



Universidad del Desarrollo
Facultad de Ingeniería

Determinantes sociales en la prevalencia de enfermedades crónicas en Estados Unidos

Comportamientos de riesgo en la población adulta de Estados Unidos

POR: CÉSAR RIVERA SERRANO

Proyecto de grado presentado a la Facultad de Ingeniería de la Universidad del
Desarrollo para optar al grado académico de Magíster en Data Science

PROFESOR GUÍA:

Dra. Loreto Bravo

Diciembre 2023

SANTIAGO

AGRADECIMIENTO

Agradezco profundamente a mi familia, en especial a mi esposa Gabriela Robles, quien con su apoyo incondicional ha permitido que este proyecto se transforme en una realidad. Su confianza en mis capacidades ha sido una fuente constante de motivación. También quiero agradecer a mis hijos Sol y Dante Rivera, quienes han cedido parte de su tiempo conmigo para acompañarme en este viaje.

Finalmente, quiero agradecer a todos aquellos que directa o indirectamente han contribuido a la realización de este proyecto. Cada uno de ustedes ha dejado una huella imborrable en mi camino académico.

TABLA DE CONTENIDO

RESUMEN.....	1
1. INTRODUCCIÓN	3
2. TRABAJO RELACIONADO	6
2.1. ESTRUCTURA SOCIAL Y HABITUS:.....	6
2.2. EL HABITUS Y LOS DETERMINANTES SOCIALES DE LA SALUD (DSS).....	9
2.3. DEFINICIÓN Y CLASIFICACIÓN DE LOS DETERMINANTES SOCIALES DE LA SALUD (DSS)	11
2.4. RELACIÓN ENTRE DSS Y SALUD	16
2.5. RELEVANCIA GEOGRÁFICA Y TEMPORAL DE LOS DSS	17
2.6. METODOLOGÍAS COMUNES PARA ESTUDIAR LOS DSS	17
3. HIPÓTESIS Y OBJETIVOS.....	21
3.1. HIPÓTESIS.....	21
3.2. OBJETIVOS	21
4. METODOLOGÍA	22
4.1. DATOS	22
4.2. METODOLOGÍA.....	55
5. RESULTADOS	59
5.1.1. MODELO DE RIESGO DE TENER UN MAL ESTADO DE SALUD FÍSICO Y MENTAL.....	59
5.1.2. MODELO DE RIESGO DE PADECER DIABETES.	61
5.1.3. MODELO DE COBERTURA DE SALUD.....	64
5.1.4. MODELO DE RIESGO DE MULTIMORBILIDAD	66
5.1.5. MODELO DE RIESGO DE ADICCIÓN AL TABACO.	68
5.1.6. MODELO DE RIESGO DE ADICCIÓN AL ALCOHOL.	71
5.1.7. MODELO DE ESTADO DE SALUD GENERAL.....	73

5.1.8.	BOSQUES ALEATORIOS.....	76
6.	CONCLUSIONES	78
7.	BIBLIOGRAFÍA.....	80
8.	ANEXOS.....	82
8.1.	CONTENIDO DEL CONJUNTO DE DATOS.....	82
8.2.	CÓDIGO DESARROLLADO EN PYTHON.....	84

Resumen

El enfoque principal de este proyecto se centra en la propuesta de que los determinantes sociales desempeñan un papel crucial en la aparición de enfermedades crónicas y conductas de riesgo entre los adultos en los Estados Unidos de 2017 a 2021. El objetivo principal es identificar y medir el alcance de qué factores como el empleo, los ingresos y el nivel educativo afectan la salud pública.

Los datos utilizados en este estudio se obtuvieron del Sistema de Vigilancia de Factores de Riesgo del Comportamiento (BRFSS), una encuesta integral que recopila datos relacionados con la salud de personas que residen en los Estados Unidos. Para analizar el conjunto de datos, se construyó un conjunto de modelos de clasificación, identificando y priorizando las variables pertinentes en función de su importancia y aplicabilidad para la investigación.

Se generaron 6 árboles de decisión como modelos de clasificación de la relación de algunas variables sociodemográficas (Sexo, estado civil, nivel educacional, propiedad del hogar, empleo, menores de 18 años viviendo en el hogar, ingreso, zona (cercanía de la vivienda a la zona metropolitana), raza, embarazo) relacionados con en torno a distintos riesgo tales como tener un mal estado de salud físico y mental, padecer diabetes, acceso a cuidados de salud, tener multimorbilidades, adicción al tabaco, adicción al alcohol; un séptimo árbol de decisión se construyó para evaluar la relación entre las variables sociodemográficas y el estado general de salud. Finalmente se generó un bosque aleatorio que determinó que el empleo y los ingresos tienen el impacto más significativo en la salud como determinantes sociales, seguidos de cerca por el nivel educativo. El modelo demostró una fuerte capacidad de predicción de los resultados de salud, con una tasa de precisión del 70,1%, lo que enfatiza aún más la importancia de estas variables para determinar el bienestar general.

Los hallazgos indicaron que la salud de la población está fuertemente influenciada por los determinantes sociales. Se identificaron factores relacionados con la economía y la educación como indicadores importantes de enfermedades crónicas y conductas de riesgo. El proyecto reveló una fuerte correlación entre el estatus social y el bienestar, lo que enfatiza aún más la necesidad de implementar políticas dirigidas a estos dominios para mejorar la salud de la población general. Las implicaciones de estos hallazgos son directamente relevantes para el desarrollo de políticas de salud pública, lo que indica que las intervenciones deben priorizar la mejora de los factores socioeconómicos para mejorar efectivamente la salud general de la población. El estudio enfatiza el papel fundamental

de los enfoques integrados que integran la atención médica con la mejora de las circunstancias sociales.

1. Introducción

A medida que avanzamos como sociedad, hemos ido profundizando el entendimiento de que las relaciones entre individuos inciden en nuestra forma de vida; que desde nuestras percepciones y las relaciones que se establecen a partir de ellas, modificamos y construimos perspectivas que impactan en el mundo. Un ejemplo claro de esta forma transformación de la realidad, es visible a partir de los distintos elementos que conforman al estado y que basan sus objetivos y actuaciones desde las políticas públicas, incidiendo en la vida de los ciudadanos (es decir el conjunto de actores relaciones que forman parte del país en el que ese Estado se asienta). Durante el siglo XX, se hace evidente que muchas de estas relaciones ya no solo afectan a un conjunto concreto de individuos; sino que, en mayor o menor medida, a cada uno de nosotros en todo el mundo. En este contexto; una serie de organizaciones mundiales se crearon, enfocadas en enfrentar el hambre, la guerra y la enfermedad, entre otras calamidades, intencionando el horizonte utópico de posibilidades para la humanidad. En consecuencia, en 1945 los representantes de 51 países sentaron las bases para la creación de la organización de naciones unidas, cuyo objetivo consiste en mantener la paz y la seguridad internacional, así como defender los derechos humanos; fomentando las relaciones de amistad entre las naciones, impulsando la cooperación internacional para la solución de los problemas económicos, sociales, culturales y humanitarios. La gran mayoría de estas organizaciones mundiales, se crearon en el momento histórico posterior a la segunda guerra mundial, momento en que el hambre, las enfermedades y las atrocidades cometidas durante el período de la guerra se hicieron evidentes y se repudiaron en forma generalizada. Fue en este período que nacieron organizaciones tales como la Organización de las Naciones Unidas para la Alimentación y la Agricultura, el banco mundial, el fondo monetario internacional, el organismo internacional de energía atómica, la organización meteorológica mundial, La organización mundial de la salud, etc.¹

Entre las organizaciones mencionadas en el punto anterior, cobra vital importancia la organización mundial de la salud OMS, organización internacional especializada en salud pública de las Américas, que, junto con los países miembros de la región, colabora y facilita los mecanismos para combatir las enfermedades, los padecimientos crónicos y para fortalecer los sistemas de salud ante situaciones de emergencia. Fortaleciendo y apoyando el derecho al acceso a la salud de los individuos a través del fomento de la cooperación entre países y trabajo colaborativo con los ministerios de salud, agencias de gobiernos, organizaciones de la sociedad civil, agencias internacionales, universidades,

¹ Organización de las Naciones Unidas. (s. f.). Historia de las Naciones Unidas, disponible en <https://www.un.org/es/about-us/history-of-the-un>

organismos de la seguridad social, grupos comunitarios y otros con el fin de que las personas puedan tener una vida más larga y sana.²

En la búsqueda del objetivo mencionado en el párrafo anterior, La OPS, una institución vital para promover la salud y el bienestar en las Américas se estableció a principios del siglo XX. Sus orígenes se remontan a la Primera Conferencia Internacional Americana en 1889, cuyo objetivo era fomentar la colaboración entre las naciones en cuestiones de salud. Como resultado, en 1902 se creó la precursora de la OPS, la Oficina Sanitaria Internacional. Los principales objetivos de la OPS giran en torno a mejorar el bienestar y elevar la calidad de vida de las personas que residen en las Américas. Como división regional de la Organización Mundial de la Salud (OMS), la OPS colabora estrechamente con sus países miembros en la lucha contra las enfermedades, la mejora de la pureza del agua y el saneamiento y el avance de estrategias de salud pública justas y eficientes. Además, la OPS asume un papel fundamental al abordar crisis y brotes de salud brindando apoyo técnico y facilitando la cooperación internacional. A lo largo de su existencia, la OPS ha desempeñado un papel fundamental en la eliminación de enfermedades como la viruela y la polio en la región, logrando avances sustanciales en la lucha contra enfermedades transmisibles y no transmisibles, mejorando la salud materna e infantil y fortaleciendo los sistemas de salud. La inquebrantable dedicación de la OPS para garantizar la equidad en salud y mejorar el acceso a la atención médica solidifica su condición de institución vital que impulsa el progreso de la salud pública en el hemisferio occidental.

La OPS toma en consideración un sinnúmero de elementos que inciden en la salud de las personas, desde prácticas tales como vacunaciones, hábitos de alimentación saludables, práctica de ejercicios, etc. Elementos cuyo origen son de carácter concreto, en relación directa y evidente con las enfermedades. Sin embargo, existen una serie de elementos de carácter social que inciden también en el estado de salud.³

Según la Organización Mundial de la Salud, los determinantes sociales de la salud (DSS) como "las circunstancias en que las personas nacen crecen, trabajan, viven y envejecen,

² "Brief History of WHO", Columbia University Case Consortium, disponible en: https://ccnmtl.columbia.edu/projects/caseconsortium/casestudies/112/casestudy/www/layout/case_id_112_id_776.html

³ Organización Panamericana de la Salud. (2017). Documentos Básicos de la Organización Panamericana de la Salud, Decimonovena edición. Recuperado de <https://www.paho.org/es/documentos/documentos-basicos-organizacion-panamericana-salud-decimonovena-edicion>

incluido el conjunto más amplio de fuerzas y sistemas que influyen sobre las condiciones de la vida cotidiana". Estas fuerzas y sistemas incluyen políticas y sistemas económicos, programas de desarrollo, normas y políticas sociales y sistemas políticos. Las condiciones anteriores pueden ser altamente diferentes para varios subgrupos de una población y pueden dar lugar a diferencias en los resultados en materia de salud. Es posible que sea inevitable que algunas de estas condiciones sean diferentes, en cual caso se consideran desigualdades, tal como es posible que estas diferencias puedan ser innecesarias y evitables, en cual caso se consideran inequidades y, por consiguiente, metas apropiadas para políticas diseñadas para aumentar la equidad.

En este proyecto se intentará armar un modelo que permita interpretar que determinantes sociales principales inciden en el estado de salud; También permitiría clasificar una prioridad de atención a partir de estos condicionantes.

Para este proyecto, se utilizará datos provenientes de la encuesta anual "Behavioral Risk Factor Surveillance System"⁴ una de las encuestas de salud realizada en los Estados Unidos; esta es una de las encuestas más grandes realizadas en el mundo e inicia en 1986. Es una encuesta telefónica, que hacia el 2009 incorpora los teléfonos celulares. El cuestionario incorpora una serie de preguntas en torno a determinado tópicos tales como frecuencia en la que se realiza ejercicios, salud mental, salud física, diabetes, problemas cardiovasculares, etc. Incorporando a su vez, preguntas sobre el sexo, nivel de educación, ingreso, niveles de acceso a servicios de salud entre otras variables sociodemográficas. Para este proyecto se utilizaron los resultados de las encuestas desde el 2017 hasta el 2021

⁴ Centers for Disease Control and Prevention. (2023). Behavioral Risk Factor Surveillance System Annual Data. Recuperado de https://www.cdc.gov/brfss/annual_data/annual_data.htm

2. Trabajo Relacionado

2.1. Estructura social y habitus:

En la teoría sociológica de Bourdieu, el mundo social no sólo se compone de estructuras objetivas sino también de representaciones, percepciones y visiones. De esta forma, existe una correspondencia entre las estructuras sociales y las estructuras mentales; en la que los sistemas simbólicos contribuyen a construir el mundo y, por tanto, las categorías de la percepción del mundo social son el producto de la incorporación de las estructuras objetivas del marco social; Lo que implica que toda visión de mundo opera bajo las coacciones estructurales objetivas, estableciendo una correspondencia entre las estructuras material y mentales. “En el trabajo de Pierre Bourdieu, lo social muestra una doble existencia: se manifiesta tanto en las estructuras independientes de la conciencia y de la voluntad de los agentes individuales, grupales, clases o sectores (estructuras objetivas), como en los compendios de percepción, de pensamiento, de acción que componen socialmente nuestra subjetividad (estructuras subjetivas). Las estructuras objetivas tienen la capacidad de dirigir y obligar las prácticas sociales y las representaciones que de las mismas se hacen los individuos o agentes sociales”⁵

En concreto, la correspondencia entre estructura material y estructura mental significa que el mundo material se ve reflejado en el mundo mental; pero también, que el mundo mental se ve reflejado en el mundo material. Se puede ejemplificar este proceso en el tratamiento de una enfermedad. El mundo material indica que la persona tiene una enfermedad en tanto que manifiesta síntomas concretos que pueden ser observados (tos, fiebre, etc.) estos síntomas hacen que la persona que los sufre se acerque a un médico, quien observa los síntomas para luego categorizarlos desde el mundo mental (percibe los síntomas y esboza una teoría en la que los síntomas coinciden con cierta teoría médica que los explica, esta teoría se encuentra en el mundo mental). El médico, entonces, basado en esa teoría mental, establece un determinado tratamiento médico (estableciendo una correspondencia desde el mundo mental hacia el mundo material) estableciendo una correspondencia entre ambas estructuras.

Ambas estructuras (mentales y materiales) se encuentran inscritas en las estructuras sociales, las que le otorgan un espacio en el cual desarrollarse, Así, el tratamiento descrito en el párrafo anterior, se encuentra inscrito al interior de las estructuras sociales, puesto que desarrollará una serie de acciones que solo podrán tener significado al estar inscritas

⁵ Barrera Sánchez, O. (2011). El cuerpo en Marx, Bourdieu y Foucault. Iberóforum. Revista de Ciencias Sociales de la Universidad Iberoamericana., 121-137. Pag.127.

en un marco social determinado (en el mundo científico moderno, se recetara una serie de medicamentos destinados a sanar al paciente) pero estas acciones cambiarían radicalmente, en otro marco social (un sanador ugandés recetaría el hablar con el espíritu de una serpiente pitón, tal como su marco social lo indica).

Adicionalmente a lo anterior, esta correspondencia entre mundo material y mundo mental toma asidero en el cuerpo; puesto que las acciones que permiten relacionar el mundo mental y el mundo material solo pueden ocurrir a partir del cuerpo. Así, en el ejemplo de la enfermedad mencionada, es por medio del cuerpo que se toman remedios o se sufren operaciones; así como en el ejemplo del sanador ugandés, es por medio del cuerpo por el cual se realiza el ritual para hablar con el espíritu que vive en una pitón y así, toda incidencia en el mundo material solo puede ocurrir por medio del cuerpo.

Es el cuerpo en el que las correspondencias entre el mundo material y el mundo mental son elaboradas como sistemas de disposiciones duraderas y transferibles, como estructuras estructuradas predispuestas para funcionar como estructuras estructurantes, es decir, como principios generadores y organizadores de prácticas y representaciones que pueden estar objetivamente adaptadas a su fin. “Así, en el trabajo de Pierre Bourdieu el cuerpo humano es considerado o leído como un producto social y, por tanto, irrumpido por la cultura, por relaciones de poder, las relaciones de dominación y de clase.”⁶

Es a partir de la reproducción de estas estructuras, que al devenir en estructurantes se convierten en una práctica que va más allá del acto consciente (no hacemos consciente el cruzar una determinada calle cuando el semáforo enciende la luz verde, ni analizamos el mecanismo que hace que tomemos una aspirina cuando nos duela la cabeza) las estructuras devienen en estructurantes en tanto y en cuanto se establecen como una serie de reglas que rigen nuestra cotidianeidad y nos llevan a naturalizar una serie de prácticas que se establecen desde como objetivamente reguladas y regulares. Todas estas prácticas, que finalmente se establecen más allá de la conciencia y más allá del individuo, es a lo que Bourdieu denominó como “habitus”.

El “habitus”, es decir, la serie de representaciones que relacionan el mundo material y el mundo mental y que ocurren en forma de acciones en el cuerpo, que además están relacionadas con las acciones de otros sujetos, vinculando las prácticas sociales y las prácticas de los agentes (actores sociales), funcionando en todo momento como una matriz

⁶ Barrera Sánchez, O. (2011). El cuerpo en Marx, Bourdieu y Foucault. Iberóforum. Revista de Ciencias Sociales de la Universidad Iberoamericana., 121-137. Pag.129.

de percepciones, de apreciaciones y de acciones. El conjunto de percepciones, apreciaciones y acciones, constituyen el entramado social, que se reproduce en los distintos aspectos de las vidas de las personas, la forma que adquiere el trabajo, la crianza, las relaciones sociales y todos los aspectos que adquiere la vida.

Por lo que podríamos decir que la realidad material se ve constantemente estructurada a partir de las percepciones sociales; que estas percepciones se organizan y reproducen constantemente en la cotidianidad, es decir que se transforman en prácticas sociales y como tales, inciden en los aspectos más concretos de la realidad, inscribiéndose en el cuerpo a partir de las acciones habituales que realiza el individuo inscrito en ese marco social, reproduciendo las prácticas deseables y validadas por el entorno social inmediato, en primer lugar, y por el entorno social general, en segunda instancia.

Finalmente se puede afirmar que las prácticas sociales inciden en el cuerpo; pues es en el cuerpo donde aparecen las marcas de la práctica social determinada, (como ejemplo, realizar un determinado deporte repercutirá en aspectos tan concretos y observables como son el bajar de peso, el sufrir determinadas lesiones según el deporte que se practique, el distribuir los tiempos de determinada forma, modificando los momentos que se realice alguna acción concreta, en cómo se recorre la ciudad, en la cantidad de alimento que se ingiera, etc.) y, por tanto, podemos asegurar que existe una concordancia, que no está determinada por la obediencia a una regla, sino que interiorizada por el sujeto y reproducida a su entorno inmediato mediante sus actos.

El “habitus” como generador de prácticas sociales, es decir, incidiendo en la realidad material mediante la conducta humana, establece un vínculo indisoluble entre lo biológico y lo social; desde la forma y condiciones en la que se realiza un trabajo, hasta los hábitos alimentarios, deportivos, de esparcimiento, etc. Así, todo lo que el ser humano produce, crea, expresa, piensa, reproduce o transforma; lo hace como ente social, alterando los ambientes físicos y culturales con el fin de satisfacer las necesidades productivas y reproductivas; en estas prácticas se constituyen distintos tipos de prácticas sociales que pueden ser agrupadas en campos; estos campos poseen cierta independencia entre unos y otros, pero pueden interactuar bajo los aspectos más generales de las prácticas sociales; a estos campos solemos denominarlos como económico, político, científico. De esta forma, si bien el campo político es independiente del campo económico, a la vez se encuentran fuertemente relacionados; de la misma forma se relacionan con el campo religioso, científico y cualquier otro aspecto de las prácticas sociales.

Uno de los campos mencionados en el párrafo anterior, es el campo de la salud, que al igual que los demás, también es independiente de los demás, pero a su vez se encuentra

profundamente influenciado por los campos económico, político, religioso, científico, deportivo, cultural, etc.

Este sistema de relaciones estructurales constituidos por los esquemas de percepción, evaluación y acción, que producen prácticas sociales actuadas en el campo social, internalizadas y naturalizadas mediante el habitus, y al mismo tiempo, reproducidas por él, como un principio unificado.

Es decir que, según Pierre Bourdieu, el habitus es un conjunto de tendencias duraderas y transferibles que se arraigan en los individuos a través de sus experiencias de vida. Estas tendencias moldean cómo las personas perciben, valoran y actúan en el mundo.

2.2. El habitus y los Determinantes Sociales de la Salud (DSS)

El concepto de habitus destaca cómo las circunstancias sociales y personales se internalizan y luego afectan la forma en que los individuos ven e interactúan con el mundo que los rodea. Mientras que, por su parte, los Determinantes Sociales de la Salud se refieren a las circunstancias en las que los individuos nacen, se desarrollan, habitan, trabajan y maduran. Estos abarcan también el sistema de salud. La distribución del dinero, el poder y los recursos en todos los niveles, incluidos el global, el nacional y el local, da forma a estas circunstancias.

El habitus refleja la internalización de las estructuras sociales dentro de los individuos, incluidos los determinantes sociales de la salud. Estos determinantes son las condiciones y estructuras precisas que impactan la salud. Las condiciones socioeconómicas, el acceso a la atención sanitaria y la educación, y el entorno comunitario y laboral se internalizan y pasan a formar parte del habitus de un individuo.

Los hábitos y disposiciones de una persona pueden tener un profundo impacto en sus actitudes hacia la salud, la forma en que afrontan los problemas de salud y sus interacciones con el sistema de salud. De este modo, una persona que ha experimentado un acceso limitado a una educación y atención médica de calidad durante su educación puede llegar a considerar ciertas enfermedades como inevitables o mostrarse escéptica respecto de las instituciones de atención médica.

Bourdieu postuló que el habitus es instrumental en la preservación de las estructuras sociales ya que los individuos actúan en base a sus predisposiciones interiorizadas. En el

ámbito de la salud, esto podría traducirse en que las personas de comunidades marginadas experimenten perpetuamente resultados de salud inferiores debido a nociones y acciones preconcebidas arraigadas en sus hábitos.

Para crear intervenciones que sean más eficientes, es importante comprender la conexión entre el habitus y los determinantes sociales de la salud. Por ejemplo, si el objetivo es mejorar la salud de una comunidad marginada, simplemente proporcionar servicios médicos no es suficiente. También es imperativo abordar las tendencias internalizadas (habitus) que pueden estar impidiendo que las personas busquen asistencia médica o cumplan con los consejos médicos.

En resumen, la noción de habitus en la filosofía de Bourdieu ofrece una comprensión profunda de cómo las estructuras y condiciones sociales, incluidos los determinantes sociales de la salud, se asimilan en los individuos e influyen en sus actitudes, acciones y, en última instancia, en su bienestar. Reconocer esta conexión podría ser crucial para formular estrategias de salud pública que sean más eficientes y justas.

Estos conceptos, en el que la relación del cuerpo se ve mediada por el poder, sus instituciones y las formas que de control que cada una de ellas adquiere; así como las prácticas que generan en la modernidad es definida por Foucault a partir de “un ajuste de las prácticas de confesión cristianas, el cuerpo comienza a ser el baluarte de los pecados, razón por la cual debe ser interrogado y puesto en palabras por cada individuo. El sexo enclaustra una verdad que debe ser descubierta a través de técnicas que lo objetivan. Pero lo seductor es que la clave de esta extracción se encuentra en la interpretación por parte de aquel que escucha; capacitado y autorizado para descifrar el código encriptado de la verdad sexual. Así, el cuerpo se inserta en unas relaciones de poder al confesarse el sujeto ante aquellos que ostentan las claves interpretativas: los sacerdotes, los médicos, los psiquiatras y los científicos sociales.”⁷

El campo de la medicina ha experimentado importantes avances. El desarrollo de un cuerpo depende en gran medida de aspectos fundamentales. Los efectos de este asunto son tanto individuales como sociales. No se puede subestimar la importancia de las humanidades y las ciencias. El campo de las ciencias suele abarcar una amplia gama de cuestiones sociales. La salud tiene el potencial de proporcionar mayores beneficios. El elemento de autoconciencia se vuelve crucial cuando nos enfrentamos a la aplicación

⁷ Barrera Sánchez, O. (2011). El cuerpo en Marx, Bourdieu y Foucault. Iberóforum. Revista de Ciencias Sociales de la Universidad Iberoamericana., 121-137. Pag.134.

práctica del conocimiento, un concepto que tiene importancia para todos los individuos. A medida que la hegemonía continúa prevaleciendo, la necesidad de ello se vuelve cada vez más imperativa. La lógica biomédica tiene suma importancia en este campo. El tema de la biopolítica y su relación con el poder ha sido objeto de importantes discusiones y análisis. El campo de la medicina también es un tema de gran importancia. Se ha dedicado una importante cantidad de esfuerzo al desarrollo del personaje, destacando su importancia. A costa de estos últimos, el productor prioriza sus intereses. Se destacan las características opresivas dentro del campo de la medicina. La ideología del marxismo experimentó un crecimiento e influencia significativos a lo largo de la década de 1970. Sobre todo, estos análisis han tenido un impacto significativo. La salud pública es un campo que implica la realización de investigaciones. Investigaciones relativas a la utilización de diversos implementos. El marco teórico de Foucault proporciona herramientas conceptuales para comprender y analizar diversos fenómenos. En la actualidad, tanto la seguridad social como la medicina desempeñan papeles importantes en la sociedad⁸

2.3. Definición y Clasificación de los Determinantes Sociales de la Salud (DSS)

La Organización Mundial de la Salud explica que el lugar donde uno nace crece, pasa la vida, trabaja y envejece tiene un impacto en su salud. Los Determinantes Sociales de la Salud (DSS) llaman la atención sobre este hecho. Estas circunstancias se ven afectadas por la distribución de recursos, poder y dinero a nivel global, nacional y local y están determinadas por los estándares, sistemas y tendencias actuales. La OMS señala que el DSS se ocupa de las desigualdades en salud que están presentes tanto dentro como entre las naciones. Mediante la combinación de factores económicos, ambientales, conductuales y sanitarios, la salud no está determinada únicamente por la genética y la biología. Comprender cómo influyen los factores sociales y culturales en la salud puede ayudarnos a comprender mejor las desigualdades en salud a través de DSS. Es crucial reconocer el impacto que tienen factores aparentemente no relacionados con la salud en el bienestar individual y comunitario. Paulo Marchiori Buss y Alberto Pellegrini Filho, autores del artículo "A Saúde e seus Determinantes Sociais", cuyo objetivo es examinar las conexiones entre la salud y sus determinantes sociales, brindan una visión concisa de los diferentes paradigmas explicativos sobre el proceso salud/enfermedad en las sociedades

⁸ Estrada-Mesa, D. A., & Cardona-Arias, J. A. (2018). La medicina en la obra de Michel Foucault: Meta-síntesis. *Civilizar Ciencias Sociales y Humanas*, 18, 223-236.

desde mediados del siglo XIX. Se explora una gama de posibles intervenciones y programas de políticas, destinados a abordar las disparidades en salud que surgen del DSS. La Comisión Nacional sobre Determinantes Sociales de la Salud, establecida en marzo de 2006, ha esbozado sus objetivos, estrategias e iniciativas clave. Su objetivo principal es facilitar la investigación sobre los determinantes sociales de la salud (DSS) y proponer políticas que promuevan la equidad en salud. Además, la comisión busca involucrar a varios sectores de la sociedad en discusiones y acciones relacionadas con DSS y la reducción de las disparidades en salud.⁹

Parece haber un claro enfoque en el concepto de DSS. Existe un acuerdo general en que la solución fundamental para abordar este problema profundamente arraigado reside en las disparidades que existen actualmente entre los diferentes grupos de la sociedad. Y en que es necesario intervenir en tales determinantes, que en gran medida se deben a la concentración del poder político y económico, así como el control de la economía por parte de minorías privilegiadas producto de las fuerzas del capitalismo y el fenómeno de la globalización financiera.

