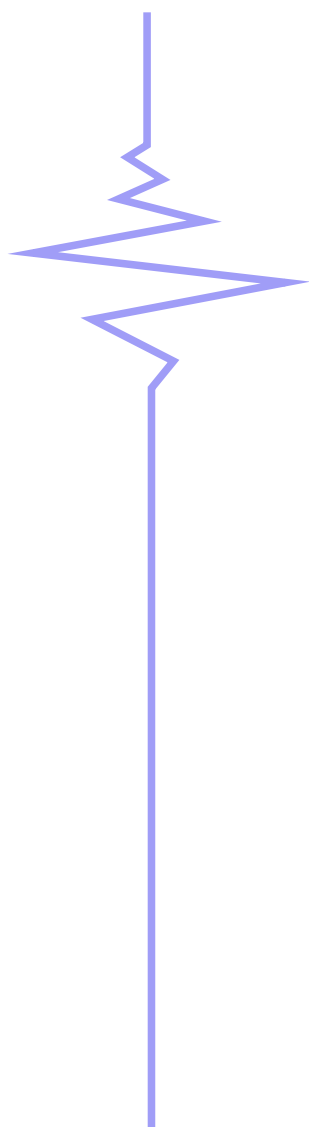
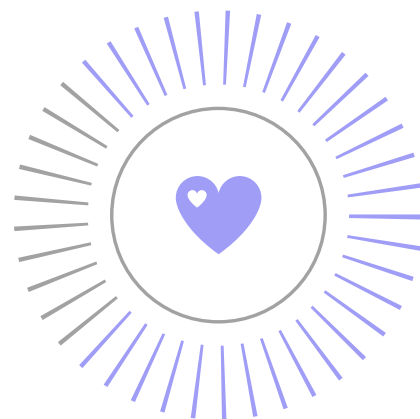


PROPENSIÓN A LA **HIPERTENSIÓN ARTERIAL**

GRUPO 3 **MUNDOSE - DATA SCIENCE 2302**

Dominguez Paula Lucía
Bonino Mauricio Edmundo
Gabirondo Lucas Martín
Lope de Barrios Nicolás
Rinaldi Juan Francisco





Por qué elegimos la (HTA)

Es el factor de riesgo de mayor peso en el desarrollo de enfermedades cardiovasculares, ya que está fuertemente asociada a la ocurrencia de diversas patologías como la enfermedad cerebrovascular, la enfermedad coronaria, la insuficiencia cardíaca, entre otras

19%
de las defunciones
a nivel mundial
son causadas por HTA

Objetivo



Determinar la relación entre distintos factores de riesgo (como consumo de tabaco, alcohol, alimentación, actividad física, entre otros) y la propensión a la hipertensión arterial (HTA), en personas de la República Argentina.

Objetivos Específicos

Explorar la base de datos publicada

Seleccionar variables a analizar

Caracterizar las variables seleccionadas

Analizar de forma bivariada los factores de riesgo con la prevalencia de HTA

Determinar y comprender la relación que se establece entre las variables analizadas y la HTA

Descripción del dataset

indec

29.224 filas
287 variables

Nuevo Dataset

26.780 filas
11 variables
✖ ns/nc

Encuesta Nacional de Factores de Riesgo (ENFR)
Instituto Nacional de Estadística y Censos (INDEC)
Ministerio de Salud de la Nación.
Año 2018

