Nate Bitting
PREDICT 450 - Marketing Analytics
Solo 1 – App Happy General Attitudinal Post Hoc Segmentation Analysis

Overall Objective

App Happy is in search of a segmentation scheme leveraging survey data of 1,663 respondents conducted by the Consumer Spy Corporation (CSC). We will examine multiple clustering techniques leveraging this data and will discuss the statistical methods used in our general attitudinal post hoc segmentation analysis. Lastly, given App Happy would like to transition to the entertainment app space, we will provide recommendations for what type of entertainment apps to market to each customer segment. We will also provide suggestions for additional market research that may be required and classification model selection options to allocate new customers into the suggested customer segments derived from our analysis.

Survey Structure

The survey conducted by CSC is broken down into multiple components including application and device preferences (5 questions), demographics (8 questions), technology use behaviors (12 questions), personality traits (12 questions), and buying habits (16 questions) of the sample customers. According to the documentation, the survey questionnaire was based on some preliminary qualitative research that included focus groups and one-on-one interviews. Our assumption is that the respondents who participated in the survey are representative of App Happy's target customer base.

Customer Segmentation Scheme Methodology

Given the objective was to conduct a general attitudinal post hoc segmentation analysis, we first had to identify what questions we would use as our basis variables. Upon reviewing the survey, we concluded that questions 24, 25 and 26 were the most logical choice to assess the customers' attitudes. First, question 24 focused primarily on the customers' attitudes towards technology use. Second, question 25 provided us insight into the personality traits of the customers. Finally, question 26 provided us insights to the customers' opinions around their buying behavior. All three questions contained multiple sub-questions leveraging Likert scales, including the following options: Strongly Agree,

Agree, Somewhat Agree, Somewhat Disagree, Disagree, and Strongly Disagree. Each of these options were converted to numeric variables ranging from 1 (Strongly Agree) to 6 (Strongly Disagree).

After we determined the basis variables to use for our segmentation scheme, we then had to explore what clustering algorithms to leverage. Our first clustering algorithm we chose was *k-means* from the cluster package in R. The assumptions used for this approach were to treat the basis variables as continuous given we were leveraging central tendency measures in the k-means algorithm. The second algorithm we considered was *hierarchical clustering* using hclust from the stats package in R. In order to use our basis variables in this approach we had to first calculate the distance matrix of all variables leveraging the Euclidean distance measurement.

K-means Clustering Approach

The methodology used to determine the number of clusters for the k-means algorithm included the use of plotting the within sum of squares against the number of clusters as well as the NbClust package in R.

Figure 1

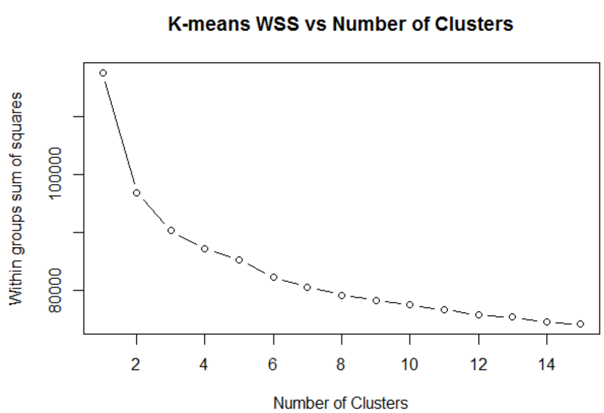
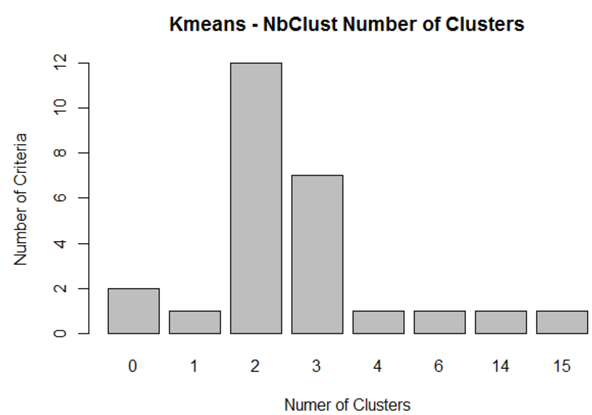
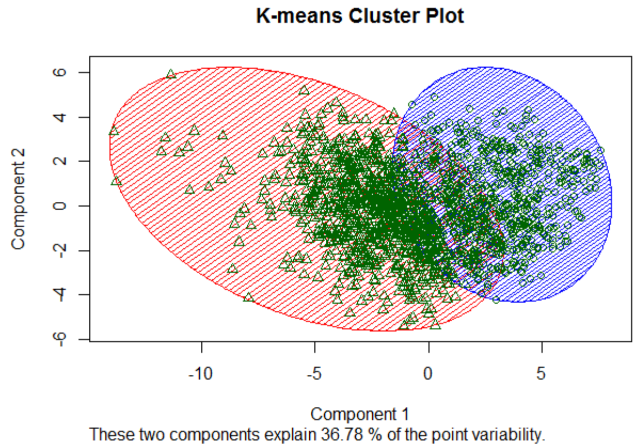


