MEVOD ANALYSIS

CONTENT

GitHub Link

Executive Summary

As competition continues to threaten the streaming market for Mevod in the Middle East, it becomes crucial to re-evaluate and expand upon critical marketing KPIs. By utilizing various marketing analytics aspects, we can better tap into the localized market and provide more native customer content and enhanced customer service

Business Insights

[American] Competitors are looming, but none can tailor to the Arabic market as well as Mevod can:

- Multinational entertainment company, HBO, is attempting to enter the OOT market
- Starz already has a foothold in the OOT market

Data Background

Data was collected from 4 different sources that contained important information for our analysis. This data was generalized based on 4 topics:

- Subscriber Data
- Customer Service Representative Data
- Digital Channel Spend Data (Attribution & Allocation)
- Engagement Analytics Data

Initial pre-analysis determined that data was collected on about 227,000 subscribers, including 2,200,000 customer service requests and 2,500,000 tracked engagements. Customer service requests and engagements included duplicate entries from varying subscribers.

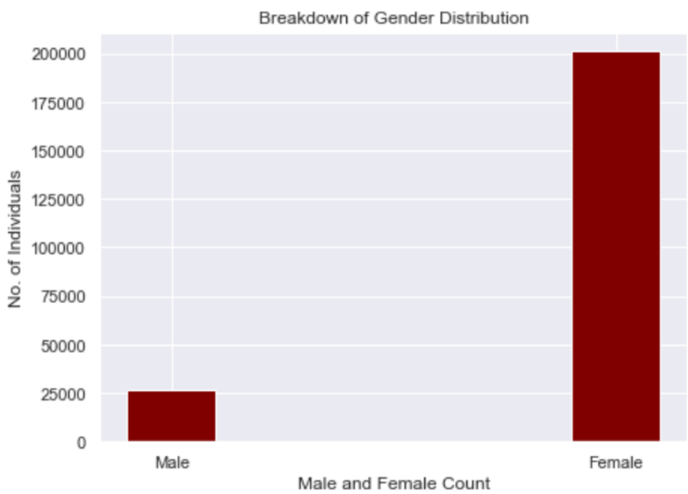
Digital Channel Spend was also collected and analyzed over a 8-month period: from July-2019 up until March-2020.

Due to the complexity of the data collected, which was parsed and collected manually through survey integration, we often had to deal with improper or "dirty" data, so it was more effective to drop inconclusive entries.

What to know about Mevod Users

AVERAGES FOUND (TOTAL POPULATION):

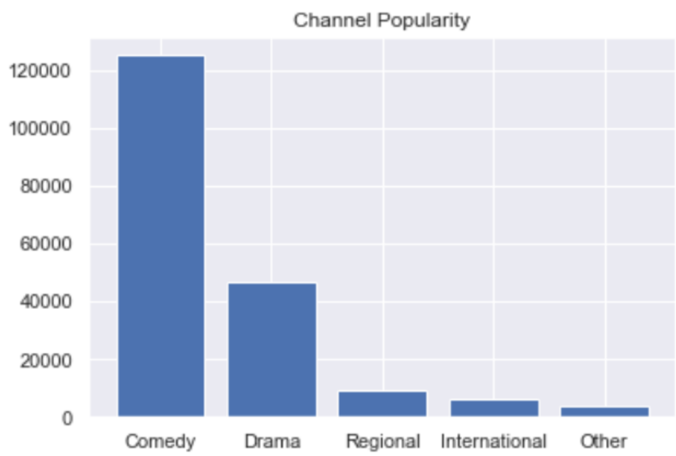
- NUMBER OF WEEKLY UTILIZED SERVICES: 3.00 SERVICES
- WEEKLY CONSUMPTION HOURS: 28 HOURS
- NUMBER OF IDEAL STREAMING SERVICES: 2.00 SERVICES



