Task5-notes

Simpson's paradox

In statistics, an effect that occurs when the marginal association between two categorical variables is qualitatively different from the partial association between the same two variables after controlling for one or more other variables.

Simpson's paradox is not simply an obscure phenomenon of interest only to a small group of statisticians.

Simpson's Paradox is not only a surprising mathematical fact; it serves as a lens through which to understand the role of probabilities in data analysis, causal inference, and decision-making.

Why does Simpson's Paradox happen?

Simpson's Paradox happens because disaggregation of the data (e.g., splitting it into subgroups) can cause certain subgroups to have an imbalanced representation compared to other subgroups. This might be due to the relationship between the variables, or simply due to the way that the data has been partitioned into subgroups.

Simpson's Paradox is a tricky issue, but a good analyst or data scientist can handle it with the right tools and knowledge. We hope our new work can help others deal with this issue in an easier and more efficient manner.