

Matrimonios en Guatemala.

RESUMEN

Objetivos

- Determinar si existe una tendencia respecto a la diferencia de edad entre los casados, cual es la media de esta.
- Determinar que genero tiende a ser mayor de edad al momento de contraer matrimonio.
- Determinar y analizar una media entre la diferencia de edad de las parejas.
- Determinar cuál es la edad correcta en un hombre para que un matrimonio sea "bueno" con distintos parámetros y puntajes como la escolaridad de las parejas.

En Guatemala se dan matrimonios entre mujeres y hombres donde la edad de estos puede tener una gran diferencia. Esto puede ser ocasionado por muchas razones, puede ser por amor, interés propio, obligados por terceros, etc. Esto empieza a volverse un problema cuando alguno de las parejas es menor de edad.

El Renap asegura que esta cantidad de matrimonios donde hay algún menor de edad se está disminuyendo al pasar los años, especialmente luego del decreto impuesto en el 13-2017 impuesto en el 2017 prohibiendo contracción de matrimonio en menores. Sin embargo, todavía existen estos casos, incluso, casos donde menores de edad se encuentran embarazadas y existe una gran diferencia de edad con el marido. Esto puede generar problemas o condiciones de salud física y/o mental en la embarazada y por lo general genera más pobreza ya que estos casos se dan más que todo en sectores pobres.

Se utilizarán herramientas de análisis de datos para buscar tendencias y analizar bien los datos proporcionados por el gobierno de Guatemala respecto a los matrimonios, con información de, escolaridad, edad, departamento, etc. Respecto a las parejas.

Marco Teórico

Algoritmos a utilizar:

- K-MEANS: Utilizamos el algoritmo K-means , este es utilizado para encontrar el número de clústeres y número de centroides más óptimos para generar dado un set de datos.
- Método del Codo: Este utiliza la distancia media de las observaciones. Es decir, se fija en las distancias intra-clusters. Cuanto más grande es el número de clusters k, la varianza intra-clusters tiende a disminuir. Y cuando es menos son más compactos.
- Método Silueta: Este mide la calidad del agrupamiento o clusterings. Mide la distancia de separación entre los clusters.
- (Clustering) Agrupamiento jerárquico:
El agrupamiento jerárquico es un método de análisis de grupos puntuales, el cual busca construir una jerarquía de grupos. Estrategias para agrupamiento jerárquico generalmente caen en dos tipos:
 - Aglomerativas: Este es un acercamiento ascendente: cada observación comienza en su propio grupo, y los pares de grupos son mezclados mientras uno sube en la jerarquía.
 - Divisivas: Este es un acercamiento descendente: todas las observaciones comienzan en un grupo, y se realizan divisiones mientras uno baja en la jerarquía.

Clústeres

Centroides

Descripción de Datos y análisis exploratorio:

Los datos disponibles poseen varias bases de datos con distintos clústeres de datos en ellas, estadísticas respecto a los matrimonios y nacimientos en distintos lapsos de tiempo, colegiaturas, etc. brindadas por el gobierno mediante una página web. Para realizar los análisis necesarios se utilizarán las bases de datos de los años de 20 a 20 , esto debido a que por el problema y los objetivos planteados. Debemos analizar los datos previos al decreto 13-2017 (impuesto el 2017) y post el decreto 13-2017 para analizar correctamente los datos en sus lapsos de tiempo necesarios y lograr cumplir nuestros objetivos. Lo primero a realizar es la unión y organización de las bases de datos a utilizar, en estos casos serían los matrimonios y nacimientos de 20 a 20+, organización refiere adición y extracción de columnas necesarios, modificación de los nombres necesarios. Se trabajará con las variables cuantitativas numéricas.

Se realizara de igual manera varias ponderaciones para lograr transformar variables cualitativas como departamento, escolaridad , en variables cuantitativas para así poder utilizar el algoritmo de RNA predicción.

Análisis Exploratorio:

