Introducción a la Analítica Empresarial

Proyecto - Black Friday

2022

Milagros Correa, Emilia Regent, Mia Robaina, Ignacio Vilar del Valle

****

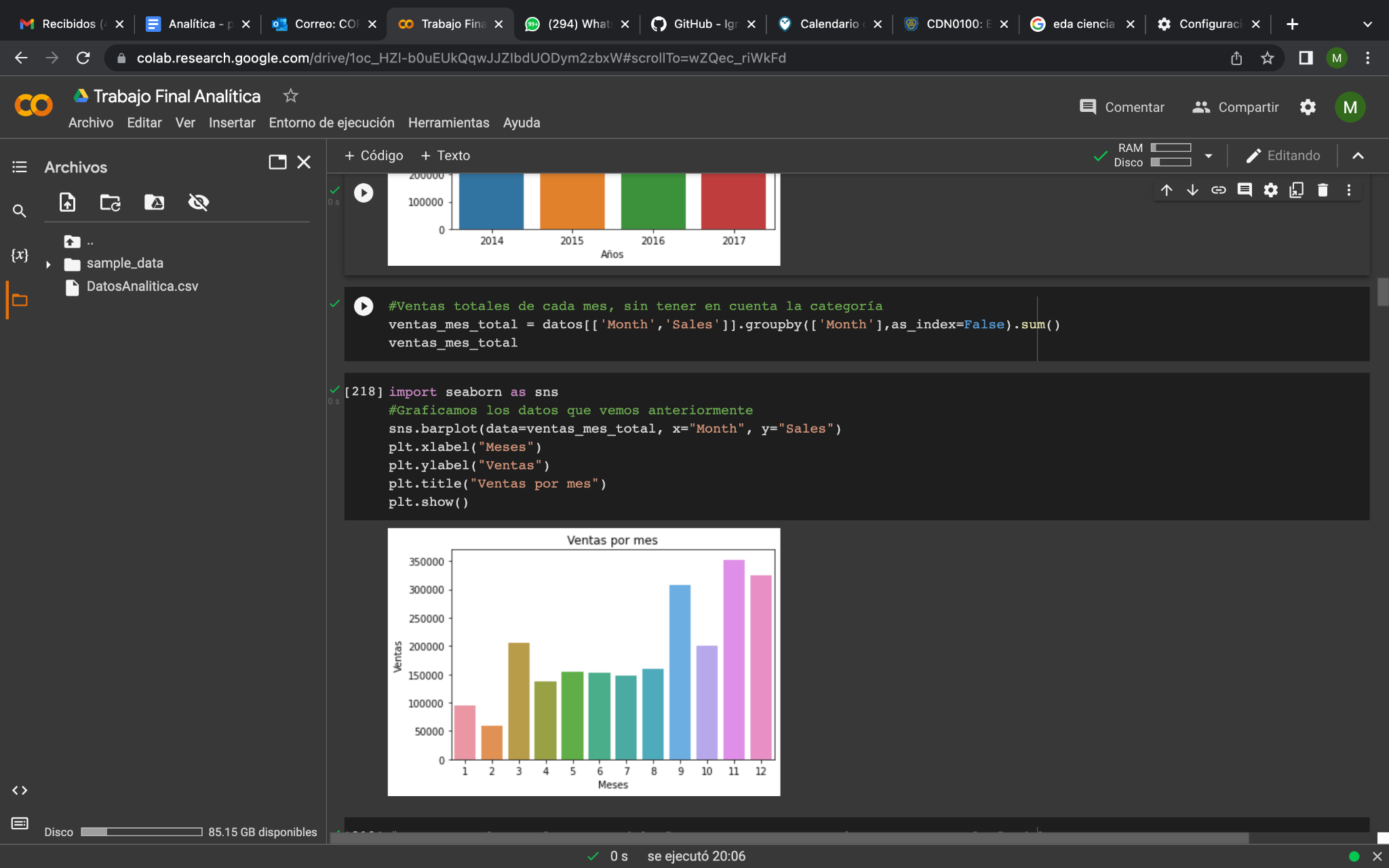
**Desarrollo del Proyecto**

El proyecto que nos tocó desarrollar trata sobre las ventas durante la temporada del Black Friday. La idea era poder predecir los valores de ventas para los próximos Black Fridays con el uso de modelos predictivos y ver si hay grandes diferencias positivas en las ventas en ese día.

Para poder trabajar sobre esto, nos fue necesario indagar más el tan famoso “Black Friday” y así conocer cómo funciona, cómo se vende, en qué países está más presente. Luego de esto, nos enfrentamos a nuestro primer proyecto con un software de analítica de datos, para lo cual tuvimos que poner en práctica las habilidades aprendidas durante el curso, afinar aquellas que no tuviésemos tan claras y profundizar más para poder desarrollar un mejor trabajo.

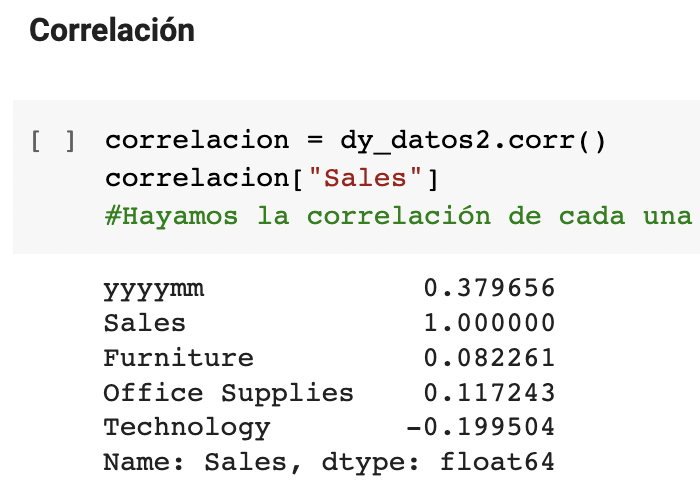
Para comenzar con el proyecto, realizamos una búsqueda de datasets para ver valores sobre las ventas de una empresa. Inicialmente, la consigna del trabajo fue planteada sobre la empresa Amazon, pero no pudimos dar con ningún dataset de dicha empresa. Debido a esto, continuamos la búsqueda y encontramos un dataset de otra empresa, el cual nos pareció lo suficientemente completo para poder realizar el proyecto, con valores que podrían representar los cambios en ventas de Amazon, solo que a menor escala.

Una vez obtenido el Dataset, nos enfrentamos al Data Cleaning. En el dataset encontramos columnas, filas y datos que nos eran irrelevantes para nuestro proyecto. Por eso, fue necesario modificarlo, eliminando aquello que no íbamos a utilizar y reordenando la información para que se adaptara a lo que necesitábamos. Nos quedamos con las columnas de datos que nos daban información sobre el tipo de artículos, el número de ventas y las fechas de cada una de las compras. Esto significó hacer un recorte minucioso de los datos y unir y organizarlos de forma adecuada, ya que había que tener cuidado de mantener la coherencia del Dataset.

Lo siguiente que realizamos es explorar los datos que tuvimos realizando un EDA. Un EDA justamente es un análisis exploratorio de datos que, con el uso de gráficos, tablas y otros elementos visuales, busca explorar y analizar un conjunto de datos. Realizamos esto con el objetivo de aprender, investigar acerca de nuestro dataset y de cómo podíamos enfrentarnos a nuestro problema. En esta etapa relacionamos variables, hicimos tablas separando columnas para compararlas. Agrupamos y sumamos las ventas en meses, años y por categoría. Visualizamos la información mejor para ver a qué nos estábamos enfrentando. A la derecha podemos ver un ejemplo de lo realizado en el EDA. Aquí, usando la herramienta ‘.groupby’ sumamos las ventas de cada uno de los meses de todos los años y luego la graficamos. De este tipo hicimos varios hasta considerar que ya comprendimos el dataset a tal nivel que podríamos comenzar a predecir los datos. 

Luego realizamos un filtrado de datos para poder trabajar con ellos más adelante al momento de la correlación y los modelos. Para eso creamos un dataset auxiliar donde creamos una columna nueva que incluía los años y los meses. Luego, a partir de ella, filtramos los datos y creamos otro dataset con solo los datos de cada año del mes de noviembre. Este era el mes relevante para nosotros ya que es el mes en el que se lleva a cabo el Black Friday año trás año. Además, solamente incluimos los datos de los días en los que se efectuaron los Black Fridays de estos años. En 2014 se realizó un 28, en 2015 un 27, en 2016 un 25 y en 2017 un 24. Finalmente, obtuvimos un dataset con los datos de los artículos por categoría vendidos en los Black Friday de los años pasados. Ese era el dataset con el que continuaríamos el trabajo y realizaríamos las predicciones a futuro.

