

INFORME SOBRE LA PREDICCIÓN DEL USO DEL TRANSPORTE PÚBLICO EN ISLAS BALEARES

Background

Islas Baleares

El trabajo a realizar consiste en la elaboración de las predicciones del número de billetes de transporte público que se expedirán en las Islas Baleares en enero de 2017.

Islas Baleares es la duodécima comunidad autónoma con más habitantes de España. Dicho de otra forma, es la quinta que menos habitantes tiene. Sin embargo, por su reducido territorio, es la quinta con mayor densidad de población. 1.221.000 personas conviven en un territorio de apenas 4.992 kilómetros cuadrados. Muchas de estas personas son de origen europeo no español que pasan gran parte del año en su vivienda disfrutando del clima y el ocio que reporta este territorio. Son una de las dos comunidades autónomas españolas que están fuera del territorio de la Península Ibérica. La otra son las Islas Canarias.

En este caso, esta comunidad está formada por cuatro grandes islas (Mallorca, Menorca, Ibiza y Formentera) o otros islotes que no tienen población, como Dragonera, Conejera, la isla del Aire, la isla de Colom, la de Espalmador o la de Espardell.

Su capital es Palma, situada en la isla de Mallorca. Donde viven 423.000 personas.

La isla con más población es Mallorca, con un total de 880.000 habitantes, seguida de Ibiza, con 145.000 personas y, en tercer lugar, Menorca, que cuenta con 91.920 habitantes.

Islas Baleares es una comunidad autónoma uniprovincial y tiene dos idiomas cooficiales: el castellano, que en el año 2014 lo hablaba el 50% de los baleares, y el catalán, que lo habla el 36% de los habitantes de estas islas. Uno de cada diez habitantes de las islas habla ambas lenguas.

Transporte público

El transporte público es cada vez más importante en una sociedad que camina hacia la reducción de emisiones contaminantes mediante medios de transporte privados menos agresivos hacia el medio ambiente, como puede ser el coche o moto eléctricos o la bicicleta, pero también de un transporte público que es vital para los desplazamientos largos (como puede ser viajar en tren entre Madrid y Barcelona) o desplazamientos que no se pueden hacer a pie por otras circunstancias, como pueden ser las malas condiciones meteorológicas de un determinado día.

Incluso durante la crisis sanitaria del covid 19, uno de los lugares donde se extremaban las precauciones y donde las autoridades públicas centraron su atención fue en el transporte público, donde las aglomeraciones de gente que no tiene nada en común podía disparar la propagación del virus en las grandes ciudades.

Turismo

En el transcurso del año 2021, las Islas Baleares destacaron como la región autónoma más frecuentada por turistas internacionales en España, superando a destinos tan emblemáticos como Cataluña y Canarias.

Este conjunto de islas ubicado en el Mediterráneo occidental ha experimentado un notable ascenso en popularidad, consolidándose como uno de los destinos turísticos más codiciados no solo a nivel nacional, sino también a lo largo y ancho de toda Europa, siendo especialmente aclamado por los visitantes procedentes de Alemania.

El renombre de Mallorca entre los alemanes ha alcanzado tal magnitud que en ocasiones se refieren a la isla como su "decimoséptimo Estado federal", evidenciando así la profunda conexión que han establecido con este enclave paradisíaco. El clima mediterráneo que caracteriza a la región, junto con sus impresionantes calas de aguas cristalinas, convergen para crear un escenario idílico que lo posiciona como el destino vacacional por excelencia. Este atractivo conjunto de factores ha contribuido de manera significativa a consolidar a las Islas Baleares como un destino turístico de primer orden, cuya belleza natural y encanto cultural continúan cautivando a viajeros de todo el continente europeo.

Durante el año 2022, último del que a día de hoy hay datos completos y que se puede considerar como el primero tras la recuperación completa del turismo post pandémico, el número de turistas que llegaron a las islas baleares fue de tres millones y medio de visitantes españoles y otros trece millones provenientes de otros lugares del planeta. Es decir, en total, 16 millones y medio de turistas, lo que supone multiplicar por quince la población que habita en Mallorca, Menorca, Ibiza y Formentera.

Los datos de 2023, computados hasta el 30 de noviembre, según el Instituto de Turismo de España, conocido como TURESPAÑA, fue de 14,24 millones de turistas, el mejor dato anual para esa fecha desde que hay registros. Turespaña es un organismo autónomo de la Administración General del Estado encargado de promocionar España en el exterior, como destino turístico y lo hace mediante el marketing y el impulso de la estrategia turística española, la cual coordina y lidera.

Además, la condición insular de la comunidad autónoma Balear implica que los viajeros, tanto nacionales como internacionales (especialmente provenientes de Alemania, Francia y Reino Unido), no pueden desplazarse hasta estos destinos en coche, lo cual limita sus opciones de movilidad una vez llegan a su isla elegida. Esta circunstancia dispara los alquileres de coches respecto a otros destinos peninsulares (donde, al menos los turistas españoles, portugueses y franceses sí que se desplazan con su propio automóvil) y también en el uso del transporte público, que, como comprobaremos, es mucho mayor durante la temporada turística que durante el resto del año. Algo que sin duda, no ocurre en las grandes ciudades. Esto supone un cambio de paradigma por tanto para quien vaya a trabajar en TIB (Transports de les Illes Balears) y esté acostumbrado a realizar estas labores en una gran ciudad como podría ser Madrid, en la EMT (Empresa Municipal de Transportes), donde el uso del tren de cercanías, el Metro y el autobús urbano, cuando decrece, es en los meses de verano.

Literature review

La estacionalidad turística en determinados puntos turísticos revoluciona las pautas de movilidad del conjunto de una región, según Gutiérrez y Miravet (2016). Afirman que los resultados de la investigación señalan la identificación de estructuras urbanas jerárquicas y centros regionales que exhiben adaptabilidad y cambios a lo largo de las estaciones. El estudio destaca la importancia del turismo como un impulsor esencial de procesos profundos de reorganización territorial, ejerciendo un papel crucial en la configuración de dinámicas metropolitanas en regiones turísticas del Mediterráneo. Un Mediterráneo al que pertenece la costa tarraconense, a la que hacen referencia Gutiérrez y Miravet, pero también las Islas Baleares, que son las que estamos estudiando. Este fenómeno subraya, afirman, la significativa influencia del turismo en la transformación del entorno territorial y en la articulación de procesos urbanos en dichas áreas.

Mediante el uso de series temporales entre 1966 y 2019, Albaladejo Pina, I., González Martínez, M. I. y Martínez-García, M. P. (2023) afirmaron que las Islas Baleares mostraron “cierto estancamiento” tras el boom del turismo de entre 1960 y 1970. Un estancamiento del que se recuperaron a partir del año 1990 y que viene marcado por una tendencia lineal que indica un crecimiento menos acelerado que el de dos de las principales comunidades autónomas peninsulares en lo que a turismo se refiere: Cataluña y Andalucía.

Metodología

Para realizar este estudio lo primero que se llevó a cabo fue una exploración de los datos. Unos datos que venían ordenados por año, mes y día de la semana y día de mes. Además, los datos venían en catalán, uno de los dos idiomas cooficiales de las Islas Baleares.

El primer paso tras leer los datos fue atribuirles un mes y un cuatrimestre, para elaborar una tabla donde apareciera el mes, el cuatrimestre, y los billetes vendidos. Esto se llevó a cabo mediante un mapeo y un bucle for.

