MSBA 321: Big Data Ecosystem

Sweta Kumari

Department of Business Analytics, Golden Gate University

August 25, 2020

Table of Contents

Introduction

San Francisco Airport (SFO) is the 7th busiest airport in the United States and 24th busiest airport in the world by passenger count. It is a major gateway to both Europe and Asia; it is also a maintenance hub for United Airlines. Due to the services/facilities, the number of connectivity of several regions, SFO is the commercial center of Northern California. To improve the passengers' experience, the county of San Francisco has created several road maps (flysfo, n.d.). This paper gives insight for the SFO airport that can be used by Airport divisions who work for the improvement plan of SFO airport.

Data Source

The data source is retrieved from the eLearning website of the Golden Gate University provided by the professor. This dataset is originally released by the county of San Francisco for public usage. The aim of releasing the data about SFO airport is to analyze the data and provide insight that can help in improving the San Francisco airport functionality.

The dataset describes the "San Francisco International Airport Report on Monthly Passenger Traffic Statistics by Airline" (DataSF, 2020). It has 22.121k rows with 12 columns. The goal of the analysis is to give insight for the SFO airport with the help of statistical analysis.

Statistical Analysis and Approach

The statistical analysis is performed based on the business questions such as which is the busiest terminal for the International arrival/departure. All queries are performed in the Hive and results are retrieved from the same. To enhance the user experience, the data visualization tool (Tableau) is used. Descriptive statistics are used to answer business questions.

Descriptive Statistics

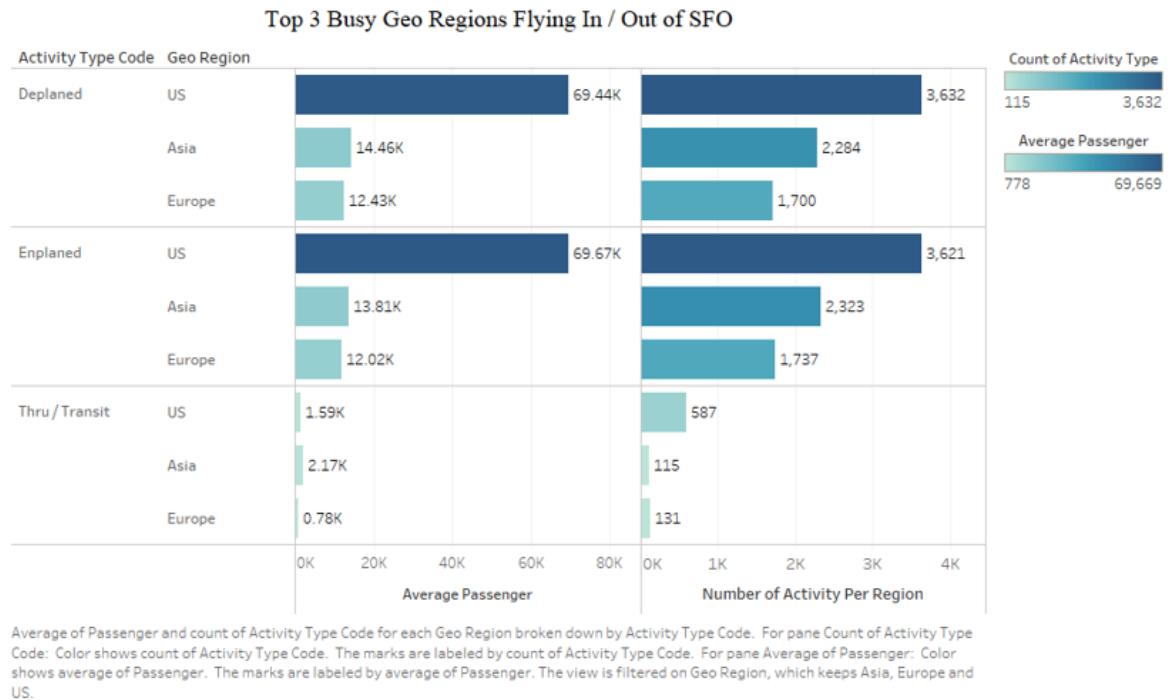1. What were the top 3 busiest Geo Regions Flying in and out of SFO?

We analyze the busiest regions based on the average number of passengers traveled in and out of SFO airport. There are three types of actions mentioned in a relation to flight which include boarding a flight / "Enplaned", getting off a flight / "deplaned" and transiting to another location / "Thru/Transit". Below is the query and result:

```
SELECT * FROM
(SELECT activity_type_code, geo_region AS Busiest_Region,
ROUND((AVG(CAST(REPLACE(passenger, ',', '') AS INT))/1000),2)
AS Average_Passengers_in_Thousands,
COUNT(geo_region) AS Count_of_Flights,
ROUND((SUM(CAST(REPLACE(passenger, ',', '') AS INT))/1000000),2)
AS Total_Passengers_in_Millions,
ROW_NUMBER() OVER (PARTITION BY activity_type_code
ORDER BY AVG(CAST(REPLACE(passenger, ',', '') AS INT)) DESC) AS Rank_Of_Busy_Region
FROM sfo_monthly_air_traffic_passengers_data
GROUP BY geo_region, activity_type_code)
Region WHERE Rank_Of_Busy_Region <=3;
```

	region.activity_type_code	region.busiest_region	region.average_passengers_in_thousands	region.count_of_flights	region.total_passengers_in_millions	region.rank_of_busy_region
1	Deplaned	US	69.44	3632	252.21	1
2	Deplaned	Asia	14.46	2284	33.03	2
3	Deplaned	Europe	12.43	1700	21.13	3
4	Enplaned	US	69.67	3621	252.27	1
5	Enplaned	Asia	13.81	2323	32.09	2
6	Enplaned	Europe	12.02	1737	20.88	3
7	Thru / Transit	Asia	2.17	115	0.25	1
8	Thru / Transit	US	1.59	587	0.93	2
9	Thru / Transit	Australia / Oceania	0.92	115	0.11	3

Top 3 Busy Geo Regions Flying In / Out of SFO:

In the below graph, Top 3 regions are shown based on the average number of passengers with each type of passenger activity.

Top 3 Busy Geo Regions Flying In / Out of SFO

Average of Passenger and count of Activity Type Code for each Geo Region broken down by Activity Type Code. For pane Count of Activity Type Code: Color shows count of Activity Type Code. The marks are labeled by count of Activity Type Code. For pane Average of Passenger: Color shows average of Passenger. The marks are labeled by average of Passenger. The view is filtered on Geo Region, which keeps Asia, Europe and US.

Interpretation:

Deplaned/ Enplaned Activity: The top 3 busiest regions are US, Asia, and Europe having maximum average passengers with the maximum number of flights deplaned/enplaned.

According to ClaimCompass, SFO is a major gateway to both Europe and Asia (ClaimCompass, 2020). The above analysis supports the stats given by ClaimCompass.

