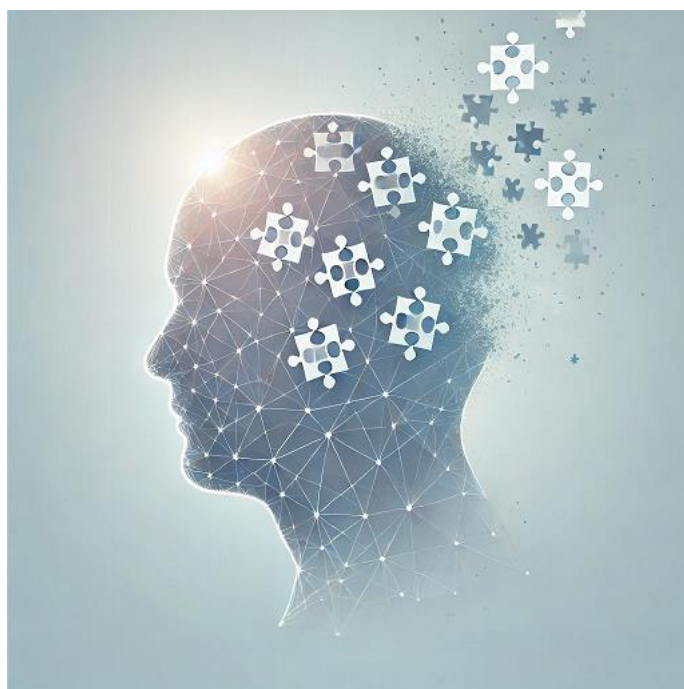


Análisis Exploratorio de Datos (EDA) sobre la enfermedad del Alzheimer.



Memoria descriptiva y conclusiones basadas en los hallazgos exploratorios.

Marta Tébar

Diciembre 2024

1. Contexto del Problema

- **Descripción general del problema:**

El **Alzheimer** es una enfermedad que afecta a un 4% de la población en España entre 75 y 79 años, y aumenta hasta un 34% en los mayores de 85 años [1], de ahí la gran importancia que tiene su análisis e investigación para mejorar la **prevención, detección temprana y tratamiento**. A esto hay que sumarle que la población está envejeciendo, se espera que en 2050 haya más de un 16% de personas mayores de 65 años [2], por lo que es de esperar que aumente la incidencia del Alzheimer en el futuro.

Esta enfermedad suele diagnosticarse examinando los antecedentes médicos, haciendo evaluaciones del estado mental y del humor, pruebas físicas y neurológicas, análisis de sangre y pruebas de imagen que puedan descartar otras enfermedades como la demencia [3]. Por ejemplo, existen pruebas de rápidas de cribado como el Mini-Mental State Examination (MMSE), que es la más usada y consiste en una serie de preguntas y realización de acciones que valoran la memoria, orientación temporal y espacial, y concentración del paciente con una puntuación de 0 a 30 [4].

Este EDA tiene como **objetivo** principal analizar las variables que pueden influir más en pacientes con Alzheimer, y que puedan llegar a servir para mejorar su diagnóstico e incluso tratamiento. Puede ayudar a plantear nuevos enfoques de investigación y conocimiento sobre esta enfermedad. Además, en un futuro este análisis podría ser utilizado como punto de partida para desarrollar modelos predictivos de *Machine* y *Deep Learning* que optimicen la detección temprana y guíen intervenciones más efectivas en el manejo de esta enfermedad.

- **Características del dataset:**

El dataset elegido para la realización del EDA es una **base de datos sintética** que ha sido obtenida de la plataforma *kaggle*, debido a que, además de tener un gran número de pacientes, tiene una gran cantidad de variables que se adecúan a las utilizadas habitualmente en el diagnóstico, mencionadas previamente. Además, reúne pacientes diagnosticados y no diagnosticados con la enfermedad, lo que permitirá realizar en el futuro el análisis predictivo ya comentado.

Concretamente, tiene 2149 pacientes diferentes como **filas**, de los cuales 1389 son sanos y 760 son diagnosticados de Alzheimer, y 35 variables distintas como **columnas**, que se pueden clasificar en:

- ID para identificarlos y actuar como índice.
- Variables demográficas: edad, sexo, etnia, nivel de educación.
- Factores de la vida diaria: Body Mass Index (BMI), si el paciente fuma, el alcohol que consume semanalmente, el ejercicio físico que hace por semana, la calidad de la dieta y del sueño.
- Historial médico: si tiene antecedentes familiares de la enfermedad, si padece enfermedades cardiovasculares, si tiene diabetes, si tiene depresión, si ha sufrido un traumatismo craneal y si tiene hipertensión.
- Medidas clínicas: presión arterial sistólica o diastólica, el colesterol total, colesterol de lipoproteínas de baja densidad (LDL), colesterol de lipoproteínas de alta densidad (HDL) y triglicéridos.
- Evaluaciones cognitivas y funcionales: MMSE (Mini-Mental State Examination), evaluación funcional, quejas de memoria, problemas de conducta y ADL (Activities of Daily Living).

- Síntomas: confusión, desorientación, cambios de personalidad, dificultad en completar tareas y olvido.
- La columna target que indica si el paciente está diagnosticado de Alzheimer o no.
- Información sobre el doctor que es confidencial.

En este caso no hizo falta realizar limpieza de valores nulos ni duplicados, puesto que no había, aunque se comprobó que todo era correcto para no incluir errores que pudieran afectar al posterior análisis.

2. Hipótesis

Una vez presentado el contexto y con el dataset elegido, es posible plantearse unas hipótesis que, mediante el EDA, puedan ser comprobadas o rechazadas. Después del posterior análisis, podrían obtenerse diferentes conclusiones que ayuden a crear un plan de acción futuro sobre la enfermedad del Alzheimer.

- **Hipótesis 1:** ¿hay alguna tendencia en los datos demográficos en los pacientes de Alzheimer? Sobre todo podría ser la edad o, incluso, el nivel educacional.
- **Hipótesis 2:** ¿el estilo de vida influye en el desarrollo de Alzheimer?
- **Hipótesis 3:** ¿los datos del historial médico influyen en el desarrollo de Alzheimer?
- **Hipótesis 4:** ¿las mediciones clínicas sirven para diagnosticar la enfermedad? ¿y las evaluaciones cognitivas?
- **Hipótesis 5:** ¿están estos síntomas del dataset relacionados con el Alzheimer?

3. Análisis

Previo a la realización del análisis, se realizaron las comprobaciones pertinentes que indicaban que los datos estaban efectivamente limpios y listos para continuar. Además, se simplificaron las medidas clínicas, de forma que únicamente se utilizaron el colesterol total y la presión arterial media que fue una variable creada a partir de dos ya existentes. También se eliminó la columna del doctor al tratarse de datos confidenciales.

Con esto, se realizó el análisis en primer lugar **univariante**, para poder ver la distribución general de los pacientes del dataset, y, con las conclusiones obtenidas, se llevó a cabo también un análisis **bivariante** teniendo en cuenta la columna target de diagnóstico, siendo este análisis el más relevante para poder resolver las preguntas que se habían planteado inicialmente. Para terminar y tener una visión general, se realizó un análisis **multivariante**, observando las correlaciones entre variables y viendo si se alinean con los resultados obtenidos previamente.

I. Univariante

En primer lugar, se realizó un análisis univariante de cada una de las variables del dataset, incluyendo pacientes diagnosticados y no diagnosticados con la enfermedad. De esta forma, se podría observar la **distribución** en general de cada una de las columnas.

Por grupos de variables ya mencionados previamente, estas son las descripciones y hallazgos principales:

- Variables demográficas.

La edad de los pacientes es bastante uniforme yendo desde los 60 a los 90 años, aunque hay un pico en las personas más mayores, la media es 75 años. Por tanto, el estudio se centra en personas de avanzada edad, que son las que más sufren esta enfermedad y sus consecuencias.

Respecto al sexo, hay prácticamente el mismo número de pacientes mujeres que hombres, lo cual es bueno para el estudio puesto que no incluye un sesgo respecto al género.

Si se tiene en cuenta la etnia, hay casi el triple de pacientes caucásicos que del resto como se puede ver en la Figura 1, por lo que es una variable que no se tendrá en cuenta a lo largo del estudio.

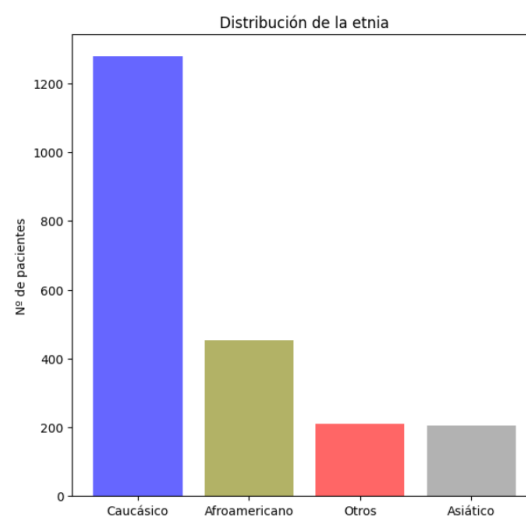


Figura 1: Distribución de las etnias de los pacientes.

En relación al nivel de educación, la mayoría de los pacientes tienen la escuela secundaria, y según aumenta el nivel educacional hay un menor número de personas, hay incluso más pacientes sin estudios que con estudios superiores. Esto indica que podría ser una variable interesante en el análisis bivalente para ver si esto puede influir en la enfermedad.

- Factores del estilo de vida.

Todas las variables que se presentan en esta categoría tienen distribuciones uniformes entre los pacientes y no presentan outliers, por lo que habrá pacientes de todo tipo y no nos permiten ver ninguna tendencia relevante en ninguna de ellas. El BMI de los pacientes es de media 27.65, consumen 10 litros de alcohol por semana de media, hacen 4.92 horas de deporte de media por semana, la media de la calidad de la dieta es 5/10 y de la calidad del sueño 7/10 (va de 4 a 10 el rango de valores, como se puede ver en la Figura 2). En general, todos los valores numéricos están sobre la media de sus respectivos rangos, lo que indica también una desviación estándar alta.

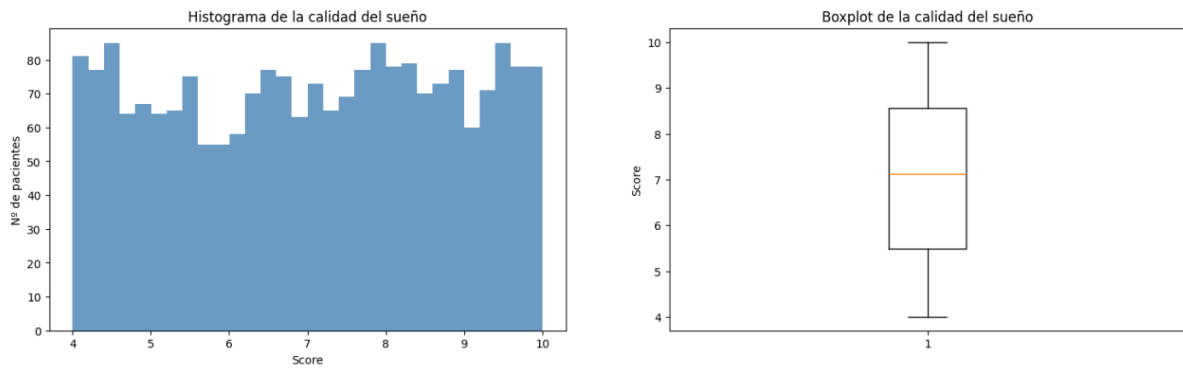


Figura 2: Histograma de la calidad del sueño (izquierda) y boxplot (derecha).

Además, se obtiene que más del doble de los pacientes no fuman. Se deberá realizar un futuro análisis separando a las personas diagnosticadas de las que no lo están para ver si ahí puede existir alguna tendencia.

- Historial médico.

En general, el análisis mostró que la mayoría de los pacientes no presentan ni antecedentes familiares (se puede ver en la Figura 3), ni ninguna de las enfermedades incluidas (depresión, hipertensión, diabetes, enfermedad cardiovascular y traumatismo craneal previo). Así que será necesario ver si, en concreto, los pacientes diagnosticados de Alzheimer son más propensos a sufrirlas.

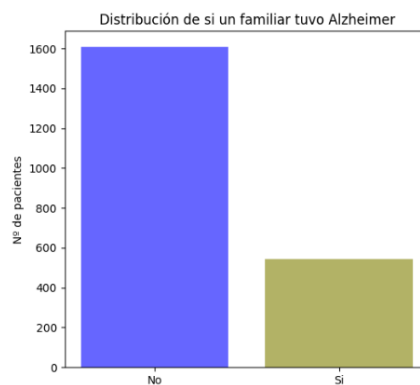


Figura 3: Distribución de los antecedentes familiares de Alzheimer de cada paciente.