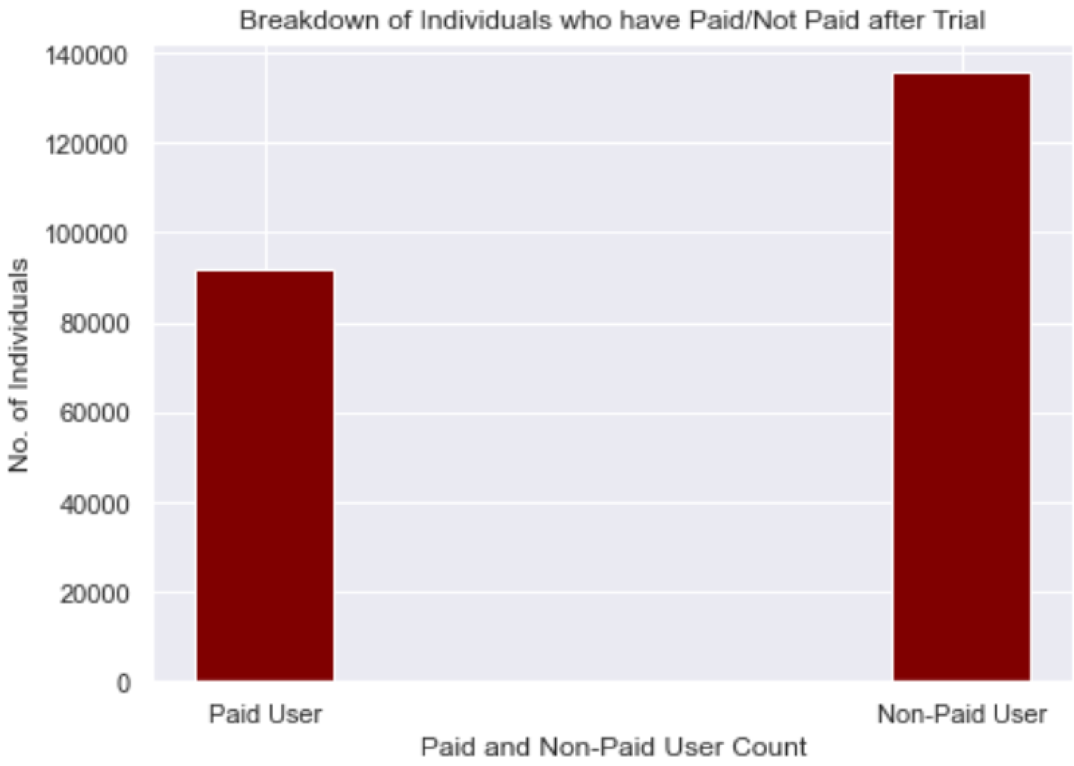
Most Popular Products

TOP FREQUENCIES (TOTAL POPULATION):

- PACKAGE TYPE: **BASE**
- PREFERRED GENRE: **COMEDY**
- INTENDED USE: **ACCESS TO EXCLUSIVE CONTENT**
- PLAN TYPE: **BASE UAE 14 DAY TRIAL**
- PAYMENT TYPE: **STANDARD CHARTER**
- OPERATING SYSTEM: **IOS**



Background: Paid & Non-Paid (after Trial)



Background (Paid)

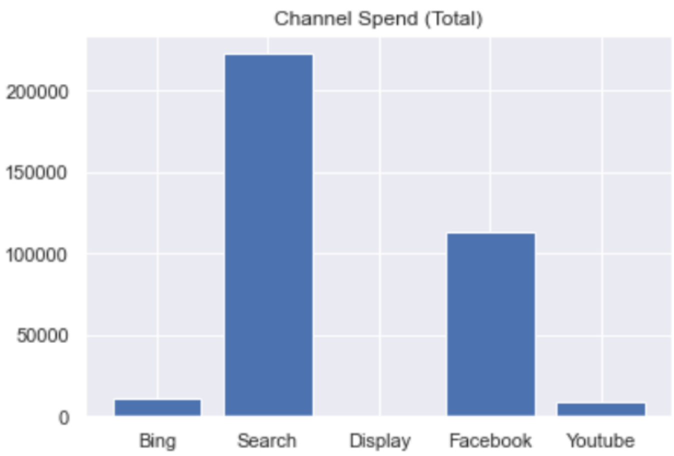
DEMOGRAPHIC BACKGROUND:
MALE #: **12071**
FEMALE #: **79803**

TOP FREQUENCIES (TOTAL POPULATION):
- PACKAGE TYPE: **BASE**
- PREFERRED GENRE: **COMEDY**
- INTENDED USE: **ACCESS TO EXCLUSIVE CONTENT**
- PLAN TYPE: **BASE UAE 14 DAY TRIAL**
- PAYMENT TYPE: **STANDARD CHARTER**
- OPERATING SYSTEM: **IOS**
- NUMBER OF PLAN TYPES TESTED: **11**

Background (Non-Paid)

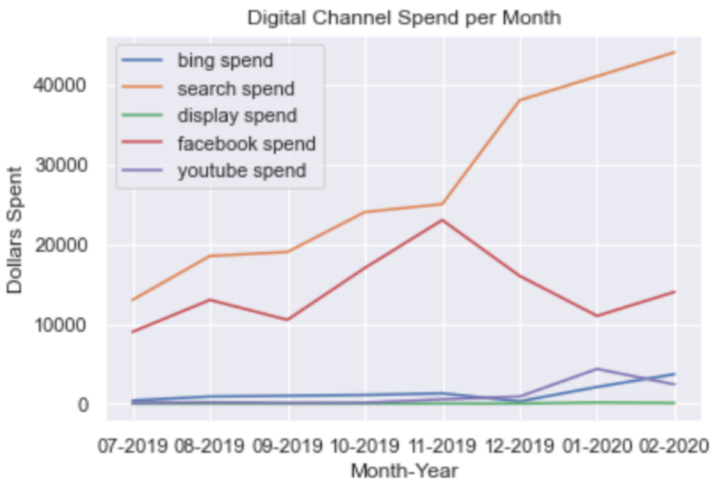
DEMOGRAPHIC BACKGROUND:
- MALE #: **14386**
- FEMALE #: **121099**

TOP FREQUENCIES (TOTAL POPULATION):
- PACKAGE TYPE: **BASE**
- PREFERRED GENRE: **COMEDY**
- INTENDED USE: **ACCESS TO EXCLUSIVE CONTENT**
- PLAN TYPE: **BASE UAE 14 DAY TRIAL**
- OPERATING SYSTEM: **IOS**
- NUMBER OF PLAN TYPES TESTED: **7**



