APANPS5900

SOLVING REAL WORLD PROBLEMS WITH ANALYTICS

Cost Structure Optimization for HubSpot's Emerging Markets

Proof of Concept

Instructor: Pieter De Leenheer

Date: 10/6/2018

Written by: Mengjuan Xing (mx2184), Fang Yang (fy2230), Chenlu Wang (cw3057), Quanlei Chi (qc2246), Sarah Garman (sag2213), Gian Wirjawan (gw2368), Jiaying Wu (jw3644)

APAN 5900 Proof of Concept

Introduction

Our proof of concept aims to deliver the following:

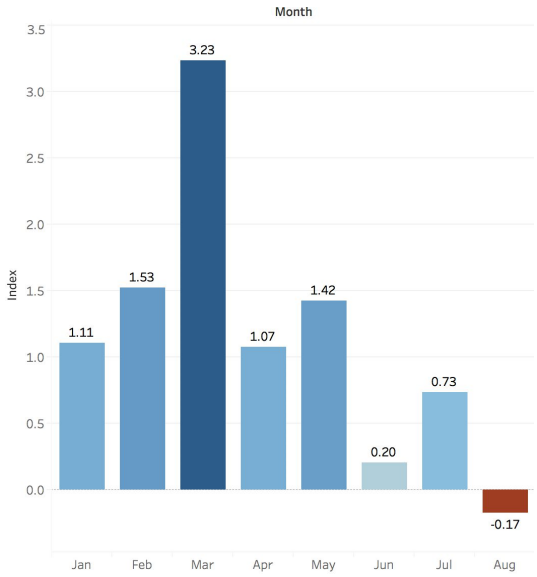
- Expand our early analyses of Hubspot
- Address each research question directly concluding in results or “proofs”, and provide preliminary strategies as a result of our findings

The challenge is to provide strategies on how Hubspot can boost its LTV:CAC ratio to at least 4 times or greater. The research questions to tackle this challenge are the following:

1. Are the independent variables: Deal Discount, Salary, Commission, and Sales Representative Headcount, correlated to HubSpot’s health index (LTV:CAC Ratio)?
2. Are there any fluctuations in HubSpot’s revenue over time?
3. Do different company sizes impact HubSpot’s revenue differently?
4. Does the business revenue vary over different industries?

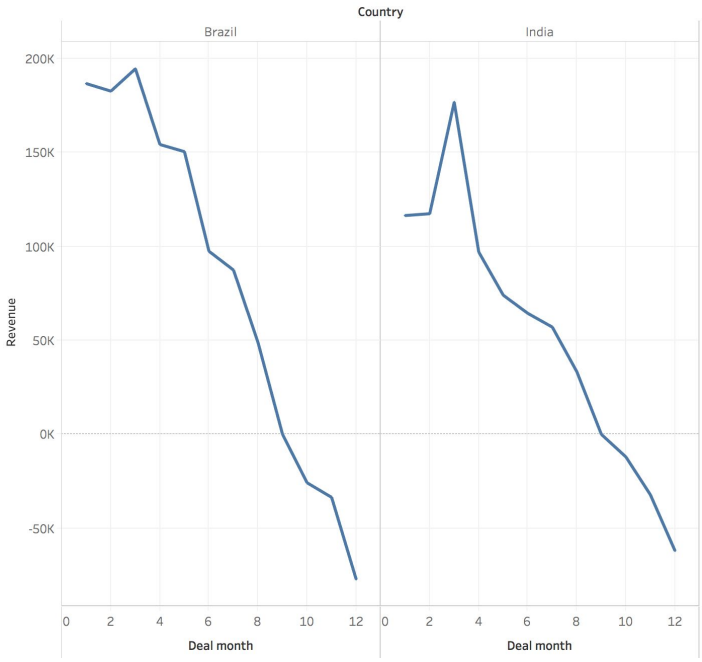
Proof 1: Fluctuations in Revenue and Health Index Over Time

Early analysis indicates that there is a fluctuation in Hubspot’s health index and life-time value of the customer (January 2018 - August 2018).



From the chart on the left (charts created in Tableau), it is clear that the LTV:CAC "health index" varies by month, with March having the highest ratio and August having the lowest with a negative value. The chart on the right compliments the findings from the chart on the left; if there is a significant margin between the CAC and the LTV, the health index will be lower. This margin represents a mismatch where Hubspot is spending more to acquire customers than the revenue obtained from them.

Exploring the patterns more deeply, we explored the long-term revenue trend ***by country for the years 2010 until 2018***, as seen below.

