Modeling Car Insurance Claim Outcomes

Goals: Using variables/features to fit logistic models and to identify the best feature

Technologies: Python, Pandas, Numpy, logit (from statsmodels.formula.api), seaborn, matplotlib.pyplot

Description:

1. Read in data from pandas.read_csv()
2. Applied .info() and .head() to check missing values and data type conversion if needed
3. Visualized missing data columns using histogram and imputed NA with .mean()
4. Converted object data columns into ordinal categorical variables based on dummies code provided by stakeholders (a for loop was created to decrease repetitive work)
5. Used .drop() to drop undesired columns during fit process
6. Created second loop to fit logit model and used .pred_table() to get confusion matrix and determined best features and best accuracy
7. Delivered result