

DATA SCIENCE – PROYECTO FINAL

DATASET – HÁBITOS DE COMPRAS



LUSVIL BARRETO

INTRODUCCIÓN

Para el mundo del Retail(Tiendas por departamento) es todo un reto mantener la lealtad de sus consumidores ya que en la actualidad existe mucha competencia en el rubro, por ello es muy importante el análisis sobre comportamiento del consumidor y hábitos de compra por que esto proporciona información completa sobre las preferencias, tendencias y patrones de los consumidores durante sus experiencias de compra, lo que permitirá tomar decisiones acerca de los consumidores, ayudando a las empresas a diseñar estrategias de marketing específicas, optimizar las ofertas de productos y mejorar la satisfacción general del cliente.

DATASET

El Data set a analizar posee datos de compras de clientes en diferentes sucursales alrededor de varias ciudades de una misma tienda de retail, a continuación, se presenta un diccionario de datos que describirá a cabalidad cada columna del DataSet.

Nombre	Descripción
Customer ID	Un identificador único asignado a cada cliente
Age	Edad del cliente
Gender	Género del cliente
Item Purchased	Nombre del producto
Category	Categoría del producto
Purchase Amount (USD)	Valor del producto
Location	Ubicación geográfica de la compra
Size	Tamaño (si corresponde) del artículo comprado
Color	Color asociada con el artículo comprado
Season	Estacional del artículo comprado
Review Rating	Evaluación numérica
Subscription Status	Indica si el usuario esta suscrito
Shipping Type	Método utilizado para entregar el artículo
Discount Applied	Indica si se aplicó algún descuento
Promo Code Used	Indica si se utilizó un código promocional
Previous Purchases	Proporciona información sobre el número o frecuencia
Payment Method	Especifica el modo de pago
Frequency of Purchases	Indica la frecuencia con la cual el cliente compra

HIPÓTESIS

- ¿Cuál es la sucursal con mayores ventas y cual sucursal la más baja?

Hipótesis: Se espera que las ciudades con mayor número de clientes sean los que tengan las mayores ventas.

- ¿Cuál es la distribución de ventas entre las mujeres y los hombres?

Hipótesis: Se espera que las mujeres sean a las que más se les vende.

- ¿Cuál es el método de pago preferido por los clientes?

Hipótesis: Se espera que las personas en primavera y otoño compren más productos debido al cambio de temporada.

- ¿Hay temporadas en las que las ventas son mejores?

Hipótesis: Se espera que en cada temporada se vendan más artículos relacionadas a las mismas

- ¿Existe relación entre las categorías de productos y ciertos grupos de edad o géneros?

Hipótesis: Se espera que las personas más jóvenes tiendan a comprar más artículos relacionados a la categoría Accessories, ya que se espera que son los que pueden estar dispuestos a gastar de más.

- ¿Existe relación entre las ventas y edad?

Hipótesis: Se espera que las personas de edad promedio sean los clientes los que más compren debido al nivel adquisitivo.

