Joseph Nebres

### First choice: dataset 1 (P.S.: I’m a Life Insurance Agent and would really love to work on this one!)

1. https://www.kaggle.com/competitions/prudential-life-insurance-assessment/data

2. This dataset provides data that describes people's lifestyles and how their lifestyles affect their level of risk when applying for life insurance.

3. The target is “Response” which is an ordinal level of risk assessment.

4. Each row represents a person.

5. This is a regression problem, since the level of risk is ordinal, I believe. .

6. 128, I’m hoping I can change the data to be at the 25 column mark, since the columns seem to be pretty vague.

7. 59381 for the training data

8. This dataset was used in a competition with a cash prize on Kaggle, but the competition ended 7 years ago. I hope this data is public and safe to use for educational purposes. Also, they separated the training data with the test data in two separate CSV files. I don’t know if I need to glue them together. This dataset has a lot of features, but I think I can remove some of them to fit within the 25 mark. I was hoping to use the features from Product\_info1 to BMI and the target as ‘Response’, this way I would have 11 features. Please let me know how lenient this can be, I am really passionate about life insurance. I would actually show my knowledge of how machine learning can help underwriting! Here is the dataset for reference (I uploaded it on Google): https://docs.google.com/spreadsheets/d/e/2PACX-1vQjrqP-rni9ezWD60lZS4QJWiQWO8Fu69iOQnBVs9nGosi6aD30PfoXYJBCrA-LIJ5Ek\_giJQUf5OdG/pub?gid=517654057&single=true&output=csv

### Second choice: dataset 2

1. https://www.kaggle.com/datasets/sagnik1511/car-insurance-data

2. A company’s shared data of annual car insurance that can help predict customer behaviors.

3. The target is the ‘outcome’ column, in other words if a custom has claimed their loan.

4. What does one row represent? (A person? A business? An event? A product?)

5. This is a classification problem.

6. 18

7. 10000

8. No challenges, just need to clean up the data. There are missing values and a column that isn’t needed for machine learning.