



SMOKING

- Nombre: Nazareno Pera.
- Curso: Data Science II: Maching Learning para la Ciencia de Datos.
- Proyecto: Smoking .
- Comision: 60865.
- Profesor: Fernando Carabedo.
- Tutor: Noelia Ferrero.
- Modelo: Clasificacion.

INTRODUCCION

El tabaquismo es una adicción y la nicotina es una sustancia adictiva. Cada cigarrillo suministra entre 1 y 2 mg de nicotina. Esta llega al cerebro en unos 10 segundos, lo que contribuye a su gran poder adictivo, mayor al de la cocaína y la heroína.

La definición de adicción contempla: tolerancia (se debe ir aumentando la dosis para tener el mismo efecto); síndrome de abstinencia (cuando no se consume) y consumo compulsivo, a pesar de conocer las consecuencias negativas que tiene en la salud.

El síndrome de abstinencia por falta de nicotina incluye generalmente irritabilidad, somnolencia, fatiga, dificultad para concentrarse, trastornos del sueño, aumento de peso, urgencia de fumar. La mayoría de estos síntomas alcanzan su máxima intensidad entre 24 a 48 horas luego del último cigarrillo y van disminuyendo lentamente en un período de 2 a 4 semanas.

Datos extraídos:

El cigarrillo produce daños en casi todos los órganos del cuerpo y es la principal causa de muerte a largo plazo. A pesar de la disminución de las tasas de tabaquismo, se estima que fumar contribuye a unas 480.000 muertes anuales. Los fumadores mayores de 60 años tienen un índice de mortalidad dos veces más alto que quienes nunca han fumado y se estima que mueren aproximadamente 6 años antes. Dejar de fumar genera beneficios inmediatos para la salud, y es posible recuperar todas o algunas de las reducciones de la expectativa de vida según la edad a la que la persona deje de fumar. En el mundo mueren mas de 8 millones de personas por año a causa del cigarrillos.

El consumo de tabaco genera tres formas de dependencia:

Física: Provocada directamente por la nicotina, responsable del síndrome de abstinencia.

Psicológica: Fumar se asocia a diferentes situaciones cotidianas y parece imposible cambiar esta relación.

Social: Para muchas personas, fumar sigue siendo un imperativo social.

OBJETIVO

- El Objetivo de este proyecto es poder aplicar un modelo de machine learning de aprendizaje supervisado y a través de las mis variables categóricas o numéricas, poder llevar a cabo un algoritmo donde me permita clasificar y poder ver si las personas fumadoras presentar mayores problemas de salud que las personas que no consumen nicotina .
- Modelo que use: Use algoritmos de clasificación porque son útiles cuando se quiere predecir una variable objetivo que pertenece a un conjunto discreto de variables categóricas o numéricas.

ESTRUCTURA DEL TRABAJO

Mi proyecto se divide en 5 partes:

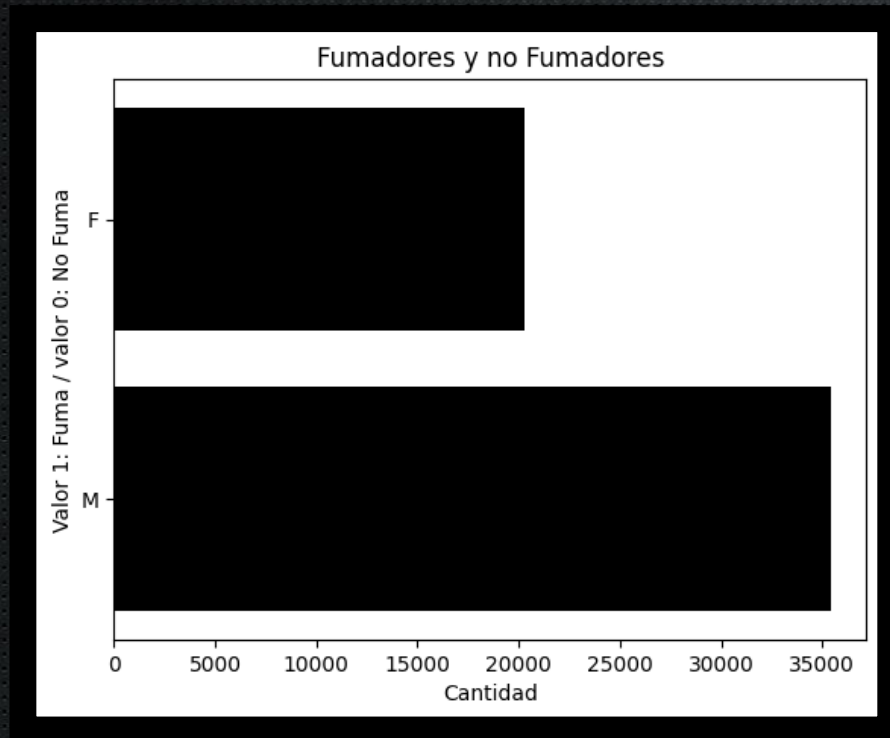
- Introducción del Trabajo: Se explica un poco de que trata el dataset, se muestra los sitios de donde extraje la información y se habla de la hipótesis que planteo y sobre mi modelo que use.
- Análisis: Se muestra el análisis y la conclusión de cada grafico.
- Modelo de entrenamiento: Se muestra el modelo que use para entrenar mi modelo de machine learning.
- Pca: Se aplica pca para ver las reducción de dimensionalidad y para la mejora del rendimiento del modelo.
- Conclusión: Desarrollo una breve conclusión sobre mi proyecto de data science.

