

A COMPREHENSIVE STUDY OF SAN FRANCISCO AIRPORT'S PASSENGER FLOW

Executive Summary

This project examines passenger traffic at San Francisco International Airport (SFO) during three critical periods: pre-COVID (January 2018-December 2019), pandemic (January 2020-December 2022), and post-COVID (January 2022-December 2022). The study used Tableau's visualization tool to understand the pandemic's impact on passenger flow, assess airline resilience, and track the airport's recovery.

The project uses two datasets: the San Francisco International Airport dataset, which provides monthly passenger traffic statistics, and the Los Angeles International Airport dataset, which contains vital attributes essential for the analysis.

A meticulous data cleaning process was carried out, including removing NA records and filtering data for specified periods. The SFO dataset has 7735 records after cleaning, while the Los Angeles dataset has 2172 records.

The project investigates five hypotheses, yielding significant results. Notably, the pandemic reduced SFO passenger counts by 50%, affecting airlines and terminal utilization. Despite top airline rankings remaining stable, travel preferences shifted, as evidenced by varying destination popularity. Surprisingly, a comparison with Los Angeles International Airport revealed parallel trends.

Finally, this thorough examination sheds light on the intricate dynamics of passenger travel at SFO, providing valuable insights into the impact of COVID-19 on airlines and the travel industry.

Index

<i>Executive Summary.....</i>	<i>2</i>
<i>1. Data Description</i>	<i>4</i>
<i>2. Data Cleaning</i>	<i>5</i>
<i>3. General Introduction.....</i>	<i>7</i>
<i>4. Insights and Findings.....</i>	<i>8</i>
<i>4.1 Hypothesis 1</i>	<i>8</i>
<i>4.2 Hypothesis 2</i>	<i>9</i>
<i>4.3 Hypothesis 3</i>	<i>10</i>
<i>4.4 Hypothesis 4</i>	<i>11</i>
<i>4.5 Hypothesis 5</i>	<i>12</i>
<i>5. Conclusion.....</i>	<i>13</i>

1. Data Description

San Francisco International Airport Dataset, the primary dataset, gives monthly passenger traffic statistics from San Francisco, and the airline reports the data.

The key attributes in the dataset are:

Activity Period	Time Period- Month and Year
Operating Airline	Name of the operating Airline
GEO Summary	Gives if the Flight is Domestic or International
GEO Region	Gives the origin of the flight- Continent
Activity Type Code	Gives the action performed by the passenger- Enplaned, Deplaned or Transit
Price Category Code	Type of the fare for that flight- Low or Normal Fare
Terminal	Terminal number of the flight in the Airport
Boarding Area	Boarding Area in the Airport
Passenger Count	Gives the total count of passengers

The Los Angeles International Airport Dataset is our secondary dataset, which gives monthly passenger traffic statistics from Los Angeles.

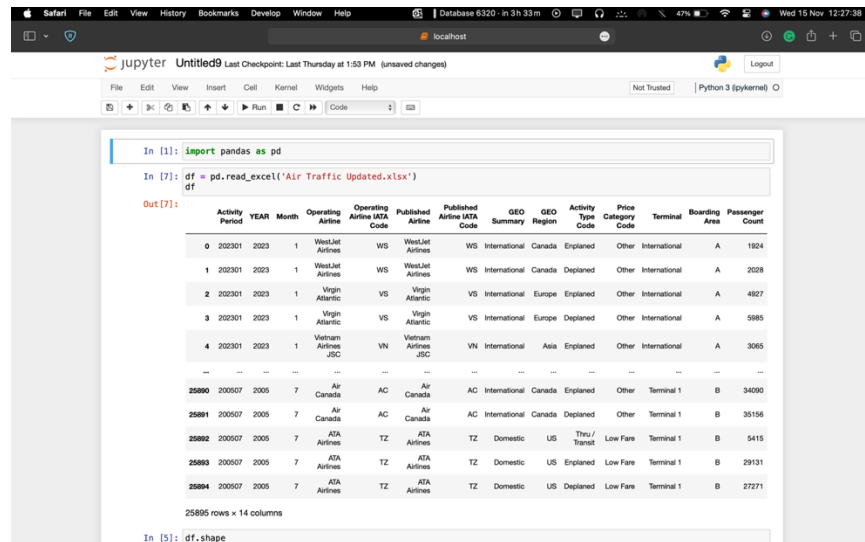
The key attributes in the dataset are:

ReportPeriod	Time Period- Month and Year
Terminal	Terminal number of the flight in the Airport
Arrival_Departure	Gives the action performed by the passenger- Arrival or Departure
Domestic International	Gives if the Flight is Domestic or International
Passenger_Count	Gives the total count of passengers

There is an additional data set that gives monthly data on positive cases of COVID-19 in SFO.