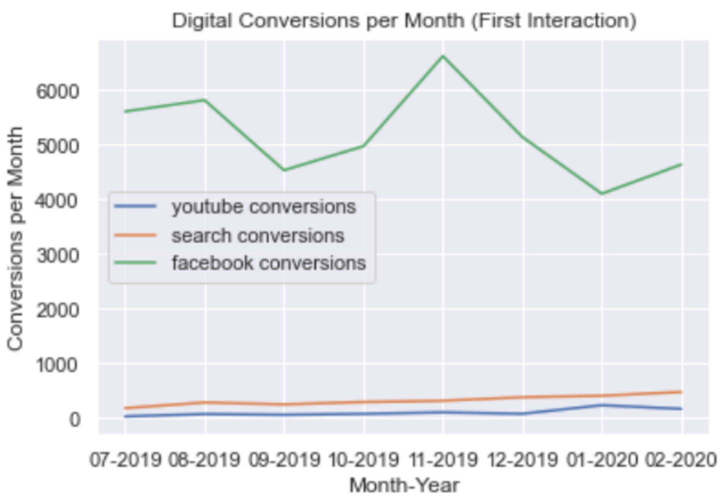
Attribution & Allocation

Goal: *Optimize conversions through advertising channels while also discovering the most cost effective channels to advertise with by using Cost of Acquisition calculations*



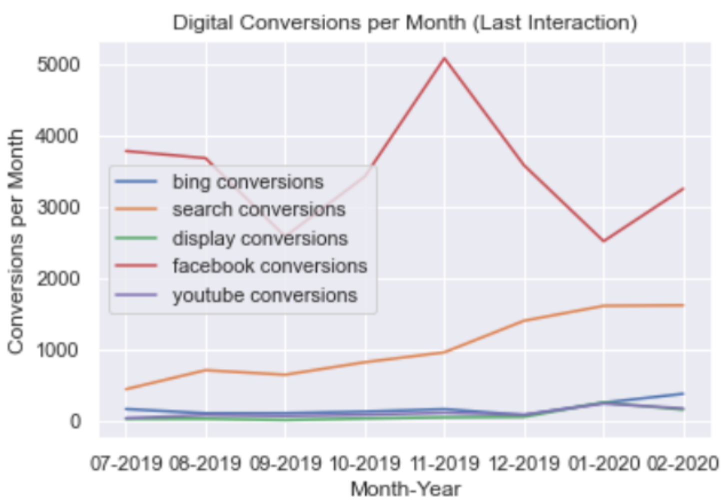
Digital Spend Patterns

- Spending increases by about $14,000 during the month of November, 2019 on "Search" spend, but decreases by about $6,000 for "Facebook" spend
- Spending stays relatively consistent for "Bing" and "Youtube" spend until December, 2019, where spending then increases
- Spending is unchanged for "Display"



Conversions (First Interactions)

- First Interaction assigns conversion credit to the first channel that a "Paid" customer used to discover our brand, and it was derived from the "Attributions_Survey" column in our dataset
- Because it was difficult to categorize varying discovery methods into the 5 digital channel attribution models, we will disregard this graph in our analysis



Conversions (Last Interactions)

- Last Interaction assigns conversion credit to the last channel that a "Paid" customer used to discover our brand.
- Notice how sharply Facebook conversions increase as spend vamps up during the months of Sept-Nov. The conversion rate of Facebook increased faster than the amount of money spent on the channel was increasing

Attribution & Allocation

Based on the average "Cost of Acquisition (CAC)" of our LAST INTERACTION conversions, we see some interesting insights (these values are derived from filtering our conversion variable (paid_TF) to "True". Below are two graphs displaying the CAC per channel over a monthly interval, as well as throughout the entire 8-month period

	Average CAC per Channel (LAST INTERACTION)				
	Channel 1	Channel 2	Channel 3	Channel 4	Channel 5
Marginal Spend	Bing	Search	Display	Facebook	Youtube
July - 2019	$ 2.52	$ 29.75	$ 0.57	$ 2.38	$ 90.00
August - 2019	$ 9.00	$ 26.28	$ 0.91	$ 3.53	$ 36.00
September - 2019	$ 9.90	$ 29.69	$ 2.17	$ 4.07	$ 100.00
October - 2019	$ 8.94	$ 29.41	$ 0.73	$ 4.97	$ 32.50
November - 2019	$ 8.28	$ 26.18	$ 0.69	$ 4.53	$ 36.67
December - 2019	$ 4.00	$ 27.18	$ 0.63	$ 4.47	$ 32.14
January - 2020	$ 8.47	$ 25.51	$ 0.59	$ 4.37	$ 41.13
February - 2020	$ 9.92	$ 27.26	$ 0.62	$ 4.31	$ 37.23
Conversions Total	1336	8171	567	27889	225

	Average CAC per Channel (LAST INTERACTION)				
Marginal Spend	Bing	Search	Display	Facebook	Youtube
Whole Period (Jul - Mar)	$ 8.08	$ 27.23	$ 0.65	$ 4.07	$ 38.80

Analysis

By breaking down the CAC into monthly intervals, we can see some patterns that aren't as apparent in the aggregated calculations. For example:

- Why does Bing CAC in July-2019 appear to be so much lower than the other Bing CAC months?
- How come channels such as Search and Facebook appear more uniform in terms of monthly CAC distributions?
- Why were CAC's in September for most of the channels above noticeably higher?

Attribution & Allocation

Conclusions

The intuition would be to spend each monthly allocation of digital channel spend budget on the channel with the lowest CAC ("Display"). However, we have to be wary of other factors as well.

- Social and Search Channels, although rather expensive in terms of CAC, allow for enhanced reach and brand awareness due to the prevalence of such platforms. By utilizing social or search channels, randomness of the "long-tail" can impact how people find and encounter new brands based on interests and search intent. Typing in "arabic movie service" in Google may bring in more engagements than expected to Mevod.
- Sharp increases in CACs for Bing, Display and YouTube are alarming in any month prior to December, 2019, because there was little to no increase in spending on these channels.
- Even a little more ad spend being delegated towards YouTube does little to improve conversions or CAC.
- Getting a conversion rate would be helpful in better understanding the strength of each channel's advertising campaign.
- Calculating the marginal CAC would be the next useful step in better optimizing Mevod's attribution and allocation model.

