SalesHero Analysis

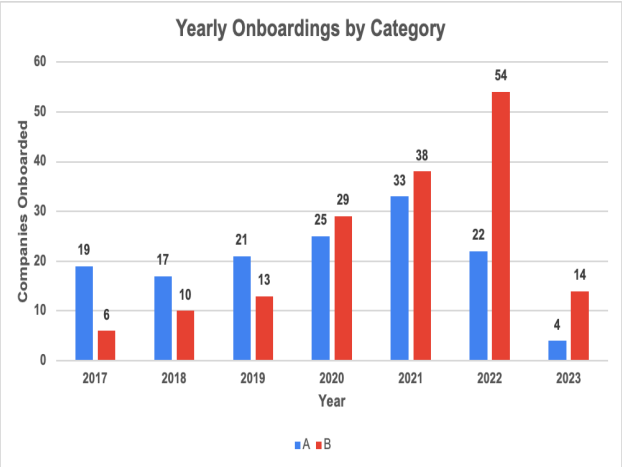
This is a report of an analysis that answers a few business questions about a fictional B2B SaaS company called SalesHero that sells software to help salespersons do their job. SalesHero has two categories of products, namely A and B, and it sells Admin License and SalesRep License with different functionalities. ***CLICK HERE*** ***to see the python workbook of the analysis on Google Colab.***

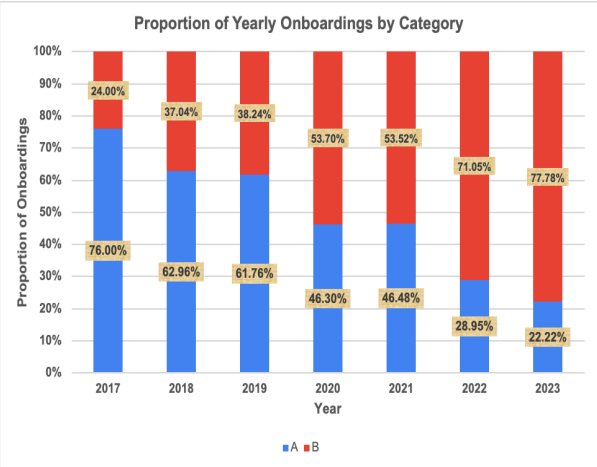
Q1. What is the mix of customer profiles in SalesHero base and is it shifting over time?

Ans:- We can look at the data from multiple perspective,

Onboarding

If we look at the data from an onboarding perspective (i.e. first sign up), then over the years the proportion of companies from **Category B** have been increasing. We are looking at the data from the year 2017 because prior to that, the total count was not high enough for a proper analysis.





As we can see, out of the two categories, the proportion of **Category B** companies in total onboardings has increased from **~24%** in 2017 to **~71%** in 2022. The increase seems to continue in 2023 with the current year numbers showing a **~78%** onboarding of **Category B** companies. In aggregate, the **Category B** companies account for **~54%** of the total onboardings since 2017.

If we look across the industries, the highest proportion of onboarding happens from the **Information Technology and Services** (IT) sector. Because of a small dataset spread across many industries, the Year-on-Year trend jumps around quite a bit. But what seems clear is that

