

Análisis del Rendimiento Cognitivo Humano

Basado en diferentes factores de estilo de vida

Data Scientist :

Nicolas Felipe Mogollon Granda

Tipo de proyecto :

investigativo

Introduccion:

El estilo de vida moderno ha provocado cambios significativos en la salud cognitiva de las personas, especialmente en jóvenes y adultos. Factores de la vida diaria como la duración del sueño, los niveles de estrés, la alimentación, el tiempo frente a pantallas, el ejercicio físico y el consumo de cafeína se dice que podrían influir notablemente en habilidades del rendimiento cognitivo general y

otros factores como la memoria, el tiempo de reacción también podrían influir.

Este proyecto, desde un enfoque social, busca identificar qué variables del estilo de vida o habilidades de memoria y motricidad mental tendrían una relación estadísticamente significativa con el rendimiento cognitivo de un individuo. Para ello, se aplicarán técnicas estadísticas como el análisis de varianza (ANOVA), el coeficiente de regresión y la correlación. El objetivo es evidenciar qué aspectos, variables impactan más en el desarrollo de las capacidades cognitivas.

Escenario : Social

El proyecto se enmarca en un contexto social. Comprender e identificar las variables que afectan la salud mental y el bienestar general de las personas es fundamental. Detectar los factores asociados a permitirá proponer medidas colectivas orientadas a mejorar la calidad de vida de la población.

Pregunta De Investigacion:

¿Cuales factores tendrian relacion con el rendimiento cognitivo ?

Justificación del Estudio:

La relevancia de esta investigación radica en la necesidad de reconocer qué podría incidir en el rendimiento cognitivo de las personas. Con esta información, será posible desarrollar intervenciones y propuestas de mejora enfocadas en los factores influyentes, fomentando así una mejor salud cognitiva en la sociedad.

Analisis Tabla De Contenido

Dataset	5
Limpieza y Preparación de Datos	5
Analisis bi-variado --- Cognitive score	6
Age-Cognitive score	6
Sleep duration-Cognitive score	7
Daily screen time-Cognitive score	8
Caffeine-intake-Cognitive score	9
Reaction time – Cognitive score	10
Memory test score-Cognitive score	11
Conclusion Bi variado graficas	12
Correlacion	13
Validacion de supuestos--- Anova	15
Normalidad:	15
Conclusion normalidad.....	18
Homogeniedad de varianzas.....	19
ANOVA	20
Two-Way ANOVA.....	21
ANOVA Welch.....	23
Coeficientes De Regresion	24
Interpretación de los Coeficientes de Regresión	25
Validacion Del Modelo de Regresion	27
Prueba De Hipotesis “Coeficientes De Regresion”	28
METODOLOGIA	30
Perspectiva Adoptada para el Análisis	30
¿El producto final da solución al problema planteado inicialmente?	33
LINK.....	33
CONCLUSION ANALISIS	34

Dataset

Age	Gender	Sleep_Duration	Stress_Level	Diet_Type	Daily_Screen_Time	Exercise_Frequency	Caffeine_Intake	Reaction_Time	Memory_Test_Score	Cognitive_Score
57	Female	6.5	3	Non-Vegetarian	6.5	Medium	41	583.33	65	36.71
39	Female	7.6	9	Non-Vegetarian	10.8	High	214	368.24	58	54.35
26	Male	8.2	6	Vegetarian	5.7	Low	429	445.21	49	32.57
32	Male	7.8	9	Vegetarian	8.3	Low	464	286.33	94	70.15
50	Male	9.7	2	Non-Vegetarian	11.3	Medium	365	237.65	62	87.54
...
55	Male	7.8	9	Non-Vegetarian	9.4	Low	157	312.34	55	50.80
35	Female	9.8	6	Vegetarian	6.5	Medium	130	290.17	86	94.89
32	Female	7.3	10	Non-Vegetarian	7.5	Low	462	249.57	72	63.51
34	Male	8.7	8	Vegetarian	7.2	Medium	96	319.48	82	81.43
41	Female	7.9	7	Vegetarian	2.0	Medium	27	263.21	42	80.39

80000 rows × 13 columns

“los datos provienen de kaggle el dataset de estudio en cuestion se llama [Human Cognitive Performance Analysis](#)”

Limpieza y Preparación de Datos

Volumen de datos: El conjunto analizado contiene 80,000 registros y 11 variables

Datos completos: No se identificaron valores faltantes, garantizando la integridad de la información.

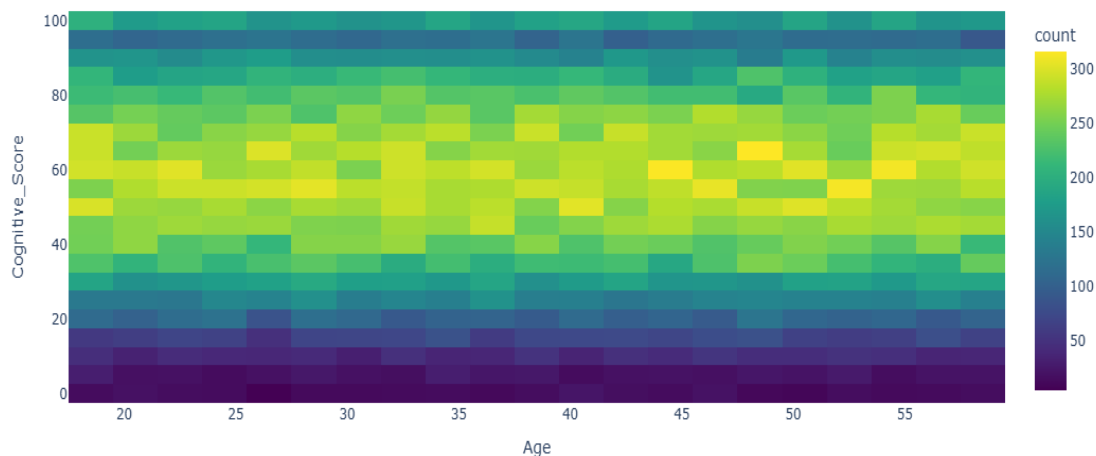
Calidad de registros: Se confirmó que no existen filas duplicadas, asegurando que cada entrada sea única y confiable.

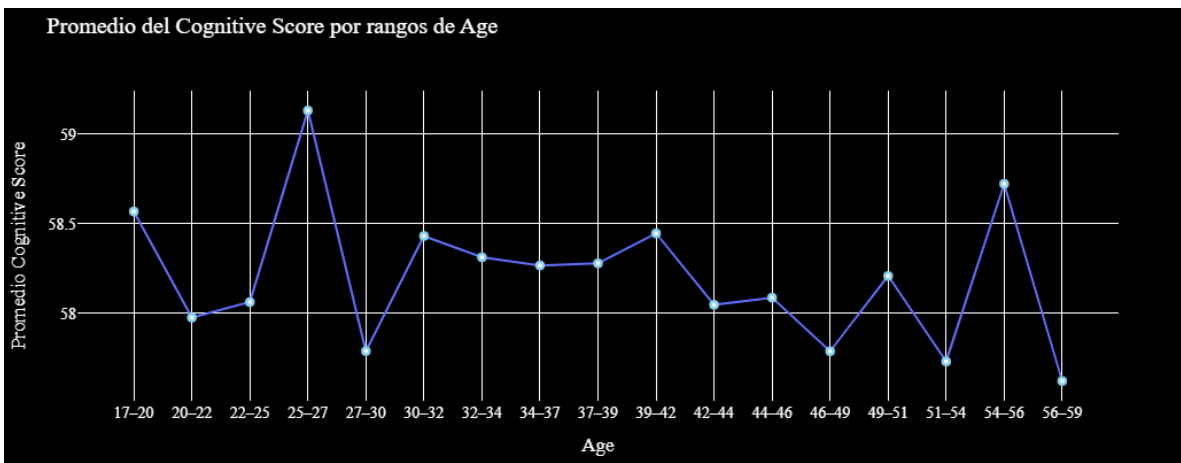
Enfoque en variables relevantes: Se eliminaron las columnas "user_id" y "AI_Predicted_Score" por no aportar al objetivo del estudio.

