Exercises: Imbalanced Data Classification

Exercise 1 Consider the provided dataset for churn prediction.

1. Explore and pre-process the dataset.
2. Use cross-validation, and build a model to predict churn rate. It can be any kind of model, or it could be a model that you haven't seen in the class.
3. Once the model is built,
    - Show the confusion-matrix
    - Analyze precision, recall and f1-score of the two classes.
4. Improve f1-score of the minority class using
    - Under-sampling
    - Over-sampling by duplication the minority class
    - Over-sampling by SMOTE
    - Compare performance from the three techniques