“La salud pública, que en su dimensión de disciplina científica aboga por las transformaciones sociales de salud y bienestar colectivo, el tema de la política no ha sido ajeno. De hecho, en su fundamentación teórica desde sus orígenes, muchos académicos de esta disciplina han subsumido la política como la categoría más relevante en el análisis de las condiciones de salud y bienestar de la población. Es en esta categoría donde implícita y explícitamente se configuran las demás dimensiones de la estructura social con sus diferentes tensiones, fricciones y transmutaciones”¹⁰. De este modo, al presentar sus propuestas políticas al público, La promoción y protección de la salud pública es un aspecto crucial para mantener el bienestar general. Reconociendo la necesidad de integrar y expresar efectivamente los diversos ámbitos de progreso social que abarcan los establecimientos públicos.

A principios de este siglo, Suecia inició un proceso Innovaciones en el desarrollo de políticas de salud Públicos en el enfoque de los determinantes Salud social. En lo que respecta a su fórmula, el estado ha identificado una serie de objetivos estructurales a alcanzar a manejarlos de manera muy simple, a diferencia Otros países. El aspecto central es la hipótesis como objetivo. Reducir las desigualdades en salud (en un país con cultura

⁹ Buss, P. M., & Pellegrini Filho, A. (2007). A Saúde e seus Determinantes Sociais. *Physis* vol. 17. 77-93

¹⁰ Mejía O, L. M. (2013). Los Determinantes Sociales de la Salud: base teórica de la salud pública. *Revista Facultad Nacional de Salud Pública*, 31, 28-36. Universidad de Antioquia. Medellín, Colombia, Pag. 34

de igualdad social y grandes logros cohesión social) para luego basar sus políticas en abordar estas razones.

Se identificaron 6 áreas de acción: Incrementar la cohesión/capital social, incluyendo la reducción de la discriminación contra los grupos minoritarios promoviendo la democracia local. Promover mejores condiciones laborales mediante la Reducción del estrés crónico y promover la participación de los empleados en su entorno trabajo, así como la promoción de esquemas flexibles de trabajo. Mejorar las condiciones de los niños y jóvenes a través de mecanismos generalizados de apoyo social y familiar, promoviendo en las escuelas el cuidado de salud. Mejorar el entorno físico a través de la coordinación e intervención en medio ambiente y salud. Finalmente, promover un estilo de vida saludable.¹¹

Guanilo De la torre Ugarte y Alfredo Oyola-García, señalan en su artículo “Los determinantes sociales de la salud: una propuesta de variables y marcadores/indicadores para su Medición” que El marco conceptual sobre los determinantes sociales de la salud propuesto por la Organización Mundial de la Salud combina elementos de varios modelos y pretende cumplir con todos los criterios anteriores. En este marco, se presentan de manera didáctica dos conjuntos de determinantes con el fin de comprender mejor los aspectos que determinan el proceso salud-enfermedad de los grupos sociales: los determinantes estructurales y los determinantes intermedios. El primer grupo está formado por aspectos relacionados con las estructuras sociales que determinan las desigualdades en salud. Es decir, pueden aportar aspectos de un grupo social que son más susceptibles a sufrir problemas de salud o, por el contrario, brindar mayores oportunidades para recuperar o mantener la salud. Ambos dependen del estatus social, ya que los grupos sociales con mayor acceso a recursos materiales, sociales, educativos, etc., pueden presentar circunstancias menos vulnerables o tener más probabilidades de recuperar o mantener la salud. Las situaciones de confrontación surgen cuando el estatus social de un grupo limita el acceso a dichos recursos. Por tanto, el primer grupo de DSS incluye aspectos relacionados con el contexto socioeconómico y político, así como aspectos relacionados con los mecanismos que determinan el estatus socioeconómico de los grupos sociales. Se consideran determinantes del entorno socioeconómico y político todos los mecanismos sociales, económicos y políticos que generan, configuran y mantienen un estatus socioeconómico desigual o una estratificación social, al asignar diferentes posiciones sociales a los individuos, que a su vez determinan el entorno. y desarrollarse.

¹¹ Villar, E. (2007). Los Determinantes Sociales de Salud y la lucha por la equidad en Salud: desafíos para el Estado y la sociedad civil. *Saúde e Sociedade*, 16(3), 7-13

Su elemento central es la gobernanza. El mecanismo que determina el estatus socioeconómico de los grupos sociales está relacionado con la educación y el ingreso, donde el nivel educativo afecta la ocupación y este último afecta el ingreso. Cabe señalar que los determinantes estructurales influyen en la salud a través de los determinantes intermedios (el segundo grupo y los DSS propuestos en el marco conceptual). El segundo grupo de determinantes se refiere a las condiciones materiales, el entorno psicosocial o social, los factores de comportamiento y los sistemas de salud en los que las personas viven y se desarrollan. En este sentido, los sistemas de salud tienen la responsabilidad de desempeñar un papel activo en la reducción de las desigualdades, no sólo a través del acceso equitativo a la atención (acceso y cobertura), sino también a través de la planificación y ejecución de programas de salud pública. Gestión transversal. Al cruzar estos dos determinantes (estructural e intermedio), encontramos que la cohesión y el capital social son puntos centrales en la política de salud, refiriéndose a la creación de relaciones de colaboración entre ciudadanos e instituciones. Es decir, el Estado debe asumir la responsabilidad de establecer instituciones más flexibles que brinden espacios concretos para la participación ciudadana, lo que significa una redistribución del poder para que las comunidades tengan mayor influencia en la toma de decisiones y la formulación de políticas. afectando su bienestar y calidad de vida. Es importante enfatizar que los determinantes intermedios afectarán la salud en mayor o menor medida dependiendo del estatus social, ya que los individuos, dependiendo de su estatus social, estarán diferentemente expuestos a condiciones que pueden dañar su salud y, por tanto, tendrán diferentes vulnerabilidades. Determina diferentes consecuencias para la salud, incluidas las económicas y sociales, así como diferentes resultados para la salud. Los determinantes estructurales se diferencian de los factores intermediarios en que son mecanismos sociales que producen sistemáticamente DSS (por tanto, grupos sociales más vulnerables que otros) y distribuyen la desigualdad entre los grupos sociales, por lo que también se les llama determinantes sociales.¹²

Una clasificación comúnmente aceptada para los determinantes sociales abarca varios métodos, entre ellos:

- **Salud y estatus socioeconómico:** El nivel educativo, los ingresos económicos y la ocupación de un individuo son algunos de los determinantes de su estatus socioeconómico, que se ha descubierto que tiene un impacto directo en su salud.

¹² De la Torre-Ugarte-Guanilo, M., & Oyola-García, A. (2014). Los determinantes sociales de la salud: una propuesta de variables y marcadores/indicadores para su medición. *Revista Peruana de Epidemiología*, 18(1), 1-6. Sociedad Peruana de Epidemiología.

Generalmente, las personas con un nivel socioeconómico más alto tienden a tener mejor salud.

- **Salud y Sistemas de Salud:** La salud física y mental puede verse afectada por la cohesión comunitaria y los sistemas de apoyo social, además de las normas y valores culturales. Las relaciones sociales y la cantidad y calidad de las interacciones sociales juegan un papel dentro de estos contextos sociales y comunitarios. La determinación de los resultados de salud en las poblaciones depende de varios factores, uno de los cuales es la calidad, accesibilidad y organización de los servicios de salud. El determinante Salud y Sistemas de Salud se centra en este aspecto ya que el acceso a servicios de salud preventivos y curativos de calidad es esencial para promover una vida saludable.
- **Entorno Físico:** La presencia de espacios verdes y recreativos, la calidad del aire y el agua y viviendas seguras y asequibles son parte del entorno físico. Un entorno que sea a la vez seguro y limpio puede reducir significativamente los riesgos de enfermedades y fomentar estilos de vida saludables.
- **Comportamientos y Estilos de Vida:** Aunque aspectos como la alimentación, la actividad física, fumar o beber están determinados en gran medida por otros DSS, juegan un papel vital en la salud de las personas.

De estas categorizaciones surgen algunos de los principales determinantes sociales de la salud, tales como:

1. **Ingresos y estatus social:** las personas con mayores ingresos y estatus social suelen tener mejor salud. Un nivel socioeconómico más alto se asocia con mejores condiciones de vida, nutrición, educación y acceso a servicios de salud.
2. **Educación:** cuanto más educadas sean las personas, más probabilidades tendrán de disfrutar de buena salud. La educación apoya el desarrollo de habilidades para la vida y fomenta comportamientos más saludables.
3. **Empleo y condiciones laborales:** El empleo proporciona una fuente de ingresos, mejorando así la calidad de vida, mientras que las malas condiciones laborales aumentan los riesgos para la salud.
4. **Acceso a servicios de salud adecuados:** El acceso a servicios de salud de calidad es fundamental para prevenir enfermedades y tratar las condiciones existentes.
5. **Cultura:** Las creencias, normas y valores culturales pueden influir en los comportamientos relacionados con la salud, las percepciones de enfermedades y las decisiones de atención médica.

6. **Género:** Las desigualdades de género impactan el acceso a recursos y servicios, la exposición a ciertos riesgos y la capacidad de tomar decisiones relacionadas con la salud.
7. **Redes de apoyo social:** las relaciones y redes sociales pueden influir en la salud mental y el bienestar. Sentirse incluido y apoyado es fundamental para la salud.
8. **Entorno físico:** El acceso a agua potable, aire limpio, viviendas seguras y alimentos nutritivos, así como la exposición a contaminantes, tienen un impacto directo en la salud física.
9. **Hábitos y comportamientos personales:** como fumar, dieta, ejercicio, bebida, etc.
10. **Factores biológicos y genéticos:** Estos factores influyen en la susceptibilidad a determinadas enfermedades.

2.4. Relación entre DSS y Salud

La salud de una población no es sólo el resultado de factores genéticos o decisiones personales, sino que está fuertemente influenciada por una variedad de condiciones y entornos en los que las personas viven y se desarrollan.

Los determinantes sociales de la salud (DSS) constituyen estas condiciones, cuyos efectos se presentan a lo largo de la vida de una persona, desde el nacimiento hasta la vejez. Las enfermedades crónicas, como las cardiovasculares, la diabetes y ciertos tipos de cáncer, están estrechamente asociadas con el DSS. (Ej. las personas con menores niveles de educación o ingresos tienden a tener tasas más altas de estas enfermedades debido a factores como la falta de acceso a alimentos saludables, vivir en ambientes adversos o la falta de acceso a atención médica preventiva oportuna).

Los comportamientos de riesgo como fumar, el consumo excesivo de alcohol, una dieta poco saludable y la inactividad física se ven afectados por el DSS. Las normas y valores culturales, la publicidad y el marketing, la disponibilidad y el costo de los productos y servicios y las políticas públicas influyen en la prevalencia de estos comportamientos entre diferentes poblaciones.

Los DSS no actúan de forma aislada; más bien, interactúan para formar una red con efectos complejos sobre la salud. Por ejemplo, el nivel educativo puede influir en los tipos de empleo que una persona puede obtener, lo que a su vez afecta sus niveles de ingresos, su acceso a la atención sanitaria y sus condiciones de vivienda, lo que tiene como resultado una serie de consecuencias para la salud.

2.5. Relevancia Geográfica y Temporal de los DSS

Los determinantes sociales de la salud no son uniformes en todas las regiones o áreas geográficas. Existen diferencias significativas en DSS entre áreas urbanas y rurales y entre diferentes estados o regiones.

- **Áreas urbanas versus áreas rurales suburbanas:** En general, las áreas urbanas tienden a ofrecer mayores oportunidades de empleo, educación y acceso a servicios de salud. Sin embargo, también pueden presentar desafíos específicos, como altos niveles de contaminación, estilos de vida sedentarios y estrés asociado al ritmo de vida. Por otro lado, las zonas rurales pueden enfrentar desafíos como la falta de servicios de salud especializados, la falta de infraestructura y una diversidad insuficiente de oportunidades educativas y laborales.
- **Diferencias entre estados o regiones:** cada estado o región puede tener diferentes políticas, economías y culturas que pueden afectar el DSS. Por ejemplo, los países que invierten mucho en educación y salud probablemente tengan mejores indicadores de salud que los países que invierten menos en estas áreas. Con el tiempo, los DSS pueden cambiar debido a factores como el desarrollo económico, cambios políticos, avances tecnológicos y tendencias socioculturales.
- **Tendencias de largo plazo:** Factores como el crecimiento económico, la globalización y el desarrollo tecnológico han transformado los DSS durante décadas, afectando aspectos como la educación, el empleo y el acceso a la información.
- **Eventos específicos:** Situaciones como crisis económicas, desastres naturales o epidemias pueden tener efectos directos y de largo plazo en los DSS, afectando el bienestar y la salud de la población.

Utilizando datos del BRFSS, se pueden identificar tendencias y cambios en el DSS entre 2017 y 2021. Esta información puede proporcionar información valiosa sobre cómo evolucionan estos determinantes en períodos de tiempo relativamente cortos y cómo afectan la salud de un individuo. Población en diferentes regiones y en general.

2.6. Metodologías comunes para estudiar los DSS

Al reconocer que los determinantes sociales de la salud tienen un impacto significativo en el bienestar de la población, las autoridades y organizaciones de salud los incluyen en el desarrollo de políticas.

Comprender cómo factores como el nivel educativo, la situación económica o el entorno laboral afectan la salud es fundamental para diseñar políticas que aborden las causas profundas de las desigualdades en salud.

Los impactos de los DSS en la salud no pueden abordarse únicamente desde el sector de la salud. Se necesita una acción coordinada y multisectorial que abarque áreas como la educación, la vivienda, el transporte y la economía. Por ejemplo, una mejor educación puede tener un impacto positivo en la salud al mejorar las condiciones de trabajo y crear conciencia sobre comportamientos saludables.

Dado que los DSS pueden exacerbar las desigualdades en salud, las políticas y estrategias deben priorizar a los grupos vulnerables. Esto significa dirigir recursos, intervenciones y programas a quienes enfrentan mayores barreras sociales, económicas y culturales para su salud.

La naturaleza cambiante de los DSS y su impacto en la salud requiere una evaluación y un seguimiento continuos. Herramientas como BRFSS son fundamentales para comprender las tendencias actuales, identificar áreas de preocupación y evaluar la eficacia de las intervenciones implementadas.

El artículo "Asociaciones entre los determinantes sociales de la salud y la depresión: un análisis del sistema de vigilancia de factores de riesgo conductual (BRFSS) de 2017" examina cómo los determinantes sociales de la salud (SDOH) influyen en la prevalencia de la depresión. Este estudio utilizó datos del BRFSS de 2017 para analizar la relación entre varios SDOH, incluida la inseguridad económica, la seguridad del vecindario y la inseguridad alimentaria, y las tasas de depresión autoinformadas. El estudio concluyó que existe una fuerte correlación entre la SDOH adversa y una mayor prevalencia de depresión, lo que enfatiza la importancia de considerar estos factores sociales en las estrategias de prevención y tratamiento de la depresión.¹³ Estudios como el mencionado indican que no sólo las políticas públicas sino también la atención sanitaria personal debería considerar los DSS. Los profesionales de la salud deben estar capacitados para comprender y considerar los DSS al evaluar y tratar a los pacientes, reconociendo que factores fuera del entorno médico tradicional pueden tener un impacto significativo en la salud de un individuo.

¹³ Wilkinson, L.L.; Long-Daniels, A.; Appah, M.; Zhai, Y.; Watson, D.M.; Walker, K.; Young-Bilbo, K.; Aboagye, A.; Tucker, C.; McCarthy, S. The Association between Social Determinants of Health and Depressive Disorders: A 2017 Behavioral Risk Factor Surveillance System (BRFSS) Analysis. *Psychiatry Int.* **2023**, *4*, 147-159. <https://doi.org/10.3390/psychiatryint4020017>

Por otro lado, el artículo "Predicting Self-Rated Health Across the Life Course: Health Equity Insights from Machine Learning Models" examina la salud autopercebida y su relación con diversos factores sociales y de salud. El estudio utilizó el Sistema de Vigilancia de Factores de Riesgo del Comportamiento (BRFSS) y algoritmos de aprendizaje automático para analizar cómo las condiciones socioeconómicas, los comportamientos de salud, el acceso a la atención médica y otros factores predicen la autoevaluación de la salud en diferentes grupos demográficos. Los hallazgos resaltan la importancia de los determinantes sociales de la salud, especialmente en la mediana edad, y sugieren enfoques prometedores para utilizar el aprendizaje automático para promover la equidad en salud. El estudio también aborda las limitaciones y posibles aplicaciones de estos modelos en la investigación y la práctica de la salud pública.¹⁴

Estos proyectos demuestran la utilidad del BRFSS para estudiar cómo los determinantes sociales impactan en diversos aspectos de la salud y el bienestar en los Estados Unidos, utilizando técnicas avanzadas de análisis de datos como el aprendizaje automático.

Adicionalmente, en el artículo de Javier Mora Pineda "Modelos de salud predictivos basados en aprendizaje automático" analiza el impacto significativo del aprendizaje automático en la salud pública. Se centra en cómo estas tecnologías informáticas modernas permiten análisis predictivos y reconocimiento de patrones más precisos en los datos de salud. Los autores exploran un cambio de paradigma en la investigación médica en el que el aprendizaje automático ayuda a generar modelos predictivos confiables a partir de grandes cantidades de datos, incluidos registros clínicos no estructurados. Se destacan varios ejemplos de aplicación en diferentes campos médicos, enfatizando la capacidad de estos modelos para mejorar la toma de decisiones clínicas y la eficiencia de la gestión de recursos sanitarios.¹⁵

14 Clark, C. R., Ommerborn, M. J., Moran, K., Brooks, K., Haas, J., Bates, D. W., & Wright, A. (Año). Predicting Self-Rated Health Across the Life Course: Health Equity Insights from Machine Learning Models. *J Gen Intern Med.* 2021 May;36(5):1181-1188. doi: 10.1007/s11606-020-06438-1. Epub 2021 Feb 23. PMID: 33620624; PMCID: PMC8131482.

15 Javier Mora Pineda (2022). *Revista médica Clínica Las Condes: Modelos predictivos en salud basados en aprendizaje de maquina (Machine learning)*

Finalmente, el artículo de Ángela María Segura Cardona "Uso de árboles de decisión en salud pública" explora cómo se utiliza el método de análisis CART (Árbol de clasificación y regresión) en salud pública. Se destaca la eficiencia de los árboles de decisión para segmentar poblaciones y generar subgrupos con características similares que puedan ser utilizados para intervenciones de salud pública. Este artículo revisa la investigación sobre el uso de CART en una variedad de entornos, incluidas enfermedades crónicas e infecciosas, y destaca su potencial para identificar poblaciones en riesgo y para la toma de decisiones de salud pública. Aquí, Se analizaron estudios que utilizan árboles de decisión para el diagnóstico y seguimiento de enfermedades, mostrando la efectividad de los árboles de decisión para segmentar poblaciones e identificar factores de riesgo. Además, se analiza cómo esta tecnología puede facilitar la implementación de políticas de salud más específicas y eficientes, lo que conducirá a una mejor asignación de recursos y una toma de decisiones basada en datos concretos.¹⁶

¹⁶ Segura Cardona, Angela María. "Aplicación de árboles de decisión en la salud pública". *Revista CES Salud Pública*, vol. 3, no. 1, 2012, pp. 94-103.

3. Hipótesis y Objetivos

3.1. Hipótesis

Los determinantes sociales tienen un impacto significativo en la prevalencia de enfermedades crónicas y comportamientos de riesgo relacionados con la salud en la población adulta de Estados Unidos entre 2017 y 2021

3.2. Objetivos

Objetivo general: Evaluar la influencia de los determinantes sociales en la prevalencia de enfermedades crónicas y comportamientos de riesgo relacionados con la salud en la población adulta de Estados Unidos entre 2017 y 2021 utilizando los datos del Sistema de Vigilancia de Factores de Riesgo del Comportamiento (BRFSS)

Objetivos específicos:

- Catalogar y describir las variables del BRFSS que representen determinantes sociales de la salud.
- Desarrollar un modelo estadístico que permita predecir la prevalencia de enfermedades crónicas y comportamientos de riesgo en función de diferentes determinantes sociales.

4. Metodología

4.1. Datos

El Sistema de Vigilancia de Factores de Riesgo del Comportamiento (BRFSS) es el principal sistema nacional de encuestas telefónicas relacionadas con la salud que recopila datos estatales sobre los residentes de los Estados Unidos con respecto a sus comportamientos de riesgo relacionados con la salud, afecciones crónicas de salud y uso de servicios preventivos. Establecido en 1984 con 15 estados, BRFSS ahora recopila datos en los 50 estados, así como en el Distrito de Columbia y tres territorios de los Estados Unidos. BRFSS completa más de 400,000 entrevistas con adultos cada año, lo que lo convierte en el sistema de encuestas de salud más grande realizado continuamente en el mundo.

Al recopilar datos de riesgo de salud conductual a nivel estatal y local, BRFSS se ha convertido en una herramienta poderosa para enfocar y desarrollar actividades de promoción de la salud. Como resultado, los usuarios de BRFSS han exigido cada vez más datos y han pedido más preguntas sobre la encuesta. Actualmente, existe un amplio patrocinio de la encuesta BRFSS, que incluye la mayoría de las divisiones del Centro Nacional para la Prevención de Enfermedades Crónicas y la Promoción de la Salud de los CDC; otros centros de los CDC; y agencias federales, como la Administración de Recursos y Servicios de Salud, la Administración sobre el Envejecimiento, el Departamento de Asuntos de Veteranos y la Administración de Servicios de Salud Mental y Abuso de Sustancias.

Para este proyecto se obtuvo los últimos 5 años disponibles (2017 a 2021) en la dirección del BRFSS¹⁷, disponibles en un archivo por año en formato SAS. Sin embargo, de la totalidad de variables disponibles, solo se conservaron aquellas que hacen referencia a los tópicos de identificación del registro, de acceso a salud, calidad de vida, Condiciones crónicas de salud, variables sociodemográficas, uso de tabaco, consumo de alcohol y algunas variables calculadas¹⁸

Los datos provienen de la página del “Centers for disease control and prevention” y se encuentran concentradas por años en el siguiente link.

¹⁷ https://www.cdc.gov/brfss/annual_data/annual_data.htm

¹⁸ En el anexo puede verificarse el detalle de las variables utilizadas y sus posteriores recategorizaciones

https://www.cdc.gov/brfss/annual_data/annual_data.htm

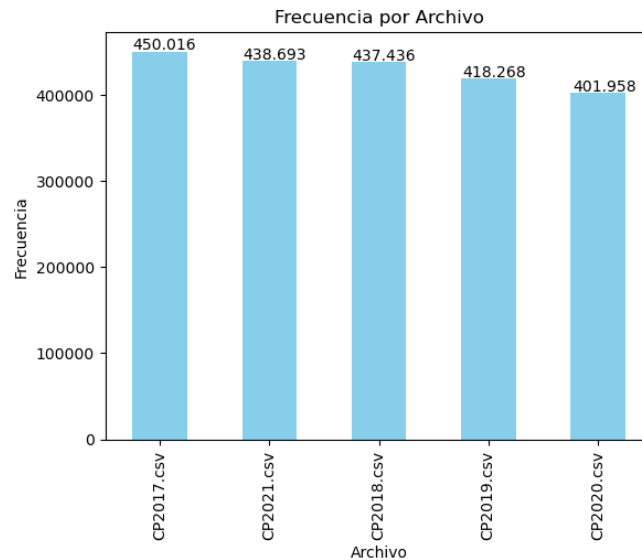
el detalle de cada año es el siguiente:

AÑO	DATA	CODEBOOK
2017	https://www.cdc.gov/brfss/annual_data/2017/files/LLCP2017XPT.zip	https://www.cdc.gov/brfss/annual_data/2017/pdf/codebook17_llcp-v2-508.pdf
2018	https://www.cdc.gov/brfss/annual_data/2018/files/LLCP2018XPT.zip	https://www.cdc.gov/brfss/annual_data/2018/pdf/codebook18_llcp-v2-508.pdf
2019	https://www.cdc.gov/brfss/annual_data/2019/files/LLCP2019XPT.zip	https://www.cdc.gov/brfss/annual_data/2019/pdf/codebook19_llcp-v2-508.HTML
2020	https://www.cdc.gov/brfss/annual_data/2020/files/LLCP2020XPT.zip	https://www.cdc.gov/brfss/annual_data/2020/pdf/codebook20_llcp-v2-508.pdf
2021	https://www.cdc.gov/brfss/annual_data/2021/files/LLCP2021XPT.zip	https://www.cdc.gov/brfss/annual_data/2021/pdf/codebook21_llcp-v2-508.pdf

Dado que este proyecto armará un modelo que permita interpretar que determinantes sociales principales inciden en el estado de salud; de la totalidad de variables que contiene la encuesta, se utilizó solo aquellas relevantes para el estudio, es decir aquellas relacionadas con el estado de salud, con los comportamientos de riesgo para la salud y de determinantes sociales¹⁹

Se utilizó Python para convertir los archivos originales en formato SAS, a archivos de CSV, que fueron unificado en un solo dataset, este dataframe contiene 2.146.371 registros, distribuidos de la siguiente forma:

¹⁹ Ver el conjunto de datos en el punto 13.1 de los anexos



El proyecto construye 6 dimensiones:

1. Riesgo de tener un mal estado de salud físico y mental.
2. Riesgo de padecer diabetes.
3. Posibilidad de acceder a cuidados de salud.
4. Riesgo de tener multimorbilidades.
5. Riesgo de ser adicto al tabaco.
6. Riesgo de ser adicto al alcohol.

Estas 6 dimensiones se utilizarán para construir un método de clasificación que permita identificar los determinantes sociales que probablemente puedan incidir en un aumento o disminución de algunas de las características de las dimensiones; adicionalmente, la combinación de estas 6 dimensiones construirá un índice que se definirá como “estado de salud”, donde podrá elaborarse un sistema de clasificación general.

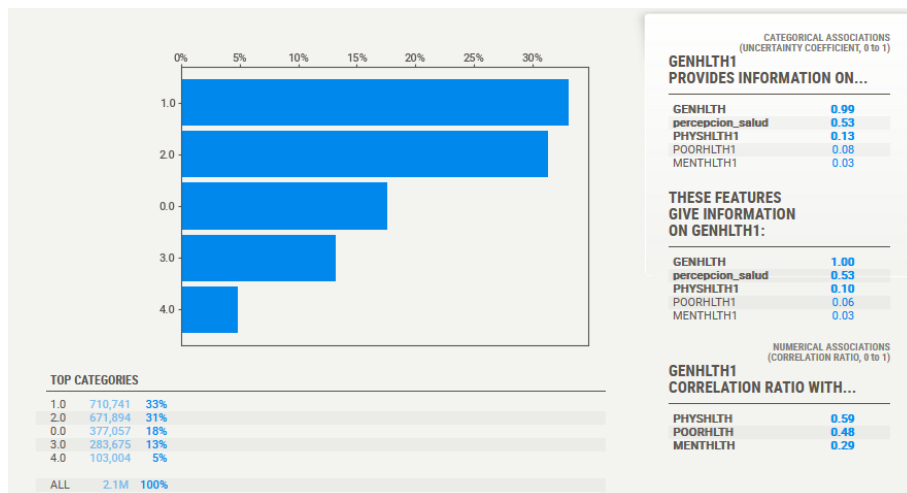
Para generar los gráficos de barras que se mostraran en este punto, se utilizó la librería de Python “Sweetviz”.

