

# Credit card churning

We define that a credit card is churned by a customer if all cards of that customer become inactive for three consecutive months. You, as a data scientist, are assigned to build a model and predict customer credit card churning signals for the next three months (Jan 2018 – Mar 2018).

You are given the list of customers, their demographics, financial history and miscellaneous data. Below are the specification of your inputs and how we evaluate the results.

**Few hints regarding user pool for this question.**

- Every user has their credit card activated by the time the data was given
- Every user has transaction with a card at least once in year 2017
- Each user was limited to a maximum of three cards per person (Selecting ones with most transactions)

## Provided Files:

- File name: **y\_train.csv**
  - o Output for selected user ids
  - o Size: 494 rows

Field Name	Data Type	Description
user_id	INT	Dummy user id
Label	INT	0–Active, 1–Inactive

- File name: **demo.csv**
  - o Personal information of customers
  - o Size: 52,762 rows

Field Name	Data Type	Description
user_id	INT	Dummy user id
account_start_date	TIMESTAMP	First account start date

birth_year	INT	Birth year
gender	STRING	F-Female, M-Male
marital_status	INT	1-Single, 2-Married, 3-Divorce/widow
family_income_segment_code	INT	Family income segment in ordinal scale (1 as lowest income segment)
individual_income_segment_code	INT	Individual income segment in ordinal scale (1 as lowest income segment)

- File name: **card\_info.csv**

- o Credit card information

- o Size: 60,296 rows

Field Name	Data Type	Description
user_id	INT	Dummy user id
bill_cyc	INT	Bill cycle day of month
cr_lmt_amt	INT	Credit limit amount
card_no	INT	Dummy credit card number

- File name: **cc\_txn.csv**

- o Credit card transaction log

- o Period: 12 months

- o Size: 3,223,075 rows

Field Name	Data Type	Description
mcc	STRING	Merchant category
txn_dt	TIMESTAMP	Transaction timestamp
user_id	INT	Dummy user id
txn_amt	FLOAT	Transaction amount
card_no	INT	Dummy credit card number

The category code descriptions are as follows:

No.	Main Category	No.	Main Category
1	Fashion and Apparel	9	Sports
2	Health and beauty	10	Children
3	Food and Beverage	11	Services
4	Appliance and Electronics	12	Education
5	Office supplies, books and gift shop	13	Pet
6	Automotive shops and Vehicles	14	Travel
7	Entertainment	15	Accommodation
8	Home and Garden	16	Others

- File name: **sa\_bal.csv**
  - o Saving account balance aggregated by months
  - o Period: 12 months
  - o Size: 633,144 rows

Field Name	Data Type	Description
user_id	INT	Dummy user id
mm	INT	Month of year
max_sa_bal	FLOAT	Maximum saving account balance in the month

- File name: **dtxn.csv**
  - o Incoming and outgoing transactions aggregated by months (exclude credit card transactions)

- o Period: 12 months
- o Size: 490,599 rows

Field Name	Data Type	Description
user_id	INT	Dummy user id
mm	INT	Month of year
amt_in	FLOAT	Amount of money inbound
amt_out	FLOAT	Amount of money outbound

## Evaluation

The problem focuses on anticipating the customers who are likely to stop using our creditcards. The business team will handle the rest to maintain those customers. Although precision is also important, the evaluation optimizes these two factors with higher emphasis on recall. Hence, we use F-2 score of label 1:

$$F_2 = 5 * \frac{precision * recall}{(4 * precision) + recall}$$

## Objective

Let you present data exploration and preparation Churn Prediction Challenge and build a predictive model to classify users between 2 classes: 0–Active and 1–Inactive based their activity and payment history along all their lifetime on service. And evaluate by using F-2 score of label 1 of the test set (use y\_train.csv split data 70/30).

Remark: Some of the data provided maybe incomplete, inconsistent, missing, noisy, erroneous, etc. as can occur in the real-world setting. It is the participants' task to recognize such cases as the challenge intentionally posed by the problem designer.