BEYOND ANALYSIS

COR-241

Gaurav Kumar
Sumanth Balabhadruni
Yoga Ramachandran

DATASET OVERVIEW

Online Skill Based Gaming Data

1. To estimate customer value and extrapolate the existing value into future
2. 96298 customers and 22 features
3. 2 target variables viz- Customer Value & Temporal valuation
4. Root Mean Square Error (RMSE) - Evaluation metrics

STEPS INVOLVED TO SOLVE THE PROBLEM

1. Dataset Distribution
2. Exploratory Data analysis
3. Feature Engineering

4.Data Preprocessing
5.Modelling and Metrics
6. Feature Importance & Insights

DATASET DISTRIBUTION

Includes the information of all the players competing for virtual money

Practice Game

Includes information of deposit and withdrawal transactions

Transaction

Includes the information of all the players competing for real money

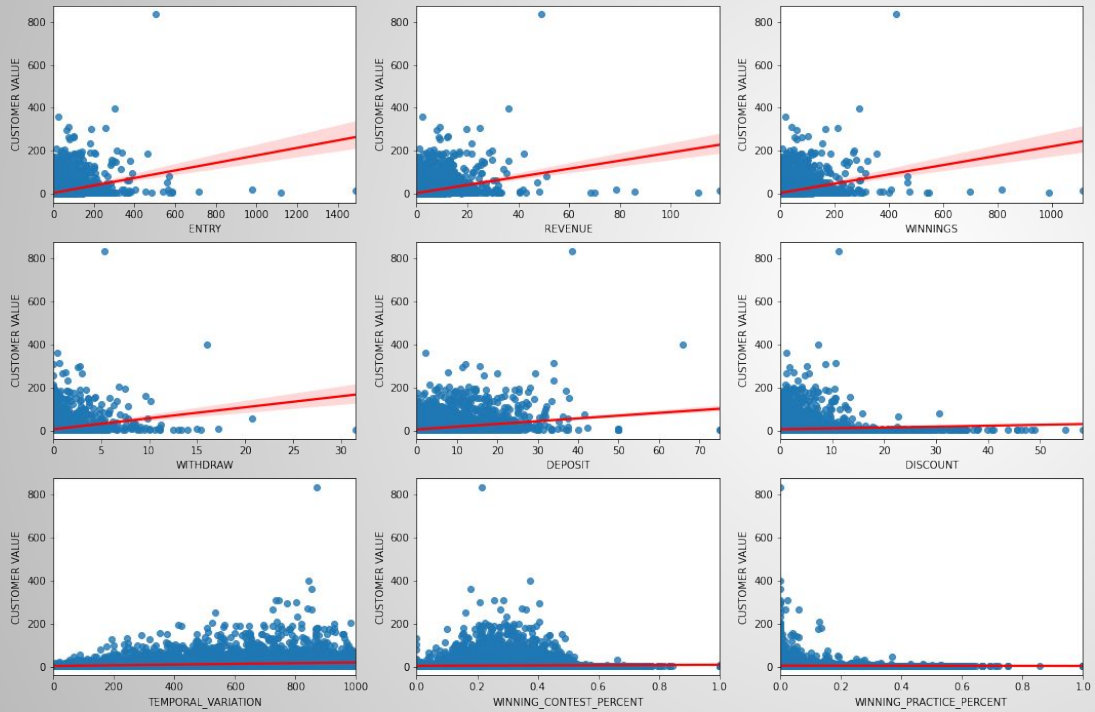
