADORA

Instituto Europeo de Estudios Empresariales

1.- RESUMEN DEL CONTENIDO

Este trabajo corresponde al Proyecto Fin de Master sobre Data Science y Análisis de Datos. Dentro de las múltiples posibilidades que esta disciplina nos aporta se ha elegido un proyecto de investigación e innovación, para poder profundizar en algunas áreas del Data Science, en concreto es un trabajo de investigación desde la perspectiva de correlación. Se buscan las relaciones entre dos o más variables que en este caso son la temperatura media de una zona en concreto y la incidencia de nuevos casos de SARS-CoV-2 en esa zona.

Para poder estudiar esta correlación se han elegido seis provincias españolas, con características muy diversas que son: León, Huelva, Cantabria, Lérica, Ciudad Real y Castellón.

El objetivo principal de este proyecto es determinar si existe y como es la relación entre las variables mencionadas para en un futuro poder determinar las acciones a tomar sobre el control de la pandemia en función de la temperatura ambiental.

2.- ÍNDICE

1.- RESUMEN DEL CONTENIDO	1
2.- ÍNDICE	2
3.- INTRODUCCIÓN	3
4.- CONTENIDO DEL TRABAJO.....	4
4.1.- Delimitación del tema de estudio	4
4.2.- Descripción del método.....	5
4.3.- Exposición de resultados	7
5.- CONCLUSIONES.....	29
6. PLANIFICACIÓN Y PRESUPUESTO	33
6.1.- Planificación del proyecto	33
6.2.- Presupuesto.....	35
7.- MEJORAS FUTURAS DEL PROYECTO	38
8.- BIBLIOGRAFÍA.....	39
9.- ANEXOS	41
Anexo I - Los tipos de clima en España.....	41

3.- INTRODUCCIÓN

En el último año y medio nos hemos visto envueltos en la pandemia mundial extendida a todos los países que no se ha dado nunca en la historia. La palabra SARS – CoV2 nos acompaña desde que nos despertamos hasta que nos acostamos y ha cambiado nuestros hábitos de vida, la forma de relacionarnos con las personas que nos rodean, la forma de trabajar y un largo etcétera.

Vivimos constantemente preocupados por infinidad de números y estadísticas, que marcan las **restricciones que son aplicadas a la población** tanto en el trabajo, como en su vida social. Algunos de estos datos que más nos suenan son el número de personas contagiadas por Covid diariamente o la incidencia que estos contagios suponen a 14 días o el número de personas hospitalizadas por esta causa.

Otro de las palabras que más hemos oído y utilizado los últimos meses, es “ola”, podríamos definirlas como episodios de contagios masivos sin control en la población. En la aparición de estas olas podemos destacar diversos factores entre los que hay que destacar la época del año, la interacción social y cumplimiento de las medidas de control. Los dos últimos son aspectos sobre los que no tenemos capacidad de análisis ya que corresponden a comportamientos individuales impredecibles. En cambio, sobre la época del año en la que se producen estas olas, y las condiciones climáticas que corresponden a cada zona si podemos **realizar un análisis de correlación**.

Este proyecto versa sobre la relación que tienen estas olas o episodios con la temperatura media que hay en las poblaciones

Entre los objetivos principales de este proyecto está la posibilidad de **tomar medidas de control** sobre la población en función de la climatología de la zona.

4.- CONTENIDO DEL TRABAJO

4.1.- Delimitación del tema de estudio

Dentro del amplio espectro de utilidades que existe en el mundo del **Data Science** podemos destacar como principales objetivos: la interpretación y análisis de resultados, la búsqueda de tendencias, correlaciones y patrones a partir de un análisis de datos concretos, recopilación de grandes conjuntos de datos estructurados y no estructurados de fuentes dispares, limpieza y validación de los datos para garantizar la precisión, integridad y uniformidad, diseñar y aplicar modelos y algoritmos para explotar las grandes bases de datos, interpretación de los datos para descubrir soluciones y oportunidades, así como comunicar los hallazgos con medios de visualización de datos acorde con el nivel del receptor de la información.

Vamos a trabajar en este proyecto intentado llevar a cabo varios de estos objetivos para un aprovechamiento máximo de los conocimientos adquiridos en esta materia.

Como metodología principal del proyecto, realizaremos el tratamiento de la información mediante **un proceso ETL** (Extract-Transform-Load), se trata de un proceso que permite a las organizaciones mover datos desde múltiples fuentes, reformatearlos y cargarlos en otra base de datos (datamart o datawarehouse) con el objeto de analizarlos.

Para la realización de la primera etapa de nuestro modelo ETL utilizaremos la minería de datos, ya que necesitamos recabar los **datos climatológicos** de las seis provincias que hemos elegido para el muestreo y también los datos de los **nuevos contagios** diarios de estas seis provincias. Posteriormente realizaremos la transformación y adecuación de estos datos para, por último, realizar la carga de los mismo en una herramienta de Big Data como es R, lenguaje de programación que nos facilita el cálculo estadístico y la generación de gráficos.

El estudio de la correlación se llevará a cabo entre las dos variables establecidas: Temperatura media, y nº de contagios diarios. Este trabajo sería aplicable a todo el territorio español, pero hemos decidido delimitar el estudio a seis provincias, comenzando por León, porque es la provincia en la que resido, y eligiendo las otras cinco provincias, Huelva, Cantabria, Lérida, Ciudad Real y Catellón, por tener una población provincial similar y estar situadas en puntos distantes, así como tener condiciones climáticas y culturales diferentes y

pertenecer a diversas Comunidades Autónomas donde se han establecido diferentes medidas administrativas de control de la pandemia.

En esta tabla se muestra un resumen de las características principales de las seis **provincias** elegidas.

PROVINCIA	POBLACIÓN PROVINCIAL	ALTITUD MEDIA	COMUNIDAD AUTÓNOMA	SITUACIÓN
LEÓN	455.066	912	Castilla y León	Centro-Norte
HUELVA	530.386	19	Andalucía	Sur- Océano Atlántico
CANTABRIA	583.904	52	Cantabria	Norte – Mar Cantábrico
LLEIDA	436.911	185	Cataluña	Noreste
CIUDAD REAL	492.065	628	Castilla – La Mancha	Centro
CASTELLÓN	576.941	43	Comunidad Valenciana	Este-Mediterráneo

Estos datos los hemos obtenido del sitio web del *[Instituto Nacional de Estadística]*

Al igual que ha sido necesario limitar las provincias de estudio para este proyecto, también hemos marcado unos **límites temporales** a la recogida de datos, que van desde el 1 de julio de 2020 hasta el 30 de junio de 2021. Estos límites corresponden a un año completo, partiendo de la fecha en la que se terminó el confinamiento estricto de los meses de marzo, abril y mayo de 2020, donde los datos no son del todo precisos, ya que no existía interacción social.

4.2.- Descripción del método

Como ya hemos comentado vamos a realizar un proceso de ETL (Extracción, Transformación y Carga (Load)) de datos, modelo que nos permite basar las decisiones en los datos extraídos de diversas fuentes de datos, como es el caso de este proyecto.

Extracción

Para la obtención de los **datos climatológicos** hemos recurrido a la página de la Agencia Estatal de Meteorología en su sección Open Data [Aemet2021] después de registrarnos, nos envían una clave para poder descargar los datos. En la página de Acceso General, apartado de Valores climatológicos, Climatologías diarias, seleccionando las provincias para nuestro estudio, así como la fecha inicio y fecha fin del estudio, y podemos descargarnos dichos datos.

En el caso de **los datos de incidencia de Covid** recurrimos a las páginas de información de la correspondiente Comunidad Autónoma de las provincias objeto de estudio, que son las siguientes:

- León [*Datos abiertos JCyL. (agosto 2021)*]
- Huelva [*Junta de Andalucía (agosto 2021)*]
- Cantabria [*Servicio Cántabro de Salud (agosto 2021)*]
- Lleida [*Datos Covid Cataluña (agosto 2021)*]
- Castellón [*Datos abierto, gobierno de Valencia (agosto 2021)*]
- Ciudad Real [*Sanidad Castilla – La Mancha (agosto 2021)*]

En todas ellas nos ofrece la posibilidad de descargar los datos en archivos Excel o .csv.

Transformación

En esta etapa se realizan las labores necesarias para **la adecuación de los datos** que entre otras han sido la selección de ciertas columnas, revisar que corresponden con las fechas marcadas para este proyecto, dividir una columna en varias, ordenar los datos, comprobar la existencia de datos faltantes o erróneos y unir datos de múltiples fuentes.

Carga (Load)

Realizamos la carga de datos por provincias en el **programa RStudio** versión 1.4.1106, para poder realizar análisis estadísticos y obtener gráficos, con el fin de poder extraer conclusiones.

La carga de los datos se realiza tanto desde archivos Excel, como archivos de texto. En el propio programa generamos dataframes con los datos y realizamos las operaciones para la obtención de gráficos.

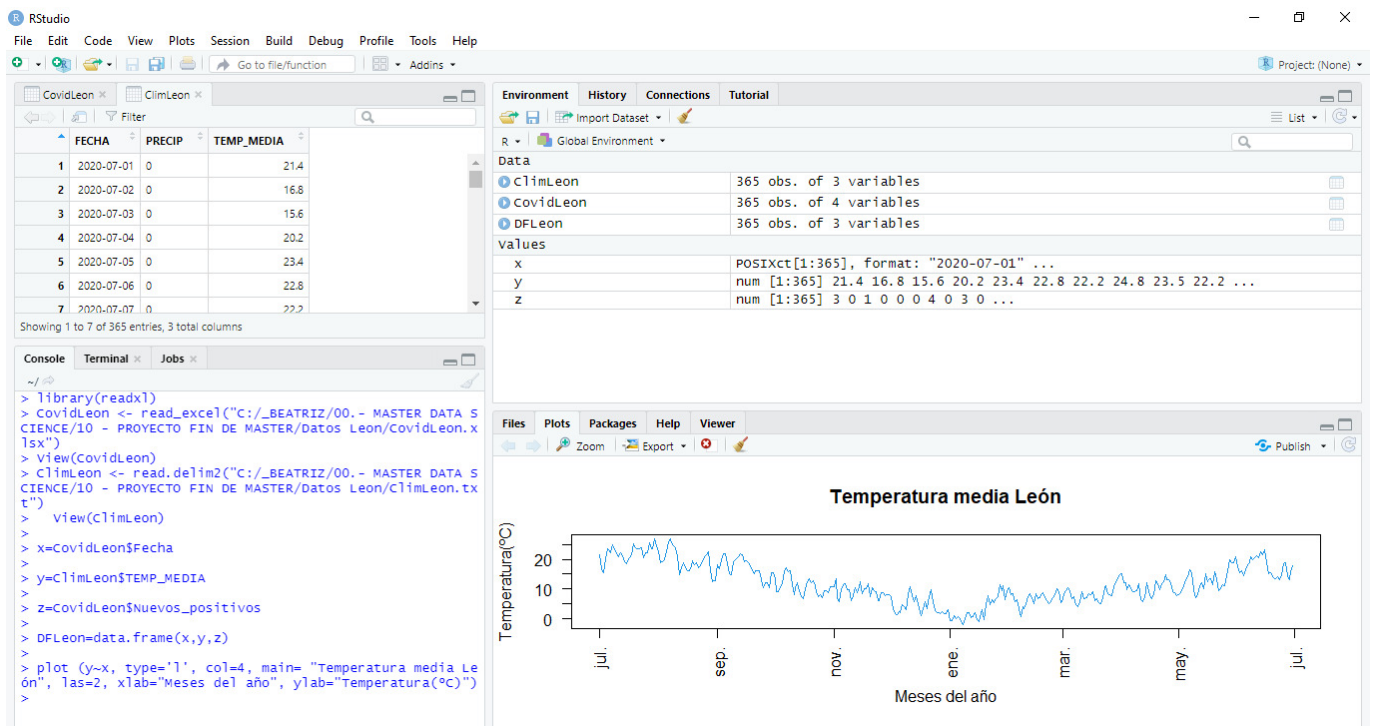
4.3.- Exposición de resultados

En la aplicación RStudio se cargan los archivos de datos y se procesan con el fin de obtener unos gráficos que nos muestren la información de forma más sencilla y concisa.

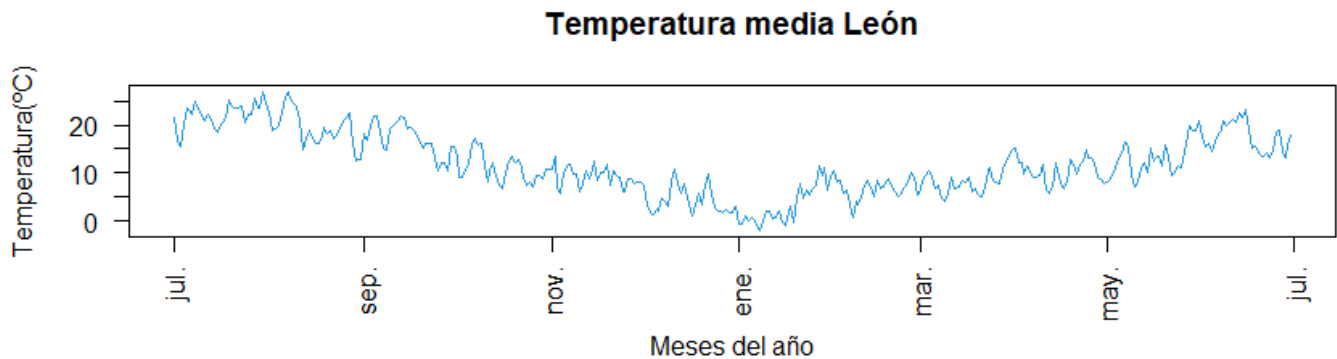
LEÓN

En esta imagen se puede ver como se ha cargado el archivo de texto “ClimLeon” y el archivo de Excel “CovidLeon”. Se asignan valores de ‘x’ para la columna de las fechas, ‘y’ para la columna de las temperaturas medias por día y ‘z’ para el número de casos positivos de Covid por día.

Finalmente **se genera un dataframe** con estas tres variables DFLeon(x, y, z), para, a partir del mismo, poder generar los gráficos que relacionen las variables.

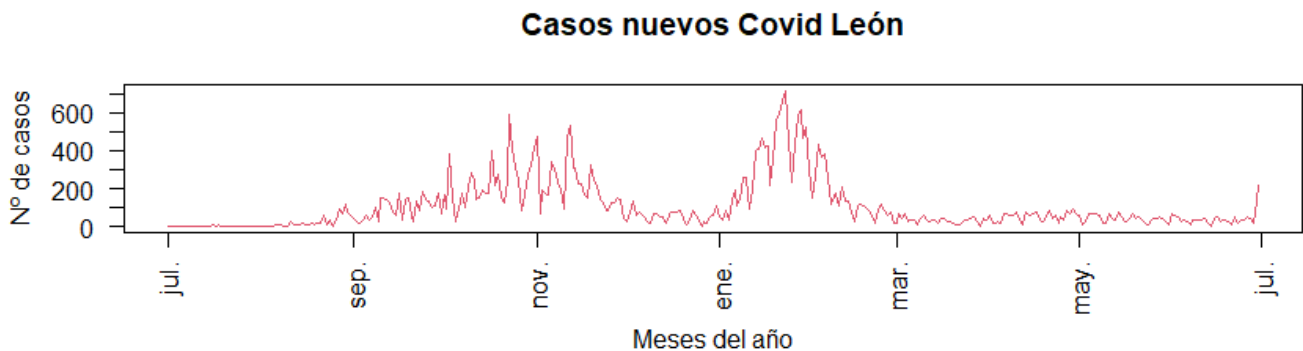


En el siguiente gráfico se observa la **variación de la temperatura media diaria** desde el 1 de Julio de 2020 al 30 de junio de 2021, en León.