DEPREG:cualitativa nominal

MUPREG:cualitativa nominal

MESREG:cuantitativa discreta

AÑOREG:cuantitativa discreta

AÑOOCU:cuantitativa discreta

CLAUNI:cualitativa nominal

NUNUHO:cuantitativa discreta

NUNUMU:cuantitativa discreta

EDADHOM:cuantitativa discreta

EDADMUJ:cuantitativa discreta

NACHOM:cualitativa nominal

NACHMUJ:cualitativa nominal

ESCHOM:cualitativa ordinal

ESCMUJ:cualitativa ordinal

CIUHOM:cualitativa nominal

CIUMUJ:cualitativa nominal

DEPOCU:cualitativa nominal

MUPOCU:cualitativa nominal

DIAOCU:cuantitativa discreta

MESOCU:cualitativa discreta

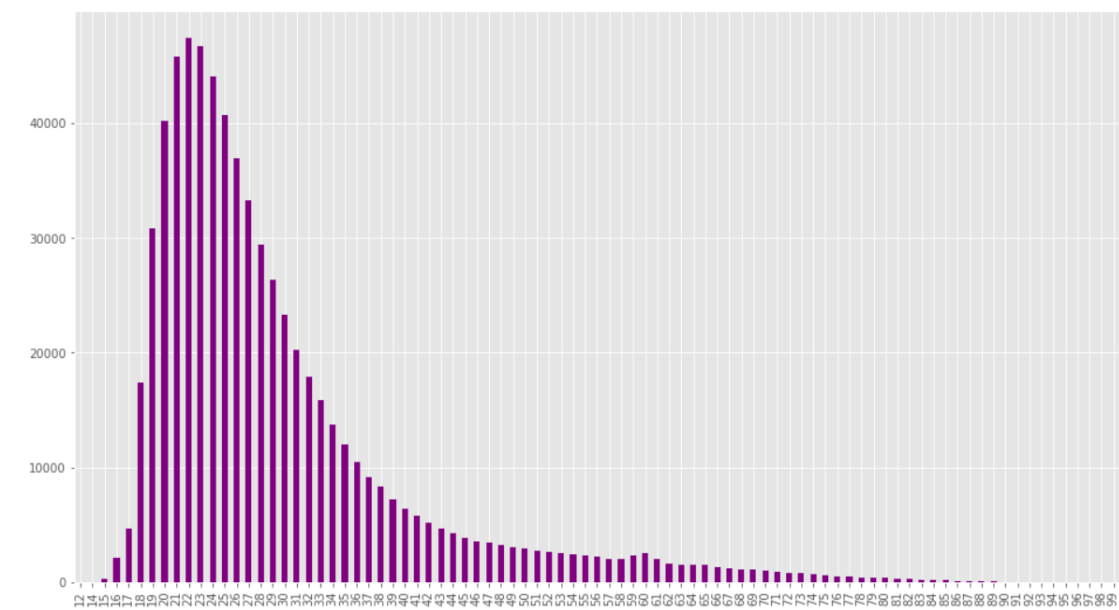
GETHOM: cualitativa nominal

GETMUJ: cualitativa nominal

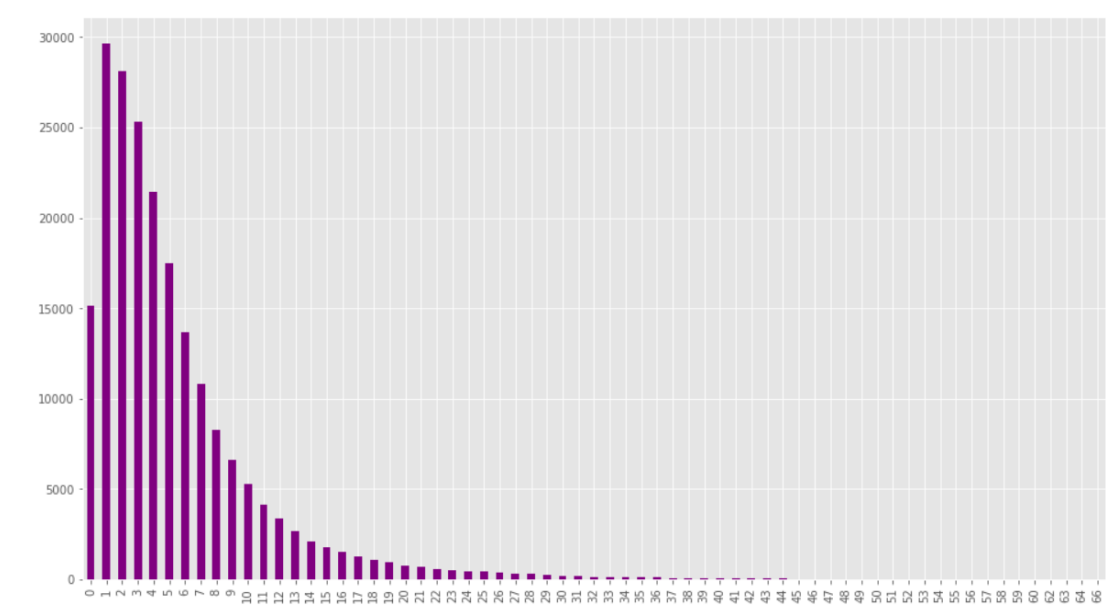
Discusión de Resultados

a