Budgetary Allocation

Monthly CAC's are more impacted by seasonal changes that can skew results in directions that are not helpful to our analysis, so I did my allocation modeling based on the *aggregate* conversions and channel spend.

Allocation of Budget	Bing	Search	Display	Facebook	Youtube	Total Conversions
conversion %	3%	21%	1%	73%	1%	38188
budget %	3%	63%	0.103%	31.89%	2%	
Total	$ 5,338.44	$ 74,166.67	$ 3,660.00	$ 268,365.89	$ 4,365.00	$ 355,896.00

- Based on a ratio of budget $ /conversion %, I allocated various amounts of money toward channels
- Channels that had high CACs, such as Bing, Search and YouTube had their spending cut, while channels with low CACs or higher CACs/high conversions, had their budgets increased
- I kept the entire period budget the same at: **$355,896**
- I would suggest experimenting with YouTube spending a bit because of the volatility that exists within the monthly CAC values, as well as the fact that many media platforms are turning to YouTube to advertise their products/services

Customer Segmentation & Clustering

Previous analysis explored the demographic characteristics of our main "Paid" customer population However, we are interested in seeing if elements of similarity or homophily can help to better target and market Mevod's products to an ever-expanding market.

Goal: *Understand various customer segments (groups) that Mevod can create in order to better target these segments*

K-means clustering to segment the Market

K-means segmentation is a power machine learning algorithm that can be used to uncover "latent" or undiscovered groups within our data that may offer some transparency in terms of marketing objectives. The clusters below have been filtered to reflect our Paying Customers (paid_TF=True), and dirty data columns were dropped.

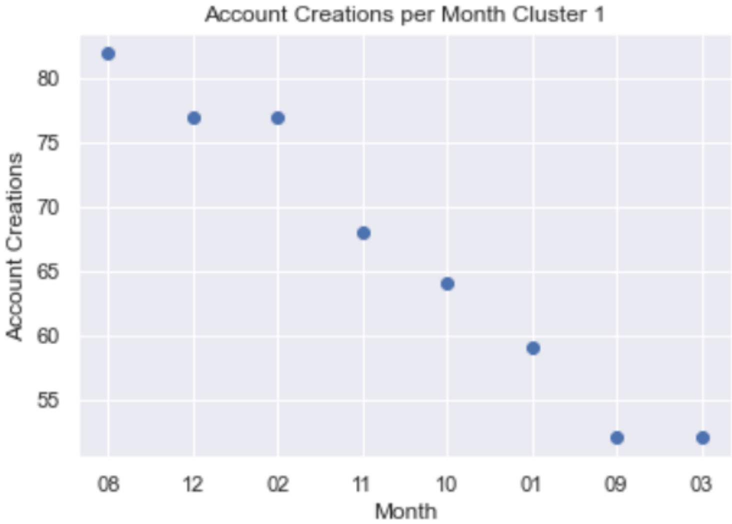
Through this, we found ***3*** segmented cluster...

	Cluster 1	Cluster 2	Cluster 3
Segment Characterization	"Avid Fans"	"Minimalists"	"The Casual Watcher"
# of Weekly Services Utilized	3.58	2.79	3.07
Weekly Consumption Hours	35.46	26.75	26.96
# Ideal Streaming Services	2.01	1.96	2.05
Average Age	45.11	51.84	47.93
Size of Cluster	531	1577	1070

	Counts per Payment Type		
	Cluster 1	Cluster 2	Cluster 3
Payment Type CBD	35	95	42
Payment Type Najim	13	55	22
Payment Type Paypal	208	563	415
Payment Type RAKBANK	79	268	180
Payment Type Standard Charter	196	596	411
	Counts per Trial Type		
	Cluster 1	Cluster 2	Cluster 3
Plan - Base / EUR / 14 day	0	0	0
Plan - Base / UAE / 14 day	531	1577	1070
Plan - High / UAE / 14 day	0	0	0