2. How does SFO international flights compares to its domestic flights over the past 10 years? How has it been evolving over the past 4 decades?
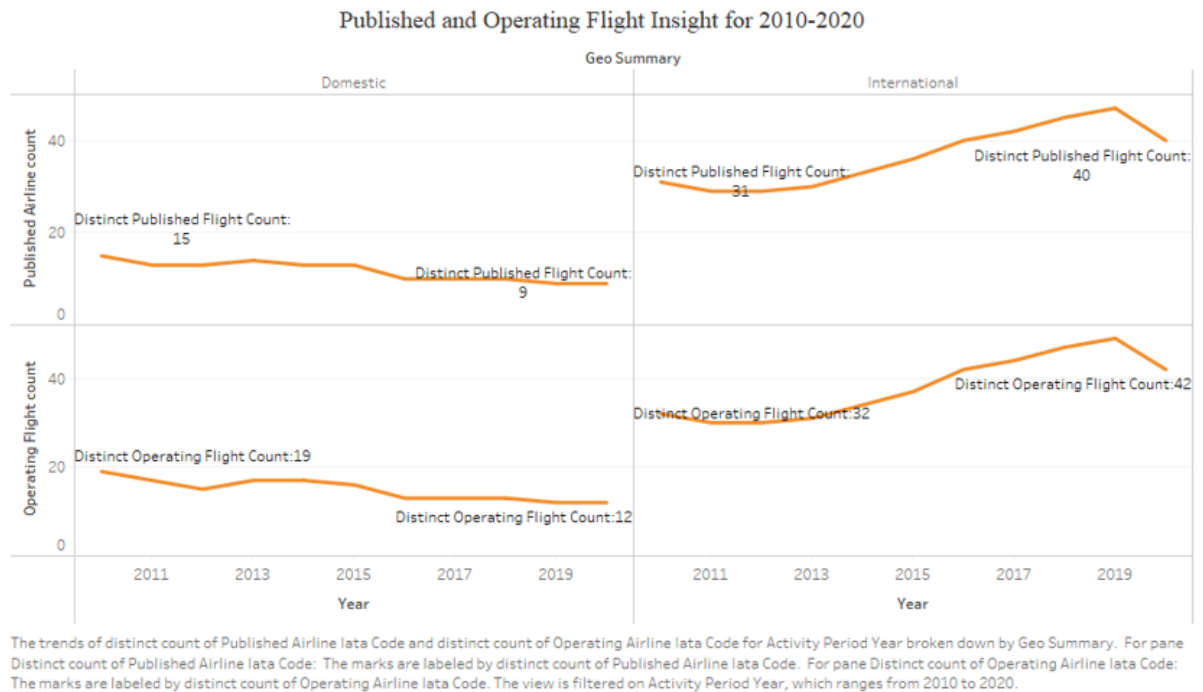
The query for operating as well as published airlines:

```
SELECT SUBSTR(activity_period,1, 4) AS Year, geo_summary,
COUNT(distinct(published_airline_iata_code)) AS Published_flight_count,
COUNT(distinct(operating_airline_iata_code)) AS Operating_flight_count
FROM sfo_monthly_air_traffic_passengers_data
WHERE CAST(SUBSTR(activity_period,1, 4) AS INT) >=2010
GROUP BY SUBSTR(activity_period,1, 4), geo_summary;
```

Below are the few lines of result:

	year	geo_summary	published_flight_count	operating_flight_count
1	2010	Domestic	16	20
2	2010	International	32	33
3	2011	Domestic	14	18
4	2011	International	30	31

Published and Operating Flight Insight for 2010-2020:

Published and Operating Flight Insight for 2010-2020

The trends of distinct count of Published Airline Iata Code and distinct count of Operating Airline Iata Code for Activity Period Year broken down by Geo Summary. For pane Distinct count of Published Airline Iata Code: The marks are labeled by distinct count of Published Airline Iata Code. For pane Distinct count of Operating Airline Iata Code: The marks are labeled by distinct count of Operating Airline Iata Code. The view is filtered on Activity Period Year, which ranges from 2010 to 2020.

Interpretation: Operating Airlines are subset of Published airlines (DataSF, 2020). The number of operating/published airlines have been decreased for domestic flights while it is increased for international flights. In other words, International airlines have evolved in last 10 years to facilitate the international passengers. We explore the passengers for domestic and international flights for further insight.
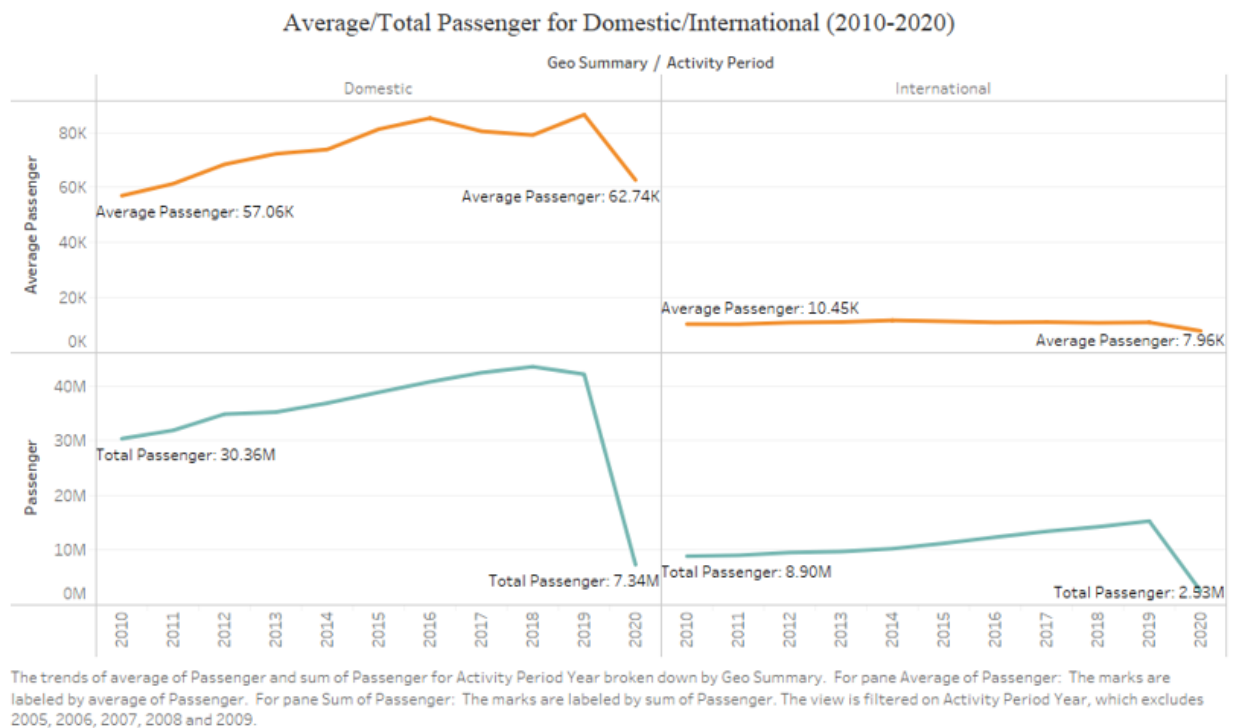
The query for average/total number of passengers for domestic and international flights for past 10 years:

```
SELECT SUBSTR(activity_period,1, 4) AS Year, geo_summary,
ROUND((AVG(CAST(REPLACE(passenger, ',', '') AS INT))/1000),2)
AS Average_Passengers_in_Thousands,
ROUND((SUM(CAST(REPLACE(passenger, ',', '') AS INT))/1000000),2)
AS Total_Passengers_in_Millions
FROM sfo_monthly_air_traffic_passengers_data
WHERE CAST(SUBSTR(activity_period,1, 4) AS INT) >=2010
GROUP BY SUBSTR(activity_period,1, 4), geo_summary;
```

Below are the few lines of the result:

	year	geo_summary	average_passengers_in_thousands	total_passengers_in_millions
1	2010	Domestic	57.06	30.36
2	2010	International	10.45	8.9
3	2011	Domestic	61.41	31.87
4	2011	International	10.4	9.06

Average/Total Passenger for Domestic/International (2010-2020):

