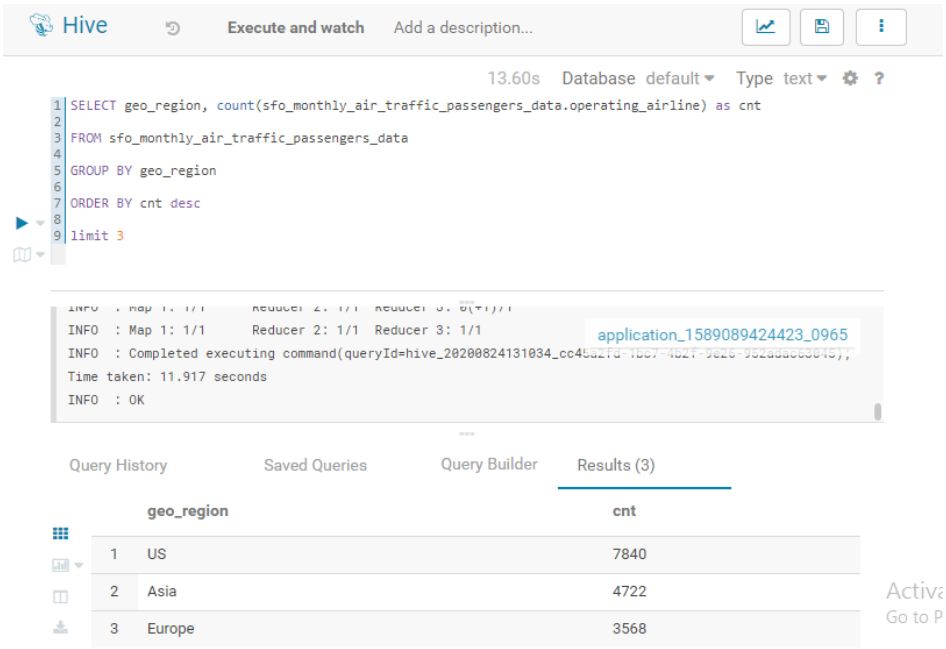