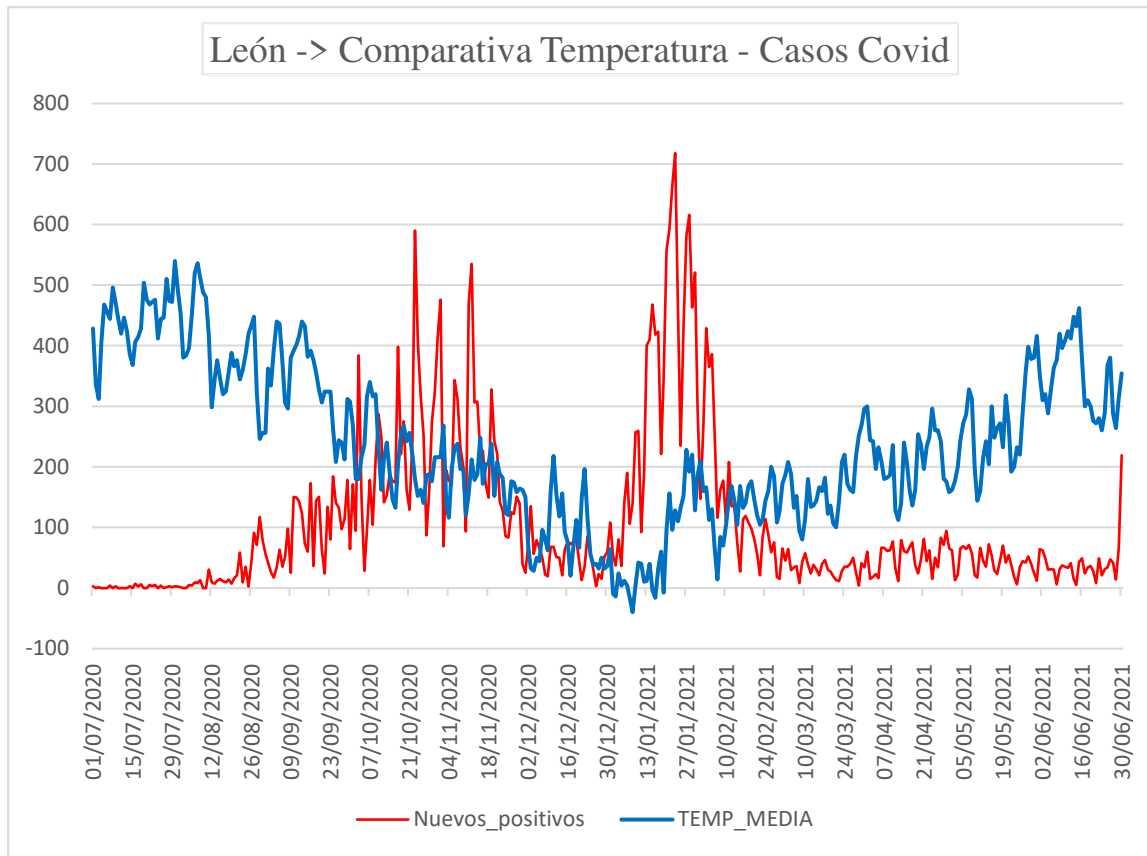
Como podemos ver en el Anexo I de los Tipos de Clima en España, la temperatura media diaria se corresponde con una curva típica atribuible al **Clima Mediterráneo Interior**. En este clima los veranos son frescos (con una media de 22° C en el mes más cálido) y los inviernos son considerablemente fríos, con entre 6°C y -3° C en el mes más frío del año.

Ahora se genera el **gráfico del número de casos nuevos de Covid por día** en el año seleccionado.



Vemos en este gráfico que se han producido dos “olas” principalmente con sus puntos álgidos al final del mes de octubre de 2020 y al final del mes de febrero de 2021.

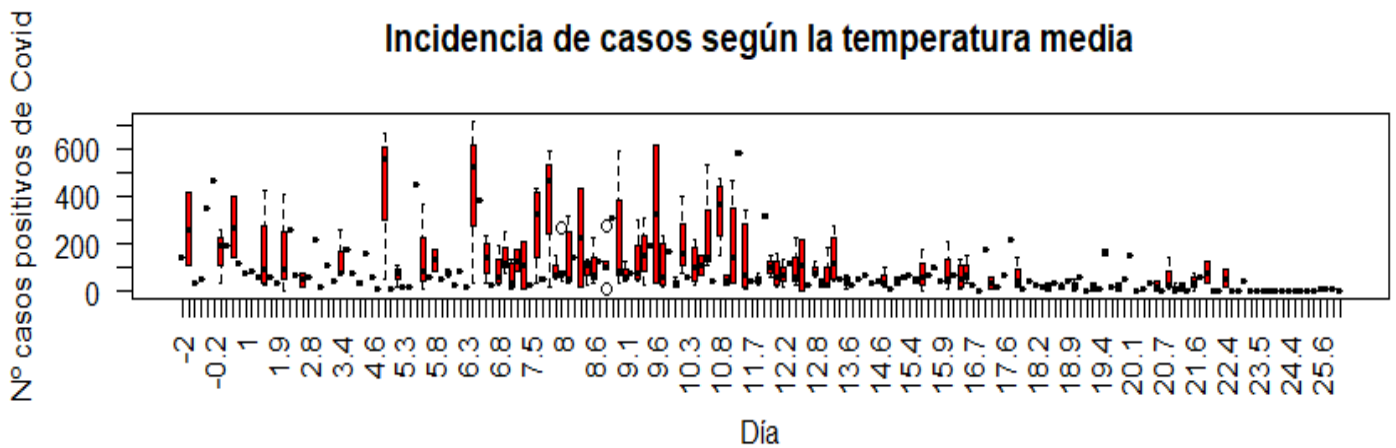
A continuación, vamos a unificar en la misma gráfica las dos anteriores, que tiene los valores de temperatura media (azul) y el número de casos nuevos de Covid diarios (rojo). La temperatura tiene una escala ampliada (x20) para que se pueda observar con claridad los valores y su comparación con la variable del número de nuevos contagios.



En esta gráfica se observa la relación inversa que existe entre la temperatura y el número de casos de Covid, ya que entre los meses de julio de 2020 y septiembre de 2020 y entre febrero de 2021 y junio de 2021, se concreta que, **“a mayor temperatura, menor incidencia de nuevos casos”**.

En los meses de noviembre a enero coincidiendo con la época en la que las temperaturas medias son bajas, **por debajo de los 10 °C**, es el momento en el que se producen más contagios, en este caso concentrados mayoritariamente en dos “olas”.

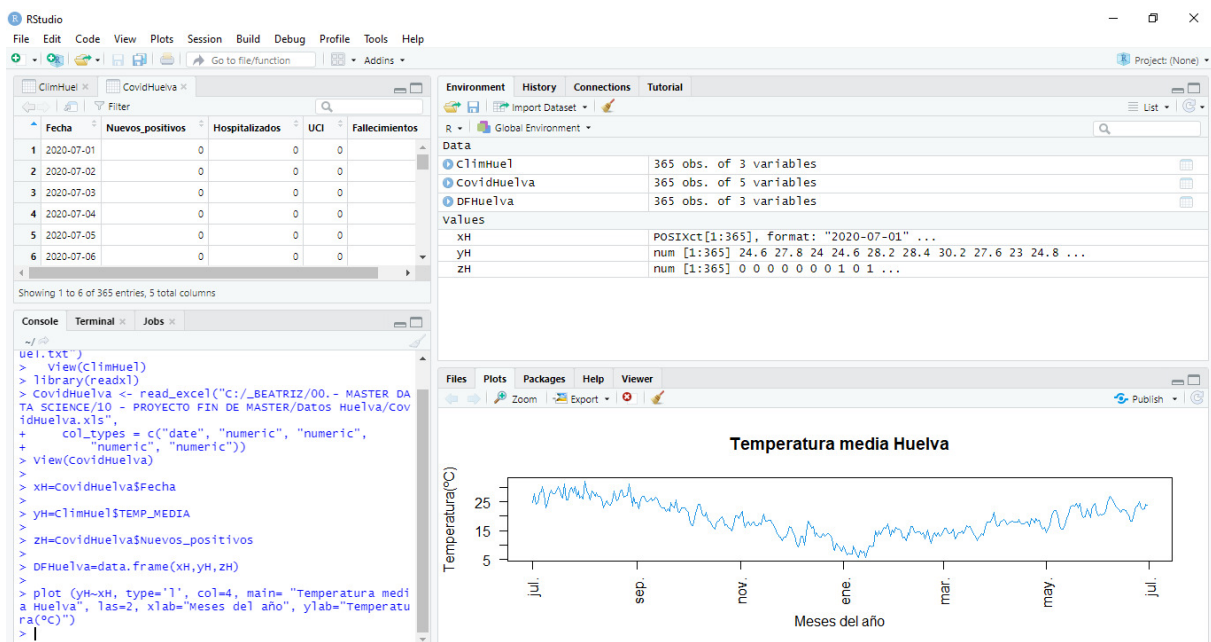
Finalmente se ha generado el gráfico de la relación entre el número nuevo de casos Covid que se producen según la temperatura media del día (Boxplot)



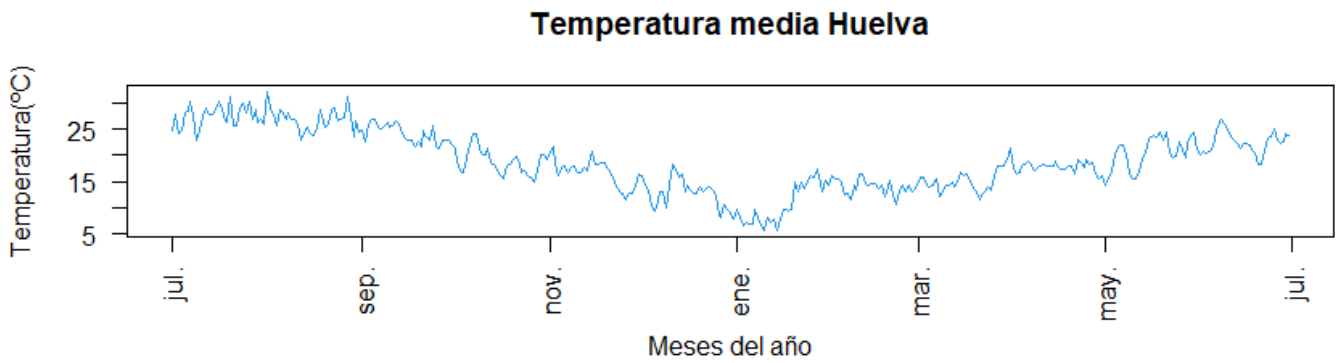
En este gráfico vemos las temperaturas medias en el eje x, ordenadas de menor a mayor y los gráficos en rojo nos representan el número de casos nuevos diarios. Como podemos observar **los días que están por debajo de los 12°C de media tienen una mayor incidencia de casos**, que los que están por encima de esa temperatura media.

HUELVA

Mediante la observación y estudio de los datos descargados para la provincia de Huelva, realizamos las mismas operaciones en R.

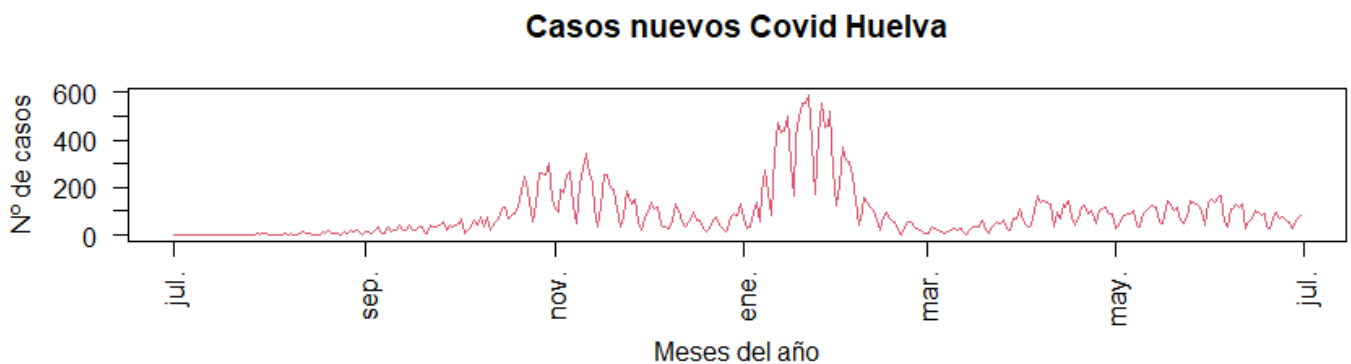


Se comienza obteniendo el gráfico de la temperatura media diaria desde el 1 de julio de 2020 al 30 de junio de 2021.

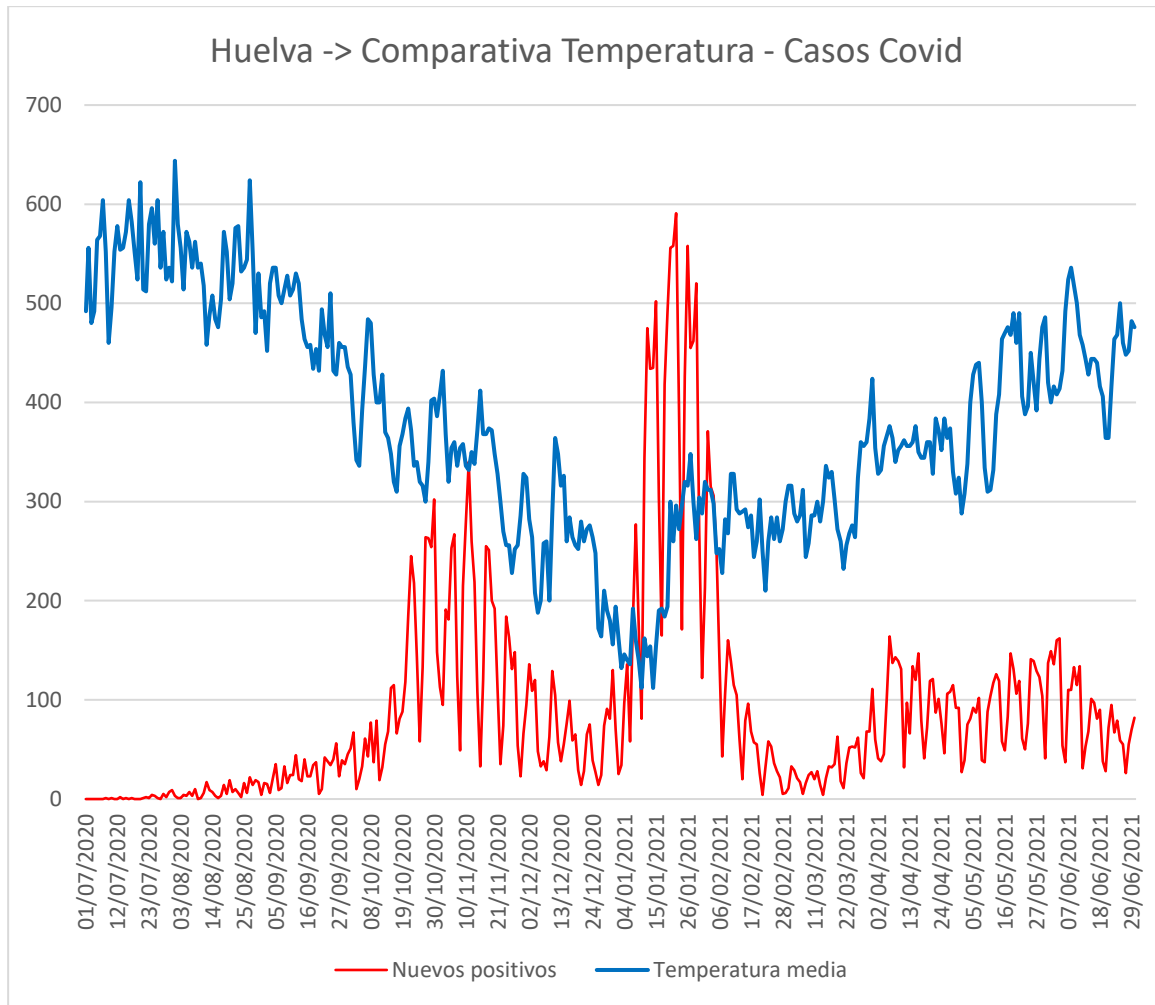


Vemos que este gráfico nos muestra la variación de un **Clima Mediterráneo Costero**, que no tienen tantas variaciones entre máximas y mínimas como en el interior peninsular y los veranos no son demasiado calurosos, por la influencia del mar. Por su parte, los inviernos son suaves y la media del mes más frío no baja de los 10° C (ver Anexo I).

Respecto al gráfico del número de casos nuevos de Covid por día en el año seleccionado, vemos que se han producido dos episodios de alta incidencia entorno al mes de noviembre y al mes de enero.