2. Data Cleaning

We imported the SFO dataset into Jupyter Notebooks and checked the number of records using Python. The total records are found to be around **25895** records.



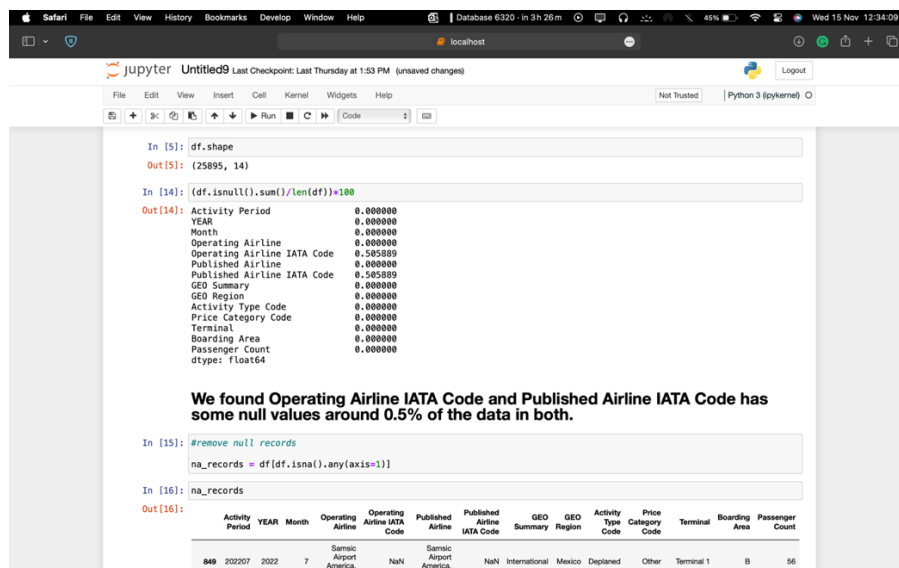
```
In [1]: import pandas as pd
In [7]: df = pd.read_excel('Air Traffic Updated.xlsx')
df
Out[7]:
```

	Activity Period	YEAR	Month	Operating Airline	Operating Airline IATA Code	Published Airline	Published Airline IATA Code	GEO Summary	GEO Region	Activity Type Code	Price Category Code	Terminal	Boarding Area	Passenger Count
0	202301	2023	1	WestJet Airlines	WS	WestJet Airlines	WS	International	Canada	Enplaned	Other	International	A	1824
1	202301	2023	1	WestJet Airlines	WS	WestJet Airlines	WS	International	Canada	Deplaned	Other	International	A	2028
2	202301	2023	1	Virgin Atlantic	VS	Virgin Atlantic	VS	International	Europe	Enplaned	Other	International	A	4827
3	202301	2023	1	Virgin Atlantic	VS	Virgin Atlantic	VS	International	Europe	Deplaned	Other	International	A	5885
4	202301	2023	1	Vietnam Airlines JSC	VN	Vietnam Airlines JSC	VN	International	Asia	Enplaned	Other	International	A	3065
...
25890	200907	2005	7	Air Canada	AC	Air Canada	AC	International	Canada	Enplaned	Other	Terminal 1	B	34090
25891	200907	2005	7	Air Canada	AC	Air Canada	AC	International	Canada	Deplaned	Other	Terminal 1	B	35156
25892	200907	2005	7	ATA Airlines	TZ	ATA Airlines	TZ	Domestic	US	Thru/Transit	Low Fare	Terminal 1	B	5415
25893	200907	2005	7	ATA Airlines	TZ	ATA Airlines	TZ	Domestic	US	Enplaned	Low Fare	Terminal 1	B	29131
25894	200907	2005	7	ATA Airlines	TZ	ATA Airlines	TZ	Domestic	US	Deplaned	Low Fare	Terminal 1	B	27271

