Team 5: Learner Crew

Natalie Chen, Daniel Chung, Su Park, Kylene Ye

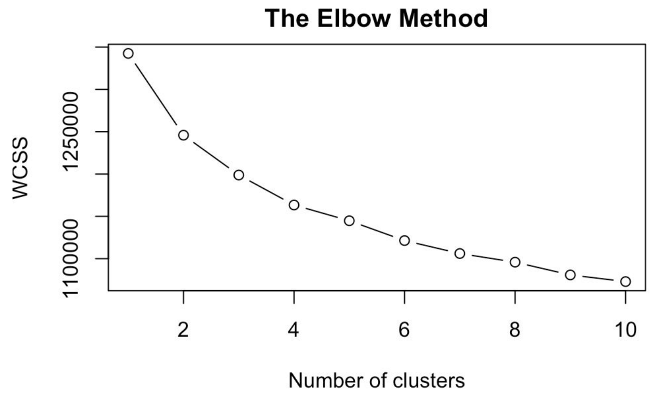
Introduction

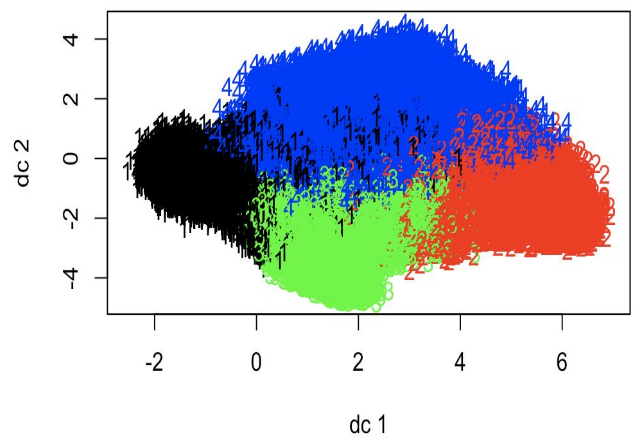
Our client is a boutique eyewear company with 3 main products, which respectively target 3 distinct audience groups - women in 18-34, men in 18-34, and adults in 18-34. As their agency (Mindshare), our goal was to deliver a tailored audience engagement strategy that will increase the sales volume by 15 % while keeping the media budget under $70MM.

Phase 1: Audience Analysis

We were provided 3 different data sets, which we utilized for different phases of our analysis. In Phase 1, we used the Audience Listening data set to segment the target audience based on various criterion including demographics, psychographics, communication behaviors and media use. We began by examining the dataset and looking for missing values, in case the data required further cleaning. While running multiple iterations of K-means clustering (using R Tidyverse Library), we searched for patterns in the dataset by filtering based on certain demographic profiles and removed variables that revealed less meaningful insights on media consumption and clothes purchasing habits. We chose K-means over K-modes clustering - despite the fact that the Audience Listening dataset is entirely consisted of categorical variables - because the responses were either 0 or 1, hence getting the average value of response for each column would help us distinguish each group better down to decimals. Using the Elbow method to estimate the optimal number of clusters, we finally came up with 4 main groups - Boomers and Gen X, Millennials, the Elderly, and the unidentified/under-represented.

Since these groups didn’t necessarily match the provided target audience profiles, we decided to further analyze unique behavioral characteristics for each group to see how we can more efficiently market our client’s products. One insight was that Boomers and Gen X have a higher interest in outdoors than Millenials and preferred traditional print media. Another was that Millennials were more trendy, interested in corporate social initiatives, liked customization, and most likely to own a SMART TV. These factors were taken into consideration in our Media plan.



