In [2]:
```
import pandas as pd
from sqlalchemy import create_engine
import json
import sqlite3

airports=pd.read_csv('airports.csv')
```

In [3]:
```
# Data cleaning steps for CSV containing airport ID and location

# Remove duplicate rows based on the ID
airports = airports.drop_duplicates(subset='airport_id', keep='first')

# Drop rows with missing values
airports = airports.dropna()

# Strip leading and trailing whitespaces from all string columns
airports = airports.apply(lambda x: x.str.strip() if x.dtype == 'O' else x)

# Display the cleaned DataFrame
print("\nCleaned DataFrame:")
print(airports)

# Save the cleaned DataFrame back to a new CSV file
airports.to_csv('cleaned_airports.csv', index=False)
```

```
Cleaned DataFrame:
     airport_id          city state                                name
0         10165   Adak Island    AK                                Adak
1         10299     Anchorage    AK  Ted Stevens Anchorage International
2         10304         Aniak    AK                       Aniak Airport
3         10754        Barrow    AK     Wiley Post/Will Rogers Memorial
4         10551        Bethel    AK                       Bethel Airport
..          ...           ...   ...                                 ...
360       11233      Cheyenne    WY  Cheyenne Regional/Jerry Olson Field
361       11097          Cody    WY                Yellowstone Regional
362       11865      Gillette    WY            Gillette Campbell County
363       12441       Jackson    WY                        Jackson Hole
364       14543  Rock Springs    WY      Rock Springs Sweetwater County

[365 rows x 4 columns]
```

In [4]:
```
with open('airline_operator_risk.json', 'r') as json_file:
    data = json.load(json_file)

# Convert JSON data to DataFrame
df_airline = pd.json_normalize(data)

# Display  original DataFrame
print("Original DataFrame:")
print(df_airline)

# Data cleaning steps
# Drop rows with missing values
df_airline = df_airline.dropna()

# Strip leading and trailing whitespaces from string columns
df_airline = df_airline.apply(lambda x: x.str.strip() if x.dtype == 'O' else x)

# Perform an SQL transformation (example: filtering based on a condition)
conn_airline = sqlite3.connect(':memory:')  # Create an in-memory SQLite database
with conn_airline:
    df_airline.to_sql('airline_data', conn_airline, index=False, if_exists='replace')  #
```

```
# Example SQL query: Select rows where av_fleet_age is greater than 10
sql_query_airline = "SELECT * FROM airline_data WHERE av_fleet_age > 10"
df_filtered_airline = pd.read_sql(sql_query_airline, conn_airline)

# Display the cleaned DataFrame after the SQL transformation
print("\nCleaned DataFrame after SQL transformation:")
print(df_filtered_airline)

# Save the cleaned DataFrame back to a new JSON file
df_filtered_airline.to_json('cleaned_airline_data.json', orient='records', lines=True)
```

Original DataFrame:

	models	aircraft	is_iosa_certified	accidents_5y	fatalaccidents_5y
0	1	1	False	0	0
1	1	9	False	0	0
2	1	2	False	0	0
3	2	2	False	0	0
4	1	2	False	0	0
..	...	...	...	...	...
295	12	16	False	0	0
296	2	13	True	1	0
297	1	1	False	0	0
298	2	2	False	0	0
299	3	41	False	0	0

	operatorName	operatorCode	countryName
0	Advisors Aviation LLC	BKM	United States
1	Honda	HON	United States
2	Telford Aviation Inc	TEL	United States
3	National Center for Atmospheric Research	SIQ	United States
4	Pathfinder Aviation LLC	PAF	United States
..	...	...	...
295	Aeroways Inc	TKK	United States
296	Omni Air International	OAE	United States
297	Big Island Air	BIG	United States
298	Freedom Aviation	FAW	United States
299	Alpha Flying Inc - PlaneSense	CNS	United States

	countryCode	av_fleet_age	aircraft_over_25y	is_international	routes
0	USA	15.000000	0	False	NaN
1	USA	7.111111	0	False	NaN
2	USA	21.000000	1	False	NaN
3	USA	26.000000	1	False	NaN
4	USA	41.500000	2	False	NaN
..	...	...	...	...	...
295	USA	13.187500	4	False	NaN
296	USA	20.076923	3	True	8.0
297	USA	41.000000	1	False	NaN
298	USA	12.500000	0	False	NaN
299	USA	5.731707	0	False	NaN

	connections	destinations	annual_flights	annual_international_flights
0	NaN	NaN	NaN	NaN
1	NaN	NaN	NaN	NaN
2	NaN	NaN	NaN	NaN
3	NaN	NaN	NaN	NaN
4	NaN	NaN	NaN	NaN
..	...	...	...	...
295	NaN	NaN	NaN	NaN
296	4.0	6.0	202.0	202.0
297	NaN	NaN	NaN	NaN
298	NaN	NaN	NaN	NaN
299	NaN	NaN	NaN	NaN