25895 rows x 14 columns

```
In [5]: df.shape
```

The next step was to check the number of NA records present in the dataset; we found that around 0.5% of the data had NA records.



```
In [5]: df.shape
Out[5]: (25895, 14)
In [14]: (df.isnull().sum()/len(df))*100
Out[14]:
```

Activity Period	YEAR	Month	Operating Airline	Operating Airline IATA Code	Published Airline	Published Airline IATA Code	GEO Summary	GEO Region	Activity Type Code	Price Category Code	Terminal	Boarding Area	Passenger Count
0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.505889	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

We found Operating Airline IATA Code and Published Airline IATA Code has some null values around 0.5% of the data in both.

```
In [15]: #remove null records
na_records = df[df.isna().any(axis=1)]
In [16]: na_records
Out[16]:
```

Activity Period	YEAR	Month	Operating Airline	Operating Airline IATA Code	Published Airline	Published Airline IATA Code	GEO Summary	GEO Region	Activity Type Code	Price Category Code	Terminal	Boarding Area	Passenger Count	
849	202207	2022	7	Samair Airport America, LLC	NaN	Samair Airport America, LLC	NaN	International	Mexico	Deplaned	Other	Terminal 1	B	56

```
131 rows x 14 columns

In [17]: # make a seprate file for null records
na_records.to_csv('na_records.csv', index=False)

In [20]: new_df = df.dropna()

In [22]: # data without null
new_df.shape
Out[22]: (25764, 14)

In [ ]: # clean csv file
new_df.to_csv('new_df.csv', index=False)

In [ ]: LA_df = pd.read_excel('Book1.xlsx')
LA_df

In [ ]: LA_df.columns

In [34]: #drop unnecessary column
LA_df.drop(columns=['Unnamed: 6', 'Unnamed: 7', 'Unnamed: 8', 'Unnamed: 9'], inplace=True)

In [37]: #check any null values
LA_df.isna().sum()
Out[37]: DataExtractDate      0
ReportPeriod      0
Terminal      0
Arrival_Departure      0
Domestic_International      0
Passenger_Count      0
dtype: int64
```

We dropped the NA records from the dataset and exported the clean CSV with **25764** records.

```
In [ ]: LA_df = pd.read_excel('Book1.xlsx')
LA_df

In [ ]: LA_df.columns

In [34]: #drop unnecessary column
LA_df.drop(columns=['Unnamed: 6', 'Unnamed: 7', 'Unnamed: 8', 'Unnamed: 9'], inplace=True)

In [37]: #check any null values
LA_df.isna().sum()
Out[37]: DataExtractDate      0
ReportPeriod      0
Terminal      0
Arrival_Departure      0
Domestic_International      0
Passenger_Count      0
dtype: int64

there are no null values.

In [ ]: #export the clean data
LA_df.to_csv('LA_data.csv', index=False)

In [ ]:
```

The LA Dataset had no NA values.

We filtered data from January 2018 to December 2022 because the analysis will be performed in three time periods: before the pandemic (January 2018-December 2019), during the pandemic (January 2020-December 2021), and after the pandemic (January 2022-December 2022). Thus, the San Francisco dataset has **7735** records, while the Los Angeles dataset has **2172** records.

3. General Introduction

In an era marked by unprecedented global challenges, the dynamics of passenger travel within the aviation industry have undergone profound transformations. Our project, "A Comprehensive Study of San Francisco Airport's Passenger Flow," delves into the intricate tapestry of these changes, focusing on San Francisco International Airport (SFO) during three pivotal periods: pre-COVID, the pandemic, and post-COVID.

As the world grappled with the far-reaching effects of the COVID-19 pandemic, the aviation sector emerged as a focal point of resilience, adaptation, and evolution. SFO, as a central international hub, became a microcosm of the industry's response to unprecedented disruptions. This project endeavors to unravel the multifaceted story embedded in the passenger traffic data, comprehensively examining the airport's journey through these transformative phases.

Our team, comprising diligent members with diverse expertise, employed Tableau's visualization tool to analyze and interpret the data. By leveraging datasets from San Francisco and Los Angeles International Airports, we sought to draw nuanced insights into the impact of the pandemic on passenger counts, terminal utilization patterns, airline preferences, and broader industry dynamics.

