Analytic plan.

# The problem.

The aim of research is build an effective model which is able to detect credit card fraud with high probability.

# Data provided.

There are base of 280 S000+ classified transactions with 28 unnamed number characteristics. However, the dataset is very unbalanced. Only ~0.1% of data is classified as fraud transactions.

# Models overview.

Since it is hard to discover data nature, we should use some general, low dependent on data type models. We should try models like Decision Trees or Logistic Regression. As a metric of good our model is we should take ROC/AUC, since our dataset is unbalanced. We should build model with at least 95%.