UNIVERSIDAD DE SAN CARLOS DE GUATEMALA

FACULTAD DE INGENIERIA

ESCUELA DE ESTUDIOS DE POSTGRADO

MAESTRIA PARA LA INDUSTRIA

CON ESPECIALIZACION EN CIENCIAS DE LA COMPUTACION

CURSO: Minería de Datos.

Por: Denys Fernando Orozco.

PROYECTO PARTE 1 DE MINERIA DE DATOS

ANÁLISIS DE LA SITUACIÓN DE LAS DEFUNCIONES Y DEFUNCIONES FETALES EN GUATEMALA

Contenido

[Introducción: 1](#_Toc182085699)

[Desarrollo: 2](#_Toc182085700)

[Librerías y funciones más importantes utilizadas en el proyecto: 2](#_Toc182085701)

[Arules 2](#_Toc182085702)

[Dyplr 3](#_Toc182085703)

[Ggplot2 3](#_Toc182085704)

[Fim4r 3](#_Toc182085705)

[Algoritmos utilizados durantes el analisis 4](#_Toc182085706)

[APRIORI 4](#_Toc182085707)

[FPGROWTH (Frequent Pattern Growth) 4](#_Toc182085708)

[KMEANS 4](#_Toc182085709)

[Casos de uso para cada algoritmo: 5](#_Toc182085710)

[Conclusiones 5](#_Toc182085711)

[Referencias 6](#_Toc182085712)

# Introducción:

En el país de Guatemala, el registro y análisis de las defunciones y defunciones fetales es parte fundamental que permite comprender una perspectiva de la salud del país, identificado los principales retos que enfrenta el sistema de salud nacional.

Estos datos vitales, recopilados a través del Registro Nacional de las Personas (RENAP) en coordinación con el Ministerio de Salud Pública y Asistencia Social (MSPAS), proporcionan información concluyente sobre las causas de mortalidad, las condiciones de vida de la población y la efectividad de las políticas en salud pública implementadas en el territorio nacional.

Al analizar las causas de muerte y los factores de riesgo asociados, podemos comprender mejor las vulnerabilidades de la población y trabajar en la prevención de enfermedades y la promoción de la salud.

# Desarrollo:

Para el desarrollo del presente análisis, se tomó en cuenta la sugerencia del manual ***“Principios y recomendaciones para un sistema de estadísticas vitales”,*** (ONU, 2014, p. 163)***;*** para la consideración de rango de la cantidad de datos a evaluar.

La información a utilizar durante el desarrollo del análisis de datos, se toma del sitio de descarga de la información pública que corresponde al sitio oficial del Instituto Nacional de Estadística de Guatemala[[1]](#footnote-1).

## Librerías y funciones más importantes utilizadas en el proyecto:

A continuación, se detallan algunos conceptos de las librerías y algunas funciones que se utilizaron en el desarrollo de la minería de datos. Se eligieron las mismas por la importancia que revisten para la ejecución del proyecto.

### Arules

Definición: Paquete especializado en minería de reglas de asociación y conjuntos de ítems frecuentes.

Principales usos:

* Análisis de canasta de mercado
* Descubrimiento de patrones frecuentes
* Generación de reglas de asociación

### Dyplr

Definición: Paquete fundamental para manipulación de datos que proporciona una gramática relacionada para trabajar con data frames.

Principales usos:

* Filtrar observaciones
* Seleccionar variables
* Transformar variables
* Agrupar datos
* Resumir datos

### Ggplot2

Definición: Sistema para crear gráficos basado en la gramática de gráficos.

Principales usos:

* Visualización de datos
* Creación de gráficos estadísticos
* Personalización de visualizaciones

### Fim4r

Definición: Implementación de algoritmos de minería de conjuntos de ítems frecuentes.

Principales usos:

* Descubrir patrones frecuentes en datos transaccionales
* Análisis de secuencias
* Minería de datos temporal

## Algoritmos utilizados durantes el analisis

A continuación, se detallan los algoritmos utilizados para el desarrollo de la minería de los datos. Se tomaron en cuenta tres algoritmos: apriori, fpgrowth y kmeans. Atendiendo a su definición, tenemos:

### APRIORI

Definición: Algoritmo para minería de reglas de asociación que identifica conjuntos de ítems frecuentes y genera reglas de asociación.

Características principales:

* Utiliza el principio "bottom-up"
* Aplica poda basada en soporte mínimo
* Genera reglas basadas en confianza mínima

### FPGROWTH (Frequent Pattern Growth)

Definición: Algoritmo para minería de patrones frecuentes que utiliza una estructura de datos en árbol (FP-tree) para almacenar información de frecuencia de items.

Características principales:

* No genera conjuntos candidatos
* Utiliza estructura de árbol compacta
* Más eficiente que Apriori para grandes conjuntos de datos

### KMEANS

Definición: Algoritmo de agrupamiento que divide n observaciones en k grupos donde cada observación pertenece al grupo con la media más cercana.

Características principales:

* Agrupamiento basado en centros
* Método iterativo
* Minimiza la suma de cuadrados dentro de grupos

### Casos de uso para cada algoritmo:

Cada algoritmo tiene sus propias ventajas y casos de uso específicos:

* Apriori: Mejor para conjuntos de datos pequeños a medianos donde la interpretabilidad es importante
* FP-Growth: Preferible para conjuntos de datos grandes donde la eficiencia es crucial
* K-means: Excelente para segmentación cuando se conoce aproximadamente el número de grupos deseados.

# Conclusiones

# Referencias

Hahsler, M., Grun, B., & Hornik, K. (29 de Septiembre de 2005). *arules - A Computational Environment for Mining Association Rules and Frequent Item Sets.* Obtenido de https://cran.r-project.org: https://cran.r-project.org/web/packages/arules/readme/README.html

Itemsets, M. A. (11 de Noviembre de 2024). *Interface to Mining Algorithms from fim4r*. Obtenido de https://search.r-project.org: https://search.r-project.org/CRAN/refmans/arules/html/fim4r.html

ONU. (2014). *Principios y recomendaciones para un sistema de estadísticas vitales.* Nueva York: ONU.

Wien, W. (11 de Noviembre de 2024). *The Comprehensive R Archive Network*. Obtenido de https://cran.r-project.org/: https://cran.r-project.org/web/packages/dplyr/index.html

Wien, W. (11 de Noviembre de 2024). *The Comprehensive R Archive Network*. Obtenido de https://cran.r-project.org/: https://cran.r-project.org/web/packages/arules/index.html

Wien, W. (11 de Noviembre de 2024). *The Comprehensive R Archive Network*. Obtenido de https://cran.r-project.org/: https://cran.r-project.org/web/packages/ggplot2/index.html

1. https://www.ine.gob.gt/vitales/ [↑](#footnote-ref-1)