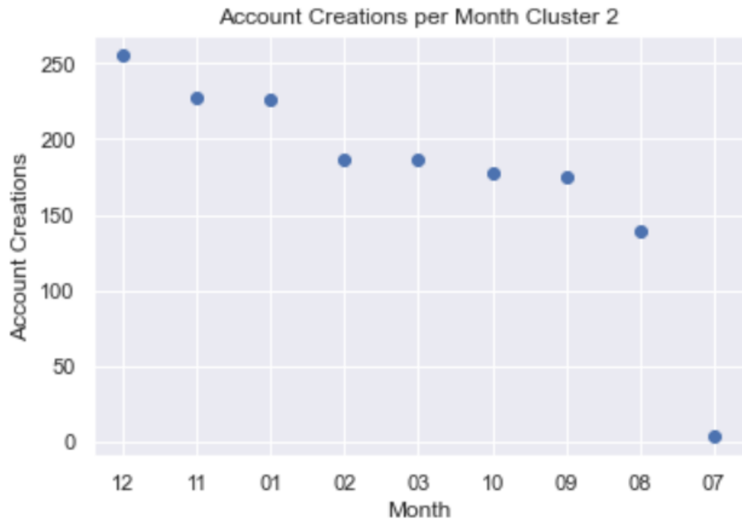
Customer Segmentation & Clustering

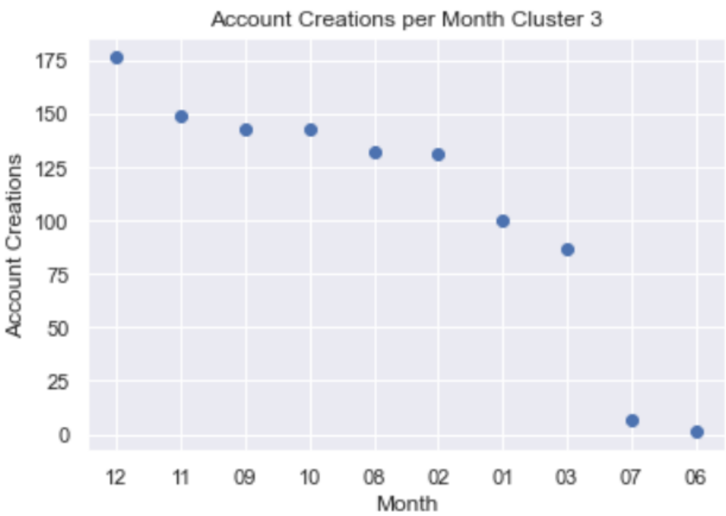
Account Creation



- Cluster 1 has far fewer individuals scattered among the range of months listed on the x-axis, with the most individuals occurring in June and the least amount of individuals occurring in September
- Account creations seem to rise as the holiday season approaches from October to December, then decreasing again in the New Year

- While holiday months such as November - January are receiving more than 200 account creations, summer months seem to be lacking account sign-ups
- July has virtually no account sign-ups from this cluster
- Streaming services tend to drive more account creations as the year proceeds past the summer months





- June and July have virtually no account sign-ups. Similar to Cluster 2, this likely indicates a dying interest in streaming over the summer months
- Cluster 2 and 3 exhibit similar patterns in terms of account sign-up popularity growing as the fall months approach and the summer months end, with November and December topping the charts for account creations

Customer Segmentation & Clustering

Marketing Initiatives - The 4 P's

Price:

- All three clusters are drawn to the same basic plans and payment methods. Remaining with the same consistent 14-day-trail period seems smart. Based on the data that was utilized, it may be increasingly smart to promote discounts and payments through the relevant payment methods that are used by all of the clusters, namely Paypal and Charter Bank. Through these mediums, extended or exclusive base plans can be offered in some sort of partnership that may better enhance the worth of a the "High" UAE 14-day plan. Paypal should be the key player as it allows for more secure and universal payments.

Product:

- For Cluster 1, the importance of product comes in the form of exclusive or continuous content. Whether this be some sort of "binge marathon" special or rapid release of original content, Cluster 1 has significantly more time invested in consuming Mevod's content and is interested in more ideal services. With this knowledge, more novel content can be promoted or shown to these people by prompting them to sign up for a more advanced plan.
- Cluster 2 seems more passive and older in demographic age than the other two clusters. They are less engaged with the service as of now, so it may be worth promoting old, classic films to this audience as a way of creating nostalgia.
- Cluster 3 is highly interested in comparing different streaming services. With this knowledge, and the goal of creating more native Turkish/Egyptian content, produce and promote culturally-representative media that is superior to other non-native competitors'.

Place:

- With the lack of tailored, Arabic media in Mevod's market region, the company should focus on brand and culturally appropriate advertisements that can appeal to this specific culture. These can include television advertisements or trailors for movies that don't include explicit content and focus on humanitarian deeds. At the same time, different age averages across the clusters likely posit different acceptances to digital media: older people are flocking towards Facebook, while younger, more tech-savvy individuals are more drawn towards graphic platforms such as SnapChat, Instagram and maybe even *Display* Ads.

Promotion:

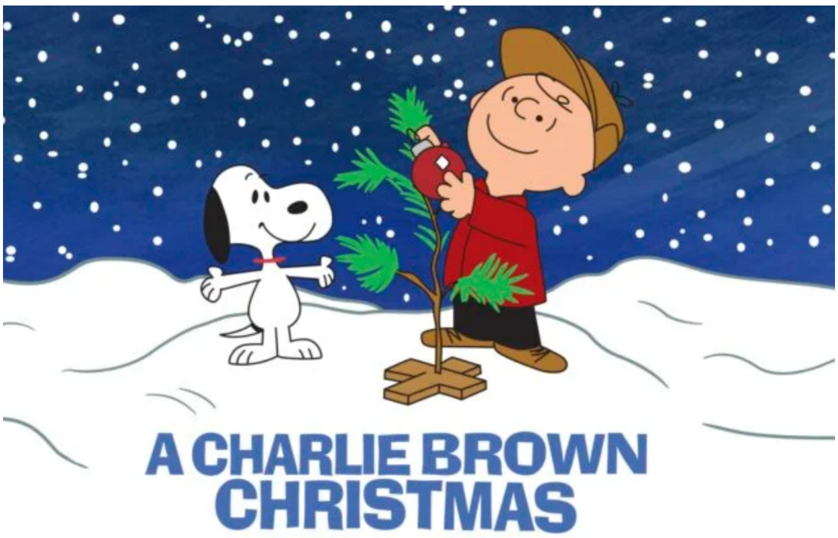
- For Cluster 1, promote new, exclusive content or "binge marathons" during the summer months. Because this cluster has much more time invested in utilizing the streaming service and are open to more ideal streaming services, Mevod can promote content when they seem to have more free time in the summer.
- For Cluster 2 & 3, the summer months seem to be almost uneventful. Thus, these clusters should be targeted with shorter, engaging but novel content around the holiday months that allows for less time to be spent on the service. At the same time, maybe consider doing a holiday (Arabic or US, depending on audience targeting preferences) discount promotion with "Paypal" or "Standard Charter" to incentivize cheaper monthly plans after a FREE 14-day-trial period.