Analisis bi-variado --- Cognitive score

En esta etapa del análisis, examinamos cada variable individualmente en relación con el rendimiento cognitivo (*cognitive score*). Para ello, realizamos gráficas de densidad y scatterplots (agrupados por intervalo y promedio debido a la cantidad de datos) comparando la variable independiente (X) y la variable dependiente de interés (*cognitive score*, Y). El objetivo fue identificar posibles patrones, tendencias que sugirieran una relación significativa entre las variables analizadas y el desempeño cognitivo

Age-Cognitive score

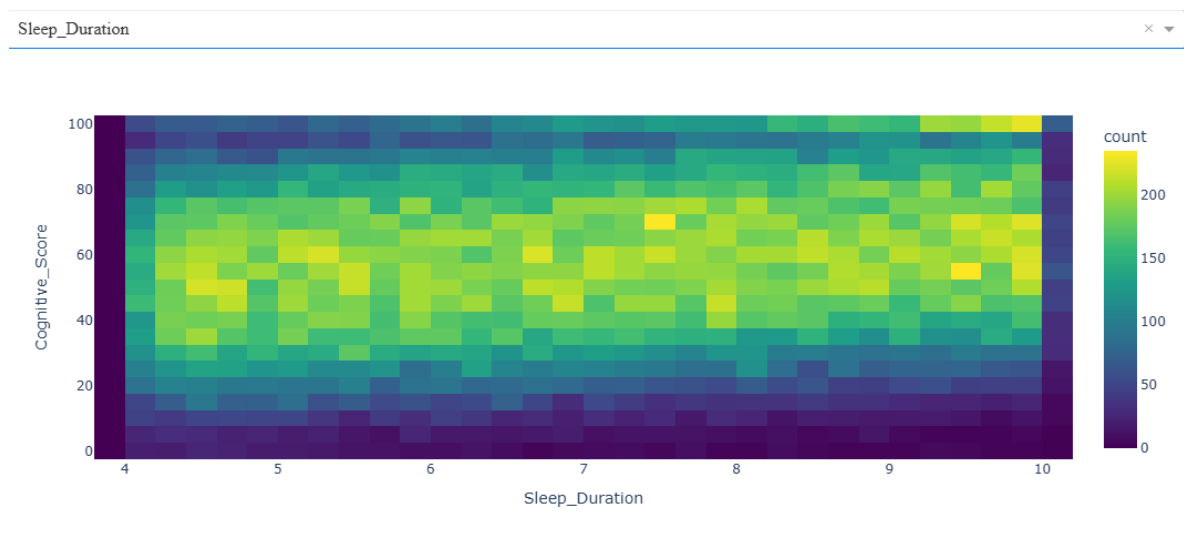


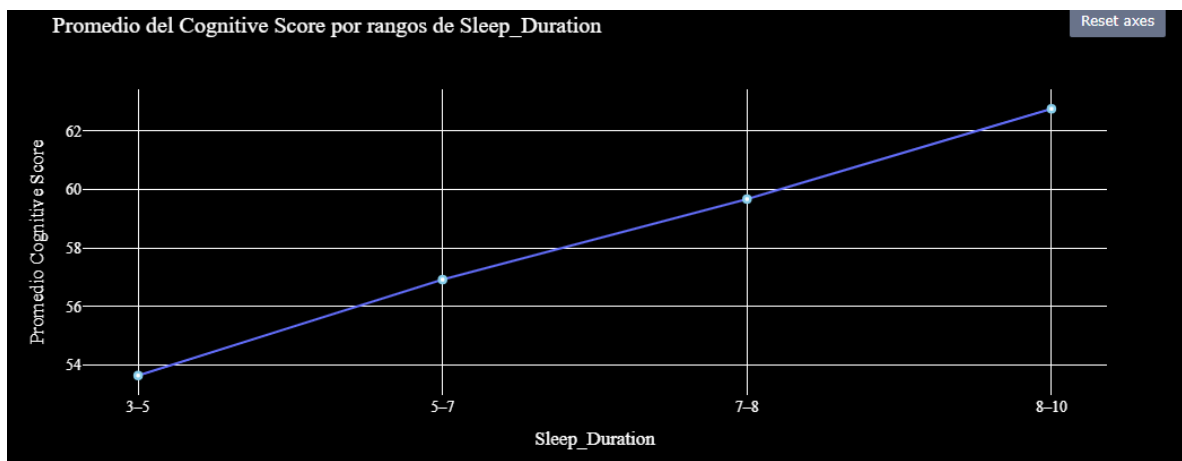


En el grafico de densidad miramos que entre estas dos variables no se ve relacion alguna ,no se muestra o encuentra una tendencia o insight claro por el cual discernir como la edad afectaria el cognitivo, mirando la edad por intervalos nos damos cuenta que del rango de 25-27 años hay un promedio cognitivo muy bueno, de 30-42 años se mantiene consistente el promedio de puntuacion cognitivo de las personas,y por ultimo habria un deteriro cognitivo en el rango de edad de 42-54 años

“el rango de puntuacion cognitivo fue de un maximo de 60 y un minimo de 57 ”