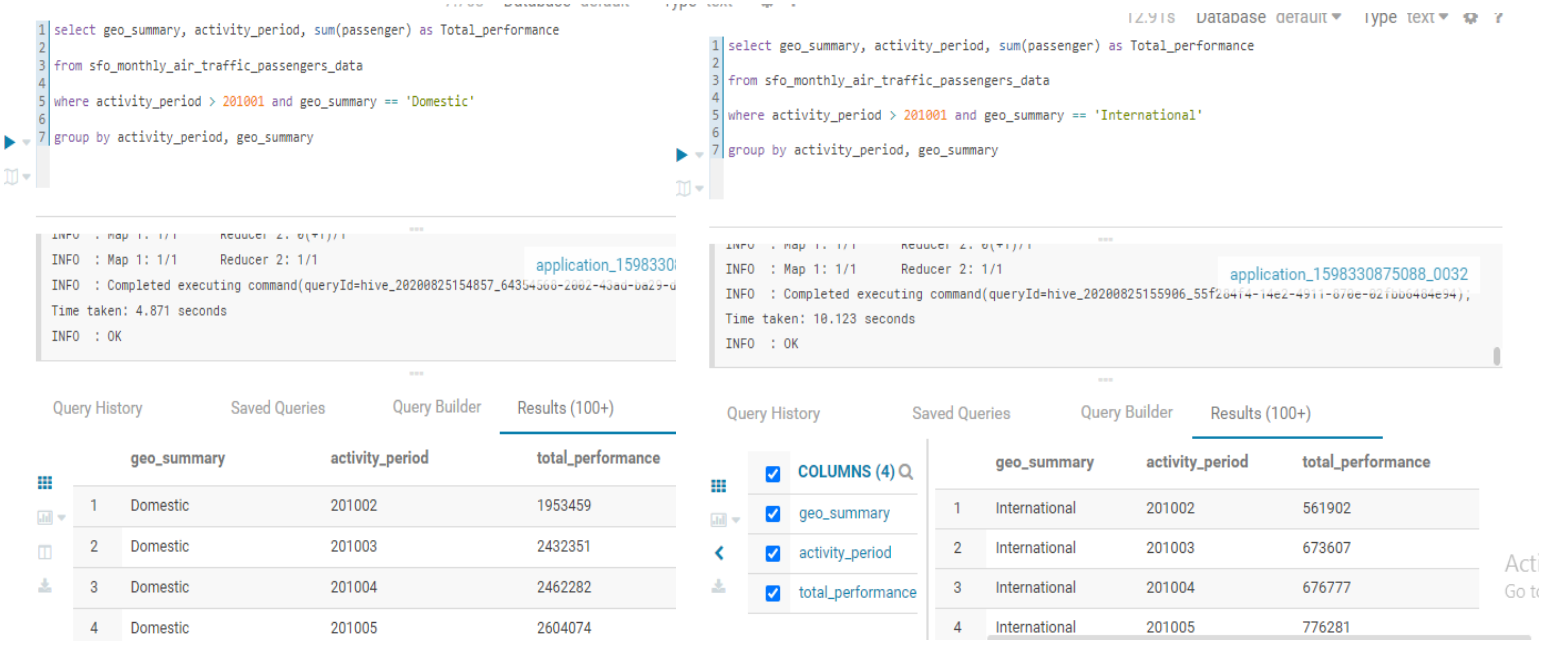
Final Project