El siguiente paso fue reordenar las fechas, de tal forma que tuvieran un orden más lógico y entendible, permutando el orden por día de la semana a orden por día del mes. De este modo, se ordenan primero por año (de 2013 a 2017), luego por mes (de enero a diciembre), y por último, por día del mes (del 1 al 28, 29, 30 o 31).

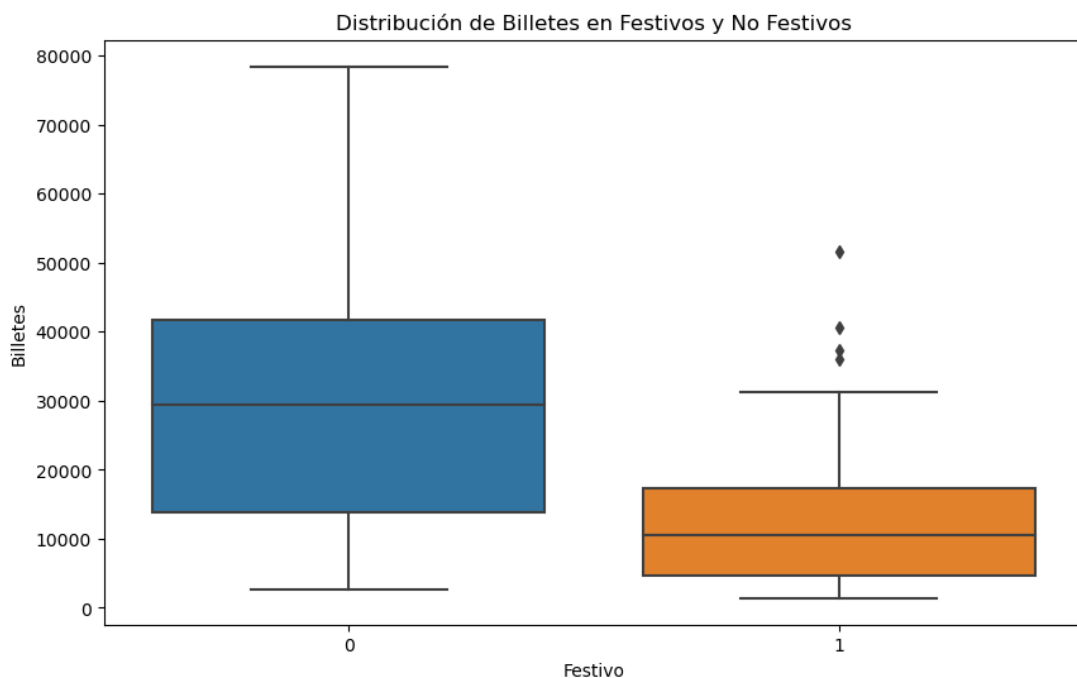
Con la biblioteca *datatime*, podemos incorporarle a la tabla también el día de la semana de cada instancia, de tal forma que se pueda, especialmente, diferenciar los días de entre semana de los sábados y los domingos.

Impacto de los festivos.

Por último, incluimos una variable dummy que son los **festivos**, para comprobar si los viajes en transporte público varían según si el día en que vamos a predecir es o no es laborable. De esta forma se nos queda un dataframe con 1461 filas (una por cada día de este período de cuatro años) y ocho columnas: Fecha, año, mes, día del mes, día de la semana, trimestre, festivo y la variable objetivo: billetes.

La columna “Fecha” la hemos construido a partir de ligar como *datatime* “Año”, “Mes” y “Día semana”.

Gráfico 1: Distribución de venta de Billetes según si el día es festivo o no.



Fuente: TIB

Decidimos pasar al índice la columna de fecha y eliminar todas las variables que ya están reflejadas en ella, quedando en el dataframe exclusivamente el día de la semana, si es festivo

en las Islas Baleares o no lo es, y el número de billetes de transporte público vendidos en ese día.

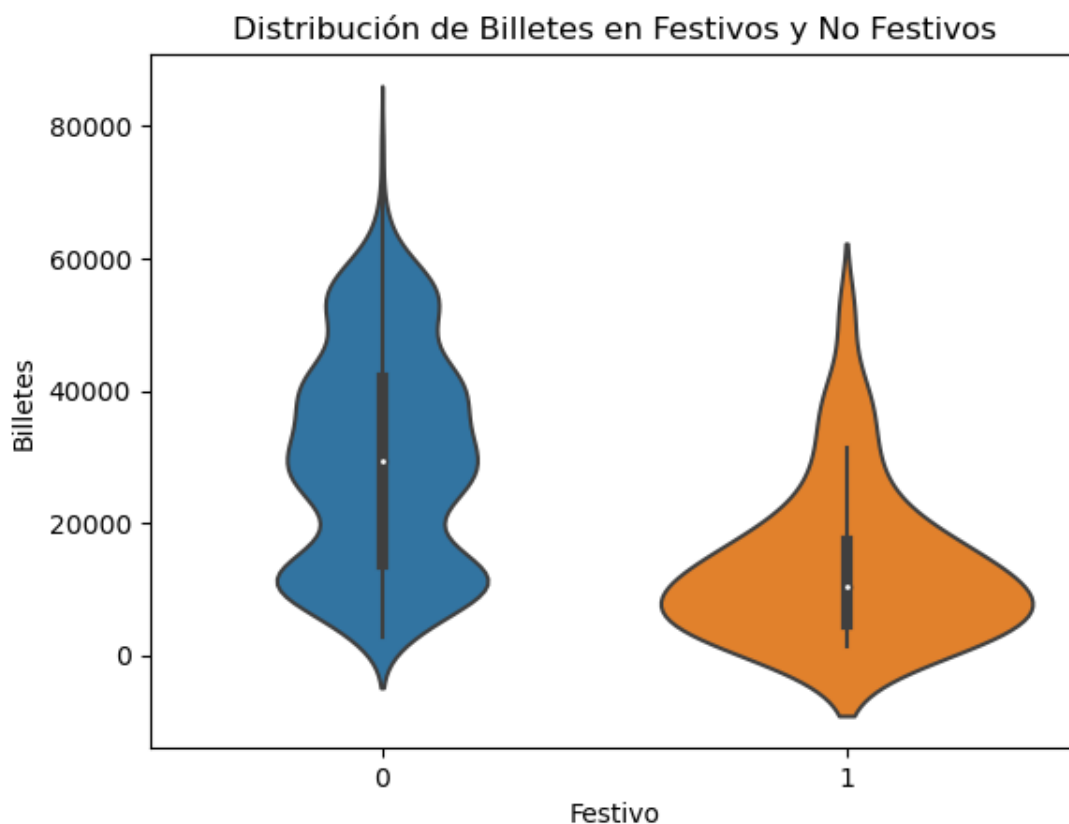
Una vez hecho esto, graficamos en base a los festivos, para comparar si es relevante que un día sea no laborable para la venta de billetes de autobús.

En el Gráfico 1 y en el Gráfico 2 podemos comprobar que, efectivamente, cuando estamos ante un día festivo, la venta de billetes cae drásticamente. Algo que, como después comprobaremos, no ocurre cuando hablamos de periodos vacacionales, lo cual, nos permitirá inferir que hay una clara distinción entre los periodos no laborables cortos (un festivo o un puente) y los periodos vacacionales, donde, como comprobaremos, la venta de billetes de autobús de servicios ferroviarios se dispara.