Figure 2



As shown in Figures 1 and 2 above, we can see that both indicate that the ideal number of clusters is two. We can also plot the visualization of the clusters themselves for each observation as shown in Figure 3 below.

Figure 3

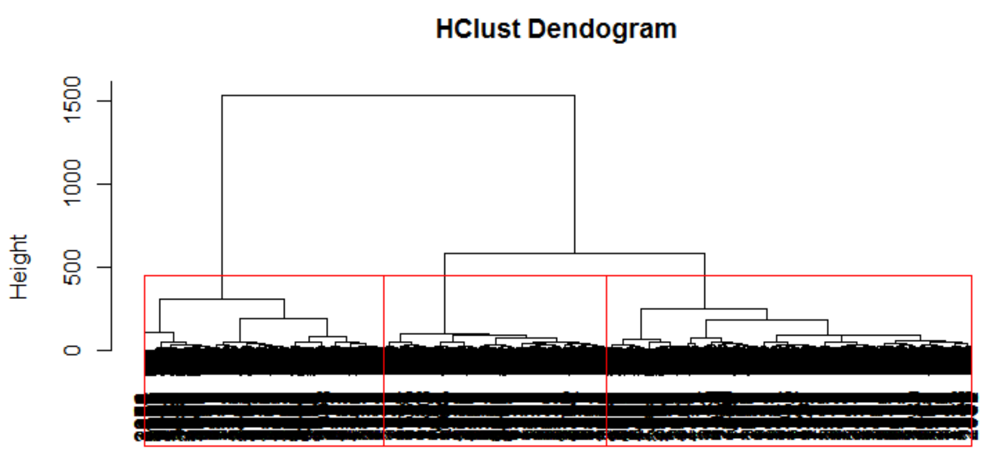


We can clearly see in Figure 3 that overlap exists between the two clusters, which indicates the k-means cluster algorithm may not be the optimal method to derive the appropriate number of clusters. To test this theory, we conducted chi-squared and t-tests between the different clusters for all questions in the survey, which resulted in little differences between the clusters. Thus, we chose to explore our second clustering algorithm: hierarchical clustering.

Hierarchical Clustering Approach

Our second clustering algorithm, as mentioned previously, is the hclust algorithm from the stats package in R. We first had to convert our attitudinal dataset into a distance matrix leveraging the Euclidean distance measurement. Upon exploring the dendogram plot, shown in Figure 4, we chose to select three clusters as it seemed to have an even separation of the customers into unique groups.

