ENJOY TRAVELING BY AIR

Ying Song

This project aims to visualize the traveling by airline of different age subgroups in 2022. The dataset is publicly available in Kaggle, which contains 98,615 synthetic samples with 15 dimensions(please check https://www.kaggle.com/datasets/iamsouravbanerjee/airline-dataset).

THE DATA PROCESSING CAN BE SIMPLY DESCRIBED AS FOLLOWS:

1. Only choose the unique passenger ID, age, and departure date variables.
2. Truncate the age variable into 9 different ranges.
3. For each date, count the traveling times of different age subgroups. Store the results in a table.
4. Resample the table by 7 days for clear visualization and then sum the results into a new table.
5. Visualize the resampled table with the stack area chart. The x-axis represents weeks in 2022, y-axis counts the number of travelers. 9 different colors denote 9 age ranges, see below.

Note: the colormap is sampled from the background picture(published online for download). And the three figures cannot be integrated into one since the last two will be blurred after image reduction.

Traveling of Different Age Subgroups by Air in 2022
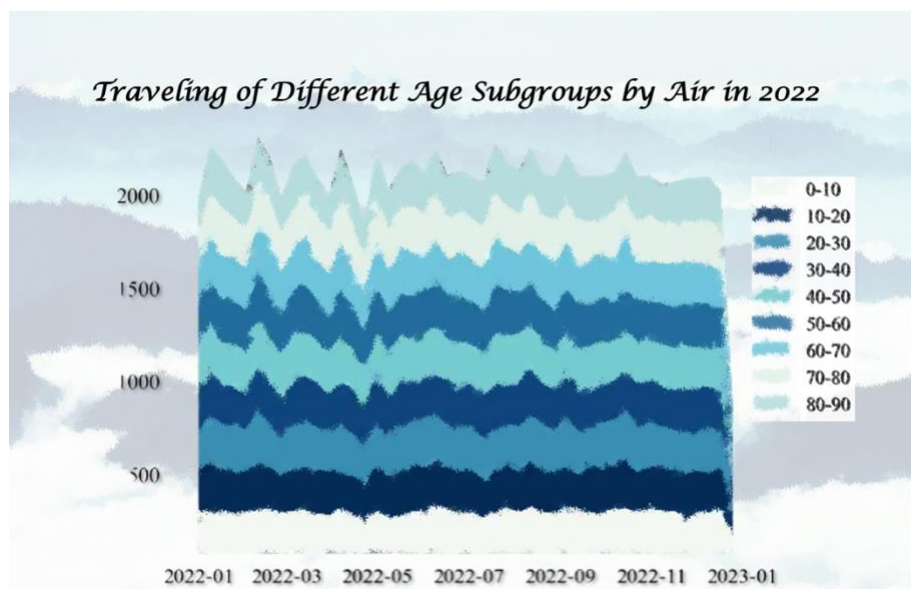
GIVEAWAYS:

1. In all, there are obvious peak and slack seasons in traveling by air. E.g., there are some peak seasons in several beginning weeks in January&February, please note that there are many holidays worldwide during this time; likewise, there are several slack seasons in April&May, the possible reason may be that people are struggling with work or study or decide to spend quality times with their loved ones.

2. One counterintuitive thing is that no matter how the whole traveling times are changing, the percentage of kids less than 10 years old that take the airline change faster over time(Since the overall changes fast but the area of kids less than 10 are close to an almost straight line). And their changes are close to teenagers and young adults around 10-20 changes. But the kids require more company from their parents or their relatives to take planes.

3. The changes of people in 20-30 and 30-40 are close; the changes of people in 40-50 and 50-60 are super close; and the changes of people beyond 60 are super close. But in general, their differences are nuanced. And people in 20-60 are more likely to take the plane than the rest.

4. Simply check the notebooks in Kaggle and then find their tables and figures fit into reality, which reveals the high quality of the synthetic airline dataset. However, when it involves 3 dimensions like in our case, everything becomes too ideal(the proportions of different age subgroups in

different months are too close). But anyway, hope everyone in the world can enjoy their peaceful leisure time and have more traveling if they want. No conflict, no war, and no loss.

THE DATA PROCESSING CAN BE SIMPLY DESCRIBED AS FOLLOWS:

1. Only choose the Country Name and Continent variables.
2. Count travelers by each country and assign the continent to each counted country.
3. Visualize the table with the sunburst chart. The outer circle represents the traveling records of different countries, where the size of different countries depends on the number of travelers, and so does the degree of color. The inner circle denotes the traveling record of different continents. The size and color have the same meanings as the inner circle.

