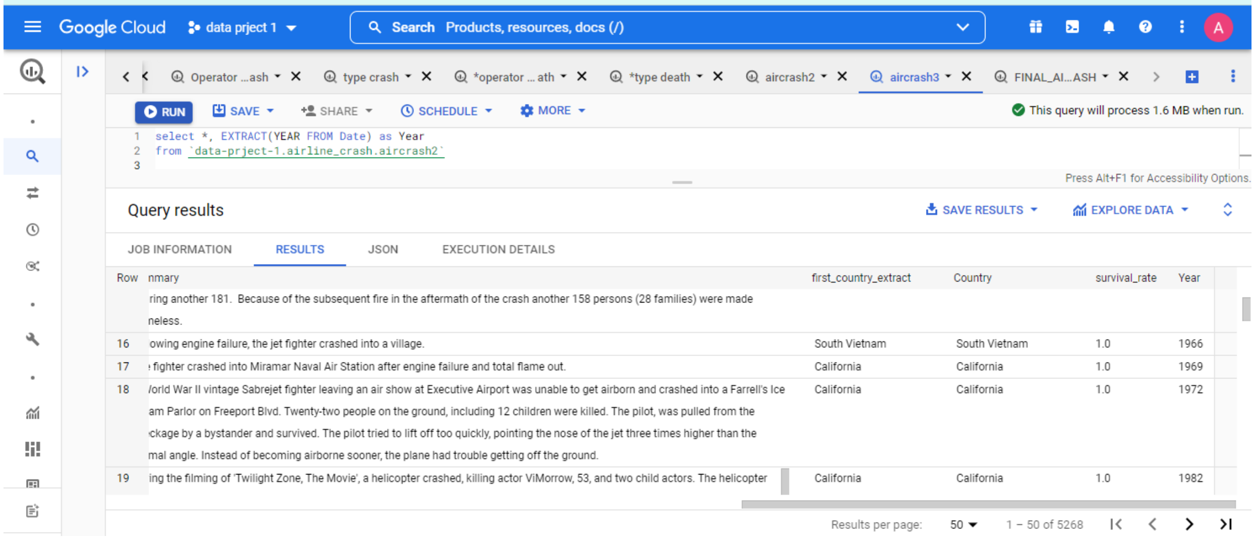
PRE-ANALYSIS

To derive more insights from the data, I created some extra columns from the existing ones using SQL. This also helped in my exploratory analysis to get initial findings from the data.

The added columns and reasons for creating them are discussed below.

Year Column

Here, I extracted each crash year from the date column to be able to visualize yearly trends.



This query will process 1.6 MB when run.

```
select *, EXTRACT(YEAR FROM Date) as Year
from `data-prject-1.airline_crash.aircrash2`
```

Press Alt+F1 for Accessibility Options.

Query results

Row	nmary	first_country_extract	Country	survival_rate	Year
	ring another 181. Because of the subsequent fire in the aftermath of the crash another 158 persons (28 families) were made neless.				
16	owing engine failure, the jet fighter crashed into a village.	South Vietnam	South Vietnam	1.0	1966
17	fighter crashed into Miramar Naval Air Station after engine failure and total flame out.	California	California	1.0	1969
18	/orld War II vintage Sabrejet fighter leaving an air show at Executive Airport was unable to get airborn and crashed into a Farrell's Ice am Parlor on Freeport Blvd. Twenty-two people on the ground, including 12 children were killed. The pilot, was pulled from the ckage by a bystander and survived. The pilot tried to lift off too quickly, pointing the nose of the jet three times higher than the mal angle. Instead of becoming airborne sooner, the plane had trouble getting off the ground.	California	California	1.0	1972
19	ing the filming of 'Twilight Zone, The Movie', a helicopter crashed, killing actor ViMorrow, 53, and two child actors. The helicopter	California	California	1.0	1982