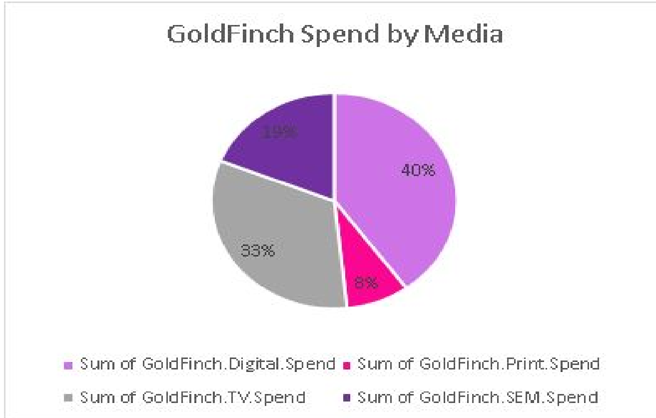


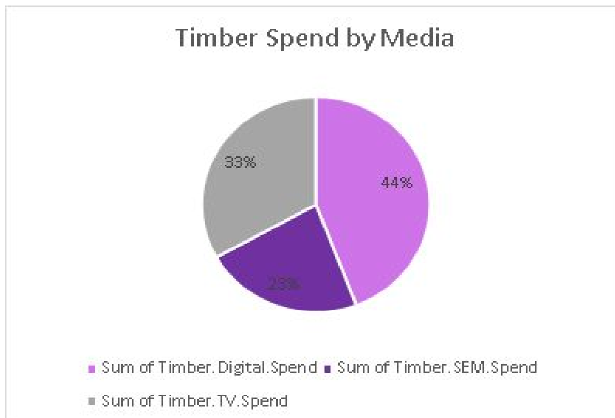
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1		Cluster	Adults.18.49.and	Men	Women	Parents	Educ..graduated.	Educ..attended.c	Educ..graduated.	Educ..did.not.gra	Educ..post.gradu	Educ..no.college	Age.18.24	Age.25.34	Age.35.44	Age.45.54
2	1	1	0.005974735	0.408586548	0.591413452	0.021509047	0.184448617	0.297029703	0.353106862	0.165414817	0.065039263	0.51852168	0.000341413	0	0.008279276	0.105155343
3	2	2	0.884878509	0.57136117	0.42863883	0.675945585	0.562789966	0.303216167	0.113076984	0.020916883	0.218133208	0.133993866	0.040968782	0.218526382	0.486356845	0.254147991
4	3	3	0.009017953	0.575577066	0.424422934	0.129229751	0.520311078	0.297096054	0.160337553	0.022255316	0.229999173	0.182592868	0.000661868	8.27E-05	8.27E-05	0.292132043
5	4	4	0.767737238	0.497047164	0.502952836	0.398268748	0.17352965	0.404902516	0.275058652	0.146509182	0.039964404	0.421567834	0.273278861	0.404740717	0.196666936	0.118760618

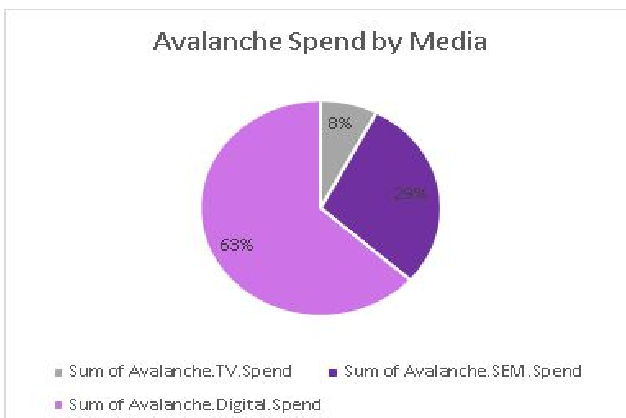
Phase 2: Media Analysis

In Phase 2, we analyzed the client’s current media buying patterns by exploring the media mix and Cost Per Acquisition metric and used the Brand Consideration Rate and NYC (Central Park) weather data to identify seasonal trends.

Various media buying patterns among the three different products were explored with pie charts.



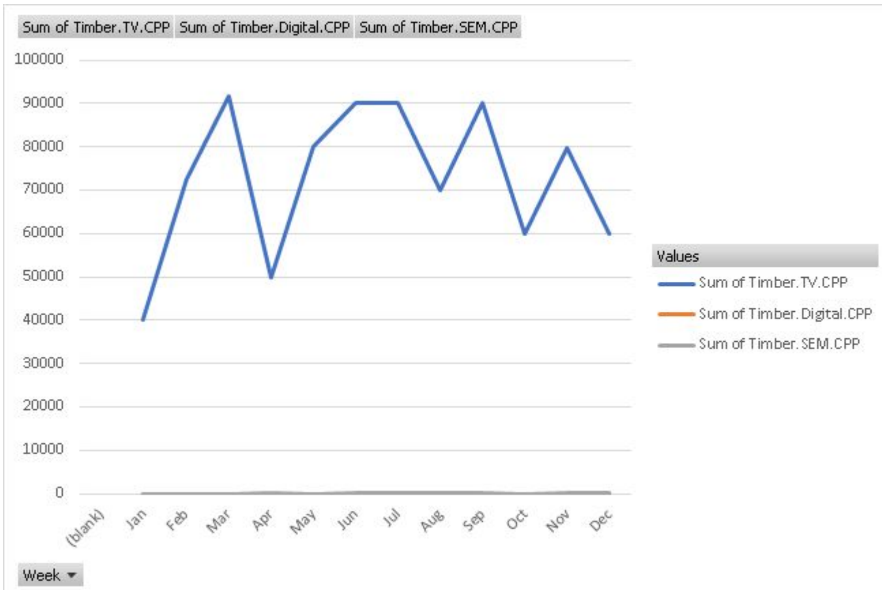




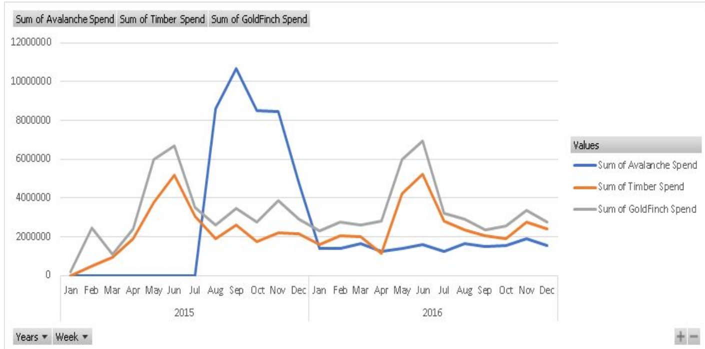
The majority of media budget for all three products was on digital. Both Gold Finch and Timber spent a significant amount on TV and SEM. Gold Finch is the only medium that advertises using Print, which comprised the lowest % of the brand's total spend. The second most popular medium for Avalanche was SEM. We can also see that Avalanche is only advertised digitally, with minimal investment in TV.

