KULLIYYAH OF INFORMATION & COMMUNICATION TECHNOLOGY

INFO 4311 DATA WAREHOUSING

SEMESTER 1, 2024/2025

SECTION 01, GROUP A

GROUP PROJECT: DATA PROFILING
Aviation Accident (2021 - 2023)

PREPARED BY:

NO.	NAME	MATRIC NO
1	MUHAMMAD IKMAL HAKIMI BIN ROSLI	2210827
2	AHMAD IMAN BIN TURIN	2123349
3	MUHAMMAD MUSLIHUDDIN BIN MUSTAFFAR	2213263
4	ARMAN NURI ANUAR	2119279

LECTURER

ASSOC. PROF. DR. LILI MARZIANA ABDULLAH

DUE

21st NOVEMBER 2024

TABLE OF CONTENTS

1.0 QUESTION 1

For the data set of your source, ANSWER the following questions (use Appendix A to fill up the answers for each data set):

a) How many rows are in each spreadsheet? (0.5 point)

b) Which column(s) can be the natural key? In other words which column or columns from the spreadsheet can act as a primary key of the data set? Example of the natural key in the "Orders" spreadsheet is OrderID. (0.5 point)

c) What does one row of each spreadsheet represent? For example, one row in the "Orders" spreadsheet represents an individual customer's order. (1 point)

Data Source	Number of Rows	Natural Key(s)	One row represents …
2021 Accident Table	9	Registration	Details of an accident
2022 Accident Table	6	Registration	Details of an accident
2023 Accident Table	4	Registration	Details of an accident

Appendix A

2.0 QUESTION 2

Once the spreadsheet has been imported into Microsoft SQL Server, USE SQL QUERIES to run a profile on your data set (for each column) following the statements below. Show the SQL queries and the results. You may compile the results for the columns in the table(s) and show them in any suitable and presentable manner. (6 points)

2.1 Value Distribution

Find the number of times each value in the column occurs (also show the value). This profile helps you identify problems in your data, such as incorrect numbers of distinct values in a column. For example, you profile a column that is supposed to contain states in Malaysia and discover more than 13 distinct values.

SQL QUERIES: 2021 Plane Accident Database

SQLQuery1.sql - L...BG5SU64\User (53))*

```sql
SELECT 'Date' AS ColumnName, CAST([Date] AS NVARCHAR(100)) AS Value, COUNT(*) AS Occurrences
FROM [PlaneAccidentDB].[dbo].[2021 Plane Accident]
GROUP BY [Date]

UNION ALL

SELECT 'Time' AS ColumnName, CAST([Time] AS NVARCHAR(100)) AS Value, COUNT(*) AS Occurrences
FROM [PlaneAccidentDB].[dbo].[2021 Plane Accident]
GROUP BY [Time]

UNION ALL

SELECT 'Location' AS ColumnName, CAST([Location] AS NVARCHAR(100)) AS Value, COUNT(*) AS Occurrences
FROM [PlaneAccidentDB].[dbo].[2021 Plane Accident]
GROUP BY [Location]

UNION ALL

SELECT 'Operator' AS ColumnName, CAST([Operator] AS NVARCHAR(100)) AS Value, COUNT(*) AS Occurrences
FROM [PlaneAccidentDB].[dbo].[2021 Plane Accident]
GROUP BY [Operator]

UNION ALL

SELECT 'Flight' AS ColumnName, CAST([Flight] AS NVARCHAR(100)) AS Value, COUNT(*) AS Occurrences
FROM [PlaneAccidentDB].[dbo].[2021 Plane Accident]
GROUP BY [Flight]

UNION ALL

SELECT 'Route' AS ColumnName, CAST([Route] AS NVARCHAR(100)) AS Value, COUNT(*) AS Occurrences
FROM [PlaneAccidentDB].[dbo].[2021 Plane Accident]
GROUP BY [Route]

UNION ALL

SELECT 'AC_Type' AS ColumnName, CAST([AC_Type] AS NVARCHAR(100)) AS Value, COUNT(*) AS Occurrences
FROM [PlaneAccidentDB].[dbo].[2021 Plane Accident]
```

100 %

Results Messages

Query executed successfully. LAPTOP-NBG5SU64\MSSQLSERVER... LAPTOP-NBG5SU64\User (53) PlaneAccidentDB 00:00:00 104 rows

SQLQuery1.sql - L...BG5SU64\User (53))*

```sql
SELECT 'AC_Type' AS ColumnName, CAST([AC_Type] AS NVARCHAR(100)) AS Value, COUNT(*) AS Occurrences
FROM [PlaneAccidentDB].[dbo].[2021 Plane Accident]
GROUP BY [AC_Type]

UNION ALL

SELECT 'Registration' AS ColumnName, CAST([Registration] AS NVARCHAR(100)) AS Value, COUNT(*) AS Occurrences
FROM [PlaneAccidentDB].[dbo].[2021 Plane Accident]
GROUP BY [Registration]

UNION ALL

SELECT 'Construction_no_line_no' AS ColumnName, CAST([Construction_no_line_no] AS NVARCHAR(100)) AS Value, COUNT(*) AS Occurrences
FROM [PlaneAccidentDB].[dbo].[2021 Plane Accident]
GROUP BY [Construction_no_line_no]

UNION ALL

SELECT 'Aboard' AS ColumnName, CAST([Aboard] AS NVARCHAR(100)) AS Value, COUNT(*) AS Occurrences
FROM [PlaneAccidentDB].[dbo].[2021 Plane Accident]
GROUP BY [Aboard]

UNION ALL

SELECT 'Fatalities' AS ColumnName, CAST([Fatalities] AS NVARCHAR(100)) AS Value, COUNT(*) AS Occurrences
FROM [PlaneAccidentDB].[dbo].[2021 Plane Accident]
GROUP BY [Fatalities]

UNION ALL

SELECT 'Ground' AS ColumnName, CAST([Ground] AS NVARCHAR(100)) AS Value, COUNT(*) AS Occurrences
FROM [PlaneAccidentDB].[dbo].[2021 Plane Accident]
GROUP BY [Ground]
UNION ALL
SELECT 'Summary' AS ColumnName, CAST([Summary] AS NVARCHAR(300)) AS Value, COUNT(*) AS Occurrences
FROM [PlaneAccidentDB].[dbo].[2021 Plane Accident]
GROUP BY [Summary];
```

100 %

Results Messages

Query executed successfully. LAPTOP-NBG5SU64\MSSQLSERVER... LAPTOP-NBG5SU64\User (53) PlaneAccidentDB 00:00:00 104 rows

RESULTS: 2021 Plane Accident Database

SQLQuery1.sql - L...BG5SU64\User (53))*

100 %

Results | Messages

	ColumnName	Value	Occurrences
1	Date	2021-01-09	1
2	Date	2021-03-02	1
3	Date	2021-03-28	1
4	Date	2021-05-21	1
5	Date	2021-06-10	1
6	Date	2021-07-04	1
7	Date	2021-07-06	1
8	Date	2021-09-12	1
9	Date	2021-12-27	1
10	Time	800	1
11	Time	1130	1
12	Time	1440	1
13	Time	1500	1
14	Time	1705	1
15	Time	1800	1
16	Time	1835	1
17	Time	1914	1
18	Time	2315	1
19	Location	El Cajon, California	1
20	Location	Kazachinskoye◆, Russia	1
21	Location	Near Butte, Alaska	1
22	Location	Near Jakarta, Indonesia	1
23	Location	Near Kaduna, Nigeria	1
24	Location	Near Pyin Oo Lwin, Myanmar	1
25	Location	Palana, Russia	1
26	Location	Patikul, Sulu, Philippines	1
27	Location	Pieri, Sudan	1
28	Operator	Aeroservice/SiLA	1
29	Operator	Kamchatka Aviation Enterprise	1
30	Operator	Med Jet	1
31	Operator	Military - Myanmar Air Force	1
32	Operator	Military - Nigerian Air Force	1
33	Operator	Military - Philippine Air Force	1
34	Operator	Soloy Helicopters◆	1
35	Operator	South Sudan Supreme Airlines◆	1
36	Operator	Sriwijaya Air	1
37	Flight	?	6
38	Flight	251	1
39	Flight	51	1
40	Flight	SJ182	1
41	Route	?	1
42	Route	Cagayan de Oro-Lumbia - Jolo	1
43	Route	Irkutsk - Kazachinskoye◆	1
44	Route	Jakarta - Pontianak	1

Query executed successfully.

SQLQuery1.sql - L...BG5SU64\User (53))*

100 %

Results | Messages

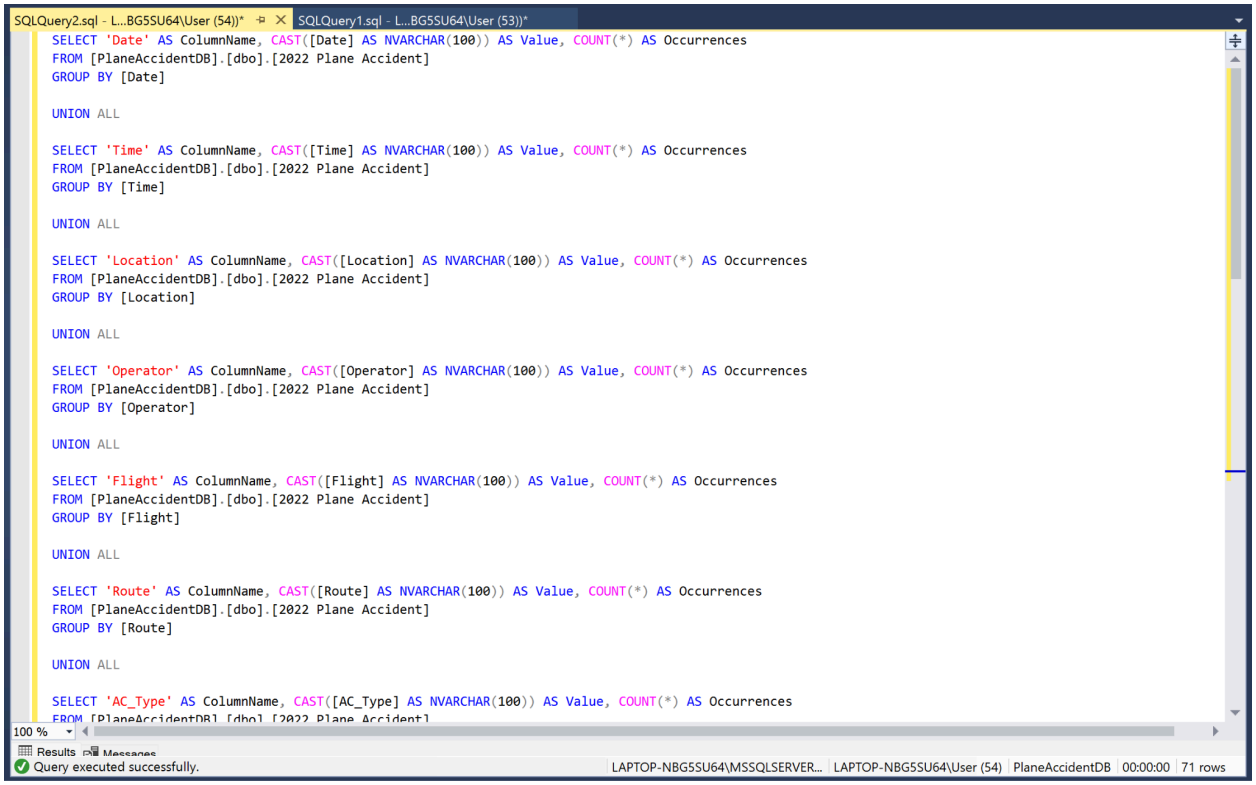
	ColumnName	Value	Occurrences
45	Route	Naypyidaw - Anisakan	1
46	Route	Petropavlovsk - Palana	1
47	Route	Pieri - Yuai	1
48	Route	Santa Ana - El Cajon	1
49	Route	Sightseeing Charter	1
50	AC_Type	Antonov An 26B-100	1
51	AC_Type	Beechcraft 1900D	1
52	AC_Type	Beechcraft B300 King Air 350i	1
53	AC_Type	Boeing 737-524	1
54	AC_Type	Eurocopter AS350B3◆Ecureuil	1
55	AC_Type	Learjet 35A	1
56	AC_Type	Let L-410UVP-E	1
57	AC_Type	Let L-410UVP-E20	1
58	AC_Type	Lockheed C-130H Hercules	1
59	Registration	4610	1
60	Registration	5125	1
61	Registration	HK-4274◆	1
62	Registration	N351SH	1
63	Registration	N880Z	1
64	Registration	NAF203	1
65	Registration	PK-CLC	1
66	Registration	RA-26085	1
67	Registration	RA-67042	1
68	Constructio...	12310	1
69	Constructio...	27323/2616	1
70	Constructio...	2916	1
71	Constructio...	35A-591	1
72	Constructio...	4598	1
73	Constructio...	5125	1
74	Constructio...	902525	1
75	Constructio...	E-325	1
76	Constructio...	FL-891	1
77	Aboard	4	1
78	Aboard	6	1
79	Aboard	10	1
80	Aboard	11	1
81	Aboard	14	1
82	Aboard	16	1
83	Aboard	28	1
84	Aboard	62	1
85	Aboard	96	1
86	Fatalities	4	2
87	Fatalities	5	1
88	Fatalities	10	1

Query executed successfully.

