Research Task: Prediction on a quantitative response variable (how fast a person can drive before the person gets a ticket) based on several explanatory variables

Data Features:

The data is tall and wide. The response variable(speed) is quantitative. It also has a large collection of explanatory variables which are both quantitative and categorical.

Analysis Strengths: It can find which explanatory variable contributes the most, a reasonable percentage of variance was explained by the model. It also used machine statistical learning, model to train, and random forests.

Analysis Weaknesses: The mean of squared residuals is high, which suggests the variance of the prediction is also high. The data is only collected from certain county which cannot really give good prediction.

Alternative Example:

Statistical Analysis on factors influencing Life Expectancy

It contains a large collection of explanatory variables which include both quantitative and categorical data.

The data is collected from the WHO data repository website and its corresponding economic data was collected from United Nation website.

<https://www.kaggle.com/kumarajarshi/life-expectancy-who/data>