

Series Temporales práctica

Final: TOSCOS

Realizado por: Eduardo Salgado Díaz del Río



Introducción:

En el fichero datos.csv se encuentran las ventas (Sales) de las diez tiendas de la cadena de supermercados TOSCOS. Son datos diarios y además de las ventas tiene información sobre el número de clientes que han entrado en cada tienda (Customers), si está abierta (Open), si tienes promociones (Promo), si es fiesta (StateHoliday) y si es día lectivo (SchoolHoliday).

Las tiendas están divididas en 3 zonas (1, 2 y 3) y tienen 3, 3 y 4 tiendas respectivamente.

Así la tienda T2c, es la tienda c de la zona 2.

El objetivo es predecir las ventas desde el 1 de agosto de 2015 al 10 de septiembre de 2015.

Objetivos de la práctica:

1. Graficar las ventas diarias de las 10 tiendas.
2. Graficar las ventas diarias de las 3 zonas.
3. Graficar las ventas totales diarias.
4. Graficar las ventas totales semanales.
5. Predecir las ventas totales diarias.
6. Predecir las ventas totales semanales.

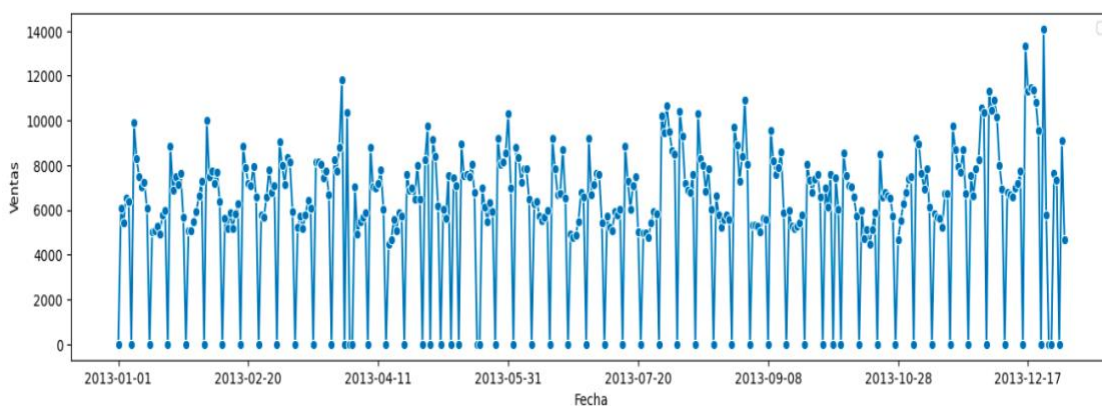
Adicionalmente se tiene que entregar dos csv's con las predicciones.

1. Graficar las ventas diarias de las 10 tiendas

En la introducción podemos observar como están divididas y subdivididas las diferentes tiendas de la empresa Tosco.

Para empezar la empresa Tosco esta dividida en tres zonas enumeradas del uno al tres, dentro de los sectores encontramos que las tiendas están detalladas, siendo en el primer sector y segundo un total de tres tiendas y para el último sector un total de cuatro tiendas.

Para poder graficar de una forma clara y concisa estas graficas ya que al ser un total de diez tiendas, el grafico quedaba ilegible. Hemos decidido separar las ventas diarias por año, es decir, cada tienda tiene un total de tres graficas.



Este seria un ejemplo de las grafica de ventas diarias para la tienda T1a en el año 2013.

Los puntos en el cero representan cuando la tienda esta cerrada, esta parte será importante posteriormente para poder desarrollar unas buenas predicciones de las ventas diarias, así mismo existen otras posibilidades de que las ventas diarias sean igual a cero, como hemos comentado anteriormente, existen columnas en la base de datos proporcionada como 'StateHoliday' o 'SchoolHoliday'.

Por otra parte, realizando el estudio del dataset encontramos como los picos de ventas de cada tienda se encuentran fuertemente correlacionados con los días que existen promociones en esa tienda, observando que son los valores atípicos del dataframe, es decir, los outliers del dataset se corresponden con la columna 'Promo'.

En conclusión, observamos que todas las tiendas tienen un incremento de ventas anual, también podemos decir que las ventas tienen unos patrones de ventas repetitivos como se pueden observar en las graficas.

2. Graficar las ventas diarias de las tres zonas

Para empezar, el calculo de las ventas diarias de las tres zonas se hace mediante un sumatorio de cada tienda en cada zona, eso nos da un valor total de las ventas diarias por sector.

