**Proyecto de Análisis de Datos**

**Nombre del Proyecto**: Aeropuerto 2000  
**Fecha de Actualización**: 01/11/2024

**1. Descripción del Proyecto:** Este proyecto tiene como objetivo crear un informe de ciencia de datos en Power BI para analizar y visualizar información clave sobre las operaciones de Aeropuertos Argentina 2000 en el período comprendido entre 2018 y 2024. El análisis abarcará datos sobre el tráfico aéreo comercial, tanto doméstico como internacional, que representa la mayoría de los vuelos en Argentina. También se examinarán aspectos de la infraestructura, los servicios ofrecidos y el compromiso de la empresa con la modernización, sostenibilidad y calidad en la experiencia del pasajero. El propósito del informe es proporcionar una visión integral de la evolución de las operaciones aeroportuarias a lo largo de estos años, aportando valor mediante la obtención, limpieza, exploración, enriquecimiento y modelado de los datos, los cuales se presentarán en tableros interactivos con Power BI**.**

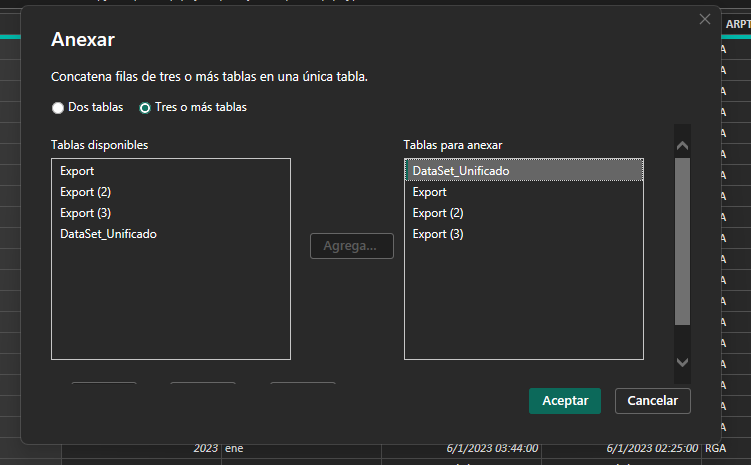
**2. Set de Datos**

* **Nombre del Dataset**: RGA 2023.xlsm, RGA 2024.xlsx, Datos RGA-marzo-2024.xlsx
* **Fuente de los Datos**: Archivo con extensión.xlsx)
* **Descripción de los Campos**: Diccionario de datos

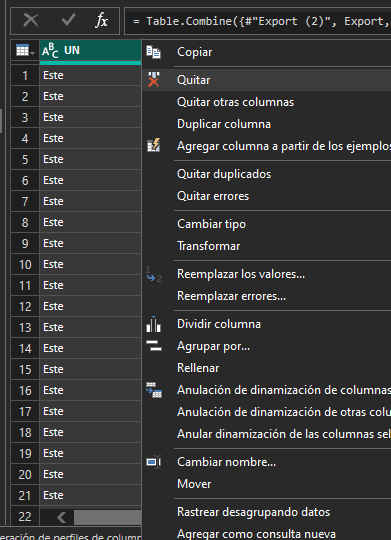
|  |  |  |  |
| --- | --- | --- | --- |
| **Columnas** | **Tipo de valor** | **Tamaño** | **Descripción** |
|  |
| ID\_OP\_FLIGHT | Entero | 7 | ID de registro de vuelos |  |
| AÑO | Entero | 4 | Años de la fecha de cuando se voló |  |
| UN | Caracter | 4 | Filtro de consulta anual |  |
| MES | Caracter | 3 | Mes de la fecha de cuando se voló |  |
| FECHA Y HORA REAL | DateTime | 19 | Fecha de cuando se voló |  |
| FECHA Y HORA PROG | DateTime | 19 | Fecha programación |  |
| ARPT | Carácter | 3 | Aeropuerto origen (código IATA de 3 a 4 caracteres) |  |
| ID LLAA | Carácter | 3 | ID aerolínea (código IATA de la aerolínea) |  |
| LLAA | Carácter | 50 | Descripción de aerolínea |  |
| NRO. VUELO | Entero | 4 | Número de vuelo |  |
| STAND | Entero | 1 | Consultas al Cliente |  |
| MOV TIPO | Carácter | 10 | Tipo de vuelo (arribo/partida) |  |
| ARPT1 | Carácter | 4 | Aeropuerto destino (código IATA de 3 a 4 caracteres) |  |
| PAIS | Carácter | 9 | Nacional |  |
| VUELO REGION | Carácter | 20 | Regionalizad del vuelo |  |
| VUELO TIPO | Carácter | 30 | Tipo de vuelo (comercial, aviación general, taxi aéreo, etc.) |  |
| VUELO CLASE | Carácter | 20 | Clase de vuelo |  |
| GATE | Carácter | 5 | Consultas al Cliente |  |
| VUELOS | Entero | 1 | Solo un valor “1” |  |
| PAX TOTALES | Entero | 4 | Pasajeros totales |  |
| PAX | Entero | 4 | Pasajeros que pagan |  |
| PAX TOB | Entero | 4 | Pasajero en tránsito (no se baja del avión en la conexión) |  |
| PAX TOC | Entero | 4 | Pasajero en transferencia (se baja del avión en la conexión) |  |
| PAX TSC | Entero | 4 | Pasajero en transferencia (se baja del avión en la conexión) |  |
| PAX INFOAS | Entero | 4 | Pasajeros infantes |  |
| PAX EXENTOS | Entero | 4 | Los Diplomáticos (es decir, los que no pagan tasa) |  |
| MATRÍCULA | Carácter | 10 | Matrícula del avión |  |
| ID ACFT | Carácter | 10 | ID equipo/modelo de aeronave |  |
| ACFT DESC | Carácter | 50 | Descripción equipo/modelo de aeronave |  |
| BODY | Carácter | 2 | Consultas al Cliente |  |
| ASIENTOS TOTALES | Entero | 4 | Asientos totales |  |
| ASIENTOS BSN | Entero | 4 | Asientos búshines |  |
| ASIENTOS ECY | Entero | 4 | Asientos economy |  |
| ASIENTOS 1ST | Entero | 4 | Asientos First class |  |
| FACTOR OCUPACIÓN | Decimal | 5,2 | % de asientos ocupados respecto a la cantidad de asientos |  |