▪ Información adicional:
Quería sumar que algunas conclusiones o justificaciones de eliminación de outliers se puede ver en el colab. También quería agregar que algunos gráficos están en el colab porque se desarrolló mucho el trabajo y no quería llenar la presentación de tantos gráficos, ya que algunos datos no son tan representativos como otros.

ANÁLISIS EXPLORATORIO DE LOS DATOS (EDA).

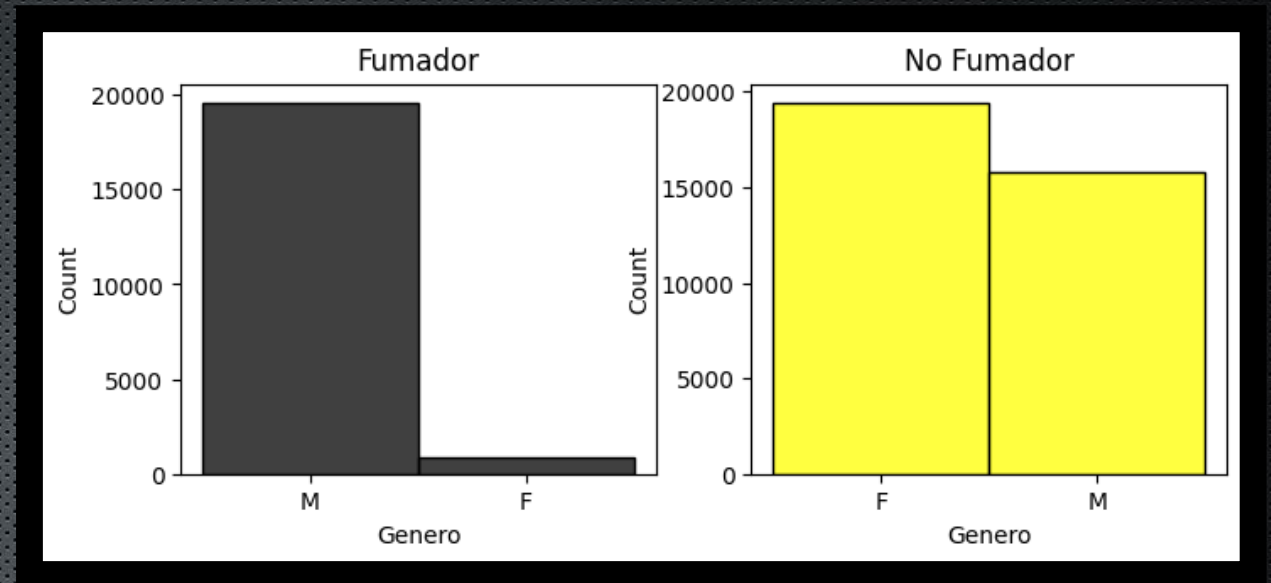
- Descripción de cada variable.
- Genero: El genero de cada sexo.
- Edad: La edad de cada persona.
- Altura: La altura de cada persona.
- Peso: El peso de cada Persona.
- Cintura: Medida de la cintura.
- Vista izquierda: La vista del ojo izquierdo marca que graduación de porcentaje tienen.
- Vista derecha: La vista del ojo derecho marca que graduación de porcentaje tienen. tienen.
- Oído izquierdo: La audición del oído izquierdo.
- Oído derecho: La audición del oído derecho.
- sistólica: Es la presión de cada persona.
- Relajación: La relajación.
- Azúcar en la sangre en ayuna: La medición de azúcar en sangre en ayuna.
- Colesterol: Medición del colesterol.
- triglicérido: La prueba de triglicéridos mide la cantidad de una grasa que hay en la sangre llamada triglicérido.
- Colesterol Bueno: Medición del colesterol Bueno.
- Colesterol Malo: Medición del colesterol Malo.
- Hemoglobina: La hemoglobina
- Urine proteína: Análisis de proteína en orina.
- Suero de creatinina: Resultado de suero de creatina.
- AST: La prueba de sangre de AST se suele incluir en un análisis de sangre de rutina para monitorear la salud de su hígado.
- ALT: La prueba de sangre de ALT es parte de una prueba sanguínea de rutina para revisar la salud del hígado.
- Gtp: El GPT en un análisis de sangre se refiere a la concentración de transaminasa GPT en el torrente sanguíneo, una enzima que se concentra mayoritariamente en el hígado.
- oral: Oral.
- Caries dental: Examen de caries dental.
- Sarro: Examen de sarro dental.
- smoking: El valor 1 representa los fumadores y el valor 0 representa los no fumadores.
- En base a lo ejecutado hasta ahora, se puede ver que el dataset consta de:
 - 27 columnas.
 - 55692 filas.
 - Datos Alfanumérico.
 - No hay datos nulos.

