|  |
| --- |
| Problem Statement |
| Food & Beverages Spend Prediction in Club Mahindra Resorts |
| Club Mahindra (Club M) makes significant revenue from Food and Beverages (F&B) sales in their resorts. The members of Club M are offered a wide variety of items in either buffet or À la carte form. Following are some benefits that the model to predict the spend by a member in their next visit to a resort will bring: |
| 1. Predicting the F&B spend of a member in a resort would help in improving the pre-sales during resort booking through web and mobile app |
| 2. Targeted campaigns to suit the member taste and preference of F&B |
| 3. Providing members in the resort with a customized experience and offers |
| 4. Help resort kitchen to plan the inventory and food quantity to be prepared in advance |
| Given the information related to resort, club member, reservation etc. the task is to predict average spend per room night on food and beverages for the each reservation in the test set. |
|  |
| Data Description |
| train.zip |
| train.zip contains train.csv and data\_dictionary.csv. |
| train.csv contains the training data with details on a set of reservations with the average spend per room night |
| Data\_Dictionary.xlsx contains a brief description of each variable provided in the training and test set. |
|  |
| test.csv |
| test.csv contains details of all reservations for which the participants need to predict the average spend on FnB per room night |
|  |
| sample\_submission.csv |
| sample\_submission.csv contains the submission format for the predictions against the test set. A single csv/zip needs to be submitted as a solution. |
|  |
| Evaluation Metric |
| Submissions are evaluated on 100 \* Root Mean Squared Error (RMSE) on the variable amount\_spent\_per\_room\_night\_scaled |
| Club Mahindra Baseline |
| To qualify for the Prizes, the RMSE of your submitted predictions on the private set or the private leaderboard score must beat the baseline RMSE set by Club Mahindra = 97 |
| In case nobody beats the baseline, the following prizes will be distributed: |
| 1st Prize: HP Pavilion Gaming Core i5 8th Gen 15.6-inch FHD Laptop |
| 2nd Prize: Apple Watch Series 3 |
| 3rd Prize: Apple Watch Series 3 |
|  |
| Public and Private Split |
| Test data is further randomly divided into Public (30%) and Private (70%) data. |
| Your initial responses will be checked and scored on the Public data. |
| The final rankings would be based on your private score which will be published once the competition is over. |
|  |
| Code & Approach Submission |
| Setting the final submission is mandatory. Without a final submission, the submission corresponding to best public score will be taken as the final submission |
| Code file and approach document are mandatory while setting final submission. For GUI based tools, please upload a zip file of snapshots of steps taken |
| Participants must adhere to the approach document format provided at this link. |
|  |
| Hackathon Rules |
| 1. Use of external datasets is not allowed |
| 2. You can only make 15 submissions per day |
| 3. The code file uploaded should be pertaining to your final submission. |
| 4. No submission will be accepted after the contest deadline. |

Challenges/Approach:

Exploratory Data Analysis to understand

* Target data distribution of Train and Test
* Descriptive stats on Memberid
* Memberid distribution of Train and Test; found that member ids in train and test are different and there is no overlap
* Missing value and datatype analysis of all the raw features
* Frequency distribution of all the categorical variables between Train and Test

Creating various Intuitive Features on the numeric data, Descriptive stats based features, Lag features

Intuitive Data sorting improved the score

Removing outliers/data clipping, Data Imputation

Cross validation strategy, used memberid based kfold

Train and Test data have different set of members ids, cv is defined in such a way Dev/Val data have different set of memberid to mimic train/test

Due to time constraint, haven’t concentrated on model tuning and ensembling

Private LB top score: 95.8128335535

My private LB score: 96.1575219761 (Rank 10)

My public LB score: 95.1948501887387