**3. Procesos ETL**

Anexar tres dataset



Quitar columnas innecesarias



Tratamientos de valores NULL

Captura de pantalla de un celular

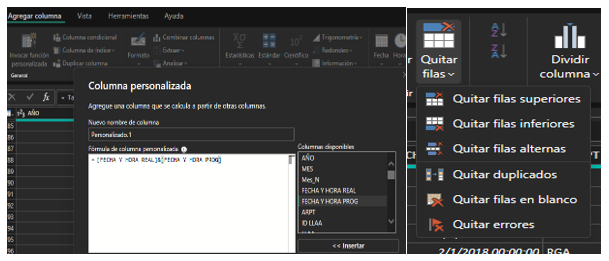
Descripción generada automáticamente

Tratamiento de valores con errores

Imagen de la pantalla de un computador

Descripción generada automáticamente con confianza baja

Eliminar datos duplicados

****

**3.1. Proceso de Extracción**

* **Fuente y Método de Extracción**: Conexión con libro de Excel.
* **Herramientas Utilizadas**: Power BI.
* **Frecuencia de Extracción**: Se realizaron 3 conexión con archivos de Excel para anexarlos y convertir un solo dataset.

**3.2. Proceso de Transformación**

**3.2.1. Limpieza de Datos**

* **Columnas Eliminadas**: Se quitaron de la estructura del dataset las columnas.
  + **UN:** Esta columna tenia valores que representaban a filtros para obtener valores del periodo anual. Valor único “ESTE”.
  + **ID\_OP\_FLIGHT:** Esta columna solo aparecía en un solo dataset y contenía valores del ID de la base de datos donde se almacenada la información.
* **Filas Filtradas**: En el conjunto de datos se identificaron 494 filas que presentaron errores cargados en todas las columnas.
* **Tratamiento de Valores Nulos**: Para las filas que presentaron valores nulos, se llevó a cabo análisis de las diferentes columnas con cálculos. Se decidió reemplazar estos valores nulos por cero, ya que este era el único valor que no afectaba las columnas con cálculos. Esta estrategia se implementó con el fin de mantener la integridad de los análisis, y se dejará pendiente para una revisión adicional para consultar con el cliente.
* **3.2.2. Transformación de Datos**
* **Modificación de Fechas:** En las columnas que contenían fechas, se realizó la conversión al tipo de valor DateTime. La columna que solo incluía el año (números) se transformó al tipo de valor Entero, mientras que la columna correspondiente al mes (3 carácter) se cambió al tipo de valor texto
* **Creación de Nuevas Columnas**: Se generaron las siguientes columnas.
  + Mes\_N: Columna de tipo Entero, que enumerar los meses del año.
  + Latitud: Columna de tipo Decimal, coordenadas para mapa.
  + Longiud: Columna de tipo Decimal, coordenadas para mapa.
  + Aeropuertos: Columna de tipo Texto, ubicación completa del aeropuerto.
  + AeropuertoLeyenda: Columna de tipo Texto, nombre del aeropuerto.
  + Ciudad: Columna de tipo Texto, lugar de ubicación del aeropuerto.
  + Personalizado: Columna de tipo Texto, fechas concatenadas para eliminación de datos duplicados.
* **Estandarización de Valores**: Los tres conjuntos de datos presentaban estructuras diferentes, lo que llevó a la creación de una estructura única que incorporara todas las columnas necesarias para conservar toda la información. En lo que respecta a los valores de textos, estaban correctamente categorizados, por lo que no fue necesario aplicar ningún tratamiento adicional, tal como “Normalización de datos”, “Corrección de tipográfica”, “Unificación de terminologías”, etc**.**
* **3.3. Proceso de Carga**
* **Destino de los Datos Transformados**: Los conjuntos de datos fueron procesados en Power BI utilizando la herramienta Power Query, lo que resultó en un archivo final llamado Dataset\_Unificado.pbix.
* **Formato de Exportación**: PowerBI.
* **Frecuencia de Carga**: No aplica.

