**Análisis De Datos**

**Tarea 3 - Tarea 3 - Algoritmos de Aprendizaje Supervisado**

**Nombre de los Estudiantes:**

**Jonathan Andres Yara**

**Nro. de grupo:**

**202016908\_45**

**Universidad Nacional Abierta y a Distancia**

**Vicerrectoría Académica y de Investigación**

**Escuela de Ciencias Básicas, Tecnología e Ingeniería**

**Ciudad Cali – Valle del Cauca**

**26/11/2024**

**Introducción de la actividad**

En la presente tarea 4 de Algoritmos de Aprendizaje no Supervisados tiene como inicio comprender y conocer conceptos técnicos sobre las diversas técnicas de aprendizaje no supervisado para así mismo desde estos saber identificar el tipo de modelo que pertenecen para poder llevar a ejecución y análisis desde una aplicación acorde a la situación del problema o necesidad que permitirá una interpretación concreta y significativa del comportamiento de los datos que se están ejecutando así mismo interpretación hacia futuro.

**Objetivos**

**Objetivo Principal**

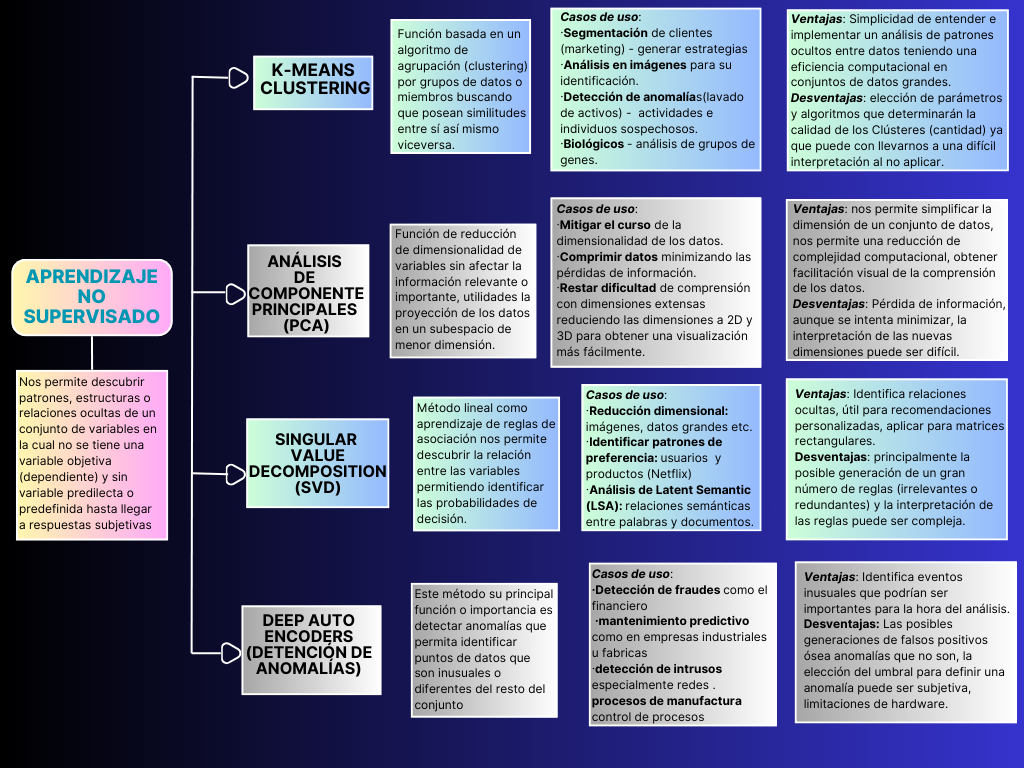
Aplicar algoritmos de Machine Learning no supervisado según el problema, empleando técnicas de Clustering.

**Objetivos generales**

* realizar de forma adecuada el cuadro sinóptico sobre los diferentes modelos de Aprendizaje no supervisado.
* consultar de forma adecuada las definiciones de los conceptos solicitados.
* implementar de forma adecuada los modelos de K-Means e Hierarchical Clustering.
* interpretar de forma adecuada los resultados obtenidos por los modelos.

**Desarrollo de la actividad**

* Cuadro sinóptico



Link acceso:

<https://www.canva.com/design/DAGVlgqwGH4/MZh-lvbsvQQyiDpUgiI1lw/edit?utm_content=DAGVlgqwGH4&utm_campaign=designshare&utm_medium=link2&utm_source=sharebutton>

Carlos Véliz. (2020). [*Aprendizaje automático. Introducción al aprendizaje profundo*](https://bibliotecavirtual.unad.edu.co/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=2600876&lang=es&site=eds-live&scope=site&ebv=EB&ppid=pp_I). El Fondo Editorial de la Pontificia Universidad Católica del Perú. https://bibliotecavirtual.unad.edu.co/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=2600876&lang=es&site=eds-live&scope=site&ebv=EB&ppid=pp\_I Cap 3

Pratap Dangeti. (2017). [*Statistics for Machine Learning*](https://bibliotecavirtual.unad.edu.co/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=1560931&lang=es&site=eds-live&scope=site&ebv=EB&ppid=pp_Cover) : Build Supervised, Unsupervised, and Reinforcement Learning Models Using Both Python and R. Packt Publishing. https://bibliotecavirtual.unad.edu.co/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=1560931&lang=es&site=eds-live&scope=site&ebv=EB&ppid=pp\_Cover Cap 8

Giuseppe Bonaccorso. (2018). [*Machine Learning Algorithms*](https://bibliotecavirtual.unad.edu.co/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=1881497&lang=es&site=eds-live&scope=site&ebv=EB&ppid=pp_Cover) : Popular Algorithms for Data Science and Machine Learning, 2nd Edition: Vol. 2nd ed. Packt Publishing. Tomado de: https://bibliotecavirtual.unad.edu.co/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=1881497&lang=es&site=eds-live&scope=site&ebv=EB&ppid=pp\_Cover Cap 9 y 11

**2.** Definiciones:

**Dendograma**

Diagrama que que permite visualizar las agrupaciones sucesivas generado por un algoritmico jerárquico de forma intuitive en la construcción de grupos permitiendo la generación de diferentes niveles de agrupamiento con la finalidad de proporcionar una información igual o más útil para la agrupación y en casos definitivos.

Los algoritmos jerárquicos son voraces (greedy) afirmando lo anterior dicho que el agrupamiento se da para tener mejores desiciones locales en cada momentpo, peor no asegura una solución global óptima.

En la figura siguiente observamos una agrupación progresiva con tipo de algoritmo glomerativo – estrategia bottom-up y existe otro tipo de algoritmo llamado divisivo – estrategia top – down donde estos se diferencian por el sentido de las flechas.

Figura 1

Dendograma

Diagrama

Descripción generada automáticamente

Nota: tipo de dendograma progresivo porque las flechas tienen el sentido de arriba hacia abajo y si fuera sido opuesto el sentido seria un dendograma dividivo.

**Índice de Calinski-Harabasz**

Métrica que nos permite evaluar el grado de agrupación de un conjunto de datos como de muestra, agrupación y valores de criterio donde para obtener una agrupación positiva se debe obtener un mayor valor del índice ósea el cociente frente la dispersión entre los diferentes clústeres.

Figura 2

Indice de Calinski - Harabasz

Gráfico, Gráfico de líneas

Descripción generada automáticamente

Nota: se visualizan el agrupamiento de los clústeres el cual generan un según comportamiento de los datos teniendo en cuenta hacia donde se inclina más.

L. Pastrán, S.Gongora (2021). Algoritmo de selección y validación del método de clusterización óptimo para datos no supervisado. https://repositorio.utp.edu.co/server/api/core/bitstreams/9c69e97a-7742-4553-8c4f-df3609245065/content

Minguillón, J. Casas, J. y Minguillón, J. (2017). [*Minería de datos: modelos y algoritmos*](https://elibro-net.bibliotecavirtual.unad.edu.co/es/ereader/unad/58656). Editorial UOC. https://elibro-net.bibliotecavirtual.unad.edu.co/es/ereader/unad/58656. Cap 7 y 8

* Con el dataset anterior diseñar los modelos de Clustering (agrupación): K-Means e hierarchical clustering. Para cada algoritmo realizar los siguientes pasos:

1. Realizar un análisis exploratorio de los datos para identificar relaciones entre variables, valores atípicos, tendencias, etc.

Imagen 1

Visualización de datos

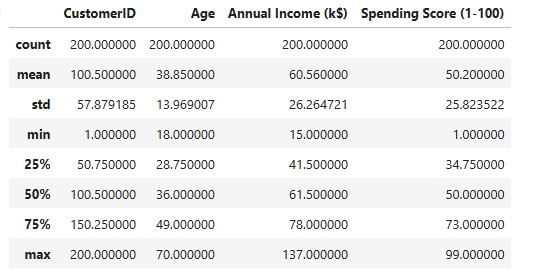
Tabla

Descripción generada automáticamente

*Nota: se visualizan los datos de los clientes para conocer el tipo de dato de los mismos donde solo se observa que la única variable de tipo texto es Gender (Género).*

Imagen 2

Análisis descriptivo



*Nota: se observa la descripción de los datos como la edad promedio de los clientes es de 38,5, el promedio del salario de los clientes está en el 60,5 y la puntación de gastos tiene un gasto medio 50.2 así mismo se encuentra datos indispensables para el análisis como desviación estándar, mínimo y mayor dato de cada variable, cuartiles de 25% al 75%.*

Imagen 3

Análisis de existencia de datos faltantes

Texto, Carta

Descripción generada automáticamente

*Nota: Se puede identificar que ninguna variable posee datos faltantes.*

Imagen 4

Histograma de la variable Age

Gráfico, Histograma

Descripción generada automáticamente

*Nota: Podemos observar las edades más frecuentes de los clientes son entre 32 y 35 años de ahí les sigue las edades 18 y 22 años posterior a estas los comprendidos entre 45 y 49 años, pero las menos frecuentes se encuentran entre los 59 y 67 años.*

Imagen 5

Histograma de la variable Annual Income (k$)

Gráfico, Histograma

Descripción generada automáticamente

*Nota: Podemos observar los ingresos mensuales más frecuentes en clientes están aproximadamente entre 72 y 80 salarios de ahí les sigue los salarios entre 58 y 62 posterior a estas los comprendidos entre 45 y 58 salarios, pero los salarios menos frecuentes se encuentran entre los 85 y 96 igual entre 121 y 130 así mismo entre 130 y 138 salarios.*

Imagen 6

Histograma de la variable Spending Score (1-100)

Gráfico, Histograma

Descripción generada automáticamente

*Nota: la puntuación de gastos más frecuentes entre los clientes es aproximadamente entre 48 y 52 puntos de ahí les sigue los puntajes con igual frecuencia de 41 y 48 puntos posterior a estas los comprendidos entre 52 y 59 puntos, pero los puntos menos frecuentes se encuentran entre los 21 y 28 igual entre 66 y 70 así mismo entre 79 y 84 puntos.*

Imagen 7

Identificación de datos atípicos en la variable Age

Gráfico, Gráfico de cajas y bigotes

Descripción generada automáticamente

*Nota: se puede observar que las edades de clientes más distantes (alejados) frente a la media se encuentran en el extremo derecho ósea que los datos atípicos se encuentran en el extremo superior derecho.*

Imagen 8

Identificación de datos atípicos en la variable Income (k$)

Gráfico, Histograma

Descripción generada automáticamente

Nota: *se puede observar que los clientes con ingresos anuales más distantes (alejados) frente a la media se encuentran hacia el extremo superior derecho en pocas palabras los outliers están en el extremo derecho.*

Imagen 9

Identificación de datos atípicos en la variable Annual Income (k$)

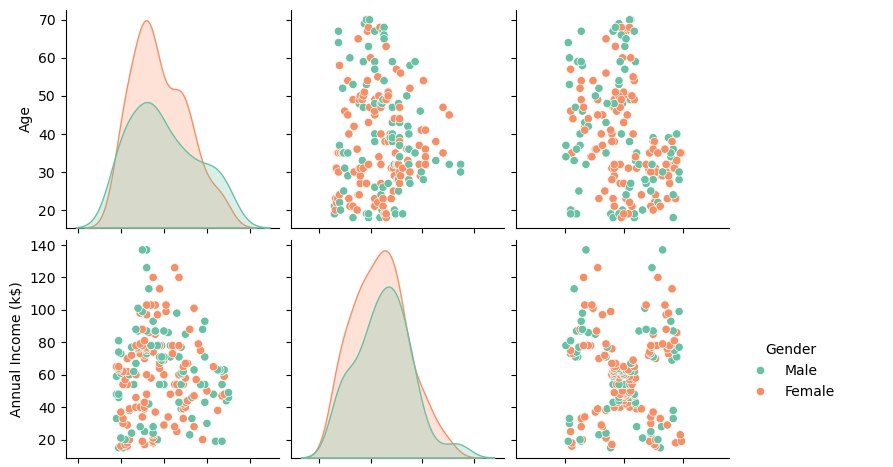
Gráfico, Gráfico de cajas y bigotes

Descripción generada automáticamente

*Nota: se puede observar que los clientes con puntuación de gastos más alejados de la media se encuentran hacia la extrema izquierda por lo tanto hacia ese extremo se direccionan los datos atípicos.*

Imagen 10

Gráfica de dispersión



Gráfico, Gráfico de dispersión

Descripción generada automáticamente

*Nota: podemos observar primero que todo que las mujeres son más frecuentes en este análisis y que no se logra ver una relación línea fuerte entre las variables pares en la gráfica de dispersión frente a los datos, pero con una mínima tendencia quizás de relación entre la variable ingresos anuales y las puntuaciones del gasto porque se logra ver un poco más agrupados a comparación con los otros análisis pares.*

Imagen 11

Matriz de correlación

Gráfico, Gráfico de barras

Descripción generada automáticamente

*Nota: Podemos observar que no existe una correlación e incluso se aprecia negativa entre las variables numéricas como edad con ingresos mensuales con un -0.33 (tendiendo a negativos) así mismo le pasa entre las variables edad y puntos obtenidos con un -0.01 y las únicas variables que tienen posible relación con 0.01 entre las variables ingresos anuales y los puntos obtenidos de los clientes*.

**Análisis exploratorio**

Primero que todo observamos que no existen datos faltantes dentro de las variables por lo tanto no ahí que realizar imputación de datos ya que poseen consistencia y completitud de los datos de los clientes sobre cada variable en este caso 200.

**Variable Age**:

Tenemos como análisis que el promedio de edad de los clientes es 38 .850 años con una variabilidad de las edades frente a la media es de 13.96 años (hacia arriba y abajo) lo que nos indican que los clientes posibilitan estar en edades jóvenes como también más adultas alrededor del promedio así mismo la edad mínima es de 18 años y máxima es 70 años por consiguiente se obtuvo que el Q1 nos determina que el 25% de las personas tienen menos de 28.75 años y Q2 que el 50% tienen menos de 36 años en cambio en Q3 el 75% de los clientes tienen menos de 49 años.

En el histograma observamos que los clientes con edades más frecuentes son entre 32 y 35 años de ahí les sigue las edades 18 y 22 años posterior a estas los comprendidos entre 45 y 49 años, pero las menos frecuentes se encuentran entre los 59 y 67 años

ahora analizando la media con la mediana podemos concluir que es una simetría positiva porque primeramente la media (38.85 años) es un poco más alta que la media (36 años) por el cuál las edades altas provocan esta tendencia y así mismo que se deben encontrar mayor número de clientes con edades inferior a la media (36 años), pero se analiza que los datos (edades) atípicos se encuentran en el extremo derecho por que se encuentran más distantes (alejados) de la media.

**Variable Annual Income (k$):**

Existe clientes con Ingreso anual mínimo anual 15 y máximo ingreso anual 137 donde el promedio del ingreso anual de los clientes se sitúa en 60.56 con una variabilidad de los datos frente a este del 26.26 (arriba y abajo) donde nos conlleva analizar que los ingresos anuales se sitúan tanto menores ingresos anuales como altos, por consiguiente se obtuvo que el Q1 nos determina que el 25% de las personas tienen ingresos anuales inferiores a 41.50 y Q2 que el 50% tienen ingresos inferiores a 61.50 en cambio en Q3 el 75% de las personas tienen ingresos anuales inferiores a 78.

En el histogramapodemos observar que los clientes con ingresos anuales más frecuentes están aproximadamente entre 72 y 80 salarios de ahí les sigue los clientes con los salarios anuales entre 58 y 62 posterior a este los clientes comprendidos con 45 y 58 salarios anuales, pero los clientes con salarios anuales menos frecuentes se encuentran entre los 85 y 96 igual entre 121 y 130 así mismo entre 130 y 138 salarios.

Como podemos analizar la media y la mediana se aproximan ósea que casi los datos tuvieron una tendencia casi simétrica, pero está sesgada levemente hacia izquierda (negativa) porque los salarios bajos de los clientes afectaron un poco la media a pesar de esto los datos atípicos (Outliers) se encuentran más hacia extremo superior derecho.

**Spending Score (1-100):**

Frente a la puntuación de gasto de los clientes tenemos como puntuación mínima 1 y máxima 99 por consiguiente se obtuvo una media de 50.20 siendo justo el medio de los puntos y desviación estándar de 25.82 (arriba y abajo) afirmando que los puntajes de los clientes se encuentran tanto por debajo o por encima de la media, ahora el Q1 (25%) representan los puntajes de los clientes que son inferiores a 34.75 puntos, Q2 (50%) puntajes de gastos inferiores a 50 puntos y Q3 (75%) puntajes de gastos inferiores a 73 puntos.

El histograma nos refleja que los clientes con la frecuencia más alta en puntuación de gastos aproximadamente están entre 48 y 52 puntos posterior a este rango le sigue los clientes con igual frecuencia entre 41 y 48 puntos así mismo entre 52 y 59 puntos, pero los clientes con rangos con frecuencia con menos puntos se encuentran entre los 21 y 28 igual entre 66 y 70 así mismo entre 79 y 84 puntos*.*

Ahora analizando la media y la mediana obtuvimos una tendencia simétrica ya que la media se encuentra en 50.2 y la media en 50 por el cuál no están sesgados a en una dirección significativa (~~izquierda – derecha~~)encontrando un balance general en puntos tanto en los altos y bajos a pesar de esto los puntajes atípicos se presentan inferiores a 34.75 puntos (Q1).

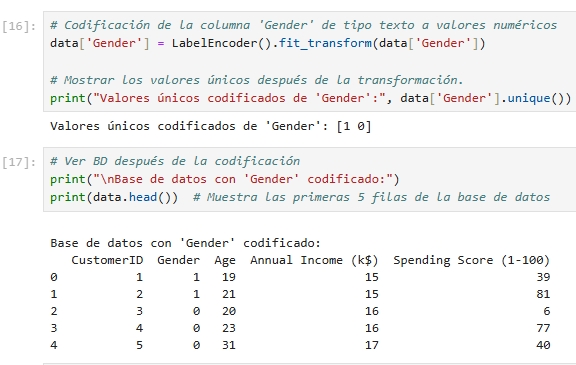
Por último, la gráfica de dispersión primero que todo podemos observar que existen más clientes del género femenino así mismo los gráficos de dispersión resalta que no existe una relación fuerte entre las variables pares analizadas, pero existe quizás una mínima tendencia de relación entre las variables pares ingresos anuales y puntuación de gastos porque ahí más datos agrupados a comparación con otros pares que se les realizó el análisis reafirmándolo la matriz de correlación dos los otros dos pares tienen relación negativa.

* Preprocesar los datos limpiándolos, tratando valores faltantes y transformándolos según sea necesario.

Como anteriormente pudimos observar en la imagen 3 no existen datos faltantes por el cuál no tenemos la necesidad de convertirlos, pero lo que si realizamos fue convertir en dato numérico los datos de la variable Gender ponderado 1 como masculino y 0 femenino esto nos permite tener una DB más uniforme.

Figura 12

Convertir variable Gender



*Nota: podemos observar desde el código hasta la ilustración con una tabla de los datos convertido de texto a numéricos obteniendo una BD más uniforme.*

* Seleccionar las características más relevantes para entrenar el modelo utilizando selección de características.

Figura 13

Selección de características relevantes

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

*Nota: se observa que se escogieron las características Gender, Age, Annual Income, Spending Score para realizar el estudio K-Means de Clustering.*

* Entrenar el modelo configurando los diferentes hiperparámetros.

Lo primero que se realizó fue que ya conociendo que existían valores atípicos en el análisis exploratorio donde esos datos muy alejados a la hora de diseñar los modelos de Clustering (agrupación) distorsionaban el resultado dificultando el mismo porque no se tenía una representación significativa cuando se cambiaba la cantidad de Clusters afectando los análisis estos Outliers.

Se aplico una sensibilidad de 1.0 porque con 1.5 no teniendo diferencia significativa con este rango Inter cuartil así mismo excluyendo los valores atípicos ya que solo eliminaba 2 valores con 1.5 de sensibilidad en cambio con 1.0 de sensibilidad se obtuvieron 8 Outliers y la eliminación de los mismos bajo estos parámetros nos permitió un mejor análisis.

Imagen 14

BD sin Outliers

Tabla

Descripción generada automáticamente

*Nota: se puede observar que la DataFrame original posee 200 datos, pero al eliminar los datos atípicos calculando el rango cuartil (IQR) se detectaron 8 valores atípicos quedando solo 192 valores.*

Por consiguiente, normalizamos la BD sin los valores atípicos como se muestra a continuación.

Imagen 15

Normalización de datos sin Outliers

Texto

Descripción generada automáticamente con confianza bajava

Nota: se evidencia la debida transformación de los datos en cada una de las variables aplicables para el análisis.

* Evaluar el desempeño del modelo con métricas como Coeficiente de Silhouette, Índice de Calinski - Harabasz, etc.

Realizamos estás con 4 métricas para determinar el número de Clusters óptimos

**Métrica de inercia (método de codo):** Evalúa la cohesión interna de los Clusters, es decir, qué tan cerca están los puntos dentro de un mismo Clusters. El valor de inercia disminuye a medida que aumenta el número de Clusters, y el "codo" en la gráfica indica el número óptimo de Clusters, donde la mejora en la reducción de inercia se vuelve menos significativa.

Acorde al análisis anterior utilizamos para la métrica de codo 12 Clusters para poder hallar una mejor interpretación del codo como punto de cambio brusco de dentro del gráfico ya que con diez no se reflejaba bien si 6 o 7 encontrándonos según la imagen siguiente que el codo se refleja en el codo 7 ósea que sería subjetivamente el número de Clustering correcto tocaría analizar las otras métricas para llegar a misma conclusión .

Imagen 16

Métrica Codo

Gráfico, Gráfico de líneas

Descripción generada automáticamente

Nota: se puede visualizar que se encuentra un cambio significativo en el Clusters 7

**Métrica de Coeficiente de Silhouette:** Evalúa la separación entre clusters y la cohesión dentro de ellos. Un valor cercano a 1 indica que los clusters están bien definidos y los puntos están correctamente asignados.

Imagen 17

Coeficiente de Silhouette con unión de líneas

Gráfico, Gráfico de líneas

Descripción generada automáticamente

Nota: Nos representa el Coeficiente de Silhouette trazado con los 12 Clusters para el análisis de comportamiento entre el rango 1 a 12 Clusters

Imagen 18

Coeficiente de Silhouette 7 Clusters

Gráfico, Gráfico de dispersión

Descripción generada automáticamente

Nota: El punto que nos representa el 7 Clúster que tiene una ponderación 3.48 como coeficiente de Silhouette donde es aceptable ósea moderado.

**Métrica de Índice Calinski-Harabasz:** nos permite calcular la relación entre la dispersión entre Clusters y dentro de ellos donde si poseemos un valor más alto significa Clusters bien separados y son más compactos internamente.

Imagen 19

Índice Calinski-Harabasz trazado de línea

Gráfico, Gráfico de líneas

Descripción generada automáticamente

Nota: Nos representa el Índice Calinski-Harabasz trazado entre un rango entre 1 y 12 Clusters con el fin de determinar su punto más alto en determinado cambio significativo en este caso 72.662 una ponderación positiva u significativa en los 7 Clusters igual que coeficiente de Silhouette.

Imagen 20

Índice Calinski-Harabasz

Gráfico, Gráfico de dispersión

Descripción generada automáticamente

*Nota: El punto que nos representa el 7 Clúster que tiene una ponderación 72.662 como Índice Calinski-Harabasz siendo relativamente alto.*

**Figura 21**

**Visualización de componente principales 2D PCA -T-SNE**

Gráfico, Gráfico de dispersión

Descripción generada automáticamenteGráfico, Gráfico de dispersión

Descripción generada automáticamente

Nota: podemos visualizar en las gráficas tanto en PCA Y T-SNE se visualiza una buen Clustering los datos ya que T-SNE hace ver cómo se conserva la distribución en 7 Clusters ya que este preserva las distancias entre ellos y en PCA a pesar de que prevalece como una dimensión mayor se logra identificar los centroides para cada Clustering y un comportamiento moderado así mismo como lo estipulo el codo.

