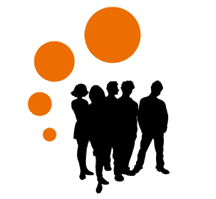
Universidad de Castilla-La Mancha

Facultad de Ciencias Económicas y Empresariales



Proyecto Datathon 2021

Un análisis del impacto del coronavirus sobre el comercio de bienes agrícolas.

Presentado por:

Pedro José Piqueras Martínez

Carlos Caballero Picazo

Felicidad María Lozoya

Albacete, 2021

# **INDICE**

1.INTRODUCCIÓN 3

2. OBTENCION Y TRATAMIENTO DE LOS DATOS 4

3. ANALISIS Y RESULTADOS. 4

4. MODELIZACIÓN DEL COVID Y LOS PRECIOS DE EXPORTACIÓN 11

5. CONCLUSIONES.. 12

6. PERSPECTIVAS DE FUTURO.. 13

## **1.INTRODUCCIÓN**

Con motivo de la competición organizada por Cajamar presentamos este trabajo consistente en un análisis descriptivo y concluyente sobre la evolución del precio y el valor de las exportaciones españolas de bienes agrícolas. Adicionalmente presentamos modelos para el pronóstico de la evolución del coronavirus y de los precios de un bien agrícola en particular.

En el aspecto metodológico del proyecto, hemos empleado el lenguaje de programación R como herramienta de trabajo. Para facilitar la reproducción del código, hacerlo más legible y sencillo hemos empleado las herramientas que proporciona el paquete “tidyverse”, es decir “pipelines” y “tibbles” los primeros facilitan enlazar operaciones sobre una base de datos y los segundos son una versión mejorada del “data.frame” de R base que entre otros aspectos de rendimiento proporcionan una salida de consola más visible.

En lo referente a la programación hemos estructurado el proyecto en diferentes scripts. El script librerías como su nombre indica incluye la llamada a los paquetes externos a R base que empleamos a lo largo del proyecto para no repetir este código en el resto de script y tener las librerías ubicadas para cuando se requieran. El archivo funciones son una serie de rutinas sencillas desarrolladas con la finalidad de simplificar el resto del código o realizar cálculos que son recurrentes a lo largo del trabajo. Por otra parte el documento “ObtencionTratamiento” incluye el proceso de lectura de las bases de datos y las transformaciones pertinentes para adaptarlas al formato requerido por las funciones de R así como el tratamiento de los valores ausentes y atípicos, sobre este documento nos referiremos más adelante en el apartado correspondiente. El script “AnalisisResultados” como su nombre hace referencia incluye el análisis de los datos de los que disponemos. Finalmente “ModelosCovid” y “PronosticoPrevision” son archivos que incluyen las líneas de código con que programamos los modelos de coronavirus y precios respectivamente. Para facilitar la navegación y lectura el código de cada script está separado por bloques.

Aunque el proyecto que hemos realizado sea intensivo en el uso de técnicas estadísticas y econométricas no por ello dejamos de lado un aspecto fundamental del análisis económico que consiste en la observación y el razonamiento; por ello a lo largo del trabajo tratamos de partir de las técnicas más sencillas (Como números índice) a las más complejas (Como redes neuronales) en búsqueda del equilibrio entre eficiencia y tecnología.

El hecho de participar en este evento nos ha aportado experiencia en el trabajo en equipo y en el análisis de datos que nos lleva a agradecer poder formar parte de esta competición.

## **2. OBTENCION Y TRATAMIENTO DE LOS DATOS**

En este apartado trataremos el proceso de transformación de las bases de datos para adaptarlas al formato requerido para la aplicación de las funciones asociadas al lenguaje R y sus paquetes externos, además de la solución de problemas como es el tratamiento de valores ausentes, valores atípicos y la heterogeneidad de los datos. En este sentido tratamos de transformar las bases de datos de forma que sean “anchas” y no “largas” es decir, tengamos una fila por observación y una columna por variable. Adicionalmente convertimos variables cualitativas como es la región o el tipo de producto en factores y las variables numéricas con forma de fechas en fecha.

Para el tratamiento de los valores atípicos empleamos la distancia de Mahalanobis[[1]](#footnote-1) para detectarlos, representándola gráficamente eliminamos aquellas observaciones con una distancia inusual en comparación al resto. En cuanto a valores ausentes, dependiendo del caso, los imputamos mediante el algoritmo de los k-Vecinos más cercanos o eliminamos las observaciones y variables correspondientes. A continuación trataremos los aspectos característicos de cada base de datos en este proceso, en el script “ObtencionTratamiento” están las líneas de código empleadas con esta finalidad.

# **3. ANALISIS Y RESULTADOS.**

En este apartado trataremos el proceso de análisis realizado en el marco de los objetivos de la competición, en vista general, mostraremos los resultados en una escala regional descendente, desde nivel mundial a particular de un producto y país.

Puesto que tratar a la vez todas las regiones y todos los productos es una tarea ardua y poco fructífera hemos optado por hacer un análisis general desde dos perspectivas, regiones y productos y un análisis específico para los productos y países más relevantes en términos cuantitativos.