1. **Riesgo de tener un mal estado de salud físico y mental:** Este indicador se construyó considerando como el encuestado percibe su salud general, su salud mental. Su salud física y la cantidad de días que tuvo que dejar alguna actividad habitual debido a algún problema físico o mental. Cada una de estas variables fueron recategorizadas, de forma tal de poder obtener 5 categorías desde 0 como

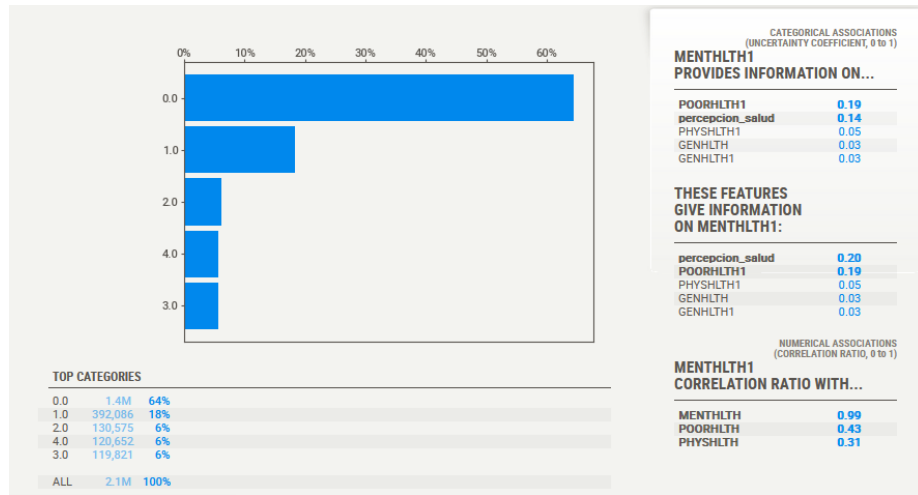
el mejor estado físico y/o mental posible, hasta 4 como el peor estado físico y/o mental posible (0=excelente estado, 1=muy buen estado, 2= buen estado, 3=mal estado y 4=muy mal estado). La suma de estas variables conformó el indicador final, el que se recategorizó siguiendo las categorías mencionadas anteriormente.

Las variables individuales recategorizadas se suman para formar un índice compuesto. Este índice proporciona una medida integral de la percepción del estado de salud físico y mental, basada en varios aspectos de la salud y el bienestar.

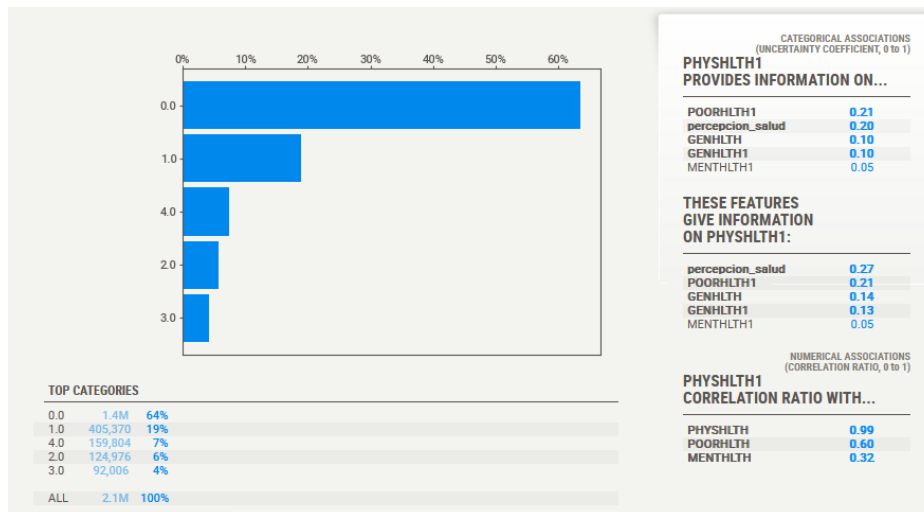
1.1. GENHLTH: Cual diría que es su estado de salud general



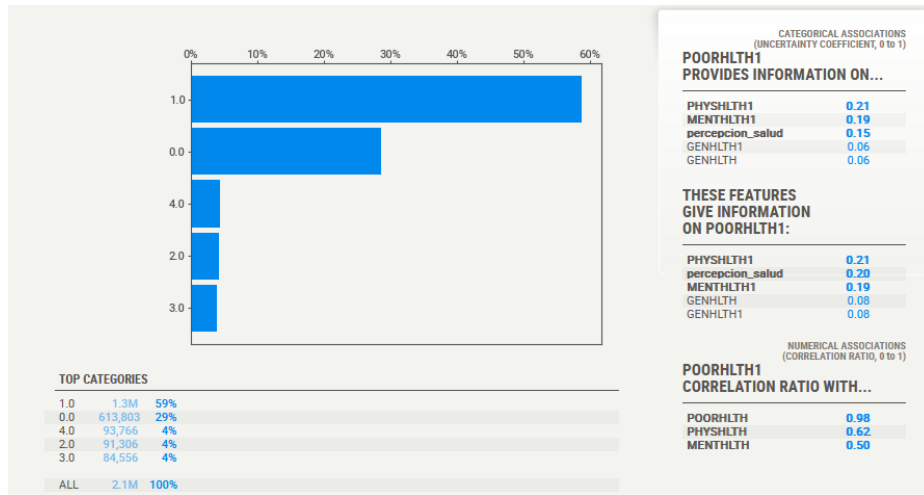
1.2. **MENTHLTH:** Pensando en su salud mental, incluyendo stress, depresión y problemas emocionales, por cuantos días durante los pasados 30, diría que su salud mental no es buena.



1.3. **PHYSHLTH:** Pensando en su salud física, incluyendo enfermedades y accidentes, por cuantos días durante los pasados 30, diría que su salud mental no es buena.

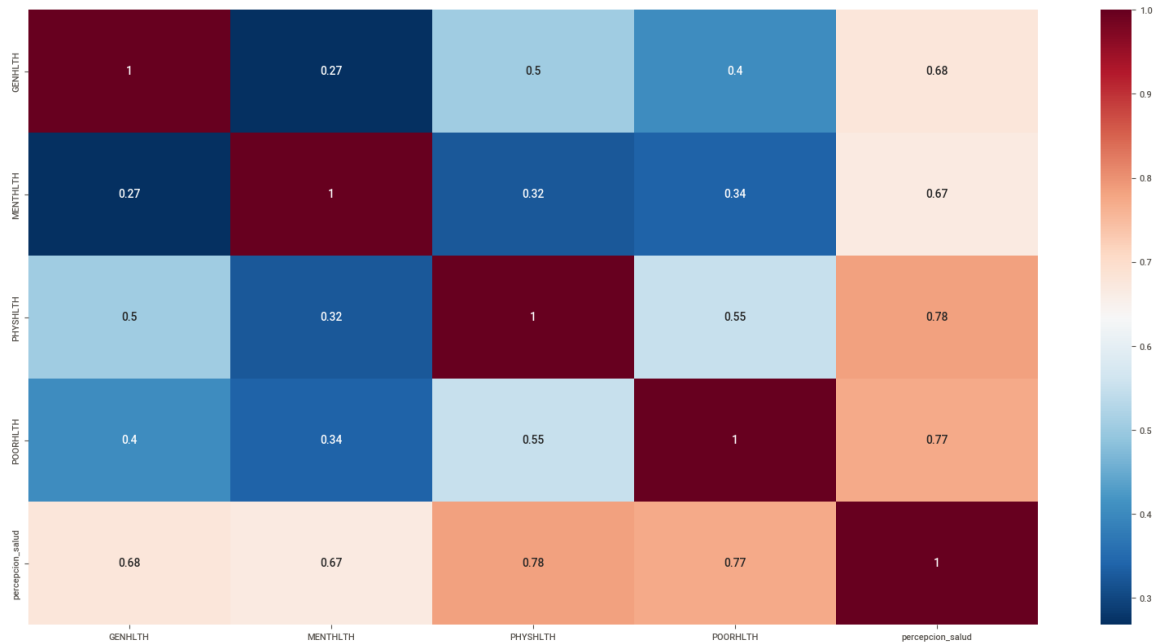


1.4. **POORHLTH:** Durante los pasados 30, por cuantos días diría que una mala salud física o mental le impidió hacer sus actividades usuales de autocuidado, trabajo o recreación



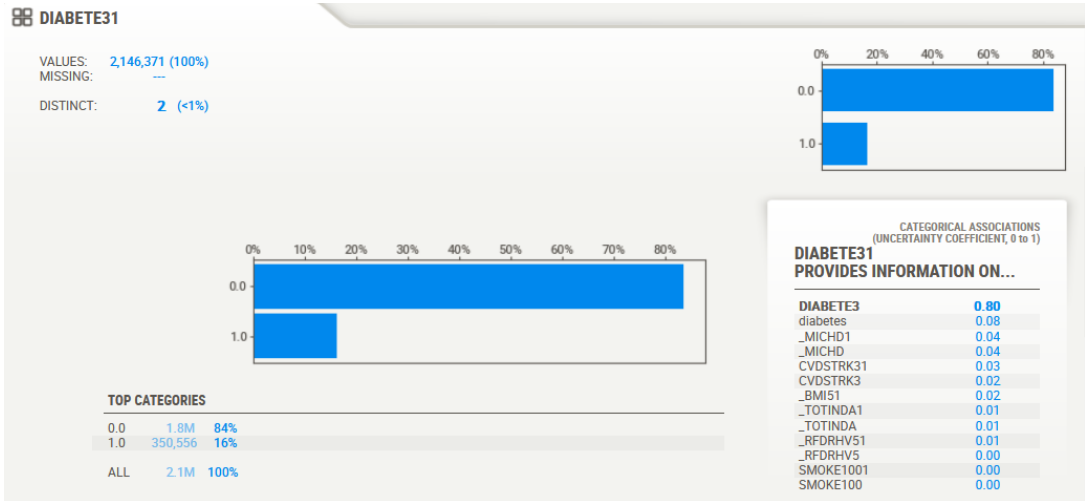
La suma y posterior recategorización de la variable obtenida dio origen al indicador “percepcion_salud”, que resume en una sola variable las percepciones indicadas por los encuestados en torno a su estado de salud física y mental; así como del impedimento de realizar alguna actividad específica debido a ella.

Se obtuvo una matriz de correlaciones entre la variable “percepción_salud” y las variables originales que componen el indicador, antes de ser recategorizadas, mostrando que la asociación entre ambas tiende a ser lo suficientemente fuerte como para suponer que el indicador creado es lo suficientemente válido y fiable como para representar correctamente la percepción

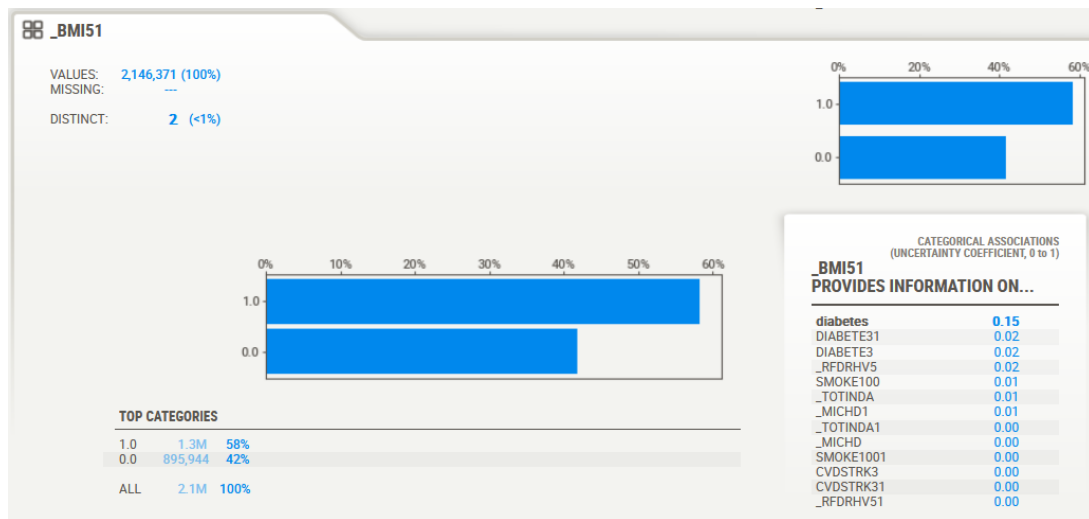


2. **Riesgo de padecer diabetes:** Este indicador se construyó considerando las preguntas de la encuesta que informan sobre el posible diagnóstico de diabetes, el índice de masa corporal, el haber fumado cierta cantidad de cigarrillos, el posible diagnóstico de un ataque cardiovascular, el tener una enfermedad cardíaca, el haber hecho actividad física durante los pasados 30 días, el haber bebido más de 14 bebidas alcohólicas durante la semana en hombres y 7 en mujeres. En este indicador, solo el índice de masa corporal fue recodificado cambiando el nivel de medición de la categoría (25 o menos= sin incidencia en el riesgo de padecer diabetes, 26 o más= con incidencia en el riesgo de padecer diabetes), En el resto de las variables se recodificaron las categorías invirtiendo la escala (0= sin incidencia en el riesgo de padecer diabetes, 1 = con incidencia en el riesgo de padecer diabetes), de forma que coincida la orientación de las respuestas al sumar las variables en la construcción del indicador.

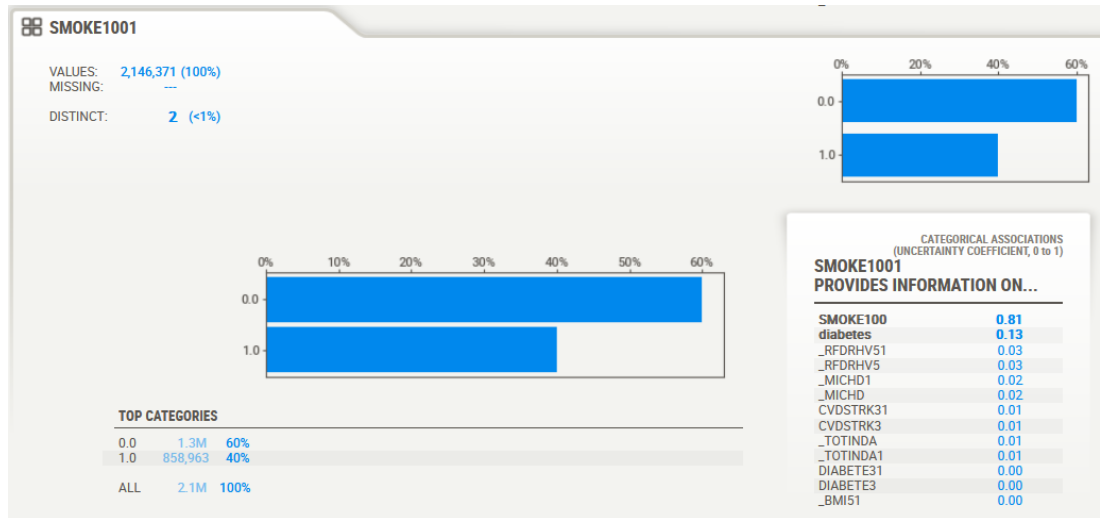
2.1. **DIABETE3:** Alguna vez le han dicho que tiene diabetes



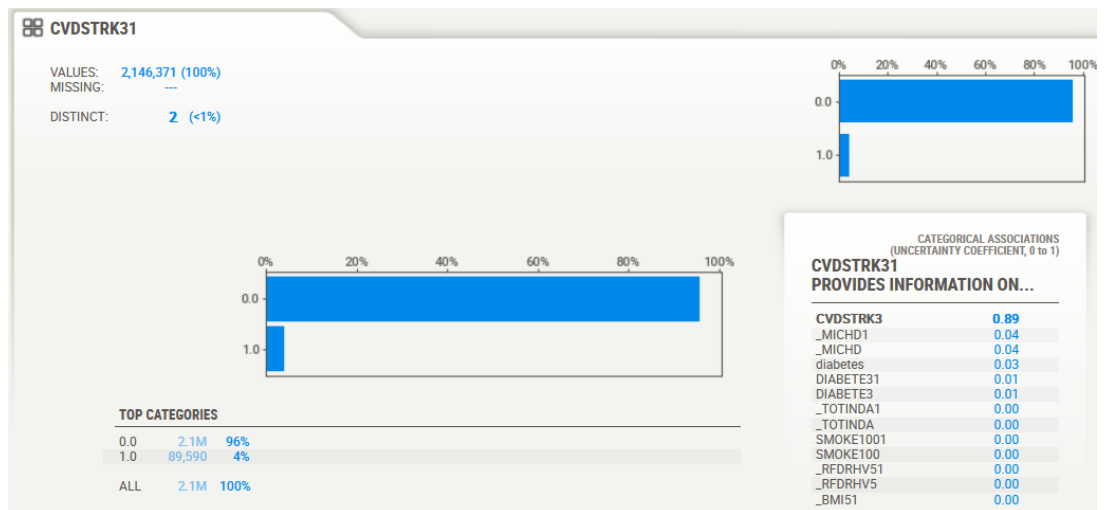
2.2. BMI: índice de masa corporal



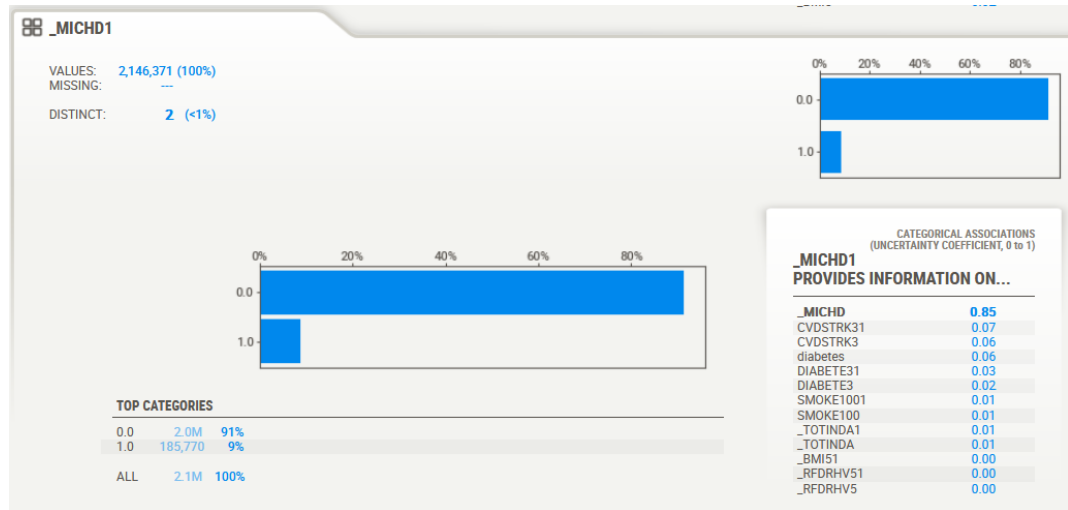
2.3. SMOKE100: Ha fumado al menos 100 cigarrillos en su vida.



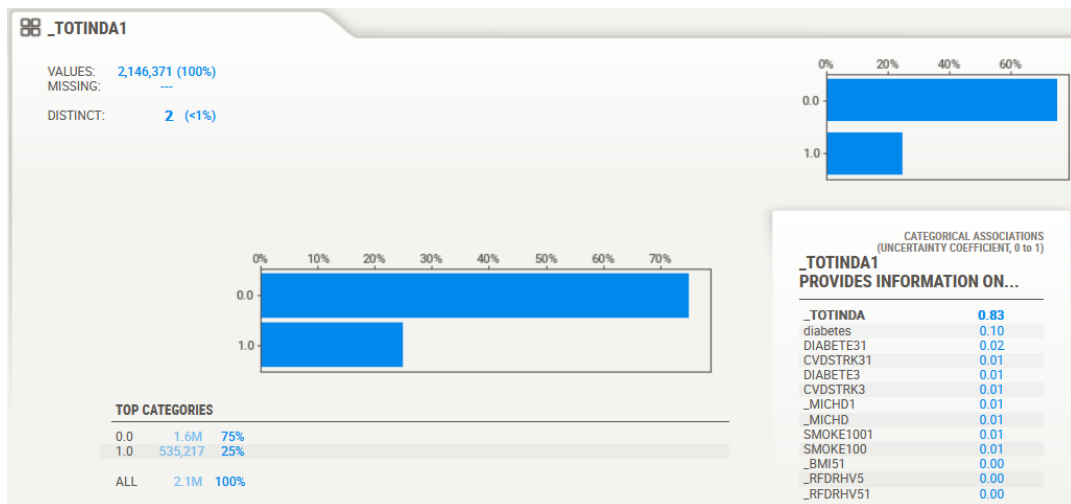
2.4. CVDSTRK3: Ha tenido un accidente cerebro vascular



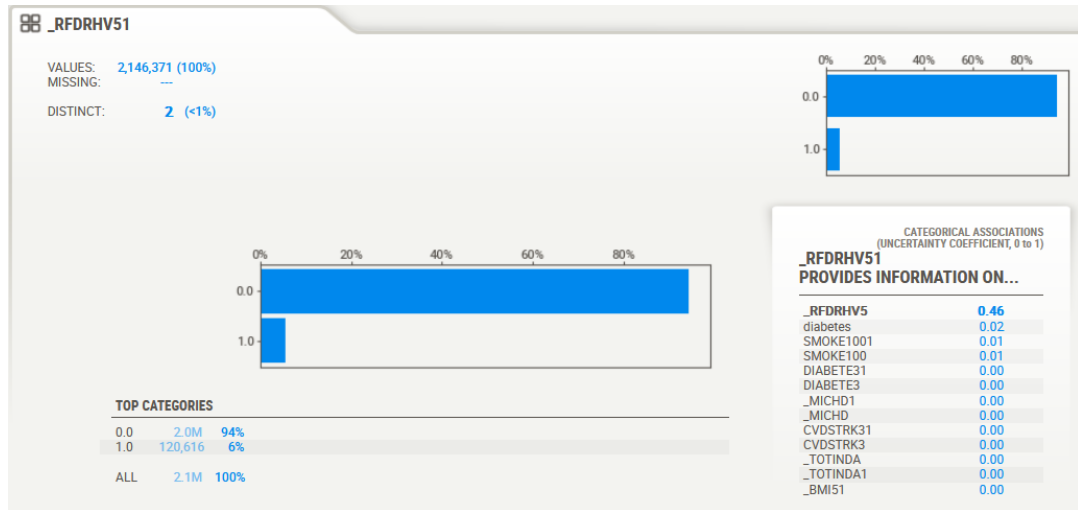
2.5. MICH1: Encuestados que han tenido una enfermedad coronaria o infarto al miocardio



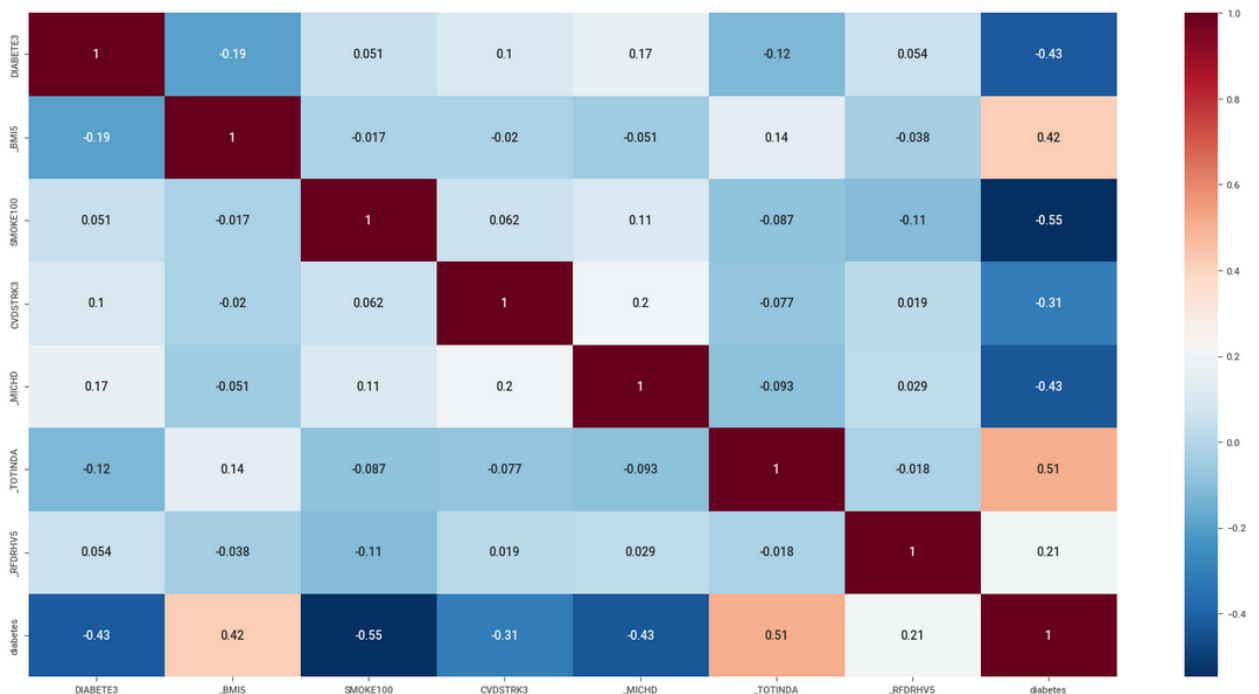
2.6. **TOTINDA:** Encuestados adultos que reportan haber hecho ejercicios o actividad física (distinta a las actividades de su trabajo habitual) en los últimos 30 días



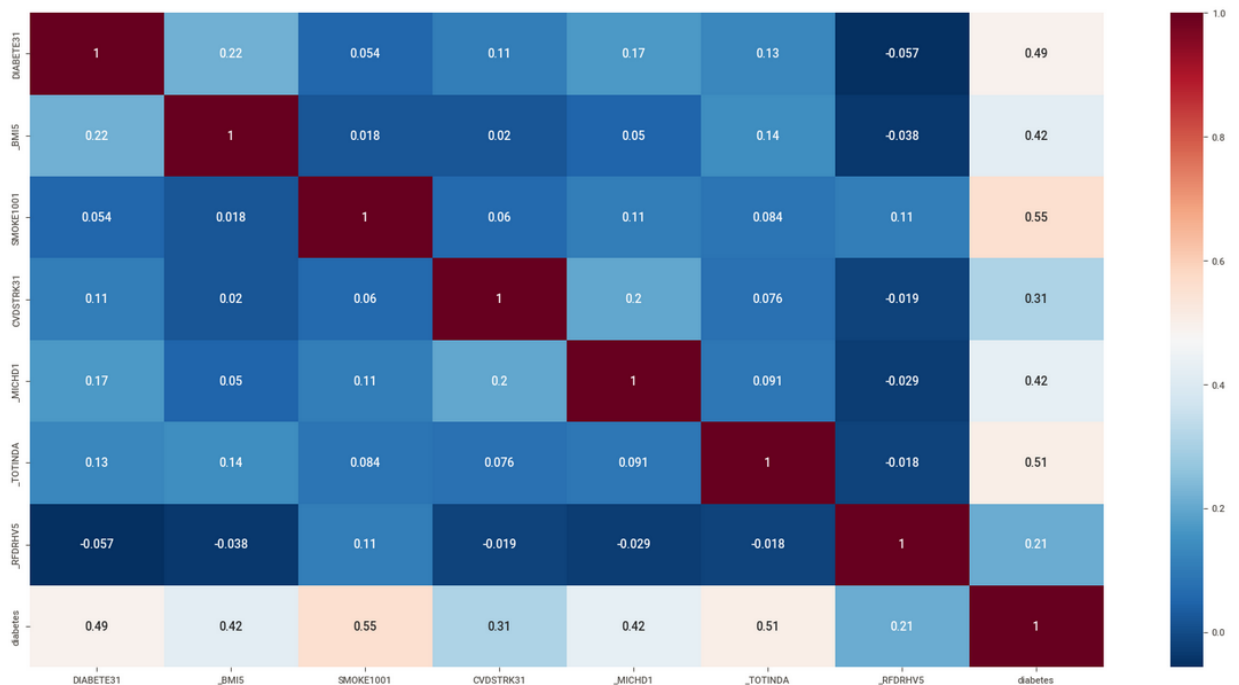
2.7. **RFDRHV5:** Hombres que tuvieron 14 o más bebidas alcohólicas y mujeres que tuvieron 7 o más bebidas alcohólicas durante la semana



En este caso, la matriz de correlaciones se obtenida para verifica la asociación entre variables, previo a su recategorización, y el indicador final muestra asociaciones inversamente proporcionales.

