

Customer Segmentation and Life Time Value for Flash Deal Site

Steffen Hartleib

sysdate()

The Data Set

Online purchase data from 2010 to 2016

553087 observations (line items), 19 variables

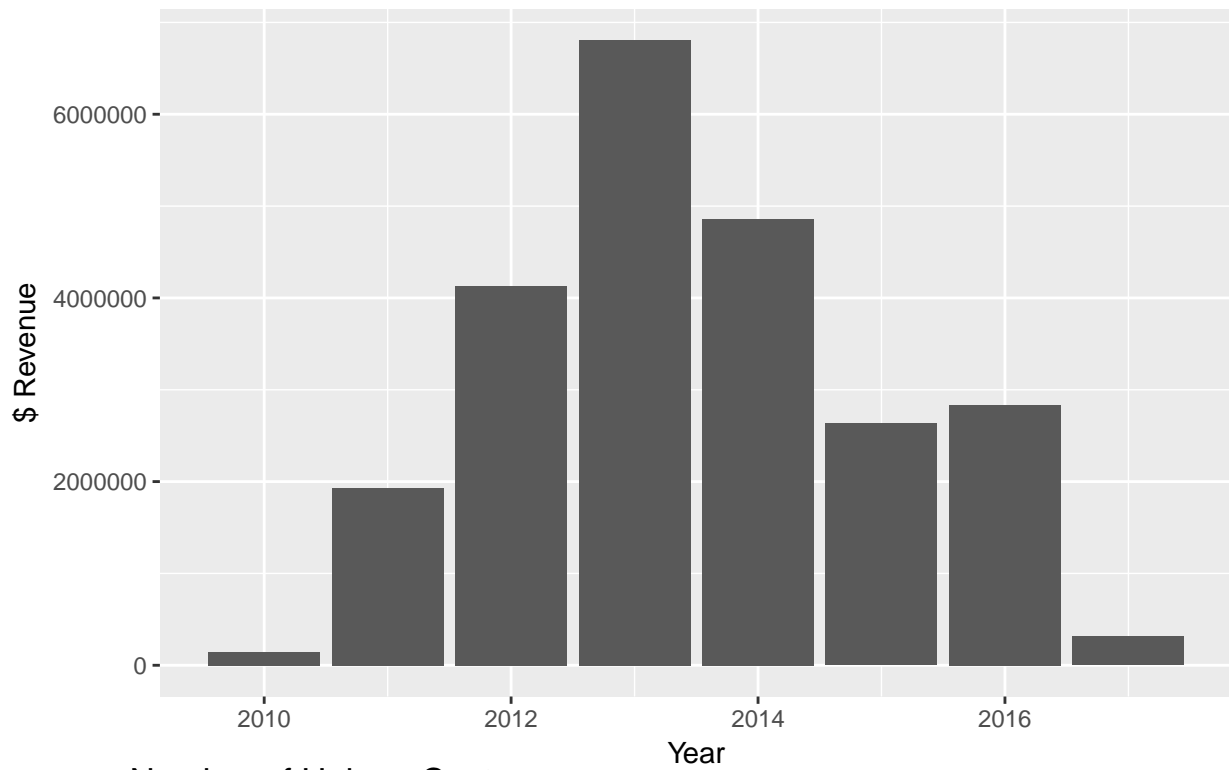
Questions

1. Which customer segments are driving revenue?
2. What is the average life time value of a customer?

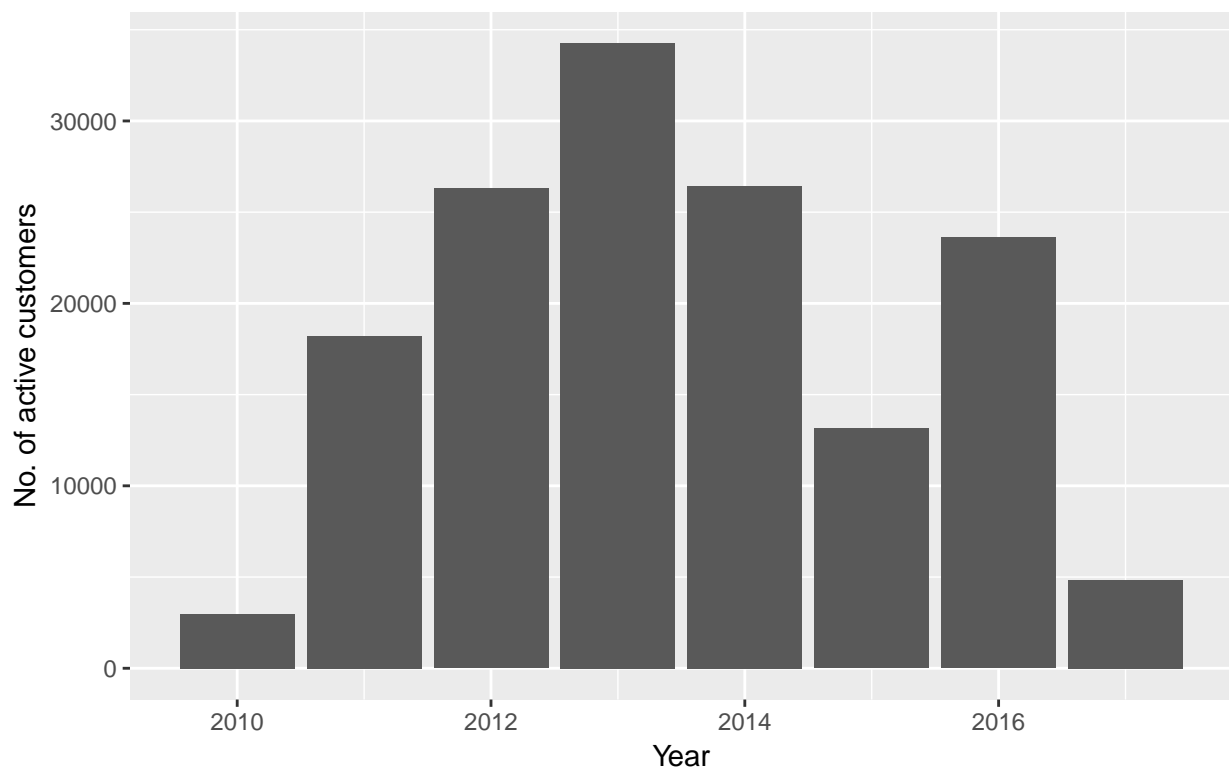
Let's look at year over year revenue and customer trends

```
## [1] TRUE
```

