BUNK 742: Advanced Marketing Analytics

Project #2: Analyzing Print Ad Designs Using Eye-Movement Data

Group #1
Shiyun Shen, Yingzheng Xue, Michael Rees, Zihan Cai, Jana Haas, Ziying Yan

October 22nd, 2019

Honor Pledge:
We pledge on our honor that we have not given or received any unauthorized assistance on this assignment.

Executive Summary

We were given this assignment to improve brand recognition for L’Oreal by conducting a detailed analysis that turns the data gathered from the eye-tracking research into actionable information that L’Oreal can use when creating their magazine ads. To sort through and analyze the data we created a binomial logit model, poisson regression model, and multiple regression model to examine how different variables affect the overall advertising and brand memory of a viewer. We found that overall the most effective advertisements include a bigger brand size and pictorial size that are located on the right of the page and towards the end of the magazine.

Introduction and Background

L’Oreal, founded in 1919, has been one of the leading companies in the beauty industry. In this ever-changing field; however, profound history does not speak for itself. Plenty of new brands emerge every year. In order to defend its leading position, L’Oreal must constantly improve their marketing strategy. Maximizing brand recall is a great starting point. L’oreal is interested in finding out the best way to attract viewers’ attention to their brand versus other brands and maximizing their memory of the advertisements and their brand after viewing it.

L’Oreal is struggling to identify where the picture should be located on the page and in the magazine along with the size of the brand and the image that accompanies it. In order to do so we will be conducting an analysis that examines data collected from an eye tracker using three different types of models. An eye-tracker tracks the movement of the eye, it measures when the eyes pause (fixations) and when there are quick jumps to different areas of the advertisement (saccades). Fixations are used to analyze the advertisement and its effectiveness. One of the main problems with all of the eye tracking data collected is that it cannot be analyzed efficiently. Our job is to analyze the data and answer several questions that observe a number of variables, all of these factors are measured while looking at a magazine, as this is L'oreal’s biggest struggle today.

Data and Methodology

The data we were given are the results of an eye-movement study. Eighty eight participants were asked to page through an issue of the magazine *Cosmopolitan* which included thirty five advertisements. Their eye-movement was recorded with infra-red corneal reflection eye tracking methodology using two forms: brand fixation and picture fixation. Apart from the IDs of respondents and advertisements, the dataset also includes basic information about the advertisements including page number, page position, brand size and the size of pictorial element. The results were provided in the form of recall accuracy and recall time.

For our methodology, aside from some of the variables mentioned, we created a new dummy variable called L’Oreal to test the overall impact the brand L’Oreal had on its respondents if there is any at all. If the brand is L’Oreal it becomes 1 if it refers to other brands it becomes 0. The interaction effect between brand size and the size of pictorial elements were also taken into account in all our models, for the reason that ad sizes are limited in magazines, and one is too large or too small would reasonably alter the effect of the other.

We utilized the binary logit model, poisson regression model, and multiple regression model, general linearized models that use a great variety of variables. To be more specific, binary logit models, poisson regression, and multiple regression are developed to handle the relationship of a dependent dummy variable, count variables, and continuous variables respectively. In our attempt to assess the effect on brand fixation and picture fixation, we involved the poisson regression model. These two are count variables, and the model included, brand_size, pic_size, page_pos, page_num, and brand_size*pic_size. Likewise, the binary logit model was used to appraise the influence on recall accuracy. As for continuous variable recall time, we employed the multiple regression model. Compared to the two poisson regression models above, variables brand_fix and pic_fix are added in the rest models. (See Appendix Figure 11) According to the result of the binary logit model for recall accuracy, brand_fix and pic_fix have a negative effect on recall accuracy. However, the fact that these two parameters have a positive effect on brand_fix and pic_fix, while brand_fix and pic_fix also have a positive effect on recall accuracy, suggests the probability of multicollinearity in the model, so we remove variables above one by one to compare the BIC. The smallest BIC demonstrates that the model without variable brand_size is of the highest quality when comparing with others. (See Appendix Figures 9 and 10) Consequently, we dismiss brand_size in this model.

