

# Case Study Report

## Background / Introduction

Supahands is a data labeling company in the AI/ML industry. Recently, Supahands is testing out the launch of DataBolt, a brand-new platform that will enable the company to widen its product offerings. So that the company will need only the best of the SupaAgents working in the current platform to perform the labeling for DataBolt.

## Objective(s)

- To generate insight from the given data for helping the SupaAgent choice
- To analyze the insight and narrow down the top 50 from 1,962 SupaAgents to be invited to DataBolt.

## Methodology

The result of this study is based on those parameters:

1. Productive workers in **Payment per Month** value
2. **Total payout** per SupaAgent
3. **Number of unique projects** as tiebreaker.
4. **Duration** on joining as SupaAgents

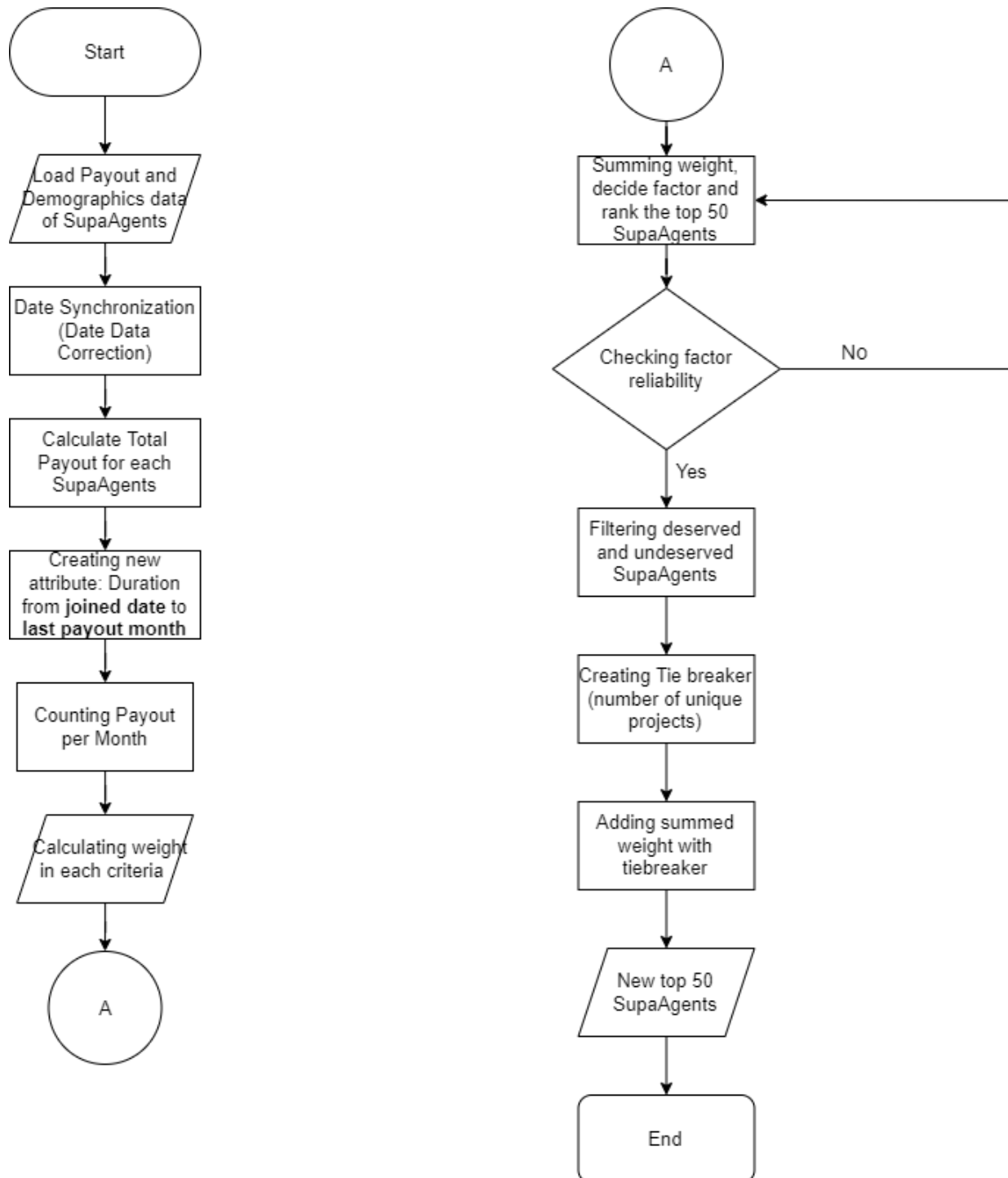
This study has several limitations and assumptions stated below:

1. **Payout per Month (PpM)** metric is generated from the **Total Payout** of each agent per “**duration**” as SupaAgent. “**Duration**” as SupaAgent is calculated from **Date Joined** to the **last payout month** (assumed to 31 December 2021).
2. PpM metric is assumed to measure **SupaAgent productivity**
3. **Payout** is assumed linear to the **difficulty (quality)** or **workload (quantity)** of the project (more payment received means more difficult or much more projects done).
4. **Productivity (PpM)** is considered **more important** than **difficulty/quantity** (Total Payout). I assume **quantity** is linear with **capability** and **experience** while **productivity** shows **how quick the works can be done**. Thus, the reasons behind the preference are:

- a. As the slight information I have got, there is an assessment test on the SupaAgent recruitment. So that the capability of the soon-to-be SupaAgents have been tested there. **We already have our capable members**; all we need to do is **seeking the better one in productive terms**.
  - b. Because SupaAgent is a freelance type of work (not a full-time workers) so that their **experience** and **duration on joining** as SupaAgent is almost **negligible**.
5. **Age** factor is also **insignificant** as it is assumed that the top 50 is still in freelance work (not a full-time). If they would be recruited as a **full-time employee**, then **younger** SupaAgents are **more loved** as they still got so many years on their career.
  6. All projects is assumed giving the same type of work. So that the **number of unique projects** is not highly considerable. But it is used as a tiebreaker if previous calculation gives tie result.
  7. **Duration** is not calculated in weight because it is found that the trend is similar to the **Total Payout**. But duration under 2 months is considered **too early** so that duration is used as a parameter after the weight ranking.
  8. PpM Metric is calculated by this formula:
 
$$\text{Payout per Month} = \frac{\text{Total Payout of each SupaAgent}}{\text{Duration}}$$
  9. Duration is calculated by this formula:
 
$$\text{Duration} = \text{Joining Date} - \text{Last Payout Date}$$
  10. All PPM Factor, Payout Factor, Duration Factor are calculated by this formula:
 
$$r_{ij} = \frac{x_{ij}}{\text{Max}(x_{ij})}$$
  11. The tiebreaker(T) is calculated by ranking the number of unique project (R). Then the calculation of the tiebreaker is done as below:
 
$$x = \frac{R}{1000}$$

$$T = 1 - x$$
  12. So that the highest rank of the R will get biggest T. As if there is a tie, this T factors will be added and break the tie.
  13. The ranking is made based on the factor value multiplied with each of their weight. PPM weight is given at 62.5%, Payout weight at 37.5%, and the Duration at 0%.

Here is included the flowchart of steps taken during the process



**Results / Impact / Outcome**

Here is provided the top 50 agents based on the made analysis.

