

Análisis de la historia de accidentes aéreos entre 1908 y 2023:

Autor: Esteban Mohr

Institución: Coderhouse

Curso: Data Analytics:

Fecha de emisión: 09 de mayo 2024

Tabla de Contenidos

Introducción:	3
Descripción de la temática de los Datos:	3
Hipótesis:.....	3
Dataset:	4
Objetivo del proyecto.....	4
Alcance del Proyecto	4
Usuario final y nivel de aplicación.....	4
Diagrama entidad-relación.....	5
Listado de tablas.....	5
Listado de columnas por tablas.....	6
Carga de Datos en Power BI y ajustes iniciales	7
Transformación de Datos	9
Creación de tabla calendario.....	11
Columnas Calculadas.....	11
Medidas Calculadas.....	11
Modelo de datos en Power BI.....	13
Visualizaciones Creadas	14
Solapa Portada:	14
Solapa Resumen:	14
Solapa Análisis Temporal:	15
Solapa Análisis Lugar:	15
Solapa Análisis Ruta/Operador:	16
Solapa Detalles Accidentes:	16
Conclusiones	17
Futuras Líneas	18

Introducción:

El transporte aéreo ha experimentado una evolución notable a lo largo de la historia, pasando de ser una actividad de riesgo a convertirse en uno de los medios de transporte más seguros del mundo. Esta transformación no ha sido obra del azar, sino el resultado de décadas de avances tecnológicos, regulaciones rigurosas y lecciones aprendidas a partir de tragedias pasadas.

En este contexto, el análisis detallado de datos sobre accidentes aéreos ofrece una ventana reveladora sobre los patrones, tendencias y factores que han influido en la seguridad de la aviación.

En este informe, se detallan diferentes insights obtenidos a partir de observaciones del tablero de datos desarrollado en Power BI, que demuestra la efectividad de las regulaciones aeronáuticas y la mejora continua en la seguridad en la aviación. Estas conclusiones no solo reflejan el pasado, sino que también proyectan un futuro prometedor en el cual los viajes en avión continúan siendo el medio de transporte más seguro y confiable disponible.

Descripción de la temática de los Datos:

Se recopilan datos de la historia de accidentes de aviones y fatalidades en todo el mundo desde 1908 hasta 2023. El conjunto de datos encapsula información invaluable para investigadores, entusiastas de la aviación y expertos en seguridad interesados en comprender la dinámica, tendencias y patrones de incidentes de aviación a lo largo de más de un siglo.

Hipótesis:

Se desea comprender la evolución del comportamiento de los accidentes aéreos a lo largo de la historia de la aviación. En concreto en base a los datos se desea conocer lo siguiente:

- Cuáles han sido los accidentes aéreos que más vidas humanas se ha llevado durante la historia de la aviación.
- Descubrir cuales modelos de aviones han sufrido una mayor cantidad de accidentes.
- Conocer cuáles son los operadores aéreos que más accidentes han sufrido, la frecuencia de sus accidentes y la cantidad de fallecimientos acumulados en sus accidentes.
- Descubrir si existen áreas geográficas (lugares, estados o países) en las cuales han acontecido más cantidad de accidentes y/o muertes en relación con otras.
- Analizar si existen rutas u horarios en los cuales se han presentado accidentes de forma más frecuente.

Posteriormente y en base a los análisis realizados se concluirá en qué condiciones existe un riesgo mayor de producirse un accidente y se generarán propuestas de cómo se podrían prevenir futuros accidentes.

Dataset:

El dataset utilizado se encuentra en el archivo dataset_accidentes_aereos_Esteban_Mohr.csv, el cual contiene 4998 filas y 17 columnas. Las columnas del dataset son las siguientes:

- **Date:** Fecha del accidente
- **Time:** Hora del accidente
- **Location:** Lugar del accidente
- **Operator:** Operador del vuelo
- **Flight #:** Número de vuelo
- **Route:** Ruta de vuelo
- **AC Type:** Tipo/Modelo de aeronave
- **Registration:** Número de matrícula del avión
- **cn/ln:** Número de serie/fabricación del avión
- **Aboard:** Cantidad de personas a bordo
- **Aboard Passangers:** Cantidad de pasajeros a bordo
- **Aboard Crew:** Cantidad de personas de tripulación a bordo
- **Fatalities:** Cantidad total de fallecidos
- **Fatalities Passangers:** Número de pasajeros fallecidos
- **Fatalities Crew:** Número de fallecidos en tripulación
- **Ground:** Número de fallecidos en tierra
- **Summary:** Breve descripción del accidente

Objetivo del proyecto

El objetivo de este proyecto es crear un Dashboard en Power BI que permita facilitar encontrar las respuestas a las preguntas planteadas en la sección 2 de este documento.