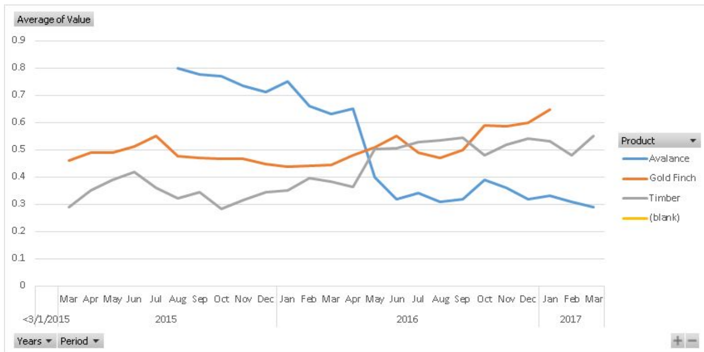
When we take a closer look at media spread, we see that Digital and SEM are the most efficient for all products, while TV is the least efficient. One hypothesis we had was that TV's relative inefficiency may be due to heavy reliance on Linear TV, instead of Programmatic TV which allows targeting based on time, household IP and Geos, since linear TV cannot track impressions.

Note: Digital CPP (in orange) is so low that it cannot be seen on the graph.



We also took Brand Consideration Rate into account in our media mix. We noticed that there was a significant drop in Avalanche's brand consideration percentage around April 2016. When we investigated, we noticed that this drop corresponded with a pause in TV spend in January 2016. By April, skiing season in North America ends & the brand is no longer under consideration for many skiers. This lead us to hypothesize that a mixed media approach is crucial for brand consideration.



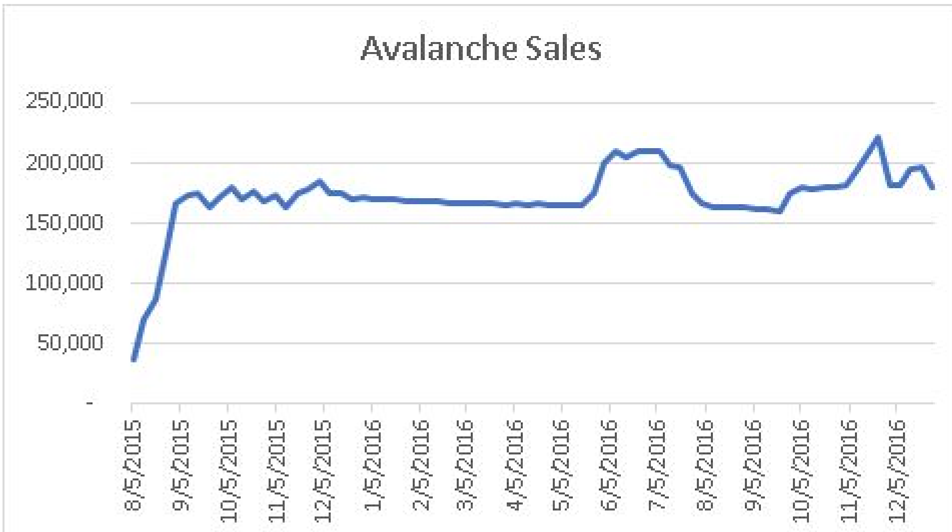


We also explored seasonality and weather information in relation to sales, which we hypothesized would be higher in places that were hotter and sunnier with multivariate regression in the Glm package in R.

$*p<.05$ $**p<.01$ $***p<.001$

		Golden Finch	Timber	Avalanche
New York City (Central Park)	Average Temperature	Positive ***	Positive ***	Positive ***
	Wind	Negative *	Negative *	Negative
	Average Humidity	Positive *	Positive	Positive
	Sunshine (hrs/day)	Positive	Positive	Negative ***

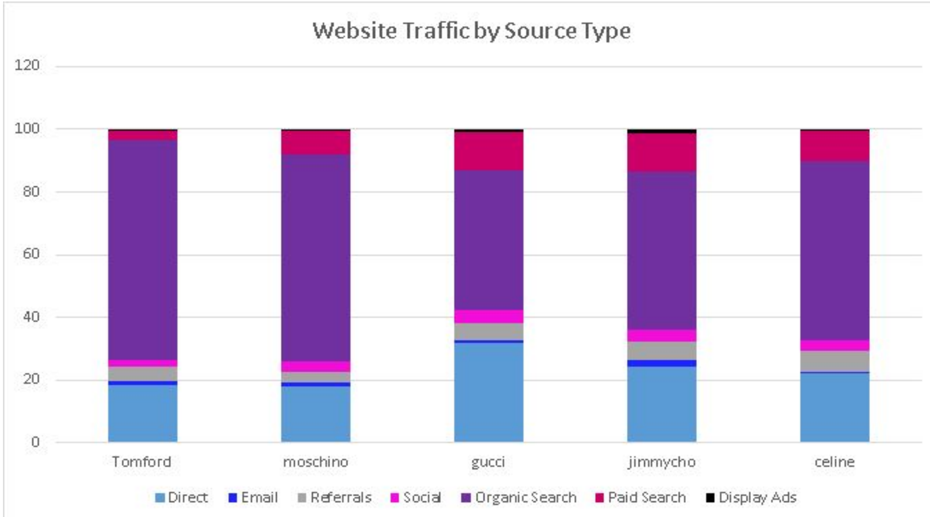
From the graph, we can see that the higher the average temperature, the greater the sales for all three days. We were puzzled as to why this would be true for Avalanche because this is counterintuitive to our hypothesis, since ski goggles are used in the winter. However, we noticed that on average, Avalanche Sales started in the mid-late summer, when temperatures are higher, in preparation for ski season, so this would make sense. Furthermore, wind was negatively correlated for Golden Finch and Timber, while humidity was positively correlated for Golden Finch. We also noticed that interestingly, number of sunshine hours per day were significantly negatively correlated with Sales. This might be because sales tend to increase around late fall and early winter, when sunlight hours are much lower.