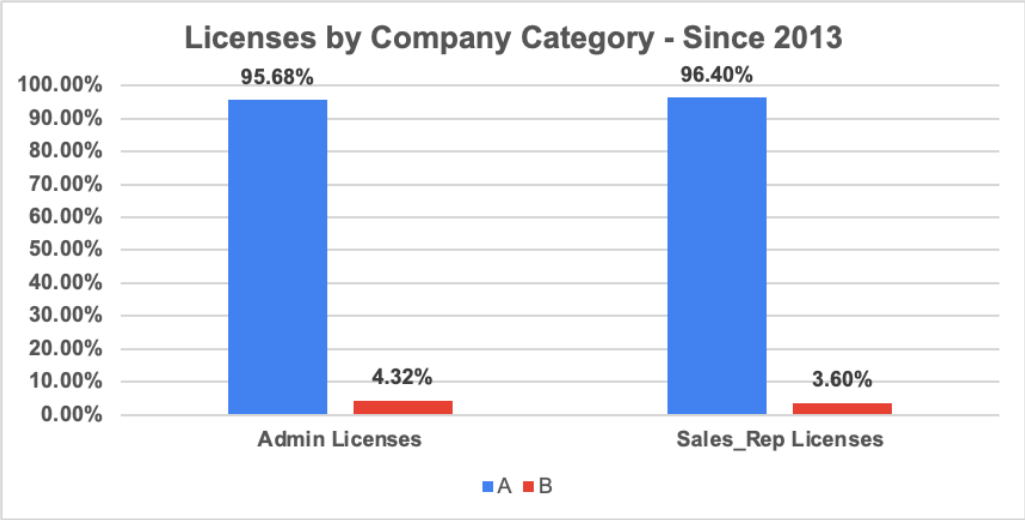
the **IT sector** accounts for **over a half to two-thirds** of the onboardings in any given year. Since 2017, the **IT sector** has accounted for **~52%** of the total onboardings and the **Healthcare sector** stands at second with **~13%** of the onboardings.

Industry	2017	2018	2019	2020	2021	2022	2023	Grand Total
Information Technology and Services	68.00%	14.81%	32.35%	55.56%	49.30%	65.79%	66.67%	**52.13%**
Healthcare	16.00%	14.81%	29.41%	9.26%	9.86%	7.89%	16.67%	**12.79%**
Finance	4.00%	18.52%	5.88%	5.56%	8.45%	7.89%	5.56%	**7.87%**
Communications	4.00%	7.41%	5.88%	3.70%	4.23%	6.58%	5.56%	**5.25%**
Professional Services	4.00%	7.41%	5.88%	5.56%	5.63%	3.95%	0.00%	**4.92%**
Education & Training	4.00%	0.00%	2.94%	1.85%	8.45%	2.63%	5.56%	**3.93%**
Retail	0.00%	7.41%	8.82%	1.85%	2.82%	1.32%	0.00%	**2.95%**
Manufacturing	0.00%	7.41%	2.94%	5.56%	2.82%	0.00%	0.00%	**2.62%**
Real Estate	0.00%	3.70%	0.00%	7.41%	1.41%	2.63%	0.00%	**2.62%**
CPG	0.00%	14.81%	2.94%	3.70%	0.00%	0.00%	0.00%	**2.30%**
Travel & Tourism	0.00%	0.00%	0.00%	0.00%	2.82%	1.32%	0.00%	**0.98%**
Agriculture & Forestry	0.00%	3.70%	0.00%	0.00%	0.00%	0.00%	0.00%	**0.33%**
Automobile	0.00%	0.00%	0.00%	0.00%	1.41%	0.00%	0.00%	**0.33%**
Automotive	0.00%	0.00%	0.00%	0.00%	1.41%	0.00%	0.00%	**0.33%**
NGO	0.00%	0.00%	0.00%	0.00%	1.41%	0.00%	0.00%	**0.33%**
Transportation	0.00%	0.00%	2.94%	0.00%	0.00%	0.00%	0.00%	**0.33%**
Grand Total	**100.00%**	**100.00%**	**100.00%**	**100.00%**	**100.00%**	**100.00%**	**100.00%**	**100.00%**

So the **IT sector** companies and **Category B** companies account for the major chunk of the total onboardings, but their trends have been different. While Category-wise the contribution of **Category B** companies have increased over the years, Industry-wise the **IT sector** has remained as the major contributor to the onboardings since 2017.

License Sale

Now if we look at the data from the License Sale perspective, then the **Category A** companies dominate the numbers as they account for **~96%** of the total **Admin Licenses** and **~96%** of the total **Sales-Representative Licenses** since 2013.



