Online Shop Customer Sales

Data Analysis Project

***Introducción:***

En el mundo actual, donde la tecnología ha revolucionado la forma en que compramos y consumimos productos, el análisis de datos desempeña un papel fundamental en la comprensión de los patrones de comportamiento de los consumidores. Este proyecto se centra en un conjunto de datos de ventas recopiladas por una tienda en línea durante el año 2021. Los datos han sido obtenidos de Kaggle, una plataforma reconocida por proporcionar conjuntos de datos de alta calidad para análisis.

***Descripción del Conjunto de Datos:***

El conjunto de datos contiene información valiosa sobre las transacciones de ventas realizadas por la tienda en línea. Cada registro en el conjunto de datos representa a un cliente individual y se compone de diversas variables que capturan diferentes aspectos del comportamiento del cliente y sus compras. Estas variables incluyen:

*Customer\_id:* Un identificador único para cada cliente.

*Age:* La edad del cliente.

*Gender*: El género del cliente, codificado como 0 para masculino y 1 para femenino.

*Revenue\_Total:* El total de ventas realizado por cada cliente.

*N\_Purchases:* El número de compras realizadas hasta la fecha por cada cliente.

*Purchase\_DATE:* La fecha de la última compra realizada en formato DD-MM-YY.

*Purchase\_VALUE*: El valor de la última compra en euros.

*Pay\_Method:* El método de pago utilizado, codificado como 0 para billeteras digitales, 1 para tarjetas, 2 para PayPal y 3 para otros métodos.

*Time\_Spent:* El tiempo que cada cliente pasó en el sitio web, medido en segundos.

*Browser:* El navegador web utilizado por el cliente, codificado como 0 para Chrome, 1 para Safari, 2 para Edge y 3 para otros navegadores.

*Newsletter:* Si el cliente se ha suscrito al boletín informativo, codificado como 0 para no suscrito y 1 para suscrito.

*Voucher:* Si el cliente ha utilizado un cupón de descuento en su compra, codificado como 0 para no utilizado y 1 para utilizado.

***Objetivos del Proyecto:***

El objetivo principal de este proyecto de análisis de datos es explorar y descubrir patrones interesantes en los comportamientos de compra de los clientes en la tienda en línea durante el año 2021. Mediante el uso de herramientas analíticas y visuales, se buscará responder a preguntas clave como:

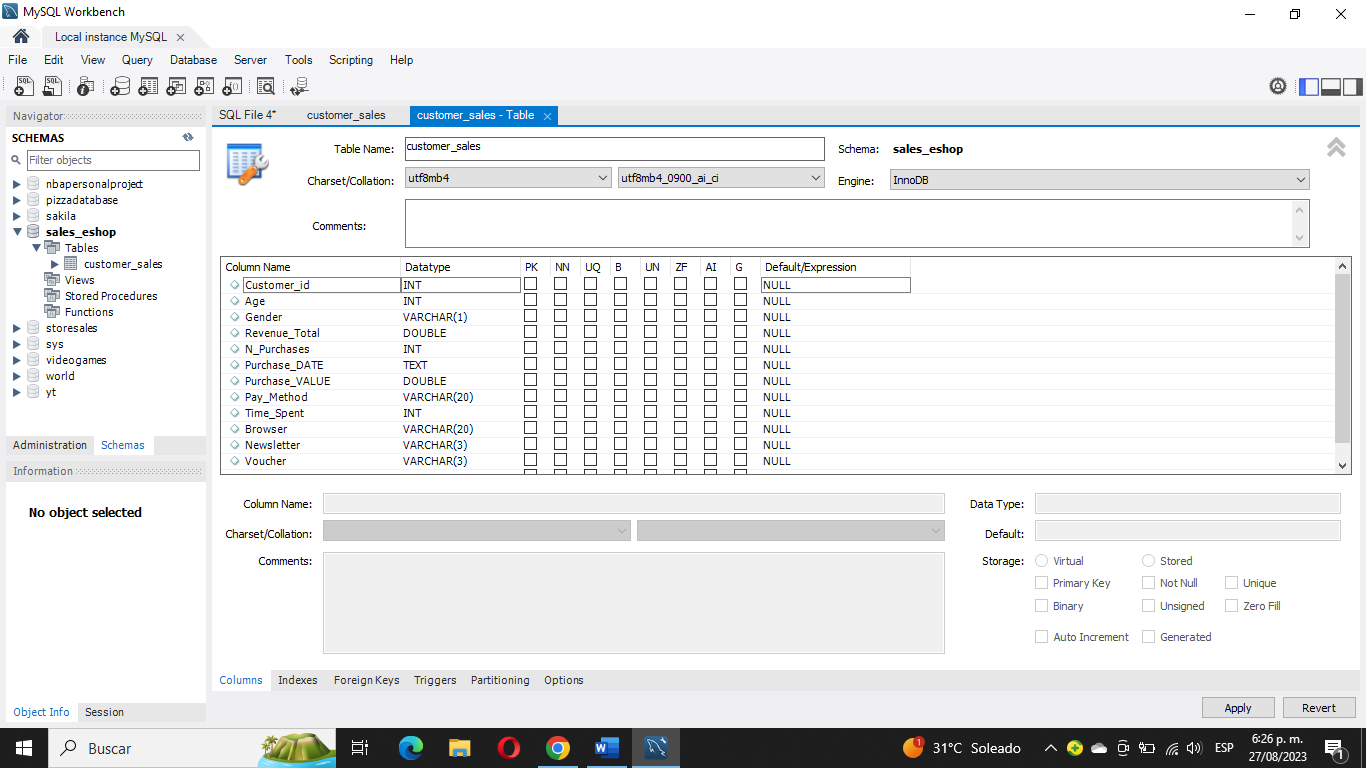
* ¿Cuáles son los patrones de compra según la edad y el género de los clientes?
* ¿Cuáles son los métodos de pago más populares y cómo se relacionan con el valor de las compras?
* ¿Cómo se distribuyen las compras en función del navegador utilizado por los clientes?
* ¿Existe alguna relación entre el tiempo que un cliente pasa en el sitio web y el valor de sus compras?
* ¿Cómo afecta la suscripción al boletín y el uso de cupones de descuento a las tendencias de compra?

***ABC del proyecto***

*Primer paso: Limpieza de datos*

Como es posible ver en la introducción del proyecto, la base de datos “codifica” los valores de la misma con números en vez de texto con el propósito de hacerlo más intuitivo con el analista, sin embargo esto solo complica la visualización del informe y de las gráficas a aquellas personas que no tienen idea de lo que significa o que, en su defecto, se confunden con la leyenda; es por esto que el primer paso es modificar algunas columnas para hacer más fácil la lectura del informe y la visualización de los datos al espectador promedio.

La primera acción de limpieza que realicé en esta base de datos fue cambiar las conversiones de los datos en donde modifiqué las siguientes columnas: *gender, Pay\_Method, Browser, Newsletter* y *Voucher*, que pasaron de “int” a “varchar”.



Esto fue realizado de manera manual en el editor de MySQL sin la ejecución de ningún código.

Ahora es el turno de limpiar los datos y colocar los valores escogidos, en esta ocasión sí fue necesario ejecutar un código largo pero fácil y muy reconocido, para cambiar los datos de solo números enteros a un texto más legible y entendible.

***Nueva Leyenda:***

Lo que deriva en que se cambien las condiciones del principio a las siguientes:

*Customer\_id:* Un identificador único para cada cliente.

*Age:* La edad del cliente.

*Gender*: El género del cliente, codificado como “M” para masculino y “F” para femenino.

*Revenue\_Total:* El total de ventas realizado por cada cliente.

*N\_Purchases:* El número de compras realizadas hasta la fecha por cada cliente.

*Purchase\_DATE:* La fecha de la última compra realizada en formato DD-MM-YY.

*Purchase\_VALUE*: El valor de la última compra en euros.

*Pay\_Method:* El método de pago utilizado, codificado como “ E-Wallet” para billeteras digitales, “Card” para tarjetas, “PayPal” para PayPal y “Other” para otros métodos.