**4. Integración de Procesos**

* **Resumen de Integración**: Describa cómo se consolidaron los datos de los diferentes procesos y los criterios de unificación.
* **Nombre del ETL Final**:
* **Ubicación del Archivo Final**:

**5. Herramientas y Estructura de Archivos**

**5.1. Herramientas Utilizadas**

* **Librerías/Software**: Power BI, Cookie Cutter, Git
* **Propósito**:
  + **Power BI**: El proceso ETL realizado en Power BI ha integrado y transformado tres conjuntos de datos sobre vuelos, regiones y pasajeros en un único dataset, optimizando su análisis y visualización. Se han limpiado y estructurado los datos, eliminando información irrelevante y gestionando valores nulos para garantizar la confiabilidad del dataset. Se añadieron columnas adicionales, como latitud y longitud para el análisis geoespacial y una columna de ordenación mensual para observar tendencias, nombre de aeropuertos y ubicaciones. El resultado es un dataset final de 7,972 registros y 38 columnas, limpio y funcional, ideal para la exploración de datos y la toma de decisiones estratégicas.
  + **CooKieCutter**: Estamos utilizando la herramienta para estructurar de manera eficiente los archivos que diseñamos, facilitando la creación de plantillas de proyectos estandarizadas. Esto nos permite establecer una organización coherente de directorios y archivos desde el inicio, acelerando el desarrollo y asegurando el cumplimiento de las mejores prácticas.
  + **Git**: Utilizamos la herramienta como sistema de control de versiones para gestionar y rastrear los cambios en nuestro código y documentos a lo largo del tiempo. Nos permite colaborar de manera efectiva entre el equipo, facilitando la incorporación de aportes y el seguimiento de las modificaciones realizadas en cada documentos y archivos. Además, nos ayudar a dejar trazabilidad, ya que proporciona un historial completo de cambios, lo que facilita la reversión a versiones anteriores en caso de ser necesario.

**5.2. Organización de Archivos**

* **Carpeta de Datos Crudos**: Nombre y descripción de los archivos de datos originales.
* **Carpeta de Procesos**: Incluir subcarpetas y nombres de archivos relevantes, como scripts de ETL, reportes trimestrales, etc.
* **Visualización**: Archivos o enlaces relacionados con visualización (e.g., reportes de Power BI).

**6. Visualización de Datos**

**6.1. Objetivo de la Visualización**

* **Resumen del Propósito**: Defina el objetivo de las visualizaciones y su audiencia (e.g., clientes, analistas).

**6.2. Visualizaciones Implementadas**

* **Descripción**: Liste cada tipo de visualización, su estructura, y la información que muestra.
  + *Tarjeta de Métrica 1*: (e.g., Cantidad de Trozas, Promedio de Largo).
  + *Gráficos*: Describa brevemente los gráficos y el motivo de su elección.

**6.3. Segmentación de Datos**

* **Campos Clave**: Detalle los campos que permiten la segmentación (e.g., Año, Calidad, Tipo de producto).

**7. Observaciones Finales y Recomendaciones**

* **Resumen de Resultados**: Incluir observaciones clave y hallazgos.
* **Recomendaciones para Mejora de Datos**: Propuestas para mejorar la calidad del proceso de ETL y la consistencia de datos.
* **Desafíos y Soluciones**: Liste los problemas encontrados y cómo se resolvieron.

**8. Otras modificaciones: dejo ejemplo**

Incluya cualquier información adicional relevante (e.g., código de scripts, referencias a documentación externa, diagramas de flujo).

* **Modificaciones durante el desarrollo del Sprint 2**
* **Cambio de Roles**
* **Si el proyecto también incluye el uso de una API, puedes agregar una sección específica para documentar su uso en el proceso ETL.**