El Gráfico 1, que es un diagrama de cajas, nos muestra como la mediana de los días laborables está prácticamente al nivel del percentil superior de los días festivos. Es decir, un día normal, en las Islas Baleares, se usa tanto el transporte público como los días festivos en los que más se usó durante el periodo entre 2013 y 2016, en los que se registra una venta de billetes de alrededor de 30.000.

Por otro lado, la mediana de los días festivos, es decir, un día normal dentro de uno de los días ni laborables ni de fin de semana, es de 10.000 billetes vendidos. Esta cifra se encuentra por debajo del primer cuartil de los días laborables del periodo analizado. Dicho de otra forma: tres de cada cuatro días laborables se venden más billetes de autobús y tren que cualquier día normal que sea festivo.

Gráfico 2: Distribución de venta de Billetes según si el día es festivo o no.



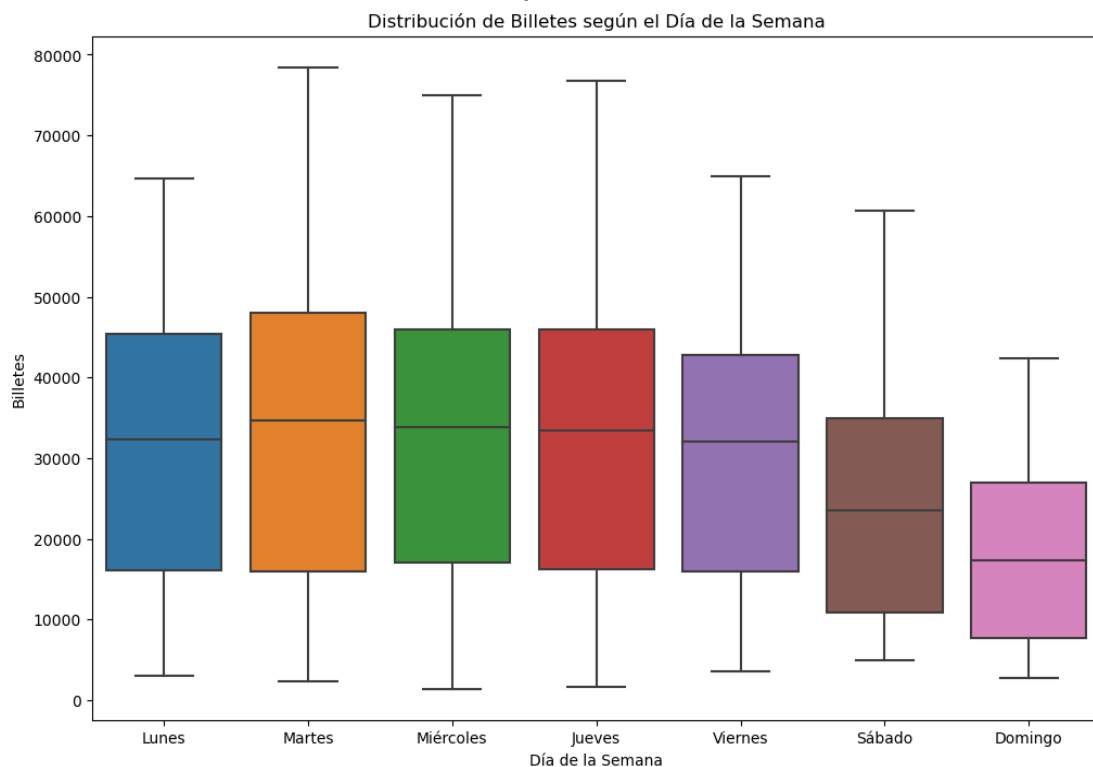
Fuente: TIB

Impacto de los fines de semana.

Además de los festivos, es preciso analizar el impacto del fin de semana dentro de la venta de billetes. Para ello, en el Gráfico 3 podemos comprobar cómo, mientras durante la semana laboral la venta de billetes se mantiene estable (entre los lunes y los viernes la venta se mantiene estable superando ligeramente los 30.000 billetes vendidos diariamente con un rango intercuartílico que va desde los 45.000 del cuartil superior hasta 17.500 del cuartil inferior).

Mientras tanto, los sábados esta venta cae respecto a los días de diario, y esta caída se acentúa los domingos, de tal forma que los sábados se venden, como mediana, alrededor de 22.500 billetes de transporte público. Por su parte, los domingos esta venta cae del nivel de los 20.000 y se sitúa, aproximadamente (en términos de mediana) al nivel del cuartil inferior de los días de entre semana: en torno a 17.500 billetes.

Gráfico 3: Distribución de venta de billetes por día de la semana.



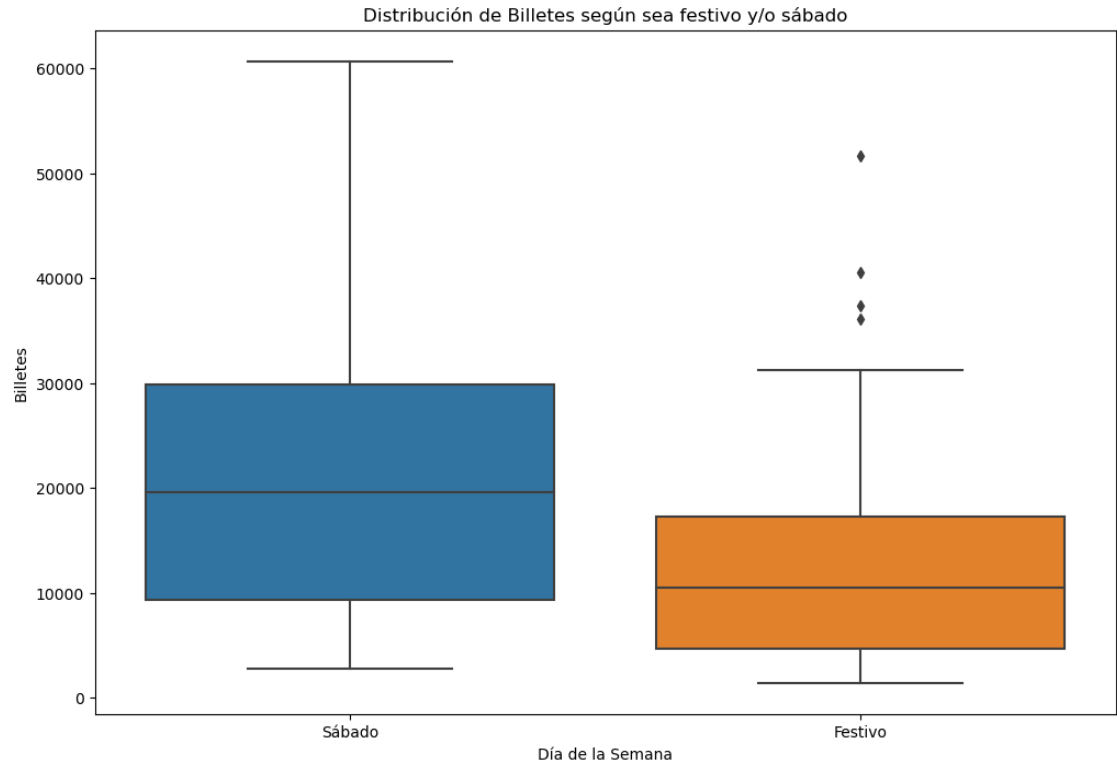
Fuente: TIB

¿Qué afecta más? ¿Un festivo o un fin de semana?