The trends of average of Passenger and sum of Passenger for Activity Period Year broken down by Geo Summary. For pane Average of Passenger: The marks are labeled by average of Passenger. For pane Sum of Passenger: The marks are labeled by sum of Passenger. The view is filtered on Activity Period Year, which excludes 2005, 2006, 2007, 2008 and 2009.
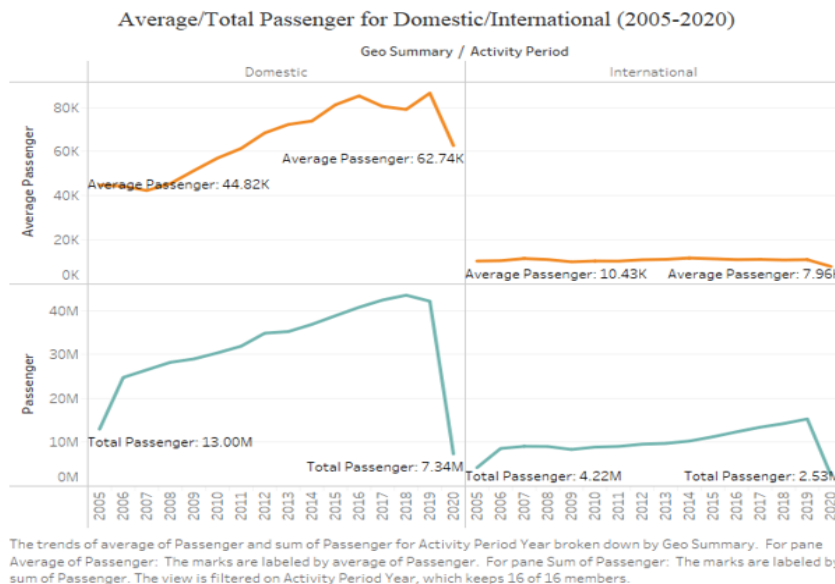
Interpretation: Though Operating flights are decreased for domestic flights, the average/total number of passengers is increased until 2019. On the other hand, Operating flights are increased for international flights, the average number of passengers is not increased. It gives insight about the busy domestic flights. As there is a smaller number of airlines for domestic, the

number of passengers has shown increment which shows the domestic flights have not been

evolved much compared to international flights. Since we have not data for 4 decades to

investigate about the domestic/international flight, we further explore on 15 years of the SFO

data.

Published and Operating Flight Insight for 2005-2020:

The trends of distinct count of Published Airline Iata Code and distinct count of Operating Airline Iata Code for Activity Period Year broken down by Geo Summary. For pane Distinct count of Published Airline Iata Code: The marks are labeled by distinct count of Published Airline Iata Code. For pane Distinct count of Operating Airline Iata Code: The marks are labeled by distinct count of Operating Airline Iata Code. The view is filtered on Activity Period Year, which ranges from 2005 to 2020.

Average/Total Passenger for Domestic/International (2005-2020):

The trends of average of Passenger and sum of Passenger for Activity Period Year broken down by Geo Summary. For pane Average of Passenger: The marks are labeled by average of Passenger. For pane Sum of Passenger: The marks are labeled by sum of Passenger. The view is filtered on Activity Period Year, which keeps 16 of 16 members.

Interpretation: Similar trend has been observed for Domestic and International for the period of 2005-2020. There is a higher number of airlines for international, but the number of passengers has shown little increment. It shows that international flights have evolved much compared to domestic flights. Since, the domestic airlines are busy with higher number of passengers, this would be an area of improvement that the SFO airport division can investigate it.

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

We analyze the busy terminal based on the average number of passengers traveled for each type of the passenger activity in relation to SFO arrived from or departed to a location within the United States/ "Domestic", or outside the United States / "International". Below is the query and result:

```
SELECT * FROM
(SELECT geo_summary, terminal, COUNT(terminal),
ROUND((AVG(CAST(REPLACE(passenger, ',', '') AS INT))/1000),2)
AS Average_Passengers_in_Thousands,
ROUND((SUM(CAST(REPLACE(passenger, ',', '') AS INT))/1000000),2)
AS Total_Passengers_in_Millions,
ROW_NUMBER() OVER
(PARTITION BY geo_summary
ORDER BY AVG(CAST(REPLACE(passenger, ',', '') as INT)) DESC)
AS Rank_Of_Busy_Terminal
FROM sfo_monthly_air_traffic_passengers_data
GROUP BY geo_summary, terminal)
terminal WHERE Rank_Of_Busy_Terminal<=3;
```
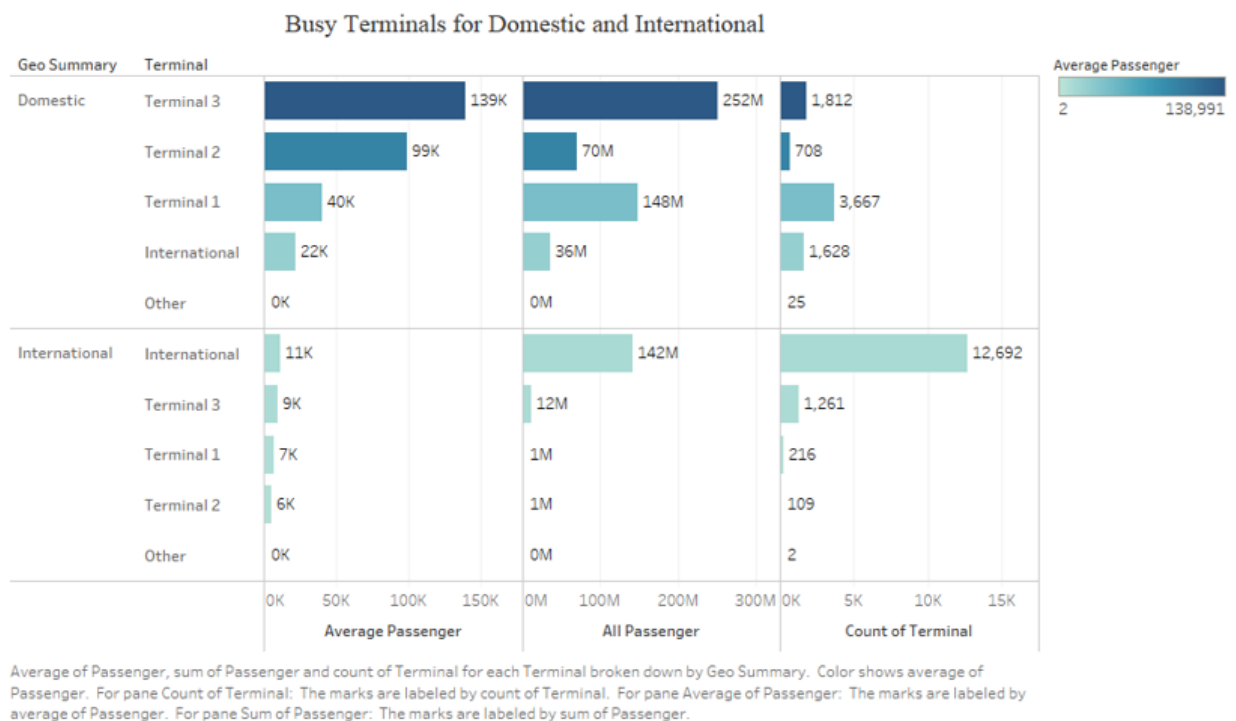
	terminal.geo_summary	terminal.terminal	terminal._c2	terminal.average_passengers_in_thousands
1	Domestic	Terminal 3	1812	138.99
2	Domestic	Terminal 2	708	98.61
3	Domestic	Terminal 1	3667	40.29
4	International	International	12692	11.18
5	International	Terminal 3	1261	9.48
6	International	Terminal 1	216	6.73

terminal.total_passengers_in_millions	terminal.rank_of_busy_terminal
251.85	1
69.81	2
147.76	3
141.92	1
11.95	2
1.45	3

Interpretation: Terminal 3 is the busiest terminal with the average/total number of passengers is highest for domestic while International is the busiest terminal for International arrival/departure.

Busy Terminals for Domestic and International:

In below graph, all terminals are shown for both type of the passenger activity in relation to SFO arrived from or departed to a location within the United States/ "Domestic", or outside the United States / "International".

