This dataset contains infusion soft user's buying behavior. A general dataset looks like:

```
      Date    Amount  CompanyID  Product  orderID
0 2009-12-04  1700.0      70353     PPOP        4
1 2010-11-15   700.0      70355     PPOP        9
2 2010-11-22   616.0      70371     PPOP       10
3 2010-12-02   910.0      69401     PPOP       25
4 2010-12-04   811.0      70353     PPOP       26
```

So we have date wise, company wise, product wise order and corresponding purchase amount provided on the dataset. A cohort analysis will provide us retention insight.

Steps:

From each purchase date, we prepared order period which is simply a YYYY-MM-DD date. Then for each company, we found out its minimum purchase date. We called it cohort group. The outcome after adding order period and cohort group column looks like:

```
   CompanyID       Date   Amount  Product  orderID  OrderPeriod  CohortGroup
0      70353  2009-12-04   1700.0    PPOP         4      2009-12      2009-12
1      70355  2010-11-15    700.0    PPOP         9      2010-11      2010-11
2      70371  2010-11-22    616.0    PPOP        10      2010-11      2010-11
3      69401  2010-12-02    910.0    PPOP        25      2010-12      2010-12
4      70353  2010-12-04    811.0    PPOP        26      2010-12      2009-12
5      70367  2011-01-14    700.0    PPOP        27      2011-01      2011-01
6      70391  2011-01-25   1890.0    PPOP        34      2011-01      2011-01
7      70383  2011-02-07   6075.0    PPOP        52      2011-02      2011-02
8      70373  2011-03-30  28665.0    PPOP        59      2011-03      2011-03
9      70363  2011-06-13   9000.0    PPOP        68      2011-06      2011-06
```

We then calculated number of total companies, total orders, and total amount purchased within each cohort group. We finally got a dataset with cohort period like this:

CohortGroup	OrderPeriod	TotalUsers	TotalOrders	Amount	CohortPeriod
2009-12	2009-12	1	1	1700.0	1
	2010-12	1	1	811.0	2
	2011-12	1	1	1560.0	3
2010-11	2010-11	2	2	1316.0	1
	2011-11	2	2	1050.0	2
2010-12	2010-12	1	1	910.0	1
	2011-12	1	1	6630.0	2
	2012-12	1	1	3314.0	3
	2013-12	1	1	3498.0	4
	2014-12	1	1	3653.0	5
	2015-12	1	2	6393.0	6
	2016-12	1	2	5290.0	7
	2017-12	1	2	6719.0	8
2011-01	2011-01	2	2	2590.0	1
2011-02	2011-02	1	1	6075.0	1
2011-03	2011-03	1	1	28665.0	1
2011-06	2011-06	2	2	9875.0	1
	2012-06	1	1	750.0	2
	2013-06	1	1	750.0	3
	2014-06	1	1	750.0	4

To find retention analysis, we calculated each cohort group size. Then finally we generated our retention dataset based on each company's first purchase data and onwards, their purchasing history.

Key Findings:

- We can see from the data that fewer users tend to purchase as time goes on.
- For example: User retention for period 2015-12 ended at 25%.
- Strongest cohort correlation observed on 2015-05, 2015-12 etc.
- Other attributes (besides first purchase month) do these companies share which might be causing them to stick around? How were the majority of these users acquired? Was there a specific marketing campaign that brought them in? Did they take advantage of a promotion at sign-up? etc question can be asked.

CohortPeriod	1	2	3	4	5	6	7	8	9	10
CohortGroup										
2015-01	100.0	100.0	83.3	83.3	NaN	NaN	NaN	NaN	NaN	NaN
2015-02	100.0	85.7	71.4	71.4	42.9	NaN	NaN	NaN	NaN	NaN
2015-03	100.0	14.3	100.0	85.7	57.1	NaN	NaN	NaN	NaN	NaN
2015-04	100.0	83.3	83.3	66.7	16.7	NaN	NaN	NaN	NaN	NaN
2015-05	100.0	33.3	100.0	33.3	100.0	33.3	66.7	NaN	NaN	NaN
2015-06	100.0	85.7	57.1	42.9	14.3	NaN	NaN	NaN	NaN	NaN
2015-07	100.0	71.4	71.4	57.1	NaN	NaN	NaN	NaN	NaN	NaN
2015-08	100.0	71.4	57.1	57.1	NaN	NaN	NaN	NaN	NaN	NaN
2015-09	100.0	16.7	66.7	66.7	66.7	NaN	NaN	NaN	NaN	NaN
2015-10	100.0	80.0	60.0	40.0	NaN	NaN	NaN	NaN	NaN	NaN
2015-11	100.0	12.5	87.5	50.0	50.0	NaN	NaN	NaN	NaN	NaN
2015-12	100.0	12.5	87.5	12.5	62.5	12.5	62.5	25.0	NaN	NaN
2016-01	100.0	90.0	60.0	50.0	NaN	NaN	NaN	NaN	NaN	NaN
2016-02	100.0	100.0	66.7	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2016-03	100.0	77.8	55.6	11.1	NaN	NaN	NaN	NaN	NaN	NaN
2016-04	100.0	50.0	50.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2016-05	100.0	100.0	88.9	11.1	NaN	NaN	NaN	NaN	NaN	NaN
2016-06	100.0	12.5	62.5	12.5	50.0	NaN	NaN	NaN	NaN	NaN
2016-07	100.0	100.0	60.0	20.0	NaN	NaN	NaN	NaN	NaN	NaN
2016-08	100.0	100.0	50.0	16.7	16.7	NaN	NaN	NaN	NaN	NaN
2016-09	100.0	25.0	100.0	75.0	25.0	NaN	NaN	NaN	NaN	NaN
2016-10	100.0	100.0	83.3	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2016-11	100.0	100.0	100.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN