

# Análisis exploratorio de datos

...

# Conjunto de Datos de Salud y Prescripciones de Pacientes con Demencia.

Este conjunto de datos sirve como una fuente rica para el análisis, proporcionando una visión multifacética de los factores que pueden contribuir al inicio y la progresión de la demencia.

Es un recurso valioso para los investigadores que desean explorar la compleja interacción entre el estilo de vida, la genética y los resultados de salud.

Proveniente de PUBMED, fuentes de investigación en línea, NHS, Google Scholar y consultas con profesionales de la salud.

	Diabetic	AlcoholLevel	HeartRate	BloodOxygenLevel	BodyTemperature	Weight	MRI_Delay	Prescription	Dosage in mg	Age	...	Smoking_Status	APOE_ε4
0	1	0.084974	98	96.230743	36.224852	57.563978	36.421028	NaN	NaN	60	...	Current Smoker	Negative
1	0	0.016973	78	93.032122	36.183874	56.832335	31.157633	Galantamine	12.0	61	...	Former Smoker	Positive
2	0	0.009000	89	93.566504	37.326321	59.759066	37.640435	NaN	NaN	69	...	Former Smoker	Negative
3	0	0.086437	60	93.906510	37.030620	58.266471	50.673992	Donepezil	23.0	78	...	Never Smoked	Negative
4	1	0.150747	67	97.508994	36.062121	67.705027	27.810601	Memantine	20.0	77	...	Never Smoked	Positive
...	...	...	...	...	...	...	...	...	...	...	...	...	...
995	1	0.081825	87	93.851963	36.495134	50.380106	42.318663	Donepezil	10.0	88	...	Never Smoked	Positive
996	1	0.145249	97	94.522391	36.270804	94.006484	52.812568	NaN	NaN	80	...	Never Smoked	Negative
997	1	0.073692	65	98.578388	37.065703	80.088613	13.640229	NaN	NaN	67	...	Never Smoked	Positive
998	0	0.037347	71	91.298580	37.037202	95.322210	17.445715	Memantine	20.0	62	...	Never Smoked	Positive
999	0	0.085969	90	95.522825	36.026748	57.671455	30.011838	NaN	NaN	80	...	Never Smoked	Positive
1000 rows × 24 columns													
	Physical_Activity	Depression_Status	Cognitive_Test_Scores	Medication_History	Nutrition_Diet	Sleep_Quality	Chronic_Health_Conditions	Dementia					
	Sedentary		No	10	No	Low-Carb Diet	Poor	Diabetes	0				
	Moderate Activity		No	1	Yes	Low-Carb Diet	Poor	Heart Disease	1				
	Moderate Activity		No	8	No	Mediterranean Diet	Poor	Heart Disease	0				
	Mild Activity		Yes	5	Yes	Balanced Diet	Poor	Hypertension	1				
	Mild Activity		No	0	Yes	Low-Carb Diet	Good	Diabetes	1				
	...		...	...	...	...	...	...	...				
	Mild Activity		No	5	No	Balanced Diet	Poor	Diabetes	1				
	Moderate Activity		No	9	Yes	Low-Carb Diet	Poor	Diabetes	0				
	Sedentary		No	8	Yes	Balanced Diet	Good	Diabetes	0				
	Sedentary		Yes	2	No	Low-Carb Diet	Good	NaN	1				
	Mild Activity		No	10	Yes	Mediterranean Diet	Good	Heart Disease	0				

# Datos Numéricos

- Nivel de Alcohol
- Frecuencia Cardíaca
- Nivel de Oxígeno en Sangre
- Temperatura Corporal
- Peso
- Retraso de la RMN (tiempo de retraso para obtener una resonancia magnética)
- Dosis en mg
- Edad
- Puntuaciones de Pruebas Cognitivas

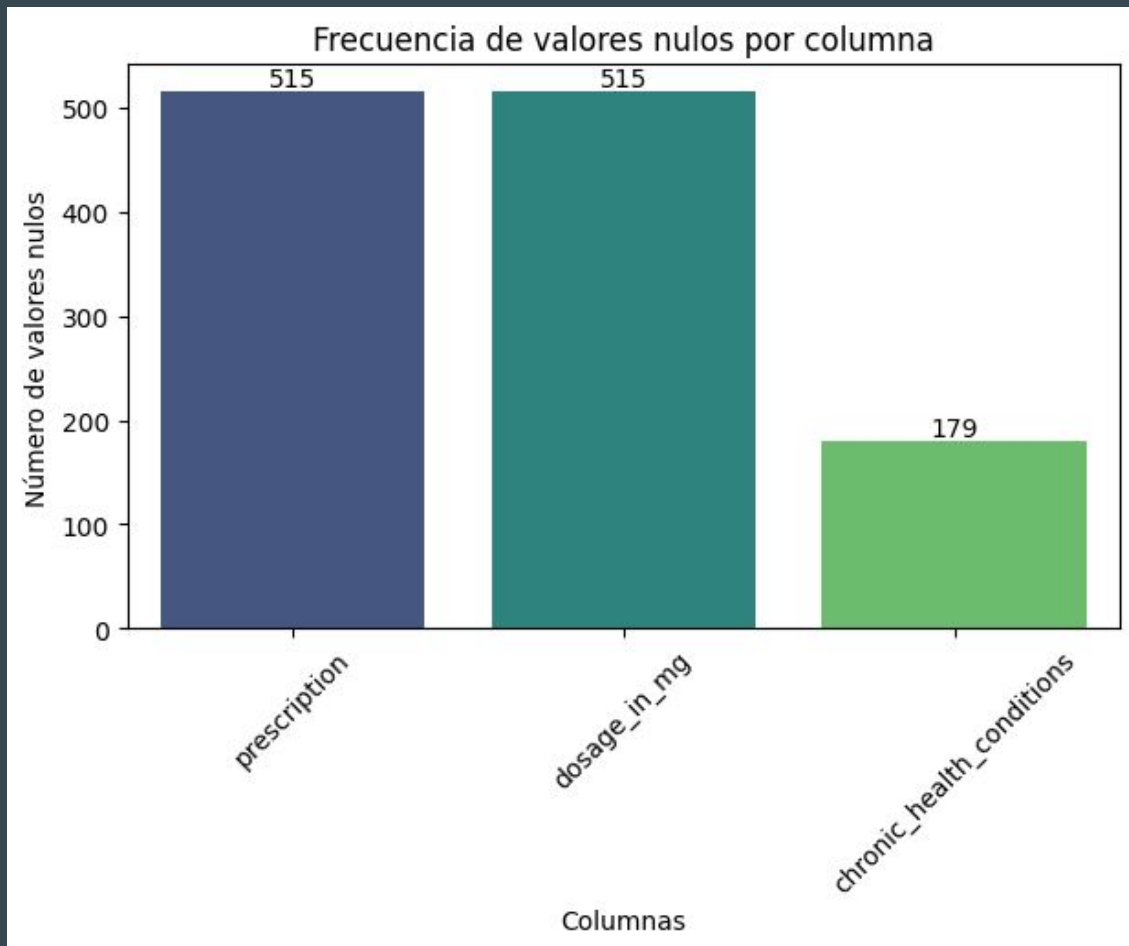
# Datos Categóricos

- Diabético
- Prescripción
- Nivel de Educación
- Mano Dominante
- Género
- Antecedentes Familiares (*antecedentes familiares de demencia*)
- Estado de Tabaquismo
- APOE\_ε4 (*presencia de variante genética asociada con el Alzheimer*)
- Actividad Física
- Estado de Depresión
- Historial de Medicación
- Dieta Nutricional
- Calidad del Sueño
- Condiciones Crónicas de Salud
- Demencia

# Missing Values

Al iterar por cada atributo del conjunto de datos, se encontraron valores faltantes en los atributos:

- prescription
- dosage\_in\_mg
- chronic\_health\_conditions



Tras analizar el origen de los valores faltantes detectados se verificó si existía alguna relación entre las variables.

```
## Verificar si hay dosis nulas donde hay prescripción
missing_dosage = m_df[m_df['prescription'].notnull() & m_df['dosage_in_mg'].isnull()]
print(f"¿Existen dosis nulas en caso de prescripción?:\n{(not missing_dosage.empty)}")
```

```
¿Existen dosis nulas en caso de prescripción?:
False
```

Considerando que no existen dosis nulas en caso de prescripción, entendemos que los valores nulos corresponden al caso de no prescripción.

También consideramos que en caso de valor nulo para *chronic\_health\_conditions*, el individuo no presenta ninguna condición crónica.

Es por esto que tomamos la decisión de:

- Convertir los valores nulos a 'None' para los casos de *prescription* y *chronic\_health\_conditions*.
- Dejar los valores nulos en el caso de *dosage\_in\_mg*, para evitar que al agregar algún valor nos genere ruido a la hora de realizar estadísticas.

# Outliers

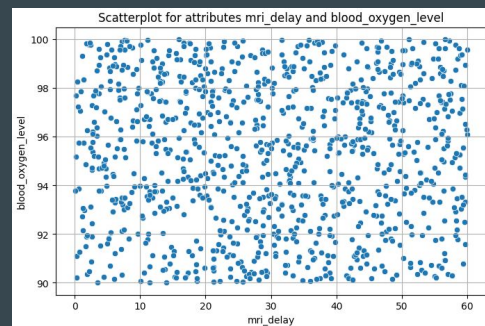
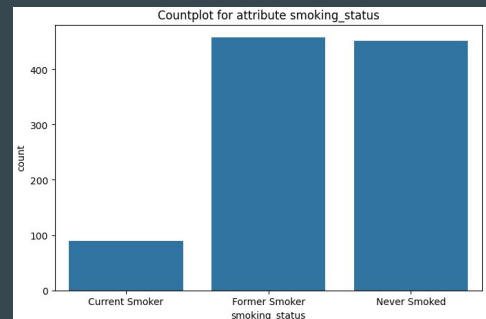
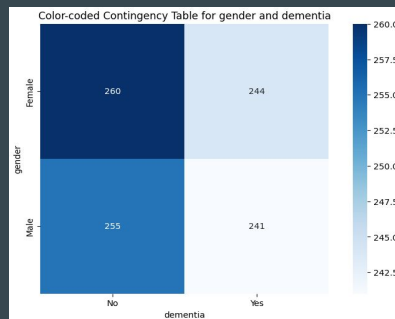
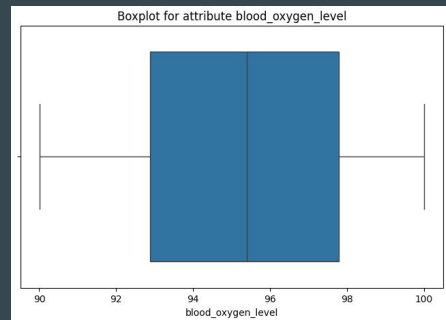
Para la detección de outliers se realizaron análisis univariado y bivariado.

## *Análisis Univariado*

- Variables cuantitativa: boxplots
- Variables cualitativas: gráficos de barras

## *Análisis Bivariado*

- Numérico-numérico: scatterplots
- Categórico- numérico: boxplots
- Categórico-categórico: tablas de contingencia

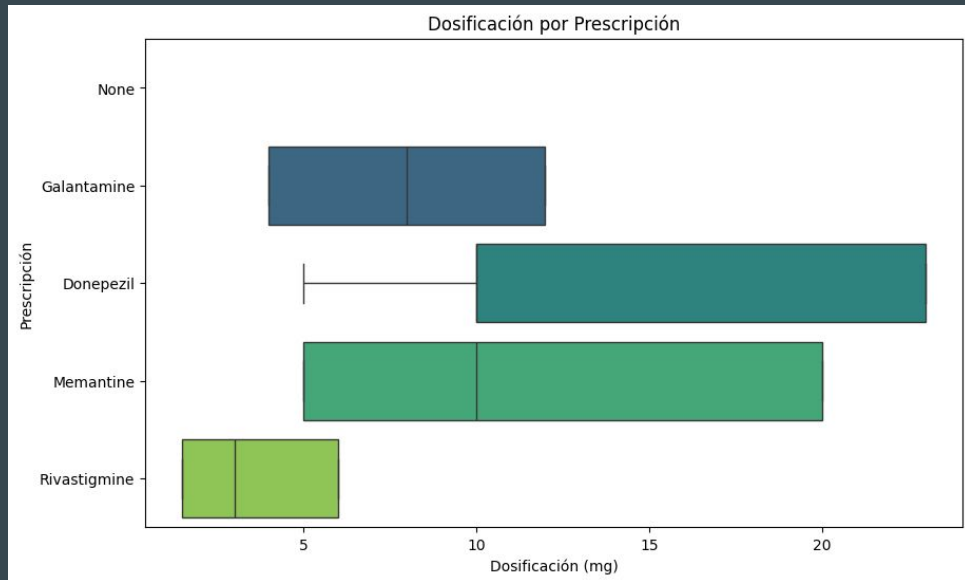
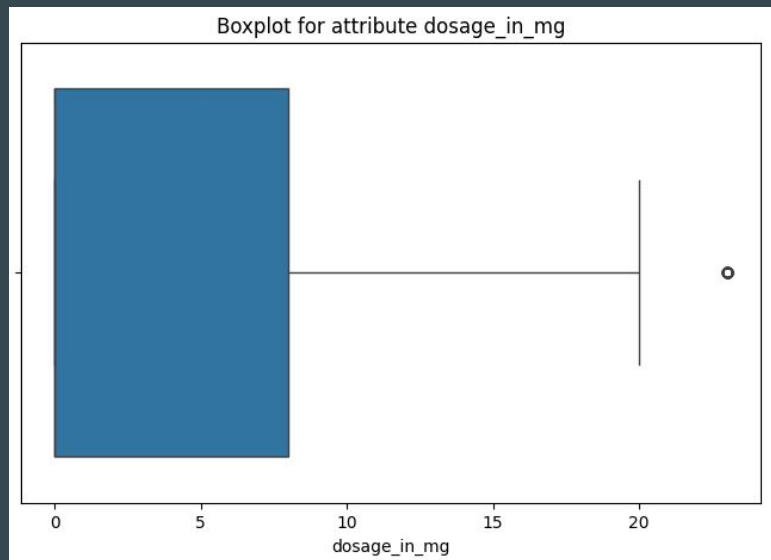




# Análisis Univariado

En el análisis univariado sólo se registran outliers para el atributo *dosage\_in\_mg*.

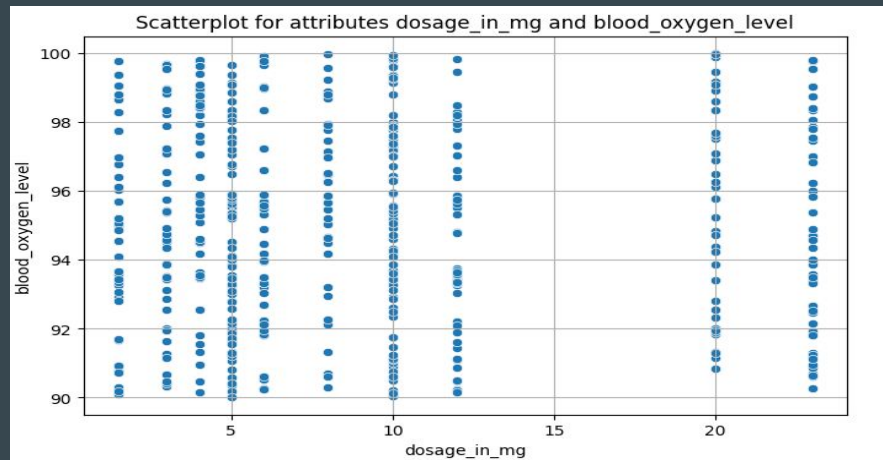
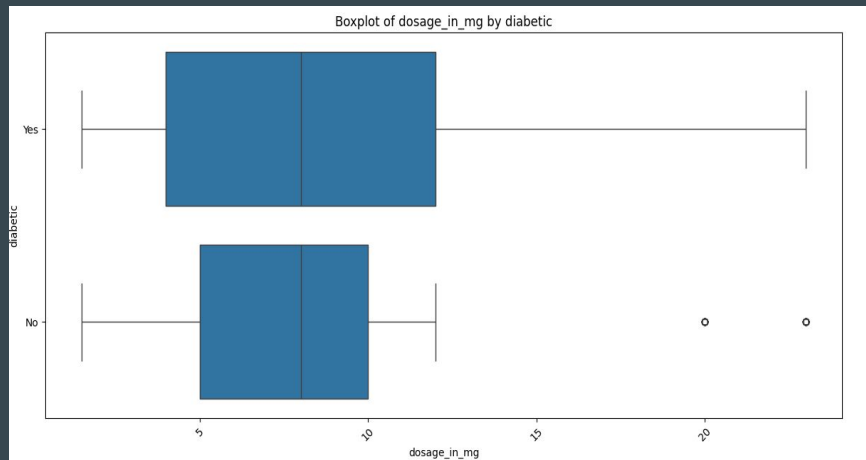
En este caso los outliers no parecen ser valores erróneos, dado que en relación con el atributo *prescription*, parece estar dentro de los valores esperados para la prescripción de 'Donepezil'.



# Análisis Bivariado

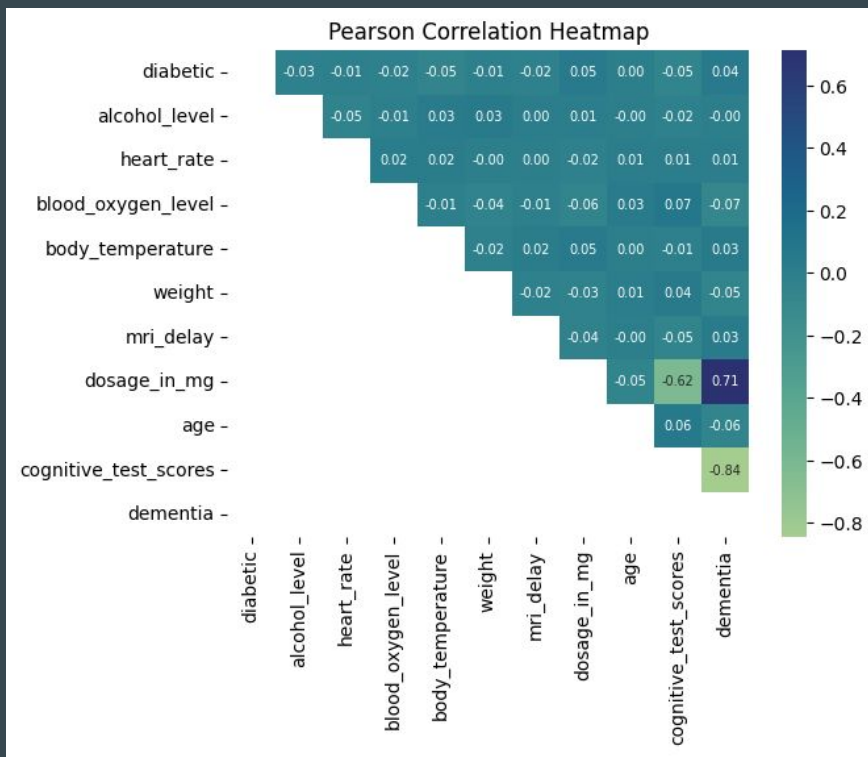
En el análisis Bivariado, se registraron outliers para algunas columnas en relación con `dosage_in_mg`.

Al igual que en el caso anterior, los valores de dosis para la prescripción de Donepezil se encuentran por fuera de los valores esperados, en comparación con las demás prescripciones.



# Correlación

Para hacer un análisis de correlación, se filtraron los datos numéricos y se obtuvieron los siguientes valores graficados en un mapa de calor:



- **Correlación negativa** entre:
  - *dementia*
  - *cognitive\_test\_scores*.
- **Correlación negativa** entre:
  - *dosage\_in\_mg*
  - *cognitive\_test\_scores*.
- **Correlación positiva** entre:
  - *dementia*
  - *dosage\_in\_mg*.

# Correlación

- **Correlación negativa** entre *dementia* y *cognitive\_test\_scores*.
- **Correlación negativa** entre *dosage\_in\_mg* y *cognitive\_test\_scores*.
- **Correlación positiva** entre *dementia* y *dosage\_in\_mg*.

Analizando los datos, encontramos que existe una relación directa entre los atributos *dementia*, *prescription* y *dosage\_in\_mg*.

Ya que en todos los casos de *dementia* positiva, existe una prescripción y por ende una dosis asociada.