

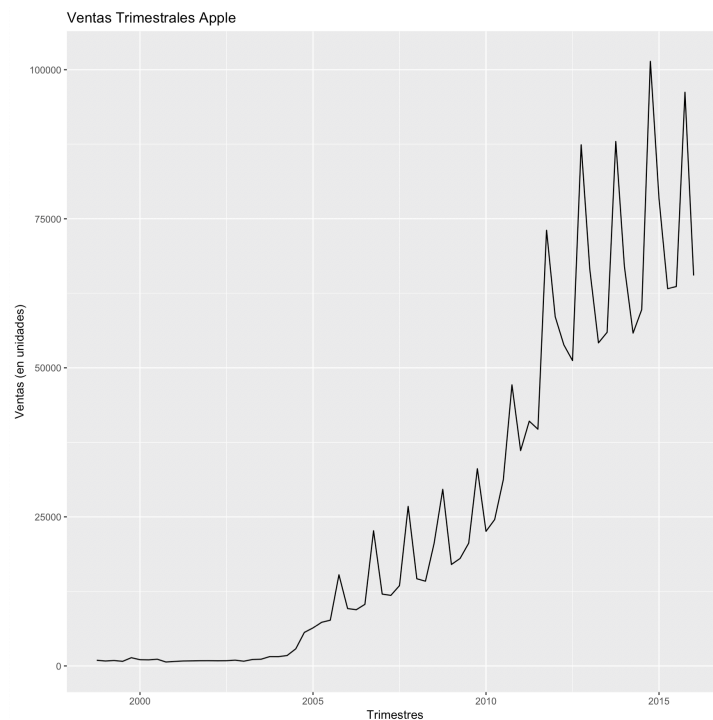
## **INFORME PREDICCIÓN VENTAS DE APPLE**

### **Abstract**

*En este informe se parte de una base de datos que recoge el número (en millones) de unidades vendidas de cada producto (iPod, iPad, iPhone y Mac) agrupadas por trimestre desde el último trimestre de 1998 hasta el primero de 2016. A partir de esta información y tras un brevísimo análisis exploratorio se proponen distintos modelos predictivos y se comprueba su validez para dicha base de datos.*

Al observar la base de datos existen ciertas anomalías que saltan a la vista. En primer lugar existen una gran cantidad de huecos en la base de datos. Esto se debe a que no todos los productos considerados existen desde septiembre de 1998 que es cuando empiezan las ventas de esta base de datos. El iPod es presentado el 23 de Octubre de 2001 y por eso no aparecen sus ventas hasta el último trimestre de dicho año. Seis años después en Junio de 2007 sale a la venta el primer iPhone y por último el 3 de Abril de 2010 sale a la venta el último producto considerado en este estudio, el iPad. A partir de 2014 aparecen también huecos en el iPod. Esto se debe a que Apple decide en dicha época dejar de presentar las ventas de iPod por separado y pasa a presentarlas integradas en las de iPhone<sup>1</sup>. Esto se debe probablemente a la enorme pérdida de relevancia del iPod en los ingresos generales de Apple. A finales de 2014 representaba apenas el 1% de los ingresos por ventas de Apple (habiendo llegado en 2006 a suponer un 40% de estos mismos). A sabiendas de todo esto parece razonable rellenar todos los huecos antes citados con ceros.

En el código adjunto se construyen algunas gráficas para la detección de outliers. En ellas se aprecia una tendencia creciente de ventas pero no se detectan datos atípicos de donde cabe suponer que la carga de datos se ha realizado de manera correcta (**Figura 1**).



**Figura 1.** Ventas de Apple entre 1999 y 2016.

<sup>1</sup> <https://www.businessinsider.com/apple-ipod-rise-fall-chart-2017-7?IR=T>

En este momento se plantean dos opciones a la hora de construir los modelos de predicción; por una parte podría realizarse una predicción individual por producto (descartando el iPod pues como ya sabemos sus ventas se reflejan al final de los datos integradas en las del iPhone). Otra opción sería calcular el número de unidades vendidas (sumando iPod, iPhone, iPad y Mac) y realizar la predicción sobre estos nuevos datos. Si se deseara una predicción aun más precisa se podría realizar una pequeña investigación sobre las ganancias a nivel volumen económico en lugar de cantidad de ventas unitarias. Así la predicción sobre el modelo agregado sería mucho más realista pues no notaría la perturbación producida por considerar las ventas de iPod como iPhone y ofrecería información mucho más útil (en términos generales la venta de un Mac produce mucho más dinero que la de un iPad, por ejemplo).