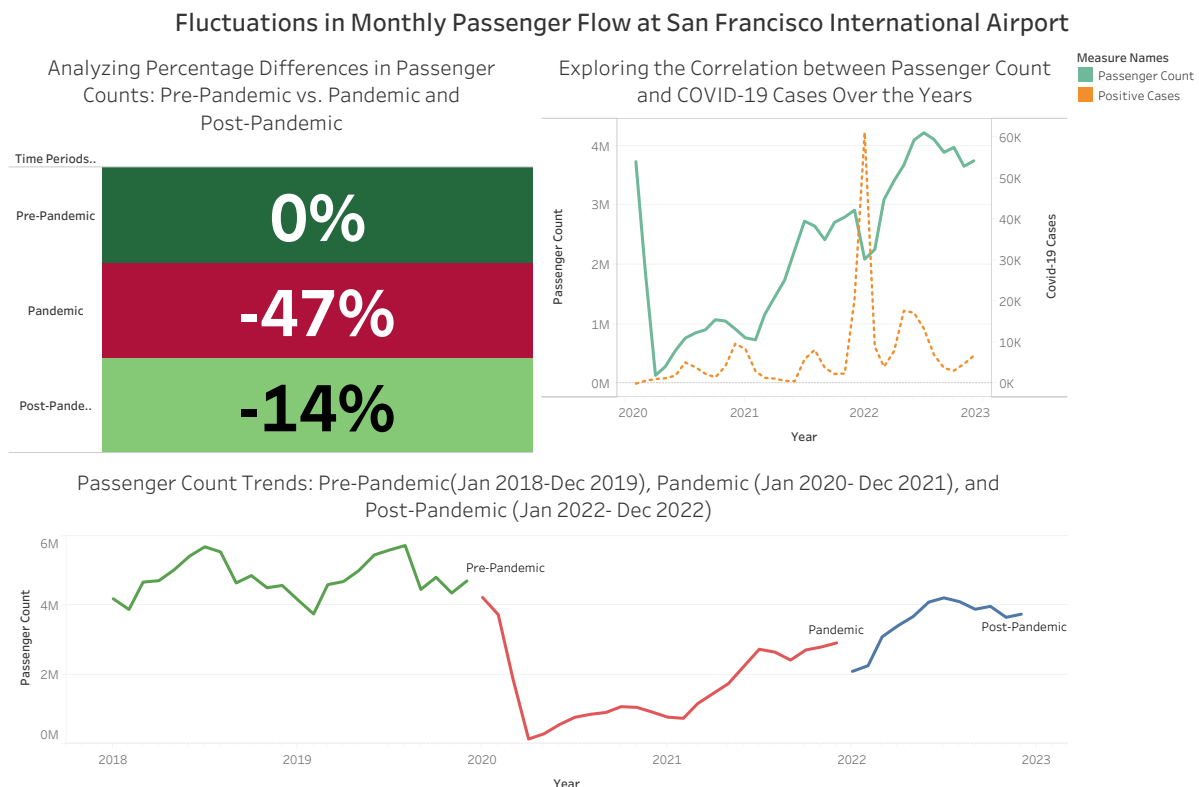
This study aims not only to illuminate the challenges posed by the pandemic but also to underscore the resilience and adaptability of the aviation industry. Our exploration findings will contribute valuable insights into the aviation sector and serve as a foundation for informed decision-making and strategic planning. As we embark on this journey through data-driven analysis, our goal is to offer a comprehensive understanding of SFO's passenger flow dynamics, navigating through the uncertainties of the past and paving the way for a more resilient and adaptive future.

4. Insights and Findings

4.1 Hypothesis 1

We hypothesize that COVID-19 had an impact on monthly passenger flow at San Francisco International Airport, resulting in a 50% drop during the pandemic (Jan 2020-Dec 2021) and a subsequent recovery in terms of numbers post-pandemic (Jan 2022-Dec 2022) as that of Pre-Covid (Jan 2018-Dec 2019)

Visualization:



Findings:

- We begin by comparing the total number of passengers flown from and to SFO over three periods: pre-pandemic, pandemic, and post-pandemic. The visualization above shows that COVID-19 has significantly impacted the passenger count.
- We can see a steep drop in the count in April 2020, which continues until December 2021, with an improvement that continues as we head into the post-pandemic phase. Compared to pre-pandemic levels, there is a 50% decrease in passenger count during and a 14% decrease after the pandemic.

- There is a strong correlation when total COVID cases are compared to the number of passengers. For example, in January 2022, there was a steep decrease in passenger count but a steep increase in the number of cases. As a result, COVID-19 significantly impacted the airline industry.

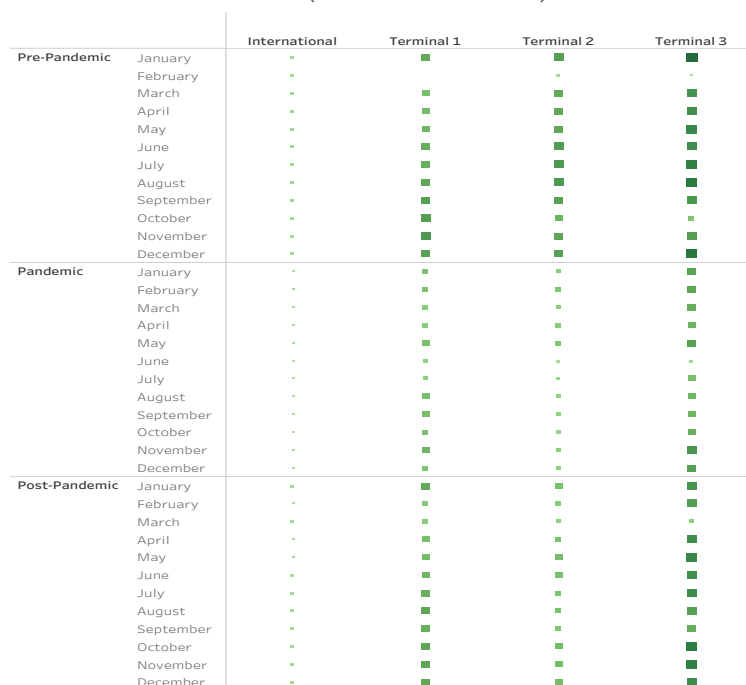
4.2 Hypothesis 2

We hypothesize that the heatmap of terminal utilization at San Francisco International Airport will exhibit seasonal variations, reflecting travel patterns, with certain months experiencing more than 30% passenger numbers after Covid (Jan 2022-Dec 2022) as compared to Pre-Covid (Jan 2018-Dec 2019). Additionally, we expect to observe a substantial reduction of 60% in passenger activity across all terminals during pandemic months compared to pre-pandemic months

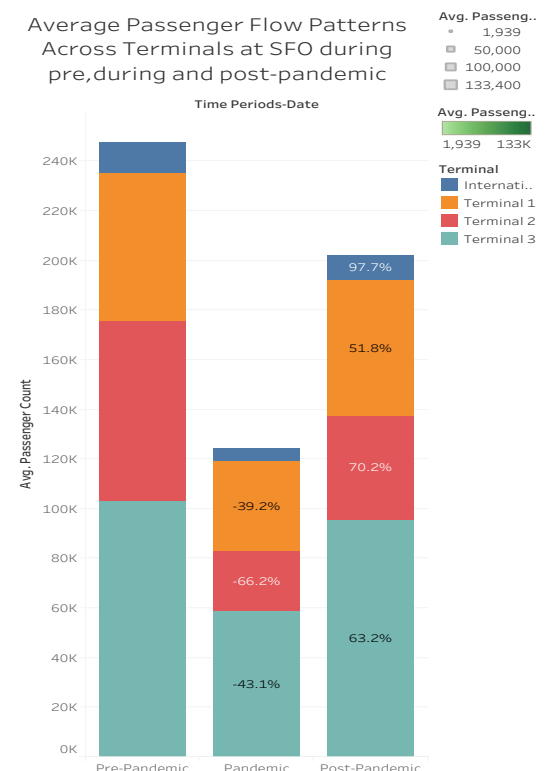
Visualization:

Analysis of Passenger Flow Patterns Across Terminals at SFO: Pre(Jan 2018 - Dec 2019), During (Jan 2020 - Dec 2021) and Post-Pandemic Periods(Jan 2022 - Dec 2022)

Passenger Flow Patterns Across Terminals at SFO: Pre(Jan 2018 - Dec 2019), During (Jan 2020 - Dec 2021) and Post-Pandemic Periods(Jan 2022 - Dec 2022)



Average Passenger Flow Patterns Across Terminals at SFO during pre,during and post-pandemic



Findings:

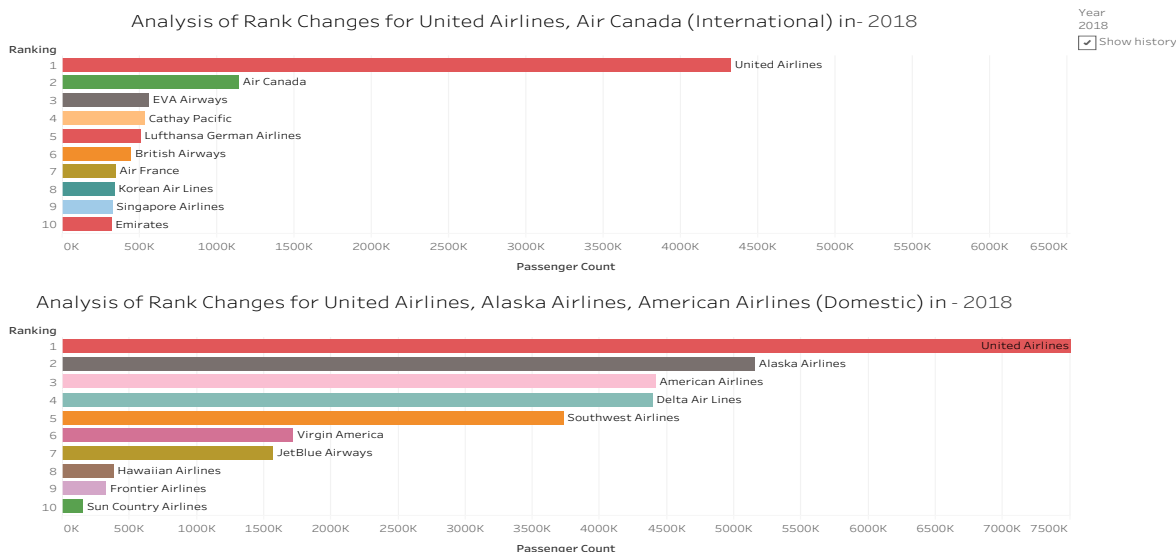
- Flow at each terminal post-pandemic is lower than during the pre-pandemic period. This suggests that passengers still hesitate to travel, even after the pandemic has subsided. This indicates that it had a significant impact on air travel.
- The number of passengers is typically higher during the summer months (June, July, and August) than during the winter months (November, December, and January).
- Passengers at all terminals have increased since the start of the post-pandemic phase. However, the increase has slowed at Terminal 1, which is still below its pre-pandemic passenger count.
- The proportion of passengers using Terminal 2 increased from 66% to 69%, while the proportion of passengers using Terminal 3 increased from 44% to 68%.
- The most significant decrease was seen at Terminal 2, by 67%.
- International Terminal experienced significant decreases in passenger numbers after COVID.

4.3 Hypothesis 3

We hypothesize that the pandemic has changed SFO travelers' top airline preferences, specifically United Airlines and Air Canada for international flights and United Airlines, Alaska Airlines, and American Airlines for domestic flights. We'd examine how their ranks have evolved from pre-covid (Jan 2018-Dec 2019), throughout the pandemic (Jan 2020-Dec 2021), and post-covid (Jan 2022-Dec 2022)

Visualization:

The Impact of the Pandemic on SFO Travelers' Airline Preferences: An Analysis of Rank Changes across Pre-COVID (Jan 2018- Dec 2019), COVID(Jan 2020-Dec 2021), and Post-COVID(Jan 2022- Dec 2022) Periods for Airlines



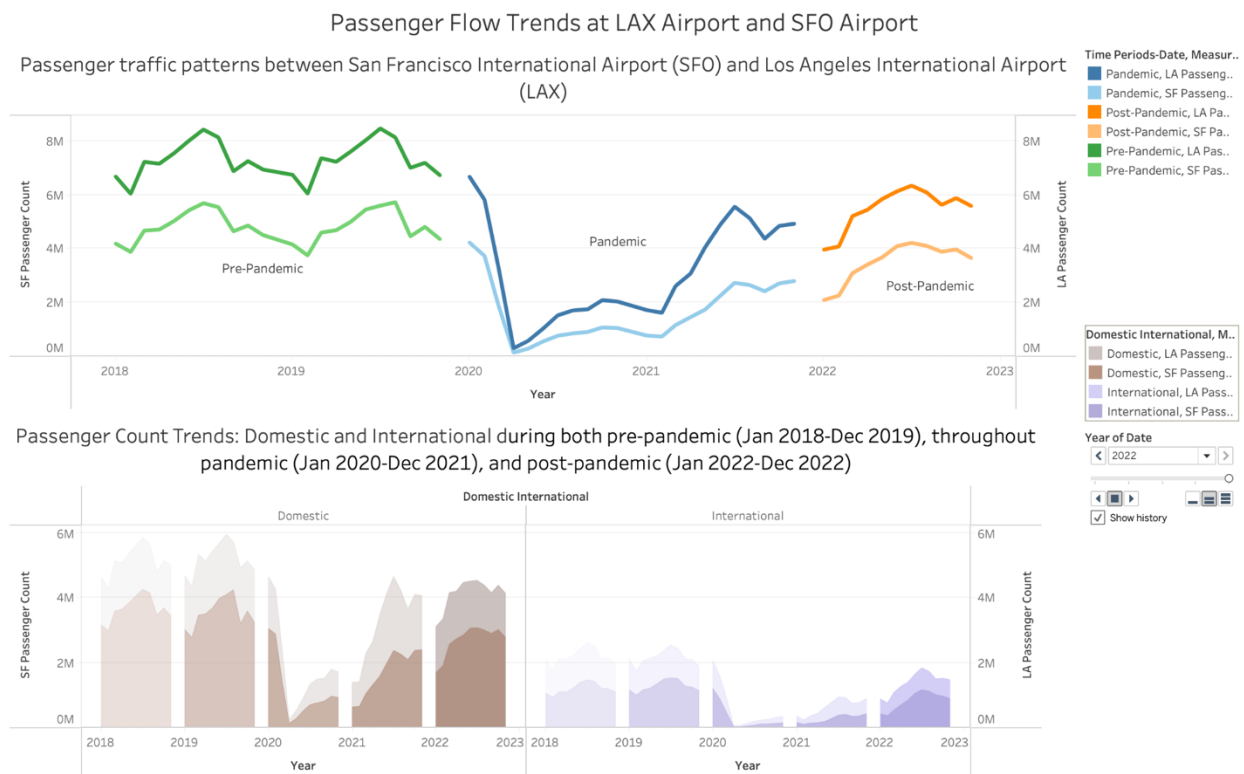
Findings:

- In the International airline industry, United Airlines and Air Canada consistently hold the top two positions. Other airlines often change rankings yearly, and some drop out of the top 10 altogether.
- Similarly, United Airlines and Alaska Airlines are consistently ranked first and second in the domestic airline sector. American and Delta Airlines tend to switch between the third and fourth positions. Most airlines generally maintain their places within the top 10 rankings.
- In summary, the pandemic didn't have much impact on rankings.

4.4 Hypothesis 4

We propose a hypothesis suggesting that passenger traffic patterns are similar between San Francisco International Airport (SFO) and Los Angeles International Airport (LAX). Specifically, we anticipate that LAX Airport will experience a higher passenger count overall and that domestic flights outweigh international flights in frequency during both pre-covid (Jan 2018-Dec 2019), throughout the pandemic (Jan 2020-Dec 2021), and post-pandemic (Jan 2022-Dec 2022)

Visualization:



Findings:

- Trends across SFO and LAX are similar across the years
- Our analysis reveals a consistent preference for domestic flights over international ones at both airports, surpassing international ones by 80.21% across all three periods.
- The visual data shows a clear impact of COVID-19, with a sharp drop in passengers from April 2020 to December 2021, followed by a gradual recovery.
- The pandemic caused a 50% decrease in passenger numbers; post-pandemic, there was a 14% dip compared to pre-pandemic levels.

