Outline:

* Identify the necessary libraries
* Functions: 1) Find and identify different methods for outlier detection, anomaly detection,

outlier treatment, data imputation (if needed as data is clean, data imputation is

not needed but we need to decide collectively);

2) Feature Selection

3) Confusion Matrix

4) AUC curve

* Reading Data, Data Visualization by plotting and checking for imbalance(according to the file given the data is unbalanced so we just need to show that)
* Decide and select ways/techniques of re-balancing classes: 1) Oversampling, Under sampling, Near miss (maybe more ways exist I just listed which I found, the reference of which is given below)

(<https://www.analyticsvidhya.com/blog/2020/07/10-techniques-to-deal-with-class-imbalance-in-machine-learning/>)

2)Check “data imputation” is needed or not (I think not needed as the data is clean i.e no missing values)---> we need to check this

* Feature selection and Model Selection
* Hyperparameter Tuning
* Conclusion