Descartada esta última opción por su complejidad debido al limitado tiempo del que se dispone para realizar este informe se estudian a continuación las dos primeras opciones:

**Nota:** En el código adjunto se presentan diversas construcciones de modelos. Las he realizado más a título personal para comprender más en profundidad cómo funcionan las distintas opciones que para obtener los resultados de este informe. Por tanto aunque no se hará mención a ellos, en el código, aparecen comentados en detalle.

### Modelo agregado

Partiendo de los datos iniciales se calcula mediante una suma el número de productos vendidos trimestralmente por Apple. Una vez se dispone de este dato se construye una nueva serie temporal en la que se basará este primer modelo predictivo (**Figura 1**).

Partiendo de estos datos se plantean dos familias de modelos: ETS (Algoritmo Suavizado Exponencial) y ARIMA (Media móvil autoregresiva integrada). Así pues buscamos el modelo óptimo de cada una de las familias y los comparamos para ver cuál de los dos nos ofrece una mejor predicción.

A la hora de elegir modelo se empleará como medida de calidad su puntuación AIC (siendo mejor el modelo cuya puntuación sea menor). En este caso el modelo óptimo perteneciente a la familia ETS presenta una AIC de 1268.758 y un AICc de 1271.973 mientras que el modelo óptimo de la familia ARIMA presenta un AIC de 1176.26 y un AICc de 1177.82 por lo que parece aconsejable elegir este segundo. Si se omite el último año del modelo, se entrenan ambos y se comparan los resultados con las predicciones se obtienen los siguientes errores:

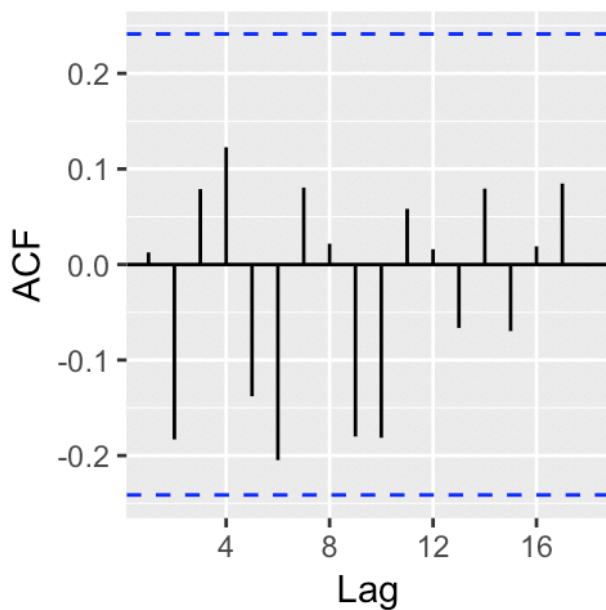
	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
<b>ETS</b>	522.7466	3158.95	2,007.231	1.475937	16.62762	0.4218715	0.0127147
<b>ARIMA</b>	67.82741	3.157.60	2,040.512	-0.3609418	14.52064	0.4288664	-0.1048625