Para lograr un análisis general por regiones hemos condensado los datos de productos en un índice construido sobre una media ponderada de las exportaciones (O valor según corresponda) según el peso del total exportado de cada producto en cada país sobre el total de exportaciones del país pertinente, de esta forma en lugar de una observación por momento temporal para cada producto en cada país tenemos una única observación por momento temporal y país. El valor 100 del índice será el primer valor no nulo de la media ponderada. Esto nos permite un vistazo general y la realización de gráficos que de lo contrario se asemejarían a telarañas.

Para el análisis general por productos tomaremos la suma del total de exportaciones del producto en cada momento temporal y para poder comparar temporal y categóricamente construimos un índice tomando como base cien el primer valor no nulo.

En el análisis específico filtramos las bases de datos en un único país y producto, para poder apreciar mejor su relación con la evolución del coronavirus procedemos a desestacionalizar la serie de datos de productos mediante la descomposición temporal STL, ya que los productos agrícolas poseen remarcada estacionalidad debido a que su oferta depende de la fecha del año dado que cada producto tiene su propio ciclo biológico.

En cuanto a los datos de coronavirus decidimos emplear los datos de defunciones, por dos razones, son más difíciles de manipular y son más fáciles de comparar, esto hace referencia a que en los primeros momentos del Covid-19 se hacían menos test que en la actualidad.

Un aspecto a remarcar es el empleo del coeficiente de variación para medir la volatilidad del comercio de cada producto. Empleamos este coeficiente y no la desviación típica para facilitar la comparación espacial y por producto.

El valor de las exportaciones a nivel mundial de las que disponemos de información oscila entre un mínimo de 0 y un máximo de 666541 millones de euros, con un valor promedio de 37755 millones. Datos que obtenemos de Merca Barna.

El precio presenta un mínimo de 7 céntimos y un máximo de 5,89 euros mientras que el promedio de los precios agrícolas se sitúa en 1,32 euros.

El producto que presenta una mayor volatilidad a nivel internacional (A nivel nacional es el comercio de Lechuga Iceberg con Alicante) es el comercio de verduras con Francia, con un coeficiente de variación de 3,86 seguido por el comercio de Rábanos con Holanda, con una volatilidad de 3.31. El producto con un comercio menos volátil es el Melón Galia con Argentina con una volatilidad prácticamente nula sobre el valor y de 0.038 sobre el precio. Si miramos la volatilidad en la base de datos “ComercioExterior” son las exportaciones españolas de guisantes y cerezas con Bulgaria los productos con una mayor volatilidad.

La Figura 6, a nuestro juicio la más interesante en este apartado de análisis internacional nos muestra mediante un gráfico de caja y bigotes la distribución estadística del índice de comercio por países. Nos permite observar que la mayor parte de las regiones toman valores inferiores a 100 a lo largo de la serie, mientras otras toman valores por encima y por debajo, siendo las regiones que menos aquellas que solo toman valores del índice por encima de 100.

Regiones como Bélgica, Chile, Francia, Sudáfrica o Valladolid muestran resultados positivos con un índice que se distribuye generalmente por encima de 100. En cambio regiones como Argentina, Castellón, Colombia, Kenia o EEUU muestran resultados particularmente negativos. En un punto intermedio nos encontramos con Badajoz, Canadá o Segovia. Remitiéndonos a estos casos señalados podemos analizar tanto las tasas de variación interanuales por mes como la tasa de variación del total de comercio entre 2019 y 2020.

En la Figura 2 podemos apreciar que entre 2019 y 2020 las exportaciones totales de forma general se reducen con la excepción de países como Francia, Alemania o Luxemburgo. Siendo Chipre el caso más destacado de reducción de volumen de exportaciones (Exportaciones de España hacia estos países)

Figura 1. Distribución estadística del índice de comercio por regiones.[[2]](#footnote-2)

