



Facultad de Ingeniería en Sistemas, Universidad Fidélitas

SC-602 Data Warehouse y bases de datos

Proyecto Final

Tema:

Migración en Centroamérica

Integrantes:

Gabriel Mora Torres

Ana Salas Palma

Valeria Matamoros Navarro

Anthony Salas Villalobos

Profesor: Miguel Corea Toruño

25 agosto del 2022

Introducción

La cantidad de migrantes procedentes de Centroamérica ha aumentado significativamente durante los últimos 30 años, en un 137 por ciento entre 1990 y 2020, de 6,82 millones a casi 16,2 millones, pero ¿por qué se da tanta migración?, se especifican tres causas principales por las que los centroamericanos han sido obligados a salir de su país de origen: la pobreza, la inseguridad y violencia y el cambio climático.

Nunca el tema migratorio había adquirido tanta relevancia y trascendencia como la que asume hoy.

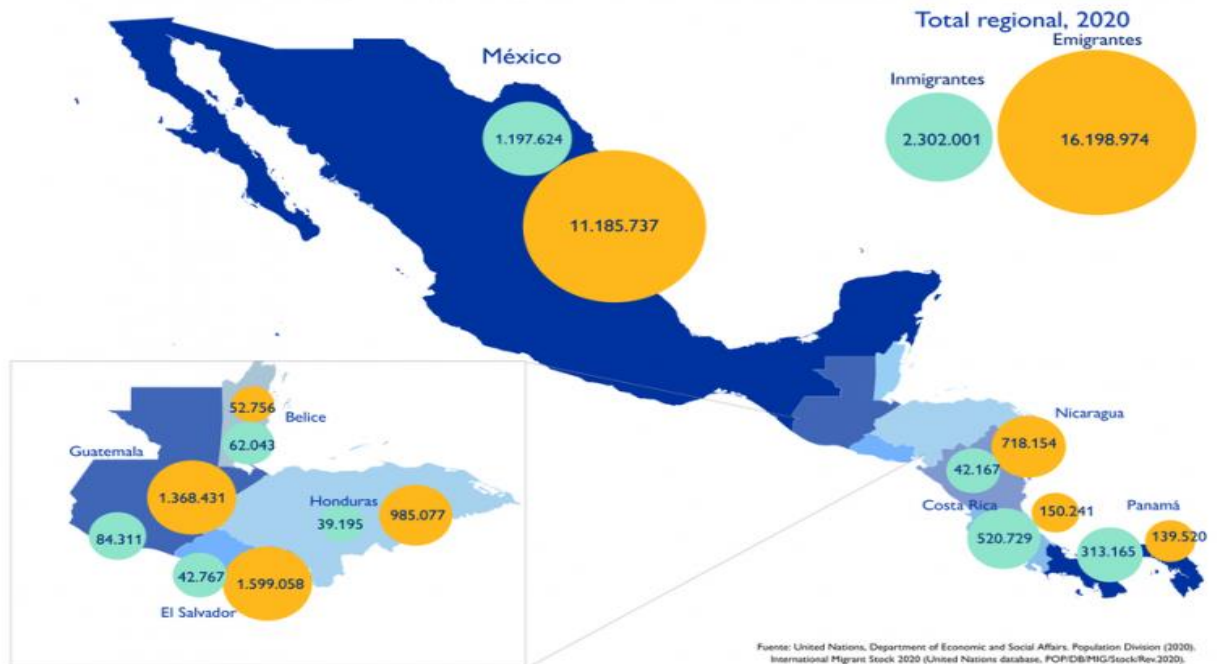
Entre el 2000 y el 2010, incrementó 59% el número de personas que nacieron en los países del norte de Centroamérica vive en un lugar distinto al que nació. En el caso de Honduras, el crecimiento fue de 94% y los porcentajes siguen creciendo día a día.

Planteamiento del problema

En los últimos años se ha incrementado las inmigraciones en distintos países de Centroamérica, los patrones que comúnmente se aprecian son: 1) movimientos de población de hacia países del norte como lo son Estados Unidos y Canadá en los cuales los inmigrantes deben abandonar sus respectivos países y 2) flujos de migración intrarregional, esto sucede cuando tienen como destino otros países de la región (ejemplo México, Panamá o Costa Rica).

En el año 2020 debido a la pandemia, el número de migrantes aumentó. Inclusive muchos países que han sido destino de migrantes han tenido que tomar distintas medidas, un ejemplo muy conocido es el muro que creó Estados Unidos en el gobierno de Donald Trump para poder evitar que los números siguieran creciendo.

Población total de inmigrantes y emigrantes en Centroamérica y México, 2020



En la imagen anterior podemos ver distintos datos de lo que fueron el número aproximado de los emigrantes e inmigrantes en distintos países y se puede ver que el mayor país con mayor inmigración y migración es México, en el 2020.

Justificación

Muchas son las razones para que los migrantes decidan buscar nuevas oportunidades, muchos deben dejar atrás sus empleos, amistades, familia y hogares, en busca para tener una vida más estable.

Uno de los motivos por lo cual deciden abandonar su país es por amenazas hacia ellos, sus familias o negocios, los cuales, al sentirse en peligro por la inseguridad, prefieren dejar todos sus bienes atrás. A esto se le suma que la tasa de homicidios en estos países generalmente es muy alta, por ejemplo:

Tasas de homicidios en 2017 - (Por cada 100.000 habitantes)	
El Salvador	60, equivalente a 3,947 víctimas
Honduras	43, equivalente a 3,791 víctimas
Guatemala	26, equivalente a 4,409 víctimas

Tasa de feminicidio en 2017 - (Homicidios de mujeres de 15 años y más por razón de género cada 100.000 mujeres)	
El Salvador	10.2
Honduras	5.8
Guatemala	2.6

Y otro punto importante son los fenómenos climáticos como las sequías, huracanes o terremotos, los cuales pueden llegar a destruir hogares en un país que la pobreza ya está muy marcada, además de impactar de una manera muy agresiva la producción ganadera y agrícola.

