# Report

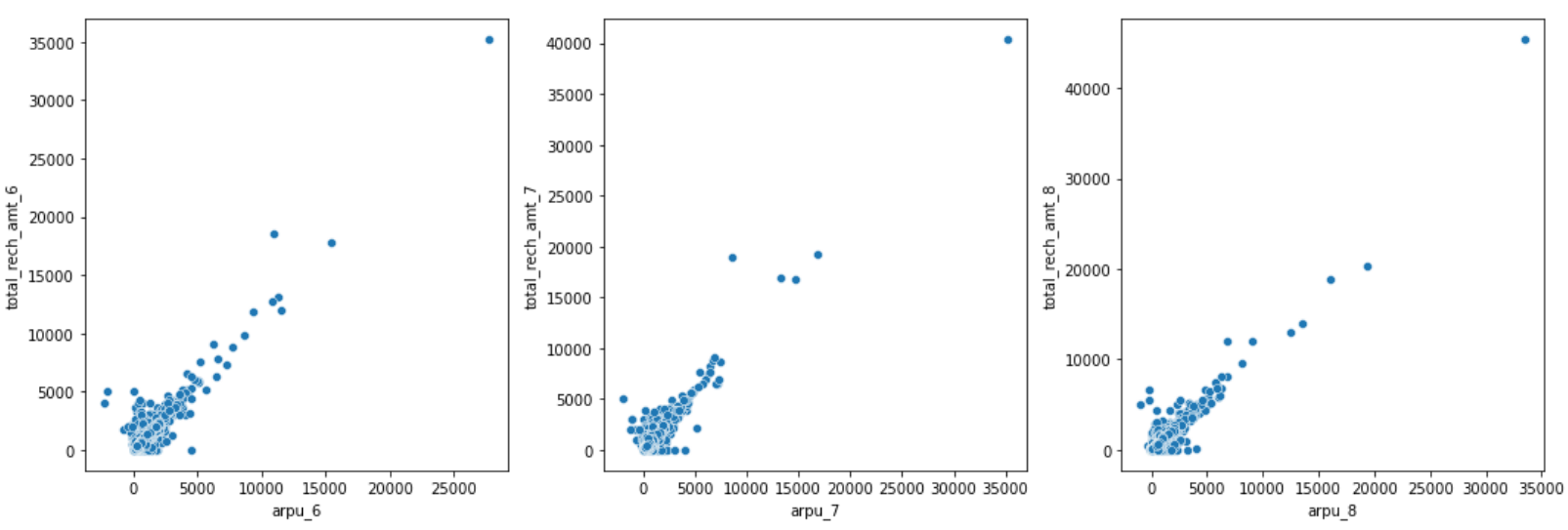
### **Executive Summary**

*This assignment is for predicting churn in a telecom company using the data of 10,000 customers provided by the company. The assignment only targets to reduce the churn of high-value customers. Conducted feature engineering, data visualization, modelling of several machine learning models, and building a neural network, to achieve the said task. Finally, we calculated misclassification cost.*

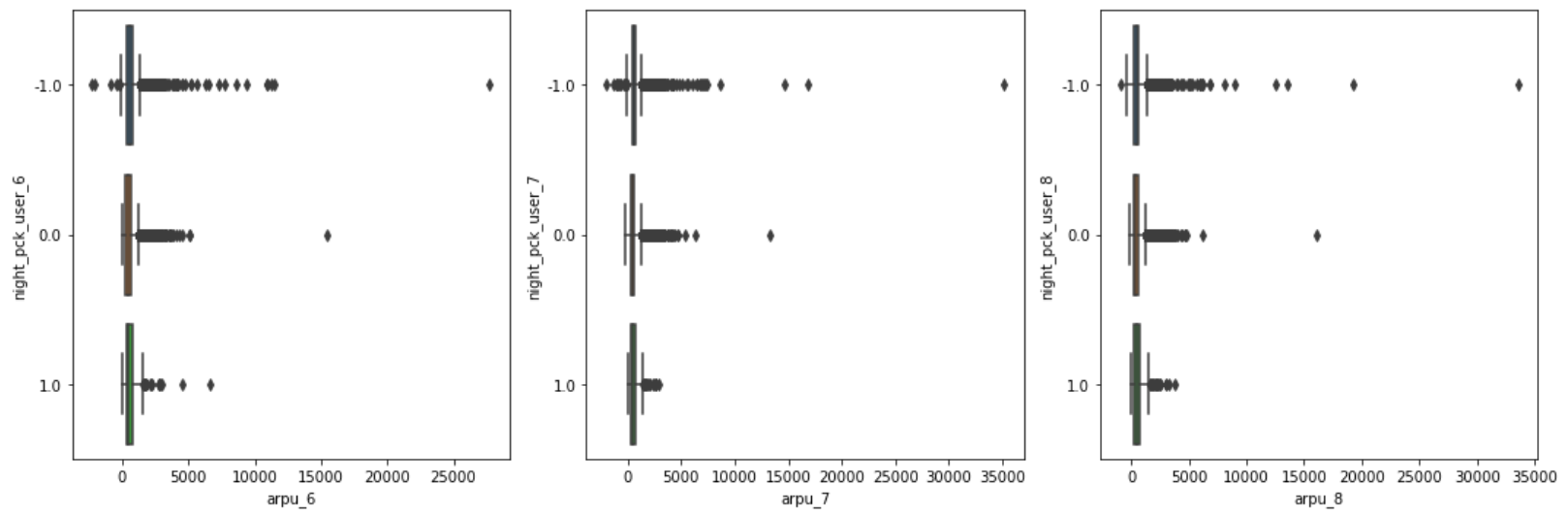
*We concluded that the best model for classifying churn is Random Rainforest with Random Oversampling. This model achieved the testing accuracy of 98.6%, F1 score of 99%, precision of 99%, and recall value of 96%.*