[300 rows x 17 columns]

Cleaned DataFrame after SQL transformation:

	models	aircraft	is_iosa_certified	accidents_5y	fatalaccidents_5y
0	1	13	0	0	0
1	1	36	1	2	0
2	2	3	0	0	0
3	2	60	1	0	0
4	1	2	0	0	0
5	1	14	1	0	0
6	2	5	0	2	1

7	1	9	0	1	0
8	1	1	1	1	0
9	4	37	1	1	0
10	3	21	0	0	0
11	6	684	1	26	0
12	3	6	0	0	0
13	1	52	1	1	0
14	11	591	1	22	0
15	1	626	0	9	1
16	2	10	1	0	0
17	1	1	0	0	0
18	5	11	0	1	0
19	3	13	0	0	0
20	4	462	1	3	0
21	3	183	1	5	0
22	1	13	0	0	0
23	3	3	0	0	0
24	5	100	0	0	0
25	2	12	0	0	0
26	1	6	0	0	0
27	1	8	0	0	0
28	2	8	0	0	0
29	2	22	0	0	0
30	1	57	1	0	0
31	7	502	1	12	0
32	3	36	0	0	0
33	2	2	0	0	0
34	9	34	0	0	0
35	2	93	0	0	0
36	4	9	0	0	0
37	3	18	1	0	0
38	9	61	0	0	0
39	2	182	1	1	0
40	1	29	1	0	0
41	2	6	0	0	0
42	4	29	0	0	0
43	2	13	1	1	0

	operatorName	operatorCode	countryName	\
0	ACE Air Cargo	AER	United States	
1	CommutAir	UCA	United States	
2	Aloha Air Cargo	AAH	United States	
3	Atlas Air	GTI	United States	
4	Sky Lease Cargo	KYE	United States	
5	ABX Air	ABX	United States	
6	Air Sunshine	RSI	United States	
7	Lynden Air Cargo	LYC	United States	
8	GoJet Airlines	GJS	United States	
9	Kalitta Air	CKS	United States	
10	Kenmore Air	KEN	United States	
11	American Airlines	AAL	United States	
12	Iliamna Air Taxi	IAR	United States	
13	Piedmont Airlines	PDT	United States	
14	Delta Air Lines	DAL	United States	
15	Southwest Airlines	SWA	United States	
16	Polar Air Cargo	PAC	United States	
17	Innovation First Transportation Inc	BUG	United States	
18	Everts Air Alaska	VTS	United States	
19	IBC Airways	CSQ	United States	
20	SkyWest Airlines	SKW	United States	
21	JetBlue Airways	JBU	United States	
22	Warbelow's Air Ventures	WAV	United States	
23	Executive Express Aviation LLC	LTD	United States	
24	Cape Air	KAP	United States	

25	Taquan Air	TQN	United States
26	Eastern Airlines	DYA	United States
27	Amerijet International	AJT	United States
28	Northern Air Cargo	NAC	United States
29	Tradewind Aviation	GPD	United States
30	Air Wisconsin	AWI	United States
31	United Airlines	UAL	United States
32	Southern Airways Express	FDY	United States
33	40-Mile Air	MLA	United States
34	Bering Air	BRG	United States
35	Allegiant Air	AAY	United States
36	Ryan Air	RYA	United States
37	Silver Airways	SIL	United States
38	Era Helicopters LLC	ERH	United States
39	Republic Airways	RPA	United States
40	Sun Country Airlines	SCX	United States
41	Aztec Airways	AZY	United States
42	Grant Aviation	GUN	United States
43	Omni Air International	OAE	United States