Key Findings

From the results of Poisson Regression for brand fixation count, it can be concluded that every variable, except page_num whose p-value (p=0.59) is over 0.05, has a significant effect on brand fixation. Among all these significant variables, if brand_size increases by 1 square inch, the number of brand fixations would increase by 22%. While one square inch increase in picture size would result in a 1.8% increase in brand fixation. As for the position of the ad, being on the right side of the page is 79% more likely to attract brand fixation than on the left, holding other variables constant. The next variable Loreal, advertisement appearing on L'Oreal is expected to decrease the probability of brand fixation by 59%, controlling for other variables. The last variable we use to control for interaction between brand size and picture size and to control for this possible effect on the overall model. Finally, holding other variables are 0, the estimated brand fixation rate will be 0.19(See Appendix Figure 5). For brand fixation count, brand size has a stronger effect than picture size. However, brand size is not significant in pictorial fixation count. We can conclude that people's eyes are more attracted by a picture and then whether they will notice the brand is related to the size of the brand. For the recall accuracy, brand fixation has a stronger effect than pictorial fixation, which means more attractive brands that increase fixations lead to better recall of the brand.

According to Poisson Regression for Pictorial Fixation Count, all the variables involved have significant effects on pic_fix except brand_size. To be more specific, if pictorial element size increases by 1 square inch, the fixation on pictorial element will witness a 13.7% growth. And being on the left side of the page indicates that the ad has a higher rate of 0.15% to attract fixation on its pictorial element. Additionally, as the number of pages rises, the possibility of pic_fixation is 0.15% weaker by every page. The pic fixation on L'oreal is 27.49% less than other brands. (See Appendix Figure 6). The last estimated pictorial fixation rate is 1.27 when controlling other variables are 0.

In the binary model, brand_fix variable and pic_fix variable are significant in the model because the p-values of both are smaller than 0.05. Also, the effects of the pic_size variable, page_pos variable and page_num are all significant when it comes to the Binary Logit Model for Accuracy=1. Also, the interaction effect between brand_size and pic_size is also significant. And when the dummy variable L'Oreal is added, its effect is also significant. It means whether the brand is L'Oreal will make a big difference in respondents' recall memory. To be more specific, all the variables in the model contribute to recall accuracy and help the memory when holding other variables constant. And, one more fixation on the brand increases the odds of recalling the ads by 6.24%. One more fixation on the picture increases the odds of recalling the ads by 3.84%. In addition, if the picture size increases by 1 squared-inch, the odds of recalling accuracy will drop by 1.17%. If page number adds one page, the recall accuracy will increase by 0.78%. As for the position of the advertisement, appearing on the right side of the page is 16.93% more likely to attract brand fixation than on the left (See Appendix Figure 7).

A multiple regression model is adopted to assess the effect on recall_time, another important factor to measure the effectiveness of the ads. There are five significant explanatory variables: brand_size, pic_size, page_number, brand_size*pic_size, and L'Oreal. With one squared-inch rise in brand size, recall time is 8.55% longer. Similarly, the larger the picture size, the longer the recall time with 0.70% increase for one squared-inch growth. Additionally, the variable of page number shows a recency effect, indicating that if the ad appears one page behind, 0.33% less time it takes for participants to remember it. Speaking of our brand, participants need 24.89% time fewer than others to remember the advertisement. When the brand_size and pic_size is zero, the ad which is not L'oreal is on page zero, and there's no interaction effect between brand_size and pic_size involved, the recall time of the ad 7.82 (See Appendix Figure 8)

Conclusions and Recommendations

Some ads yielded 0 fixations, so there may have been errors in the collection of the data, with some participants not paying any attention whatsoever to some ads. This may also just be due to poorly designed ads. Without detailed information, we are not capable of reaching further conclusions. The next limitation we noticed was prior recognition of the brand. If someone knows the brand, the fixation length could be impacted and skew results. This could also affect the recall results if they already would have recognized the brand before viewing the ad. We are trying to establish a non-influenced conclusion that can be utilized by all brands, along with being specific to L'Oreal. The final limitation refers to how the results of the multiple regression model showed potential mistakes in the data set. If there are mistakes in the data set, the results and conclusion we have come upon could be skewed or inaccurate.