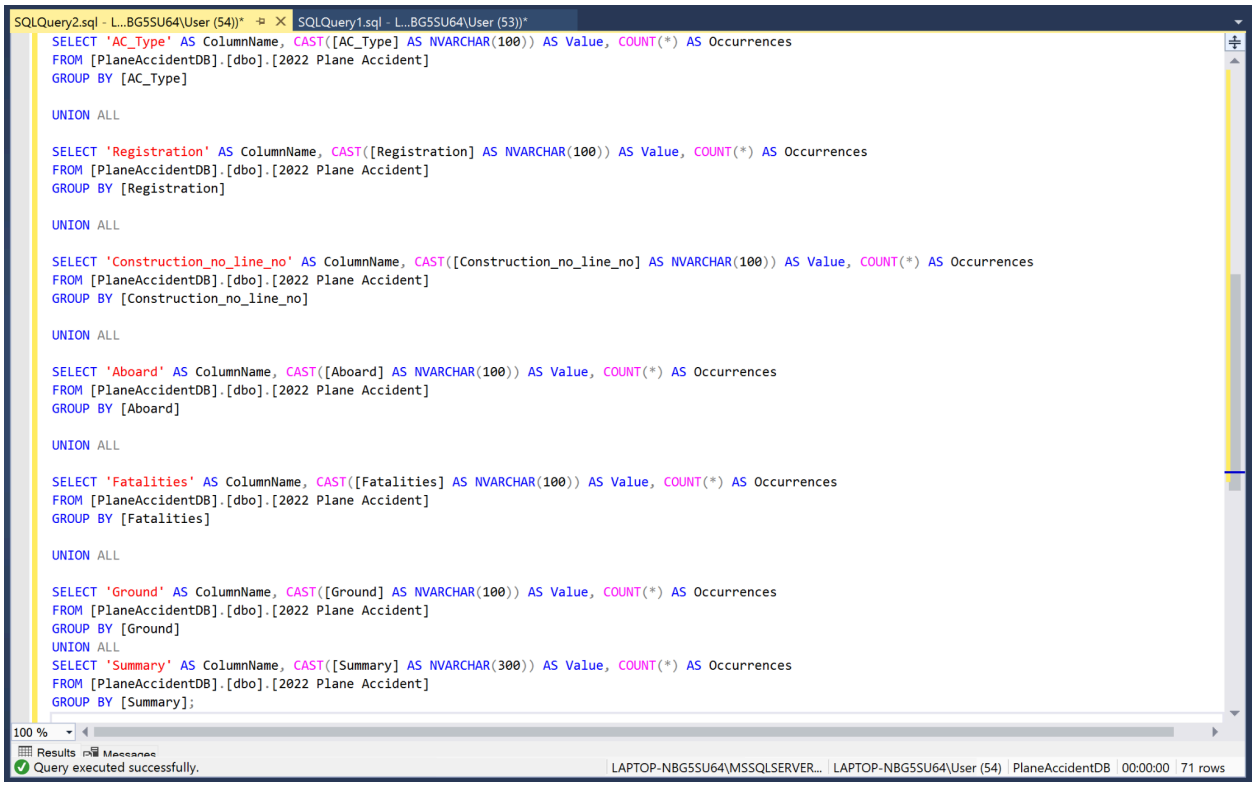
89	Fatalities	11	1
90	Fatalities	12	1
91	Fatalities	28	1
92	Fatalities	50	1
93	Fatalities	62	1
94	Ground	0	8
95	Ground	3	1
96	Summary	One of the engines on the airc...	1
97	Summary	Sriwijaya Air flight 182 was clim...	1
98	Summary	The air ambulance flight was c...	1
99	Summary	The aircraft was on it's second ...	1
100	Summary	The passenger plane crashed...	1
101	Summary	The plane was carrying militar...	1
102	Summary	The sightseeing helicopter cra...	1
103	Summary	While attempting to land at Jol...	1
104	Summary	While on final approach, in po...	1

Query executed successfully.

SQL QUERIES: 2022 Plane Accident Database



```sql
SELECT 'Date' AS ColumnName, CAST([Date] AS NVARCHAR(100)) AS Value, COUNT(*) AS Occurrences
FROM [PlaneAccidentDB].[dbo].[2022 Plane Accident]
GROUP BY [Date]

UNION ALL

SELECT 'Time' AS ColumnName, CAST([Time] AS NVARCHAR(100)) AS Value, COUNT(*) AS Occurrences
FROM [PlaneAccidentDB].[dbo].[2022 Plane Accident]
GROUP BY [Time]

UNION ALL

SELECT 'Location' AS ColumnName, CAST([Location] AS NVARCHAR(100)) AS Value, COUNT(*) AS Occurrences
FROM [PlaneAccidentDB].[dbo].[2022 Plane Accident]
GROUP BY [Location]

UNION ALL

SELECT 'Operator' AS ColumnName, CAST([Operator] AS NVARCHAR(100)) AS Value, COUNT(*) AS Occurrences
FROM [PlaneAccidentDB].[dbo].[2022 Plane Accident]
GROUP BY [Operator]

UNION ALL

SELECT 'Flight' AS ColumnName, CAST([Flight] AS NVARCHAR(100)) AS Value, COUNT(*) AS Occurrences
FROM [PlaneAccidentDB].[dbo].[2022 Plane Accident]
GROUP BY [Flight]

UNION ALL

SELECT 'Route' AS ColumnName, CAST([Route] AS NVARCHAR(100)) AS Value, COUNT(*) AS Occurrences
FROM [PlaneAccidentDB].[dbo].[2022 Plane Accident]
GROUP BY [Route]

UNION ALL

SELECT 'AC_Type' AS ColumnName, CAST([AC_Type] AS NVARCHAR(100)) AS Value, COUNT(*) AS Occurrences
FROM [PlaneAccidentDB].[dbo].[2022 Plane Accident]
```

```sql
SELECT 'AC_Type' AS ColumnName, CAST([AC_Type] AS NVARCHAR(100)) AS Value, COUNT(*) AS Occurrences
FROM [PlaneAccidentDB].[dbo].[2022 Plane Accident]
GROUP BY [AC_Type]

UNION ALL

SELECT 'Registration' AS ColumnName, CAST([Registration] AS NVARCHAR(100)) AS Value, COUNT(*) AS Occurrences
FROM [PlaneAccidentDB].[dbo].[2022 Plane Accident]
GROUP BY [Registration]

UNION ALL

SELECT 'Construction_no_line_no' AS ColumnName, CAST([Construction_no_line_no] AS NVARCHAR(100)) AS Value, COUNT(*) AS Occurrences
FROM [PlaneAccidentDB].[dbo].[2022 Plane Accident]
GROUP BY [Construction_no_line_no]

UNION ALL

SELECT 'Aboard' AS ColumnName, CAST([Aboard] AS NVARCHAR(100)) AS Value, COUNT(*) AS Occurrences
FROM [PlaneAccidentDB].[dbo].[2022 Plane Accident]
GROUP BY [Aboard]

UNION ALL

SELECT 'Fatalities' AS ColumnName, CAST([Fatalities] AS NVARCHAR(100)) AS Value, COUNT(*) AS Occurrences
FROM [PlaneAccidentDB].[dbo].[2022 Plane Accident]
GROUP BY [Fatalities]

UNION ALL

SELECT 'Ground' AS ColumnName, CAST([Ground] AS NVARCHAR(100)) AS Value, COUNT(*) AS Occurrences
FROM [PlaneAccidentDB].[dbo].[2022 Plane Accident]
GROUP BY [Ground]
UNION ALL
SELECT 'Summary' AS ColumnName, CAST([Summary] AS NVARCHAR(300)) AS Value, COUNT(*) AS Occurrences
FROM [PlaneAccidentDB].[dbo].[2022 Plane Accident]
GROUP BY [Summary];
```

RESULTS: 2022 Plane Accident Database

SQLQuery2.sql - L...BG5SU64\User (54))* | SQLQuery1.sql - L

100 %

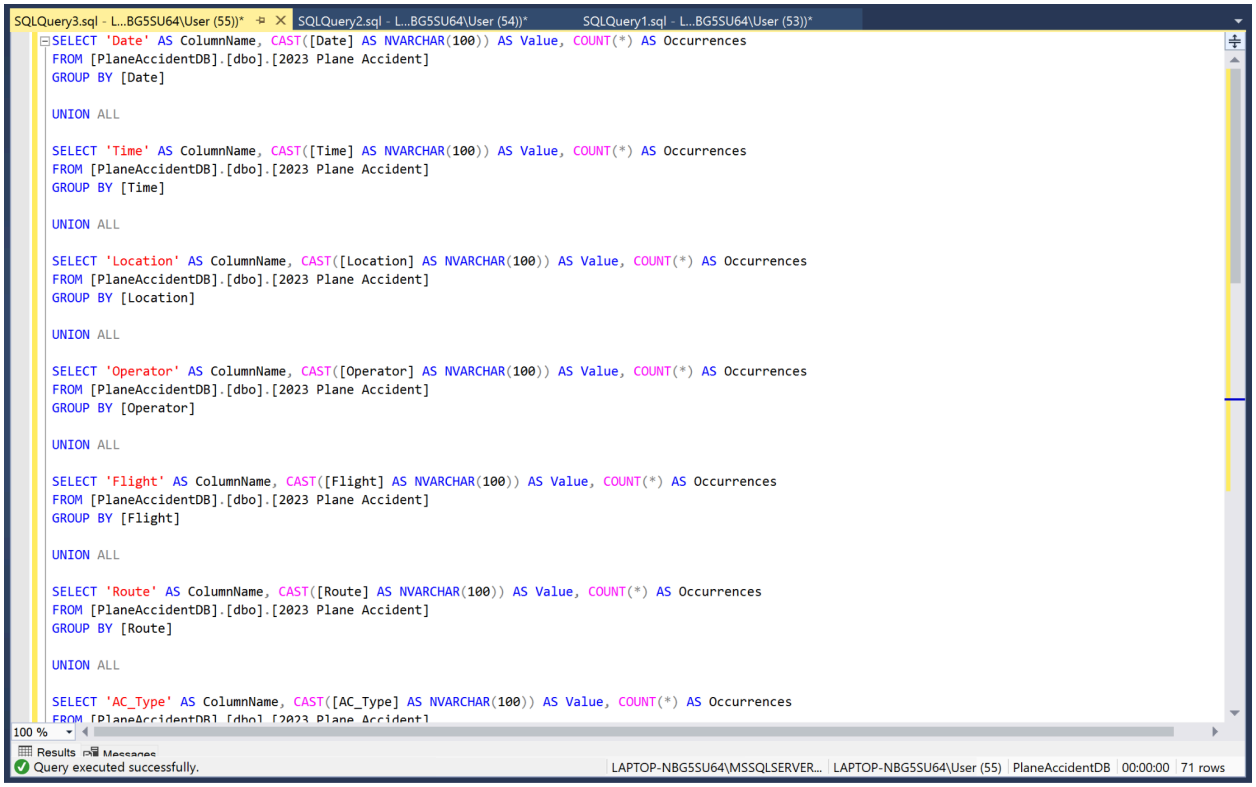
Results | Messages

	ColumnName	Value	Occurrences
1	Date	2022-03-22	1
2	Date	2022-05-29	1
3	Date	2022-07-16	1
4	Date	2022-11-06	1
5	Date	2022-11-18	1
6	Date	2022-11-21	1
7	Time	622	1
8	Time	853	1
9	Time	1010	1
10	Time	1015	1
11	Time	1511	1
12	Time	2247	1
13	Location	Bukoba, Tanzania	1
14	Location	Eleftheroupolis, Greece	1
15	Location	Lima, Peru	1
16	Location	Medell◆n, Colombia	1
17	Location	Near Lete Pass	1
18	Location	Wuzhou, Guangxi, China	1
19	Operator	AeroPaca SAS	1
20	Operator	China Eastern Airlines	1
21	Operator	LATAM	1
22	Operator	Meridian	1
23	Operator	Precision Air	1
24	Operator	Tara Air	1
25	Flight	?	2
26	Flight	LA2213	1
27	Flight	MEM3032	1
28	Flight	MU 5735	1
29	Flight	PW494	1
30	Route	Dar es-Salaam -Bukoba	1
31	Route	Kunming - Guangzhou	1
32	Route	Lima - Juliaca	1
33	Route	Medell◆n - Pizarro	1
34	Route	Nis- Amman	1
35	Route	Pokhara - Jomsom	1
36	AC_Type	Airbus 320-271N	1
37	AC_Type	Antonov An-12	1
38	AC_Type	ATR 42-500	1
39	AC_Type	Boeing 737-89P WL	1
40	AC_Type	de Havilland Canada DHC-6 Twin Otter 300	1
41	AC_Type	Piper PA-31-350 Navajo Chieftain	1
42	Registration	5H-PWF◆	1
43	Registration	9N-AET	1
44	Registration	B-1791	1
45	Registration	CC-BHB	1
46	Registration	HK-5121◆	1
47	Registration	UR-CIC◆	1
48	Constructio...	1347701	1
49	Constructio...	31-7652004	1
50	Constructio...	41474/5453	1
51	Constructio...	619	1
52	Constructio...	7864	1
53	Constructio...	819	1
54	Aboard	8	2
55	Aboard	22	1
56	Aboard	43	1
57	Aboard	108	1
58	Aboard	132	1
59	Fatalities	0	1
60	Fatalities	8	2
61	Fatalities	19	1
62	Fatalities	22	1
63	Fatalities	132	1
64	Ground	0	5
65	Ground	2	1
66	Summary	Flying at 29,098 feet, the jetliner started d...	1
67	Summary	The Airbus A320 collided with a fire truck ...	1
68	Summary	The cargo plane carrying eight crew me...	1
69	Summary	The commuter plane crashed while on a ...	1
70	Summary	The plane was chartered to carry a team ...	1
71	Summary	While on final approach to Bukoba Airport...	1

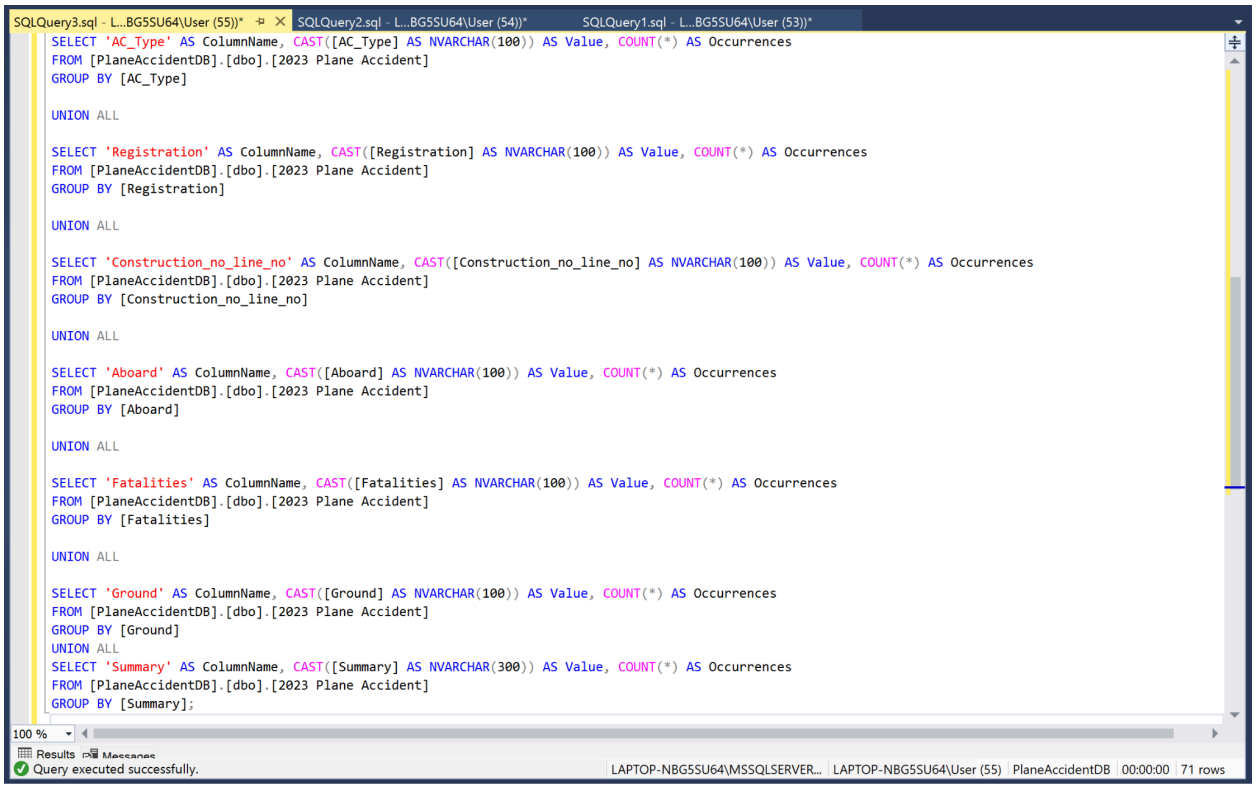
Query executed successfully.

Query executed successfully.

SQL QUERIES: 2023 Plane Accident Database



```
SELECT 'Date' AS ColumnName, CAST([Date] AS NVARCHAR(100)) AS Value, COUNT(*) AS Occurrences
FROM [PlaneAccidentDB].[dbo].[2023 Plane Accident]
GROUP BY [Date]

UNION ALL

SELECT 'Time' AS ColumnName, CAST([Time] AS NVARCHAR(100)) AS Value, COUNT(*) AS Occurrences
FROM [PlaneAccidentDB].[dbo].[2023 Plane Accident]
GROUP BY [Time]

UNION ALL

SELECT 'Location' AS ColumnName, CAST([Location] AS NVARCHAR(100)) AS Value, COUNT(*) AS Occurrences
FROM [PlaneAccidentDB].[dbo].[2023 Plane Accident]
GROUP BY [Location]

UNION ALL

SELECT 'Operator' AS ColumnName, CAST([Operator] AS NVARCHAR(100)) AS Value, COUNT(*) AS Occurrences
FROM [PlaneAccidentDB].[dbo].[2023 Plane Accident]
GROUP BY [Operator]

UNION ALL

SELECT 'Flight' AS ColumnName, CAST([Flight] AS NVARCHAR(100)) AS Value, COUNT(*) AS Occurrences
FROM [PlaneAccidentDB].[dbo].[2023 Plane Accident]
GROUP BY [Flight]

UNION ALL

SELECT 'Route' AS ColumnName, CAST([Route] AS NVARCHAR(100)) AS Value, COUNT(*) AS Occurrences
FROM [PlaneAccidentDB].[dbo].[2023 Plane Accident]
GROUP BY [Route]

UNION ALL

SELECT 'AC_Type' AS ColumnName, CAST([AC_Type] AS NVARCHAR(100)) AS Value, COUNT(*) AS Occurrences
FROM [PlaneAccidentDB].[dbo].[2023 Plane Accident]
```



```
SELECT 'AC_Type' AS ColumnName, CAST([AC_Type] AS NVARCHAR(100)) AS Value, COUNT(*) AS Occurrences
FROM [PlaneAccidentDB].[dbo].[2023 Plane Accident]
GROUP BY [AC_Type]

UNION ALL

SELECT 'Registration' AS ColumnName, CAST([Registration] AS NVARCHAR(100)) AS Value, COUNT(*) AS Occurrences
FROM [PlaneAccidentDB].[dbo].[2023 Plane Accident]
GROUP BY [Registration]

UNION ALL

SELECT 'Construction_no_line_no' AS ColumnName, CAST([Construction_no_line_no] AS NVARCHAR(100)) AS Value, COUNT(*) AS Occurrences
FROM [PlaneAccidentDB].[dbo].[2023 Plane Accident]
GROUP BY [Construction_no_line_no]

UNION ALL

SELECT 'Aboard' AS ColumnName, CAST([Aboard] AS NVARCHAR(100)) AS Value, COUNT(*) AS Occurrences
FROM [PlaneAccidentDB].[dbo].[2023 Plane Accident]
GROUP BY [Aboard]

UNION ALL

SELECT 'Fatalities' AS ColumnName, CAST([Fatalities] AS NVARCHAR(100)) AS Value, COUNT(*) AS Occurrences
FROM [PlaneAccidentDB].[dbo].[2023 Plane Accident]
GROUP BY [Fatalities]

UNION ALL

SELECT 'Ground' AS ColumnName, CAST([Ground] AS NVARCHAR(100)) AS Value, COUNT(*) AS Occurrences
FROM [PlaneAccidentDB].[dbo].[2023 Plane Accident]
GROUP BY [Ground]
UNION ALL
SELECT 'Summary' AS ColumnName, CAST([Summary] AS NVARCHAR(300)) AS Value, COUNT(*) AS Occurrences
FROM [PlaneAccidentDB].[dbo].[2023 Plane Accident]
GROUP BY [Summary];
```