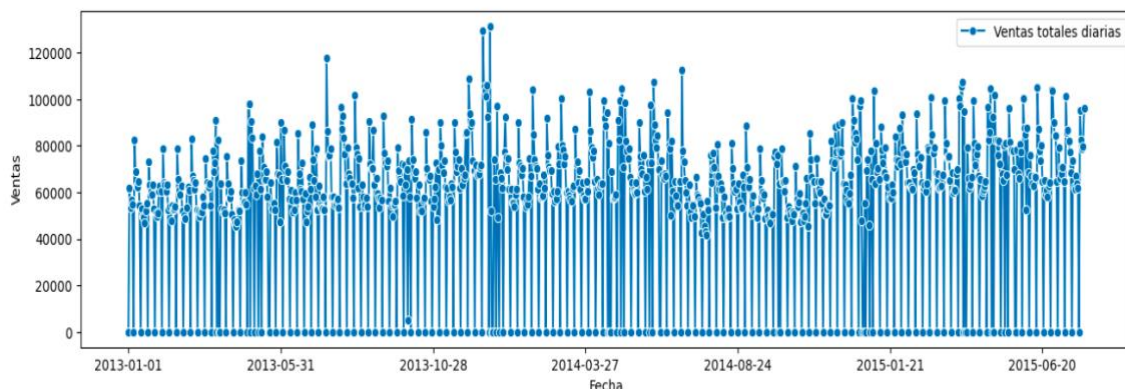
Continuando lo anterior en el primer sector, es el que menos ventas diarias totales tiene en comparación con los otros tres sectores teniendo una media entorno a los 17000\$ el primer año, aumentado un poco su rendimiento en ventas diarias los años venideros.

El segundo sector, es el mas estable de los tres situándose su media aproximada entorno a los 25000\$ a lo largo del tiempo, sigue unos patrones de ventas repetitivos.

El tercer sector, es el que mas ventas diarias totales consigue de media, aunque hay que resaltar que el verano de 2014 sus ventas bajan considerablemente, esto es debido a que la tienda T3b cerro en verano, por tanto, las ventas diarias totales bajaron. También cabe recordar que las ventas totales diarias de este sector son mayores puesto que tiene una tienda adicional.

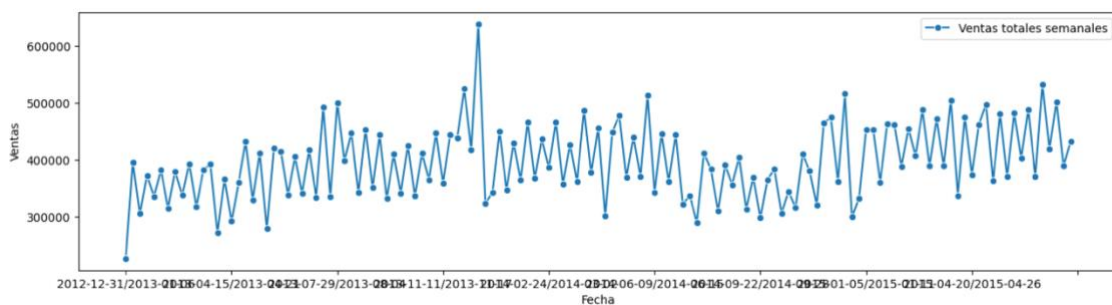
3. Graficar las ventas totales diarias

Para poder graficar las ventas totales diarias realizamos un sumatorio de las ventas totales por zonas, obteniendo así las ventas podemos observar en la grafica que se muestra que sigue una tendencia de la serie temporal un tanto monótona, el problema es que las ventas totales diarias en el año 2014 bajan considerablemente, esto es debido a lo comentado anteriormente en el sector tres que una de las tiendas del sector esta cerrada durante ese periodo, el grafico no se puede observar con claridad es por ello que lo hemos ampliado año a año para poder observarlo mejor.



4. Graficar las ventas totales semanales

Las ventas totales semanales son obtenidas al agrupar las ventas diarias en semanas, observamos como no son tan repetitivas como las diarias. En este punto de la práctica la problemática con los ceros deja de tener la importancia que tenia anteriormente, puesto que todas las tiendas cierran los domingos. Podemos observar también como durante el año 2013 mas o menos se mantiene constante las ventas semanales, siendo este el segundo mejor año de la serie, en contrapunto de este el año 2014 es el que menos ventas totales semanales obtiene. Para finalizar el año 2015, es el que mejores ventas tiene de los tres. La serie temporal de las ventas totales semanales no es tan repetitiva como las diarias, si es cierto que sigue teniendo algunas tendencias parecidas a lo largo de los años.