Industry-wise, with **~64%** of the **Admin Licenses** and **~62%** of the **Sales-Rep Licenses**, the **IT sector** accounts for the majority of Licenses sold since 2013.

Industry	**Admin Licenses (Since 2013)**	**Sales_Rep Licenses (Since 2013)**
Information Technology and Services	64.19%	62.41%
Healthcare	18.34%	18.70%
Retail	4.76%	6.02%
Finance	4.87%	4.35%
Communications	2.57%	3.35%
CPG	1.89%	1.79%
Manufacturing	1.68%	1.62%
Professional Services	0.74%	0.73%
Transportation	0.37%	0.54%
Education & Training	0.29%	0.22%
Real Estate	0.17%	0.13%
Agriculture & Forestry	0.05%	0.07%
Automotive	0.03%	0.04%
Travel & Tourism	0.03%	0.02%

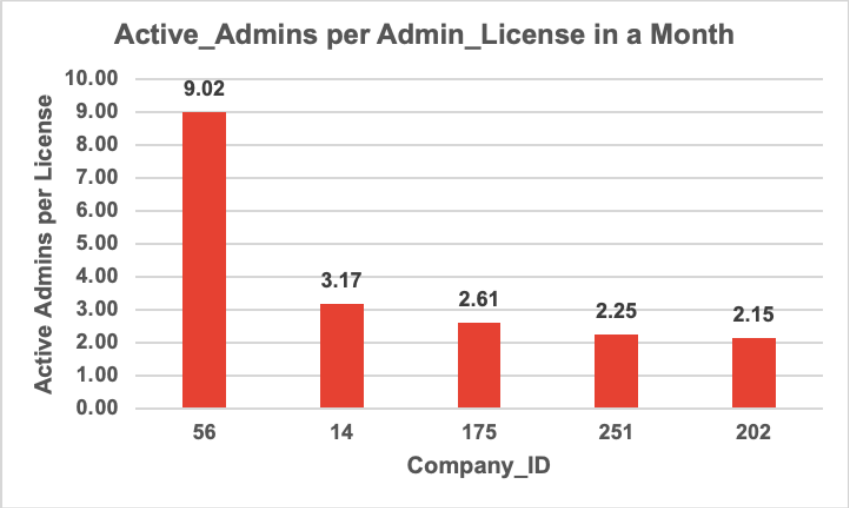
Automobile	0.01%	0.01%
NGO	0.02%	0.01%
Grand Total	**100.00%**	**100.00%**

Q2. Which customers are most utilizing the licenses they bought and is there anything common amongst them?

Ans:- We can look at how many courses were created per Admin License and how many courses were completed per Sales_Represnetative License. But this metric does not necessarily give the utility of licenses bought because it could be the case that just a handful of Admins are creating the majority of courses in the company, and similarly just a few Sales_Reps are completing the majority of the courses.

So a better measure of license use might be **Active Users per License per Month**.

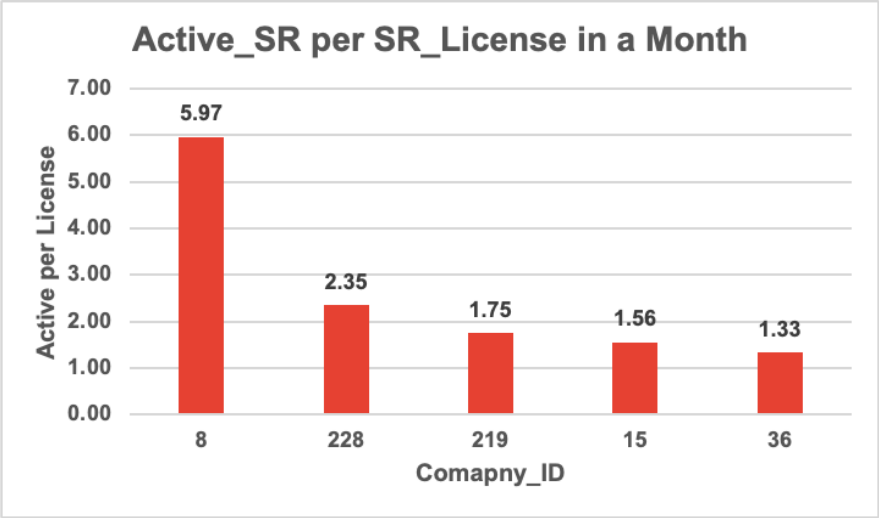
The data shows the top 5 companies that have the highest **Active Admin Users per Admin License per Month**.