Real Game

Includes information of revenue to the organisation

Revenue

EXPLORATORATORY DATA ANALYSIS

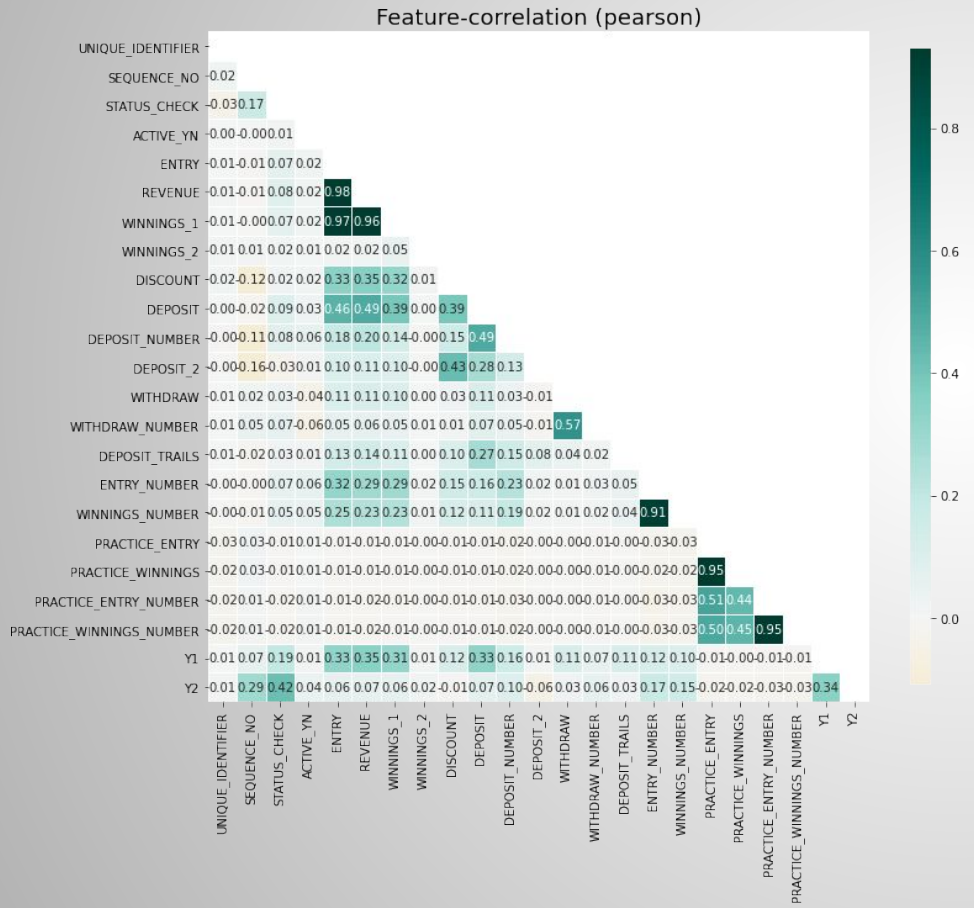
❑ REGRESSION PLOTS



- The variables entry, revenue, withdrawal and winning show effective positive relationship w.r.t customer value
- Discount, deposit and temporal variation show slight positive relationship w.r.t customer value
- The percentage of winning contests and winning practice show minimal positive relationship w.r.t customer value

EXPLORATORATORY DATA ANALYSIS

❑ CORRELATION MAP



Key Observations:

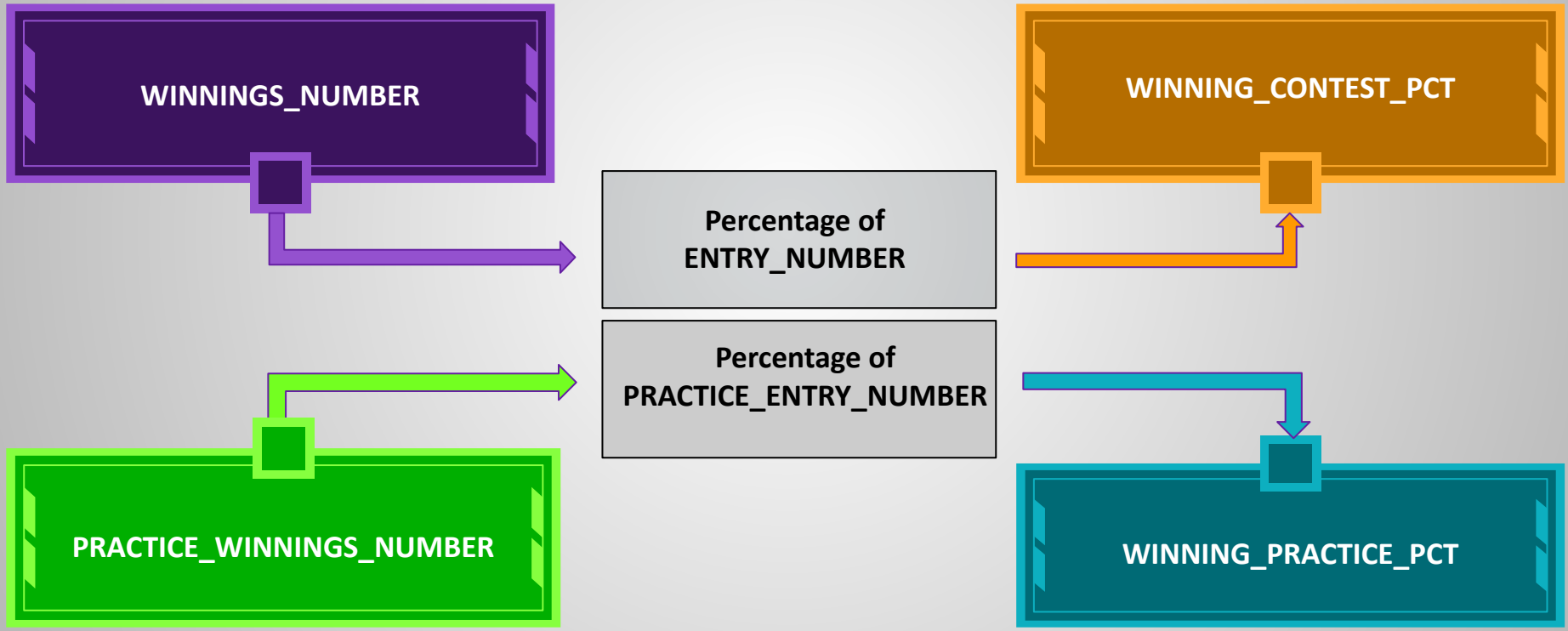
- Winnings number and entry number are highly correlated (r = 0.91)
- Practice winnings number and practice entry number are highly correlated (r = 0.95)
- Customer Value (Y1) and temporal variation (Y2) are correlated(r=0.34)