Rank	SupaAgent ID	Nationality	Education	Occupation	Birth date	Age	AgeGrouping	Date Joined
1	X1VSD	Philippines	Bachelor's Degree	Customer Service	12/24/1982	39	35-44	8/27/2019
2	X1XF0	Philippines	Bachelor's Degree	Prefer not to say	01/02/1989	33	25-34	10/11/2019
3	X1ZEF	Philippines	Bachelor's Degree	Manager	11/22/1979	42	35-44	12/9/2019
4	X8HON36	Philippines	Bachelor's Degree	Teacher/Educator	04/04/1997	25	25-34	8/9/2021
5	X2914E	Philippines	Bachelor's Degree	Prefer not to say	10/05/1997	24	Under 25	10/20/2020
6	X2AE7	Indonesia	Bachelor's Degree	Business owner	05/08/1991	31	25-34	10/26/2020
7	X2914J	Philippines	Bachelor's Degree	Prefer not to say	04/15/1996	26	25-34	10/20/2020
8	XLUPGOT	Philippines	Bachelor's Degree	Prefer not to say	11/08/1996	25	25-34	8/2/2021
9	X1X7E	Malaysia	High School or equivalent	Student	05/10/1996	26	25-34	10/4/2019
10	X4KPZ5L	Philippines	Bachelor's Degree	Teacher/Educator	05/29/1976	46	45 and above	8/26/2021
11	X1KT7	Philippines	Bachelor's Degree	Others, please specify:	11/28/1997	24	Under 25	9/20/2018
12	X1SIX	Philippines	Bachelor's Degree	Teacher/Educator	11/20/1992	29	25-34	5/23/2019
13	X3540	Philippines	Bachelor's Degree	Business owner	05/06/1981	41	35-44	8/23/2018
14	X1R3Y	Philippines	Bachelor's Degree	Prefer not to say	06/12/1984	38	35-44	4/3/2019
15	X1Y72	Philippines	Bachelor's Degree	Homemaker / Stay-at-Home Parent	08/26/1987	34	25-34	10/24/2019
16	X2AQ4	Philippines	Others, please specify:	Prefer not to say	10/12/1990	31	25-34	11/3/2020
17	X1V19J	Philippines	Bachelor's Degree	None	04/03/1984	38	35-44	8/30/2019

18	X1X9T	Philippines	High School or equivalent	None	08/10/1997	24	Under 25	9/30/2019
19	X2C1AQ	Indonesia	Bachelor's Degree	Others, please specify:	09/12/1993	28	25-34	1/28/2021
20	X2756	Malaysia	Bachelor's Degree	Accountant	07/27/1989	32	25-34	2/6/2018
21	XGTZH3H	Philippines	Others, please specify:	Others, please specify:	06/21/1998	24	Under 25	3/26/2021
22	X20PB	Philippines	Bachelor's Degree	Healthcare professional	09/12/1977	44	35-44	1/14/2020
23	XPKPQXT	Philippines	Bachelor's Degree	Others, please specify:	09/17/1991	30	25-34	5/10/2021
24	X1X52	Malaysia	Bachelor's Degree	Student	11/22/1996	25	25-34	9/25/2019
25	X2DDD	Philippines	Diploma or equivalent	Others, please specify:	01/25/1991	31	25-34	1/11/2021
26	X1Y4KGG	India	Bachelor's Degree	Engineer	03/06/1992	30	25-34	3/25/2021
27	X1663	Malaysia	Bachelor's Degree	None	12/07/1977	44	35-44	3/23/2017
28	XISLHII	Philippines	Bachelor's Degree	Engineer	12/30/1996	25	25-34	6/25/2021
29	XHL3YWD	0	0	0	01/01/1900	122	Invalid Age	10/15/2021
30	X20CK	Philippines	High School or equivalent	Homemaker / Stay-at-Home Parent	07/26/1985	36	35-44	1/8/2020
31	X1752	Philippines	Bachelor's Degree	Prefer not to say	01/31/1986	36	35-44	3/27/2017
32	X2CIO	India	High School or equivalent	Student	04/20/1996	26	25-34	12/21/2020
33	X24BL	Philippines	Bachelor's Degree	Sales	10/07/1986	35	35-44	5/12/2020
34	XWD2YS6	Philippines	Prefer not to say	Student	09/06/2001	20	Under 25	9/24/2021
35	XHI1K3I	Philippines	Bachelor's Degree	Prefer not to say	04/23/1997	25	25-34	8/16/2021
36	X2C16E	Philippines	Bachelor's Degree	None	05/07/1998	24	Under 25	1/21/2021
37	X2TUOAM	Philippines	Others, please specify:	Student	01/17/1998	24	Under 25	10/11/2021
38	X2917	Philippines	Bachelor's Degree	None	04/12/1990	32	25-34	3/19/2018