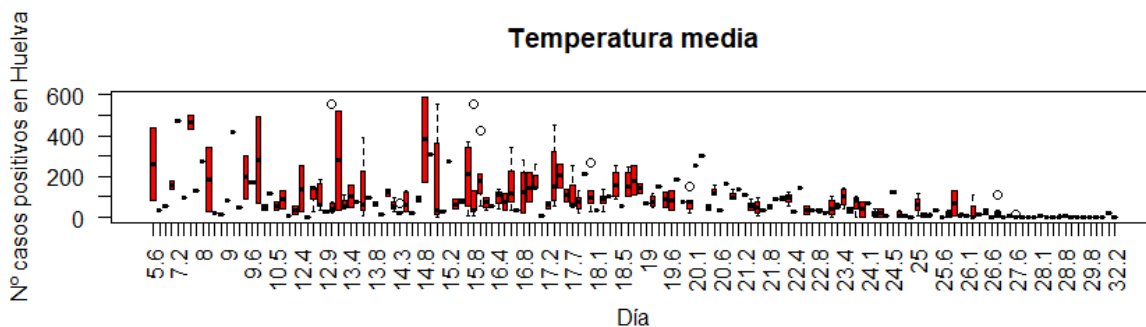


Para poder **comparar la influencia de las dos variables** unificamos en el mismo gráfico la temperatura media y los casos de Covid. La temperatura tiene una escala ampliada (x20) para que se pueda observar en la misma gráfica.



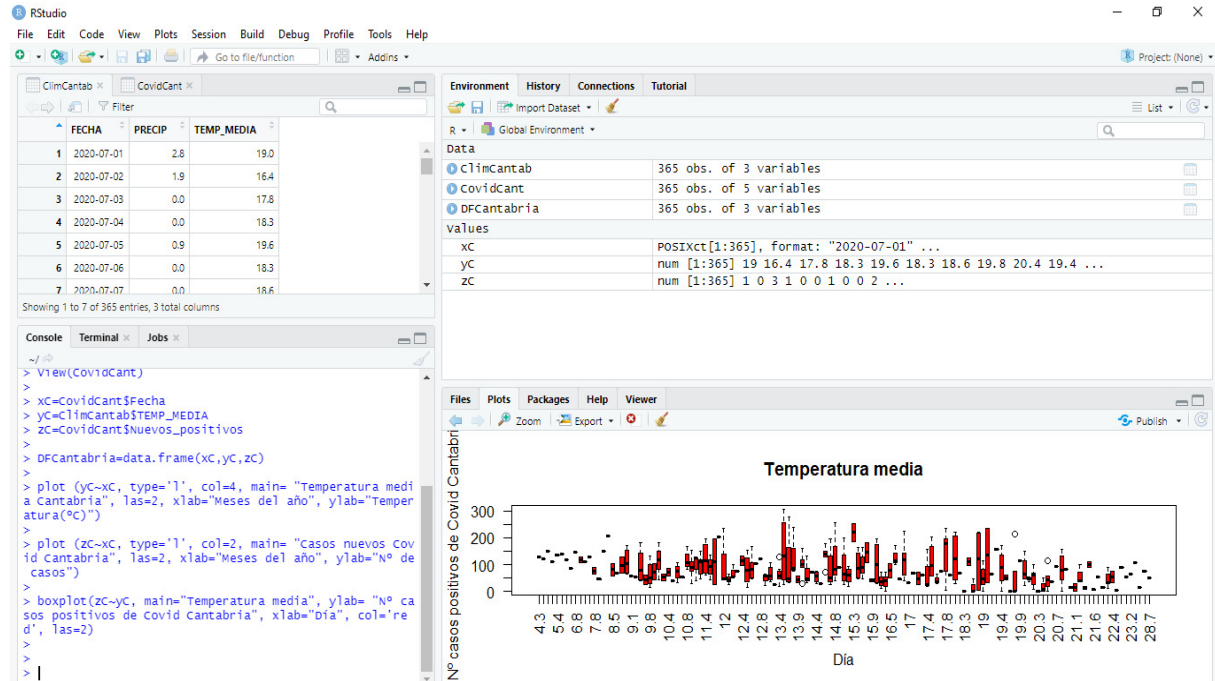
Se observa en el gráfico, que en los meses de julio de 2020 a septiembre de 2020, así como desde mediados de febrero de 2021 a junio de 2021, donde **las temperaturas son altas, la incidencia de Covid se mantiene en niveles bajos**. En cambio, en los meses de octubre a enero donde la temperatura media está por debajo de los 15°C la incidencia de Covid es mayor.

En este gráfico vemos la relación entre el número nuevo de casos Covid que se producen según la temperatura media del día (Boxplot). Y vemos **que la incidencia mayor se produce en los días que tienen una temperatura media menor a 17° C.**

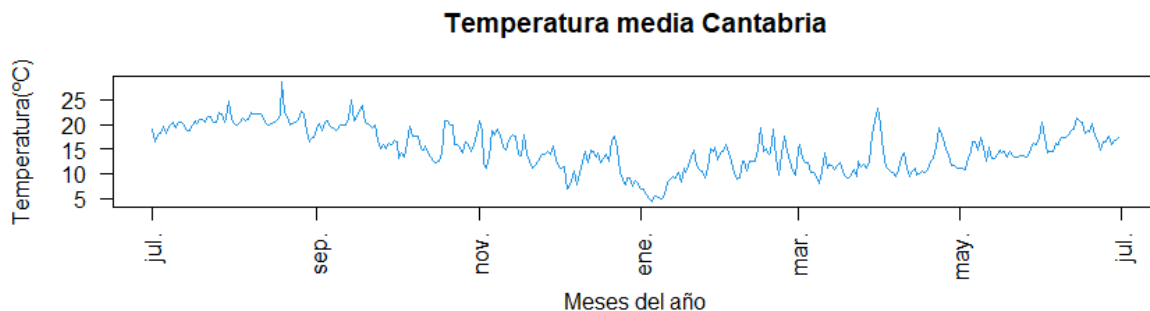


CANTABRIA

Vemos ahora los cálculos realizados para la provincia de Cantabria.

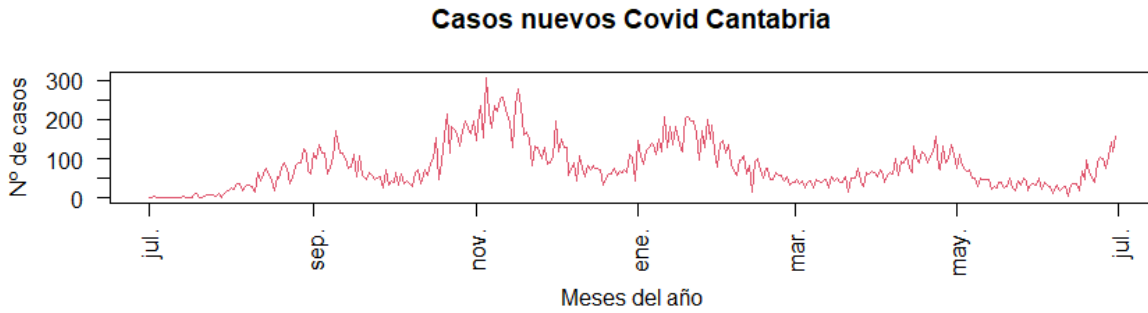


Comenzando por el gráfico de la temperatura media a lo largo de todo el año de estudio.



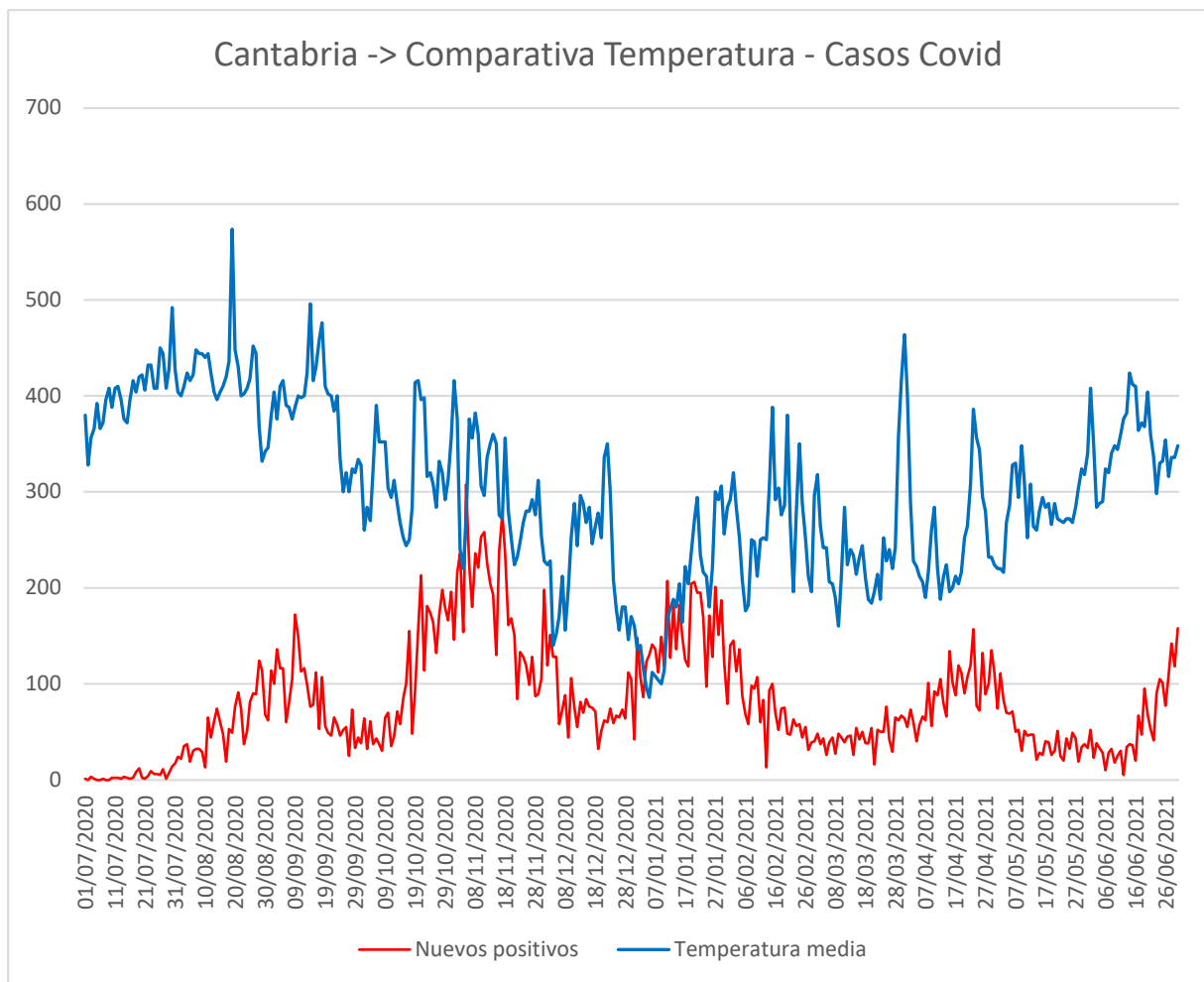
Dado que Cantabria se encuentra en la zona de costa al norte, su **Clima** es de tipo **oceánico**, donde las temperaturas suelen ser suaves debido a la cercanía del mar, lo que también redundaría en una mayor humedad. En invierno oscilan entre los 12 °C y los 15 °C y en verano rondan los 20-25 °C en cuanto a las medias (ver Anexo I).

Ahora vemos el gráfico del número de casos nuevos de Covid por día en el año seleccionado



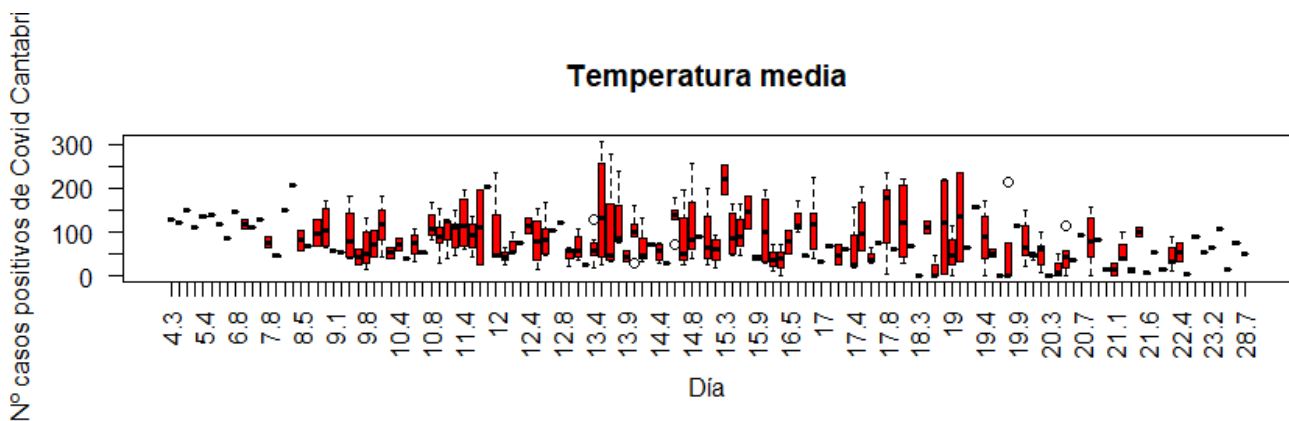
Comprobamos que en este caso se han producido 4 “olas” a lo largo del año, en septiembre, noviembre, enero y mayo. Han sido de menor intensidad que las ocurridas en las provincias de León y Huelva.

Vamos a comprobar en el **gráfico que unifica la temperatura media y los casos de Covid** cuál es su relación. La temperatura tiene una escala ampliada (x20) para que se pueda observar en la misma gráfica.



Al tener una temperatura más constante a lo largo del año, y no tener tanta amplitud térmica, se producen episodios de contagio de menor intensidad pero más a menudo. Solo en los meses de máximo calor, como julio, agosto y mayo y junio podemos observar que con mayor temperatura media se produce menor número de contactos.

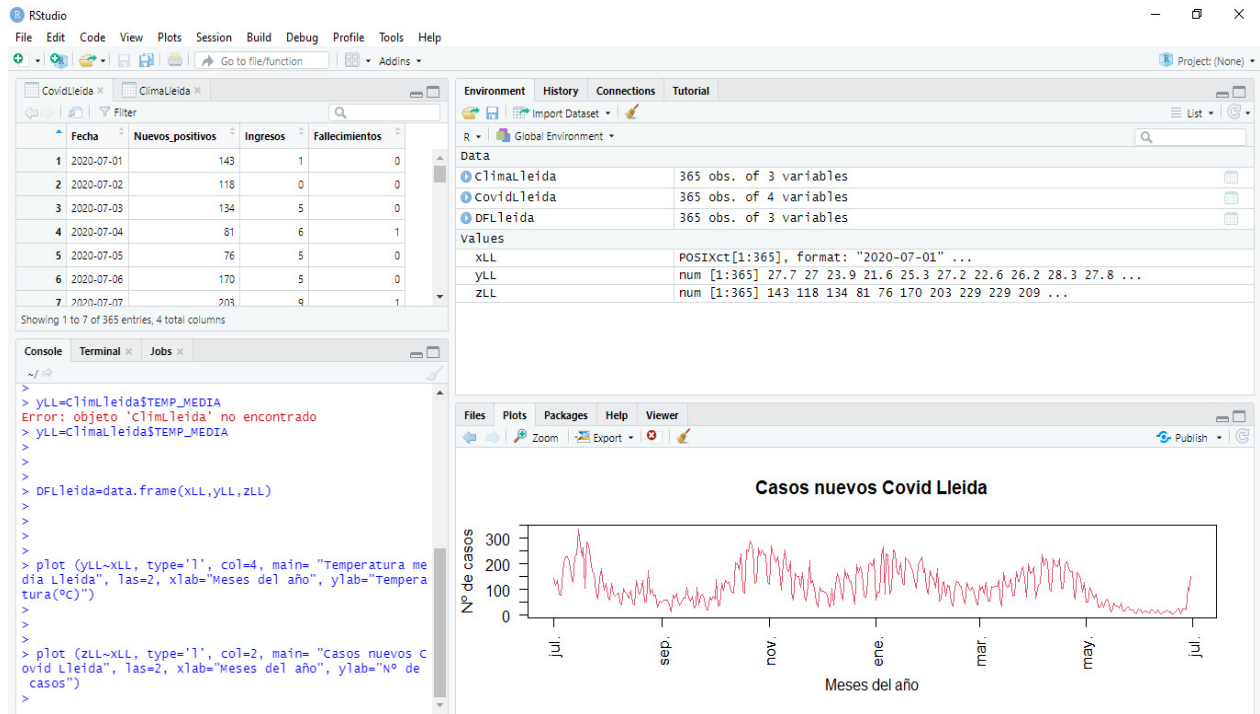
Finalmente estudiamos el gráfico que muestra la relación entre el número nuevo de casos Covid que se producen según la temperatura media del día (Boxplot), ordenadas las temperaturas medias de menor a mayor.