5. Predecimos las ventas totales diarias

La predicción de las ventas totales diarias las calculamos mediante el algoritmo de predicción para las series temporales de Facebook Prophet.

En este caso vamos a crear tres diferentes Prophet, el primero modelo es el base para sacar las predicciones, para elegir el mejor modelo de Prophet utilizaremos la métrica de r^2_score , para este primer modelo obtenemos un r^2 de 0.60, observamos como esta métrica puede ser mejorable.

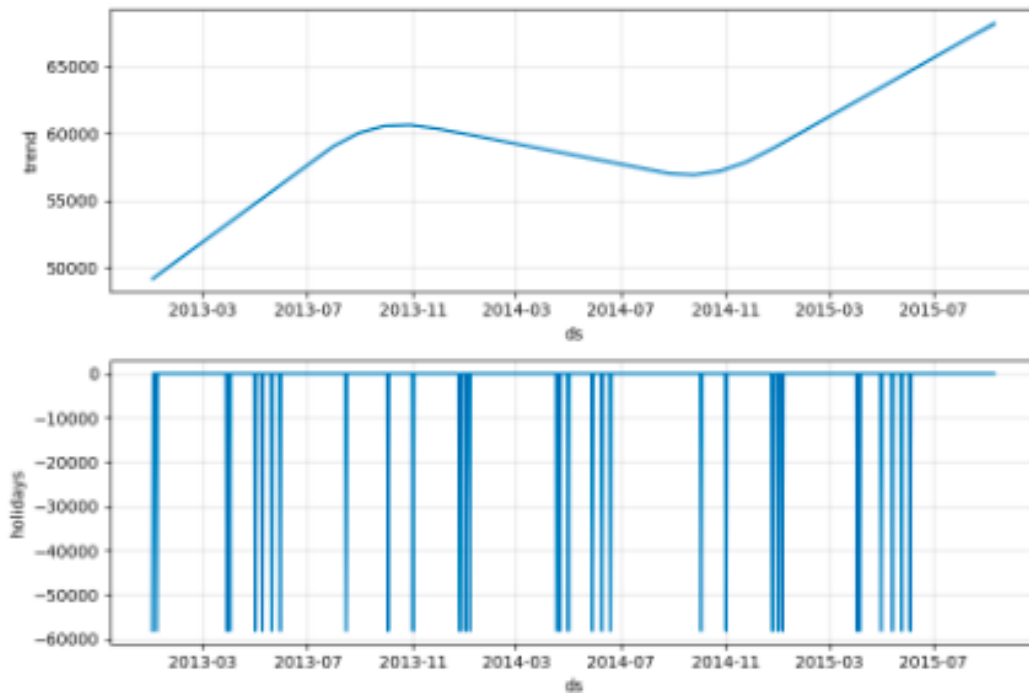
Para ello, en el segundo modelo vamos a añadir estacionalidades semanales, mensuales y anuales, observamos como el modelo empeora un poco respecto al r^2_score .