- Mediciones clínicas.

En la presión arterial media se puede observar una distribución normal, habiendo un máximo entre 100-110 mmHg, que está en el rango de lo considerado normal, así que hay pocos pacientes que presenten valores muy altos o muy bajos como se puede ver en la Figura 4.

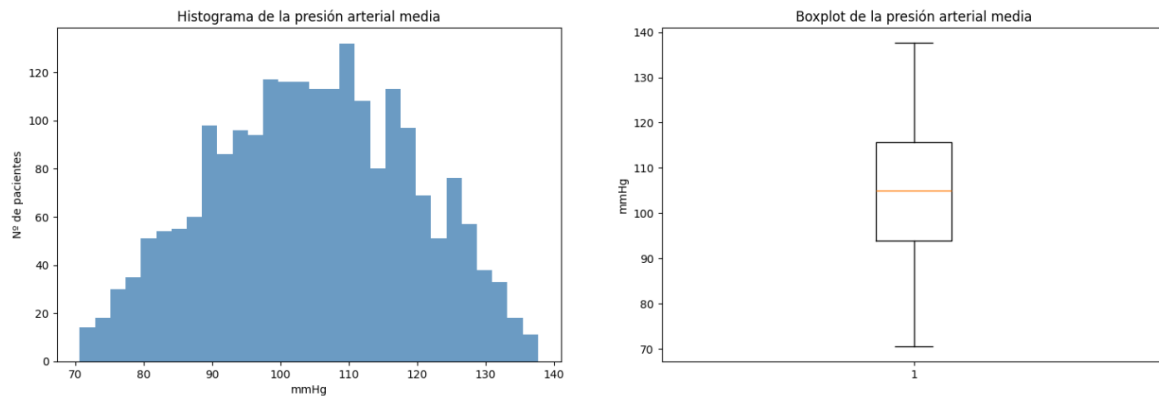


Figura 4: Histograma de la presión arterial media (izquierda) y boxplot (derecha).

Respecto al colesterol, se presenta una distribución uniforme, indicando que hay pacientes con un colesterol muy alto, pero no crea ninguna tendencia. La media es 225.19, siendo algo normal en personas de avanzada edad, y presenta una alta desviación estándar.

- Evaluaciones cognitivas y funcionales.

En MMSE (como se puede ver en la Figura 5), evaluaciones funcionales y actividades diarias, se ven puntuaciones que siguen una distribución uniforme, las medias son 14.75, 5.08, 4.98, respectivamente, que corresponden con la mitad de los rangos de puntuación de cada uno de ellos. Esto significa que hay pacientes con resultados muy bajos y que pueden estar relacionados con la enfermedad del Alzheimer, y pacientes con resultados muy buenos que podrían ser los no diagnosticados. También indica una variabilidad alta de los datos.

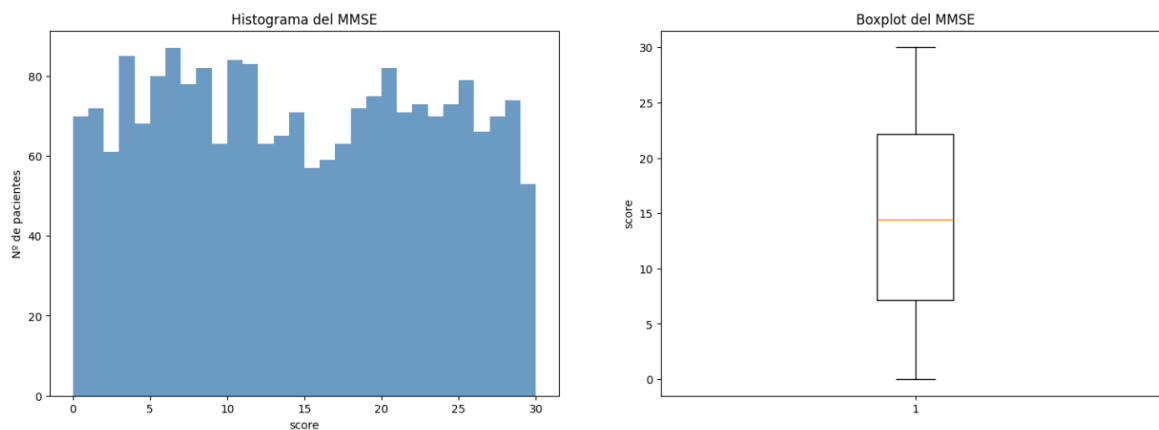


Figura 5: Histograma del MMSE (izquierda) y boxplot (derecha).

La mayoría de los pacientes no tienen quejas de memoria ni problemas de comportamiento, por lo que habrá que seguir investigando si las personas afectadas están diagnosticadas o no.

- Síntomas.

Con respecto a los síntomas, en todos ellos vemos que la mayoría de pacientes no los sufren. El síntoma que más se repite es la falta de memoria, y no llegan a la mitad de los pacientes, como se puede ver en

la Figura 6. En este caso también es interesante separar los diagnosticados de Alzheimer de los que no, para poder conocer los síntomas más comunes.

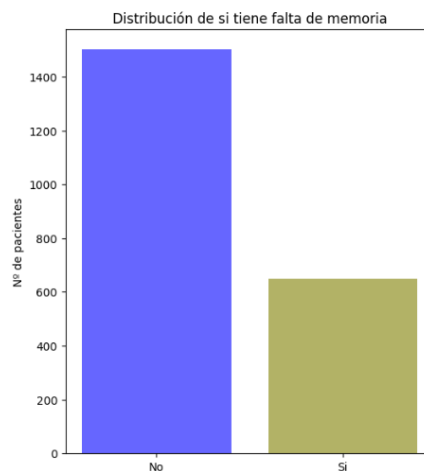


Figura 6: Distribución de la falta de memoria en los pacientes.

Como conclusiones generales de este análisis univariante podemos decir que, al mezclar pacientes diagnosticados con los no diagnosticados, nos lleva a distribuciones de las variables en su mayoría uniformes que nos indican una gran variabilidad entre los datos, que quizá con el futuro análisis bivalente puedan adquirir sentido y relación con la enfermedad del Alzheimer.

II. Bivalente

El análisis bivalente es capital en este estudio: nos indicará diferencias en todas las variables entre los pacientes diagnosticados de Alzheimer y los no diagnosticados. Todas las gráficas que se van a analizar pueden ser obtenidas en datos absolutos o relativos, siendo estos últimos más útiles, ya que, como se ha dicho, hay más pacientes no diagnosticados y no se estarían sacando las conclusiones correctas si se utilizan únicamente medidas absolutas.