Figure 4



To confirm if these groups were in fact unique from each other, we applied a generalized linear model to each variable against the class variable to estimate the Akaike Information Criterion (AIC); the lower the AIC value, the higher the predictive ability of the model. We then sorted the AIC values for each question in ascending order to investigate questions that had better predictive power based on the class in which they were assigned. The second approach we used to confirm differences between clusters for each questions was by performing an ANOVA between all three clusters and assess the p-values to determine if we could reject the null hypothesis of all three clusters being the same. These two approaches helped to narrow down important questions to investigate during the segment profiling phase. The hierarchical clustering algorithm provided much more clarity over k-means, therefore we chose to leverage this method for segmenting the customers.

Segment Profiling

The segment profiling approach we used was to explore the survey questions with the perspective of knowing App Happy intends to develop apps for the entertainment space. We first wanted to understand what type of apps their customers currently use, understand their demographics, as well as exploring their attitudes towards technology use, personality traits, and buying habits. Cluster 1 is the largest segment, while clusters 2 and 3 are smaller and therefore have more unique attributes.

Commonalities Among All Clusters

Before we explore the differences between the three segments, we felt it pertinent to cover the similarities across all clusters. We will not go over all of the similarities, but those we found to be the most useful for App Happy to be aware of as they enter this new entertainment app market. From a mobile platform perspective, the majority of their customers use Apple and Android devices with very few using other platforms. Secondly, the majority of customers are heavy users of Facebook and YouTube.

All clusters had a high level of interest in music and sound identification, gaming, and social networking apps, which complements the market App Happy intends to enter. Most customers prefer to use very specific news publication applications (i.e. NYT, WSJ) as opposed to generic news sources (i.e. Yahoo! News), so it is our recommendation to avoid making any news-related apps. From a demographic perspective, all tend to be white/Caucasian, similar marital status distribution, have some level of college education, and are evenly distributed between men and women across all segments.

Cluster 1 Profile

Cluster 1 represented 44% of the entire survey population and the majority of question responses had a Gaussian distribution. When analyzing the demographic-specific information provided, we found the majority fell into the $30-90K income range and the age range was varied with the majority falling into the 18-24 age category. They prefer to use free applications and have little to no attachment to brand names or luxury items. This cluster tends to disagree that just because you earn a lot of money, you have to spend it. They do not believe it is worth spending the money for additional app features. They are not impulse buyers and tend to do most of their shopping offline. They disagree that they cannot get enough apps, which indicates that they would prefer to have fewer individual apps that are capable of addressing a broader set of use cases. All in all, this customer segment would most likely use apps that are more of a “one stop shop” that combines the capabilities of multiple apps and are free to download.

Cluster 2 Profile

Cluster 2 represented 29% of all survey respondents and proved to be the most unique segment as their characteristics and attitudes deviated significantly from the other clusters. From an application preference perspective, they had the highest usage of entertainment, TV-related, and gaming apps compared to clusters 1 and 3. They visited more of the websites indicated in the survey more frequently than the other clusters, possibly indicating a higher level of curiosity for different types of applications or content. From a demographics perspective, this group has the highest percentage of people in the 18-24 age group and have a higher proportion of younger children under 6 years old. The majority of customers in this segment fall into the $20-60K income range, but had the highest aptitude for paying for apps and additional app features.

Cluster 2 considers themselves to always be up-to-date with the latest technology developments, have to have the latest and greatest, leverage technology to provide them with more control, and have a heavy use of social networking sites to keep in touch with friends and family. They also strongly agree that the internet and sites like Facebook make it much easier to avoid seeing or speaking to family and friends. This indicates that they are early adopters, or innovators, when it comes to new technology. Music is a very important part of their life and they really enjoy learning about TV shows even when they are not watching them either through TV apps or sites like IMDB.

From a personality perspective, they were consistent in how they assessed themselves as being thought leaders, prefer to be in charge while making decisions, and are first among their friends and family to try out new technology. Additionally, they consider themselves risk takers, always active and on the go, and optimistic. These general attitudes and personality traits indicate Cluster 2 would be a prime candidate as a segment in which to experiment with new app ideas with early access programs and offers.