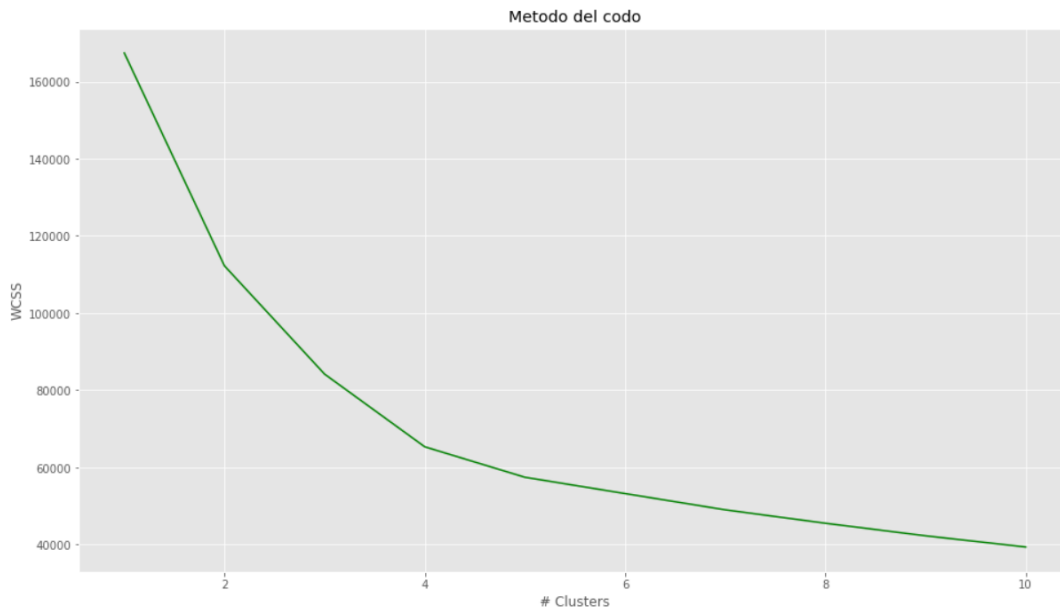
Primero se realizaron varios gráficos para iniciar el análisis de los datos así observando las edades de los matrimonios en las parejas, diferencia de edades, departamentos de dominancia respecto a los casados, escolaridades, etc.



Podemos observar que la edad más común para contraer Matrimonio es a los 20-25 años , sin embargo, existen matrimonios de menores de edad, principalmente a la edad de 17 años, siendo el hombre todavía NO mayor de edad, también matrimonios a los 15-16 años.



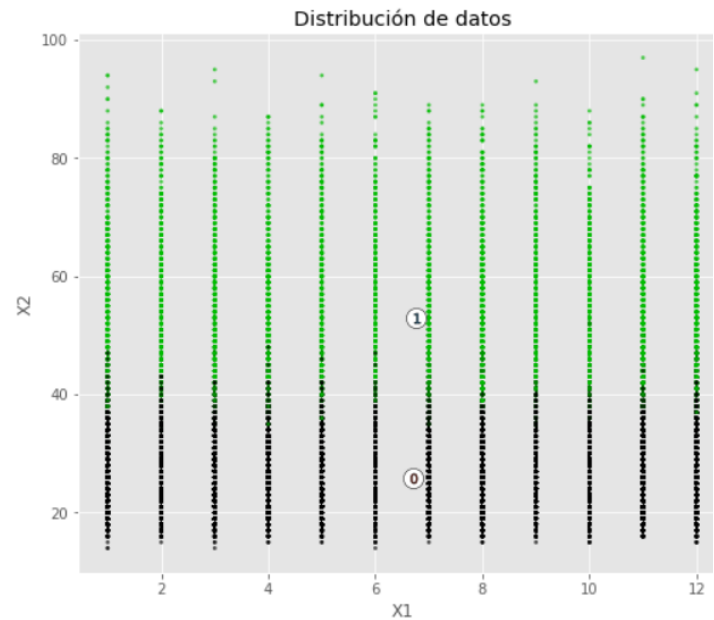
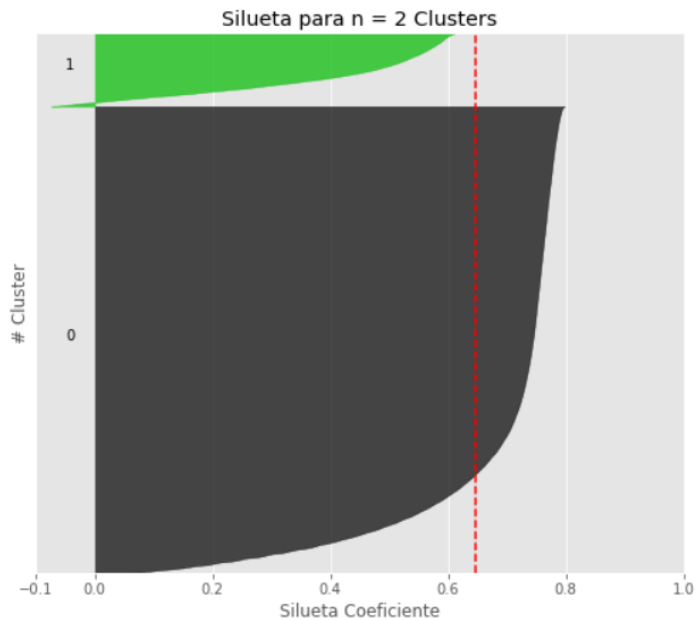
En este gráfico podemos observar la diferencia de edades entre los casados, la mayoría de matrimonios tienen una diferencia entre 1-5 años.



Utilizamos KMEANS, y podemos observar como el declive en la pendiente se da en el valor de 2, siendo este la mejor cantidad de clusters a utilizar gracias al método de Codo.

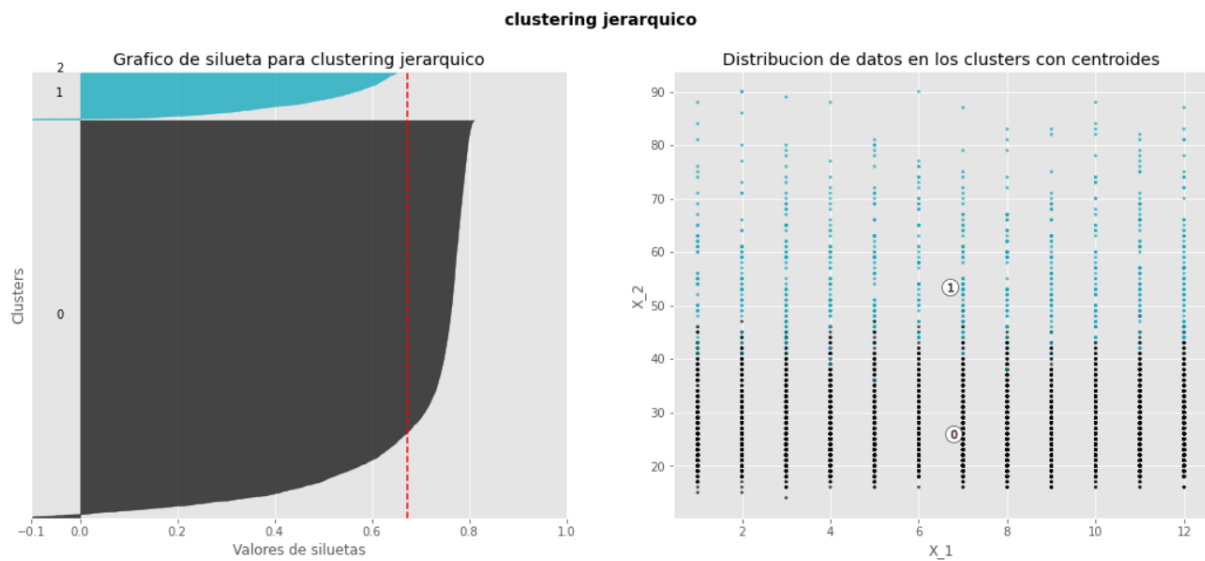
Con: 2 Clusters La cantidad promedio de la silueta es : 0.6470934954694808
 Con: 3 Clusters La cantidad promedio de la silueta es : 0.45092961450524927
 Con: 4 Clusters La cantidad promedio de la silueta es : 0.3554927285661972
 Con: 5 Clusters La cantidad promedio de la silueta es : 0.2930803134322536
 Con: 6 Clusters La cantidad promedio de la silueta es : 0.2600252843572982

