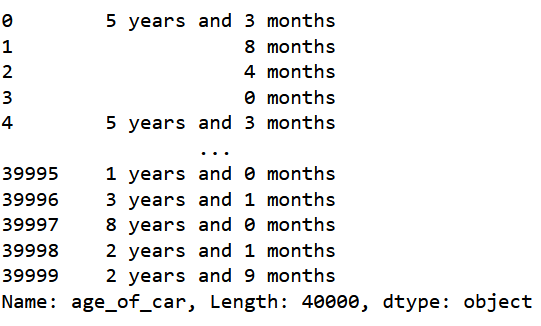
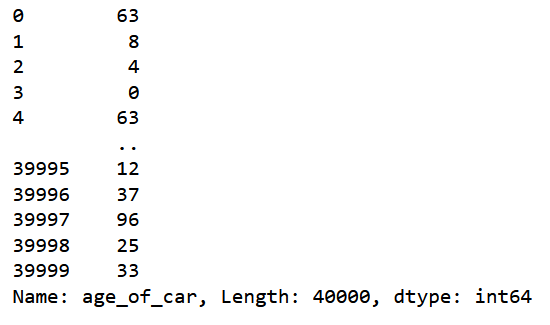
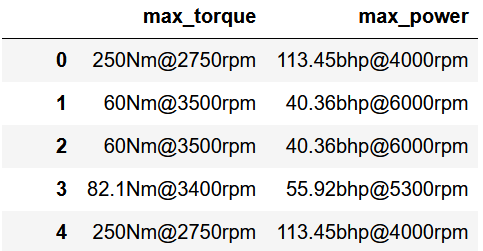
**Preprocessing:**

1. Drop Unnamed: 0 and policy\_id column
2. Convert age\_of\_car data to number of months

1. Replace Yes/No values of dataset to 1/0
2. Convert datatype of **make** as object
3. Compute area using length, width and height
4. Extract Nm ,rpm values from max\_torque and bhp, rpm from max\_power



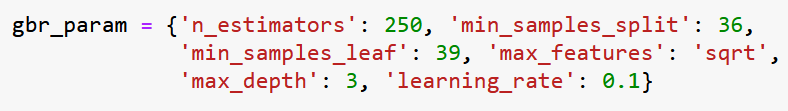
1. Drop features with correlation >0.97, therefore we dropped

['is\_central\_locking', 'is\_driver\_seat\_height\_adjustable', 'airbags', 'is\_rear\_window\_washer', 'is\_ecw', 'gear\_box']

