**Lesson 04 Code Talk**

The data set that we worked with had full names of all passengers. Where you work, what are the precautions taken to ensure that people are not easily identifiable?

Unfortunately, I have not worked directly with labels directly related to people, so I am curious on how people’s data is handled to ensure they are not easily identifiable. The only example that I have when I had to mask (somehow) some information to protect the identity of the client/customer was when presenting data about a project or product to a potential new client for the GE Shop Services for turbojet/turbofan engines. We needed to show the data/trends and tools we were able to offer to them as well as some of the advantages our current customers experience such as: improvements on engine performance & improvements on fuel consumption. All this by making sure that the data coming from each of the customers was not identifiable (labels, locations and/or anything that could potentially lead to identify the customers/airlines). This are some of the process that I remember (kind of) we followed: Change the label/names and location of the engines and instead plot/present everything in terms of ambient conditions, in other cases we would remove all together name labels and/or location labels and if possible take a random sample from all the data points available removing any labels.