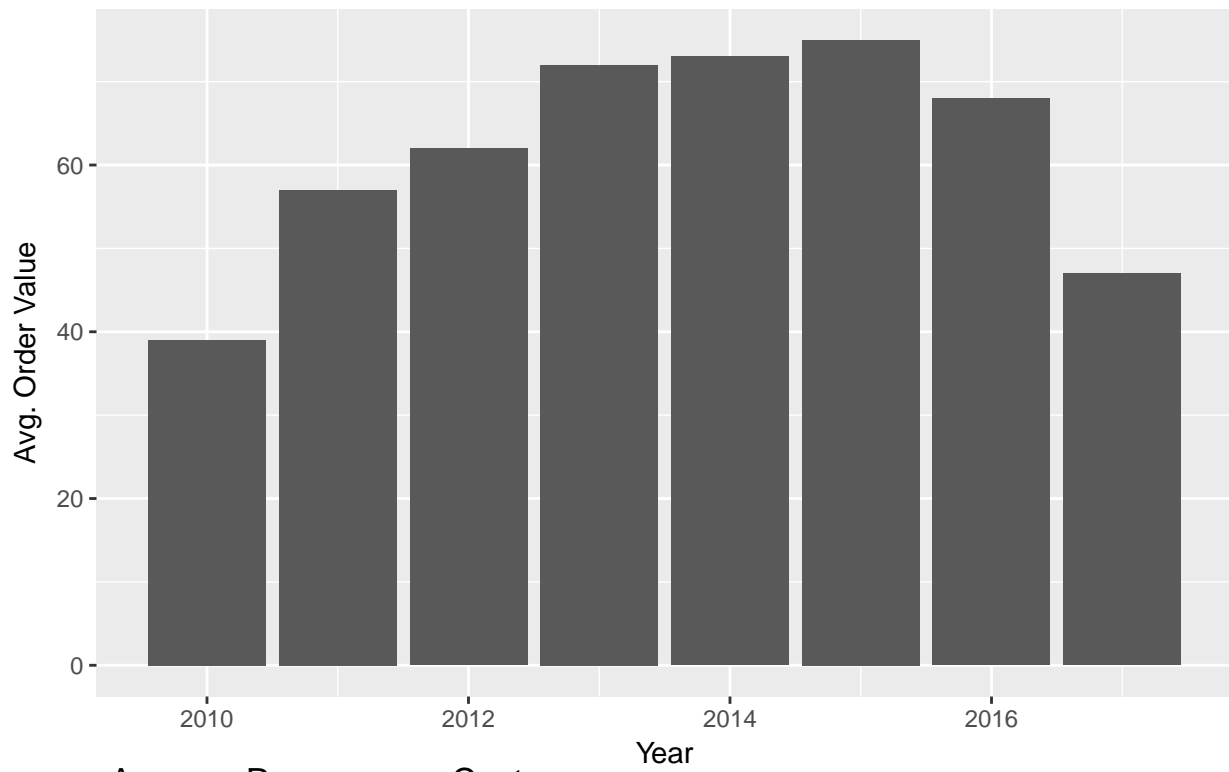
Yearly Revenue:



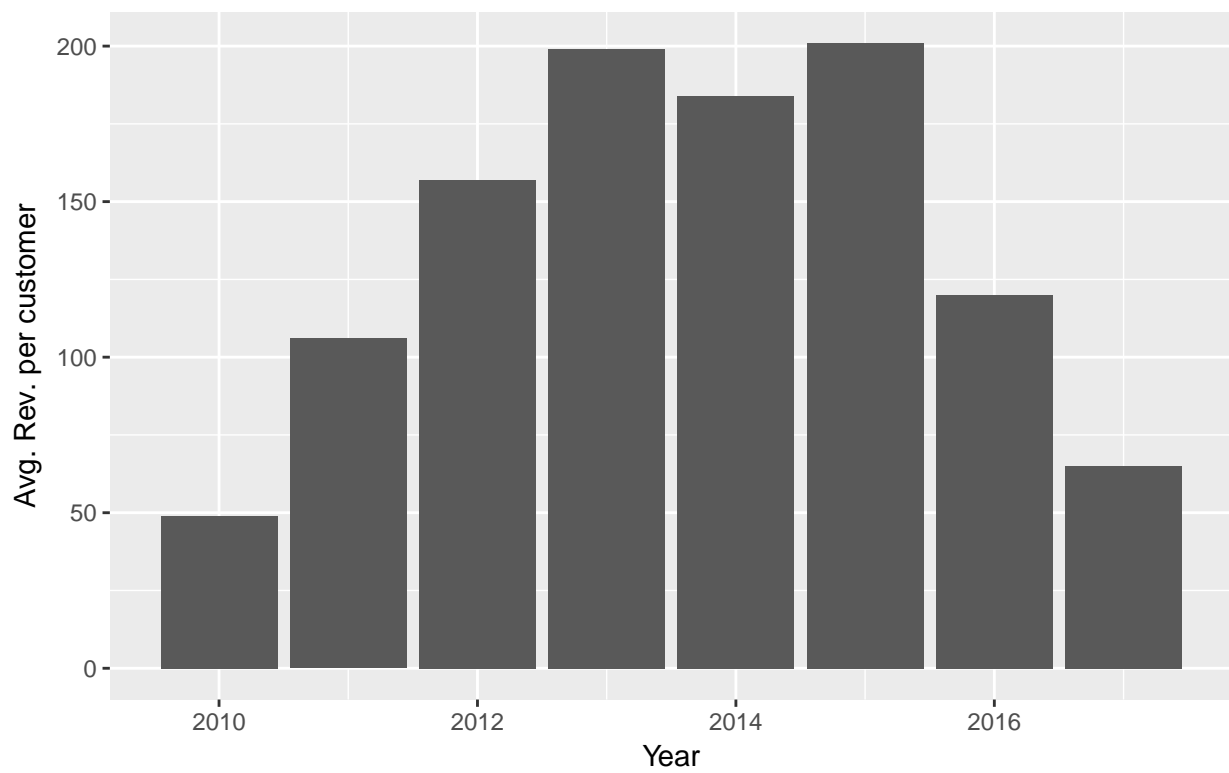
Number of Unique Customers:



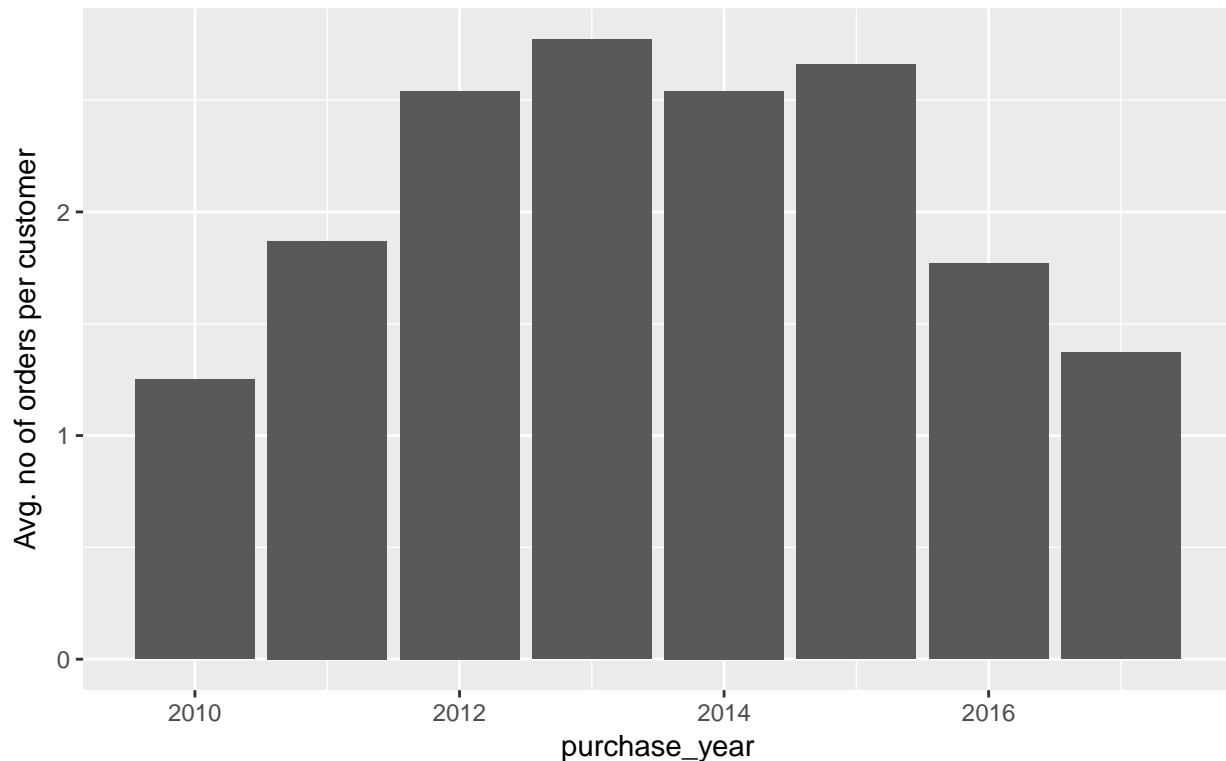
Average Order Value:



Average Revenue per Customer:



Average Number of Orders per Customer:



=> In 2016 the number of customers increased significantly. But Total Revenue rose only slightly. Because, on average, customer placed fewer orders and spent less per order. This is pattern is typical for customer acquisition campaigns driven by discounts. The challenge will be to convert these new customers from bargain shoppers to valuable long term customers.

