

Memoria

Análisis de Riesgos de enfermedades en una Aseguradora

1. Contexto

En este análisis, se busca explorar los factores que influyen en el riesgo cardiovascular, inflamatorios, de diabetes y del consumo de tabaco en una aseguradora a partir de una base de datos de clientes. Se ha filtrado un subconjunto de datos correspondientes a personas con riesgo alto para estos problemas de salud, con el fin de identificar posibles patrones y generar recomendaciones para mejorar las políticas de prevención dentro de la aseguradora.

Variables de interés:

- **Riesgo Cardiovascular**
- **Riesgo Inflamatorio**
- **Fumadores**
- **Riesgo de Diabetes**
- **Factores demográficos:** Género, edad, nivel de ingresos, raza y país de nacimiento.

2. Hipótesis

Las hipótesis planteadas en este análisis son:

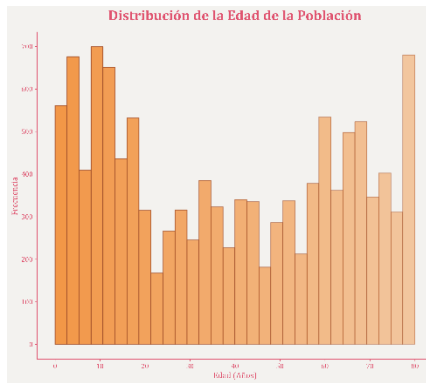
1. Existe una relación significativa entre las características demográficas (como género, nivel de pobreza, raza) y el riesgo de las diferentes enfermedades
2. Las personas con riesgo cardiovascular elevado tienen un mayor índice de enfermedades asociadas, como la diabetes y enfermedades inflamatorias.

3. Análisis Univariante

Objetivo: Analizar la distribución de cada variable individualmente para comprender la naturaleza de los datos.

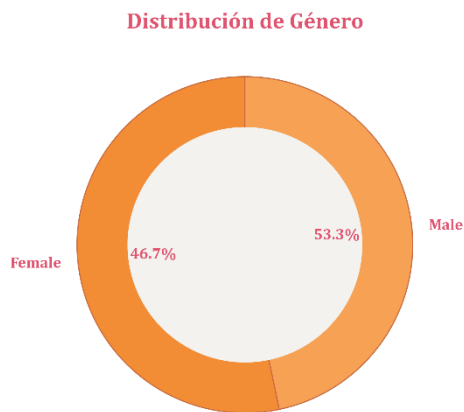
Análisis de Edad:

- Se muestra la distribución de la edad en los individuos de la población
- **Visualización:**



Análisis de Género:

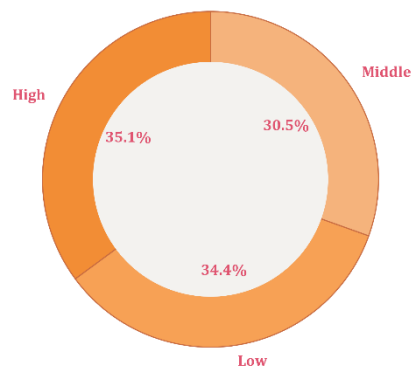
- Se analiza la proporción de hombres y mujeres.
- **Visualización:** Pie chart



Análisis de Nivel de ingresos:

- La variable de ingresos se examina para entender la distribución de los pacientes en diferentes categorías socioeconómicas.
- **Visualización:** Pie chart

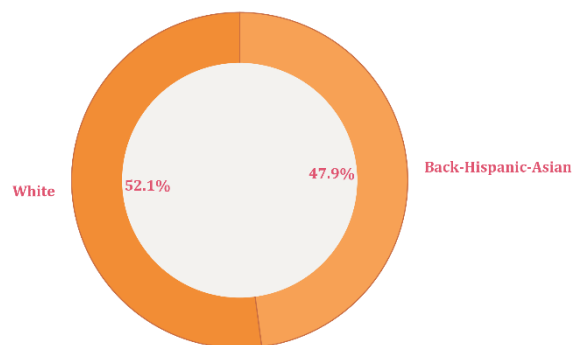
Distribución de Ingresos/Pobreza



Análisis de la raza de origen:

- La variable de raza se examina para entender la distribución de los pacientes.
- **Visualización:** Pie chart