Alcance del Proyecto

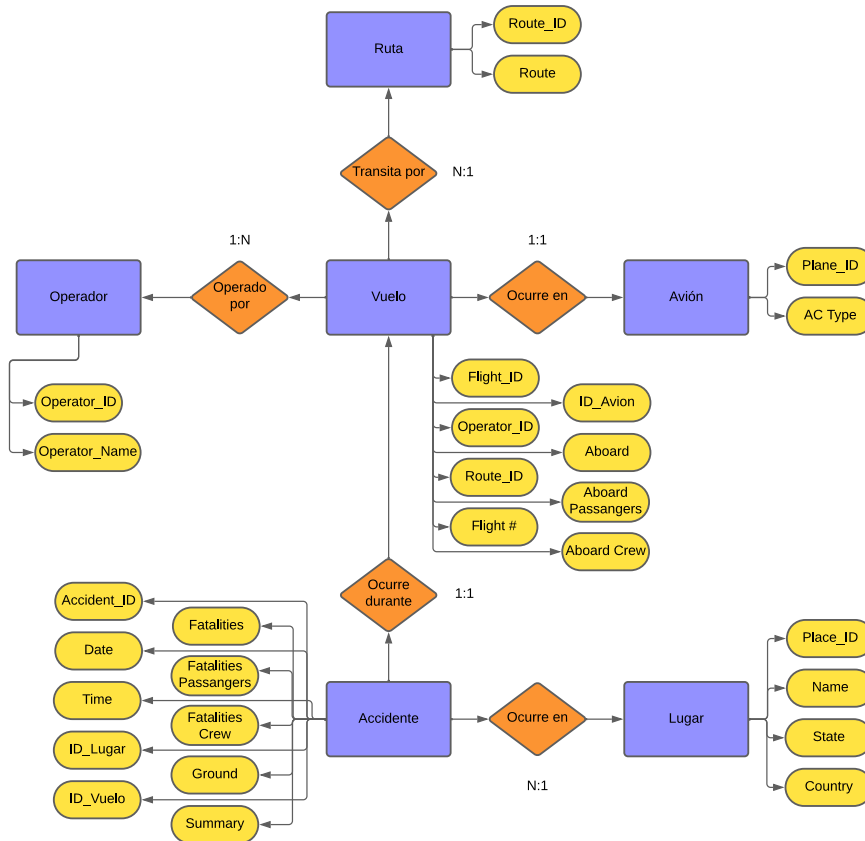
El alcance de este proyecto implica la creación de la base de datos y la creación del dashboard en Power BI. El dashboard se considerará como terminado cuando contenga todas las visualizaciones necesarias para que el usuario final pueda responder a las preguntas planteadas en la sección 2 de este documento.

Usuario final y nivel de aplicación

Los usuarios finales de esta aplicación son de tipo estratégicos, tales como gerentes de aerolíneas que desean evaluar qué condiciones pueden ser riesgosa para la operación en el largo plazo o miembros de organizaciones reguladoras de aviación en busca de la creación de normas aeronáuticas que aseguren la seguridad en los vuelos.

Diagrama entidad-relación

En la siguiente imagen se observa el diagrama de entidad-relación del modelo de datos creado a partir del Dataset:



Listado de tablas

- **Accidente:** Contiene la fecha, hora,
 - PK: Accident_ID
 - FK: ID_Lugar
 - FK: ID_Vuelo
- **Lugar:** Contiene todos los lugares en los cuales ha habido un accidente. Los registros de lugares están separados por nombre, estado (solo para Estados Unidos y Canadá) y país.
 - PK: Location_ID
- **Vuelo:** Contiene los datos del vuelo en el cual ocurrió el accidente, tal como el operador, ruta, número de vuelo, modelo de avión y personas a bordo.
 - PK: Flight_ID
 - FK: Operator_ID
 - FK: Route_ID
 - FK: Plane_ID
- **Operador:** Contiene todos los operadores que han tenido algún accidente aéreo. Entiéndase Operador como la empresa u organización responsable del vuelo.

- PK: Operator_ID
- **Ruta:** Contiene todas las rutas en las cuales ha sucedido algún accidente aéreo.
 - PK: Route_ID
- **Avión:** Contiene todos los modelos de aviones en los cuales han ocurrido accidentes.
 - PK: Plane_ID

Listado de columnas por tablas

A continuación, se presentan las tablas creadas en la base de datos (archivo “Base de Datos.xls”) con sus respectivas tablas, sus tipos de datos y si corresponde se indica si la columna es llave primaria (PK) o foránea:

Accidente		
Columna	Tipo de dato	Tipo de Clave
Accident_ID	Int	PK
Date	Date	
Time	Time	
ID_Lugar	Int	FK
ID_Vuelo	Int	FK
Fatalities	Int	
Fatalities Passangers	Int	
Fatalities Crew	Int	
Ground	Int	
Summary	Varchar	

Lugar		
Columna	Tipo de dato	Tipo de Clave
Location_ID	Int	PK
Name	Varchar	
State	Varchar	
Country	Varchar	

Vuelo		
Columna	Tipo de dato	Tipo de Clave
Flight_ID	Int	PK
Operator_ID	Int	FK
Route_ID	Int	FK
Flight #	Int	
Plane_ID	Int	FK
Aboard	Int	
Aboard Passangers	Int	
Aboard Crew	Int	

