Cohort analysis is a subset of behavioral analytics that takes the data from a given data set (e.g. an EMRS, an e-commerce platform, web application, or online game) and rather than looking at all users as one unit, it breaks them into related groups for analysis. These related groups, or cohorts, usually share common characteristics or experiences within a defined time-span.

Whatever the key metrics are in your particular business, a cohort analysis lets you see how those metrics develop over the customer lifetime as well as over what might be called product lifetime:

If you read the chart above *horizontally*, you can see how your retention develops over the customer lifetime, presumably something that you can link to the quality of your product, operations, and customer support. Reading it *vertically* shows you the retention at a given lifetime month for different customer cohorts. This might be called product lifetime; and, especially if you look at early lifetime months, it can be linked to the quality of your onboarding experience and the performance of your customer success team.

This chart has like two really big main benefits so one we can **compare different cohorts or groups of users at the same stage in their life cycle** so that would be looking down from top to bottom on the chart so like we can see for all the cohorts what percentage are coming back to the site three months after they signed up, and we can compare how we're doing with our improvement in product and user experience so hopefully that's increasing in our products becoming better and more enjoyable.

The second big benefit is we can **see the long term relationship** that we have with a given user group for each cohort about 25 percent are coming back after 1 month 6 percent coming after two months and we can see how long people from a given cohort are coming back to the site.

customer
- customer_id (smallint)
- store_id (tinyint)
- first_name (varchar)
- last_name (varchar)
- email (varchar)
- address_id (smallint)
- active (tinyint)
- create_date (datetime)
- last_update (timestamp)

```sql
/* --Customer's First Rental */
DROP TEMPORARY TABLE IF EXISTS first_rental;
CREATE TEMPORARY TABLE first_retal;

SELECT
    customer_id,
    min(rental_date) as first_time
FROM
    rental
GROUP BY 1
;
```

customer_id	first_time
1	05/25/2005 11:30:37
2	05/27/2005 00:09:24
3	05/27/2005 17:17:09
4	06/15/2005 09:31:28
5	05/29/2005 07:25:16
6	05/25/2005 08:43:32

```sql
/* --Cohort Size */
DROP TEMPORARY TABLE IF EXISTS cohort_size;
CREATE TEMPORARY TABLE cohort_size;

SELECT
   left(first_time,7) as month,
   count(customer_id) as num
FROM
   first_rental
GROUP BY 1
;
```

month	num
2005-05	520
2005-06	78
2005-07	1

payment
- payment_id (smallint)
- customer_id (smallint)
- staff_id (tinyint)
- rental_id (int)
- amount (decimal)
- payment_date (datetime)
- last_update (timestamp)

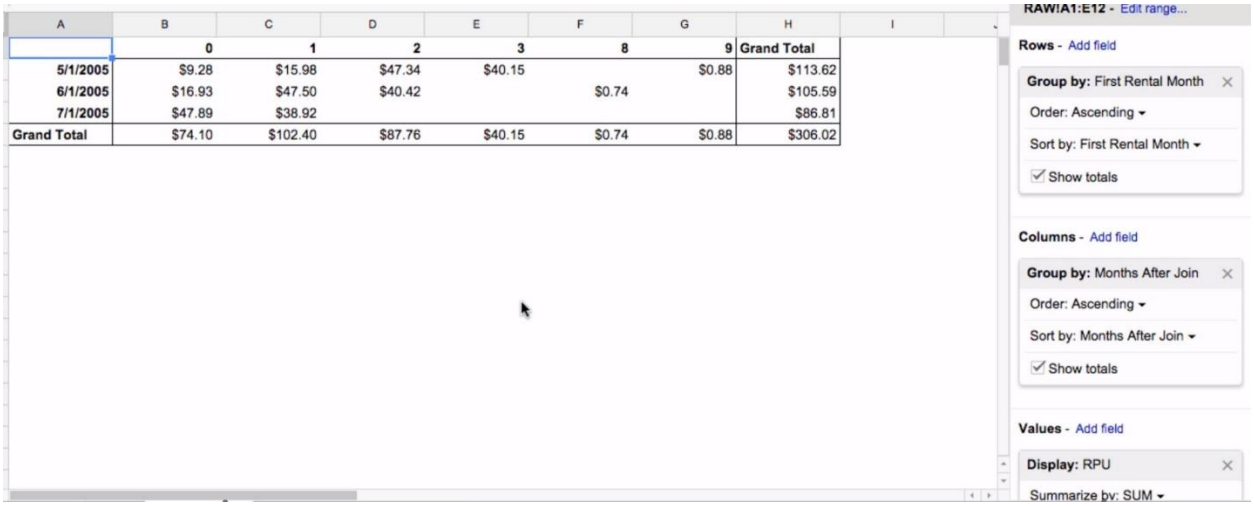
```sql
/* --Revenue Per User for Each Cohort and month */
DROP TEMPORARY TABLE IF EXISTS cohort;
CREATE TEMPORARY TABLE cohort;
SELECT
    data_format(f.first_time, '%Y%m') as cohort_formatted,
    data_format(r.rental_date, '%Y%m') as rental_date_formatted,
    cs.num as cohort_size,
    sum(p.amount) as month_rev,
    sum(p.amount)/cs.num as RPU
FROM
    rental r
    JOIN first_rental f ON r.customer_id = f.customer_id
    JOIN cohort_size cs ON cs.month = left(f.first_time,7)
    JOIN payment p ON p.rental_id=r.rental_id
GROUP BY 1,2
;
```

cohort_formatted	rental_date_formatted	cohort_size	month_rev	RPU
200505	200505	520	4823.44	9.275846
200505	200506	520	8309.17	15.979173
200505	200507	520	24615.91	47.338288
200505	200508	520	20878.69	40.151327
200505	200602	520	456.40	0.877692
200506	200506	78	1320.72	16.932308
200506	200507	78	3705.11	47.501410
200506	200508	78	3152.53	40.417051
200506	200602	78	57.78	0.740769
200507	200507	1	47.89	47.890000
200507	200508	1	38.92	38.920000

```sql
/* --Prettify */
SELECT
  left(STR_TO_DATE(cohort_formtted, '%Y%m'),7) as 'First Rental Month',
  period_diff(rental_data_formatted, cohort_formatted) as 'Month After Join',
  cohort_size,
  RPU
FROM
   cohort
GROUP BY 1,2
;
```

First Rental Month	Months After Join	cohort_size	RPU
2005-05	0	520	9.275846
2005-05	1	520	15.979173
2005-05	2	520	47.338288
2005-05	3	520	40.151327
2005-05	9	520	0.877692
2005-06	0	78	16.932308
2005-06	1	78	47.501410
2005-06	2	78	40.417051
2005-06	8	78	0.740769
2005-07	0	1	47.890000
2005-07	1	1	38.920000

Copy to Excel use Pivot



	0	1	2	3	8	9	Grand Total
5/1/2005	$9.28	$15.98	$47.34	$40.15		$0.88	$113.62
6/1/2005	$16.93	$47.50	$40.42		$0.74		$105.59
7/1/2005	$47.89	$38.92					$86.81
Grand Total	$74.10	$102.40	$87.76	$40.15	$0.74	$0.88	$306.02