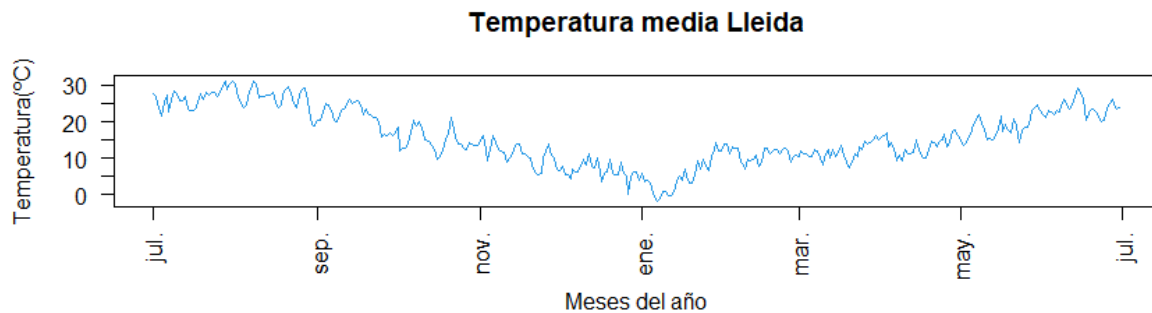
En este gráfico vemos que no se cumple la constante de los anteriores, ya que **no se produce una caída de contagios hasta que la temperatura media no está por encima de los 20°C**, probablemente porque la amplitud térmica de las medias de las temperaturas es menor que en las provincias de León y Huelva.

LLEIDA

Partiendo de los datos de la provincia de Lleida, realizamos los cálculos correspondientes.

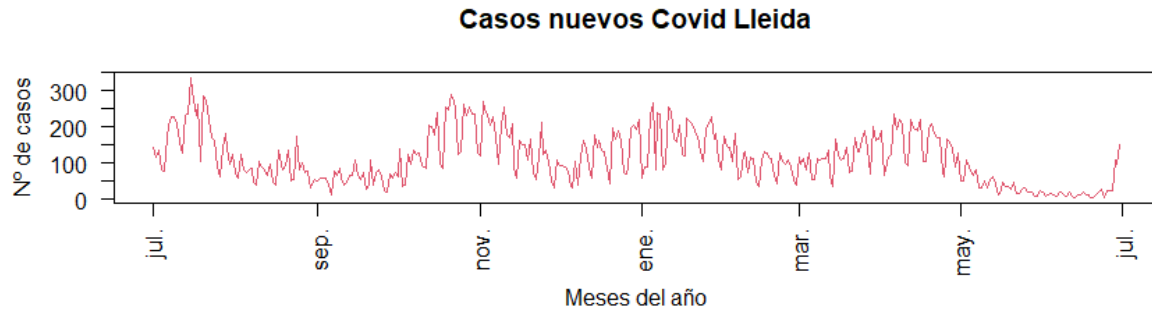


Se puede observar la distribución de la media de temperaturas a lo largo del año.

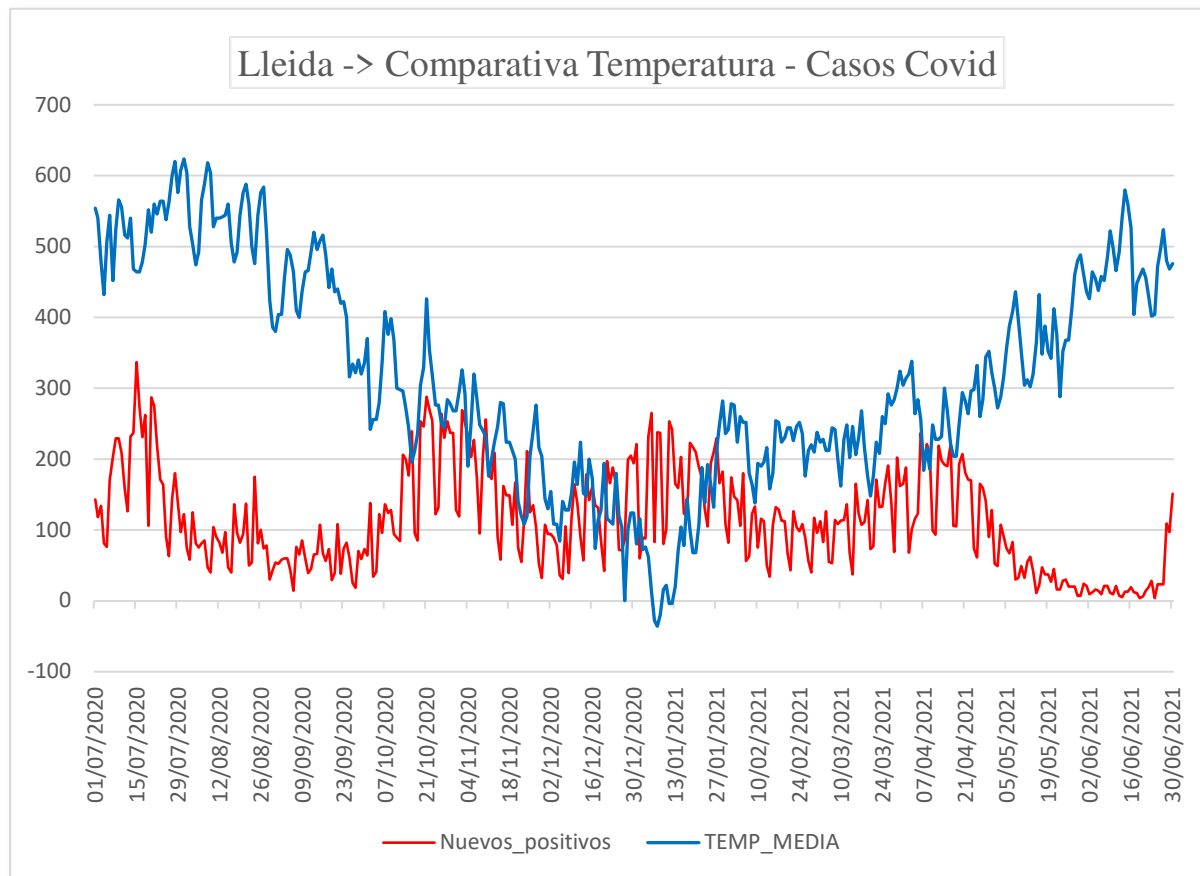


Dentro de la provincia de Lleida tenemos varios tipos de Clima, ya que en la zona norte es clima es de Montaña, y también tiene parte de la provincia en clima Oceánico y otra parte en clima Mediterráneo de interior. Esta peculiaridad hace que las temperaturas no sean similares en toda la provincia (ver Anexo I).

Obtenemos la distribución de los casos de Covid a lo largo del año estudiado, viendo que se han producido cuatro olas principales, en los meses de julio y noviembre de 2020, así como enero y abril de 2021.

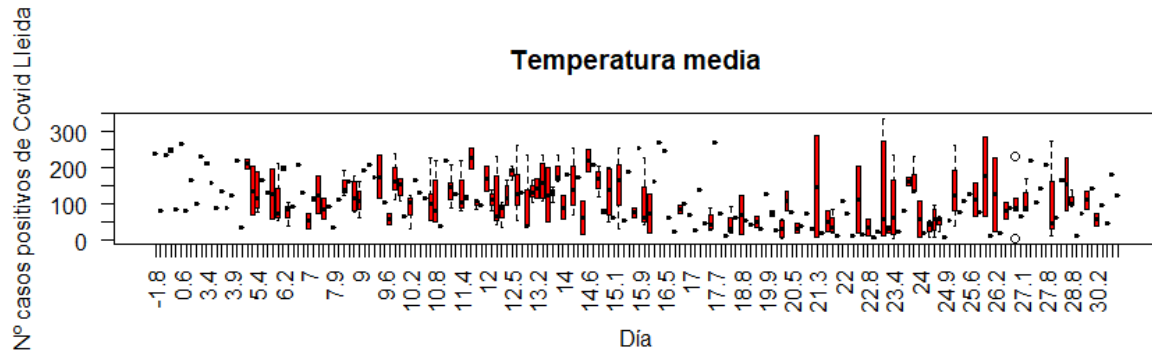


En el mismo gráfico vemos la temperatura y la incidencia de casos Covid.



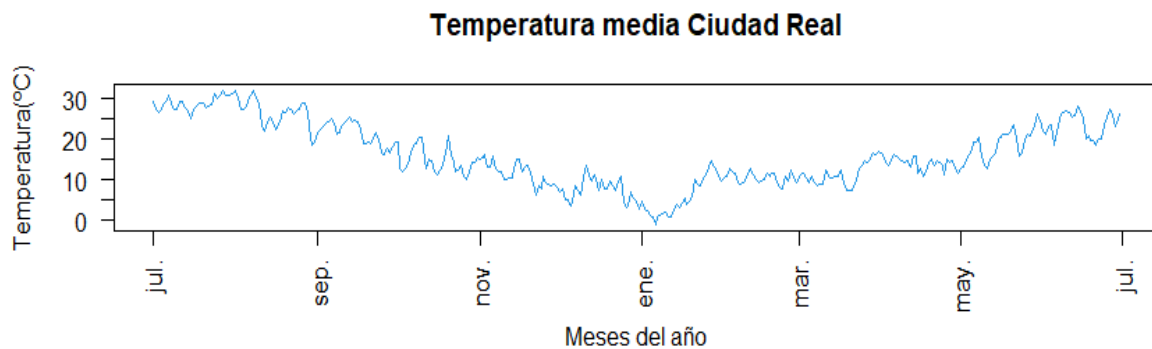
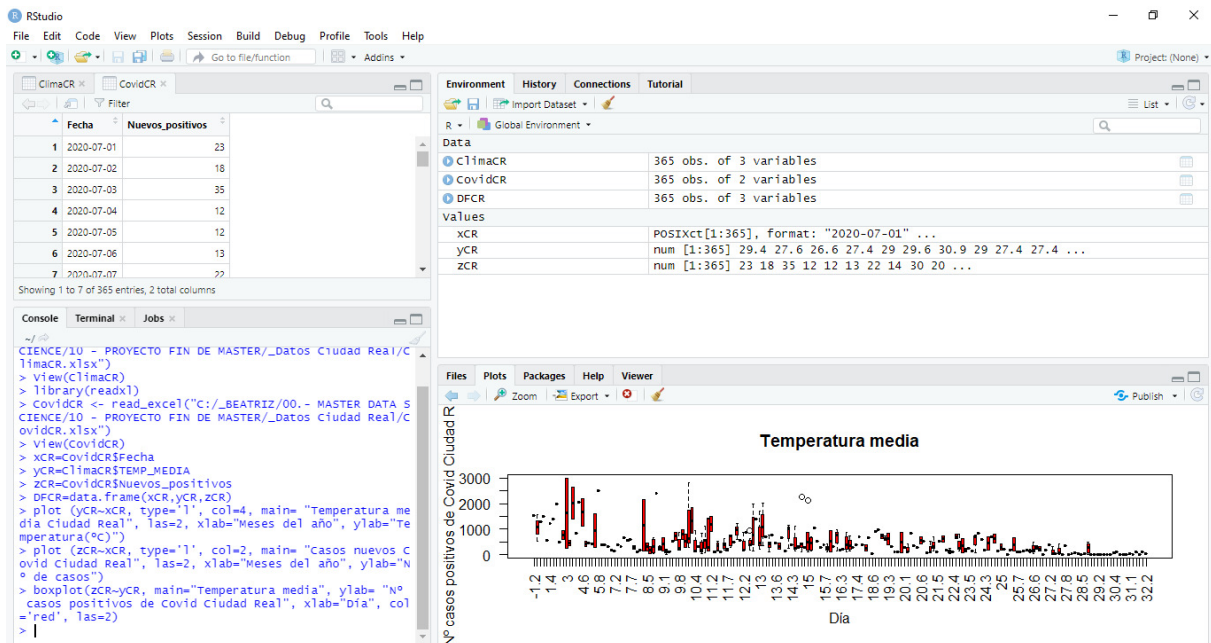
Podemos observar que al igual que en los casos anteriores, **la incidencia de nuevos casos en baja en los meses en los que la temperatura es mas alta**, como Julio a Septiembre de 2020 y a partir de abril de 2021.

También se estudia los casos Covid que se producen en función de la temperatura media y vemos que **no existe un claro patrón, debido a la gran variedad de tipos de climas** concentrados en una misma provincia.

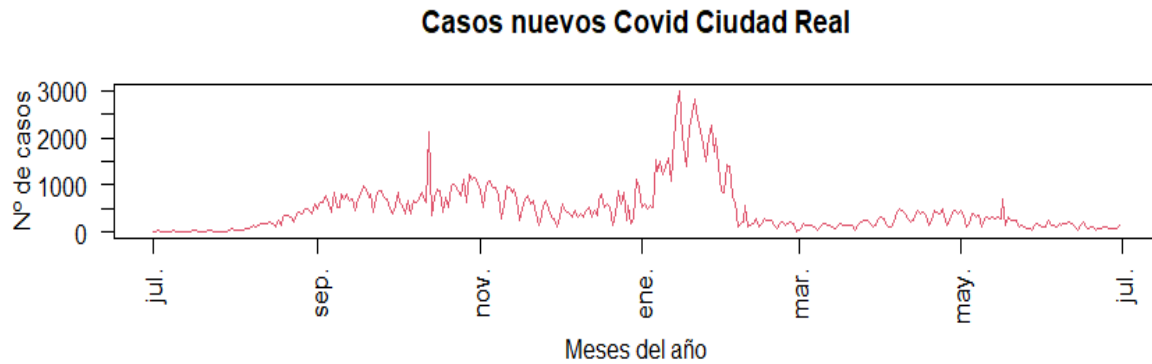


CIUDAD REAL

Se toman ahora los datos para la provincia de Ciudad Real.

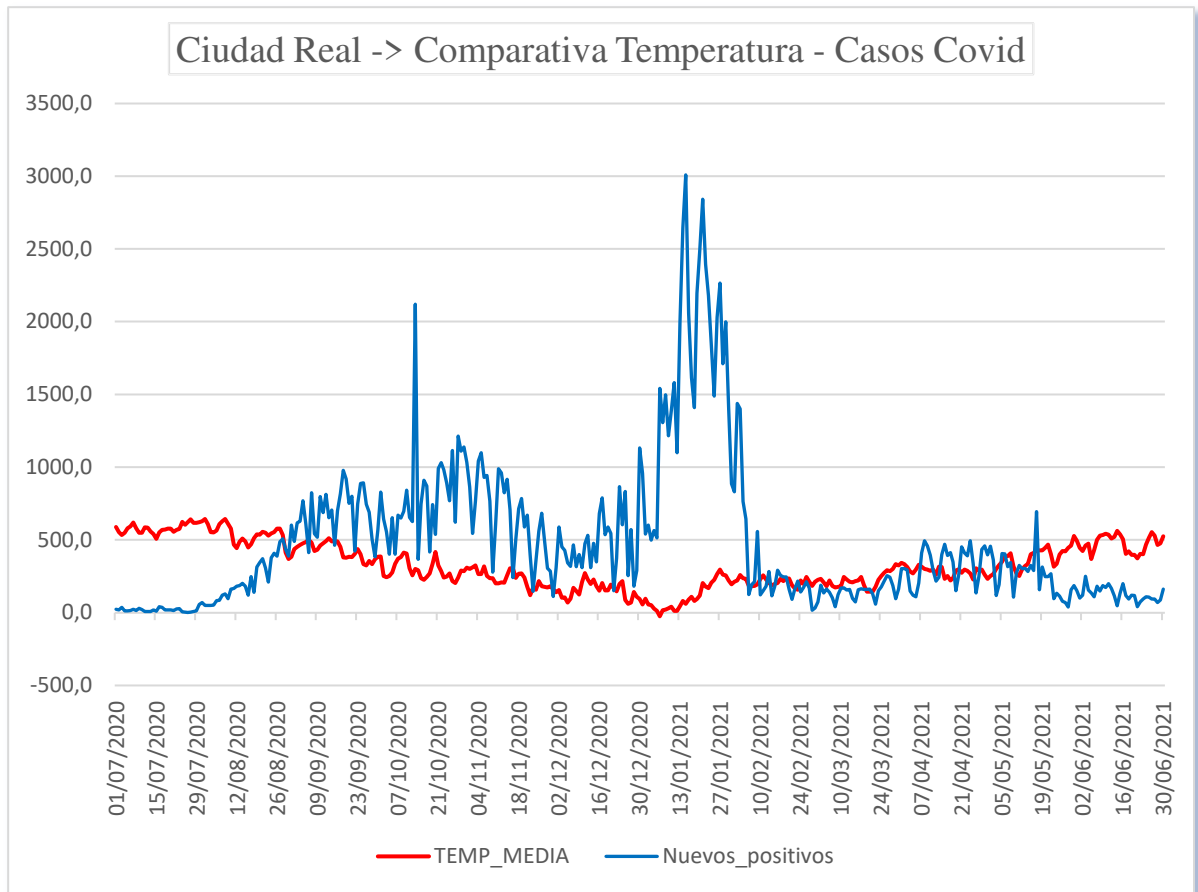


El gráfico de temperaturas, nos muestra una gráfica usual del clima característico de la provincia de Ciudad Real que es Clima Mediterráneo Interior, en el que los veranos son frescos, y los inviernos bastante fríos (ver Anexo I).