ANÁLISIS BIVARIADO



Cantidad de fumadores

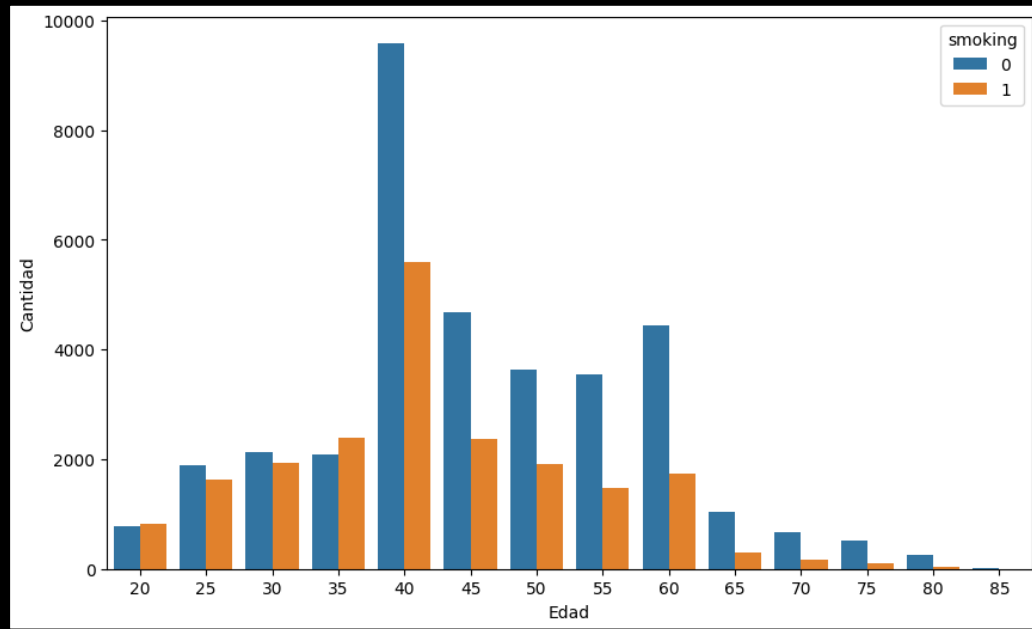
- En este grafico se puede ver la cantidad de fumadores y no fumadores. El Valor 1 representa los fumadores y el valor 0 representa los no fumadores, en esta visualización se ve como los fumadores pueden llegar al 55 % del total de las personas que no fuman.



Géneros

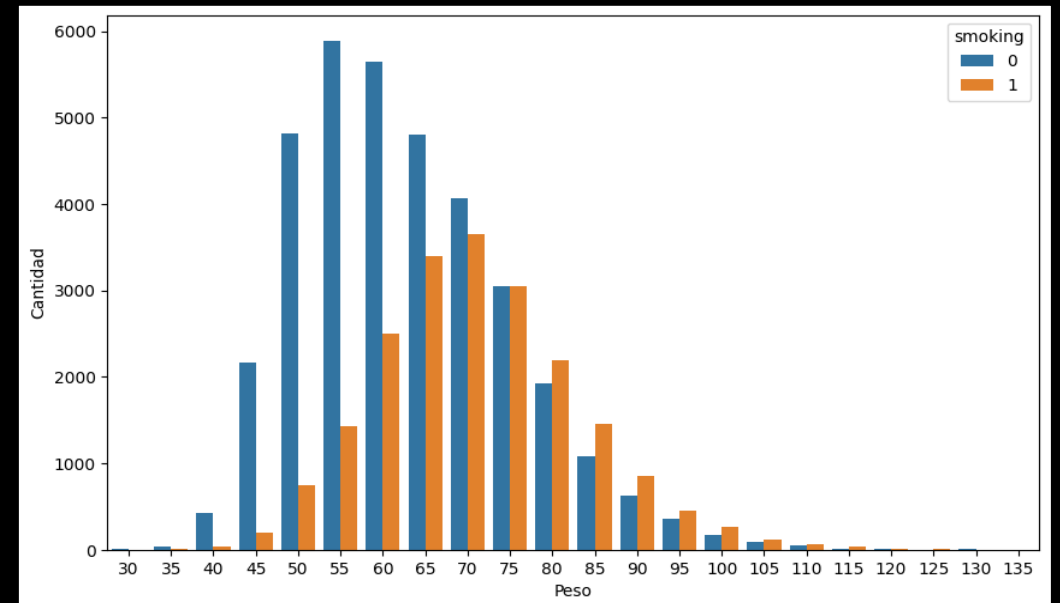
- En este grafico se puede mostrar como la mayor cantidad de fumadores son de genero masculino y el genero femenino es mucho menor.
- La cantidad de fumadores femeninos no llegan ni al 10% del genero masculino fumador.
- Se podría llegar a la conclusión que el genero masculino es mayormente el genero que mas fuma.