	countryCode	av_fleet_age	aircraft_over_25y	is_international	routes	\
0	USA	30.923077	13	0	26.0	
1	USA	15.694444	0	0	76.0	
2	USA	24.333333	2	0	15.0	
3	USA	22.766667	25	1	21.0	
4	USA	16.500000	0	1	19.0	
5	USA	32.285714	14	1	48.0	
6	USA	36.600000	5	1	52.0	
7	USA	41.444444	9	0	5.0	
8	USA	17.000000	0	1	531.0	
9	USA	22.378378	17	1	8.0	
10	USA	59.714286	20	1	7.0	
11	USA	10.195906	0	1	1601.0	
12	USA	30.666667	3	0	3.0	
13	USA	19.500000	0	0	140.0	
14	USA	12.534687	38	1	1576.0	
15	USA	12.052716	0	1	1475.0	
16	USA	17.100000	1	1	47.0	
17	USA	32.000000	1	0	29.0	
18	USA	25.818182	5	0	25.0	
19	USA	30.538462	10	1	8.0	
20	USA	10.666667	0	1	1349.0	
21	USA	11.207650	0	1	520.0	
22	USA	41.076923	13	0	7.0	
23	USA	26.666667	1	0	61.0	
24	USA	32.480000	82	1	83.0	
25	USA	64.166667	12	0	3.0	
26	USA	31.666667	6	1	12.0	
27	USA	30.125000	8	1	145.0	
28	USA	28.375000	8	0	26.0	
29	USA	13.454545	0	1	22.0	
30	USA	17.894737	0	1	360.0	
31	USA	14.085657	33	1	2352.0	
32	USA	15.055556	8	1	18.0	
33	USA	28.500000	1	0	3.0	
34	USA	18.352941	13	0	15.0	
35	USA	13.763441	0	1	667.0	
36	USA	31.888889	7	0	9.0	
37	USA	11.888889	0	1	111.0	
38	USA	14.163934	9	0	36.0	
39	USA	10.131868	0	1	804.0	
40	USA	14.206897	0	1	94.0	
41	USA	39.333333	6	1	3.0	
42	USA	25.413793	18	0	13.0	

```
43          USA     20.076923                    3                    1     8.0

    connections  destinations  annual_flights  annual_international_flights
0          21.0          14.0          3883.0                           0.0
1          38.0          29.0         42487.0                           0.0
2           8.0           7.0         15053.0                           0.0
3          16.0          13.0          1416.0                        1416.0
4          19.0          18.0           544.0                         444.0
5          30.0          24.0         13426.0                        4809.0
6          28.0           9.0         26351.0                       26351.0
7           4.0           4.0           996.0                           0.0
8         271.0          89.0         90229.0                        5153.0
9           5.0           8.0           240.0                         210.0
10          4.0           4.0          8510.0                        2296.0
11        849.0         295.0       1134936.0                      213447.0
12          2.0           2.0           942.0                           0.0
13         70.0          60.0         94960.0                           0.0
14        827.0         289.0       1063049.0                      152257.0
15        738.0         100.0       1321055.0                       31875.0
16         37.0          20.0          8487.0                        6359.0
17         17.0          15.0         42984.0                           0.0
18         17.0          13.0          4713.0                           0.0
19          4.0           5.0           196.0                         196.0
20        689.0         207.0        642142.0                       38455.0
21        260.0         101.0        341393.0                       85214.0
22          5.0           4.0          8858.0                           0.0
23         32.0          19.0          6974.0                           0.0
24         43.0          31.0        113014.0                       37425.0
25          2.0           2.0          6206.0                           0.0
26          6.0           7.0          1677.0                        1677.0
27        118.0          54.0         10211.0                       10077.0
28         16.0          10.0          3656.0                           0.0
29         11.0          11.0         10481.0                        6838.0
30        180.0          77.0        147372.0                       12112.0
31       1390.0         331.0        716586.0                      175582.0
32          9.0          10.0          6385.0                          76.0
33          2.0           2.0           838.0                           0.0
34         10.0           6.0         26331.0                           0.0
35        334.0         106.0         75369.0                          34.0
36          5.0           5.0         16508.0                           0.0
37         57.0          40.0         38872.0                        9431.0
38         19.0          13.0         26826.0                           0.0
39        408.0         114.0        223493.0                       20121.0
40         52.0          38.0         19316.0                        3733.0
41          3.0           3.0           480.0                         480.0
42          7.0           7.0         55903.0                           0.0
43          4.0           6.0           202.0                         202.0
```

In [5]:

```python
with open('airline_incidents.json', 'r') as json_file:
    data = json.load(json_file)

# Convert JSON data to DataFrame
df_incidents = pd.json_normalize(data)

# Display the original DataFrame
print("Original DataFrame:")
print(df_incidents)

# Data cleaning steps
# Drop rows with missing values
df_incidents = df_incidents.dropna()

# Strip leading and trailing whitespaces from string columns
df_incidents = df_incidents.apply(lambda x: x.str.strip() if x.dtype == 'O' else x)
```

```
# Perform an SQL transformation (example: filtering based on a condition)
conn_incidents = sqlite3.connect(':memory:')  # Create an in-memory SQLite database
with conn_incidents:
    df_incidents.to_sql('incidents_data', conn_incidents, index=False, if_exists='replac

# Example SQL query: Select rows where FlightPhase is 'Approach'
sql_query_incidents = "SELECT * FROM incidents_data WHERE FlightPhase = 'Approach'"
df_filtered_incidents = pd.read_sql(sql_query_incidents, conn_incidents)

# Display the cleaned DataFrame after the SQL transformation
print("\nCleaned DataFrame after SQL transformation:")
print(df_filtered_incidents)

# Save the cleaned DataFrame back to a new JSON file
df_filtered_incidents.to_json('cleaned_incidents_data.json', orient='records', lines=Tru
```