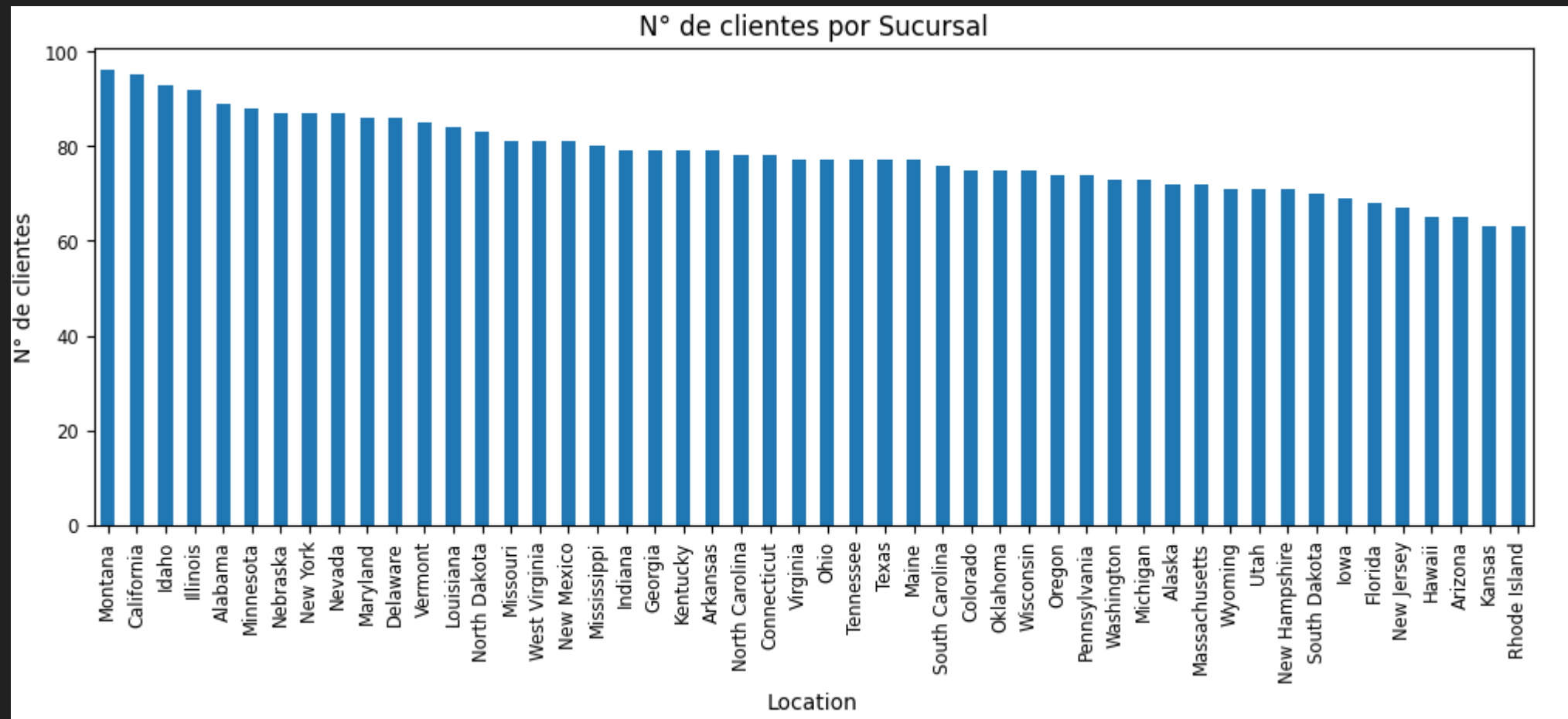
VISUALIZACIONES

A continuación se muestran el listado de las ventas por sucursales , donde podemos observar:

- Montana se mantiene en el top siendo el que mas dinero recaudo, esto concuerda con el numero de ventas que fue menor , lo mismo pasa para las que tienen ventas menores.
- Hay que considerar que las que tuvieron mejores promedios en los montos recaudados por las ventas fueron Arizona y Pennsylvania

