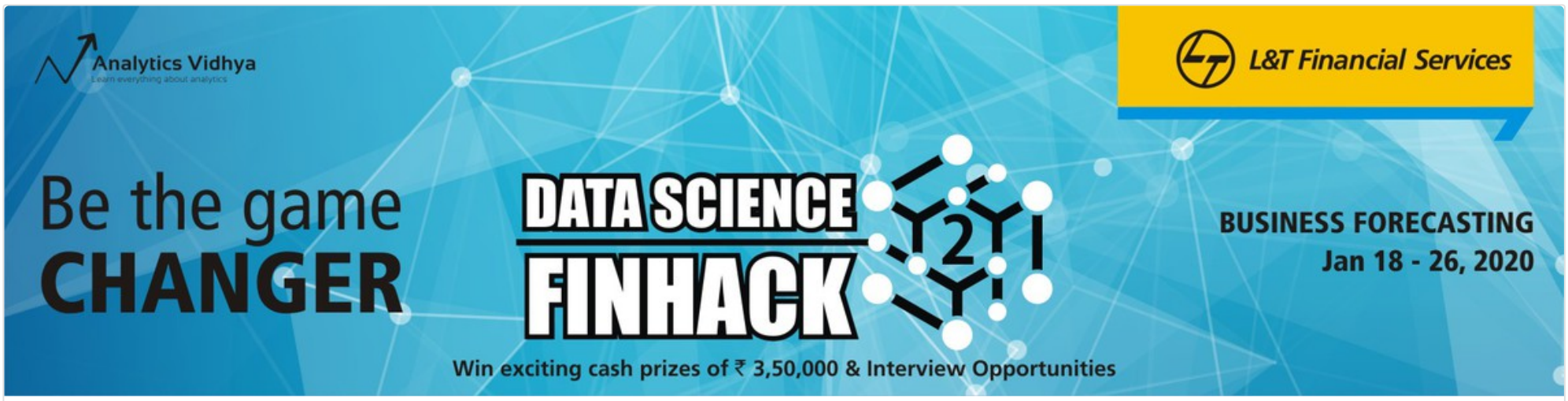


LTFS Data Science FinHack 2

Online 18-01-2020 12:00 AM to 26-01-2020 11:59 PM

5882
Registered

592
Number of Teams

Cash prizes worth INR 3.5 lacs & Interview opportunities with LTFS
Prizes

ENDS IN

4 DAYS 4 HOURS 52 MIN

Registered

About | Problem Statement | Solution Checker | My Submissions | Leaderboard | Team

LTFS receives a lot of requests for its various finance offerings that include housing loan, two-wheeler loan, real estate financing and micro loans. The number of applications received is something that varies a lot with season. Going through these applications is a manual process and is tedious. Accurately forecasting the number of cases received can help with resource and manpower management resulting into quick response on applications and more efficient processing.

You have been appointed with the task of forecasting daily cases for **next 3 months for 2 different business segments** aggregated at the **country level** keeping in consideration the following major Indian festivals (inclusive but not exhaustive list): Diwali, Dussehra, Ganesh Chaturthi, Navratri, Holi etc. (You are free to use any publicly available open source external datasets). Some other examples could be:

- Weather
- Macroeconomic variables

Note that the external dataset must belong to a reliable source.

Data Dictionary

The train data has been provided in the following way:

- For business segment 1, historical data has been made available at branch ID level
- For business segment 2, historical data has been made available at State level.

Train File

Variable	Definition
application_date	Date of application
segment	Business Segment (1/2)
branch_id	Anonymised id for branch at which application was received
state	State in which application was received (Karnataka, MP etc.)
zone	Zone of state in which application was received (Central, East etc.)

Feedback

case_count	(Target) Number of cases/applications received

Test File

Forecasting needs to be done at country level for the dates provided in test set for each segment.

Variable	Definition
id	Unique id for each sample in test set
application_date	Date of application
segment	Business Segment (1/2)