ANÁLISIS BIVARIADO



Edad

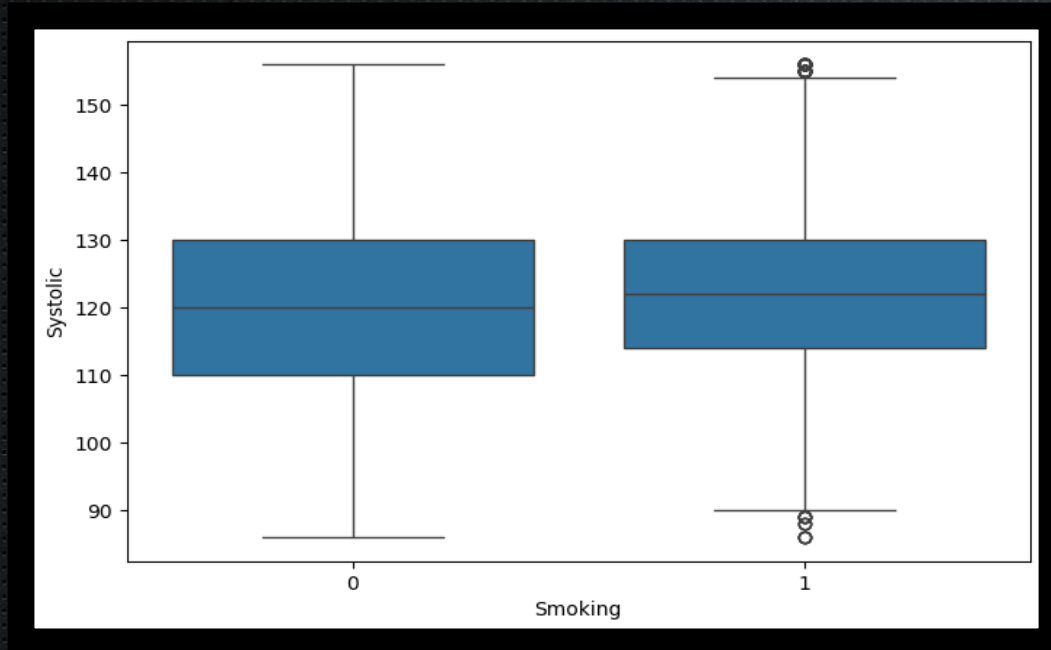
- Esta grafico se puede mostrar como el porcentaje de fumadores por edades y no fumadores, relativamente no cambia mucho entre si. De igual manera el mayor volumen de fumadores parece ser en los 40 años y después va disminuyendo.
- Conclusión se podría decir que tanto no fumadores y fumadores, no hay un cambio entre volúmenes de edades.



Peso

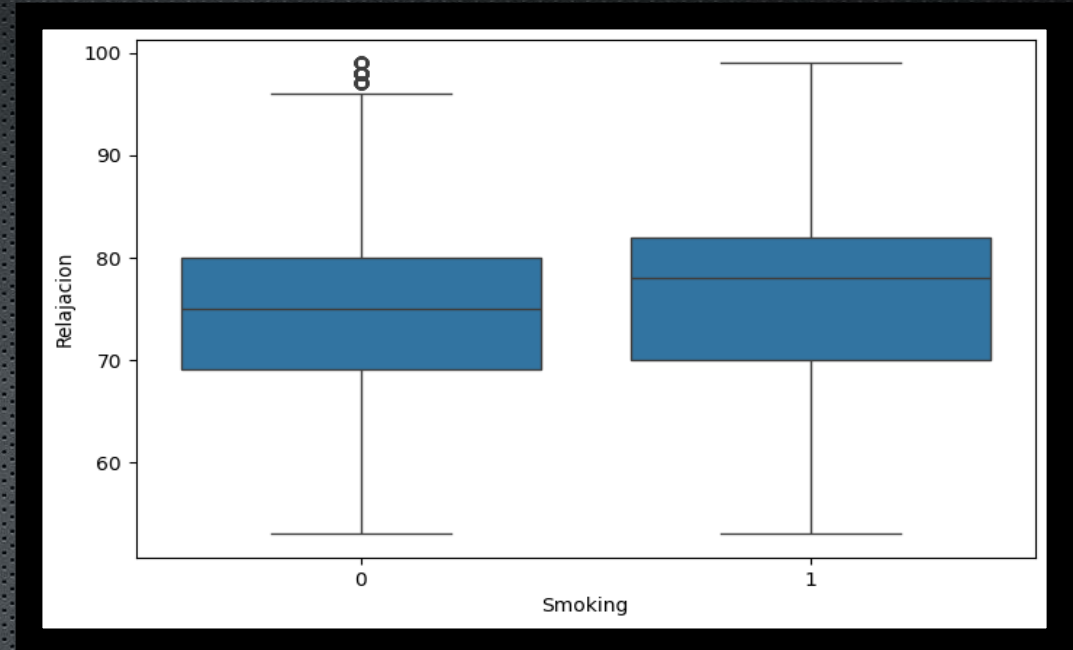
- En este grafico se puede ver el peso de los fumadores y no fumadores.
- En los dos gráficos entre fumadores y no fumadores el peso es relativamente el mismo, excepto se puede ver un volumen mayor de 80 y 85 kilos en adelante en personas fumadoras .
- Conclusión en ambos gráficos no representan un mayor volumen de uno a otro, no hay tantas diferencias entre variables.

