



Contenido

DESCRIPCIÓN DEL PROYECTO.	3
OBJETIVOS.....	3
NOSOTROS.....	4
MISIÓN.	4
VISIÓN.	4
METODOLOGIA.....	4
ORQUESTACION AIRFLOW	5
PIPELINE.	7
KPIs.....	8
Tasa de Emigración	8
Tasa de inmigración	9
Tasa de Homicidios por cada 100K habitantes	10
La tasa de remesas	10
Modelo Machine Learning.....	11

DESCRIPCIÓN DEL PROYECTO.

MIGRATIO es una consultora contratada por una ONG internacional para estudiar los flujos migratorios y su impacto en distintos aspectos de los países analizados. El objetivo principal es analizar las causas y consecuencias de las migraciones humanas, considerando factores políticos, económicos, sociales y ambientales en los países de origen y destino, así como los conflictos armados, los desastres naturales y las políticas migratorias.

El proyecto se centra en la creación de una solución integral de datos, incluyendo el desarrollo de un data pipeline, una base de datos, un data warehouse, un dashboard y modelos predictivos. Estos componentes permitirán comunicar de manera eficaz la relación entre los flujos migratorios y los aspectos socioeconómicos y geopolíticos de los países analizados.

OBJETIVOS.

- Analizar los flujos migratorios y su impacto en los países receptores y emisores.
- Identificar las causas y consecuencias de las migraciones, considerando factores políticos, económicos, sociales y ambientales.
- Estudiar las condiciones políticas, económicas y sociales de los países de origen y destino, así como los conflictos armados, los desastres naturales y las políticas migratorias.
- Crear una solución de datos completa que incluya un data pipeline, una base de datos, un data warehouse, un dashboard y modelos predictivos.
- Comunicar eficazmente la relación entre los flujos migratorios y los aspectos socioeconómicos y geopolíticos de los países analizados.

NOSOTROS

MISIÓN.

Nuestra misión es proporcionar información de calidad relacionada con temas migratorios para respaldar la toma de decisiones informadas por parte de organizaciones internacionales, gobiernos y otras partes interesadas.

VISIÓN.

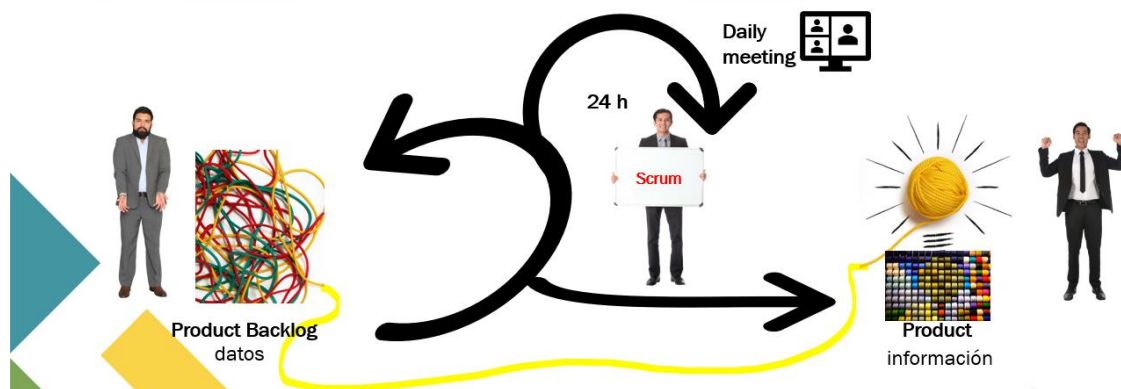
Nuestra visión es convertirnos en la consultora especializada líder en temas migratorios, reconocida por la calidad de los servicios que brindamos y por nuestra contribución a la comprensión y gestión de los flujos migratorios a nivel global.

METODOLOGIA.

Scrum es una metodología ágil de gestión de proyectos que se utiliza principalmente para el desarrollo de software. Se basa en la idea de que los proyectos cambian constantemente en el mundo real, y este cambio debe ser planificado y aceptado. Scrum utiliza un enfoque iterativo e incremental para optimizar la predictibilidad y el control del riesgo. Los proyectos se dividen en ciclos de trabajo llamados "sprints", que suelen durar de una a cuatro semanas.

Cada sprint comienza con una reunión de planificación, donde el equipo selecciona los elementos de una lista de tareas pendientes, conocida como "backlog", para trabajar durante el sprint. Durante el sprint, el equipo se reúne diariamente en las "reuniones de scrum" para discutir el progreso y los obstáculos. Al final del sprint, el equipo revisa su trabajo en la "revisión del sprint", demostrando lo que ha completado. Luego, en la "retrospectiva del sprint", el equipo reflexiona sobre su proceso y hace planes para mejorar en el próximo sprint. Este ciclo se repite hasta que el proyecto esté completo.