39	X1Q4J	Malaysia	Master's Degree	None	08/22/1988	33	25-34	3/4/2019
40	X1U93	Malaysia	Bachelor's Degree	Homemaker / Stay-at-Home Parent	04/26/1967	55	45 and above	7/10/2019
41	X27X9	Philippines	Master's Degree	Teacher/Educator	03/29/1982	40	35-44	8/19/2020
42	X3523	Philippines	Bachelor's Degree	Business owner	04/18/1978	44	35-44	8/10/2018
43	XOS5TFE	Philippines	High School or equivalent	Prefer not to say	01/28/2003	19	Under 25	10/13/2021
44	X2PPBXT	0	0	0	01/01/1900	122	Invalid Age	4/21/2021
45	X2BZR	Philippines	Bachelor's Degree	Others, please specify:	04/02/1974	48	45 and above	12/9/2020
46	X29JS	Philippines	Diploma or equivalent	Others, please specify:	11/30/1980	41	35-44	10/6/2020
47	X1ZHE	Philippines	Others, please specify:	Clerk/Secretary	11/06/1973	48	45 and above	12/13/2019
48	X28CT	Philippines	Master's Degree	Teacher/Educator	12/19/1989	32	25-34	9/14/2020
49	X1WRX	Philippines	Bachelor's Degree	Others, please specify:	03/18/1984	38	35-44	9/6/2019
50	XIF1XVY	Philippines	Bachelor's Degree	Customer Service	05/07/1987	35	35-44	3/16/2021

## Discussion

As mentioned in methodology, the ranking is made by those formulas. To check the weight giving reliability, I did 2 comparison on several hyphotesis:

1. Comparing the rank with 66-33 weight of PpM-Total Payout and 62.5-37.5 weight

I ranked top 50 SupaAgents with different weight on the factor. First the weight of Total Payout is given 0.333 and then it is compared with the one with 0.375 weight. Here it is found 1 differences as stated below:

Inside Weight of 0.333 <b>but</b> outside of 0.375						
SupaAgent ID	Total Payout	Duration(Month)	PayoutPerMonth	PPM Rank	0.333 Rank	0.375 Rank
XFJ936M	59.44	0.833333333	71.328	33	47	51
Inside Weight of 0.375 <b>but</b> outside of 0.333						
SupaAgent ID	Total Payout	Duration(Month)	PayoutPerMonth	PPM Rank	0.333 Rank	0.375 Rank
X1WRX	1306.19	28.23333333	46.264	49	51	49

Comparing both agent, I prefer "**X1WRX**" rather than the "**XFJ936M**". Even though **XFJ936M** getting more PpM, its duration is not even a month yet. Also **X1WRX** PPM Rank is still under the top 50 and its total payout is much more than **XFJ936M** whose total payout is even below its own PpM. So in this comparison, **X1WRX** is more preferred than **XFJ936M**. Also by this comparison, i choose **62.5-37.5** as my weight of PpM-TotalPayout as its filtering the less preferred agent.