En este apartado se verá si las respuestas uniformes previas se debían a la mezcla de ambos grupos y existen **relaciones significativas** entre ellas y el diagnóstico, o si, por el contrario, esto no sucede. Para comprobar la independencia de las variables se han utilizado: test de Mann-Whitney U (en variables numéricas sin distribución normal) y chi-cuadrado (entre variables categóricas). Esto puede ayudar a focalizar que variables son más relevantes para poder profundizar y sacar conclusiones.

- Variables demográficas vs diagnóstico.

Como el sexo de los pacientes era prácticamente equitativo entre hombres y mujeres, y la etnia podría haber introducido sesgo en el análisis, solo se realizó el análisis bivalente con la edad y el nivel educativo de los pacientes.

En primer lugar, se observó que las distribuciones de edades entre grupos (diagnosticados y no diagnosticados) eran prácticamente iguales, tanto a través de histograma, como de boxplot. Además, el p-valor fue mayor que 0.05, indicando que la relación entre las variables de edad y diagnóstico no es significativa. Es decir, sea cual sea la edad del paciente, no se relaciona con tener Alzheimer o no.

Después, se vio que la distribución de los niveles educativos en ambos grupos era similar, mostrado en la Figura 7, en el grupo de no diagnosticados había más pacientes con estudios superiores y menos sin estudios, coherente con los resultados obtenidos en el análisis univariante, pero no era una relación significativa, se obtuvo un $p\text{-valor} > 0.05$. Por esto podemos saber que el nivel educativo no influye en sufrir Alzheimer.

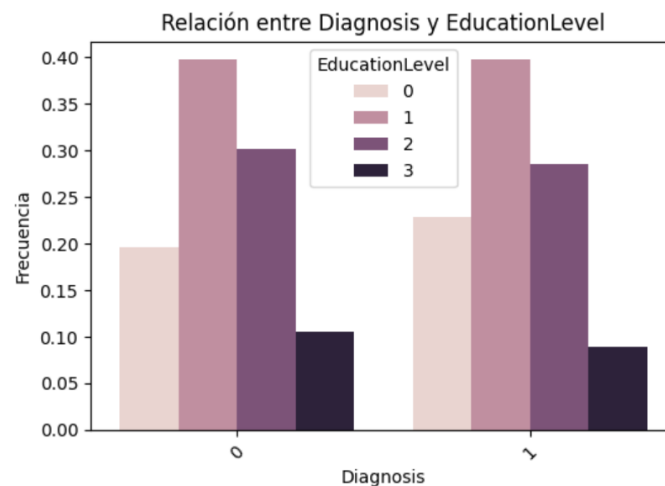


Figura 7: distribución de los niveles educativos según si el paciente está diagnosticado o no.

- Factores del estilo de vida vs diagnóstico.

Respecto al BMI, vemos una distribución similar para ambos grupos, salvo en los valores altos de esta variable, donde se forma claramente un pico en los pacientes diagnosticados de Alzheimer, como se puede ver en la Figura 8. Sin embargo, el $p\text{-valor}$ que se obtiene es mayor que 0.05, por lo que esta peculiaridad no es suficiente como para afirmar que el BMI tiene una relación significativa con el diagnóstico de la enfermedad.

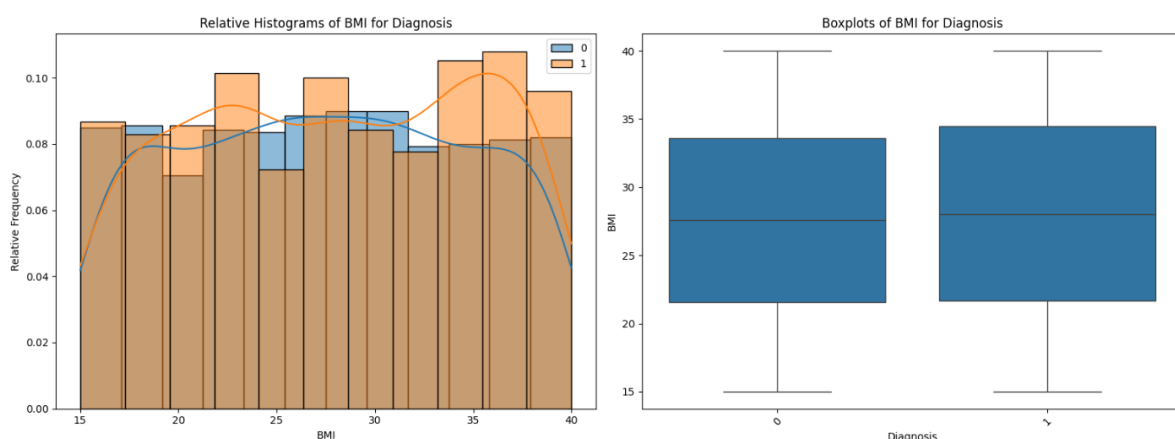


Figura 8: Histograma del BMI por grupos de pacientes diagnosticados o no (izquierda) y su respectivo boxplot (derecha).

A continuación, se observaron las distribuciones de fumadores en los pacientes, viendo que seguían la misma proporción en ambos grupos, y no presentaron relación significativa con el diagnóstico. Se podría decir que si el paciente fuma o no es independiente de la enfermedad.

Siguiendo con la distribución de la cantidad de alcohol consumido por los pacientes, las distribuciones que se presentan por grupos son bastante uniformes, y no llegan a ser significativas. En este caso, también se puede decir que esta variable no tiene relación con el diagnóstico de la enfermedad.

Las horas de actividad física que realizan los pacientes de ambos grupos por semana también tienden a seguir la misma distribución, e incluso se ve que hay más pacientes diagnosticados que hacen deporte, pero no es una relación significativa. Simplemente podría estar indicando que, al estar diagnosticados de la enfermedad, se les ha recomendado hacer más ejercicio y por eso está ligeramente por encima.

Algo similar pasa con la calidad de la dieta, aunque no se pueda asegurar que las variables tienen relación significativa, se observa que la tendencia de la puntuación de esta variable de los pacientes diagnosticados es ligeramente mejor que la de los no diagnosticados. Puede significar lo mismo que se comentaba previamente, que se les haya recomendado comer más sano y por eso están ligeramente por encima.

