Air Traffic Passenger Statistics

Ajwad Almajnuni

Abstract :

The goal of analyzing Air Traffic Passenger Statistics data is because there is a company that wants to present an advertising campaign for a specific product inside the airport, so the data will be analyzed according to the number of passengers to attract the largest possible number of this product.

Design :

San Francisco International Airport (SFO) is an international airport in San Mateo County, 13 miles (21 km) south of downtown San Francisco, California. It has flights to points throughout North America and is a major gateway to Europe and Asia.

San Francisco International Airport is the largest airport in the San Francisco Bay Area and the second busiest airport in California after Los Angeles Airport. In 2017, it was the seventh busiest airport in the United States and 24th busiest airport in the world by passenger numbers. Therefore, this company wants to present its product advertisement at this airport to attract the largest possible number of people. So I answer these questions:

- The five most visited countries in 2020?
- What is the expected increase in the number of passengers by years ?
- What is the most month in which the number of passengers was high ?

Data :

San Francisco International Airport is the largest airport in the San Francisco Bay Area and the second busiest airport in California after Los Angeles Airport. In 2017, it was the seventh busiest airport in the United States and the 24th busiest airport in the world by passenger numbers. This describes the San Francisco International Airport (SFO) passenger statistical data available for download through DataSF. The data is self-reported by the airlines and is only available on a monthly level and the data is updated quarterly. The form has 46,670 rows and 12 columns, the sample data I chose is the last 3 years (2019, 2020, 2021).

Algorithms :

1-Load the dataset (Air Traffic Passenger Statistics)
2- Data exploration
- display the 10 rows in first
- display the 10 rows in end
- shape of data frame
- find the data type of each column

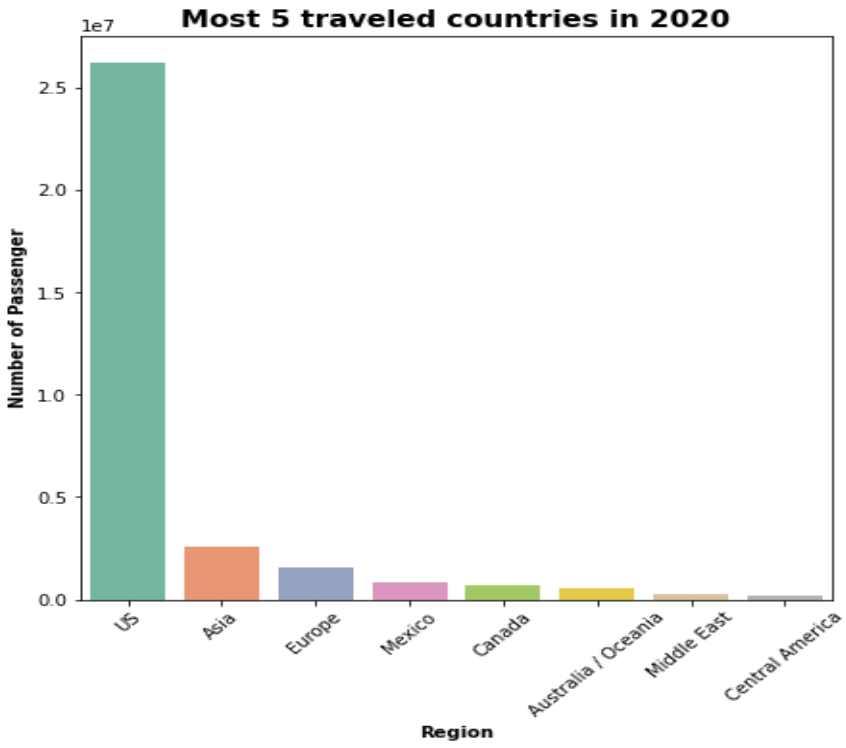
3- Data cleaning:
- Stripping columns from whitespace .
- drop nulls .
- drop rows that is not needed .
- Extracting the MONTH and YEAR from Activity Period column .
- Adding other columns needed for analysis the data :
  1- Month column
  2- Year column