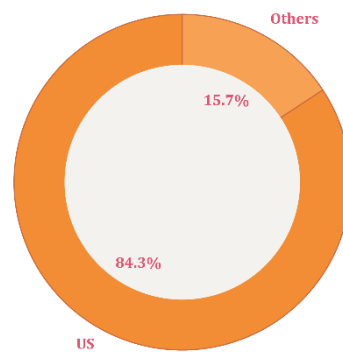
Distribución de Razas



Análisis del País de nacimiento:

- La variable de país de nacimiento se examina para entender la distribución de los pacientes.
- **Visualización:** Pie chart

Distribución de País de Nacimiento

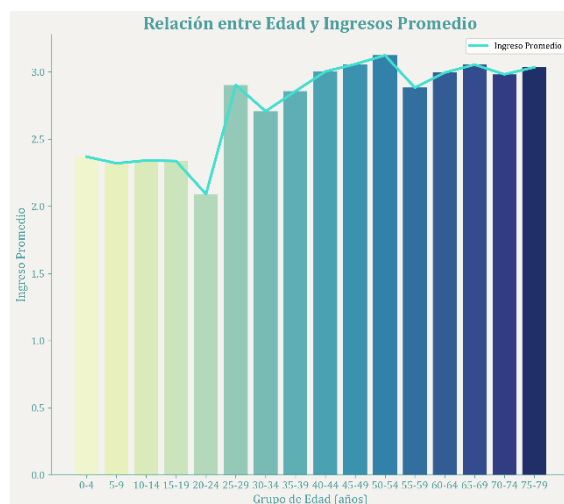


4. Análisis Bivalente

Objetivo: Evaluar la relación entre dos variables y determinar si hay alguna correlación o diferencia significativa.

Análisis entre Edad y Nivel de Pobreza:

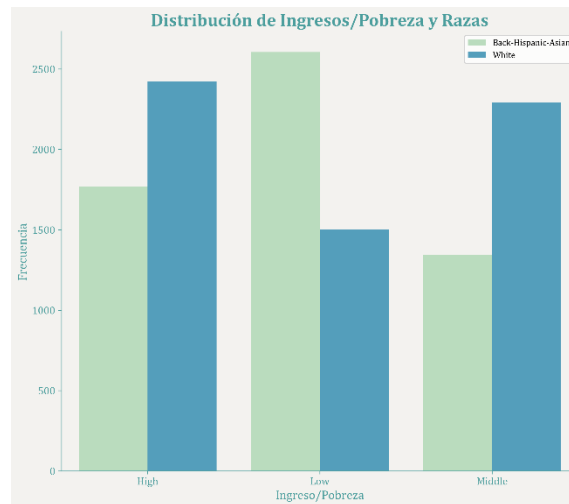
- Se explora si existe una diferencia significativa en el nivel de pobreza en función de la edad de los individuos.
- **Visualización:** Histograma que muestra la distribución de la edad en función del nivel de pobreza. Se puede utilizar para encontrar los grupos de edad con mayores ingresos, interesantes para la compañía.



Análisis entre Raza y Nivel de Pobreza:

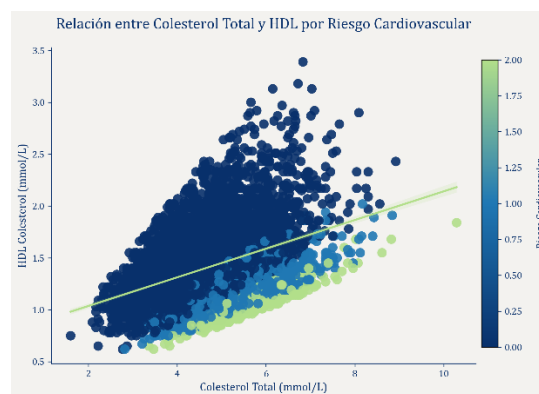
- Se analiza si la raza tiene alguna influencia en la distribución de la pobreza.

- **Visualización:** Barplot para visualizar la proporción de individuos de diferentes razas en función de su nivel de pobreza.