Based on the exploratory analysis conducted in Phase 1 & 2, we crafted out an enhanced media roll-out plan for Phase 3. We designed a multi-touch attribution model for each product that assigns different variable weights to each media channel based on the target audience's media consumption patterns.

Phase 3: Media Strategy

To flesh out our competitive analysis and determine assumptions needed for our multi-touch attribution model, we examined how various competing luxury sunglasses brands (Tomford, Moschino, Gucci, Jimmy Cho, and Celine) are gathering traffic to their websites, since these metrics are easy to track. We used data collected from similarwebpro.com to discern that Display Ads seem to be the initial touch point in the consumer journey, while Search (Organic & Paid) is the last touchpoint before visiting the website.



	Tomford	Moschino	Gucci	Jimmycho	Celine
Direct	18.5%	17.8%	31.8%	24.1%	22.1%
Email	1.2%	1.2%	0.8%	2.4%	0.6%
Referrals	4.4%	3.5%	5.4%	5.8%	6.7%
Social	2.5%	3.6%	4.1%	3.6%	3.5%
Organic Search	70.3%	66.1%	45.0%	50.8%	57.2%
Paid Search	3.0%	7.3%	11.9%	12.2%	9.5%
Display Ads	0.3%	0.5%	0.9%	1.1%	0.5%

From this information, we made various assumptions in order to create our multi-touch attribution model using the Markov Chains Concept. We did this in order to understand the importance of each channel so we could create the budget plan. The aim of the model is to calculate the weight of channels that leads to conversion (not null). The model was made using the R Channel Attribution Library.

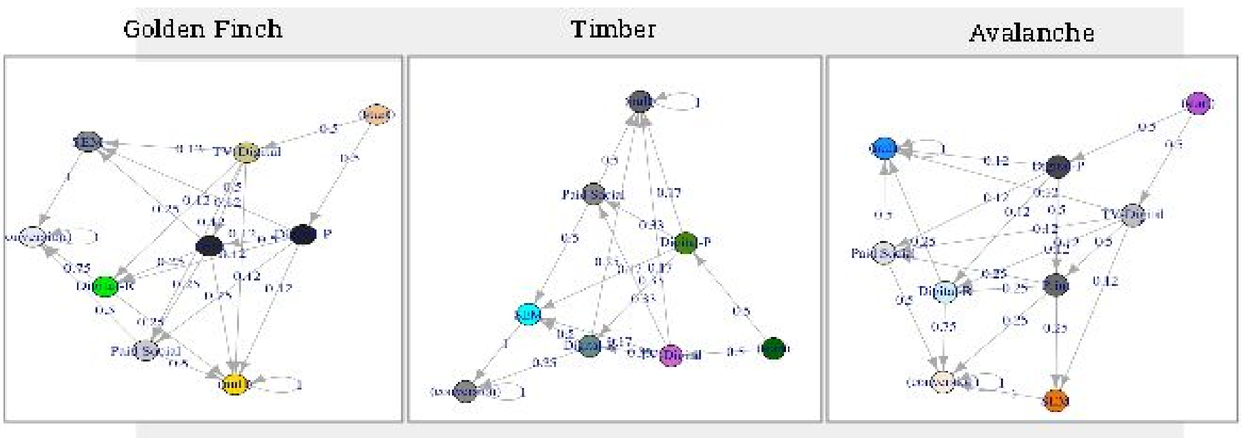
Using Markov chains allow us to switch from heuristic models to probabilistic ones. We can represent every customer journey (sequence of channels/touchpoints) as a chain in a directed Markov graph where each vertex is a possible state (channel/touchpoint) and the edges represent the probability of transition between the states (including conversion.) By computing the model and estimating transition probabilities we can attribute every channel/touchpoint.

The process for making the Markov chain attribution model was as follows:

- Frame possible channels/touchpoints.
- Order the channels and then link them into paths. The various channels include:
  - 1: Digital-P (prospecting)

  - 2: TV-Digital
  - 3: Print
  - 4: Paid Social
  - 5: SEM
  - 6: Digital-R (retargeting)
- Golden Finch and Avalanche utilize all channels but with different importance weights.
- Timber has no Print touch point, since we decided to eliminate it from the media mix based on our previous analysis.

We then generated multiple paths with the assumptions listed below and ran the model based on the Markov chains concept. We extracted the results of attribution and graphed the results.