**Figura 2.** Comparativa de modelos ETS y ARIMA trabajando en el agregado.

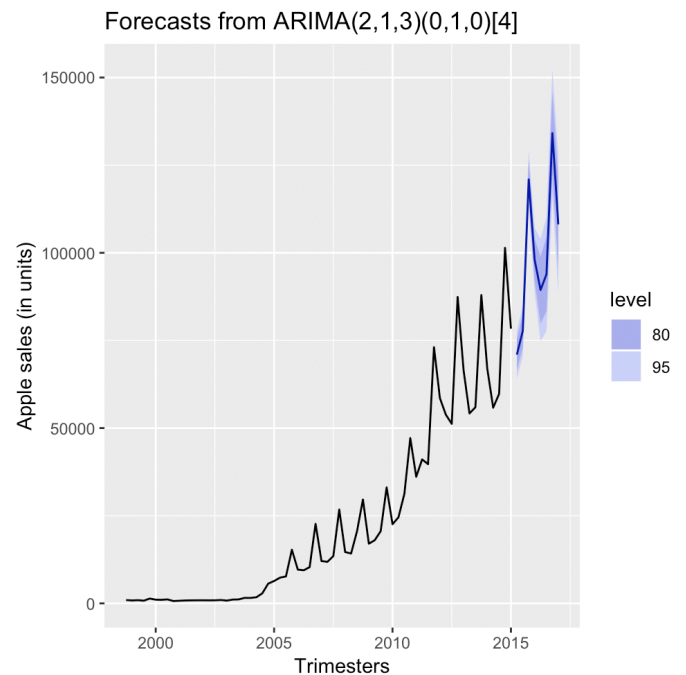
Se puede observar a partir de la **Figura 2** que el modelo ARIMA presenta mejores resultados en prácticamente todas las medidas de errores luego se confirma que en este caso es más conveniente emplear dicho modelo para la predicción.

A continuación se comprueba hasta qué punto son fiables estas predicciones. El modelo ARIMA obtenido es un ARIMA de orden 2, con una diferenciación y una media móvil de orden 3 en su parte no estacionaria y de orden 0 con una diferenciación y media móvil de orden 0 en su parte estacionaria.

Como se puede apreciar en la **Figura 3** todas las autocorrelaciones están dentro de los límites del umbral, lo que indica que los residuos se comportan como ruido blanco y por tanto que la predicción es válida. En la **Figura 4** se muestra como evolucionarían las ventas en los dos años siguientes a la recogida de datos.



**Figura 3.** Residuos ARIMA



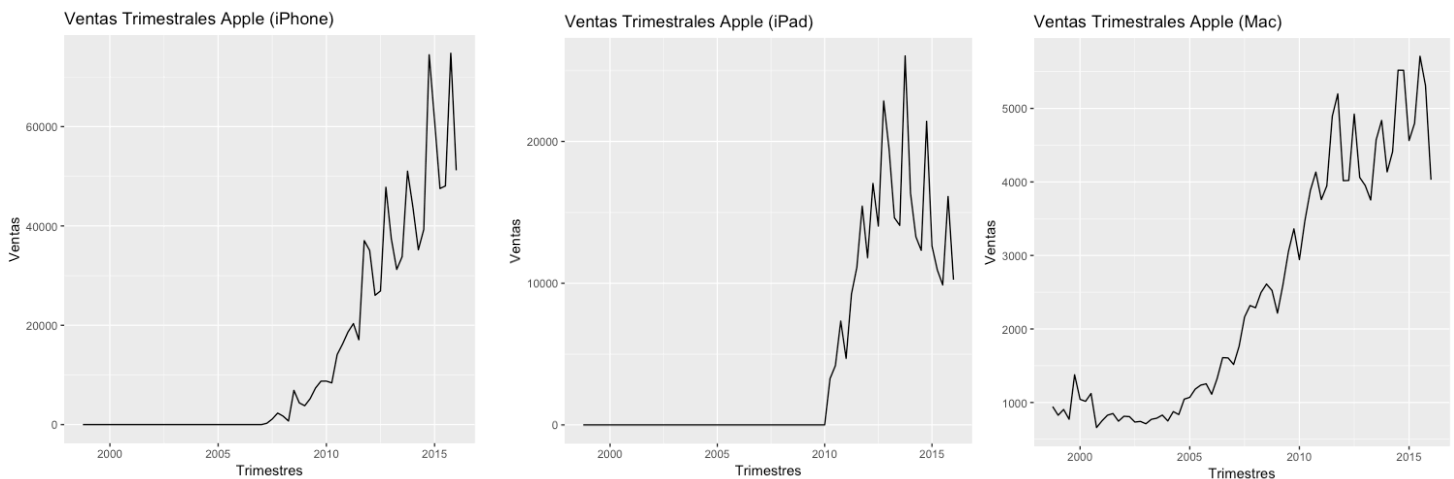
**Figura 4.** Predicción de ventas a dos años

Si se compara esta gráfica con la representación realizada por Statista parece que la predicción es bastante acertada.