Sleep duration-Cognitive score



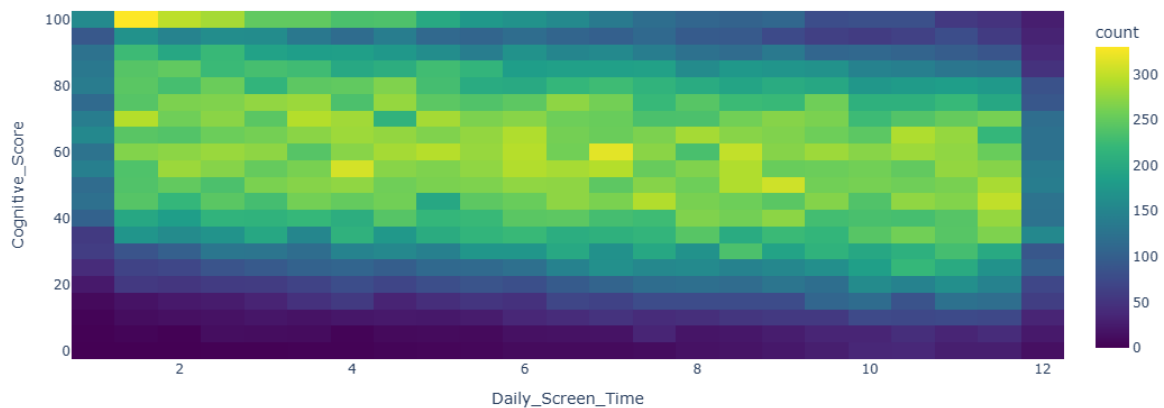


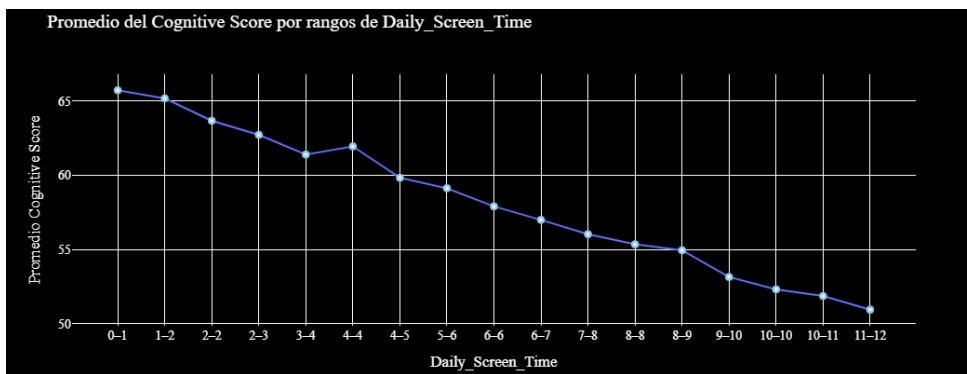
Aquí si se puede ver claro como las horas de sueño afectarían al puntaje del cognitive score si miramos a detalle la grafica de densidad miramos que hay un insight claro a partir de las 9-10 horas de sueño las personas que caen en este intervalo tienen mejor rendimiento cognitivo aunque la relacion no es muy fuerte si no mas bien tenue hay un gran incremento cognitivo justo en este intervalo , ya si miramos la grafica

scatter nos confirma lo que vimos anteriormente y fue en que entre mas horas de sueño mejor cognitivo “hay una linealidad entre la relacion de estas 2 variables “

“el rango de puntuacion cognitivo fue de un maximo de 63 y un minimo de 53 ”

Daily screen time-Cognitive score

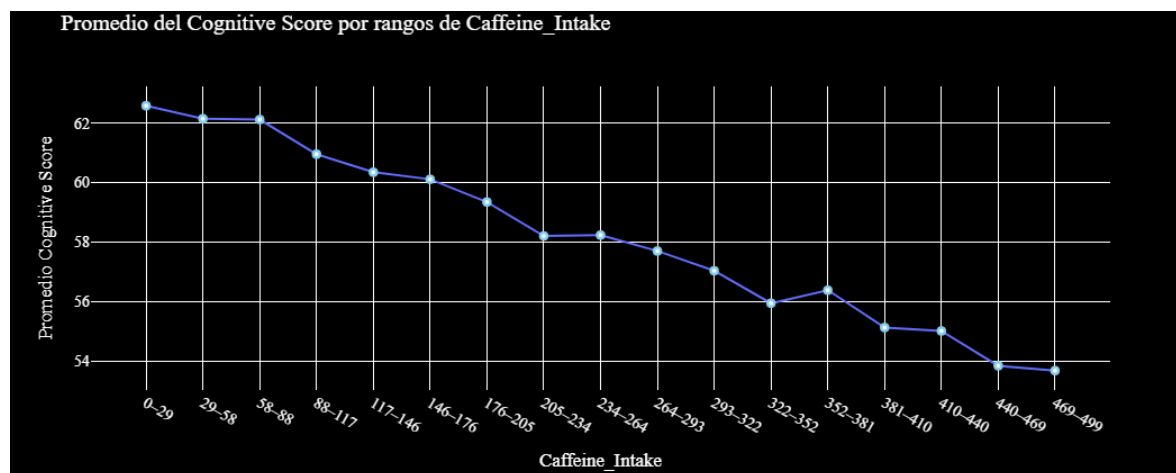
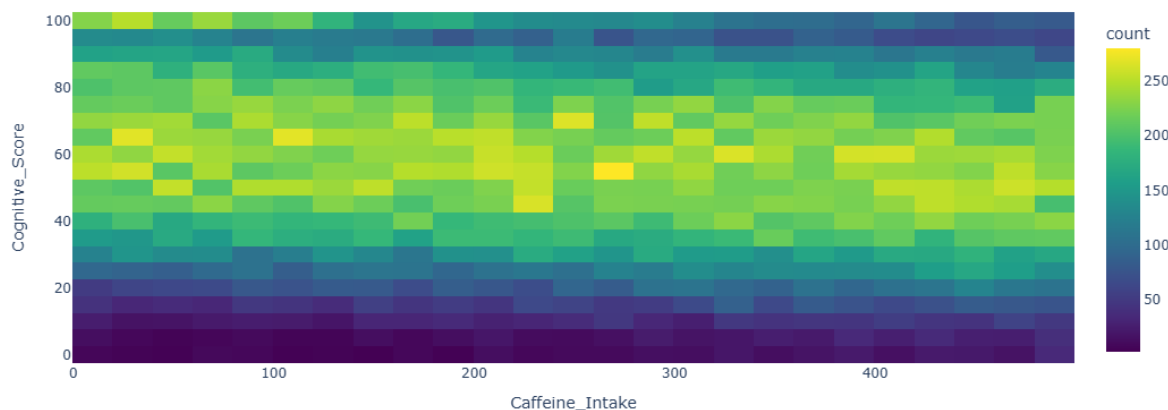




Aquí hay una relacion muy buena entre el tiempo en pantalla en dispositivos tecnologicos y el rendimiento cognitivo, las personas que solo tienen un máximo de 2 horas diarias tienen un puntaje cognitivo alto ya este rendimiento se va deteriorando conforme se le dedique mas tiempo a estos dispositivos, en el grafico de densidad se ve que a partir de las 9 horas hay una gran diferencia en rendimiento con los que apenas usan estos dispositivos solo 2 horas como máximo, ya promediando el puntaje por intervalo de tiempo en pantalla (grafico scatter) vemos mas claramente esta relacion.

“el rango de puntuacion cognitivo fue de un máximo de 65 y un mínimo de 51 ”

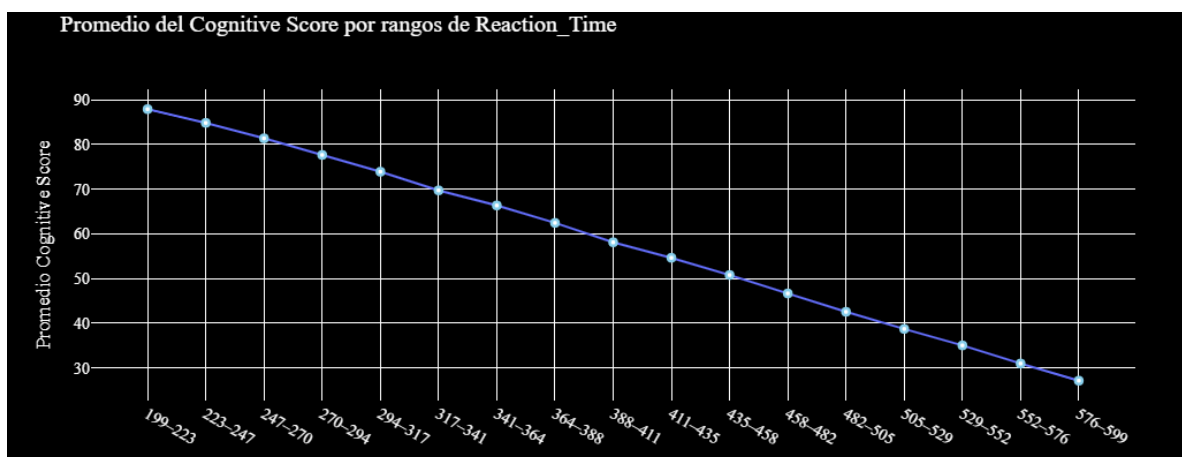
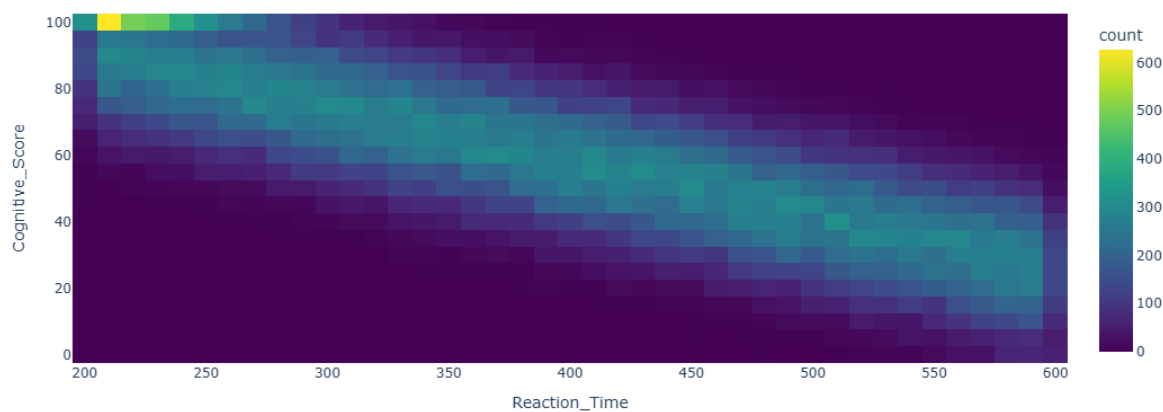
Caffeine-intake-Cognitive score



La cantidad de cafeina tomada tambien muestra una relacion lineal con la variable cognitive score, se puede ver que si la cafeina tomada diariamente es menor de 140 habria mejor rendimiento en el cognitivo ingerir mas resultaria bajando el rendimiento cognitivo, se puede ver que el peor puntaje cognitivo que se pudo alcanzar fue 53 y el intervalo seria desde 440-499 lo cual en si es muy exagerado este consumo de cafeina esto explicaria el deterioro cognitivo pero a niveles normales lo maximo que bajaria de puntuacion seria hasta 58-56

“el rango de puntuacion fue de un maximo de 62 y un minimo de 53 ”

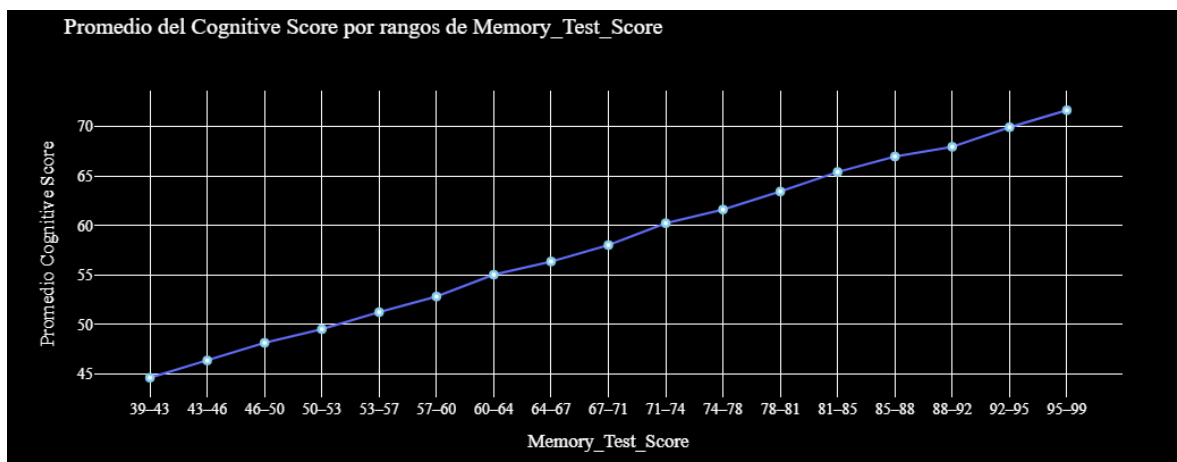
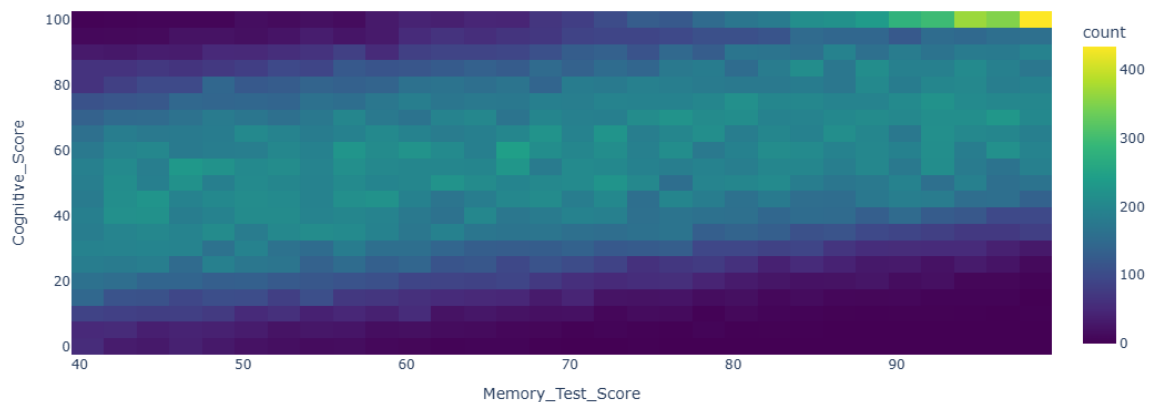
Reaction time – Cognitive score



En la grafica de densidad vemos que a valores muy bajos en la puntuacion sacada en el test de reaction time es equivalente a un cognitivo score que supera los 100 la relacion que tiene esta variable con el cognitive score es muy fuerte se puede ver como a mas tiempo que la persona se tarde haciendo el test bajaria demasiado su cognitive score en el grafico scatter vemos que en promedio por intervalo de puntuacion de reaction time el minimo valor que se llego fue de 27 dando un rango muy grande entre la puntuacion del cognitive score según el reaction time

“el rango de puntuacion fue de un maximo de 87 y un minimo de 27 ”

Memory test score-Cognitive score



En la grafica de densidad vemos que a valores muy altos en la puntuacion sacada en el test de memory es equivalente a un cognitive score que supera los 100 la relacion que tiene esta variable con el cognitive score es igual de fuerte que la variable reaction time, se puede ver como a mas puntaje sacado en el test de memoria seria equivalente a un mejor cognitive score aunque esta variable no tiende a ser tan agresiva como la variable reaction time en cuanto que a menor puntaje sacado en el test de memory bajaria mucho el cognitive score es una variable con fuerte relacion con el cognitive score

“el rango de puntuacion fue de un maximo de 71 y un minimo de 44 ”

Conclusion Bi variado graficas

En este primer acercamiento de analisis hacia el rendimiento cognitivo podemos ver que la variable age fue la unica variable que no mostro un patron o algo que nos dejara ver una relacion con el rendimiento cognitivo ,las variables como reaction time y memory test fueron las variables con una alta relacion o entrelazamiento con la variable cognitive score, las variables como sleep duration,caffeine intake,daily screen time mostraron relacion pero no es una relacion muy fuerte o clara mas bien moderada entre esas 3 variables caffeine intake fue la variable que mostro una relacion mas tenue en comparacion con las otras 2

Correlacion



Variable	Coefficiente de correlación (r)	Valor p	Significativo (p < 0.05)
Age	-0.006	0.091	False
Sleep_Duration	0.151	0.000	True
Daily_Screen_Time	-0.199	0.000	True
Caffeine_Intake	-0.123	0.000	True
Reaction_Time	-0.818	0.000	True
Memory_Test_Score	0.364	0.000	True

Entre las correlaciones de las variables numericas la unica que no tuvo una significancia en terminos estadisticos fue la de age esto afirmando lo que vimos en el analisis bi variado en el cual no vimos ninguna relacion en cuanto age y cognitive score,el resto de variables cumplen con ser significativas a la hora de una correlacion- relacion con cognitive score (esto fue confirmado con la prueba t-student de correlacion), la variable numerica que tuvo mejor magnitud de correlacion con cognitive score fue reaction_time con -0.82 lo que indica que entre mas se demore el individuo en reaccionar al test menos puntuacion en el cognitivo tendria (hay una relacion inversa entre las dos variables).

“Aunque la magnitud de correlacion de las otras variables con cognitive score es muy tenue como tenemos una muestra suficientemente grande, eso haria que hasta las correlaciones débiles se detecten como estadísticamente significativas.”

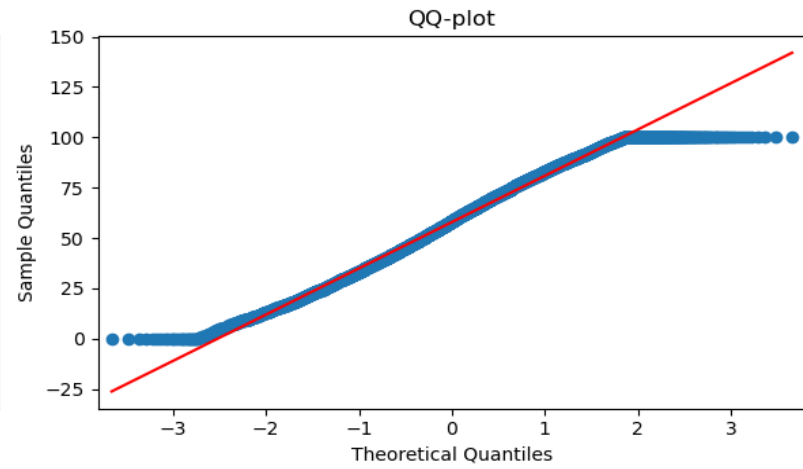
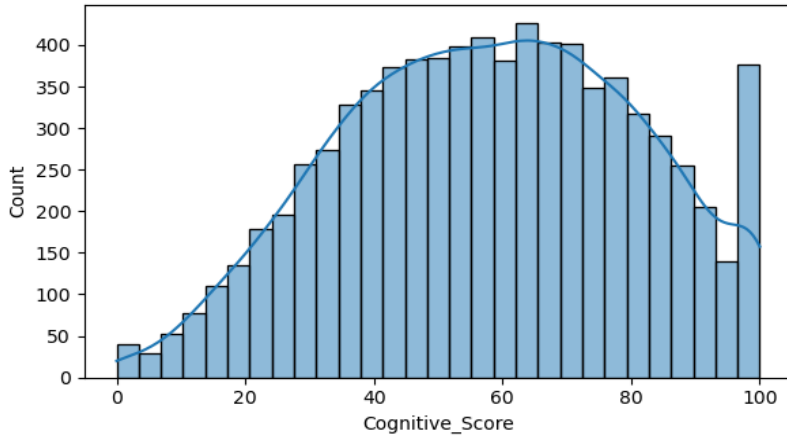
$$t = \sqrt{\frac{n - 2}{1 - r^2}}$$

Fórmula de transformación
del coeficiente de correlación
de Pearson r a un estadístico t
(t -student)

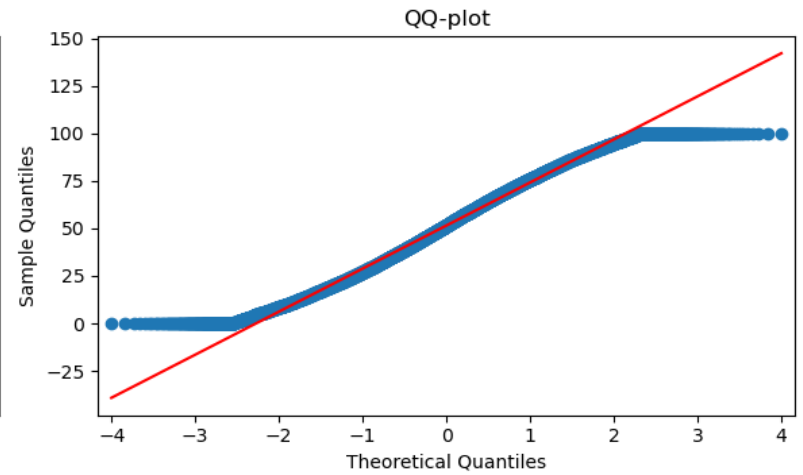
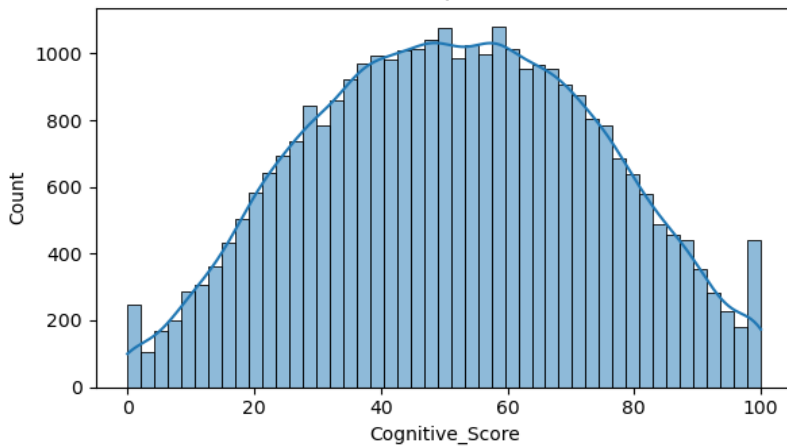
Validacion de supuestos--- Anova

Normalidad:

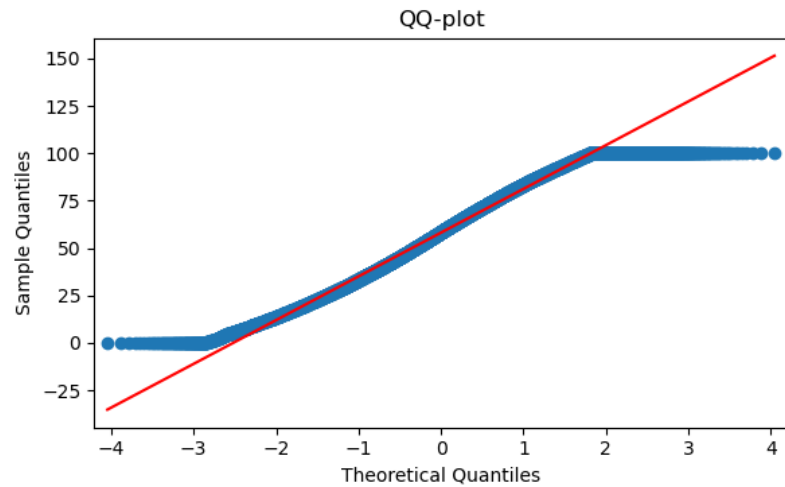
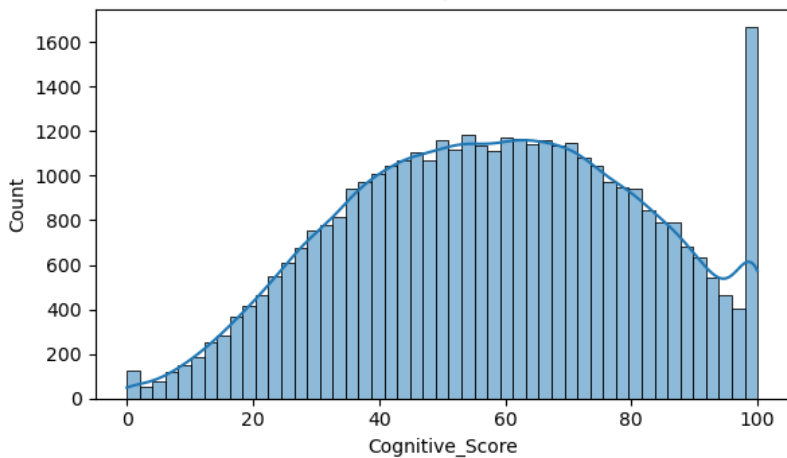
Diet_Type=Vegan
Asimetría=-0.11, Curtosis=-0.71



