**KEY POINTS**

* Based on our problem statements, we can proceed with simple and quick statistics about the variables (eg. Mean, SD, etc of given variable).
* We can then combine the statistics into one main problem.
* Best would be to do clustering on the dataset.
* We can define our own variables. Eg. What does it mean by fatal and non-fatal? Fatal could mean facing life and death situation, while non-fatal could mean able to recover after a duration of hospitalisation.
* Proceeding with how to tackle the problem:

1. Basic statistics on variables.
2. Cluster them together. Avoid using k-means.
3. Perform another set of basic statistics so that we can label the different clusters.
4. If time allows, can do up on which accidents don’t fit the in the data. Can do this just by calculating the accidents that are furthest distance away from the cluster centre.
5. Be very careful about categorical values, since they do not contain the distances needed to do clustering. Have to encode them to do so.
6. All of this is based on unsupervised learning.
7. Some models that are recommended include: DBScan, Hierarchical clustering (Dendogram).

For basic statistics, analyse according to:

Koushan: 1-6

TQ: 7-12

Tai: 13-18

Nas: 19-23