Información sobre las **Enfermedades No Transmisibles (ENT)**

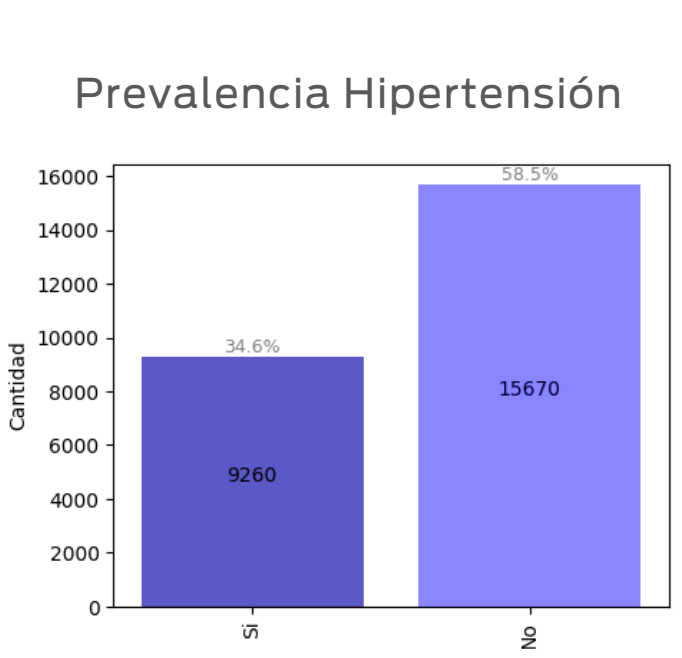
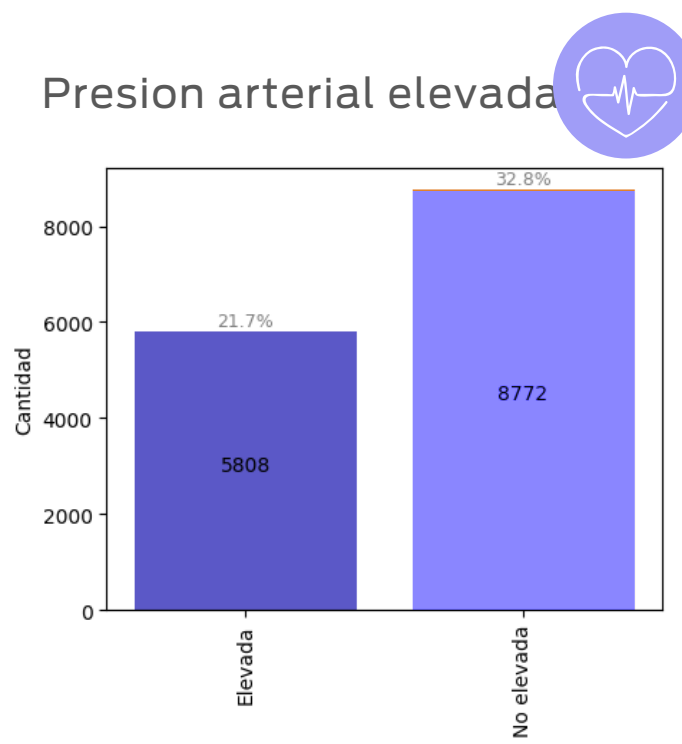
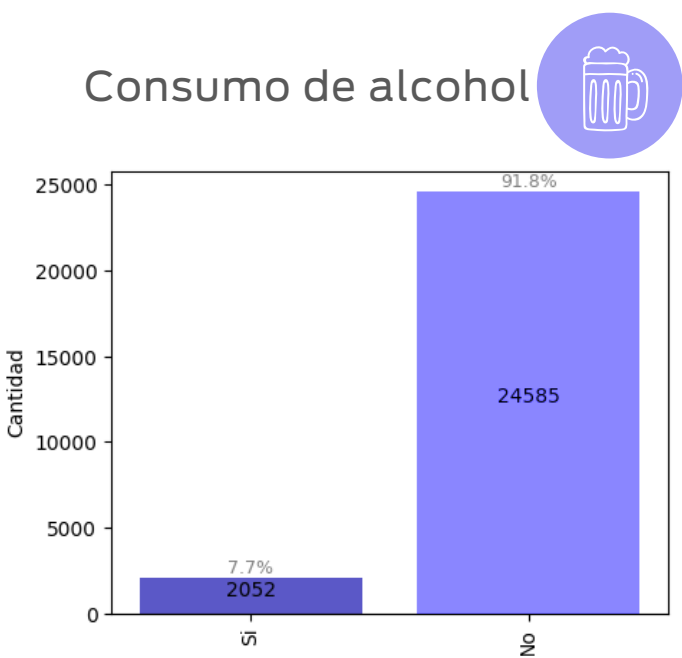
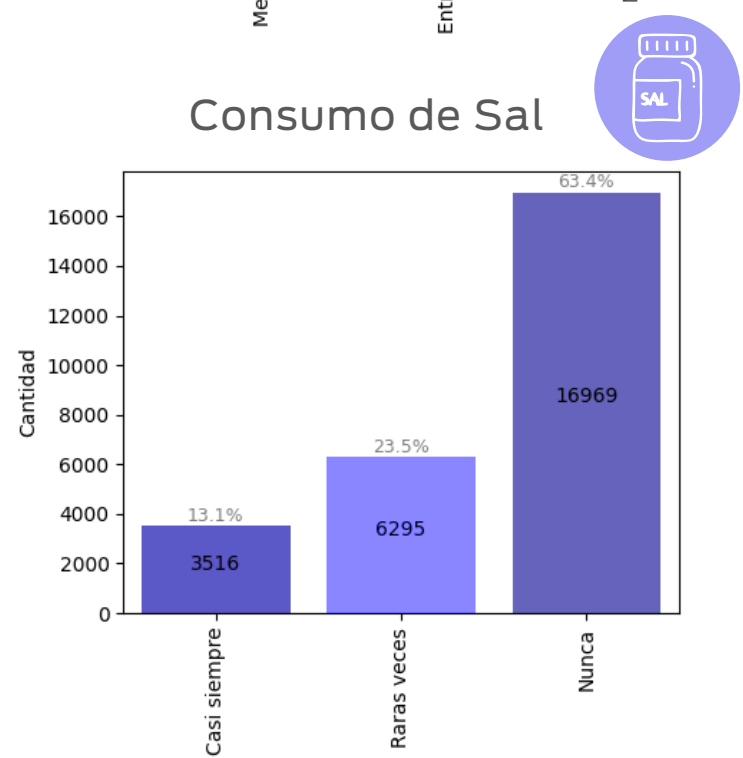
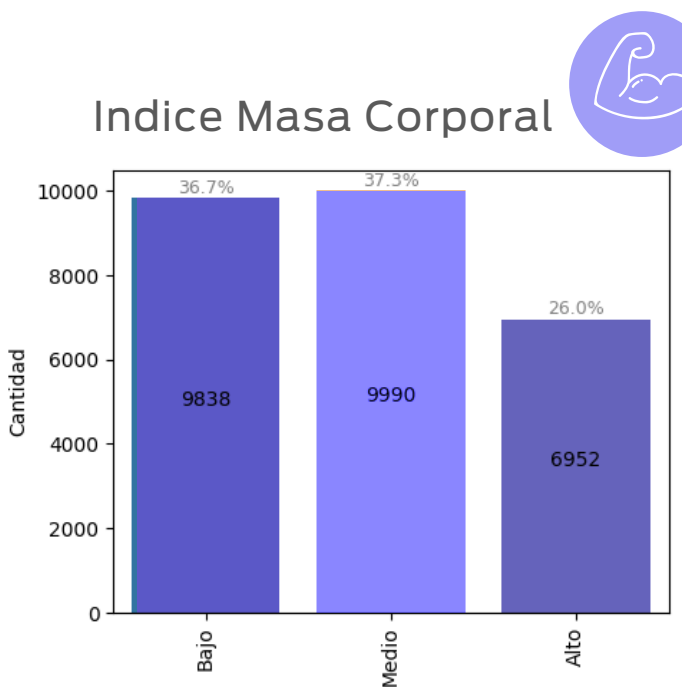
