Universidad de San Carlos de Guatemala
Facultad de Ingeniería
Seminario de Sistemas 2
Sección A

Edgar Mauricio Gómez Flores
2011-14340
28/02/2025

Introducción

El objetivo de este proyecto es desarrollar un proceso ETL (Extract, Transform, Load) que procese información de vuelos para un modelo de inteligencia de negocios. La aplicación se ejecuta por consola, en el lenguaje Python, y se apoya en SQL Server para la capa de almacenamiento y consultas analíticas.

Para correr la aplicación se requiere:

- SQL Server Express
- SQL Management Studio (opcional)
- Python 3.10 o superior

Se utilizó `pipreqs` para generar el archivo `requirements.txt` (`pipreqs –encoding=iso-8859-1`) y poder instalar las dependencias utilizando `pip install -r requirements.txt`.

Arquitectura General de la Solución

La solución se basa en varios módulos de Python que trabajan conjuntamente para ofrecer una aplicación en consola. A nivel conceptual, cada módulo cumple una función dentro del proceso ETL:

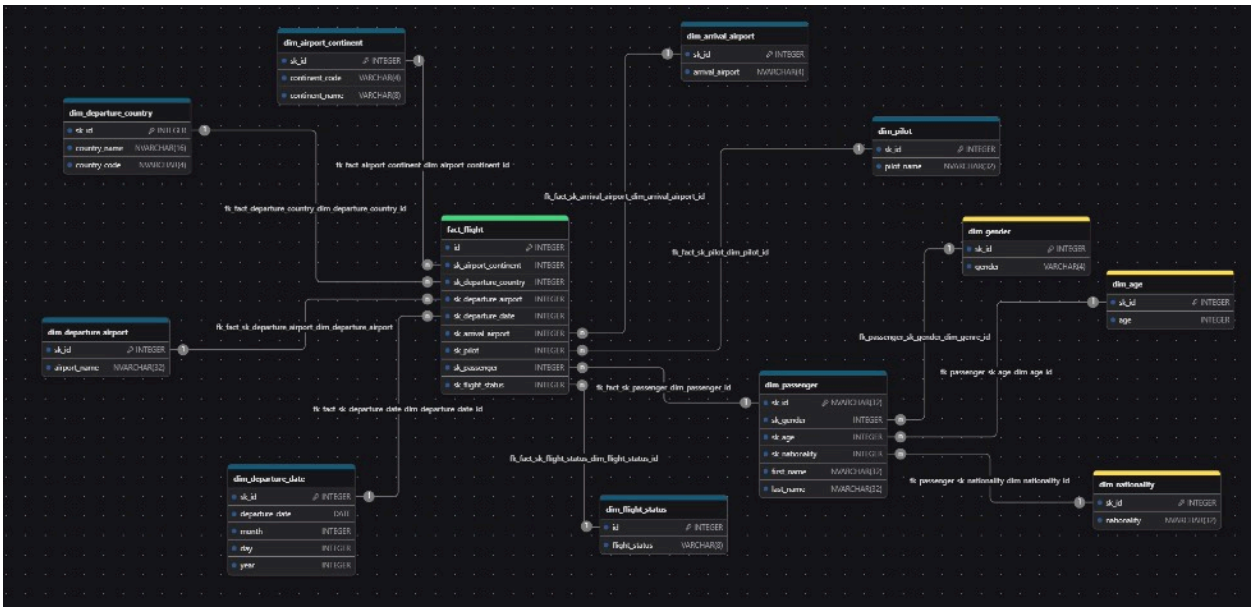
- *database.py*
  Contiene los datos necesarios para gestionar la conexión a la base de datos.
- *extract.py*
  Implementa la lógica para la extracción de datos desde el archivo CSV.
- *transformation.py*
  Se encarga de transformar y limpiar la información extraída utilizando

Edgar Mauricio Gómez Flores
2011-14340
28/02/2025

dataframes de Pandas. Aplica transformaciones a las fechas para estandarizarlas.

- *load.py*
  Se encarga de cargar los datos transformados en la base de datos. Invoca métodos para insertar y utiliza la libreta tqdm para mostrar barras de progreso.
- *model.py*
  Define los queries para crear o destruir el modelo de base de datos.
- querys.py
  Contiene las consultas en SQL que extraen información del modelo.
- ETL.py
  Integra todo el proceso en un flujo (extract → transform → load) en un menu de consola amigable.

