Background:
In this exercise, you'll be calculating the Lifetime Value (LTV) of a cohort of customers from a raw data set. This is not just some "test" but rather is an analysis that we often perform to help guide biz intel and customer acquisition strategies.

LTV: The LTV is a metric used to value customers that a business acquires. It is a fundamental KPI that is used to analyze the health of our business's marketing efforts. In SaaS and online retail, investors and operators will often compare the LTV to the cost to acquire a customer (CAC) as a way to measure the ROI of advertising spend.

At a high level, the lifetime value of a customer is equivalent to the purchases a customer makes over a given timeline. The LTV should always be defined by a particular time horizon (e.g., 6 month, 12 month, 24 month) so it can be compared (apples to apples) to other businesses. The LTV of a cohort of customers is calculated by grouping customers into a cohort, and then tracking their purchases over time.

The Exercise

See the linked **data file**:

- this is order data from 2/1/2015 through 3/30/2016 (approximately 14 months)
- note that the data here has been manipulated slightly for obvious security reasons
- you can assume that all purchasers in month 1 of the data set (2/15) are new customers

Part 1:

I have pasted an example of the chart we would like you to create. Please create a similar chart for the LTV of the February cohort. You can use all the data in the data set (so 2/15 through 3/16 - 13 month time horizon after the end of 2/15).

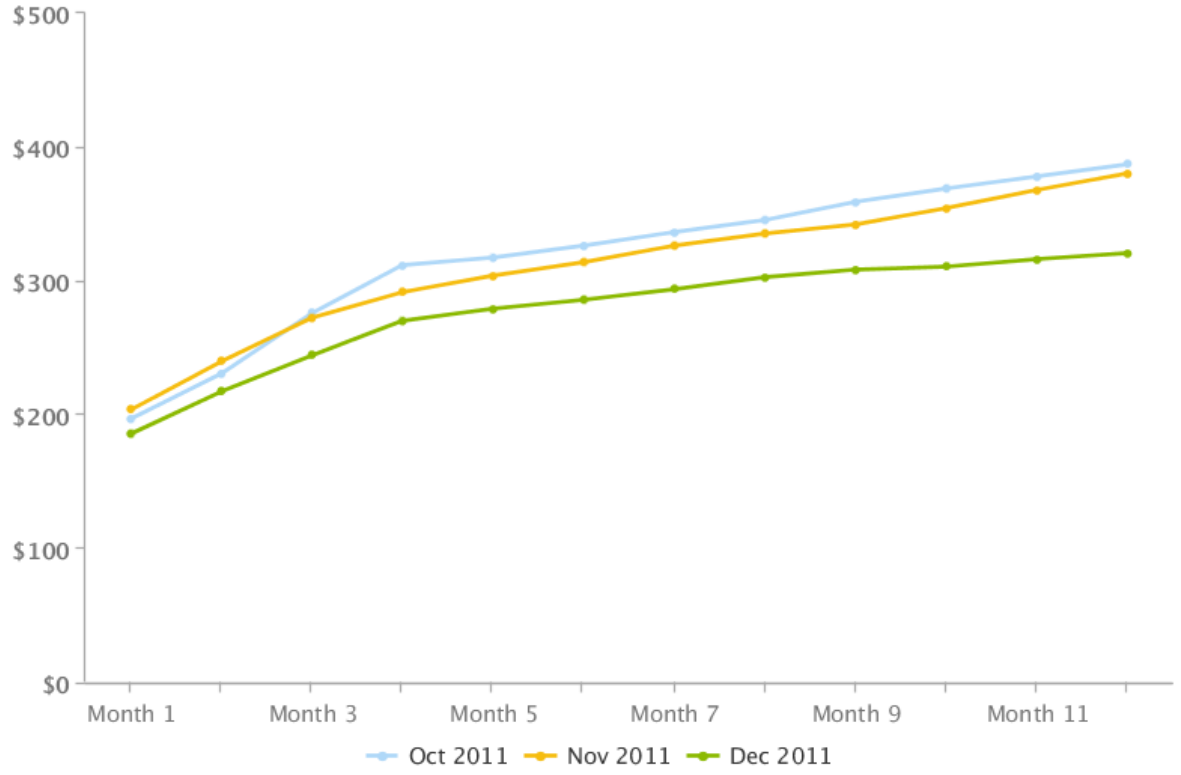
- the cohort LTV is segmented by customer type
- % of customers is the % of all customers in the cohort that are the customer type
- % of Rev is the % of gross revenue of the entire cohort (only the cohort's revenue) over the entire time horizon
- LTV (GMV) is the average LTV of the customer type

Very important!! The LTV should be calculated for only one cohort of customers (in this case February) over a specific timeline. Some people fall into the "trap", where they take a dataset of customers (say 12 months of data), find the number of customers, sum of purchases per customers, etc, and present it in the format below. This is flawed because the customers in the later months in that data set do not have a chance to repurchase. Therefore, you always need to think of LTV (and calculate it) over a specific time horizon and track the purchases of a particular group of customers over time.

22 Month LTV (February 2014 Cohort)			
Customer Type	% of customers	% of Rev	LTV (GMV)
All Customers	100%	100%	$264.75
1 time Purchasers	72.81%	47.16%	$171.46
>1 purchase	27.19%	52.84%	$514.60
>2 purchases	11.30%	33.06%	$774.69
>3 purchases	6.03%	23.59%	$1,035.46
>4 purchases	3.57%	18.09%	$1,342.12
>5 purchases	2.29%	15.22%	$1,756.74
>6 purchases	1.44%	12.21%	$2,237.85
>7 purchases	1.10%	10.50%	$2,516.80
>8 purchases	0.76%	8.38%	$2,901.96
>9 purchases	0.59%	7.82%	$3,481.96

Part 2:

You probably have seen a lot of charts like the one below that calculate the LTV of multiple cohorts over month 1, month 2, month 3, and so on. Please calculate a line graph of the LTV for customers acquired in months Feb '15 through Jul '15 similar to the one below using the same data set.



Note that there are multiple ways to think about the time horizon. One way is to classify Feb '15 purchases made by Feb '15 customers as month 1, and Mar '15 purchases made by Feb '15 customers as month 2. In this methodology, the time horizon of the customer isn't exact, but it's based on the cohort month. Another way is to calculate the time horizon by customer -- so if a customer purchases on Feb 15, then purchases again on March 10, both purchases would fall into month 1. For the sake of time, please use the first methodology. Let us know if this isn't clear.

Lastly, if you run into issues (w/ the data, instructions, etc) please feel free to email. We are happy to answer any questions you have. We don't care if your work is messy, scrappy, etc, we should be able to follow so long as you ***leave all worksheets and formulas*** (there may be times where you have to hardcode values, but leave formulas where you can...)