Exercise_Frequency=Low
Asimetría=-0.01, Curtosis=-0.70



Gender=Male
Asimetría=-0.09, Curtosis=-0.74



“Tome de cada variable categorica una etiqueta para optimizar espacio pero en si todas las etiquetas de cada variable categorica siguen la misma o muy parecida distribucion”

En los histogramas se muestra que hay una ligera inclinacion a la derecha lo que impide la normalidad La distribución del puntaje cognitivo presenta una acumulación notable de valores en 100, que es tanto su valor máximo como su moda. Esta concentración en el extremo derecho genera una ligera asimetría negativa, afectando la simetría esperada en una distribución normal.

En los qq plots vemos que lo que se supone que seria “la normalidad” según cada quantile de las variables “linea roja”, la “ linea azul “ representaria nuestros datos, como vemos cuando comparamos los datos de la variables “linea azul” con la “linea roja” que representa la normalidad, en las colas de nuestra variable se pierde la normalidad de nuestros datos , las colas de nuestra variable no siguen la distribucion normal esperada “linea roja”.

En cuanto la asimetria nos dio de -0.01 hasta -0.11 si entendemos los rangos de la asimetria :

“0: Simetría perfecta (como la normal)”

“>0: Cola derecha más larga (sesgo positivo)”

“<0: Cola izquierda más larga (sesgo negativo)”

Y considerando que la normalidad estable estaria entre $[-1,1]$

Diriamos que en cuanto la asimetria las variables no se salen mucho de lo considerado normal ya que no salen del rango establecido que admitiria una normalidad

En cuanto el curtois que mide el pico de la distribución de la variable o Mide cuántos valores están muy cerca de la media (formando un "pico") nos dio desde -0.71, -0.74 si entendemos los rangos de curtois :

(3) pico de la distribución normal

Curtosis alta (>3): Datos muy concentrados en la media (distribución "picuda")

Curtosis baja (<3): Datos más dispersos lejos de la media (distribución "aplanada")

“el rango aceptable es de 2-4 o en datasets largos de 1-5 entonces se podría decir que mi variable no lleva un pico en la distribución que sea considerada el pico de una distribución normal ”

En esta validación de supuestos , se optó por no emplear tests estadísticos tradicionales de normalidad (Shapiro-Wilk, Kolmogorov-Smirnov o Anderson-Darling) debido a tres razones fundamentales:

1. Sensibilidad en muestras grandes ($n=80,000$):

Los tests de normalidad son extremadamente sensibles a desviaciones mínimas cuando el tamaño muestral es grande, rechazando la hipótesis nula. Con tamaños muestrales como el nuestro, estos tests pierden utilidad diagnóstica.

2. Evidencia gráfica y descriptiva suficiente:

Se priorizó el análisis visual (QQ-plots e histogramas) junto con medidas descriptivas (asimetría y curtosis), que proporcionan información más interpretable en contextos con grandes volúmenes de datos:

- Los QQ-plots mostraron desviaciones localizadas principalmente en los extremos (0 y 100)
- La asimetría (-0.11 a -0.01) estuvo consistentemente dentro del rango $[-1, 1]$

- La curtosis (-0.71 a -0.74) indicó distribuciones más planas

3. Robustez del ANOVA con n grande:

El Teorema del Límite Central garantiza que, con muestras suficientemente grandes ($n > 30$ por grupo), las medias muestrales siguen una distribución normal independientemente de la forma de la distribución original.

todos los grupos cumplen $n > 30$ (TLC aplica):

Gender	Diet_Type	Exercise_Frequency	
Female	Non-Vegetarian	High	4674
		Low	9247
		Medium	9062
	Vegan	High	768
		Low	1538
		Medium	1477
	Vegetarian	High	2379
		Low	4576
		Medium	4683
Male	Non-Vegetarian	High	4560
		Low	9206
		Medium	9225
	Vegan	High	801
		Low	1504
		Medium	1488
	Vegetarian	High	2272
		Low	4513
		Medium	4753
Other	Non-Vegetarian	High	415
		Low	792
		Medium	805
	Vegan	High	54
		Low	126
		Medium	122
	Vegetarian	High	191
		Low	394
		Medium	375
dtype: int64			

Conclusion normalidad

El Teorema del Límite Central garantiza la robustez del ANOVA en nuestro estudio, ya que todos los grupos superan ampliamente el umbral de $n > 30$ (desde 54 hasta 9,247 observaciones). Esto hace que el análisis sea válido incluso cuando los gráficos (QQ-plot/histogramas) muestren desviaciones de la normalidad, pues el TLC asegura que las medias muestrales -el foco del ANOVA- siguen una distribución normal."

Homogeniedad de varianzas

Utilizamos el test estadístico de Levene para mirar la homogeneidad de las varianzas entre cada grupo (variable)

```
Levene test para Gender - p-value: 0.4879  
Levene test para Diet_Type - p-value: 0.7990  
Levene test para Exercise_Frequency - p-value: 0.0029
```

Considerando el valor de significancia de 0.05

H_0 = variable homogénea

H_1 = variable no homogénea

la única variable que rechazaría H_0

“ $p_value < \text{significancia}$ ”

Sería exercise frequency lo cual significaría que las categorías dentro de esta variable no tienen una varianza homogénea lo cual para esta variable en especial utilizaríamos ANOVA de Welch el cual no asume varianzas iguales

ANOVA

Análisis Two-Way ANOVA//Anova Welch — Cognitive Score

Se utilizó un análisis de varianza de dos vías (*Two-Way ANOVA*) para evaluar el efecto conjunto de dos variables independientes categóricas sobre la variable dependiente *cognitive score*, (*Two-Way ANOVA*) no solo permite mirar el impacto individual de cada variable, sino también las posibles interacciones entre las dos variables en relación con la varianza del rendimiento cognitivo.. Para la variable que no tiene varianza homogénea “exercise frequency” usamos (Anova Welch)

- **Variables independientes:** (*Gender, Diet_type*)<- (*Two-Way ANOVA*), (*Exercise_frequency*)<- (Anova Welch)
- **Variable dependiente:** *Cognitive score*

Este enfoque fue tomado para entender cómo estas características combinadas pueden influir en las diferencias observadas en el desempeño cognitivo.

Two-Way ANOVA

	sum_sq	df	F	PR(>F)	eta_squared
C(Gender)	5.215795e+02	2.0	0.490462	0.612346	0.000012
C(Diet_Type)	1.051966e+03	2.0	0.989205	0.371877	0.000025
C(Gender):C(Diet_Type)	4.625575e+02	4.0	0.217480	0.928835	0.000011
Residual	4.253306e+07	79991.0	NaN	NaN	0.999952

Aquí podemos ver en la tabla generada los resultados de cada variable y la combinacion de las dos variables en cuanto el rendimiento cognitivo

¿Qué tanta variabilidad aporta gender,diet_type hacia el rendimiento cognitivo?