Original DataFrame:

	Date	StateOfOccurrence	Location	Model	Registration	Operator	StateOfRegistry	FlightPhase	Class	Fatalities	...	TypeDesignator	Helicopter
0	2008-01-25T00:00:00.000Z	USA	Newark , New Jersey	AIRBUS A320	N462UA	United States United Air Lines Inc.	USA	Take-off	Serious incident	0	...	A320	False
1	2008-01-28T00:00:00.000Z	USA	Chicago-Midway Airport, IL (MDW)	DC-9-32	N605NW	Northwest Airlines	USA	Standing		0	...	TAYD	False
2	2008-02-18T00:00:00.000Z	USA	Austin	MCDONNELL DOUGLAS	N935DL	United States Delta Air Lines, Inc.	USA	Approach	Serious incident	0	...	DC93	False
3	2008-02-25T00:00:00.000Z	USA	Jackson Hole, WY	AIRBUS A320	N442UA	United States United Air Lines Inc.	USA	Landing	Incident	0	...	A320	False
4	2008-04-15T00:00:00.000Z	USA	Pueblo , Colorado	AIRBUS A320	N497UA	United States United Air Lines Inc.	USA	En route	Incident	0	...	A320	False
..	...	...	...	...	...	...	...	...	...	...	...	...	...
176	2016-07-13T00:00:00.000Z	COL	AD Bogota	AIRBUS A330 200	N279AV	Peru	USA	Take-off	Serious incident		...	BE20	False
177	2016-07-26T00:00:00.000Z	USA	Houston-Sugar Land Regional Airport, TX (SGR)	EMBRAER EMB505 PHENOM 300	N362FX	United States Flight Options (Johnson City, Ny)	USA	Landing	Incident	0	...	E55P	False
178	2016-08-11T00:00:00.000Z	USA	en croisiÃ¨re	AIRBUS A320	N632JB	United States	USA	En route	Serious incident		...	A320	False
179	2016-10-26T00:00:00.000Z	USA	Depuis AD Seatlle	AIRBUS A330 200	N855NW	United States	USA	Take-off	Serious incident		...	BE20	False
180	2016-10-28T00:00:00.000Z	USA	Fort Lauderdale International Airport, FL, FLL...	MCDONNELL DOUGLAS MD10 10	N370FE	United States Federal Express Corporation (Lit...	USA						

```
180     Landing  Serious incident             ...            DC10       False

     Airplane Engines EngineType  StateOfOperator  Official  Risk  \
0        True       2        Jet              USA             SCF
1        True       1     Piston              USA
2        True       2        Jet              USA              RS
3        True       2        Jet              USA              RS
4        True       2        Jet              USA            TURB
..        ...     ...        ...              ...       ...   ...
176      True       2  Turboprop              USA              RS
177      True       2        Jet              USA              RS
178      True       2        Jet              USA            TURB
179      True       2  Turboprop              USA             SCF
180      True       3        Jet              USA             SCF

              OccCats  Year
0            [SCF-NP]  2008
1                  []  2008
2              [BIRD]  2008
3                [RE]  2008
4              [TURB]  2008
..                ...   ...
176              [RE]  2016
177       [RE, LOC-G]  2016
178            [TURB]  2016
179    [F-NI, SCF-PP]  2016
180  [SCF-NP, F-POST]  2016

[181 rows x 24 columns]

Cleaned DataFrame after SQL transformation:
                       Date StateOfOccurrence  \
0   2008-02-18T00:00:00.000Z               USA
1   2008-10-19T00:00:00.000Z
2   2008-11-16T00:00:00.000Z               USA
3   2009-03-14T00:00:00.000Z
4   2009-04-19T00:00:00.000Z               USA
5   2009-06-30T00:00:00.000Z               USA
6   2009-07-11T00:00:00.000Z               USA
7   2009-09-02T00:00:00.000Z               USA
8   2009-09-13T00:00:00.000Z               BHS
9   2009-10-26T00:00:00.000Z
10  2009-12-25T00:00:00.000Z               USA
11  2010-01-10T00:00:00.000Z               USA
12  2011-07-23T00:00:00.000Z               BHS
13  2011-10-01T00:00:00.000Z               BHS
14  2013-05-08T00:00:00.000Z               DEU
15  2015-04-06T00:00:00.000Z               CRI

                                          Location                  Model  \
0                                           Austin      MCDONNELL DOUGLAS
1                                         Edmonton         EMBRAER ERJ170
2                   Philadelphia International Air  DE HAVILLAND DHC8 300
3                                        Vancouver            AIRBUS A319
4                                            Miami         BOEING 757 200
5                       KLGA - New York/La Guardia         BOEING 737 800
6                                       Los Angeles         BOEING 737 800
7                                  approach to BWI            AIRBUS A320
8                                             MYGF          ATR ATR72 200
9                                  75 km off Tokyo         BOEING 777 300
10  near Detroit-Metropolitan Wayne County Airport...       Airbus A330-323X
11                                        Newark            AIRBUS A319
12                                          MYNN         BOEING 737 800
13                                          MYGF          ATR ATR72 200
```

```
14                                    En route                    BAE
15                                                        BOEING 757 200

   Registration                                          Operator  \
0        N935DL                United States Delta Air Lines, Inc.
1        N642RW  United States Shuttle America (Windsor Locks, Ct)
2        N326EN  United States Piedmont Airlines, Inc. D/b/a He...
3        N836UA                United States United Air Lines Inc.
4        N690AA               United States American Airlines Inc.
5        N907AN               United States American Airlines Inc.
6       NAXXXXX               United States American Airlines Inc.
7        N456UA                United States United Air Lines Inc.
8        N431AT  United States Executive Airlines, Inc., D/b/a ...
9        N777AN               United States American Airlines Inc.
10       N820NW                                   Northwest Airlines
11       N816UA                United States United Air Lines Inc.
12       N905AN               United States American Airlines Inc.
13       N417AT  United States Amr American Eagle, Inc. (Dallas...
14       HB-IYS                   Switzerland Swiss World Airways
15       N978FD  United States Federal Express Corporation (Lit...

   StateOfRegistry FlightPhase             Class Fatalities  ...  \
0              USA    Approach  Serious incident       None  ...
1              USA    Approach          Incident       None  ...
2              USA    Approach          Incident       None  ...
3              USA    Approach          Incident       None  ...
4              USA    Approach          Incident       None  ...
5              USA    Approach          Incident       None  ...
6              USA    Approach          Incident       None  ...
7              USA    Approach          Incident       None  ...
8              USA    Approach          Incident       None  ...
9              USA    Approach          Incident       None  ...
10             USA    Approach                         None  ...
11             USA    Approach  Serious incident       None  ...
12             USA    Approach          Incident       None  ...
13             USA    Approach          Incident       None  ...
14             CHE    Approach  Serious incident       None  ...
15             USA    Approach          Incident       None  ...

    TypeDesignator  Helicopter  Airplane Engines EngineType  StateOfOperator  \
0             DC93           0         1       2        Jet              USA
1             E170           0         1       2        Jet              USA
2             VNOM           0         1       1        Jet              USA
3             A319           0         1       2        Jet              USA
4             ST75           0         1       1     Piston              USA
5             B37M           0         1       2        Jet              USA
6             B38M           0         1       2        Jet              USA
7             A320           0         1       2        Jet              USA
8             BE20           0         1       2  Turboprop              USA
9              C17           0         1       4        Jet              USA
10            A319           0         1       2        Jet              USA
11            A319           0         1       2        Jet              USA
12            B37M           0         1       2        Jet              USA
13            M200           0         1       1     Piston              USA
14              J5           0         1       1     Piston              USA
15            ST75           0         1       1     Piston              USA

    Official  Risk OccCats  Year
0              RS    None  2008
1             OTH    None  2008
2              RS    None  2008
3             SCF    None  2009
4             OTH    None  2009
5              RS    None  2009
```

```
6              OTH    None  2009
7              SCF    None  2009
8              SCF    None  2009
9             TURB    None  2009
10                    None  2009
11              RS    None  2010
12             SCF    None  2011
13           F-NI    None  2011
14             OTH    None  2013
15                    None  2015

[16 rows x 24 columns]
```