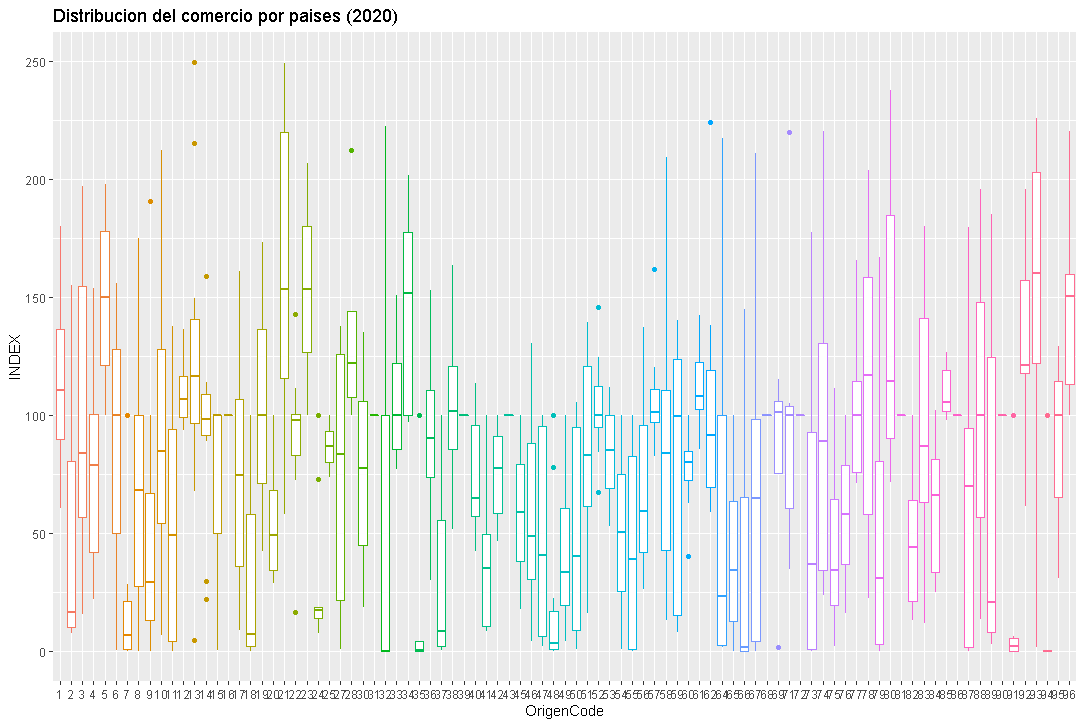
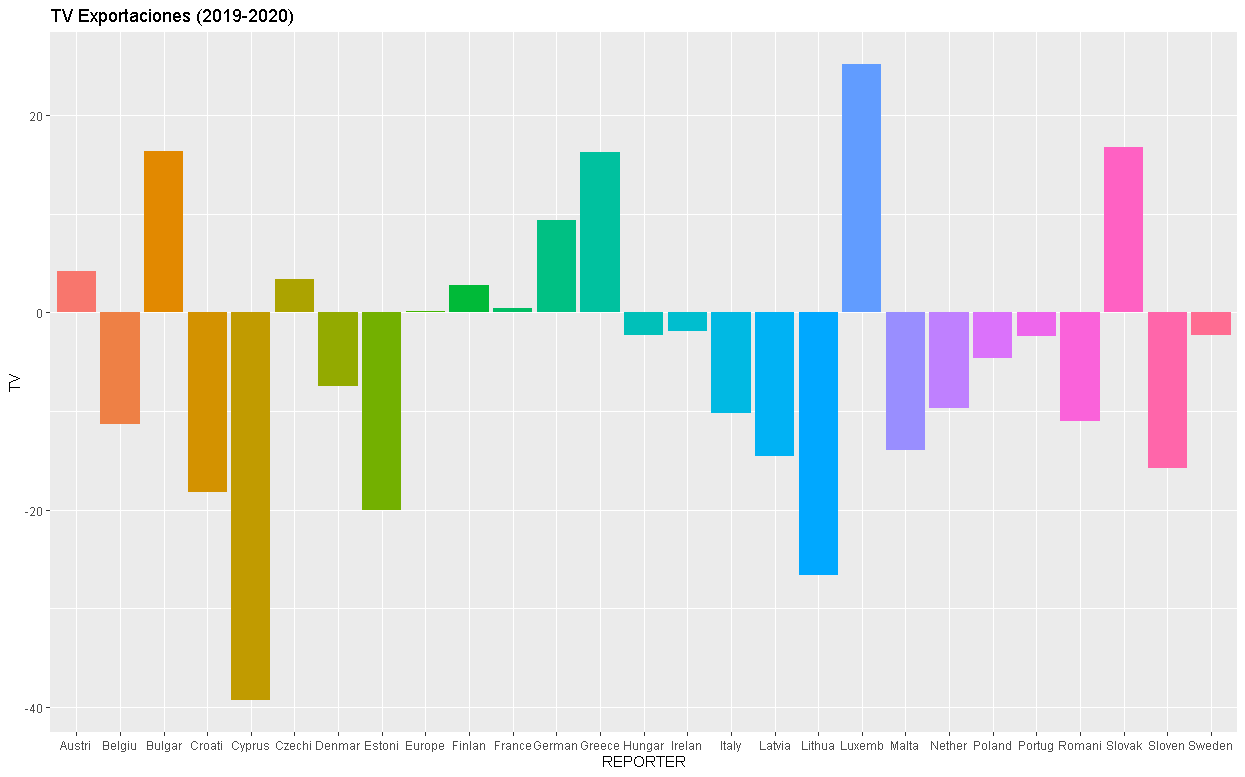


Figura 2. Tasa de variación del total de exportaciones entre 2019 y 2020 en la Unión Europea