Let's segment the customers to see what's behind these trends.

Simple Customer Segmentation by Recency, Frequency and Monetary Value (RFM)

We need to calculate three variables per customer:

1. Recency: Number of days since the last purchase
2. Frequency: Number of orders placed, life to date
3. LTD Revenue: Product Revenue net of discounts life to date

Here are the summary statistics of each variable:

	Recency	LTD.Revenue	No.of.Purchases
Minimum	0	1	1
Average	1,042	224	3
Median	1,115	70	1
Max	2,292	222,182	2,863

=> Look at LTD Revenue! The mean is over three times the median. And one customer spent over \$200k. Clearly a few high rollers are pulling up the the average. (A bit like Bill Gates sitting down at your neighborhood bar...)

Let's define our customer segements as follows. I'm using round numbers since this is a managerial analysis.

Segment	Recency	LTD Revenue
Active New	1st purchase last year	
Active Low	purchase last year	< \$100
Active Med	purchase last year	>= \$100 & < \$1000
Activ High	purchase last year	> \$1000
Inactive Low	no purchase last year	< \$100
Inactive Med	no purchase last year	>= \$100 & < \$1000
Inactive High	no purchase last year	> \$1000

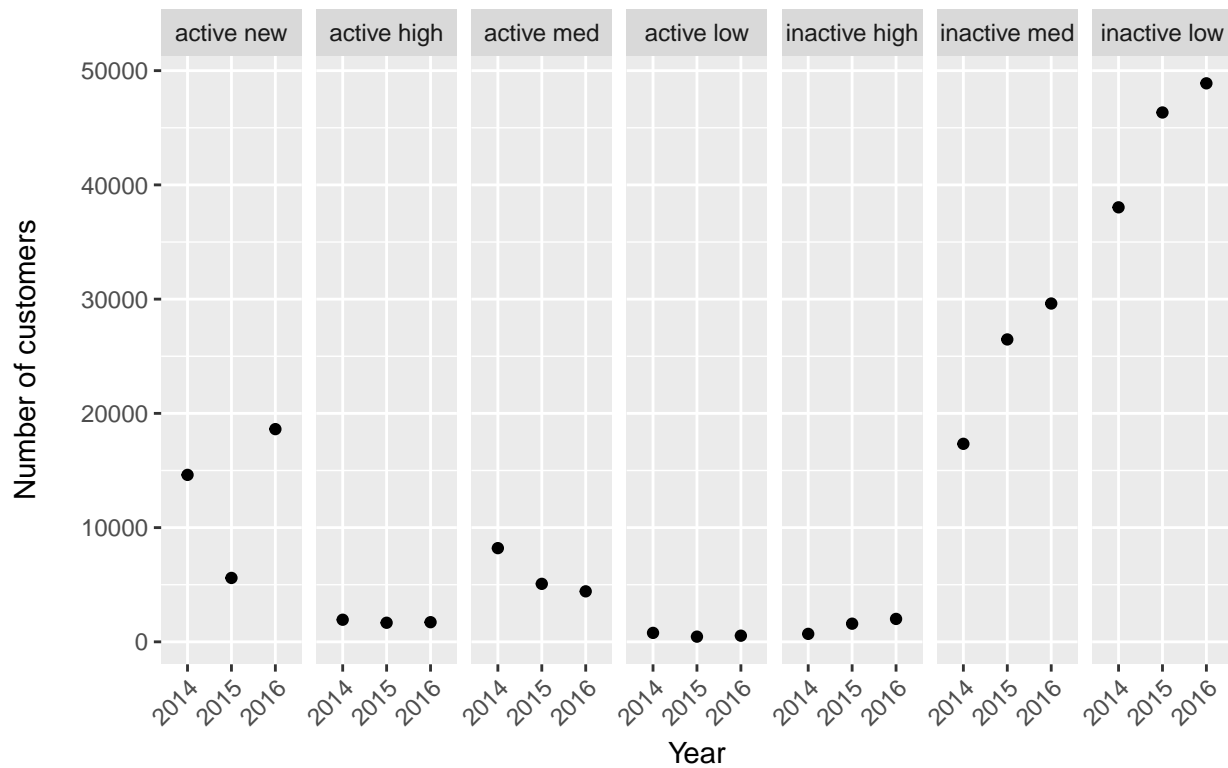
Summary of the segments:

segment	Customers	PercOfTlCustomers	RevTD	Rev365	PercOfTlRev365
active new	18621	0.18	1410602	1410602	0.48
active high	1719	0.02	5432010	950284	0.32
active med	4420	0.04	1813677	576158	0.20
active low	538	0.01	38624	16103	0.01
inactive high	2008	0.02	4770592	0	0.00
inactive med	29617	0.28	7817822	0	0.00
inactive low	48892	0.46	2371343	0	0.00

How neat, a perfect 80/20 ratio: In 2016 80% of revenue came from exactly 20% of customers (New Active + Active High). 52% of revenue came from just 6% of customers (Active High + Active Medium). And 32% of revenue came from the top 2% of customers (Active High). This site has some seriously loyal high rollers. Great!

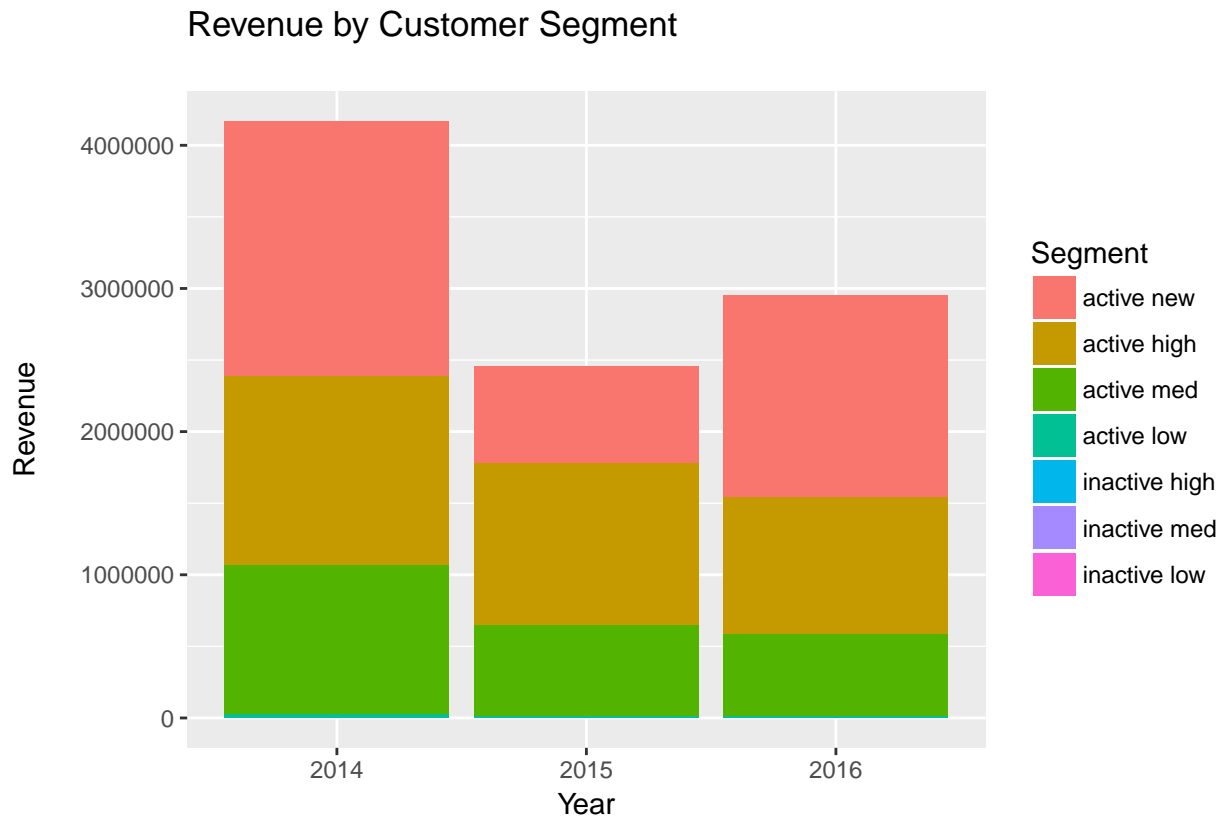
Here is how the customer segments changed over time:

Number of customers per segment over the last three years



Active New customers more than doubled last year. Active High/Med/Low segments stayed almost flat. So customer acquisition efforts are paying off and customer retention is working. The challenge will be to retain these new active customers, and to convert them to higher priced offers.

Here is how revenue per segment stacks up, over time:



New Active customers drove the increase in revenue.

What's the value of the current data base?

This isn't the same as Customer Life Time Value (LTV). LTV tells us how much a customer spends on average in her lifetime. That's very useful to determine how much we can spend to acquire a new customer. Here we're asking how much more revenue we can expect from the customers in the data base.

Let's make a Transition Matrix

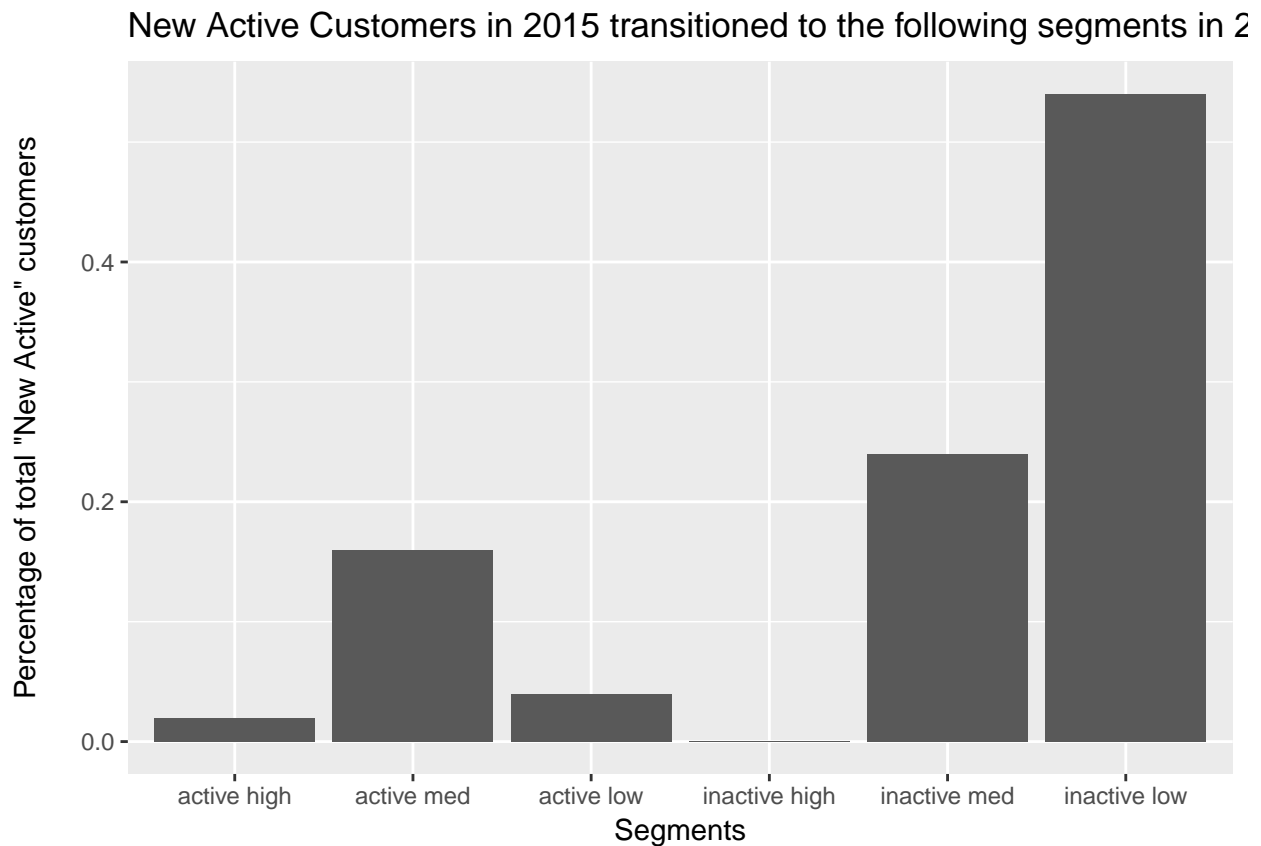
It shows the probabilities of customers changing segments between 2015 and 2016, e.g. how many "Active High" customers in 2015 were also Active High customers in 2016? etc. (the answer is 67%, not bad!)


```

| | active new| active high| active med| active low| inactive high| inactive med| inactive low| |-----|-----|
-:|-----:|-----:|-----:|-----:|-----:|-----:| | active new | 0| 0.02| 0.16| 0.04| 0.00| 0.24| 0.54|
|active high | 0| 0.67| 0.00| 0.00| 0.33| 0.00| 0.00| |active med | 0| 0.06| 0.34| 0.00| 0.00| 0.60| 0.00| |active low |
0| 0.00| 0.19| 0.06| 0.00| 0.00| 0.75| |inactive high | 0| 0.09| 0.00| 0.00| 0.91| 0.00| 0.00| |inactive med | 0| 0.00|
0.04| 0.00| 0.00| 0.95| 0.00| |inactive low | 0| 0.00| 0.01| 0.01| 0.00| 0.00| 0.98| [1] "table" ***

```

Let's quickly visualize where last year's New Active customers ended up:



22% of New Active customers are still Active the next year. Most of them are in the Active Medium segment. Curious to see what next year will look like...

