1. Missing Vaule:

(1)，Drop all of the missing value

(2)，Mean and Mode(with fisher criteria and sfs)

recall = 0.372,Precision = 0.17770,AUC = 0.6524

(3)，Multiple Imputation(with fisher criteria and sfs)

Recall = 0.373,Precision = 0.17777,AUC = 0.6525

**Conclusion: Multiple Imputation is a better choice**.

1. Feature Selection:

Missing Value: Multiple Imputation

Existing Methods: Correlation Test, Time Consistency Test, Fisher Criteria, Stepwise forward Selection, PCA and Lasso.

Testing Models: Logistic Regression

(1)，Time Consistency + Correlation Test

(2)，Time Consistency + Correlation Test + Fisher Criteria(top20) + SFS(top6)

(3)，Time Consistency + Correlation Test + Fisher Criteria(top180) + SFS(top100)

(4)，Correlation Test + PCA

(5)，Correlation Test + Lasso

(6)，PCA

(7)，Lasso

Shrink two features down to 0, recall = 0.72, precision = 0.15, AUC = 0.78, may indicates that correlation test is not necessary?

(8)，Fisher Criteria(top20) + SFS(top6)

Recall = 0.373,Precision = 0.17777,AUC = 0.6525

(9)，Fisher Criteria(top180) + SFS(top100)

**Conclusion:**

1. Model Selection(With cross validation)
2. Logistic Regression
3. SVM
4. Random Forest
5. Xgboost
6. LGBM