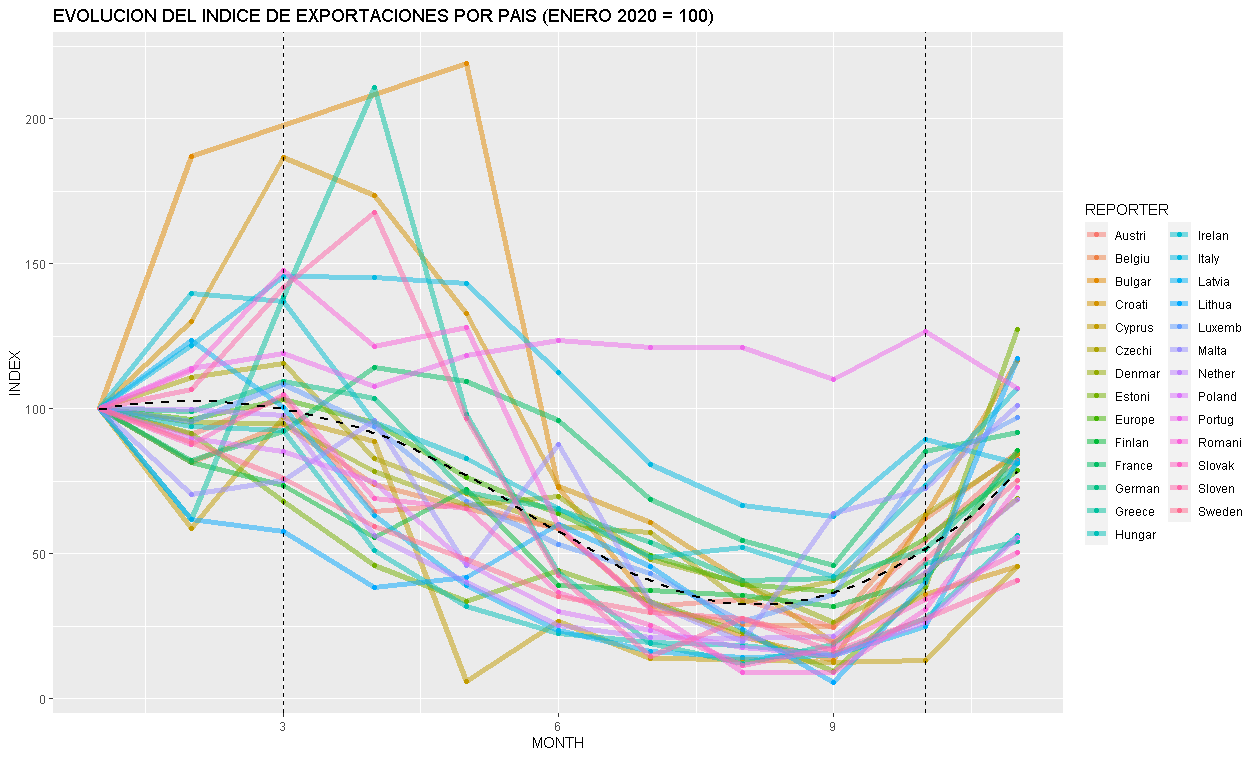
En el Cuadro 1 mostramos las 10 mayores caídas de las exportaciones totales entre 2019 y 2020 por mes, es entre verano y finales de verano cuando más se ven afectadas las exportaciones.

Cuadro 1. Caída de las exportaciones por mes entre 2019 y 2020 en la UE



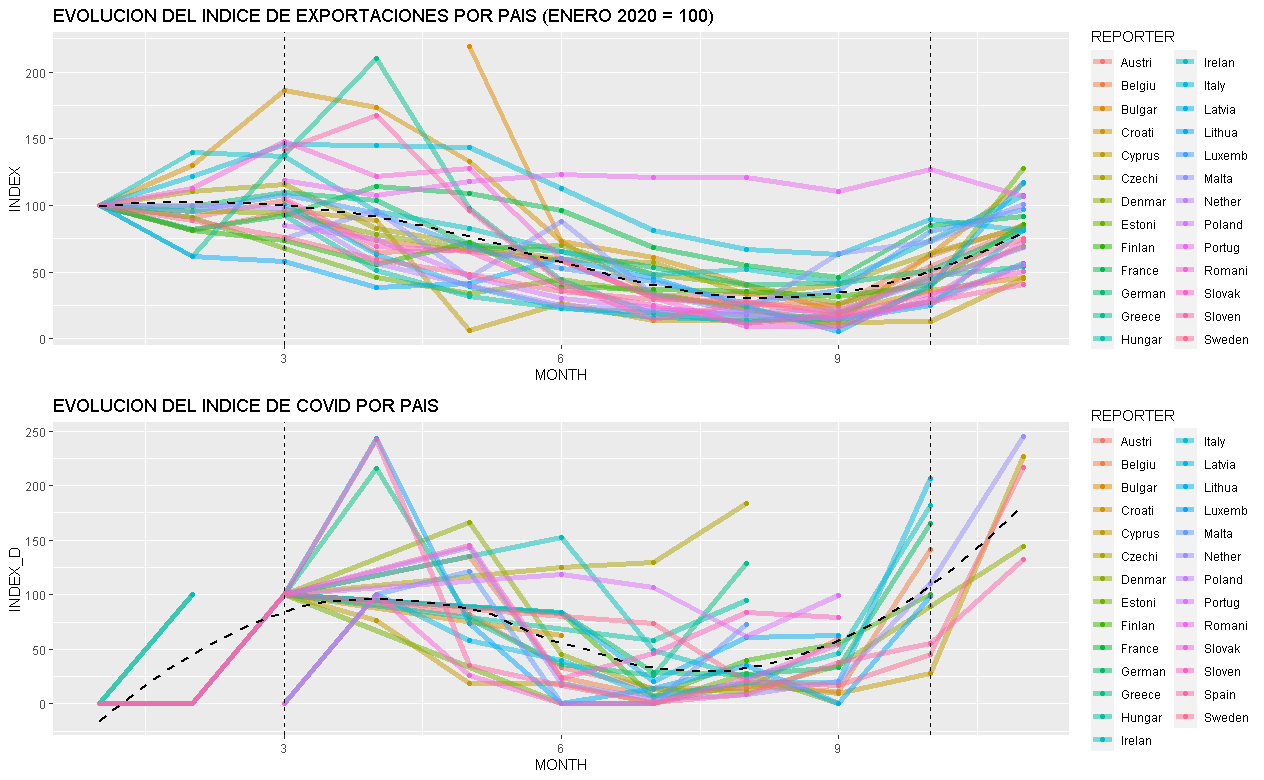
En la Figura 3 mostramos la evolución del índice de exportaciones por país en la UE, podemos apreciar como coincidiendo con la crisis sanitaria las exportaciones sufren una tendencia a la baja, mientras que en cambio la segunda ola de la pandemia no provoca un efecto tan contundente:

Figura 3. Evolución del Índice de exportaciones por producto en la UE



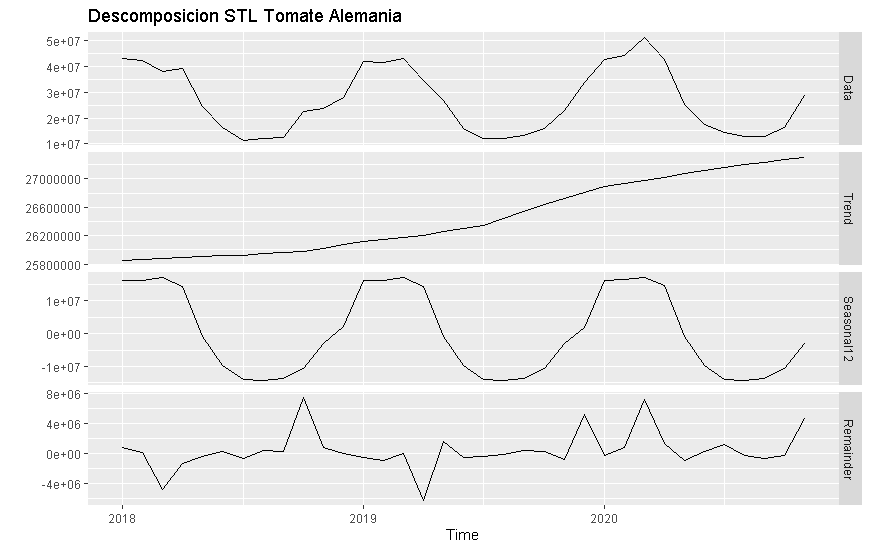
La misma gráfica podemos compararla con la evolución de un índice de coronavirus tomando como base 100 la primera muerte no nula (Para poder hacer una comparación espacial), de esta forma apreciamos que el impacto de la primera ola es mayor, como observamos en la Figura 4.

Figura 4. Comparativa de la evolución de las exportaciones y el coronavirus.



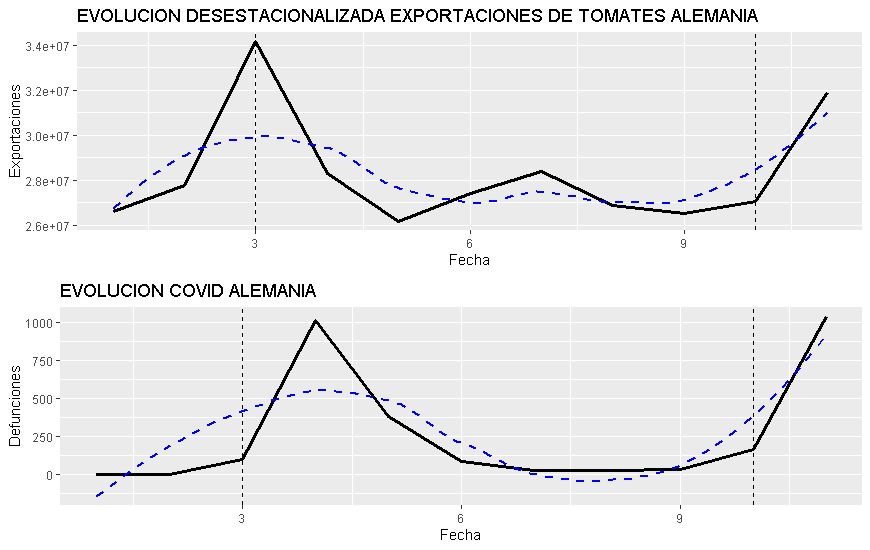
Para un análisis específico nos centramos en las exportaciones de tomate a Alemania por ser el producto y región más relevantes respectivamente, con esta finalidad, si queremos apreciar la relación entre el coronavirus y el valor de las exportaciones, tenemos que desestacionalizar esta última serie, para lo cual empleamos la descomposición temporal STL (Figura 5). Podemos apreciar que por la propia naturaleza del tomate tenemos una amplia estacionalidad ligada al ciclo natural de este vegetal

Figura 5. Descomposición temporal de las exportaciones de tomate hacia Alemania.