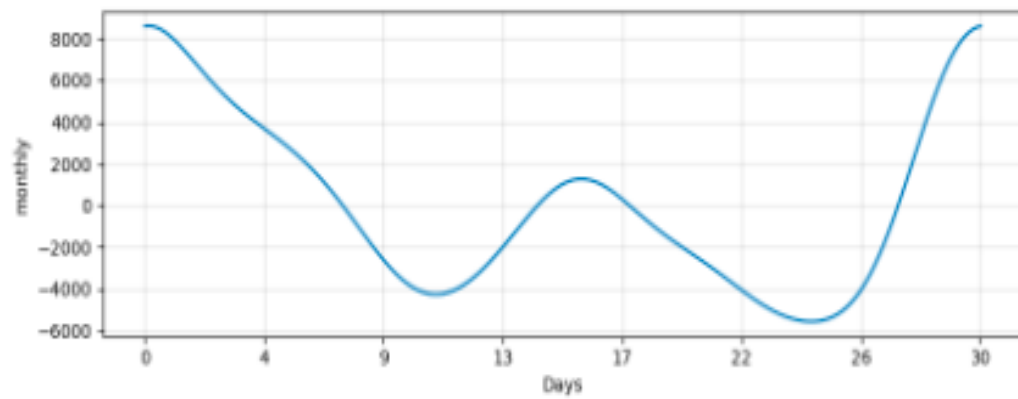
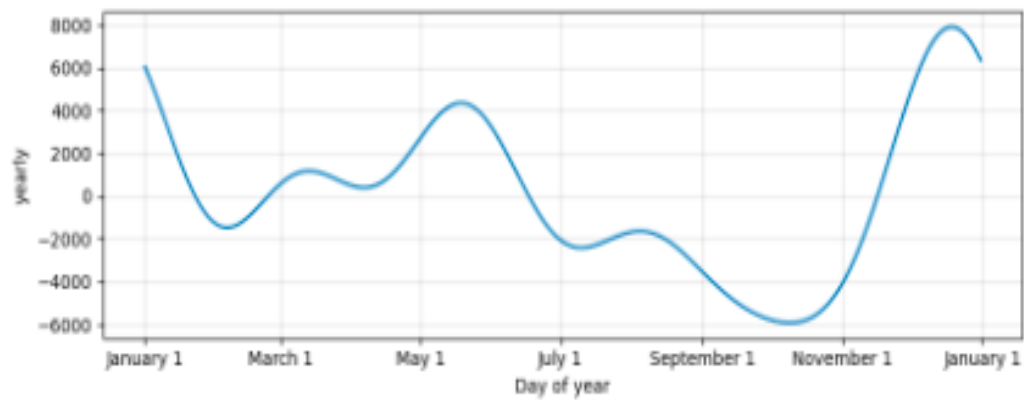
Para finalizar las practica de las predicciones de ventas totales diarias, añadiremos una nueva columna al modelo, añadimos estacionalidades mensuales y anuales, además de estas dos nuevas variables al modelo, también añadiremos la columna

holidays a la predicción puesto que este modelo ofrece la posibilidad de añadirla, y esto hace que mejore el r^2 a un total de 0.75.

Es por ello, esta va a ser el modelo que vamos a utilizar para realizar las predicciones diarias hasta el 10 de septiembre de 2015.

Esta sería la estacionalidad mensual y anual del modelo con la variable holiday añadida.





2015-08-01	76669.45486666624
2015-08-02	0.0
2015-08-03	92633.81291114188
2015-08-04	81223.19118130491
2015-08-05	76766.51663014678
2015-08-06	76230.92274748108
2015-08-07	78236.06626973883
2015-08-08	69908.21953954236

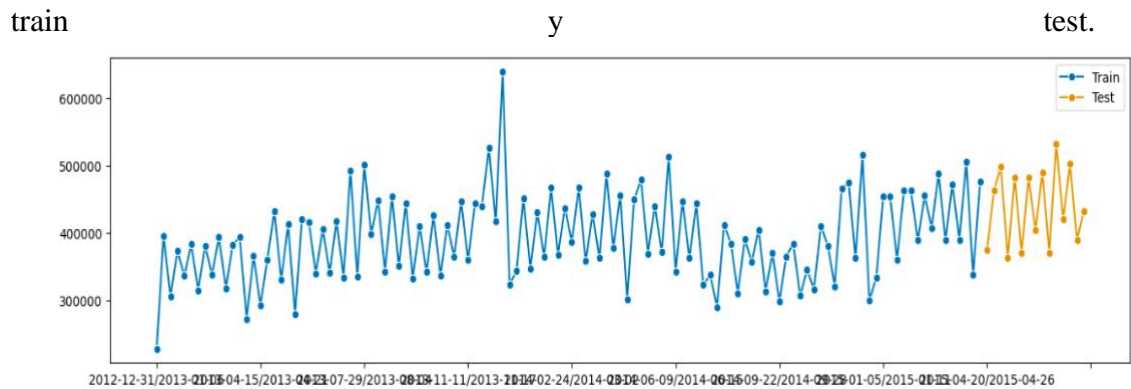
2015-08-09	0.0
2015-08-10	83211.02023844069
2015-08-11	71870.4429150164
2015-08-12	68049.27070004406
2015-08-13	68830.86429806452
2015-08-14	72931.15283101305
2015-08-15	67417.74045991774
2015-08-16	0.0
2015-08-17	86543.71910005737
2015-08-18	76497.47563033045
2015-08-19	72331.6312943735
2015-08-20	71605.21356539014
2015-08-21	73601.8050133665
2015-08-22	65623.15500512657
2015-08-23	0.0
2015-08-24	80297.65905337164
2015-08-25	69636.57678519885
2015-08-26	66159.38731608831
2015-08-27	67100.4105379794
2015-08-28	71842.84787447807
2015-08-29	67997.36083941224
2015-08-30	0.0
2015-08-31	91821.501497203
2015-09-01	83118.83173808144
2015-09-02	79256.30485343713
2015-09-03	78192.89160868578
2015-09-04	79731.3703658313