Esquema de la base de datos



Universidad de San Carlos de Guatemala
Facultad de Ingeniería
Seminario de Sistemas 2
Sección A

Edgar Mauricio Gómez Flores
2011-14340
28/02/2025

Descripción del Esquema de Base de Datos (flights_db.sql)

La base de datos denominada "flights_db" sigue un modelo relacional simplificado según lo dispuesto en la práctica. Aunque la definición exacta de las tablas aparece en el script "flights_db.sql", a continuación se describe la lógica general típica en un esquema de vuelos:

Tabla dimDate (Dimensión de fechas):
Contiene atributos como fecha_id, dia, mes, anio, entre otros.

Tabla dimAirport (Dimensión de aeropuertos):
Guarde información relevante de aeropuertos: código, nombre, país, ciudad, continente, etc.

Tabla dimPassenger (Dimensión de pasajeros):
Datos demográficos del pasajero (género, nacionalidad, edad, etc.).

Tabla dimFlight (Dimensión de vuelos):
Información del número de vuelo, aerolínea, origen, destino, etc.

Tabla factFlight (Hechos de vuelos):
Vincula llaves foráneas a las dimensiones anteriores.
Incluye métricas como la cantidad de pasajeros, estado del vuelo, etc.

Uso de la Aplicación en Consola

La aplicación ofrece un menú interactivo que se despliega en la terminal que luce de la siguiente forma:

Universidad de San Carlos de Guatemala
Facultad de Ingeniería
Seminario de Sistemas 2
Sección A

Edgar Mauricio Gómez Flores
2011-14340
28/02/2025

```
1. Borrar Modelo
2. Crear Modelo
3. Extraer datos
4. Transformar datos
5. Cargar datos
6. Realizar consultas
7. Salir

Opción:
```

A continuación, se describe cada opción:

Borrar modelo
Ejecuta sentencias para eliminar las tablas del esquema (DROP TABLE ...) o el esquema completo. Se recomienda usar esta opción con cuidado, ya que borra toda la información.

Crear modelo
Ejecuta el script detallado en flights_db.sql para ejecutar las sentencias CREATE TABLE y definir las relaciones.

Extraer datos
Invoca a las funciones en extract.py para tomar los datos desde el archivo origen.

Transformar datos
Llama a los métodos de transformation.py para limpiar y homogeneizar la información.

Universidad de San Carlos de Guatemala
Facultad de Ingeniería
Seminario de Sistemas 2
Sección A

Edgar Mauricio Gómez Flores
2011-14340
28/02/2025

Cargar datos
Hace uso de load.py para almacenar los registros transformados en las tablas respectivas dentro de la base de datos de seminario.

Realizar consultas
Utiliza querys.py para realizar consultas.

Salir
Termina la ejecución de la aplicación.

Consultas de Validación y Ejemplo de Resultados

- Datos en todas las tablas

```
Entradas en continentes de aeropuerto:  6

Entradas en paises de salida:  235

Entradas en aeropuerto de salida:  9062

Entradas fecha de salida:  364

Entradas aeropuerto destino:  9024

Entradas pilotos:  98605

Entradas genero:  2

Entradas edad:  90

Entradas nacionalidades:  240

Entradas pasajero:  98617

Entradas estado de vuelo:  3

Entradas hecho vuelo:  98619
```

Universidad de San Carlos de Guatemala
Facultad de Ingeniería
Seminario de Sistemas 2
Sección A

Edgar Mauricio Gómez Flores
2011-14340
28/02/2025

- Porcentaje de pasajeros por género

```
Opción: 2
mujeres: 49.7064460
hombres: 50.293554
```

- Nacionalidades con su mes año de mayor fecha de salida

fecha	nationality	1-2022	2-2022	3-2022	4-2022	5-2022	...	7-2022	8-2022	9-2022	10-2022	11-2022	12-2022
0	Afghanistan	30	31	38	35	31	...	33	34	28	32	34	29
1	Aland Islands	3	2	4	3	1	...	2	0	1	2	0	1
2	Albania	38	25	45	38	43	...	37	32	37	38	29	40
3	Algeria	2	0	0	0	1	...	0	0	0	0	0	0
4	American Samoa	1	1	4	2	3	...	1	2	1	4	1	0
..	...	...	...	...	...	...	...	...	...	...	...	...	...
235	Wallis and Futuna	2	2	0	2	0	...	0	0	1	0	0	1
236	Western Sahara	1	0	0	0	2	...	1	0	1	0	0	0
237	Yemen	37	35	25	40	32	...	30	34	22	42	29	33
238	Zambia	9	9	4	5	14	...	5	5	10	6	5	6
239	Zimbabwe	10	10	10	8	13	...	7	6	6	11	7	8