Silueta K-Means con = 2 Clusters

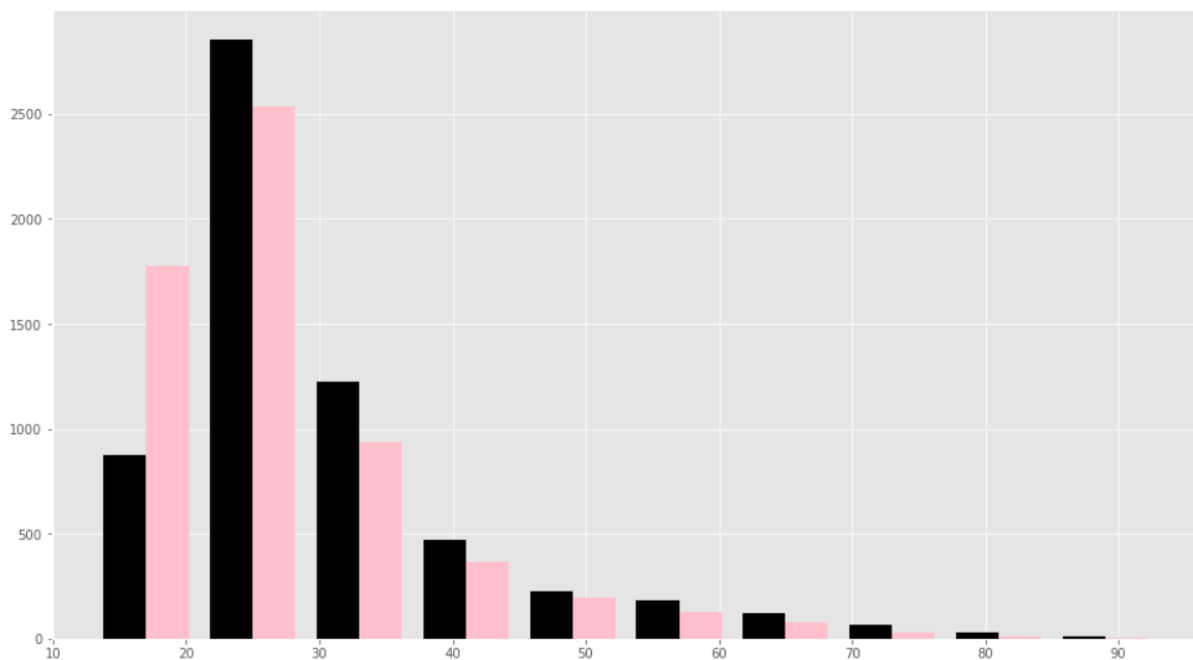


Podemos observar los coeficientes de silueta de nuestros distintos escenarios con distinta cantidad de clusters, notamos que en 2 clusters tenemos un coeficiente de 0.64, siendo este un buen

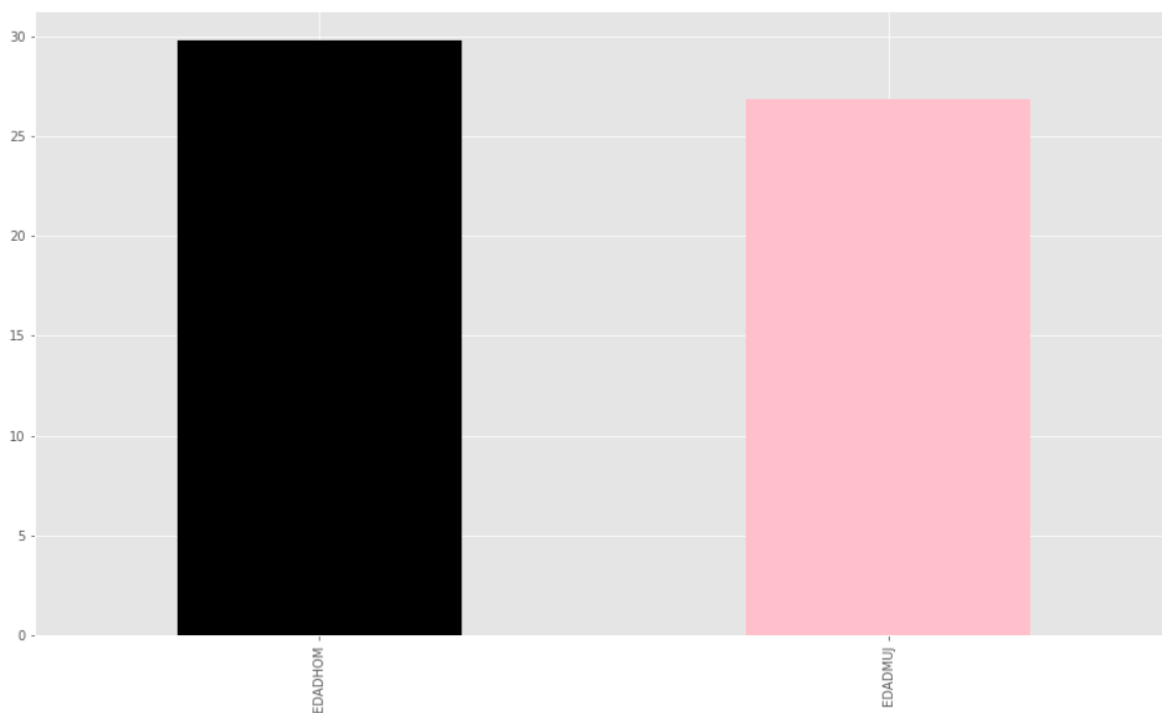
resultado del promedio. Dado este resultado más el metodo de codo podemos reafirmar que utilizar 2 clústeres es lo mas optimo



