

## CASSANDRA PS1 WRITE UP

### TEAM DATA\_FREAKZ

Private score:25.68181

Public score :26.30408

At first, we analyzed data set profoundly to know that Description is useless and we dropped it. Since invoice date and due date are in date format, we converted them to separate columns into month, date and day for both invoice and due date. We added ID in keeping vendor\_name in mind. We also added Due\_day, Due weekday, created\_day, created batch attributes correlating the given data in result gave us better score. We also added late attribute on basis of whos late on due payment. We also dropped amount, settled, outstanding which felt usefull and gave us better scores.

Finally we used XGBOOST model which gave us better score comparatively by using multi layered ensembling.

THANK YOU.