[240 rows x 13 columns]

- COUNT de vuelos por país

```
208.Thailand: 500
209.Timor-Leste: 84
210.Togo: 23
211.Tonga: 69
212.Trinidad and Tobago: 23
213.Tunisia: 105
214.Turkey: 704
215.Turkmenistan: 54
216.Turks and Caicos Islands: 68
217.Tuvalu: 16
218.Uganda: 154
219.Ukraine: 398
220.United Arab Emirates: 158
221.United Kingdom: 1371
222.United States: 22104
223.United States Minor Outlying Islands: 27
224.Uruguay: 163
225.Uzbekistan: 128
226.Vanuatu: 327
227.Venezuela, Bolivarian Republic of: 730
228.Viet Nam: 368
229.Virgin Islands, British: 31
230.Virgin Islands, U.S.: 52
231.Wallis and Futuna: 17
232.Western Sahara: 39
233.Yemen: 186
234.Zambia: 222
235.Zimbabwe: 148

press Enter to continue
```

Universidad de San Carlos de Guatemala
Facultad de Ingeniería
Seminario de Sistemas 2
Sección A

Edgar Mauricio Gómez Flores
2011-14340
28/02/2025

- Top 5 aeropuertos con mayor número de pasajeros

```
Opción: 5
C:\Users\Pat0\Documents\SS2_1S2025_2011
onnectable (engine/connection) or databa
 Please consider using SQLAlchemy.
  df = pd.read_sql_query(query, databas

 Aeropuerto mas transitado de salidas:

          airport_name  total
0    San Pedro Airport     43
1  Santa Maria Airport     38
2   Böblingen Flugfeld     36
3    Santa Ana Airport     35
4      Maliana airport     32
C:\Users\Pat0\Documents\SS2_1S2025_2011
onnectable (engine/connection) or databa
 Please consider using SQLAlchemy.
  df = pd.read_sql_query(query, databas

 Aeropuerto de destino mas transitados:

  arrival_airport  total
0             MOE     43
1             HTM     38
2             MNR     36
3             DSV     35
4             AYP     31

press Enter to continue
```

- COUNT dividido por estado de vuelo

	flight_status	Total
0	Cancelled	32942
1	Delayed	32831
2	On Time	32846

Universidad de San Carlos de Guatemala
Facultad de Ingeniería
Seminario de Sistemas 2
Sección A

Edgar Mauricio Gómez Flores
2011-14340
28/02/2025

- Top 5 de los países más visitados

```
        country_name  total
0      United States  22104
1          Australia   6370
2             Canada   5424
3             Brazil   4504
4  Papua New Guinea   4081
```

-

- Top 5 de los continentes más visitados

```
   continent_name continent_code  total
0  North America             NAM  32033
1           Asia              AS  18637
2        Oceania              OC  13866
3         Europe              EU  12335
4         Africa              AF  11030
```

- Top 5 de edades dividido por género que más viajan

```
Top edades de mujeres:
    age  gender  total
0   39  Female    616
1   72  Female    587
2   67  Female    585
3   57  Female    583
4   29  Female    580
C:\Users\Pat0\Documents
onnectable (engine/conn
 Please consider using
  df = pd.read_sql_quer

Top edades de hombres:
    age gender  total
0   89   Male    615
1   46   Male    604
2   24   Male    597
3   29   Male    590
4   27   Male    590
```

Universidad de San Carlos de Guatemala
Facultad de Ingeniería
Seminario de Sistemas 2
Sección A

Edgar Mauricio Gómez Flores
2011-14340
28/02/2025

- COUNT de vuelos por MM-YYYY

	month	year	total
0	1	2022	8416
1	2	2022	7653
2	3	2022	8431
3	4	2022	7959
4	5	2022	8496
5	6	2022	8128
6	7	2022	8451
7	8	2022	8544
8	9	2022	8149
9	10	2022	8415
10	11	2022	8053
11	12	2022	7924