RESULTS: 2023 Plane Accident Database

SQLQuery3.sql - L...BG5SU64\User (55))* SQLQuery2.sql - L

100 %

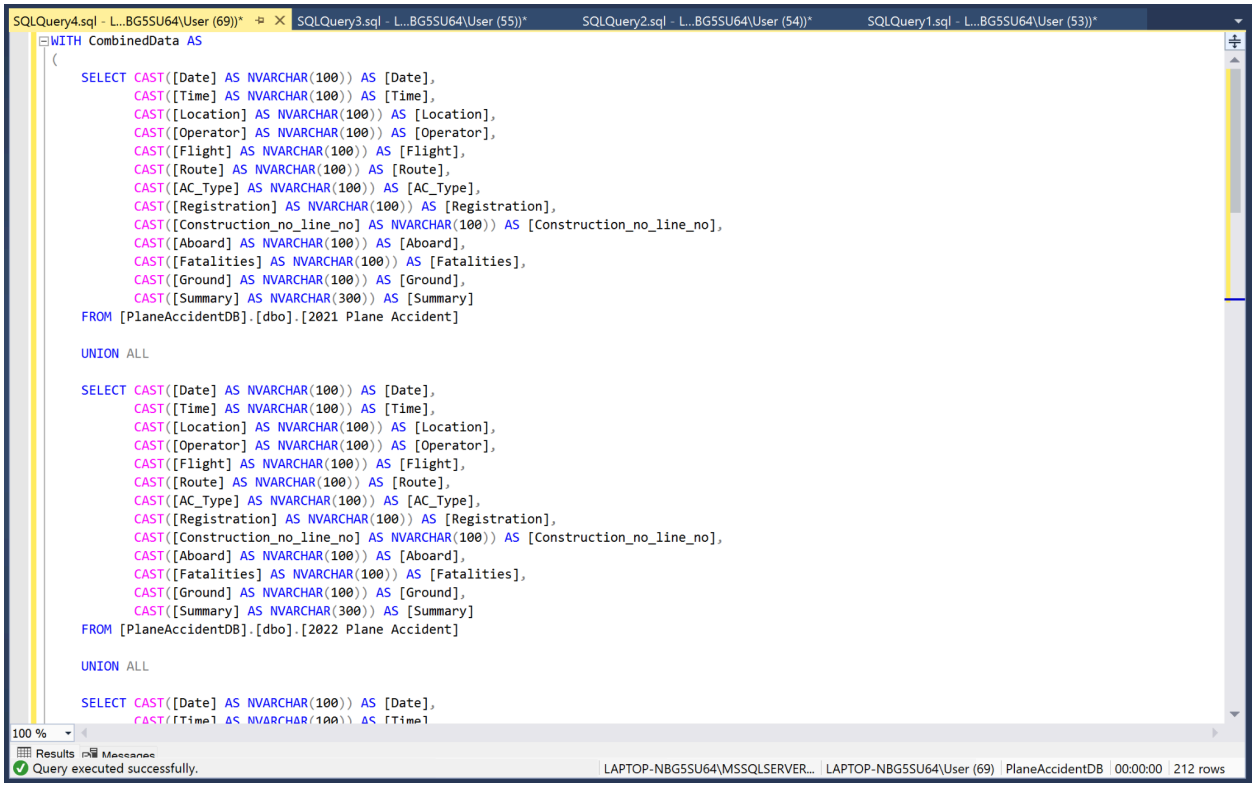
Results Messages

	ColumnName	Value	Occurrences
1	Date	2022-03-22	1
2	Date	2022-05-29	1
3	Date	2022-07-16	1
4	Date	2022-11-06	1
5	Date	2022-11-18	1
6	Date	2022-11-21	1
7	Time	622	1
8	Time	853	1
9	Time	1010	1
10	Time	1015	1
11	Time	1511	1
12	Time	2247	1
13	Location	Bukoba, Tanzania	1
14	Location	Eleftheroupolis, Greece	1
15	Location	Lima, Peru	1
16	Location	Medell◆n, Colombia	1
17	Location	Near Lete Pass	1
18	Location	Wuzhou, Guangxi, China	1
19	Operator	AeroPaca SAS	1
20	Operator	China Eastern Airlines	1
21	Operator	LATAM	1
22	Operator	Meridian	1
23	Operator	Precision Air	1
24	Operator	Tara Air	1
25	Flight	?	2
26	Flight	LA2213	1
27	Flight	MEM3032	1
28	Flight	MU 5735	1
29	Flight	PW494	1
30	Route	Dar es-Salaam -Bukoba	1
31	Route	Kunming - Guangzhou	1
32	Route	Lima - Juliaca	1
33	Route	Medell◆n - Pizarro	1
34	Route	Nis- Amman	1
35	Route	Pokhara - Jomsom	1
36	AC_Type	Airbus 320-271N	1
37	AC_Type	Antonov An-12	1
38	AC_Type	ATR 42-500	1
39	AC_Type	Boeing 737-89P WL	1
40	AC_Type	de Havilland Canada DHC-6 Twin Otter 300	1
41	AC_Type	Piper PA-31-350 Navajo Chieftain	1
42	Registration	5H-PWF◆	1
43	Registration	9N-AET	1
44	Registration	B-1791	1
45	Registration	CC-BHB	1
46	Registration	HK-5121◆	1
47	Registration	UR-CIC◆	1
48	Constructio...	1347701	1
49	Constructio...	31-7652004	1
50	Constructio...	41474/5453	1
51	Constructio...	619	1
52	Constructio...	7864	1
53	Constructio...	819	1
54	Aboard	8	2
55	Aboard	22	1
56	Aboard	43	1
57	Aboard	108	1
58	Aboard	132	1
59	Fatalities	0	1
60	Fatalities	8	2
61	Fatalities	19	1
62	Fatalities	22	1
63	Fatalities	132	1
64	Ground	0	5
65	Ground	2	1
66	Summary	Flying at 29,098 feet, the jetliner started d...	1
67	Summary	The Airbus A320 collided with a fire truck ...	1
68	Summary	The cargo plane carrying eight crew me...	1
69	Summary	The commuter plane crashed while on a ...	1
70	Summary	The plane was chartered to carry a team ...	1
71	Summary	While on final approach to Bukoba Airport...	1

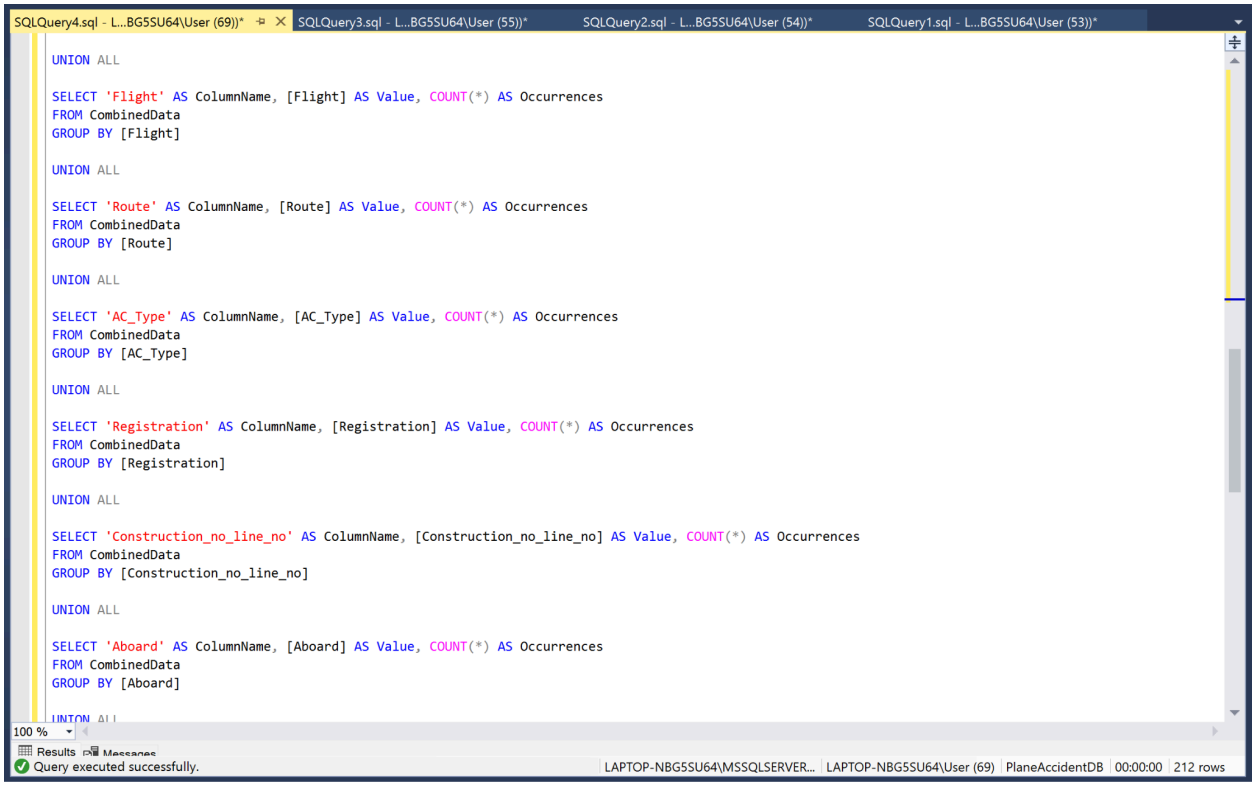
Query executed successfully.

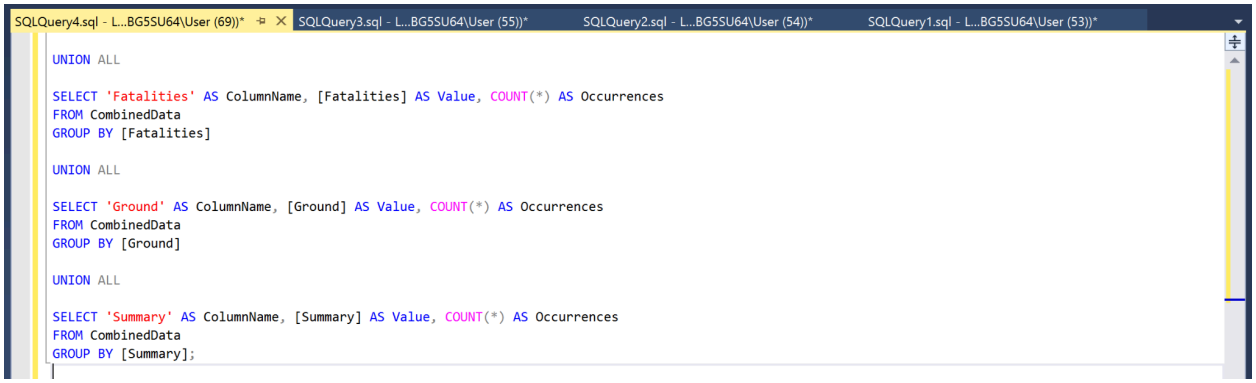
Query executed successfully.

SQL QUERIES: Combination of 2021, 2022 & 2023 Plane Accident Database



```
WITH CombinedData AS
(
    SELECT CAST([Date] AS NVARCHAR(100)) AS [Date],
           CAST([Time] AS NVARCHAR(100)) AS [Time],
           CAST([Location] AS NVARCHAR(100)) AS [Location],
           CAST([Operator] AS NVARCHAR(100)) AS [Operator],
           CAST([Flight] AS NVARCHAR(100)) AS [Flight],
           CAST([Route] AS NVARCHAR(100)) AS [Route],
           CAST([AC_Type] AS NVARCHAR(100)) AS [AC_Type],
           CAST([Registration] AS NVARCHAR(100)) AS [Registration],
           CAST([Construction_no_line_no] AS NVARCHAR(100)) AS [Construction_no_line_no],
           CAST([Aboard] AS NVARCHAR(100)) AS [Aboard],
           CAST([Fatalities] AS NVARCHAR(100)) AS [Fatalities],
           CAST([Ground] AS NVARCHAR(100)) AS [Ground],
           CAST([Summary] AS NVARCHAR(300)) AS [Summary]
    FROM [PlaneAccidentDB].[dbo].[2021 Plane Accident]

    UNION ALL

    SELECT CAST([Date] AS NVARCHAR(100)) AS [Date],
           CAST([Time] AS NVARCHAR(100)) AS [Time],
           CAST([Location] AS NVARCHAR(100)) AS [Location],
           CAST([Operator] AS NVARCHAR(100)) AS [Operator],
           CAST([Flight] AS NVARCHAR(100)) AS [Flight],
           CAST([Route] AS NVARCHAR(100)) AS [Route],
           CAST([AC_Type] AS NVARCHAR(100)) AS [AC_Type],
           CAST([Registration] AS NVARCHAR(100)) AS [Registration],
           CAST([Construction_no_line_no] AS NVARCHAR(100)) AS [Construction_no_line_no],
           CAST([Aboard] AS NVARCHAR(100)) AS [Aboard],
           CAST([Fatalities] AS NVARCHAR(100)) AS [Fatalities],
           CAST([Ground] AS NVARCHAR(100)) AS [Ground],
           CAST([Summary] AS NVARCHAR(300)) AS [Summary]
    FROM [PlaneAccidentDB].[dbo].[2022 Plane Accident]

    UNION ALL

    SELECT CAST([Date] AS NVARCHAR(100)) AS [Date],
           CAST([Time] AS NVARCHAR(100)) AS [Time],
           CAST([Location] AS NVARCHAR(100)) AS [Location],
           CAST([Operator] AS NVARCHAR(100)) AS [Operator],
           CAST([Flight] AS NVARCHAR(100)) AS [Flight],
           CAST([Route] AS NVARCHAR(100)) AS [Route],
           CAST([AC_Type] AS NVARCHAR(100)) AS [AC_Type],
           CAST([Registration] AS NVARCHAR(100)) AS [Registration],
           CAST([Construction_no_line_no] AS NVARCHAR(100)) AS [Construction_no_line_no],
           CAST([Aboard] AS NVARCHAR(100)) AS [Aboard],
           CAST([Fatalities] AS NVARCHAR(100)) AS [Fatalities],
           CAST([Ground] AS NVARCHAR(100)) AS [Ground],
           CAST([Summary] AS NVARCHAR(300)) AS [Summary]
    FROM [PlaneAccidentDB].[dbo].[2023 Plane Accident]
)
SELECT 'Date' AS ColumnName, [Date] AS Value, COUNT(*) AS Occurrences
FROM CombinedData
GROUP BY [Date]

UNION ALL

SELECT 'Time' AS ColumnName, [Time] AS Value, COUNT(*) AS Occurrences
FROM CombinedData
GROUP BY [Time]

UNION ALL

SELECT 'Location' AS ColumnName, [Location] AS Value, COUNT(*) AS Occurrences
FROM CombinedData
GROUP BY [Location]

UNION ALL

SELECT 'Operator' AS ColumnName, [Operator] AS Value, COUNT(*) AS Occurrences
FROM CombinedData
GROUP BY [Operator]

UNION ALL
```