ANÁLISIS BIVARIADO



Systolic

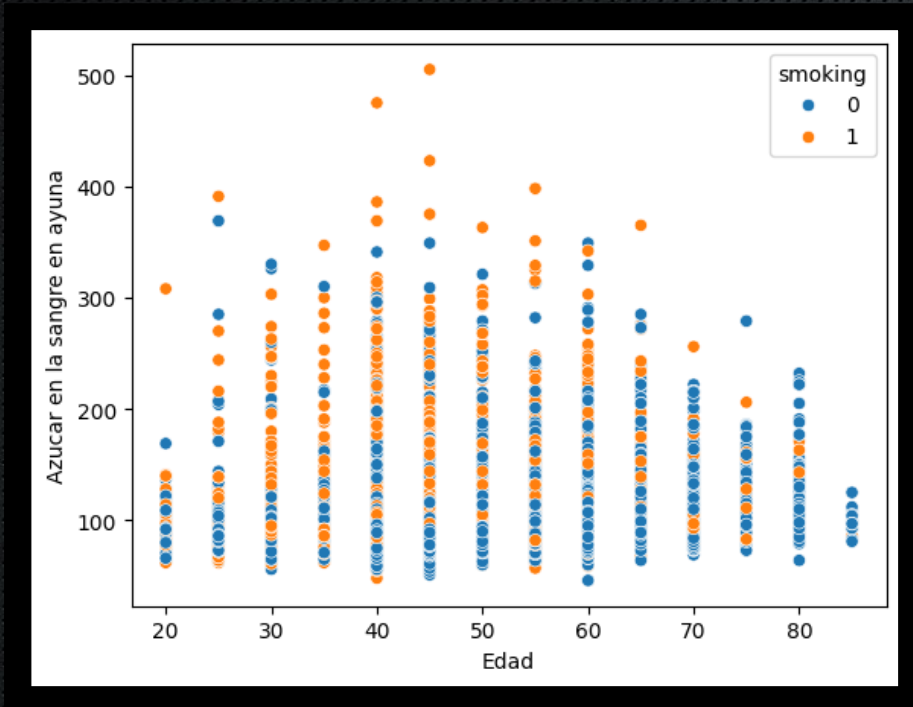
- En este grafico de boxplot se eliminaron outliers para reducir el volumen de ambas variables y quedaron datos atípicos en la variables fumadora en la parte superior mostrando que hay casos en que tienen presión mas alta y mostrando un mayor volumen de datos en la caja en la media .
- Se recomienda mantener una presión arterial normal de 120/80 mmHg para adultos.
- Conclusión hay cambios notorios y un fuerza mayor en la variable 1.



Relajación

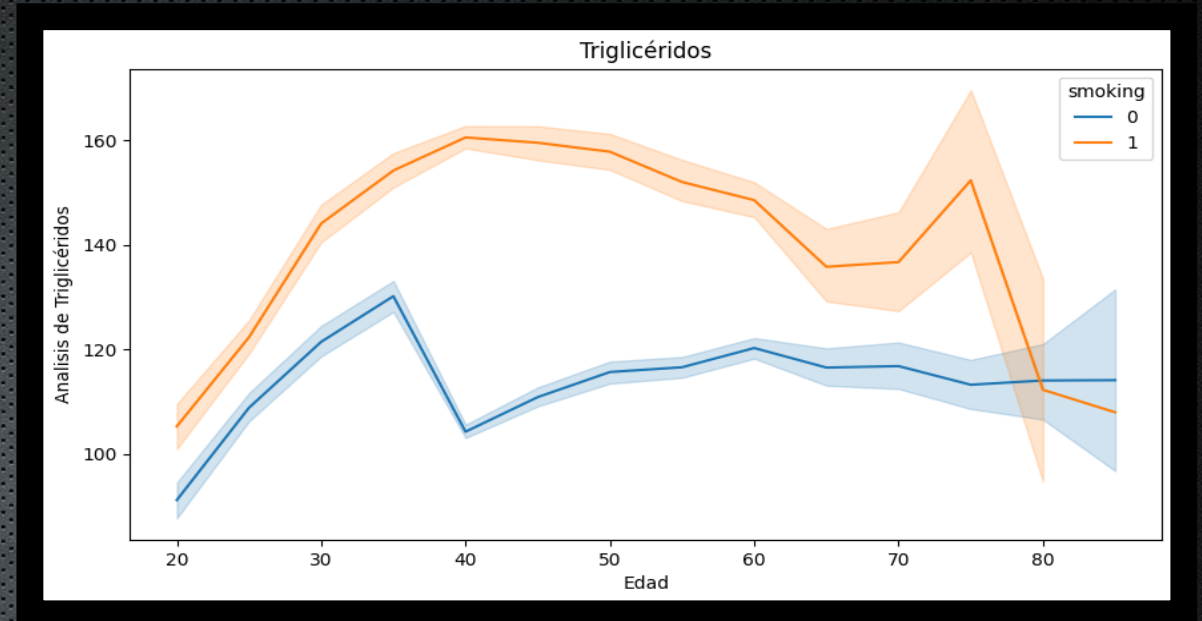
- En este gráfico se muestra cómo a las personas fumadoras les cuesta más relajarse que las personas no fumadoras, en la variable 1 se puede ver en la caja un mayor volúmenes de datos y mayor frecuencia superior .
- Conclusión los fumadores les cuesta mas relajarse por el consumo excesivo de tabaco y en cambio las personas que no consumen tabaco suelen estar mas tranquilos por no recibir sustancias en su cuerpo que provoquen esos cambios.

