Nghi Huynh

Big Data 2022: Q&A:

1) **Why did you participate in this Big Data Challenge? Why did you choose mental health and suicide rates to analyze?**

   Because we wanted to apply, challenge, and push our data science skills to deal with messy data in the real world and develop ideas and solutions for sustainable health economics. We found that there is an ongoing pressure in many countries, especially in low- and middle countries, when it comes to healthcare systems. However, most of the research only focuses on healthcare in general but does not specifically address mental healthcare. Therefore, we wanted to analyze more about the impacts of socioeconomic factors on mental health and suicide rates to provide appropriate starting points for governments to effectively incorporate those factors into the economic evaluation and healthcare assessment.

2) **Why did you choose DBA clustering technique? Why is Dynamic Time Warping better than Euclidean metrics?**

   There are many clustering techniques to analyze the hidden patterns in our data. However, those techniques typically use generic similarity metrics like Euclidean distance, which aren't explicitly designed for time-series data like ours. Moreover, time-series data have information in the order of data points, so those generic similarity metrics don't usually work. That's why we chose DBA clustering, which is based on the centroid K-means clustering technique. But instead of using the Euclidean distance, it uses the DTW distance to capture the temporal orderings of time-series data.

3) **What do 3 clusters vs 2 clusters represent? Why is it relevant? Cyclic link in detail?**

   High and lower-middle income groups: 3 clusters

   High income:

   - The first 2 clusters group all socioeconomic factors together, while the last cluster only groups suicide-related factors. This result suggests no direct link between suicide rates, mental health, and socioeconomic factors in this group.

   Lower-middle income group:

   - The first two clusters group socio-economic and a few mental health factors together, while the last cluster specifically groups suicide and mental health factors with health expenditure. This result suggests these factors are somewhat related.

   Upper-middle and low-income groups: 2 clusters

- There are only 2 clusters in the upper-middle and low-income groups. When we analyzed the distributions, we found that one cluster is more about socio-economic factors. In contrast, the other cluster is a mix of suicide rates, mental health, and socio-economic factors. This result suggests a significant link between socio-economic factors, mental health, and suicide rates.

4) **Why do you think middle- and low-income groups are more vulnerable during economic recessions?**

   There are many socioeconomic factors such as unemployment rates, inflations, and other stressors that might increase suicide rates during economic recessions. And due to the relationship between suicide rates, mental health, and socioeconomic factors we established in our results, those middle- and low-income countries are more vulnerable to economic recessions. It's like a domino effect. Once those countries get hit by one aspect, it's more likely that other factors will collapse. In this case, once an economic recession occurs, it will increase unemployment rates and mental disorder burdens and, as a result, elevate suicide rates.

5) **How do you quantize mental health?**

   We quantized it based on mental health prevalence and disorder burdens rates

6) **How do you probe it a bit more? Can the vicious cycle between mental health, suicide rates and socioeconomic factors be broken? How?**

   Besides moving forward to the national and regional levels to see the predictions, we'll explore the influences of other social factors such as no stable job, violence, security on suicide rates, and mental disorder burden predictions.

   Moreover, we will specifically look for data or studies on the impacts of poverty alleviation and mental health interventions on health economics in middle- and low-income countries.