Metodología usada: SCRUM



ORQUESTACION AIRFLOW

El objetivo principal de este proceso es facilitar la extracción, transformación y carga de datos de manera automatizada. Para ello, se utilizan las siguientes tecnologías:

Docker: es una plataforma de código abierto que permite automatizar la implementación de aplicaciones dentro de contenedores. Los contenedores de Docker encapsulan las aplicaciones y todas sus dependencias en un entorno aislado, lo que garantiza la portabilidad y consistencia del entorno de ejecución. En nuestro proyecto, Docker se utiliza para empaquetar y desplegar los componentes del sistema de ETL de manera eficiente y reproducible.

Airflow: es una plataforma de código abierto para programar, monitorear y orquestar flujos de trabajo (workflows) en el contexto del procesamiento de datos. Permite definir tareas y sus dependencias en un formato legible, lo que facilita la creación y gestión de flujos de trabajo complejos. En nuestro proyecto, Airflow se utiliza para orquestar el proceso de ETL, coordinando la ejecución de las diferentes etapas y tareas involucradas en el flujo de datos.



En este proyecto, hemos utilizado Docker y Airflow para orquestar el proceso de ETL de manera eficiente y automatizada. Mediante el uso de Docker, logramos empaquetar y desplegar los componentes del sistema de manera consistente y portátil. Airflow, por su parte, nos permitió definir y coordinar las tareas del flujo de datos, incluyendo la carga incremental de datos y la limpieza de archivos CSV para seleccionar los países de interés. Estas tecnologías han facilitado la implementación y escalabilidad del proyecto, mejorando la productividad y reduciendo la posibilidad de errores manuales.

Al utilizar Docker, pudimos asegurar la reproducibilidad y consistencia del entorno de ejecución, lo que simplificó el despliegue del sistema en diferentes entornos y garantizó que todas las dependencias estuvieran correctamente configuradas.

Airflow, por otro lado, nos brindó la capacidad de programar y orquestar flujos de trabajo complejos, permitiéndonos definir las dependencias entre tareas y controlar la ejecución de manera automatizada. Esto nos permitió ahorrar

tiempo y recursos al evitar la ejecución manual de cada tarea y garantizar la correcta secuencia de las etapas del proceso ETL.

La implementación de la carga incremental de datos nos permitió optimizar el tiempo de ejecución al cargar solo los datos nuevos o modificados desde la última ejecución exitosa del flujo de datos. Esto resultó en una mayor eficiencia y un menor consumo de recursos, ya que evitamos procesar y cargar datos redundantes.

La limpieza de archivos CSV basada en los códigos de país de tres letras nos permitió seleccionar únicamente los datos relevantes para nuestro trabajo, eliminando registros innecesarios y mejorando la calidad de los datos cargados. Esto aseguró que trabajáramos con la información pertinente y redujo el riesgo de errores causados por datos irrelevantes o inconsistentes.

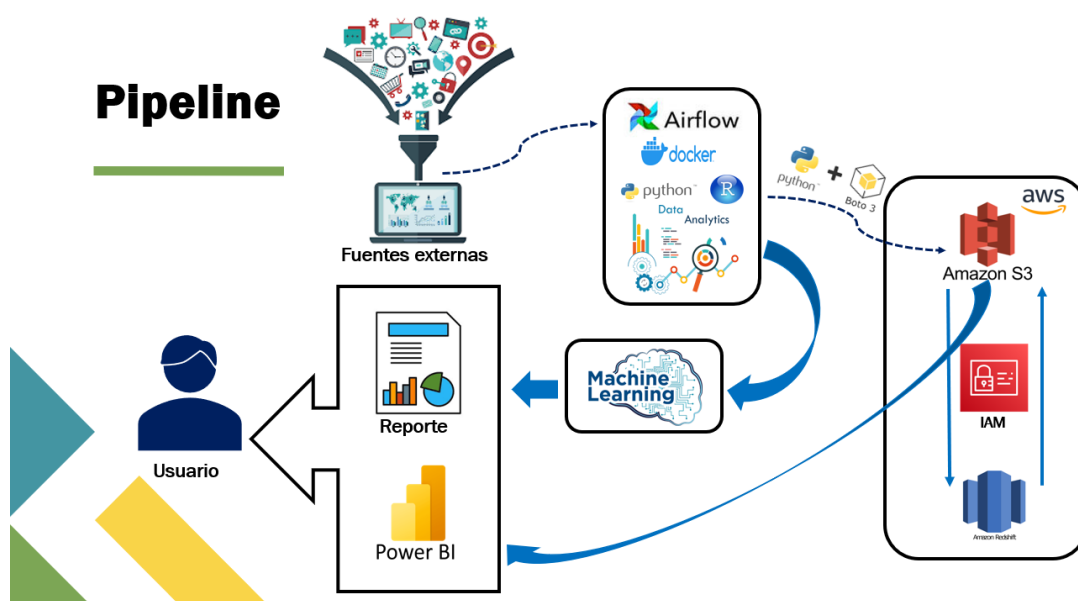
En resumen, la combinación de Docker y Airflow en nuestro proyecto de orquestación de ETL ha proporcionado una solución eficiente y automatizada para la extracción, transformación y carga de datos. Estas tecnologías han mejorado la escalabilidad, la productividad y la confiabilidad del proceso, al tiempo que nos han permitido implementar estrategias de carga incremental y limpieza de datos para garantizar la calidad y relevancia de la información utilizada en nuestro trabajo.