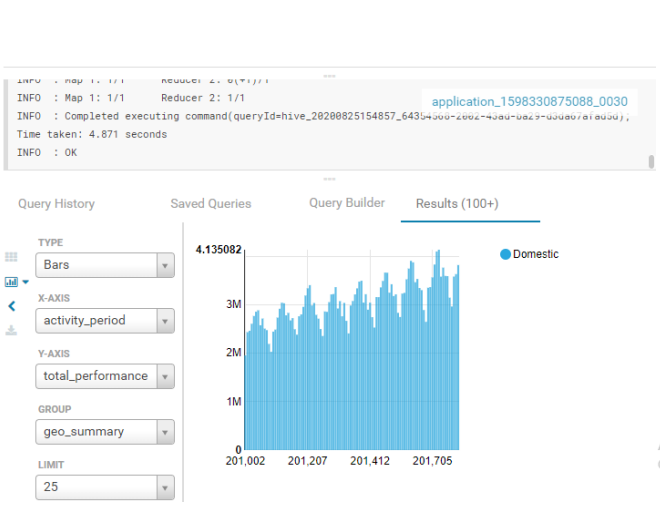
1. What where the top 3 busiest Geo Regions Flying in and out of SFO?

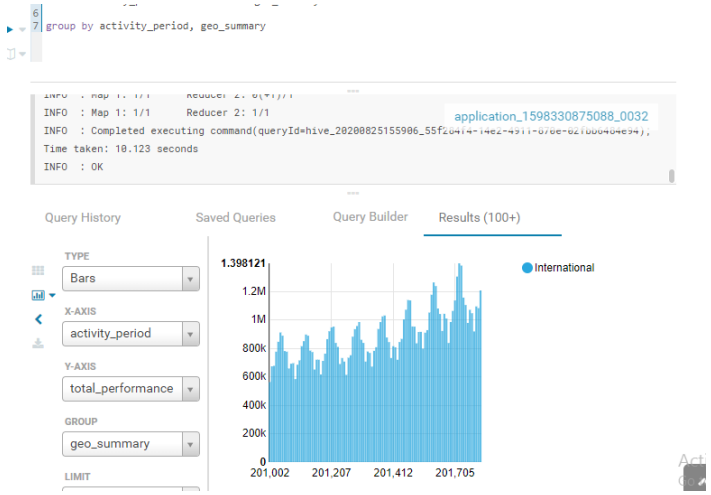


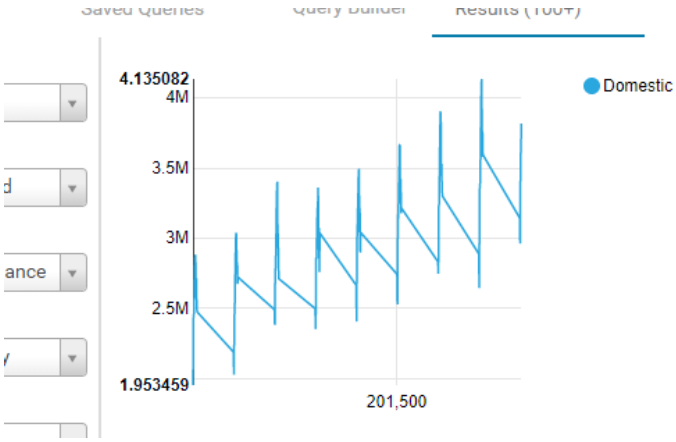
Answer: The top 3 busiest Geo Regions Flying in and out of SFO were US, Asia and Europe according to our results.

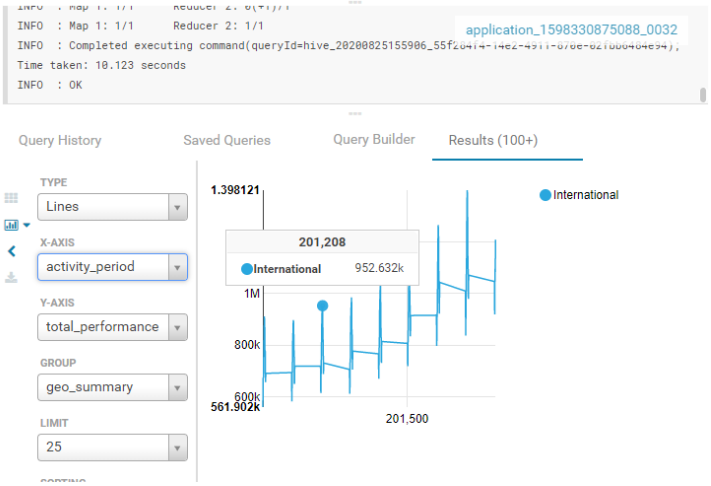
2. How do SFO international flights compare to its domestic flights over the past 10 years? How has it been evolving over the past 4 decades?