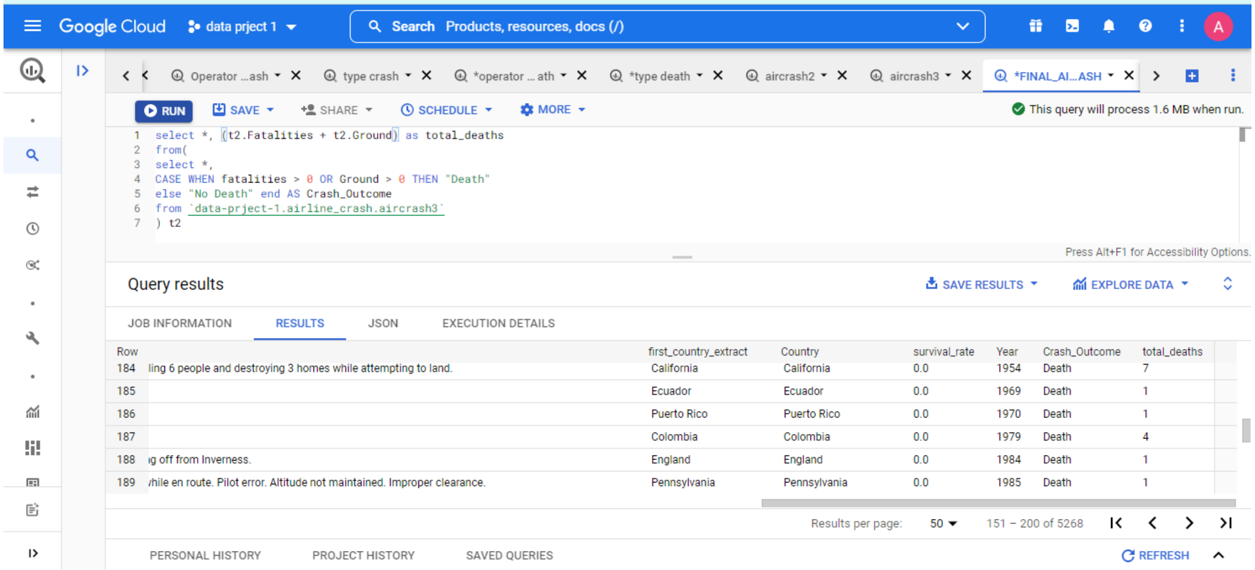
Results per page: 50 1 – 50 of 5268

Month Column

Extracted each rows month name from the date column. This was done during data transformation in power query.

Total Deaths

This column shows the total number of deaths for each crash. This column was derived by adding the fatalities and ground columns. This was needed to get the death trend over the years.



This query will process 1.6 MB when run.

```
select *, (t2.Fatalities + t2.Ground) as total_deaths
from(
select *,
CASE WHEN fatalities > 0 OR Ground > 0 THEN "Death"
else "No Death" end AS Crash_Outcome
from `data-prject-1.airline_crash.aircrash3`
) t2
```

Press Alt+F1 for Accessibility Options.

Query results

Row		first_country_extract	Country	survival_rate	Year	Crash_Outcome	total_deaths
184	ling 6 people and destroying 3 homes while attempting to land.	California	California	0.0	1954	Death	7
185		Ecuador	Ecuador	0.0	1969	Death	1
186		Puerto Rico	Puerto Rico	0.0	1970	Death	1
187		Colombia	Colombia	0.0	1979	Death	4
188	ıg off from Inverness.	England	England	0.0	1984	Death	1
189	/hile en route. Pilot error. Altitude not maintained. Improper clearance.	Pennsylvania	Pennsylvania	0.0	1985	Death	1

