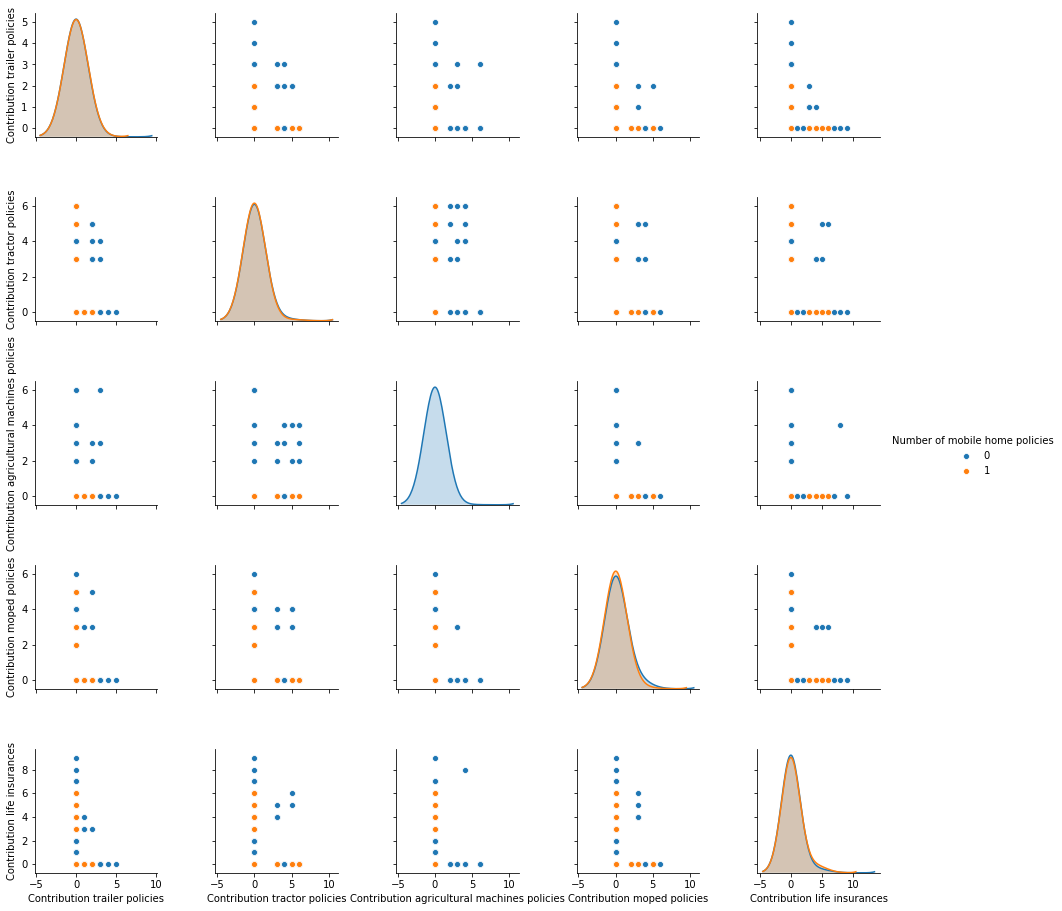
Data Story Telling

Starting from [13], I plotted a horizontal bar chart with the bars represent the average probability that bucket had a caravan insurance policy. So far it looks like that there have not been any variable that immediate stands out with significant separation visually with the KDE histograms.

Example of Pair plot and KDE histogram:



Also, the data is highly imbalance, which seems logical as these policies are not often bought like personal auto or property. About 94% of the observations in my training data have no caravan insurance, so it means pick a random zip block, 94% chance there’s no one with a caravan Insurance.

Number of mobile home policies Count %

0 94.02

1 5.98

Name: Number of mobile home policies, dtype: float64

Majority of the Cramer-V plot suggest that majority of the independent variable are independent as shown in the heatmap. (0 means absolutely no correlation to 1 which is perfectly correlated)

For example:

