**Customer Segmentation**

The objective of this analysis is to understand the mobile app research survey dataset, and to provide insightful ways to segment and target a specify audience.

The survey contains a series of questions that vary from demographics to detailed questions on mobile apps usage. In other to understand and segment better an audience, the demographic questions were removed to avoid biased segmentation. This step only assumes that would be no reason to, for example, differentiate between a 24-year-old respondent with a 26-year-old respondent.

In an attempt to create a targeted strategy using the results of this survey, approximately 1500 observations were made. This report presents insights into factors that helped clustering the defined groups together.

**ANALYSIS**

Before generating insights from this analysis, the dataset was explored, to comprehend the data and ensure that it is consistent, usable and of good quality. The quality of insights generated by the analysis is highly reliant on the quality of the data. The data did not need any cleaning, all values were present and did not show any signs of anomalies. The only thing worth note is that the question 10 (“Do you own any of the following smartphones…?”) had no respondents for response 10, therefore it did not have to be included in the analysis.

**ALGORITHMS**

Two different unsupervised learning algorithms were used to divide the data and create insightful targeted audiences for a better understanding of the customers.

Principal component analysis (PCA) is a technique used to emphasize variation and bring out strong patterns in a dataset. It's often used to make data easy to explore and visualize. It was used to first segment and understand which were the variables better divided the respondants, then, it was used in combination with K-Means to segment and create customer clusters, for better marketing and personalization.

K-Means clustering is a type of unsupervised learning, which is used when you have unlabeled data (i.e., data without defined categories or groups). The goal of this algorithm is to find groups in the data, with the number of groups represented by the variable K. The algorithm works iteratively to assign each data point to one of K groups based on the features that are provided. Data points are clustered based on feature similarity. The results of the K-means clustering algorithm are:

1. The centroids of the K clusters, which can be used to label new data
2. Labels for the training data (each data point is assigned to a single cluster)

Rather than defining groups before looking at the data, clustering allows you to find and analyze the groups that have formed organically.

Each centroid of a cluster is a collection of feature values which define the resulting groups. Examining the centroid feature weights can be used to qualitatively interpret what kind of group each cluster represents.

PCA and K-Means were later combined to create a easy to understand and target audience. By understanding the variables that better define each group, it is a lot easier to target the said audiences.

**INSIGHTS**

There were various very interesting insights generated by this analysis. But for the sake of this exercise, I will focus on two that most caught my attention.

The first target audience I would focus on would be a group that I would call “Apple Lovers”. This group consists of customers that have iPhone, use their phone for social networking, Facebook in special, have a above average income, and like new gadgets. This group could be target by using Facebook ads, and by creating an excited new app, that can control gadgets, and help user save time in daily activities. Those users are generally younger, and the majority is female. Creating a targeted marketing strategy to would be great, but they seem to be price sensitive, since the spending on apps is below average, this would mean that the app would most likely have to be free but include a premium feature in other to have a return on investment. This group consists of about 9% of the population.

The second group would be “Android Lovers”. As the name might suggest, this group uses android phones more often, uses their devices to listen to music and access social networks. They seem to be professionals, since they are using LinkedIn and Twitter the most. This group is more inclined to take advices from others when the subject is technology or electric products. For this second group, I would advise creating a professional app, that would target this audience. Using LinkedIn and Twitter influencer to talk about the app would seem like a good idea, seem this group is more incline to listen to people that know about technology. They seem not to be as price sensitive as the average of the population, therefore further analysis on price of the app would have to be done in order to price it correctly, but this app can be paid.

Creating a target campaign is not an easy process, but the first step is to identify those audiences. This process so far, is only based on the survey responses provided in this dataset, which only creates a hypothesis that has to be tested and explored.

**CONCLUSION**

Targeting the groups mentioned above would create more confidence in the marketing strategy, making this approach a lot more personalized than creating a regular marketing approach. Unsupervised learning is a great tool to take out the human biased out of the equation and focus on clusters of customers that are truly similar.