2. Comparing the deserved and undeserved agent.

Because of the ranking are mainly determined by PpM, I was comparing **chosen agent** with **PpM below top 50 PpM rank** and **unchoosen agent** with **PpM above top 50 PpM rank** as shown below

Possibly Deserve (Weighted Rank under 50 <b>but</b> PpM Rank above 50)						
SupaAgent ID	Total Payout	Duration(Month)	PayoutPerMonth	PPM Rank	Weighted Rank	Switch Priority
XFJ936M	59.44	0.833333333	71.328	33	51	3
XHI1K3I	276.63	4.566666667	60.576	40	56	2
X29JS	765.48	15.03333333	50.919	42	54	1
X27X9	807.91	16.63333333	48.572	47	57	1
X28CT	724.27	15.76666667	45.937	50	62	1
X9BO4YQ	75.32	1.533333333	49.122	45	70	3

Possibly Undeserve (Weighted Rank above 50 <b>but</b> PpM Rank under 50)						
SupaAgent ID	Total Payout	Duration(Month)	PayoutPerMonth	PPM Rank	Weighted Rank	Switch Priority
X1752	2539.25	58	43.780	53	31	5
X1837	2478.02	57.5	43.096	54	35	4
X3523	1860.33	41.3	45.044	51	42	6
X2247	2163.82	54.7	39.558	62	41	2
X2956	1917.58	50.26666667	38.148	63	46	1
X1J191	1669.16	40.7	41.011	59	48	3

More in priority
Medium
Less in priority

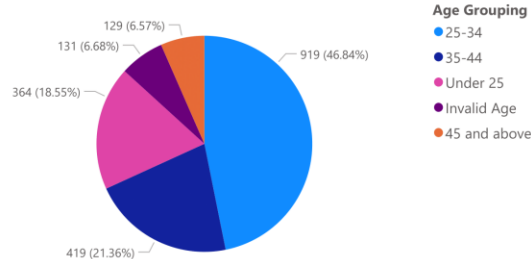
As we're seeking the most productive one in terms of PpM, every of undeserved agent listed above are actually can be switched with the one inside the deserve agent list.

But there is further analysis on the possibly deserve as describe on methodology number 7. From those six agents, I decide "XFJ936M" and "X9BO4YQ" are still undeserved because of the duration of the joining agent is still too short (less than 2 months). The other 4 agents have more than 2 months, three of those are more than a year and categorized more deserve agents.

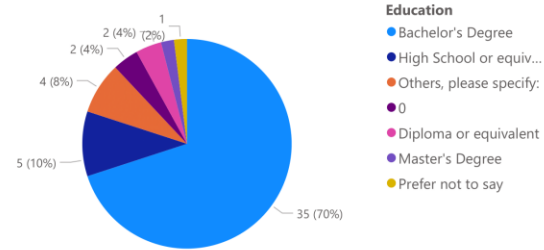
So, it is decided to take the 4 out of 6 possibly deserve agent as they are all have a good duration on joining as SupaAgent. So that the top 50 from the weighing is revised with this analysis by changing 4 undeserved agents with the 4 deserved one (Indicated with the green and yellow fill).



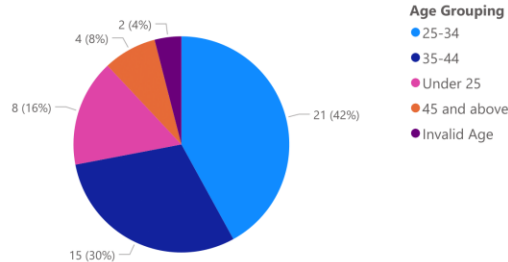
(ALL) Count of SupaAgent ID by Age Grouping



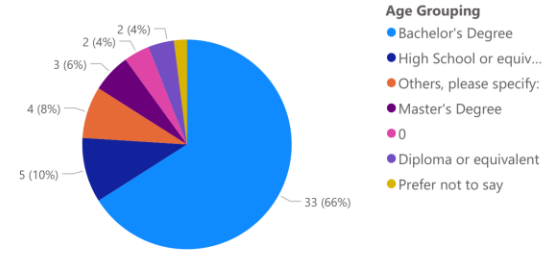
(ALL) Count of SupaAgent ID by Education