Operator

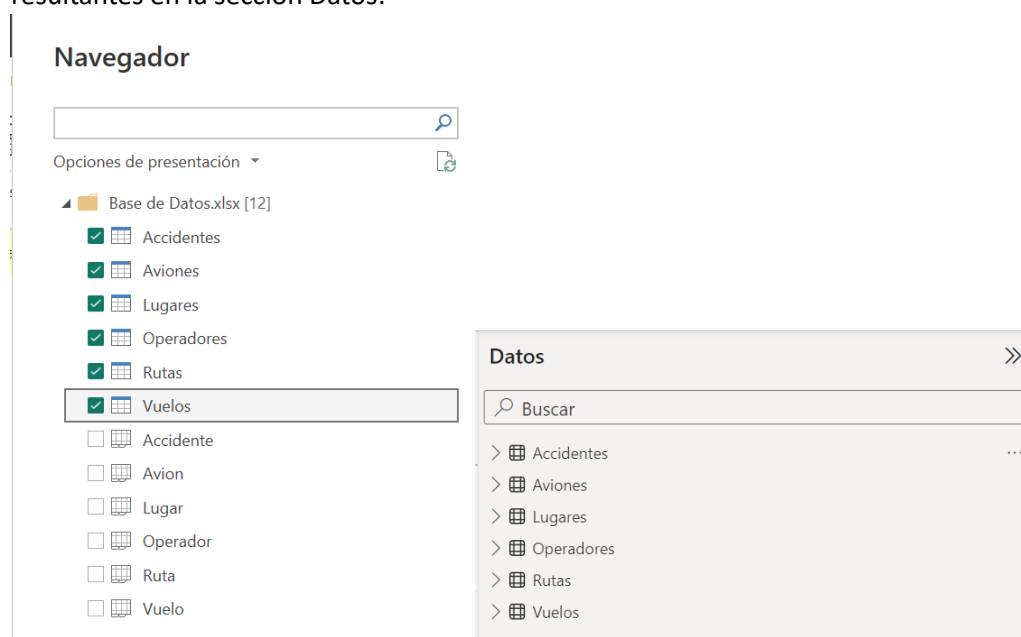
Columna	Tipo de dato	Tipo de Clave
Operator_ID	Int	PK
Operator_Name	Varchar	

Ruta		
Columna	Tipo de dato	Tipo de Clave
Route_ID	Int	PK
Route	Varchar	

Avion		
Columna	Tipo de dato	Tipo de Clave
Plane_ID	Int	PK
AC_Type	Varchar	

Carga de Datos en Power BI y ajustes iniciales

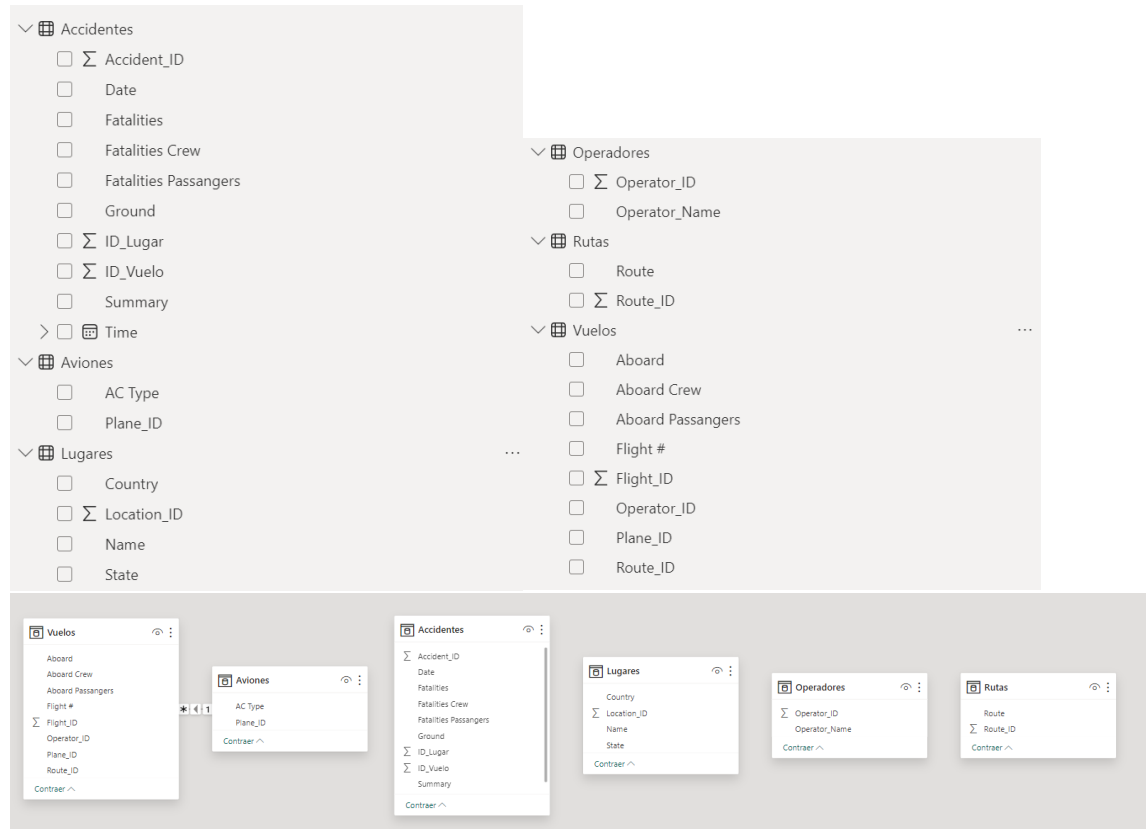
Antes de cargar los datos en Power BI se vuelve a ingresar a la base de datos (archivo "Base de Datos.xls") para formatear los datos de cada una de las páginas como tablas para que estas puedan ser detectadas por Power BI. Habiendo realizado esto se procede a crear el archivo que contiene el informe de Power BI llamado "Análisis Accidentes Aereos.pbix" y se importan los datos de la base de datos. En las siguientes imágenes se observa cómo se seleccionan las tablas del archivo Excel y las tablas resultantes en la sección Datos:



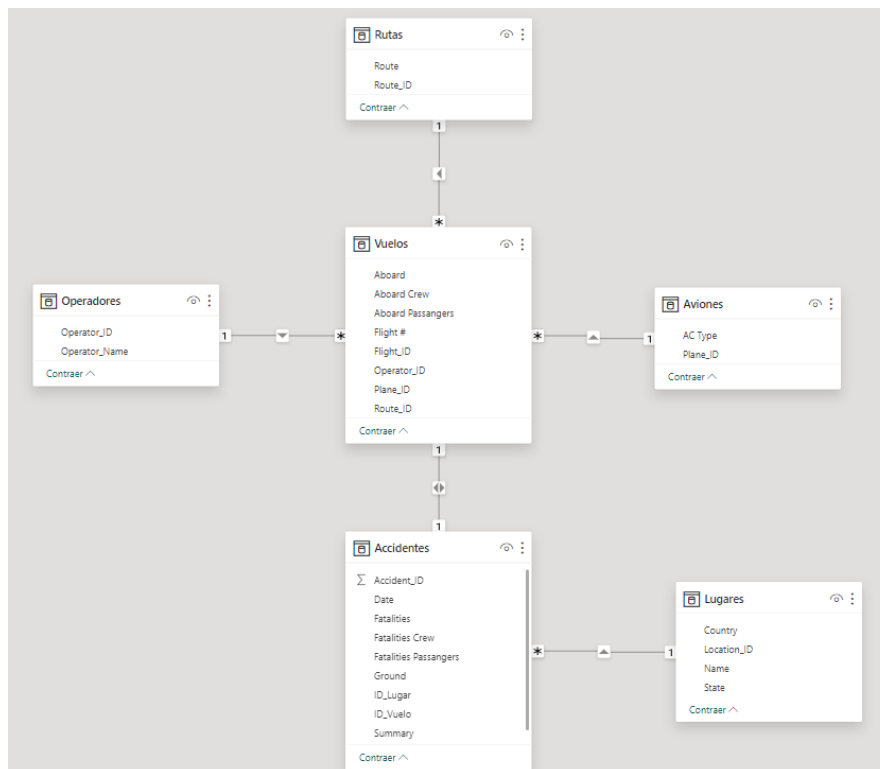
Inicialmente las relaciones no fueron correctamente debido a que existen relaciones a través de claves foráneas donde los nombres de las columnas no son exactamente las

mismas, por ejemplo, en la tabla Accidentes se tiene la columna ID_Lugar y en la tabla Lugares se tiene la columna Location_ID.

En las siguientes imágenes se observa el estado inicial de las diferentes columnas y del modelo relacional.



Debido a la inconsistencia de los nombres, como arreglo temporal se ajustan manualmente las relaciones, quedando el modelo relacional de la siguiente forma:



Como todas las relaciones creadas son del tipo uno es a muchos (1:N) no es necesario crear ninguna tabla puente.

Transformación de Datos

A continuación, se describen las diferentes transformaciones realizadas a los datos de las diferentes tablas de la base de datos:

a. Tabla Accidentes

- i. Se cambia el tipo de la columna *Accident_ID* a número entero
- ii. Se cambia el tipo de la columna *Date* a tipo Fecha con configuración regional para formato de fecha de estados unidos. De esta manera se interpreta bajo el formato Mes-Día-Año (MM-DD-YYYY).
- iii. Se cambian el tipo de las columnas *ID_Lugar*, *ID_Vuelo*, *Fatalities*, *Fatalities Passengers*, *Fatalities Crew* y *Ground* a número entero.
- iv. Se cambia el valor de *Ground* para el accidente con ID 4330 a 0. Esto se debe a que este accidente junto al accidente con ID 1329 corresponden a los dos accidentes aéreos del atentado a las torres gemelas en 2001. La cantidad de fallecidos en tierra para ambos registros de accidente están duplicadas por lo cual para limpiar los datos es necesario realizar esta transformación. Para realizar esta transformación se crea una nueva columna temporal con la expresión:

```
= Table.AddColumn("#Tipo cambiado1", "temp", each if [Accident_ID] = 4330 then 0 else [Ground])
```

Con esta expresión la columna *temp* tiene los mismos elementos de *Ground*, pero el registro con *Accident_ID* igual a 4330 se le cambia el valor a 0. Luego se elimina la columna *Ground* original y se le cambia el nombre de la columna *temp* a *Ground*. Además, a la nueva columna *Ground* se le cambia el tipo a número entero.