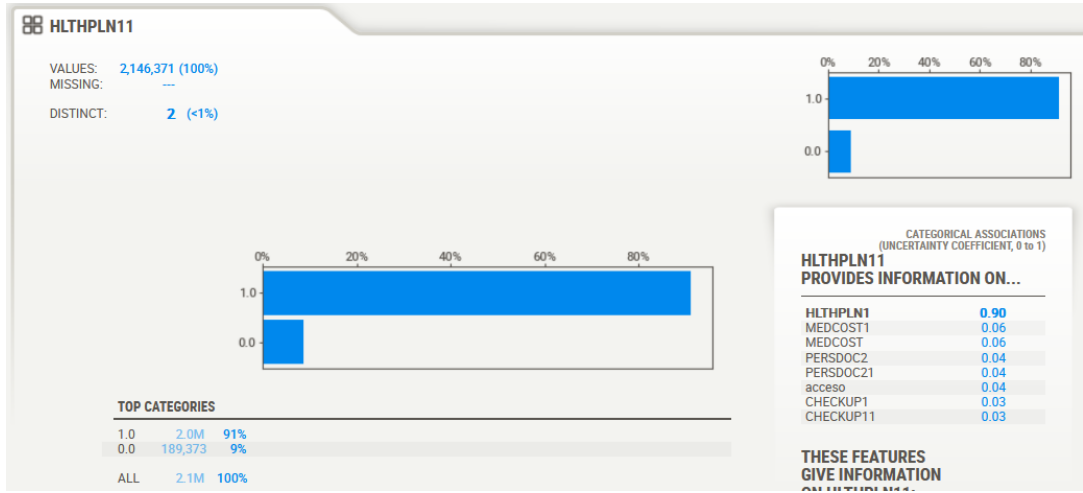


Estas asociaciones negativas, adquieren un valor positivo al utilizar las mismas variables, pero recategorizadas, de modo tal que invierten su escala y por tanto las asociaciones negativas, pasan a convertirse en asociaciones positivas

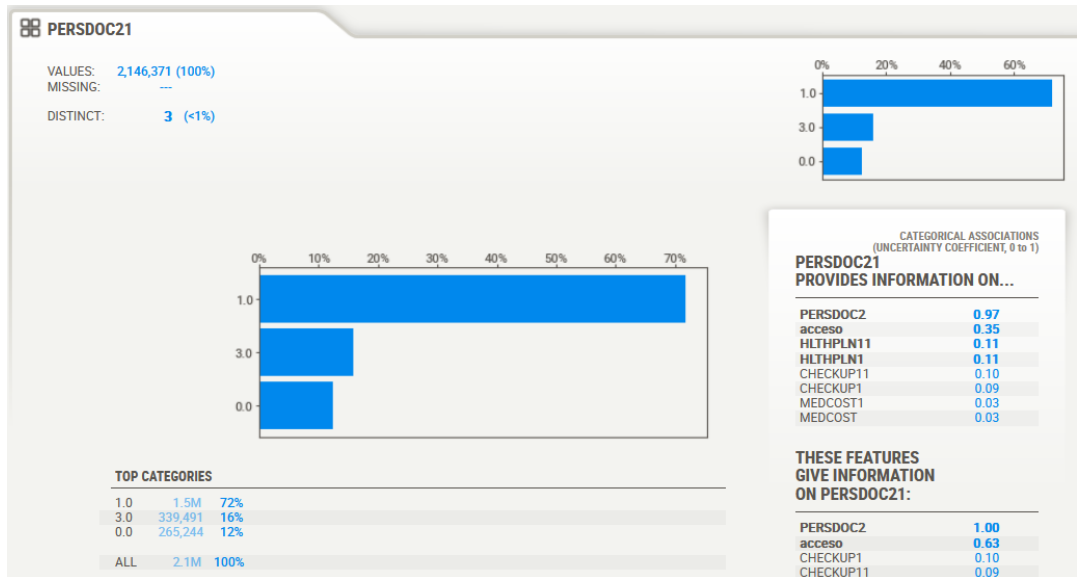


3. **Posibilidad de acceder a cuidados de salud:** Este indicador se construyó considerando las preguntas sobre cobertura y acceso a salud, al igual que en los indicadores anteriores, las preguntas fueron recodificadas y luego sumadas en un indicador que recoge el grado de probabilidad de tener cobertura de salud, las variables que conforman la dimensión fueron recodificada en variables binarias (0 = no tiene acceso, 1 = tiene acceso), con excepción de la variable que mide el tiempo desde la última visita a un doctor

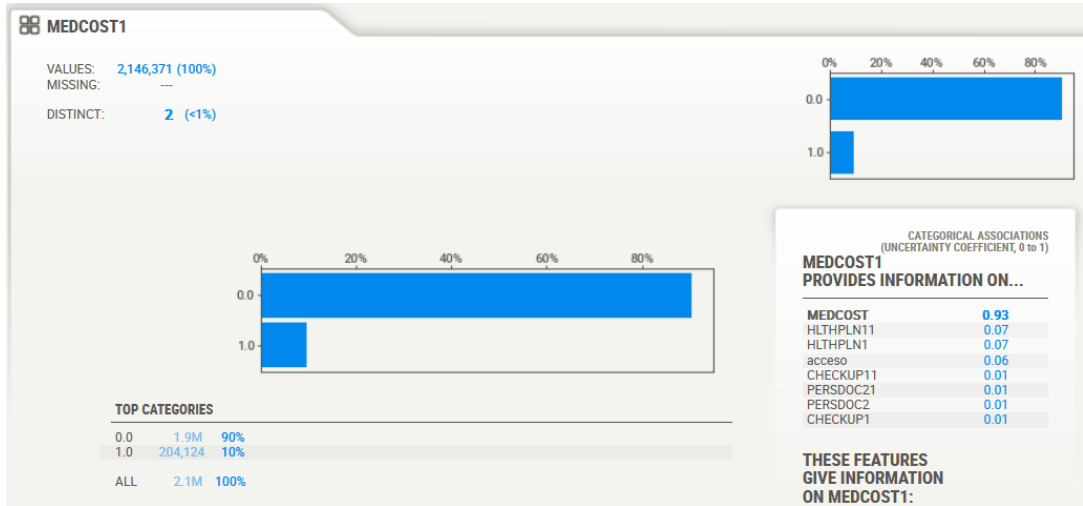
3.1. **HLTHPLN1:** Tiene algún tipo de cobertura de salud, incluyendo seguro de salud, planes de prepago o planes de gobierno (medicare)



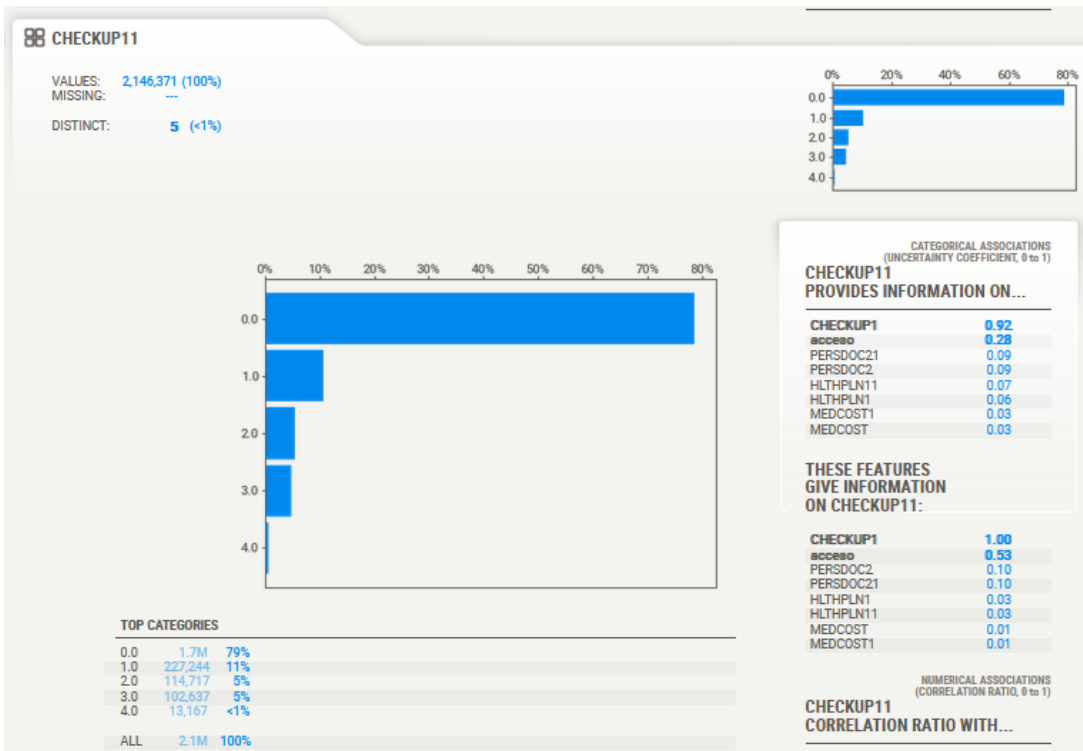
3.2. **PERSDOC2:** Tiene alguna persona a la que considera como su doctor personal o proveedor de cuidados



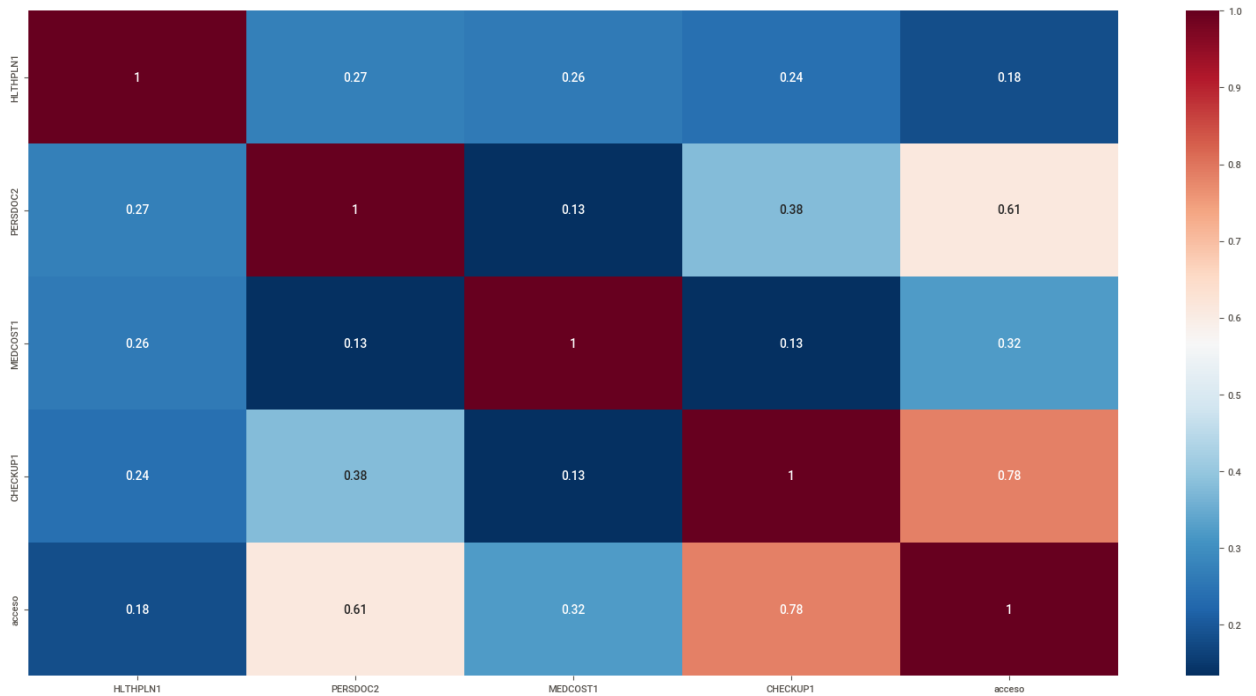
3.3. **MEDCOST:** Hubo algún momento en los 12 meses pasados en los que necesito ver a un doctor, pero no pudo debido a su costo



3.4. CHECKUP1: Cuanto tiempo hace desde su última visita a un doctor para un chequeo médico de rutina



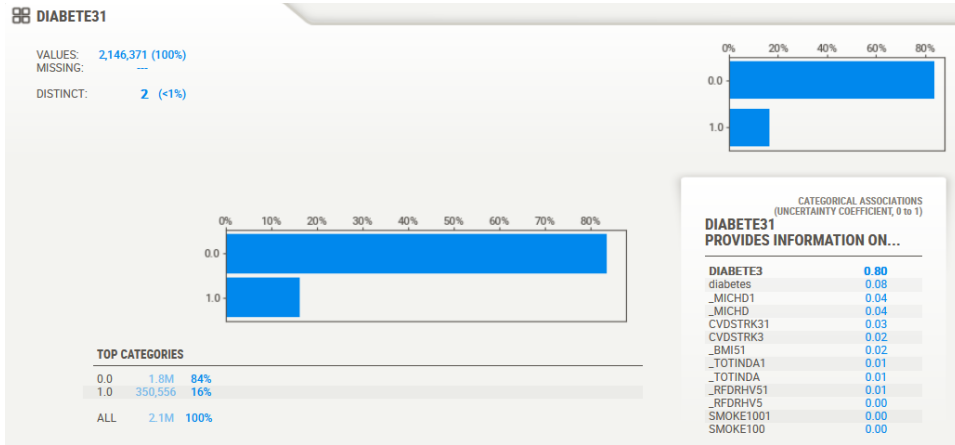
Nuevamente se obtuvo una matriz de correlaciones con las variables originales, con excepción de aquellas que, sin cambiar el nivel de la escala, fueron recategorizadas con el fin de coincidir el sentido de esta.



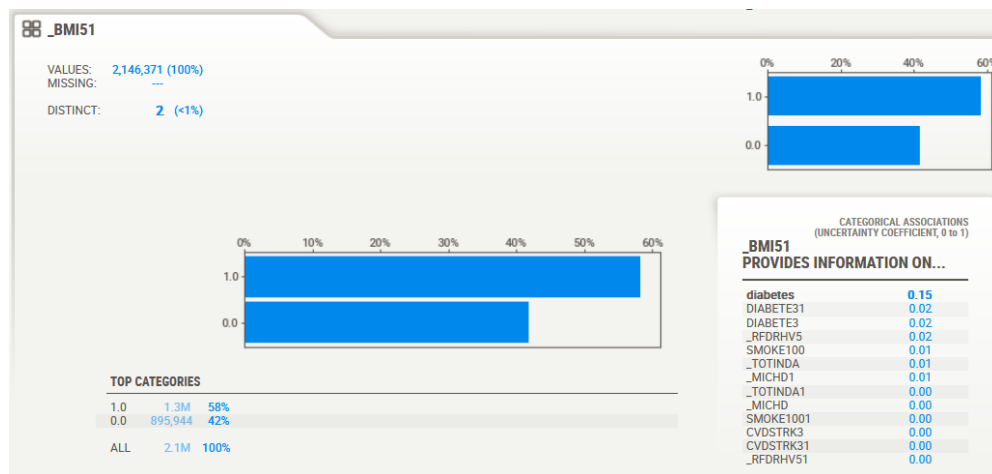
- Riesgo de tener multimorbilidades:** al igual que los indicadores anteriores, este indicador se construyó considerando las preguntas de la encuesta que informan sobre enfermedades crónicas, de forma tal de identificar de cuantas sufre el encuestado y por tanto estimar su gravedad. Cada una de las variables que integran la dimensión fueron recodificada como una variable binaria (0 = no tiene la enfermedad, 1 = tiene la enfermedad).

Las variables consideradas para construir este indicador fueron:

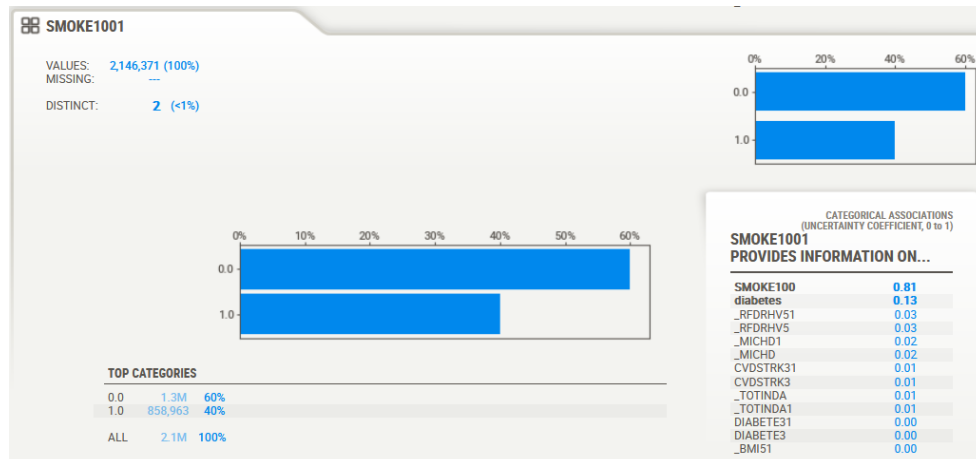
- CVDINFR4:** Alguna vez ha tenido un ataque al corazón, también llamado infarto al miocardio.



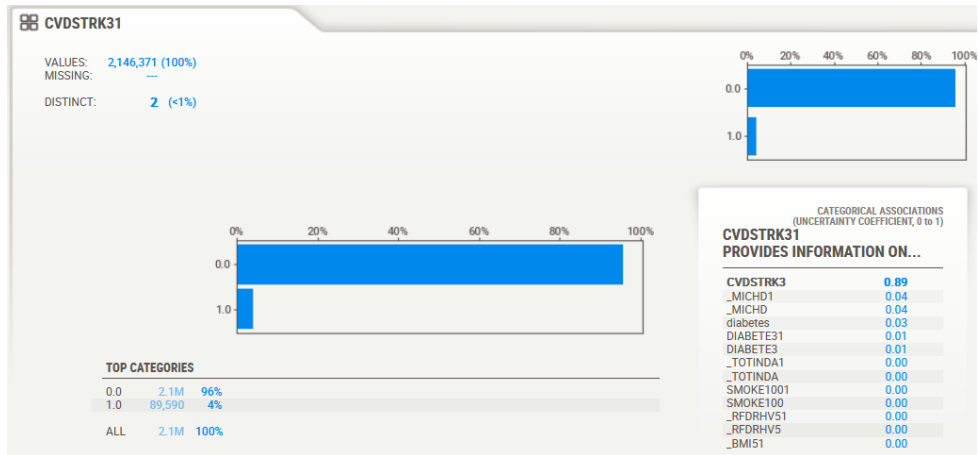
4.2. CVDCRHD4: Alguna vez ha tenido angina de pecho, también llamada enfermedad coronaria.



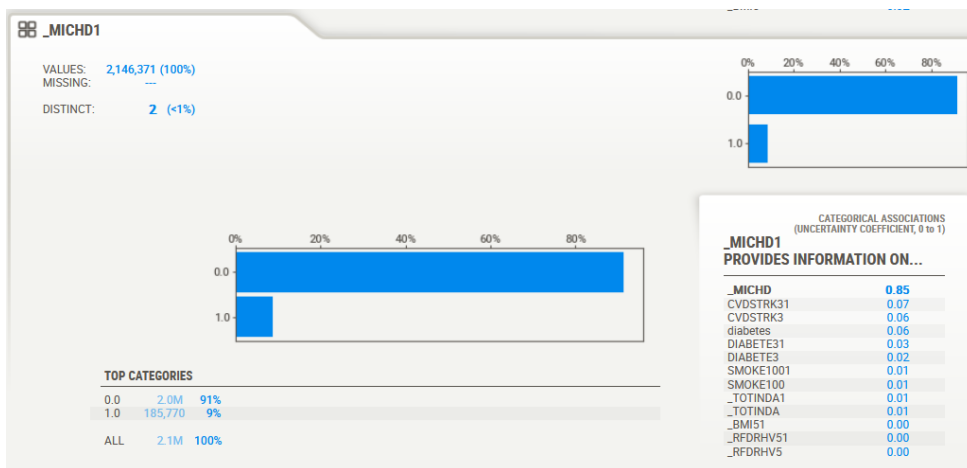
4.3. CVDSTRK3: Ha tenido un accidente cerebro vascular



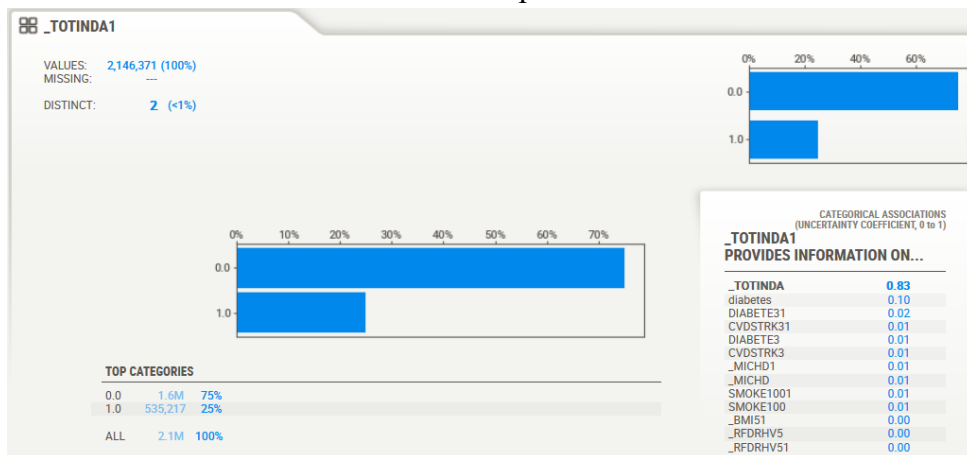
4.4. ASTHMA3: Ha tenido asma



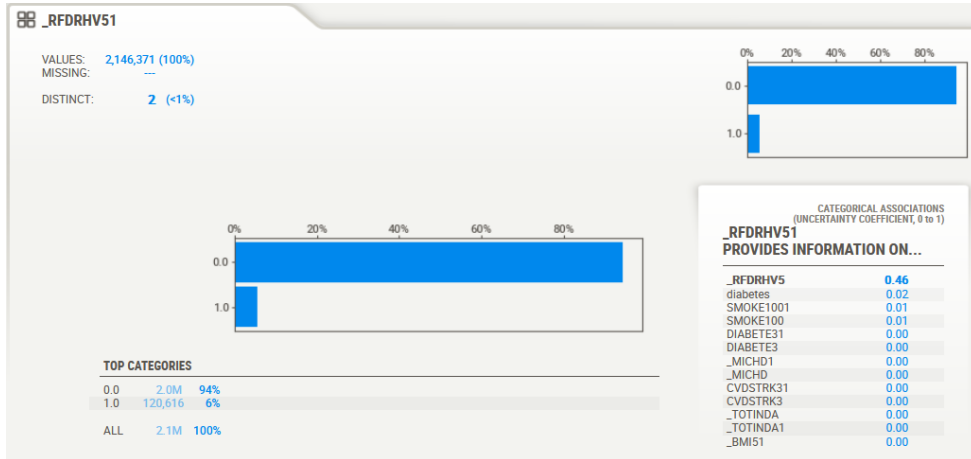
4.5. ASTHNOW: Aún tiene Asma



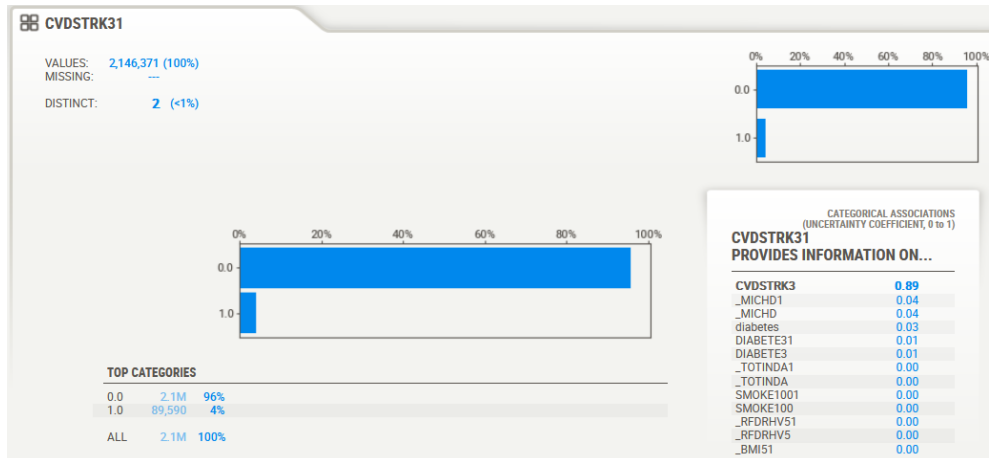
4.6. CHCSCNCR: ha tenido cáncer de piel



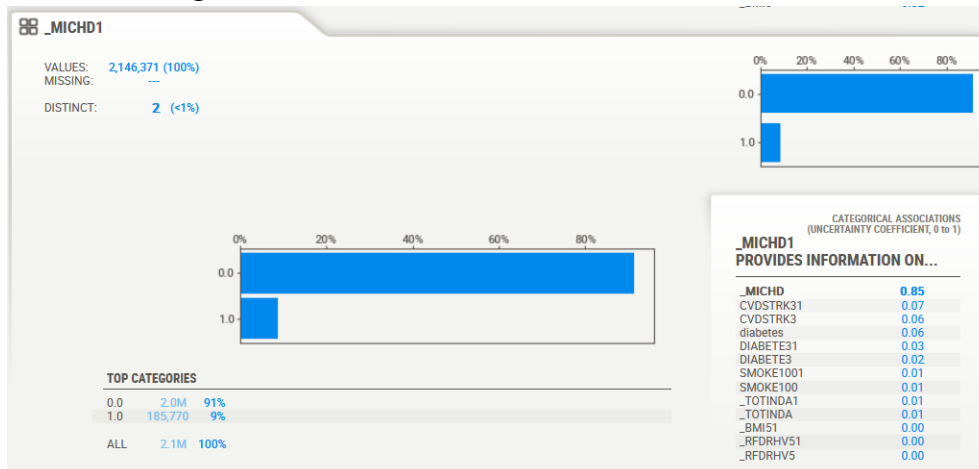
4.7. CHCOCNCR: Ha tenido algún otro tipo de cáncer



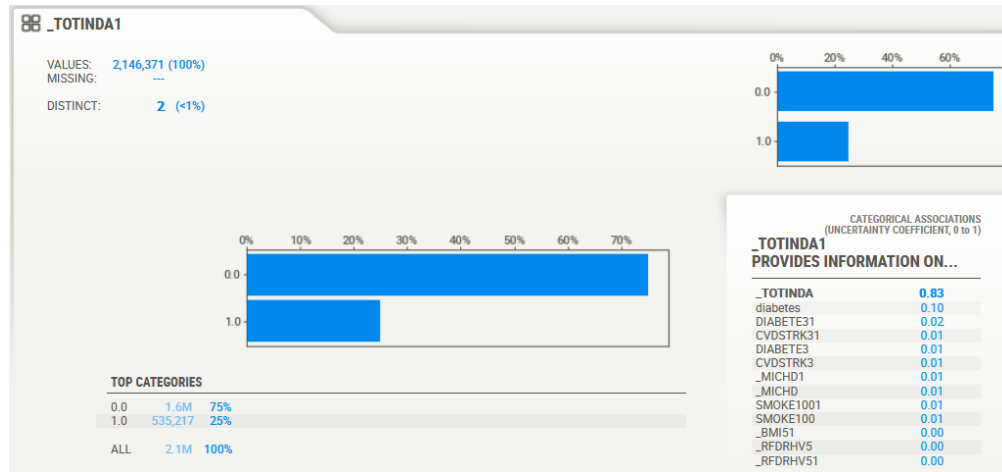
4.8. CHCCOPD1: Tiene EPOC, enfisema o bronquitis crónica



