

Q1 :- Choose the most suitable evaluation metric and state why you chose it.

The evaluation metric for this competition is Root Mean Squared Logarithmic Error.

The RMSLE is calculated as

$$\epsilon = \sqrt{\frac{1}{n} \sum_{i=1}^n (\log(p_i + 1) - \log(a_i + 1))^2}$$

Where:

ϵ is the RMSLE value (score)

n is the total number of observations in the (public/private) data set,

p_i is your prediction of trip duration, and

a_i is the actual trip duration for i .

$\log(x)$ is the natural logarithm of x .

As Our Data Have the Listed Below Property :-

- targets having exponential growth, such as population counts, average sales of a commodity over a span of years etc
- we care about **percentage errors** rather than the **absolute value of errors**.
- there is a wide range in the target variables and
- we **don't want to penalize big differences** when **both the predicted and the actual are big numbers**.
- we want to penalize **under estimates** more than **over estimates**.

Note :- Questions Are Properly Solved in the .ipynb Files Or Jupiter Notebook file and Save as pdf via github View of the respective file Due to nbconvert error on Jupyter .