Vintage Analysis

Data Science Assignment - I

06.03.2019 - 08.03.2019

—

Pramod Yadav

Indian Institute of Information Technology , Sricity

Problem Statement

1 . Create a vintage curve of 30+ DPD (Days Past Due) from the following datafiles -
contract data and payment history .

Milestones

1. Group the loan data according to its origination point i.e. the 'Contract Date' and
    label each as a Cohort group .

2. For each Cohort Group plot and analyse the sum of 'Due Amount' per period i.e.
    for each emi month of the loan in the following cases -
    a) The whole loan data and payment history .
    b) Consider loans which have 'Days Past Due' greater than 30 (30+ DPD).

Assumptions

Following were the assumptions and some roadblocks encountered during solving the
assignment -

1. Due to lack of in-depth knowledge of Financial terms and practices , a little or
    feature engineering has been done on the data .
2. There were very few missing data . Hence , the treatment of the missing values
    were left to Pandas.
3. After referring to numerous examples on Internet related to loan which compute
    a sort of loss function in Vintage Analysis , the function used was sum total of due
    amount as Principal for each loan was not available .
4. The Vintages were month wise and were based on the the Contract_date - the
    date on which loan was booked .

Procedure

I. Data Preparation

The data from 'contract.csv' and 'payment_history.csv' were 'inner joined' together on the basis of common 'loan_id' to create a central dataset consisting of each loan id and it's payment history each month since the date of booking.

II. Aggregation

The data was aggregated on the basis of Vintages (the contract date) and the in further calculation of total loans to emi per month and to calculate total due amount from various loans each month , the data was further grouped on monthly basis for each cohort and thereby after referred as Cohort Period .

Following sections shows a glimpse of Cohorts and respective calculations on them.

III. Tables

The following shed light on the results obtained from aggregation of the data-

1 . No of loans per cohort group per period - (in case -I and Case - II)

```
CohortGroup                    CohortGroup
2016-09     1                  2016-09      3
2016-10     1                  2016-10     11
2016-11     8                  2016-11     15
2016-12    10                  2016-12     33
2017-01     7                  2017-01     19
2017-02    14                  2017-02     41
2017-03    21                  2017-03     59
2017-04     9                  2017-04     46
2017-05    17                  2017-05     69
2017-06    14                  2017-06     60
2017-07    16                  2017-07     57
2017-08    12                  2017-08     65
2017-09     8                  2017-09     40
2017-10    16                  2017-10     63
2017-11    10                  2017-11     66
2017-12    30                  2017-12    131
2018-01    10                  2018-01    102
2018-02    11                  2018-02    173
2018-03    10                  2018-03    191
2018-04     1                  2018-04    161
                               2018-05     79
Name: no_of_loans, dtype: int64   Name: no_of_loans, dtype: int64
```

Figure - Image on Left - Cohorts for solving 30+dpd , Image on right - All Cohorts.

2 . No of loans per cohort per period (for all the loans)