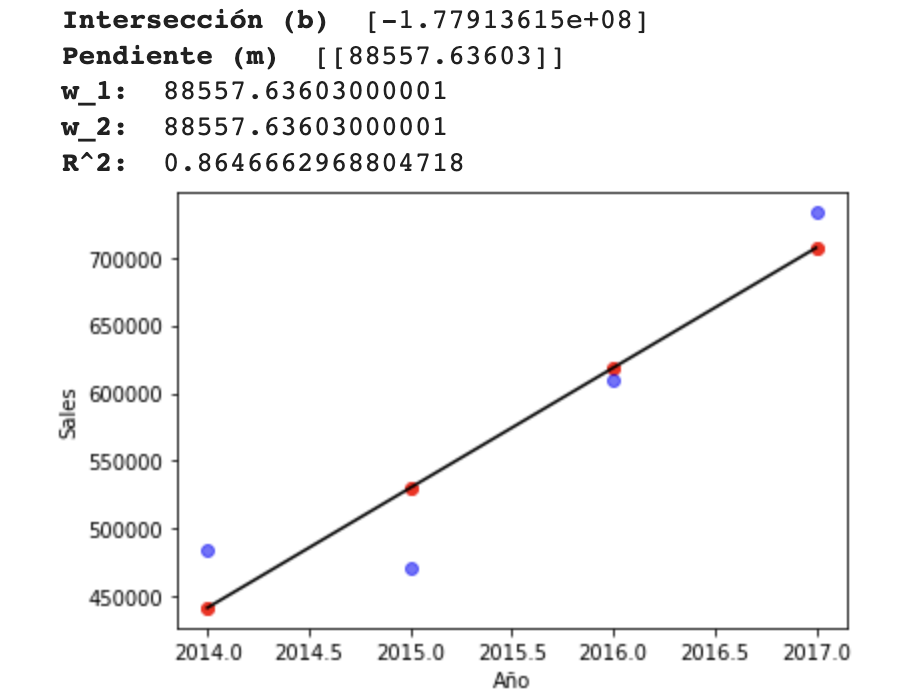
También realizamos una suma de las ventas por categoría en cada uno de estos días. Esto lo realizamos con el objetivo de crear otro dataset igual al anterior, pero incluyendo información de días normales de ventas para la empresa. Realizamos un dataset con los jueves de la semana anterior al Black Friday que consideramos es un día bastante común para todos. A partir de esto realizado, podemos obtener conclusiones sobre si las ventas de Black Friday realmente son mayores que en un día cualquiera.

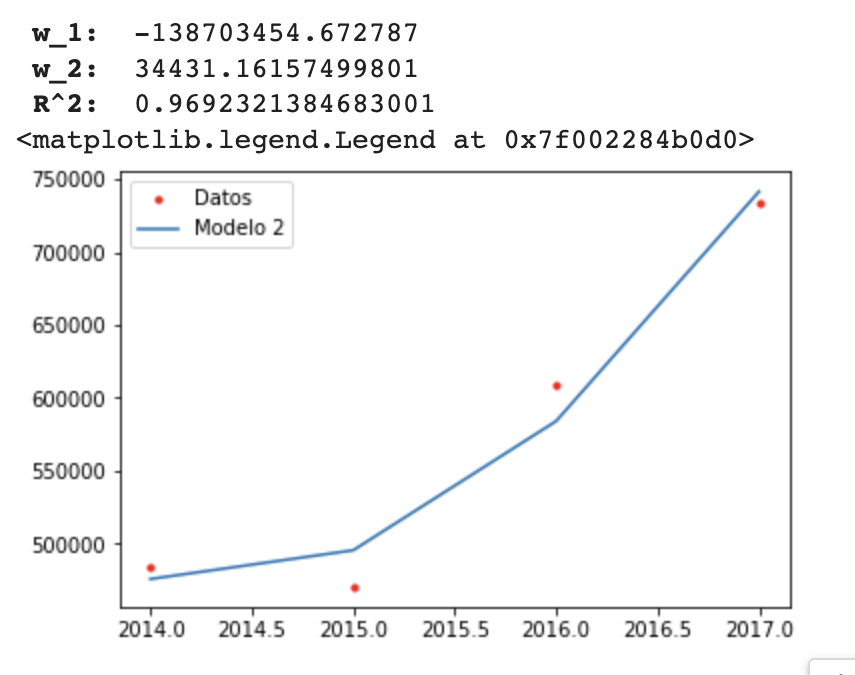
Ahora sí, comenzando con la parte de correlaciones, primero obtuvimos nuestra recta de regresión lineal. Esta es una recta que se ajusta lo mejor posible a los valores y que a partir de ella podemos obtener una estimación de cualquier valor que necesitemos. Además, podemos utilizarla para ver las correlaciones entre las variables usadas con la variable ventas. Esto nos sería útil para ver qué variable es más influyente en las ventas. 