Customer Segmentation & Clustering

Marketing Examples:

Cluster 3
"Need a reason to watch a movie? Try family, fun, snacks and a million streaming options."

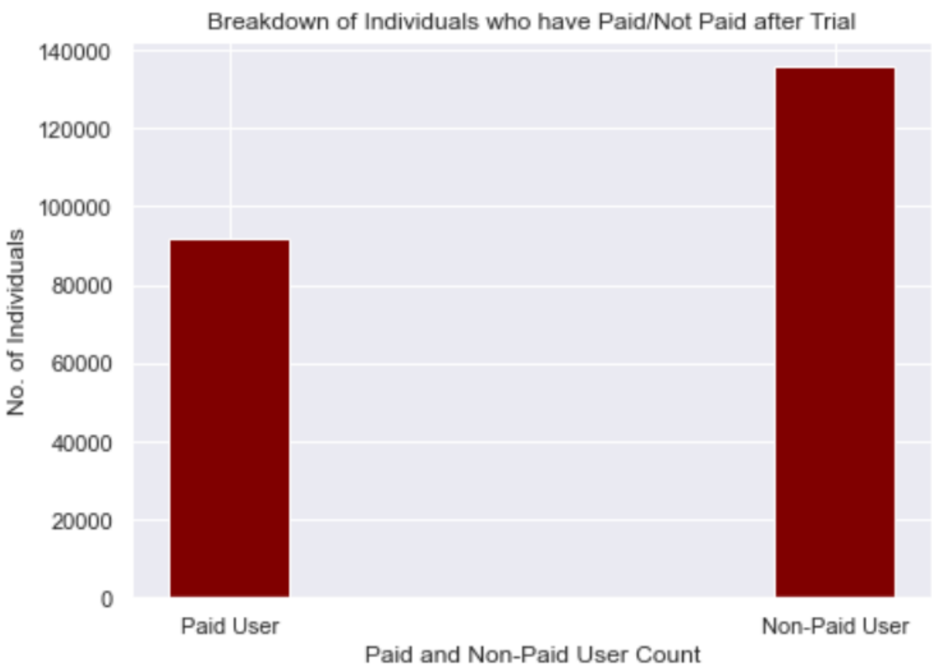
Cluster 1
"Exclusivity at its finest, oh, and all of the perks that come with it too."



Cluster 2
"Feel that little kid inside of you with a throwback classic."

Churn Modeling

Goal: *With the help of machine learning algorithms such as linear regression and logistic regression, we can help identify which customers of Mevod's are likely or not likely to churn. From here, we can plan our strategies accordingly to prevent this loss of business.*



How we defined *Churn*

In terms of simplicity, we utilize the variable "paid_TF" as the ultimate test of churn. This is because this variable will determine if the user has or has not paid for an actual subscription service offered by Mevod after a trial period. At the same time, it is easy to see the discrepancies that exist within the chart above. For example, how can we help prevent over 130,000 from leaving Mevod's service without paying? This can be influenced by several things such as discounts, acceptance rate of churn and the thresholds that we classify churners based on. We want to make sure that we are able to retain as many people as we can and keep them as paying customers without offering unnecessary discounts and hurting our bottom line.

To do this, we begin by creating various linear and logistic regression models to predict which consumers are likely to be churners, based on a variety of ENGAGEMENT variables and SUBSCRIBER factors. By combining these datasets together, we were able to better understand certain aspects of covariance in the hopes of making better predictions.

Churn Modeling

The Models: Linear Regression

```
RESULTS OF LOGIT MODEL FITTING

                            OLS Regression Results
==============================================================================
Dep. Variable:                paid_TF   R-squared:                       0.940
Model:                            OLS   Adj. R-squared:                  0.940
Method:                 Least Squares   F-statistic:                 2.797e+04
Date:                Mon, 13 Dec 2021   Prob (F-statistic):               0.00
Time:                        17:27:42   Log-Likelihood:                 49350.
No. Observations:               37641   AIC:                        -9.866e+04
Df Residuals:                   37619   BIC:                        -9.847e+04
Df Model:                          21
Covariance Type:            nonrobust
=====================================================================================================
                                       coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------------------------
app_opens                           -0.0238      0.014     -1.703      0.089      -0.051       0.004
cust_service_mssgs                  -0.0469      0.011     -4.283      0.000      -0.068      -0.025
num_videos_more_than_30_seconds      0.1700      0.012     13.966      0.000       0.146       0.194
num_videos_rated                     0.1474      0.019      7.643      0.000       0.110       0.185
num_series_started                  -0.0599      0.009     -6.536      0.000      -0.078      -0.042
num_weekly_services_utilized        -0.0139      0.010     -1.325      0.185      -0.034       0.007
weekly_consumption_hour              0.0263      0.010      2.601      0.009       0.006       0.046
num_ideal_streaming_services        -0.0013      0.016     -0.079      0.937      -0.034       0.031
retarget_TF                          0.0051      0.002      2.664      0.008       0.001       0.009
age                                 -0.0141      0.063     -0.224      0.823      -0.138       0.110
male_TF                              0.0044      0.001      3.106      0.002       0.002       0.007
monthly_price                       -0.1223      0.052     -2.341      0.019      -0.225      -0.020
discount_price                      -0.0688      0.010     -6.745      0.000      -0.089      -0.049
creation_until_cancel_days           0.8709      0.011     79.096      0.000       0.849       0.893
cancel_before_trial_end              0.0176      0.002     10.622      0.000       0.014       0.021
revenue_net                          1.3338      0.015     90.034      0.000       1.305       1.363
join_fee                            -0.0038      0.001     -2.550      0.011      -0.007      -0.001
refund_after_trial_TF                0.9036      0.002    404.281      0.000       0.899       0.908
plan_type_base_eur_14_day_trial      0.0015      0.082      0.019      0.985      -0.160       0.163
plan_type_base_uae_14_day_trial     -0.0033      0.065     -0.051      0.959      -0.131       0.124
plan_type_high_aud_14_day_trial     -0.1273      0.061     -2.079      0.038      -0.247      -0.007
plan_type_high_sar_14_day_trial     -0.0912      0.079     -1.150      0.250      -0.247       0.064
plan_type_high_uae_14_day_trial      0.0606      0.082      0.741      0.459      -0.100       0.221
const                               -0.3860      0.039     -9.956      0.000      -0.462      -0.310
==============================================================================
Omnibus:                    20734.186   Durbin-Watson:                   2.010
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          2788343.426
Skew:                          -1.656   Prob(JB):                         0.00
Kurtosis:                      45.034   Cond. No.                     7.84e+15
==============================================================================
```