In [12]:

```python
with open('delay_index.json', 'r') as json_file:
    data = json.load(json_file)

# Flatten the nested JSON structure
df_delay_indexes = pd.json_normalize(data['delayIndexes'])

# Display the original DataFrame
print("Original DataFrame:")
print(df_delay_indexes)

# Data cleaning steps
# Drop rows with missing values
df_delay_indexes = df_delay_indexes.dropna()

# Strip leading and trailing whitespaces from string columns
df_delay_indexes = df_delay_indexes.apply(lambda x: x.str.strip() if x.dtype == 'O' else

# Display the cleaned DataFrame
print("\nCleaned DataFrame:")
print(df_delay_indexes)

# Save the cleaned DataFrame back to a new JSON file
df_delay_indexes.to_json('cleaned_delay_indexes.json', orient='records', lines=True)
```

```
Original DataFrame:
      rawScore  normalizedScore                dateStart  \
0         0.00              0.0  2024-02-17T17:30:00.000Z
1         1.20              0.5  2024-02-17T17:30:00.000Z
2         0.00              0.0  2024-02-17T17:30:00.000Z
3         0.00              0.0  2024-02-17T17:30:00.000Z
4         0.00              0.0  2024-02-17T17:30:00.000Z
...        ...              ...                       ...
1614      6.15              0.5  2024-02-19T22:00:00.000Z
1615      2.24              0.5  2024-02-19T22:00:00.000Z
1616      0.00              0.0  2024-02-19T22:00:00.000Z
1617      0.00              0.0  2024-02-19T22:00:00.000Z
1618      0.00              0.0  2024-02-19T22:00:00.000Z

                      dateEnd  flights  observations  canceled  onTime  \
0     2024-02-17T20:30:00.000Z        0             0         0       0
1     2024-02-17T20:30:00.000Z       10            10         0       9
2     2024-02-17T20:30:00.000Z        2             2         0       2
3     2024-02-17T20:30:00.000Z        2             0         0       0
4     2024-02-17T20:30:00.000Z        0             0         0       0
...                        ...      ...           ...       ...     ...
1614  2024-02-20T01:00:00.000Z       68            63         0      60
1615  2024-02-20T01:00:00.000Z       24            24         0      22
1616  2024-02-20T01:00:00.000Z        4             4         0       4
1617  2024-02-20T01:00:00.000Z        1             1         0       1
1618  2024-02-20T01:00:00.000Z        3             3         0       3

      delayed15  delayed30  ...  airport.utcOffsetHours  airport.latitude  \
0             0          0  ...                    -5.0         47.167778
1             0          0  ...                    -5.0         38.174397
2             0          0  ...                    -9.0         56.114041
3             0          0  ...                    -5.0         26.152778
4             0          0  ...                    -6.0         37.044444
...         ...        ...  ...                     ...               ...
1614          0          1  ...                    -6.0         44.883016
1615          1          1  ...                   -10.0         20.892883
1616          0          0  ...                    -6.0         32.454707
1617          0          0  ...                    -9.0         65.829167
1618          0          0  ...                    -9.0         64.736090

     airport.longitude airport.elevationFeet airport.classification  \
0           -88.486389                  1095                      4
1           -85.735990                   474                      1
2          -133.123301                     0                      4
3           -81.775278                     8                      4
4          -100.960556                  2883                      4
...                ...                   ...                    ...
1614        -93.210922                   841                      1
1615       -156.438629                    54                      2
1616        -93.828383                   258                      3
1617       -144.063889                   610                      4
1618       -156.936112                   128                      4

     airport.active                                airport.weatherUrl  \
0              True  https://api.flightstats.com/flex/weather/rest/...
1              True  https://api.flightstats.com/flex/weather/rest/...
2              True  https://api.flightstats.com/flex/weather/rest/...
3              True  https://api.flightstats.com/flex/weather/rest/...
4              True  https://api.flightstats.com/flex/weather/rest/...
...             ...                                               ...
1614           True  https://api.flightstats.com/flex/weather/rest/...
1615           True  https://api.flightstats.com/flex/weather/rest/...
1616           True  https://api.flightstats.com/flex/weather/rest/...
1617           True  https://api.flightstats.com/flex/weather/rest/...
```

```
1618          True  https://api.flightstats.com/flex/weather/rest/...

                     airport.street1 airport.postalCode airport.street2
0                                NaN                NaN             NaN
1                  600 Terminal Drive              40209             NaN
2                                NaN                NaN             NaN
3            160 Aviation Drive North          34104-3568             NaN
4                                NaN          67905-2199             NaN