ANÁLISIS BIVARIADO



Azúcar en la sangre en ayuna

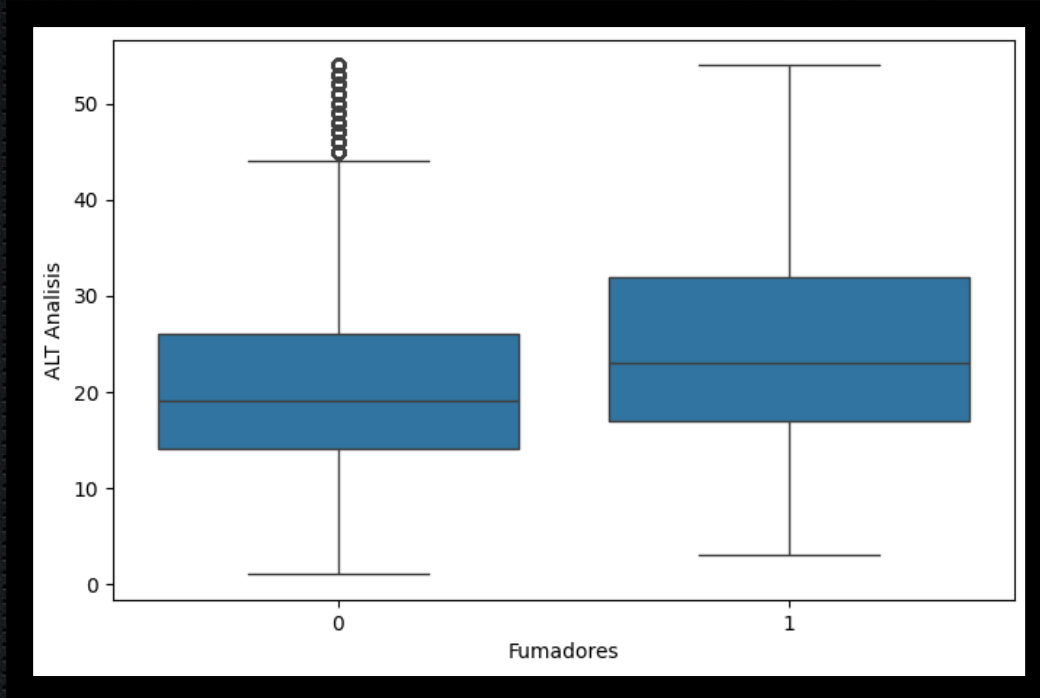
- Se puede ver como las personas que fuman les aumenta los niveles de azúcar en la sangre y eso es muy malo, tal problema es provocado por el tabaco.
- En grafico se puede llegar a la conclusión que los niveles de azúcar en sangre en ayuna son relativamente mas altos que las personas que no consumen tabaco, esto es debido al cambio que produce consumir tabaco.



Triglicéridos

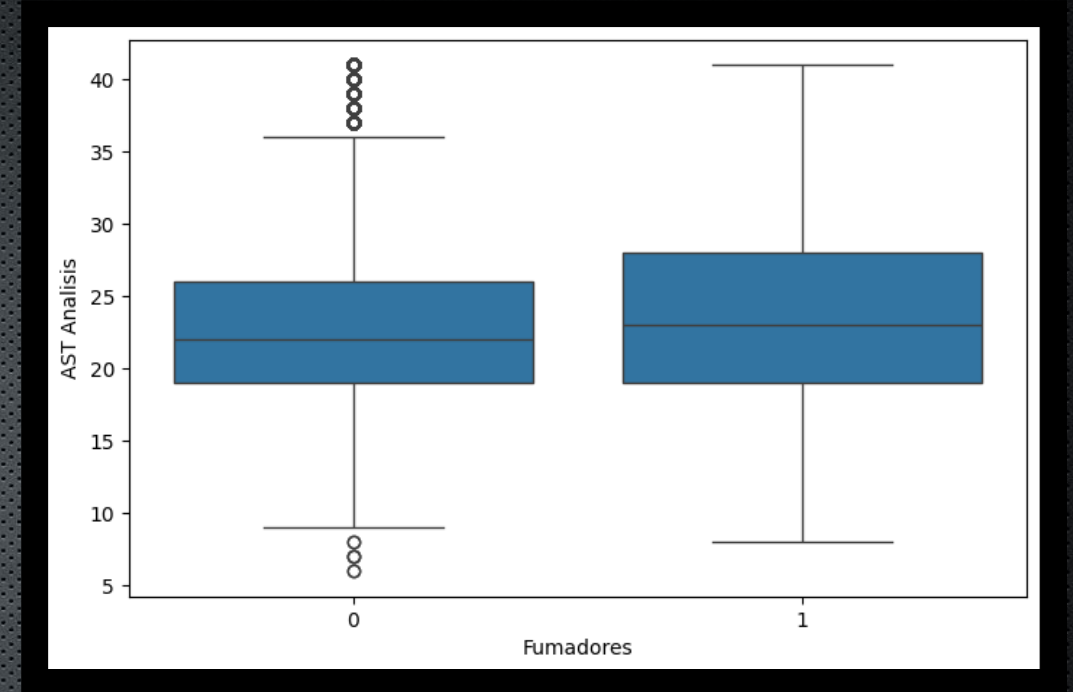
- Valor normal: menos de 145 mg/dL.
- En grafico se puede ver los altos niveles de triglicéridos de los consumidores de tabaco y en cambio las personas que no fuman los niveles son bajos medianamente o normal.
- Los niveles altos de triglicéridos aumenta el riesgo de sufrir un accidente cerebrovascular, ataque cardíaco y cardiopatías.
- Conclusión en este grafico, se puede ver como los niveles triglicéridos son altos y fuera de lo normal en personas consumidoras de tabaco.

