

## KADEN DATA ANALYST ASSESSMENT

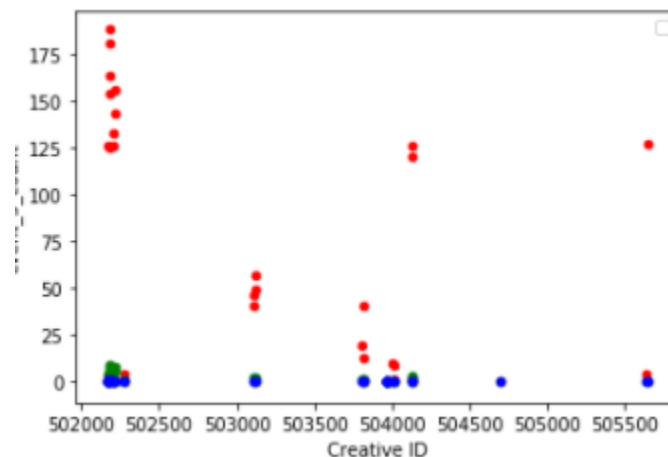
GitHub repository containing code used to perform analysis:

<https://github.com/LukaGrujicic/Kaden-Data-test/blob/master/Kaden%20Data%20Test.ipynb>

- The objective from the Advertiser is to obtain more app installs (and post install events as secondary (Events 1-5 in the report)). Using the report attached in the email and definitions as guidance, how would you analyze / categorize the data?  
Note: The goal of this campaign is to achieve a low CPI (Cost per install).
- Need to see and understand which creative IDs are drawing in low eCPI and on which platforms/Softwares. This is key to understanding the demographic interacting with the ad being shown (iOS vs Android, Tablet vs Phone).
  - We need to figure out which creative IDs generate the highest number of post install events. Using the describe() function Python gives us a quick overlook at the data so we can start making analytical decisions

	index	Creative ID	Wins	Win Rate	Cost	Clicks	Conversions	Click Rate	event_1_count	event_2_count
count	323823.000000	323823.000000	323823.000000	323823.000000	323823.000000	323823.000000	323823.000000	323823.000000	323823.0	323823.000000
mean	161911.000000	503046.460591	11.032249	10.122337	0.009543	0.017593	0.006565	0.088781	0.0	0.006695
std	93479.792447	1001.488317	254.384219	25.363956	0.233369	0.410099	0.208846	2.332849	0.0	0.219146
min	0.000000	502167.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.000000
25%	80955.500000	502183.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.000000
50%	161911.000000	503107.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.000000
75%	242866.500000	504005.000000	1.000000	1.717900	0.000734	0.000000	0.000000	0.000000	0.0	0.000000
max	323822.000000	505644.000000	21744.000000	133.333300	23.898793	52.000000	30.000000	100.000000	0.0	26.000000

- Here is the plot of creative IDs vs Event Frequency where Event 2 is in Red, Event 4 is in Green and Event 5 is in Blue. Events 1 and 3 are not here as they do not occur at all in the data set



2. Is there a rule which you would apply to actively limit or even stop spend on current and future sites for lack of performance? How many impressions would you require on any given site for statistical relevance and the application of that rule?
  - Given that the average click rate is at 0.088% and the average win rate is 11.03%, a good rule to apply would be anything that has a win rate > 8.096% and a click rate of < 0.044% should be stopped. There is no impressions column in the data set, however if there was, the approach would be to look at the distribution of the impressions to find the mean and the variance as well as the standard deviation of the data. The rule would first be anything over 8.096 Win rate, with Impressions greater than average number of impressions +1 standard deviation, and a click rate of <0.044 should be stopped.
  
3. Adtech Data Analysts must find a balance between the stated goal of the advertiser and the platform goal of driving the most spend possible. The goal of this advertiser is to be at a <\$1.5 effective CPI for iOS. What adjustments would you make to increase the competitiveness of the campaigns and allow for potential revenue growth?
  - Invest in creative ids that generate high post install events (502182, 502179 etc..)
  - Average eCPI is 0.74 for non-zero eCPI's, we have room to play with as the goal is to be under 1.5. More win percentages can result in more post install events and lower cost further
  - Look into the most effective strategy (increasing bid to increase win rate of ad space, may be diminishing returns in showing more people the same ad have to find optimum point).
  - Stop showing ineffective Creative IDs via the rule above which will reallocate funds into more effective campaigns
  
4. Which creative ids is currently driving the strongest post-install performance (Events 1-5). How would you weigh the application of the creative sets moving forward?
  - Event 1:
    - Zero total
  - Event 2
    - Has the highest rate at 2168 Total (96.6% of Total events)
    - Results in a high number of post install wins and as a result should be weighted more than the rest followed by 4 and 5 then 1 and 3 with very little weighting
  - Event 3
    - Zero total
  - Event 4
    - 73 Total (3.25% of Total Events)
  - Event 5
    - 2 Total (0.15% of Total Events)

### Creative IDs Sorted by Amount of Post Install Events

- These are the top 5 Creative IDs and their associated apps that are driving the highest post install performance.