Figura 22

Comparación con Clustering K=7 y K=8 POR PCA

Gráfico, Gráfico de dispersión

Descripción generada automáticamente

*Nota: Podemos identificar que con K=7 se agrupan mejor los Clustering mientras que con K=8 se empieza a ver más dispersión y esto puede conllevar a que datos en el Clustering alterar el análisis y cada vez que no aumentamos k pues se van a dispersar tanto los datos como Clustering y no se logra el objetivo que es el mejor agrupamiento posible valores.*

Figura 23

Dendrograma

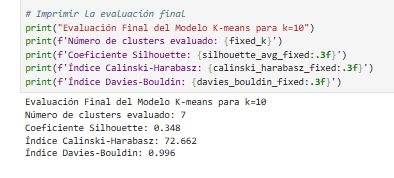
Gráfico

Descripción generada automáticamente

*Nota: Podemos observar dos grandes divisiones entre naranja y verde donde principalmente se divide en 2 Clusters posterior a estos le siguen 6 Clusters secundarios podemos decir que existen dos grupos con gran generador de gasto unos según con mejores ingresos y otros con menores ingresos y los que lo siguen unos con más generadores que gastos que otros y viceversa igual la condición económica.*

Figura 24

Evaluación final



*Nota: la mejor opción de Clustering es con 7 Clusters donde el coeficiente de Silhouette es aceptable (Moderado) con un 0.348 y el índice Calinski – Harabasz 72.662 siendo positiva razonable, aunque coeficiente de Silhouette sea aceptable.*

* Realizar las diferentes gráficas que permitan visualizar los resultados del modelo 7. Interpretar, analizar y documentar los resultados obtenidos.

De acuerdo a todo lo anterior podemos concluir que desde inicio que se inició con el análisis descriptivo se obtuvo como resultado variables con poca relación como se ilustro en la matriz de correlación esto iba a significar un mejor análisis e interpretación apara la hora de realizar un Clustering con la BD por eso primero se intentó con 10 Clusters donde se obtenía un buen coeficiente de silueta pero resaltando que al aumentar los mismo no es que se obtuviera un aumento significativo haciendo que cada vez grupos más pequeños y valores más distante permitiendo que quizás datos o valores se queden por fuera por eso por todo lo anterior mencionado hace más difícil interpretar, igual que el índice de Kalinski Harabasz buen resultado por la integración de los valores con el centroide pero puede suceder lo mencionado en el coeficiente de silueta además como se representa mejor es con 7 porque se agrupan bien y la separación de los Clusters es recomendable porque no se dispersan tanto los datos y se logra evidenciar los mismos con el centroide.

Afirmando lo objetivo e intuido con el codo pues salió siendo algo coherente con las similitudes sobre el número de Clusters necesarios donde las gráficas de dispersión y el Dendrograma que oscila entre los 6 y 7 Clusters por eso tiene una mejor tendencia de agrupación (Clustering) y representación con 7 Clusters.

7. Interpretar, analizar y documentar los resultados obtenidos.

Podemos interpretar el análisis con todo lo anterior dicho quedando que cada Clusters se puede interpretar de la siguiente manera:

**Cluster 1: Jóvenes de ingresos altos con alto Spending Score osea se tiene** Personas jóvenes con un ingreso alto y un comportamiento de gasto elevado donde se puede focalizar el marketing Ofrecer productos premium, de lujo o experiencias exclusivas (como viajes personalizados, tecnología de alta gama).

**Cluster 2: Jóvenes de ingresos bajos y alto Spending Score son aquellas** Personas jóvenes con ingresos limitados, pero que tienden a gastar en productos o servicios de moda o tendencia donde se podría focalizar en campañas en descuentos, promociones y etc.

**Cluster 3:** Adultos mayores con ingresos altos y bajo el Spending Score son aquellos adultos mayores que tienden a tener hábitos de gasto más mesurado ósea conservadores este le puede servir productos de servicios , asesoría financiera entre otros.

**Cluster 4: Adultos mayores con ingresos bajos, bajo Spending Score** ósea sonPersonas mayores con ingresos bajos y un bajo nivel de gasto permitiendo obtener productos de más bajos costos p, baja calidad e imitaciones.

**Cluster 5:** Jóvenes con ingresos medios y un Spending Score medio ósea que estos mantienen un equilibrio ante lo devengado y el gasto de estos pueden tener un equilibrio así mismo las estrategias serian productos asequibles y en métodos de pagó

**Cluster 6:** Personas de todas las edades, ingreso medio, bajo Spending Scre: Personas más prudentes, que tienden a priorizar el ahorro independientemente del ingreso así mismo permitiendo adquirir productos ya sea contado o método de endeudamiento

**Cluster 7:** Clientes mixtos de ingresos altos y bajo Spending Score estos son los clientes con mejores ingresos, pero con gastos mesurados donde sería genial generar una estrategia de permita llamar mucho la atención de estos clientes de su interés