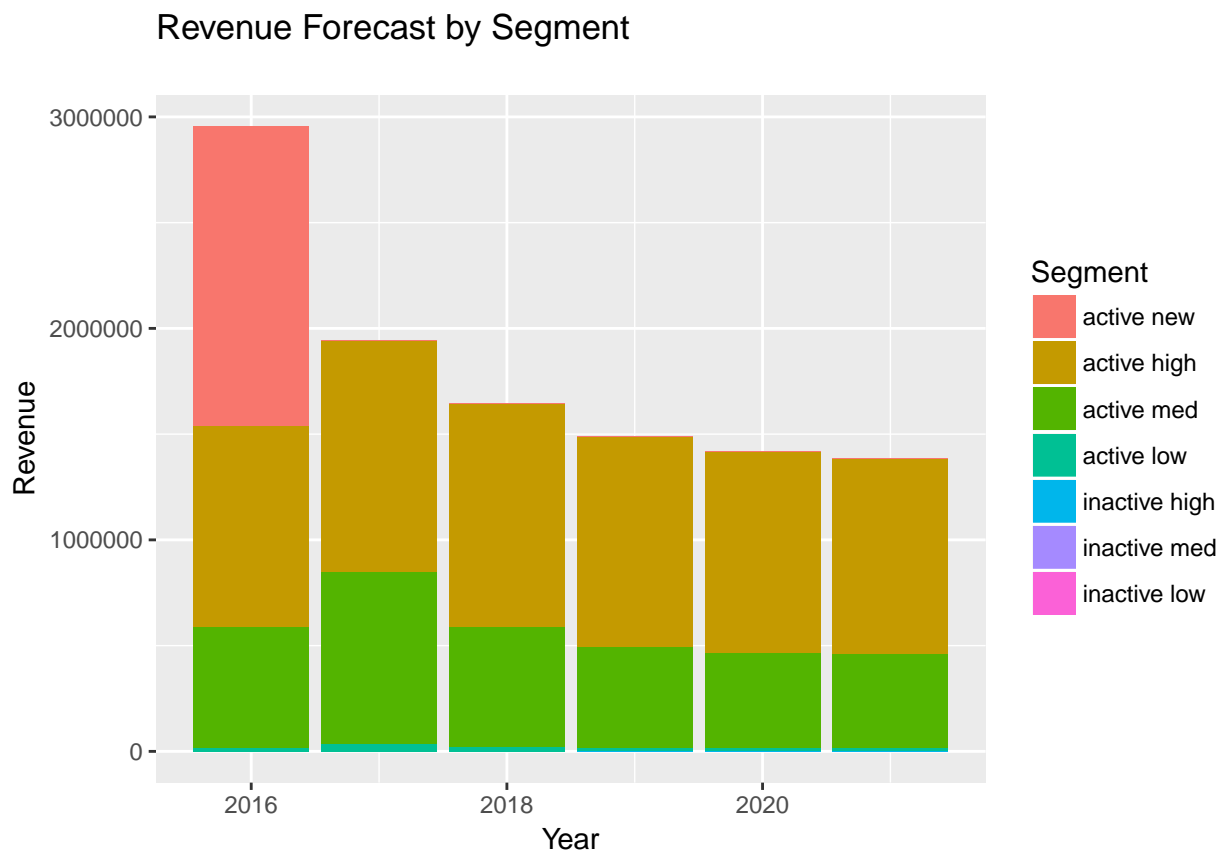
How do we forecast the size of each segment over the next 5 years?

Let's assume customers will continue to transition from segment to segment in the same proportions as the did from 2015 to 2016. That way we can multiply this year's segments by the transition matrix to predict next year's segments. Then we'll multiply next year's (forecasted) segments by the same transition matrix to get the following year. And so on.

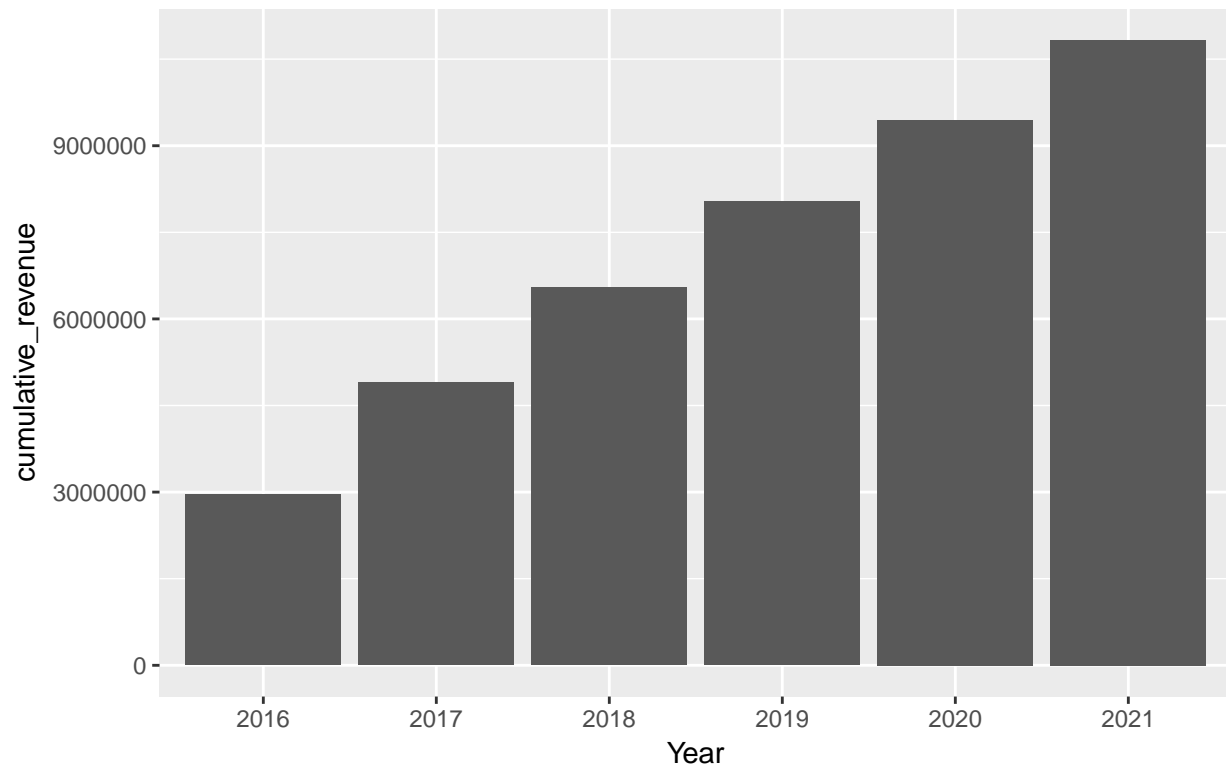
```
##
##
## |           | 2016| 2017| 2018| 2019| 2020| 2021|
## |:-----:|-----:|-----:|-----:|-----:|-----:|-----:|
## |active new | 18621| 0.00| 0.0000| 0.0000| 0.0000| 0.0000|
## |active high | 1719| 1970.07| 1910.9352| 1796.6880| 1713.0461| 1671.4594|
## |active med  | 4420| 6257.98| 4362.2784| 3680.0317| 3464.2545| 3394.6546|
## |active low  | 538| 1266.04| 659.6924| 621.1322| 612.1353| 604.7566|
## |inactive high | 2008| 2394.55| 2829.1636| 3205.1475| 3509.5913| 3759.0333|
## |inactive med | 29617| 35257.19| 37249.1185| 38004.0296| 38311.8471| 38474.8074|
## |inactive low | 48892| 58373.00| 58155.0700| 57486.7379| 56802.8523| 56125.8968|
```