PIPELINE.

El modelo desarrollado para el pipeline, incluyendo lo implementado en el punto anterior es el que puede verse en la ilustración siguiente. Los datos se toman de fuentes externas, se realiza una primera inspección y modificaciones en Python. Luego se ingresa al proceso orquestado por AIRFLOW. Con la salida, y mediante un script de Python se sube a un bucket de S3 en AWS previamente creado. Se utiliza IAM para gestionar permisos y se crea una tabla en Redshift con el mismo formato para poder tomar la información de esas tablas. Ahí se

pueden realizar queries y guardar esa salida en un archivo que de forma paralela se archiva en el mismo bucket de S3.

A continuación, mediante script generamos una URL temporal para poder importar esas tablas en Power Bi para poder continuar el proceso de diseño de los indicadores.



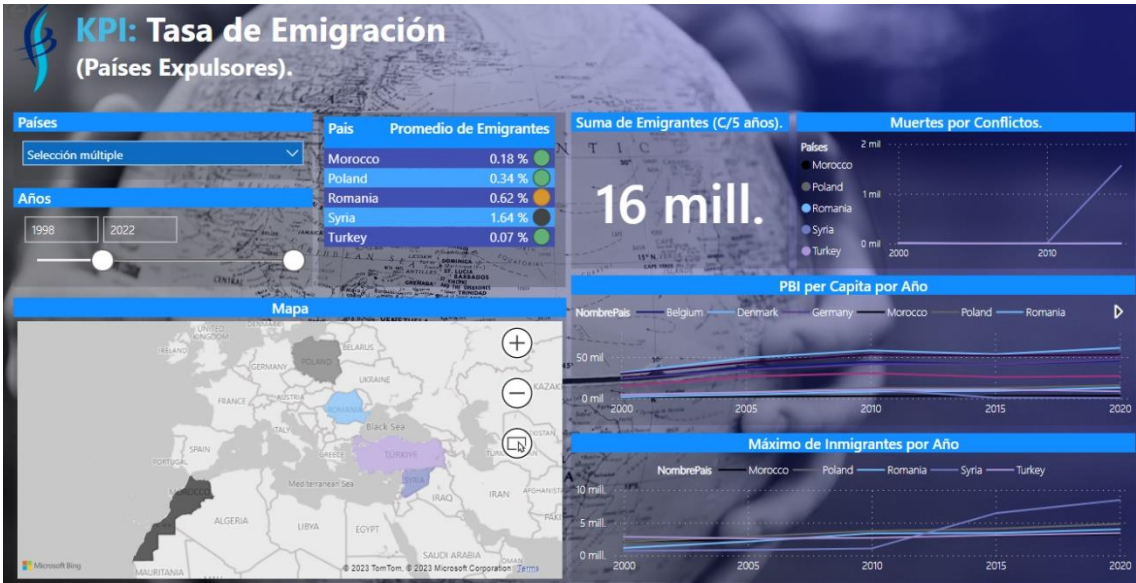
KPIs

Migratio propone a la ONG una serie de KPIs o Indicadores Clave de Desempeño que pueden ser muy relevantes para tomar decisiones políticas en relación con la migración en los distintos estados. Los KPIs son medidas cuantitativas que permiten evaluar el desempeño de un proceso o actividad en relación con objetivos específicos. En el contexto de la migración, los KPIs pueden ayudar a evaluar la efectividad de las políticas migratorias, así como a identificar áreas de mejora y oportunidades de crecimiento:

Tasa de Emigración

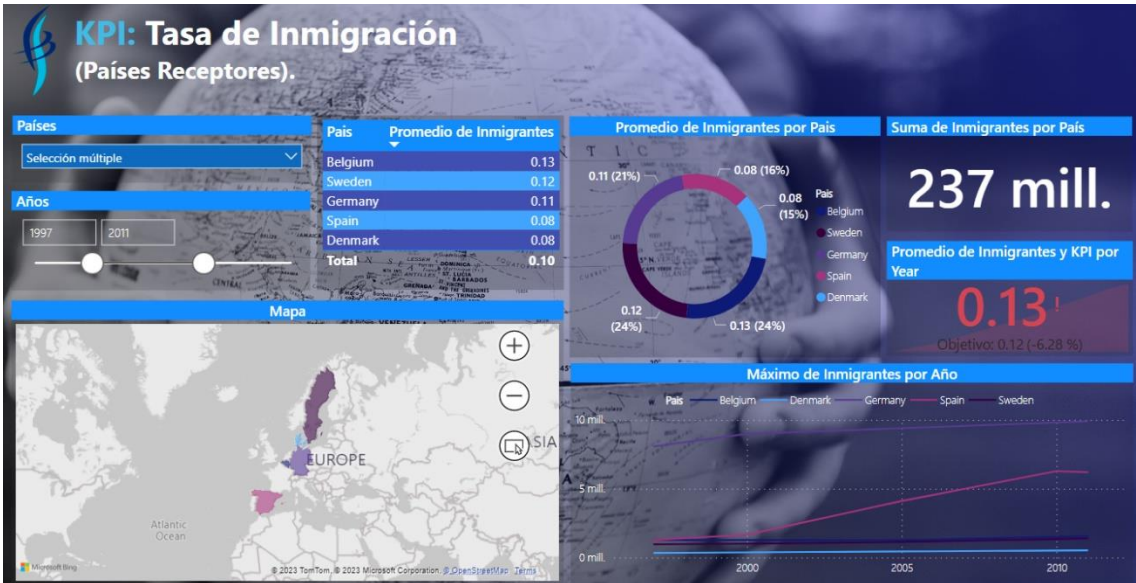
Se refiere a la proporción de personas que emigran desde un país o región en un período de tiempo determinado. Esta medida puede ayudar a evaluar la calidad de vida en un lugar y las oportunidades disponibles para las personas. Por ejemplo, si la tasa de emigración es alta, puede indicar que las personas no

están encontrando oportunidades suficientes en su lugar de origen y están buscando mejores oportunidades en otro lugar.



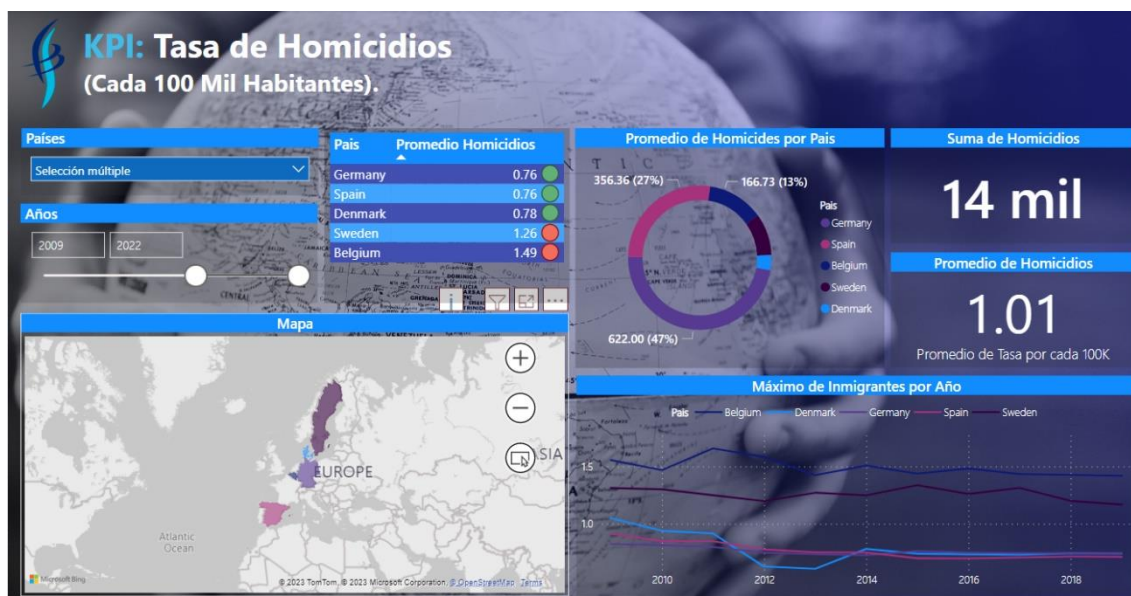
Tasa de inmigración

Se refiere a la proporción de personas que inmigran a un país o región en un período de tiempo determinado. Esta medida puede ayudar a evaluar la atracción de un lugar para las personas de otras regiones o países. Por ejemplo, si la tasa de inmigración es alta, puede indicar que el lugar ofrece buenas oportunidades de empleo, educación y calidad de vida.



