Competition Link: https://datahack.analyticsvidhya.com/contest/amexpert-2019-machine-learning-

You are provided with the following files in train.zip:

**train.csv:** Train data containing the coupons offered to the given customers under the 18 campaigns

|  |  |
| --- | --- |
| **Variable** | **Definition** |
| id | Unique id for coupon customer impression |
| campaign\_id | Unique id for a discount campaign |
| coupon\_id | Unique id for a discount coupon |
| customer\_id | Unique id for a customer |
| redemption\_status | (target) (0 - Coupon not redeemed, 1 - Coupon redeemed) |

**campaign\_data.csv:** Campaign information for each of the 28 campaigns

|  |  |
| --- | --- |
| **Variable** | **Definition** |
| campaign\_id | Unique id for a discount campaign |
| campaign\_type | Anonymised Campaign Type (X/Y) |
| start\_date | Campaign Start Date |
| end\_date | Campaign End Date |

**coupon\_item\_mapping.csv:** Mapping of coupon and items valid for discount under that coupon

|  |  |
| --- | --- |
| **Variable** | **Definition** |
| coupon\_id | Unique id for a discount coupon (no order) |
| item\_id | Unique id for items for which given coupon is valid (no order) |

**customer\_demographics.csv:** Customer demographic information for some customers

|  |  |
| --- | --- |
| **Variable** | **Definition** |
| customer\_id | Unique id for a customer |
| age\_range | Age range of customer family in years |
| marital\_status | Married/Single |
| rented | 0 - not rented accommodation, 1 - rented accommodation |
| family\_size | Number of family members |
| no\_of\_children | Number of children in the family |
| income\_bracket | Label Encoded Income Bracket (Higher income corresponds to higher number) |

**customer\_transaction\_data.csv:** Transaction data for all customers for duration of campaigns in the train data

|  |  |
| --- | --- |
| **Variable** | **Definition** |
| date | Date of Transaction |
| customer\_id | Unique id for a customer |
| item\_id | Unique id for item |
| quantity | quantity of item bought |
| selling\_price | Sales value of the transaction |
| other\_discount | Discount from other sources such as manufacturer coupon/loyalty card |
| coupon\_discount | Discount availed from retailer coupon |

**item\_data.csv:** Item information for each item sold by the retailer

|  |  |
| --- | --- |
| Variable | Definition |
| item\_id | Unique id for item |
| brand | Unique id for item brand |
| brand\_type | Brand Type (local/Established) |
| category | Item Category |

**test.csv:** Contains the coupon customer combination for which redemption status is to be predicted

|  |  |
| --- | --- |
| **Variable** | **Definition** |
| id | Unique id for coupon customer impression |
| campaign\_id | Unique id for a discount campaign |
| coupon\_id | Unique id for a discount coupon |
| customer\_id | Unique id for a customer |

*\*Campaign, coupon and customer data for test set is also contained in train.zip*

**sample\_submission.csv:** This file contains the format in which you have to submit your predictions.

To summarise the entire process:

* Customers receive coupons under various campaigns and may choose to redeem it.
* They can redeem the given coupon for any valid product for that coupon as per coupon item mapping within the duration between campaign start date and end date
* Next, the customer will redeem the coupon for an item at the retailer store and that will reflect in the transaction table in the column coupon\_discount.