(Top 50) Count of SupaAgent ID by Age Grouping

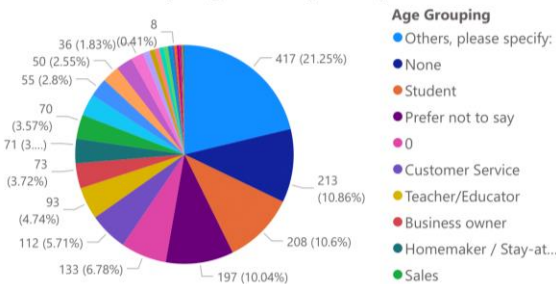


(Top 50) Count of SupaAgent ID by Education

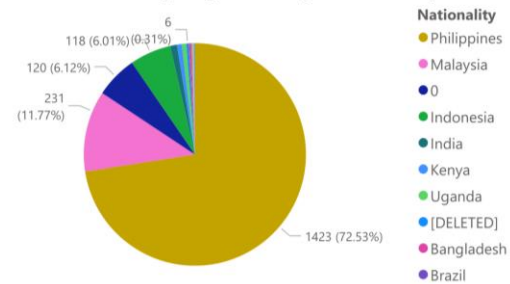


The Top 50 SupaAgents' education and age distribution are similar to the entire dataset distribution. The age for the Top 50 are ranging mostly in the age of 25-34 years. The education of the Top 50 SupaAgents mostly are Bachelor's Degree and the rest is almost equally distributed on others education.

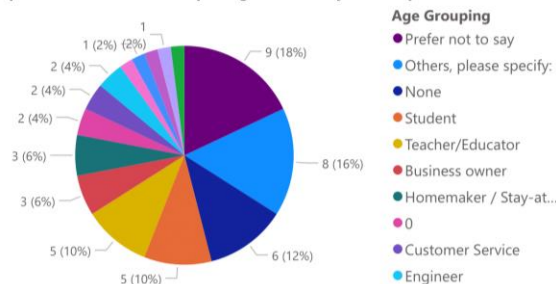
(ALL) Count of SupaAgent ID by Occupation



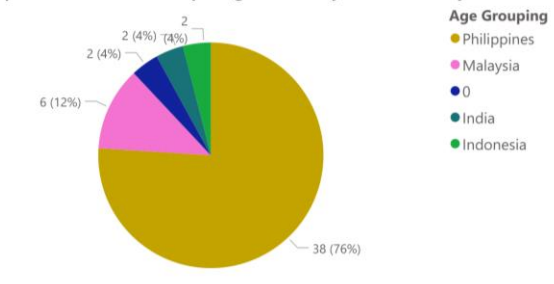
(ALL) Count of SupaAgent ID by Nationality



(Top 50) Count of SupaAgent ID by Occupation



(Top 50) Count of SupaAgent ID by Nationality



The Top 50 SupaAgents' occupation and nationality distribution are also similar to the entire dataset. Occupations are distributed mostly on the "Prefer not to say", "Others", "None", "Student". I assume that people who register as SupaAgents are the unemployed people and seeking for income. Nationality of the Top 50 Agents is mostly distributed at "Philippines" as the entire dataset is also showing the same distribution.

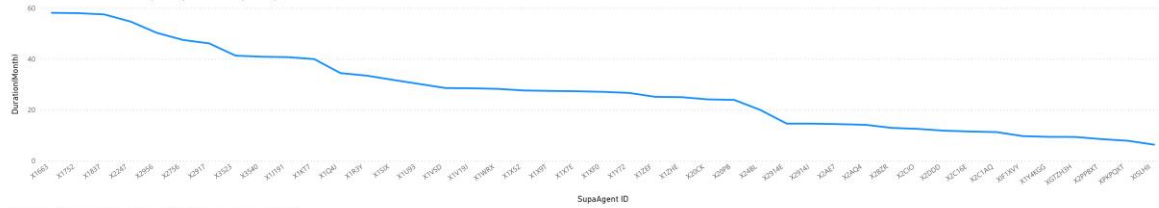
On the graph below also shows the relation between the duration and the total payout when it is sorted by duration and total payout descendingly.