### **Business Insights and Recommendations**

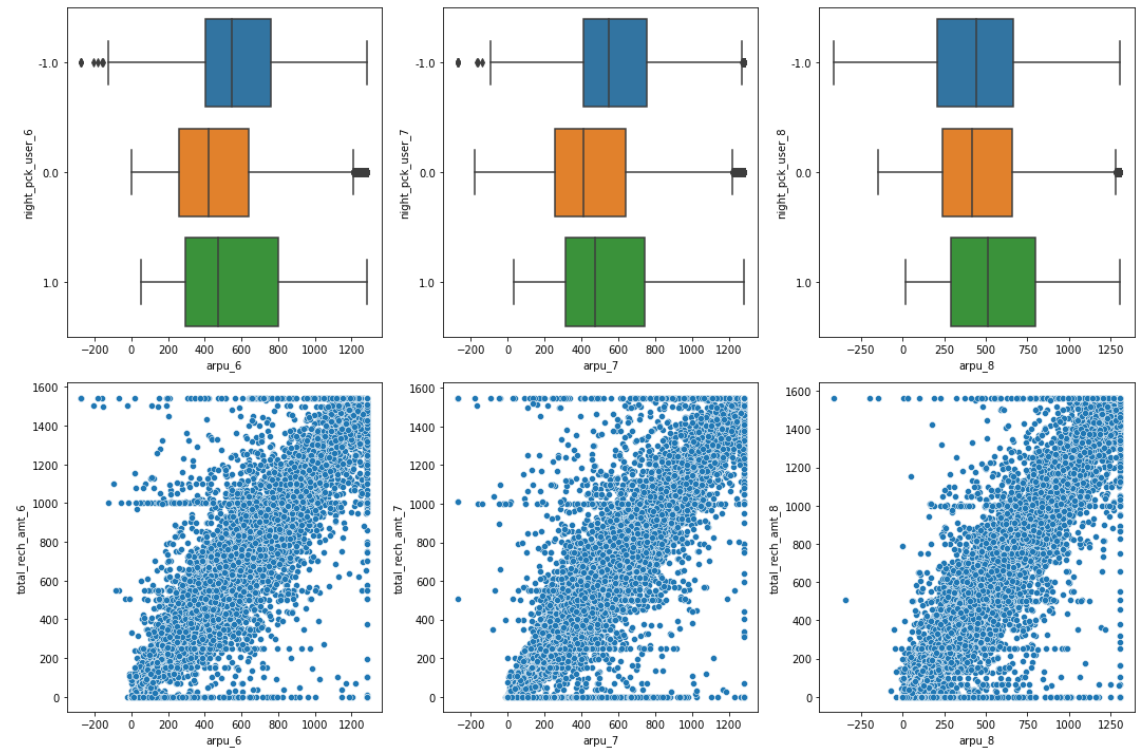
*Out of 10,000 customers, only about 3000 were high-value customers, i.e., the customers who have recharged with an amount more than or equal to the 70th percentile of the average recharge amount in the first two months.*

*Below is arpu vs total recharge data on a scatter plot before outliers’ treatment.*

*Below is arpu vs night pack users’ data on a box plot before outliers’ treatment.*

**

*Below is arpu vs total recharge data on a scatter plot & arpu vs night pack users’ data on a box plot after outliers’ treatment.*

**

*The misclassification cost of this model is 1147 on the testing dataset.*

*We can leverage NLP techniques, recent advancements in Large Language Models (LLMs), and Generative AI can provide valuable insights and help predict customer churn. We can use NLP models for sentiment analysis which can help identify customers' emotions and satisfaction levels, which are strong indicators of churn risk.*

*We can also leverage topic modeling to understand customer interactions better. Topic modeling can identify recurring issues and concerns, providing insights for service improvement.*