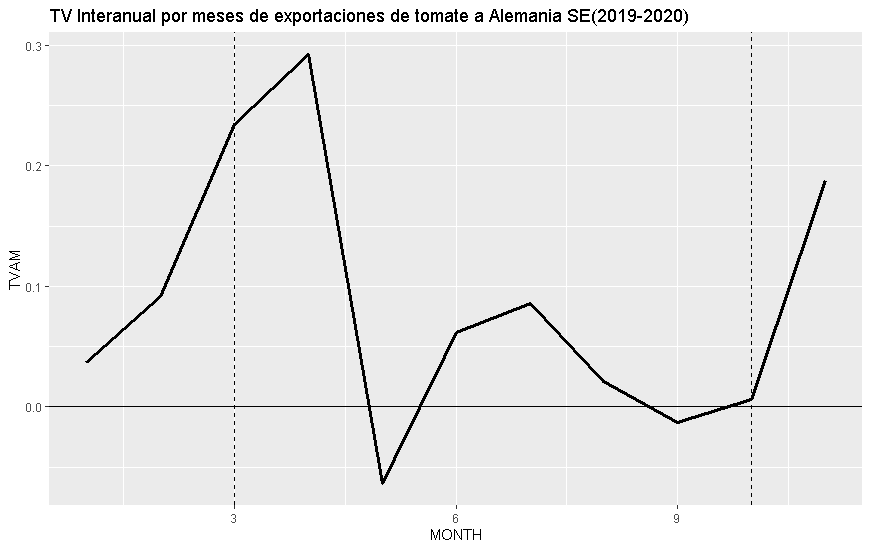
De esta forma tras eliminar la estacionalidad podemos comparar gráficamente la evolución de las exportaciones de tomate y de las defunciones por coronavirus:

Figura 6. Comparativa de la evolución de las exportaciones de tomate y el Covid-19.



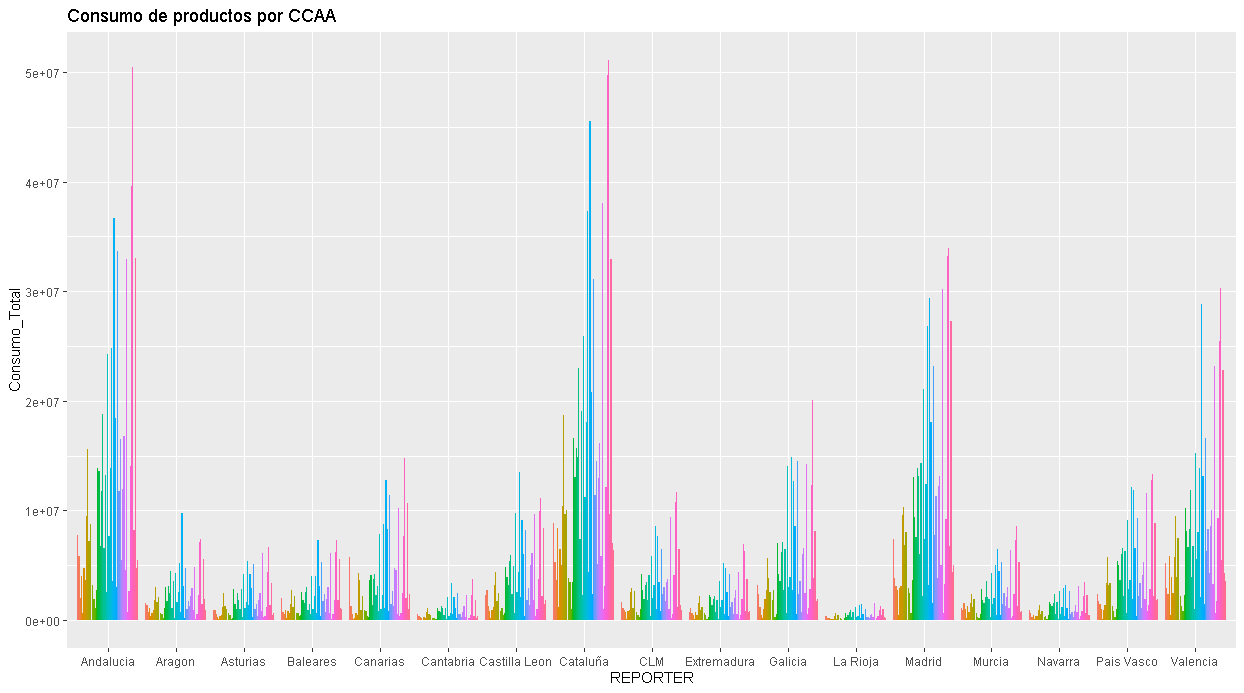
Si tenemos en cuenta que las medidas y el impacto vienen a consecuencia de los casos que suceden previamente en el tiempo a las defunciones podemos apreciar como el máximo de las exportaciones que se produce en marzo es un cambio de tendencia a consecuencia de la crisis sanitaria global que por una parte reduce la renta alemana y con ello sus compras a España y por otra parte implica restricciones en la frontera que pueden dañar el comercio. Esto lo podemos apreciar mejor en la Figura 7:

Figura 7. TV interanual por meses de las exportaciones de tomate hacia Alemania.