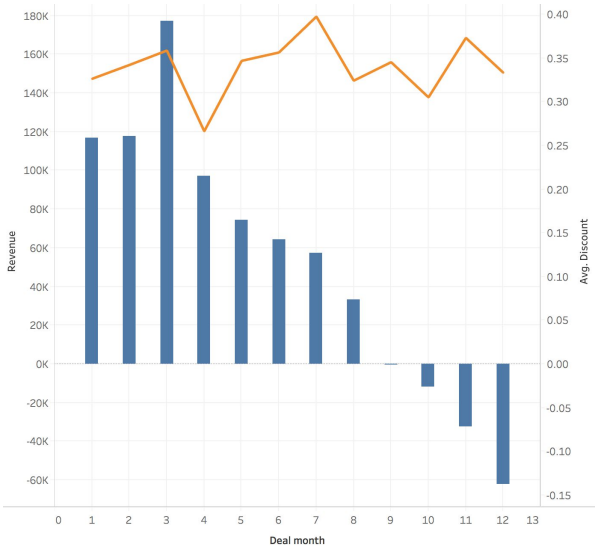


Over the course of 8 years, in both countries, there is a similar pattern of revenue where the revenue peaks in March before a significant decrease from April until the end of the year. These analyses expands on earlier analysis and prove that:

- March is the peak month for the LTV:CAC Ratio
- The revenue dropped 100% within 6 months after the peak month

Brazil Trend of Revenue vs Discount Rate

India Trend of Revenue vs Discount Rate



The graphs of revenues over time against the deal discount rate show that after March, the discount rates were either increased or maintained at a high level. Despite the higher average discount rates, revenues in both countries continued to drop. There is little-to-no impact on revenue from discount rate when looking at different countries as a whole.

We ran a statistical significance test to address our hypotheses for this research question.

- Statistical Calculation:
  - Time range: 2018 Jan-Aug
  - Method: ANOVA F test

```
> aov.time = aov(Revenue~Deal_month, data = sales_data)
> summary(aov.time)
              Df    Sum Sq Mean Sq F value Pr(>F)
Deal_month     1 8.300e+07 8.3e+07   83.84 <2e-16 ***
Residuals   1433 1.419e+09 9.9e+05
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

  - p-value = <2e-16 << 0.05 -> reject the H0
  - Yes, the business revenues are significantly different over 2018 Jan-Aug.

Proof 2: Correlation of Variables

Are the independent variables: Deal Discount, Salary, Commission, and Sales Representative Headcount, correlated to HubSpot's health index?

Limitations: 1) It is statistically meaningless to calculate the coefficient because we calculate the dependent variable by the independent variables, which brings us the multicollinearity. Even if we calculate the statistics, the correlation coefficient and the result of F-/T - Test is not reliable. 2) After looking at the dataset and attempting to put together a model, we found that salary and commission only represent one data entry as a generalized representation of each country and is used for cost only. They are not integrated with the customer and sales datasets; therefore, this model does not help us address this question.

Solution: We ran segmentation models by company size, country, and industry.

Proof 3: Company Sizes Impact

Do different company sizes impact HubSpot's revenue differently?

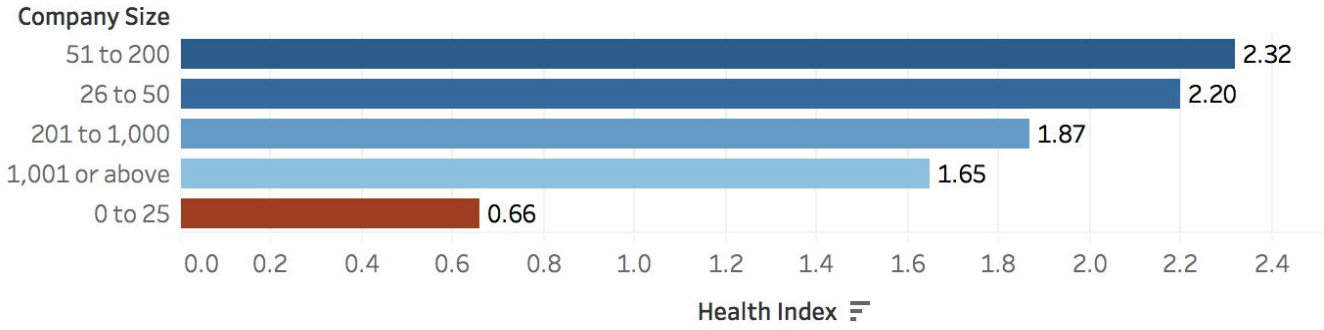
First, we recategorized company sizes.

- Statistical Significance Calculation
  - Time range: 2018 Jan-Aug
  - Method: ANOVA F test