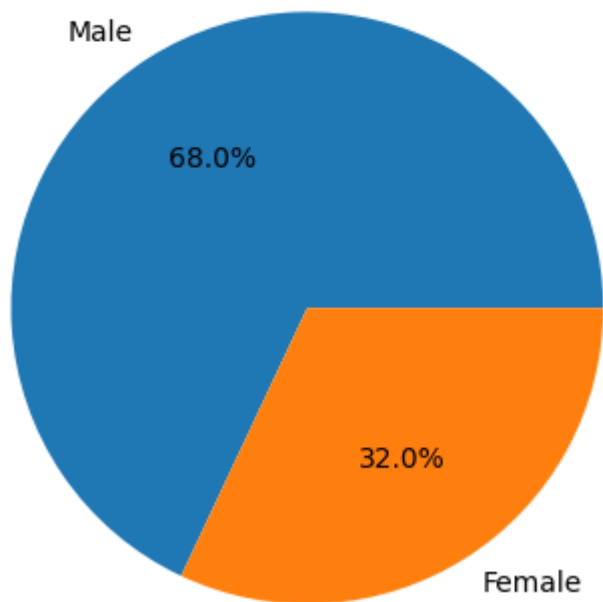
```
Location
Montana 96
California 95
Idaho 93
Illinois 92
Alabama 89
Minnesota 88
Nebraska 87
New York 87
Nevada 87
Maryland 86
Delaware 86
Vermont 85
Louisiana 84
North Dakota 83
Missouri 81
West Virginia 81
New Mexico 81
Mississippi 80
Indiana 79
Georgia 79
Kentucky 79
Arkansas 79
North Carolina 78
Connecticut 78
Virginia 77
Ohio 77
Tennessee 77
Texas 77
Maine 77
South Carolina 76
Colorado 75
Oklahoma 75
Wisconsin 75
Oregon 74
Pennsylvania 74
Washington 73
Michigan 73
Alaska 72
Massachusetts 72
Wyoming 71
Utah 71
New Hampshire 71
South Dakota 70
Iowa 69
Florida 68
New Jersey 67
Hawaii 65
Arizona 65
Kansas 63
Rhode Island 63
Name: count, dtype: int64
```

```
Regional Purchase Amount Stats:
                                mean  median  sum
Location
Montana 60.250000 64.0 5784
Illinois 61.054348 65.0 5617
California 59.000000 57.0 5605
Idaho 60.075269 62.0 5587
Nevada 63.379310 66.0 5514
Alabama 59.112360 56.0 5261
New York 60.425287 62.0 5257
North Dakota 62.891566 64.0 5220
West Virginia 63.876543 66.0 5174
Nebraska 59.448276 58.0 5172
New Mexico 61.901235 63.0 5014
Minnesota 56.556818 53.0 4977
Pennsylvania 66.567568 70.0 4926
Mississippi 61.037500 64.0 4883
Alaska 67.597222 68.5 4867
Vermont 57.176471 54.0 4860
Louisiana 57.714286 55.5 4848
Virginia 62.883117 62.0 4842
Arkansas 61.113924 58.0 4828
Maryland 55.755814 52.0 4795
Tennessee 61.974026 61.0 4772
Delaware 55.325581 52.5 4758
North Carolina 60.794872 62.5 4742
Texas 61.194805 62.0 4712
Missouri 57.913580 61.0 4691
Indiana 58.924051 60.0 4655
Ohio 60.376623 66.0 4649
Georgia 58.797468 62.0 4645
Washington 63.328767 63.0 4623
Michigan 62.095890 63.0 4533
Utah 62.577465 65.0 4443
South Carolina 58.407895 59.0 4439
Kentucky 55.721519 53.0 4402
Maine 56.987013 57.0 4388
Massachusetts 60.888889 64.0 4384
Oklahoma 58.346667 60.0 4376
Arizona 66.553846 68.0 4326
Wyoming 60.690141 60.0 4309
Oregon 57.337838 54.0 4243
South Dakota 60.514286 62.0 4236
Connecticut 54.179487 48.5 4226
Colorado 56.293333 51.0 4222
New Hampshire 59.422535 59.0 4219
Iowa 60.884058 60.0 4201
Wisconsin 55.946667 55.0 4196
Rhode Island 61.444444 63.0 3871
New Jersey 56.746269 51.0 3802
Florida 55.852941 56.0 3798
Hawaii 57.723077 55.0 3752
```