Average of Passenger, sum of Passenger and count of Terminal for each Terminal broken down by Geo Summary. Color shows average of Passenger. For pane Count of Terminal: The marks are labeled by count of Terminal. For pane Average of Passenger: The marks are labeled by average of Passenger. For pane Sum of Passenger: The marks are labeled by sum of Passenger.

Interpretation:

Domestic: Though count of terminals is highest for Terminal 1, Terminal 3 is the busiest terminal among all terminal that are assigned for domestic arrival/departure with the average passenger as 139K and the total number of passengers is 252 Million.

International: International terminal is the busiest terminal among all terminal that are assigned for International arrival/departure with the average passenger as 11k and the total number of passengers is 142 Million. The other two terminals which are busy are Terminal 3 and Terminal 1. Let's check why Terminal 3 and International terminal are the busiest terminal for Domestic and International arrival/departure respectively.

We explore the busy airlines for domestic as well as international. Below is the query and result of the busy airlines based on average number of passengers.

```sql
SELECT * FROM
(SELECT geo_summary, operating_airline_iata_code,
ROUND((AVG(CAST(REPLACE(passenger, ',', '') AS INT))/1000),2)
AS Average_Passengers_in_Thousands,
ROUND((SUM(CAST(REPLACE(passenger, ',', '') AS INT))/1000000),2)
AS Total_Passengers_in_Millions,
COUNT(operating_airline_iata_code) AS Count_of_Operating_Airline,
ROW_NUMBER() OVER (PARTITION BY geo_summary
ORDER BY AVG(CAST(REPLACE(passenger, ',', '') AS INT)) DESC) AS Rank_Of_Busy_Airline
FROM sfo_monthly_air_traffic_passengers_data
GROUP BY operating_airline_iata_code, geo_summary)
Airline WHERE Rank_Of_Busy_Airline<=3;
```

	airline.geo_summary	airline.operating_airline_iata_code	airline.average_passengers_in_thousands	airline.total_passengers_in_millions
1	Domestic	VX	141.73	36.99
2	Domestic	UA	126.63	205.64
3	Domestic	AA	108.32	50.26
4	International	AC	22.49	10.39
5	International	LH	19.63	6.95
6	International	UA	18.25	50.47

airline.count_of_operating_airline	airline.rank_of_busy_airline
261	1
1624	2
464	3
462	1
354	2
2765	3

The average/total passenger of the busy airlines in terminal 3:

```
SELECT geo_summary, operating_airline_iata_code,
ROUND((AVG(CAST(REPLACE(passenger, ',', '') AS INT))/1000),2)
AS Average_Passengers_in_Thousands,
ROUND((SUM(CAST(REPLACE(passenger, ',', '') AS INT))/1000000),2)
AS Total_Passengers_in_Millions
FROM sfo_monthly_air_traffic_passengers_data
WHERE geo_summary='Domestic' and terminal='Terminal 3'
GROUP BY geo_summary, operating_airline_iata_code
ORDER BY Average_Passengers_in_Thousands DESC LIMIT 3;
```

	geo_summary	operating_airline_iata_code	average_passengers_in_thousands	total_passengers_in_millions
1	Domestic	UA	170.49	191.63
2	Domestic	AA	136.17	19.06
3	Domestic	OO	112.18	40.38

Interpretation: The top 3 busiest airlines for domestic are Virgin America, United

Airlines, and American Airlines. If we check the busy airlines in terminal 3, we observe that the

two busiest airlines United Airlines and American Airlines use terminal 3. Also, for United

Airlines, most of the passenger's activity (enplaned/deplaned) is in Terminal 3. This is the

reasons why terminal 3 is busy. It increases the waiting time of the passengers at terminal 3.

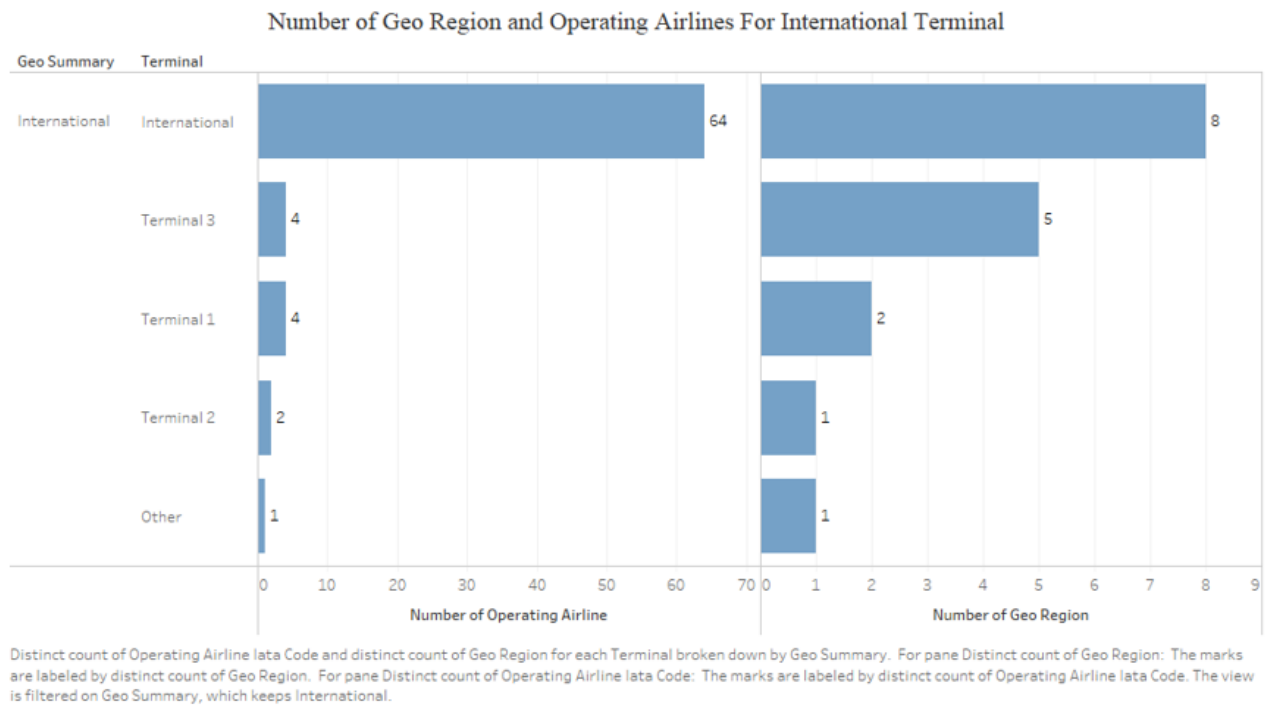The average/total passenger of the busy airlines in International terminal:

```
SELECT geo_summary, operating_airline_iata_code,
ROUND((AVG(CAST(REPLACE(passenger, ',', '') AS INT))/1000),2)
AS Average_Passengers_in_Thousands,
ROUND((SUM(CAST(REPLACE(passenger, ',', '') AS INT))/1000000),2)
AS Total_Passengers_in_Millions
FROM sfo_monthly_air_traffic_passengers_data
WHERE geo_summary='International' and terminal='International'
GROUP BY geo_summary, operating_airline_iata_code
ORDER BY Average_Passengers_in_Thousands DESC LIMIT 3;
```

	geo_summary	operating_airline_iata_code	average_passengers_in_thousands	total_passengers_in_millions
1	International	UA	22.24	44.09
2	International	AC	21.84	7.03
3	International	LH	19.63	6.95

Interpretation: The top 3 busiest airlines for International are Air Canada, Lufthansa

German Airlines, and United Airlines. If we check the busy airlines in International terminal, we

observe that the two busiest airlines United Airlines and Air Canada use International terminal. Also, for United Airlines, most of the passenger's activity (enplaned/deplaned) is in International terminal. This is the reasons why International terminal is busy. Also, it increases the waiting time of the passengers at International terminal. We further explore on geo regions and operating airlines:

Number of Geo Regions and Operating Airlines for International Terminal:

Distinct count of Operating Airline Iata Code and distinct count of Geo Region for each Terminal broken down by Geo Summary. For pane Distinct count of Geo Region: The marks are labeled by distinct count of Geo Region. For pane Distinct count of Operating Airline Iata Code: The marks are labeled by distinct count of Operating Airline Iata Code. The view is filtered on Geo Summary, which keeps International.