Common features among them are,

1. All of the companies have greater than 1 ratio, i.e. the Licenses are being shared by multiple Admins in the company.

Now, below the data shows the top 5 companies that have the highest **Active Sales_Representative Users per Sales_Representative License per Month**.

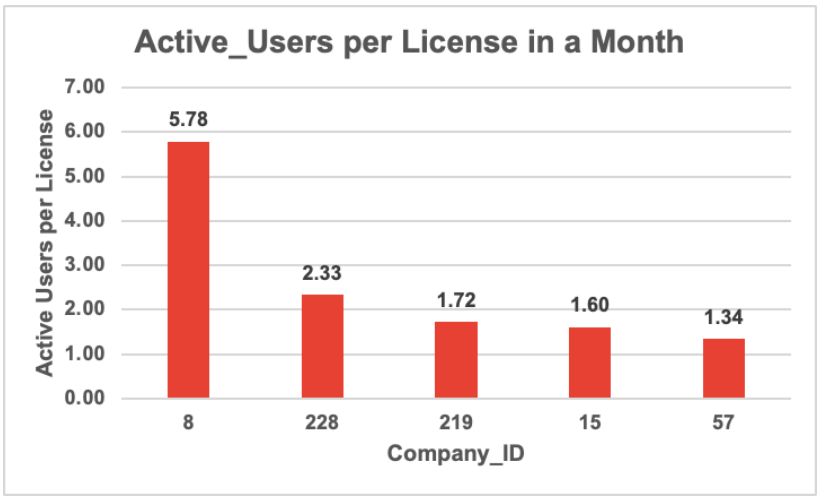


Common features among them are,

1. All 5 of them belong to Category A. But this is because, except for 1 company, all the rest bought licenses during the years when most of the companies that signed up were from Category A. So the base rate of Category A was high.
2. All of the companies have greater than 1 ratio, i.e. the Licenses are being shared by multiple Sales_Representatives in the company.

Now we can look at **Active Users per License per Month** by combining the Admins and Sales_Rep data. This gives an overall utilization metric because a company can have a few Admins but many Sales_Rep or vice versa. So looking at the complete user base might be useful.

Here are the top 5 companies with highest **Active Users per License per Month**,



We can see that 4 out 5 companies in the top 5 are the same as the ones in Active Sales_Rep ranking. This is because, out of the total Licenses sold, most of the Licenses are for Sales_Rep.

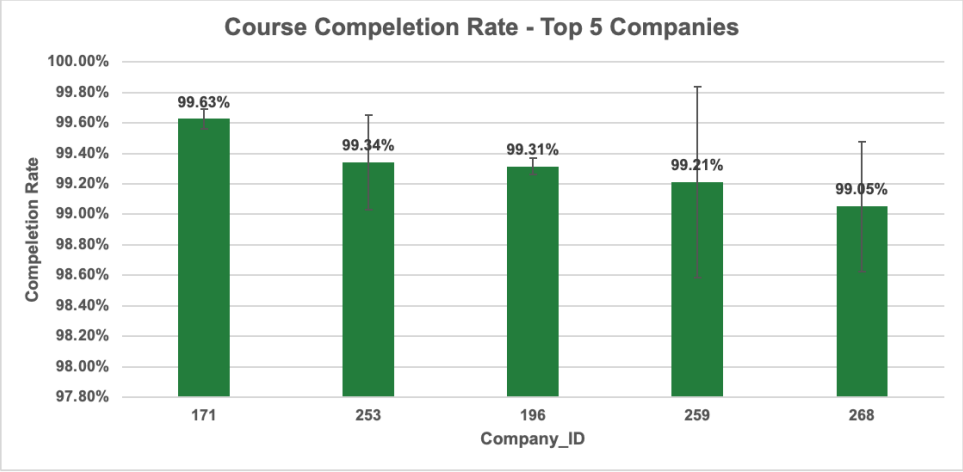
Q3. The end-users of which customers are most and least engaged?

