



(/teams/black-friday/create)



(/contest/black-friday/lb)



Registered

Starts at	Tue Jul 19 2016 00:00:00 GMT+0530 (India Standard Time)
Closes on	Sat Dec 31 2016 23:59:59 GMT+0530 (India Standard Time)
Mode	Online
# Participants	1586
Prizes	Knowledge and Learning

Rules

- The hackathon starts at 00:00 (UTC + 5:30) on 19th Jul, 2016 and closes at 23:59 on 31st Dec, 2016 (UTC + 5:30)
- One person cannot participate with more than one user accounts.
- You are free to use any tool and machine you have rightful access to.
- You can use any programming language or statistical software.
- You are free to use solution checker as many times as you want.

Resources

- Data Exploration is the core of machine learning competition. To understand more about it, click here (<http://www.analyticsvidhya.com/blog/2016/01/guide-data-exploration/>).
- You can refer introduction course to participate in machine learning competition on DataCamp (<https://www.datacamp.com/community/open-courses/introduction-to-python-machine-learning-with-analytics-vidhya-hackathons>) designed by Analytics Vidhya.



- You can participate in workshop "Experiments with Data (<http://datahack.analyticsvidhya.com/contest/experiments-with-data-4>)" to start your data science journey using Excel, R or Python



Problem Statement

A retail company "ABC Private Limited" wants to understand the customer purchase behaviour (specifically, purchase amount) against various products of different categories. They have shared purchase summary of various customers for selected high volume products from last month.

The data set also contains customer demographics (age, gender, marital status, city_type, stay_in_current_city), product details (product_id and product category) and Total purchase_amount from last month.

Now, they want to build a model to predict the purchase amount of customer against various products which will help them to create personalized offer for customers against different products.

Data

Variable	Definition
User_ID	User ID
Product_ID	Product ID
Gender	Sex of User
Age	Age in bins
Occupation	Occupation (Masked)
City_Category	"Category of the City (A,B,C)"
Stay_In_Current_City_Years	Number of years stay in current city
Marital_Status	Marital Status
Product_Category_1	Product Category (Masked)
Product_Category_2	Product may belongs to other category also (Masked)
Product_Category_3	Product may belongs to other category also (Masked)
Purchase	Purchase Amount (Target Variable)



Your model performance will be evaluated on the basis of your prediction of the purchase amount for the test data (test.csv), which contains similar data-points as train except for their purchase amount. Your submission needs to be in the format as shown in "SampleSubmission.csv".



We at our end, have the actual purchase amount for the test dataset, against which your predictions will be evaluated. Submissions are scored on the root mean squared error (RMSE). RMSE is very common and is a suitable general-purpose error metric. Compared to the Mean Absolute Error, RMSE punishes large errors:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2},$$

Where \hat{y} is the predicted value and y is the original value.

Please note :

- Public leaderboard is based on 30% of the test dataset, while 70% of the dataset is used for Private Leaderboard.
- The final results would be declared only on Private Leaderboard

📁 Test File (/contest/black-friday/media/test_file/test-comb.csv)

📁 Train File (/contest/black-friday/media/profile_photos/train-corrected.zip)

📁 Sample Submissions (/contest/black-friday/media/sample_submission/Sample_Submission_Tm9Lura.csv)

Solution Checker

Code File

No file chosen

Solution File*

(.csv or .zip file)

No file chosen

Solution Description

(max : 180 chars)*

Solution Description (for your interest)

Add Solution





[Join Slack Live Chat](#)

ABOUT US

For those of you, who are wondering what is “Analytics Vidhya”, “Analytics” can be defined as the science of extracting insights from raw data. The spectrum of analytics starts from capturing data and evolves into using insights / trends from this data to make informed decisions.

Latest Posts



3 Compelling reasons why you must compete in Data Hackathon 3.x

KUNAL JAIN

(<https://www.analyticsvidhya.com/blog/2015/09/compelling-reasons-compete-data-hackathon/>)



Learn to use Forward Selection Techniques for Ensemble Modeling

TAVISH SRIVASTAVA

(<https://www.analyticsvidhya.com/blog/2015/09/selection-techniques-ensemble-modelling/>)

Quick Links

[Home \(https://www.analyticsvidhya.com\)](https://www.analyticsvidhya.com)

[About Us \(https://www.analyticsvidhya.com/about-me/\)](https://www.analyticsvidhya.com/about-me/)

[Our Team \(https://www.analyticsvidhya.com/about-me/team/\)](https://www.analyticsvidhya.com/about-me/team/)

[Privacy Policy \(https://www.analyticsvidhya.com/privacy-policy/\)](https://www.analyticsvidhya.com/privacy-policy/)

[Refund Policy \(https://www.analyticsvidhya.com/refund-policy/\)](https://www.analyticsvidhya.com/refund-policy/)

[Terms of Service \(https://www.analyticsvidhya.com/terms/\)](https://www.analyticsvidhya.com/terms/)