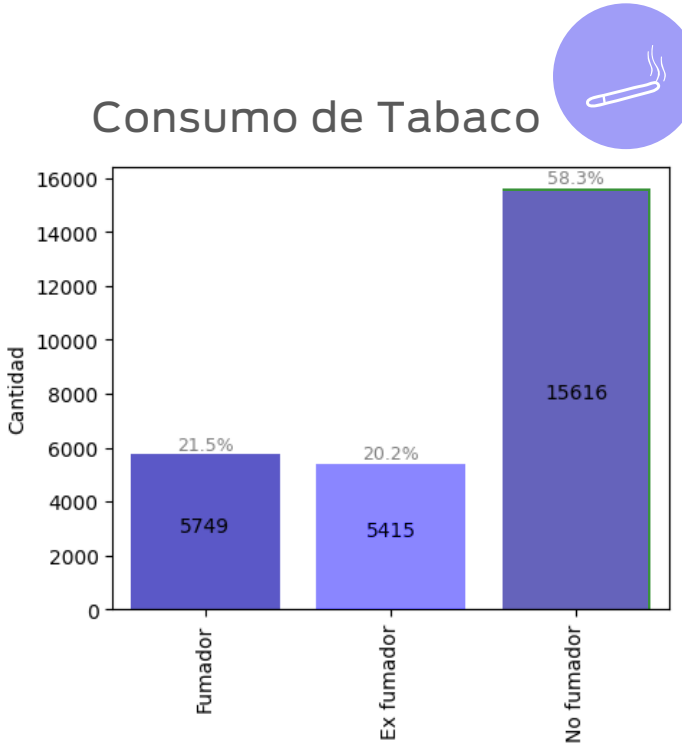
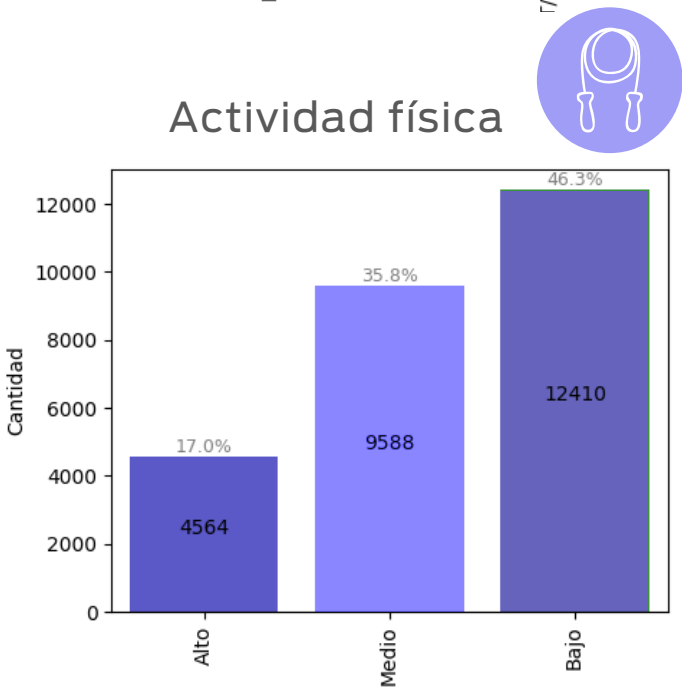
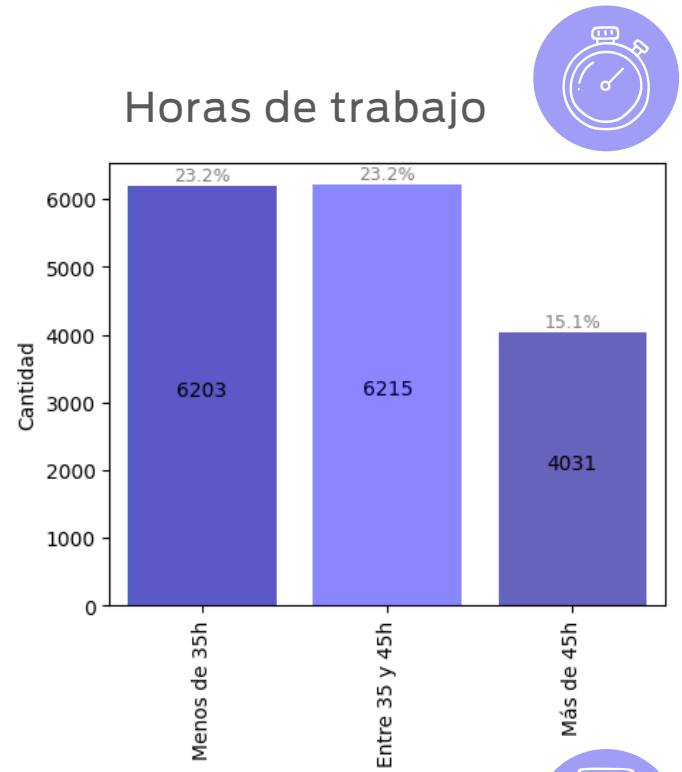
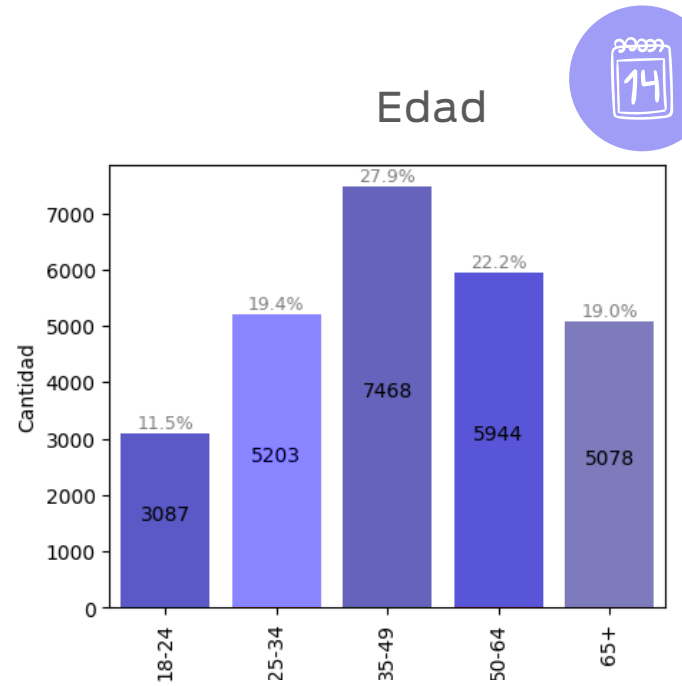
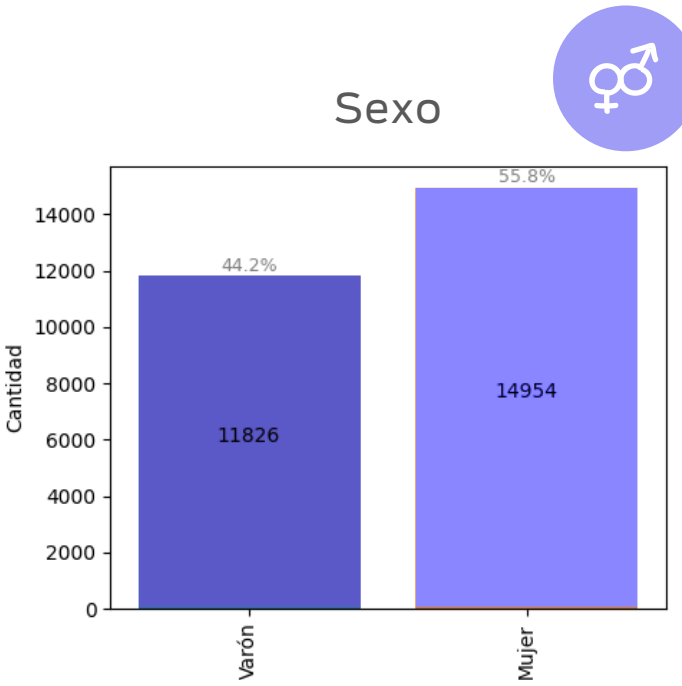
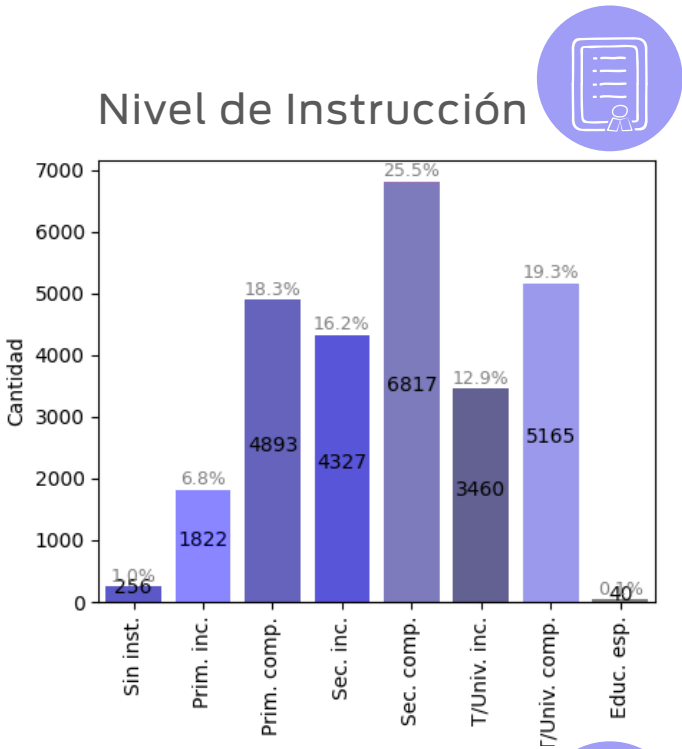
Variables predictoras

