Week 9 submission
Project: Bank Marketing Campaign

Group Name: Evolve Data
Name: Dmitry Sharukhin
Email: sharuhinda@gmail.com
Country: Russia
College/Company: Finval GC
Specialization: Data Science

Problem description
ABC Bank wants to sell its term deposit product to customers and before launching the product they want to develop a model which help them in understanding whether a particular customer will buy their product or not (based on customer's past interaction with bank or other Financial Institution).
Bank wants to use ML model to shortlist customer whose chances of buying the product are more so that their marketing channel (tele marketing, SMS/email marketing etc.) can focus only to those customers.
This will save resource and their time (which is directly involved in the cost (resource billing)).

The task is to create binary classifier to forecast the probability of customer's agreement to open term deposit

Github repo link: https://github.com/sharuhinda/bank_marketing_campaign/tree/review

EDA notebook:
https://github.com/sharuhinda/bank_marketing_campaign/blob/review/2_eda.ipynb

Reviews thread: https://github.com/sharuhinda/bank_marketing_campaign/pull/2

The Weight of Evidence was chosen as a primary approach to evaluate and clean dataset. Considering the dataset is highly imbalanced it allowed not to drop but retain missing-like data and encode it based on positive/negative outcome ratio.

For the purpose to apply this approach correctly 2 additional methods were applied:
- binning was performed on features where the possible number of outcomes (values) was too large ("pdays" feature)
- outliers cutoff for the numerical features with long right tales ("campaign" and "previous" features)

Finally simple imputations were made to "marital", "job" and "education" features using most_frequent strategy.