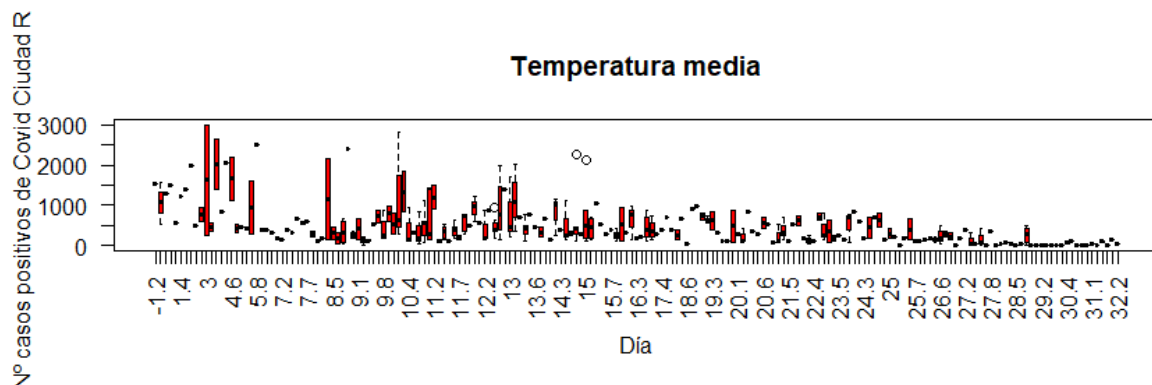


En el gráfico de nuevos casos Covid vemos que tuvieron un pequeño episodio de incidencia aunque prolongado en el tiempo, durante los meses de Septiembre a Diciembre de 2020. También se observa un episodio de gran intensidad en Enero y Febrero de 2021.

Observando el gráfico que **compara las temperaturas medias con la incidencia** podemos concluir que para esta zona climática solo se disminuye la incidencia del Covid con la temperatura en los meses de máximo calor como son: junio, julio y agosto.

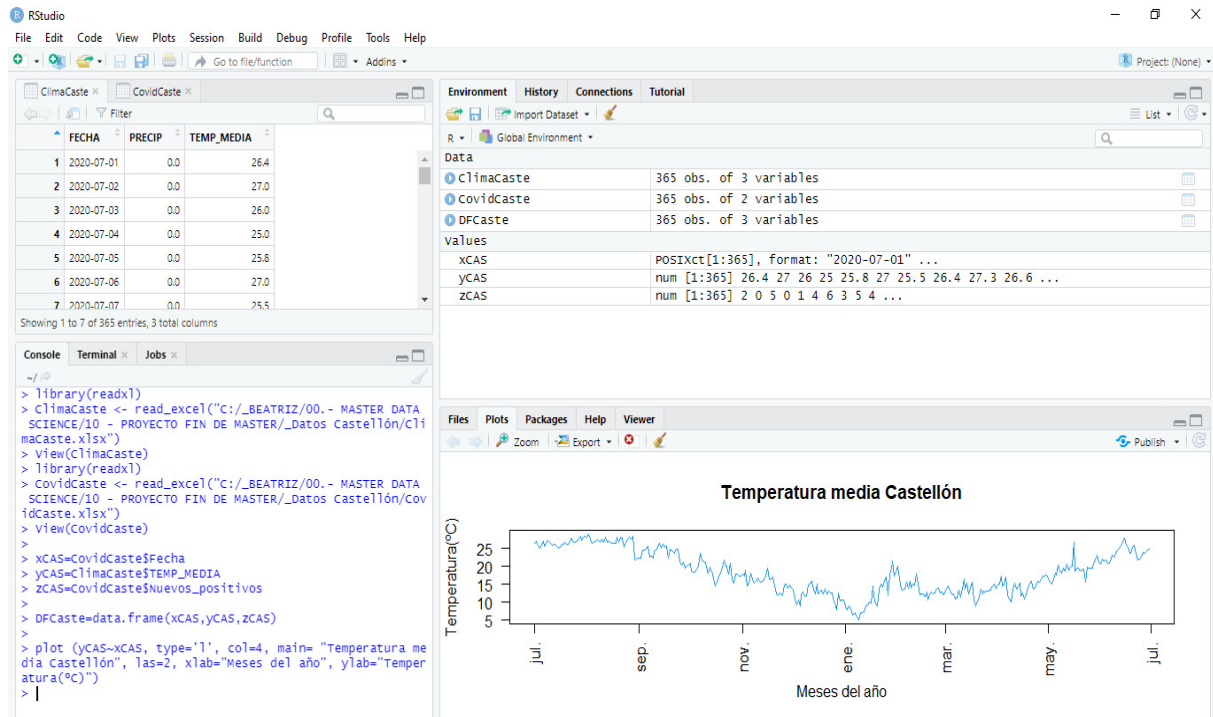


En el gráfico de número de casos por temperatura media, vemos que por debajo de los 13°C el número de casos nuevos aumenta.

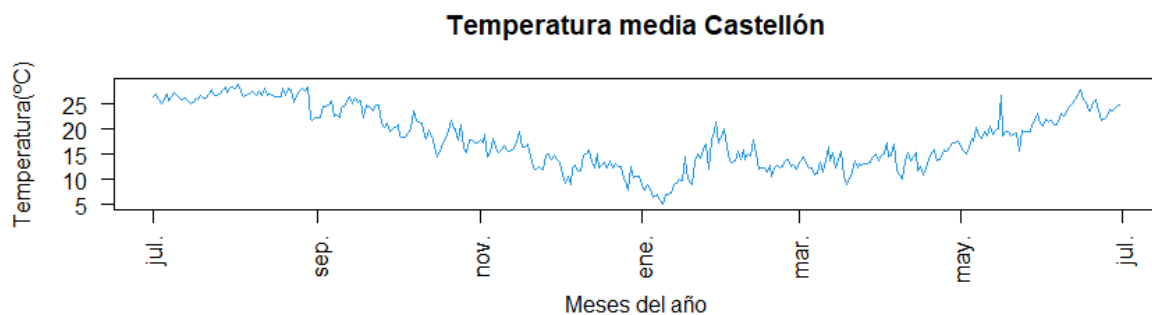


CASTELLÓN

Para la provincia de Castellón estudiamos los datos obtenidos.

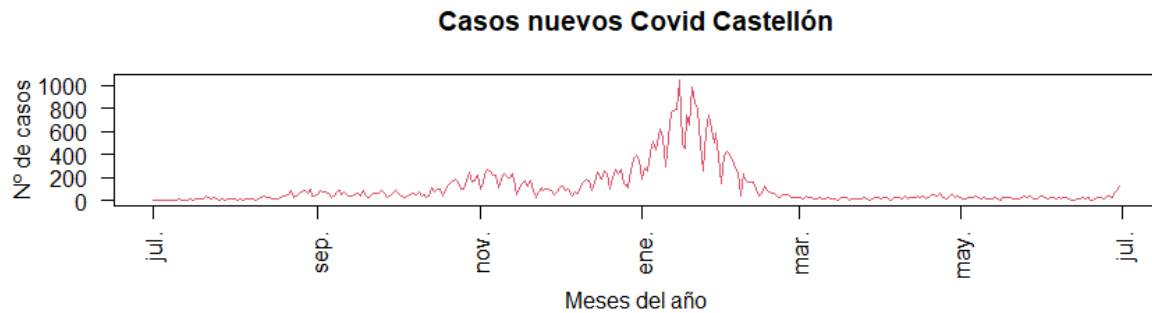


El clima en Castellón, es **Mediterráneo de costa**, y como ya hemos comentado no tiene tantas variaciones entre máximas y mínimas como en el interior de la península, los veranos no son excesivamente calurosos, al igual que lo inviernos no son extremos.

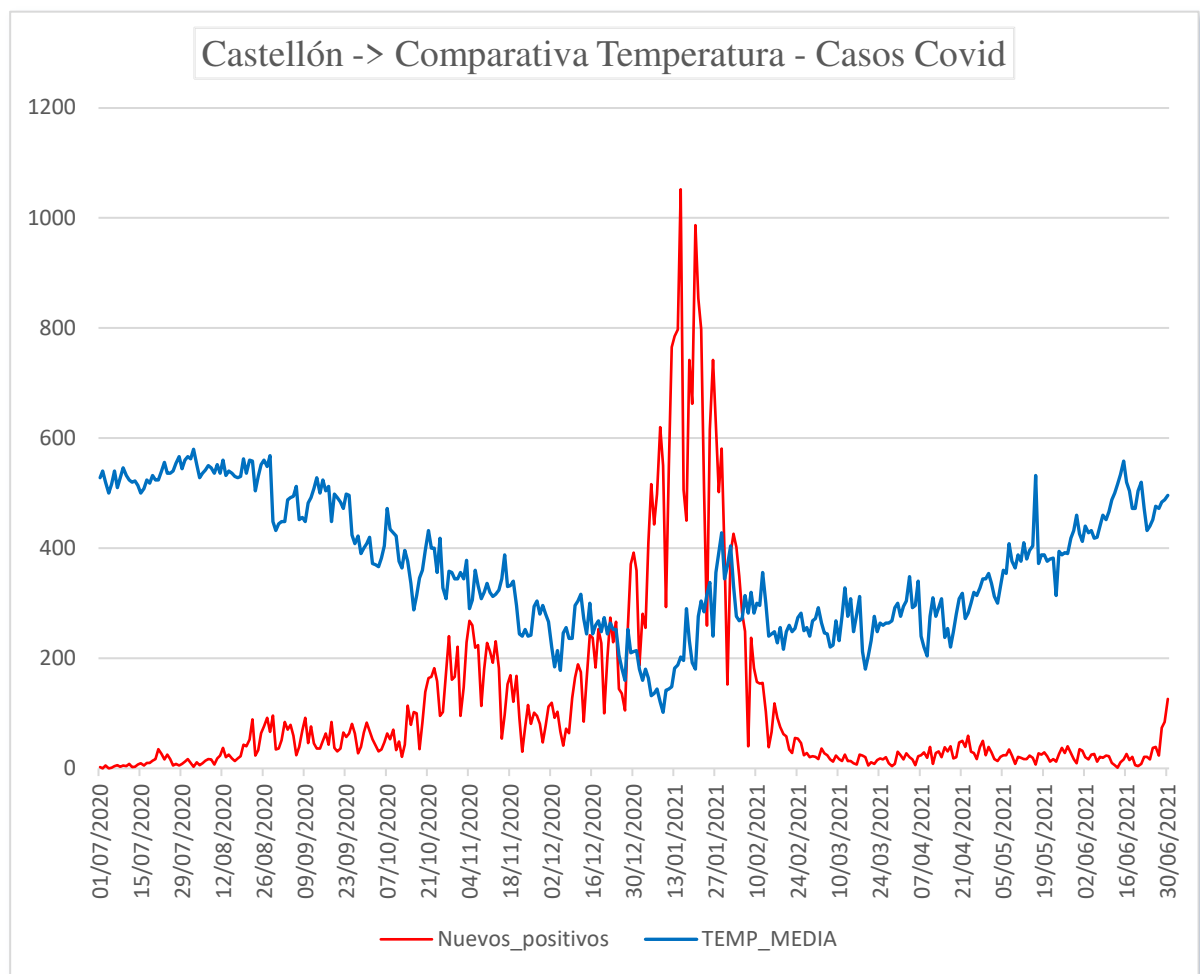


El clima en Castellón, es Mediterráneo de costa, y como ya hemos comentado no tiene tantas variaciones entre máximas y mínimas como en el interior de la península, los veranos no son excesivamente calurosos, al igual que lo inviernos no son extremos, no suelen tener temperaturas por debajo de los 10°C (ver Anexo I).

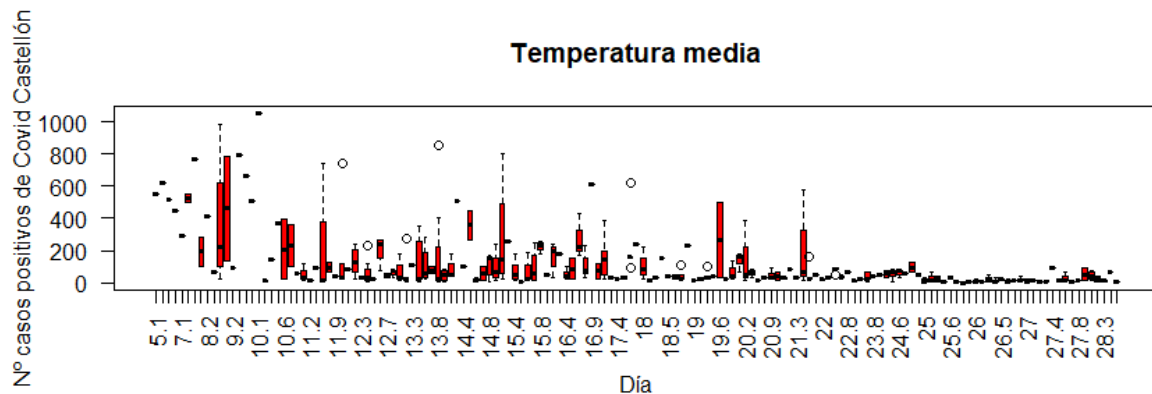
Respecto al gráfico del número de casos nuevos de Covid por día en el año seleccionado, vemos que se han producido dos episodios, uno de muy baja intensidad en el mes de noviembre de 2020 y otro de muy alta incidencia en enero de 2021.



Al estudiar esta gráfica de comparativa, concluimos que **solo existe relación directa entre la temperatura media y los nuevos casos de covid, desde mediados de octubre de 2020 hasta mediados del mes de febrero de 2021.**



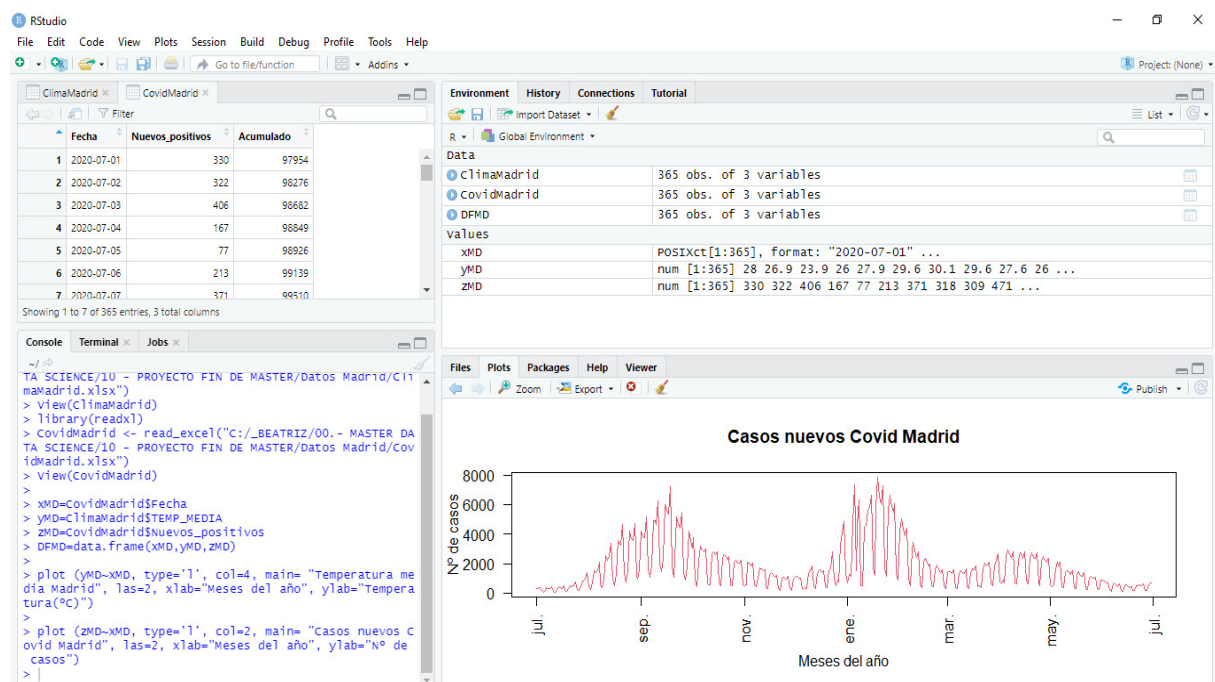
En este grafico se observa como para días cuya temperatura está por debajo de los 17 grados, la incidencia de nuevos casos de Covid autmenta.