Para saber si afecta más un festivo o un fin de semana, primero debemos observar que ningún festivo oficial de las Islas Baleares cae de domingo. Cuando ocurre algo así, lo que suelen hacer las Comunidades Autónomas habitualmente es, o colocar el festivo al día siguiente (como por ejemplo hicieron con el Día de Todos los Santos (1 de noviembre) de 2015 que al ser domingo lo colocaron al día siguiente,, o como ocurrió con el Día de la Hispanidad (12 de octubre) de 2014, que al caer de domingo obviaron esa fiesta y la recuperaron el día 26 de diciembre, que se celebra San Esteban, una celebración arraigada en las zonas de cultura catalanoparlante. En definitiva, que en el periodo de cuatro años estudiado hubo 47 festivos que se repartieron entre todos los días de la semana menos el domingo. Si a esto le sumamos los sábados y los

domingos, tendremos un total de 459 días. En el Gráfico 4, por tanto a analizamos una de las cuestiones relevantes a la hora de preparar una oferta de transporte público, y es que los sábados, la compra de billetes de transporte público duplica prácticamente la de cualquier festivo.

Gráfico 4: Distribución de Billetes según sea festivo y/o sábado



Fuente: TIB

Siguiendo con la metodología, pasamos a limpiar el dataframe.

Tabla 1: Dataframe en el que aparecen todos los billetes vendidos entre 2013 y 2016

Billetes	
Fecha	
2013-01-01	2338
2013-01-02	9376
2013-01-03	9190
2013-01-04	9523
2013-01-05	5492
...	...
2016-12-27	23928
2016-12-28	23960
2016-12-29	24279
2016-12-30	22763
2016-12-31	5346

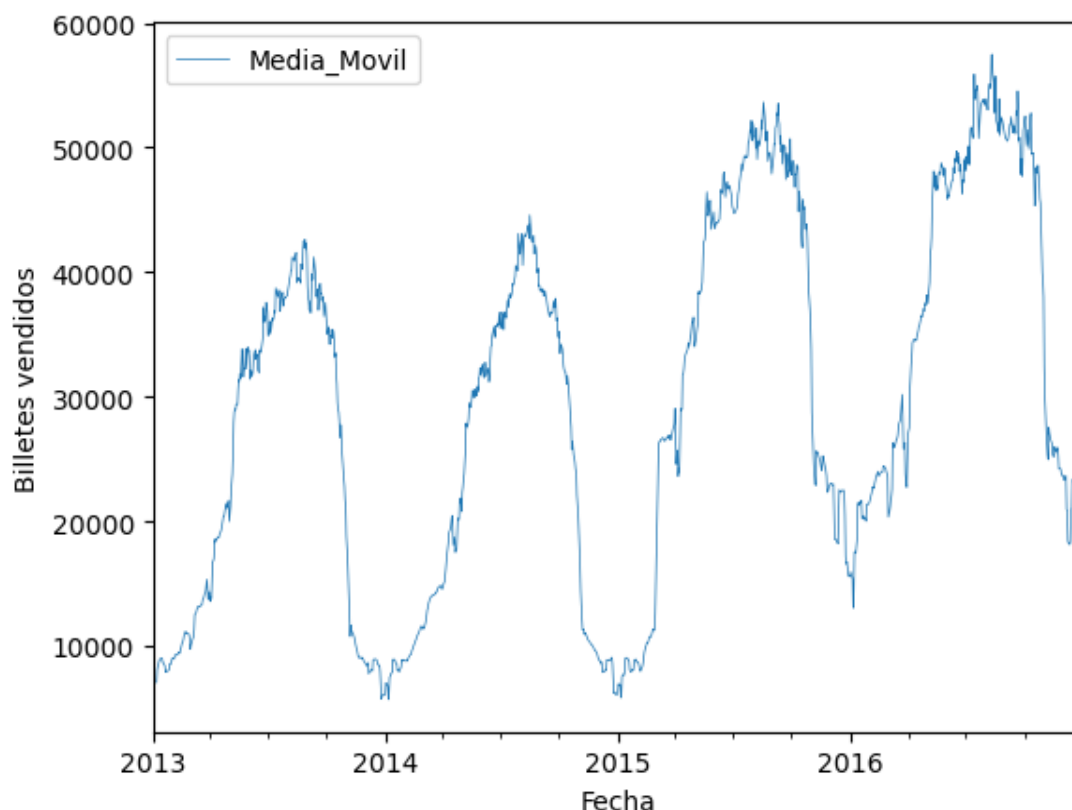
1461 rows × 1 columns

Fuente: TIB

De esta forma, solo nos queda como índice las fechas y como única columna los billetes vendidos por día, tal y como se puede observar en la Tabla 1. Procedemos por tanto a elaborar

una media móvil de siete días que suavice el efecto semanal para poder comprobar cómo se comporta la serie. En el Gráfico 5 observamos que en la parte central de cada año (durante los meses de calor) es cuando más viajeros acuden al transporte público.

Gráfico 5: Billetes diarios vendidos con una media móvil de 7 días



Fuente: TIB

Gracias a la media móvil de siete días utilizada en la Gráfico 5, en la que se elimina el efecto de la periodicidad estacional semanal, observamos con extrema nitidez el comportamiento cíclico anual que tiene la serie temporal de las ventas de billetes para el transporte público en las Islas Baleares.

Estacionariedad en media y en varianza

Al analizar si la serie temporal es estacionaria en media o en varianza, lo primero que debemos ver es la evolución de los datos gráficamente. Para ello podemos utilizar el Gráfico 5, en el que vemos como, tanto los picos como los valles de los ciclos más recientes son superiores a los más antiguos. Es decir, la media no se mantiene estable si no que va creciendo, por tanto, podemos afirmar que **no es estacionaria en media**.

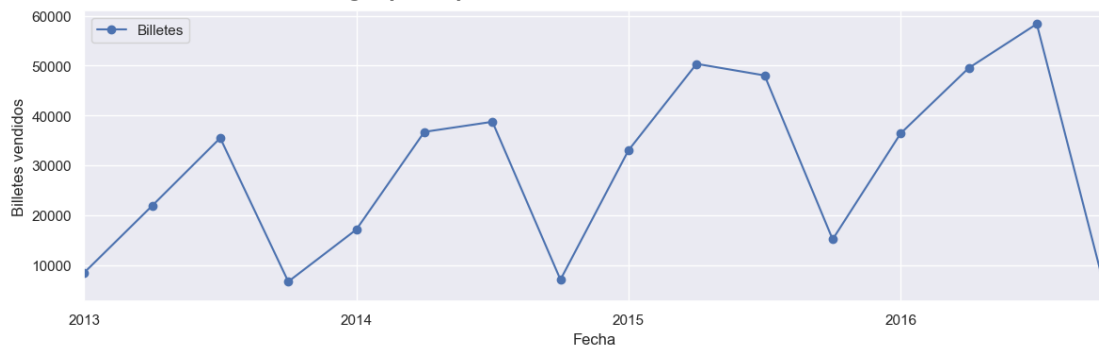
Respecto a la estacionalidad en varianza, a simple vista parece que mengua al crecer la serie histórica (normalmente, cuando una serie temporal crece en media, si no es estacionaria, suele crecer también en varianza, por tanto, estaría ocurriendo un hecho poco frecuente). Recurrimos, para comprobarlo, a la prueba de Dickey-Fuller para realizar la comprobación.