Ans:- To check the most and least engaged end-users, we can measure the **Course Completion Rate** of the end-users for each customer, i.e. company. The **Course Completion Rate** will be calculated as **(completions/starts)*100**.

Note that since we are talking about measuring engagement, we will not be looking at how many end-users actually start the course after getting an invite. We are only trying to measure how engaged the end-users are after they start the course.

The **Company_ID 171** seems to have the end-users that are most engaged as their **Course Completion Rate** is **~99.63%** with **~42.5k** courses started. There are 4 companies that had a **Completion Rate** of 100%. But we can ignore them as their course start count is quite low (<50). So we should have the most confidence in considering the **Company_ID 171** as the customer with the most engaged end-users. **Company_ID 171** is a **Category A** company from the **Transportation** Industry.

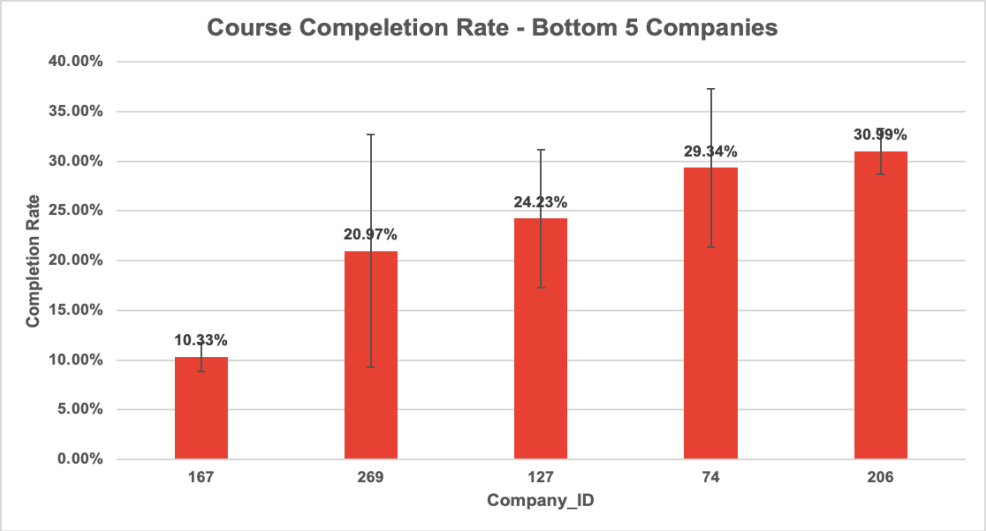
The chart below shows the top 5 companies based on the **Course Completion Rate**,



Note that although **Company_ID 171** has the highest **Completion Rate**, the **Error Bars** based on the 95% Confidence Intervals of a t-distribution show that the **Company_ID 253** and **259** could also have the ***true*** **Completion Rate** as high or higher than **Company_ID 171**.

At the other end of the rankings, the **Company_ID 167** seems to have the end-users that are least engaged as their **Course Completion Rate** is **~10.33%** with **~2.25k** courses started. This company is a **Category A** company from the **Healthcare** industry.

