ZS

2017 Young Data Scientist Challenge
Submission Guide

SALES + MARKETING

Impact where it matters.

YDS 2017 Submission Details

The final submission should consist of 3 parts:

1) Section A – Approach document (Submission file format : PPT/XLSX/DOC)
2) Section B – Supporting codes for final submission (Submission file format: Any programming language)
3) *Section C – Model prediction (Submission file format: *.csv)*

[See Submission Guidelines for more information]

Any submission not having all 3 components will be considered as incomplete submission

Submitting "Section A"

In Section A, please provide your findings and justification on the below topics -

1) *A 250-word write-up outlining your approach towards the problem statement*

2) *Quality checks performed / Errors found*: Please report any potential errors / inaccuracies in the training dataset

3) *Data preprocessing steps:* Please provide potential data aggregations / transformations performed including how the quality checks from section 1 are dealt with

4) *Feature generation*: Please provide list of features that were created and share the how and why the feature was created

5) **Key observations / Trends:** Please provide any information on your key observations / insights / trends that you may gleaned from the datasets

6) **Model choice explanation:** Please justify the usage of a particular model for the dataset provided. This should ideally include the reasoning and comparison justifying the choice

7) **Expected error for submission**: please provide the expected error for submission file in Section C

8) **Top 5 most significant variables in model:** please provide the most significant feature in the dataset and how were they determined

Please submit your thoughts in word/PPT/Excel on the portal

The above list is for directional purposes. Please include other findings identified while solving case

Submitting "Section B"

- Please submit the **source code for the training and prediction**

- Upload a *zip* file (max allowed size is *2MB*)

- The submitted file must have a *README* file with a detailed description about how to run the model to predict the dependent variable and generate the ***<userID>.csv*** file.

- There is no limit on execution time, but the code should generate the output file***:<UserID>.csv***

- Please ensure that the submitted code files are well-commented and structured for easy comprehension

Submitting "Section C"

- Create new submission file containing "ID" from test set and "Premium" column for each ID (see below table for details)

- Predictions for the submission dataset are to be submitted via "My Submission" page on the ZS YDS web platform

"submission_format.csv" File Structure

PID	Event1	Event2	Event3	Event4	Event5	Event6	Event7	Event8	Event9	Event10
1028890										
1021443										
1016194										
1017298										
1006098										
1010331										
1022799										
1031711										
1000151										
1010910										
1005009										
1017379										

Update highlighted columns with your prediction

Note:

In case you face any error while submitting the model prediction for evaluation even after following the submission format, share the following with your query:

- Snapshot of the error faced
- Submission csv file

Evaluation Engine consider only .csv file in the abovementioned format as a valid submission.
In case of any submission related errors, please reach out to the support team at
YDS_ZS_Support@zsassociates.com