Interpretation: In International terminal, there are large number of regions as well as operating airlines that have passenger activity i.e flights enplaned/deplaned. Other terminal does have smaller number of geo regions and operating airlines for International departure/arrival.

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

We analyze the busy airline based on the average number of passengers traveled for each type of the passenger activity in relation to SFO arrived from or departed to a location within the United States/ "Domestic", or outside the United States / "International".

Below is the query and result for the busiest airlines for Domestic and International:

```
SELECT * FROM
(SELECT geo_summary, operating_airline_iata_code,
ROUND((AVG(CAST(REPLACE(passenger, ',', '') AS INT))/1000),2)
AS Average_Passengers_in_Thousands,
ROUND((SUM(CAST(REPLACE(passenger, ',', '') AS INT))/1000000),2)
AS Total_Passengers_in_Millions,
COUNT(operating_airline_iata_code) AS Count_of_Operating_Airline,
ROW_NUMBER() OVER (PARTITION BY geo_summary
ORDER BY AVG(CAST(REPLACE(passenger, ',', '') AS INT)) DESC) AS Rank_Of_Busy_Airline
FROM sfo_monthly_air_traffic_passengers_data
GROUP BY operating_airline_iata_code, geo_summary)
Airline WHERE Rank_Of_Busy_Airline<=3;
```

	airline.geo_summary	airline.operating_airline_iata_code	airline.average_passengers_in_thousands	airline.total_passengers_in_millions
1	Domestic	VX	141.73	36.99
2	Domestic	UA	126.63	205.64
3	Domestic	AA	108.32	50.26
4	International	AC	22.49	10.39
5	International	LH	19.63	6.95
6	International	UA	18.25	50.47

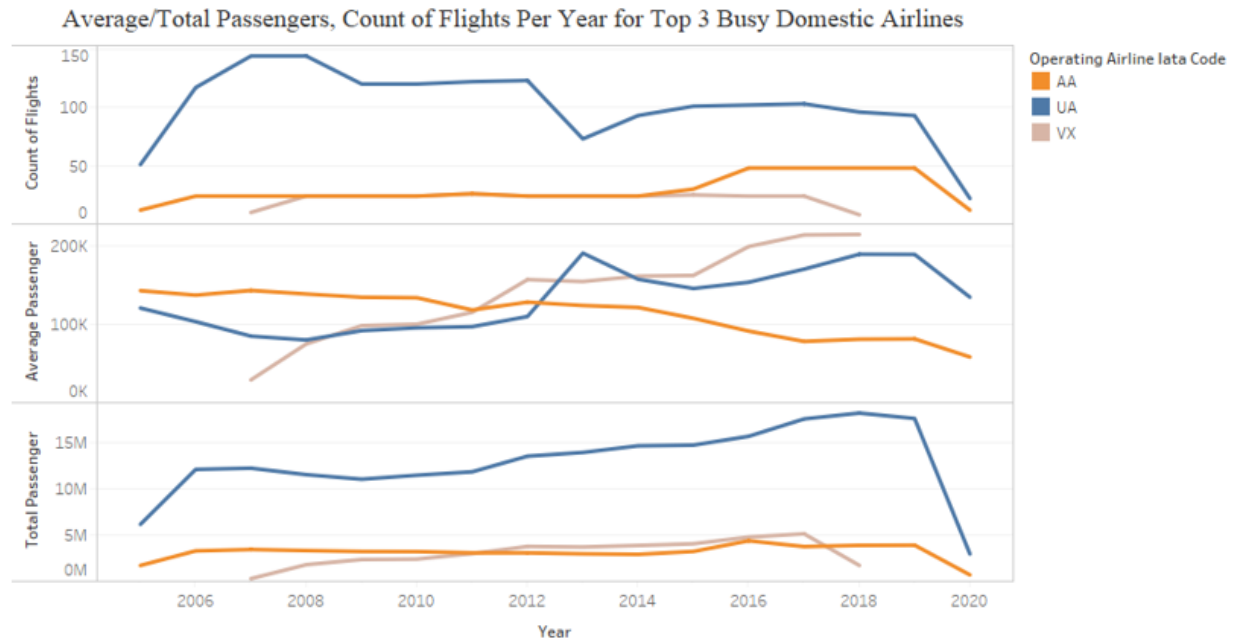| airline.count_of_operating_airline	airline.rank_of_busy_airline
261	1
1624	2
464	3
462	1
354	2
2765	3

Interpretation: Virgin America is the busiest airline for domestic while Air Canada is the busiest airline for international. It is due to the count of flights are less, but the average number of passengers is high. For instance: United Airlines has the highest number of flights with the highest number of passengers traveled. But United Airlines is not the busiest airline as the average number of passengers is less. Below the graph shows the same interpretation for top 3 busy airlines domestic as well as international:

Average/Total Passengers, Count of Flights Per Year for Top 3 Busy International Airlines:
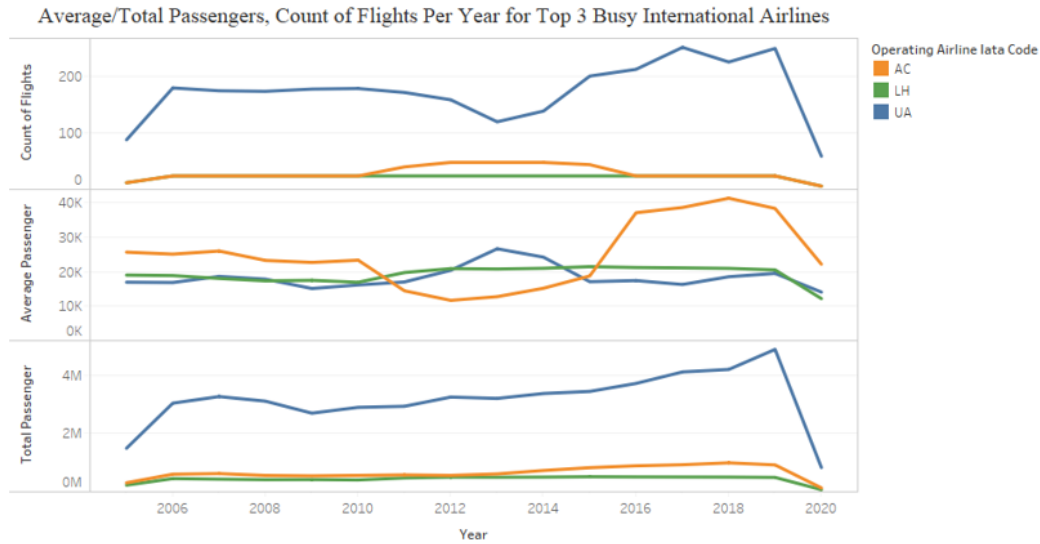
Average/Total Passengers, Count of Flights Per Year for Top 3 Busy International Airlines

The trends of count of Operating Airline Iata Code, average of passenger and sum of passenger for Activity Period Year. Color shows details about Operating Airline Iata Code. The data is filtered on Geo Summary, which keeps International. The view is filtered on Operating Airline Iata Code and count of Operating Airline Iata Code. The Operating Airline Iata Code filter keeps AC, LH and UA. The count of Operating Airline Iata Code filter ranges from 6 to 252.