SQLQuery4.sql - L...BG5SU64\User (69))*
SQLQuery3.sql - L...BG5SU64\User (55))*
SQLQuery2.sql - L...BG5SU64\User (54))*
SQLQuery1.sql - L...BG5SU64\User (53))*
UNION ALL
SELECT 'Flight' AS ColumnName, [Flight] AS Value, COUNT(*) AS Occurrences
FROM CombinedData
GROUP BY [Flight]
UNION ALL
SELECT 'Route' AS ColumnName, [Route] AS Value, COUNT(*) AS Occurrences
FROM CombinedData
GROUP BY [Route]
UNION ALL
SELECT 'AC_Type' AS ColumnName, [AC_Type] AS Value, COUNT(*) AS Occurrences
FROM CombinedData
GROUP BY [AC_Type]
UNION ALL
SELECT 'Registration' AS ColumnName, [Registration] AS Value, COUNT(*) AS Occurrences
FROM CombinedData
GROUP BY [Registration]
UNION ALL
SELECT 'Construction_no_line_no' AS ColumnName, [Construction_no_line_no] AS Value, COUNT(*) AS Occurrences
FROM CombinedData
GROUP BY [Construction_no_line_no]
UNION ALL
SELECT 'Aboard' AS ColumnName, [Aboard] AS Value, COUNT(*) AS Occurrences
FROM CombinedData
GROUP BY [Aboard]
100 %
Results
Query executed successfully.
LAPTOP-NBG5SU64\MSSQLSERVER...
LAPTOP-NBG5SU64\User (69)
PlaneAccidentDB
00:00:00
212 rows


SQLQuery4.sql - L...BG5SU64\User (69))*
SQLQuery3.sql - L...BG5SU64\User (55))*
SQLQuery2.sql - L...BG5SU64\User (54))*
SQLQuery1.sql - L...BG5SU64\User (53))*
UNION ALL
SELECT 'Fatalities' AS ColumnName, [Fatalities] AS Value, COUNT(*) AS Occurrences
FROM CombinedData
GROUP BY [Fatalities]
UNION ALL
SELECT 'Ground' AS ColumnName, [Ground] AS Value, COUNT(*) AS Occurrences
FROM CombinedData
GROUP BY [Ground]
UNION ALL
SELECT 'Summary' AS ColumnName, [Summary] AS Value, COUNT(*) AS Occurrences
FROM CombinedData
GROUP BY [Summary];

RESULTS: Combination of 2021, 2022 & 2023 Plane Accident Database

SQLQuery4.sql - L...BG5SU64\User (69))*

100 %

Results | Messages

	ColumnName	Value	Occurrences
1	Date	2021-01-09	1
2	Date	2021-03-02	1
3	Date	2021-03-28	1
4	Date	2021-05-21	1
5	Date	2021-06-10	1
6	Date	2021-07-04	1
7	Date	2021-07-06	1
8	Date	2021-09-12	1
9	Date	2021-12-27	1
10	Date	2022-03-22	1
11	Date	2022-05-29	1
12	Date	2022-07-16	1
13	Date	2022-11-06	1
14	Date	2022-11-18	1
15	Date	2022-11-21	1
16	Date	2023-01-15	1
17	Date	2023-09-16	1
18	Date	2023-10-29	1
19	Date	2023-11-01	1
20	Time	NULL	1
21	Time	1010	1
22	Time	1015	1
23	Time	1050	1
24	Time	1130	1
25	Time	1415	1
26	Time	1440	1
27	Time	1500	1
28	Time	1511	1
29	Time	1705	1
30	Time	1800	1
31	Time	1835	1
32	Time	1914	1
33	Time	2247	1
34	Time	2315	1
35	Time	622	1
36	Time	630	1
37	Time	800	1
38	Time	853	1
39	Location	Barcelos, Brazil	1
40	Location	Bukoba, Tanzania	1
41	Location	El Cajon, California	1
42	Location	Eleftheroupolis, Greece	1
43	Location	Kazachinskoye, Russia	1
44	Location	Lima, Peru	1

Query executed successfully.

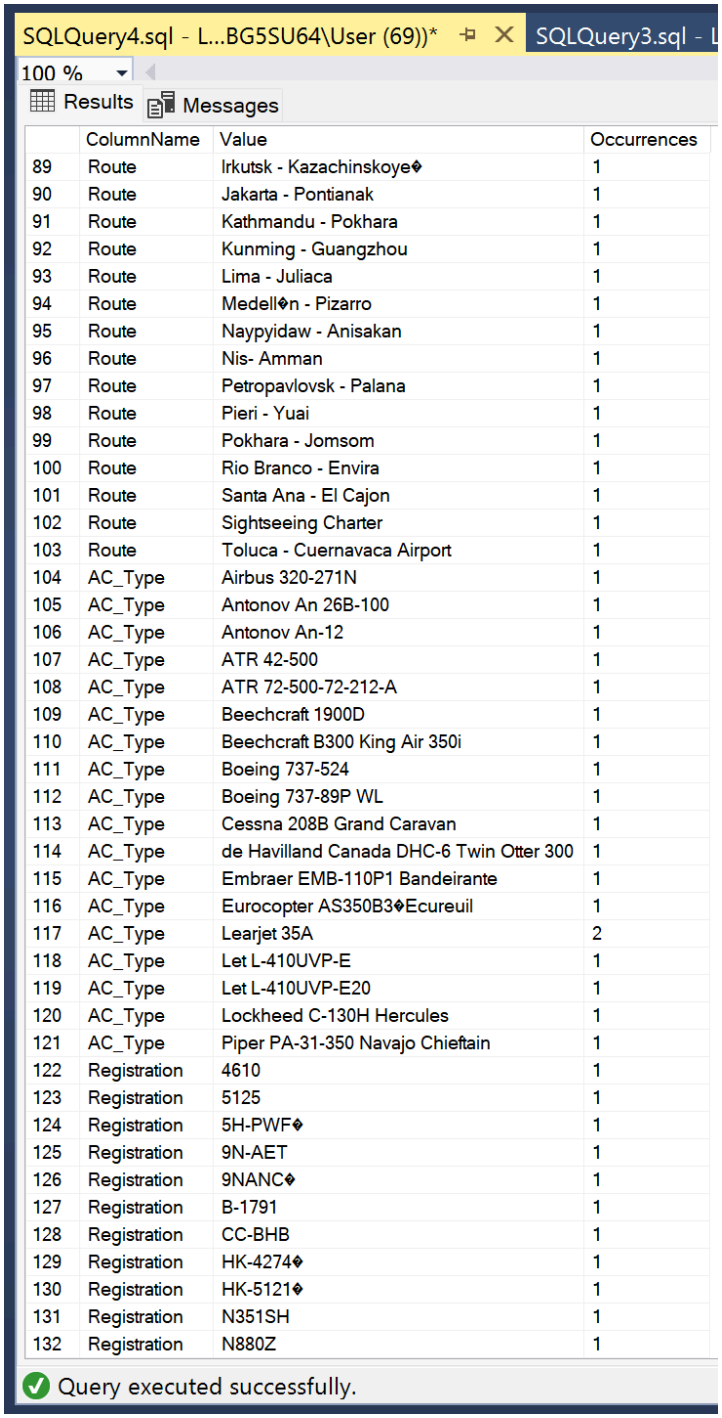
SQLQuery4.sql - L...BG5SU64\User (69))*

100 %

Results | Messages

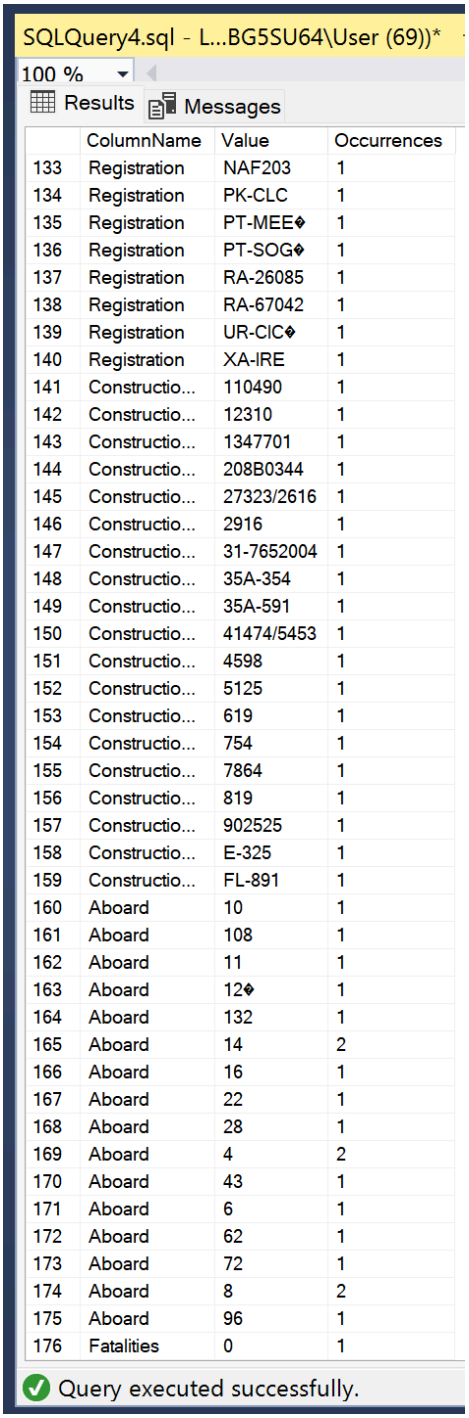
	ColumnName	Value	Occurrences
45	Location	Medell◆n, Colombia	1
46	Location	Near Butte, Alaska	1
47	Location	Near Jakarta, Indonesia	1
48	Location	Near Kaduna, Nigeria	1
49	Location	Near Lete Pass	1
50	Location	Near Pyin Oo Lwin, Myanmar	1
51	Location	Palana, Russia	1
52	Location	Patikul, Sulu, Philippines	1
53	Location	Pieri, Sudan	1
54	Location	Pokhara, Nepal	1
55	Location	Rio Branco, Brazil	1
56	Location	Temixco, Morelos, Mexico	1
57	Location	Wuzhou, Guangxi, China	1
58	Operator	AeroPaca SAS	1
59	Operator	Aeroservice/SiLA	1
60	Operator	ART Taxi Aero	1
61	Operator	China Eastern Airlines	1
62	Operator	Jet Rescue Air Ambulance	1
63	Operator	Kamchatka Aviation Enterprise	1
64	Operator	LATAM	1
65	Operator	Manaus Aerotaxi	1
66	Operator	Med Jet	1
67	Operator	Meridian	1
68	Operator	Military - Myanmar Air Force	1
69	Operator	Military - Nigerian Air Force	1
70	Operator	Military - Philippine Air Force	1
71	Operator	Precision Air	1
72	Operator	Soloy Helicopters◆	1
73	Operator	South Sudan Supreme Airlines◆	1
74	Operator	Sriwijaya Air	1
75	Operator	Tara Air	1
76	Operator	Yeti Airlines	1
77	Flight	?	11
78	Flight	251	1
79	Flight	51	1
80	Flight	LA2213	1
81	Flight	MEM3032	1
82	Flight	MU 5735	1
83	Flight	PW494	1
84	Flight	SJ182	1
85	Flight	YT691	1
86	Route	?	2
87	Route	Cagayan de Oro-Lumbia - Jolo	1
88	Route	Dar es-Salaam -Bukoba	1

Query executed successfully.