Finalmente, respecto a la calidad del sueño vemos una clara diferencia en las puntuaciones bajas en la Figura 9. Los pacientes que sufren Alzheimer duermen peor que los no diagnosticados, de media 6.9h y 7.12h, respectivamente. En este caso, las variables sí están significativamente relacionadas, puesto que el p-valor obtenido es 0.009, menor que 0.05.

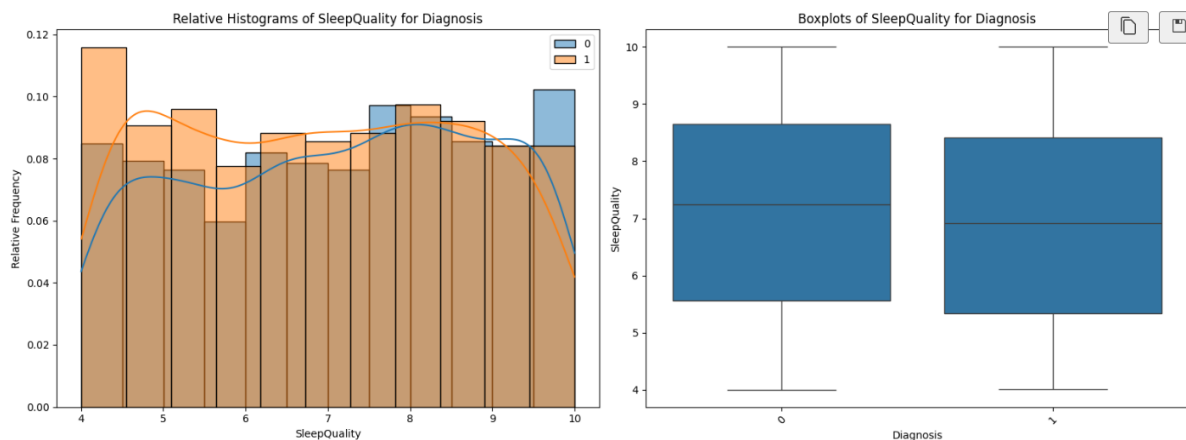


Figura 9: Histograma de la calidad del sueño por grupos de pacientes diagnosticados o no (izquierda) y su respectivo boxplot (derecha).

- Historial médico vs diagnóstico.

Todas las variables de este grupo son independientes del diagnóstico, puesto que no se han encontrado diferencias significativas en ninguno de los casos. Respecto a los antecedentes familiares no es algo impactante, puesto que no es hereditario, aunque la genética puede aumentar el riesgo [4], esto no sucede en los pacientes que componen este dataset. La diabetes tampoco influye en el diagnóstico, puesto que, de los pacientes con diabetes, la mayoría no tienen Alzheimer. Esto mismo sucede con la depresión, la mayoría de pacientes con depresión no tienen Alzheimer y no está significativamente relacionado. Ídem para los pacientes que han sufrido un traumatismo craneal, es independiente de sufrir la enfermedad.

Sin embargo, respecto a las enfermedades cardiovasculares y a la hipertensión, aunque sigue habiendo más personas no diagnosticadas de Alzheimer que las sufren, se puede ver que la proporción se ajusta algo más, como en la Figura 10. Aunque no llega a ser una relación significativa, el p-valor es de 0.11, y supone un cambio con respecto al resto de variables de esta categoría. Además, la hipertensión provoca enfermedades cardiovasculares, por lo que entre ellas también se relacionan.

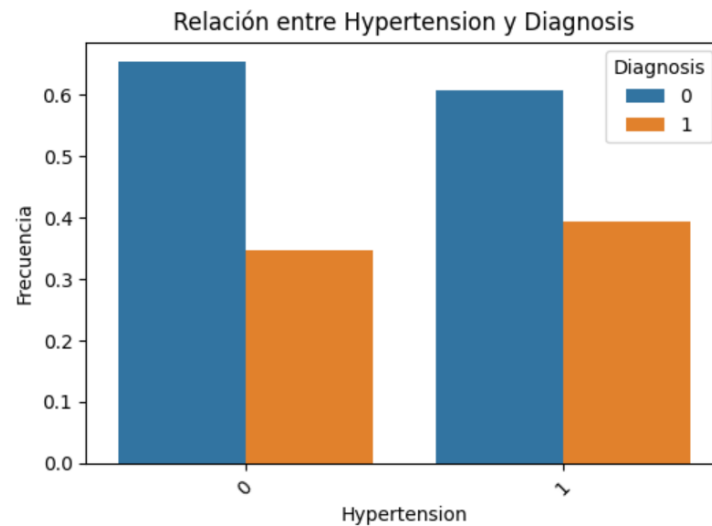


Figura 10: Distribución de la hipertensión entre pacientes diagnosticados y no diagnosticados.

- Mediciones clínicas vs diagnóstico.

Respecto a la tensión arterial media sigue una distribución normal en ambos grupos como se puede ver en la Figura 11, por lo que no se hallaron relaciones significativas. En este caso se utilizó t-student para obtener el p-valor, justamente por tener esta distribución.

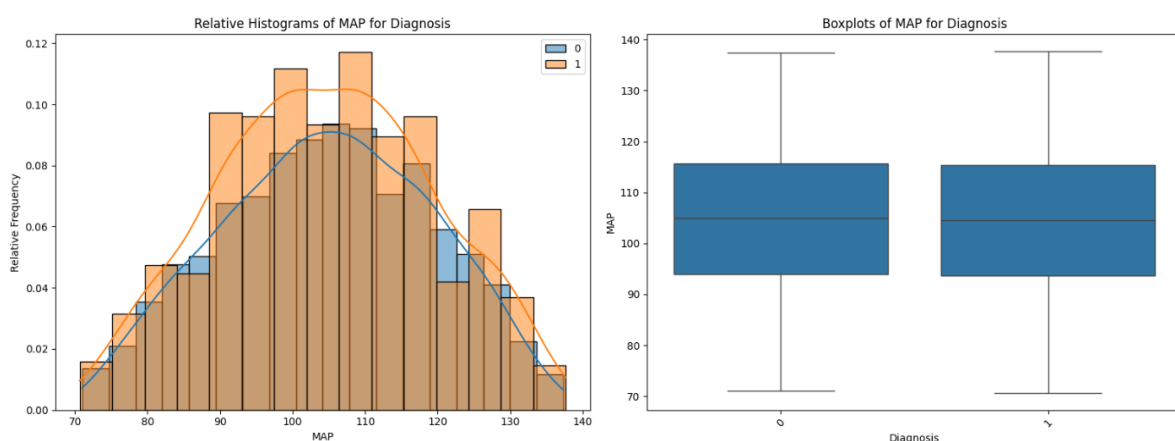


Figura 11: Histograma de MAP por grupos de pacientes diagnosticados o no (izquierda) y su respectivo boxplot (derecha).