En un breve resumen de esta sección, se puede decir que se ha construido un modelo ARIMA para los datos agregados que realiza unas predicciones bastante fiables desde un punto teórico (debido al comportamiento de los residuos) y desde un punto de vista práctico (en este caso, se dispone de datos reales de ventas de Apple con los que se puede comparar esta predicción).

### Modelo desagregado

Recuerde el lector que en este caso se realizará una predicción para cada producto para luego extraer conclusiones generales. Se puede observar que existe una gran diferencia entre el desarrollo de cada producto (**Figura 5**) por lo que esta idea no parece descabellada.



**Figura 5.** Distribución de ventas de cada producto.

A continuación se realizara para cada producto un análisis análogo al realizado anteriormente para los volúmenes totales de ventas.

Comenzamos por el iPad. Al probar las dos familias de modelos se comprueba que el modelo  $ARIMA(3,1,0)(2,0,0)[4]$  es el óptimo presentando un AIC de 1147.2 (frente a un AIC de 1269.992 del mejor modelo ETS). Una vez calculados los errores se pueden observar enormes diferencias entre los dos modelos lo que corrobora la aseveración anterior (para conocer los errores al detalle consultar el código). El gráfico ACF (se puede consultar en el código) muestra además que los residuos se comportan como ruido blanco por lo que la predicción parece bastante adecuada.

Una vez estudiado el iPad veamos el iPhone. De nuevo el modelo  $ARIMA(0,0,1)(0,1,0)[4]$  presenta un AIC de 566.61 frente a los 658.28 del modelo ETS. De nuevo esto se refleja en los errores de cada modelo. Al realizar el gráfico ACF de nuevo se observa que los residuos se comportan como ruido blanco.

Por último se realiza el estudio de modelos para predicción de las ventas de Mac. De nuevo en este caso parece más adecuado un modelo ARIMA, en concreto,  $ARIMA(0,1,1)(1,1,2)[4]$  pues presenta un AIC de 867.54 frente al AIC de 1002.849 del mejor modelo ETS. El gráfico ACF de nuevo asegura que los residuos se comportan como ruido blanco y la comparación de errores asevera una vez más que es más adecuado el modelo ARIMA.

Por último para cerrar este informe se comparan los resultados de ambas predicciones (la del total de ventas y la de producto por producto) con los resultados reales de Apple:

	2016 Q2	2016 Q3	2016 Q4	2017 Q1	2017 Q2	2017 Q3
<b>Agregado</b>	89406.53	93868.00	134131.74	108101.12	96692.50	99088.89
<b>Productos separados</b>	62136.85	64236.36	108596.97	86108.96	65685.27	67922.38

Comparando estos resultados con los datos reales se observa que el cálculo por productos totales en lugar de separado ofrece resultados mucho más adecuados por lo que parece que el mejor modelo de los aquí estudiados sería el ARIMA para el total de los productos.

### Conclusiones

En este informe se plantean dos líneas de trabajo. En una primera parte se busca el mejor modelo para predecir las ventas a futuro a partir de las realizadas anteriormente en total (es decir, sin distinguir cuánto aporta cada producto a la venta). En la segunda parte se disgregan los productos y se plantean modelos predictivos para cada uno. En esta ocasión se han considerado solo modelos de las familias ETS y ARIMA. Tras comparar los diferentes resultados se llega a la conclusión de que el mejor modelo es el ARIMA(2,1,3)(0,1,0)[4] que se basa en el total de los productos siendo este el que menos diferencia presenta con las ventas reportadas por Apple.

Una posible idea para mejorar este modelo sería como se menciona anteriormente estimar el coste medio de cada producto para poder estimar el volumen de ganancia de la compañía lo cual sería una predicción mucho más interesante pues saber el número de productos vendidos no permite estimar de manera clara las ganancias pues existen productos más rentables que otros para la compañía.

### Código

El código sobre el que se fundamenta este informe se encuentra disponible en mi repositorio de GitHub al que se puede acceder a través del siguiente enlace. Varios profesores nos han sugerido que comentemos el código en inglés en lugar de en español así que he optado por hacerlo, si le supone un problema no dude en hacérmelo saber.