Respondiendo directamente a la pregunta el 99% de la variabilidad de cognitive score no es atribuida ni a gender o diet_type ni a la combinacion de los dos osea es casi nulo el aporte de estas 2 variables al score del rendimiento cognitivo esto lo podemos ver en la tabla en la columna eta_squared que nos muestra por cada variable el aporte que hace hacia el cognitive score en residual vemos que lo que explica a cognitive score el 99 % no es atribuido por estas dos variables

La formula para mirar la variabilidad que explica cada variable con respecto a la variable dependiente (y):

$$\eta^2 = \frac{\text{sum sq}}{\text{ss total}}$$

“ss total” hace referencia a la suma de la columna sum sq

En cuanto la prueba de hipotesis vemos que el p value es en la tabla $PR(>F)$ si tomamos un nivel de significancia del 0.05

$p_value < 0.05$

se rechaza la hipotesis nula H_0 entonces se diria que esa variable en cuestion tiene una gran significancia en la variacion de la variable dependiente cognitive score

$P_value > 0.05$

no se rechaza la hipotesis nula entonces se diria que esa variable en cuestion NO tiene una gran significancia en la variacion de la variable dependiente cognitive score

Como podemos ver H_1 se cumple en Gender, diet_type (Gender--diet_type) dandonos a entender que si usamos la prueba de hipotesis y no miramos la variabilidad tambien llegamos a la misma conclusion de decir que estas variables no explican el cognitive score

ANOVA Welch

	Source	ddof1	ddof2	F	p-unc	np2
0	Exercise_Frequency	2	43256.806	2607.501	0.000	0.061

El p- value que nos sirve para hacer nuestra prueba hipotesis es lo que nos indica la columna p-unc en este caso es de 0.000 para la variable exercise frequency si usamos un nivel de significancia del 0.05

$p_value < 0.05$

H1 = se acepta

$p_value > 0.05$

H0 = se acepta

$0.000 < 0.05$

Exercise frequency influye estadisticamente en la variabilidad de cognitive score si queremos ver cuanta variabilidad total la variable exercise frequency explica del cognitive score miramos la columna np2 aquí nos dice que el 6 % de la variabilidad total del cognitive score es explicada por la variable exercise frequency

Coeficientes De Regresion

Contexto:

Al ser *Cognitive_Score* una variable numérica continua, podemos hacer uso de los coeficientes de regresión para mirar en detalle cómo cada variable independiente influye individualmente sobre el cognitive score. Estos coeficientes representan el cambio estimado en la variable dependiente (*Cognitive_Score*) por cada unidad de cambio en la variable independiente, manteniendo las demás constantes.

En particular:

- Un **coeficiente positivo** indica que a medida que la variable independiente aumenta, también lo hace el *Cognitive_Score*.
- Un **coeficiente negativo** sugiere una relación inversa: a mayor valor de la variable independiente, menor será el *Cognitive_Score*.
- La **magnitud del coeficiente** refleja la intensidad de esta relación.
- Además, el valor **p** asociado a cada coeficiente nos indica si ese efecto es estadísticamente significativo (generalmente se considera significativo si $p < 0.05$).

Este análisis permite no solo saber si hay relación, como lo muestra la ANOVA, sino también **cuánto y en qué dirección afecta cada categoría o nivel** de la variable explicativa al resultado cognitivo.

Resultados:

	features	coeficientes
4	num_Reaction_Time	-18.878063
5	num_Memory_Test_Score	8.380495
11	label_encoding_Exercise_Frequency	7.590830
6	num_Stress_Level	-5.531730
2	num_Daily_Screen_Time	-4.591885
1	num_Sleep_Duration	3.344942
3	num_Caffeine_Intake	-2.785432
10	one hot_Diet_Type_Vegetarian	0.037447
9	one hot_Diet_Type_Vegan	-0.013600
8	one hot_Gender_Other	-0.010122
7	one hot_Gender_Male	0.004627
0	num_Age	-0.004381

Interpretación de los Coeficientes de Regresión

Los coeficientes estimados permiten interpretar cómo varía el *Cognitive_Score* ante un cambio de una desviación estándar en cada variable numérica estandarizada (o un cambio de categoría en variables categóricas codificadas):

- **num__Reaction_Time (-18.88):** Un aumento de una desviación estándar en el tiempo de reacción se asocia con una disminución de aproximadamente 18.88 puntos en el *Cognitive_Score*. Es decir, mientras más se demora una persona en reaccionar durante una prueba cognitiva, peor será su desempeño cognitivo general.
- **num__Memory_Test_Score (8.38):** Por cada aumento en una desviación estándar en el puntaje del test de memoria, el *Cognitive_Score* incrementa

- en 8.38 puntos. Esto indica que un mejor desempeño en memoria se traduce en mayor capacidad cognitiva general.
 - **label_encoding__Exercise_Frequency (7.59):** Un incremento en la frecuencia/intensidad del ejercicio físico equivale a un aumento de 7.59 puntos en el *Cognitive_Score*. Esto sugiere una relación positiva entre la actividad física y el rendimiento cognitivo.
 - **num__Stress_Level (-5.53) y num__Daily_Screen_Time (-4.59):** Ambas variables presentan relaciones inversas. Un aumento de una desviación estándar en los niveles de estrés o en el tiempo diario frente a pantallas se asocia con una reducción del *Cognitive_Score* en 5.53 y 4.59 puntos respectivamente.
 - **num__Sleep_Duration (3.34):** Un incremento en la duración del sueño está relacionado con un aumento de 3.34 puntos en el *Cognitive_Score*, lo cual resalta la importancia del descanso para el funcionamiento cognitivo.
-

Variables con Baja Influencia en el Rendimiento Cognitivo

Las siguientes variables presentan coeficientes muy cercanos a cero, lo que indica una influencia mínima o nula en el *Cognitive_Score* en este modelo:

- **num__Caffeine_Intake (-2.79)**
- **one hot__Diet_Type_Vegetarian (0.037)**
- **one hot__Diet_Type_Vegan (-0.0013)**
- **one hot__Gender_Other (-0.0101)**
- **one hot__Gender_Male (0.0046)**

- num__Age (-0.0044)

Dado su bajo peso, estas variables no parecen ser factores relevantes para inferir el rendimiento cognitivo en este análisis.

Validacion Del Modelo de Regresion

OLS Regression Results						
=====						
Dep. Variable:	Cognitive_Score	R-squared:	0.993			
Model:	OLS	Adj. R-squared:	0.993			
Method:	Least Squares	F-statistic:	7.619e+05			
Date:	Tue, 06 May 2025	Prob (F-statistic):	0.00			
Time:	22:05:50	Log-Likelihood:	-1.3271e+05			
No. Observations:	64000	AIC:	2.654e+05			
Df Residuals:	63987	BIC:	2.656e+05			
Df Model:	12					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

const	44.4633	0.022	1997.551	0.000	44.420	44.507
x1	-0.0044	0.008	-0.576	0.565	-0.019	0.011
x2	3.3449	0.008	439.622	0.000	3.330	3.360
x3	-4.5919	0.008	-603.549	0.000	-4.607	-4.577
x4	-2.7854	0.008	-366.105	0.000	-2.800	-2.771
x5	-18.8781	0.008	-2480.977	0.000	-18.893	-18.863
x6	8.3805	0.008	1101.543	0.000	8.366	8.395
x7	-5.5317	0.008	-727.025	0.000	-5.547	-5.517
x8	0.0046	0.016	0.298	0.766	-0.026	0.035
x9	-0.0101	0.039	-0.257	0.797	-0.087	0.067
x10	-0.0136	0.026	-0.520	0.603	-0.065	0.038
x11	0.0374	0.017	2.204	0.027	0.004	0.071
x12	7.5908	0.010	747.681	0.000	7.571	7.611
=====						
Omnibus:	53443.017	Durbin-Watson:	1.997			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	2819070.391			
Skew:	-3.709	Prob(JB):	0.00			
Kurtosis:	34.656	Cond. No.	11.7			
=====						

Aquí los parámetros clave a considerar son, primero, el **R-squared**, que indica qué proporción de la variabilidad de la variable dependiente queda explicada por las variables independientes; y segundo, la columna “**P > |t|**”, que contiene los p-values con los que realizamos la prueba de hipótesis para determinar si cada coeficiente de regresión de cada variable independiente es estadísticamente significativo.