	ColumnName	Value	Occurrences
89	Route	Irkutsk - Kazachinskoye•	1
90	Route	Jakarta - Pontianak	1
91	Route	Kathmandu - Pokhara	1
92	Route	Kunming - Guangzhou	1
93	Route	Lima - Juliaca	1
94	Route	Medell•n - Pizarro	1
95	Route	Naypyidaw - Anisakan	1
96	Route	Nis- Amman	1
97	Route	Petropavlovsk - Palana	1
98	Route	Pieri - Yuai	1
99	Route	Pokhara - Jomsom	1
100	Route	Rio Branco - Envira	1
101	Route	Santa Ana - El Cajon	1
102	Route	Sightseeing Charter	1
103	Route	Toluca - Cuernavaca Airport	1
104	AC_Type	Airbus 320-271N	1
105	AC_Type	Antonov An 26B-100	1
106	AC_Type	Antonov An-12	1
107	AC_Type	ATR 42-500	1
108	AC_Type	ATR 72-500-72-212-A	1
109	AC_Type	Beechcraft 1900D	1
110	AC_Type	Beechcraft B300 King Air 350i	1
111	AC_Type	Boeing 737-524	1
112	AC_Type	Boeing 737-89P WL	1
113	AC_Type	Cessna 208B Grand Caravan	1
114	AC_Type	de Havilland Canada DHC-6 Twin Otter 300	1
115	AC_Type	Embraer EMB-110P1 Bandeirante	1
116	AC_Type	Eurocopter AS350B3•Ecureuil	1
117	AC_Type	Learjet 35A	2
118	AC_Type	Let L-410UVP-E	1
119	AC_Type	Let L-410UVP-E20	1
120	AC_Type	Lockheed C-130H Hercules	1
121	AC_Type	Piper PA-31-350 Navajo Chieftain	1
122	Registration	4610	1
123	Registration	5125	1
124	Registration	5H-PWF•	1
125	Registration	9N-AET	1
126	Registration	9NANC•	1
127	Registration	B-1791	1
128	Registration	CC-BHB	1
129	Registration	HK-4274•	1
130	Registration	HK-5121•	1
131	Registration	N351SH	1
132	Registration	N880Z	1

Query executed successfully.



	ColumnName	Value	Occurrences
133	Registration	NAF203	1
134	Registration	PK-CLC	1
135	Registration	PT-MEE•	1
136	Registration	PT-SOG•	1
137	Registration	RA-26085	1
138	Registration	RA-67042	1
139	Registration	UR-CIC•	1
140	Registration	XA-IRE	1
141	Constructio...	110490	1
142	Constructio...	12310	1
143	Constructio...	1347701	1
144	Constructio...	208B0344	1
145	Constructio...	27323/2616	1
146	Constructio...	2916	1
147	Constructio...	31-7652004	1
148	Constructio...	35A-354	1
149	Constructio...	35A-591	1
150	Constructio...	41474/5453	1
151	Constructio...	4598	1
152	Constructio...	5125	1
153	Constructio...	619	1
154	Constructio...	754	1
155	Constructio...	7864	1
156	Constructio...	819	1
157	Constructio...	902525	1
158	Constructio...	E-325	1
159	Constructio...	FL-891	1
160	Aboard	10	1
161	Aboard	108	1
162	Aboard	11	1
163	Aboard	12•	1
164	Aboard	132	1
165	Aboard	14	2
166	Aboard	16	1
167	Aboard	22	1
168	Aboard	28	1
169	Aboard	4	2
170	Aboard	43	1
171	Aboard	6	1
172	Aboard	62	1
173	Aboard	72	1
174	Aboard	8	2
175	Aboard	96	1
176	Fatalities	0	1

Query executed successfully.

177	Fatalities	10	1
178	Fatalities	11	1
179	Fatalities	12	2
180	Fatalities	132	1
181	Fatalities	14	1
182	Fatalities	19	1
183	Fatalities	22	1
184	Fatalities	28	1
185	Fatalities	4	3
186	Fatalities	5	1
187	Fatalities	50	1
188	Fatalities	62	1
189	Fatalities	72	1
190	Fatalities	8	2
191	Ground	0	17
192	Ground	2	1
193	Ground	3	1
194	Summary	The crew while in initial climb, encounter...	1
195	Summary	A business jet operating as an air ambula...	1
196	Summary	Before approach, the pilot requested a ch...	1
197	Summary	Flying at 29,098 feet, the jetliner started d...	1
198	Summary	One of the engines on the aircraft failed 1...	1
199	Summary	Sriwijaya Air flight 182 was climbing throu...	1
200	Summary	The air ambulance flight was completing ...	1
201	Summary	The air taxi crashed in heavy rain while at...	1
202	Summary	The Airbus A320 collided with a fire truck ...	1
203	Summary	The aircraft was on it's second approach i...	1
204	Summary	The cargo plane carrying eight crew me...	1
205	Summary	The commuter plane crashed while on a ...	1
206	Summary	The passenger plane crashed into the to...	1
207	Summary	The plane was carrying military personne...	1
208	Summary	The plane was chartered to carry a team ...	1
209	Summary	The sightseeing helicopter crashed after ...	1
210	Summary	While attempting to land at Jolo Airport, th...	1
211	Summary	While on final approach to Bukoba Airport...	1
212	Summary	While on final approach, in poor weather ...	1

Query executed successfully.

2.2 Length Distribution

Find the minimum and maximum lengths of string values in the selected columns. This profile helps you identify problems in your data, such as values that are valid. For example, you profile a column of Malaysia states codes that should be two characters and discover values longer than two characters.

SQL QUERIES: 2021 Plane Accident Database

```
SELECT
    'Date' AS ColumnName,
    MIN(LEN([Date])) AS MinLength,
    MAX(LEN([Date])) AS MaxLength
FROM [dbo].[2023 Plane Accident]
UNION ALL
SELECT
    'Time' AS ColumnName,
    MIN(LEN([Time])) AS MinLength,
    MAX(LEN([Time])) AS MaxLength
FROM [dbo].[2023 Plane Accident]
UNION ALL
SELECT
    'Location' AS ColumnName,
    MIN(LEN([Location])) AS MinLength,
    MAX(LEN([Location])) AS MaxLength
FROM [dbo].[2023 Plane Accident]
UNION ALL
SELECT
    'Operator' AS ColumnName,
    MIN(LEN([Operator])) AS MinLength,
    MAX(LEN([Operator])) AS MaxLength
FROM [dbo].[2023 Plane Accident]
UNION ALL
SELECT
    'Flight #' AS ColumnName,
    MIN(LEN([Flight #])) AS MinLength,
    MAX(LEN([Flight #])) AS MaxLength
FROM [dbo].[2023 Plane Accident]
UNION ALL
SELECT
    'Route' AS ColumnName,
    MIN(LEN([Route])) AS MinLength,
    MAX(LEN([Route])) AS MaxLength
FROM [dbo].[2021 Plane Accident]
UNION ALL
SELECT
    'AC Type' AS ColumnName,
    MIN(LEN([AC Type])) AS MinLength,
    MAX(LEN([AC Type])) AS MaxLength
FROM [dbo].[2021 Plane Accident]
UNION ALL
SELECT
    'Registration' AS ColumnName,
    MIN(LEN([Registration])) AS MinLength,
    MAX(LEN([Registration])) AS MaxLength
FROM [dbo].[2021 Plane Accident]
UNION ALL
SELECT
    'Construction no / line no' AS ColumnName,
    MIN(LEN([Construction no / line no])) AS MinLength,
    MAX(LEN([Construction no / line no])) AS MaxLength
FROM [dbo].[2021 Plane Accident]
UNION ALL
SELECT
    'Summary' AS ColumnName,
    MIN(LEN([Summary])) AS MinLength,
    MAX(LEN([Summary])) AS MaxLength
FROM [dbo].[2021 Plane Accident];
```

RESULTS: 2021 Plane Accident Database

	ColumnName	MinLength	MaxLength
1	Date	27	27
2	Time	3	4
3	Location	12	26
4	Operator	7	31
5	Flight #	1	5
6	Route	1	28
7	AC Type	11	29
8	Registration	4	8
9	Constructio...	4	10
10	Summary	70	692

SQL QUERIES: 2022 Plane Accident Database

```sql
SELECT
    'Date' AS ColumnName,
    MIN(LEN([Date])) AS MinLength,
    MAX(LEN([Date])) AS MaxLength
FROM [dbo].[2022 Plane Accident]
UNION ALL
SELECT
    'Time' AS ColumnName,
    MIN(LEN([Time])) AS MinLength,
    MAX(LEN([Time])) AS MaxLength
FROM [dbo].[2022 Plane Accident]
UNION ALL
SELECT
    'Location' AS ColumnName,
    MIN(LEN([Location])) AS MinLength,
    MAX(LEN([Location])) AS MaxLength
FROM [dbo].[2022 Plane Accident]
UNION ALL
SELECT
    'Operator' AS ColumnName,
    MIN(LEN([Operator])) AS MinLength,
    MAX(LEN([Operator])) AS MaxLength
FROM [dbo].[2022 Plane Accident]
UNION ALL
SELECT
    'Flight #' AS ColumnName,
    MIN(LEN([Flight #])) AS MinLength,
    MAX(LEN([Flight #])) AS MaxLength
FROM [dbo].[2022 Plane Accident]
UNION ALL
SELECT
    'Route' AS ColumnName,
    MIN(LEN([Route])) AS MinLength,
    MAX(LEN([Route])) AS MaxLength
FROM [dbo].[2022 Plane Accident]
UNION ALL
SELECT
    'AC Type' AS ColumnName,
    MIN(LEN([AC Type])) AS MinLength,
    MAX(LEN([AC Type])) AS MaxLength
FROM [dbo].[2022 Plane Accident]
UNION ALL
SELECT
    'Registration' AS ColumnName,
    MIN(LEN([Registration])) AS MinLength,
    MAX(LEN([Registration])) AS MaxLength
FROM [dbo].[2022 Plane Accident]
UNION ALL
SELECT
    'Construction no / line no' AS ColumnName,
    MIN(LEN([Construction no / line no])) AS MinLength,
    MAX(LEN([Construction no / line no])) AS MaxLength
FROM [dbo].[2022 Plane Accident]
UNION ALL
SELECT
    'Summary' AS ColumnName,
    MIN(LEN([Summary])) AS MinLength,
    MAX(LEN([Summary])) AS MaxLength
FROM [dbo].[2022 Plane Accident];
```

RESULTS: 2022 Plane Accident Database

	ColumnName	MinLength	MaxLength
1	Date	27	27
2	Time	3	4
3	Location	10	23
4	Operator	5	22
5	Flight No	1	7
6	Route	10	21
7	AC Type	10	40
8	Registration	6	8
9	Constructio...	3	10
10	Summary	75	781