Important things to notice above the linear regression results are the variables in the "coeff" column. These are the individual correlations that exist among the variables on the outcome of churn or not churn, in our case, pay or not pay. For the example, revenue_net and creation_until_cancel_days are highly correlated with the outcome variable, and thus have a large influence on its value. On the other hand, certain values such as monthly_price or discount_price are negatively correlated with the outcome variable.

At a significance level of 0.05, we can take a look at the P>|t| column, and wherever this value is less than alpha, this association is statistically significant. Changes in these variables are associated with changes in churn/paid rates.

With an r-squared of 0.94, this model fits our data well.

Churn Modeling

The Models: Logistic Regression

```
RESULTS OF LOGIT MODEL FITTING

                           Logit Regression Results
==============================================================================
Dep. Variable:                paid_TF   No. Observations:                37641
Model:                          Logit   Df Residuals:                    37617
Method:                           MLE   Df Model:                           23
Date:                Mon, 13 Dec 2021   Pseudo R-squ.:                  0.9974
Time:                        17:28:49   Log-Likelihood:                -25.976
converged:                      False   LL-Null:                       -10169.
Covariance Type:            nonrobust   LLR p-value:                     0.000
==============================================================================================
                                     coef    std err          z      P>|z|      [0.025      0.975]
----------------------------------------------------------------------------------------------
app_opens                         -7.7464     13.429     -0.577      0.564     -34.066      18.574
cust_service_mssgs                -4.1495     13.861     -0.299      0.765     -31.317      23.018
num_videos_more_than_30_seconds   33.0282     13.368      2.471      0.013       6.827      59.230
num_videos_rated                  14.7399     55.151      0.267      0.789     -93.354     122.834
num_series_started               -17.9031      9.489     -1.887      0.059     -36.502       0.696
num_weekly_services_utilized      10.2747     11.332      0.907      0.365     -11.936      32.485
weekly_consumption_hour          -24.9607     14.933     -1.672      0.095     -54.228       4.307
num_ideal_streaming_services     -19.4106     17.365     -1.118      0.264     -53.445      14.623
retarget_TF                       -0.3981      3.507     -0.114      0.910      -7.273       6.476
age                               14.3159    694.025      0.021      0.984   -1345.949    1374.580
male_TF                            1.7718      2.429      0.730      0.466      -2.988       6.532
monthly_price                      2.1012        nan        nan        nan         nan         nan
discount_price                    -4.6727   2.64e+07  -1.77e-07      1.000   -5.18e+07    5.18e+07
creation_until_cancel_days        26.8248      3.834      6.996      0.000      19.310      34.340
cancel_before_trial_end           16.1635    616.445      0.026      0.979   -1192.047    1224.374
revenue_net                       36.4482      5.886      6.193      0.000      24.912      47.984
join_fee                          -0.5029      1.907     -0.264      0.792      -4.240       3.235
refund_after_trial_TF             37.3719   1626.276      0.023      0.982   -3150.071    3224.815
plan_type_base_eur_14_day_trial  -24.9085        nan        nan        nan         nan         nan
plan_type_base_uae_14_day_trial  -10.9560   3.11e+06  -3.53e-06      1.000   -6.09e+06    6.09e+06
plan_type_high_aud_14_day_trial   -5.2364        nan        nan        nan         nan         nan
plan_type_high_sar_14_day_trial   -2.4769        nan        nan        nan         nan         nan
plan_type_high_uae_14_day_trial  -10.5695        nan        nan        nan         nan         nan
const                             -8.0018        nan        nan        nan         nan         nan
==============================================================================================
```