```
> aov.size = aov(Revenue~CompanySize, data = sales_data)
> summary(aov.size)
             Df    Sum Sq Mean Sq F value  Pr(>F)
CompanySize   4 3.113e+07 7781451   4.529 0.00128 **
Residuals   736 1.265e+09 1718249
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
694 observations deleted due to missingness
```

  - p-value = 0.00128 << 0.05 -> reject the H0
  - Yes, the business revenues are significantly different over different company sizes.

Emerging Markets LTV:CAC Ratio Health Index by Company Size

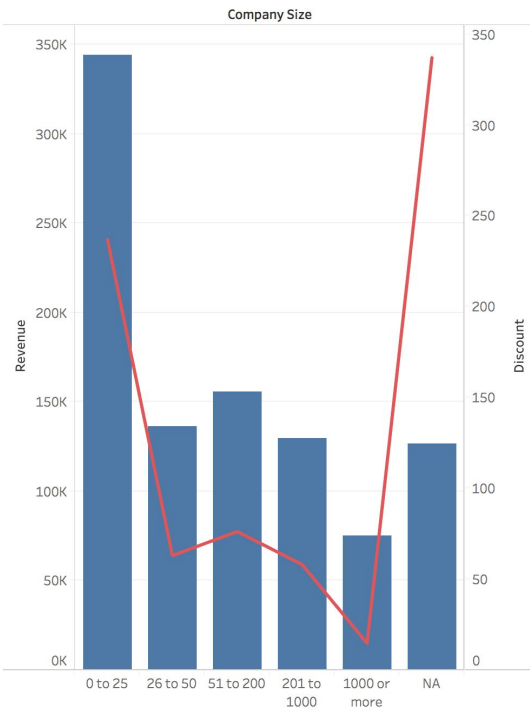


Our preliminary analysis looks at the overall health index by company size. As seen by the graph above, the two highest categories of company sizes that provide the highest ratios are:

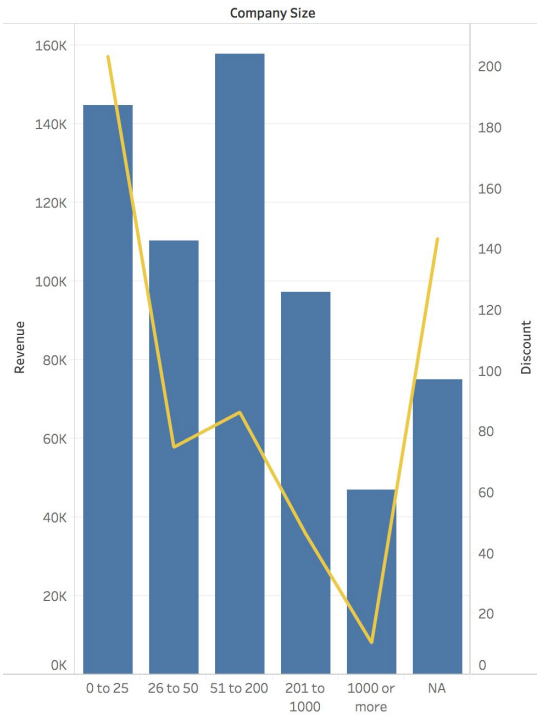
- Company Size: 51 to 200 employees (2.32)
- Companie Size: 26 to 50 employees (2.20)

Initial Strategy: This provides us a baseline that Hubspot should focus its sales workforce and advertising spending on medium-sized businesses that have between 26 to 200 employees.

Brazil



India



We further explored the relationships between company sizes and revenue among the two countries (Brazil and India). This further analysis describes slight differences between the two countries.

- **Brazil**
    - Brazil generates the most revenue from small companies **(0 to 25 employees)**
    - The 2nd and 3rd largest revenue sources are segments "**26 to 50" employees and "51 to 200" employees**, however, the gap with the largest sources is large.
- **India**
    - India generates its largest revenue stream from **"51 to 200" employees and "0 to 25" employees.**
    - India has a more **evenly distributed revenue streams across all company sizes**, unlike Brazil, there is not significant differences in revenue among company sizes.

Another finding from the above graphs is that the discount rates is not proportional to the success of the revenue. This can be seen in India, where the revenues are performing higher than the discount rates issue (for instance, for **26 to 50 employees**). A consideration is that Hubspot can lower discount rates for the **0 to 25 employees** segment and instead issue higher discount rates for the **26 to 50 employees segment**.

Proof 4: Performance by Country

Does the business revenue vary over different countries?

■ Statistical Calculation

- Time range: 2018 Jan-Aug
- Method: Welch Two Sample t-test