GIVEAWAYS:

Traveling Records among Different Countries in Different Continents
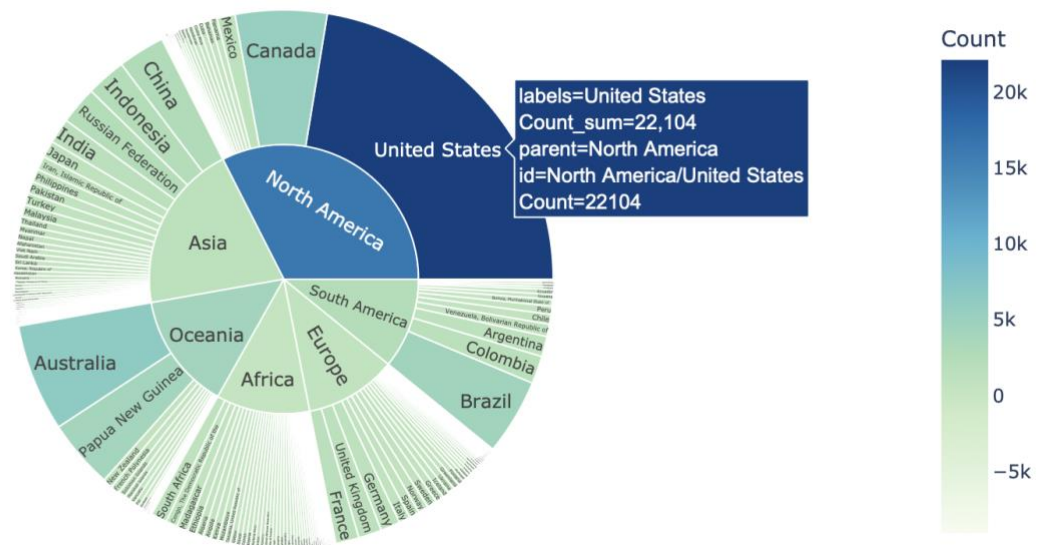

1. Note that the traveling records are entry records. The US is the most welcome traveling destination, the second one is Australia, the third one is Canada, the fourth is Brazil and the fifth is Pagua New Guinea. The most welcome continent is North America, which contains the two most welcome countries. The second is Asia, where China, Indonesia, Russia, and India are the most welcome countries, the third is Oceania, which also contains the two favorite countries. The next one is South America, which only contains one favorite country.
2. Besides the above-mentioned countries and continents, the most welcome country in Africa is South Africa, and the rest are almost the same; and the most welcome countries in Europe are France, UK, and Germany, the rest are almost the same.

THE DATA PROCESSING CAN BE SIMPLY DESCRIBED AS FOLLOWS:

1. Only choose the Nationality variables.
2. Count travelers by each country.
3. Visualize the table with the pie chart. The different circle represents the traveling numbers of different countries, where the size of different countries depends on the number, and so does the degree of color.

Which Country's Citizens Love Travelling Most
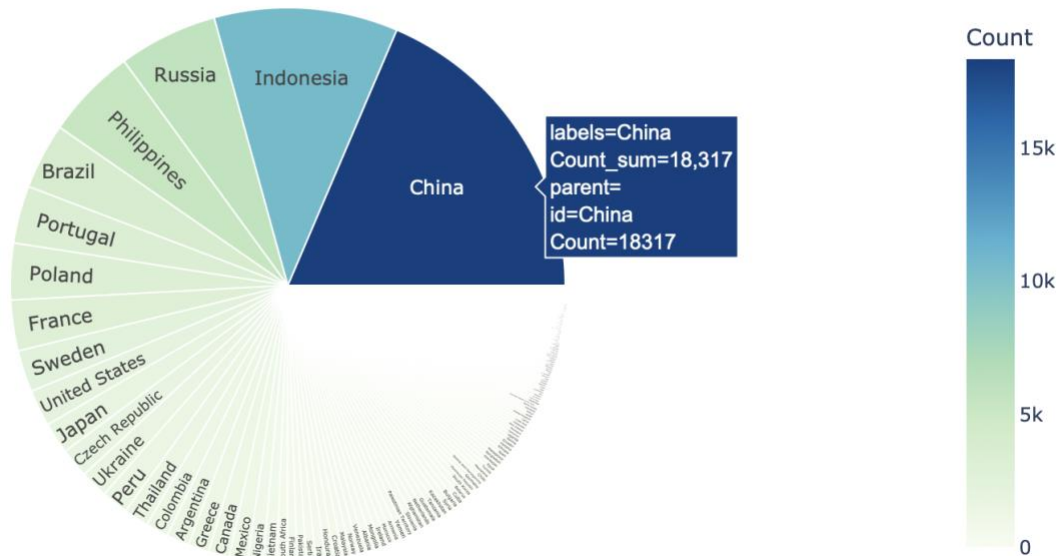
1. Chinese love traveling most, the following are Indonesian, Russian, and Philippinese, which all belong to Asia. The next is Brazil from South America, Portugal, Poland, France and Sweden from Europe, US from North America, and Japan from Asia.
2. The rest countries enjoy almost the same numbers for traveling.

WHY THE PRESENTED FIGURES ARE SIGNIFICANT:

1. This aligns with the research. After the COVID-19 pandemic, the traveling industry revives. It is high time to find traveling patterns and leverage these patterns for some specific purposes, e.g., boosting economics, improving personal experiences, and fostering communications, etc.
2. To explore traveling industries worldwide, we need to identify which time is most favorite by travelers, if possible, policymakers can provide more traveling benefits during the proper time, e.g., providing peak-time bonuses for airport staff.
3. Figuring out which country and continent are the most welcome destinations, and which country loves traveling most helps us better understand traveling patterns.

Note: the data and code are provided in:
https://github.com/yings0930/INFSCI2415.git