...                              ...                ...             ...
1614           6040 28th Avenue South              55450             NaN
1615                                               96732
1616  5103 Hollywood Avenue, Suite 300              71109
1617                             NaN               99733             NaN
1618                             NaN               99741             NaN

[1619 rows x 36 columns]

Cleaned DataFrame:
      rawScore  normalizedScore                 dateStart  \
17        0.00             0.00  2024-02-17T17:30:00.000Z
22       13.64             1.25  2024-02-17T17:30:00.000Z
23        0.00             0.00  2024-02-17T17:30:00.000Z
24        0.00             0.00  2024-02-17T17:30:00.000Z
27        0.00             0.00  2024-02-17T17:30:00.000Z
...        ...              ...                       ...
1600      0.00             0.00  2024-02-19T22:00:00.000Z
1607      1.02             5.00  2024-02-19T22:00:00.000Z
1612      0.00             0.00  2024-02-19T22:00:00.000Z
1615      2.24             0.50  2024-02-19T22:00:00.000Z
1616      0.00             0.00  2024-02-19T22:00:00.000Z

                      dateEnd  flights  observations  canceled  onTime  \
17    2024-02-17T20:30:00.000Z        1             1         0       1
22    2024-02-17T20:30:00.000Z       59            58         0      49
23    2024-02-17T20:30:00.000Z        1             1         0       1
24    2024-02-17T20:30:00.000Z        2             2         0       2
27    2024-02-17T20:30:00.000Z        6             5         0       5
...                        ...      ...           ...       ...     ...
1600  2024-02-20T01:00:00.000Z        0             0         0       0
1607  2024-02-20T01:00:00.000Z        1             1         0       0
1612  2024-02-20T01:00:00.000Z        0             0         0       0
1615  2024-02-20T01:00:00.000Z       24            24         0      22
1616  2024-02-20T01:00:00.000Z        4             4         0       4

      delayed15  delayed30  ...  airport.utcOffsetHours  airport.latitude  \
17            0          0  ...                    -5.0         43.170556
22            5          0  ...                    -5.0         42.366460
23            0          0  ...                    -5.0         37.326944
24            0          0  ...                    -6.0         32.454707
27            0          0  ...                   -10.0         19.714565
...         ...        ...  ...                     ...               ...
1600          0          0  ...                    -6.0         40.967222
1607          0          0  ...                    -5.0         41.265278
1612          0          0  ...                    -6.0         32.386667
1615          1          1  ...                   -10.0         20.892883
1616          0          0  ...                    -6.0         32.454707

     airport.longitude airport.elevationFeet airport.classification  \
17          -86.237500                   628                      4
22          -71.020176                    19                      1
23          -79.201111                   938                      4
24          -93.828383                   258                      3
27         -155.039628                    38                      3
...                ...                   ...                    ...
```

```
1600        -98.306667                 1847                      4
1607        -72.888333                   14                      4
1612        -94.715278                  365                      4
1615       -156.438629                   54                      2
1616        -93.828383                  258                      3

     airport.active                                  airport.weatherUrl  \
17             True  https://api.flightstats.com/flex/weather/rest/...
22             True  https://api.flightstats.com/flex/weather/rest/...
23             True  https://api.flightstats.com/flex/weather/rest/...
24             True  https://api.flightstats.com/flex/weather/rest/...
27             True  https://api.flightstats.com/flex/weather/rest/...
...             ...                                               ...
1600           True  https://api.flightstats.com/flex/weather/rest/...
1607           True  https://api.flightstats.com/flex/weather/rest/...
1612           True  https://api.flightstats.com/flex/weather/rest/...
1615           True  https://api.flightstats.com/flex/weather/rest/...
1616           True  https://api.flightstats.com/flex/weather/rest/...

                     airport.street1 airport.postalCode airport.street2
17                  99 Sinclair Drive              49441
22               One Harborside Drive          02128-2909
23          4308 Wards Road, Suite 100              24502
24    5103 Hollywood Avenue, Suite 300              71109
27                                                  96720
...                                ...                ...             ...
1600               3743 Sky Park Road              68801
1607                  155 Burr Street              06512
1612                                               75603
1615                                               96732
1616  5103 Hollywood Avenue, Suite 300              71109

[336 rows x 36 columns]
```