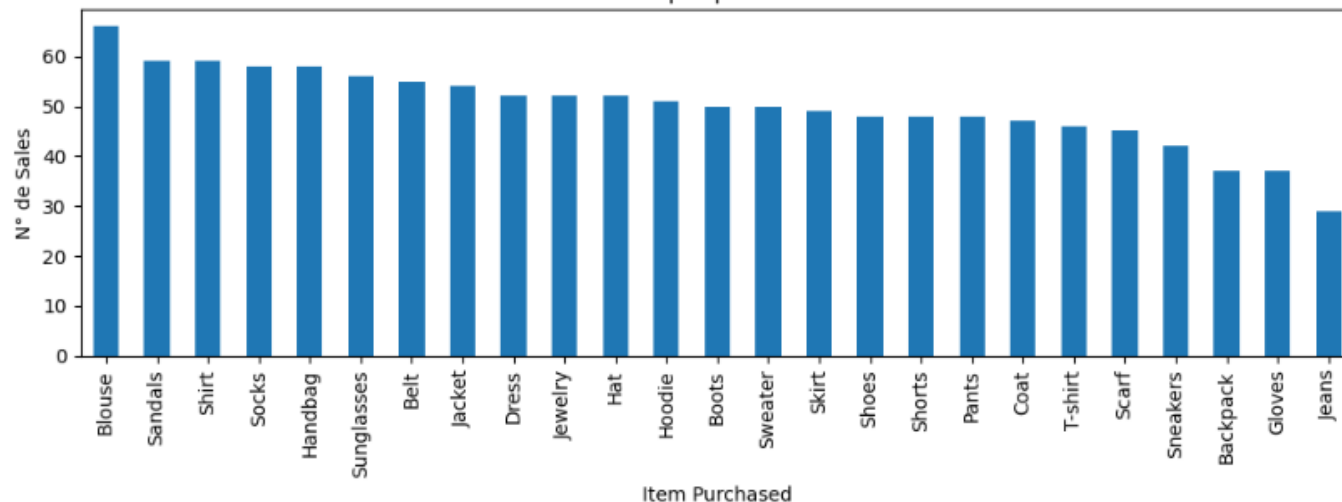
- Montana tiene la mayor cantidad de clientes.
- Rhode Island, podría ser foco importante para campañas de marketing que permitan a un nuevos compradores.

Porcentaje de distribución por género

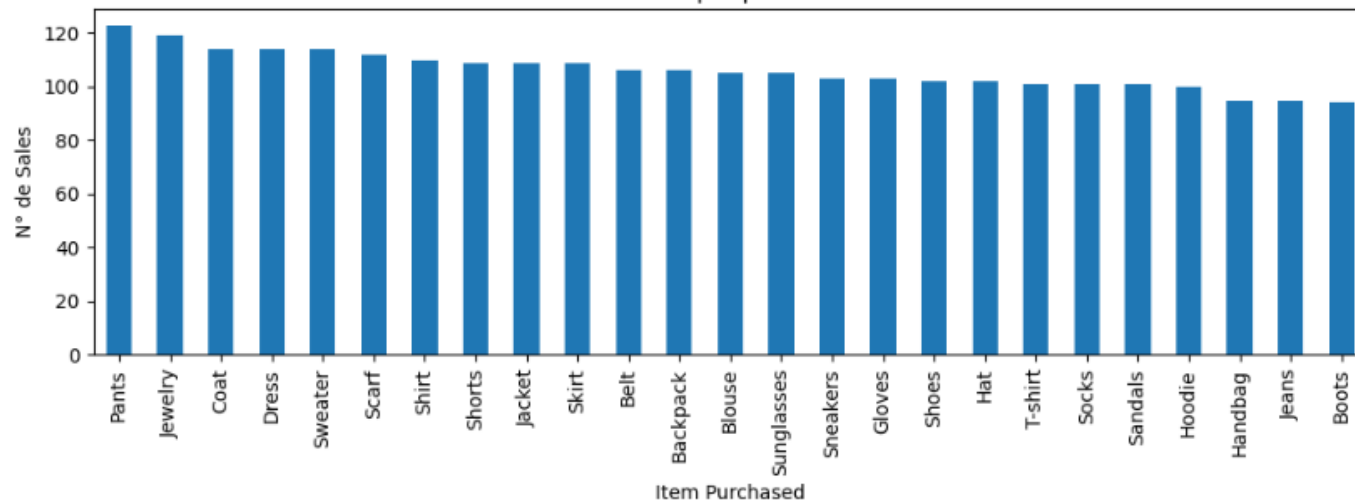


- Los hombres contribuyen a más de la mitad de las compras totales, representando el 68%, mientras que las mujeres representan sólo el 32% de las compras totales.