```
> with(sales_data, t.test(Revenue ~ Country))

        Welch Two Sample t-test

data:  Revenue by Country
t = -2.8264, df = 1324.5, p-value = 0.004779
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -260.52845  -47.04469
sample estimates:
mean in group Brazil  mean in group India
            379.5106             533.2972
```

- p-value = 0.004779 << 0.05 -> reject the H0
- **Yes**, the mean of business revenue vary over different countries.

Performance by Country (India vs Brazil)

Overall Health Index: 2.08

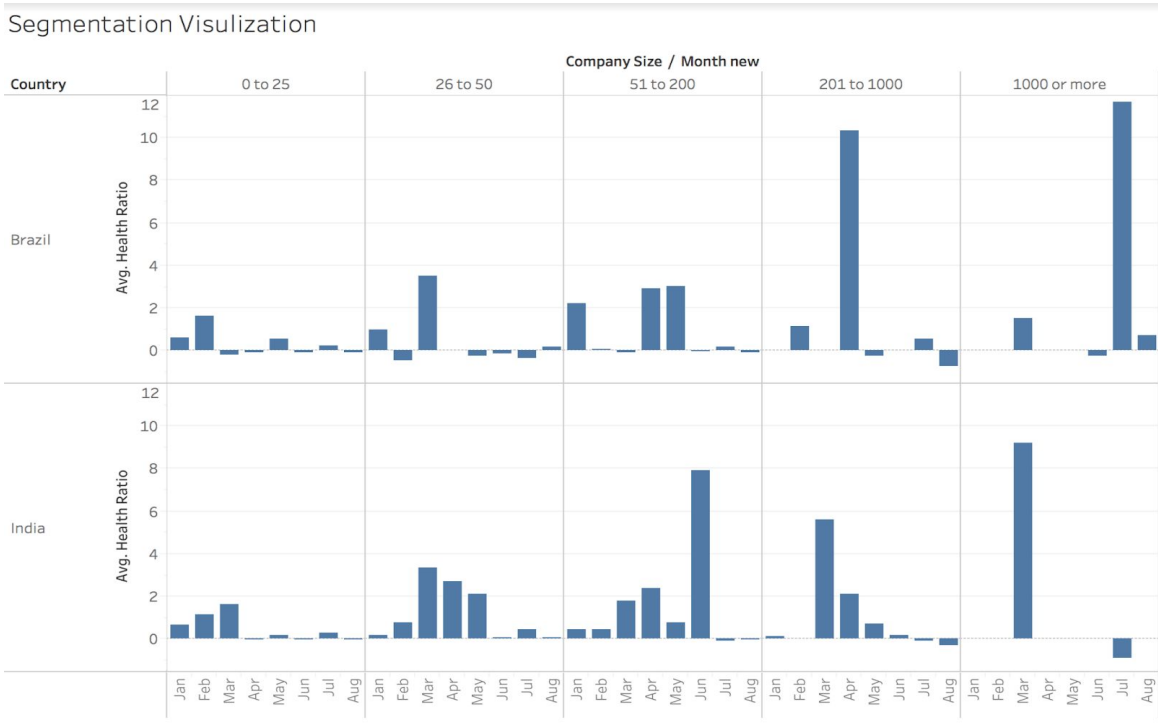
India Health Index: 3.11

Brazil Health Index: 1.50

The overall LTV:CAC ratio of both countries is 2.08 (2018 YTD). India's health index is much healthier at 3.11 compared to Brazil's health index (greater than 2x the index).

	ARPC	Gross M.	Retention	LTV	CAC	INDEX
India	554.09	73%	12.70%	463.87	149.11	**3.11**
Brazil	535.09	61%	-4.51%	312.35	208.67	**1.50**

The table details each element of the LTV:CAC ratio by country (based on the LTV formula). The similar patterns across the following elements: ARPC, Gross Margin, LTV, and CAC (to a certain extent). However, the main difference that is causing the discrepancy in LTV:CAC ratio between the 2 countries is the retention rate. There is a significant gap between India (12.70%) and Brazil (-4.51%). We also find that the CAC for Brazil (208.67) is 40% greater than the CAC for India (149.11), even though Brazil yields a lower LTV. Therefore, this particular finding pertains to a strategy for Brazil to restructure their sales workforce (or even cut sales and advertising spending), while improving the retention for both countries.



Average of Health Ratio for each Month new broken down by Company Size vs. Country.

The above dashboard clearly shows the country's performance by company size and month. Brazil has a few big deals but lack of small deals, which indicates that Brazil's market is not as emerging as that of India. Brazil should focus on the market force and user conversion, and India should focus on getting better deals.

Proof 5: Revenue Patterns by Industry

Does the business revenue vary over different industries?

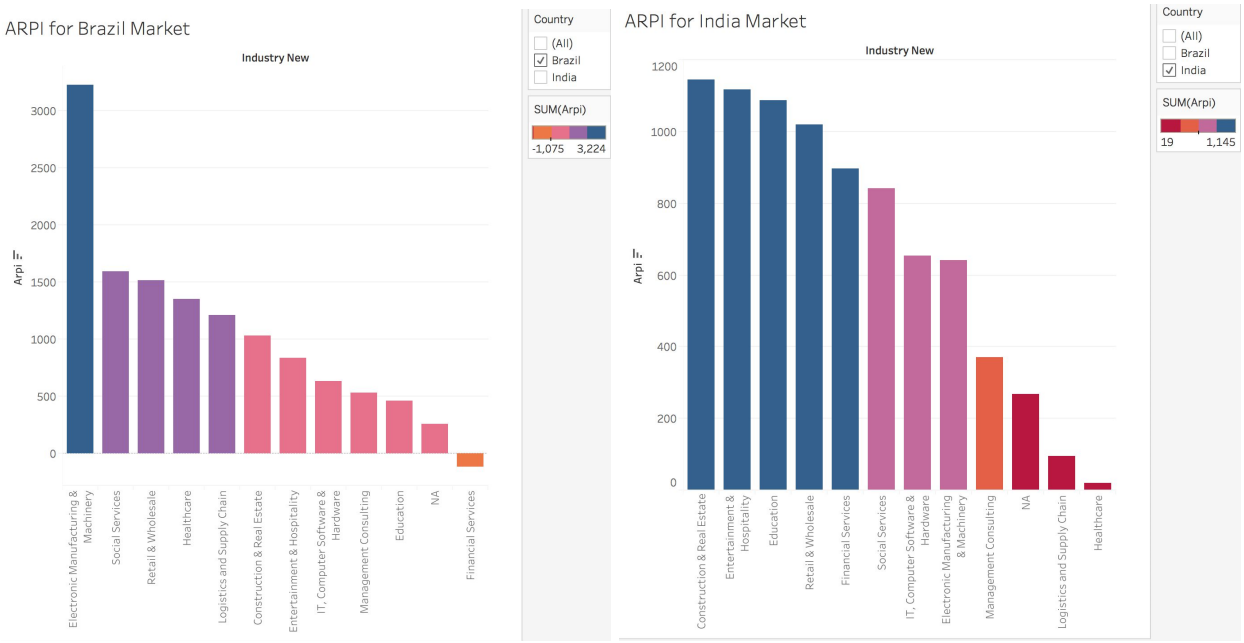
- We categorized industries into 9 categories
- Statistical calculation
  - Time range: 2018 Jan-Aug
  - Method: ANOVA F test