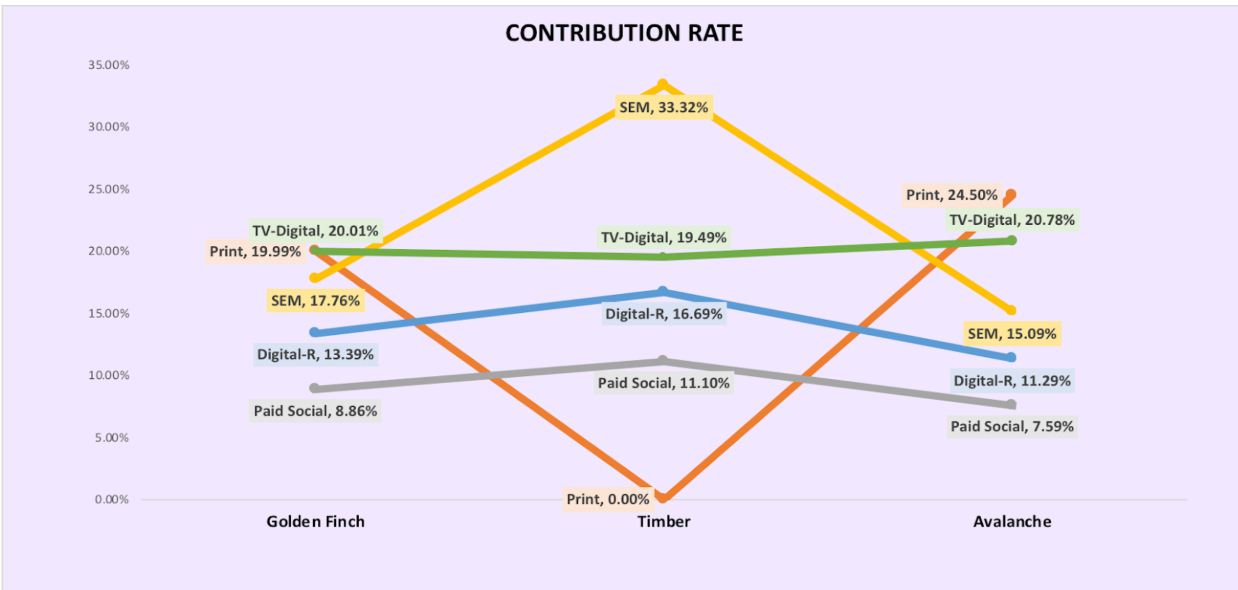


Assumptions we made in our model included:

1. Consumers cannot move vertically between initial contact channels (ex. Cannot move from Digital Prospecting to Digital TV since they are both the first contact point in the conversion funnel).
2. Cannot reverse paths, such as going from Print to TV - Digital (going back up the funnel).
3. We assumed that Digital was the most efficient medium, followed by SEM, Print, then TV, based off our previous sales analysis.
4. Cross-media exposure is more powerful of a conversion tool than only seeing an ad on one device. Using cross-media exposure, consumers convert at 1.4x the rate than if they just viewed from a single device view (see appendix for source).
5. We assume that only SEM & Cross-Device exposure are effective enough to end the user journey in two steps (see appendix for source).

6. Three-step paths will result in conversion.
7. For Avalanche, print (reputable digital magazines/online sports communities) with detailed descriptions of product functionalities is a key conversion driver.
8. Millenial women consume print more frequently than millenial men.
    a. Source: Audience Analysis

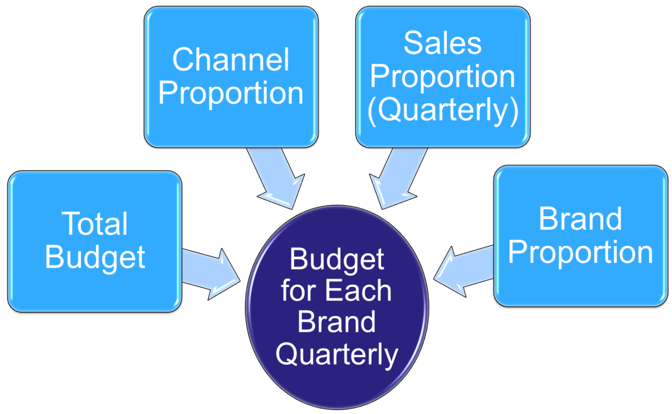
T



Based on the previous assumptions, we created the above suggestions for media mix for each product.

Budget plan

Out budget plan consists of several different variables:



We started out with the $70M we were allocated and worked backwards in order to determine quarterly allocation of each media budget. We then multiplied the $70M by the applied proportions of each variable. The three other variables include:

1. Channel Proportion
    - Each media channels' contributions to the conversion based on the attribution model mentioned above.
2. Brand Sales Portion (Quarterly)

    - The sales proportion of a given quarter within a brand (quarterly sales/total yearly sales volume per brand)
3. Brand Proportion
    - Sales proportion of the brand (Yearly sales by brand/total yearly sales)

For example, to calculate budget for Timber on Digital (Prospecting) in Q1:

= $70 Mil *Channel Portion*Brand Sales Portion(Quarterly) * Brand Portion

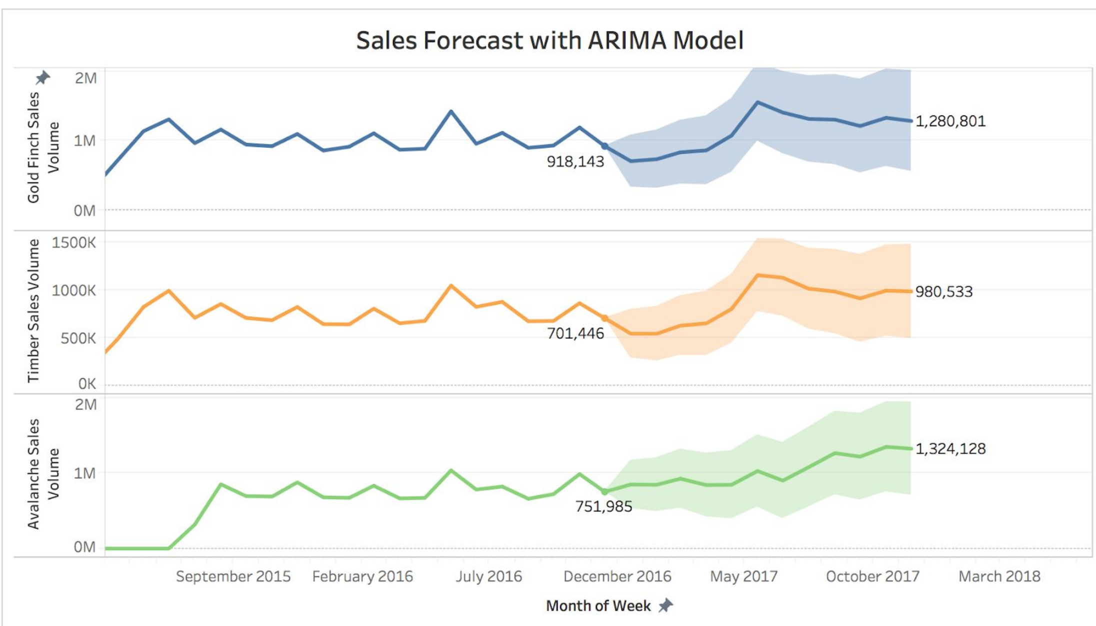
= $70 Mil * 0.194(Digital-P)* 0.115(1st Quarter Sales within Timber) * 0.316(Timber Portion)

= $470 K

TIMBER	Q1	Q2	Q3	Q4	Total
Digital-P	$470,002.03	$1,094,195.52	$1,479,893.41	$1,324,214.88	$4,368,305.84
Print	$0.00	$0.00	$0.00	$0.00	$0.00
Paid Social	$268,992.46	$626,232.07	$846,975.43	$757,877.20	$2,500,077.16
SEM	$807,178.21	$1,879,163.73	$2,541,558.59	$2,274,197.38	$7,502,097.91
Digital-R	$404,206.82	$941,019.94	$1,272,724.29	$1,138,839.08	$3,756,790.13
TV-Digital	$472,026.68	$1,098,909.03	$1,486,268.41	$1,329,919.26	$4,387,123.38
Total	$2,422,406.20	$5,639,520.29	$7,627,420.12	$6,825,047.80	**$22,514,394.41**

Sales Prediction

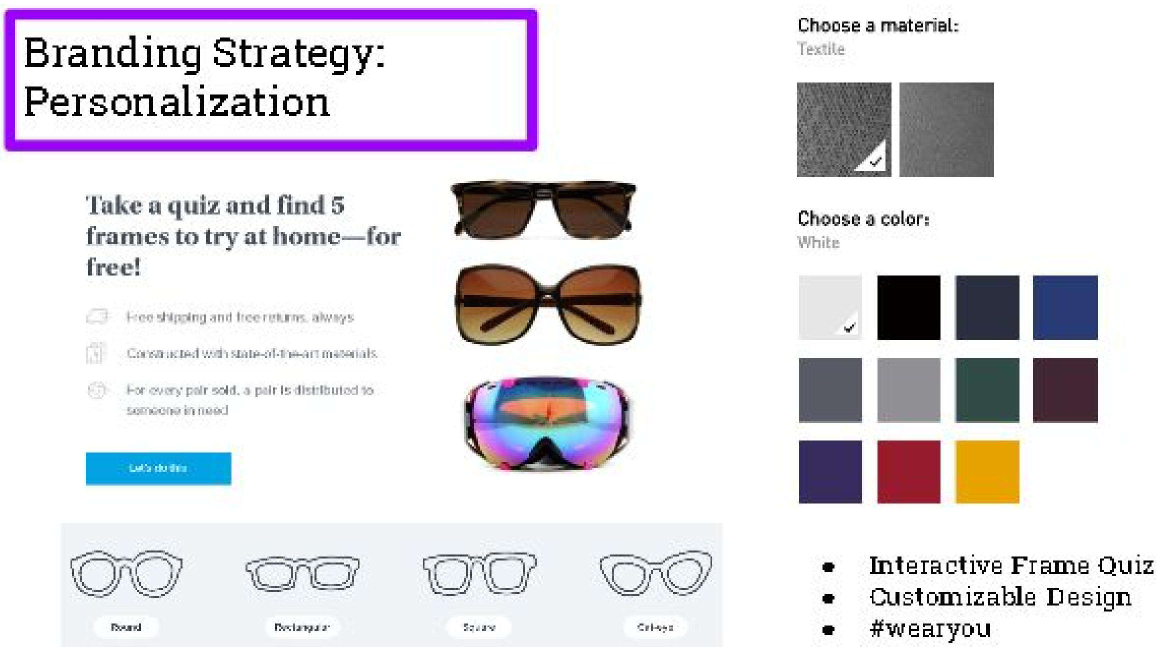
We used both R timeseries library and Tableau's forecasting feature to design an ARIMA model to predict the total sales volume for each product for the next 12 months (up to December 2017). With this prediction, we calculated the increase in sales from December 2016 to December 2017, Golden Finch is projected to increase by 39.50%, Timber by 39.79% and Avalanche by 76.08%.



Phase 4: Conclusion

For our client’s multimedia mix, there are three stages of conversion - Awareness, Consideration, and Action. Within each stage, there are various metrics that our client can keep track of in order to measure the effectiveness of their marketing campaign. For the awareness stage, the various KPIs that can be used to measure customer awareness include number of impressions, video completion rate, number of unique users, and net promoter score. For the consideration stage, metrics include click-thru rate, video watch time, and items in bag. For the action stage, metrics include registrations, calls, and sales.

Various creatives that we created for our included various marketing strategies that our client could implement to attract the target audiences for each product.