- v. Se elimina la columna *Time*, ya que estos datos no serán usados en el tablero.
- vi. Se cambia el nombre de todas las columnas a nombres en español y más intuitivos.

b. Tabla Aviones

- i. Se cambia el tipo de la columna *Plane_ID* a número entero y el tipo de la columna *AC Type* a texto.
- ii. Se aplica la funcionalidad de limpiar texto en la columna *AC Type*.
- iii. Se cambia el nombre de todas las columnas a nombres en español y más intuitivos.

c. Tabla Lugares

- i. Se cambia el tipo de la columna *Location_ID* a número entero y el tipo de las columnas *Name*, *State* y *Country* a texto.
- ii. Se aplica la funcionalidad de limpiar texto en las columnas *Name*, *State* y *Country*.
- iii. Se cambia el nombre de todas las columnas a nombres en español y más intuitivos.

d. Tabla Operadores

- i. Se cambia el tipo de la columna *Operator_ID* a número entero y el tipo de la columna *Operator_Name* a texto.
- ii. Se aplica la funcionalidad de limpiar texto en la columna *Operator_Name*.
- iii. Se cambia el nombre de todas las columnas a nombres en español y más intuitivos.

e. Tabla Rutas

- i. Se cambia el tipo de la columna *Route_ID* a número entero y el tipo de la columna *Route* a texto.
- ii. Se aplica la funcionalidad de limpiar texto en la columna *Route*.
- iii. Se cambia el nombre de todas las columnas a nombres en español y más intuitivos.

f. Tabla Vuelos

- i. Se cambia el tipo de la columna *Flight #* a texto y el tipo del resto de las columnas a número entero.
- ii. Se elimina la columna *Flight #* ya que estos datos no serán utilizados en el análisis.
- iii. Se cambia el nombre de todas las columnas a nombres en español y más intuitivos.

Creación de tabla calendario

Para crear la tabla calendario se crea una tabla calculada utilizando la siguiente expresión, con la cual se crean las fechas de todos los días desde el 17 de septiembre de 1908 hasta el 29 de octubre de 2023:

```
Calendario = CALENDARAUTO()
```

Luego se crean columnas calculadas para extraer el día, el mes, el año y el día de semana de cada una de las fechas utilizando las siguientes expresiones:

```
Day = DAY(Calendario[Date])
```

```
Month = MONTH(Calendario[Date])
```

```
Year = YEAR(Calendario[Date])
```

```
Week_Day = WEEKDAY(Calendario[Date])
```

Además, para este tablero se requiere tener no solo las fechas en formato numérico, sino en también en formato de texto. Las siguientes expresiones crean columnas calculadas para extraer el mes, y el día de semanas en formato de texto.

```
Month_Text = FORMAT(DATE(2000, Calendario[Month], 1), "MMMM")
```

```
Week_Day_Text = FORMAT(Calendario[Week_Day], "dddd")
```

Columnas Calculadas

Se crea una columna calculada adicional a las creadas en la tabla calendario, esta es la columna *Total Fallecidos* la cual representa la cantidad total de fallecidos de un accidente de avión sumando los fallecidos en el avión y los fallecidos en tierra. Esta columna se calcula utilizando la siguiente expresión.

```
Total Fallecidos = Accidentes[Fallecidos a Bordo] + Accidentes[Fallecidos en Tierra]
```

Medidas Calculadas

En esta sección se describen las diferentes medidas calculadas para ser utilizadas en el tablero. Cada una de estas medidas se encuentran en la columna calculada creada bajo el nombre Medidas

1) Max_Fallecidos_Accidente:

Esta medida representa el accidente que más fallecidos ha generado en la historia, la cual se calcula utilizando la siguiente expresión:

```
Max_Fallecidos_Accidente = MAX(Accidentes[Total Fallecidos])
```

2) Promedio_Fallecidos_Por_Accidente:

Esta medida representa el promedio de fallecidos por accidente, la cual se calcula utilizando la siguiente expresión:

```
Promedio_Fallecidos_Por_Accidente = AVERAGE(Accidentes[Total Fallecidos])
```

3) Total_Fallecidos:

Esta medida representa el número total de fallecidos provocados por accidentes aéreos en toda la historia entre 1908 y 2023. Esta medida se calcula utilizando la siguiente expresión:

```
Total_Fallecidos = SUM(Accidentes[Total Fallecidos])
```

4) Total_Modelos:

Esta medida representa el total de diferentes modelos de avión que han sido parte de la historia de los accidentes aéreos. Esta medida se calcula utilizando la siguiente expresión:

```
Total_Modelos = DISTINCTCOUNTNOBLANK(Aviones[Modelo de Avión])
```

5) Total_Operadores:

Esta medida representa el total de diferentes operadores que han sufrido accidentes aéreos. Esta medida se calcula utilizando la siguiente expresión:

```
Total_Operadores = DISTINCTCOUNTNOBLANK(Operadores[Nombre Operador])
```

