

Diseño de una Bodega de Datos en ambientes de Inteligencia De Negocios para el Análisis de Indicadores claves de Enfermedades Cardiacas.

Nombres y apellidos

Elvia Rosa Méndez Escobar

C.C. No. 1004507962

Código estudiantil: 201911214814

Correo institucional: elvia.mendez@unisimon.edu.co

Thalía Carolay Hernández Matías

C.C. No. 1001897011

Código estudiantil: 2019114100231

Correo institucional: Thalia.hernandez@unisimon.edu.co

Deyler José Freile Bolaño

C.C. No. 1001881274

Código estudiantil: 201911213973

Correo institucional: Deyler.freile@unisimon.edu.co

Johnner Enrique Freile Bolaño

C.C. No 1001881273

Código estudiantil: 201911213975

Correo institucional: Jhoner.freile@unisimon.edu.co

Trabajo de Investigación del Programa de ingeniería de sistema e ingeniería industrial.

Tutor(es):

Juan Carlos Calabria Sarmiento

RESUMEN

En la investigación se tomó como punto de partida una base de dato “**heart_2020_cleaned**”, esto se realizó con la finalidad de identificar el número de riesgo cardiovascular al que padecen las personas, por tal motivo se analizaron aspectos que tienen relación directa para que se presenten estas situaciones. Se

tuvieron en cuenta las características como raza, sexo, presencia de enfermedad cardiovascular, edad, masa corporal, actividades sanas y no sanas y antecedentes de enfermedades cardíacas lo anterior es útil porque a través de estos datos se analizaron aspectos que pueden incidir para que se presenten este tipo de complicaciones o enfermedades cardíacas.

Según el informe que presento la revista uninorte la ciudad de barraquilla presenta qué la población objeto de estudio esta entre los 18 y 40 años que están expuesta, lo que indica que tienen probabilidad de padecer complicaciones a nivel cardiovascular, a pesar de que en menor proporción no tienen los factores para que sean predisponentes de riesgo cardiovascular.

Las causas de este problema radican en la extensa cantidad de datos que se generan de esta temática, siendo un poco dificultoso trabajar con estos. Además de la complejidad de los diversos datos, también se puede decir que la falta de un sistema que no solo tome los datos para registrarlos en un formato el cual es de difícil comprensión para un usuario sin el conocimiento del tema, si no la construcción de un sistema que ayude a entender a la persona, y que esta pueda interpretar la información dada y sacar sus propias conclusiones.

Con los datos anteriores, se planteó una propuesta para dar solución a la anterior problemática, diseñando una bodega de datos, hacer un proceso de ETL (Extracción, Transformación y Carga) utilizando DataWereHouse para el almacenamiento de los datos y realizar análisis por medio de herramientas de la inteligencia de negocio, y con ayuda de las norma ISO 25012 se revisará la calidad de los datos por medio de una guía para llegar a una toma de decisión más acertada.

ABSTRACT

In the investigation, a "heart_2020_cleaned" database was taken as a starting point, this was done with the purpose of identifying the number of cardiovascular risk that people suffer, for this reason aspects that are directly related were analyzed so that they are presented these situations. Characteristics such as race, sex, presence of cardiovascular disease, age, body mass, healthy and unhealthy activities, and

history of heart disease were taken into account. The foregoing is useful because through these data, aspects that can affect present this type of complications or heart disease.

According to the report presented by the uninorte magazine, the city of Barranquilla shows that the population under study is between 18 and 40 years of age who are exposed, which indicates that they are likely to suffer complications at the cardiovascular level, despite the fact that to a lesser extent they do not have the factors to be predisposing to cardiovascular risk.

The causes of this problem lie in the extensive amount of data that is generated on this subject, being a bit difficult to work with. In addition to the complexity of the various data, it can also be said that the lack of a system that not only takes the data to record it in a format that is difficult to understand for a user without knowledge of the subject, but also the construction of a system that helps to understand the person, and that he can interpret the information given and draw his own conclusions.

With the previous data, a proposal was proposed to solve the previous problem, designing a data warehouse, doing an ETL (Extraction, Transformation and Loading) process using Data warehouse for data storage and performing analysis through tools of business intelligence, and with the help of Iso standards, the quality of the data will be reviewed to reach a decision by means of statistical graphs.

KeyWords:

DataWereHouse, enfermedades cardiacas, riesgos, base de Datos, inteligencia de Negocio, ETL (Extracción, Transformación y Carga)

REFERENCIAS

1. González SA, Castiblanco MA, Arias-Gómez LF, Martinez-Ospina A, Cohen DD, Holguin GA, et al. Results from Colombia's 2016 Report Card on Physical Activity for Children and Youth. J Phys Act Health. 2016;13(11 Suppl 2): 129-136. <https://doi.org/10.1123/jpah.2016-0369>
2. World Health Organization. Global health risks: mortality and burden of disease attributable to selected major risks. Geneva; 2009. https://www.who.int/healthinfo/global_burden_disease/GlobalHealthRisks_report_full.pdf