The chart below shows the bottom 5 companies based on the **Course Completion Rate**,



Again note that although **Company_ID 167** has the lowest **Completion Rate**, the **Error Bars** based on the 95% Confidence Intervals of a t-distribution show that the **Company_ID 269** could also have the ***true*** **Completion Rate** as low or lower than the **Company_ID 167**.

We have only looked at the **Completion Rates** without considering the years. It could be the case that different years had different UI/UX experience for the end-users, thus potentially making the comparison slightly misleading.

So below is a list showing companies with the highest **Completion Rates** each year (after removing course counts < 50),

customer_sign_year	company_id	completion_rate
2013	267	93.01%
2014	228	72.86%
2015	268	99.05%
2016	171	99.63%
2017	253	99.34%
2018	259	99.21%

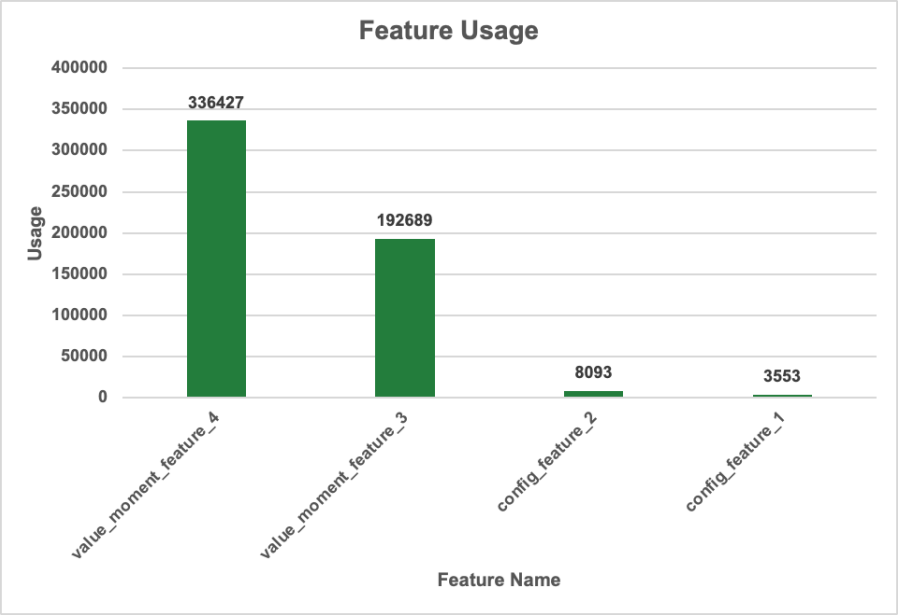
2019	196	99.31%
2020	202	97.48%
2021	90	97.55%
2022	297	97.18%
2023	311	98.19%

And here is a list showing companies with the lowest **Completion Rates** each year (after removing course counts < 50),

