* EDA :
  1. Pandas profiling
  2. Seaborn pair plot and correlation heat map
* Findings
  1. is\_timer- always 0
  2. click rate – highly skewed target values
  3. categorical variables – with integer and text values
* pre processing :
  1. times\_of\_day – encoding with pandas get\_dummies
  2. power transformer
  3. np.log transform target
  4. StandardScalar transform
* Modelling:

|  |  |  |  |
| --- | --- | --- | --- |
| No | model | R2 train | R2 test |
| 1 | XGBRegressor | 0.99 | 0.5 |
| 2 | RandomForestRegressor | 0.93 | 0.56 |
| 3 | LinearRegression | 0.31 | 0.29 |
| 4 | (model1 + model2 + model3 ) /3 | 0.85 | 0.46 |
| 5 | .15model1 + .2model2 + .65model3 | 0.6 | 0.46 |