Conclusion:

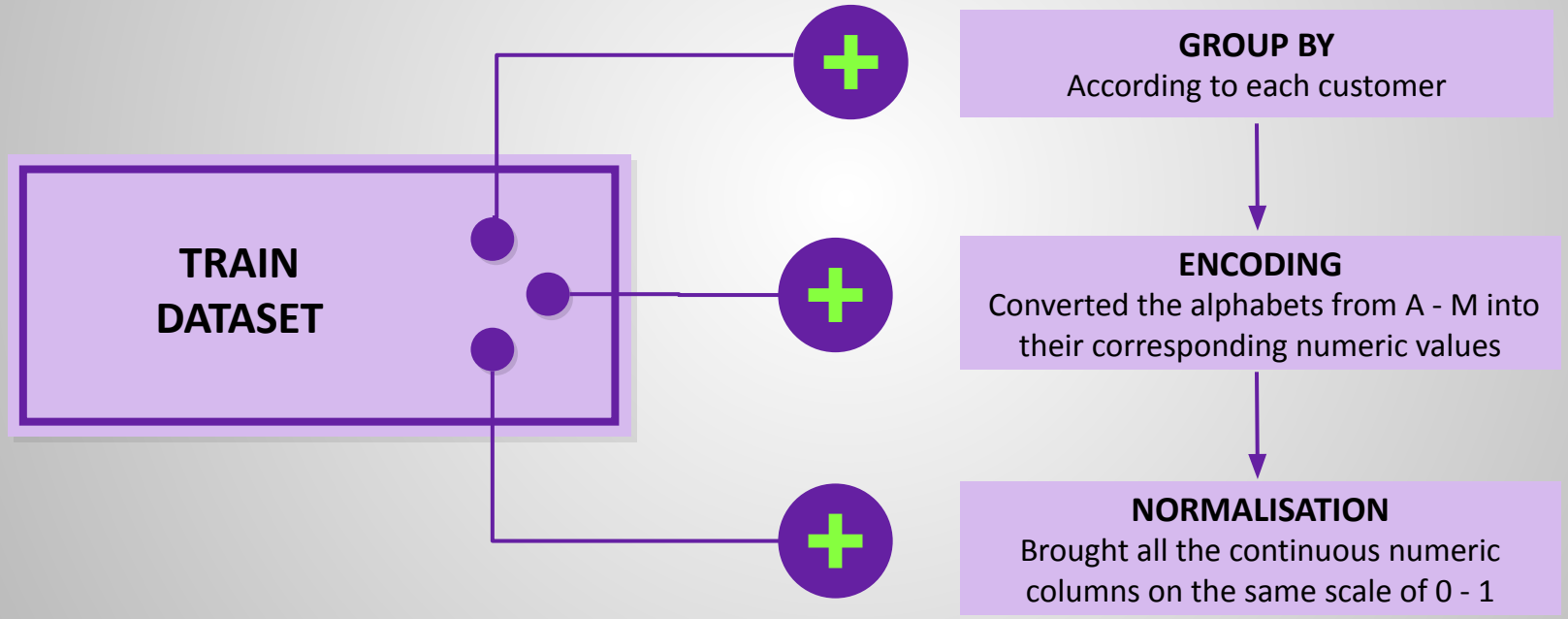
- The above two high correlations have been dealt with further in the feature engineering
- Y2 can be used to predict Y1