2015-09-05	71517.29055812638
2015-09-06	0.0
2015-09-07	85931.0078429447
2015-09-08	74422.8972700594
2015-09-09	69364.04633032579
2015-09-10	68106.52503697488

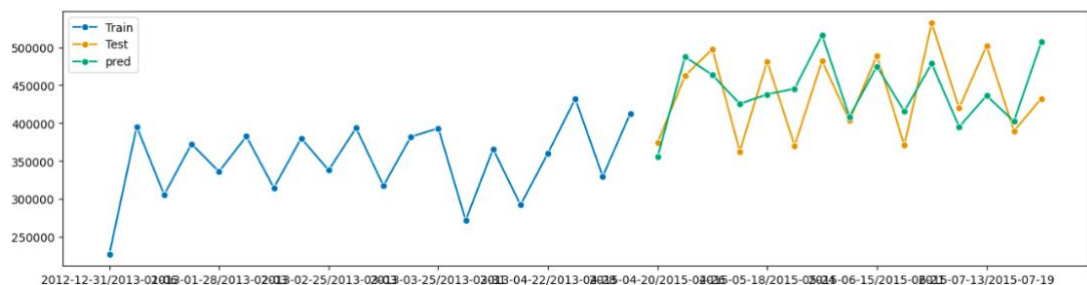
Estas son las predicciones hasta el 10 de septiembre de 2015, como se puede observar los días que son cero, como hemos comentado anteriormente es debido a que son domingos, y por tanto las tiendas están cerradas y no se puede comprar en las tiendas. Después de estas observaciones vamos a observar un poco diariamente, podemos ver como las ventas diarias mas alta son los lunes, y que la mas baja son los sábados.

6. Predecimos las ventas totales semanales

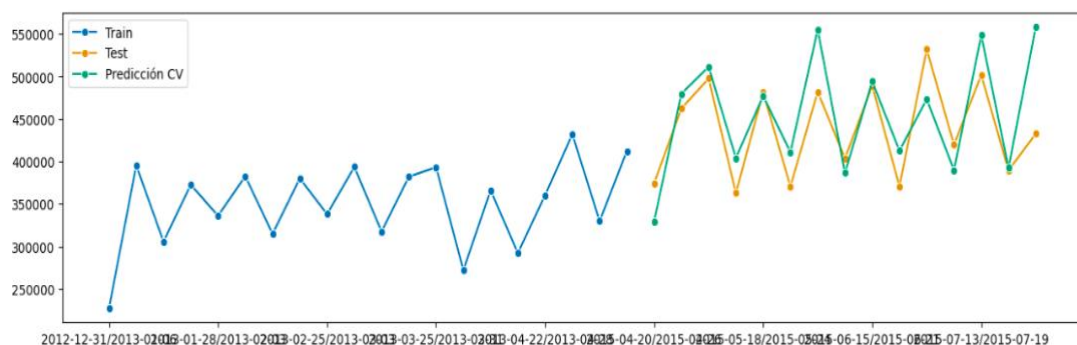
Hemos decidido separar los datos entre train y test destinando un 20% de los datos para test, en la siguiente grafica podeos observar como quedaría la serie temporal con



Para las ventas semanales hemos decidido realizar el modelo de predicción ETS, en este caso hemos utilizado un modelo ETS con una estacionalidad aditiva, en este caso hemos obtenido un MAPE de un 9,11%.



Para seguir con el análisis del modelo hemos buscado aplicar la validación cruzada para mejorar el modelo, mediante la realización de un grid en búsqueda que minimizar el error del modelo y así tener una predicción que se asemeje mas a la real. Con esta cross validation en el modelo ETS, obteniendo como resultado un valor de MAPE de un 8,71%.



Observamos como mediante la utilización de la validación cruzada el modelo de predicción entiende mejor los mínimos, aunque la diferencia con el MAPE sea poca al graficarla observamos como tiene una clara mejora.

Este modelo de ETS con validación cruzada será el que utilicemos para realizar las predicciones a cinco semanas.

430688.0592088179
475763.55437603325
376603.501891187
483049.51249772403
385460.5802139144

Estas son las predicciones mensuales para las siguientes semanas, en total se predican cinco nuevas semanas.