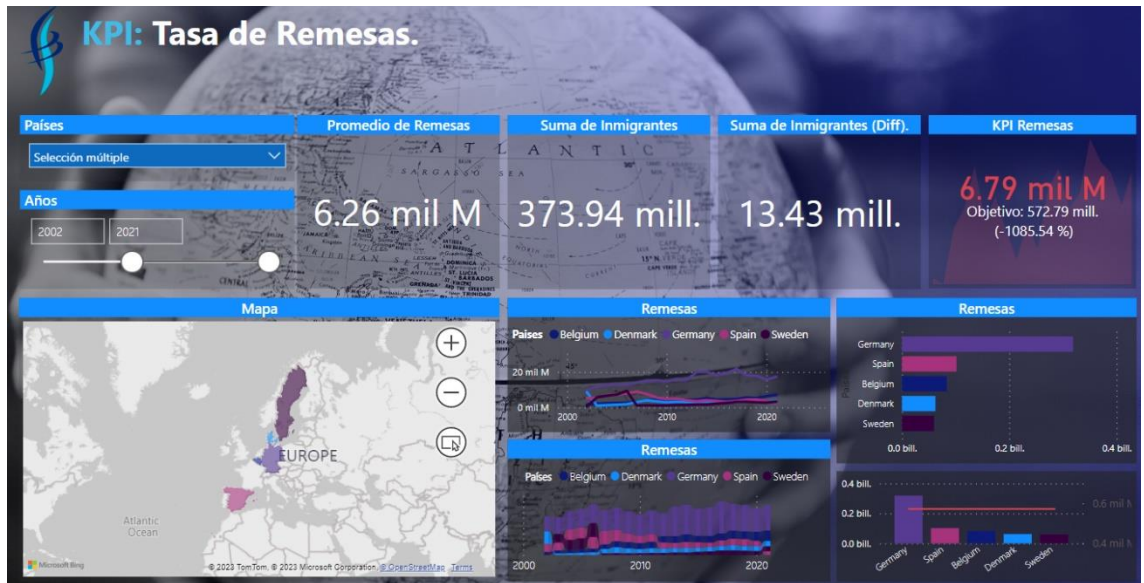
Tasa de Homicidios por cada 100K habitantes

Se refiere a la cantidad de homicidios que ocurren en una región o país en un período de tiempo determinado, en relación con la cantidad de habitantes. Esta medida puede ayudar a evaluar la seguridad de un lugar y su estabilidad política. Por ejemplo, si la tasa de homicidios es alta, puede indicar un entorno inseguro y desfavorable para la vida cotidiana y la inversión.



La tasa de remesas

Se refiere al dinero que las personas envían a sus hogares en su lugar de origen desde otro lugar donde están trabajando o viviendo. Esta medida puede ayudar a evaluar el impacto económico de la migración en la región o país de origen. Por ejemplo, si la tasa de remesas es alta, puede indicar que la migración está contribuyendo a la estabilidad financiera de las familias en el lugar de origen, pero también puede significar una fuga de cerebros y talentos de la región.



Es importante tener en cuenta que estas medidas no son una evaluación completa y exhaustiva de la situación migratoria de un lugar, pero pueden servir como puntos de referencia útiles para evaluar la calidad de vida, la seguridad, la atracción y el impacto económico de la migración.

Modelo Machine Learning

El modelo utilizado se basa en el algoritmo de Gradient Boosting Regressor que sigue los siguientes pasos:

- Comienza con un conjunto de datos de entrada.
- El primer modelo se entrena y hace predicciones.
- Se calculan los errores residuales (la diferencia entre las predicciones y los valores reales).
- El siguiente modelo se entrena utilizando estos errores residuales como objetivo.
- Este proceso se repite n veces, cada vez entrenando un nuevo modelo en los errores residuales del modelo anterior.

- El modelo final es la suma de las predicciones de todos los modelos individuales.
- Este modelo final se utiliza para hacer las predicciones de salida.