Comenzamos utilizando la herramienta de dummies que ofrece Python. Con esta podemos cambiar variables categóricas y crear ficticias. De esta forma creamos un dataframe con una columna para cada categoría con valores binarios (0 y 1) donde dice en qué fila aparece cada una de las categorías. Este data frame es necesario para alimentar muchos de los estimadores, como lo son la correlación y los modelos. Luego hicimos la correlación y obtuvimos resultados de los que luego sacaríamos conclusiones.

La primera conclusión que obtuvimos, es que las correlaciones son bastante bajas en general, e incluso algunas negativas. No podríamos afirmar que hay una fuerte relación entre las ventas y las categorías o incluso los años y meses de ventas. De todas formas, vemos que esta última variable mencionada, es la que tiene más correlación con un 0,379656. Por lo que nos pareció la más adecuada para seguir investigando y para utilizar en los modelos.

Luego empezamos con los modelos para predecir las ventas. El primer modelo que utilizamos fue el modelo con más de una variable. Relacionamos las variables independientes de 'yyyymm', 'Technology', 'Furniture' y 'Office Supplies' con la variable dependiente sales. Este modelo predijo un valor de R².

Para continuar, realizamos modelos con una sola variable. Para los modelos utilizamos las variables ventas y años por lo mencionado en las conclusiones de la correlación El primero que realizamos fue el modelo de grado 1 y obtuvimos resultados. Luego el modelo de grado 2 y luego el de grado 5. Cada uno de los modelos nos dió un gráfico con una recta de ajuste a los datos. Además obtuvimos un valor de R² para cada uno de ellos. A partir de eso sacamos conclusiones. 

En el modelo de grado 1, obtuvimos un valor de R² de aproximadamente 0,86. Si bien es una correlación moderadamente alta, creímos que al aumentar el grado del modelo, la recta se ajustaría mejor. En el modelo grado 2 ya obtuvimos un R² más alto, como habíamos pensado. El R² nos dió aproximadamente 0,97. Sin duda en este modelo los valores se ajustaban de una mejor manera por lo que la predicción podría ser mucho mejor. Finalmente intentamos con el modelo de grado 5 para ver si se ajustaba aún más. El R² nos dió 0,999. Esto fue demasiado ajustado a los datos. Concluimos que en este modelo se dió un sobreajuste en los datos. Por lo que no sería tan efectivo predecir usando este modelo. 

Habiendo sacado estas conclusiones, decidimos utilizar el modelo dos para predecir las ventas para el 2018. Estas nos dieron un número de ventas mayor al del año anterior como pensábamos que iba a pasar. Dado a que las ventas del Black Friday son más altas que un día cualquiera, incluso en una cantidad de 1000 unidades por categoría, y sumando el contexto de que cada día somos más tecnológicos era esperable que las ventas en este día siguieran subiendo. Luego pronosticamos las ventas del 2019 y las de 2021 que también nos dió que iban a seguir aumentando. Es un número demasiado alto que quizá se aleja mucho de la realidad, pero son consecuencias de que el modelo utilizado es más preciso con datos más cercanos.

**Conclusión**: Habiendo analizado todas las variables y resultados obtenidos podemos decir que el Black Friday si es un factor causante de un aumento de ventas, a su vez todo el proceso nos fue de gran utilidad para adentrarnos en lo que fue un nuevo mundo para nosotros, tomar decisiones basado en lo analiticamente procesado y trabajar en equipo de manera eficiente.