En la otra medida clínica, el colesterol total, ambos grupos tenían distribuciones similares y uniformes, por lo que no se hallaron relaciones significativas que puedan relacionar que un paciente tenga mayor colesterol con el desarrollo del Alzheimer.

- Evaluaciones cognitivas y funcionales vs diagnóstico.

En todas las variables de esta categoría encontramos p-valores menores a 0.05, por lo que todas ellas están relacionadas de forma significativa con el diagnóstico de la enfermedad. Todo esto tiene sentido debido a que son métodos rápidos de cribaje muy utilizados, sencillos, y, según este estudio, íntimamente relacionados con el diagnóstico final.

Respecto al MMSE, en la Figura 12 se ve claramente que, a partir de la puntuación 23/30, el grupo con mejores notas es el de pacientes no diagnosticados, y los pacientes con Alzheimer no suelen alcanzar esas altas puntuaciones, indicando que algo está pasando. Sin embargo, en las notas bajas, el grupo dominante es el de los pacientes diagnosticados. El valor medio de MMSE de los no diagnosticados es 16.26, y el de los diagnosticados 11.99. Además, el p-valor es 6.62×10^{-28} , por lo que la relación entre variables es significativa. Un valor bajo en MMSE, puede indicar que el paciente sufre Alzheimer.

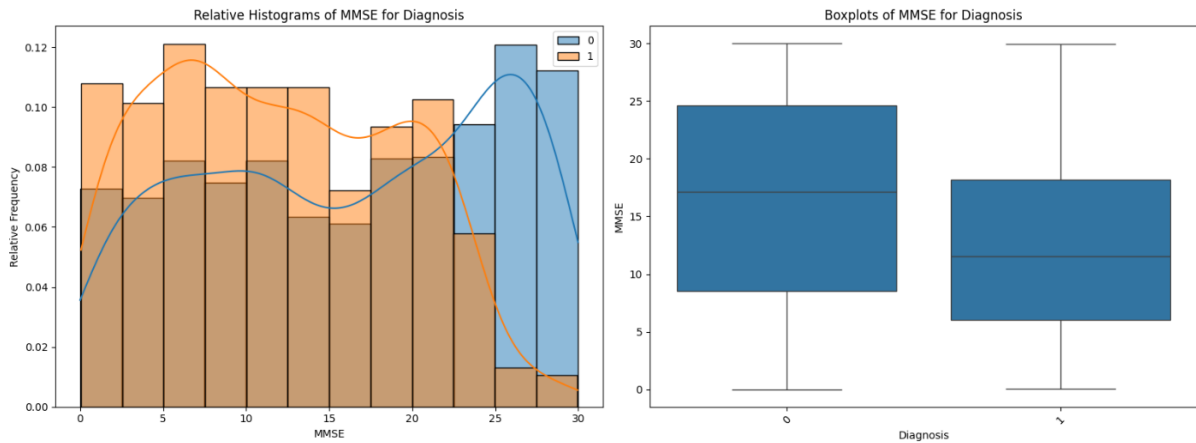


Figura 12: Histograma de MMSE por grupos de pacientes diagnosticados o no (izquierda) y su respectivo boxplot (derecha).

A continuación, con las evaluaciones funcionales sucede lo mismo. En las puntuaciones altas, el grupo dominante es el de no diagnosticados, con una media de 5.86, al contrario de lo que pasa con las puntuaciones bajas, en la que dominan los pacientes diagnosticados con una media de 3.65. Además, el cambio en la tendencia sucede en la puntuación 5, por lo que se podría hacer un binning de esta variable en ese punto para separar posibles pacientes que sufran la enfermedad. El p-valor obtenido fue 8.99×10^{-65} , por lo que la relación entre las variables, como era evidente, es significativa.

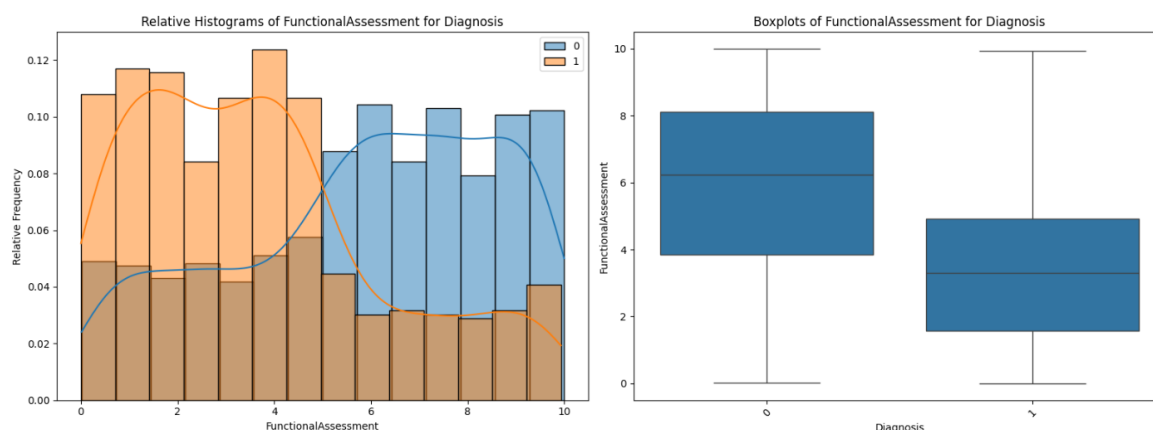


Figura 12: Histograma de evaluación funcional por grupos de pacientes diagnosticados o no (izquierda) y su respectivo boxplot (derecha).

Una situación similar sucede con la valoración de las actividades de la vida diaria, en la mitad de rango que corresponde al 5, se produce un cambio de tendencia. Las mejores puntuaciones son obtenidas por los pacientes no diagnosticados, como se preveía. En este caso también se obtiene un p-valor bajo, de 6.05×10^{-53} , que indica que las variables están significativamente relacionadas. La media de las puntuaciones de los no diagnosticados es de 5.7, y de los diagnosticados es de 3.65.

Para continuar, los pacientes que sufren las quejas de memoria y los problemas de comportamiento, son en mayor proporción los diagnosticados con Alzheimer, se ve el ejemplo de la variable de quejas de memoria en la Figura 13. Además, ambos p-valores son menores de 0.05, siendo 1.52×10^{-45} y 4.73×10^{-25} , respectivamente, por lo que tener problemas de memoria y comportamiento es propio de este grupo de pacientes de forma significativa.

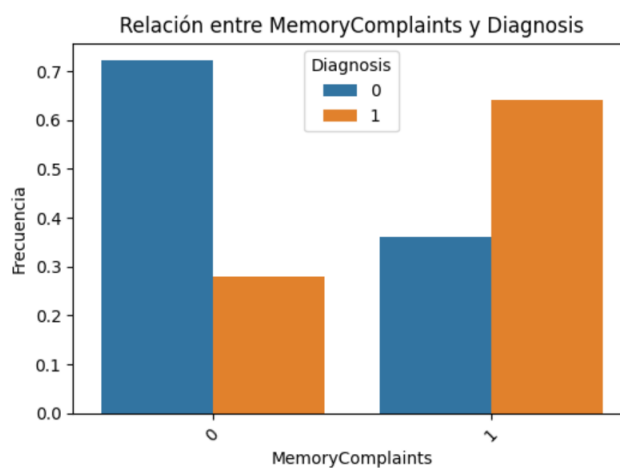


Figura 13: Distribución de las quejas de memoria entre pacientes diagnosticados y no diagnosticados.

En esta categoría se evidencia que las personas con Alzheimer obtienen peores puntuaciones en las evaluaciones cognitivas y funcionales, y que además sufren de memoria y problemas de comportamiento. Estas variables son usadas en el cribaje de los pacientes y, por ello, tienen sentido los resultados.

- Síntomas vs diagnóstico.

Todas las variables de esta categoría obtuvieron p-valores mayores que 0.05, lo que supone que no están significativamente relacionados con sufrir Alzheimer. En todos los casos (confusión, desorientación, cambios de personalidad, dificultad para hacer tareas y olvido), se obtiene una distribución similar, en los pacientes que las sufren, la mayoría son no diagnosticados. Con esto, no se puede decir que los síntomas tengan relación con la enfermedad.

Para seguir analizándolos y poder llegar a entenderlo, se comprobó si estos síntomas podrían estar relacionados con la edad, ya que podría ser una nueva hipótesis, y en ninguno de los casos se obtuvo una relación significativa. Por esto, no se puede decir que los pacientes que sufren estos síntomas tengan algo que ver con padecer la enfermedad.

Como conclusiones de este análisis bivalente tenemos que, al separar los grupos de pacientes por diagnóstico, se obtienen las respuestas a las preguntas iniciales. Se destaca la relación de la calidad del sueño y de las evaluaciones cognitivas y funcionales con la enfermedad del Alzheimer.

III. Multivariante

Este análisis ayudará a tener una visión general de todas las variables y las correlaciones que existen entre ellas. Es necesario porque al haber muchas columnas, al final se puede perder la perspectiva, y es interesante mirar todo en conjunto para completar las conclusiones del estudio.

Como se puede ver con el *heatmap* de las correlaciones en la Figura 14, las variables no están correlacionadas o tienen valores muy bajos. Las que más pueden relacionarse con el diagnóstico son las valoraciones cognitivas y funcionales, lo cual es coherente con los resultados obtenidos en el apartado anterior. La mayor correlación con el diagnóstico se halla en la variable evaluaciones funcionales con el valor -0.36 , que además es coherente con el análisis bivalente, en el que tenía un p-valor muy bajo. Además, son valoraciones que se suelen utilizar en cribaje y ayuda al diagnóstico, por lo que tiene sentido y era lo esperado. En cambio, aquí no se aprecia correlación con la calidad del sueño, que sí se había obtenido previamente, el valor en su matriz de correlación es -0.056 .