Average/Total Passengers, Count of Flights Per Year for Top 3 Busy Domestic Airlines:

Average/Total Passengers, Count of Flights Per Year for Top 3 Busy Domestic Airlines

The trends of count of Operating Airline Iata Code, average of passenger and sum of passenger for Activity Period Year. Color shows details about Operating Airline Iata Code. The data is filtered on Geo Summary, which keeps Domestic. The view is filtered on Operating Airline Iata Code and count of Operating Airline Iata Code. The Operating Airline Iata Code filter keeps AA, UA and VX. The count of Operating Airline Iata Code filter ranges from 6 to 252.

5. What is the average passenger count per airline? Who has the max number of passengers

and why do you think that is?

The query for the average passenger and the total number of passengers:

```
SELECT operating_airline_iata_code,
ROUND((SUM(CAST(REPLACE(passenger, ',', '') AS INT))/1000000),2)
AS Total_Passengers_in_Millions,
COUNT(operating_airline_iata_code) AS Count_of_Flights,
ROUND((AVG(CAST(REPLACE(passenger, ',', '') AS INT))/1000),2) AS Average_Passenger_in_Thousands
FROM sfo_monthly_air_traffic_passengers_data
GROUP BY operating_airline_iata_code ORDER BY Average_Passenger_in_Thousands desc;
```
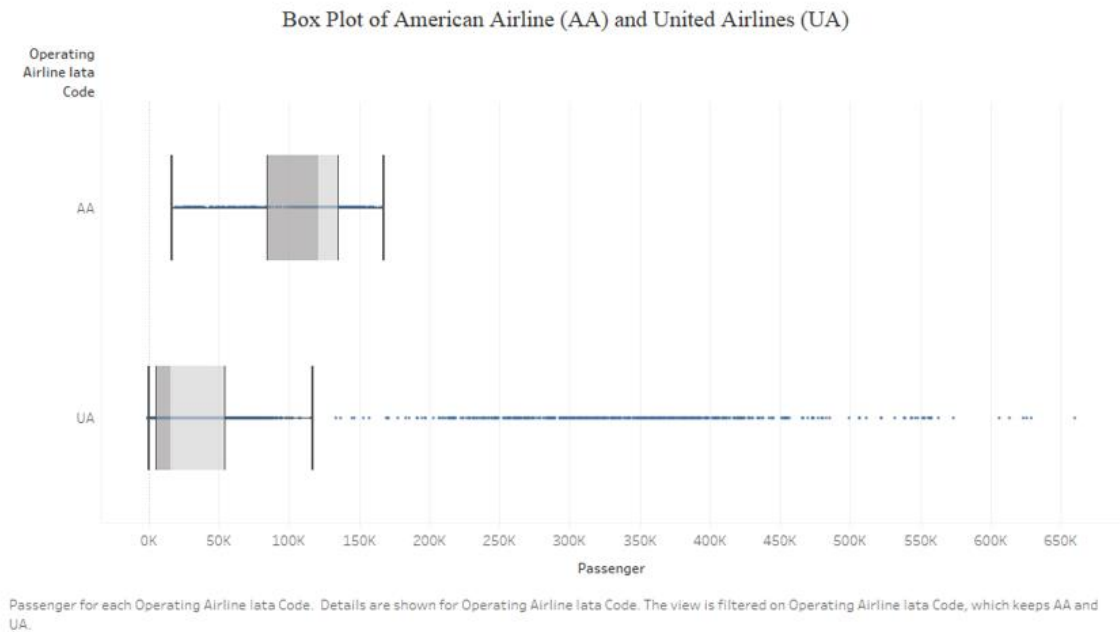
Few lines of the query results are:

	operating_airline_iata_code	total_passengers_in_millions	count_of_flights	average_passenger_in_thousands
1	AA	50.26	464	108.32
2	WN	39.1	446	87.68
3	DL	42.29	493	85.78
4	VX	37.84	462	81.91
5	UA	256.12	4389	58.35
6	US	16.82	304	55.32
7	OO	51.64	1386	37.26
8	B6	13.89	380	36.54
9	AS	27.65	977	28.3
10	NW	6.27	240	26.11

Interpretation: The average number of passengers is highest for American Airlines while

the total number of passengers is highest for United Airlines. The number of flights mentioned

monthly in the dataset is much lesser for American Airlines than United Airlines. We can

explore if there is an outlier for United Airlines.

Box Plot of American Airline (AA) and United Airlines (UA):

Box Plot of American Airline (AA) and United Airlines (UA)

Passenger for each Operating Airline Iata Code. Details are shown for Operating Airline Iata Code. The view is filtered on Operating Airline Iata Code, which keeps AA and UA.

Interpretation: The box plot of United Airlines is highly right-skewed where the mean is greater than the median. It has a longer tail on the right side. The median is close to the first quartile. Most of the data is distributed on the left side. On the other hand, the box plot of American Airlines is left-skewed where the median is greater than mean. It has a longer tail on the left side. The Median is close to the third quartile. Most of the data is distributed on the right side. It explains that though the number of flights (mentioned monthly) and the total number of passengers are maximum for United Airlines, the number of passengers is less for most of the flights of United Airlines. Hence, the average passenger is less, but the total number of passengers is maximum due to the highest number of flights for United Airlines.

6. 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?

    The query and result for Top 3 years for maximum number of passengers:

```
SELECT SUBSTR(activity_period,1, 4) AS Top_3_Year,
ROUND((SUM(CAST(REPLACE(passenger, ',', '') AS INT))/1000000),2)
AS Total_Passengers_in_Millions
FROM sfo_monthly_air_traffic_passengers_data
GROUP BY SUBSTR(activity_period,1, 4)
ORDER BY Total_Passengers_in_Millions DESC LIMIT 3;
```

	top_3_year	total_passengers_in_millions
1	2018	57.75
2	2019	57.42
3	2017	55.82

The query and result for Bottom 4 years for maximum number of passengers:

```
SELECT SUBSTR(activity_period,1, 4) AS Bottom_4_Year,
ROUND((SUM(CAST(REPLACE(passenger, ',', '') AS INT))/1000000),2)
AS Total_Passengers_in_Millions
FROM sfo_monthly_air_traffic_passengers_data
GROUP BY SUBSTR(activity_period,1, 4)
ORDER BY Total_Passengers_in_Millions LIMIT 4;
```
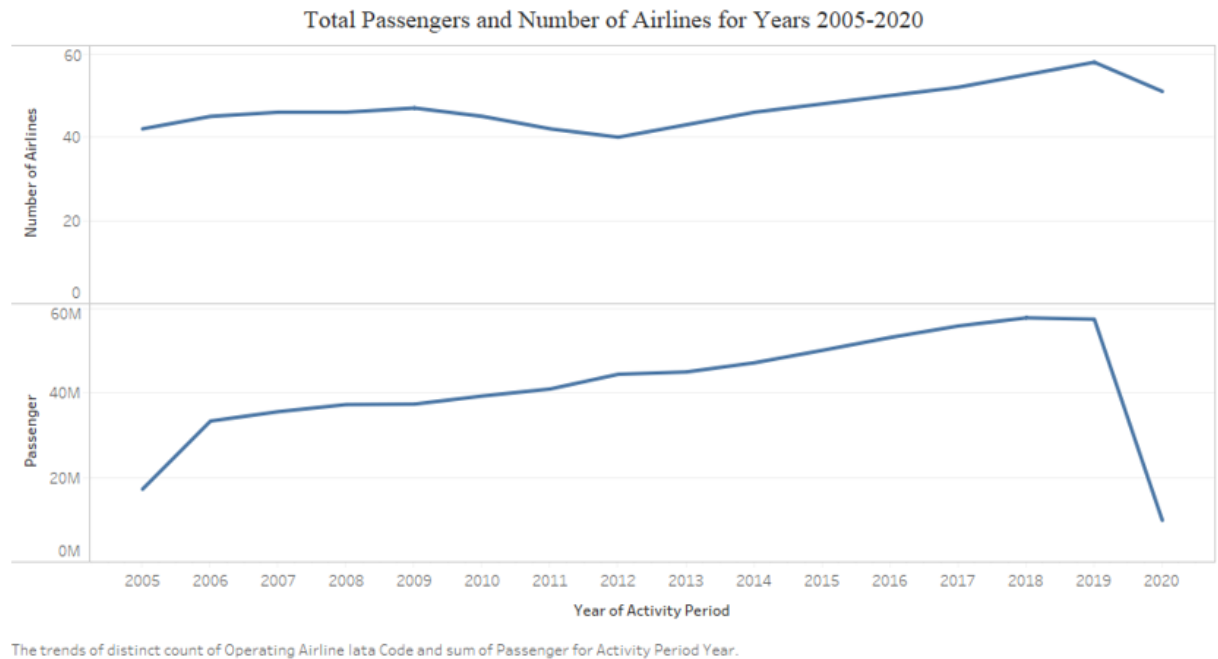
	bottom_4_year	total_passengers_in_millions
1	2020	9.87
2	2005	17.22
3	2006	33.33
4	2007	35.55