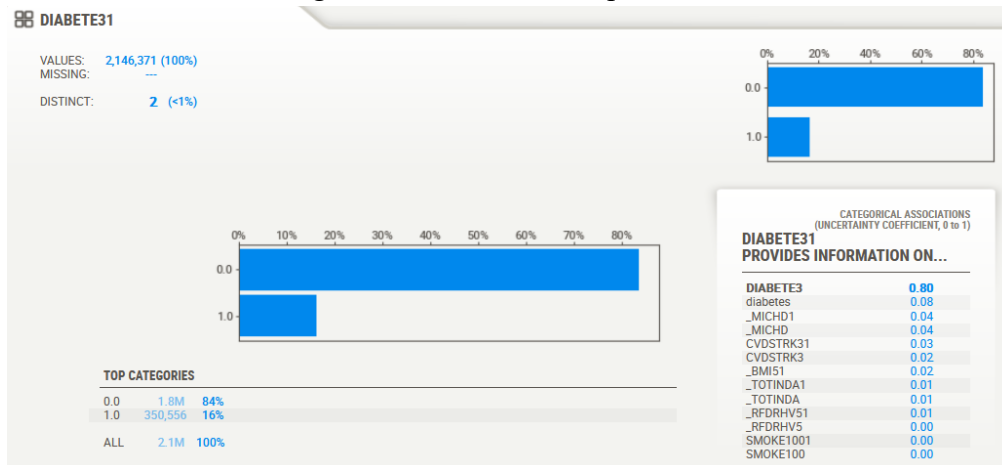
4.9. HAVARTH3: Ha tenido alguna forma de artritis, artritis reumática, lupus o fibromialgia



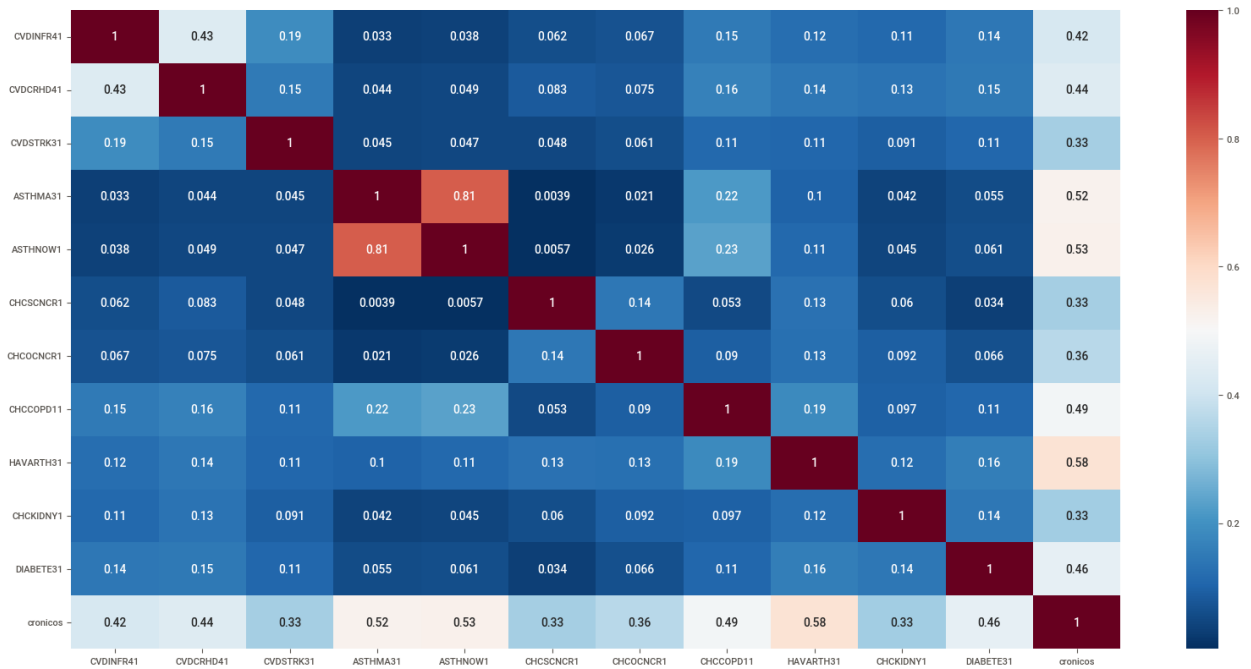
4.10. CHCKIDNY: ha tenido enfermedades al riñón



4.11. DIABETE3: Alguna vez le han dicho que tiene diabetes

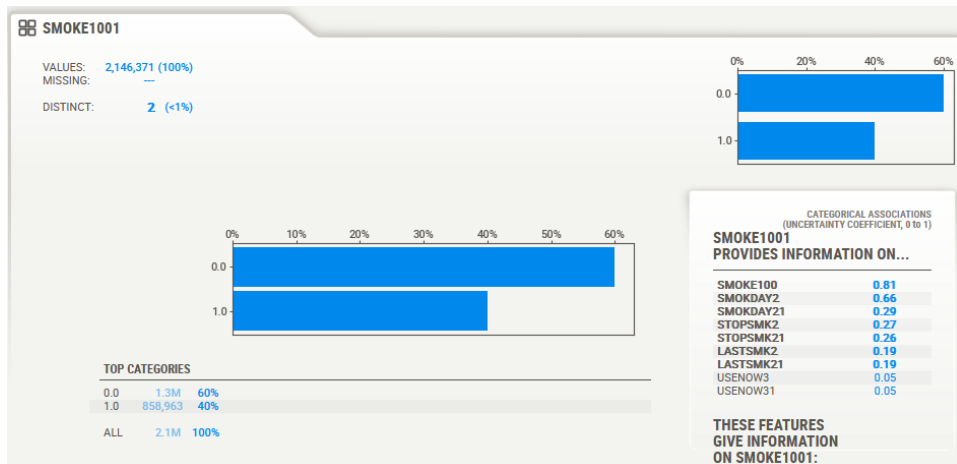


Una vez más se revisa la correlación entre las variables con el fin de validar la pertinencia del indicador creado.

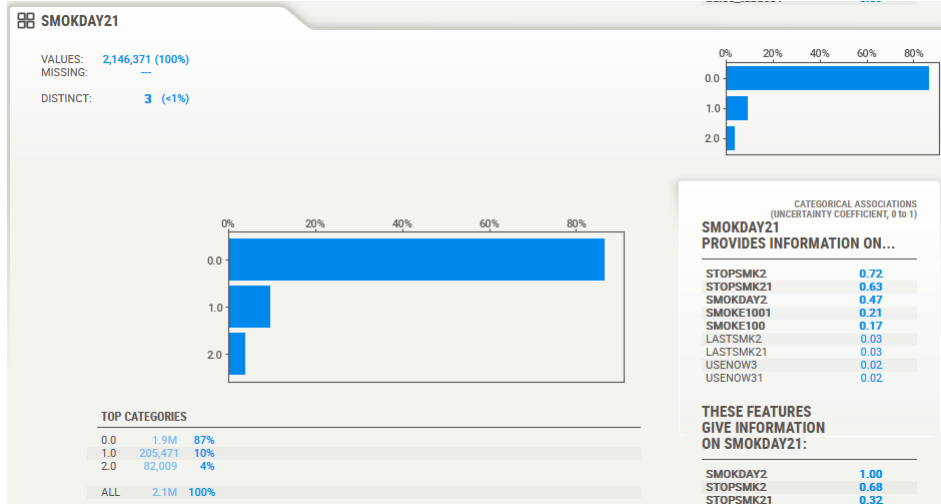


5. **Riesgo de ser adicto al tabaco:** En este indicador confluyen las preguntas de la encuesta destinadas a conocer los comportamientos del encuestado en torno al uso del tabaco. Es un indicador valioso que se enfoca en evaluar la probabilidad de adicción al tabaco en los individuos, basándose en sus comportamientos y respuestas a preguntas relacionadas con el uso del tabaco. Esta dimensión es crucial para estudios sobre salud pública y comportamientos de riesgo. Las variables consideradas para construir este indicador fueron:

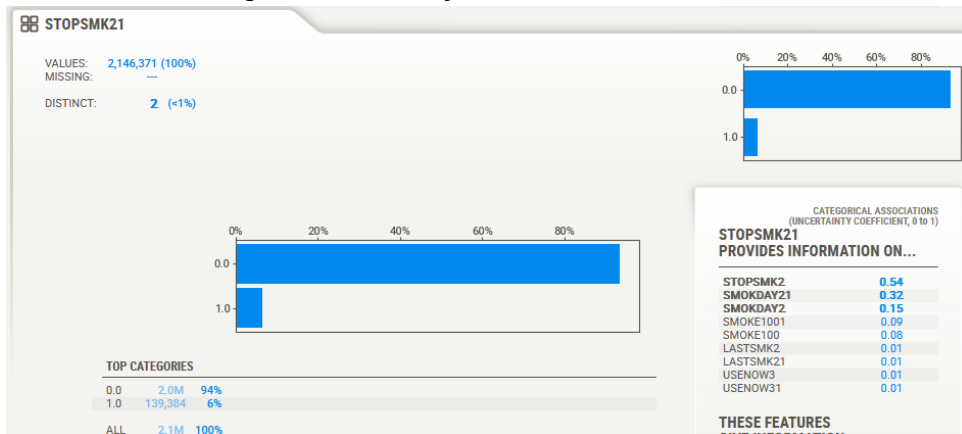
5.1. SMOKE100: Ha fumado al menos 100 cigarrillos en tu vida



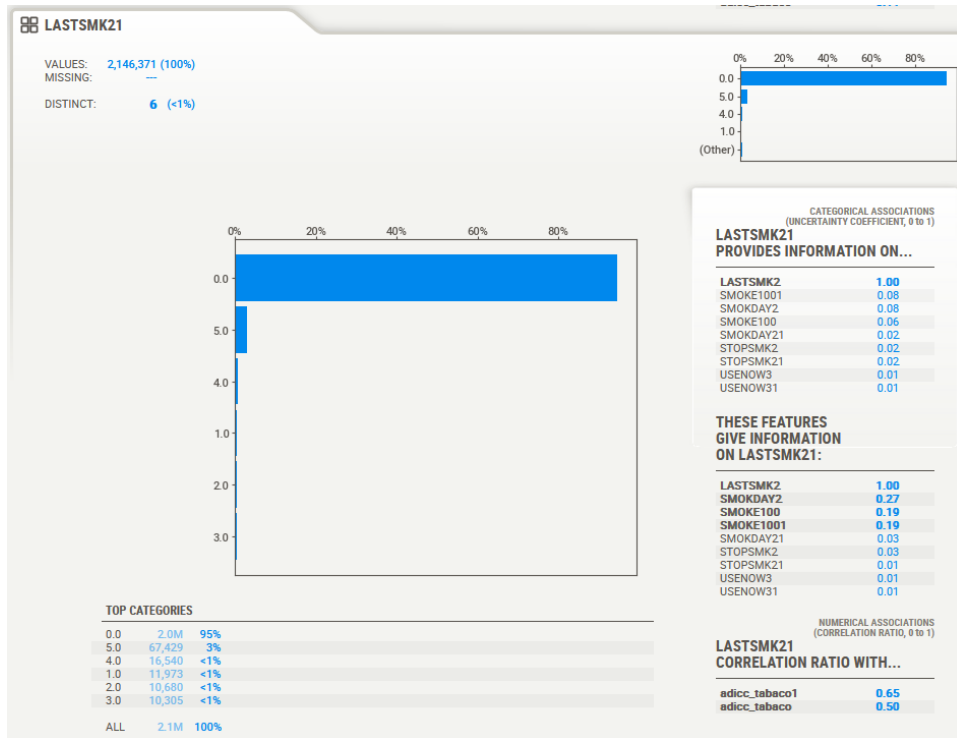
5.2. SMOKDAY2: fumas cigarrillos todos los días, algunos días o nunca



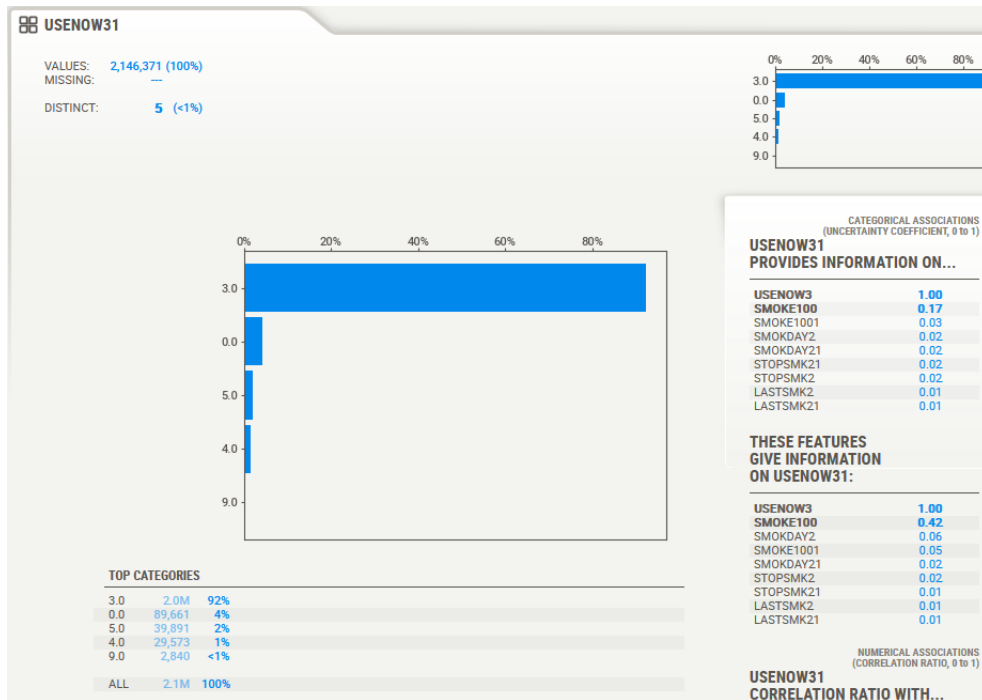
5.3. STOPSMK2: Durante los pasados 12 meses ha dejado de fumar por 1 o más días debido a que intentaba dejar de fumar



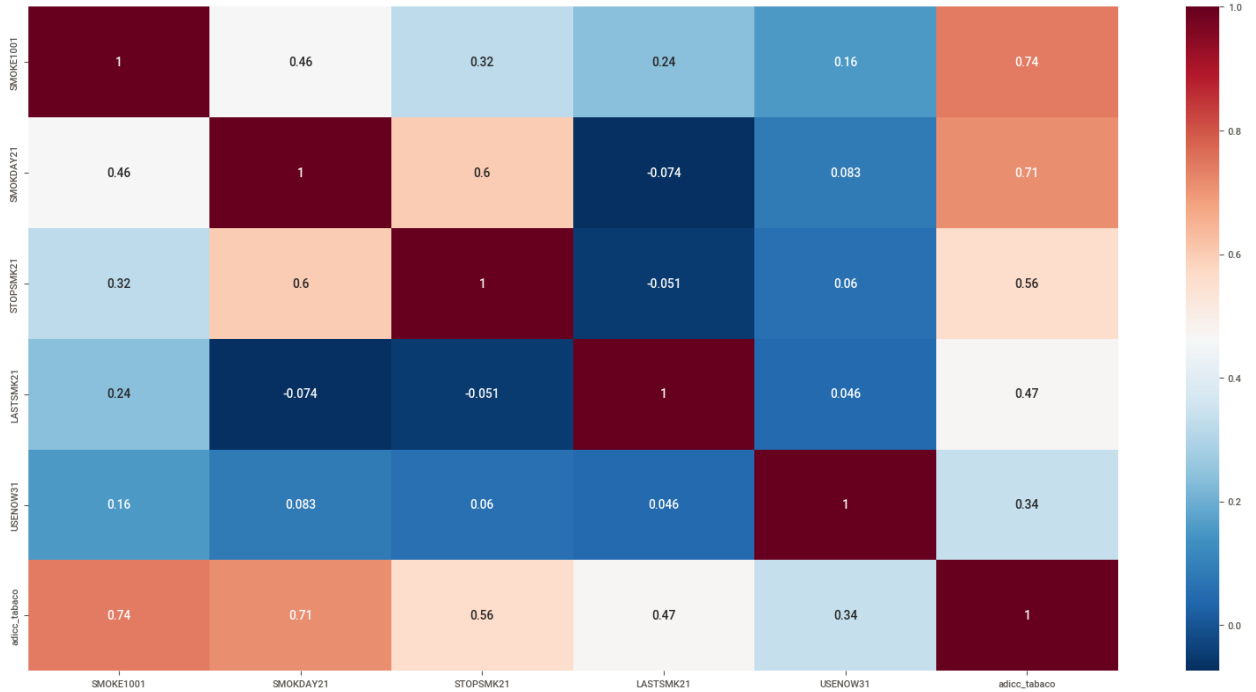
5.4. LASTSMK2: Cuanto tiempo hace desde el último cigarrillo fumado, considerando incluso una o dos fumadas



5.5. USENOW3: Mascar tabaco, usas snuff o snus todos los días, algunos días o nunca



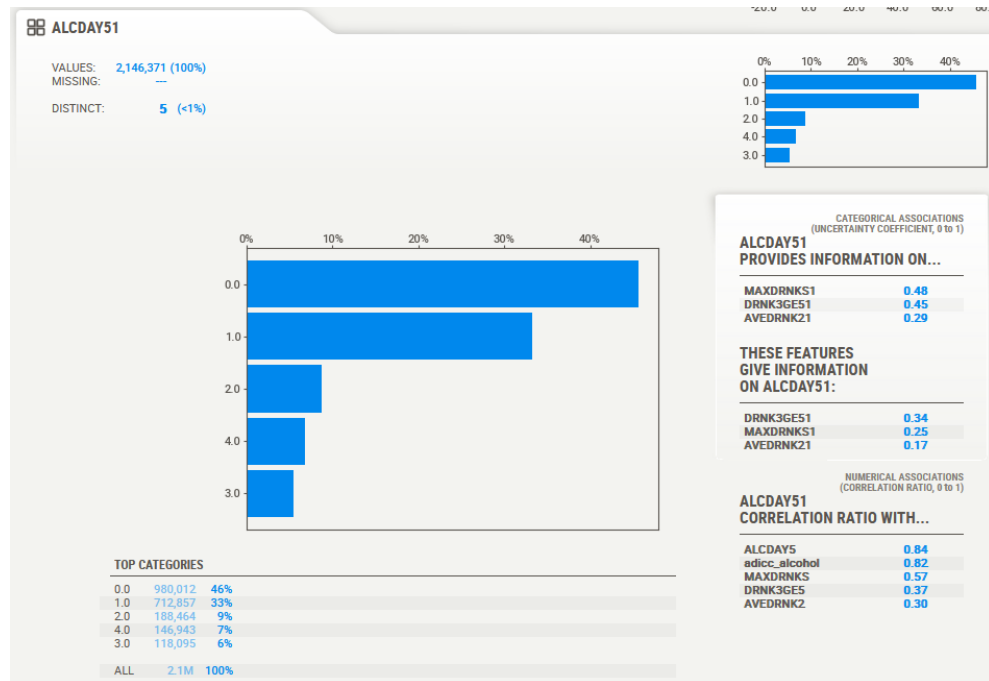
La matriz de correlaciones muestra lo siguiente:



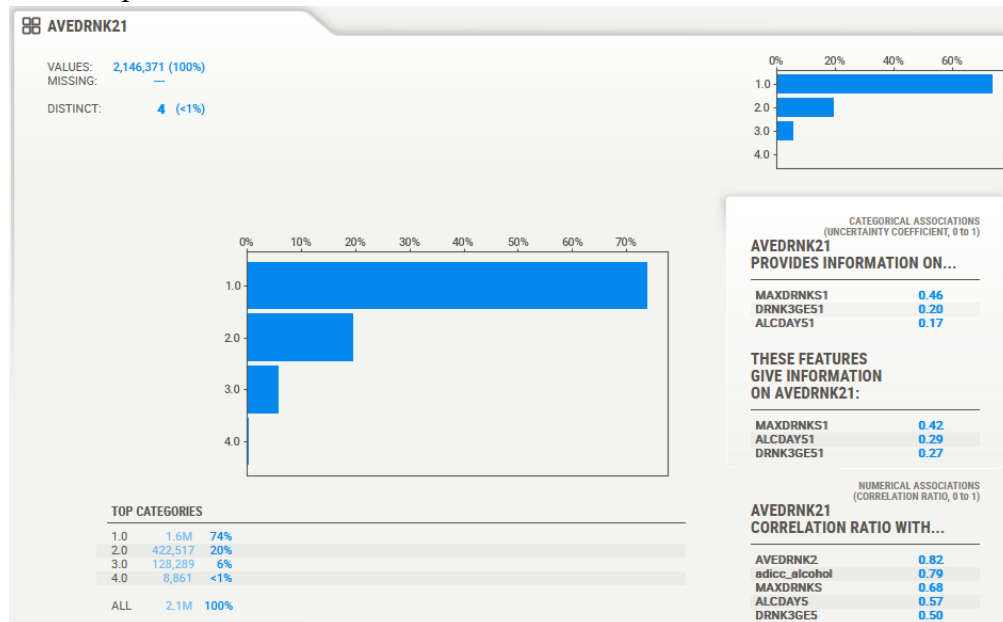
6. **Riesgo de ser adicto al alcohol:** En este indicador confluyen las preguntas de la encuesta destinadas a conocer los comportamientos del encuestado en torno al uso del alcohol. Se centra en identificar hábitos de consumo y otras señales que puedan sugerir una tendencia hacia la adicción al alcohol.

Las variables consideradas para construir este indicador fueron:

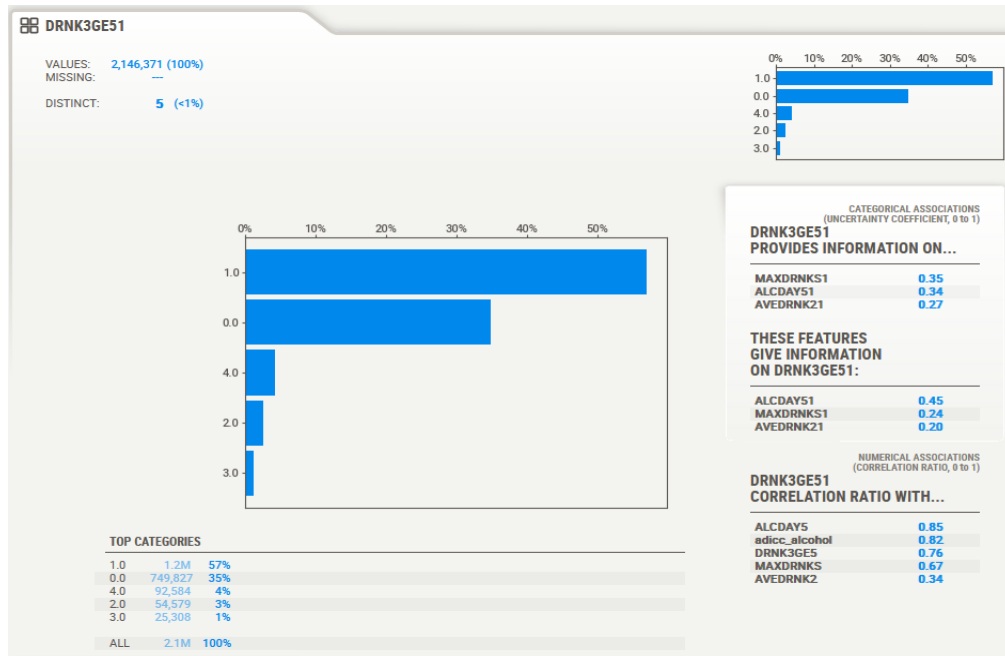
6.1. **ALCDAY5:** Durante los últimos 30 días, por semana o por mes, ha tomado al menos una bebida alcohólica como vino, cerveza, malta o licor



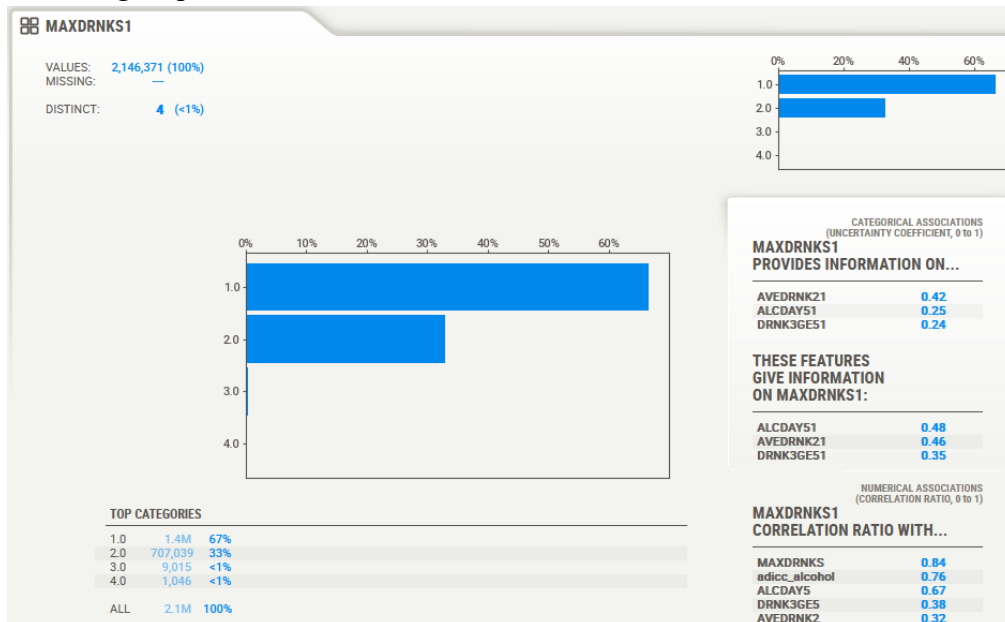
6.2. **AVEDRNK2:** Si 1 trago es equivalente a 12 onzas de cerveza, a 5 onzas de vino, o un shot de licor. Durante los 30 últimos días cuantos tragos ha tomado en promedio



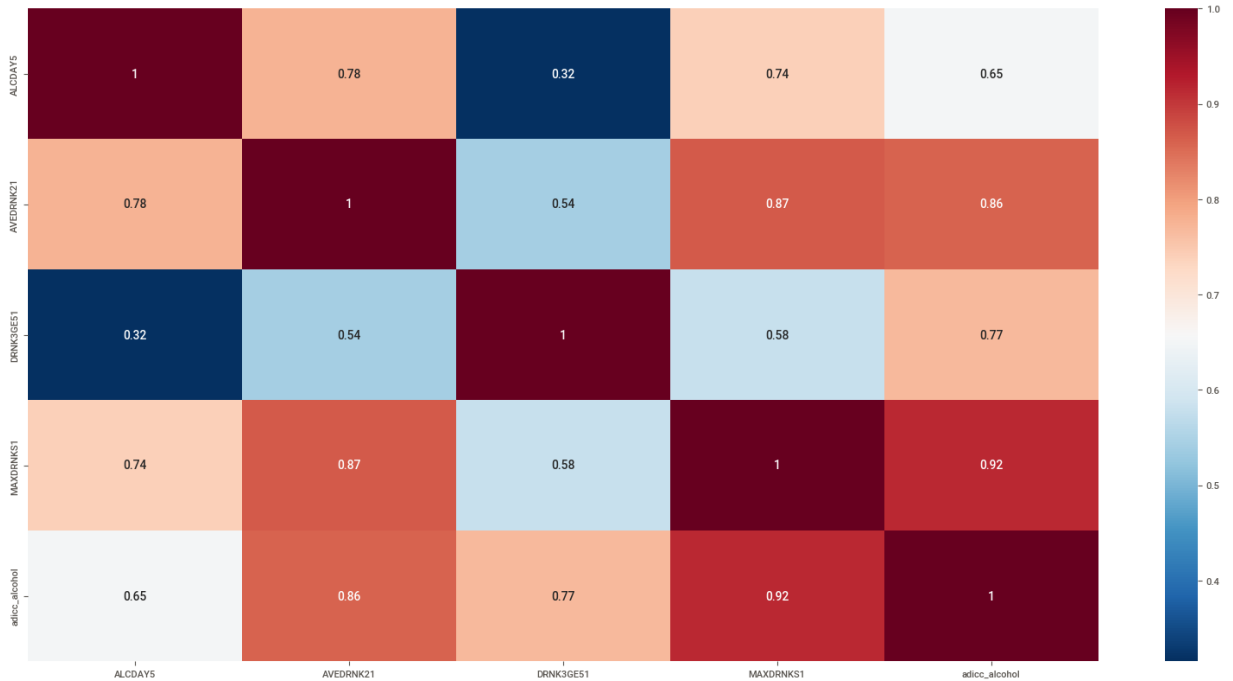
6.3. **DRNK3GE5**: Considerando todo tipo de bebidas alcohólicas, cuantas veces ha tomado más de 5 o más tragos en hombres o 4 y más en mujeres, durante los últimos 30 días.