Now we can forecast revenue per segment for the next 5 years

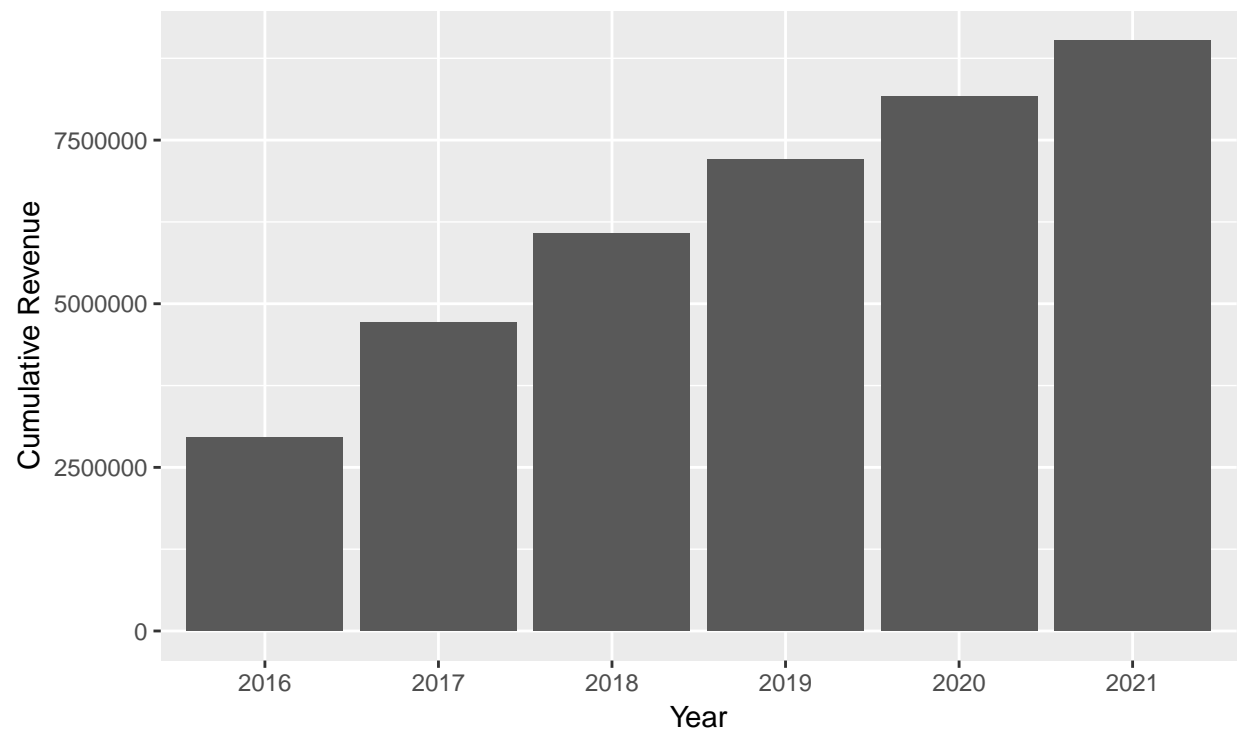
(assuming that, on average, revenue per segment stays the same)



Cumulative Revenue Projected



Present Value of Cumulative Revenue
Discounted at 10%



Total net present value of the data base in 2020 is:

##	
##	
##	X2020
##	:----- :-----
##	Net Present Value 6,069,183

And let’s add the Customer Life Time Value

Customer Life Time Value		Total Net Present Value of Data Base
2021	57.36	6,069,183

##	2020
##	Net Present Value 6,069,183

##	
##	
##	X2020
##	:----- :-----
##	Net Present Value 6,069,183