Since we have data until the month of march i.e first quarter for 2020, I chose to analyze the bottom 4 years for the number of passengers in SFO.

Interpretation: 2018, 2019, and 2017 have the largest number of passengers while 2005, 2006, and 2007 have the lowest number of passengers. Besides, 2020 is an exceptional year due to the COVID-19 impact. Let's check how the number of airlines and total passengers varied from 2005 to 2020.

Total Passengers and Number of Airlines for Years 2005-2020:

The below graph shows the total number airlines and total passengers for 15 years.

Total Passengers and Number of Airlines for Years 2005-2020

The trends of distinct count of Operating Airline Iata Code and sum of Passenger for Activity Period Year.

Interpretation: The number of airlines is increased from 2005 to 2020. In other words, many airlines have been evolved during these periods. This is one of the reasons why the number of passengers was maximum for the year 2017-2019.

In 2005-2007, the US faced the recession and crisis that slowed down the US economy sharply at the end of 2005 (The New York Times, 2006). Consumer spending slowed abruptly at the end of 2005 and the intensity of the economic slowdown reduced yearly growth to 3.5 percent compared to 2004. This economic slowdown lasted till 2009. The airline industry has also been impacted by the economic recession, which has reduced the public's ability to purchase air travel (U.S. Department of Transportation, 2012).

For 2017-2019, the US economy picked up the growth in 2017 and it remained strong till 2019 before COVID-19 pandemic. This economic growth has affected the airline industry in positive ways and increased the purchasing power of consumers that results in a higher number of passengers.

7. 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

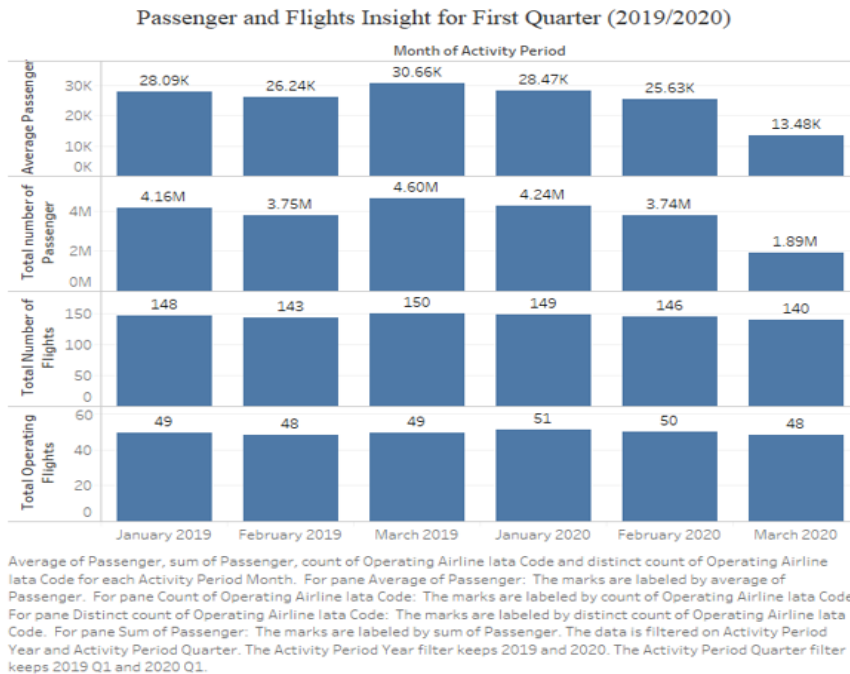   say about the impact of COVID19 on the travel industry?

   The query and result for average/total passengers, total number of flights and number of

   operating flights in first quarter of 2019 and 2020

```
SELECT SUBSTR(activity_period,5, 2) AS Month, SUBSTR(activity_period,1, 4) AS Year,
ROUND((AVG(CAST(REPLACE(passenger, ',', '') AS INT))/1000),2) AS Average_Passengers_in_Thousands,
ROUND((SUM(CAST(REPLACE(passenger, ',', '') AS INT))/1000000),2) AS Total_Passengers_in_Millions,
COUNT(operating_airline_iata_code) AS Total_number_of_flight,
COUNT(DISTINCT(operating_airline)) AS Operating_Flight_Count
FROM sfo_monthly_air_traffic_passengers_data
WHERE activity_period in ('202001','202002','202003','201901','201902','201903')
GROUP BY activity_period
ORDER BY Month;
```

	month	year	average_passengers_in_thousands	total_passengers_in_millions	total_number_of_flight	operating_flight_count
1	01	2019	28.09	4.16	148	49
2	01	2020	28.47	4.24	149	51
3	02	2019	26.24	3.75	143	48
4	02	2020	25.63	3.74	146	50
5	03	2019	30.66	4.6	150	49
6	03	2020	13.48	1.89	140	48

Passenger and Flights Insight for First Quarter (2019/2020):

The below graph shows the total/average number of passengers, the total number of

operating flights with the distinct count of operating flights for January, February, March for
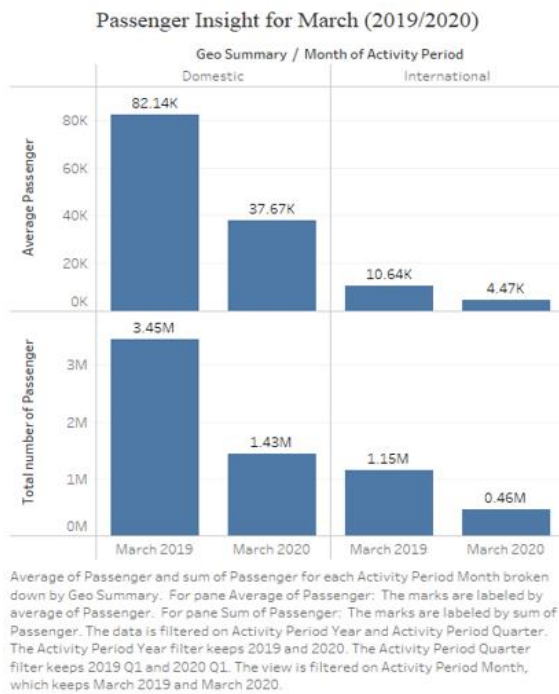
year 2019 and 2020.

Passenger and Flights Insight for First Quarter (2019/2020)

Average of Passenger, sum of Passenger, count of Operating Airline Iata Code and distinct count of Operating Airline Iata Code for each Activity Period Month. For pane Average of Passenger: The marks are labeled by average of Passenger. For pane Count of Operating Airline Iata Code: The marks are labeled by count of Operating Airline Iata Code. For pane Distinct count of Operating Airline Iata Code: The marks are labeled by distinct count of Operating Airline Iata Code. For pane Sum of Passenger: The marks are labeled by sum of Passenger. The data is filtered on Activity Period Year and Activity Period Quarter. The Activity Period Year filter keeps 2019 and 2020. The Activity Period Quarter filter keeps 2019 Q1 and 2020 Q1.