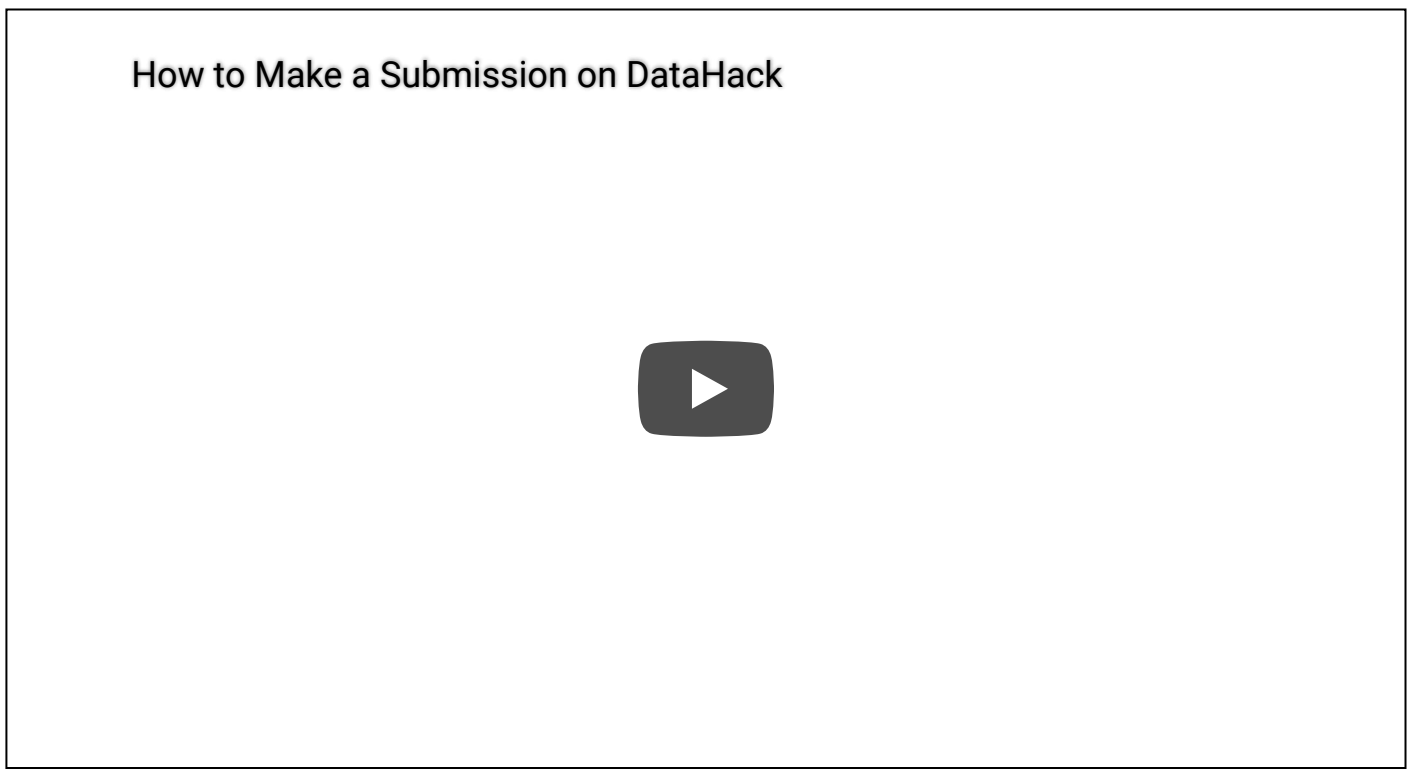
Sample Submission

This file contains the exact submission format for the forecasts. Please submit csv file only.

Variable	Definition
id	Unique id for each sample in test set
application_date	Date of application
segment	Business Segment (1/2)
case_count	(Target) Predicted values for test set

Note that you can only make **5 submissions** per day

How to Make a Submission?



Evaluation

Evaluation Metric

The evaluation metric for scoring the forecasts is **MAPE (Mean Absolute Percentage Error) M** with the formula:

$$\mathrm{M} = \frac{100\%}{n} \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right|$$

Where A_t is the actual value and F_t is the forecast value.

The Final score is calculated using MAPE for both the segments using the formula:

$$Final\ Score = 0.5 * MAPE_{Segment\ 1} + 0.5 * MAPE_{Segment\ 2}$$

Important Notes

- Note that feasibility of implementation of top solutions will be considered while adjudging winners
- The solution must produce satisfactory results for both the business segments

Public and Private Split

Test data is further divided into Public (1st Month) and Private (Next 2 months)

- 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.