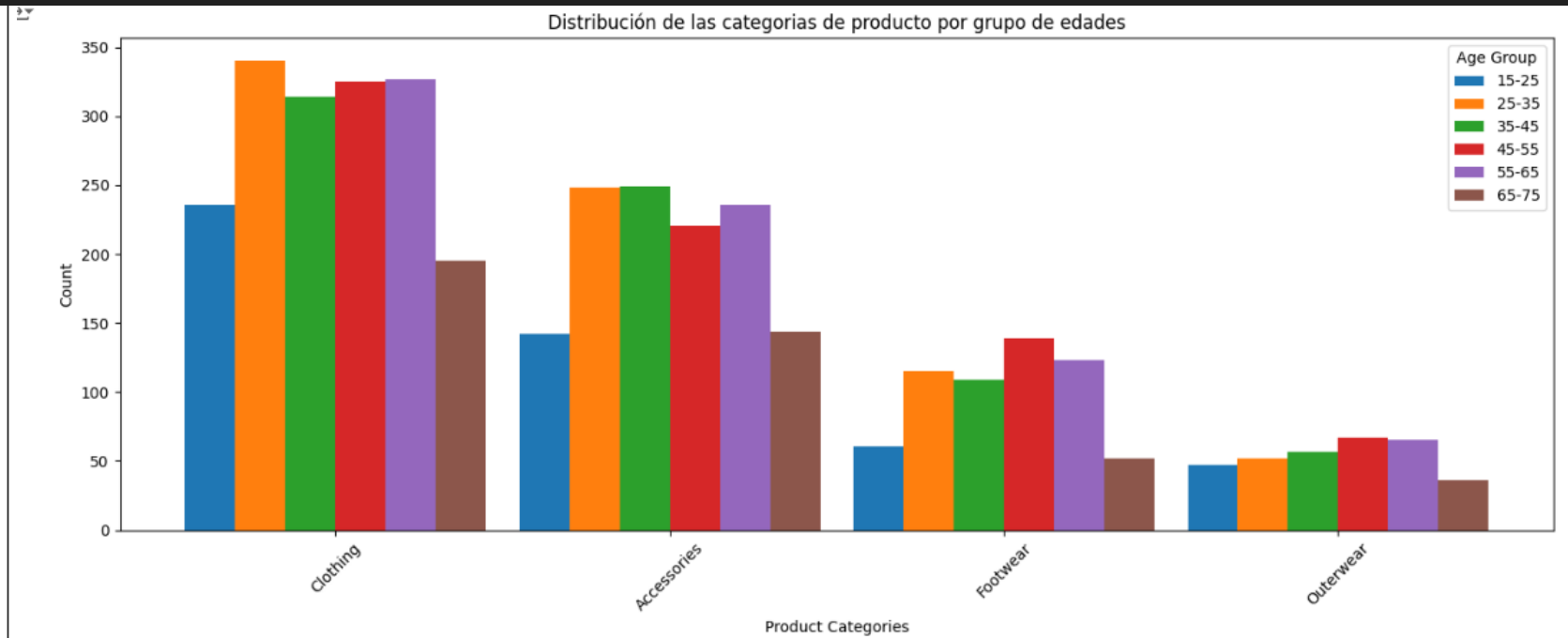
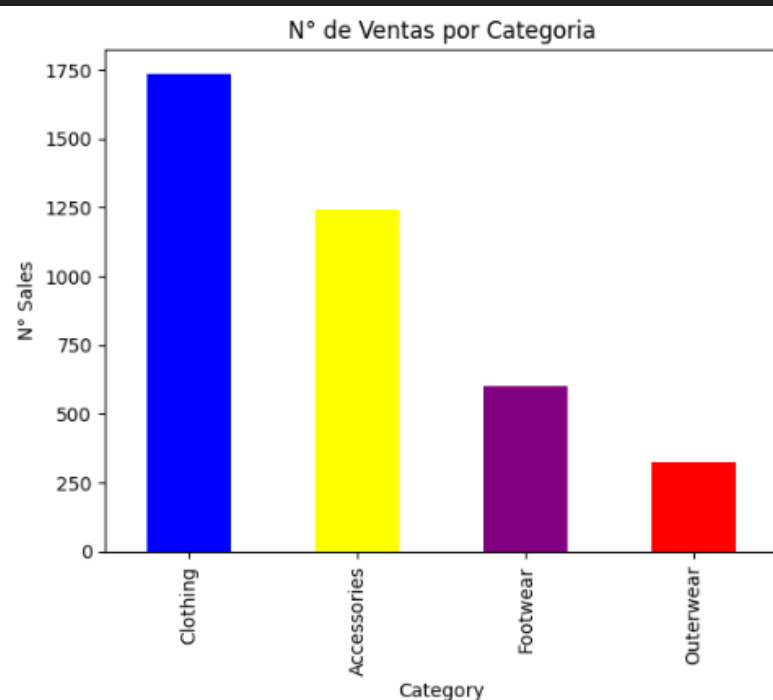
N° de Ventas por producto - Female