Despite these limitations, we are still able to give L'Oreal along with all of the other companies well-researched recommendations. For L'Oreal, we recommend that they improve their advertisements by increasing the brand size and picture size, along with placing the advertisement towards the end of the magazine - the page number should be higher. This will help with pictorial and brand fixation as well as recall time, increasing brand recognition generally. Through the use of the poisson model to increase fixation count or reaching the goal of greater attention, both the brand size and picture size need to increase. Additionally, the brand or picture should be located on the right side of the page and it should be located closer to the end of the magazine, meaning that the page number ish high. Based on the Binary Logit model, to increase recall accuracy or if enhancing brand memory is the goal, we recommend that the picture and brand are placed on the right side of the page as this helps people remember what they have looked at more - increase memory. Lastly, based on the generalized linear model, in order to increase recall time, brand size and picture size should increase. In the future for this study, we recommend that there is great expansion on why the r-squared for recall time is so low. This will increase our confidence in the models we are using and the recommendation we give.

Appendix

Figure 1: Summary of Models and Variables

Model Type	Dependent Variable	Variable Type	Independent Variable Considered	Independent Variable Used	Independent Variable Removed	Removing Reason
Poission Regression	brand_fix	count	brand_size pic_size page_num page_pos brand_size*pic_size LOreal	brand_size pic_size page_pos brand_size*pic_size LOreal	page_num	Insignificant
Poission Regression	pic_fix	count	brand_size pic_size page_num page_pos brand_size*pic_size LOreal	pic_size page_pos page_num brand_size*pic_size Loreal	brand_size	Insignificant
Binary Logit Model	recall_accu	dummy	brand_fix pic_fix brand_size pic_size page_num page_pos brand_size*pic_size LOreal	brand_fix pic_fix pic_size page_num page_pos brand_size*pic_size LOreal	brand_size	With lowest BIC
Multiple Regression Model	recall_time	continuous	brand_fix pic_fix brand_size pic_size page_num page_pos brand_size*pic_size LOreal	brand_size pic_size page_num brand_size*pic_size LOreal	brand_fix pic_fix page_pos	Insignificant

Figure 2: Descriptive Statistics

The MEANS Procedure

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
brand_fix	BRAND.FIXATIONS	3080	1.7136364	2.7259490	0	32.0000000
pic_fix	PICTORIAL.FIXATIONS	3080	4.9126623	4.9636387	0	59.0000000
brand_size	BRAND.SURFACE	3080	6.8809143	5.6980995	1.8150000	25.1660000
pic_size	PICTORIAL.SURFACE	3080	67.5134571	13.4098524	24.0220000	88.0900000
page_num	PAGE.NUMBER	3080	56.8000000	40.6167970	2.0000000	127.0000000
page_pos	LEFT.RIGHT.LOCATION	3080	0.5714286	0.4949520	0	1.0000000
Loreal		3080	0.0285714	0.1666257	0	1.0000000

Figure 3: Correlation Matrix

Pearson Correlation Coefficients, N = 3080
Prob > |r| under H0: Rho=0

	brand_fix	pic_fix	brand_size	pic_size	page_num	page_pos	Loreal
brand_fix **BRAND.FIXATIONS**	1.00000	0.30909 <.0001	0.35204 <.0001	0.05317 0.0032	0.05052 0.0050	0.20533 <.0001	-0.07279 <.0001
pic_fix **PICTORIAL.FIXATIONS**	0.30909 <.0001	1.00000	-0.05076 0.0048	0.19879 <.0001	0.06624 0.0002	0.02838 0.1153	-0.07670 <.0001
brand_size **BRAND.SURFACE**	0.35204 <.0001	-0.05076 0.0048	1.00000	0.07060 <.0001	0.05344 0.0030	0.15370 <.0001	-0.07522 <.0001
pic_size **PICTORIAL.SURFACE**	0.05317 0.0032	0.19879 <.0001	0.07060 <.0001	1.00000	-0.06478 0.0003	-0.07561 <.0001	-0.19093 <.0001
page_num **PAGE.NUMBER**	0.05052 0.0050	0.06624 0.0002	0.05344 0.0030	-0.06478 0.0003	1.00000	0.08246 <.0001	-0.21875 <.0001
page_pos **LEFT.RIGHT.LOCATION**	0.20533 <.0001	0.02838 0.1153	0.15370 <.0001	-0.07561 <.0001	0.08246 <.0001	1.00000	0.14852 <.0001
Loreal	-0.07279 <.0001	-0.07670 <.0001	-0.07522 <.0001	-0.19093 <.0001	-0.21875 <.0001	0.14852 <.0001	1.00000

