A brief on the approach, which you used to solve the problem.

Created a datasets with around 124 features, built different models and took ensemble mean of predictions from all the models as the final submission. Having reliable CV strategy was the key to finishing top in the competition.

Which data-preprocessing / feature engineering ideas really worked? How did you discover them?

- As the problem statement was to predict whether the user with redeem the coupon or not depending on historical transaction, features created from them(historical transactions) turned out to be really important.
- The important key here was to also subset the transactions to avoid leakage. So, for each row only the transactions before the startDate was used to create features for the same.
- Aggregation features was created for the following groups. The idea was to create valuable features at different level and angle that could lead to customers redeeming the coupon.
  - Customer and StartDate Group
  - Coupon and StartDate Group
  - Customer, Coupon and StartDate Group
  - Daily level aggregation for Customer and Coupons was also used.
  - Brands (Extracted from the Coupon) and StartDate Group
- More other levels of aggregated features was also tried with the same intuition but was not useful so dropped in the final dataset.
- Creating features without any leakage was key to finishing top in the competition.

How does your final model look like? How did you reach it?

- Datasets: D1 with around 124 features
- Models : LightGBM , XGBoost and Catboost
- Mean ensemble of the predictions from all the models.

What are the key takeaways from the challenge, if any?

- Creating valuable features was the key to finishing top in the competition.
- Avoiding leakage in the feature when using historical data was important for models to not overfit.

According to you, what are the 5 things a participant must focus on while solving such problems?

- Understanding the problem statement and the dataset.
- Spend some time on EDA and get a sense of features you are working with by studying the feature importance using any baseline model on initial dataset. This will help in understanding the data and creating powerful features.
- Leakage analysis prior to building features using historical data.
- Build a reliable CV strategy.
- And again always keep trying, keep thinking.