*Time\_Spent:* El tiempo que cada cliente pasó en el sitio web, medido en segundos.

*Browser:* El navegador web utilizado por el cliente, codificado como “Chrome” para Chrome, “Safari” para Safari, “Edge” para Edge y “Others” para otros navegadores.

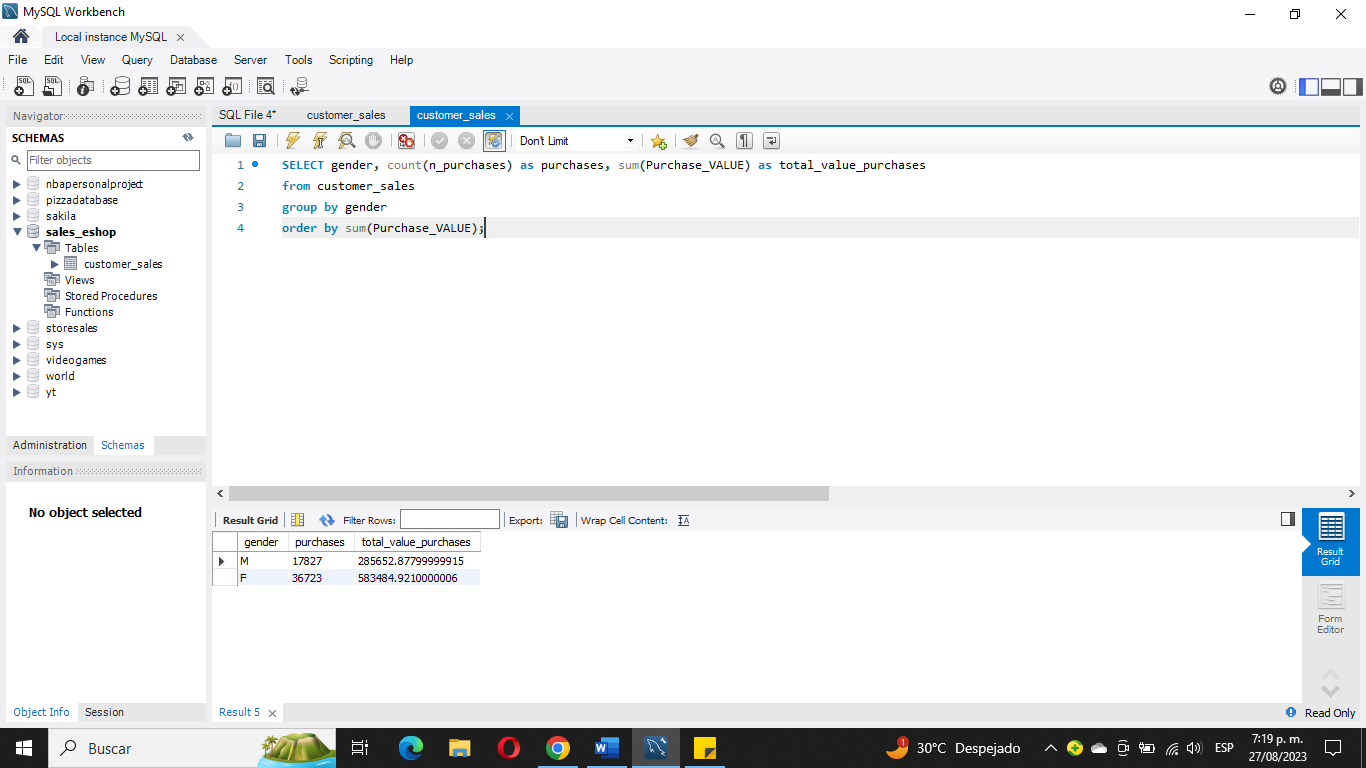
*Newsletter:* Si el cliente se ha suscrito al boletín informativo, codificado como “No” para no suscrito y “Yes” para suscrito.

*Voucher:* Si el cliente ha utilizado un cupón de descuento en su compra, codificado como “Not used” para no utilizado y “Used” para utilizado.

#### Patrones de compra de los clientes (según edad y género)

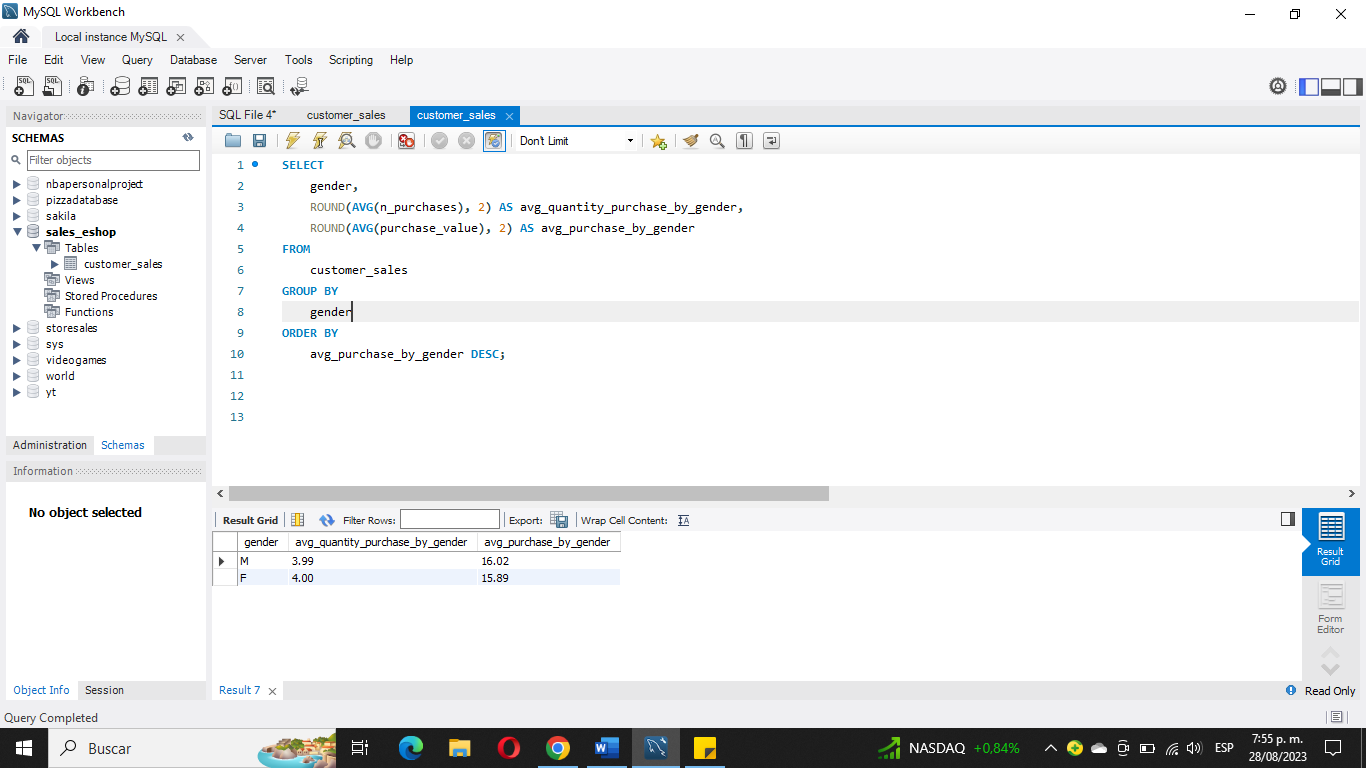
*Según género:*

El sitio web estudiado presenta una muy fuerte tendencia, más de dos tercios de quienes hacen sus compras son mujeres, y que estas, para sorpresa de nadie, son quienes más dinero gastan en esta tienda.



Con un total de 54.550 compras y 866.137,8 euros en compras, las mujeres ocupan dos tercios tanto de las compras como del total del dinero gastado en la tienda, se puede observar la fuerte tendencia de consumo del público femenino.