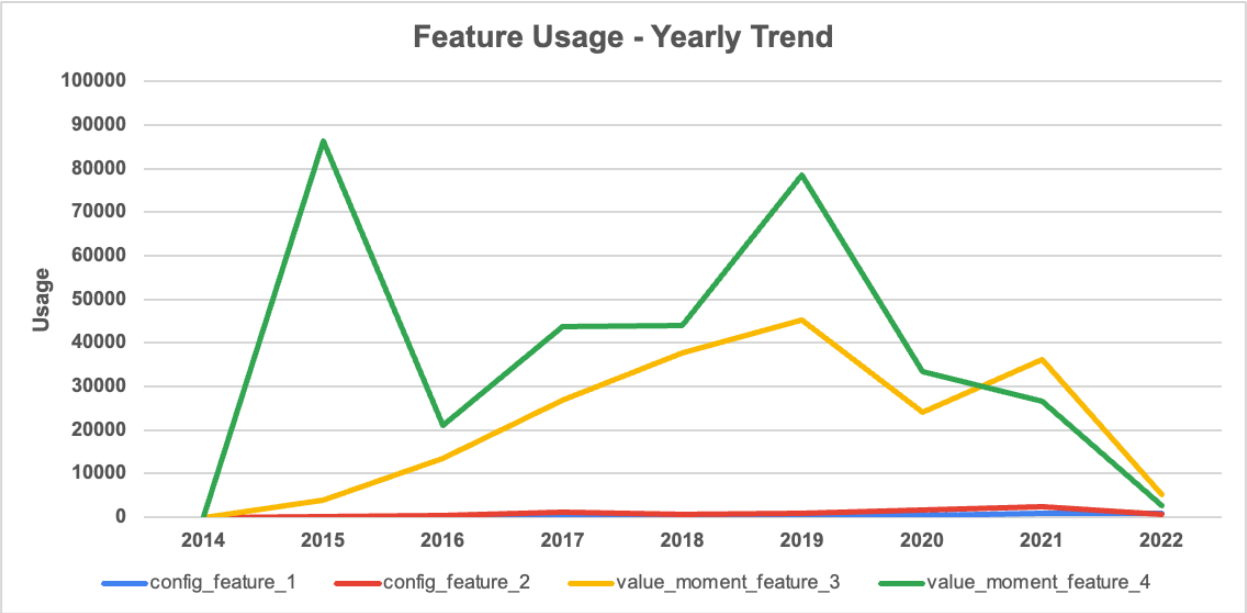
customer_sign_year	company_id	completion_rate
2013	267	93.01%
2014	228	72.86%
2015	74	29.34%
2016	66	54.69%
2017	250	55.06%
2018	110	36.19%
2019	175	41.54%
2020	104	37.16%
2021	167	10.33%
2022	127	24.23%
2023	311	98.19%

Q4. Which are the most and least used features?

Ans:- If we look at the data from inception, the most used feature is the **value_moment_feature_4** with a usage of ~3.4 lakhs (~62% of the total), and the least used feature is **config_feature_1** with a usage of ~3.5k times (~0.7% of the total). But the data shows that config_features get used much less often than value_moment_features. I am assuming this is because config_features are used by the Admins while creating courses, and value_moment_features are used by the Sales_Rep while attending courses. Since there are a much higher number of Sales_Rep than Admins, this explains the drastic discrepancy between the usage of config_features and value_moment_features



We can also look at yearly trends of feature usage. Since 2023 is not finished, we can look at the trend till 2022. We can see that still the **value_moment_feature_4** dominates most of the years. From mid of 2020 the trends for **value_moment_feature_3** and **value_moment_feature_4** crossed over, making **value_moment_feature_3** go above **value_moment_feature_4** in usage. Also it is worth noting that the usage counts dropped drastically in 2022. It could be a data issue or it could be because the usage has actually gone down due to some product changes.



Q5. What are the top 3-5 recommendations based on all of the above insights?

Ans:- A few recommendations are as follows,

I. In keeping with the principle of diversification, we should try to get more companies to onboard from industries other than IT. The IT sector has accounted for over a half (~52%) of the total onboardings since 2017. In case of a slowdown in the IT sector, we could face some slowdown as well. By getting more companies from other industries, like Retail or Finance, we can create a balance of portfolio that acts against any bad impact from any individual sector.

II. Some of the companies are utilizing a few licenses for multiple users. For example, the **Company_ID 56** has **9.02 Active Admin Users per Admin License per Month,** meaning on average ~9 Admins are sharing just one license in a month. If we develop a system to restrict how many users can use a single license (let's say 2), then maybe we can increase the license sale by making the companies buy more licenses to keep up with their needs.

III. We should try to improve the usability of the **config_feature_1** within the config_features and **value_moment_feature_3** within the value_moment_features. Improving the usability of the less used features may lead to engagement for the users, thereby leading to higher customer satisfaction. It is to be noted that the metric to improve should be something like **features_used per course completed** along with **satisfaction_scores**. If we just try to increase the proportion of, say **config_feature_1**, in total feature usage, it might indicate that the usage of other features have gone down.

END