CohortGroup CohortPeriod	2016-09	2016-10	2016-11	2016-12	2017-01	2017-02	2017-03	2017-04	2017-05	2017-06	...	2017-08	2017-09	2017-10	2017-11	2017-12	2018-01	2018-02	2018-03	2018-04	2018-05
1	3.0	11.0	15.0	33.0	19.0	41.0	59.0	46.0	69.0	60.0	...	65.0	40.0	63.0	66.0	131.0	102.0	173.0	191.0	161.0	79.0
2	2.0	10.0	15.0	32.0	19.0	41.0	59.0	46.0	66.0	60.0	...	65.0	40.0	62.0	66.0	130.0	101.0	173.0	189.0	161.0	79.0
3	2.0	10.0	15.0	32.0	19.0	41.0	59.0	46.0	65.0	60.0	...	63.0	39.0	62.0	66.0	130.0	101.0	173.0	188.0	160.0	NaN
4	2.0	10.0	15.0	31.0	19.0	41.0	59.0	46.0	65.0	60.0	...	63.0	39.0	61.0	66.0	130.0	98.0	172.0	186.0	NaN	NaN
5	2.0	10.0	15.0	31.0	19.0	39.0	57.0	46.0	62.0	60.0	...	62.0	39.0	61.0	65.0	129.0	98.0	169.0	NaN	NaN	NaN
6	2.0	10.0	15.0	30.0	18.0	39.0	57.0	44.0	62.0	58.0	...	62.0	38.0	61.0	64.0	128.0	97.0	NaN	NaN	NaN	NaN
7	2.0	9.0	15.0	29.0	18.0	38.0	57.0	43.0	62.0	58.0	...	61.0	38.0	59.0	64.0	128.0	NaN	NaN	NaN	NaN	NaN
8	2.0	9.0	15.0	29.0	18.0	36.0	57.0	42.0	62.0	58.0	...	61.0	38.0	58.0	64.0	NaN	NaN	NaN	NaN	NaN	NaN
9	2.0	9.0	15.0	29.0	18.0	35.0	57.0	42.0	62.0	58.0	...	59.0	38.0	58.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
10	2.0	8.0	15.0	28.0	17.0	35.0	56.0	41.0	62.0	56.0	...	59.0	36.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
11	2.0	8.0	15.0	27.0	17.0	35.0	55.0	41.0	60.0	55.0	...	58.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
12	2.0	8.0	15.0	27.0	17.0	35.0	55.0	40.0	58.0	54.0	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
13	2.0	8.0	15.0	26.0	17.0	35.0	55.0	39.0	53.0	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
14	2.0	8.0	15.0	26.0	17.0	35.0	55.0	38.0	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
15	2.0	8.0	15.0	26.0	17.0	35.0	55.0	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
16	2.0	8.0	15.0	26.0	17.0	35.0	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
17	2.0	8.0	15.0	26.0	17.0	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

3 .Total due amount per cohort per period (for all the loans):

CohortGroup CohortPeriod	2016-09	2016-10	2016-11	2016-12	2017-01	2017-02	2017-03	2017-04	2017-05	2017-06	...	2017-08	2017-09	2017-10	2017-11
1	0.0	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.0	9922.0	...	0.0	0.0	0.0	0.0
2	0.0	0.0	16452.0	7722.0	18719.0	0.0	12233.00	10766.00	26605.0	8622.0	...	4877.0	14562.0	20897.0	38011.0
3	0.0	0.0	7589.0	17714.0	8719.0	3650.0	51760.00	41558.08	25385.0	13347.0	...	63685.0	40577.0	88995.0	22414.0
4	0.0	3870.0	54182.0	23660.0	27438.0	23009.0	109459.00	26944.00	57635.0	39783.0	...	58149.0	86760.0	69282.0	41368.0
5	0.0	0.0	60308.0	13580.0	29984.0	48535.0	127342.00	91517.00	71542.0	44443.0	...	96490.0	67758.0	93608.0	79550.0
6	0.0	0.0	111773.0	29852.0	69739.0	50870.0	163507.00	160787.00	71206.0	112181.0	...	90041.0	101447.0	223277.0	87146.0
7	0.0	8152.0	131967.0	31359.0	76549.0	83058.0	221718.00	98536.00	113169.0	92464.0	...	106111.0	163029.0	232935.0	110624.0
8	5192.0	3871.0	189166.0	59662.0	68598.0	112340.0	253361.00	177452.00	125321.0	98262.0	...	133790.0	153217.0	291720.0	154931.0
9	5191.0	3871.0	208791.0	72838.0	118228.0	131727.0	384596.00	173925.00	161292.0	161767.0	...	156277.0	188723.0	337906.0	NaN
10	5191.0	13202.0	203724.0	77204.0	126165.0	215525.0	351706.00	254214.00	300423.0	120205.0	...	194731.0	218641.0	NaN	NaN
11	10381.0	3998.0	313240.0	116275.0	174859.0	198423.0	388158.96	377236.00	237023.0	133471.0	...	197615.0	NaN	NaN	NaN
12	10382.0	7869.0	303785.0	98300.0	183919.0	216724.0	547362.00	386572.00	269930.0	157412.0	...	NaN	NaN	NaN	NaN
13	10573.0	7868.0	331965.0	104809.0	221227.0	279197.0	514161.00	475814.00	255490.0	NaN	...	NaN	NaN	NaN	NaN
14	20956.0	11738.0	435203.0	147901.0	268442.0	274106.0	564544.00	468650.00	NaN	NaN	...	NaN	NaN	NaN	NaN
15	20956.0	11738.0	410643.0	138236.0	265308.0	303303.0	626175.00	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN
16	20956.0	11737.0	451633.0	200892.0	281623.0	323659.0	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN
17	26148.0	11737.0	453534.0	194689.0	290797.0	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN

17 rows × 21 columns

IV. Results

The following shed light on the results obtained -

1 . Vintage Curve for all the loans highlighting total due amount per cohort period.
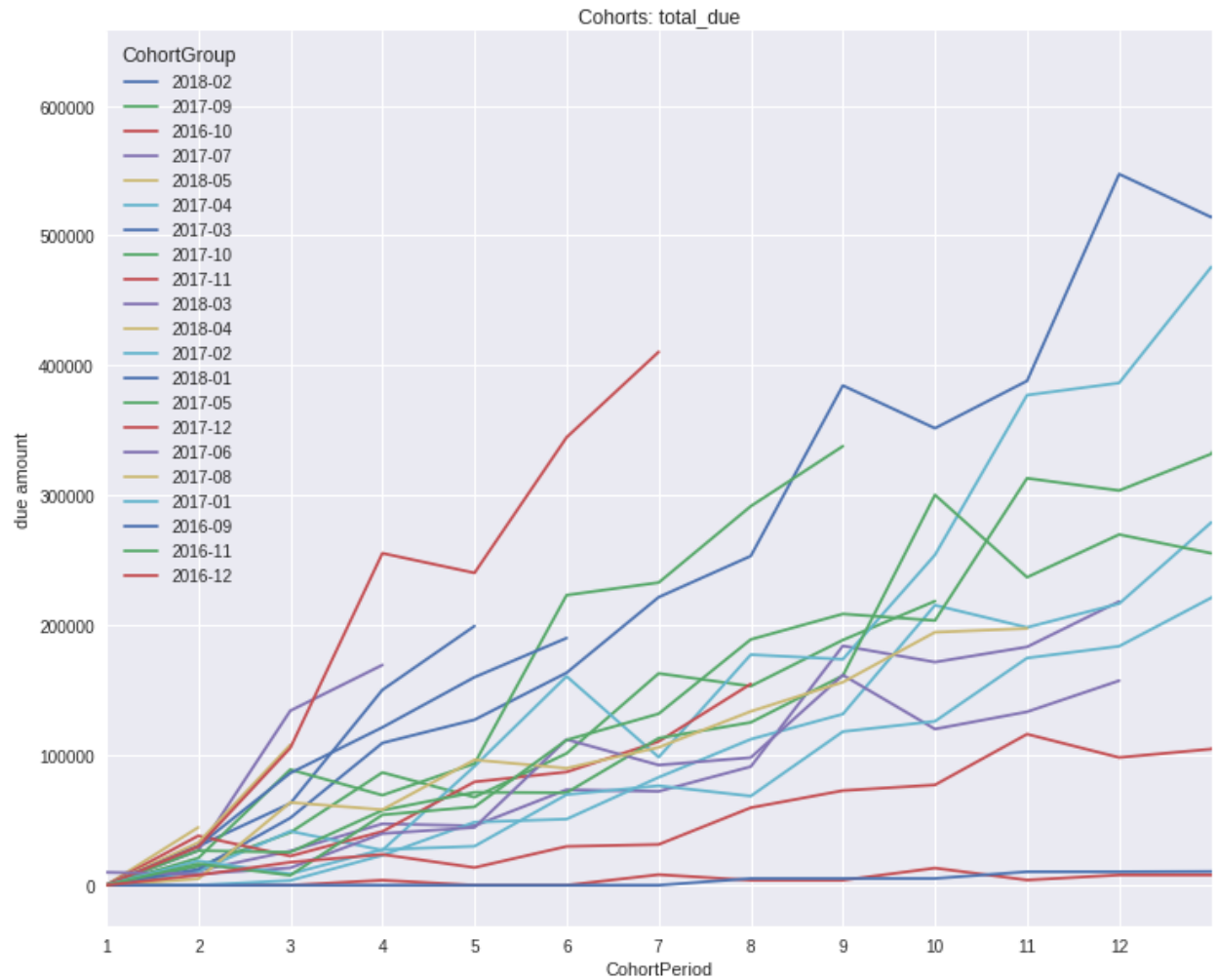

Fig : Vintage Curve for total dpp per cohort per period.