- nivel_instruccion (*Nivel de instrucción*)
- bhch03 (*Sexo biológico*)
- rango_edad (*Rango de edad*)
- bisl06 (*¿Cuántas horas semanales trabaja habitualmente en todos sus empleos/ ocupaciones?*)
- nivel_actividad_fisica (*Nivel de actividad física*)
- consumo_tabaco_100 (*Condición de fumador*)
- imc_categorias (*Índice de masa corporal agrupado*)
- bial02 (*Habitualmente, ¿le agrega sal a los alimentos una vez que están cocidos o al sentarse a la mesa...*)
- consumo_regular_riesgo (*Consumo regular de alcohol de riesgo*)
- ta_elevada (*Presión arterial elevada*)

Variable a predecir

- Prevalencia de presión arterial elevada

Análisis univariado

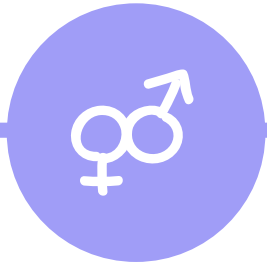


Análisis univariado



25%

SECUNDARIO COMPLETO



55%

MUJERES

44%

VARONES

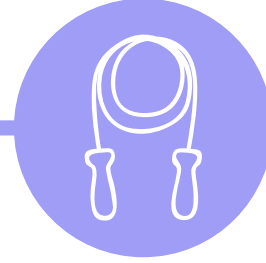


27%

35-49 AÑOS

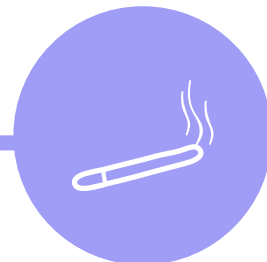
41%

+50 AÑOS



46%

POCA ACTIVIDAD FISICA



41%

FUMA O FUMÓ



7.5%

CONSUME ALCOHOL
REGULARMENTE



63%

NO AGREGA SAL



26%

IMC >= 30

74%

IMC < 30

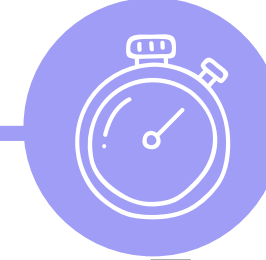


22%

PRESIÓN ELEVADA
AUTOREPORTADA

35%

PREVALENCIA HTA



46%

TRABAJO SEMANAL
<45 HORAS

15%

TRABAJO SEMANAL
>45 HORAS

Análisis bivariado

Prevalencia Hipertensión

Nivel de Instrucción

nivel_instruccion_cat	Sin inst.	Prim. inc.	Prim. comp.	Sec. inc.	Sec. comp.	T/Univ. inc.	T/Univ. comp.	Educ. esp.	All
prevalencia_hipertension_cat									
Si	0.0	4.0	8.0	5.0	8.0	3.0	6.0	0.0	35.0
No	0.0	3.0	8.0	9.0	16.0	9.0	13.0	0.0	59.0
All	1.0	6.0	17.0	14.0	24.0	12.0	19.0	0.0	93.0

Edad

rango_edad_cat	18-24	25-34	35-49	50-64	65+	All
prevalencia_hipertension_cat						
Si	2.0	4.0	8.0	10.0	12.0	35.0
No	8.0	14.0	19.0	11.0	7.0	59.0
All	10.0	17.0	26.0	22.0	19.0	93.0

Actividad física

nivel_actividad_fisica_cat	Alto	Medio	Bajo	All
prevalencia_hipertension_cat				
Si	4.0	12.0	18.0	34.0
No	11.0	22.0	25.0	58.0
All	16.0	33.0	43.0	92.0