In [8]:

```python
import json
import pandas as pd
import sqlite3

# Load data from the JSON file
with open('airline_accidents.json', 'r') as json_file:
    data = json.load(json_file)

# Convert JSON data to DataFrame
df_accidents = pd.json_normalize(data)

# Display the original DataFrame
print("Original DataFrame:")
print(df_accidents)

# Data cleaning steps
# Drop rows with missing values
df_accidents = df_accidents.dropna()

# Strip leading and trailing whitespaces from string columns
df_accidents = df_accidents.apply(lambda x: x.str.strip() if x.dtype == 'O' else x)

# Perform any additional data cleaning steps if needed

# Perform an SQL transformation (example: filtering based on a condition)
# Note: Replace 'YourColumnName' with the actual column you want to filter on
conn_accidents = sqlite3.connect(':memory:')  # Create an in-memory SQLite database
with conn_accidents:
    df_accidents.to_sql('accident_data', conn_accidents, index=False, if_exists='replace
```

```
# Example SQL query: Select rows where Fatalities is greater than or equal to 0
sql_query_accidents = "SELECT * FROM accident_data WHERE CAST(Fatalities AS INTEGER) >=
df_filtered_accidents = pd.read_sql(sql_query_accidents, conn_accidents)

# Display the cleaned DataFrame after the SQL transformation
print("\nCleaned DataFrame after SQL transformation:")
print(df_filtered_accidents)

# Save the cleaned DataFrame back to a new JSON file
df_filtered_accidents.to_json('cleaned_airline_accidents.json', orient='records', lines=
```