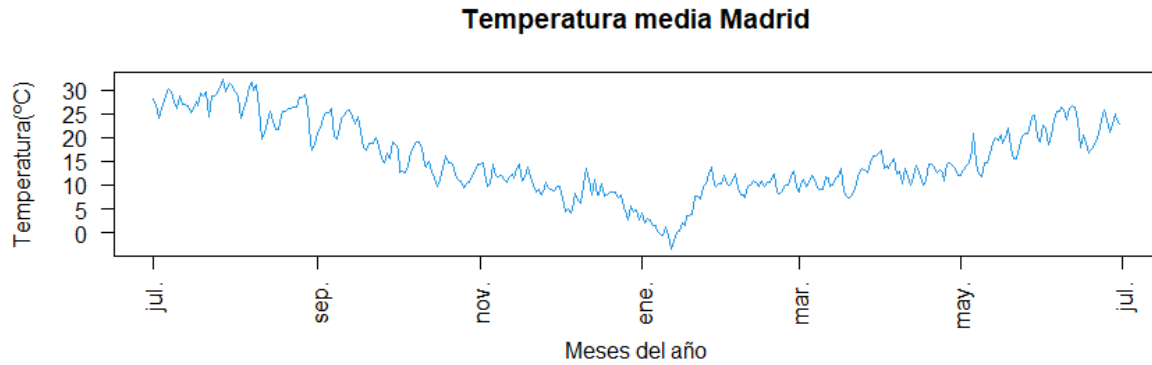


Una vez que hemos comprobado nuestro modelo en seis provincias repartidas por toda la geografía española, pasamos a **comprobar si este modelo** de correlación puede funcionar en **ciudad grades** y más pobladas como **Madrid y Barcelona**.

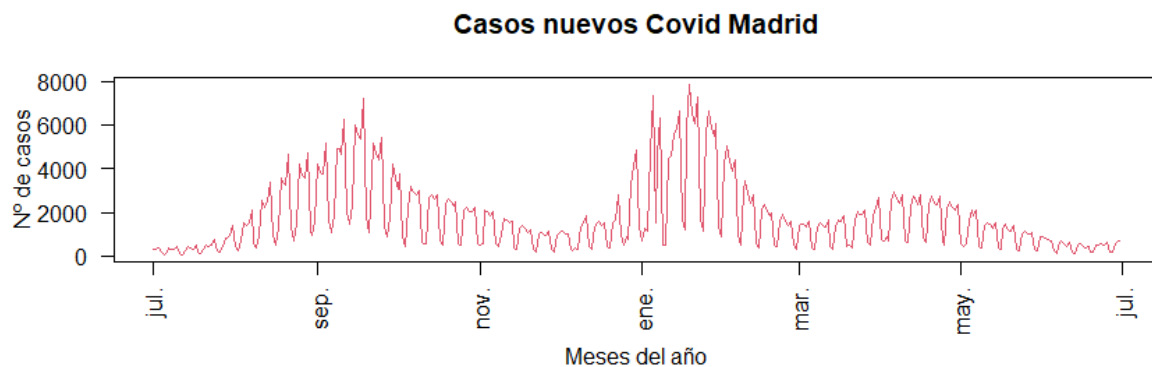
MADRID

Obtenemos los datos de incidencia de Covid en Madrid [Comunidad de Madrid, Servicio de Salud (agosto 2021)], así como los datos meteorológicos como en las provincias estudiadas anteriormente.



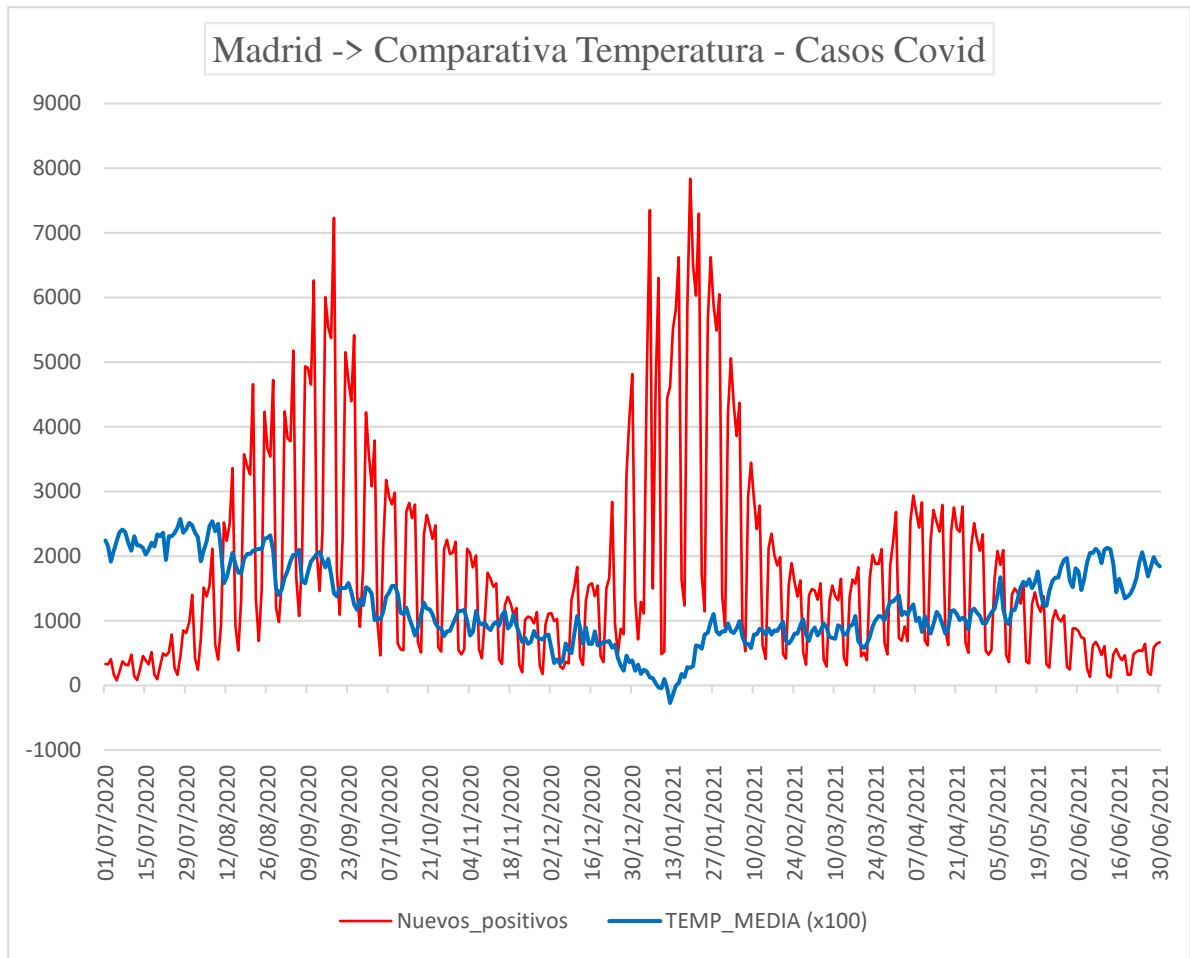


Comprobando el mapa de los Climas de España vemos que Madrid corresponde a un **Clima Mediterráneo Interior**, con inviernos fríos y veranos con temperaturas muy altas. En este gráfico vemos como la amplitud térmica en Madrid es muy amplia teniendo temperaturas extremas tanto en verano como en invierno (ver Anexo I).

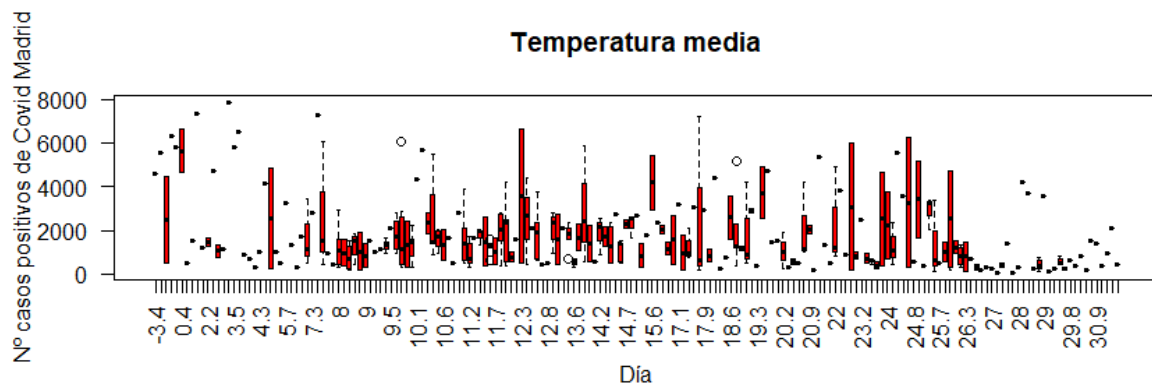


En la gráfica de nuevos casos de Covid, vemos tres “olas”, dos con alta incidencia en Septiembre de 2020 y febrero de 2021, y una de menor incidencia en el abril de 2021.

Observando el gráfico a continuación que relaciona los casos de Covid con la incidencia ocurrida, vemos que solo **en los meses de temperaturas más altas como son junio y julio se puede concluir que la elevada temperatura hace que la incidencia del Covid disminuya.**

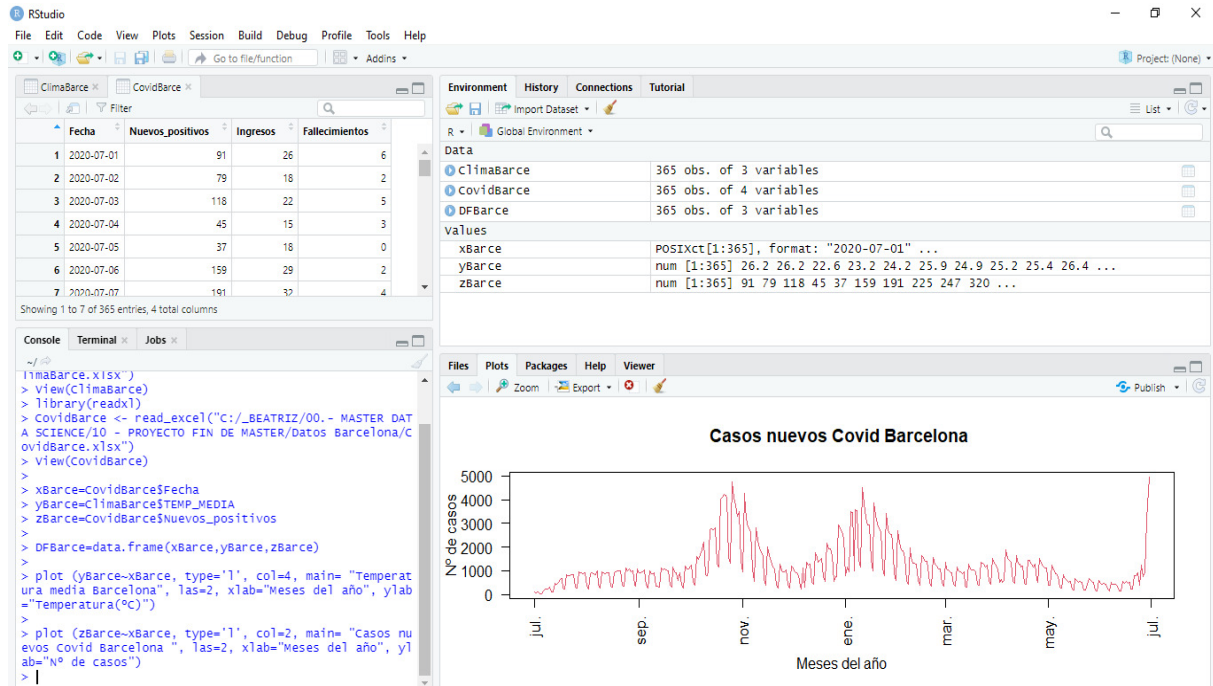


El grafico que relaciona la temperatura media con el número de casos positivos, nos lleva a la misma conclusión que el gráfico anterior, que **solo para días en los que la temperatura esté por encima de los 26 °C podemos decir que la incidencia de casos nuevos se ve atenuada.**

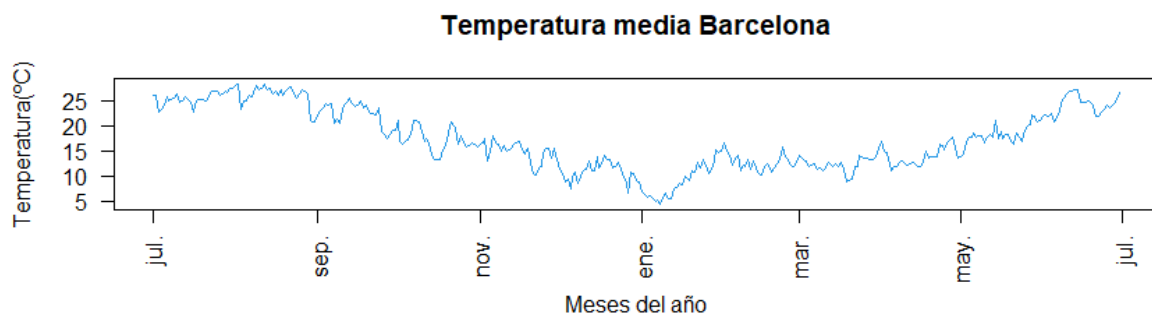


BARCELONA

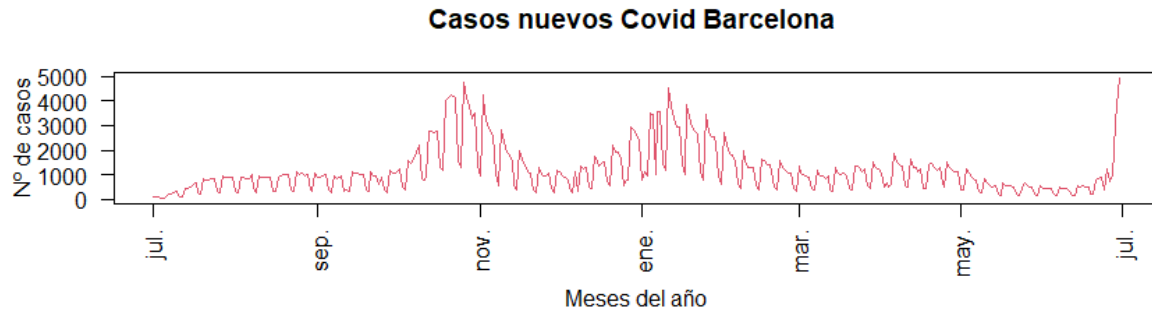
Comprobemos ahora para los datos de Barcelona [Datos covid de Cataluña (agosto 2021)], cual sería el resultado de nuestro estudio de correlación.