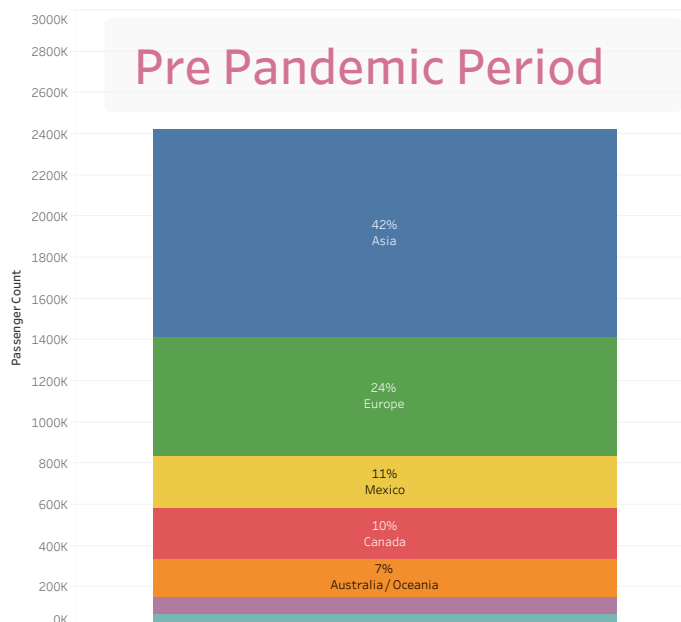
4.5 Hypothesis 5

We hypothesize that during the Pandemic (Jan 2020-Dec 2021), there will be a 50% decrease in SFO passengers traveling to Asia and that post-pandemic (Jan 2022-Dec 2022), the passenger count will be like that of pre-pandemic (Jan 2018-Dec 2019). Furthermore, we hypothesize that there is a change in the percentages of passengers arriving and departing before, during, and after the pandemic, indicating changes in passengers' travel preferences.

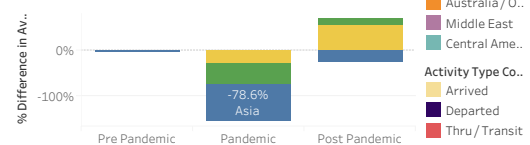
Visualization:

Analyzing Passenger Count Variations at San Francisco Airport Across Pre (Jan 2018-Dec 2019) , During (Jan 2020-Dec 2021), and Post-Pandemic (Jan 2022-Dec 2022) Eras for Insight into Travel Preference Shifts

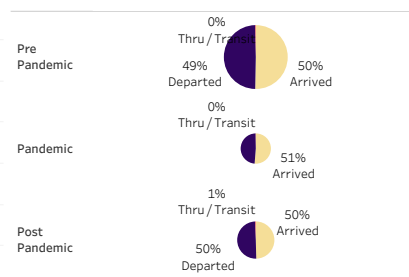
Analyzing Travel Preference Trends Across Continents and Sub-Continents from San Francisco Airport



Evaluating Average Passenger Count Percentage Changes in Top 3 Continents/Sub-Continents Post-Pandemic



Assessing Proportional Changes in International Passenger Count for Arrival and Departure Pre, During, and Post-COVID



Findings:

- Travel preferences of SFO passengers have shifted significantly.
- The visualization shows that Asia was the most popular travel destination, followed by Europe and Mexico.
- However, things changed dramatically during the Pandemic, when the average passenger count for Asia dropped by nearly 80%, and passengers preferred to travel to Europe over Asia.
- However, post-pandemic customer count, there has been an increase in the average count by nearly 50% of Mexico, indicating that there has been an evident change in passenger traveling patterns because of the pandemic.
- In addition, when we examine the trends in the proportion of people arriving and departing SFO over the three periods, we see that the pattern of arrival and departure is nearly identical, indicating that COVID-19 did not affect the arrival-to-departure ratio.

5. Conclusion

In conclusion, our comprehensive study unraveled a nuanced narrative of passenger travel at SFO. The pandemic's reverberations were evident, with a substantial decline in passenger counts during the pandemic and a subsequent recovery, albeit with notable shifts in travel preferences and behaviors. Terminal utilization patterns showcased lingering hesitancy post-pandemic, indicating the lasting impact on air travel.

Airline rankings, surprisingly stable through the pandemic, illustrated the industry's resilience. The comparative analysis with Los Angeles International Airport uncovered parallel trends, emphasizing broader industry dynamics.

Notably, our exploration of passenger destinations revealed a significant shift during the pandemic, impacting the popularity of travel to Asia. However, a resurgence in trips to Mexico post-pandemic indicated evolving passenger preferences.

This project contributes valuable insights to the aviation sector and underscores the industry's resilience in adapting to unprecedented challenges. As we conclude, the findings pave the way for informed decision-making, strategic planning, and continued vigilance in monitoring passenger trends to navigate the uncertainties that lie ahead.