1. Using pipeline and column transformer, we applied SimpleImputer, StandardScaler and OneHotEncoder to process the data

**Machine learning algorithms:**

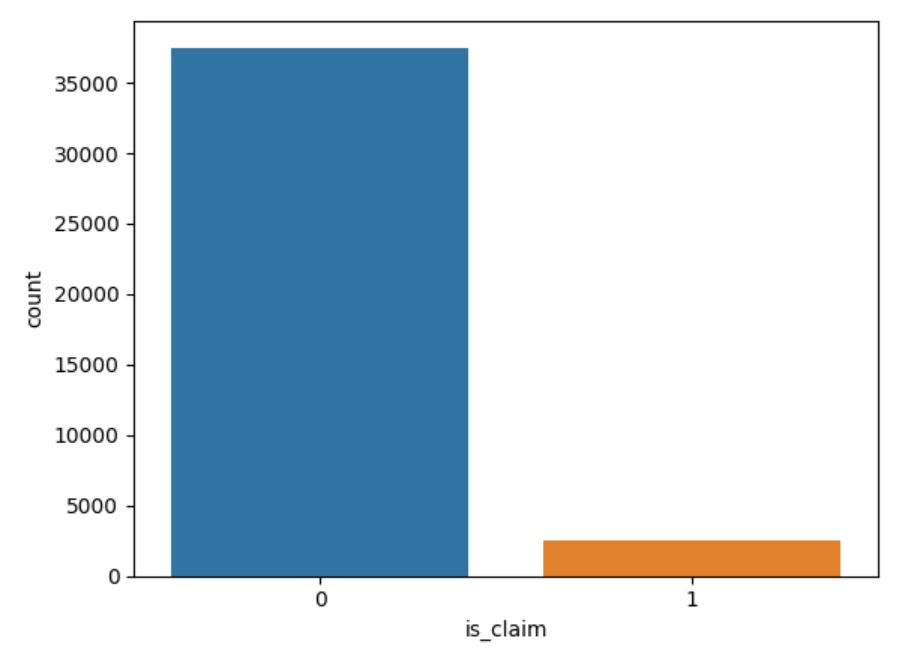
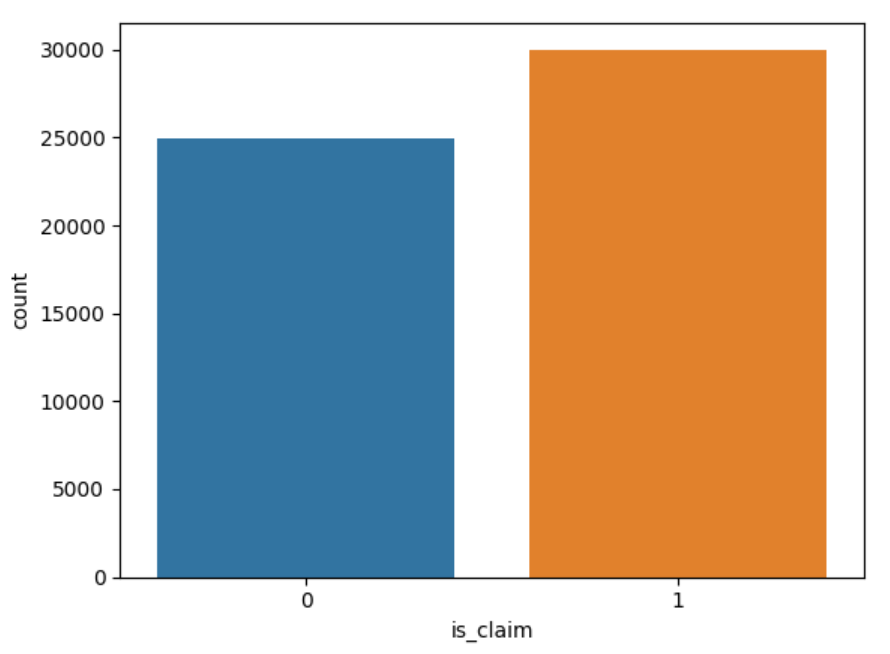
1. Tried various machine learning models that sklearn supports and found that machine learning algorithms is performing best for regression task whereas XGBClassifier is the best for classification task
2. To further optimize the model we tried, hyperparameter tuning using **RandomizedSearchCV** and found this as best parameters for regression



And for classification task:



1. Also, data was highly imbalanced as shown below, so we applied Over-sampling using SMOTE and cleaning using ENN Combine over and under-sampling using SMOTE and Edited Nearest Neighbours.

**Discuss how your solution provides business value to an organisation working in the insurance industry**

1. Regression, as well as our classification model, is optimised based on a given dataset. Also, the data is balanced before training for classification tasks to avoid biases towards a particular claim status.
2. Evaluating risk and optimizing pricing are crucial for car insurance providers. Through precise prediction of insurance claim likelihood, insurers can effectively determine the risk associated with covering a specific vehicle or driver. This, in turn, allows them to:
   * Set competitive premiums for low-risk customers, making them more attractive to this segment and helping to retain their business.
   * Properly price higher-risk policies, ensuring profitability and protecting themselves from potential losses.
3. These models can detect suspicious behavior or inconsistencies in reported incidents, enabling insurers to:
   * Detect fraudulent claims early in the process.
   * Intervene promptly to investigate and validate claims.
   * Prevent payment of fraudulent claims, thus minimizing financial losses.
4. Timely and accurate prediction of insurance claims enables insurers to streamline claims processing and provide faster payouts to policyholders. This enhances the overall customer experience
5. Predictive models can help insurance companies optimize resource allocation by prioritizing claims with a higher likelihood of validity or severity. This allows insurers to allocate their claims processing resources more efficiently, reducing processing times and operational costs. Additionally, by automating routine tasks and decision-making processes through predictive analytics, insurers can free up human resources to focus on more complex and value-added activities.