```
> aov.industry = aov(Revenue~Industry_New, data = sales_data)
> summary(aov.industry)
              Df    Sum Sq Mean Sq F value Pr(>F)
Industry_New  10  30056902 3005690     1.7 0.0805 .
Residuals    277 489777070 1768148
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
1147 observations deleted due to missingness
```

  - p-value = 0.0805 > 0.05 -> fail to reject the H0
  - **No**, the business revenues are **NOT** significantly different over industries.

However, when we ran further analyses, we found some industries that are statistically significant on the ratio LTV/CAC: Information Technology and Services (which accounts for the largest proportion), Legal Services, and Testing Laboratories.

Introducing of new attribute, average revenue per industry (ARPI)



The left side is a graph showing the average revenue over industries in emerging markets. Based on the ranking of the ARPI, we could recommend below industries that HubSpot could spend more/ less resources on.

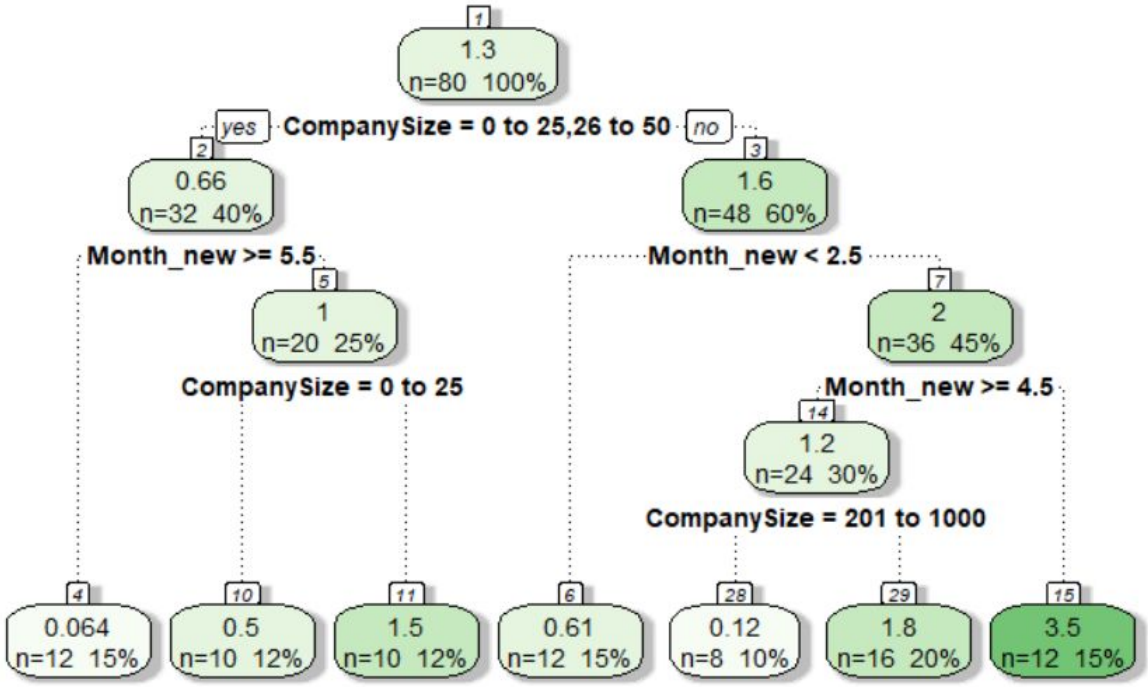
Brazil market

- Highest ARPI: **Electronic Manufacturing & Machinery/ Social Services/ Retail & Wholesale**
- Lowest ARPI: **Financial Service/ Education/ Management Consulting**

India market

- Highest ARPI: **Construction & Real Estate/ Entertainment & Hospitality/ Education**
- Lowest ARPI: **Healthcare/ Logistics and Supply Chain/ Management Consulting**

Appendix



From this decision tree model, we can see that the company size variable at the very root of the tree, and it has the most impact on the health ratio compared to month and country. The threshold happens at 50 employees. If the company size is over 50 employees, it tends to have a higher health ratio to HubSpot than those with less than 50 employees. To moving forward, we recommend sales strategy that focus more companies which have more than 50 employees.