Alcance

- a. Metodología: Lo que se busca aplicar para fines de este proyecto es lograr conseguir un modelo que logre administrar de manera correcta y eficiente a los migrantes que existen a lo largo de Centroamérica

Fases

1. Recopilación de datos: Se tomará las medidas necesarias para poder recopilar datos antiguos y nuevos para integrarlos al nuevo modelo a crear y administrarlos de mejor manera
 2. Planificar: Aquí se enfocará en qué manera se pueden organizar los datos recopilados de manera que sea fácil, rápido y cómodo para los usuarios que estarán utilizando el modelo.
 3. Comunicar: Esta fase va de la mano con la fase anterior ya que se ocupará mucho dialogo para poner de acuerdo en que se va a hacer para llevar a cabo la creación del modelo
 4. Creación del modelo
- b. Recursos necesarios:
 1. Microsoft SQL Management Studio
 2. Pentaho
 3. Power BI
 - c. Factores de éxito

1. Buena comunicación
2. Buen relación cliente/equipo de desarrollo
3. Compromiso de todos aquellos involucrados en el proyecto

Objetivos

General:

- Desarrollar un modelo de Data Warehouse durante el transcurso del segundo cuatrimestre del año 2022, que permita el análisis de la información relacionada con migración en Centroamérica por medio de las cifras obtenidas de los distintos datasets.

Específicos

- Identificar las principales causas de muerte de los ciudadanos al migrar entre el periodo del 2000 al 2022.
- Dar a conocer datos estadísticos sobre la cantidad de migraciones que se realizan en Centroamérica entre los años 2000-2022, que ayuden a comprender más la importancia que tiene este tema.
- Recopilar datos de distintas fuentes con el fin de transfórmalos en información valiosa que ayuden a determinar una serie de medidas que se puedan aplicar para mejorar las condiciones y el proceso de migración en los próximos años.
- Determinar la cantidad de población que migran entre 2007-2017.

Requerimientos

Se presentan los siguientes requerimientos con respecto al desarrollo del proyecto, sin embargo, estos no limitan los posibles nuevos requerimientos o modificaciones en los ya existentes que se puedan presentar a raíz de como avanza el proyecto:

1. Se debe de mostrar reportes sobre la cantidad de personas que migran por medio de un rango de años.
2. Se debe de mostrar reportes sobre la cantidad de personas que migran filtrado por diferencia de género.
3. Se debe de mostrar reportes sobre la cantidad de personas que migran filtrado por país.

4. Se debe cargar datos para poder obtener la información necesaria para los reportes.
5. Se debe crear tabla en base de datos a raíz de la información necesaria y almacenada.
6. Se debe crear un ETL el cual almacene los datos necesarios para el análisis.

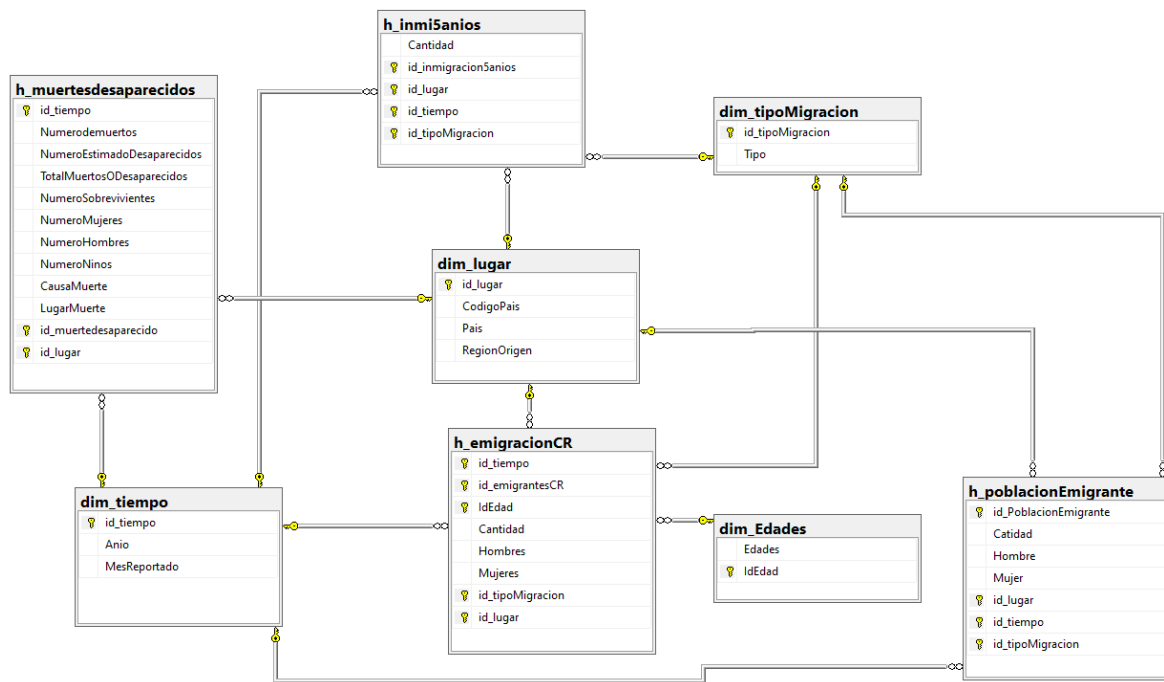
Riesgos

1. No contar con respaldos de la información utilizada y creada.
2. Mala planificación por parte de los integrantes del grupo.
3. Falta de incumplimiento en algunos requerimientos.
4. Pocos datos del tema elegido.
5. Bases de datos redundantes, no transparentes y sin documentar.
6. Existencia de datos poco confiables.
7. Infraestructura deficiente para el manejo de datos.
8. Mal manejo en la corrección de errores.
9. Uso incorrecto de los datos en la data base.
10. Falta de experiencia en las herramientas a utilizar.