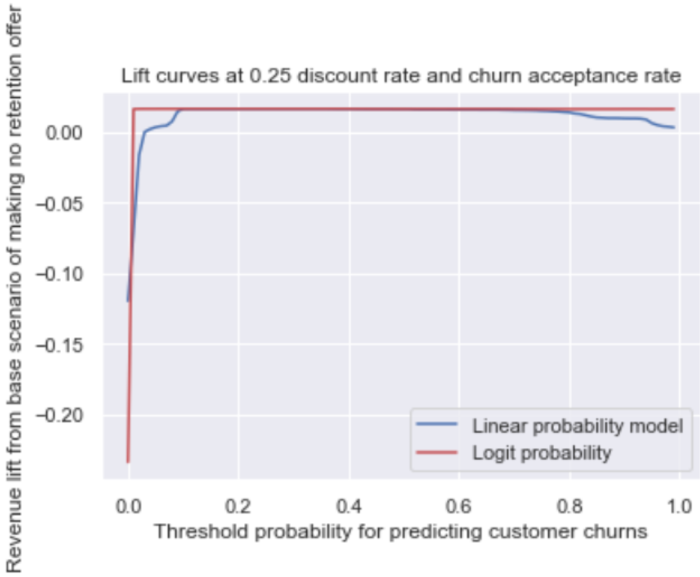
Interpreting the logistic regression graph often involves looking at the "coeff" column as well. When scanning the numbers, certain negative numbers indicate a sort of negative correlation to the outcome, while certain positive numbers indicate a positive correlation. For example, looking at the negative coeff in front of weekly_consumption_hour, we can deduce that more hours consumed on Mevod are less indicative of churning/not-paying.

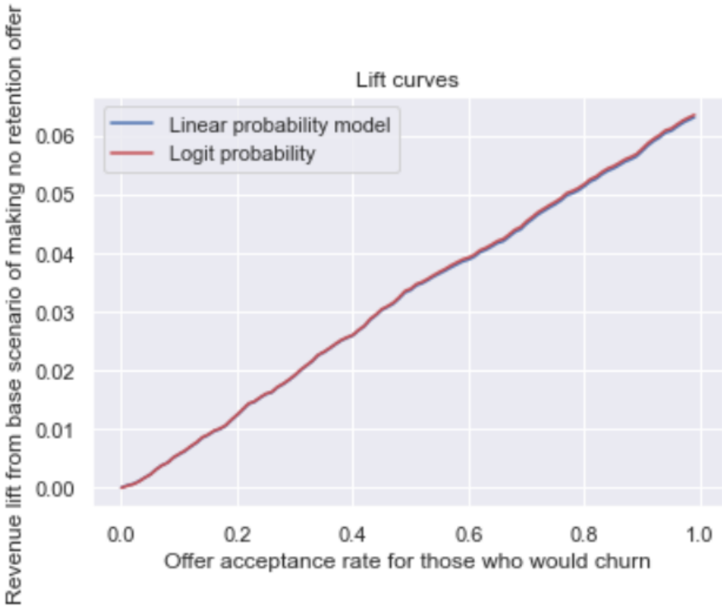
The magnitude of coefficients also needs to be accounted for. For example, the coeff for num_videos_more_than_30_seconds is twice as big as the coeff for cancel_before_trial_ends, indicating that the first value is twice as big in terms of leading to churn as the second value.

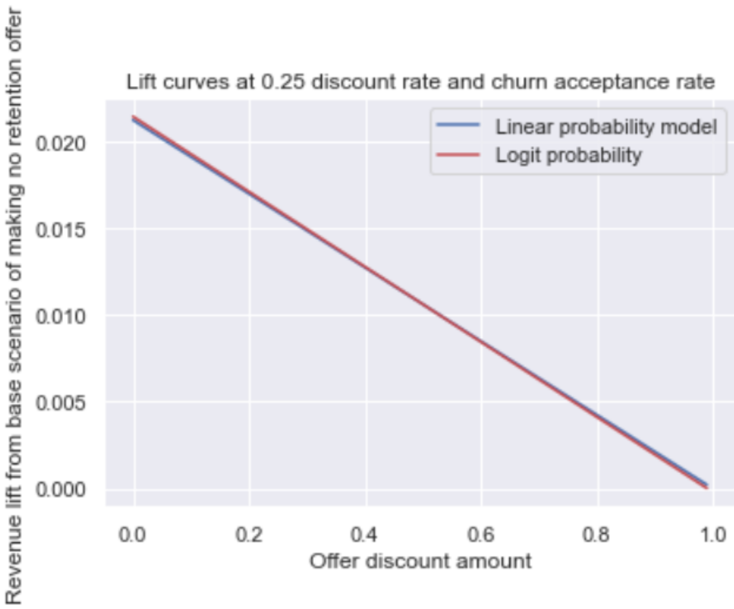
The standard error measures how far the observed value falls from the regression line, or how wrong the model is on average using values/units from the response model.

Churn Modeling

The Conclusions:







Churn Modeling

The Conclusions:
Assume 1-1 ratio of churn acceptance rate and discount amount in model

Price set at: **$15**
Optimal Churn Acceptance Rate: **0.25**
Optimal Discount Amount: **0.25**
Price at Discount Rate: **$11.25**

Maximum Revenue Lift Achievable: **0.016**

By proposing the offers:

Logistic Model Improvement:
Accept Offer: 321 not churned
Decline Offer: 15,811 churned
Additional Revenue: $3,611.25

Linear Model Improvement:
Accept Offer: 315 not churned
Decline Offer: 15,817 churned
Additional Revenue: $3,543.75

Final Remarks:

To improve our churn model we can:

- Create an AB test to better see how the relationship unfolds between acceptance rate and discount rate
- Make sure the model is properly assembled so as to not have an AUC so close to 1
- More sophisticated churn modeling can be done by using decision trees that rely on gini indexes to formulate predictions
- Discounting and CLV will offer a more robust metric to better enhance the revenue model
- Uplift modeling can be useful to see how the churners/non-churners are likely to respond to a discount or any other marketing campaign offered.