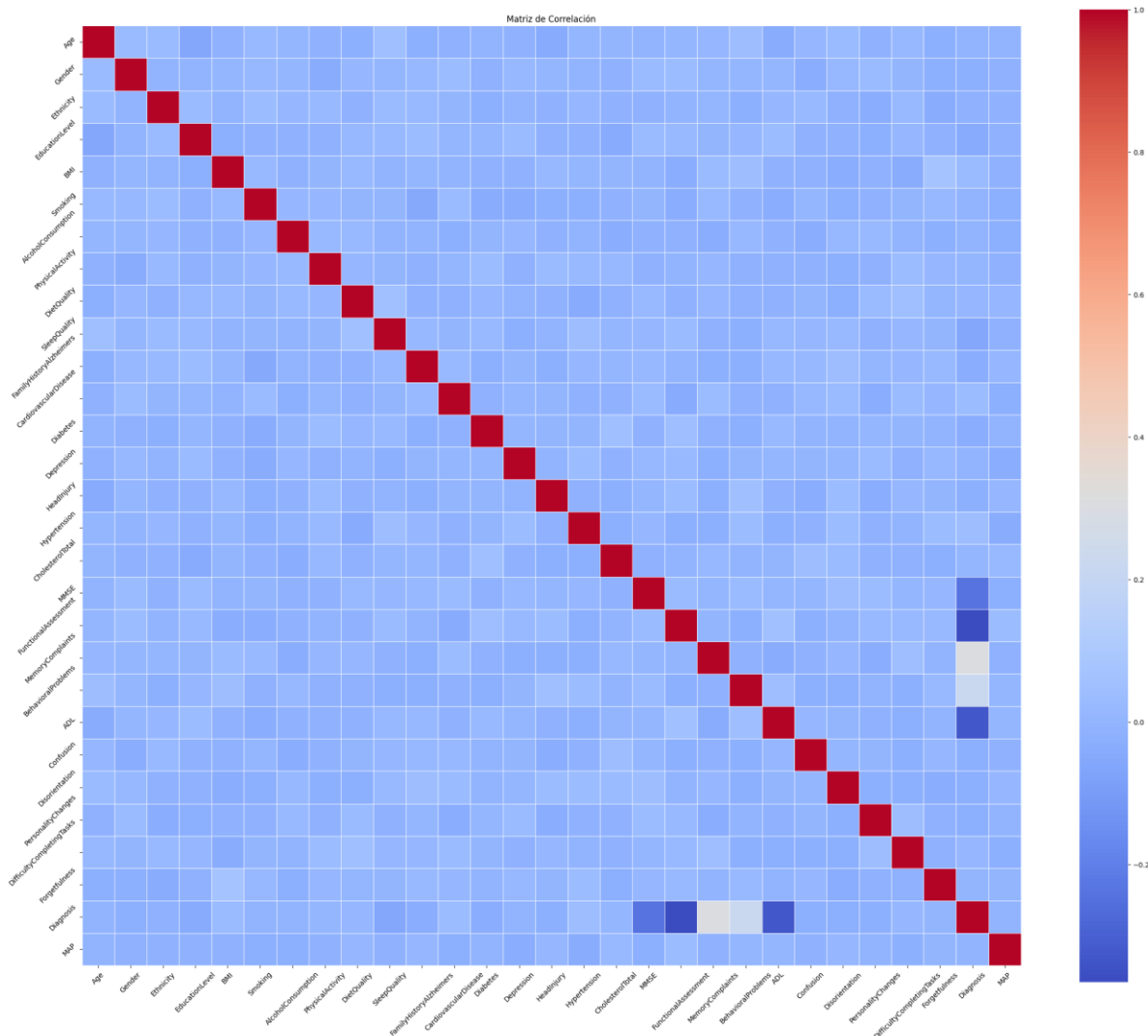


Figura 14: Representación de la matriz de correlación del dataframe mediante heatmap.

4. Conclusiones

- **Resumen de hallazgos clave:**

En primer lugar, en el análisis univariante se vio que los pacientes eran personas de edad avanzada con la mayoría de las variables uniformemente distribuidas y sin outliers, por lo que el análisis verdaderamente relevante fue el bivalente, donde se comparó cada variable con el target, que era el diagnóstico de la enfermedad.

En este, se encontró importancia significativa en la calidad del sueño, las personas con Alzheimer obtuvieron peores puntuaciones. Además, todas las evaluaciones cognitivas y funcionales propuestas fueron significativas, por lo que la puntuación obtenida va completamente relacionada con el diagnóstico de la enfermedad. Se evidencia la efectividad de su uso clínico para el cribaje de pacientes previo al diagnóstico. También deja claro que estos pacientes tienen más problemas de memoria y de comportamiento que los que no están diagnosticados.

Para verlo de una manera general, se realizó el análisis multivariante, donde volvió a visualizarse la importancia de estas valoraciones para el diagnóstico.

Con todo esto, el EDA ha servido para hallar un nuevo foco de investigación con la calidad del sueño de los pacientes, además de remarcar la importancia de las evaluaciones clínicas previas a un diagnóstico, por su rapidez y efectividad, y que, combinadas a otras técnicas, como la imagen médica, pueden llevar a cabo un diagnóstico preciso.

- **¿Se corroboraron las hipótesis iniciales?**

- Hipótesis 1: Los datos demográficos no tienen relación con el Alzheimer. La tendencia general que seguían era de personas mayores, caucásicas y con estudios secundarios.
- Hipótesis 2: Sobre el estilo de vida, la única variable que influye en el Alzheimer es la calidad del sueño, el resto no tuvieron una relación significativa.
- Hipótesis 3: Ninguna de las variables del historial médico influye en el desarrollo del Alzheimer, además la mayoría de los pacientes de este dataset no las cumplían.
- Hipótesis 4: Las mediciones clínicas no sirven para cribaje ni diagnóstico del Alzheimer, todo lo contrario, sucede con las evaluaciones cognitivas y funcionales. Son uno de los métodos más fiables para ayudar al diagnóstico, indicando las capacidades del paciente.
- Hipótesis 5: Los síntomas no tenían relación significativa con el desarrollo del Alzheimer, ni tampoco con la edad avanzada de los pacientes.

- **Análisis de limitaciones:**

Una de las principales limitaciones de este análisis es que los datos son sintéticos, debería de repetirse el análisis con datos reales. De este modo, se podrían trasladar de forma más fiable las conclusiones a situaciones del día a día. Además, se debería tener la misma o similar cantidad de datos tanto de pacientes diagnosticados de Alzheimer como de no diagnosticados, de forma que elimine posibles sesgos.

Por otra parte, este análisis no sirve por sí mismo para poder hacer un diagnóstico, podría juntarse a otras pruebas realizadas a estos pacientes, como a la imagen clínica. En conjunto se podrían crear modelos mucho más potentes que podrían, en el futuro, llevar a cabo diagnósticos de forma rápida y fiable.

5. Plan de acción

En primer lugar, se abre un nuevo campo de investigación con la importancia de la calidad del sueño. Se podría estudiar individualmente y ver hasta qué punto puede servir mejorarla en estos pacientes con Alzheimer. Además, si fuera algo efectivo en ellos, se podría invertir en dispositivos que midan calidad del sueño para tenerlo monitorizado y conseguir más datos que ayuden a las investigaciones. Podría llegar a ser una visión relativamente sencilla e innovadora.

Por otra parte, se realiza la importancia de las evaluaciones cognitivas y funcionales, por eso son las más usadas actualmente. Se demuestra que son certeras y, además, son rápidas y fáciles de hacer para el doctor, ya no solo en el Alzheimer, si no en otras enfermedades que afecten a la cognición, como podría ser la demencia.

Los pasos futuros del análisis del Alzheimer consisten en realizar el análisis con datos de pacientes reales, y poder combinarlo con otras medidas ya mencionadas. Además, la construcción de modelos predictivos para comprobar si estos pueden ayudar al doctor a realizar diagnósticos y reducir el tiempo y tasa de error. También podría estudiarse si se pueden distinguir grados de severidad del Alzheimer en los pacientes.

En definitiva, es un análisis que ha respondido a las preguntas planteadas inicialmente. Ha ofrecido respuestas esperadas sobre las evaluaciones de los pacientes, pero también sorprendentes como la relación con la calidad del sueño. También se debe tener en cuenta que las variables sin relación significativa también son capaces de aportar mucho, como la no relación de los síntomas mencionados con el Alzheimer.

6. Referencias

1-Ministerio de Sanidad. *Plan Integral de Alzheimer y otras Demencias 2019*. Recuperado de: https://www.sanidad.gob.es/profesionales/saludPublica/docs/Plan_Integral_Alzheimer_Octubre_2019.pdf

2-Naciones Unidas. *Global Issues: Ageing*. Recuperado de: <https://www.un.org/es/global-issues/ageing#:~:text=Esto%20significa%20que%20se%20prev%C3%A9,ni%C3%B1os%20menores%20de%2012%20a%C3%B1os>.

3-Alzheimer's Association. *Diagnóstico de Alzheimer y demencias*. Recuperado de: <https://www.alz.org/alzheimer-demencia/diagnostico>

4-Fundación Pasqual Maragall. *Mini Mental Test*. Recuperado de: <https://blog.fpmaragall.org/mini-mental-test>