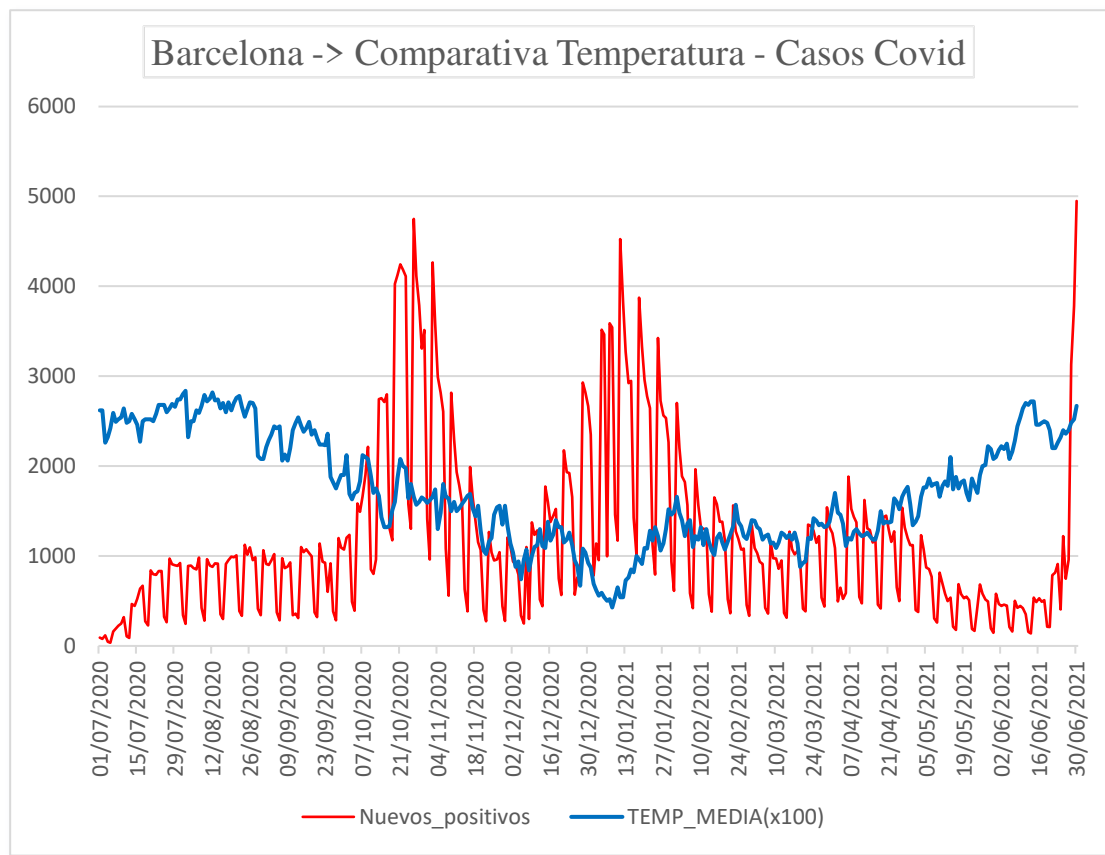
La **temperatura en Barcelona** se representa con una menor amplitud térmica, como se ve en la gráfica, los días por debajo de 10 °C son muy pocos a lo largo del año, y tampoco hay numerosos días con la temperatura por encima de 25 °C. Esta zona corresponde a un Clima Mediterráneo Costero (ver Anexo I).



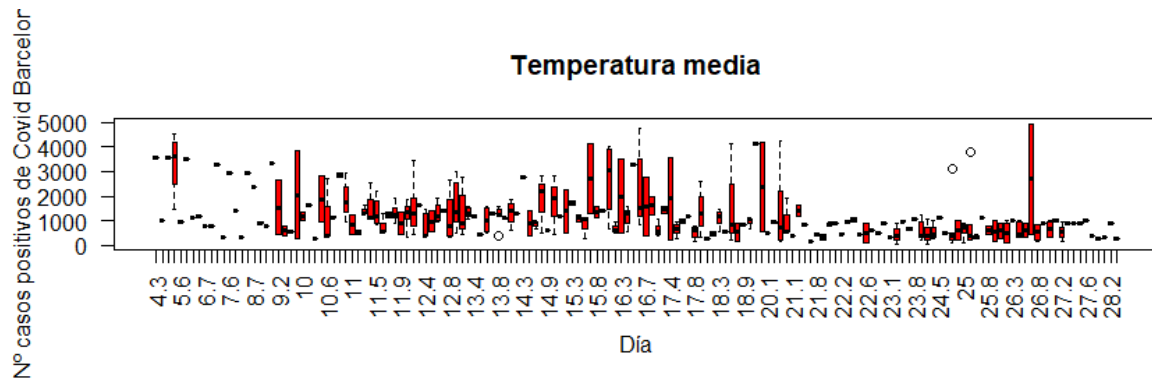
Si nos fijamos en la gráfica de incidencia de Covid en Barcelona, podemos concluir que han sucedido dos episodios de contagio masivo en los meses de noviembre y enero, y un episodio casi sin consideración en abril de 2021.



Comparando la temperatura media con la incidencia de nuevos casos de Covid, se extraen los resultados: **en los meses de temperaturas mas altas, que son de julio a septiembre de 2020 y de mayo a junio de 2021 la incidencia de covid se mantiene en niveles bajos.** Por lo tanto, se cumple el modelo de correlación presentado.



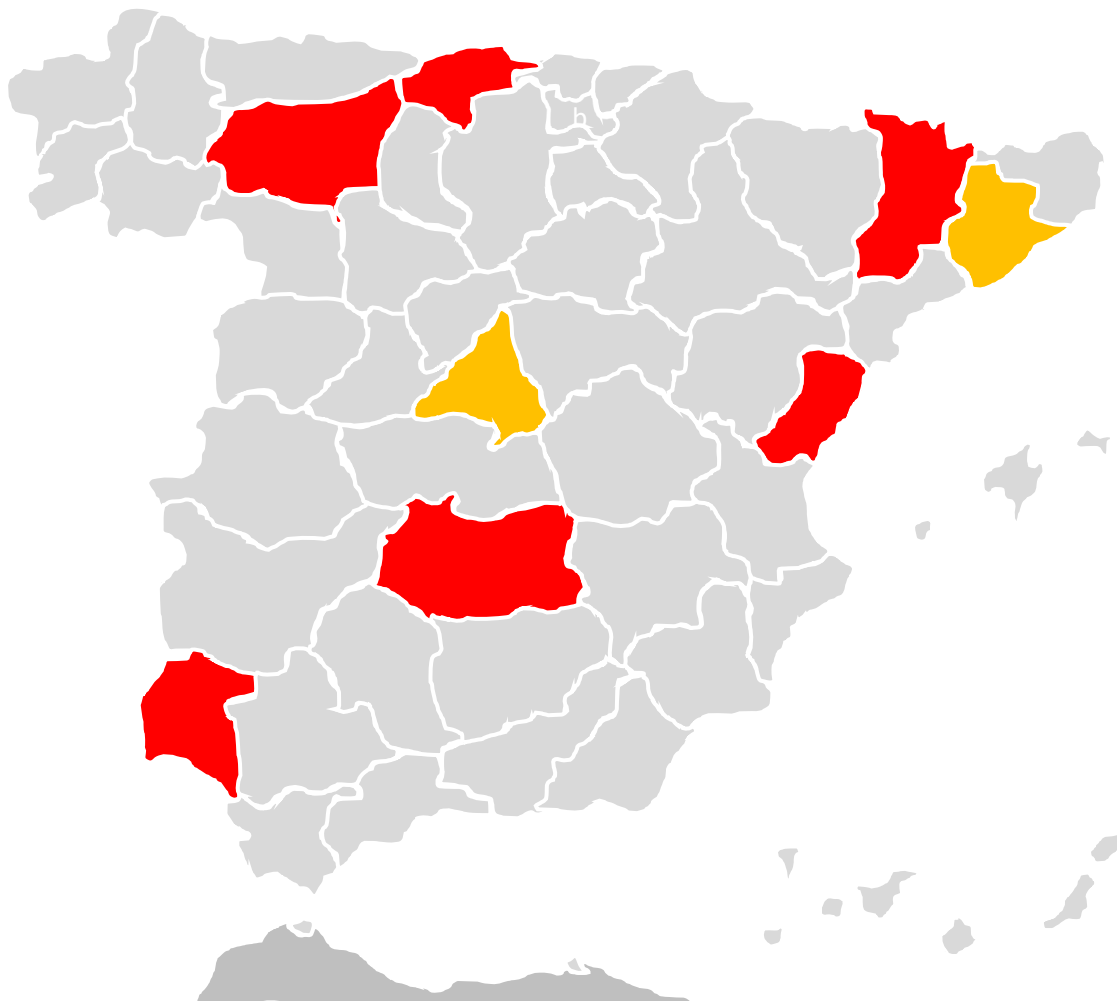
Por último en el gráfico de representación del número de casos que tenemos según la temperatura media del día, vemos que **por encima de los 20°C desciende mucho la posibilidad de contagio.**



Como vemos para los casos de **Madrid y Barcelona**, el **modelo de correlación** no se cumple con el grado de exactitud que en provincias con menos población. Factores como la densidad de población, la interacción entre los miembros de estas poblaciones, incluso la contaminación hacen que en estos territorios el modelo nos de un límite de funcionamiento para una temperatura superior a los 20°C.

5.- CONCLUSIONES

En el siguiente mapa, vemos en **color rojo las provincias de estudio**: León, Cantabria, Lleida, Castellón, Ciudad Real y Huelva y el reparto que suponen a lo largo y ancho de la geografía española. También vemos las provincias de Madrid y Barcelona donde se ha probado el **modelo de correlación** que hemos establecido en este proyecto.



Una vez hemos explotado los datos originales y generado varios gráficos para la comprensión de los resultados podemos extraer una serie de **conclusiones**.

Vemos en esta tabla un resumen de los resultados obtenidos.

PROVINCIA	Población provincial	Tipo de Clima	Media temp. Máx.	Media temp. Mín.	Amplitud térmica	Temp. umbral del estudio	Días por encima de la temperatura umbral
LEÓN	455.066	Med. interior	22,2	-0,7	22,9	12	163
HUELVA	530.386	Med. costero	28,4	7,4	21	17	217
CANTABRIA	583.904	Oceánico	21,5	3,8	17,7	20	70
LLEIDA	436.911	Oceánico + Med. interior + Montaña	28,3	0,3	28	- -	
CIUDAD REAL	492.065	Med. interior	30,2	1,2	29	13	215
CASTELLÓN	576.941	Med. costero	26,7	6,2	20,5	17	193
MADRID	6.752.763	Med. interior	29,8	1,9	27,9	26	51
BARCELONA	5.638.094	Med. costero	27,8	4,5	23,3	21	129

[Aemet.Datos climatológicos]Recogido el 15 de Septiembre de 2021

Podemos determinar que **la temperatura tiene una relación con la incidencia del Covid en la población**, cumpliéndose la máxima de que **“a mayor temperatura media ambiental, menor incidencia de Covid”**.

Esta premisa se cumple con mayor probabilidad **cuando la amplitud térmica es mayor** en la zona de estudio, es decir, cuando las medias de las temperaturas varían más entre el verano y el invierno, condición que se da en provincias como **León y Ciudad Real** que tienen un clima mediterráneo de interior.

En el caso de las provincias de **Huelva, Castellón y Barcelona**, que tienen un clima mediterráneo costero, podemos comprobar que para temperaturas umbral obtenidas del estudio de 17, 17 y 19 °C respectivamente, hay un número de días del año que se sobrepasa esa temperatura que es muy elevado, lo que quiere decir que se podrían flexibilizar las medidas de control de expansión del virus.

En el estudio de la provincia de **Cantabria**, con clima oceánico no podemos determinar con tanta rotundidad que la temperatura alta influye en el descenso del número de casos, ya que, ni tienen inviernos tan fríos como en León, ni tienen veranos tan calurosos como en Huelva. Su temperatura media anual se mantiene más constante y es más difícil determinar la influencia de la temperatura en el número de casos de contagio nuevos.

Es de mencionar el caso de la provincia de **Lleida**, en la que no podemos aplicar el modelo con rotundidad, probablemente a causa de que hay tres tipos de climas distintos en la provincia, y el gráfico de los contagios en función de la temperatura media del día no nos puede arrojar un valor umbral como si ocurre en el resto de las provincias.

Finalmente, comentar que la provincia de **Madrid** tiene una serie de características que la hacen diferente, como puede ser que las temperaturas medias máximas se ven incrementadas por tratarse de una provincia con más de seis millones de habitantes, existe una gran superficie de asfalto que influye en el clima de la zona, sobre todo en las temporadas cálidas. También hay que mencionar la gran interacción que existe entre los habitantes por tratarse de una gran ciudad y la contaminación que existe en una ciudad como Madrid, que también contribuye a que la gravedad de los contagios sea mayor.

Desde otro punto de vista, hay que comentar que el **grado de interacción social** que se produce en verano y en invierno no es igual. Si bien, en invierno se producen menos contactos sociales, con menor número de personas y en espacios interiores, en verano, al contrario, se producen las reuniones con más frecuencia, multitudinarias y en espacios exteriores y abiertos. Esto refuerza la hipótesis de que si en la época en la se realizan mayor número de reuniones y éstas afectan a mayor número de personas se producen menos contagios, es porque estas reuniones se realizan en el exterior donde las temperaturas, por ser los meses más calurosos del año, son mayores (con mucha probabilidad por encima de los 20°C).

Este estudio se ha realizado atendiendo al fenómeno meteorológico de la temperatura media diaria, pero cabría destacar que existen **otros fenómenos que también podrían tener**

relación en la transmisión y contagio del virus. Podríamos hablar de la humedad en el ambiente como posible transmisor del virus o incluso de la velocidad del viento en una determinada zona como condición limitante para su no propagación.

El objetivo de este proyecto es determinar si se podrían elaborar una serie de normas y restricciones en función de la temperatura media de un lugar y la conclusión a la que podemos llegar es que si podríamos **establecer medidas** como:

- ♦ En los meses, de noviembre a febrero, donde la temperatura media diaria está entre los 12 y los 20 °C, se limitará al nivel máximo la interacción social, ya que sabemos que estas temperaturas medias se traducen en un mayor nivel de contagio.
- ♦ En los meses de septiembre, octubre, marzo y abril, cuando las temperaturas son templadas las medidas de control y restricción de movilidad e interacción podrían ser de intensidad media, promoviendo que las reuniones de personas se realizasen en las horas centrales del día, donde se garantiza que la temperatura está por encima de los 15°C.
- ♦ Finalmente en los meses de mayo, junio, julio y agosto, donde la temperatura media sobrepasa los 20°C, se puede permitir más interacción social, sin limitación de horario, y se promoverán las actuaciones al aire libre, no en espacios cerrados, con aire acondicionado.

Aunque este estudio nos arroja unos datos con bastante exactitud sobre la correlación entre la temperatura y la incidencia de contagio del Covid, no podemos olvidar que posiblemente el factor que mayor relevancia tiene en esta incidencia de contagio sea el **comportamiento humano**, sobre el que se imponen medidas de control y restricción pero que, él mismo como individuo, tienen la última palabra en cumplirlas y respetarlas o por el contrario en obviarlas. Por esta razón somos conscientes que es necesaria la colaboración de todos para poder controlar y finalizar con esta pandemia.

6. PLANIFICACIÓN Y PRESUPUESTO

En este capítulo se exponen y se justifican el periodo de tiempo empleado y el coste económico que supone la realización de este proyecto.

Se realiza la **planificación del proyecto** con todas las tareas realizadas y con esta información se calculan los recursos utilizados en cada una de las tareas, para finalmente asignar costes a cada una de estas tareas.

6.1.- Planificación del proyecto

Se establecen las fases del proyecto y se le asignan un periodo temporal en el que se llevarán a cabo. **Las fases del proyecto** son las siguientes:

- Fase inicial: definición del problema, establecimiento de objetivos, planificación, estudio de trabajos similares.
- Fase de Análisis: especificación de requisitos, instalación del software necesario, investigación y profundización de conocimiento del lenguaje de programación.
- Fase de Desarrollo (Proceso ETL): extracción de los datos de varias fuentes de datos, transformación de dichos datos para unificarlos y finalmente la carga de datos en la aplicación para la obtención de gráficos.
- Fase de redacción de la Memoria: redacción de la documentación, generación de diagramas, elaboración de la planificación y el presupuesto, redacción de conclusiones.

Para tener una visualización completa de la planificación del proyecto creamos un **Diagrama de Gantt**. En el mismo se indican las distintas fases en intervalos temporales por semana, con su fecha de inicio, los días de duración, la línea de progreso del proyecto.

DIAGRAMA DE GANTT				JULIO						AGOSTO			
Descripción	Fecha inicio	Días	1ª s	2ª s	3ª s	4ª s	1ª s	2ª s	3ª s	4ª s			
Fase Inicial													
Investigación preliminar	01/07/2021	5											
Estudio inicial	08/07/2021	5											
Fase de Análisis													
Especificación de requisitos	08/07/2021	5											
Fase de Desarrollo (proceso ETL)													
Extracción	15/07/2021	7											
Transformación	26/07/2021	10											
Carga	09/08/2021	10											
Fase de redacción de la Memoria													
Desarrollo de la memoria	19/07/2021	30											

6.2.- Presupuesto

El presupuesto constituye la suma de todos los costes originados para la realización del proyecto.

Estos costes se pueden determinar de muy diferentes metodologías, y abarcan desde los costes de los recursos humanos necesarios para elaborar este proyecto de investigación en el tiempo marcado hasta los costes de recursos materiales necesarios.

Definición de los recursos humanos:

En un proyecto de implementación existen diferentes perfiles profesionales, entre los que destacamos los siguientes:

- ♦ *Director del proyecto:* Es la persona encargada de distribuir las tareas, de fijar planificación en función de las fases y dirige y supervisa el proyecto a lo largo de su desarrollo. En este caso este perfil es asumido por el tutor del proyecto, al que se le ha asignado según baremos actuales un coste de 26 euros/hora*.
- ♦ *Data Scientist:* profesional encargado de la investigación inicial y la extracción, transformación y carga de los datos, así como la redacción de la memoria. En este caso este perfil es llevado a cabo por la alumna. El coste por hora de un profesional de estas características es de 16 €/ hora*.

* *Los costes han sido obtenidos del sitio web [indeed21]*

Según el diagrama de Gantt, obtenemos las horas empleadas en el presente proyecto y calculamos cuál es su coste:

- Director del proyecto: 14 horas de supervisión.
- Analista programador: trabajo de 4 horas diarias los días laborables durante 6 semanas (30 días).

Recursos Humanos	Nº Horas	Coste/hora (€)	Coste (€)
<i>Director del proyecto</i>	14	26	364
<i>Analista programador</i>	120	16	1.920
TOTAL			2284

Definición de los recursos materiales:

* *Equipos de trabajo:* Consideramos que estos equipos (ordenador de sobremesa, ordenador portátil, impresora,...) pertenecientes al alumno ya están amortizados y por lo tanto no conllevan coste alguno.

* *Licencias:* Todos programas informáticos utilizados para la realización del presente proyecto (RStudio, Microsoft Office, Microsoft Excel, etc.) no suponen coste.

Recursos Materiales	Coste (€)
<i>Equipos de trabajo</i>	0
<i>Licencias</i>	0

Definición de los costes indirectos:

Los costes indirectos son aquellos que forman parte de un proyecto pero no son generados por la producción del mismo (alquiler de local, coste conexión wifi, coste de suministros como luz y calefacción, amortización de equipos, ...). Para este proyecto se asignará un 12 % del coste de personal para estos gastos indirectos.

Costes indirectos	Coste (€)
<i>Recursos humanos</i>	2.284,00
<i>Costes indirectos (12%)</i>	274,08
TOTAL	2.558,08

Beneficio:

El beneficio de un proyecto es el proceso por el cual se asegura que el proyecto cumple con su propósito, en este caso vamos a aplicar un 16 %.

Beneficio	Coste (€)
<i>Recursos + C.I.</i>	2.558,08
<i>Beneficio (16%)</i>	409,29
TOTAL	2.967,37

Impuestos:

Según la normativa española al total de los gastos del proyecto debemos sumarle el correspondiente IVA, que es de un 21 %.

IVA	Coste (€)
<i>Recursos + C.I. + Beneficio</i>	2.967,37
<i>IVA (21%)</i>	623,14
TOTAL	3.590,51

El coste total del proyecto asciende a **3.590,51 €**.

7.- MEJORAS FUTURAS DEL PROYECTO

En este apartado cabe destacar que este estudio podría ser de mucha más amplitud, ya que solo hemos estudiado seis provincias, y lo recomendable sería estudiar todas las provincias de España, e incluso por regiones dentro una misma provincia, como podría ser el caso del Principado de Asturias, donde el clima entre la zona costera y la zona montañosa es muy diferente. Los datos que nos arrojaría un estudio completo de todo el territorio español nos ofrecerían conclusiones más relevantes.

Como ya hemos mencionado, este estudio se ha centrado en la condición climática de la temperatura media de una zona, pero existen múltiples condiciones meteorológicas que también podrían influir y que podrían darnos un mapa más completo de la influencia que tienen estas condiciones meteorológicas sobre la transmisión y contagio del Covid. Dentro de estas características se pueden destacar valores como la humedad de una región, que viene dada por la cantidad de precipitación que cae en la zona, las horas de sol diarias o también la velocidad del viento, que podría también estar determinada por la altitud del lugar.

8.- BIBLIOGRAFÍA

- Inesem Business School (2019). Máster Data Science y Análisis de Datos. Data Science (2019). España: IEditorial
- Inesem Business School (2019). Máster Data Science y Análisis de Datos. Introducción a la Programación Estadística (2019). España: IEditorial
- Agencia Estatal de Meteorología – Open Data. Ministerio para la Transición Ecológica y el Reto Demográfico. Recuperado el 22 de julio de 2021, de <https://opendata.aemet.es/centrodedescargas/inicio> y en https://www.aemet.es/es/serviciosclimaticos/datosclimatologicos/efemerides_extremos?*w=0&k=can&l=1110&datos=det&x=1110&m=13&v=todos
- Instituto Nacional de Estadística. Cifras de población - Población residente. Recuperado el 23 de Julio de 2021 de <https://www.ine.es/jaxiT3/Tabla.htm?t=31304>
- Situación epidemiológica del coronavirus en Castilla y León. Recuperado el 27 de julio de 2021 en <https://analisis.datosabiertos.jcyl.es/explore/dataset/situacion-epidemiologica-coronavirus-en-castilla-y-leon/export/?disjunctive.provincia&refine.provincia=Le%C3%B3n>
- Situación epidemiológica del coronavirus en Andalucía. Recuperado el 27 de julio de 2021 en https://www.juntadeandalucia.es/institutodeestadisticaycartografia/badea/operaciones/consulta/anual/39409?CodOper=b3_2314&codConsulta=39409
- Situación epidemiológica del coronavirus en Cantabria. Recuperado el 27 de julio de 2021 en <https://www.scsalud.es/coronavirus>
- Situación epidemiológica del coronavirus en Lleida. Recuperado el 30 de agosto de 2021 en <https://dadescovid.cat/descarregues?lang=cas>

- Situación epidemiológica en Castellón. Recuperado el 30 de agosto de 2021 en <https://dadesobertes.gva.es/va/dataset/covid-19-series-casos-pdia-positiva/resource/cb50e7d2-0c0e-46b8-a359-a0fa35998577>
- Situación epidemiológica en Ciudad Real. Recuperado el 30 de agosto de 2021 en <https://sanidad.castillalamancha.es/evolucion-de-coronavirus-covid-19-en-castilla-la-mancha>
- Situación epidemiológica en Madrid. Recuperado el 30 de agosto de 2021 en <https://www.comunidad.madrid/servicios/salud/comunicados-covid-19-normativa-notas-prensa#repositorio-historico-informes-diarios-datos-2021>
- Situación epidemiológica en Barcelona. Recuperado el 30 de agosto de 2021 en <https://dadescovid.cat/descarregues?lang=cas>
- Climas de España. Recuperado el 10 de Agosto de 2021 en <https://www.geografiainfinita.com/2018/09/clima-de-espana>
- Tutorial de manipulación de datos en R. Recuperado en 4 de Agosto de 2021 en <https://r-coder.com/inicio/>

9.- ANEXOS

Anexo I.- Los tipos de clima de España

Los tipos de clima de España

Tradicionalmente se han catalogado los tipos de climas de España en cuatro grandes grupos: oceánico, mediterráneo (con algunas variaciones), subtropical y de montaña. Cada uno de estos tipos de clima tiene su influencia concreta en un área geográfica claramente delimitada.



Mapa de los tipos de clima en España. IGN.

* El clima atlántico

El clima atlántico es el de los verdes paisajes del norte peninsular. Se extiende por el norte y noroeste de la Península, desde los Pirineos hasta Galicia. Se caracteriza por la abundancia de lluvias, que suelen superar los 1000 mm, repartidas de manera regular a lo largo del año. Los días de lluvia en estas zonas superan los 150 al año.

La distribución de las precipitaciones a lo largo del año en estas zonas es bastante regular. Son zonas que se encuentran bajo la acción continua de las borrascas del frente polar. Con todo, las precipitaciones suelen alcanzar su máximo en invierno y su mínimo en verano, debido a que el anticiclón de las azores se desplaza por esas fechas al norte.

Las temperaturas suelen ser suaves debido a la cercanía del mar, lo que también redundaría en una mayor humedad. En invierno oscilan entre los 12 °C y los 15 °C y en verano rondan los 20-25 °C en cuanto a las medias.

*** El clima mediterráneo**

El clima mediterráneo es el predominante en la mayor parte de España. No obstante, su tipicidad no es homogénea, lo que lleva a hablar de distintos tipos de clima mediterráneo, en función de la zona en la que nos encontremos.

Cuenta con lluvias irregulares, que oscilan entre los 400 mm y los 700 mm anuales, concentradas especialmente en otoño y primavera. Esto ocurre por el hecho de que el anticiclón de las azores se desplaza hacia el sur, lo que hace posible que en esas épocas del año penetren con mayor facilidad las borrascas atlánticas.

Los inviernos son cortos y suaves mientras que los veranos son largos y calurosos. La temperatura media anual ronda entre los 15 °C y los 18 °C.

El clima mediterráneo puede subdividirse en tres grandes grupos o subtipos climáticos, de acuerdo con las variaciones de temperatura y precipitación, que son distintas en cada zona de España.

Clima mediterráneo marítimo

El clima mediterráneo marítimo comprende la costa mediterránea peninsular a excepción del sureste, la costa sur del Atlántico, el archipiélago balear y Ceuta y Melilla.

En estas zonas, las precipitaciones son escasas. Se sitúan entre los 800 y los 300 mm anuales. Con todo, varían entre la costa mediterránea y la zona del sur del Atlántico (Huelva y Cádiz).

En la costa mediterránea las precipitaciones son más bien escasas ya que las borrascas atlánticas llegan muy esporádicamente y cuando lo hacen no presentan ya demasiada fuerza.

Mazizo montañoso del Montgó, en Alicante, visto desde la plana de San Antonio.

En cambio, en la costa del sur de la costa atlántica española las precipitaciones son más frecuentes por la mayor presencia de esas mismas borrascas.

En cuanto a las temperaturas, no tienen tantas variaciones entre máximas y mínimas como en el interior peninsular y los veranos no son demasiado calurosos, por la influencia del mar. Por su parte, los inviernos son suaves y la media del mes más frío no baja de los 10° C.

Clima mediterráneo de inviernos fríos

El clima mediterráneo con invierno frío (erróneamente denominado “mediterráneo continentalizado”) es el segundo tipo. Se localiza en la Meseta, la depresión del Ebro, parte del Guadalquivir y la zona del norte de la provincia de Alicante.

Se caracteriza por su aislamiento de las influencias marítimas, lo que le da ese carácter “continentalizado”. De este modo, las temperaturas son muy extremas y oscilan, de media, entre 25 °C y los -13 °C. Las amplitudes entre la diferencia de las mínimas y las máximas se sitúan sobre los 16 °C.

Los inviernos son largos y muy fríos y los veranos muy calurosos. Además, las precipitaciones son escasas, en torno a los 500 mm, y aparecen en forma de tormenta en los meses de julio y agosto.

A su vez, dentro de este tipo de clima, propio del interior de España, podemos encontrar algunas diferencias.

Submeseta norte y zonas altas de Cuenca y Teruel

De este modo, nos encontramos con el subtipo de la submeseta norte y las tierras altas de Teruel y Cuenca, donde los veranos son frescos (con una media de 22° C en el mes más cálido) y los inviernos son considerablemente fríos, con entre 6°C y -3° C en el mes más frío del año.

Submeseta sur y valles del Ebro

En el caso de la submeseta sur y los bordes de los valles del Ebro, los veranos se presentan más calurosos y la media del mes más cálido suele superar los 22° C. Los inviernos son fríos, al igual que en la submeseta norte.

Extremadura e interior de Andalucía

En Extremadura y el interior andaluz los veranos son muy calurosos, los más calurosos de todo el interior peninsular. En cambio, los inviernos son más moderados que en las dos submesetas, con temperaturas que en el mes más frío se quedan entre los 6 °C y los 10° C.

Clima mediterráneo seco subdesértico o estepario

El clima mediterráneo seco es propio del sureste de España, en las zonas de Murcia, Alicante y Almería, así como en la zona media del Valle del Ebro. Las lluvias son extremadamente escasas, menos de 300 mm al año, lo que convierte estas zonas en áreas muy áridas.

Desierto de Tabernas, en Almería. Maksym Abramov.

Esa aridez en el sureste se debe a que la zona se encuentra al abrigo de las borrascas atlánticas, gracias a que el relieve de las Cordilleras Béticas les protegen. Pero también llegan con dificultad las borrascas mediterráneas y es frecuente la influencia seca de África.

De este modo, sólo las perturbaciones que se forman en la zona del Estrecho en el Mar de Alborán provocan precipitaciones.

En el caso de la zona media del Valle del Ebro, la aridez tiene su causa en que el Sistema Ibérico protege a la zona de las borrascas atlánticas. Las posiciones de las cordilleras costero catalanas hacen lo propio con la influencia de las borrascas del mediterráneo.

En estas zonas son muy frecuentes los períodos largos de sequía. Las temperaturas se parecen a las del clima mediterráneo típico, aunque el calor en verano suele ser más intenso.

Estepa cálida y estepa fría

Las temperaturas permiten distinguir entre estepa cálida, en el caso de la zona costera del sureste, con medias anuales por encima de los 17° C, y estepa fría, que es la característica de La Mancha y Albacete así como en la zona media del valle del Ebro, con temperaturas medias anuales por debajo de esos 17° C e inviernos moderados y fríos.

*** Clima subtropical**

El clima subtropical solo aparece en el archipiélago canario debido a su cercanía con el trópico de Cáncer y la costa árida de África. Se caracteriza por la presencia de los vientos alisios y las corrientes de agua fría.

Las temperaturas son cálidas durante todo el año, entre los 22 °C y los 28 °C de media, mientras que las precipitaciones pueden ser abundantes o escasas, dependiendo de la zona, se concentran en invierno.

*** Clima de montaña**

El clima de montaña es propio de los grandes sistemas montañosos que pueblan España y comprende los territorios situados por encima de los 1.000 metros.

Es el caso de los Pirineos, el Sistema Central, el Sistema Ibérico, la cordillera Penibética y la cordillera Cantábrica. En estas zonas los inviernos son muy fríos y los veranos frescos (ningún mes supera una media de 22 °C). Las temperaturas se caracterizan por una media anual baja, siempre inferior a los 10° C.

Las precipitaciones son muy abundantes a medida que aumenta la altitud, en general, en forma de nieve. Suelen superar los 1.000 mm al año. Las vertientes de las montañas que miran al norte son más frías.

Clima de montaña de la España húmeda

Las montañas situadas en la España húmeda (Pirineos y Cordillera Cantábrica) tienen a partir de los 1.200 metros al menos el doble de precipitaciones que las partes más bajas de esas mismas zonas montañosas. De este modo, las precipitaciones suelen ser en forma de nieve en invierno.

Clima de montañas del interior peninsular

Las montañas del interior son el Sistema Ibérico, el Sistema Central, los puntos más elevados de los Montes de Toledo y de Sierra Morena. Todos ellos constituyen excepciones dentro de la España seca, con precipitaciones de entre 1.000 mm. y 1.500 mm. anuales, muy por encima de las de las zonas geográficas en que se encuentran.

Clima de montaña del sur peninsular

En el caso del sistema Bético y Penibético, la influencia de la altitud sobre el aumento de las precipitaciones sólo se percibe a alturas considerables, como es el caso de Sierra Nevada. Esto sitúa su tipicidad en línea con las de las montañas situadas en zonas intertropicales.