Figure 4: Brand Names and Frequency

The FREQ Procedure

AD.BRAND				
brand	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Alfa	88	2.86	88	2.86
Appelsientje	88	2.86	176	5.71
Arden	88	2.86	264	8.57
Armani	88	2.86	352	11.43
Astnor	88	2.86	440	14.29
Barclay	88	2.86	528	17.14
Cartoon	88	2.86	616	20.00
Chantelle	88	2.86	704	22.86
Citroen	88	2.86	792	25.71
ClearBlue	88	2.86	880	28.57
Conimex	88	2.86	968	31.43
Corona	88	2.86	1056	34.29
Daihatsu	88	2.86	1144	37.14
Dali	88	2.86	1232	40.00
Davidoff	88	2.86	1320	42.86
DelMonte	88	2.86	1408	45.71
Dior	88	2.86	1496	48.57
Fancy	88	2.86	1584	51.43
Gauloises	88	2.86	1672	54.29
Gilette	88	2.86	1760	57.14
Kapper	88	2.86	1848	60.00
Kodak	88	2.86	1936	62.86
LOreal	88	2.86	2024	65.71
Lancaster	88	2.86	2112	68.57
Lancome	88	2.86	2200	71.43
Legere	88	2.86	2288	74.29
Margriet	88	2.86	2376	77.14
MarieJo	88	2.86	2464	80.00
Nivea	88	2.86	2552	82.86
SiSi	88	2.86	2640	85.71
Silan	88	2.86	2728	88.57
Spa	88	2.86	2816	91.43
Stuyvesant	88	2.86	2904	94.29
VanderBilt	88	2.86	2992	97.14
Viva	88	2.86	3080	100.00

Figure 5: Analysis of Maximum Likelihood Parameter Estimates for Brand Fixation

Analysis Of Maximum Likelihood Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	-1.6760	0.1484	-1.9667	-1.3852	127.63	<.0001
brand_size	1	0.1992	0.0141	0.1715	0.2269	198.98	<.0001
pic_size	1	0.0182	0.0020	0.0143	0.0222	80.80	<.0001
page_pos	1	0.5823	0.0324	0.5189	0.6457	323.74	<.0001
brand_size*pic_size	1	-0.0018	0.0002	-0.0022	-0.0015	92.77	<.0001
Loreal	1	-0.8989	0.1465	-1.1859	-0.6118	37.67	<.0001
Scale	0	1.0000	0.0000	1.0000	1.0000		

Figure 6: Analysis of Maximum Likelihood Parameter Estimates for Pictorial Element Fixation

Analysis Of Maximum Likelihood Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	0.2436	0.0553	0.1351	0.3520	19.39	<.0001
pic_size	1	0.0185	0.0007	0.0171	0.0200	640.68	<.0001
page_pos	1	0.1285	0.0171	0.0951	0.1619	56.80	<.0001
page_num	1	0.0015	0.0002	0.0011	0.0019	57.72	<.0001
pic_size*brand_size	1	-0.0002	0.0000	-0.0002	-0.0001	83.89	<.0001
Loreal	1	-0.3214	0.0682	-0.4551	-0.1878	22.23	<.0001
Scale	0	1.0000	0.0000	1.0000	1.0000		