Since millenials are a big fan of customization, we thought that having them take a quiz that would determine the best frames for their face would be a major draw. Having a trending hashtag such as #wearyou would also be possible promotion strategy, emphasizing each individual's unique style.

One thing that we noticed with a lot of competing online boutique shops was that although they had a big online presence, they failed to transfer that presence outside the digital world. Our suggestion is to use influencers to help market our product or various print ads to help push consumers to a digital store, where they can take artsy photos in a unique and interactive space. This would allow for them to not only try on glasses, but increase the brand's online social presence.

In addition, content syndication would make it so that any reviews places online would automatically sync across websites and forums (ex. Posting a review on the website would cross-post to Facebook

Because Avalanche is targeting the group of people who are more family oriented and into outdoor activities, we can partner with luxurious ski resorts and car rental agencies and offer family trips to ski resorts. Also, we could also attract more customers within the family by promoting Family Discount offers.

Branding Strategy:
Rewards Program

- Members ONLY perks
  - CSR Campaign
  - Refer a friend & get a promo code
  - Free regular prescription exam
  - Free shipping & Free returns

A member's only rewards club would help incentivize sunglass and ski google users to stick with the brand and make it seem more exclusive. Members would receive free regular prescription exams with the purchase of sunglasses/ski goggles as well as enjoy free shipping and returns. Promotions such as refer a friend campaigns would help spread the brand to other similar minded consumers. Furthermore, including a CSR campaign to give a pair to someone in need every time you buy one would appeal to millennials, who are concerned about corporate sustainability and social initiatives.

In conclusion, we recommend that for Golden Finch and Timber, our client invest on cross-devices (TV/Mobile) that will enable them to keep track of various metrics in their target audience. We also suggest that they ramp up programmatic TV, radio and podcast, and music streaming (Spotify/Pandora) ads to appeal to millenials. They can also invest in influencer marketing, which millennials tend to closely follow on platforms like Instagram and Facebook. Furthermore, to appeal to Millennials' desire for custom products, we suggest that they implement a design quiz that allows users to select the ideal sunglasses for them based on their needs, face shape, preferences, and so on, which enables them to feel involved in the process.

Our recommendation for Avalanche a little bit more out of the box - we suggest that our client expand the audience pool to age 50, since many older adults prefer outdoor activities compared to Millennials. We also suggest that they invest in some print (physical & digital copies) and allocate more SEM budget to Prime-time TV, since older adults prefer using these mediums. In addition, since brand consideration dropped very harshly after Avalanche stopped advertising on TV, we suggest that the company more consistently campaign using TV and not drop it off completely, since this will likely adversely affect product consideration.

