**Black Friday: A study of sales trough consumer behaviors**

About Dataset

Dataset of 550 000 observations about the black Friday in a retail store, it contains different kinds of variables either numerical or categorical. It contains missing values.

Description

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. Specifically, here the problem is a regression problem where we are trying to predict the dependent variable (the amount of purchase) with the help of the information contained in the other variables.

Classification problem can also be settled in this dataset since several variables are categorical, and some other approaches could be "Predicting the age of the consumer" or even "Predict the category of goods bought". This dataset is also particularly convenient for clustering and maybe find different clusters of consumers within it.

Acknowledgements

The dataset comes from a competition hosted by Analytics Vidhya.

Column Description

User\_ID User ID

Product\_ID Product ID

Gender Sex of User

Age Age in bins

Occupation Occupation

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

Product\_Category\_2 Product may belongs to other category also

Product\_Category\_3 Product may belongs to other category also

Purchase Purchase amount in dollars