
Assignment - To predict amount of purchase.

Overview

The dataset here is a sample of the transactions made in a retail store. The store wants to know better the customer purchase behaviour against different products. Problem is to predict the the amount of purchase with the help of the information provided.

Data

Dataset of 550000 observations about a retail store, it contains different kinds of variables either numerical or categorical. The data set 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.

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)

Evaluation

Your model performance will be evaluated on the basis of your prediction of the purchase amount for the test data (test.csv), We, 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).

Sample format of required output

User_ID,Product_ID,Purchase

Objective

The objective is to build a script/notebook with below modules(user can have more if he wants) and send it back to us with following modules

1. Data exploration module
2. Feature manipulation / preprocessing module
3. Model building module

Disclaimer please use python as programming language.