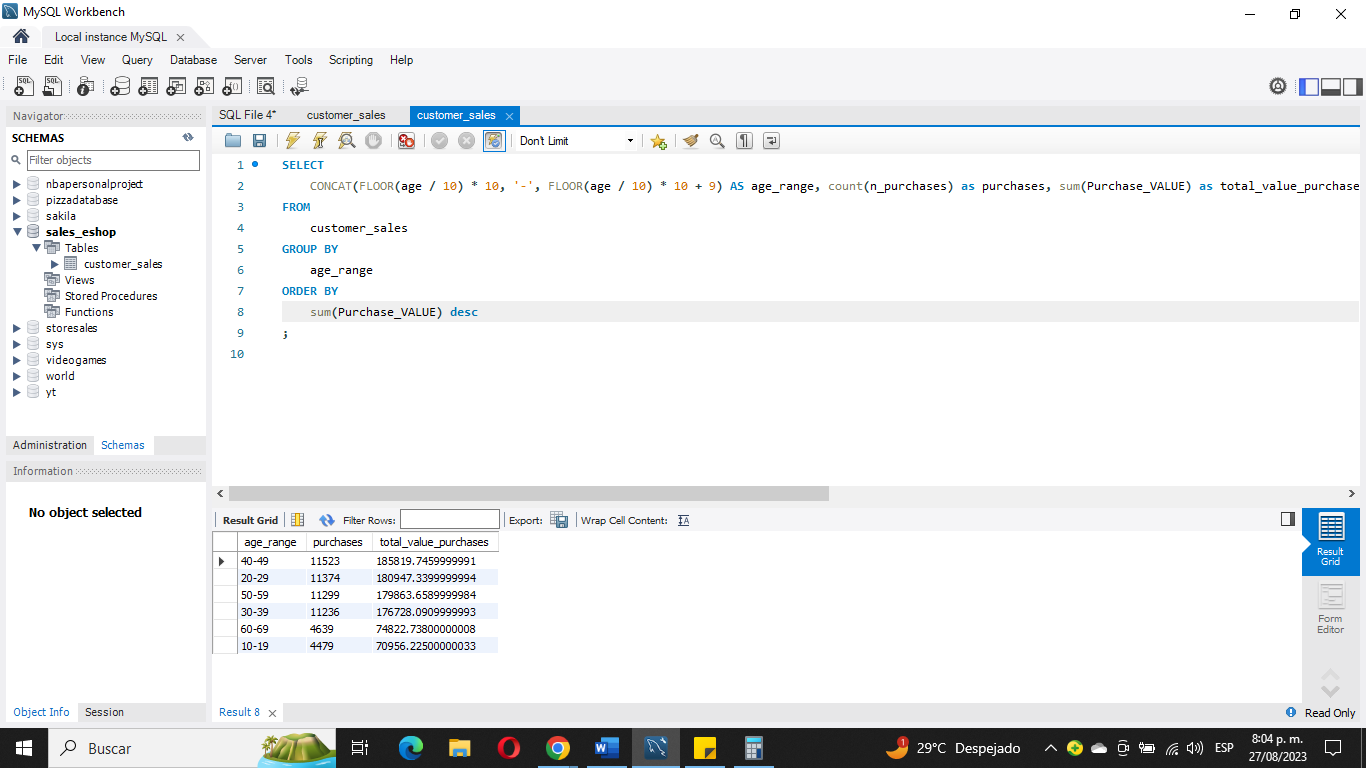
Otro patrón a notar es cuánto es el promedio que ha comprado y ha gastado cada género, así podemos ver cuál es la media de compra de acuerdo a su sexo, de acuerdo a la base de datos el resultado fue el siguiente:



A pesar de que las mujeres sean las mayores compradoras y las que más gastan dinero, la cantidad promedio de compra es la misma para ambos géneros y el gasto, sorprendentemente, es mayor para los varones; siendo 4 productos de promedio para ambos géneros, y 16,02 euros la media en una sola compra para los hombres y de 15,89 para las mujeres.

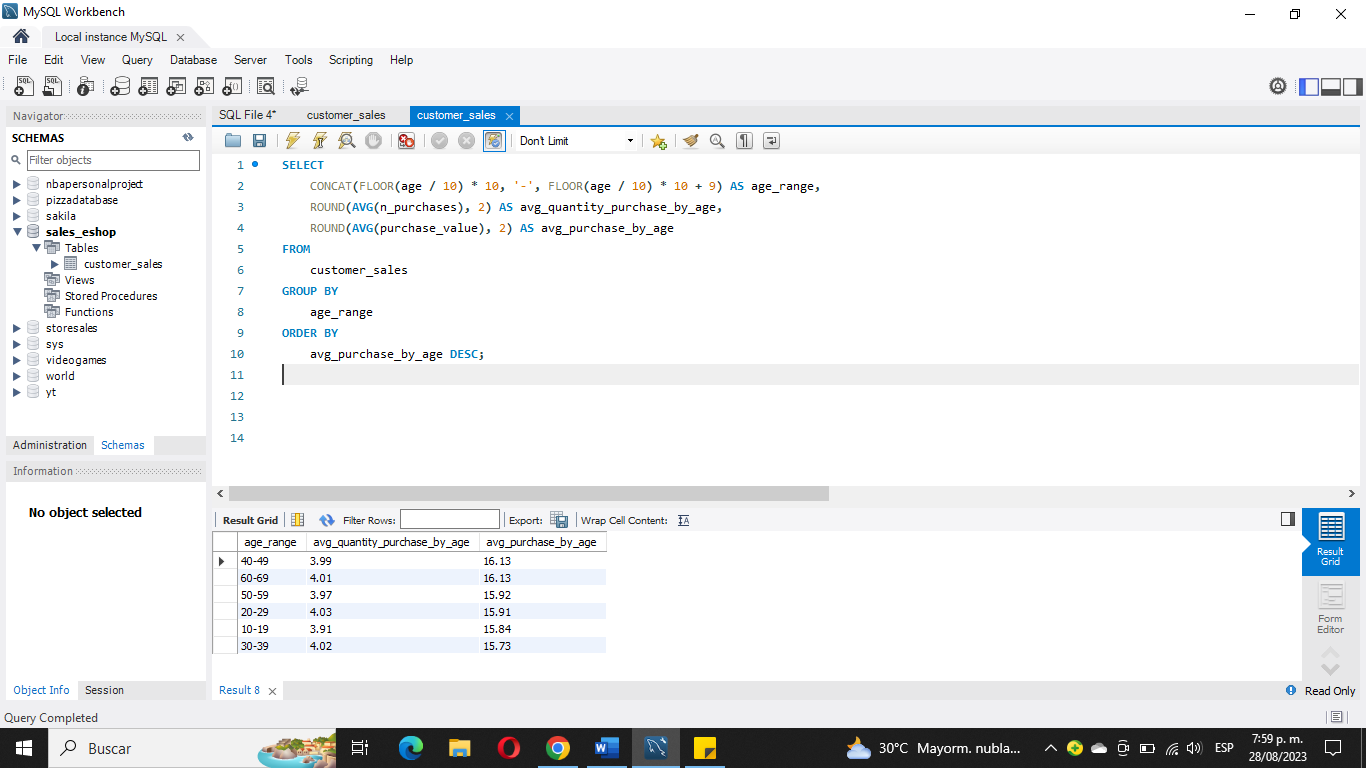
*Según edad:*

Los valores de las compras divido en rango de edades por década presenta una distribución normal, se puede ver que presenta números muy cercanos entre sí en los rangos presentadas, sin embargo, hay unos que sobresalen más que otros como se puede ver en la siguiente lista:



El rango de edades de los 40 son los que más volumen de compra ocupan, como se puede ver en la lista, sin embargo, no hay mucha diferencia con los siguientes 3 puestos de la lista, en el que hay un rango de diferencia de “tan solo” 10.000 euros. En cambio, la diferencia sí es demasiada comparada con los grupos de los 60 y de los 10, cabe resaltar que, en estos últimos grupos, los compradores son desde los 16 (edad legal para comprar en línea) hasta los 19, y desde los 60 hasta los 63 (edad máxima de la muestra); ambos con un rango de tan solo 4 años.