CLUSTER 0

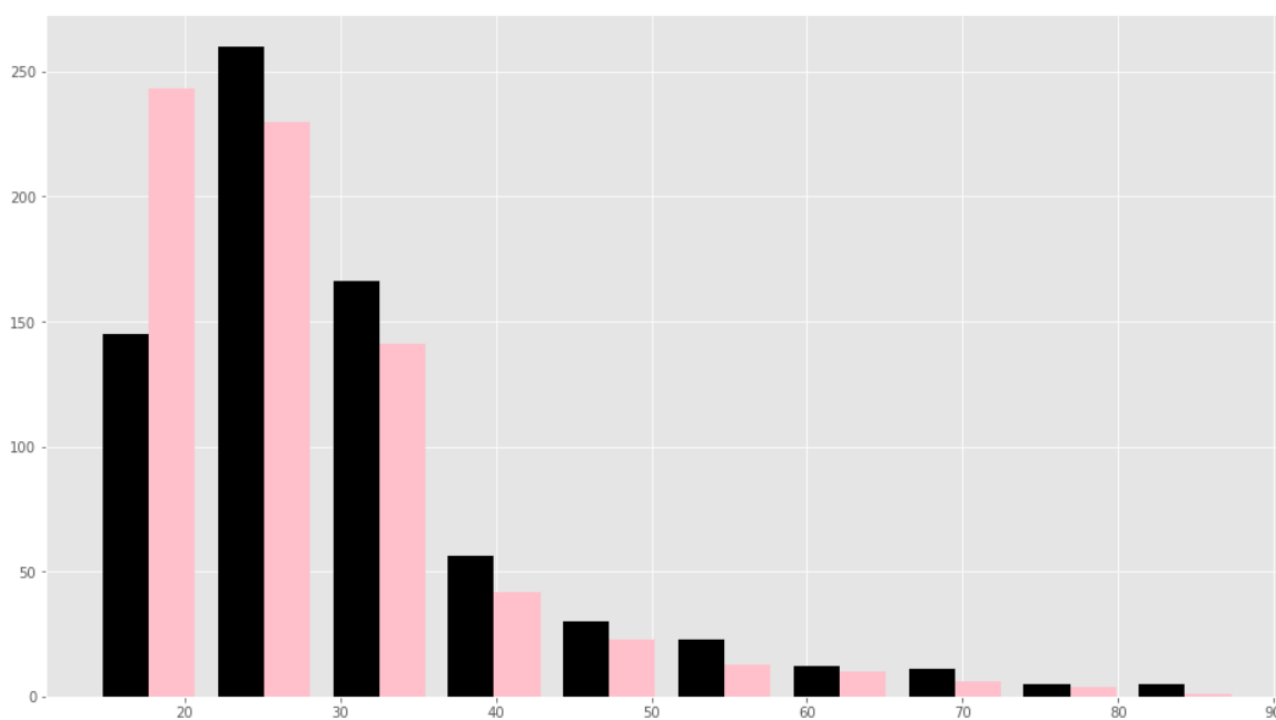


POdemos vere como los hombres (color negro) tienden a ser mayores al momento de contraer matrimonio si son mayores de 20 años, sin embargo, las mujeres (color rosado) tienden a ser mayores si es en casos de menos de 20 años.

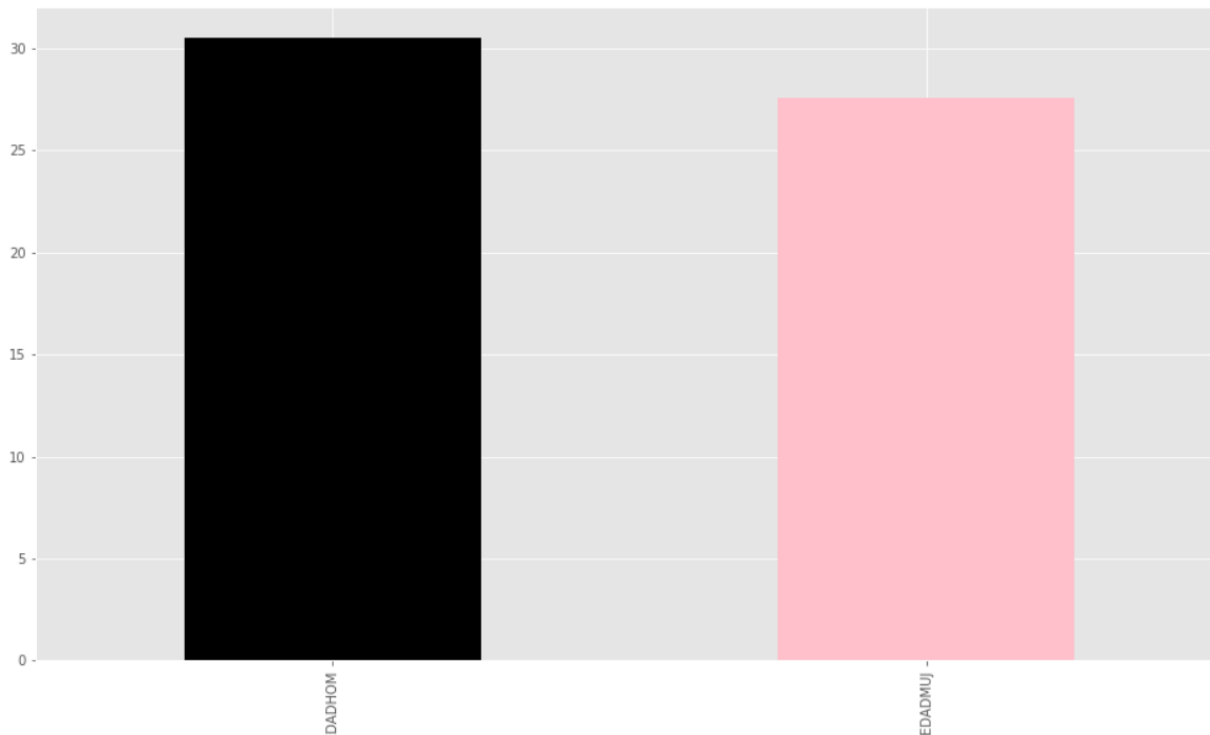


Podemos ver como el promedio de edad al contraer matrimonio es similar entre los géneros, sin embargo, el hombre si es mayor en la mayoría de matrimonios.