Note: r - Pearson Correlation Coefficient

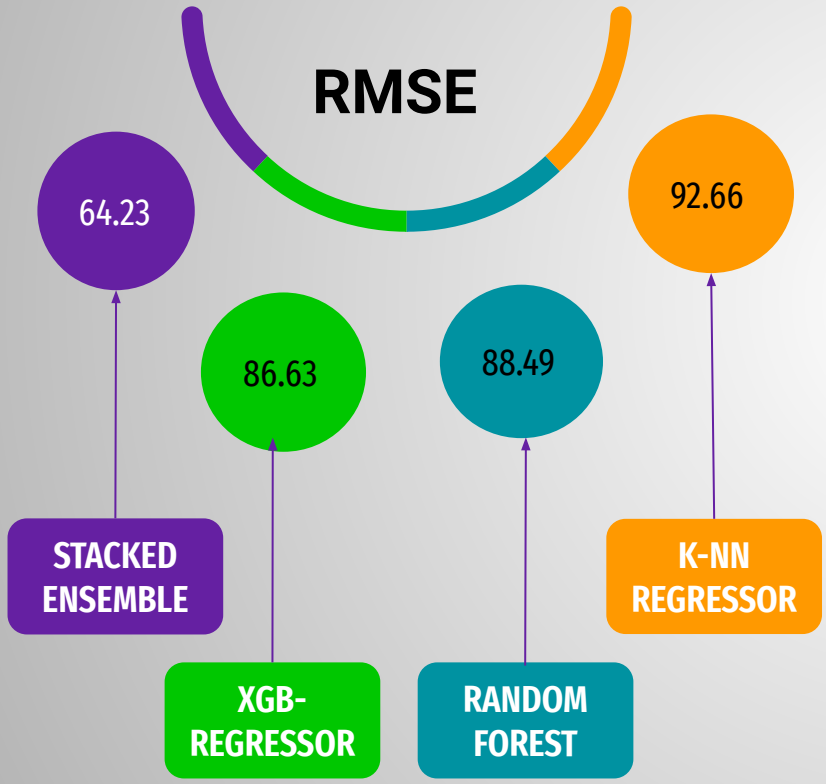
FEATURE ENGINEERING



DATA PREPROCESSING



MODELLING AND METRICS



- **Regressor Chain → Predict Y2 → Predict Y1 (with Y2 as a feature)**
- **Hyperparameter Tuning for each model using Grid Search CV**
- **Stacked model - To take the weighted average of multiple models -** KNN, Decision Tree, Random Forest and XGB Regressor
- **Mean aggregate of test data predictions for each of the customer → submission csv**

FEATURE IMPORTANCE & INSIGHTS

- Provide more offers/discounts for customers making higher deposits and frequent entries
- Attract customers having good practice winnings record by giving introductory free games
- Reach out to inactive customers having sufficient deposits or winnings

LEARNINGS & CHALLENGES

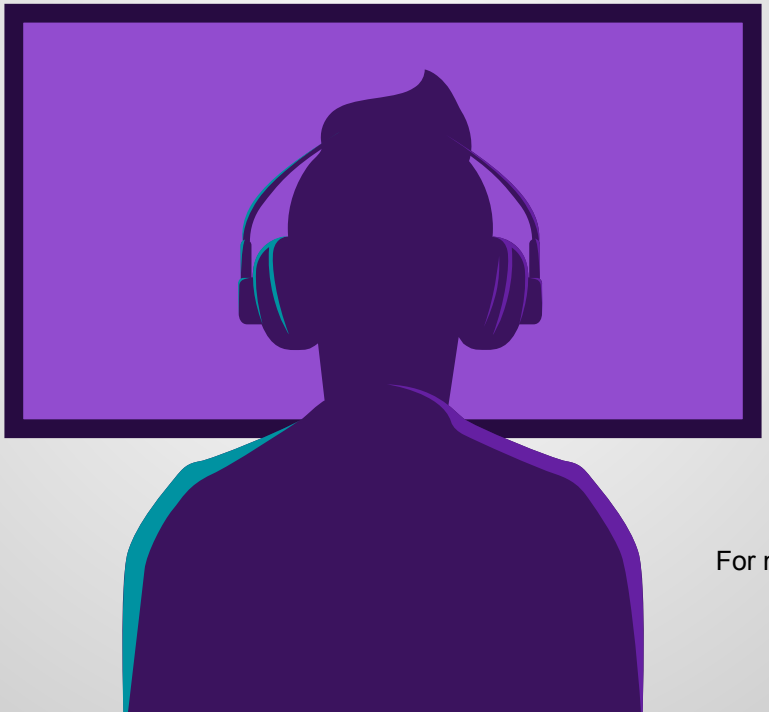
Challenges

- Absence of variation in target variables across each customer sequence
- Lack of demographic information like age of customers, recency of their action, skill level, income segment, user preferences
- Lack of KPI measures to segment the customers into distinct buckets like monthly active customers, daily active users,

Learnings

- When there is an influence between target variables, regressor chain and multi-output can be preferred models
- Stacking helps to consolidate multiple weaker models through weighted average method

THANK YOU



For more information about editing slides, please read our FAQs or visit Slidesgo School:

https://slidesgo.com/faqs and
https://slidesgo.com/slidesgo-school