SQL QUERIES: 2023 Plane Accident Database

```sql
SELECT
    'Date' AS ColumnName,
    MIN(LEN([Date])) AS MinLength,
    MAX(LEN([Date])) AS MaxLength
FROM [dbo].[2023 Plane Accident]
UNION ALL
SELECT
    'Time' AS ColumnName,
    MIN(LEN([Time])) AS MinLength,
    MAX(LEN([Time])) AS MaxLength
FROM [dbo].[2023 Plane Accident]
UNION ALL
SELECT
    'Location' AS ColumnName,
    MIN(LEN([Location])) AS MinLength,
    MAX(LEN([Location])) AS MaxLength
FROM [dbo].[2023 Plane Accident]
UNION ALL
SELECT
    'Operator' AS ColumnName,
    MIN(LEN([Operator])) AS MinLength,
    MAX(LEN([Operator])) AS MaxLength
FROM [dbo].[2023 Plane Accident]
UNION ALL
SELECT
    'Flight #' AS ColumnName,
    MIN(LEN([Flight #])) AS MinLength,
    MAX(LEN([Flight #])) AS MaxLength
FROM [dbo].[2023 Plane Accident]
UNION ALL
SELECT
    'Route' AS ColumnName,
    MIN(LEN([Route])) AS MinLength,
    MAX(LEN([Route])) AS MaxLength
FROM [dbo].[2023 Plane Accident]
UNION ALL
SELECT
    'AC Type' AS ColumnName,
    MIN(LEN([AC Type])) AS MinLength,
    MAX(LEN([AC Type])) AS MaxLength
FROM [dbo].[2023 Plane Accident]
UNION ALL
SELECT
    'Registration' AS ColumnName,
    MIN(LEN([Registration])) AS MinLength,
    MAX(LEN([Registration])) AS MaxLength
FROM [dbo].[2023 Plane Accident]
UNION ALL
SELECT
    'Construction no / line no' AS ColumnName,
    MIN(LEN([Construction no / line no])) AS MinLength,
    MAX(LEN([Construction no / line no])) AS MaxLength
FROM [dbo].[2023 Plane Accident]
UNION ALL
SELECT
    'Summary' AS ColumnName,
    MIN(LEN([Summary])) AS MinLength,
    MAX(LEN([Summary])) AS MaxLength
FROM [dbo].[2023 Plane Accident];
```

RESULTS: 2023 Plane Accident Database

	ColumnName	MinLength	MaxLength
1	Date	27	27
2	Time	1	4
3	Location	14	24
4	Operator	13	24
5	Flight #	1	5
6	Route	1	27
7	AC Type	11	29
8	Registration	6	7
9	Construction no / line no	3	8
10	Summary	87	282

Combined Length Distribution Results (2021,2022,2023) :

Column	MinLength (2021)	Maxlength (2021)	MinLength (2022)	MaxLength (2022)	MinLength (2023)	MaxLength (2023)
Date	27	27	27	27	27	27
Time	3	4	3	4	1	4
Location	12	26	10	23	10	23
Operator	7	31	5	22	13	24
Flight No	1	5	1	7	1	5
Route	1	28	10	21	1	27
AC Type	11	29	10	40	11	29
Registration	5	6	6	8	6	7
Construction no / Line no	4	11	3	10	3	8
Summary	70	692	75	781	87	282

A consistent and well-structured dataset is revealed by the length distribution analysis of 2021, 2022, and 2023, with a few significant changes that show how the data is changing over time. For comprehending the specifics of aviation crashes, the Summary column which varies from 70 to 692 characters in 2021, 75 to 781 in 2022, and 87 to 282 in 2023 remains the most important area. These discrepancies point to varying degrees of reporting detail, with shorter entries summarizing simpler events and longer ones offering in-depth descriptions of serious catastrophes. While the rising minimum length in 2023 points to an increase in giving consistent baseline facts, the declining maximum length may signal a move towards more condensed reporting.

With lengths ranging from 10 to 40, the AC Type column illustrates the range of aircraft involved in mishaps, from basic identifiers to intricate classifications. Crucial quantitative data on the human effect of accidents is provided by columns like Aboard, Fatalities, and Ground, which are dependable and consistent throughout all years. Additionally, consistent and organised are contextual columns such as Date, Location, and Operator. The Date column

consistently maintains a fixed length of 27 across all years, reflecting a uniform format that aids in tracking trends over time. Whereas the Operator column ranges from 5 to 31, representing variations in airline names, the Location column records a variety of geographic data with lengths ranging from 10 to 26. There are still data gaps, though, as fields like Flight No. and Route have placeholder entries ("?") with durations beginning at 1.

Overall the dataset effectively captures the key components that leads to aviation incidents with columns like Summary, AC Type and casualty data providing detailed insights into the nature of accidents. Supporting columns such as Date and Location provides valuable insights while incomplete data in Flight No and Route should be addressed to further improve reliability of this dataset.

2.3 Null Ratio

Find the number of null values in the columns and report the percentage of null values in the columns. This profile helps you identify problems in your data, such as an unexpectedly high ratio of null values in a column. For example, you profile a postcode column and discover an unacceptably high percentage of missing codes.

Summary of Null Values in Columns:

Tables Checked: 2021 Plane Accident, 2022 Plane Accident, 2023 Plane Accident

Nulls Identified: Only the 2023 Plane Accident table contains NULL values.

Findings:

1. 2021 Plane Accident: **No NULL** values were found in any columns but there's the “?” symbol which is identified as String.
2. 2022 Plane Accident: **No NULL** values were found in any columns but there's the “?” symbol which is identified as String.
3. 2023 Plane Accident: **Contains NULL** values in **TIME Columns**, with **25%** NULL percentage.

This indicates that the 2023 dataset may have missing or incomplete data in certain columns, while the data for 2021 and 2022 appears to be fully populated.

SQL QUERIES: 2021 Plane Accident Database

```sql
USE PlaneAccidentDB;

SELECT
    'Date' AS ColumnName,
    SUM(CASE WHEN [Date] IS NULL THEN 1 ELSE 0 END) AS NullCount,
    COUNT(*) AS TotalRows,
    (SUM(CASE WHEN [Date] IS NULL THEN 1 ELSE 0 END) * 100.0 / COUNT(*)) AS NullPercentage
FROM [2021 Plane Accident]

UNION ALL

SELECT
    'Time' AS ColumnName,
    SUM(CASE WHEN [Time] IS NULL THEN 1 ELSE 0 END) AS NullCount,
    COUNT(*) AS TotalRows,
    (SUM(CASE WHEN [Time] IS NULL THEN 1 ELSE 0 END) * 100.0 / COUNT(*)) AS NullPercentage
FROM [2021 Plane Accident]

UNION ALL

SELECT
    'Location' AS ColumnName,
    SUM(CASE WHEN [Location] IS NULL THEN 1 ELSE 0 END) AS NullCount,
    COUNT(*) AS TotalRows,
    (SUM(CASE WHEN [Location] IS NULL THEN 1 ELSE 0 END) * 100.0 / COUNT(*)) AS NullPercen
FROM [2021 Plane Accident]
```

```sql
UNION ALL

SELECT
    'Operator' AS ColumnName,
    SUM(CASE WHEN [Operator] IS NULL THEN 1 ELSE 0 END) AS NullCount,
    COUNT(*) AS TotalRows,
    (SUM(CASE WHEN [Operator] IS NULL THEN 1 ELSE 0 END) * 100.0 / COUNT(*)) AS NullPercen
FROM [2021 Plane Accident]

UNION ALL

SELECT
    'Flight' AS ColumnName,
    SUM(CASE WHEN [Flight] IS NULL THEN 1 ELSE 0 END) AS NullCount,
    COUNT(*) AS TotalRows,
    (SUM(CASE WHEN [Flight] IS NULL THEN 1 ELSE 0 END) * 100.0 / COUNT(*)) AS NullPercenta
FROM [2021 Plane Accident]

UNION ALL

SELECT
    'Route' AS ColumnName,
    SUM(CASE WHEN [Route] IS NULL THEN 1 ELSE 0 END) AS NullCount,
    COUNT(*) AS TotalRows,
    (SUM(CASE WHEN [Route] IS NULL THEN 1 ELSE 0 END) * 100.0 / COUNT(*)) AS NullPercentag
FROM [2021 Plane Accident]
```

```
UNION ALL

SELECT
    'AC_Type' AS ColumnName,
    SUM(CASE WHEN [AC_Type] IS NULL THEN 1 ELSE 0 END) AS NullCount,
    COUNT(*) AS TotalRows,
    (SUM(CASE WHEN [AC_Type] IS NULL THEN 1 ELSE 0 END) * 100.0 / COUNT(*)) AS NullPercent
FROM [2021 Plane Accident]

UNION ALL

SELECT
    'Registration' AS ColumnName,
    SUM(CASE WHEN [Registration] IS NULL THEN 1 ELSE 0 END) AS NullCount,
    COUNT(*) AS TotalRows,
    (SUM(CASE WHEN [Registration] IS NULL THEN 1 ELSE 0 END) * 100.0 / COUNT(*)) AS NullPe
FROM [2021 Plane Accident]

UNION ALL

SELECT
    'Construction_no_line_no' AS ColumnName,
    SUM(CASE WHEN [Construction_no_line_no] IS NULL THEN 1 ELSE 0 END) AS NullCount,
    COUNT(*) AS TotalRows,
    (SUM(CASE WHEN [Construction_no_line_no] IS NULL THEN 1 ELSE 0 END) * 100.0 / COUNT(*
FROM [2021 Plane Accident]
```

```
UNION ALL

SELECT
    'Aboard' AS ColumnName,
    SUM(CASE WHEN [Aboard] IS NULL THEN 1 ELSE 0 END) AS NullCount,
    COUNT(*) AS TotalRows,
    (SUM(CASE WHEN [Aboard] IS NULL THEN 1 ELSE 0 END) * 100.0 / COUNT(*)) AS NullPercenta
FROM [2021 Plane Accident]

UNION ALL

SELECT
    'Fatalities' AS ColumnName,
    SUM(CASE WHEN [Fatalities] IS NULL THEN 1 ELSE 0 END) AS NullCount,
    COUNT(*) AS TotalRows,
    (SUM(CASE WHEN [Fatalities] IS NULL THEN 1 ELSE 0 END) * 100.0 / COUNT(*)) AS NullPerc
FROM [2021 Plane Accident]

UNION ALL

SELECT
    'Ground' AS ColumnName,
    SUM(CASE WHEN [Ground] IS NULL THEN 1 ELSE 0 END) AS NullCount,
    COUNT(*) AS TotalRows,
    (SUM(CASE WHEN [Ground] IS NULL THEN 1 ELSE 0 END) * 100.0 / COUNT(*)) AS NullPercenta
FROM [2021 Plane Accident]
```

```sql
UNION ALL

SELECT
    'Summary' AS ColumnName,
    SUM(CASE WHEN [Summary] IS NULL THEN 1 ELSE 0 END) AS NullCount,
    COUNT(*) AS TotalRows,
    (SUM(CASE WHEN [Summary] IS NULL THEN 1 ELSE 0 END) * 100.0 / COUNT(*)) AS NullPercent
FROM [2021 Plane Accident];
```

RESULTS: 2021 Plane Accident Database

	ColumnName	NullCount	TotalRows	NullPercentage
1	Date	0	9	0.000000000000
2	Time	0	9	0.000000000000
3	Location	0	9	0.000000000000
4	Operator	0	9	0.000000000000
5	Flight	0	9	0.000000000000
6	Route	0	9	0.000000000000
7	AC_Type	0	9	0.000000000000
8	Registration	0	9	0.000000000000
9	Construction_no_line_no	0	9	0.000000000000
10	Aboard	0	9	0.000000000000
11	Fatalities	0	9	0.000000000000
12	Ground	0	9	0.000000000000
13	Summary	0	9	0.000000000000