Supuestos:

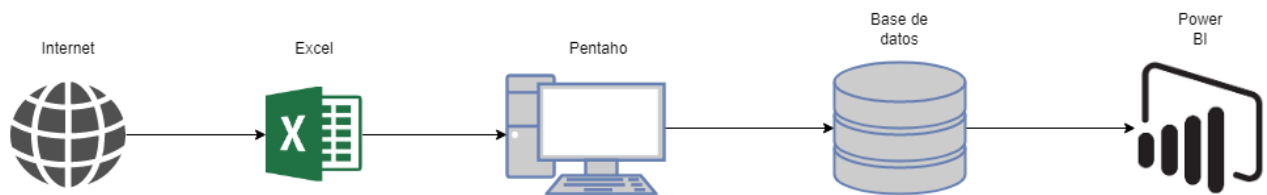
1. El equipo de trabajo está conformado por 4 miembros.
2. Se dispondrá de 15 semanas para realizar el proyecto.
3. Se utilizarán las herramientas vistas en clase como Pentaho, Power BI y Excel.
4. Se tiene como fecha límite para la entrega del proyecto el día 25 de agosto del 2022.
5. El tema principal del proyecto es migración en Centroamérica.
6. Se realizarán reuniones grupales cada semana para organizar el avance del proyecto.

Diagrama de datos



Diseño de la arquitectura

La arquitectura de bases de datos es una combinación de patrones que proporcionan la guía para la creación de database, permitir que los programadores, analistas y todo el desarrollador de software estén sincronizados donde puedan cubrir todos los objetivos y limitaciones de sus distintas aplicaciones. Se consideran el nivel más alto de diseño de arquitectura de datos porque establecen la estructura, el funcionamiento y la interacción entre las diversas partes que existen.



Herramientas:

Pentaho: es una aplicación que se utiliza para la inteligencia de negocios enfocada a la resolución de procesos utilizando la minería de datos, datawarehouse, generación de reportes entre otros.

Es de alta ayuda para la toma de decisiones y para mejorar el análisis de datos, puesto que ofrece alternativas fundamentales de una infraestructura de herramientas de análisis e informes.

Pentaho nos ofrece una amplia gama de productos útiles de la siguiente manera:

- Pentaho Reports: esta es una herramienta de renderizado capaz de crear informes programáticamente basados en perfiles XML.
- Pentaho Dashboard: esta es una plataforma integrada para proporcionar información sobre sus datos, donde puede ver informes, gráficos y más.
- Pentaho Data Mining: un conjunto de software que utiliza estrategias de aprendizaje automático, automatización y minería de datos. Cuenta con las herramientas necesarias para soportar procesos de análisis descriptivo.
- Pentaho para Apache Hadoop: es un conector de bajo nivel para facilitar el acceso a los volúmenes administrados en el proyecto Apache Hadoop.
- Servicios de análisis de Pentaho: admite lenguajes de comportamiento MDX y XML para la especificación y el análisis de la interfaz.

Además, tenemos muchas ventajas:

- Código abierto: No se requiere licencia para usar.
- Multiplataforma: Funciona tanto en Linux como en Windows, etc.
- Utilice tecnologías estándar como Java, XML o Javascript
- Dispone de entornos de desarrollo core tanto en servidores PDI como BI y otras herramientas.
- Fácil de instalar y configurar.
- Es una solución flexible que le permite crear nuevas funciones o módulos que se ajusten a las necesidades de su organización.
- Tiene un potente motor de extracción de datos y un servidor OLAP. Se enfoca en big data y bases de datos no relacionales.
- La comunidad de Pentaho es muy grande. Antes de su lanzamiento, este programa se sometió a una serie de pruebas comunitarias realizadas de forma altruista, que incluyen informes de errores y pruebas de software.

PowerBi: es una herramienta para extraer y presentar información que garantiza una buena gestión y visualización de datos para una mejor comprensión. Es un conjunto de servicios de software que interactúan con diversas aplicaciones para transformar datos en información valiosa y presentarla de forma interactiva y atractiva.

Las ventajas que tenemos al usar PBI son:

- Experiencia de usuario: este sistema básico tiene una interfaz con una excelente experiencia de usuario, divertido para la visualización y muy interactivo.
- Acceso a datos: otra de las ventajas de Power BI es que permite consultar diferentes bases de datos, incluso aquellas que son públicas y abiertas a consulta, como las del Banco Mundial o de algunos gobiernos.

- Seguridad de información: una de las ventajas de conocer Power BI es que se descubre que cuenta con un potente sistema de protección y seguridad de la información.
- Posee un costo económico muy bajo en el cual una suscripción en la nube puede costar alrededor de 10 dólares y posee una versión Desktop para usarse on-premise que es gratuita.
- Cuenta con una curva de aprendizaje relativamente sencilla ya que comparte similitudes con herramientas como Excel.
- Es muy flexible en cuanto a la cantidad de usos que se le puede dar y posee una gran conectividad con un gran número de programas entre los cuales se encuentran herramientas como Excel o las bases de Datos Oracle.