Figure 7: Analysis of Maximum Likelihood Parameter Estimates for Recall Accuracy

Analysis Of Maximum Likelihood Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	0.2436	0.0553	0.1351	0.3520	19.39	<.0001
pic_size	1	0.0185	0.0007	0.0171	0.0200	640.68	<.0001
page_pos	1	0.1285	0.0171	0.0951	0.1619	56.80	<.0001
page_num	1	0.0015	0.0002	0.0011	0.0019	57.72	<.0001
pic_size*brand_size	1	-0.0002	0.0000	-0.0002	-0.0001	83.89	<.0001
Loreal	1	-0.3214	0.0682	-0.4551	-0.1878	22.23	<.0001
Scale	0	1.0000	0.0000	1.0000	1.0000		

Figure 8: Analysis of Estimates for Recall Time

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	2.056415221	0.15710718	13.09	<.0001
brand_size	0.082076427	0.02386986	3.44	0.0006
pic_size	0.007023561	0.00214098	3.28	0.0010
page_num	-0.003298250	0.00045713	-7.22	<.0001
brand_size*pic_size	-0.001039350	0.00032964	-3.15	0.0016
Loreal	-0.286245114	0.11341732	-2.52	0.0117

Figure 9: Criteria For Assessing Goodness Of Fit for Binary Logit Model without Brand_size

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Log Likelihood		-2018.6555	
Full Log Likelihood		-2018.6555	
AIC (smaller is better)		4053.3110	
AICC (smaller is better)		4053.3579	
BIC (smaller is better)		4101.5725	

Figure 10: Summarization of BIC or Binary Logit Model

Variable Removed	BIC
None	4104.2438
brand_fix	4111.1464
pic_fix	4115.0657
page_pos	4101.9438
brand_size	4101.5725
pic_size	4115.212
page_num	4162.1821
brand_size*pic_size	4104.5958
Loreal	4117.545

The right column contains the BIC of Binary Logit Models. Each model has one variable from the left column removed. Also, the left column corresponds with the right column.

As shown in the chart, the BIC is the smallest when excluding brand_size.

Figure 11: Model Equations

Poisson Regression→

$ln(E[Y\,i]) =- 1.67 + .199 * brandsize\ .018\ * picsize + .582 * pg\ pos\ - .001 * (brandsize * picsize) - .900 * L'Oreal$

$ln(E[Y\,i]) = .244\ +\ .019\ * picsize + .219 * pgpos + .002 * pgnumb - .0002 * (picsize * brandsize) - .321 * L'Oreal$

Binary logit model →

$ln(\frac{pi}{1-pi}) =- .129 + .061 * brandfix + .377 * picfix + .156 * pgpos - .012 * picsize + .008 * pgnum + .0004 * (brandsize * picsize) - 1.19 * L'Oreal$

Multiple linear regression →

$ln(E[Y\,i]) = .056 + .082 * brandsize + .007 * picsize - .003 * pgnumb - .001 * (brandsize * picsize) - .286 * L'Oreal$

Figure 12: Descriptive Statistics explanation

The number of fixations is a count of the number of times the participant's eyes focused briefly on either the brand logo or the picture. The mean of brand_fix is 1.71 fixations and the mean of pic_fix is 4.91 fixations, these both represent the average number of times that someone fixated on either the picture or the brand. The next variable that was examined was brand_size, which had a mean of 6.88, it serves as the surface size of the brand element in inches squared. Similarly, the mean of the pic_size is 67.51, and like the brand_size it relays that the surface size of the pictorial element in inches squared. On average, the picture is clearly much larger than the brand. The following set of variables includes page_num and page_pos. The mean - 56.80 - of the page_num stands for the page number in the magazine where the advertisement appears. We also found that the mean for page_pos we can conclude that 57.14% of the advertisements were on the right side of the magazine. Finally, the added variable named L'Oreal. This is a dummy variable. L'Oreal accounted for 2.86% of the ads, or in other words appeared only once in the magazine (on page 5) in addition to all of the other advertisements only appearing once. The standard deviation represents how much the data varies from the mean. A low standard deviation means that there is not much deviation in the data while high one implies high divergence from the mean. The variables that had a high standard deviation included pic_size which had a standard deviation of 13.41 and page_num which has a standard deviation of 40.62. On the other hand, the rest of the variables had much lower standard deviations. First, brand_fix was 2.73, pic_fix was 4.96, brand_size was 5.70, and lastly, page_pos was 0.49, L'Oreal was 0.17.