N° de Ventas por producto - Male

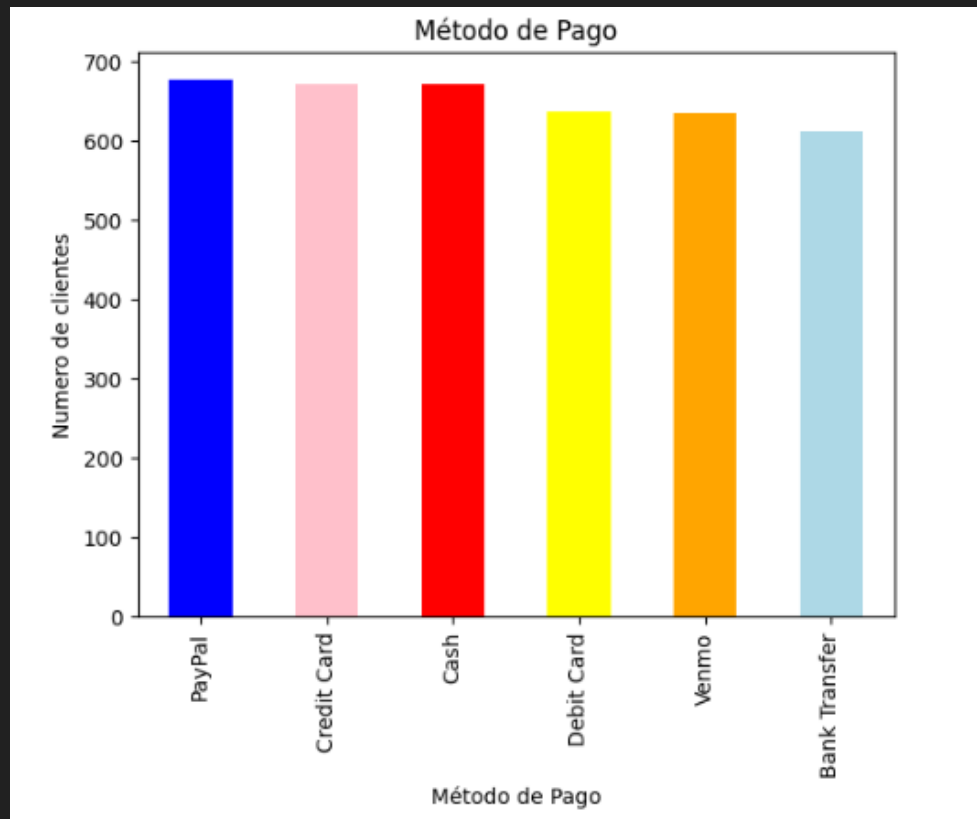


- El productos mas comprado por mujeres son las blusas(Blouse) y entre los hombres los pants