CLUSTER 1:



Podemos ver que para el cluster 1 de igual manera predominan las mismas condiciones que en el cluster 0, los hombres tienden a ser mayores cuando son mayores a 20 y las mujeres cuando son menores, cabe añadir que la mayoría de matrimonios se contraen a los 20.



De igual manera que con el cluster 0 podemos ver como el promedio de edad al contraer matrimonio es similar entre los géneros, sin embargo, el hombre si es mayor en la mayoría de matrimonios

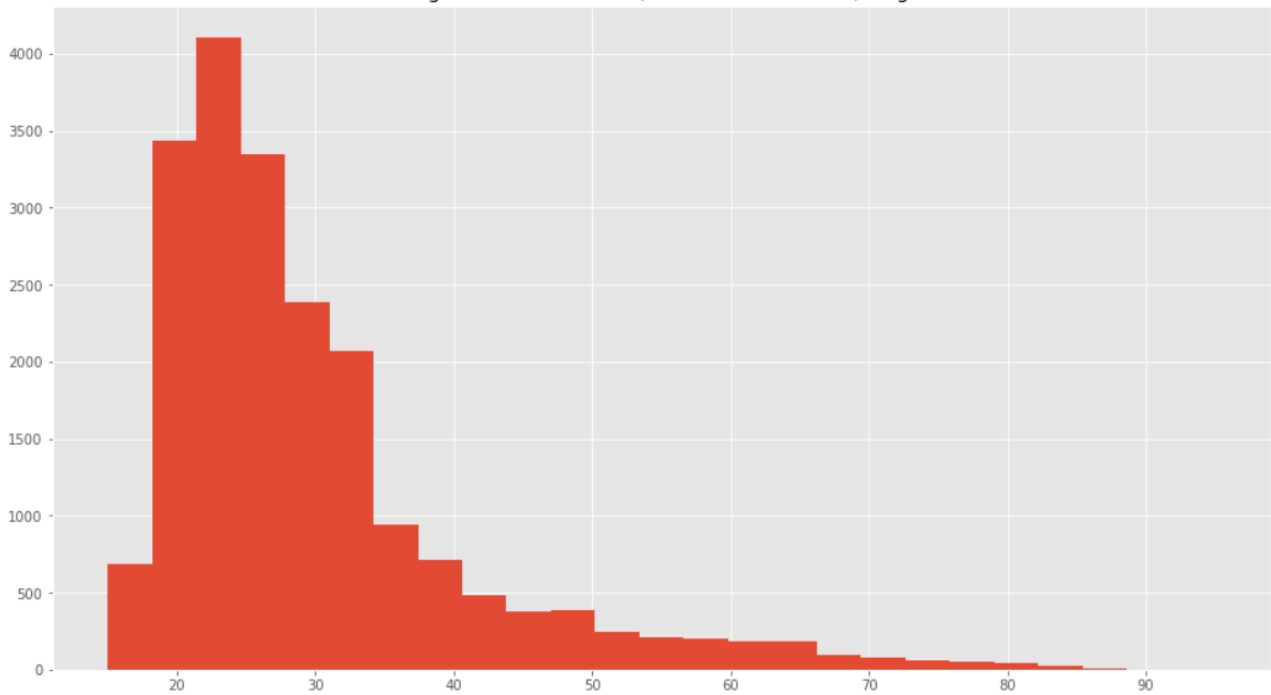
INTERPRETACIÓN DE CLUSTERS

Podemos ver como nuestro cluster 1 posee una cantidad de datos muy menor bastante significativa a diferencia del cluster 0, esto nos dice que ese cluster es tendencia a datos no típicos, si observamos los gráficos podemos ver como en nuestra muestra de datos del cluster 1 existen más matrimonios con una diferencia de edad bastante grande . Podemos asumir que el cluster 0 es más acertado al momento de decir un matrimonio bueno por así decirlo, mientras que , el cluster 1 es de matrimonios con tendencia a no ir bien. Si se desea explorar más seria idea realizar esto con los hombres y mujeres casados entre los 20 años que es la edad promedio mayor al momento de contraer matrimonio.