Interpretation: We have not observed a drastic decrease in the average and the total number of passengers along with operating flights for January and February of the first quarter of 2019 and 2020. When President Donald Trump has activated a national emergency for the COVID-19 on 13th March, it badly affected the travel industry (NCSL, 2020). As part of a national emergency, most of the aircraft were banned and stopped to avoid the spread of the coronavirus. Hence, there is a drastic decrease in the average and the total number of passengers in March. We explored further to see the effects on domestic and international travel for March. The query and result for passengers for Domestic and International activity type for March 2019 and March 2020:

```sql
SELECT SUBSTR(activity_period,5, 2) AS Month,
SUBSTR(activity_period,1, 4) AS Year, geo_summary,
ROUND((SUM(CAST(REPLACE(passenger, ',', '') AS INT))/1000000),2)
AS Total_Passengers_in_Millions,
ROUND((AVG(CAST(REPLACE(passenger, ',', '') AS INT))/1000),2)
AS Total_Passengers_in_Thousands
FROM sfo_monthly_air_traffic_passengers_data
WHERE activity_period IN ('202003','201903')
GROUP BY activity_period, geo_summary
ORDER BY geo_summary, Month;
```

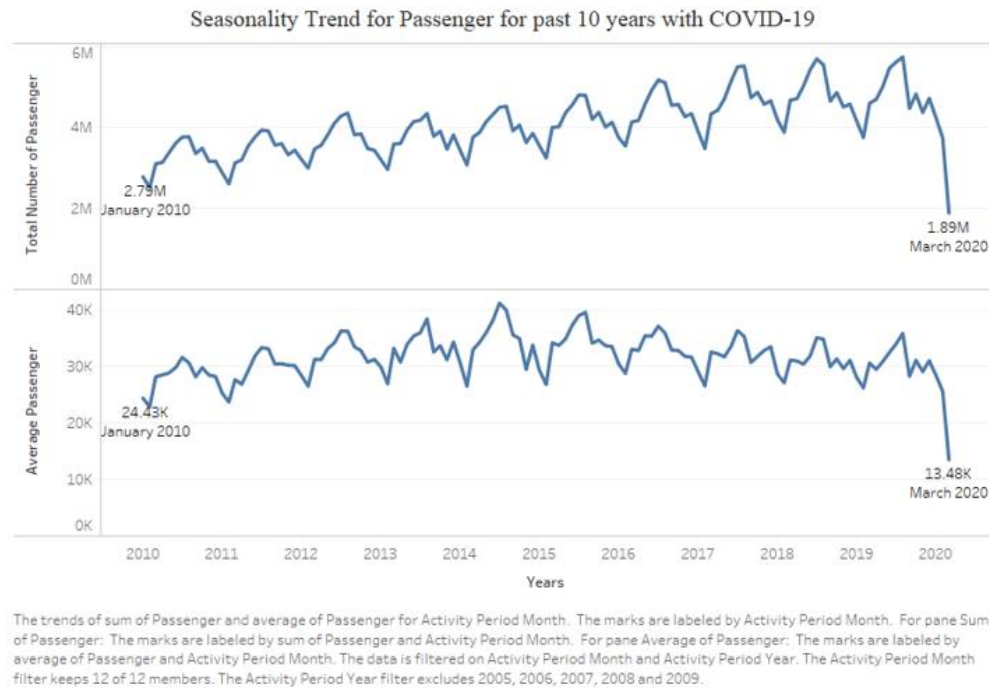|	month	year	geo_summary	total_passengers_in_millions	total_passengers_in_thousands
1	03	2019	Domestic	3.45	82.14
2	03	2020	Domestic	1.43	37.67
3	03	2019	International	1.15	10.64
4	03	2020	International	0.46	4.47

Passenger Insight for March (2019/2020):

The below graph shows the total/average passenger for Domestic/International.

Passenger Insight for March (2019/2020)

Average of Passenger and sum of Passenger for each Activity Period Month broken
down by Geo Summary. For pane Average of Passenger: The marks are labeled by
average of Passenger. For pane Sum of Passenger: The marks are labeled by sum of
Passenger. The data is filtered on Activity Period Year and Activity Period Quarter.
The Activity Period Year filter keeps 2019 and 2020. The Activity Period Quarter
filter keeps 2019 Q1 and 2020 Q1. The view is filtered on Activity Period Month,
which keeps March 2019 and March 2020.

Interpretation: There is a drastic decrease in the average and the total number of

passengers for domestic as well as international travel in March.

Seasonality Trend for Passenger for past 10 years with COVID-19:

The trends of sum of Passenger and average of Passenger for Activity Period Month. The marks are labeled by Activity Period Month. For pane Sum of Passenger: The marks are labeled by sum of Passenger and Activity Period Month. For pane Average of Passenger: The marks are labeled by average of Passenger and Activity Period Month. The data is filtered on Activity Period Month and Activity Period Year. The Activity Period Month filter keeps 12 of 12 members. The Activity Period Year filter excludes 2005, 2006, 2007, 2008 and 2009.

Interpretation: 2020 has been the worst year for the air travel industry as the average / total number of passengers is the lowest in the history of 10 years for SFO airport. Seasonality trend shows the data for past 10 years that helps to visualize how suddenly total/average number of passengers was dropped in March 2020.

Conclusion

Statistical analysis on the SFO dataset has given various insights about the airport. US, Asia, and Europe regions are the busiest region in SFO. We can suggest the Airport divisions and the county of SFO to look at the other regions and find the opportunities for the connectivity of the flights to different regions with good frequency. This can help travelers/passengers to plan the SFO visit more easily and comfortably. In addition, the terminal 3 and international terminal are the busiest terminal for domestic and international arrival/departure respectively. This is

leading to an increase in the waiting time of the passengers in the terminal. The other terminal

can have connectivity to the busiest flights so that the waiting time can be reduced and balanced

across all terminals. In addition, Virgin America is the busiest airline for domestic while Air

Canada is the busiest airline for international. Airport division can look for increasing the

frequency of the flights of these airlines to improve the passenger experience. We also observed

that the recession as well as good economic conditions deeply impacted the air travel industry.

Furthermore, the COVID-19 has worsened the air travel industry performance and reached the

lowest number of passengers in March 2020 in a history of 10 years. SFO airport needs an

improvement plan during COVID-19 that helps to improve the passenger's safety and security.

This analysis can be shared with the people who look for insight for the SFO airport

improvement.

Reference

ClaimCompass. (2020). *The Biggest and Busiest Airports in the US in 2020.* Retrieved from

https://www.claimcompass.eu/blog/biggest-busiest-us-airports/

DataSF. (2020). *Air Traffic Passenger Statistics.* Retrieved from https://data.sfgov.org/

Transportation/Air-Traffic-Passenger-Statistics/rkru-6vcg

Flysfo. (n.d.) *Strategic 5-Year Plan.* Retrieved from https://www.flysfo.com/about-sfo/the-

organization/strategic-5-year-plan

NCSL. (2020). *President Trump Declares State of Emergency for COVID-19.* Retrieved from

https://www.ncsl.org/ncsl-in-dc/publications-and-resources/president-trump-declares-

state-of-emergency-for-covid-19.aspx

The New York Times. (2006). *U.S. Economy Slowed Sharply at End of 2005.* Retrieved from

https://www.nytimes.com/2006/01/28/business/us-economy-slowed-sharply-at-end-of-

2005.html

U.S. Department of Transportation. (2012). *Office of Inspector General.* Retrieved from

https://www.oig.dot.gov/sites/default/files/Aviation%20Industry%20Performance%5E9-

24-12.pdf