6) Total_Paises:

Esta medida representa la cantidad total de países (bajo esta categoría se encuentran también lugares como mares y océanos) en los cuales han ocurrido accidentes aéreos. Esta medida se calcula utilizando la siguiente expresión:

```
Total_Paises = DISTINCTCOUNT(Lugares[País])
```

7) Total_Rutas:

Esta medida representa la cantidad total de rutas en las cuales han ocurrido accidentes aéreos.

```
Total_Rutas = DISTINCTCOUNTNOBLANK(Rutas[Ruta])
```

8) Fallecidos en Tierra:

Esta medida representa la cantidad total de fallecidos en tierra por accidentes aéreos.

```
Fallecidos en Tierra = SUM(Accidentes[Fallecidos en Tierra])
```

9) Pasajeros Fallecidos:

Esta medida representa la cantidad de total de pasajeros fallecidos por accidentes aéreos.

Pasajeros Fallecidos = $\text{SUM}(\text{Accidentes}[\text{Pasajeros Fallecidos}])$

10) Total Sobrevivientes:

Esta medida representa la cantidad de personas a bordo que sobrevivieron en accidentes aéreos.

Total Sobrevivientes = $[\text{Total Gente a Bordo}] - [\text{Total Fallecidos a Bordo}]$

11) Taza de Supervivencia:

Esta medida representa la razón entre el número total de sobrevivientes con el número total de gente a bordo de los aviones accidentados.

Taza de Supervivencia = $[\text{Total Sobrevivientes}] / [\text{Total Gente a Bordo}]$

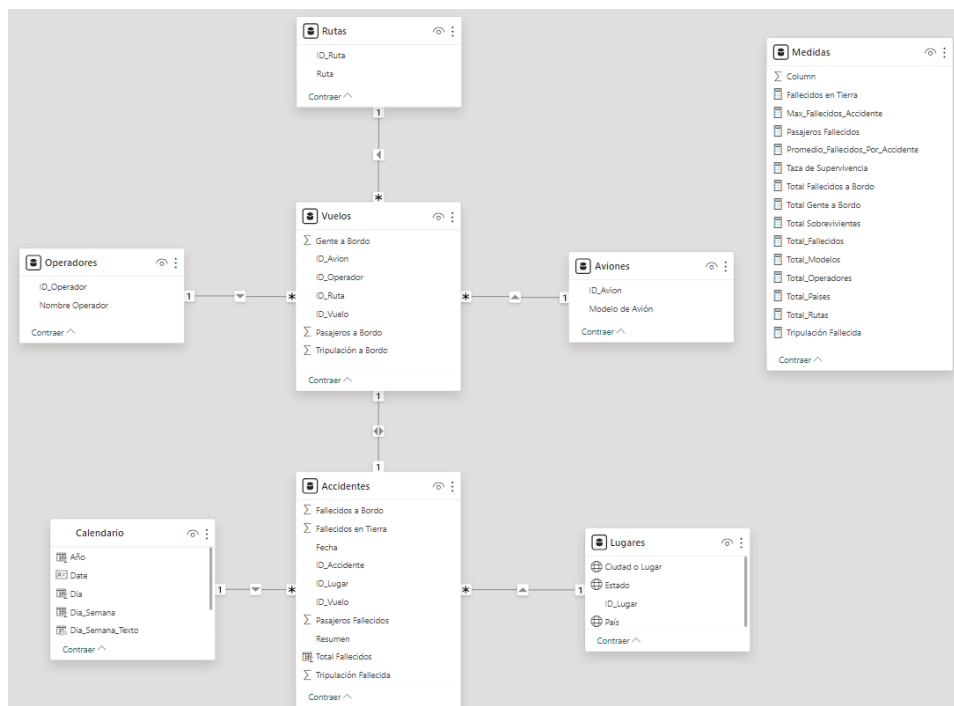
12) Tripulación Fallecida:

Esta medida representa la cantidad total de miembros de tripulación fallecidos en accidentes aéreos.

Tripulación Fallecida = $\text{SUM}(\text{Accidentes}[\text{Tripulación Fallecida}])$

Modelo de datos en Power BI

Una vez realizadas todas las transformaciones y creadas las tablas calculadas, columnas calculadas y medidas, el modelo de datos relaciones queda de la siguiente forma:



Visualizaciones Creadas

En esta sección se presentan las diferentes solapas creadas en Power BI. Cabe destacar que en las diferentes visualizaciones se tienen los nombres y descripciones de ejes por defecto. Cambiar estos nombres y el estilo de estas visualizaciones se encuentran entre las tareas pendientes a realizar para la versión final del tablero.

Solapa Portada:

Esta solapa contiene 5 botones para acceder a cada una de las solapas.