Un patrón más para tener en cuenta es el del promedio de compra y gasto en un solo pedido por rango de edad, y el resultado fue el siguiente:

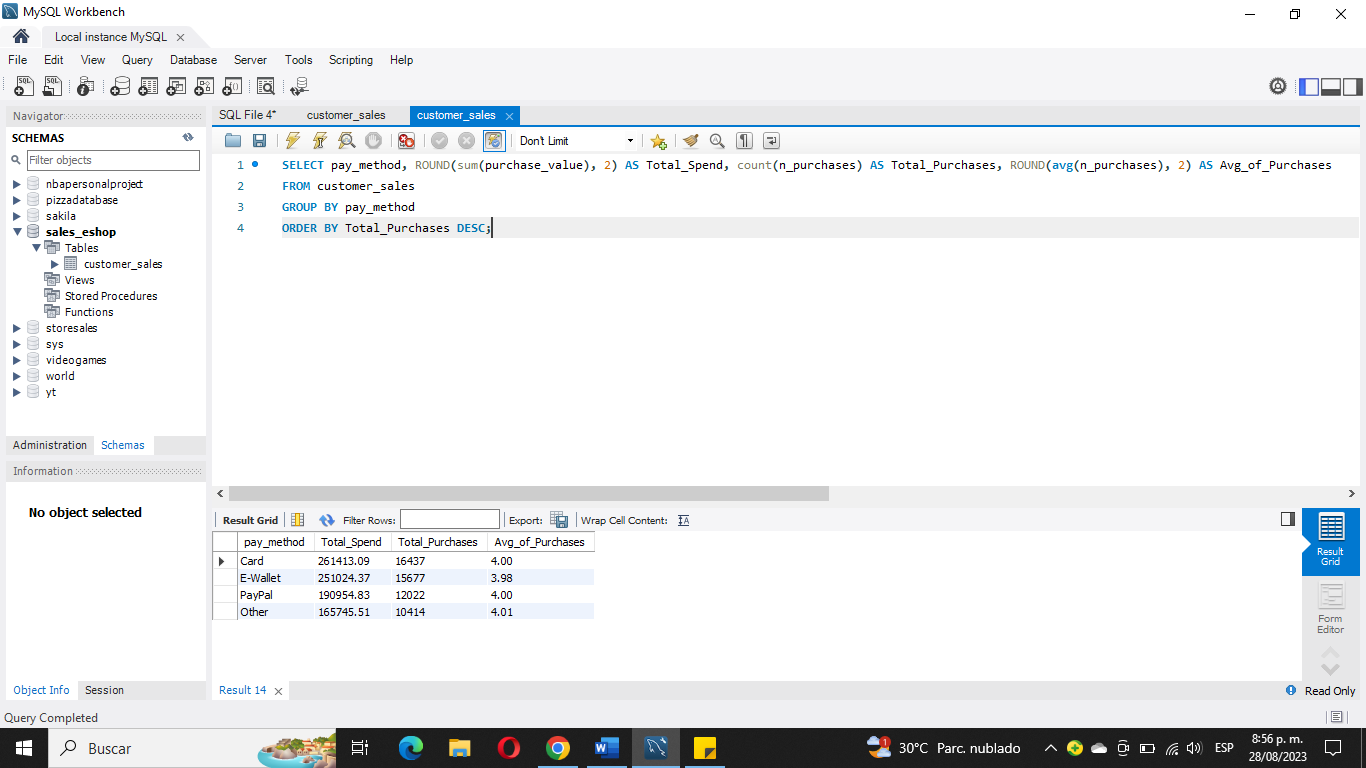


Como podemos observar en la tabla, hay unas observaciones muy curiosas, el promedio de artículos para casi todo el rango de edades está cerca de 4, aún así, el rango que más gasta es con uno de los menores promedio de artículos por compra, y es el de los 40-49. El que menos artículos promedia, a su vez, es uno de los que menos gasta y es el de los 10-19. Hay hechos como el que los del rango de 60- 69 compren mucho y gasten mucho y el que de los 30-39 compren mucho y gasten poco.

#### Métodos de pago (y su relación con las compras)

En este apartado veremos los datos que nos otorgan sobre los métodos de pagos y analizaremos si tienen alguna relación con el volumen de compra (sea en cantidad o valor).

De acuerdo a la base de datos, los resultados son los siguientes:



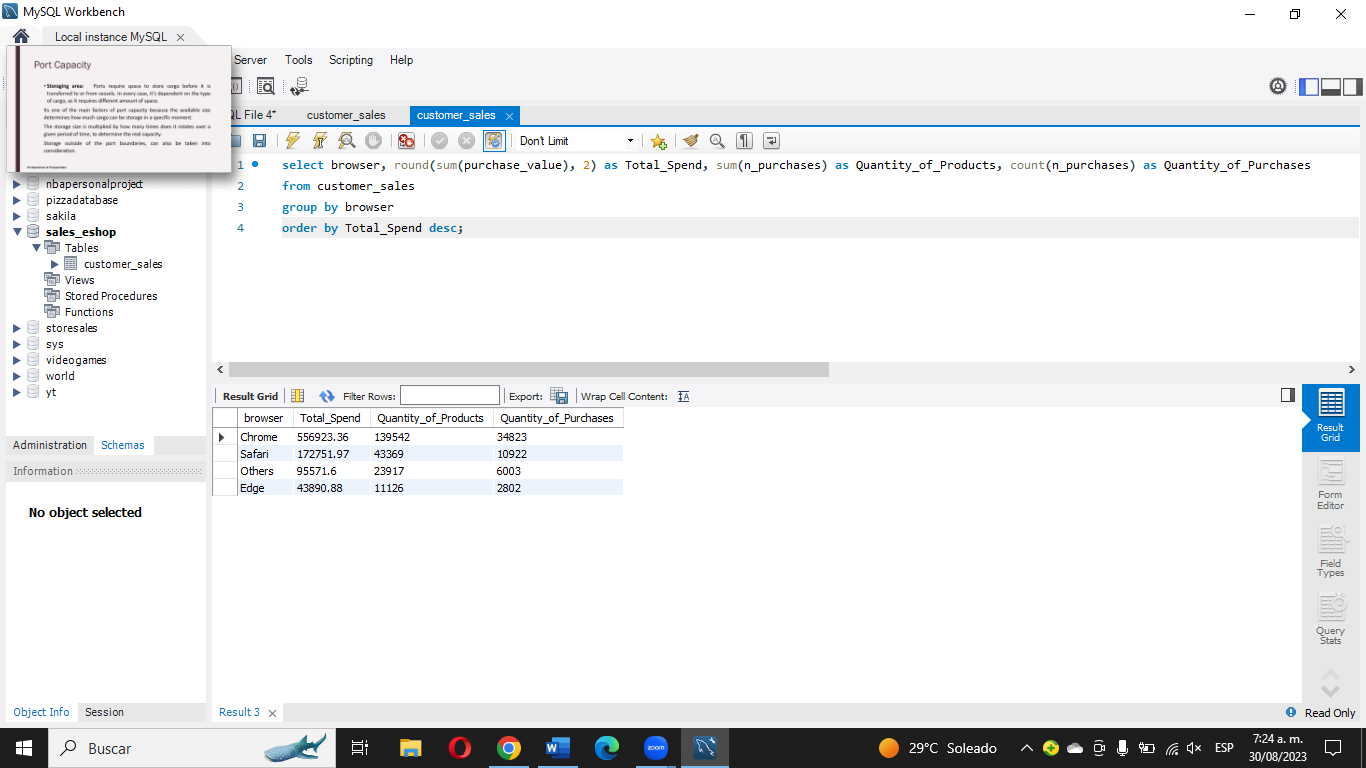
El método de pago más utilizado (Total\_Purchases) es la tarjeta (sea de crédito o débito), el segundo puesto es para las E-Wallets o billeteras electrónicas, el tercero es la plataforma PayPal y de último, están los otros métodos.

Algo a resaltar es que el volumen de compra (Avg\_of\_Purchases) es igual, en cuanto a cantidad de productos en un solo pedido, en todas las tarjetas, con un promedio de 4 productos. Sin embargo, en cuanto a cantidad de dinero gastado (Total\_Spend) por medio de estas opciones de pago, sí es posible notar una gran diferencia entre las dos primeras con el resto, se tiende a gastar más por medio de tarjetas e E-Wallets, que, de acuerdo a los resultados, más de 250.000 euros en cada uno de estos métodos, una cantidad enorme.

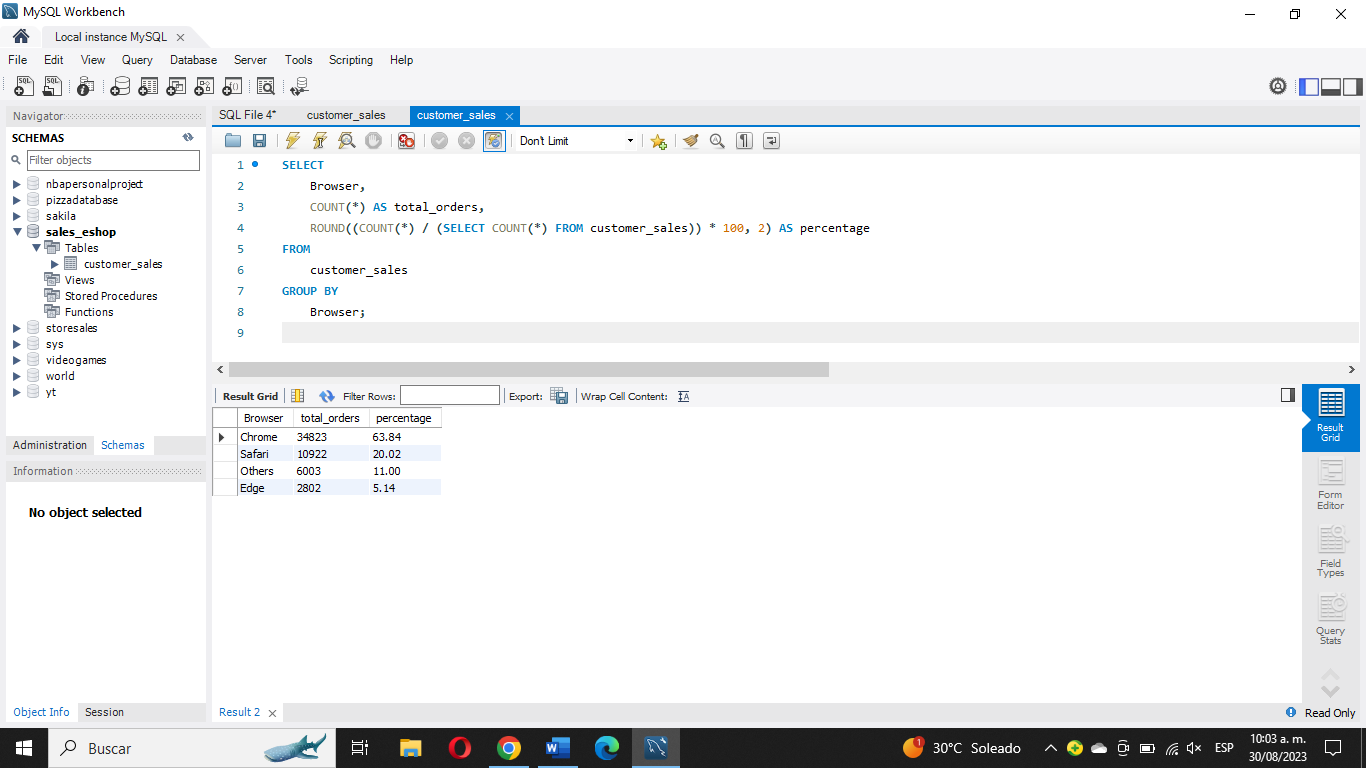
#### Navegadores (y sus patrones con el comportamiento de los clientes)

A continuación, haremos el análisis de cuánto influye el navegador web a la hora de comprar uno o muchos productos en la e-shop, veremos cuáles son los más populares, el volumen de compra por navegador y otras estadísticas para comprobar si el uso de uno u otro afecta la decisión del cliente en el momento de hacer un pedido.

Para empezar, tenemos las siguientes estadísticas:



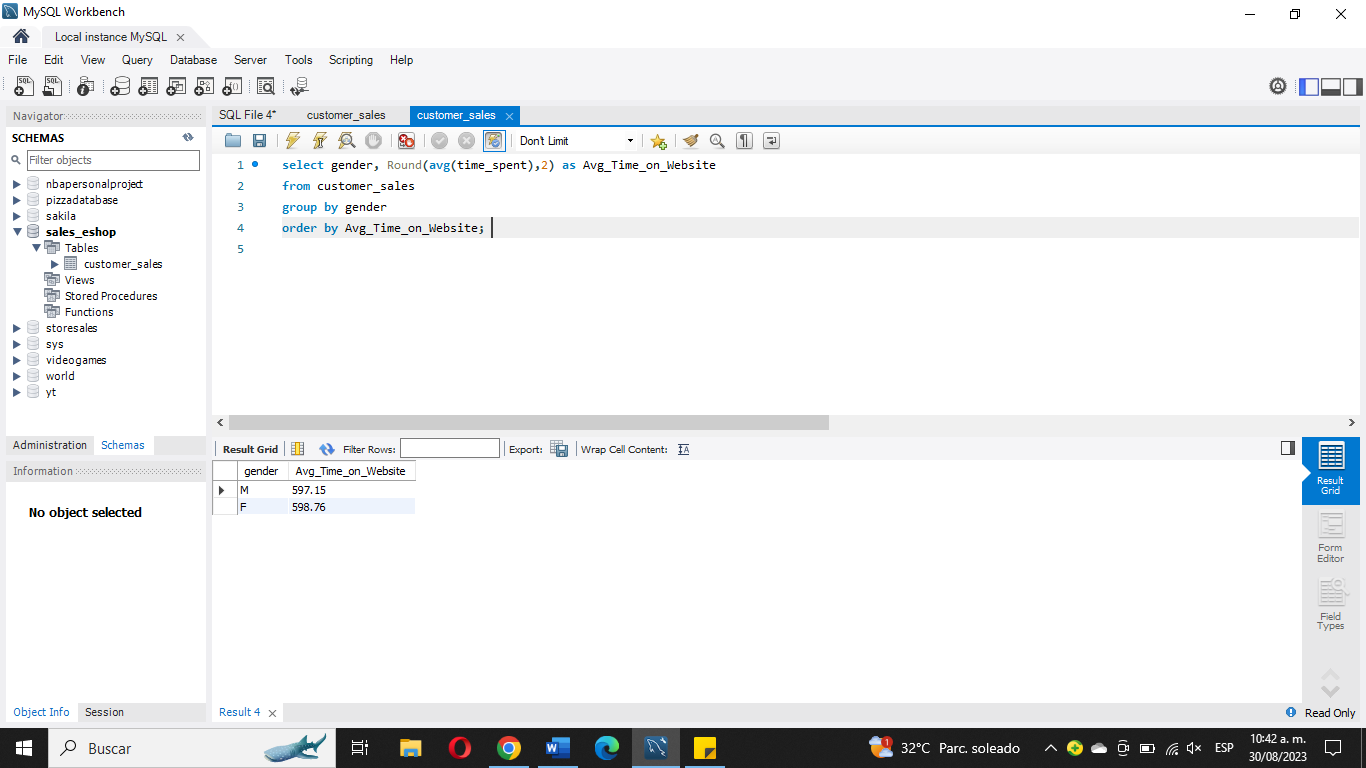
La mayor cantidad de volumen de compra se hace a través del navegador de Google Chrome, registrando más de 550.000 euros en compras, casi 140.000 productos vendidos y más de 34.000 órdenes de compra; siendo, por mucho, el navegador más utilizado por los compradores del sitio web. Safari, de Apple, se lleva el segundo lugar, pero con estadísticas mucho menores que Chrome, aún así, la gente presenta una fuerte tendencia por comprar por este medio (se puede inferir que son usuarios que compran a través de sus dispositivos Apple, sea iPhone, iPads o MacBooks). Sorpresivamente, el menos utilizado es Edge de Microsoft, incluso por debajo de los otros navegadores, registrando solo el 5% de la cantidad de dinero gastado por los clientes.



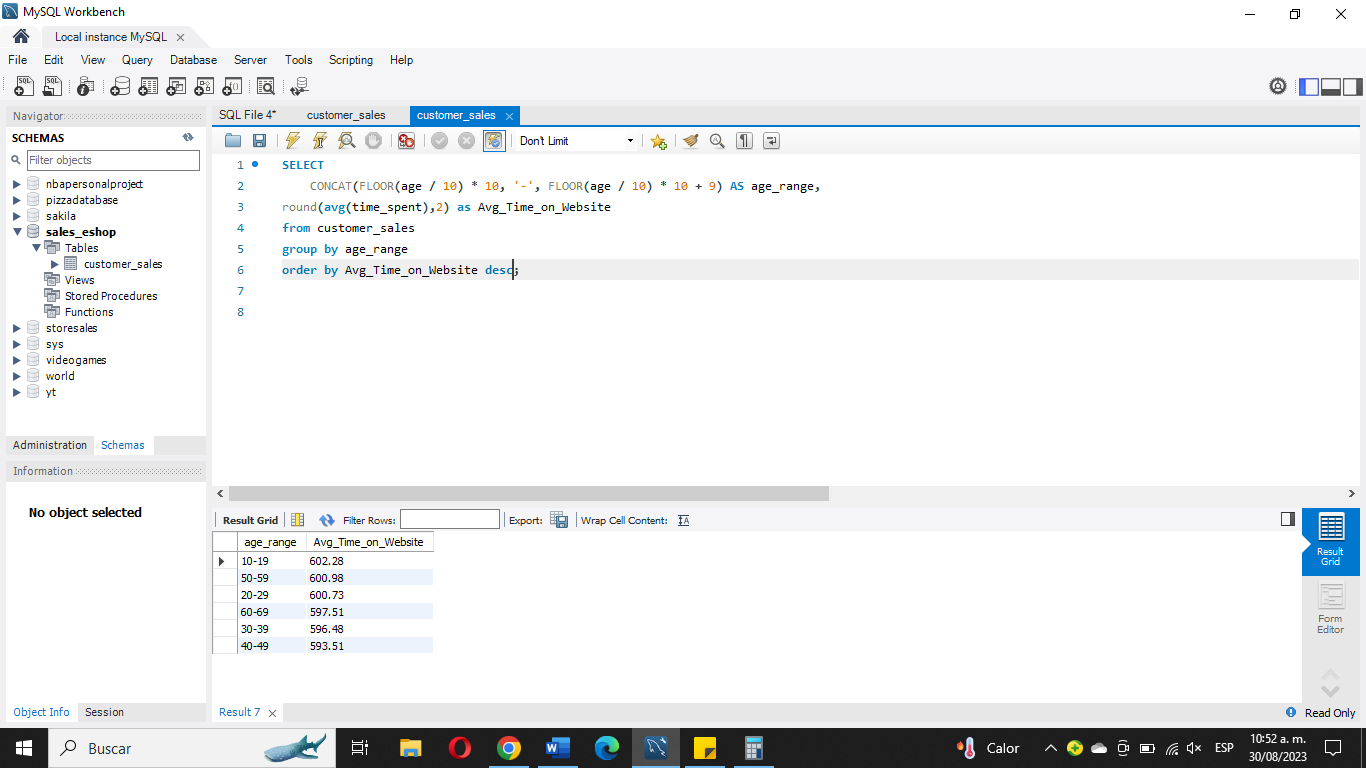
En este apartado también podemos observar el porcentaje del total de órdenes de compra por navegador, en donde Chrome representa la mayor cantidad de volumen con casi dos tercios del total.

#### Tiempo en el sitio web y relación con las compras

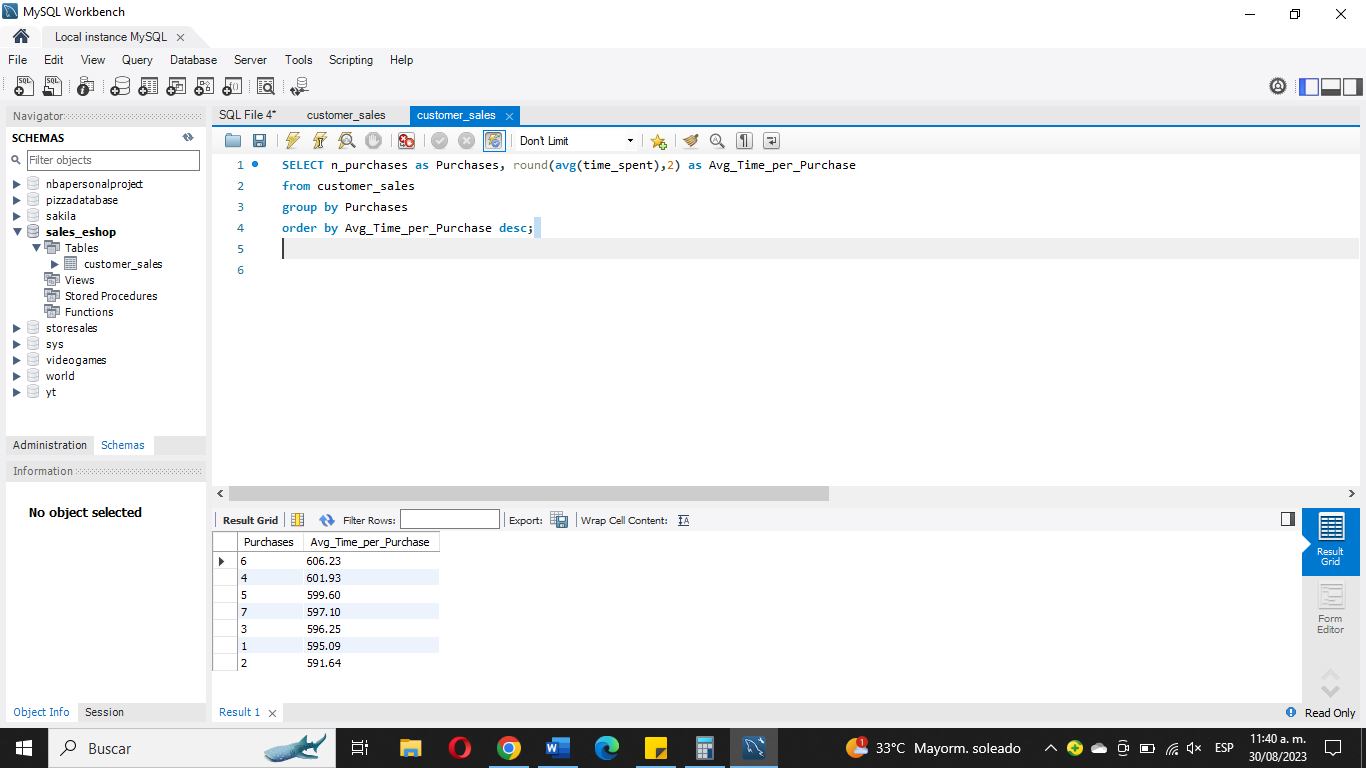
Otro apartado a analizar en este informe, es la relación entre el tiempo que los clientes pasan en el sitio web a la hora de realizar una compra, en la base de datos el tiempo se mide por segundos (segs), por lo que mantendremos la misma medida a la hora de examinar.



El promedio de tiempo en el sitio es similar para hombres como para mujeres, 598 en promedio (casi 10 minutos), un tiempo considerable.

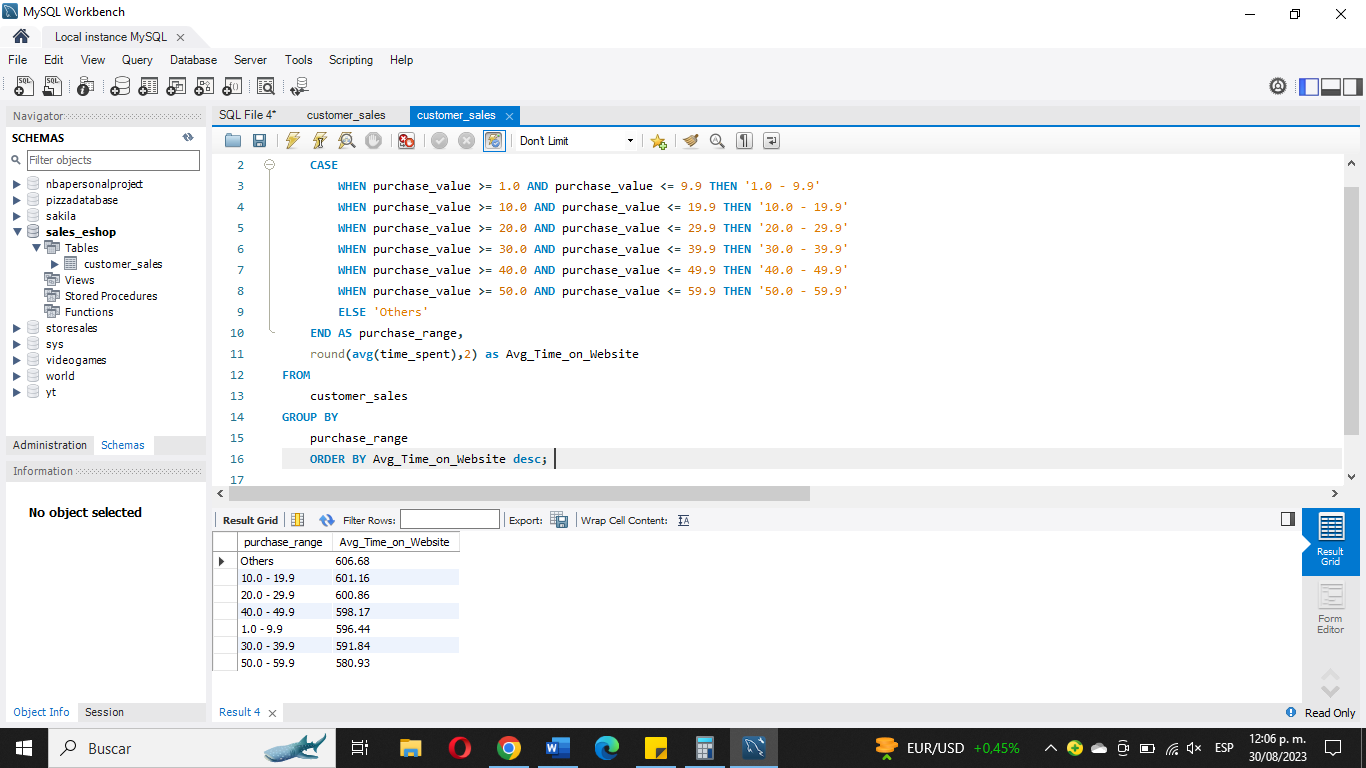


En cuanto a los rangos de edad la diferencia no es mucha entre las edades, la diferencia entre el primer y último puesto son 11 segundos, sin embargo, los de 10-19 son quienes más tiempo duran en el sitio web, 602.28 segundos (10 minutos), mientras que el último lugar, el rango de los 40-49, dura un poco menos (9 minutos, casi 10), por lo cual, en este apartado no hay mucho más que agregar.



Como era de esperarse, mientras menos productos tenga el pedido, menos tiempo tarda el cliente en comprar, sin embargo, al comprar 6 productos (que no es el máximo de pedidos por cliente) es cuando el comprador se tarda más en el sitio web, superando por casi 10 segundos que cuando compran 7 productos. Curiosamente cuando menos se tardan menos cuando piden solo 2 productos, y no cuando piden 1.

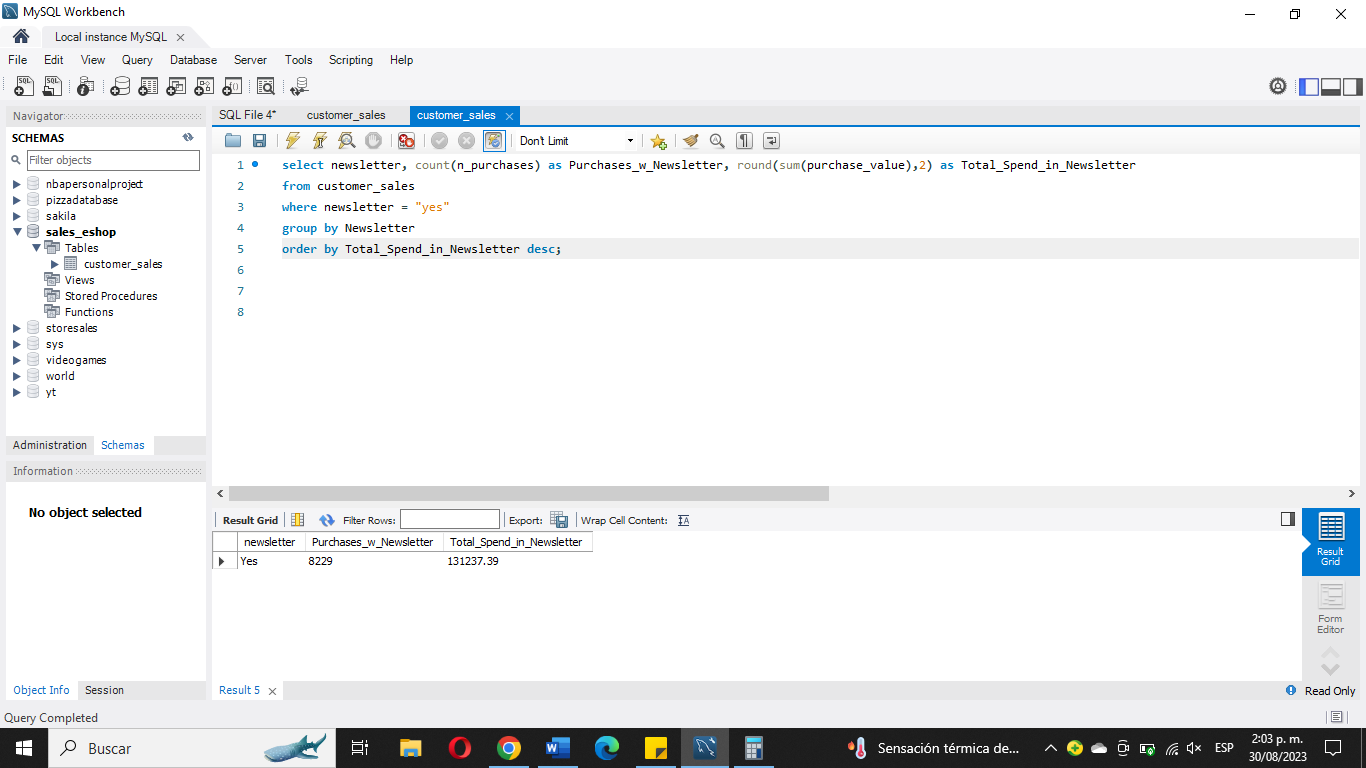
El último punto a analizar en este apartado es si existe una relación entre el precio de la compra y el tiempo invertido en el sitio web y el resultado fue el siguiente:



“Others” es un rango menor a 1 euro, y dentro de estas cantidades, casualmente, es cuando la gente más se tarda comprando en el sitio web; y para las compras más costosas (el rango de 50- 59), los clientes tardan mucho menos tiempo en la e-shop, con una diferencia de 26 segundos, la mayor presentada en rangos de tiempo, sin duda, unos patrones muy curiosos.

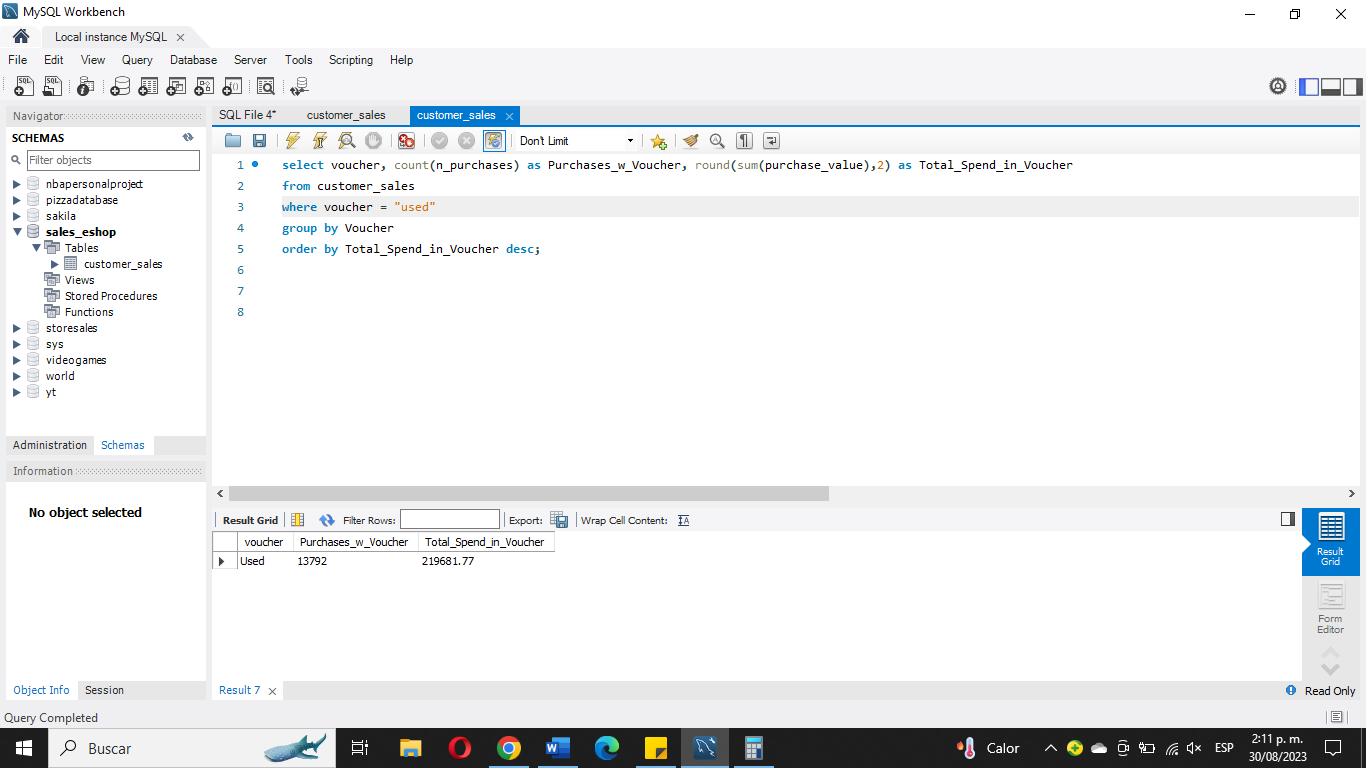
#### Uso de boletín y cupones (y su relación con las compras)

La página de web estudiada en este caso cuenta con un boletín informativo que llega al correo electrónico de los clientes y estos pueden acceder desde el enlace para acceder a nuevos productos y descuentos, también cuenta con cupones de descuento que son enviado regularmente a sus clientes para promover el consumo.



La cantidad de clientes que hicieron compra por medio del boletín fue de 8.229, un 15,08% del total de clientes de la base de datos, una cantidad que no se debe ignorar.

Por otro lado, la cantidad de clientes que compraron por medio de cupones fue la siguiente:



Casi 14.000 pedidos utilizando cupones, un cuarto del total de las órdenes de compra, sin duda el alcance de los cupones es mayor a la hora de llegar a los clientes.

En total, 4 de cada 10 compras se hacen por medios promocionales (newsletter o cupones de descuento), lo que se traduce a que los esfuerzos de la empresa por llegar y promover el consumo de sus clientes son muy eficientes y se ven reflejado en las estadísticas, y más importante aún, las ventas.

***Propuestas de mejora:***

Luego de analizar las distintas estadísticas y examinar cada una de estas en búsquedas de patrones y cómo los números reflejan el funcionamiento de la empresa en cuanto a ventas, me tomaré la libertad de realizar las siguientes propuestas para mejoría de la eficiencia y de cómo explotar las notables ventajas revisadas a lo largo de este informe.

*Según datos demográficos:*

Luego de profundizar un poco más sobre los datos que se arrojan sobre la información esencial de los clientes (edad y género), hay dos cosas que me llamaron la atención y fue que los hombres compran poco en el sitio web, pero no una cantidad tan ínfima para deducir que la página solo vende productos para mujeres; de hecho, los hombres son casi un tercio de los compradores. Es por ello que mi propuesta es la siguiente:

* Dirigir más anuncios dirigidos al público masculino, para promover el consumo por parte de este grupo. Los anuncios pueden ofrecerles ofertas, descuentos o promociones (tipo compre 1, pague 2).
* Ofrecerles productos más afines a sus gustos para llamar su atención e influenciarlos a comprar.

Ahora yendo un poco más en profundidad en el tema de la edad, todos los grupos de los rangos de edad se comportan de manera similar, todos excepto dos, los rangos de los 10-16 y el de los 60-69; esto ocurre por dos factores:

1. La cantidad de datos de estos rangos es reducida comparado a los demás.
2. Los factores externos que influyen sobre ellos a la hora de comprar por el sitio web. Los jóvenes no cuentan con tanto poder adquisitivo (así lo reflejan los datos) y, se puede deducir que los mayores no están muy familiarizados con este tipo de compras.

Por lo que se pueden llevar a cabo las siguientes propuestas:

* Similar al grupo de los hombres, ofrecer promociones llamativas a los más jóvenes, ya que estos cuentan con menor dinero para gastar, por lo que descuentos y cosas así podría motivarlos a comprar más en el momento de hacer un pedido.

*Según navegador:*

El navegador más popular de los clientes para comprar es Google Chrome, sin embargo, los clientes (sobre todo los que cuentan con dispositivos Apple) tienen mucho apego a Safari, teniendo en cuenta lo anterior, realizo las siguientes ideas:

* Invertir más recursos en Google Ads, para que con los anuncios se pueda redirigir el tráfico de los clientes a la e-shop, así conozcan los productos, y en el mejor de los casos, generen nuevas órdenes de compra.
* Con las herramientas de Google Analytics, analizar con ayuda de las métricas, el desempeño que ha tenido la página con la inversión en estos anuncios, para conocer su eficiencia y si puede ser útil para generar compras futuras.

*Según boletín y cupones:*

Como se pudo ver en este apartado, las compras por medio del Newsletter o los Voucher, son el 25% del total del volumen de compra, a pesar de esto, se necesitan más estrategias para mejorar este apartado, por lo que mis propuestas son las siguientes:

* Fomentar más a la suscripción de los boletines, esto es porque el Newsletter, a comparación de los cupones, no es tan atractivo a la hora de comprar, por lo que se deben hacer mejores trabajos de e-mail marketing para atraer a los clientes a suscribirse otorgándoles ventajas como descuento o acceso a productos en exclusiva antes que el resto, estas estrategias serían clave para añadir más suscriptores.
* Aprovechar el éxito de los cupones, los Voucher son muy atractivos para el público en general, por lo que soltar algunos de vez en cuando, puede ayudar al consumo dentro del sitio, en especial en épocas con tráfico bajo.

***Conclusión:***

Los datos son ventanas a la mente del consumidor digital. A través de este proyecto, hemos explorado transacciones de una tienda en línea en 2021 y desentrañado patrones y decisiones ocultas. Desde género y métodos de pago hasta navegadores y suscripciones, cada dato cuenta una historia.

Este análisis va más allá de los números: nos sumerge en el viaje de compra moderno. Cada dato es un eslabón en la cadena de decisiones, una gota de motivación y estrategia. Este proyecto es un recordatorio de que las elecciones digitales pintan una imagen viva de la experiencia del cliente.

En resumen, los datos no son solo puntos, son el hilo conductor de la narrativa humana en el mundo en línea. En 2021, hemos comenzado a descifrar estas historias, pero el futuro ofrece aún más revelaciones en cada clic.