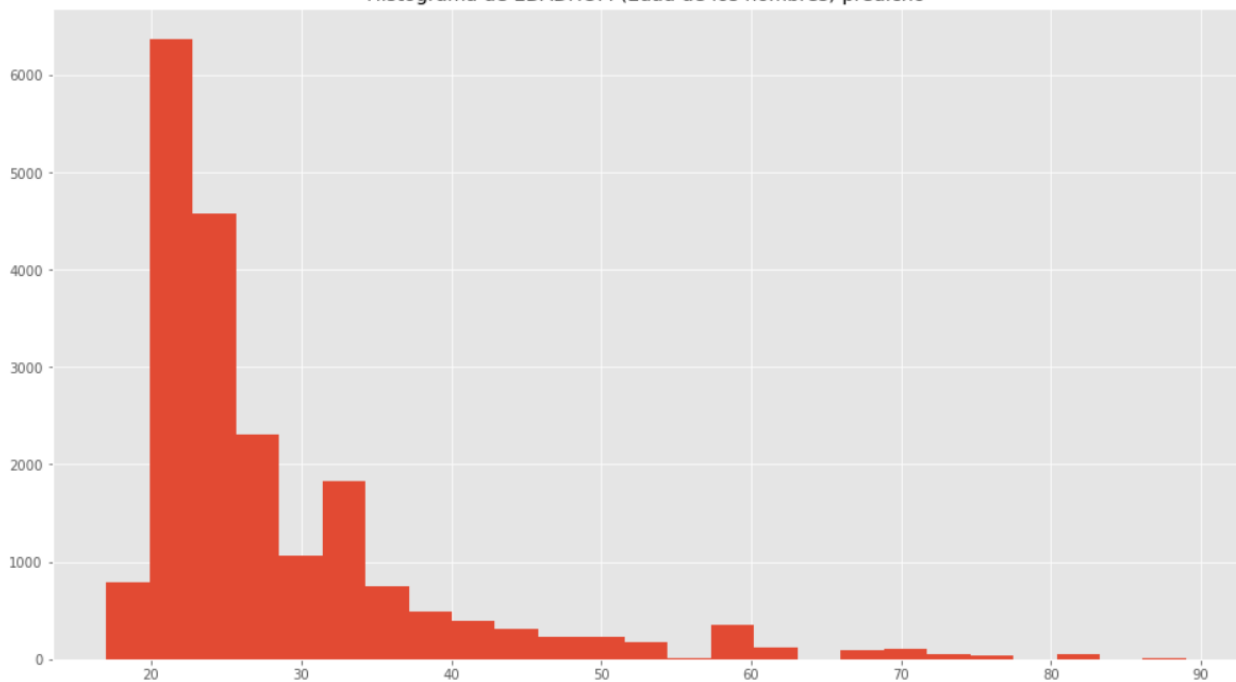
CLUSTER_0 = 'Matrimonios con tendencia correcta' CLUSTER_1 = 'Matrimonios con tendencia incorrecta'

PREDICCION RNA:

Histograma de EDADHOM (Edad de los Hombres) Original



Histograma de EDADHOM (Edad de los hombres) predicho



Luego de analizar el modelo y los resultados de la predicción, podemos observar varios factores que afirman que la predicción fue certera. Comenzando con que el R^2 tuvo un valor cercano a 1 con 0.72 lo que nos indica la clara relación entre los valores predichos y los valores de la muestra original. También podemos observar la varianza entre la predicción y los valores normales, de igual forma tiene un valor cercano a 1 con un total de 0.73, lo cual todo esto nos indica que la predicción fue meramente certera. También se realizó el despliegue de las medias de la predicción y los valores normales, con una diferencia de 1.32, los valores son 29.59 de normal y 28.27 de predicción. Al observar los histogramas desplegados podemos notar la asimetría similar entre ambos gráficos, sin embargo, hay una gran diferencia entre los valores empezando en las edades mayores a ~20-25. No obstante, la predicción fue muy certera y podemos notar que si fue una buena predicción.

CONCLUSIONES

La predicción de la edad del hombre fue realizada en base a la educación, edad de la mujer, escolaridad del hombre y escolaridad de la mujer, estas variables no poseen la mejor relación entre ellas, debido a que se utilizó la variable predictora como la edad de la mujer se puede utilizar el modelo de RNA para hacer esta predicción, de no tener esta variable predictora la predicción no sería posible. Esto se debe a que esta variable predictora ayuda a marcar la tendencia respecto a la educación. Podemos concluir que es posible realizar dicha predicción realizada de manera óptima y confiable, siempre tomando en cuenta lo mencionado anteriormente, se debe utilizar la variable predictora como la edad de la mujer para que dicho modelo sea funcional. También concluimos que la edad correcta para un hombre es de 28.27 dada la predicción y 29.59 dado sin predicción, todo esto cumpliendo nuestros objetivos.

:

A. (2019, 8 marzo). *Segmentación utilizando K-means en Python*.

MachineLearningParaTodos.com.

<https://machinelearningparatodos.com/segmentacion-utilizando-k-means-en-python/#:%7E:text=1%2D%20El%20m%C3%A9todo%20del%20código%3A&text=El%20m%C3%A9todo%20del%20código%20utiliza,intra%2Dcluster%20tiende%20a%20disminuir.>

Wikipedia. (2022, 15 mayo). *Agrupamiento jerárquico*. Wikipedia, la enciclopedia libre.

https://es.wikipedia.org/wiki/Agrupamiento_jer%C3%A1rquico

[Redes Neuronales Artificiales aplicadas al Análisis de Datos](#)