Results per page: 50 151 – 200 of 5268

Survived count

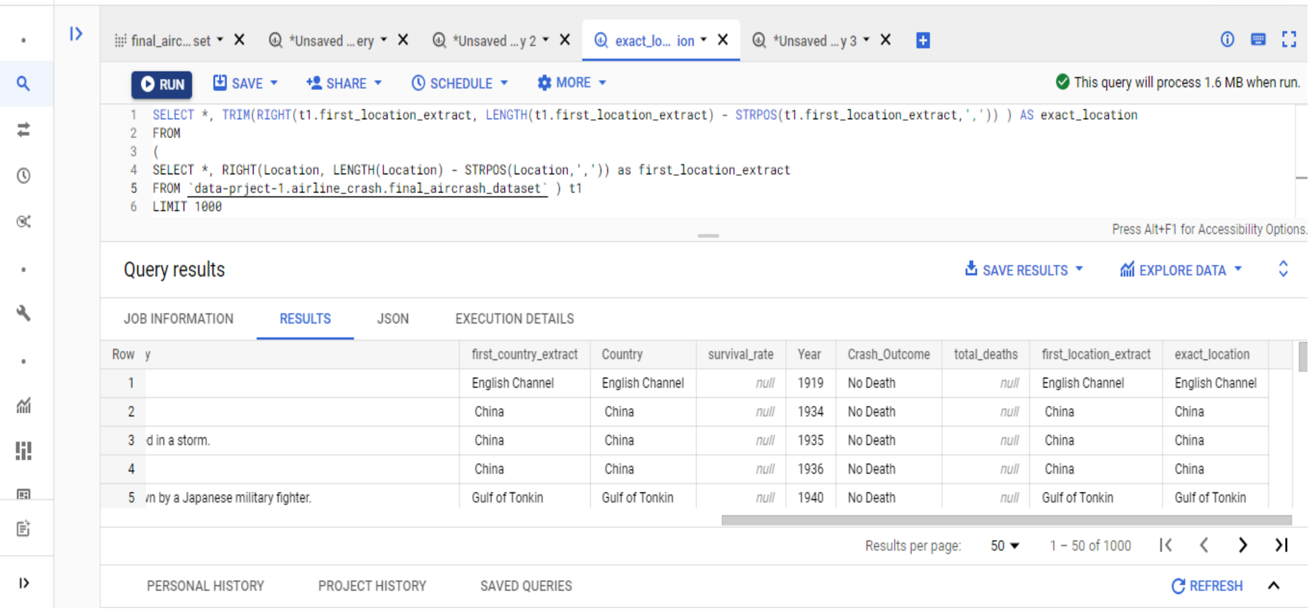
This column was created in power query. The survivors count was created by subtracting number of fatalities from number of aboard passengers and crew.

Survived or not

This column tells us whether or not the crash had survivors. The resulting column returns 0 for crash with no survivors and 1 for crash with survivors. Taking the count of these categories will help in analysis on rate of survivors.

Exact Location

This column contains the drilled down exact location of the crash. The values in this column is taken from the location column using the string position function to get the last string in the locations and the trim function to remove whitespaces.



```
SELECT *, TRIM(RIGHT(t1.first_location_extract, LENGTH(t1.first_location_extract) - STRPOS(t1.first_location_extract,',')) ) AS exact_location
FROM
(
SELECT *, RIGHT(Location, LENGTH(Location) - STRPOS(Location,',')) as first_location_extract
FROM `data-prject-1.airline_crash.final_aircrash_dataset` ) t1
LIMIT 1000
```

Row	y	first_country_extract	Country	survival_rate	Year	Crash_Outcome	total_deaths	first_location_extract	exact_location
1		English Channel	English Channel	null	1919	No Death	null	English Channel	English Channel
2		China	China	null	1934	No Death	null	China	China
3	d in a storm.	China	China	null	1935	No Death	null	China	China
4		China	China	null	1936	No Death	null	China	China
5	ın by a Japanese military fighter.	Gulf of Tonkin	Gulf of Tonkin	null	1940	No Death	null	Gulf of Tonkin	Gulf of Tonkin

NOTE

Initially, the dataset contained 5,268 rows. During data transformation in power query, 89 rows were dropped or filtered. These rows were dropped because they had null values in either of the following columns: fatalities, aboard, ground, country, operator and type. Columns not needed for the analysis were also removed. The dataset used for the analysis therefore contained 5,179 rows after cleaning.