ANÁLISIS BIVARIADO



ALT

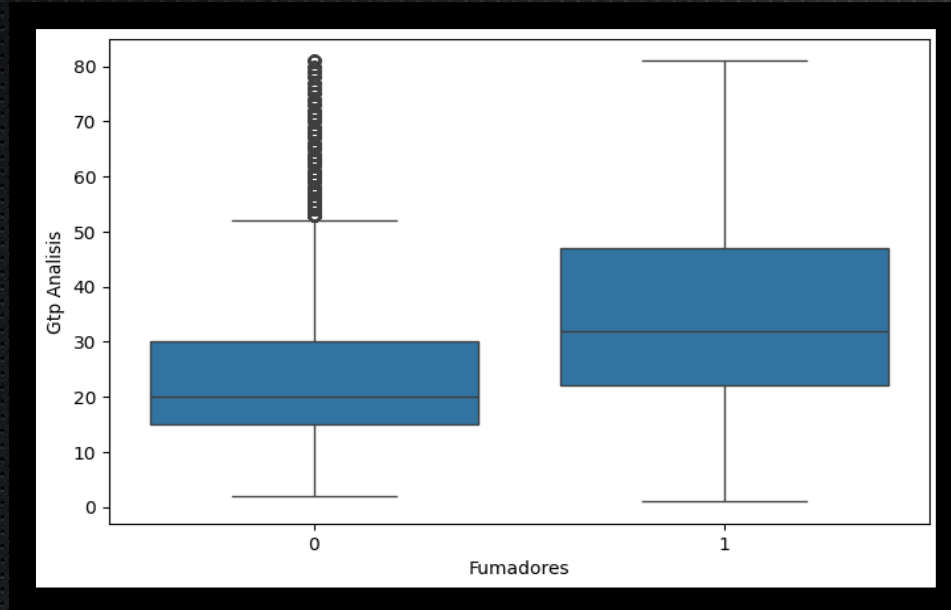
- La prueba de sangre de ALT se suele incluir en un análisis de sangre de rutina para monitorear la salud de su hígado. La prueba también puede ayudar a diagnosticar o monitorear problemas del hígado.
- Hombres: 5 - 50 U/L - No debería pasar de este resultado.
- Mujeres: 5 - 35 U/L - No debería pasar de este resultado.
- En este grafico los fumadores suelen tener los análisis un poco mas arriba de 50 U/L, pero no es mucha diferencia con las personas no fumadoras.



AST

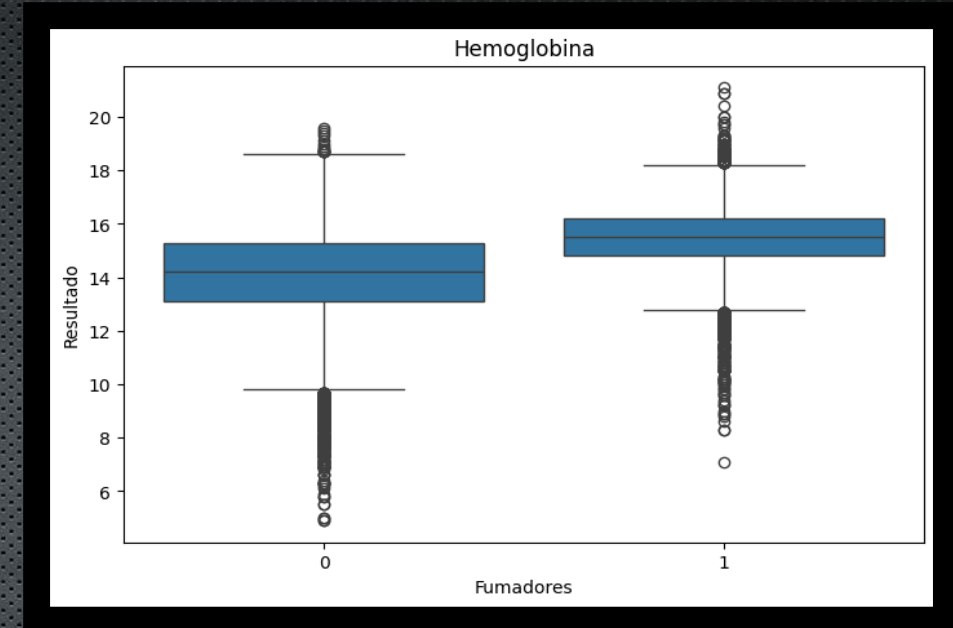
- El análisis de transaminasa (AST) en la sangre se utiliza principalmente para evaluar problemas en el hígado y en determinadas enfermedades que afectan al corazón.
- Hombres: 5 - 50 U/L - No debería pasar de este resultado.
- Mujeres: 5 - 35 U/L - No debería pasar de este resultado.
- En este análisis no se logran ver muchos cambios desmedidos unos a otros, diría que casi no hay cambios en este grafico.