Original DataFrame:

	Date	StateOfOccurrence	\
0	2021-09-23T00:00:00.000Z	USA	
1	2021-12-12T00:00:00.000Z	USA	
2	2022-04-14T00:00:00.000Z	USA	
3	2022-02-15T00:00:00.000Z	MEX	
4	2021-09-17T00:00:00.000Z	USA	
5	2022-01-22T00:00:00.000Z	USA	
6	2021-09-27T00:00:00.000Z	USA	
7	2022-02-28T00:00:00.000Z	USA	
8	2021-05-28T00:00:00.000Z	USA	

	Location	\
0	Honolulu-Daniel K. Inouye International Airpo...	
1	Las Vegas-McCarran International Airport, NV ...	
2	Denver International Airport, CO (DEN)	
3	Mexico City-Benito Juï¿½rez International Air...	
4	Asheville Regional Airport, NC (AVL)	
5	Hayden-Yampa Valley Airport, CO (HDN)	
6	Denver International Airport, CO (DEN)	
7	Washington-Ronald Reagan National Airport, DC...	
8	Dallas/Fort Worth International Airport, TX (...	

	Model	Registration	\
0	Airbus A321-271N	N208HA	
1	Airbus A320-251N	N307FR	
2	Swearingen SA227-DC Metro 23	N820DC	
3	Airbus A319-114	N354NB	
4	Canadair CRJ-200LR	N444ZW	
5	Airbus A320-232	N760JB	
6	Boeing 757-224 (WL)	N12125	
7	Embraer ERJ 170-200 LR (ERJ-175LR)	N402YX	
8	Boeing 737-823 (WL)	N834NN	

	Operator	StateOfRegistry	FlightPhase	\
0	Hawaiian Airlines	USA	Landing	
1	Frontier Airlines	USA	Landing	
2	Denver Air Connection	USA	Take-off	
3	Delta Air Lines	USA	Take-off	
4	Air Wisconsin, opf. United Express	USA		
5	JetBlue Airways	USA	Take-off	
6	United Airlines	USA	Take-off	
7	Republic Airlines, opf. American Eagle	USA	Landing	
8	American Airlines	USA	Taxi	

	Class	Fatalities	...	TypeDesignator	Helicopter	Airplane	Engines	\
0	Accident		...	A321	False	True	2	
1	Accident	0	...	A320	False	True	2	
2	Accident	0	...	SW4	False	True	2	
3	Accident	0	...	A319	False	True	2	
4	Accident	0	...	BE20	False	True	2	
5	Accident	0	...	A320	False	True	2	
6	Accident		...	ST75	False	True	1	
7	Accident	0	...	E75S	False	True	2	
8	Accident	0	...	B37M	False	True	2	

	EngineType	StateOfOperator	Official	Risk	OccCats	Year
0	Jet	USA		RS	[]	2021
1	Jet	USA		RS	[]	2021
2	Turboprop	USA		RS	[]	2022
3	Jet	USA		RS	[]	2022
4	Turboprop	USA			[]	2021
5	Jet	USA		RS	[]	2022
6	Piston	USA		RS	[]	2021

```
7        Jet              USA              RS      []  2022
8        Jet              USA              RS      []  2021

[9 rows x 24 columns]

Cleaned DataFrame after SQL transformation:
                      Date StateOfOccurrence  \
0  2021-09-23T00:00:00.000Z              USA
1  2021-09-27T00:00:00.000Z              USA

                                          Location               Model  \
0  Honolulu-Daniel K. Inouye International Airpor...     Airbus A321-271N
1            Denver International Airport, CO (DEN)  Boeing 757-224 (WL)

  Registration           Operator StateOfRegistry FlightPhase     Class  \
0       N208HA  Hawaiian Airlines             USA     Landing  Accident
1       N12125    United Airlines             USA    Take-off  Accident

  Fatalities  ...  TypeDesignator  Helicopter  Airplane Engines EngineType  \
0             ...            A321           0         1       2        Jet
1             ...            ST75           0         1       1     Piston

   StateOfOperator  Official  Risk OccCats  Year
0              USA              RS    None  2021
1              USA              RS    None  2021

[2 rows x 24 columns]
```