R-squared

El valor de R-squared fue de 0.99, lo que indica que el modelo de regresión explica el 99 % de la variabilidad de la variable dependiente a partir de las variables independientes. Este resultado sugiere que el modelo tiene un excelente poder explicativo, ya que refleja con gran precisión cómo las variables independientes están relacionadas con la variable objetivo.

Prueba De Hipotesis “Coeficientes De Regresion”

¿es significativo el resultado de los coeficientes de regresion?

	p value
Age	0.565
Sleep duration	0.000
Daily screen time	0.000
Caffeine intake	0.000
Reaction time	0.000
Memory_test_score	0.000
Stress level	0.000

Gender_male	0.766
Gender_other	0.797
Diet_Type_Vegan	0.603
Diet_Type_Vegetarian	0.027
Exercise_Frequency	0.000

Decisión con el p-valor:

- Si **p-valor < nivel de significancia** (por ejemplo, 0.05) → **Rechazamos H_0**
→ Concluimos que el coeficiente es estadísticamente significativo.
- Si **p-valor \geq nivel de significancia** → **No rechazamos H_0**
→ No hay evidencia suficiente para afirmar que ese coeficiente influye sobre la variable dependiente.

Las variables individuales o con la combinación de su etiqueta que no pudieron rechazar la hipótesis nula H_0 o sea afirmar que son estadísticamente significativas en relación con su coeficiente de regresión fueron Gender_male, Gender_other, Diet_Type_Vegan, Age **esto implica que no hay evidencia suficiente para afirmar que dicha variable, variable—combinación(etiqueta) tiene un efecto real sobre la variable dependiente**

METODOLOGIA

Perspectiva Adoptada para el Análisis

“Aplicación de Inferencia Estadística “

el approach principal que se le hizo al proyecto fue basado en la inferencia en especial la “prueba de hipotesis” (excepto el analisis bi-variado) osea verifique todos los analisis con un p-value y un nivel de significancia del 0.05 el primer analisis,verifique si la correlacion entre dos variables es significativa estadisticamente hablando a pesar de un coef de correlacion bajo ,en este caso fueron las variables numericas vs cognitive score

Hipótesis nula (H_0): No hay correlación lineal entre las dos variables en la población ($\rho < 0.05$).

Hipótesis alternativa (H_1): Existe correlación lineal ($p > 0.05$).

Variable	Coeficiente de correlación (r)	Valor p	Significativo ($p < 0.05$)
Age	-0.006	0.091	False
Sleep_Duration	0.151	0.000	True
Daily_Screen_Time	-0.199	0.000	True
Caffeine_Intake	-0.123	0.000	True
Reaction_Time	-0.818	0.000	True
Memory_Test_Score	0.364	0.000	True

Despues de hacer una validacion de supuestos para anova, con el anova tambien se miro si la variabilidad que explican las variables categoricas a la variable dependiente(numerica continua)es significativa estadisticamente mediante la prueba de hipotesis

$p_value < 0.05$

H1 = se acepta es estadisticamente significativa la variabilidad que explica la variable independiente de la variable dependiente

$p_value > 0.05$

H0 = se acepta no es estadisticamente significativa la variabilidad que explica la variable independiente de la variable dependiente

“en este caso la unica varibale categorica que explica una variabilidad relevante de anova fue exercy frequency

(0.0 < 0.05)”

Por ultimo se verifico si es relevante los resultados de los coeficientes de regresion para cada variable tambien con una prueba de hipotesis osea si la puntuacion que sumaria o bajaria a la variable cognitive score según un cambio de una desviacion estandar de la variable independiente ya que esta estandarizada seria significativa o no

p-value nivel de significancia :0.05

$h_0(p\text{-value} < 0.05)$ = es relevante el cambio de puntaje que la variable independiente aplica sobre la variable dependiente

$h_1(p\text{-value} > 0.05)$ = no hay relevancia o es muy minimo el cambio de puntaje que la variable independiente ejerce sobre la variable dependiente.

p-value =

P> t
0.000
0.565
0.000
0.000
0.000
0.000
0.000
0.000
0.000
0.766
0.797
0.603
0.027
0.000

¿El producto final da solución al problema planteado inicialmente?

Si , para empezar siempre se analizo cada variable vs la variable objetivo cognitive score se tuvo en cuenta la variabilidad que cada variable independiente explicaria de la dependiente como tambien que tanta correlacion tendrian, ya con la regresion vimos cuanto puntaje afectaria el cognitive score(dependiente) según cada cambio de la variable independiente “todo se verifico con una prueba de hipotesis”

LINK

“ este vinculo llevara a un video en el que se expone el analisis”

[Análisis del Rendimiento Cognitivo Humano Basado en diferentes factores de estilo de vida](#)

CONCLUSION ANALISIS

En cuanto a los factores que afectarían a un rendimiento cognitivo las variables que más destacaron son las que tienen que ver con el rendimiento motor mental “reaction time” y de memoria “memory test” o sea una persona que tiene un buen rendimiento en estas variables se podría decir que también tendría un buen desarrollo cognitivo mental pero si miramos en cuanto a factores de la vida diaria vimos que el ejercicio también afecta en gran medida, los niveles de estrés nos pueden llegar a afectar cognitivamente como también el tiempo en pantalla que se le dedica a los dispositivos tecnológicos en cuanto la cantidad de cafeína consumida también si se lleva a un exceso no normal podría afectar el rendimiento cognitivo ya por último, el tipo de dieta, o el género de la persona o edad no tienen nada de relación con el rendimiento cognitivo; En los tres test estadísticos que se hicieron estas variables nunca llegaron a mostrar nada de relación en cuanto al rendimiento cognitivo su influencia fue muy mínima e inexistente.

Bibliografias

1. Kutner, M. H., Nachtsheim, C. J., & Li, W. (2005). *Applied linear statistical models* (5th ed.). McGraw-Hill.
2. Vik, P. (2014). *Regression, ANOVA, and the general linear model: A statistics primer*. SAGE Publications.
3. Chatterjee, S., & Hadi, A. S. (2012). *Regression analysis by example* (5th ed.). John Wiley & Sons
4. Rencher, A. C., & Schaalje, G. B. (2008). *Linear models in statistics* (2nd ed.). Wiley.
5. Lou, Z., Zhang, X., & Wu, W. B. (2023). High-dimensional analysis of variance in multivariate linear regression. *Biometrika*, 110(1), 1–22.

6. Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to linear regression analysis* (5th ed.). Wiley.

7. Draper, N. R., & Smith, H. (1998). *Applied regression analysis* (3rd ed.). Wiley.

8. Fox, J. (2015). *Applied regression analysis and generalized linear models* (3rd ed.). SAGE Publications.

9. Seber, G. A. F., & Lee, A. J. (2012). *Linear regression analysis* (2nd ed.). Wiley.

10.

Muller, K. E., & Fetterman, B. A. (2003). *Regression and ANOVA: An integrated approach using SAS software*. Wiley.