Finally, Cluster 2 had the highest appetite to spend money on apps compared to the other clusters. Interestingly, this group identifies with being bargain shoppers, but also tend to be impulse buyers. This is somewhat counterintuitive, but can possibly be explained in that they are willing to make an impulse purchase if they feel they are getting a good deal. They have a high affinity towards brand names and luxury goods and like to show off all of their cool apps to friends and family. They are more willing to spend additional money for app features and strongly agree that their phone is a source of entertainment.

Cluster 3 Profile

Cluster 3 represented 27% of all survey respondents and had slight differences from clusters 1 and 2. Similar to Cluster 1, the customers in this segment tended to prefer using free apps. From a demographics perspective, this group stood out as the highest earners with a significant percentage in the highest income bracket of $150K or more. This made more sense as the age distribution was more even across all age groups, though slightly trending towards the younger age groups. Additional demographic information, such as location, would have also been helpful to better understand the income distribution.

Like Cluster 2, they also like having the latest and greatest gadgets and do not feel technology is overly pervasive in everyday life or think there is too much information available on the internet. They firmly believe that the internet makes it easier for them to keep in touch with friends and family. The customers in this segment also like to be in charge and would rather not be told what to do, but make decisions on their own.

From a buying behavior perspective, they have an even distribution of shopping online and offline, are indifferent towards brand names or the number of apps they have on their devices. They agree that the "cool factor" of apps is important, not necessarily the quantity. Lastly, they do have a slight attachment to luxury items and believe their mobile phone is a source of entertainment.

Recommendations

Based on the customer profiles highlighted above in each cluster, we can now have a better understanding of what type of entertainment app the company should consider developing as they enter this new consumer market. Given the majority of users from all clusters enjoy using TV-related and entertainment applications, it would be wise to consider developing an application that incorporates information about various TV shows from multiple networks into a single app including news, episode recaps, trailers, actor and writer profiles, etc. Another area to consider is a social music app that allows customers to create playlists of their favorite songs or artists that can be shared with their friends and family. App Happy could also consider incorporating some sort of music identification feature to identify songs they hear on the radio. Lastly, they could also consider entering the social gaming market as all segments showed significant interest in gaming apps. Due to the findings that all clusters had a high level of affinity towards free apps, they should consider offering a free version as well as a paid version of the apps that gives users additional features. This would ensure their apps cater to all customer segments.

Additional Research Opportunities

Although we have provided more insight to the customer profiles App Happy will market their apps to, there is always more opportunity for further research. One suggestion is to conduct further investigation of Cluster 1 given it represents the largest percentage of their customer base. Continuous research to further refine their customer segments is something they should always aspire to achieve. Secondly, the profile of Cluster 2 provided us with enough insight to know the customers in this segment are prime candidates for experimentation. App Happy should consider conducting experiments with this segment for some of their more ambitious app ideas in the entertainment space. Knowing that clusters 2 and 3 like to keep up with new technologies, it may be beneficial to conduct A/B testing with new features or add-ons to their apps for these two segments. They could do this by running early access or beta testing programs to gain further insight as to the wants and needs of their customer base.

Classification Model Selection

After establishing a segmentation scheme that proves to be useful in finding unique characteristics among different clustering of our customer base, App Happy can leverage our results to build a classification model to predict what segments new customers would fall into as they collect additional data from future survey questionnaires. The various models they can consider include multinomial logit or probit regression models, classification tree models, CHAID models, support vector machines (SVMs), and Random Forests. However, our recommendation is to leverage Random Forests to develop a typing tool to segment future customers. Random Forests have proven to be useful tools for classification problems and have many benefits including ease of interpretation, high levels of accuracy (even with missing values), can handle thousands of input variables, and are very efficient. By implementing a Random Forest model to classify new customers, App Happy will have a more successful marketing strategy geared specifically for each of its target customer segments.