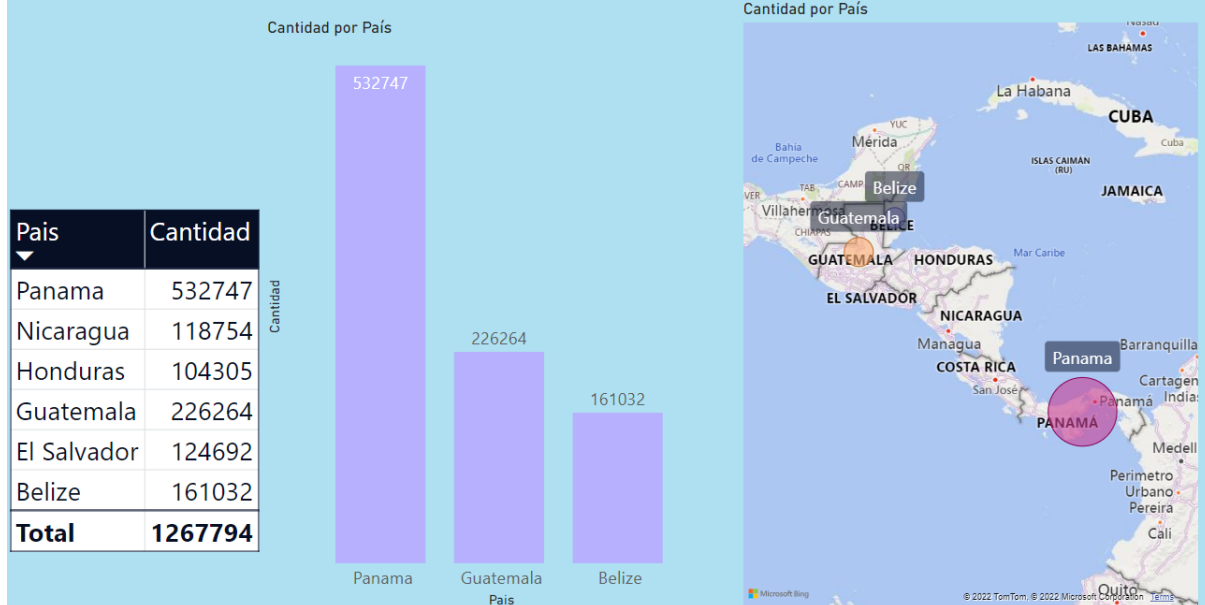
Diseño de cubos de datos:

Dim_tiempo	Dim_lugar	Dim_edades	Dim_tipoMigracion
id_tiempo Año Mes	id_lugar CodigoPais Pais RegionOrigen	id_edades Edad	id_tipoMigracion Tipo
Hechos: Cantidad de hombres, cantidad de mujeres, cantidad de niños, número de muertos, número estimado de desaparecidos y de sobrevivientes, causa y lugar de muerte			

Reportes:

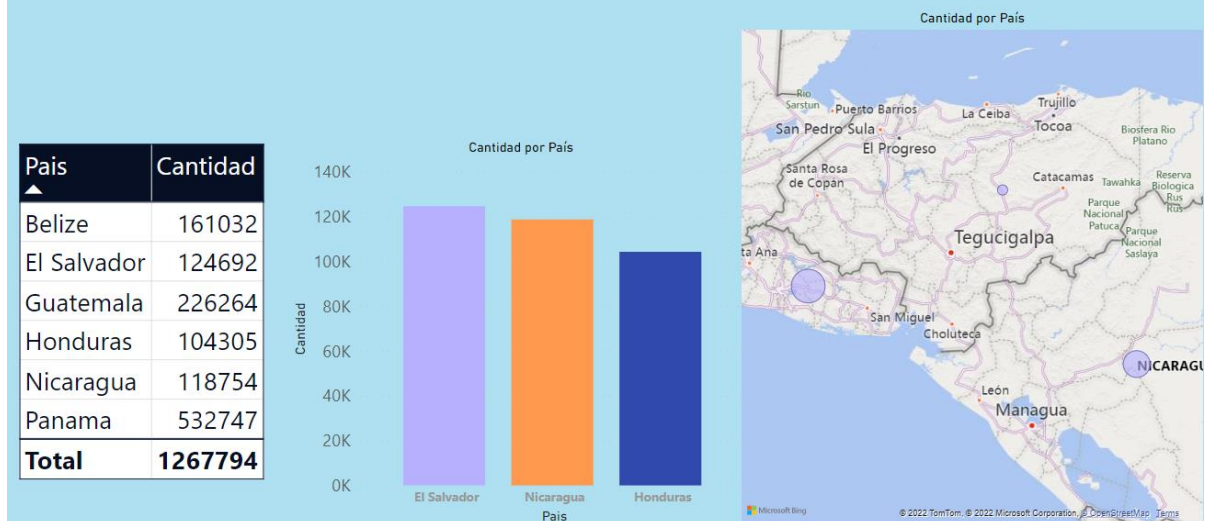
1. Países con el mayor número de migración.

Países con el mayor número de migración.