6.4. **MAXDRNKS**: Durante los últimos 30 días, cual es el mayor número de tragos que ha tenido en una ocasión



Una vez más se revisa la correlación entre las variables con el fin de validar la pertinencia del indicador creado.

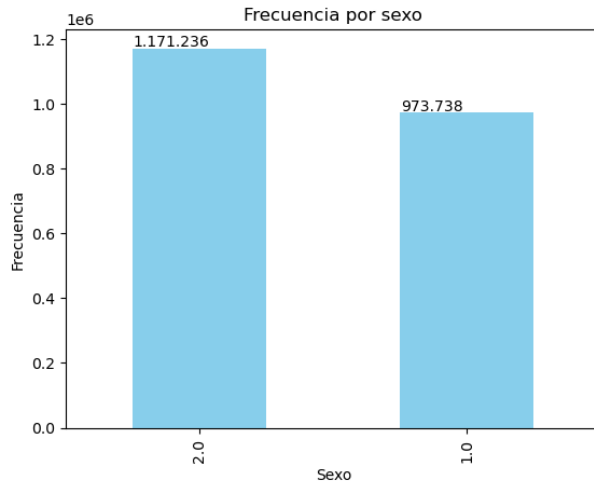


7. Variables sociodemográficas

De la encuesta se utilizaron las siguientes variables sociodemográficas, que fueron recategorizadas con el fin de darles relevancia a los aspectos explicativos de ellas:

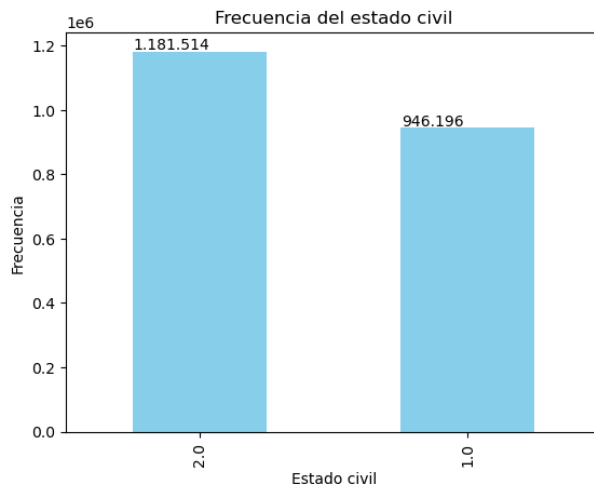
Sexo: Esta variable pide señalar el sexo del encuestado. Se recategorizaron como datos perdidos, tanto aquellos que no contestaron como aquellos que se negaron a contestar la pregunta.

SEX			
Indicate sex of respondent			
1	Male	1	Hombre
2	Female	2	Mujer
7	NAN	NAN	N/A
9	Refused	NAN	N/A



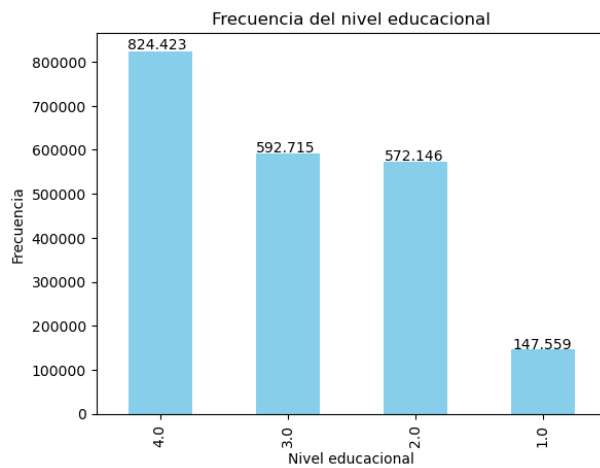
Estado Civil: Esta variable pregunta el estado marital de encuestado. Se recategorizó en solo dos opciones, se tiene pareja o no; puesto que no es relevante para el análisis determinar el estado legal delo encuestado.

MARITAL			
Are you: (marital status)			
1	Married	2	con pareja
2	Divorced	1	sin pareja
3	Widowed	1	sin pareja
4	Separated	1	sin pareja
5	Never married	1	sin pareja
6	a memeber ofan unmarried couple	2	con pareja
9	Refused	nan	N/A
blank	not asked or missing	nan	N/A



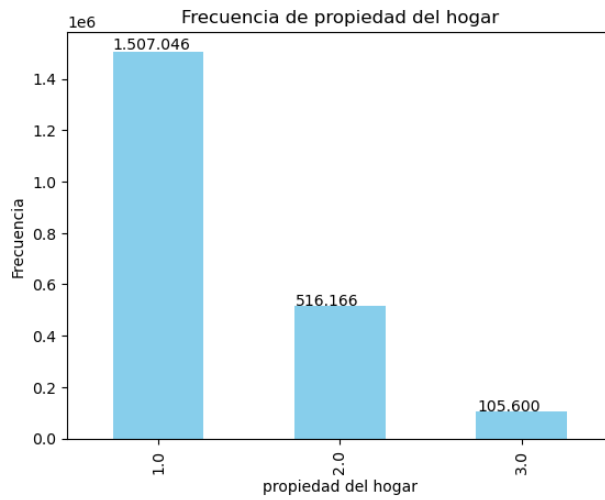
Nivel educacional: Esta variable contiene el nivel educacional del encuestado. Se recategorizó las faltas de respuestas como datos perdidos.

_EDUCAG			
Level of education completed			
1	Did not graduate High School	1	Media incompleta
2	Graduated High School	2	Media completa
3	Attended College or Technical School	3	Universitaria incompleta
4	Graduated from College or Technical School	4	Universitario completa
9	Don't know/Not sure/Missing	NAN	N/A



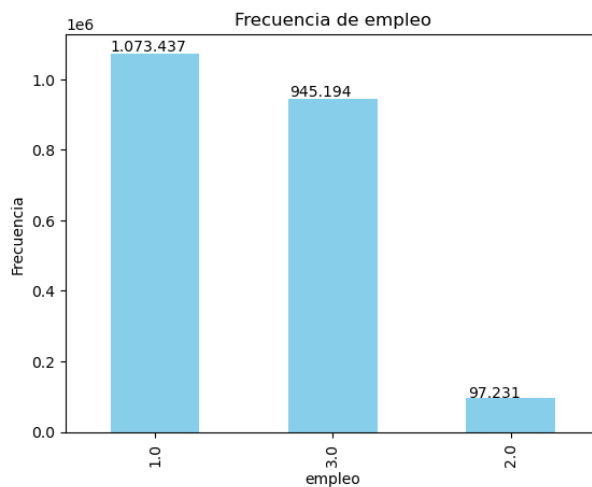
Propiedad del hogar: Esta variable contiene el estado de propiedad del hogar del encuestado; al igual que ven la variable anterior, solo se recategorizó las faltas de respuestas como datos perdidos.

RENTHOM1			
Do you own or rent your home?			
1	Own	1	Propia
2	Rent	2	Arriendo
3	Other arrangement	3	Otro
7	Don't know/Not Sure	nan	N/A
9	Refused	nan	N/A
blank	not asked or missing	nan	N/A



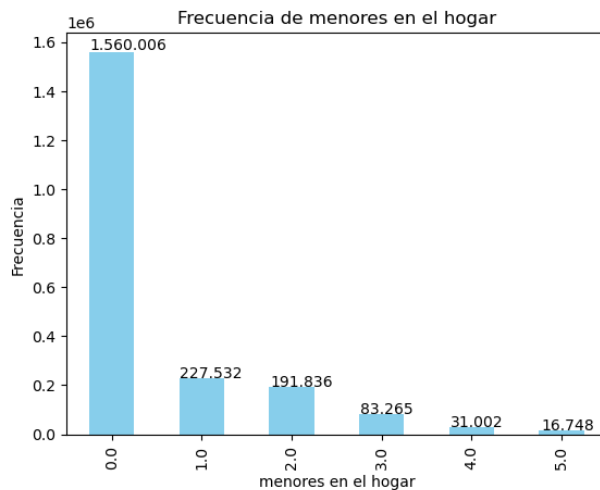
Empleo: En esta variable se pide definir el estado del empleo del encuestado. Se recategorizó en 3 opciones, el encuestado se encuentra empleado, desempleado u otro, en caso de que se defina a si mismo desde actividades distintas al trabajo.

EMPLOY1			
Are you currently...?			
1	Employed for wages	1	Empleado
2	Self-employed	1	Empleado
3	Out of work for 1 year or more	2	Desempleado
4	Out of work for < 1 year	2	Desempleado
5	A homemaker	3	Otro
6	A student	3	Otro
7	Retired	3	Otro
8	Unable to work	3	Otro
9	Refused	nan	N/A
blank	not asked or missing	nan	N/A



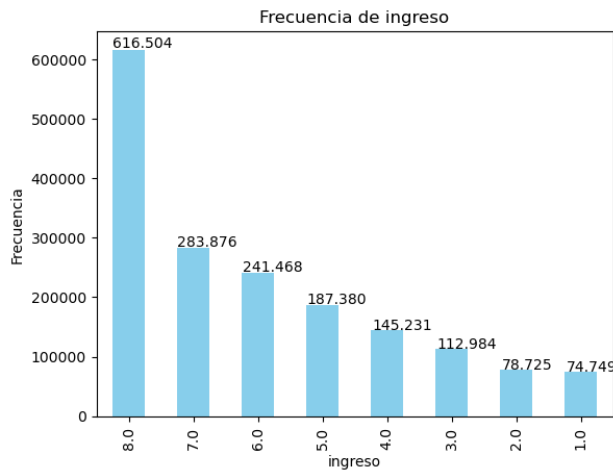
Menores en el hogar: En esta variable se solicita la encuesta indicar cuantos menores viven en el hogar, cuando la respuesta superó los 4 niños por hogar, se recategorizó en la opción “5 niños o más”, esto debido a que, al observar la frecuencia de respuestas de esta pregunta, en la mayoría de estos hogares no es habitual que haya mas de 5 niños.

CHILDREN			
How many children less than 18 years of age live in your household?			
1	Number of children	1	un niño
2	Number of children	2	dos niños
3	Number of children	3	tres niños
4	Number of children	4	4 niños
5 a 87	Number of children	5	5 niños o mas
88	None	0	None
99	Refused	nan	N/A
blank	not asked or missing	nan	N/A



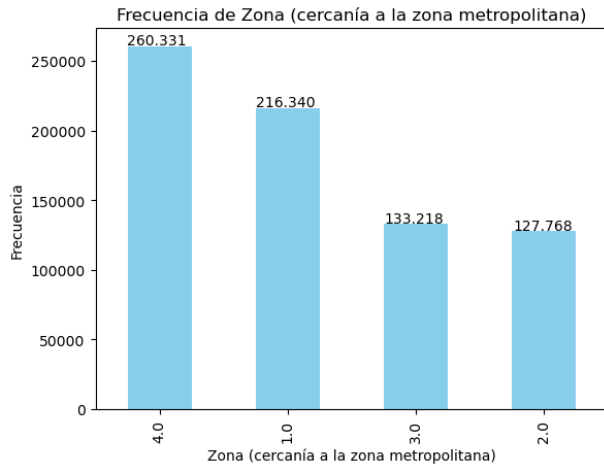
Ingreso: Esta Variable solicita indicar el ingreso anual en dólares del encuestado, Las rentas superiores a 75.000 dólares anuales, se recategorizaron en 75.000 dólares o más debido a que desde el año 2017 a 2019, las categorías llegan a un máximo de 75.000 dólares, solo en el año 2021 se incorporan las categorías posteriores. Al igual que en las variables anteriores, toda falta de respuesta se categorizó como datos perdidos.

INCOME2			
Is your annual household income from all sources: (If respondent refuses at any income level, code 'Refused'.)			
1	Less than \$10,000	10,878	2.53 3.30
2	Less than \$15,000 (\$10,000 to < \$15,000)	11,530	2.68 2.87
3	Less than \$20,000 (\$15,000 to < \$20,000)	14,960	3.48 3.65
4	Less than \$25,000 (\$20,000 to < \$25,000)	21,071	4.90 4.88
5	Less than \$35,000 (\$25,000 to < \$35,000)	43,893	10.21 9.99
6	Less than \$50,000 (\$35,000 to < \$50,000)	48,339	11.25 10.37
7	Less than \$75,000 (\$50,000 to < \$75,000)	59,408	13.82 12.67
8	Less than \$100,000? (\$75,000 to < \$100,000)	47,838	11.13 10.47
9	Less than \$150,000? (\$100,000 to < \$150,000)?	47,642	11.08 11.07
10	Less than \$200,000? (\$150,000 to < \$200,000)	19,769	4.60 5.00
11	\$200,000 or more	18,952	4.41 5.41
77	Don't know/Not sure	36,138	8.41 9.58
99	Refused	49,428	11.50 10.76
blank	not asked or missing		



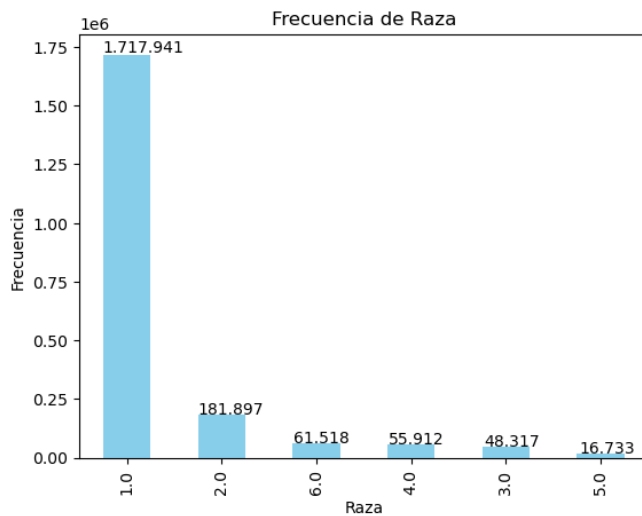
Zona (cercanía a la zona metropolitana): En esta variable se mantuvieron las mismas categorías de origen.

MSCODE			
Metropolitan Status Code			
1	In the center city of an MSA	1	centro de la zona metropolitana
2	Outside the center city of an MSA but inside the county containing the center of the city	2	cerca de la zona metropolitana
3	Inside a suburban county of the MSA	3	lejos de la zona metropolitana
5	Not in an MSA	4	fuera de la zona metropolitana
BLANK	GU, PR, VI, or cell phone	nan	N/A



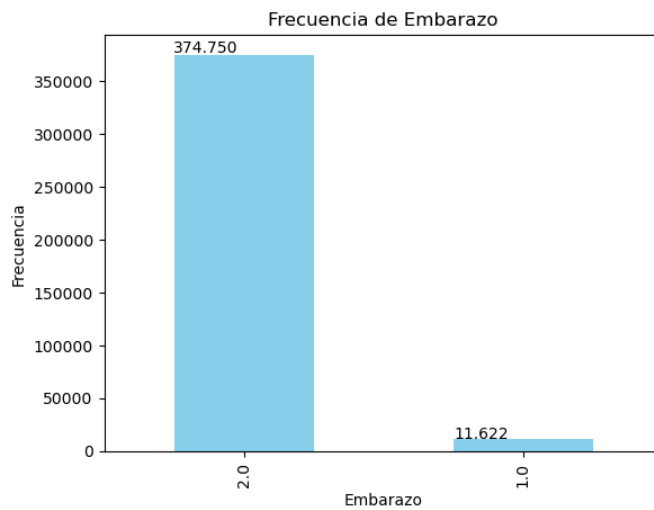
Raza: En esta variable se recategorizaron las respuestas en las que no se eligió raza (7,8,77 y 99) como datos perdidos, puesto que estas categorías impiden, eventualmente hacer un análisis sobre la raza como determinante social.

_PRACE1			
Preferred race category			
1	White	1	blanco
2	Black or African American	2	negro
3	American Indian or Alaskan Native	3	indio americano o nativo de Alaska
4	Asian	4	asiatico
5	Native Hawaiian or other Pacific Islander	5	nativo hawaian o de otras islas del pacifico
6	Other race	6	otra raza
7	No preferred race	nan	N/A
8	Multiracial but preferred race not answered	nan	N/A
77	Don't know/Not sure	nan	N/A
99	Refused	nan	N/A
BLANK	Missing	nan	N/A



Embarazo: En esta variable se recategorizaron las respuestas en las que no se eligió raza (7,8,77 y 99) como datos perdidos, puesto que estas categorías impiden, eventualmente hacer un análisis sobre la raza como determinante social.

PREGNANT			
To your knowledge, are you now pregnant?			
1 Yes		1 Si	
2 No		2 No	
7 Don't know/Not Sure	nan	N/A	
9 Refused	nan	N/A	
blank	not asked or missing	nan	N/A



4.2. Metodología

Este proyecto es analítico y descriptivo en tanto que pretende verificar como ciertos determinantes sociales afectan a la salud. Es una investigación secundaria y correlacional en tanto que opera con datos ya recolectados, busca construir un modelo clasificatorio que permita identificar la probabilidad de tener un determinado estado de salud a partir de los determinantes sociales; de forma tal que sea posible clasificar y priorizar la atención preventiva; sin embargo, y a pesar de que se está utilizando información proveniente de varios años, los análisis son transversales, puesto que no se analiza el cambio de las condiciones a lo largo del tiempo.

Este punto detalla los métodos utilizados para investigar las preguntas de investigación planteadas en el estudio. La atención se centra en el uso de técnicas de aprendizaje automático, específicamente árboles de decisión y algoritmos de bosque aleatorio, para modelar y evaluar el impacto de los determinantes sociales en la prevalencia de enfermedades crónicas y los comportamientos de riesgo relacionados con la salud en la población adulta. Utilizando datos del Sistema de Vigilancia de Factores de Riesgo del Comportamiento (BRFSS) entre 2017 y 2021

Para este proyecto se utilizarán los modelos de árboles de decisión y bosques aleatorios. Con el fin de clasificar los determinantes sociales mencionados en el punto anterior en relación con las dimensiones de salud.

La utilización de árboles de decisión, un método de aprendizaje automático supervisado es crucial en el ámbito de la ciencia de datos aplicada a la salud pública. Permite predecir y categorizar con precisión los riesgos para la salud teniendo en cuenta los determinantes sociodemográficos. Esta herramienta invaluable ayuda en la formulación de intervenciones efectivas y políticas bien informadas. Este estudio emplea específicamente árboles de decisión para examinar la correlación entre estos factores y diferentes aspectos de la salud, que abarcan riesgos de adicción tanto al tabaco como al alcohol, probabilidades de multimorbilidades y accesibilidad a la cobertura sanitaria.

El análisis que nos ocupa se presta bien a la utilización de árboles de decisión, ya que poseen la capacidad de manejar varios tipos de datos (tanto numéricos como categóricos) y son conocidos por su sencilla interpretabilidad. A diferencia de los modelos alternativos de aprendizaje automático, los árboles de decisión producen reglas explícitas y comprensibles, lo que facilita una mejor comprensión de las conexiones entre las variables y permite una toma de decisiones informada basada en los resultados del modelo.

El proceso de desarrollo de un árbol de decisión implica dividir un conjunto de datos en subconjuntos más pequeños, lo que en última instancia da como resultado una estructura de árbol que consta de nodos de decisión y nodos hoja. Los nodos de decisión sirven como puntos donde el conjunto de datos se divide en función de variables de prueba, mientras que los nodos hoja significan clasificaciones o decisiones. Básicamente, un árbol de decisiones es un modelo predictivo que se utiliza para analizar y organizar datos. La construcción de árboles de decisión en esta investigación implicó el uso de variables sociodemográficas como nodos de decisión. Estas variables abarcaron sexo, nivel educativo, ingresos, tipo de empleo, estado civil, número de hijos, propiedad del hogar, área de residencia y raza. Las variables objetivo, por su parte, fueron las dimensiones de salud, que fueron evaluadas mediante indicadores preestablecidos en el estudio.

No se puede subestimar la importancia de los árboles de decisión en el análisis de la salud. La alta interpretabilidad de los árboles de decisión es uno de sus beneficios más notables. A diferencia de modelos intrincados como las redes neuronales, los árboles de decisión ofrecen una visualización y comprensión sencillas, lo que los hace indispensables en el ámbito de la salud pública para transmitir de manera efectiva los hallazgos a los tomadores de decisiones.

Los árboles de decisión tienen la capacidad de manejar eficazmente una amplia gama de tipos de datos, incluidos datos categóricos y numéricos, lo que los hace muy adecuados para el análisis de diversos datos sociodemográficos y de salud. La capacidad de los árboles de decisión para capturar relaciones no lineales entre variables es muy ventajosa en el análisis de datos de salud, especialmente cuando se trata de conexiones intrincadas y no lineales entre los determinantes sociales y los resultados de salud.

El proceso de desarrollo del modelo de árbol de decisión comenzó con la selección y preparación de los datos apropiados. Esto implicó tareas como limpiar los datos, abordar los valores faltantes y codificar variables categóricas. Posteriormente, una parte del conjunto de datos se utilizó para entrenar el modelo, mientras que otra parte se reservó para evaluar su rendimiento y precisión. Una vez finalizado el proceso de formación, el siguiente paso fue evaluar la significancia de las variables. Esto implicó examinar los factores sociodemográficos que desempeñaban un papel crucial en la previsión de diversos riesgos para la salud.

A través de este análisis, se obtuvo una valiosa comprensión de las conexiones entre las variables; además, también se identificaron posibles vías de intervención. La aplicación

del modelo de árbol de decisiones en esta investigación tiene un inmenso potencial en el ámbito del desarrollo de políticas de salud y la formulación de intervenciones específicas.

Mediante la identificación de las variables sociodemográficas que ejercen el mayor impacto en las vulnerabilidades de la salud, los formuladores de políticas pueden asignar recursos estratégicamente y centrar sus esfuerzos en los dominios más cruciales. Sin embargo, es de suma importancia el reconocer las limitaciones asociadas con esta metodología. Si bien los árboles de decisión son modelos robustos, pueden ser susceptibles de sobreajuste, particularmente cuando se enfrentan a una amplia gama de variables. Además, las decisiones tomadas con base en los resultados del modelo deben tener en cuenta el contexto más amplio y los factores cualitativos que no abarca el modelo.

Con el fin de determinar la importancia de las variables sociodemográficas al interior del modelo y de esta forma facilitar la clasificación de la prioridad de atención preventiva a partir de determinados condicionantes sociales, se usarán los bosques aleatorios es un algoritmo altamente eficiente que se utiliza tanto en tareas de clasificación como de regresión, lo que lo hace adecuado para una amplia gama de desafíos de datos. Funciona como una colección de árboles de decisión, normalmente entrenados mediante la técnica de "ensacado". El ensacado implica fusionar numerosos modelos, específicamente árboles de decisión en este contexto, para construir un modelo compuesto más resistente y preciso.

Para mejorar la solidez del modelo y evitar el sobreajuste, es aconsejable promediar las predicciones de cada árbol individual, ya que esto reduce efectivamente la varianza del modelo, un problema común que se encuentra en los árboles de decisión independientes. Cuando se trata de gestionar grandes conjuntos de datos, este sistema tiene la capacidad de procesar miles de entradas de datos y evaluar la importancia de cada característica en el proceso de clasificación. Debido a la naturaleza no paramétrica de los árboles de decisión, destacan en el manejo eficaz de datos no lineales mediante la gestión eficaz de límites de decisión no lineales. La predicción de variables dependientes es una herramienta versátil que se puede aplicar a variables de salida categóricas como es en este caso.

Para mitigar el riesgo de sobreajuste y al mismo tiempo minimizar el sesgo, un enfoque eficaz es emplear el método de embolsado y utilizar varios árboles. Importancia de las características: esta métrica ofrece información valiosa sobre la importancia de las características en varios árboles, lo que la convierte en una herramienta valiosa para fines de selección de características. Cuando se enfrenta a datos faltantes, posee la capacidad

de mantener la precisión incluso en ausencia de una parte de los datos, lo que demuestra su resiliencia.

La paralelización se logra generando los árboles de forma independiente, lo que permite un proceso fácilmente paralelizable. Desde la predicción de enfermedades en el campo médico hasta la modelización de riesgos en finanzas, esta herramienta versátil ha demostrado ser exitosa en una amplia gama de industrias. Si bien Bosques aleatorios tiene algunas limitaciones, como su intrincado modelo que puede ser más difícil de interpretar que un único árbol de decisión y sus mayores requisitos de recursos computacionales debido a la abundancia de árboles, su amplia popularidad en la comunidad de ciencia de datos se puede atribuir a su rendimiento excepcional y su naturaleza fácil de usar, lo que le hace en una herramienta de tremenda eficiencia para los objetivos planteados en el punto 3.

Para generar los modelos planteado en los párrafos anteriores, se usó el lenguaje de programación Python.

Las bibliotecas usadas fueron las siguientes:

Mensajes y sistema:

- Warnings
- OS
- Re

Tratamiento y limpieza de datos:

- CSV
- datetime
- Pandas
- Numpy

Generación de modelos:

- Scikit-learn
- scipy

Visualización:

- Matplotlib
- Seaborn
- Sweetviz

5. Resultados

Los árboles de decisión que se obtuvieron fueron los siguientes:

5.1.1. Modelo de Riesgo de tener un mal estado de salud físico y mental.

Para este árbol de decisión se armó un dataset que contiene las siguientes variables:

1. La dimensión “Riesgo de tener un mal estado de salud físico y mental” fue recategorizada en una variable dicotómica de alto y bajo riesgo.
2. Las variables de sexo, estado civil, nivel educacional, propiedad del hogar, empleo, menores en el hogar, Ingreso, zona (cercanía a la zona metropolitana), Raza y embarazo.
3. Para mejorar el rendimiento del modelo y evitar el sobreajuste hacia alguna de las clases las clases dicotómicas de la dimensión, se realizó un balance de clases previo al entrenamiento del modelo

El árbol de decisión se entrenó con los siguientes parámetros:

- **Entrenamiento y prueba:** entrenamiento de un 70% y un set de pruebas de 30%
- **Criterio de selección (criterion):** Como criterio de selección se utilizó la entropía; es decir, la cantidad de desorden de la información, donde 0 implica un conjunto de muestras completamente homogéneas (mayor “pureza” en la selección) y 1 implica que las muestras son divididas equitativamente (mayor “impureza” en la selección). Mide la ganancia de información, o que tan bien un atributo particular separa las muestras de entrenamiento de acuerdo con la clasificación.
- **Divisor (splitter):** Se utiliza la mejor (best) estrategia de división de nodos en subnodos y no la estrategia de división aleatoria (random).