Consumo de Sal

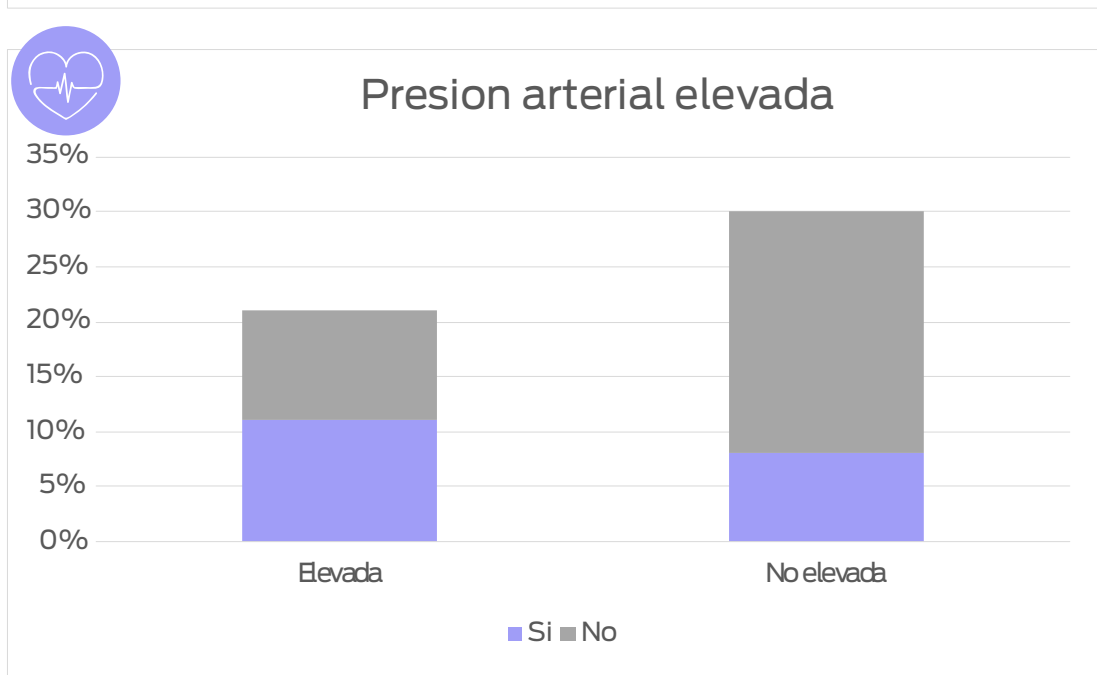
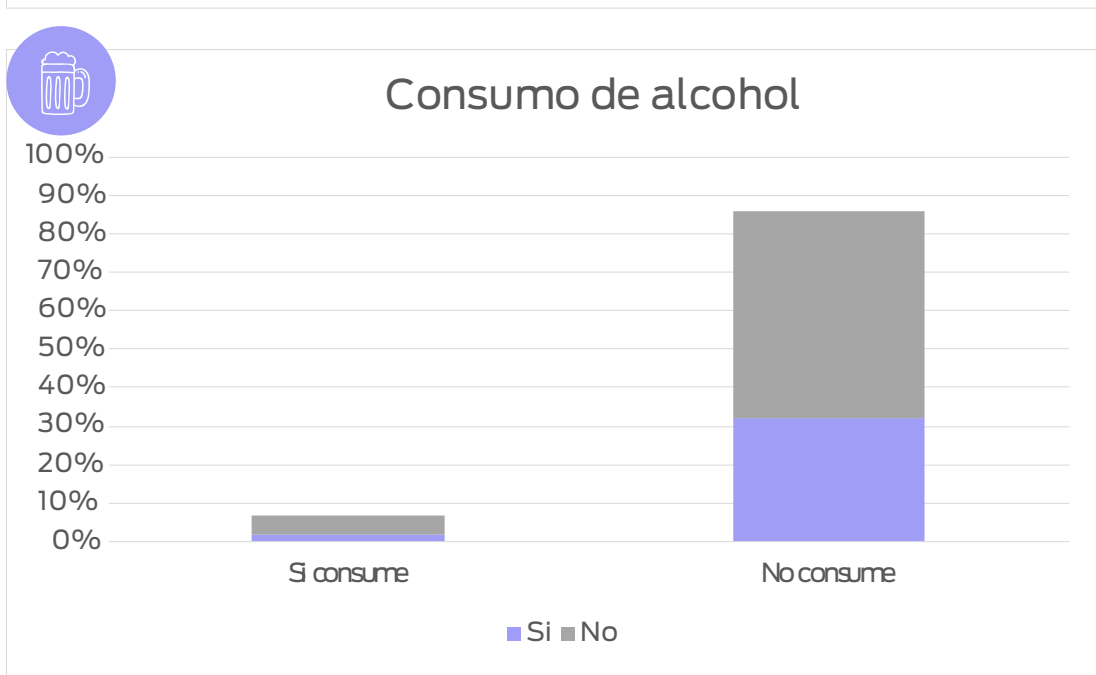
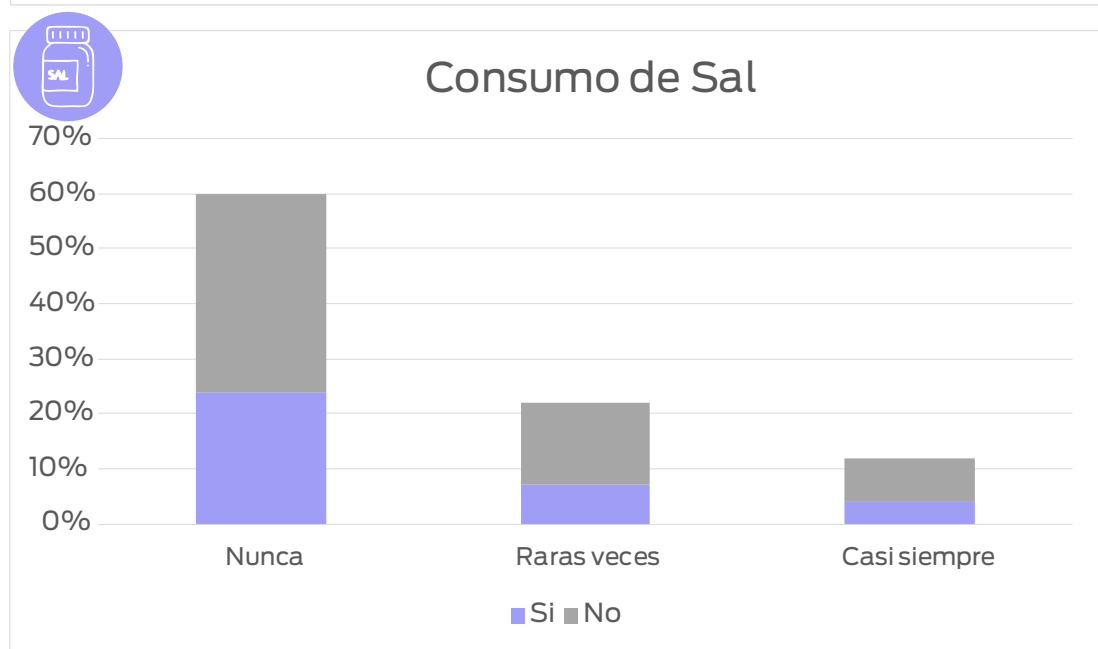
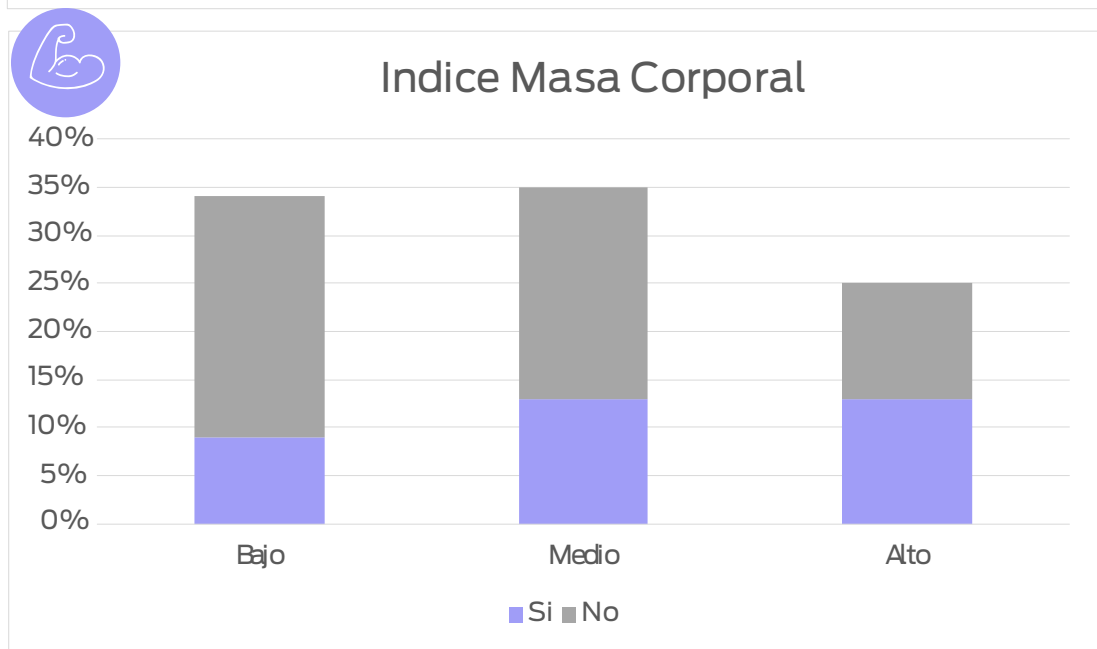
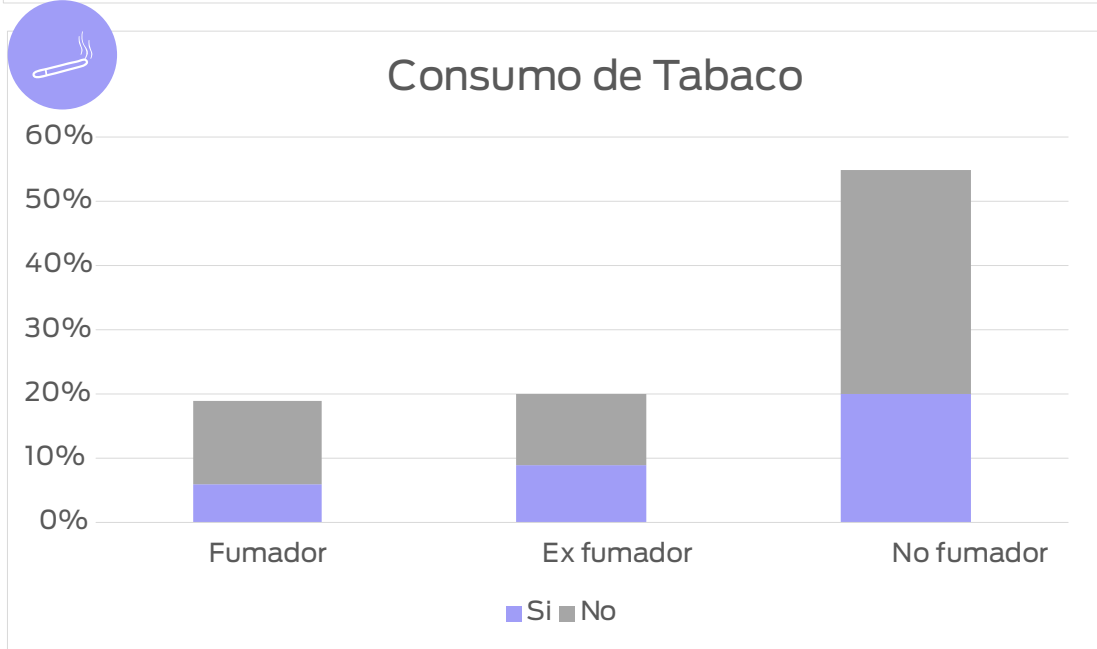
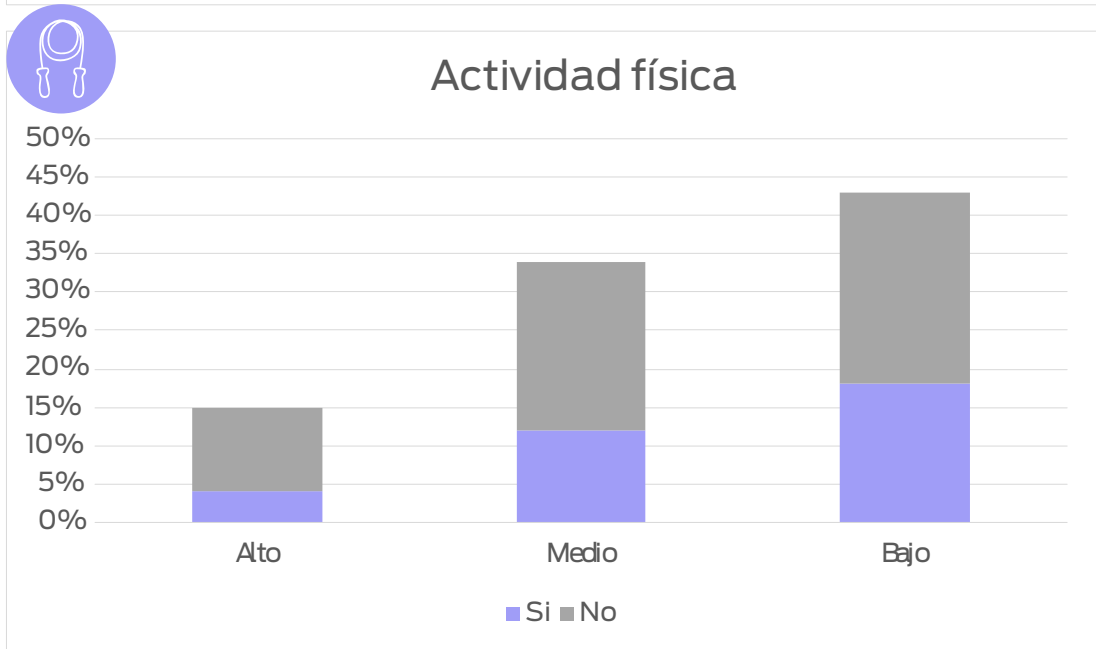
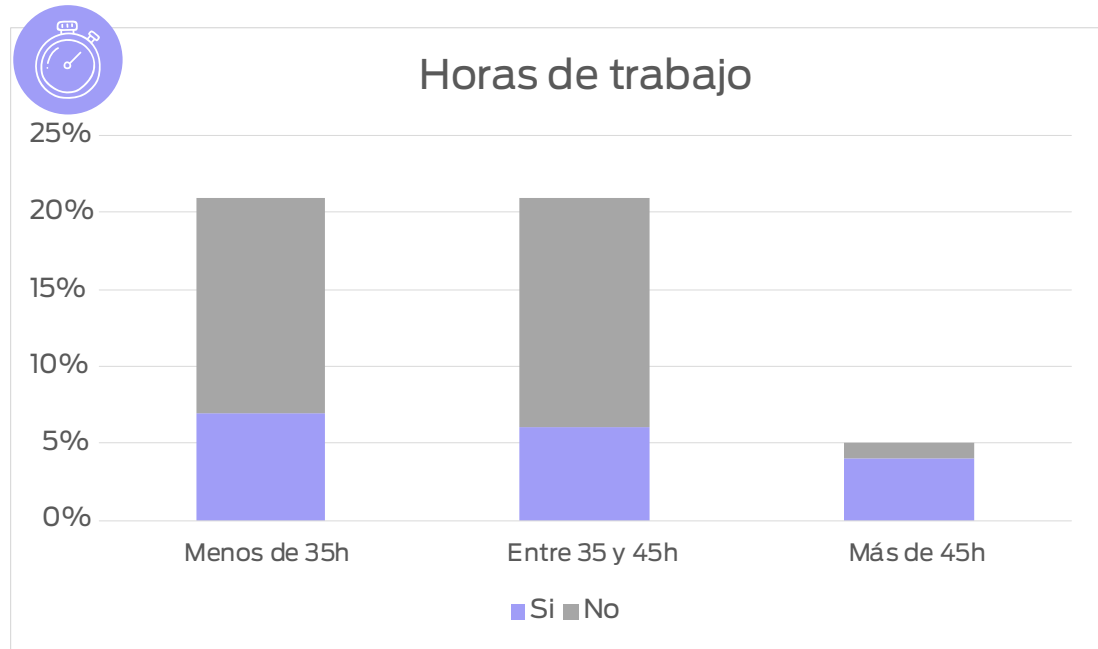
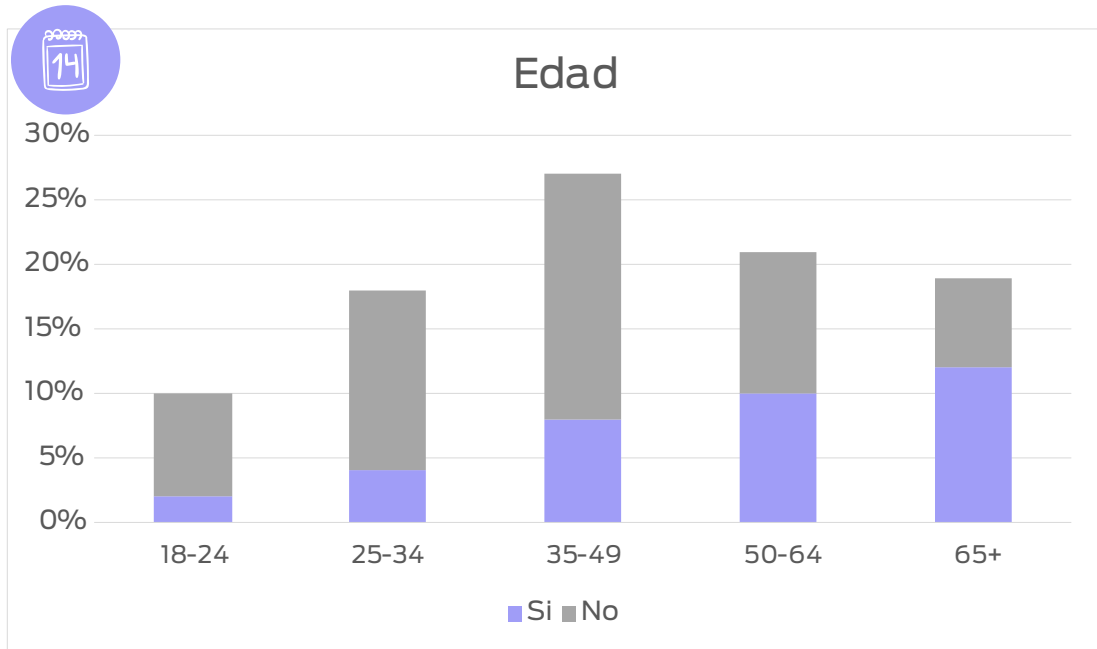
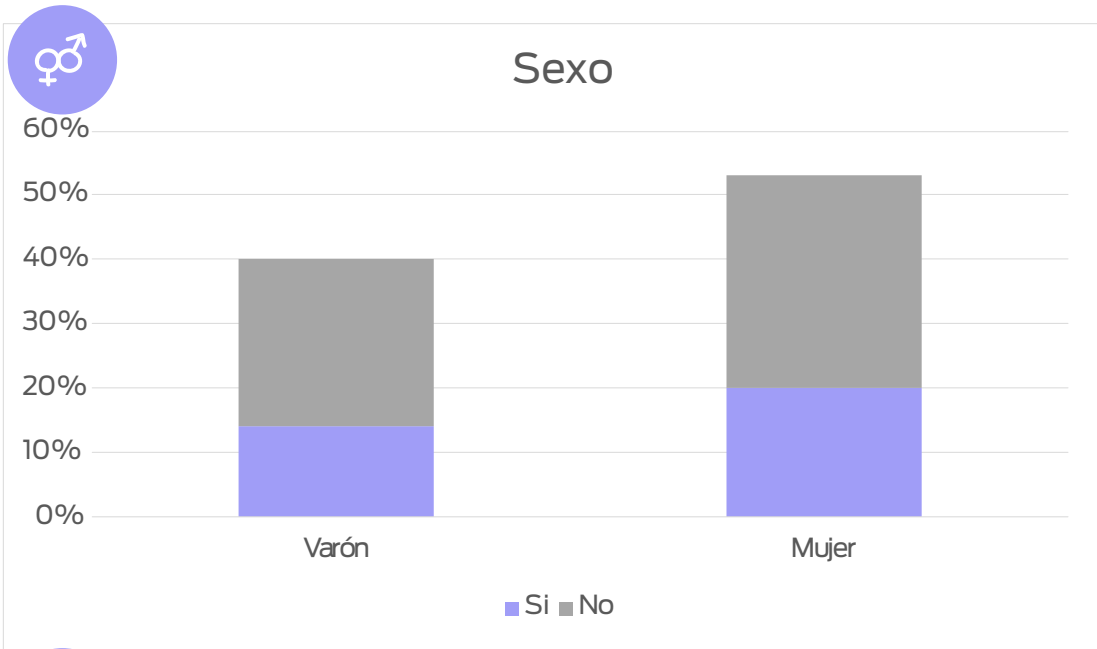