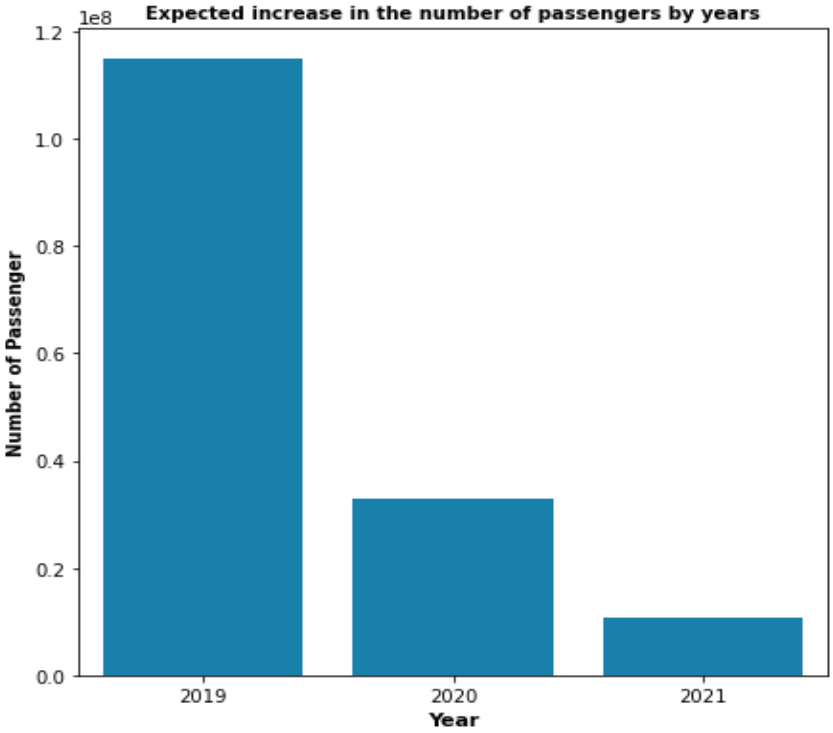
4- Data analysis
Take subset from dataset by year 2020 to easy analysis and then calculated sum to get five most visited countries in 2020 .The increase expected has been calculated in the number of passengers by years . After that I calculate the most month in which the number of passengers was high .

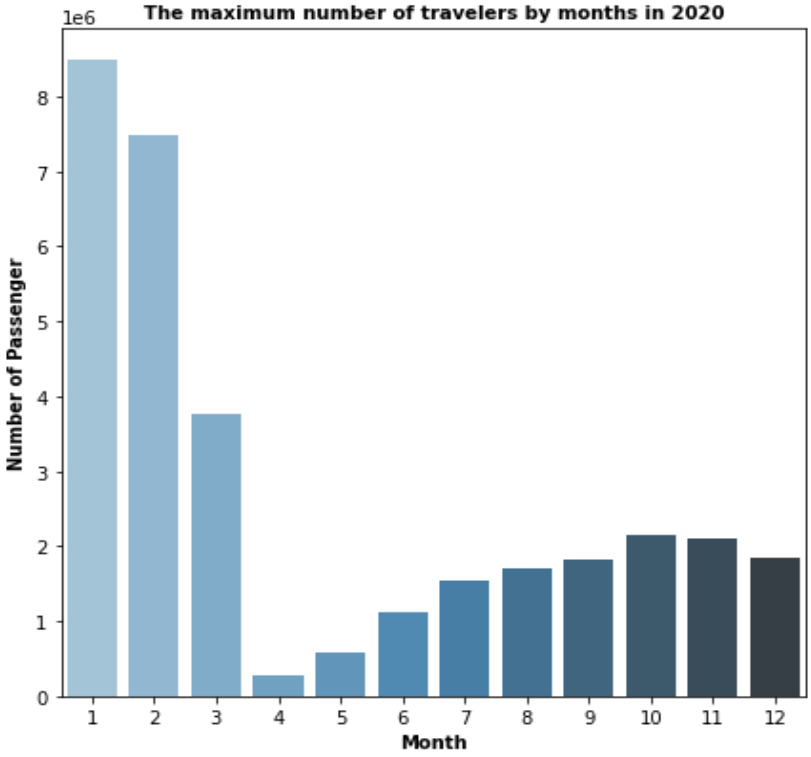
Tools :

- Python and Jupyter Notebook
- Pandas for data manipulation
- Matplotlib and Seaborn for plotting visuialization

Communication :




Expected increase in the number of passengers by years
1e8
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Number of Passenger
2019
2020
2021
Year


The maximum number of travelers by months in 2020
1e6
8
7
6
5
4
3
2
1
0
Number of Passenger
1
2
3
4
5
6
7
8
9
10
11
12
Month