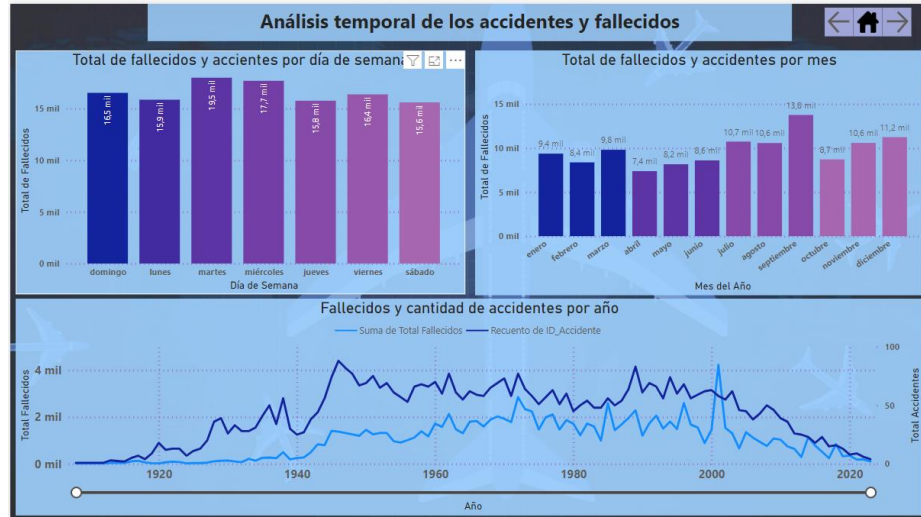
Solapa Resumen:

Contiene indicadores generales del set de datos en visualizaciones de formato tarjeta. Además, posee un gráfico de torta de la proporción de pasajeros fallecidos, tripulación fallecida y fallecidos en tierra.



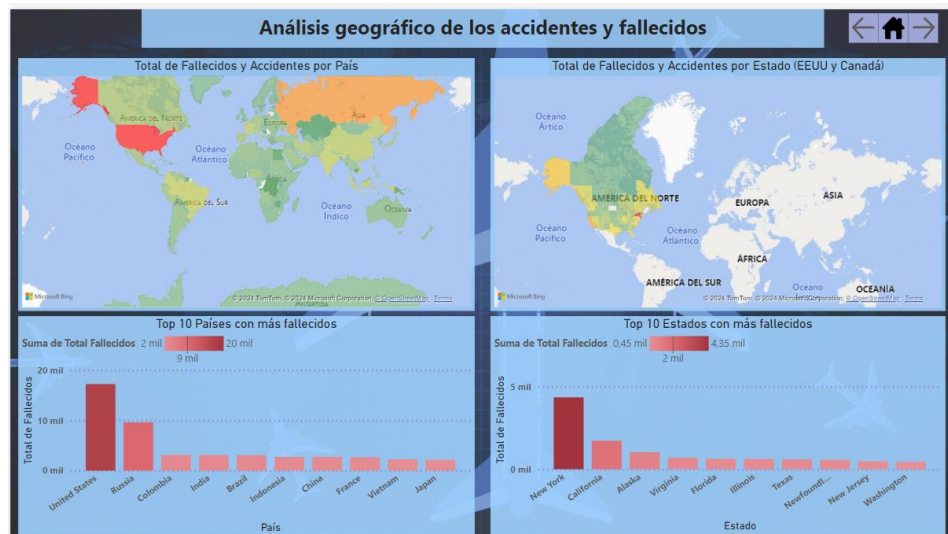
Solapa Análisis Temporal:

Esta solapa corresponde a la relación entre el tiempo con el número de accidentes y fallecidos. Contiene un gráfico de líneas que muestra la cantidad de accidentes y fallecidos en cada año y dos gráficos de columnas que indican la cantidad total de fallecidos según día de semana y mes del año respectivamente. Si se coloca el cursor sobre cualquier elemento de ambos gráficos de barras se puede observar igualmente la cantidad de accidentes según la categoría.



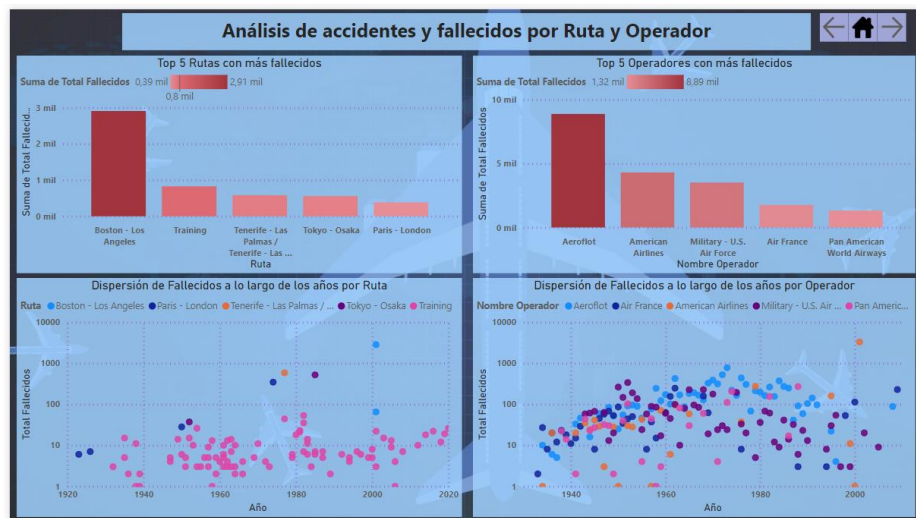
Solapa Análisis Lugar:

Esta solapa contiene un análisis geográfico de los accidentes, el cual contiene dos mapas, el primero contiene la distribución de accidentes y fallecidos por país y el segundo mapa ilustra la distribución por estados de Estados Unidos y Canadá. Además, esta solapa contiene dos gráficos de columnas de los top 10 países con mayor cantidad total de fallecidos y los top 10 estados con mayor cantidad de fallecidos. Si se coloca el cursor sobre cualquier elemento de ambos gráficos de barras o cualquier zona de los mapas se puede observar igualmente la cantidad de accidentes según lugar.