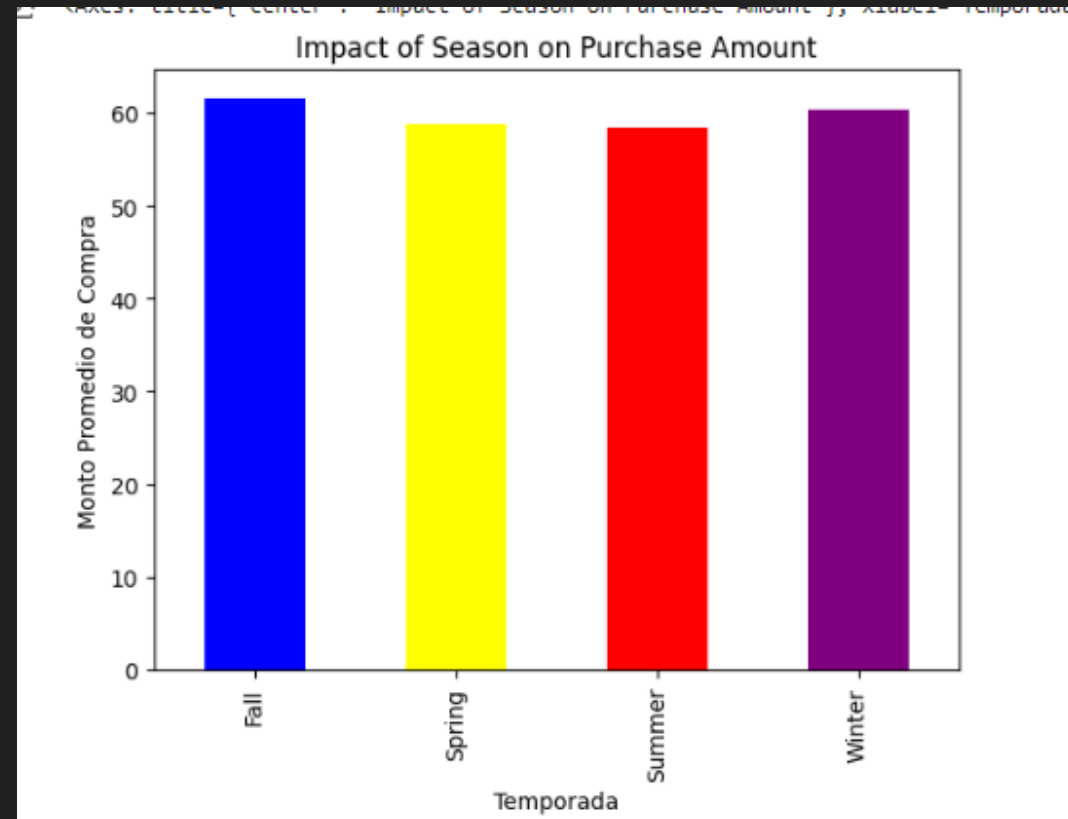


○ Clothing tiene el mejor historial de compras y la Outerwear tiene el más bajo.

○ Podemos observar , que la ropa(clothing) es la categoría mas popular entre todos los grupos de edades al igual que la ropa de abrigo(outerwear), los accesorios (Accessories) son muy populares en todos los grupos de edades excepto en los de 15-25 y 65-75 años. Sin embargo, hemos visto que en la categoría de calzado(Footwear) es más popular en el grupo de edad de 45 a 55 años.