```
Call:
lm(formula = health_ratio ~ ., data = train)

Residuals:
    Min      1Q  Median      3Q     Max
-2.4851 -1.4054 -0.6227  0.0784 12.2444

Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)               0.67382    0.93793   0.718    0.475
CountryIndia              0.12085    0.59748   0.202    0.840
CompanySize1000 or more   1.52396    0.94470   1.613    0.111
CompanySize201 to 1000    1.18108    0.94470   1.250    0.215
CompanySize26 to 50       0.69089    0.94470   0.731    0.467
CompanySize51 to 200      1.27554    0.94470   1.350    0.181
Month_new                -0.09268    0.13038  -0.711    0.479

Residual standard error: 2.672 on 73 degrees of freedom
Multiple R-squared:  0.04963,   Adjusted R-squared:  -0.02849
F-statistic: 0.6353 on 6 and 73 DF,  p-value: 0.7015
```

We attempted to run linear regression model to analyze the impact of different company sizes, countries and time pattern for the LTV:CAC ratio. From the model results above, we find that the largest coefficients impacting the index are company sizes are companies with **1000 or more employees**. However, we see that the companies with **51 to 200 employees** also has a similar impact to the index. By knowing the estimate, we also know how many times the company size segment will impact the ratio; in this case **1.52 for 1000 or more employees companies and 1.27 for 51 to 200 employees companies.** Although we find that the p value is not very significant, we still use the variables with the highest weights/coefficients to guide the strategy.