Una vez la aplicamos, se rechaza la hipótesis nula de que la serie sea estacionaria en varianza. Por tanto, **la serie tampoco es estacionaria en varianza.**

Comportamiento por trimestres

Ya hemos analizado la estacionalidad semanal, que está muy marcada, y ahora pasamos a analizar la trimestral. Para ello agrupamos la serie temporal en “quarter”, es decir, en cuartos de año, es decir, en trimestres, como se ve en el Gráfico 6.

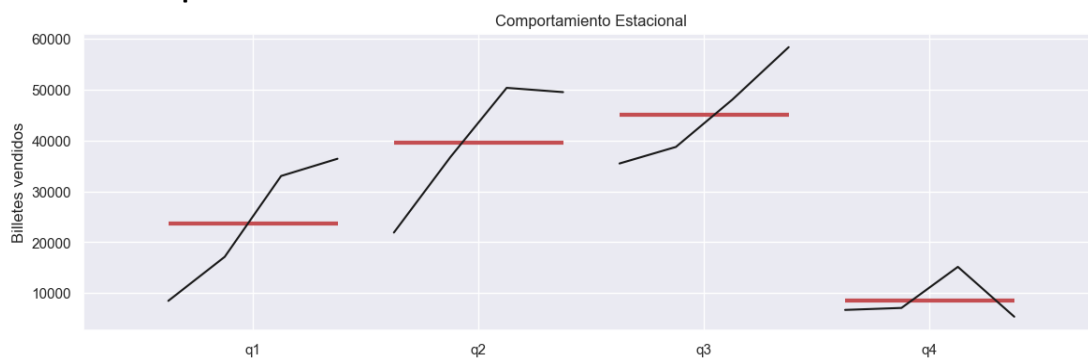
Gráfico 6: venta de billetes agrupada por cuatrimestres.



Fuente: TIB

Aquí podemos comprobar cómo el peor trimestre es el otoñal, cuando decrecen los días (además de que, dentro de España, las Islas Baleares son el territorio donde más tarde oscurece) y cómo siempre se produce una gran caída entre el final del cuatrimestre estival (que acaba el 30 de septiembre) y el comienzo del otoñal (el 1 de octubre del año siguiente).

Gráfico 7: Comportamiento estacional de la serie histórica



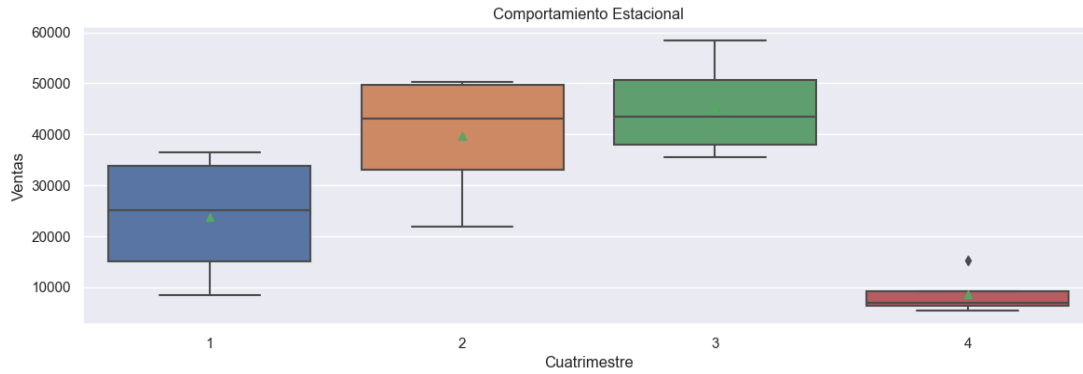
Fuente: TIB

Podemos observar cómo, en el último año, creció el primer cuatrimestre respecto al segundo, que moderó su crecimiento. Esto podría ser por el efecto Semana Santa, que unos años cae en el primer cuatrimestre del año y otros cae en el segundo.

Al depender tanto del turismo, los meses de julio, agosto y septiembre son, además de los que mayor media tienen de billetes vendidos, los que en la actualidad más crecen. Por su lado, los meses de octubre, noviembre y diciembre han decrecido relativamente (en la serie histórica todos los demás cuatrimestres crecieron de forma positiva), con una bajada en el último trimestre de 2016 que ha compensado los ligeros crecimientos de los tres años anteriores).

En el Gráfico 8 podemos observar mediante diagramas de caja cómo se comporta la venta de billetes por cuatrimestre, donde queda aún más patente la caída del uso del transporte público en los meses de octubre, noviembre y diciembre.

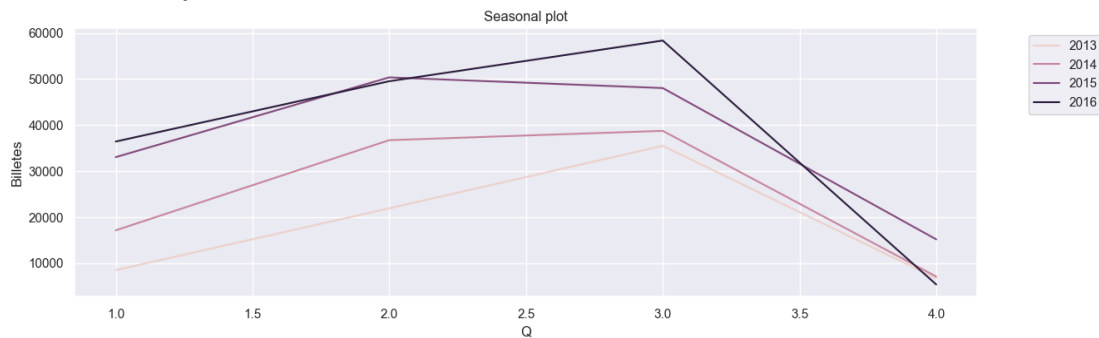
Gráfico 8: Comportamiento estacional de la serie histórica



Fuente: TIB

Si comparamos entre años, podemos en el Gráfico 9 cómo observar cómo pasamos del trimestre con más uso del transporte público de toda la serie histórica (el de verano de 2016) al trimestre con menos uso de autobuses y trenes también de toda la serie histórica (otoño de 2016). Coincide esta circunstancia con que a finales de junio de 2016 los británicos partidarios de salir de la Unión Europea ganaron el referéndum en junio de 2016.

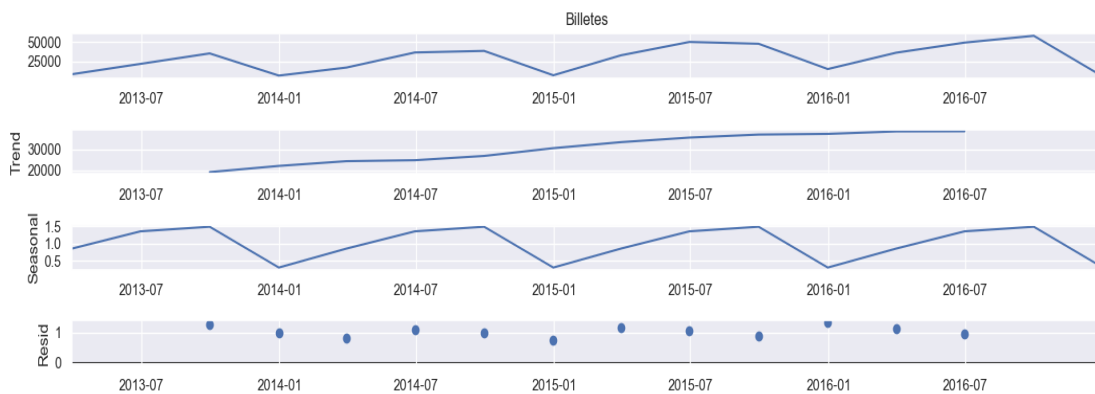
Gráfico 9: Comportamiento anual de la serie histórica



Fuente: TIB

Descomposición de la serie

Gráfico 10: Descomposición de la serie trimestral en residuos, estacionalidad y tendencia.



Fuente: TIB

En el Gráfico 10 observamos cómo, agrupada la serie en trimestres, tiene una componente de tendencia, otra de estacionalidad, y que hay alguna otra estacionalidad que pueda que se nos esté escapando, a la vista de que la serie de residuos tiene un comportamiento cíclico, por el que crecen en los meses de más horas de sol y decrecen en otoño y en invierno.

Después de esto, exportamos el dataframe diario en el que tenemos como índice las fechas y como única columna los billetes. Lo hacemos en formato csv separado por comas a partir del original que tenemos en la Tabla 1.

Paso a agrupación semanal

Como el 1 de enero de 2013 fue martes y es un registro muy antiguo con respecto a lo que queremos predecir (enero de 2017), eliminamos los primeros días de enero de ese año para conseguir una serie que nos pueda predecir por semana.

El criterio que establecemos es que queremos tener semanas completas de tal forma que los datos más recientes (los más interesantes) se agrupen por semanas completas. Es decir, como el 31 de diciembre de 2016 fue sábado, agrupamos las semanas de domingo a sábado. De esta forma no perdemos información relevante a costa de mantener datos irrelevantes o de tener datos de semanas incompletas. Es decir, como nuestras semanas para esta predicción empiezan en domingo y acaban en sábado, eliminamos los primeros cinco valores de la serie (desde el martes, 1 de enero de 2013, hasta el sábado, 5 de enero de 2013). De esta forma ya tenemos un número de instancias múltiplo de siete (número de días que tiene una semana) y podemos empezar a trabajar en términos semanales. Esto, en código Python, lo hacemos de la siguiente forma:

```
df_sin_primeras_cinco = data.iloc[5:]  
data_semanal = df_sin_primeras_cinco.resample('W-Sat').sum()
```

Después, pasamos a exportar los datos, tanto semanales como diarios.

Modelización de series diarias

Comenzamos haciendo pruebas con los modelos que más sentido tienen a la hora de predecir una serie temporal de estas características. Es decir, descartamos aquellas que pronostican varianzas o impactos en el mercado, ya que no las consideramos relevantes para este campo de estudio. La primera que elegimos es un modelo ETS, que realiza predicciones a partir del error, la tendencia y la estacionalidad (ETS: Error, Trend, Seasonality).

Para ello, convertimos los billetes en un float64 con periodo diario.

Dividimos en train y test y, si hubiéramos tenido más tiempo para realizar esto, hubiéramos validado, además de con diciembre del último año (últimos 31 días de la serie para predecir, al igual que la predicción debería ser también de 31 días), con un **20% del dataset** e incluso, por probar (habría que ver luego si los resultados son certeros) con el **mes de enero de 2016**. El motivo de probarlo con el mes de enero del año anterior es que cada mes tiene sus características y enero tiene sus propios festivos en días distintos a los de diciembre (concretamente, enero tiene como festivo el día 1, cosa que diciembre no, y no tiene ni el 8 ni

el 25, cosa que diciembre sí, además del régimen especial que tienen el 24 y el 31 en la sociedad).

Tras probar distintos seasonal periods, comprobamos que el que menor MAPE recoge (Mean absolute percentage error, que es la métrica en la que estamos centrándonos para elegir los parámetros correctos dentro de cada modelo y el modelo) es una estacionalidad con periodicidad diaria de siete. Es decir, la mayor estacionalidad se recoge en el conjunto de la semana.

Una vez con eso establecido, establecemos un bucle for que prueba con los parámetros "additive" y "multiplicative" haciendo todas las combinaciones para error, estacionalidad y tendencia. El MAPE mínimo que obtenemos es de 0,2647.

Pasamos al modelo ARIMA, en el que el bucle lo hacemos con distintas estacionalidades, probando con logaritmos y sin logaritmos (aunque con la intuición de que, al no ser una serie estacionaria ni en media ni en varianza, muy probablemente haría falta tomar logaritmos, y en la que también utilizamos el bucle for para probar con el parámetro "seasonal" True o False, con el parámetro "stationary" True o False, con los distintos "information_criterion" aic, bic, hqic y oob y con los "scoring" mse y mae.

De esta forma, obtuvimos varias combinaciones que, dejaban el MAPE en el 0.3052. Todas con estacionalidad y estacionariedad verdaderas. Sin embargo, este error es mayor que el del modelo ETS.

Acudimos por tanto al modelo TBATS, el cuál requiere una transformación del dataframe original, el cual hay que convertir en una timeseries de pandas. Lo hacemos, y dividimos en train y test con el siguiente código:

```
train, val = series.split_before(pd.Timestamp('2016-11-30'))
```

Tras probar manualmente distintos modelos, comprobamos que el mejor MAPE nos da sin usar box_cox, ni tendencia, ni tendencia amortiguada como parámetros fijos. Lo dejamos al criterio de AIC. Tras probar el modelo, obtenemos un MAPE con los parámetros establecidos de 0,3710, y sin parámetros establecidos, es decir, los que vienen por defecto en el programa, de 0,4307. Por tanto, tampoco elegimos el modelo TBATS para predecir la serie temporal de viajeros diarios.

Pasamos al modelo Neural Prophet, que intenta descomponer cada serie temporal en tendencias, temporadas y días feriados, y produce una previsión mediante la agregación de las predicciones de estos modelos y mediante el uso de redes neuronales.

En este caso metemos la periodicidad estacional dentro del bucle for, en el que introducimos todos los valores del 1 al 12, el 30 (número de días que tiene un mes), el 31(número de días que tiene un mes), 365 (número de días que tiene un año) y 366 (número de días que tiene un año). Por otra parte, probamos como función de pérdida las funciones 'MSE', 'MAE' y 'Huber'. En cuanto al modo de estacionalidad, probamos 'aditiva y multiplicativa'.

Pusimos en funcionamiento el bucle del modelo y, en este caso, como la variabilidad según se ejecutaba era muy grande, calculamos el error medio de repetir la misma predicción cinco veces. El menor error, aunque más alto que los demás por un outlier, fue el de una estacionalidad de 12, una función de pérdida MSE y una estacionalidad aditiva. Todo con un MAPE de 1,1599.

Probamos el modelo una vez y nos da un MAPE de 0,3445.

Tabla 2: Distintos modelos y error medio obtenido con el MAPE

	Modelo	MAPE Diario
0	ETS	0.2647
1	ARIMA	0.3052
3	Neural Prophet	0.3445
2	TBATS	0.3710

Fuente: TIB y elaboración propia

Elegimos, por tanto, tras analizar los datos resumidos en la Tabla 2, el modelo ETS como el mejor modelo para predecir la serie diaria.