ANÁLISIS BIVARIADO



GTP

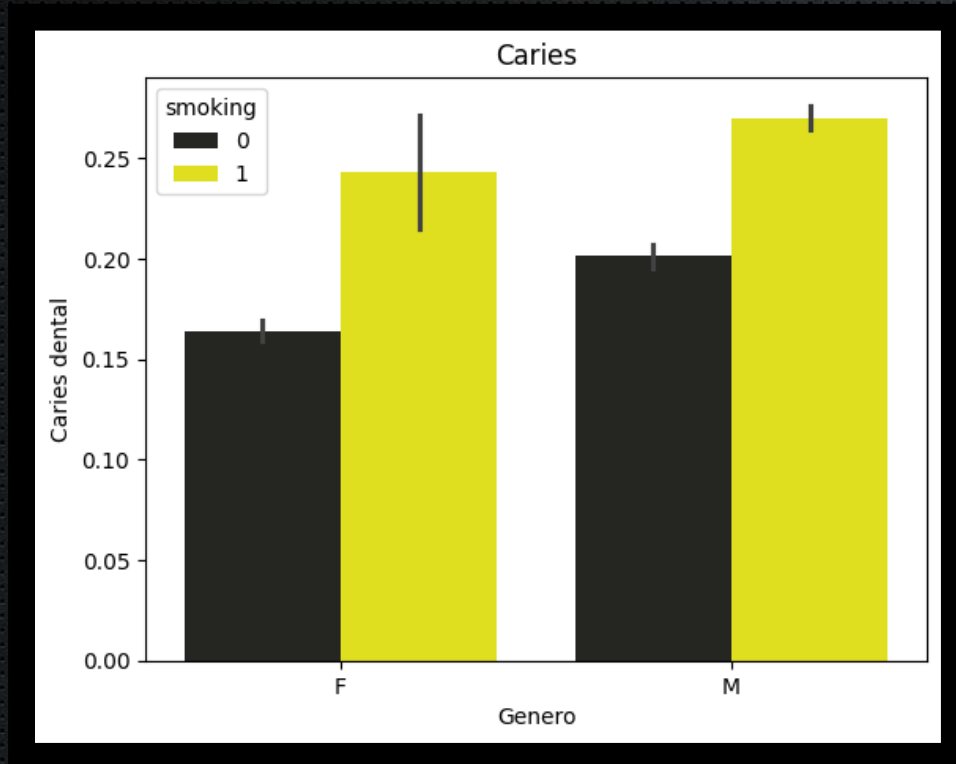
- Hombres: 5 - 50 U/L - No debería pasar de este resultado.
- Mujeres: 5 - 35 U/L - No debería pasar de este resultado.
- En este grafico se puede ver los resultados de los análisis GTP se puede notar la diferencia entre personas fumadoras y no fumadoras.
- El gtp es una prueba de Hígado y de sangre en el que si el resultado es elevado, puede traer problema en el funcionamiento de los órganos. Aunque también es posible que se trate de alguna enfermedad del páncreas o la vesícula.
- En conclusión se puede decir que las personas que son fumadoras le suelen dar mal los análisis de hígado y de sangre a causa del consumo de tabaco.



Hemoglobina

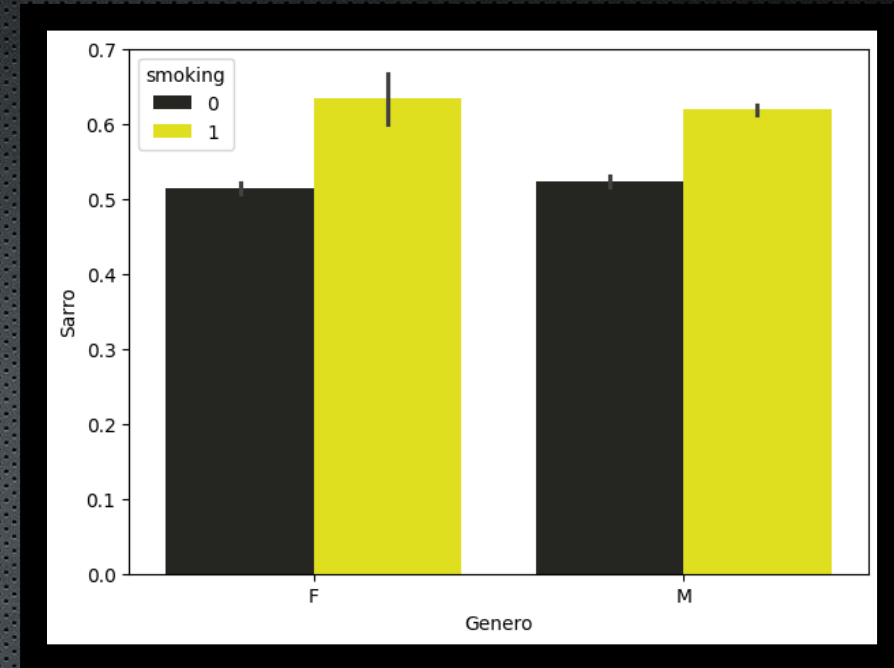
- La hemoglobina es una proteína que se encuentra en los glóbulos rojos y que es la encargada de transportar el oxígeno desde los pulmones hasta el resto de nuestro cuerpo.
- En este grafico se puede visualizar como la Hemoglobina es mucho mas alta en los fumadores a causa del consumo de tabaco.
- Conclusión de este análisis es que los fumadores mayormente tienen la hemoglobina alta por causa del consumo de tabaco y en cambio los no fumadores no suele elevarse tanto la hemoglobina y relativamente esta por el resultado normal o bajo.

ANÁLISIS BIVARIADO



Caries Dental

- El nivel de caries dental en personas que fuman es mayor, a las que personas que no fuman.
- La conclusión de este análisis se podría decir que las personas que fuman suelen tener un volumen mas grandes de caries que las personas que no consumen tabaco.



Sarro

- El sarro del tabaco puede ser un problema común para los fumadores, ya que el humo del cigarrillo puede dejar manchas amarillentas y acumularse en la superficie dental. Esto no solo afecta la apariencia de tus dientes, sino también tu salud bucal en general.
- En este grafico se puede ver como el volumen de sarro de los fumadores suele ser mas que las personas no fumadores.
- La conclusión de este grafico es que la acumulación de sarro es mayor en los fumadores a causa del consumo de tabaco y los tóxicos que tiene.