**Problem approach:**

Data do not contain any missing values and data is clean.

Found data distribution for target variable and claim\_amount both are not normally distributed. Initially used log transformation but results not much improved.

Better results obtained without major transformation , dropped “id” column as all values are unique and built pipeline with categorical variables and numerical feature variables as claim\_amount and vintage , taking all other as categorical .

**Got better results:**

using K-fold cross validation and using Gradient Boost regression algorithm and used hyper parameter tunning trail and error with values n\_estimators=85, max\_depth=3,learning\_rate=0.075