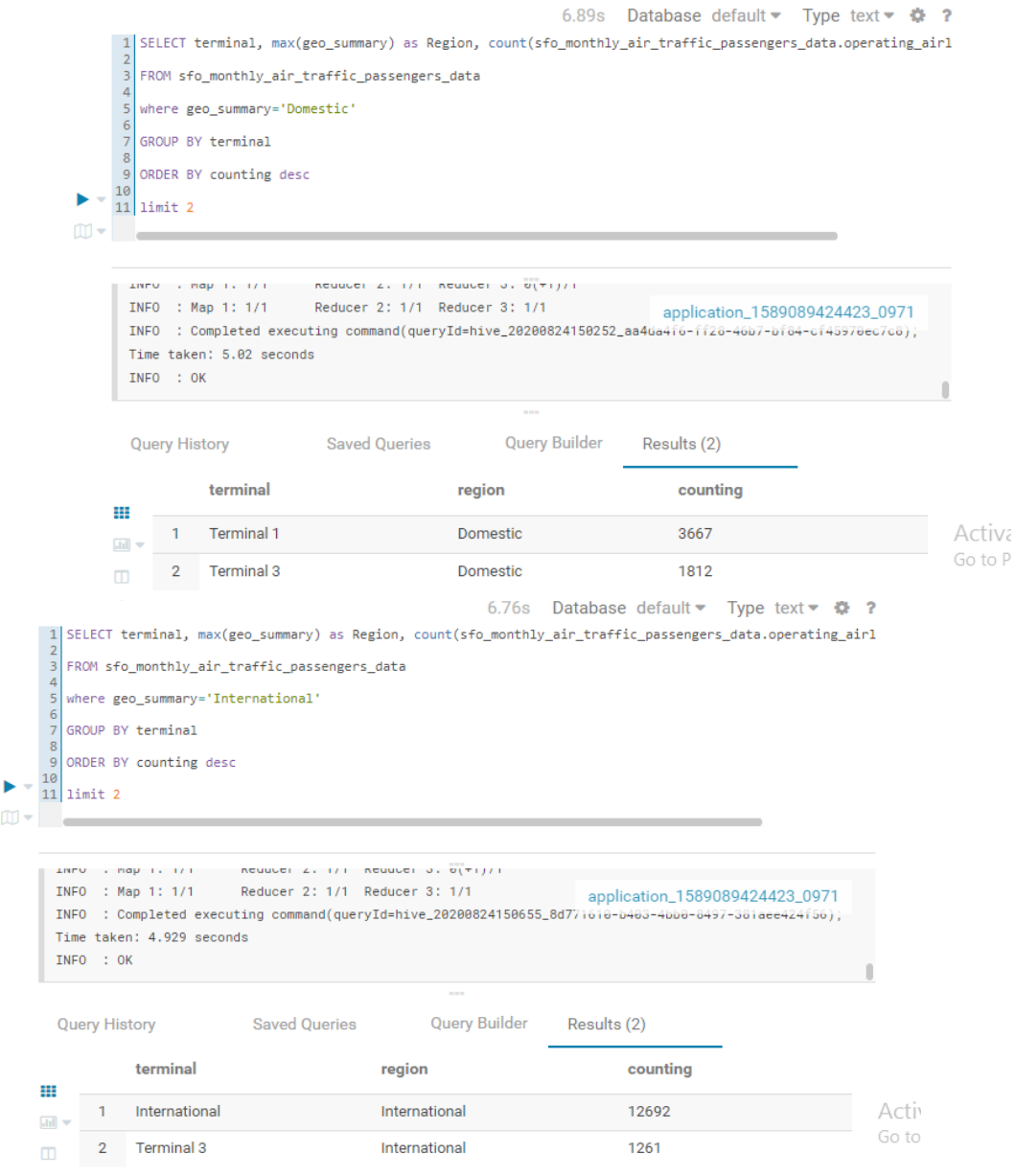






Our results show that there is a clear (seasonal) pattern at an increasing rate over the analyzed period in the number of both domestic and international flights in SFO. The pattern increases till the middle of the year and starts to decrease towards the December. It kind of reveals that the busiest time in SFO is the middle of the year , such as June and July. Moreover, interestingly, SFO has conducted more of domestic flights than international flights within the given time span. (Which I believed it quite opposite). From this result, we can also comment that in June and July, the flow of money in SF is quite tense meaning the budget of SF increases to its peak every time in these periods because more people are in the city coming in and out, they have to book a room and have some food at restaurant, potential transport rentals, amusement parks and museums, which is a lot of money for SF.

3. Which SFO Terminal is the busiest for Domestic, what about international; Why?



```
6.89s  Database default  Type text

SELECT terminal, max(geo_summary) as Region, count(sfo_monthly_air_traffic_passengers_data.operating_airl

FROM sfo_monthly_air_traffic_passengers_data

where geo_summary='Domestic'

GROUP BY terminal

ORDER BY counting desc

limit 2
```

```
INFO  : Map 1: 1/1      Reducer 2: 1/1  Reducer 3: 1/1
INFO  : Completed executing command(queryId=hive_20200824150252_aa4ba4f8-1f28-46b7-bf84-ef459f0ec7c8);
Time taken: 5.02 seconds
INFO  : OK
```

application_1589089424423_0971

Query History | Saved Queries | Query Builder | Results (2)

	terminal	region	counting
1	Terminal 1	Domestic	3667
2	Terminal 3	Domestic	1812

```
6.76s  Database default  Type text

SELECT terminal, max(geo_summary) as Region, count(sfo_monthly_air_traffic_passengers_data.operating_airl

FROM sfo_monthly_air_traffic_passengers_data

where geo_summary='International'

GROUP BY terminal

ORDER BY counting desc

limit 2
```

```
INFO  : Map 1: 1/1      Reducer 2: 1/1  Reducer 3: 1/1
INFO  : Completed executing command(queryId=hive_20200824150655_8d7f6f0-b483-4bb6-8497-381aee424f56);
Time taken: 4.929 seconds
INFO  : OK
```

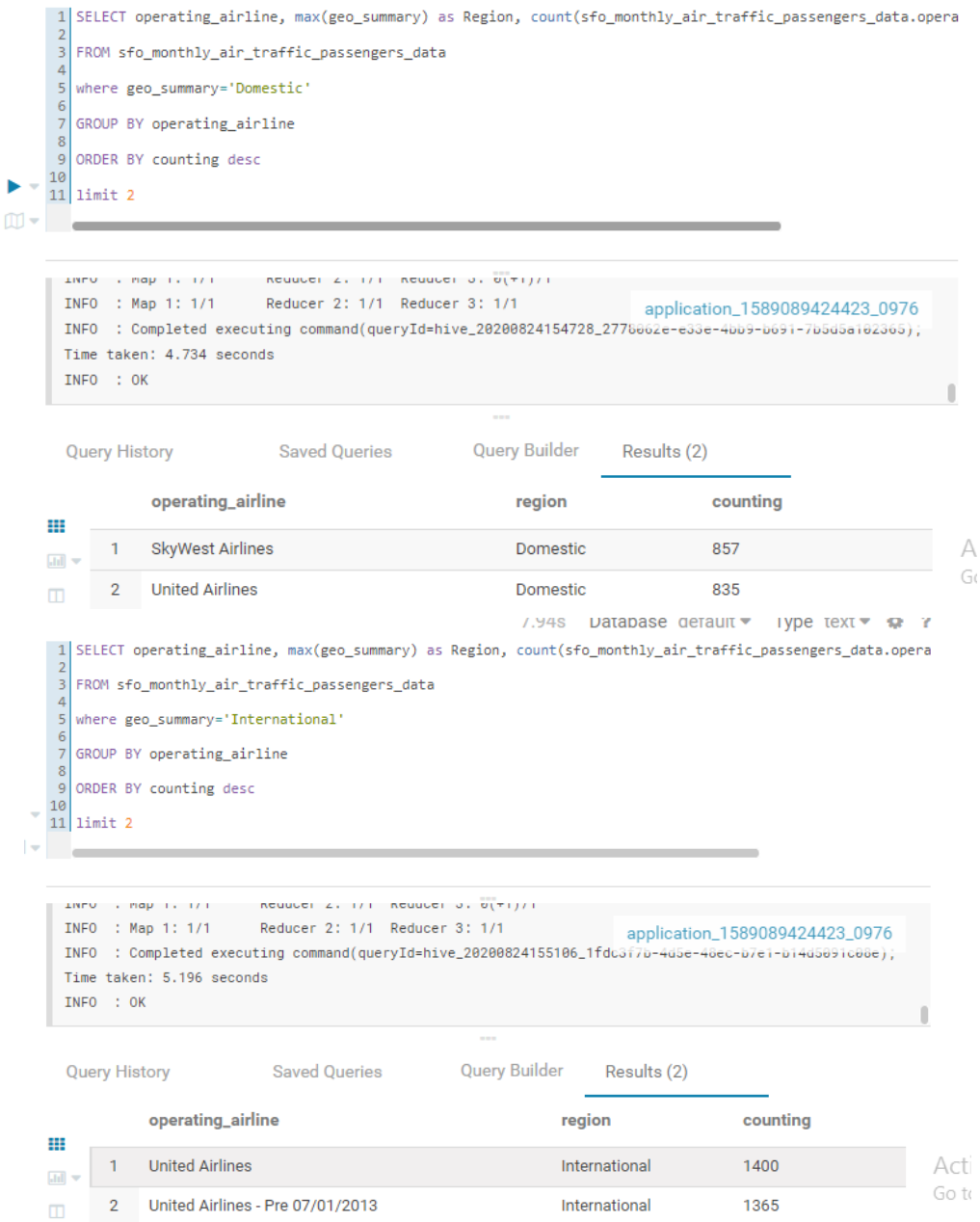
application_1589089424423_0971

Query History | Saved Queries | Query Builder | Results (2)

	terminal	region	counting
1	International	International	12692
2	Terminal 3	International	1261

Answer: the busiest terminal for international flights is terminal International and for domestic flight, it is Terminal 1. It may seem obvious why International Terminal is the busiest because as the name suggests, the main flow in and out is held through this terminal. Another reason could be that all big international airlines use this terminal or this place may have some locational advantages like near to road, bigger flying areas and etc. Regarding the domestic terminal 1, this terminal may have some locational advantages as well, when I look at the map, I saw, it is closer to Intl Terminal and to the city as well, so it is easier for passengers to get out of the airport.

4. Which Airline is the busiest for Domestic, what about international; Why?



```
SELECT operating_airline, max(geo_summary) as Region, count(sfo_monthly_air_traffic_passengers_data.opera

FROM sfo_monthly_air_traffic_passengers_data

where geo_summary='Domestic'

GROUP BY operating_airline

ORDER BY counting desc

limit 2
```

```
INFO  : Map 1: 1/1      Reducer 2: 1/1  Reducer 3: 1/1
INFO  : Completed executing command(queryId=hive_20200824154728_277b862e-e33e-4bb9-b691-7b5d5a102365);
Time taken: 4.734 seconds
INFO  : OK
```

application_1589089424423_0976

Query History | Saved Queries | Query Builder | Results (2)

	operating_airline	region	counting
1	SkyWest Airlines	Domestic	857
2	United Airlines	Domestic	835

```
SELECT operating_airline, max(geo_summary) as Region, count(sfo_monthly_air_traffic_passengers_data.opera

FROM sfo_monthly_air_traffic_passengers_data

where geo_summary='International'

GROUP BY operating_airline

ORDER BY counting desc

limit 2
```

```
INFO  : Map 1: 1/1      Reducer 2: 1/1  Reducer 3: 1/1
INFO  : Completed executing command(queryId=hive_20200824155106_1fdc5f7b-4d5e-40ec-b7e1-b14d5691c80e);
Time taken: 5.196 seconds
INFO  : OK
```

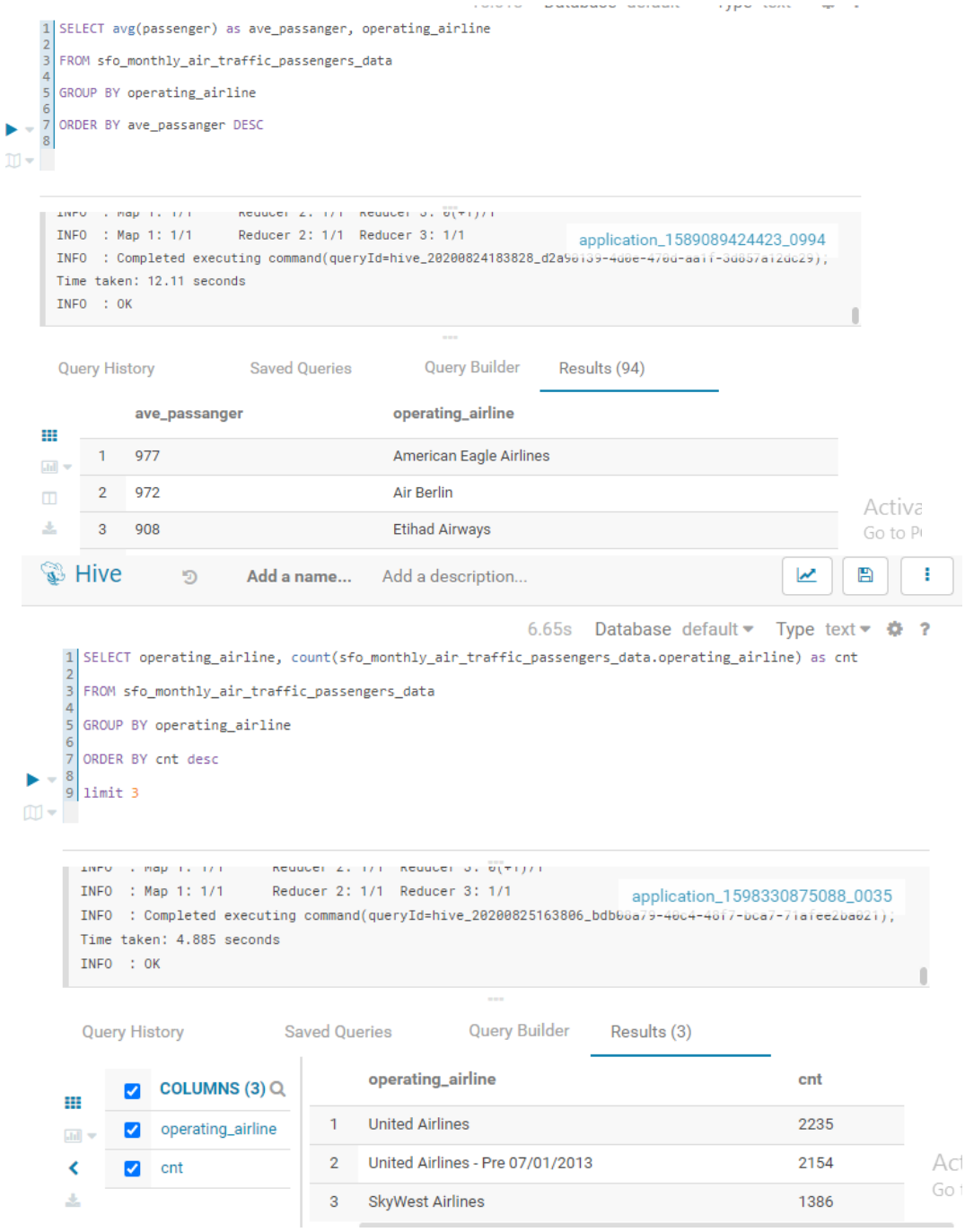
application_1589089424423_0976

Query History | Saved Queries | Query Builder | Results (2)

	operating_airline	region	counting
1	United Airlines	International	1400
2	United Airlines - Pre 07/01/2013	International	1365

Answer: The busiest operating airline for domestic use is SkyWest Airlines and for International, it is United Airlines. The reason why United Airlines is on the top could be the incentives it can offer to passengers. As the company has really big operations, it has achieved economies of scale, thus reduced cost for passengers. Coming to SkyWest Airlines, it is a popular choice for North American Population, the reason could be it is one of the oldest airlines operating in North America being founded in 1972 and it has partnership cooperation with Delta (the oldest). So, it has gained more experience and has its own reputation in the industry.

5. What is the average passenger count per airline? Who has the max number of passengers and why do you think that is?



If our query is correct, we can say that United Airlines has the max number of passengers and potential reason could be again incentives it offers to its customers like reduced flight costs or discounted prices, maybe loyalty programs it may have so that it can attract more customers.

Q6: Which where the top 3 years and the lowest 3 years where SFO had the largest number of passengers? What could be the driver behind that?

```
SELECT activity_period, max(passenger) as max_number
FROM sfo_monthly_air_traffic_passengers_data
GROUP BY activity_period
SORT BY max_number desc
limit 3
```

Execute or CTRL + ENTER

```
INFO  : Map 1: 1/1      Reducer 2: 1/1  Reducer 3: 1/1  Reducer 4: 1/1
INFO  : Completed executing command(queryId=hive_20200825175243_2dd5d78e-69ac-4865-9727-27dd65bb4c42);
Time taken: 5.109 seconds
INFO  : OK
```

application_1598330875088_0047

Query History | Saved Queries | Query Builder | Results (3)

COLUMNS (3)
- activity_period
- max_number

	activity_period	max_number
1	200907	9999
2	201502	9999
3	201201	9998

```
SELECT activity_period, min(passenger) as min_number
FROM sfo_monthly_air_traffic_passengers_data
GROUP BY activity_period
SORT BY min_number ASC
limit 3
```

```
INFO  : Map 1: 1/1      Reducer 2: 1/1  Reducer 3: 1/1  Reducer 4: 1/1
INFO  : Completed executing command(queryId=hive_20200825175433_0ff9d9d8-8d89-40a9-82eb-6526a6b6a8c4);
Time taken: 4.831 seconds
INFO  : OK
```

application_1598330875088_0047

Query History | Saved Queries | Query Builder | Results (3)

COLUMNS (3)
- activity_period
- min_number

	activity_period	min_number
1	200611	1
2	201811	1
3	200802	1

This query indicates that July, 2019; February, 2015 and January, 2012 are the top three years when SF had the largest number of passenger and the rive for that could be any big sporting event was hold during the time so more people wanted to come. Whereas, 2006, 2018, 2008 were the periods where SF had the lowest number of passengers recorder. As I was expecting from the result of second question (seasonal pattern), near new year period is slow in Airline industry in SF, maybe people prefer going somewhere else colder than SF to really feel new year. This is just subjective opinion.

Q7:Let's analyze the impact of COVID-19 on the travel industry so far. Can you compare the number of flights and number of passengers in the first quarter of 2020 (Jan+Fev+March 2020) Dataset to the previous year first quarter (Jan+Fev+March 2019). What can you say about the impact of COVID19 on the travel industry?

```
select activity_period, count(*) as flights, sum(passenger) as passengers

from sfo_monthly_air_traffic_passengers_data

where activity_period >= 202001 and activity_period <= 202003

group by activity_period
```

```
INFO  : Map 1: 1/1      Reducer 2: 1/1
INFO  : Completed executing command(queryId=hive_20200825181717_e71549aa-e93d-4c10-a265-65614ce78670);
Time taken: 9.389 seconds
INFO  : OK
```

application_1598330875088_0055

Query History | Saved Queries | Query Builder | Results (3)

COLUMNS (4)
- activity_period
- flights
- passengers

	activity_period	flights	passengers
1	202001	149	4241751
2	202002	146	3742224
3	202003	140	1887581

11.76s Database default Type text

```
select activity_period, count(*) as flights, sum(passenger) as passengers

from sfo_monthly_air_traffic_passengers_data

where activity_period >= 201901 and activity_period <= 201903

group by activity_period
```

```
INFO  : Map 1: 1/1      Reducer 2: 1/1
INFO  : Completed executing command(queryId=hive_20200825181928_65b7[illegible]);
Time taken: 6.801 seconds
INFO  : OK
```

application_1598330875088_0055

Query History | Saved Queries | Query Builder | Results (3)

COLUMNS (4)
- activity_period
- flights
- passengers

	activity_period	flights	passengers
1	201901	148	4156821
2	201902	143	3752763
3	201903	150	4599189

As it can be seen from the table, we don't see any big difference in January and February between these two years, but starting March, when the actual Covid-19 is believed a serious case, the number of flights had to be canceled and did the number of passengers decrease noticeable. As we know, the stock of airline industry has been impacted dramatically by Covid-19 and they have lost a lot of money.