Fortegra Data Challenge

A new potential partner is coming to us with a potential portfolio for us to work with. The portfolio is insurance for irrigation sprinklers, a type of farming equipment. They have been running this program for about 5 years and are looking for a new partner to work with them. We have an opportunity to analyze their data and recommend underwriting changes to improve the performance of their portfolio. We are targeting a loss rate of 60% or lower for any portfolio we work with. Loss ratio is defined as sum of total claims paid / sum of total premium earned, over a given time period.

There are 2 datasets attached: claims.csv has one row per claim filed and premiums.csv has one row per piece of equipment.

1. Clean the data how you see fit. State your assumptions. Assume that a cleaned subset of the data is better than data which is skewed, biased, or incorrect. Include conceptual validations/cleaning.
2. Calculate loss ratio by state, county, customer’s number of previous claims, deductible amount, and equipment year. Explain your results and note anything you found interesting or important.
3. Perform other pieces of EDA/analysis you think is interesting. Since time is limited, you can limit to 1-2 analyses, then list other analysis you think would be interesting to look at if you had more time.
4. Based on the analysis you’ve done, what are some recommended changes to the underwriting of this portfolio?