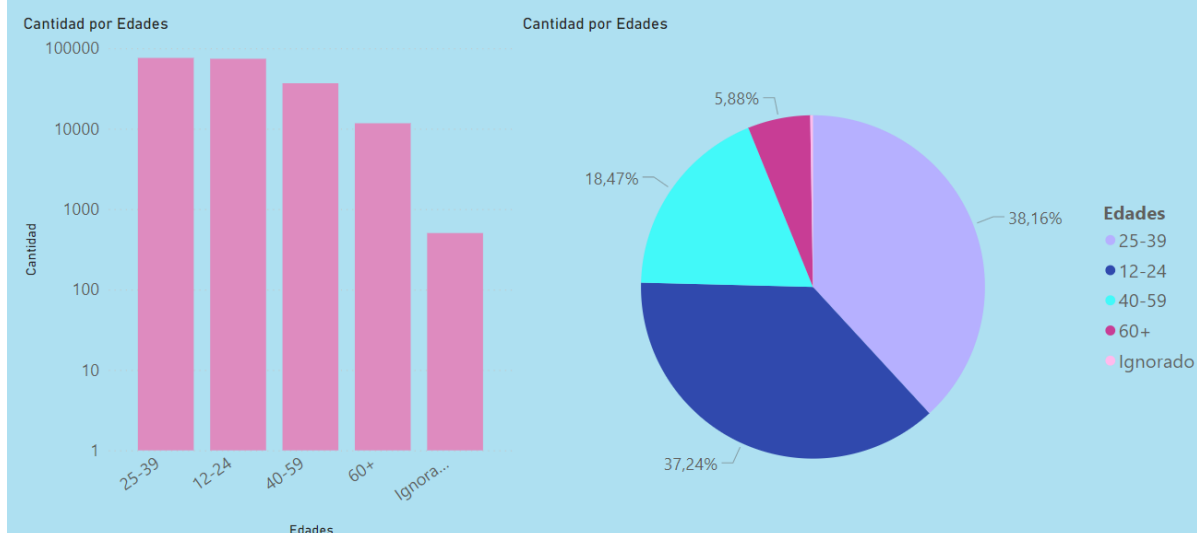
2. Países con el menor número de migración.

Países con el menor número de migración.

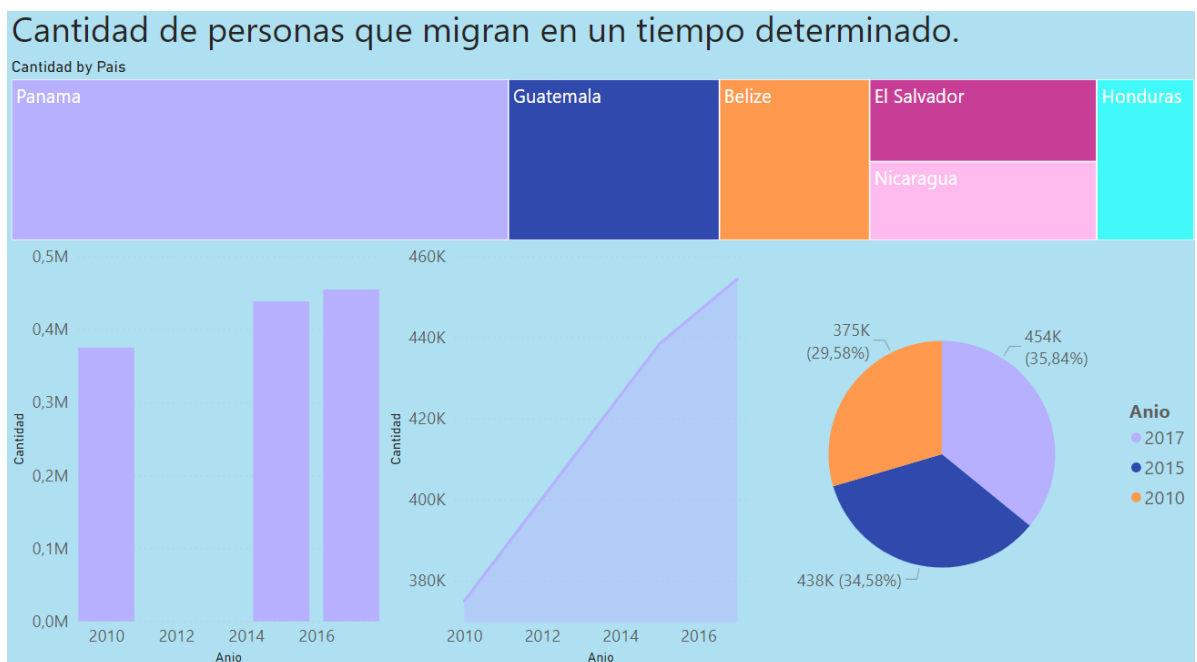


3. ¿Cuáles son las edades más frecuentes de las personas que migran?

¿Cuáles son las edades más frecuentes de las personas que migran?