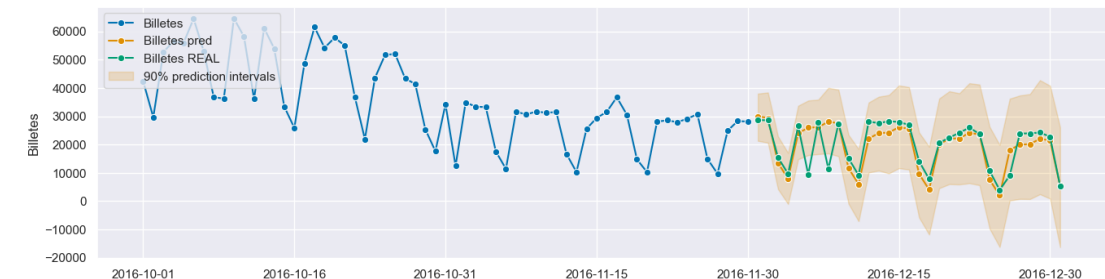
Tabla 3: Salida de resultados del modelo de pronóstico utilizando el método ETS

ETS Results				
=====				
Dep. Variable:	Billetes	No. Observations:	1430	
Model:	ETS(AMA)	Log Likelihood	-14244.966	
Date:	Mon, 08 Jan 2024	AIC	28515.933	
Time:	14:11:27	BIC	28584.384	
Sample:	01-01-2013	HQIC	28541.494	
	- 11-30-2016	Scale	26301931.972	
Covariance Type:	approx			
=====				
=====				
	coef	std err	z	P> z
				[0.025
0.975]				

Fuente: TIB y elaboración propia

Graficamos y establecemos una intervalos de confianza de 0.9 por encima y por debajo de la predicción para compararla con el test, de tal forma que se nos queda como vemos en el Gráfico 11.

Gráfica 11: Serie más predicción utilizando el método ETS



Fuente: TIB y elaboración propia

Como podemos comprobar, los únicos dos errores se producen en los martes, 6 de diciembre de 2016 y jueves, 8 de diciembre de 2016, ambos días de entre semana pero festivos, por la no utilización de una variable dummy para corregir estas predicciones.

De esta forma, comprobamos el MAPE, que ya lo teníamos, y el resto de métricas que utilizamos y podemos ver en la Tabla 4.

Tabla 4: Métricas derivadas de la aplicación del modelo ETS a la predicción de la parte test.

	Metrica	Valor
0	MAPE	0.26
1	MSE	27020314.70
2	R2 Score	0.59

Fuente: TIB y elaboración propia

Para medir la precisión de este modelo comprobamos el MAPE, el MSE y el R2 score.

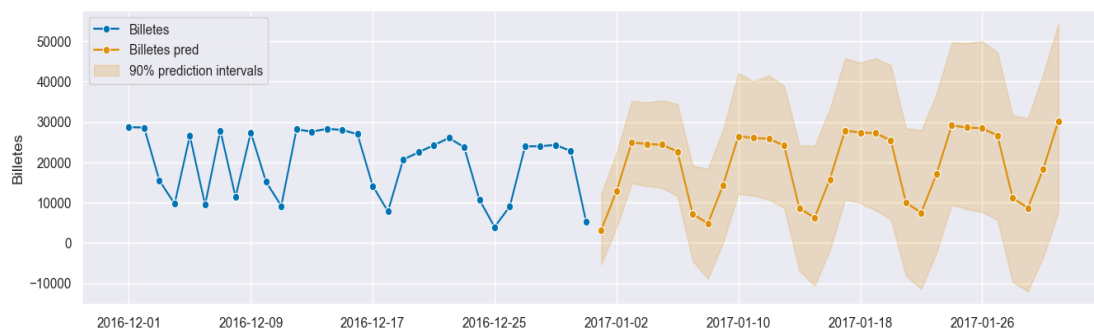
Para elegir modelo nos hemos centrado en el MAPE, que es el que proporciona una medida relativa de la precisión de un modelo de pronóstico al evaluar el porcentaje promedio de error entre las predicciones y los valores reales.

El MSE, que es la suma de los cuadrados de los errores dividido entre el número total de observaciones, penaliza fuertemente los errores de mayor calado.

Con R2 conseguimos saber cuánto de bien explica el modelo la variabilidad total de los datos. Su interpretación varía entre 0 y 1, y un valor más alto indica un mejor ajuste del modelo.

Pasamos por tanto a realizar la predicción de la misma forma que predijimos sobre test y obtenemos estos resultados que se detallan en los .html.

Gráfica 12: Salida de resultados del modelo de pronóstico utilizando el método ETS



Fuente: TIB y elaboración propia

Modelización de series semanales

Repetimos las pruebas iniciales de la modelización diaria. En este caso, es todo prácticamente igual. La principal diferencia es que aquí queda eliminada la mayor estacionalidad de los datos diarios: la estacionalidad semanal. Esto es porque al tener los datos completos semanales, ya no se establecen diferencias entre los sábados, los domingos, y el resto de la semana.

En este caso, en todos los modelos el periodo de estacionalidad es 52, es decir, el número de semanas que tiene un año. El resto, todo igual. Al eliminar esta estacionalidad y,

especialmente, al predecir solo cinco periodos (las cuatro semanas completas que tiene enero más el pico de la quinta), el error será menor porque es más fácil predecir un quinto valor que predecir un trigésimo primero que queda muy lejos en el horizonte temporal. También hubiera sido una opción, a la hora de hacer el preprocesamiento de datos, que las semanas no fuesen de domingo a sábado si no de miércoles a martes, porque enero acaba en martes y así nos ajustaríamos a lo que nos piden, comenzando la predicción el miércoles 28 de diciembre y sustituyendo los cinco primeros valores predichos por cinco valores reales.

En este caso, tras realizar todos los bucles probando distintos hiperparámetros y aplicando la teoría aprendida en clase, nos quedamos con los distintos MAPE que aparecen en la Tabla 5.

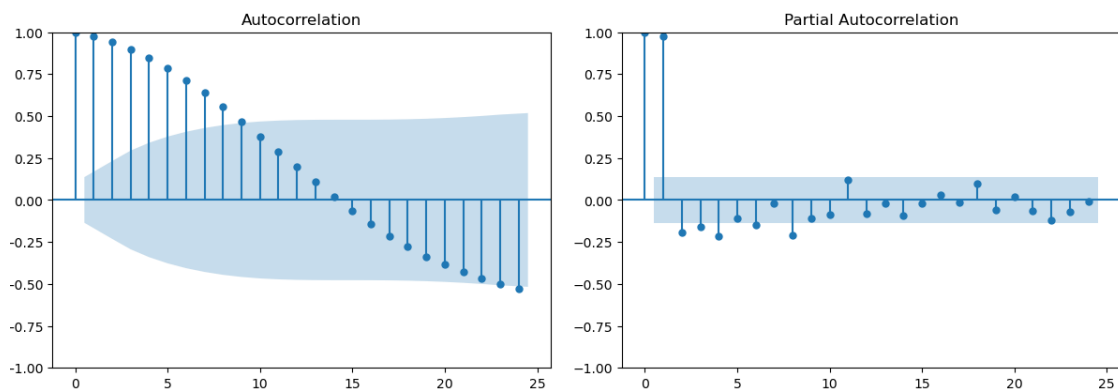
Tabla 5: Métricas derivadas de la aplicación del modelo ETS a la predicción de la parte test.