Creative ID	App Name	Software	Event Sum
502182	Musi - Simple Music Streaming	iOS 14	31
502179	Musi - Simple Music Streaming	iOS 14	30
502180	Musi - Simple Music Streaming	iOS 14	29
502211	Musi - Simple Music Streaming	iOS 14	27
502183	Musi - Simple Music Streaming	iOS 14	27

### Creative IDs Grouped by App (Multiple apps can show the same graphic)

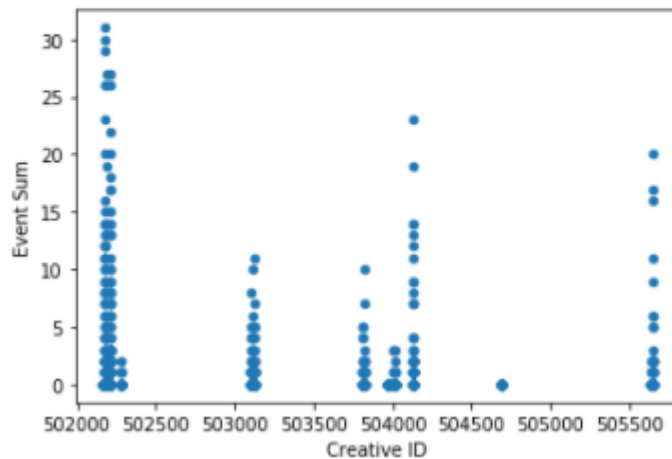
There are 148335 Creative IDs

The average Event Sum is: 0.015121178413725688

There are 147657 Creative IDs that have no post install events

0.457073516027906 % of Creative IDs result in a post install event

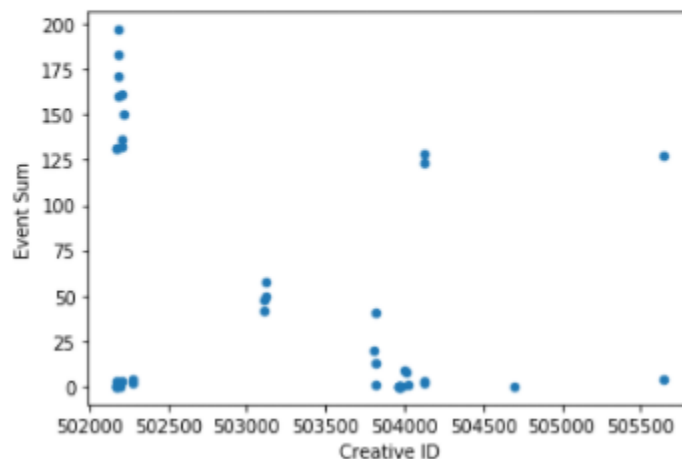
: <matplotlib.axes.\_subplots.AxesSubplot at 0x24b72e1e288>



### Unique Creative IDs

- The Creative IDs on the left side of the graph are driving the highest performance which is driving the average up but this is a good thing. -
- The Creative IDs that are performing much higher than the average should be weighted more heavily than the 14 that are generating no post install events. Though more in depth analysis is required to understand why they are not generating them.

```
There are 46 Unique Creative IDs
The average Event Sum is: 48.76086956521739
There are 14 Creative IDs that have no post install events
69.56521739130434 % of Creative IDS result in a post install event
<matplotlib.axes._subplots.AxesSubplot at 0x24b72f011c8>
```



5. Assuming the Kaden platform has spent, each day, the most allowed by the advertiser for the previous week, and that the advertiser now would like to spend an extra \$1000/day. How would you use this extra budget? Where would you apply it? Which tests would you develop and execute to gain further insight into how to develop overall revenue while maintaining post-install (Event 1-5) performance?
- Need to look at Creative IDs that are generating high post install events and invest into those that have above average post install event rates and and low CPI (this is a give and take, may have more post install events at a higher CPI but it could be worth it)
  - This should be weighted to those Creative IDs that are getting low impressions because the average bid on the ad we have is too low and as such we are not getting impressions at all. A small percent of this budget should be put into raising the average bid on creative ids with win rates less than 2% so that more data can be collected to see if the strategy is yielding results.

6. Other than noted above, which further data points could help you to better understand how to develop overall performance and revenue?
- The target demographic would be a valuable variable. For example If ads are targeted towards younger children, we should push that ad more through tablets. This would also help us understand who clicks more and how to utilize that to generate more post install actions
  - Would like to know what each post install action is to understand why Event 1 and 5 did not occur at all