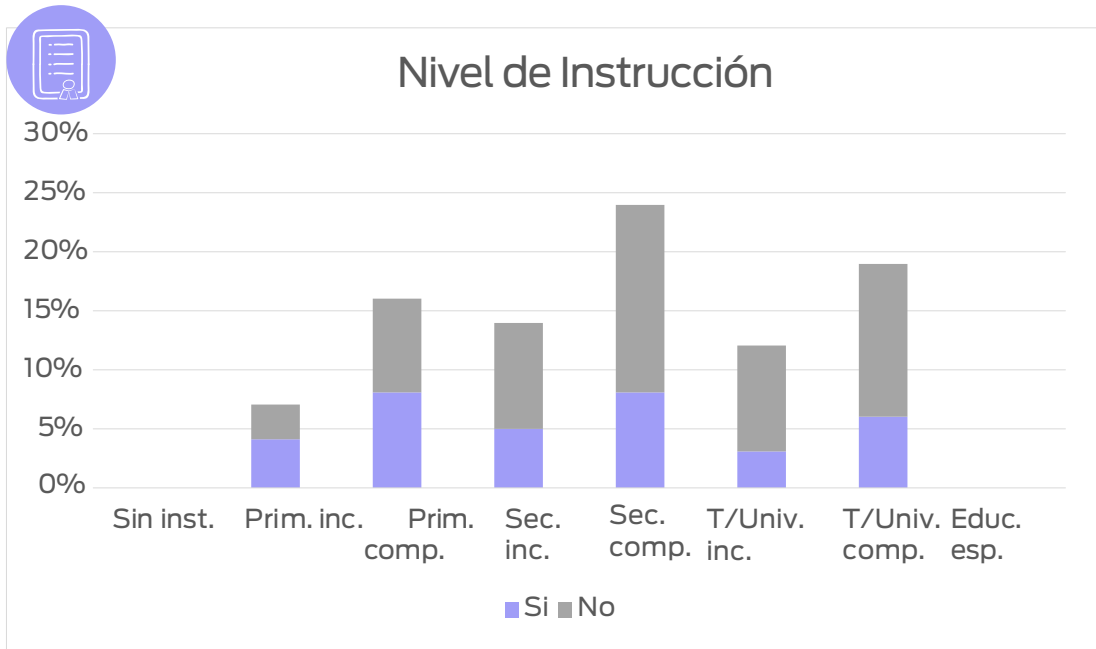
consumo_sal_cat	Casi siempre	Raras veces	Nunca	All
prevalencia_hipertension_cat				
Si	4.0	7.0	24.0	35.0
No	8.0	15.0	36.0	59.0
All	12.0	22.0	60.0	93.0

Consumo de Tabaco

condicion_fumador_cat	Fumador	Ex fumador	No fumador	All
prevalencia_hipertension_cat				
Si	6.0	9.0	20.0	35.0
No	13.0	11.0	35.0	59.0
All	19.0	20.0	54.0	93.0

Análisis bivariado

Prevalencia Hipertensión



Pruebas de independencia Chi-cuadrado

Nivel de Instrucción		
prevalencia_hipertension_cat	Si	No
nivel_instruccion_cat		
Sin inst.	125	101
Prim. inc.	950	737
Prim. comp.	2259	2271
Sec. inc.	1423	2453
Sec. comp.	2146	4163
T/Univ. inc.	812	2429
T/Univ. comp.	1536	3487
Educ. esp.	9	29
Prueba Chi-cuadrado:		
chi2 = 936.8973, p-valor = 0.00000		
Grados de libertad = 7		

Consumo de alcohol		
prevalencia_hipertension_cat	Si	No
horas_trabajo_semana_cat		
Menos de 35h	1934	3768
Entre 35 y 45h	1612	4114
Más de 45h	1117	2626
Prueba Chi-cuadrado:		
chi2 = 46.4764, p-valor = 0.00000		
Grados de libertad = 2		

Horas de trabajo		
prevalencia_hipertension_cat	Si	No
consumo_regular_alcohol_cat		
Si	652	1208
No	8562	14379
Prueba Chi-cuadrado:		
chi2 = 3.6940, p-valor = 0.05461		
Grados de libertad = 1		



Pruebas de independencia Chi-cuadrado

Se utiliza para comprobar la **independencia de dos variables categóricas**. El objetivo es analizar **si los valores característicos de la primera variable están influidos por los valores característicos de la segunda variable y viceversa**.

Si los valores observados son significativamente diferentes de los valores esperados, se puede concluir que **existe una relación entre las dos variables**.

Se formulan dos hipótesis:

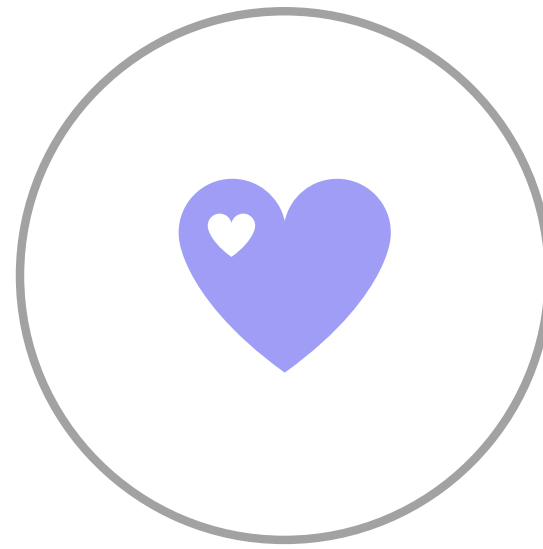
H0: No existe relación entre las características (h. nula)

H1 : Existe relación entre las características

Nosotros empleamos la función **scipy.stats.chi2_contingency** para realizar una prueba de chi-cuadrado de independencia en una tabla de contingencia. Las frecuencias observadas se calculan con las tablas de contingencia (pandas.crosstab) y las frecuencias esperadas son calculadas por chi2_contingency.

Si el **valor p** devuelto por la función es menor que el nivel de significancia establecido (0,05), se **rechaza la hipótesis nula** y se concluye que **hay una relación significativa** entre las dos variables categóricas.





Conclusiones

Concluimos que las variables seleccionadas presentan **una fuerte relación con la HTA**, a excepción del **consumo regular de alcohol en exceso**.

Se podrá extender el mismo introduciendo otras variables de la base de datos no contempladas hasta ahora.

Dado que la mayoría de las variables arrojaron **un p-valor muy cercano a 0** en las pruebas de independencia de Chi-cuadrado, se sugiere profundizar el presente trabajo a fin de analizar esta situación.

Trabajo futuro

Realizar y aplicar un modelo de **regresión logística**, que se utiliza para predecir la probabilidad de un evento binario (sí/no) en función de una o más variables independientes categóricas o continuas.



Gracias