

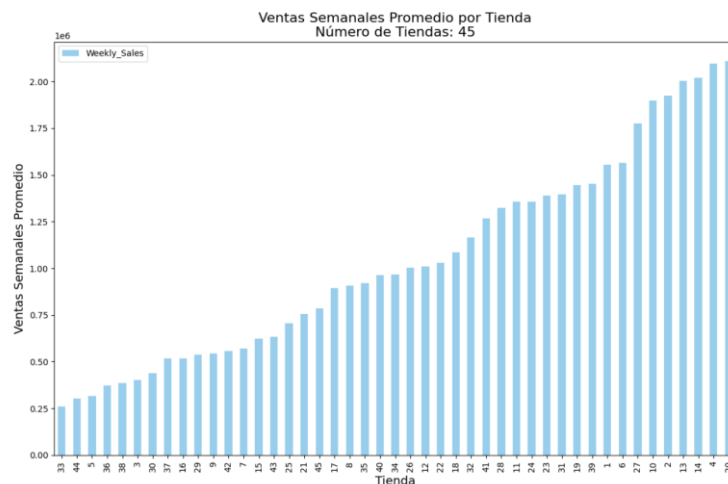
# INTERSEMESTRAL

En este dataset proporcionado se han dado los datos económicos y de ventas de una cadena de supermercados estadounidense (Walmart). En este dataset se ha podido observar que teníamos todos los datos a excepción de un dato de unemployment que hemos decidido eliminarlo ya que era solamente uno y no presentaba un gran problema para la cantidad de dataset que teníamos. Si no hubiéramos tenido tantos datos, hubiéramos decidido sustituirlo por un valor de al lado o hacer la media entre el anterior. Sin embargo, no ha hecho falta.

Los datos que teníamos eran:

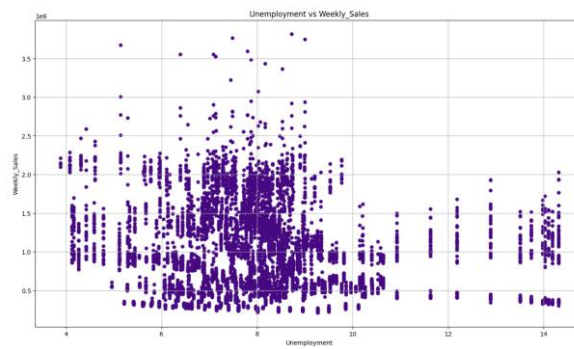
1. Store: es el id de la tienda
2. Fecha: fecha (podemos ver que siempre es por semanas (siempre en viernes)
3. Weekly\_Sales: compras de esa tienda en la semana
4. Holiday\_Flag: si es un día festivo o no.
5. Temperature: temperatura
6. Fuel\_Price: precio de la gasolina
7. CPI: Consumer Price Index
8. Unemployment: la tasa de desempleo que existe

Hemos podido observar que hay una gran diferencia entre las tiendas que presentan unas ventas muy altas y más bajas tal y como hemos podido observar en el gráfico proporcionado.



Además, cabe destacar que en el gráfico de tiempo compras hemos podido ver que durante los meses de noviembre a diciembre las ventas aumentan en gran cantidad, por lo que la primera recomendación que haría a la empresa es que aumente su abastecimiento durante esos meses, ya que es algo cíclico anual:



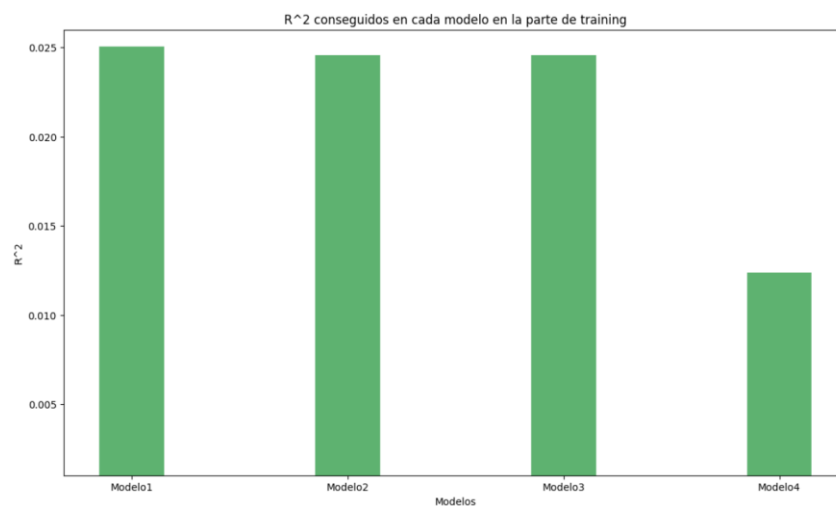


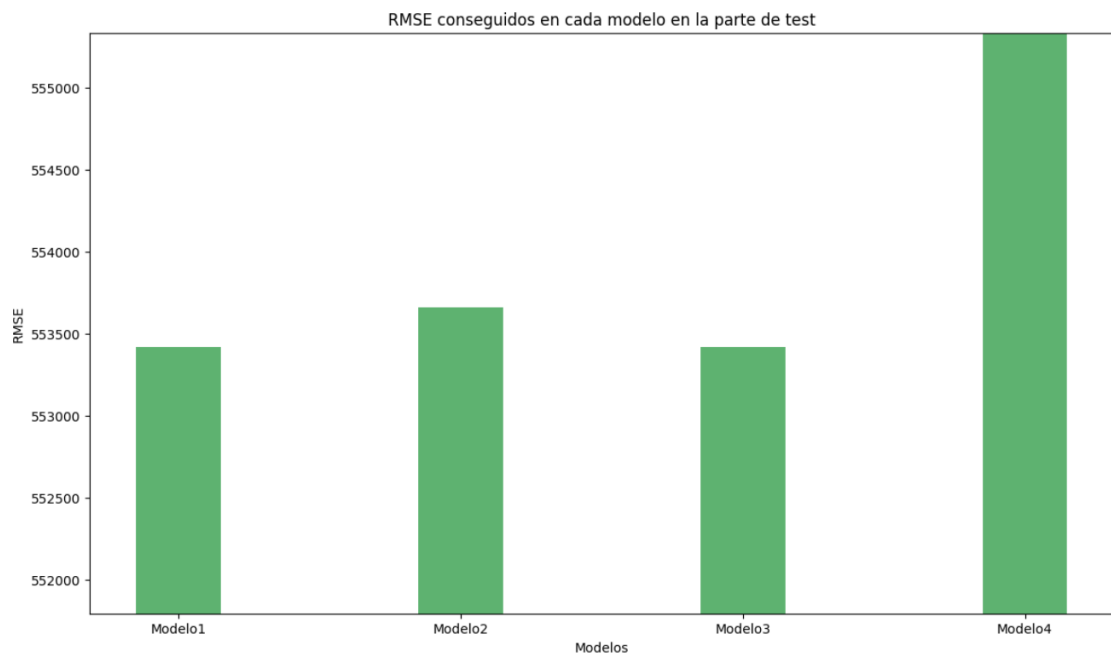
Podemos ver en este último que sí que ocurre que a menor unemployment hay más ventas por semana.

Después hemos observado que sí que existe una diferencia estadísticamente significativa entre las ventas de semanas festivas y no festivas

Ventas promedio de semanas festivas: 1122887.8923555557  
 Ventas promedio de semanas no festivas: 1041310.3793482621  
 Diferencia promedio de ventas: 81577.51300729357  
 Estadístico t: 2.6782781784408503  
 Valor p: 0.007641793439754634  
 Existe una diferencia estadísticamente significativa entre las ventas de semanas festivas y no festivas.

Podemos ver que la diferencia entre los modelos que hemos estudiado no es muy notable o es poco precisa. Por lo que lo primero que recomendaría es hacer otro tipo de modelo que no sea regresión lineal. Sin embargo podemos ver que el modelo4 todo y que tiene muy buena  $R^2$  podemos ver que en el test falla, por lo que podríamos estar teniendo overfitting. Los dos modelos que son mejores considero que son: el 1 y el 3. Finalmente, por el hecho de que es más sencillo tomaría el 3 que tiene las variables: 'Fuel\_Price', 'CPI', 'Unemployment'.





Cabe destacar que el modelo 4 es también bueno, pero peca de overfitting ya que en el training hay una gran diferencia pero luego a la hora del testing es el mismo