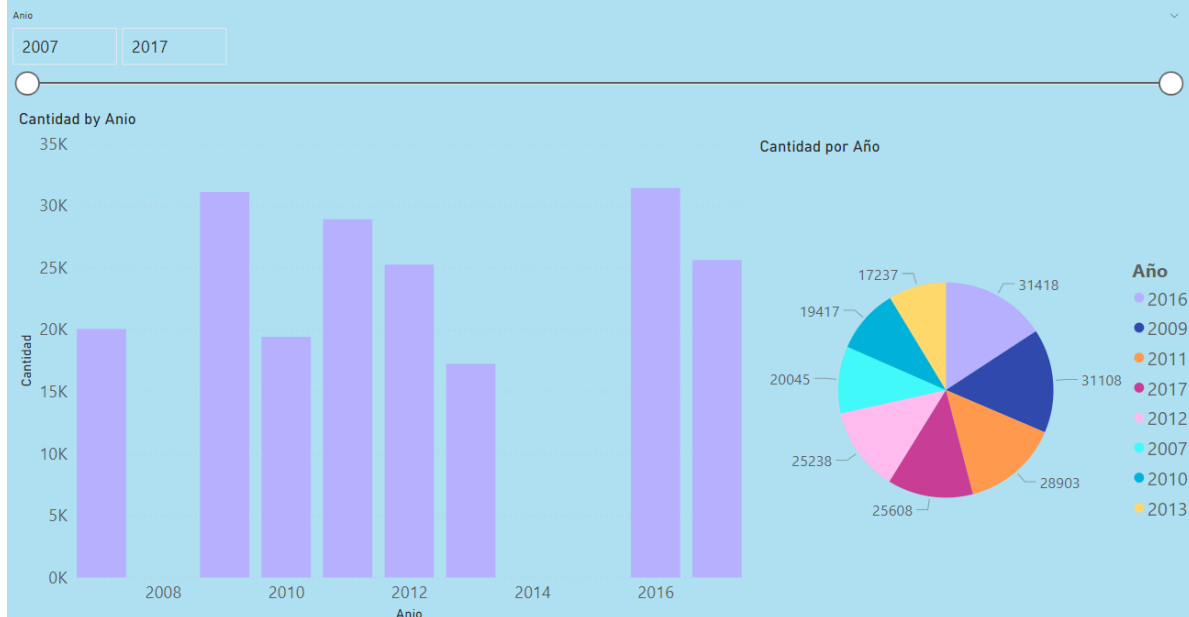


4. Cantidad de personas que migran en un tiempo determinado.



5. Migración de costarricenses a otros países

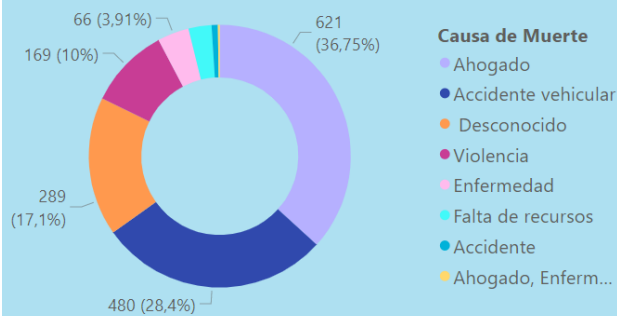
Migración de costarricenses a otros países



6. Muertes comunes durante la migración

Muertes comunes durante la migración

Total Muertos o desaparecidos by Causa de Muerte



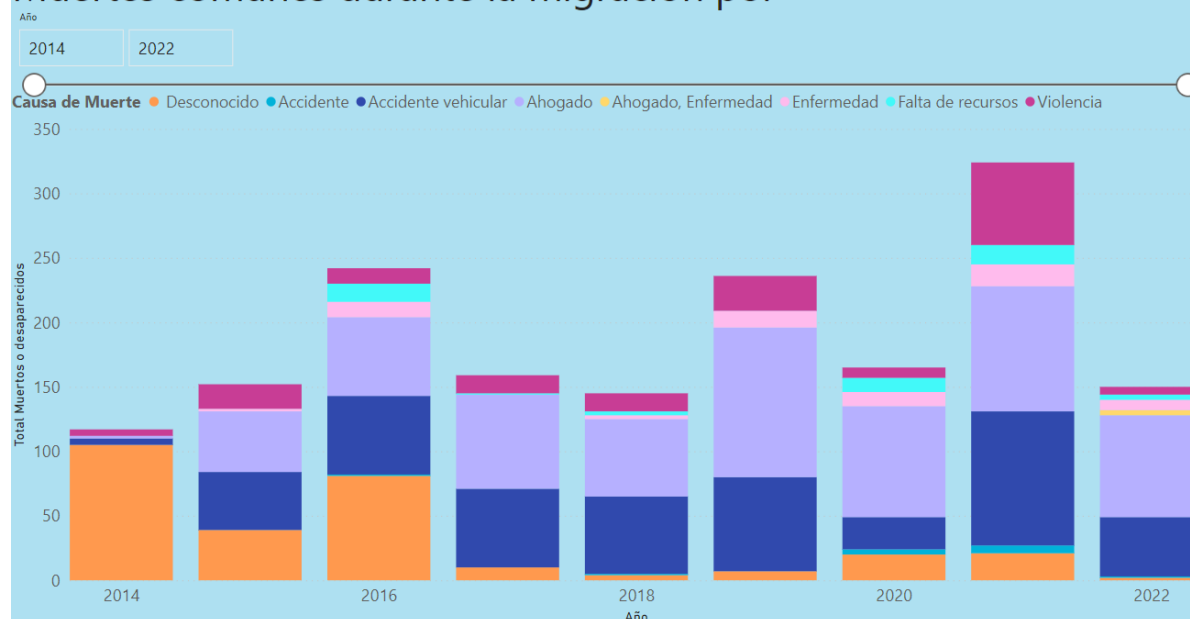
Causa de Muerte	Total Muertos o desaparecidos
Ahogado	621
Accidente vehicular	480
Desconocido	289
Violencia	169
Enfermedad	66
Falta de recursos	48
Accidente	13
Total	1686