- **Máxima profundidad (max_depth):** Se forzó una profundidad máxima de 4 nodos
- **División mínima de muestras (min_samples_split):** se fijó en 30 el número mínimo de muestras que debe tener un nodo para que pueda dividirse.
- **Compensación de balances (class_weight):** se balancearon las categorías dicotómicas de la dimensión “Riesgo de tener un mal estado de salud físico y mental”
- **Estado aleatorio (random_state):** es una semilla para generar números aleatorios reproducibles, haciendo reproducibles los resultados. Se uso 1000 como semilla

El reporte de clasificación arrojó los siguientes resultados:

```

La precisión del árbol de decisión es de 0.601
      precision    recall  f1-score   support

 0.0         0.58     0.74     0.65     353542
 1.0         0.64     0.46     0.54     353203

 accuracy                   0.60     706745
 macro avg                 0.61     0.60     0.59     706745
 weighted avg              0.61     0.60     0.59     706745

```

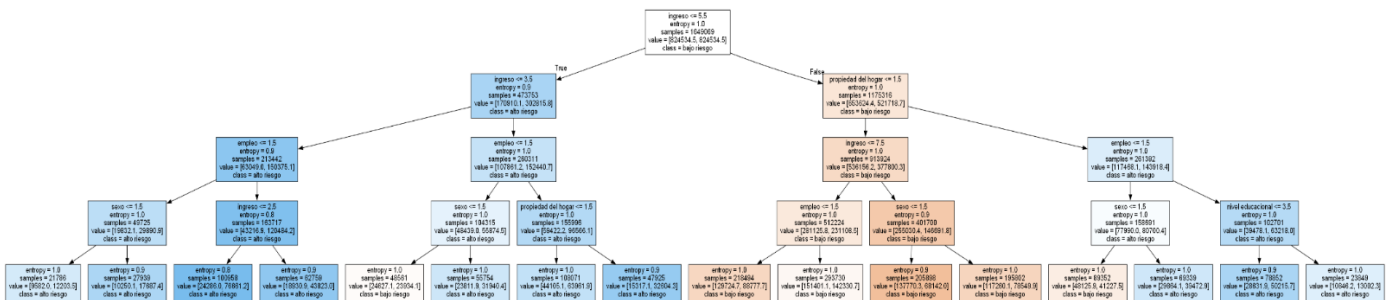
La precisión general del modelo alcanza un 60%

La clase 0, denominada como de "bajo riesgo" obtiene una precisión de 0,58, por lo que se puede inferir que aproximadamente el 58% de los casos clasificados como "bajo riesgo" por el modelo están efectivamente clasificados correctamente como "bajo riesgo". Con una recuperación (sensibilidad) de 0,74, el modelo detectó con precisión el 74 % de las instancias de "bajo riesgo" en el conjunto de datos. El F1-Score, que es la media armónica de precisión y recuperación, sirve como indicador del equilibrio entre estas dos métricas. Una puntuación F1 más alta implica un equilibrio óptimo entre precisión y recuperación. En este caso, un valor de 0,65 sugiere un nivel encomiable de desempeño en la categoría de "bajo riesgo". En el conjunto de datos de prueba, hay un total de 335.342 casos reales categorizados como "bajo riesgo".

Por otro lado, la clase de "alto riesgo" obtiene una puntuación de precisión de 0,64 significa que aproximadamente el 64 % de los casos clasificados como "alto riesgo" por el modelo se clasifican correctamente como "alto riesgo". La puntuación de recuperación de 0,46 indica que el modelo detectó con precisión el 46% de las instancias de "alto riesgo" en el conjunto de datos. Con base en una puntuación F1 de 0,54, se puede inferir que el

modelo exhibe un desequilibrio entre precisión y recuperación para la clase de "alto riesgo" en comparación con la clase de "bajo riesgo". En el conjunto de datos de prueba, hay un total de 353.203 casos clasificados como "alto riesgo".

El método Macro Avg calcula las métricas para cada clase individualmente y luego calcula el promedio, dando el mismo peso a todas las clases. Los valores de precisión, recuperación y puntuación F1 para este enfoque son 0,61, 0,60 y 0,59 respectivamente. Al considerar el soporte variable para cada clase, el método de promedio ponderado calcula métricas para clases individuales. Este enfoque aborda la cuestión del desequilibrio de clases. Los valores ponderados de precisión, recuperación y puntuación F1 son 0,61, 0,60 y 0,59 respectivamente. Dado que ambas métricas tienen en cuenta todas las clases, no sorprende que la métrica de precisión ponderada se alinee con la precisión general. De manera similar, el retiro ponderado y la puntuación F1 ponderada demuestran un desempeño que es directamente proporcional a la frecuencia de instancias en cada clase.



En el nodo raíz del árbol de decisión, la variable inicial considerada es ingresos $\leq 5,5$. Este criterio significa que el conjunto de datos se dividió primero según el nivel de ingresos a partir de aproximadamente 37.500 dólares anuales; un ingreso menor aumenta el riesgo de que la salud mental y física se vea afectada. Sin embargo, el riesgo disminuye si se tiene empleo y si se es hombre; aunque la propiedad del hogar o el arriendo también pueden incidir.

Si los ingresos son mayores a los 37.500 dólares anuales, el riesgo disminuye, aunque se ve afectado por la ausencia de empleo, o por el sexo. Aunque aspectos como un bajo nivel educacional también pueden incidir en él

5.1.2. Modelo de Riesgo de padecer diabetes.

Para este árbol de decisión se armó un dataset que contiene las siguientes variables:

1. La dimensión “Riesgo de padecer diabetes” fue recategorizada en una variable dicotómica de alto y bajo riesgo.
2. Las variables de sexo, estado civil, nivel educacional, propiedad del hogar, empleo, menores en el hogar, Ingreso, zona (cercanía a la zona metropolitana), Raza y embarazo.
3. Para mejorar el rendimiento del modelo y evitar el sobreajuste hacia alguna de las clases las clases dicotómicas de la dimensión, se realizó un balance de clases previo al entrenamiento del modelo

El árbol de decisión se entrenó con los siguientes parámetros:

- **Entrenamiento y prueba:** entrenamiento de un 70% y un set de pruebas de 30%
- **Criterio de selección (criterion):** Como criterio de selección se utilizó la entropía
- **Divisor (splitter):** Se utiliza la mejor (best) estrategia de división de nodos en subnodos y no la estrategia de división aleatoria (random).
- **Máxima profundidad (max_depth):** Se forzó una profundidad máxima de 4 nodos
- **División mínima de muestras (min_samples_split):** se fijó en 30 el número mínimo de muestras que debe tener un nodo para que pueda dividirse.
- **Compensación de balances (class_weight):** se balancearon las categorías dicotómicas de la dimensión
- **Estado aleatorio (random_state):** es una semilla para generar números aleatorios reproducibles, haciendo reproducibles los resultados. Se uso 10000 como semilla

El reporte de clasificación arrojó los siguientes resultados:

```

La precisión del bosque aleatorio es de 0.610
      precision    recall  f1-score   support

   0.0         0.63     0.53     0.57     334179
   1.0         0.59     0.69     0.64     333533

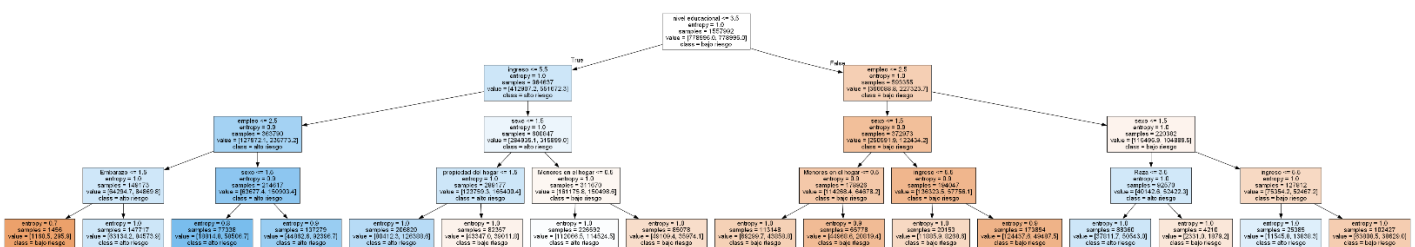
 accuracy                   0.61     667712
 macro avg              0.61     0.61     0.61     667712
 weighted avg          0.61     0.61     0.61     667712

```

La precisión general del modelo alcanza un 61%

Profundizando en los detalles de las métricas para cada clase individual, se puede observar que para la Categoría 0, la precisión del modelo es del 63% para predecir correctamente instancias como clase 0. En retrospectiva, vale la pena señalar que el modelo representó con precisión el 53% de todos los casos reales. La puntuación F1 score alcanza un 0,57, lo que representa una combinación bien equilibrada de precisión y recuperación. El conjunto de prueba contiene una cantidad sustancial de ocurrencias reales de clase 0, con un total de 334,179 instancias. Por su lado, el evento clasificado como Clase 1 predijo con precisión el 59% de ellos. En retrospectiva, vale la pena señalar que el modelo representó con precisión el 69% de todos los casos reales; el F1-Score es de 0,64, lo que representa un equilibrio armonioso entre precisión y recuperación. El conjunto de pruebas contiene un total de 333.553 apariciones reales de clase 1, lo que proporciona pruebas sustanciales de apoyo. Promedios La precisión del modelo, que mide el porcentaje de predicciones correctas, se sitúa en el 61%. La métrica Macro Avg calcula la media de precisión, recuperación y puntuación F1 sin considerar los desequilibrios de clases. En este caso, los tres valores son 0,61, lo que sugiere un desempeño bien equilibrado del modelo en ambas clases cuando no se tiene en cuenta el soporte. El promedio ponderado se calcula teniendo en cuenta la representación de cada clase en los datos. Esto significa que el rendimiento general del modelo se distribuye uniformemente entre todas las clases, como lo indica el hecho de que todas las métricas también son 0,61.

Si bien la precisión general es sólo marginalmente mejor que la de las conjeturas aleatorias cuando las clases se distribuyen equitativamente, el modelo muestra un rendimiento bien equilibrado.



En el nivel raíz del árbol, se tiene en cuenta el atributo nivel educativo $\leq 3,5$ para determinar el nodo más alto; es decir que la diabetes se relaciona fuertemente con el nivel educacional. Este criterio conduce entonces a la división del conjunto de datos en dos ramas. Cuando el nivel educativo es inferior a estudios universitarios, el árbol de decisión avanza a un nodo que tiene en cuenta el ingreso. Si el ingreso es superior

a 37.500 dólares anuales, es el sexo el que afecta el riesgo, afectado por la cantidad de menores en el hogar y por la propiedad o arriendo de este. Si el ingreso es inferior a 37.500 dólares, las variables que determinan el modelo son la tenencia de empleo, el sexo y el embarazo.

Cuando el nivel educacional es alto, los riesgos de padecer diabetes tienden a disminuir, siendo afectados por el empleo, el sexo, el nivel de ingreso y la cantidad de menores en el hogar. De esta forma, Si se es hombre, el ingreso aumenta y hay pocos menores en el hogar, el riesgo de padecer diabetes disminuye,; sin embargo, si se es mujer, un ingreso bajo y se es de raza blanca, el riesgo a padecer diabetes aumenta.

5.1.3. Modelo de cobertura de salud

En este caso el dataset estuvo compuesto de las siguientes variables:

1. La dimensión “cobertura de salud” fue recategorizada en una variable dicotómica de alto y bajo riesgo.
2. Las variables de sexo, estado civil, nivel educacional, propiedad del hogar, empleo, menores en el hogar, Ingreso, zona (cercanía a la zona metropolitana), Raza y embarazo.
3. Para mejorar el rendimiento del modelo y evitar el sobreajuste hacia alguna de las clases las clases dicotómicas de la dimensión, se realizó un balance de clases previo al entrenamiento del modelo

El árbol de decisión se entrenó con los siguientes parámetros:

- **Entrenamiento y prueba:** entrenamiento de un 70% y un set de pruebas de 30%
- **Criterio de selección (criterion):** Como criterio de selección se utilizó Gini. El índice de Gini es una métrica que cuantifica la impureza de un nodo en un árbol de decisión. Está representado por un valor numérico que se muestra en cada nodo. Un índice de Gini de 0,5 significa una distribución uniforme de clases dentro del nodo, mientras que valores más bajos indican un mayor nivel de pureza, lo que significa que predomina una clase
- **Divisor (splitter):** Se utilizó estrategia de división aleatoria de nodos en subnodos

- **Máxima profundidad (max_depth):** Se forzó una profundidad máxima de 4 nodos
- **División mínima de muestras (min_samples_split):** se fijó en 30 el número mínimo de muestras que debe tener un nodo para que pueda dividirse.
- **Compensación de balances (class_weight):** se balancearon las categorías dicotómicas de la dimensión
- **Estado aleatorio (random_state):** es una semilla para generar números aleatorios reproducibles, En este caso se usó 10.000

El reporte de clasificación arrojó los siguientes resultados:

```

La precisión del bosque aleatorio es de 0.622
      precision    recall  f1-score   support

0.0      0.63      0.61      0.62     449773
1.0      0.62      0.63      0.63     449242

 accuracy          0.62     899015
 macro avg         0.62      0.62      0.62     899015
weighted avg         0.62      0.62      0.62     899015

```

El modelo tiene una precisión general de un 62%

La precisión de la Clase 0, que puede considerarse la categoría de menor riesgo, es de 0,63. Esto significa que el 63% de los casos clasificados como "de bajo riesgo" están efectivamente clasificados como tales con precisión. La recuperación (sensibilidad) de la Clase 0 es 0,61, lo que indica que el modelo detectó con precisión el 61 % de los casos de "bajo riesgo" dentro del conjunto de datos. Al evaluar la clase 0, el F1 score alcanzo un 0,62, lo que representa una medida compuesta que abarca tanto la precisión como la recuperación. Un F1 score mayor significa un mejor equilibrio entre precisión y recuperación.

Las métricas para la Clase 1, asociada con "alto riesgo", exhiben una sorprendente similitud, con una precisión, recuperación y F1 Score de 0,62, 0,63 y 0,63 respectivamente.

El número de instancias que admiten la clase 0 es 449.773, mientras que el soporte para la clase 1 es 449.242, lo que sugiere una distribución relativamente equilibrada de datos entre las dos clases.

- Para mejorar el rendimiento del modelo y evitar el sobreajuste hacia alguna de las clases las clases dicotómicas de la dimensión, se realizó un balance de clases previo al entrenamiento del modelo

El árbol de decisión se entrenó con los siguientes parámetros:

- **Entrenamiento y prueba:** entrenamiento de un 70% y un set de pruebas de 30%
- **Criterio de selección (criterion):** Como criterio de selección se utilizó Gini
- **Divisor (splitter):** Se utilizó estrategia de división aleatoria de nodos en subnodos
- **Máxima profundidad (max_depth):** Se forzó una profundidad máxima de 4 nodos
- **División mínima de muestras (min_samples_split):** se fijó en 50 el número mínimo de muestras que debe tener un nodo para que pueda dividirse.
- **Compensación de balances (class_weight):** se balancearon las categorías dicotómicas de la dimensión
- **Estado aleatorio (random_state):** es una semilla para generar números aleatorios reproducibles, haciendo reproducibles los resultados. Se uso 5000 como semilla

El reporte de clasificación arrojó los siguientes resultados:

```

La precisión del bosque aleatorio es de 0.658
      precision    recall  f1-score   support

 0.0         0.66     0.67     0.66     437957
 1.0         0.66     0.65     0.66     438079

 accuracy                   0.66     876036
 macro avg                 0.66     0.66     876036
 weighted avg              0.66     0.66     876036

```

Con una tasa de precisión del 66%, la Clase 0 predice correctamente que el 66% de las instancias. La tasa de recuperación es del 67 %, lo que significa que el modelo detectó con precisión el 67 % de las instancias reales que pertenecen esta clase dentro

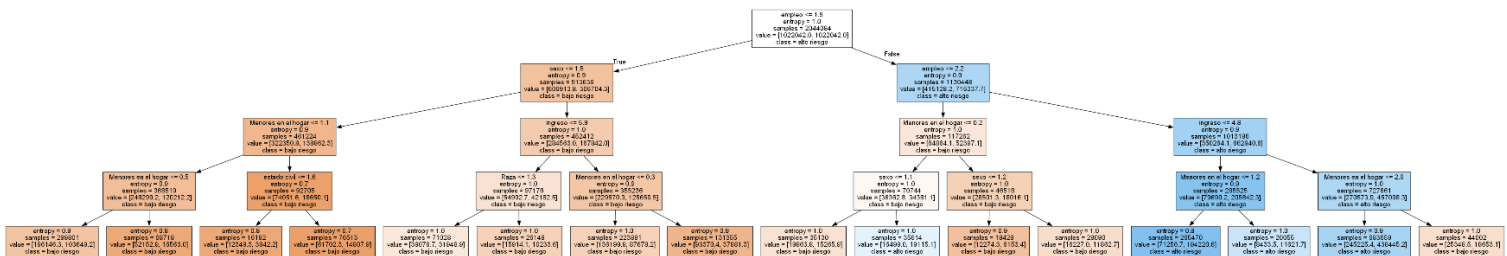
del conjunto de datos. El F1 Score es del 66%, lo que representa la combinación armoniosa de precisión y recuperación.

Con una precisión del 66 %, una retirada del 65 % y un F1 Score del 66 %, la Clase 1 demuestra métricas notablemente similares.

En el conjunto de datos de prueba, hay un total de 437.957 instancias que representan la clase 0, mientras que la clase 1 es compatible con 438.079 instancias.

Para cada métrica de precisión, recuperación y F1 Score, el promedio macro es del 66%. El proceso implica calcular estas métricas individualmente para cada clase y luego determinar el promedio, dando el mismo peso a todas las clases.

El rendimiento del modelo se mantiene estable tanto para la clase 0 como para la clase 1, aunque con una ligera ventaja en la recuperación para la clase 0. Estas medidas indican que el modelo mantiene un justo equilibrio en su capacidad para detectar ambas clases. El F1 Score confirma un equilibrio favorable entre precisión y recuperación.



La división inicial se produce evaluando la condición de empleo ≤ 1.9 . Luego, el conjunto de datos se bifurca en dos caminos separados según el resultado de esta evaluación. Cuando se está desempleado (lado derecho), el ingreso y la cantidad de menores en el hogar, así como el sexo. Un ingreso bajo, mayor cantidad de menores en el hogar y ser hombre se relaciona con un alto riesgo de padecer multimorbididades.

En el lado izquierdo, las variables que afectan al modelo son sexo, el ingreso, los menores en el hogar la raza y el estado civil; sin embargo, tener empleo se relaciona con un bajo riesgo a tener morbilidades; aunque las otras variables mencionadas también se relacionan para afectar el modelo.

5.1.5. Modelo de Riesgo de adicción al tabaco.

En cuanto al modelo de adicción al tabaco:

1. La dimensión “Riesgo de adicción al tabaco”, que fue recategorizada en una variable dicotómica de alto y bajo riesgo.
2. Las variables de sexo, estado civil, nivel educacional, propiedad del hogar, empleo, menores en el hogar, Ingreso, zona (cercanía a la zona metropolitana), Raza y embarazo.
3. Para mejorar el rendimiento del modelo y evitar el sobreajuste hacia alguna de las clases las clases dicotómicas de la dimensión, se realizó un balance de clases previo al entrenamiento del modelo.

El árbol de decisión se entrenó con los siguientes parámetros:

- **Entrenamiento y prueba:** entrenamiento de un 70% y un set de pruebas de 30%
- **Criterio de selección (criterion):** Como criterio de selección se utilizó la entropía
- **Divisor (splitter):** Se utilizó estrategia de división aleatoria de nodos en subnodos
- **Máxima profundidad (max_depth):** Se forzó una profundidad máxima de 3 nodos
- **División mínima de muestras (min_samples_split):** se fijó en 100 el número mínimo de muestras que debe tener un nodo para que pueda dividirse.
- **Compensación de balances (class_weight):** se balancearon las categorías dicotómicas de la dimensión
- **Estado aleatorio (random_state):** es una semilla para generar números aleatorios reproducibles, haciendo reproducibles los resultados. Se uso 1000 como semilla

El reporte de clasificación arrojó los siguientes resultados:

```

La precisión del bosque aleatorio es de 0.593
      precision    recall  f1-score   support

    0.0         0.59      0.64      0.61      377224
    1.0         0.60      0.55      0.57      377844

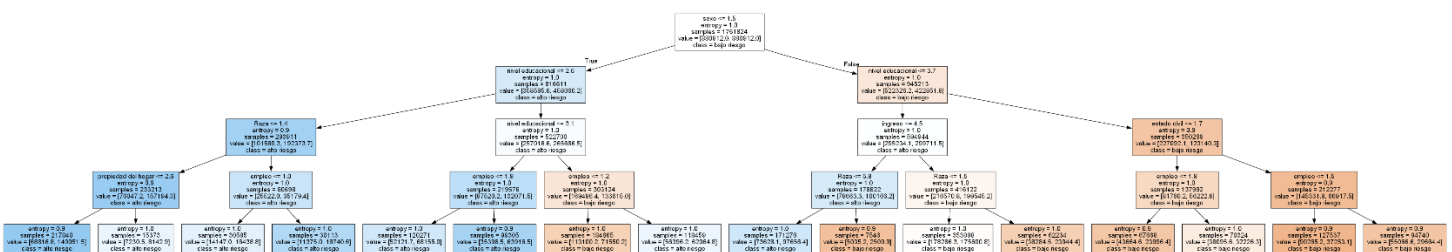
 accuracy                   0.59      755068
 macro avg                 0.59      0.59      0.59      755068
 weighted avg              0.59      0.59      0.59      755068

```

En la clase 0 (bajo riesgo) La precisión del modelo es del 59%, lo que indica que predice correctamente la clase 0 en una proporción significativa de los casos. En el conjunto de datos de prueba, el modelo reconoció con precisión el 64% de las instancias reales que pertenecen a la clase 0. El F1 Score de la clase 0 es 0,61, lo que sugiere un equilibrio satisfactorio entre precisión y recuperación dentro de esta clase en particular. En el conjunto de datos de prueba, hay un total de 377.224 instancias reales que pertenecen a esta clase.

En la Clase 1 (alto riesgo), La precisión del modelo es del 60%, lo que indica que el 60% de sus predicciones son precisas. Tras la revisión, se descubrió que el modelo reconocía con precisión el 55 % de las ocurrencias reales de clase 1 dentro del conjunto de datos de prueba. El F1 Score para la clase 1 es 0,57, lo que indica una eficacia ligeramente inferior a la hora de equilibrar la precisión y la recuperación en comparación con la clase 0, como se observa en el rendimiento del modelo. El conjunto de datos de prueba contiene una cantidad significativa de instancias reales de clase 1, por un total de 377.844.

La precisión, la recuperación y el F1 Score tienen un valor promedio de 0,59 en el promedio macro. Este tipo de promedio no considera el desequilibrio en el número de instancias entre clases y las trata por igual. Tanto el promedio ponderado como el promedio macro tienen un valor de 0,59, ya que son idénticos. Esto se debe a que el apoyo para cada clase es casi igual, lo que da como resultado que el promedio ponderado y el promedio macro se alineen perfectamente.



El riesgo a la adicción al tabaco se ve afectado principalmente por el sexo; del lado izquierdo las variables que afectan al modelo son el nivel educacional, la raza y el empleo; de forma que tal que ser hombre, tener un nivel educacional bajo, ser blanco y carecer de empleo se relaciona con un alto riesgo de ser adicto al tabaco.

Por el lado derecho del árbol, las variables que afectan al modelo son el nivel educacional el ingreso el estado civil, la raza y el empleo; de forma tal que ser mujer, tener un nivel

educacional alto, tener empleo y pertenecer a una raza que no sea la blanca se relaciona con un bajo riesgo de ser adicto al trabajo

5.1.6. Modelo de Riesgo de adicción al alcohol.

En cuanto al modelo de adicción al alcohol:

1. La dimensión “Riesgo de adicción al alcohol”, que fue recategorizada en una variable dicotómica de alto y bajo riesgo.
2. Las variables de sexo, estado civil, nivel educacional, propiedad del hogar, empleo, menores en el hogar, Ingreso, zona (cercanía a la zona metropolitana), Raza y embarazo.
3. Para mejorar el rendimiento del modelo y evitar el sobreajuste hacia alguna de las clases las clases dicotómicas de la dimensión, se realizó un balance de clases previo al entrenamiento del modelo.

El árbol de decisión se entrenó con los siguientes parámetros:

- **Entrenamiento y prueba:** entrenamiento de un 70% y un set de pruebas de 30%
- **Criterio de selección (criterion):** Como criterio de selección se utilizó la entropía
- **Divisor (splitter):** Se utiliza la mejor (best) estrategia de división de nodos en subnodos.
- **Máxima profundidad (max_depth):** Se forzó una profundidad máxima de 3 nodos
- **División mínima de muestras (min_samples_split):** se fijó en 30 el número mínimo de muestras que debe tener un nodo para que pueda dividirse.
- **Compensación de balances (class_weight):** se balancearon las categorías dicotómicas de la dimensión.
- **Estado aleatorio (random_state):** es una semilla para generar números aleatorios reproducibles, haciendo reproducibles los resultados. Se uso 1000 como semilla

El reporte de clasificación arrojó los siguientes resultados:

```
La precisión del bosque aleatorio es de 0.616
      precision    recall  f1-score   support

 0.0         0.62     0.61     0.61     298469
 1.0         0.61     0.62     0.62     297653

 accuracy                   0.62     596122
 macro avg              0.62     0.62     0.62     596122
 weighted avg           0.62     0.62     0.62     596122
```

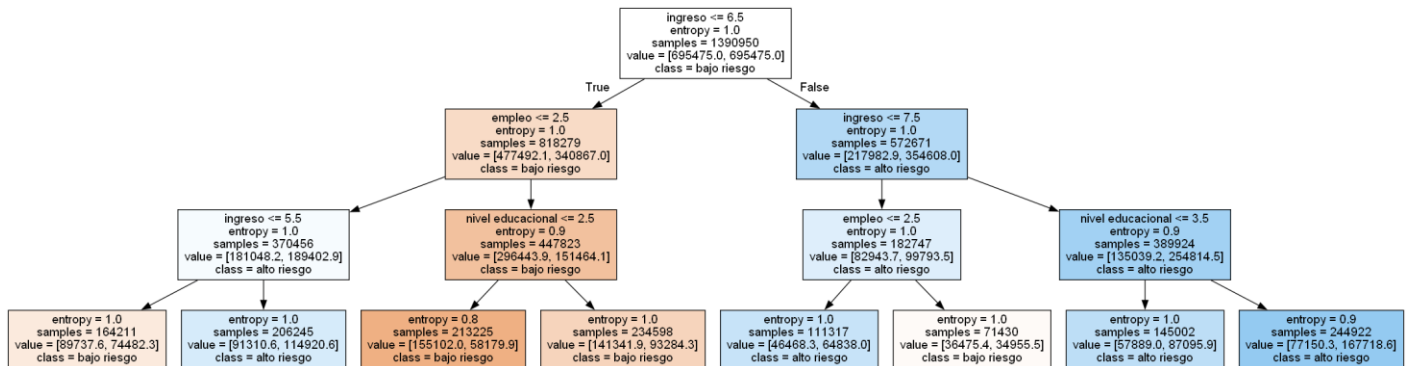
Para la Clase 0, la precisión es de 0,62, lo que indica que el 62% de las instancias clasificadas son efectivamente de esa clase. La tasa de recuperación es de 0,61, lo que indica que el 61% de todos los casos reales se identificaron con precisión. El F1Score arroja una puntuación de 0,61, lo que logra un equilibrio entre estas dos medidas.

Para la clase 1, la precisión es de 0,61, lo que significa que el 61% de los casos clasificados de la clase se identifican con precisión. La tasa de recuperación es 0,62, lo que indica que el modelo captura con éxito el 62% de todas las instancias reales.

El valor de soporte revela las ocurrencias reales de cada clase dentro del conjunto de datos de prueba, con 298.469 instancias para la clase 0 y 297.653 instancias para la clase 1, lo que indica una distribución relativamente equilibrada del tamaño de las clases.

El promedio macro, que representa el promedio no ponderado de las métricas para cada clase, es 0,62 para precisión, recuperación y F1 Score. Esto indica que el modelo demuestra un rendimiento general bien equilibrado en ambas clases. Los promedios ponderados de precisión, recuperación y F1 Score se sitúan en 0,62, lo que se alinea con el promedio macro e indica un rendimiento equilibrado del modelo en ambas clases, lo que implica tamaños de clase similares.

El rendimiento del modelo al clasificar ambas clases es moderado, con una ligera disminución en la precisión para la clase 1 en comparación con la clase 0. Las métricas demuestran un equilibrio satisfactorio entre precisión y recuperación para ambas clases.



En el nodo raíz, el conjunto de datos se divide por la variable ingresos $\leq 6,5$ (entre 35.000 y 50.000 dólares anuales). Cuando los ingresos son menores, el siguiente condicionante es tener un empleo. En el caso de que la condición no se cumpla, un ingreso y el nivel educacional son las variables que afectan al modelo.

En el lado derecho, las variables son el empleo y el nivel educacional, de modo que no tener empleo o tener un nivel educacional bajo se traduciría en un alto riesgo de ser adicto al alcohol

5.1.7. Modelo de estado de salud general.

Finalmente, Se genero un último árbol de decisión, a partir de la suma de las 6 dimensiones creadas en el punto anterior (Riesgo de tener un mal estado de salud físico y mental, Riesgo de padecer diabetes, acceso a cuidados de salud, Riesgo de tener multimorbididades, en este caso no se consideraron las adicciones como parte de un estado de salud general); esta nueva dimensión está destinada a interpretar el riesgo de padecer un mal estado de salud general.

Tal como en los modelos anteriores, el modelo de estado de salud general incorporó las variables de sexo, estado civil, nivel educacional, propiedad del hogar, empleo, menores en el hogar, Ingreso, zona (cercanía a la zona metropolitana), Raza y embarazo, se balancearon las clases, para luego ser entrenado con los siguientes parámetros:

- **Entrenamiento y prueba:** entrenamiento de un 70% y un set de pruebas de 30%
- **Criterio de selección (criterion):** Como criterio de selección se utilizó Gini

- **Divisor (splitter):** Se utiliza la mejor (best) estrategia de división de nodos en subnodos
- **Máxima profundidad (max_depth):** Se forzó una profundidad máxima de 3 nodos
- **División mínima de muestras (min_samples_split):** se fijó en 30 el número mínimo de muestras que debe tener un nodo para que pueda dividirse.
- **Compensación de balances (class_weight):** se balancearon las categorías dicotómicas de la dimensión
- **Estado aleatorio (random_state):** es una semilla para generar números aleatorios reproducibles, haciendo reproducibles los resultados. Se uso 1000 como semilla

El reporte de clasificación arrojó los siguientes resultados:

```

La precisión del árbol de decisión es de 0.695
      precision    recall  f1-score   support

 0.0         0.72     0.64     0.68     589938
 1.0         0.68     0.75     0.71     589830

 accuracy                   0.70  1179768
 macro avg                   0.70   0.70   0.69  1179768
 weighted avg                 0.70   0.70   0.69  1179768

```

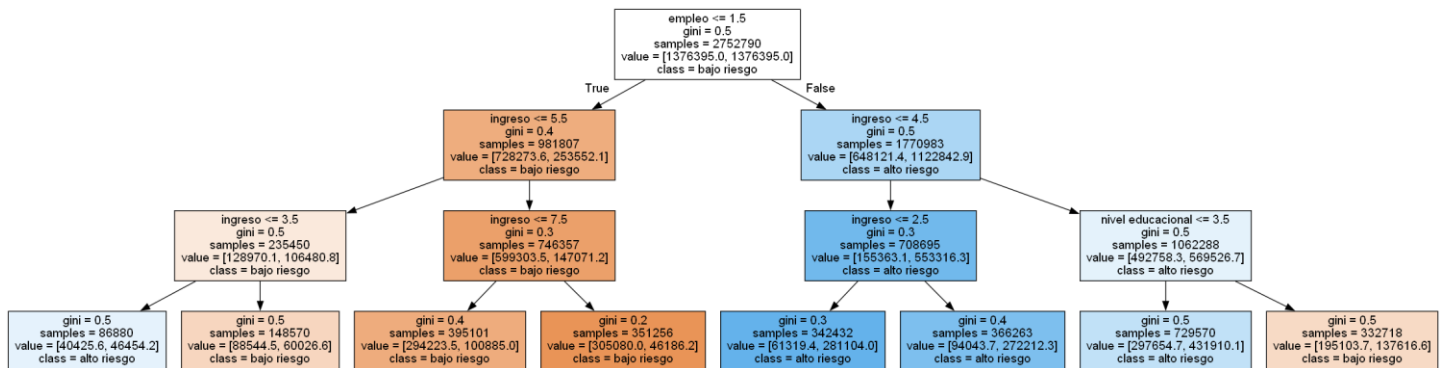
La clase 0 ("bajo riesgo") obtiene una precisión de 0,72, se puede inferir que aproximadamente el 72% de las instancias que se predijeron como clase 0 se clasifican correctamente. El recall es 0,64, indica que el modelo reconoció con precisión el 64% del total de ocurrencias de esta clase. El F1 Score alcanzó un 0,68, lo cual indica un equilibrio favorable entre precisión y recuperación, ya que esta medida combina ambos aspectos. El conjunto de datos de prueba contiene un total de 589.938 instancias de clase 0, lo que respalda firmemente este número.

La Clase 1 ("alto riesgo") alcanza una precisión de 0,68, se puede inferir que aproximadamente el 68% de los casos clasificados como clase 1 han sido identificados con precisión. Con base en el valor de recuperación de 0,75, se puede inferir que el modelo capturó con éxito el 75% de todas las instancias reales pertenecientes a la clase. El F1 Score de 0,71 indica un rendimiento ligeramente mejorado a la hora de lograr un equilibrio entre precisión y recuperación en comparación con la clase 0. El conjunto de datos de prueba contiene un total de

589.830 instancias de clase 1, lo que proporciona un fuerte respaldo para este valor en particular.

La precisión, la sensibilidad y la puntuación media de F1 se sitúan en 0,70 en el Macro Promedio. Estos promedios no ponderados dan la misma consideración a ambas clases, independientemente del número de casos dentro de cada clase. La ponderación del F1-Score, aunque ligeramente inferior a 0,69, se alinea con los promedios ponderados de precisión y recuperación, que se sitúan en 0,70. Estos promedios ponderados consideran la frecuencia de instancias para cada clase.

El rendimiento del modelo de árbol de decisión es adecuado. El modelo demuestra una precisión y recuperación equilibradas en ambas clases, lo que indica su eficacia en el manejo de instancias categorizadas como "bajo riesgo" y "alto riesgo". La métrica F1-Score es particularmente valiosa en situaciones en las que lograr un equilibrio entre precisión y recuperación es crucial. Dado que el F1-Score es alto para ambas clases, es evidente que el modelo sobresale tanto en precisión como en la identificación de casos positivos.



En la raíz del árbol, el conjunto de datos se divide a partir del empleo, hacia la derecha, no tener empleo, el ingreso en segunda instancia y el nivel educativo, de forma tal que no tener empleo, pero tener un ingreso mayor a 22.500 dólares anuales y un nivel de estudios alto disminuye el riesgo, mientras que un nivel de estudio bajo lo aumenta. Si el ingreso es menor a esa cifra, el riesgo aumenta.

En el lado izquierda del árbol, la salud se ve afectada por el hecho de tener empleo, en segunda instancia por el nivel de ingreso que se pueda poseer, un ingreso mayor a

22.500 dólares afecta positivamente al estado de salud, sin embargo, un ingreso menor le afectan negativamente.

5.1.8. Bosques aleatorios.

Como último método de clasificación y con el fin de identificar con mayor precisión cuales son los determinantes sociales que afectan al estado de salud y con que nivel de importancia lo hacen, se desarrollo un modelo de bosques aleatorios; un algoritmo eficaz para tareas de clasificación y regresión de los métodos de aprendizaje supervisado conocido como aprendizaje conjunto. Utiliza una colección de árboles de decisión que se entrenan mediante la técnica de "ensacado". El concepto fundamental detrás del ensacado es fusionar múltiples modelos, específicamente árboles de decisión, para construir un modelo combinado más sólido y preciso. Para mejorar la solidez del modelo y mitigar el riesgo de sobreajuste, se promediaron las predicciones de múltiples árboles individuales, reduciendo así la varianza del modelo, un problema común observado en los árboles de decisión independientes. También se balancearon las clases, de modo de evitar el sobreajuste del modelo a partir de la preponderancia de casos en una de ellas

El bosque aleatorio fue entrenado con los siguientes parámetros

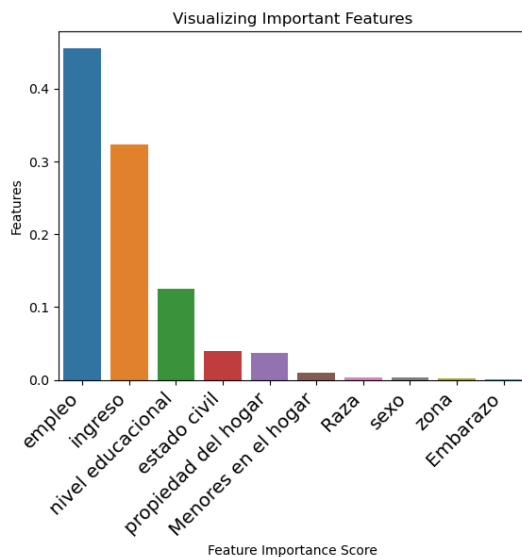
- **Entrenamiento y prueba:** entrenamiento de un 70% y un set de pruebas de 30%
- **Criterio de selección (criterion):** Como criterio de selección se utilizó Gini
- **Divisor (splitter):** Se utiliza la mejor (best) estrategia de división de nodos en subnodos
- **Máxima profundidad (max_depth):** Se forzó una profundidad máxima de 3 nodos
- **División mínima de muestras (min_samples_split):** se fijó en 10 el número mínimo de muestras que debe tener un nodo para que pueda dividirse.
- **Compensación de balances (class_weight):** se balancearon las categorías dicotómicas de la dimensión
- **Estado aleatorio (random_state):** es una semilla para generar números aleatorios reproducibles, haciendo reproducibles los resultados. Se uso 10.000 como semilla

Con una tasa de precisión del 70,1 %, el modelo de bosque aleatorio demuestra su capacidad para clasificar correctamente una mayoría significativa, alrededor del 70,1 %, de las instancias.

A continuación, la métrica de importancia de las características cuantifica el grado en que cada característica individual contribuye a la capacidad predictiva del modelo.

La precisión del bosque aleatorio es de 0.701

	feature	importance
4	empleo	0.455385
6	ingreso	0.322962
2	nivel educacional	0.124800
1	estado civil	0.040091
3	propiedad del hogar	0.037253
5	Menores en el hogar	0.010266
8	Raza	0.003979
0	sexo	0.002885
7	zona	0.001627
9	Embarazo	0.000751



La característica que tiene mayor importancia en las predicciones del modelo es el empleo, con una puntuación de importancia de 45,6%. Un valor más alto para esta característica significa que cualquier modificación que se le realice influirá en gran medida en la clasificación de la salida del modelo. Con un valor de 32,3% los ingresos son la segunda característica más importante, lo que significa su impacto significativo en el rendimiento del modelo. El nivel educativo en las decisiones del modelo es evidente con una puntuación de importancia de 12,48%, lo que destaca su contribución sustancial. Al considerar el impacto en las predicciones, es evidente que el estado civil, tienen importancias relativamente menores de 4% y un 3,7% respectivamente, lo que indica que influyen en las predicciones, pero en menor grado en comparación con el empleo y características del ingreso. Los atributos restantes, incluidos Niños en el hogar, Raza, Sexo, Área y Embarazo, tienen significativamente menos importancia, lo que sugiere que tienen una importancia mínima. impacto en la clasificación del modelo.

6. Conclusiones

La relación entre los determinantes sociales y la salud ha sido un foco central en la investigación de la salud pública durante décadas. La hipótesis de que factores como el empleo, el ingreso y la educación influyen significativamente la salud de los individuos está bien fundamentada en una amplia gama de estudios epidemiológicos y sociosanitarios. Este análisis amplía la comprensión de esta dinámica mediante la aplicación de un modelo de bosque aleatorio para evaluar la prevalencia de enfermedades crónicas y comportamientos de riesgo en la población adulta de Estados Unidos, utilizando datos del BRFSS de 2017 a 2021.

La precisión del modelo de bosque aleatorio, situada en un 70.1%, subraya su capacidad para clasificar de manera efectiva las instancias dentro del conjunto de datos. La robustez del modelo se refuerza aún más por la consistencia en las métricas de precisión y recall para las clases predichas. El análisis de la importancia de las características indica que el empleo y el ingreso están entre los predictores más influyentes, seguidos por el nivel educacional. Estos hallazgos están alineados con la literatura existente, que consistentemente destaca la influencia de la situación laboral y los recursos económicos en las opciones de estilo de vida y el acceso a la atención médica, ambos críticos para la prevención y manejo de enfermedades crónicas.

La sensibilidad del modelo para la clase asociada con un "alto riesgo" de enfermedades crónicas y comportamientos de riesgo es especialmente relevante. Un recall de 0.75

sugiere que el modelo es particularmente competente en la identificación de individuos en riesgo, lo cual es crucial para el diseño de intervenciones dirigidas y políticas de salud pública. Del mismo modo, la precisión de la clase "bajo riesgo" confirma la capacidad del modelo para discriminar adecuadamente entre perfiles de riesgo, lo que es vital para la asignación eficiente de recursos y la personalización de estrategias de salud preventiva.

Al integrar estos resultados con el entendimiento actual de los determinantes sociales de la salud, podemos afirmar que los datos proporcionan una validación sólida de la hipótesis propuesta. No solo demuestran una asociación estadística significativa entre los determinantes sociales y la salud, sino que también ofrecen insights cuantitativos sobre la magnitud de su impacto. Las implicaciones de estos hallazgos son de gran alcance, proporcionando una base empírica para la formulación de políticas que aborden las desigualdades sociales como una ruta hacia la mejora de la salud poblacional.

Es importante reconocer las limitaciones inherentes a los estudios observacionales, incluyendo la capacidad de inferir causalidad. Sin embargo, la convergencia de los hallazgos del modelo con el cuerpo de evidencia existente refuerza la validez de la hipótesis. A pesar de estas limitaciones, el modelo proporciona un marco predictivo valioso que puede ser utilizado para identificar poblaciones en riesgo y servir como una herramienta en la toma de decisiones informada para los actores de la salud pública.

En conclusión, los determinantes sociales son predictores significativos de la salud y deben ser considerados elementos clave en el desarrollo de políticas y programas de salud pública. La validación de esta hipótesis a través del análisis de datos del BRFSS refuerza la necesidad de una acción continua y centrada para abordar las barreras sociales que impiden una vida más saludable para todos los ciudadanos.

Si bien este estudio se realiza con los resultados del BRFSS, sería relevante a futuro, adecuar este proyecto a la realidad en Chile, utilizando la encuesta CACEN. A través de ella podemos obtener una comprensión integral de las circunstancias que enfrentan los hogares y las personas, particularmente aquellos que viven en la pobreza y los grupos específicos identificados como poblaciones prioritarias por las políticas sociales. Esta encuesta proporciona información detallada sobre diversos aspectos como demografía, salud, educación, vivienda, empleo e ingresos, entre otros factores. De forma tal de desarrollar un modelo de clasificación que permita priorizar las políticas públicas en salud o priorizar las atenciones de la lista de espera adaptado a la realidad chilena.

7. Bibliografía

<https://1drv.ms/f/s!Ak40qC3DPkNOmulwabfzXgjY4710Jw?e=ZDzJrW>

1. Barrera Sánchez, O. (2011). El cuerpo en Marx, Bourdieu y Foucault. *Revista de Ciencias Sociales de la Universidad Iberoamericana*, 6(11), 121-137.
2. Blas, E., Sommerfeld, J., & Kurup, A. S. (2011). Social determinants approaches to public health: from concept to practice. World Health Organization.
3. Buss, A. M., & Pellegrini Filho, A. (2007). A Saúde e seus Determinantes Sociais. *Revista Saúde Coletiva*, Rio de Janeiro, 17, 77-93.
4. Clark, C. R., Ommerborn, M. J., Moran, K., Brooks, K., Haas, J., Bates, D. W., & Wright, A. (Año). Predicting Self-Rated Health Across the Life Course: Health Equity Insights from Machine Learning Models. *J Gen Intern Med*. 2021 May;36(5):1181-1188. doi: 10.1007/s11606-020-06438-1. Epub 2021 Feb 23. PMID: 33620624; PMCID: PMC8131482.
5. De la Torre-Ugarte-Guanilo, M., & Oyola-García, A. (2014). Los determinantes sociales de la salud: una propuesta de variables y marcadores/indicadores para su medición. *Revista Peruana de Epidemiología*, 18(1), 1-6. Sociedad Peruana de Epidemiología.
6. Estrada-Mesa, D. A., & Cardona-Arias, J. A. (2018). La medicina en la obra de Michel Foucault: Meta-síntesis. *Civilizar Ciencias Sociales y Humanas*, 18, 223-236.
7. Javier Mora Pineda (2022). *Revista médica Clínica Las Condes: Modelos predictivos en salud basados en aprendizaje de maquina (Machine learning)*
8. Mejía O, L. M. (2013). Los Determinantes Sociales de la Salud: base teórica de la salud pública. *Revista Facultad Nacional de Salud Pública*, 31, 28-36. Universidad de Antioquia. Medellín, Colombia.
9. Organización Mundial de la Salud. (2009). Informe de la Secretaría: Comisión sobre Determinantes Sociales de la Salud.
10. Organización Panamericana de la Salud. (2017). Documentos Básicos de la Organización Panamericana de la Salud, Decimonovena edición. Recuperado de <https://www.paho.org/es/documentos/documentos-basicos-organizacion-panamericana-salud-decimonovena-edicion>
11. Prah Ruger, J. (2004). Health and social justice. *The Lancet*, 364(9439), 1075-1080. [https://doi.org/10.1016/S0140-6736\(04\)17056-6](https://doi.org/10.1016/S0140-6736(04)17056-6)
12. Rivera de Ramones, E. M. (2019). Estilos de vida saludable: práctica social de hábitos alimentarios. *Revista Digital de Postgrado*, 8(2).
13. Segura Cardona, Angela María. "Aplicación de árboles de decisión en la salud pública". *Revista CES Salud Pública*, vol. 3, no. 1, 2012, pp. 94-103.
14. The World Bank. (n.d.). Dying for Change: Poor people's experience of health and ill-health.

15. Villar, E. (2007). Los Determinantes Sociales de Salud y la lucha por la equidad en Salud: desafíos para el Estado y la sociedad civil. *Saúde e Sociedade*, 16(3), 7-13.
16. Wilkinson, L.L.; Long-Daniels, A.; Appah, M.; Zhai, Y.; Watson, D.M.; Walker, K.; Young-Bilbo, K.; Aboagye, A.; Tucker, C.; McCarthy, S. The Association between Social Determinants of Health and Depressive Disorders: A 2017 Behavioral Risk Factor Surveillance System (BRFSS) Analysis. *Psychiatry Int.* 2023, 4, 147-159. <https://doi.org/10.3390/psychiatryint4020017>

8. Anexos.

8.1. Contenido del conjunto de datos

COLUMNA	TIPO DATO	DESCRIPCION
_STATE	Record Identification	State FIPS Code
IDAY	Record Identification	Interview Year
IMONTH	Record Identification	Interview Year
IYEAR	Record Identification	Interview Year
DISPCODE	Record Identification	Final Disposition (1100 - COMPLETED INTERVIEW; 1200 - PARTIAL INTERVIEW)
HHADULT	Cell Phone Introduction	How many members of your household, including yourself, are 18 years of age or older?
GENHLTH	Healthy Days — Health Related Quality of Life	Would you say that in general your health is:
PHYSHLTH	Healthy Days — Health Related Quality of Life	Now thinking about your physical health, which includes physical illness and injury, for how many days during the past 30 days was your physical health not good?
MENTHLTH	Healthy Days — Health Related Quality of Life	Now thinking about your mental health, which includes stress, depression, and problems with emotions, for how many days during the past 30 days was your mental health not good?
POORHLTH	Healthy Days — Health Related Quality of Life	During the past 30 days, for about how many days did poor physical or mental health keep you from doing your usual activities, such as self-care, work, or recreation?
HLTHPLN1	Health Care Access	Do you have any kind of health care coverage, including health insurance, prepaid plans such as HMOs, or government plans such as Medicare, or Indian Health Service?
PERSDOC2	Health Care Access	Do you have one person you think of as your personal doctor or health care provider? (If 'No' ask 'Is there more than one or is there no person who you think of as your personal doctor or health care provider?'.)
MEDCOST	Health Care Access	Was there a time in the past 12 months when you needed to see a doctor but could not because of cost?
CHECKUP1	Health Care Access	About how long has it been since you last visited a doctor for a routine checkup? [A routine checkup is a general physical exam, not an exam for a specific injury, illness, or condition.]
CVDINFR4	Chronic Health Conditions	(Ever told) you had a heart attack, also called a myocardial infarction?
CVDCRHD4	Chronic Health Conditions	(Ever told) you had angina or coronary heart disease?
CVDSTRK3	Chronic Health Conditions	(Ever told) you had a stroke
ASTHMA3	Chronic Health Conditions	(Ever told) you had asthma?
ASTHNOW	Chronic Health Conditions	Do you still have asthma?
CHCSCNCR	Chronic Health Conditions	(Ever told) you had skin cancer?
CHCOCNCR	Chronic Health Conditions	(Ever told) you had any other types of cancer?
CHCCOPD1	Chronic Health Conditions	(Ever told) you have chronic obstructive pulmonary disease, C.O.P.D., emphysema or chronic bronchitis?
HAVARTH3	Chronic Health Conditions	(Ever told) you have some form of arthritis, rheumatoid arthritis, gout, lupus, or fibromyalgia? (Arthritis diagnoses include: rheumatism, polymyalgia rheumatica; osteoarthritis (not osteoporosis); tendonitis, bursitis, bunion, tennis elbow; carpal tunnel syndrome, tarsal tunnel syndrome; joint infection, etc.)
CHCKIDNY	Chronic Health Conditions	(Ever told) you have kidney disease? (Do NOT include kidney stones, bladder infection or incontinence.)
DIABETE3	Chronic Health Conditions	(Ever told) you have diabetes (If 'Yes' and respondent is female, ask 'Was this only when you were pregnant?'. If Respondent says pre-diabetes or borderline diabetes, use response code 4.)
DIABAGE2	Chronic Health Conditions	How old were you when you were told you have diabetes?
SEX	Demographics	Indicate sex of respondent
MARITAL	Demographics	Are you: (marital status)
EDUCA	Demographics	What is the highest grade or year of school you completed?

RENTHOM1	Demographics	Do you own or rent your home?
EMPLOY1	Demographics	Are you currently...?
CHILDREN	Demographics	How many children less than 18 years of age live in your household?
INCOME2	Demographics	Is your annual household income from all sources: (If respondent refuses at any income level, code Refused.)
WEIGHT2	Demographics	About how much do you weigh without shoes? (If respondent answers in metrics, put a 9 in the first column)[Round fractions up.]
HEIGHT3	Demographics	About how tall are you without shoes? (If respondent answers in metrics, put a 9 in the first column)[Round fractions down.]
PREGNANT	Demographics	To your knowledge, are you now pregnant?
DEAF	Demographics	Are you deaf or do you have serious difficulty hearing?
BLIND	Demographics	Are you blind or do you have serious difficulty seeing, even when wearing glasses?
DECIDE	Demographics	Because of a physical, mental, or emotional condition, do you have serious difficulty concentrating, remembering, or making decisions?
DIFFWALK	Demographics	Do you have serious difficulty walking or climbing stairs?
DIFFDRES	Demographics	Do you have difficulty dressing or bathing?
DIFFALON	Demographics	Because of a physical, mental, or emotional condition, do you have difficulty doing errands alone such as visiting a doctor's office or shopping?
SMOKE100	Tobacco Use	Have you smoked at least 100 cigarettes in your entire life? [Note: 5 packs = 100 cigarettes]
SMOKDAY2	Tobacco Use	Do you now smoke cigarettes every day, some days, or not at all?
STOPSMK2	Tobacco Use	During the past 12 months, have you stopped smoking for one day or longer because you were trying to quit smoking?
LASTSMK2	Tobacco Use	How long has it been since you last smoked a cigarette, even one or two puffs?
USENOW3	Tobacco Use	Do you currently use chewing tobacco, snuff, or snus every day, some days, or not at all? (Snus (Swedish for snuff) is a moist smokeless tobacco, usually sold in small pouches that are placed under the lip against the gum.)[Snus (rhymes with goose)]
ALCDAY5	Alcohol Consumption	During the past 30 days, how many days per week or per month did you have at least one drink of any alcoholic beverage such as beer, wine, a malt beverage or liquor?
AVEDRNK2	Alcohol Consumption	One drink is equivalent to a 12-ounce beer, a 5-ounce glass of wine, or a drink with one shot of liquor. During the past 30 days, on the days when you drank, about how many drinks did you drink on the average? (A 40 ounce beer would count as 3 drinks, or a cocktail drink with 2 shots would count as 2 drinks.)
DRNK3GE5	Alcohol Consumption	Considering all types of alcoholic beverages, how many times during the past 30 days did you have 5 or more drinks for men or 4 or more drinks for women on an occasion?
MAXDRNKS	Alcohol Consumption	During the past 30 days, what is the largest number of drinks you had on any occasion?
_RFHLTH	Calculated Variables	Adults with good or better health
_PHYS14D	Calculated Variables	3 level not good physical health status: 0 days, 1-13 days, 14-30 days
_MENT14D	Calculated Variables	3 level not good mental health status: 0 days, 1-13 days, 14-30 days
_HCVU651	Calculated Variables	Respondents aged 18-64 who have any form of health care coverage
_MICH1	Calculated Variables	Respondents that have ever reported having coronary heart disease (CHD) or myocardial infarction (MI)
_LTASTH1	Calculated Variables	Adults who have ever been told they have asthma
_CASTHM1	Calculated Variables	Adults who have been told they currently have asthma
_ASTHMS1	Calculated Variables	Computed asthma status
_DRDXAR1	Calculated Variables	Respondents who have had a doctor diagnose them as having some form of arthritis
_BMI5CAT	Calculated Variables	Four-categories of Body Mass Index (BMI)
_EDUCAG	Calculated Variables	Level of education completed
_INCOMG	Calculated Variables	Income categories
_SMOKER3	Calculated Variables	Four-level smoker status: Everyday smoker, Someday smoker, Former smoker, Non-smoker
_DRNKWEK	Calculated Variables	Calculated total number of alcoholic beverages consumed per week

_TOTINDA	Calculated Variables	Adults who reported doing physical activity or exercise during the past 30 days other than their regular job
_RFDRHV5	Calculated Variables	Heavy drinkers (adult men having more than 14 drinks per week and adult women having more than 7 drinks per week)
_AGEG5YR	Calculated Variables	Fourteen-level age category
_BMI5	Calculated Variables	Body Mass Index (BMI)
MSCODE	Weighting Variables	Metropolitan Status Code
_PRACE1	Weighting Variables	Preferred race category

8.2. Código desarrollado en Python

<https://1drv.ms/f/s!Ak40qC3DPkNOop0X1N3gl-XWTCIDqA?e=hdNgJK>