**Problem Statement**

Your client is an Insurance company and they need your help in building a model to predict the propensity to pay renewal premium.

You have information about past transactions from the policy holders along with their demographics. The client has provided aggregated historical transactional data like number of premiums delayed by 3/ 6/ 12 months across all the products, number of premiums paid, customer sourcing channel and customer demographics like age, monthly income and area type.

Given the information, the client wants you to predict the propensity of renewal collection.

EVALUATION CRITERIA

The probabilities predicted by the participants would be evaluated using AUC ROC score.

Public and Private Split:

Public leaderboard is based on 40% of the policies, while private leaderboard will be evaluated on remaining 60% of policies in the test dataset.

Data

train.csv

It contains training data for customers along with renewal premium status (Renewed or Not?)

|  |  |
| --- | --- |
| **Variable** | **Definition** |
| id | Unique ID of the policy |
| perc\_premium\_paid\_by\_cash\_credit | Percentage of premium amount paid by cash or credit card |
| age\_in\_days | Age in days of policy holder |
| Income | Monthly Income of policy holder |
| Count\_3-6\_months\_late | No of premiums late by 3 to 6 months |
| Count\_6-12\_months\_late | No of premiums late by 6 to 12 months |
| Count\_more\_than\_12\_months\_late | No of premiums late by more than 12 months |
| application\_underwriting\_score | Underwriting Score of the applicant at the time of application (No applications under the score of 90 are insured) |
| no\_of\_premiums\_paid | Total premiums paid on time till now |
| sourcing\_channel | Sourcing channel for application |
| residence\_area\_type | Area type of Residence (Urban/Rural) |
| premium | Monthly premium amount |
| renewal | Policy Renewed? (0 - not renewed, 1 - renewed |

test.csv

Additionally test file contains premium which is required for the optimizing the incentives for each policy in the test set.

|  |  |
| --- | --- |
| **Variable** | **Definition** |
| id | Unique ID of the policy |
| perc\_premium\_paid\_by\_cash\_credit | Percentage of premium amount paid by cash or credit card |
| age\_in\_days | Age in days of policy holder |
| Income | Monthly Income of policy holder |
| Count\_3-6\_months\_late | No of premiums late by 3 to 6 months |
| Count\_6-12\_months\_late | No  of premiums late by 6 to 12 months |
| Count\_more\_than\_12\_months\_late | No of premiums late by more than 12 months |
| application\_underwriting\_score | Underwriting Score of the applicant at the time of application (No applications under the score of 90 are insured) |
| no\_of\_premiums\_paid | Total premiums paid on time till now |
| sourcing\_channel | Sourcing channel for application |
| residence\_area\_type | Area type of Residence (Urban/Rural) |
| premium | Monthly premium amount |