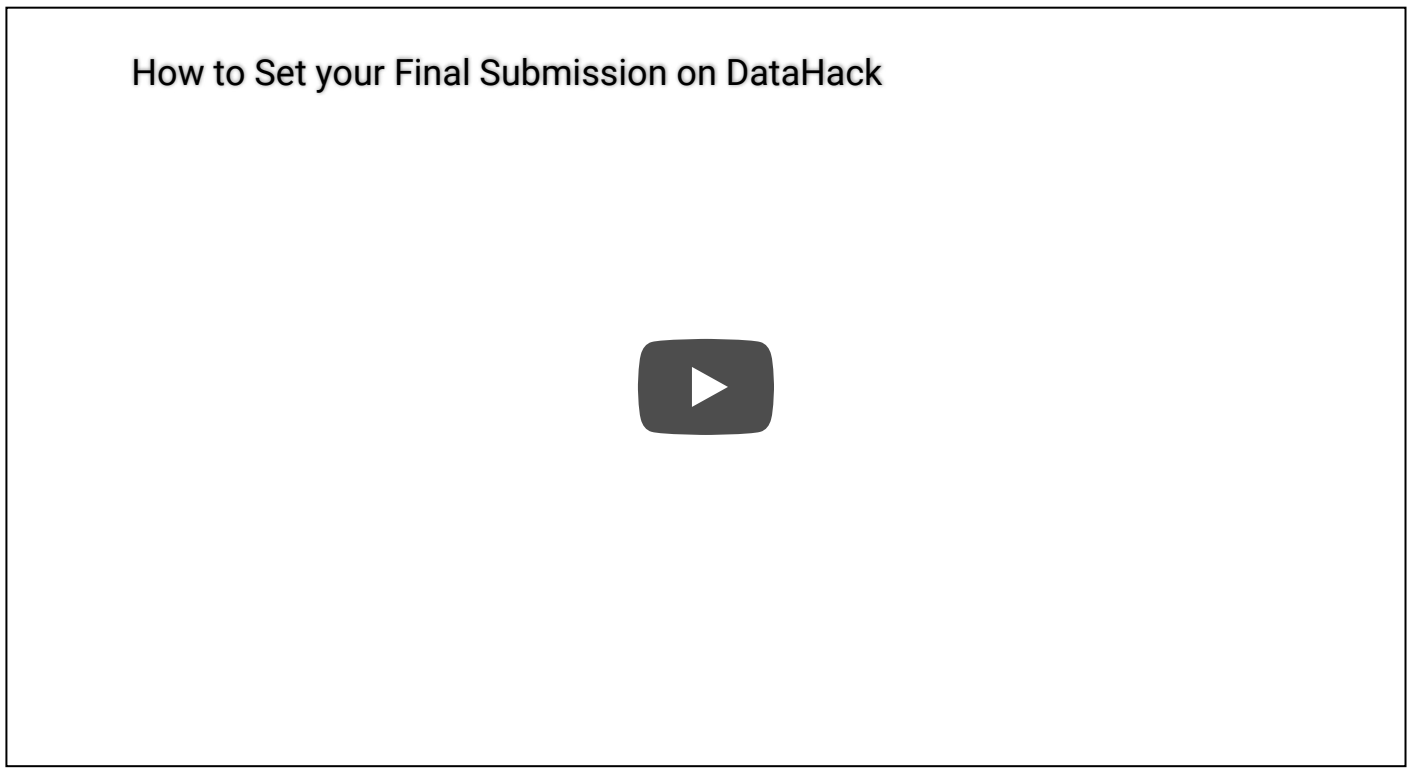
Guidelines for Final Submission

Please ensure that your final submission includes the following:

1. Solution file containing the predicted case count in the test dataset (format is given in sample submission csv)
2. Code file containing the following:

- **Code:** Note that it is mandatory to submit your code for a valid final submission
- **External Datasets:** Include any external dataset that you have used as an input variable
- **Approach:** Please share your approach to solve the problem (doc/ppt/pdf format). It should cover the following topics:
    1. A brief on the approach, which you have used to solve the problem. Mention the steps used for each segment separately.
    2. What data-preprocessing / feature engineering ideas really worked? How did you discover them?
    3. What does your final model look like? How did you reach it?

How to Set Final Submission?



Hackathon Rules

1. *The final standings would be based on private leaderboard score and presentations made in Online Interview round with LTFS & Analytics Vidhya which will be held after contest close.*

2. Setting the final submission is recommended. Without a final submission, the submission corresponding to best public score will be taken as the final submission
3. You can only make **5 submissions** per day
4. Entries submitted after the contest is closed, will not be considered
5. The code file pertaining to your final submission is mandatory while setting final submission
6. Throughout the hackathon, you are expected to respect fellow hackers and act with high integrity.
7. Analytics Vidhya and LTFS hold the right to disqualify any participant at any stage of the competition if the participant(s) are deemed to be acting fraudulently.
8. Use of multiple IDs will lead to immediate disqualification

Data

Test File | Train File | Sample Submissions