Moreover, collected insights are included in the attached Excel and PDF File. PDF File shows the report made from PowerBI as the free user cannot share it. Excel file can be accessed here on shared Google Drive :

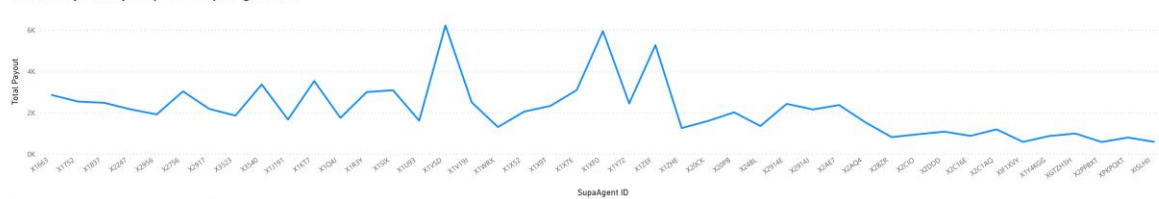
<https://docs.google.com/spreadsheets/d/1PqLCU9kbSZvXUd21FbwzJyEQrYfP398I/edit?usp=sharing&oid=109590770051333325362&rtpof=true&sd=true>

## Relation between Duration, Total Payout and Payout Per Month by Top 50 SupaAgents (Sorted by Duration)

Duration(Month) by Top 50 SupaAgent ID



Total Payout by Top 50 SupaAgent ID

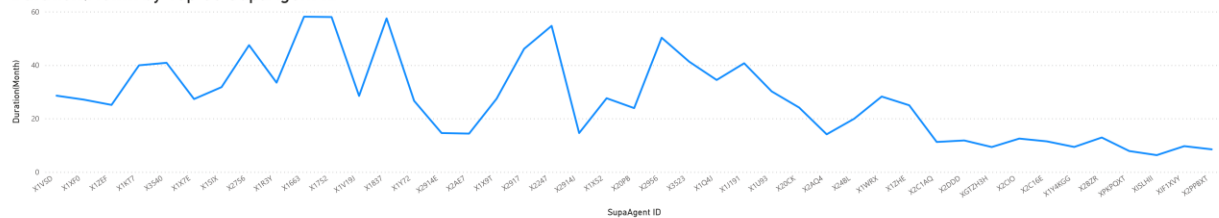


PayoutPerMonth by Top 50 SupaAgent ID

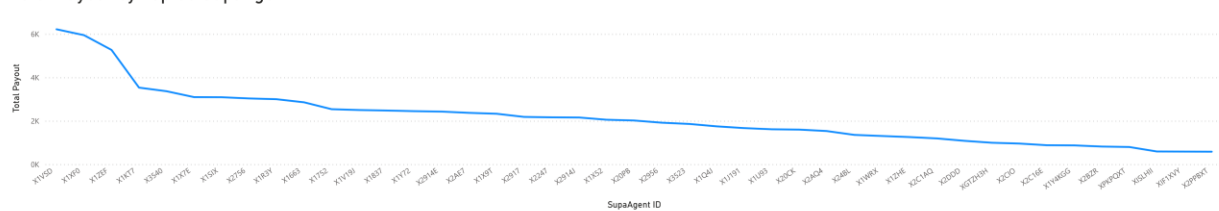


## Relation between Duration, Total Payout and Payout Per Month by Top 50 SupaAgents (Sorted by Total Payout)

Duration(Month) by Top 50 SupaAgent ID



Total Payout by Top 50 SupaAgent ID



PayoutPerMonth by Top 50 SupaAgent ID