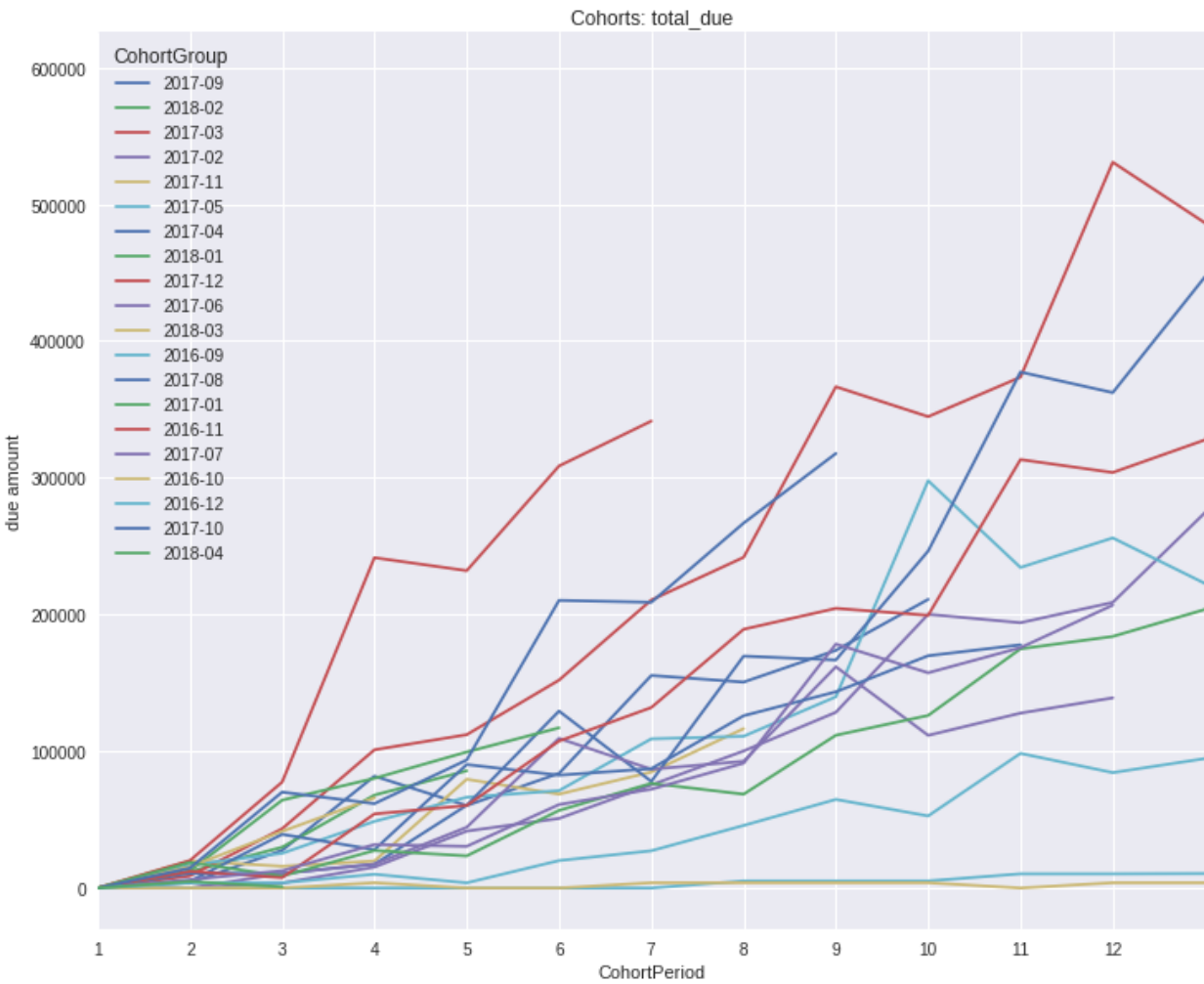2 . Vintage Curve for loan having 30+ DPP highlighting the total due amount per cohort period.

Fig : Vintage Curve for total dpp per cohort per period (for dpd > 30).