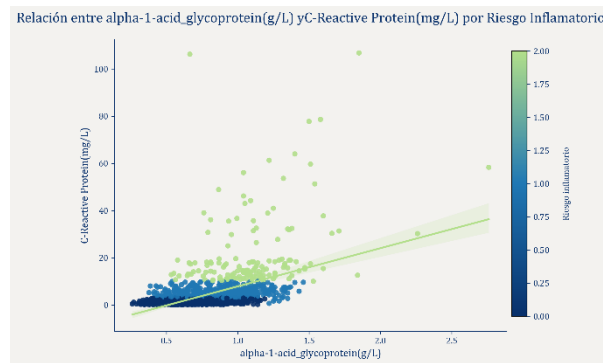
Análisis entre HDL y Colesterol total:

- Se analiza para saber si ambas variables tienen relación establecer un índice que las relacione.
- **Visualización:** Scatterplot para ver si tienen una relación positiva.



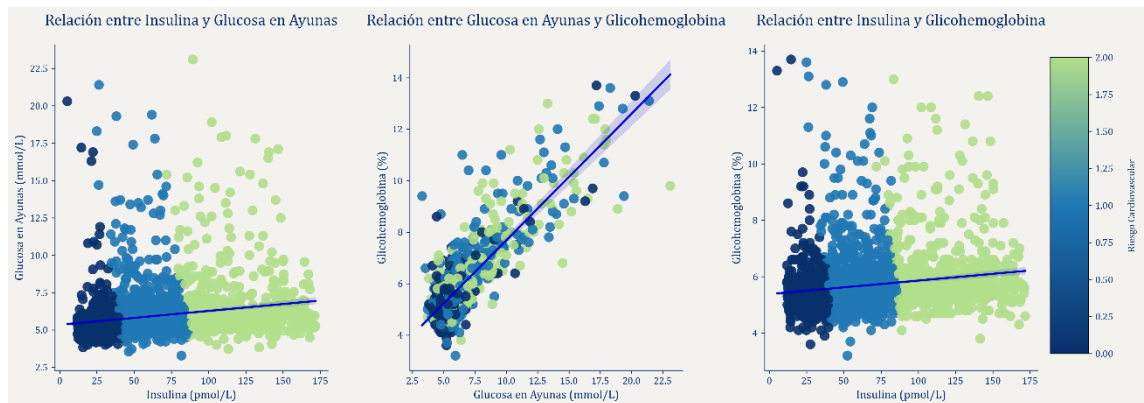
Análisis entre C-Reactin y Alpha-glicoprotein total:

- Se analiza para saber si ambas variables tienen relación establecer un índice que las relacione.
- **Visualización:** Scatterplot para ver si tienen una relación positiva.



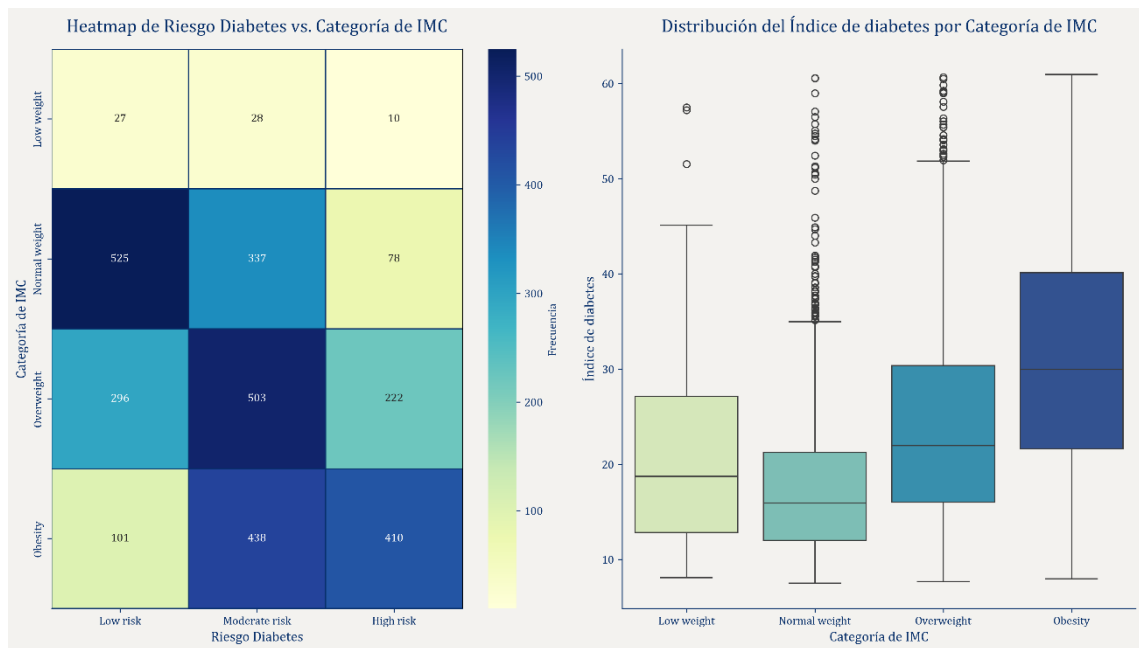
Análisis entre Insulin, Glicohemoglobin y Fasting glucose:

- Se analiza para saber si las variables tienen relación y establecer un índice que las relacione.
- **Visualización:** Scatterplot para ver si tienen una relación positiva.



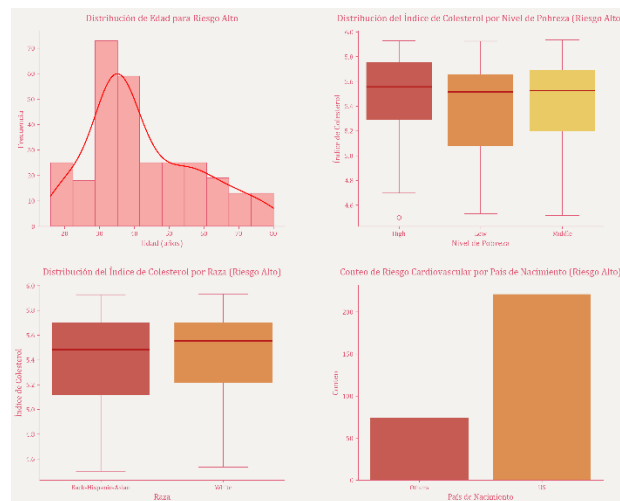
Análisis entre todas las enfermedades de riesgo y sus índices con el IMC:

- Se analiza para saber si las enfermedades de riesgo tienen relación con el factor de riesgo obesidad y viceversa.
- **Visualización:** Heatmap para ver cuánto contribuye las variables a cada una de las categorías y Boxplot para ver cuál es la categoría del IMC más afectada en función del índice de la enfermedad de riesgo.



Análisis entre todas las enfermedades de riesgo y las diferentes variables demográficas:

- Se analiza para saber que grupos demográficos pudieran estar más afectados por estas enfermedades y cuales supondría costes altos para la aseguradora.
- Visualización:** Pie chart, Histograma, Barplot y Boxplot.



5. Conclusiones

A partir de los análisis realizados, podemos destacar lo siguiente:

- Más del 60% de los clientes encuestados tienen obesidad o están cerca de padecerla, riesgo que esta creciendo a medida que pasan los años.**

2. Identificamos una serie de grupos poblacionales para cada uno de los riesgos estudiados:

- a. Cardiovascular:** Grupo de edad entre 30 y 40 años con ingresos altos.
- b. Diabéticos:** Entre los 60 y 70 años, Hombres nacidos en EEUU con raza asiática, negra o hispanica.
- c. Inflamatorias:** Grupo de edad entre los 30 y 45 años de nacionalidad EEUU con ingresos bajos.
- d. Fumadores:** Entre los 60 y 70 años, de raza blanca y nacionalidad EEUU con ingresos bajos.

6. Recomendaciones

Basado en los hallazgos, las siguientes recomendaciones son pertinentes para la aseguradora:

- 1. Intervención dirigida a la edad:** Diseñar programas preventivos especialmente dirigidos a los grupos de edad de los diferentes riesgos.
- 2. Programas de sensibilización sobre el riesgo de cada una de las enfermedades:** Crear campañas publicitarias para sensibilizar a la población de riesgo sobre las enfermedades cardiovasculares e inflamatorias.
- 3. Continuar explorando el resto de los cuestionarios facilitados a los clientes, así como otras variables demográficas descartadas en este primer estudio.**
- 4. Contactar a los grupos que se ha confirmado que pueden padecer estos riesgos e implementar un seguimiento de sus casos a modo de prevención.**