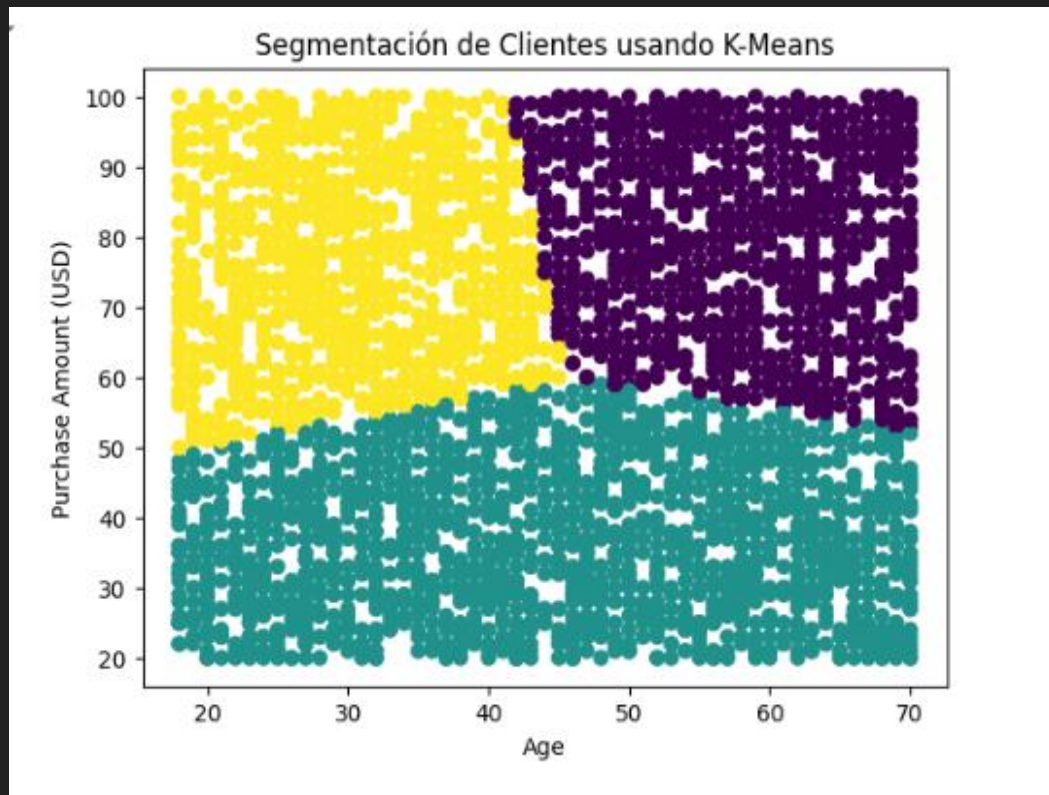
- Paypal , Credi Card y Cash son los métodos de pago mas elegidos



- Fall and Winter tienen el monto de compra promedio más alto de todas las estaciones.
- Spring and Summer no están muy por detrás de las dos temporadas más altas.

MACHINE LEARNIG

Ya que se necesitaba estudiar la relación entre las edades y las ventas se utilizo el algoritmo K-means, dando como resultado :



- K-medias nos muestra tres segmentos de clientes distintos, representados por diferentes códigos de color. El segmento 1 comprende clientes jóvenes con alto poder adquisitivo, mientras que el segmento 2 está compuesto por clientes mayores con montos de compra promedio más altos. Por el contrario, el segmento 3 abarca clientes con tendencias de gasto más bajas, que a menudo optan por artículos menos costosos.

CONCLUSIONES

- Montana tiene la mayor cantidad de clientes de todas las sucursales y Kansas y Rhode Island son las que poseen menos , lo que concuerda con la hipótesis establecida.
- Los hombres representan el 67% del total de compras y las mujeres representan el 32% del total, demostrando que los hombres en este caso fueron los que mas compraron .
- Paypal , Credi Card y Cash son los metodos de pago mas elegidos
- Los clientes realizan más compras durante el invierno(winter) y el invierno(Fall) en comparación con el verano(Summer) y la primavera(Spring).
- La ropa(clothing) es la categoría mas pupular entre todos los grupos de edades al igual que la ropa de abrigo(outerwear), los accesorios (Accessories) son muy populares en todos los grupos de edades excepto en los de 15-25 y 65-75 años. Sin embargo, hemos visto que en la categoría de calzado(Footwear) es más popular en el grupo de edad de 45 a 55 años.
- K-medias nos muestra tres segmentos de clientes distintos, representados por diferentes códigos de color. El segmento 1 comprende clientes jóvenes con alto poder adquisitivo, mientras que el segmento 2 está compuesto por clientes mayores con montos de compra promedio más altos. Por el contrario, el segmento 3 abarca clientes con tendencias de gasto más bajas, que a menudo optan por artículos menos costosos.