SQL QUERIES: 2022 Plane Accident Database

```sql
USE PlaneAccidentDB;

SELECT
    'Date' AS ColumnName,
    SUM(CASE WHEN [Date] IS NULL THEN 1 ELSE 0 END) AS NullCount,
    COUNT(*) AS TotalRows,
    (SUM(CASE WHEN [Date] IS NULL THEN 1 ELSE 0 END) * 100.0 / COUNT(*)) AS NullPercentage
FROM [2022 Plane Accident]

UNION ALL

SELECT
    'Time' AS ColumnName,
    SUM(CASE WHEN [Time] IS NULL THEN 1 ELSE 0 END) AS NullCount,
    COUNT(*) AS TotalRows,
    (SUM(CASE WHEN [Time] IS NULL THEN 1 ELSE 0 END) * 100.0 / COUNT(*)) AS NullPercentage
FROM [2022 Plane Accident]

UNION ALL

SELECT
    'Location' AS ColumnName,
    SUM(CASE WHEN [Location] IS NULL THEN 1 ELSE 0 END) AS NullCount,
    COUNT(*) AS TotalRows,
    (SUM(CASE WHEN [Location] IS NULL THEN 1 ELSE 0 END) * 100.0 / COUNT(*)) AS NullPercen
FROM [2022 Plane Accident]
```

```sql
UNION ALL

SELECT
    'Operator' AS ColumnName,
    SUM(CASE WHEN [Operator] IS NULL THEN 1 ELSE 0 END) AS NullCount,
    COUNT(*) AS TotalRows,
    (SUM(CASE WHEN [Operator] IS NULL THEN 1 ELSE 0 END) * 100.0 / COUNT(*)) AS NullPercen
FROM [2022 Plane Accident]

UNION ALL

SELECT
    'Flight' AS ColumnName,
    SUM(CASE WHEN [Flight] IS NULL THEN 1 ELSE 0 END) AS NullCount,
    COUNT(*) AS TotalRows,
    (SUM(CASE WHEN [Flight] IS NULL THEN 1 ELSE 0 END) * 100.0 / COUNT(*)) AS NullPercenta
FROM [2022 Plane Accident]

UNION ALL

SELECT
    'Route' AS ColumnName,
    SUM(CASE WHEN [Route] IS NULL THEN 1 ELSE 0 END) AS NullCount,
    COUNT(*) AS TotalRows,
    (SUM(CASE WHEN [Route] IS NULL THEN 1 ELSE 0 END) * 100.0 / COUNT(*)) AS NullPercentag
FROM [2022 Plane Accident]
```

```
UNION ALL

SELECT
    'AC_Type' AS ColumnName,
    SUM(CASE WHEN [AC_Type] IS NULL THEN 1 ELSE 0 END) AS NullCount,
    COUNT(*) AS TotalRows,
    (SUM(CASE WHEN [AC_Type] IS NULL THEN 1 ELSE 0 END) * 100.0 / COUNT(*)) AS NullPercent
FROM [2022 Plane Accident]

UNION ALL

SELECT
    'Registration' AS ColumnName,
    SUM(CASE WHEN [Registration] IS NULL THEN 1 ELSE 0 END) AS NullCount,
    COUNT(*) AS TotalRows,
    (SUM(CASE WHEN [Registration] IS NULL THEN 1 ELSE 0 END) * 100.0 / COUNT(*)) AS NullPe
FROM [2022 Plane Accident]

UNION ALL

SELECT
    'Construction_no_line_no' AS ColumnName,
    SUM(CASE WHEN [Construction_no_line_no] IS NULL THEN 1 ELSE 0 END) AS NullCount,
    COUNT(*) AS TotalRows,
    (SUM(CASE WHEN [Construction_no_line_no] IS NULL THEN 1 ELSE 0 END) * 100.0 / COUNT(*)
FROM [2022 Plane Accident]
```

```
UNION ALL

SELECT
    'Aboard' AS ColumnName,
    SUM(CASE WHEN [Aboard] IS NULL THEN 1 ELSE 0 END) AS NullCount,
    COUNT(*) AS TotalRows,
    (SUM(CASE WHEN [Aboard] IS NULL THEN 1 ELSE 0 END) * 100.0 / COUNT(*)) AS NullPercenta
FROM [2022 Plane Accident]

UNION ALL

SELECT
    'Fatalities' AS ColumnName,
    SUM(CASE WHEN [Fatalities] IS NULL THEN 1 ELSE 0 END) AS NullCount,
    COUNT(*) AS TotalRows,
    (SUM(CASE WHEN [Fatalities] IS NULL THEN 1 ELSE 0 END) * 100.0 / COUNT(*)) AS NullPerc
FROM [2022 Plane Accident]

UNION ALL

SELECT
    'Ground' AS ColumnName,
    SUM(CASE WHEN [Ground] IS NULL THEN 1 ELSE 0 END) AS NullCount,
    COUNT(*) AS TotalRows,
    (SUM(CASE WHEN [Ground] IS NULL THEN 1 ELSE 0 END) * 100.0 / COUNT(*)) AS NullPercenta
FROM [2022 Plane Accident]
```

```
UNION ALL

SELECT
    'Summary' AS ColumnName,
    SUM(CASE WHEN [Summary] IS NULL THEN 1 ELSE 0 END) AS NullCount,
    COUNT(*) AS TotalRows,
    (SUM(CASE WHEN [Summary] IS NULL THEN 1 ELSE 0 END) * 100.0 / COUNT(*)) AS NullPercent
FROM [2022 Plane Accident];
```

RESULTS: 2022 Plane Accident Database

	ColumnName	NullCount	TotalRows	NullPercentage
1	Date	0	6	0.000000000000
2	Time	0	6	0.000000000000
3	Location	0	6	0.000000000000
4	Operator	0	6	0.000000000000
5	Flight	0	6	0.000000000000
6	Route	0	6	0.000000000000
7	AC_Type	0	6	0.000000000000
8	Registration	0	6	0.000000000000
9	Construction_no_line_no	0	6	0.000000000000
10	Aboard	0	6	0.000000000000
11	Fatalities	0	6	0.000000000000
12	Ground	0	6	0.000000000000
13	Summary	0	6	0.000000000000

SQL QUERIES: 2023 Plane Accident Database

```sql
USE PlaneAccidentDB;

SELECT
    'Date' AS ColumnName,
    SUM(CASE WHEN [Date] IS NULL THEN 1 ELSE 0 END) AS NullCount,
    COUNT(*) AS TotalRows,
    (SUM(CASE WHEN [Date] IS NULL THEN 1 ELSE 0 END) * 100.0 / COUNT(*)) AS NullPercentage
FROM [2023 Plane Accident]

UNION ALL

SELECT
    'Time' AS ColumnName,
    SUM(CASE WHEN [Time] IS NULL THEN 1 ELSE 0 END) AS NullCount,
    COUNT(*) AS TotalRows,
    (SUM(CASE WHEN [Time] IS NULL THEN 1 ELSE 0 END) * 100.0 / COUNT(*)) AS NullPercentage
FROM [2023 Plane Accident]

UNION ALL

SELECT
    'Location' AS ColumnName,
    SUM(CASE WHEN [Location] IS NULL THEN 1 ELSE 0 END) AS NullCount,
    COUNT(*) AS TotalRows,
    (SUM(CASE WHEN [Location] IS NULL THEN 1 ELSE 0 END) * 100.0 / COUNT(*)) AS NullPercen
FROM [2023 Plane Accident]
```

```sql
UNION ALL

SELECT
    'Operator' AS ColumnName,
    SUM(CASE WHEN [Operator] IS NULL THEN 1 ELSE 0 END) AS NullCount,
    COUNT(*) AS TotalRows,
    (SUM(CASE WHEN [Operator] IS NULL THEN 1 ELSE 0 END) * 100.0 / COUNT(*)) AS NullPercen
FROM [2023 Plane Accident]

UNION ALL

SELECT
    'Flight' AS ColumnName,
    SUM(CASE WHEN [Flight] IS NULL THEN 1 ELSE 0 END) AS NullCount,
    COUNT(*) AS TotalRows,
    (SUM(CASE WHEN [Flight] IS NULL THEN 1 ELSE 0 END) * 100.0 / COUNT(*)) AS NullPercenta
FROM [2023 Plane Accident]

UNION ALL

SELECT
    'Route' AS ColumnName,
    SUM(CASE WHEN [Route] IS NULL THEN 1 ELSE 0 END) AS NullCount,
    COUNT(*) AS TotalRows,
    (SUM(CASE WHEN [Route] IS NULL THEN 1 ELSE 0 END) * 100.0 / COUNT(*)) AS NullPercentag
```

```
UNION ALL

SELECT
    'AC_Type' AS ColumnName,
    SUM(CASE WHEN [AC_Type] IS NULL THEN 1 ELSE 0 END) AS NullCount,
    COUNT(*) AS TotalRows,
    (SUM(CASE WHEN [AC_Type] IS NULL THEN 1 ELSE 0 END) * 100.0 / COUNT(*)) AS NullPercent
FROM [2023 Plane Accident]

UNION ALL

SELECT
    'Registration' AS ColumnName,
    SUM(CASE WHEN [Registration] IS NULL THEN 1 ELSE 0 END) AS NullCount,
    COUNT(*) AS TotalRows,
    (SUM(CASE WHEN [Registration] IS NULL THEN 1 ELSE 0 END) * 100.0 / COUNT(*)) AS NullPe
FROM [2023 Plane Accident]

UNION ALL

SELECT
    'Construction_no_line_no' AS ColumnName,
    SUM(CASE WHEN [Construction_no_line_no] IS NULL THEN 1 ELSE 0 END) AS NullCount,
    COUNT(*) AS TotalRows,
    (SUM(CASE WHEN [Construction_no_line_no] IS NULL THEN 1 ELSE 0 END) * 100.0 / COUNT(*)
FROM [2023 Plane Accident]
```

```
UNION ALL

SELECT
    'Aboard' AS ColumnName,
    SUM(CASE WHEN [Aboard] IS NULL THEN 1 ELSE 0 END) AS NullCount,
    COUNT(*) AS TotalRows,
    (SUM(CASE WHEN [Aboard] IS NULL THEN 1 ELSE 0 END) * 100.0 / COUNT(*)) AS NullPercenta
FROM [2023 Plane Accident]

UNION ALL

SELECT
    'Fatalities' AS ColumnName,
    SUM(CASE WHEN [Fatalities] IS NULL THEN 1 ELSE 0 END) AS NullCount,
    COUNT(*) AS TotalRows,
    (SUM(CASE WHEN [Fatalities] IS NULL THEN 1 ELSE 0 END) * 100.0 / COUNT(*)) AS NullPerc
FROM [2023 Plane Accident]

UNION ALL

SELECT
    'Ground' AS ColumnName,
    SUM(CASE WHEN [Ground] IS NULL THEN 1 ELSE 0 END) AS NullCount,
    COUNT(*) AS TotalRows,
    (SUM(CASE WHEN [Ground] IS NULL THEN 1 ELSE 0 END) * 100.0 / COUNT(*)) AS NullPercenta
FROM [2023 Plane Accident]
```

```sql
SELECT
    'Summary' AS ColumnName,
    SUM(CASE WHEN [Summary] IS NULL THEN 1 ELSE 0 END) AS NullCount,
    COUNT(*) AS TotalRows,
    (SUM(CASE WHEN [Summary] IS NULL THEN 1 ELSE 0 END) * 100.0 / COUNT(*)) AS NullPercent
FROM [2023 Plane Accident];
```

RESULTS: 2023 Plane Accident Database

	ColumnName	NullCount	TotalRows	NullPercentage
1	Date	0	4	0.000000000000
2	Time	1	4	25.000000000000
3	Location	0	4	0.000000000000
4	Operator	0	4	0.000000000000
5	Flight	0	4	0.000000000000
6	Route	0	4	0.000000000000
7	AC_Type	0	4	0.000000000000
8	Registration	0	4	0.000000000000
9	Construction_no_line_no	0	4	0.000000000000
10	Aboard	0	4	0.000000000000
11	Fatalities	0	4	0.000000000000
12	Ground	0	4	0.000000000000
13	Summary	0	4	0.000000000000