Solapa Análisis Ruta/Operador:

Esta solapa consiste en un análisis de la cantidad de fallecidos y accidentes según ruta u operador de vuelo. La solapa contiene dos gráficos de columnas, el primero es el top 5 de rutas con mayor cantidad de fallecidos por accidentes y la segunda es el top 5 de operadores con mayor cantidad de fallecidos por accidentes. Además, tanto para el análisis de ruta y operador se tiene un gráfico de dispersión de la cantidad de fallecidos por año por ruta/operador. En cualquiera de las visualizaciones se puede observar la cantidad de accidentes colocando el cursor sobre los elementos visuales.



Solapa Detalles Accidentes:

Esta solapa permite al usuario buscar información en detalle sobre algún accidente en concreto, para facilitar la búsqueda en la parte superior izquierda de la solapa el usuario puede filtrar los datos de las tablas utilizando un rango de fechas. Además, los indicadores de la esquina superior derecha indican la cantidad de accidentes y total de fallecidos en ese rango de tiempo.



Conclusiones

En base a las observaciones realizadas en el tablero creado se realizaron los siguientes insights :

- El atentado al world trade center en New York, Estados Unidos, el día 11 de septiembre de 2001, es el accidente aéreo en el cual ha fallecido m
- Estados unidos es el país que registra mayor cantidad de accidentes y fallecidos. Incluso si se descuenta la cantidad de fallecidos del atentado al world trade center, este sigue siendo el país que más fallecidos registra.
- Los meses de julio, agosto y septiembre, es decir, los meses del tercer trimestre del año son los que registran mayor cantidad de fallecidos por accidentes aéreos.
- Entre los años 1940 y 1946 se registra un aumento significativo en la cantidad de accidentes y fallecidos por año. Esto puede atribuirse al crecimiento de la industria aeronáutica tanto civil como militar en la época de la segunda guerra mundial. Esto en conjunto con la falta de regulación y normativa de seguridad en la época dieron origen a múltiples accidentes aéreos.
- Aeroflot, es el operador aéreo que registra mayor cantidad de fallecidos, con una suma total de 8893 fallecidos en 255 accidentes, los cuales se han registrado entre los años 1934 y 2008.
- Posterior al atentado al world trade center hasta la fecha, la cantidad de accidentes y fallecidos por accidente presentan una significativa tendencia a disminuir. Esto demuestra que las medidas de seguridad en aviación que aplicaron los países posterior al atentado han sido efectivas en disminuir la cantidad de accidentes.
- La ruta Boston – Los Angeles es la ruta que registra mayor cantidad de fallecidos, sin embargo, no se puede considerar esto como una tendencia, ya que registran 2907 fallecidos en dos accidentes, los cuales corresponden al día del atentado del world trade center. En segundo lugar, de mayores fallecidos por ruta se encuentran los vuelos de entrenamiento “training”, en el cual se registran 827 fallecidos en 93 accidentes. Además, en el gráfico de dispersión se puede observar que la cantidad de accidentes en vuelos de entrenamiento son mucho menos frecuentes hoy en día, en relación con los años 60 u 80. Esto sugiere que las regulaciones existentes para entrenamiento de pilotos han sido efectivas para prevenir accidentes

Como observación general se puede concluir que las regulaciones existentes en el mundo aeronáutico han contribuido de gran manera a prevenir muchas posibles tragedias. Considerando que la cantidad de vuelos comerciales existentes por día es de aproximadamente 120.000 vuelos, el hecho que el año 2023 se hayan registrado solamente 4 accidentes en todo el año confirmar el hecho de que los viajes en avión son el medio de transporte más seguro que existe.

Futuras Líneas

El análisis de accidentes aéreos puede ser aún más profundizado, en una futura versión de este tablero podrían añadirse los siguientes detalles que enriquecerían la información que podemos obtener de los datos:

- Añadir un gráfico de torta visualizar la proporción de accidentes ocurridos en tierra en relación con los accidentes ocurridos en agua.
- Añadir al análisis temporal los horarios más habituales en los accidentes.
- Añadir visualización con los top 5 modelos de avión que más accidentes y fallecidos registran.
- Buscar datos sobre la cantidad de vuelos a nivel mundial por año para determinar la relación entre la cantidad de vuelos y accidentes por año.
- Agregar al análisis temporal indicadores de eventos importantes, como las guerras mundiales, invención de modelos nuevos de aviones, entre otros.