Appendix: R Code

```
#Multi-touch attribution model  for Golden Finch
library(dplyr)
library(reshape2)
library(ggplot2)
library(ggthemes)
library(ggrepel)
library(RColorBrewer)
library(ChannelAttribution)
library(markovchain)
# creating a data sample
PATH1=c('Digital-P','Digital-P > Print','Digital-P > Print > Paid Social','Digital-P > Print > SEM', 'Digital-P > Print > Digital-R','Digital-P
> Paid Social', 'Digital-P > SEM','Digital-P > Digital-R','TV-Digital','TV-Digital > Print','TV-Digital > Print > Paid Social','TV-Digital >
Print > SEM', 'TV-Digital > Print > Digital-R','TV-Digital > Paid Social', 'TV-Digital > SEM','TV-Digital > Digital-R')
CONV1=c(0,0,1,1,1,0,1,0,0,0,1,1,1,0,1,1)
CONV_NULL1=c(1,1,0,0,0,1,0,1,1,1,0,0,0,1,0,0)
df1 <- data.frame(path = PATH1, conv = CONV1, conv_null = CONV_NULL1)
# calculating the model
mod1 <- markov_model(df1, var_path = 'path', var_conv = 'conv',  var_null = 'conv_null', out_more = TRUE)
# extracting the results of attribution
df_res1 <- mod1$result
df_res1
# extracting a transition matrix
df_trans1 <- mod1$transition_matrix
df_trans1 <- dcast(df_trans1, channel_from ~ channel_to, value.var = 'transition_probability')
### plotting the Markov graph ###
df_trans <- mod1$transition_matrix
df_trans
# adding dummies in order to plot the graph
df_dummy <- data.frame(channel_from = c('(start)', '(conversion)', '(null)'), channel_to = c('(start)', '(conversion)', '(null)'),
transition_probability = c(0, 1, 1))
df_trans <- rbind(df_trans, df_dummy)

# ordering channels
df_trans$channel_from <- factor(df_trans$channel_from,
                        levels = c('(start)', '(conversion)', '(null)', 'Digital-P', 'TV-Digital','Print', 'Paid Social','SEM','Digital-R'))
df_trans$channel_to <- factor(df_trans$channel_to,
                        levels = c('(start)', '(conversion)', '(null)', 'Digital-P', 'TV-Digital','Print', 'Paid Social','SEM','Digital-R'))
df_trans <- dcast(df_trans, channel_from ~ channel_to, value.var = 'transition_probability')
# creating the markovchain object
trans_matrix <- matrix(data = as.matrix(df_trans[, -1]),  nrow = nrow(df_trans[, -1]), ncol = ncol(df_trans[, -1]), dimnames =
list(c(as.character(df_trans[, 1])), c(colnames(df_trans[, -1]))))
trans_matrix[is.na(trans_matrix)] <- 0
trans_matrix1 <- new("markovchain", transitionMatrix = trans_matrix)
# plotting the graph
plot(trans_matrix1, edge.arrow.size = 0.65)

#For Timber creating a data sample
PATH2=c('Digital-P','Digital-P > Paid Social','Digital-P > Paid Social > SEM','Digital-P > SEM', 'Digital-P > Digital-R','Digital-P >
Digital-R > SEM', 'TV-Digital','TV-Digital > Paid Social','TV-Digital > Paid Social > SEM','TV-Digital > SEM', 'TV-Digital >
Digital-R','TV-Digital > Digital-R > SEM')
CONV2=c(0,0,1,1,0,1,0,0,1,1,1,1)
CONV_NULL2=c(1,1,0,0,1,0,1,1,0,0,0,0)
df2 <- data.frame(path = PATH2, conv = CONV2, conv_null = CONV_NULL2)
# calculating the model
mod2 <- markov_model(df2, var_path = 'path', var_conv = 'conv',  var_null = 'conv_null', out_more = TRUE)
# extracting the results of attribution
df_res2 <- mod2$result
df_res2
# extracting a transition matrix
df_trans2 <- mod2$transition_matrix
df_trans2 <- dcast(df_trans2, channel_from ~ channel_to, value.var = 'transition_probability')
### plotting the Markov graph ###
df_trans <- mod2$transition_matrix
```

```
df_trans
# adding dummies in order to plot the graph
df_dummy <- data.frame(channel_from = c('(start)', '(conversion)', '(null)'), channel_to = c('(start)', '(conversion)', '(null)'),
transition_probability = c(0, 1, 1))
df_trans <- rbind(df_trans, df_dummy)

# ordering channels
df_trans$channel_from <- factor(df_trans$channel_from,
                               levels = c('(start)', '(conversion)', '(null)', 'Digital-P', 'TV-Digital','Print', 'Paid Social','SEM','Digital-R'))
df_trans$channel_to <- factor(df_trans$channel_to,
                               levels = c('(start)', '(conversion)', '(null)', 'Digital-P', 'TV-Digital','Print', 'Paid Social','SEM','Digital-R'))
df_trans <- dcast(df_trans, channel_from ~ channel_to, value.var = 'transition_probability')
# creating the markovchain object
trans_matrix <- matrix(data = as.matrix(df_trans[, -1]),  nrow = nrow(df_trans[, -1]), ncol = ncol(df_trans[, -1]), dimnames =
list(c(as.character(df_trans[, 1])), c(colnames(df_trans[, -1]))))
trans_matrix[is.na(trans_matrix)] <- 0
trans_matrix1 <- new("markovchain", transitionMatrix = trans_matrix)
# plotting the graph
plot(trans_matrix1, edge.arrow.size = 0.65)

# creating a data sample for Timber
PATH2=c('Digital-P','Digital-P > Paid Social','Digital-P > Paid Social > SEM','Digital-P > SEM', 'Digital-P > Digital-R','Digital-P >
Digital-R > SEM', 'TV-Digital','TV-Digital > Paid Social','TV-Digital > Paid Social > SEM','TV-Digital > SEM', 'TV-Digital >
Digital-R','TV-Digital > Digital-R > SEM')
CONV2=c(0,0,1,1,0,1,0,0,1,1,1,1)
CONV_NULL2=c(1,1,0,0,1,0,1,1,0,0,0,0)
df2 <- data.frame(path = PATH2, conv = CONV2, conv_null = CONV_NULL2)
# calculating the model
mod2 <- markov_model(df2, var_path = 'path', var_conv = 'conv',  var_null = 'conv_null', out_more = TRUE)
# extracting the results of attribution
df_res2 <- mod2$result
df_res2
# extracting a transition matrix
df_trans2 <- mod2$transition_matrix
df_trans2 <- dcast(df_trans2, channel_from ~ channel_to, value.var = 'transition_probability')
### plotting the Markov graph ###
df_trans <- mod2$transition_matrix
df_trans
# adding dummies in order to plot the graph
df_dummy <- data.frame(channel_from = c('(start)', '(conversion)', '(null)'), channel_to = c('(start)', '(conversion)', '(null)'),
transition_probability = c(0, 1, 1))
df_trans <- rbind(df_trans, df_dummy)

########################################

#K-means Clustering
```{r}
data<-read.csv('Audience.2.csv')
```
```{r}
# Using the elbow method to find the optimal number of clusters
set.seed(6)
wcss = vector()
for (i in 1:10) wcss[i] = sum(kmeans(data, i)$withinss)
plot(1:10,
     wcss,
     type = 'b',
     main = paste('The Elbow Method'),
     xlab = 'Number of clusters',
     ylab = 'WCSS')
# 4 clusters should be optimal.
```
```{r}
# Fitting K-Means to the dataset
```
```

````
set.seed(29)
kmeans = kmeans(x = data, centers = 4)
y_kmeans = kmeans$cluster
```
```{r}
library(cluster)
library(fpc)
data(data)
```
````

Sources

Cross-device is more powerful than single-device view. Using cross-device view, consumers convert at 1.4x the rate than if you just viewed from a single device view.

a. https://www.mobilemarketer.com/ex/mobilemarketer/cms/opinion/columns/24660.html

Only SEM & Cross-Device are effective enough to end the user journey in two steps.

b. Google Analytics Click-Through-Rate Report

c. https://www.wordstream.com/blog/ws/2016/02/29/google-adwords-industry-benchmarks
````