	Modelo	MAPE semanal
1	ARIMA	0.0502
0	ETS	0.0821
2	TBATS	0.0946
3	Neural Prophet	0.3162

Fuente: TIB y elaboración propia

Nos quedamos, por tanto, con el modelo ARIMA, al que previamente le habíamos aplicado la transformación logarítmica. En la Gráfica 12 podemos ver cómo, el último retraso es el que más influye en el valor actual (y el único que afecta positivamente). Todas las demás que están fuera de la banda que se ve en la parte derecha de la Gráfica 12, afectan menos pero negativamente.

Gráfica 12: ACF y PACF semanal de la venta de billetes del TIB.



Fuente: TIB y elaboración propia

En el notebook complementario de pruebas al que antes nos referimos, hice una búsqueda con un bucle for en el que se comprobase cuáles son los parámetros que minimizan el MAPE. La prueba la hice entre todos los posibles seasonal (que es booleano), stationary (que también es booleano), information_criterion y scoring. La mejor combinación, tras analizar todas las posibles la que aparece en la Tabla 6.

Tabla 6: Salida de resultados del modelo de pronóstico utilizando el método ETS.

SARIMAX Results			
=====			
Dep. Variable:	y	No. Observations:	
203			
Model:	SARIMAX(3, 1, 1)x(2, 0, [], 52)	Log Likelihood	
-2260.598			
Date:	Mon, 08 Jan 2024	AIC	
4535.195			
Time:	17:48:58	BIC	
4558.353			
Sample:	01-12-2013	HQIC	
4544.565			
	- 11-26-2016		
Covariance Type:	opg		
=====			

Fuente: TIB y elaboración propia

Para medir la precisión de este modelo comprobamos el MAPE, el MSE y el R2 score, todos ellos en la Tabla 7.

Para elegir modelo nos hemos centrado en el MAPE, que es el que proporciona una medida relativa de la precisión de un modelo de pronóstico al evaluar el porcentaje promedio de error entre las predicciones y los valores reales.

El MSE, que es la suma de los cuadrados de los errores dividido entre el número total de observaciones, penaliza fuertemente los errores de mayor calado.

Con R2 conseguimos saber cuánto de bien explica el modelo la variabilidad total de los datos. Su interpretación varía entre 0 y 1, y un valor más alto indica un mejor ajuste del modelo. A pesar de que el MAPE sea menor aquí que en la predicción diaria, la explicabilidad es ligeramente menor (0,5927 en la predicción diaria y 0,5816 en la semanal).

Esta predicción tiene menos error que la diaria porque el horizonte temporal de predicción es menor: solo hay que predecir 5 valores y no 31.

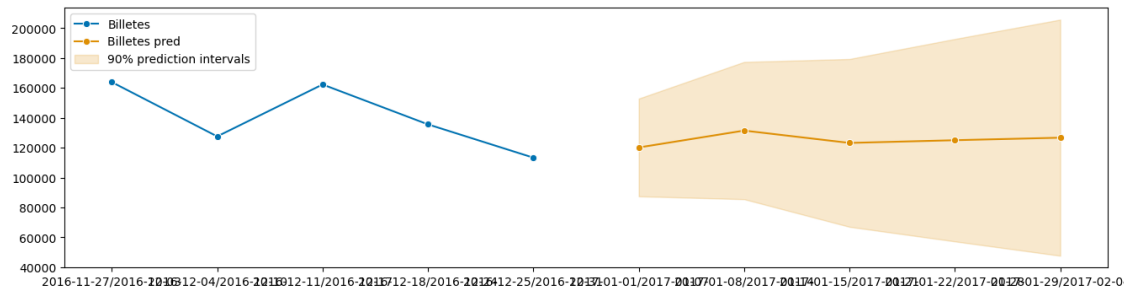
Tabla 7: Métricas de la aplicación del modelo ARIMA a la predicción test semanal.

	Metrica	Valor
0	MAPE	0.07
1	MSE	12657.02
2	R2 Score	0.59

Fuente: TIB y elaboración propia

Finalmente, la predicción que conseguimos es la que se ve en la Gráfica 13, con unas bandas situadas con un ajuste de 0,9 por arriba y por abajo.

Gráfica 12: Salida de resultados del modelo de pronóstico utilizando el método ARIMA



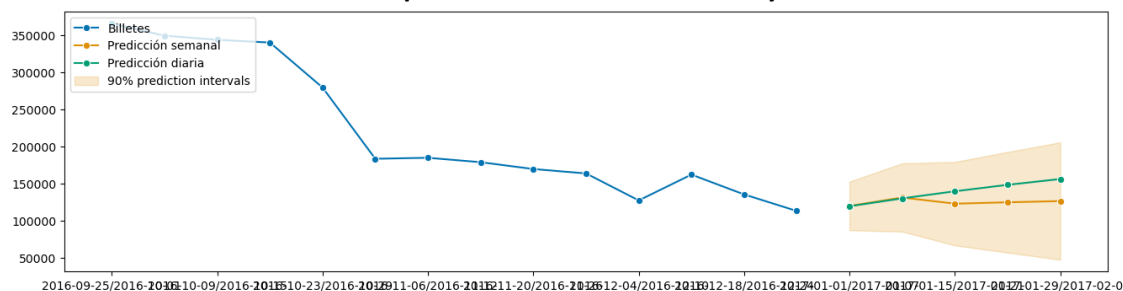
Fuente: TIB y elaboración propia

Finalmente lo pasamos a csv tal y como indica la práctica.

Discusión y conclusiones

A partir de una serie temporal completa, en la que no hubo problemas de outliers ni de valores faltantes o nulos, hemos elaborado dos predicciones. Una semanal (elaborada con un modelo ARIMA), y una diaria, elaborada con un modelo ETS. Como parece lógico, los datos no son exactamente iguales, es decir, la suma de los datos diarios no calca los datos predichos semanalmente. Sin embargo, la predicción diaria (menos fiable que la semanal) se encuentra dentro de los márgenes de la semanal, tal y como podemos ver en el Gráfico 13.

Gráfica 13: Predicción semanal a partir de los datos semanales y de los datos diarios



Fuente: TIB y elaboración propia

Esta es la forma más honesta de presentar unos datos sin manipularlos artificialmente: dos predicciones distintas, pero, que, si miramos las bandas de la más fiable, son consistentes entre sí.

Queda claro, como hemos demostrado durante todo este informe, que, a diferencia del transporte público en ciudades como Madrid, en el caso de las Islas Baleares, este es altamente dependiente del turismo por dos factores:

- 1) una población que se triplica durante las épocas vacacionales y
- 2) la dificultad, salvo alquilarlo, de tener un coche en la isla, hecho que no ocurre en viajes de españoles por la Península Ibérica.

Bibliografía

Gutiérrez, A., Miravet, D. (2016) *Estacionalidad turística y dinámicas metropolitanas: un análisis a partir de la movilidad en transporte público en el Camp de Tarragona*. Revista de Geografía Norte Grande. No. 65. <http://dx.doi.org/10.4067/S0718-34022016000300004>

Albaladejo Pina, I., González Martínez, M. I., Martínez-García, M. P. (2023). *La evolución de turismo en España: Cataluña y Andalucía versus Baleares y Canarias*. Economía Industrial Núm.426, 2022: Sostenibilidad, Innovación y Competitividad Turística págs.:37-44