Data Exploration

Dataset: Airlines Data

The airline dataset is intriguing because it offers insights into the patterns of traveling customers and what the companies can expect and improve soon. For Example, if the number of delayed or canceled flights is quite large, something must be done to prevent them from happening. I will be uncovering the facts and figures I found in my investigation of the dataset. The first easy check was the average age of the passengers traveling through the air. I found that The mean age of flight attendees is 45.3 even though the life expectancy average is only about 72 years. I expected the average to be lower because the working class from 20 to 35 mostly travel (source: google). The next finding consisted that the number of flights canceled, delayed and on-time are spread out. This means that the number of canceled flights is the same as the ones on time. This wasn't the case in 2010 when only 19% of the flights were cancelled. The next shocking discovery is that the USA has the most incoming and outgoing flights in the whole world in 22104 (about 32%) regardless of its population being 4.23% of the whole. I also found that the first names of passengers mostly start with Aa about 3% of the total which statistically should be 0.14793% which is 20 times the expected value. Names starting with A and S combine a total of 12% of the total population. Combined statistically should only be 7.69%. It is also astounding that the maximum age of a passenger on a flight is 102. Also, the fact remains that the age of passengers isn't actually normally distributed but rather uniformly distributed even throughout the number. This disproves the hypothesis of airplane industries that most travelers fall under the category of working class aged 20 to 35. North America has the most airlines flown to or by whereas South America has the least which is really ironic considering both continents have a greater number of developed regions than any other continent. I also expected Asia to be the continent with the most number of flights but is barely second last even though it contains over half (59.22%) of the world's population. The last discovery is the most interesting one. Most of the people who travel through airlines are actually Chinese Citizens (their nationality is Chinese or their home country is China). This was surprising knowing that most of the planes travel to or from the United States however, most citizens are from China. It makes sense since China currently consists of the most populous country in the world with about 1.412 billion, however, the fact that most men/women traveling in airlines are Chinese even though China has only 2000 planes going to and from. What is China plotting?

That concludes my findings on the exploration of Airplane data for 2022.