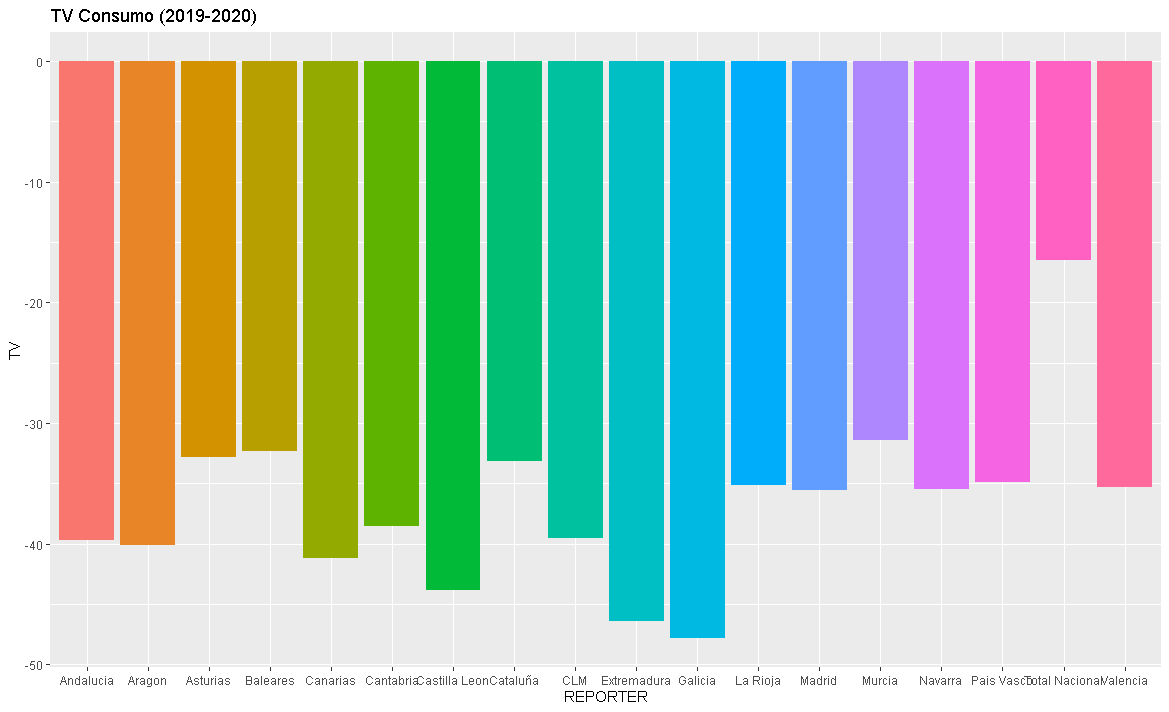
En cuanto a un análisis centrado a nivel nacional en España podemos apreciar como las CCAA más relevantes en el consumo total son Barcelona, Andalucía y Madrid.

Figura 8. Total de consumo por CCAA y producto.



En cuanto a la evolución del consumo entre 2019 y 2020 cabe apreciar que la mayor caída se produce en Galicia, Extremadura y Castilla y León.

Figura 9. Tasa de variación del consumo total por CCAA (2019-2020)

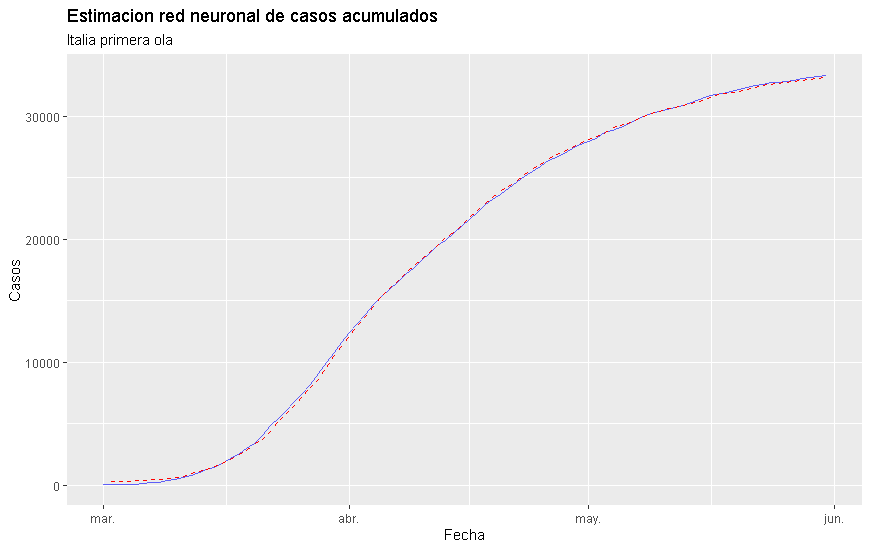


# **4. MODELIZACIÓN DEL COVID Y LOS PRECIOS DE EXPORTACIÓN**

Para esta tarea especificamos los datos a un país y producto, continuando en este caso con el tomate y empleando la base de datos “PreciosSemanales” junto a la base de datos “Covid”.

Para el ajuste de los datos de Covid empleamos 4 modelos sobre datos de Italia en la primera ola. Una regresión polinómica sobre las defunciones diarias con forma de parábola, una regresión logística sobre las muertes acumuladas, la metodología Prophet y una red neuronal. Es la red neuronal el modelo que presenta un menor MAPE (48%) y por tanto escogeríamos este.

Figura 10. Estimación de las defunciones acumuladas mediante red neuronal (Italia).

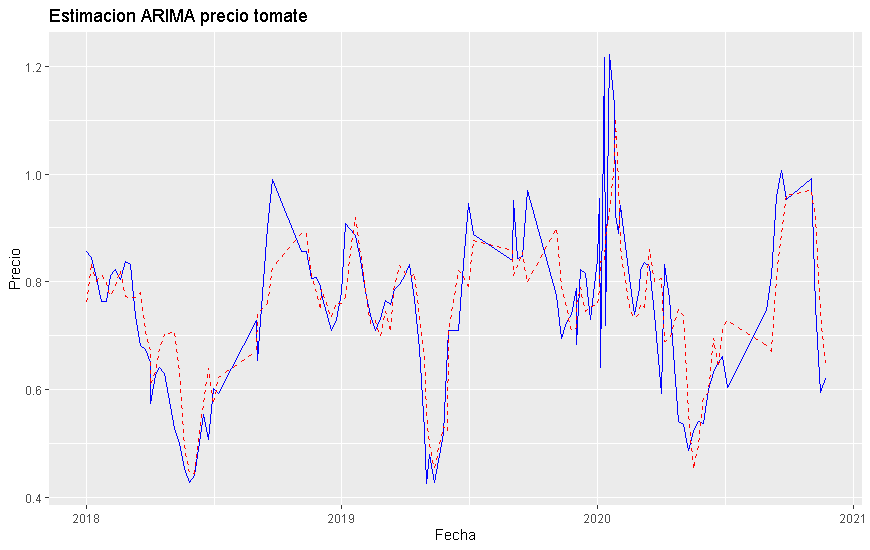


Para el ajuste de los precios del tomate probamos con dos categorías de modelos: Modelos causales y modelos no causales, los primeros emplean datos de coronavirus como variable regresor mientras los segundos no. Los modelos causales se componen de un modelo lineal, un Modelo Generalizado Aditivo (GAM) y un Árbol de regresión, mientras los no causales se trata de un modelo ARIMA y un modelo de suavizado exponencial. Como vemos en este cuadro el modelo con menor MAPE es el ARIMA:



El modelo ARIMA es un (2,0,2), se aplica sobre los datos desestacionalizados mediante STL y posteriormente se le agrega la estacionalidad, el resultado es el siguiente:

Figura 11. Estimación mediante ARIMA del precio del tomate



# **5. CONCLUSIONES.**

1. El impacto de la primera ola es superior y notorio respecto al de las siguientes olas sobre las exportaciones. Sobre esto mantenemos unas hipótesis, dado que el comienzo de la crisis sanitaria provoca una recesión esto lleva a la caída de las exportaciones, pero posteriormente como nos encontramos con un mayor conocimiento sobre el virus, las vacunas generan expectativas positivas, los controles en frontera son menos intensos y la economía no puede ir mucho peor, el impacto es menor.

2. Heterogeneidad regional. Tanto las exportaciones como el consumo tienen distinto volumen y evolución a nivel regional, hablemos de países, provincias o regiones, puesto que la estructura económica, política, así como el impacto del coronavirus y otros factores difieren geográficamente.

3. Heterogeneidad en las variedades de producto. Cada producto tiene unas características únicas en su estacionalidad, producción, demanda, normativas y competencia lo que consecuentemente implica un distinto impacto de la pandemia sobre sus variables asociadas.

# **6. PERSPECTIVAS DE FUTURO.**

Esta versión inicial del trabajo es tan solo un boceto del proyecto que tenemos en mente.

Por una parte pretendemos perfeccionar las bases de datos, homogeneizando las magnitudes valor deflactándolas, añadiendo fuentes de datos externas en particular sobre clima, salarios agrícolas y poblaciones de insectos como variables que puedan afectar a los bienes agrícolas.

Pretendemos realizar un análisis clúster con la intención de clasificar las diferentes regiones y poder hacer un análisis comparativo por conglomerados.

En el aspecto de la visualización tenemos como objetivo la realización de mapas y gráficos circos que permitan abordar mejor la dimensión espacial de la cuestión.

En cuanto al análisis pretendemos como objetivo principal del proyecto profundizar en mayor medida en las causas y consecuencias de la evolución de las variables ligadas a los bienes agrícolas de forma que podamos pasar de un análisis descriptivo a abordar el por qué de la cuestión. Por otra parte buscaremos perfeccionar los modelos con la intención de desarrollar una aplicación capaz de pronosticar a corto plazo la evolución de los precios agrícolas y otras variables, dentro de las limitaciones que implica este plan, dado que serán válidos en condiciones de estructura estable en el sistema, así mismo aquí buscaremos una aplicación de uso general y cotidiano sin demasiadas complicaciones técnicas que puedan dificultar su uso.

Finalmente buscamos perfeccionar nuestro proyecto solucionando posibles errores, mejorando y simplificando el código así como añadiendo aspectos que resulten pertinentes.

Finalmente el principal objetivo del proyecto es obtener conclusiones sobre el impacto del coronavirus en las exportaciones que puedan ser útiles para entender el fenómeno, emplearlas en otras investigaciones y así mismo buscamos hacer público el código de forma que pueda servir en un futuro a otras personas en sus investigaciones, relacionadas o no en específico a esta cuestión, para lo que intentaremos lograr la mayor eficiencia y transparencia en el mismo y así lograr que sea reproducible sin dificultades.

1. Escogemos este método de cálculo de distancia ya que tiene en cuenta la matriz de correlaciones. [↑](#footnote-ref-1)
2. En el archivo ORIGEN\_MB\_NOM podemos encontrar las referencias del código asociado a cada país [↑](#footnote-ref-2)