Total Muertos o desaparecidos by Causa de Muerte

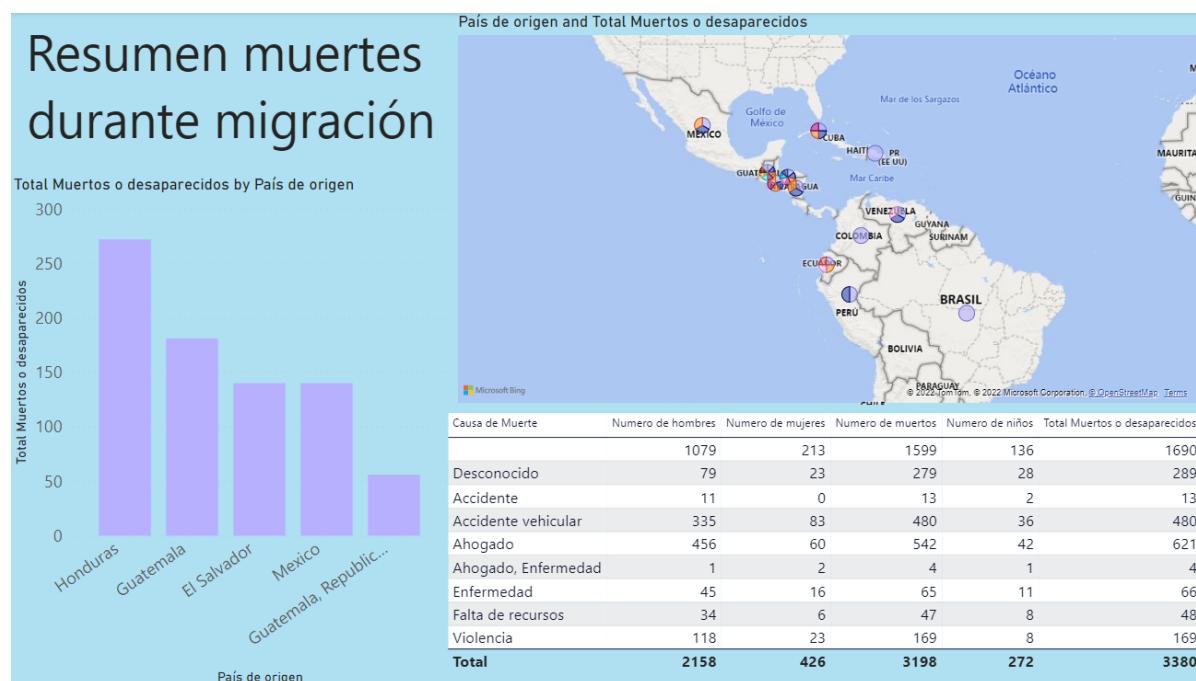


7. Muertes comunes durante la migración por año

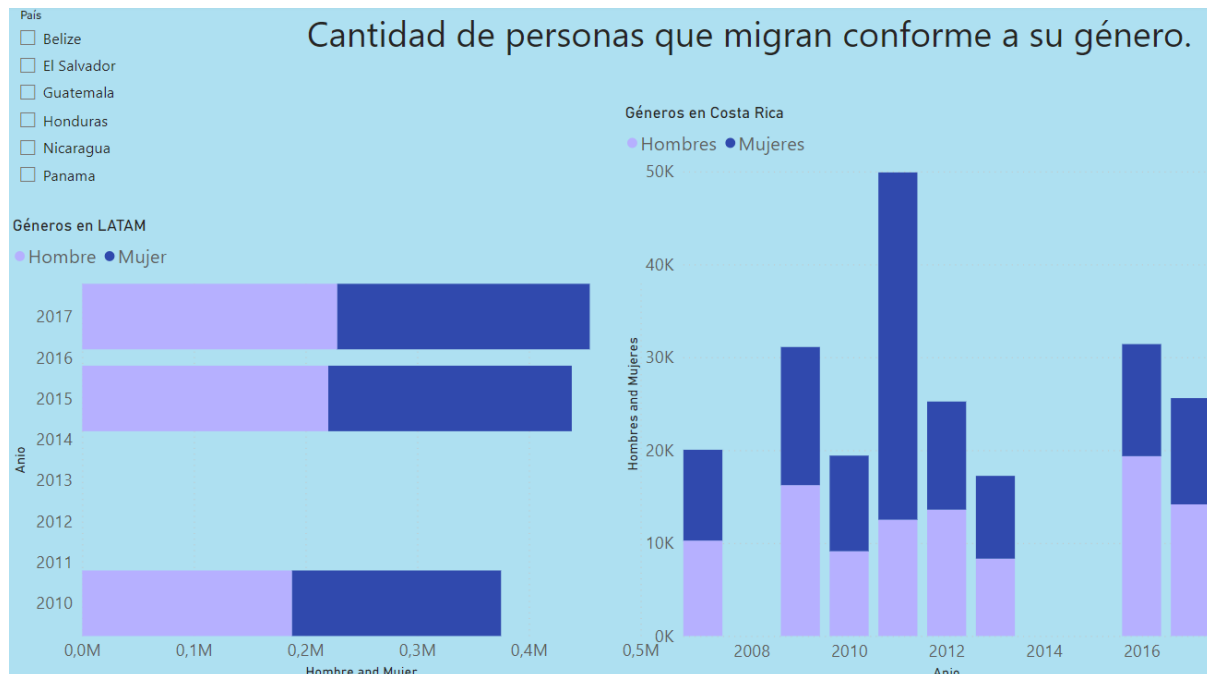
Muertes comunes durante la migración por



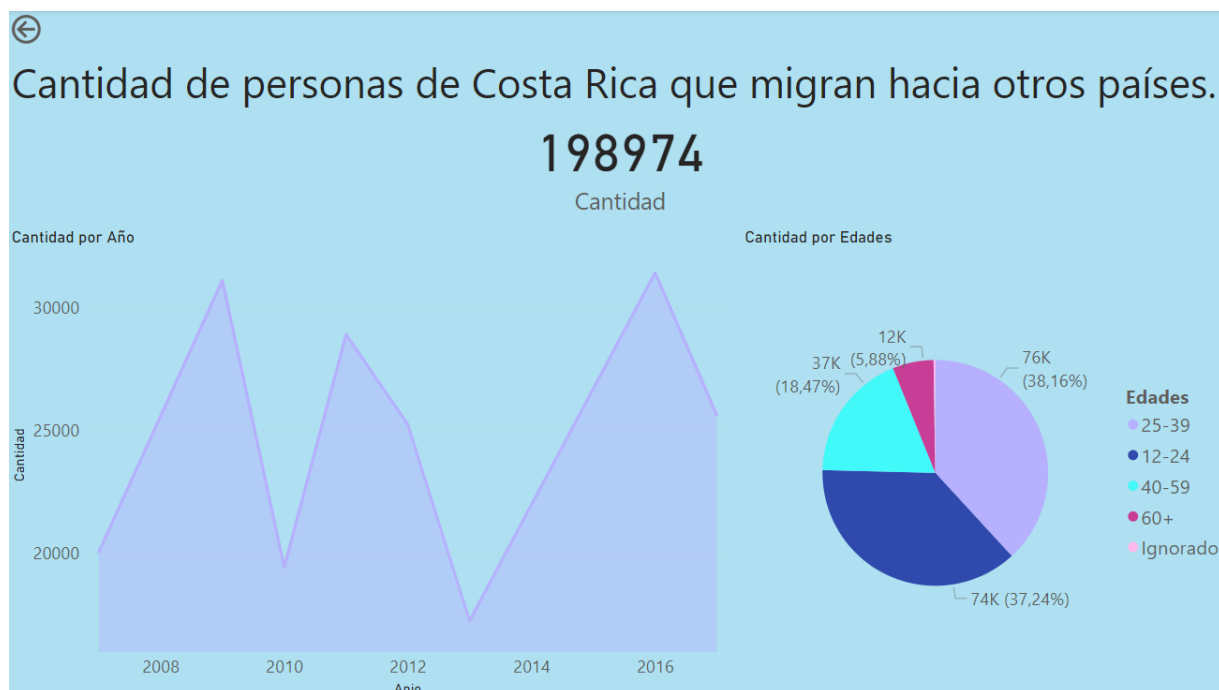
8. Resumen muertes durante migración



9. Cantidad de personas que migran conforme a su género.



10. Cantidad de personas de Costa Rica que migran hacia otros países.



Conclusiones:

Power BI cada vez más se está volviendo en una herramienta indispensable para todo tipo de aplicaciones tanto en el ámbito público como privado y con usos en diferentes campos.

El uso de Datawarehouse está sujeto a las herramientas mencionadas para capturar la información y a la capacidad de los entes que la deseen procesar para que lo hagan de la manera más adecuada para que los datos recolectados sean precisos.

Cada vez más se ve más interés en desarrollar el tema de Datawarehouse lo cual ha sido muy positivo por la expansión que provoca que cause que surja un nuevo mercado para el involucramiento de profesionales especializan en esta área lo cual también provoca demanda en carreras enfocadas a este tema.

Sin embargo, el uso de datos debe también de tomar en cuenta implicaciones legales y éticas y supone una inversión para ciertas entidades lo cual puede causar una desigualdad.

Sin duda el Datawarehouse es un tema que debe de ir evolucionando para que el provecho que se le saque sea positivo y que en los próximos años las partes involucradas se vean beneficiadas sin provocar perjuicios a la población.

Recomendaciones:

Luego de realizar las investigaciones y prácticas requeridas, nos hemos dado cuenta de que el tiempo fue bastante justo para abarcar el vasto mundo para las bases de datos enfocada al datawarehouse, por lo que se recomienda enfocar futuras interacciones en cada una de las herramientas descritas en el trabajo.

Se recomienda también buscar el consejo de una persona experta en el procedimiento de esta, para que pueda dar orientación a la hora de formular los objetivos, y además algún tipo de guía sobre los procesos que mayormente son empleados en el mercado actual.

Se recomienda, además, tomarse el tiempo de buscar ejemplos, guías o tutoriales, es decir, no limitarse a investigar artículos e información, sino también poner manos a la obra e interactuar con estos informes, ya que puede dar una visión más detallada de lo que sería un datawarehouse.

Lecciones aprendidas:

Lo primero es decir que es muy complicado dominar por completo un tema tan grande como es la parte de datawarehouse ya que se abarca mucho contenido, procedimientos, herramientas, y datos como cualquier otra rama esta puede estar actualizándose y mejorando; pero esto no debe ser excusa para dejar de aprender y tener criterio para poder decidir cuál es la mejor forma de realizar un proyecto de gran magnitud. En este proyecto nos enfocamos más que todo en la parte de datos sobre migración en Centroamérica, de nuestra parte era un tema que no teníamos mucha información antes de realizarlo y nos ayudó bastante a tener una mejor práctica para que el día de mañana tengamos las bases para poder realizar nuevamente un proyecto así.

Bibliografía:

¿Qué es Pentaho y cuáles son sus beneficios? (s. f.). ITOP. Recuperado 21 de agosto de 2022, de

<https://www.itop.es/blog/item/que-es-pentaho-y-cuales-son-sus-beneficios.html#:~:text=Pentaho%20es%20una%20plataforma%20de,%2C%20ETL%2C%20generaci%C3%B3n%20de%20informes%E2%80%A6>

Flores, A. (2022, 17 mayo). *¿Qué es Power BI? Incursiona en el análisis de datos con esta*

herramienta. <https://www.crehana.com>. Recuperado 21 de agosto de 2022, de

<https://www.crehana.com/blog/desarrollo-web/que-es-power-bi/>

Anexos:

[Datawarehouse Excel](#)

[PowerBi - Proyecto](#)

[Pentaho - SQL](#)