MKT845AE Advanced Digital Marketing Analytics Take-Home Exam 2

(Spring 2024)

INSTRUCTIONS: **ANSWERS ARE BELOW THE QUESTIONS**

1. This is an individual take-home exam. The exam includes 3 questions, which covers topics in customer profiling/cluster analysis, search analytics, and social media analytics. The total point of this exam is 50.
2. Please print the answers clearly. While answering the questions, please make sure that the flow of thinking is clearly articulated.
3. The datasets (waze, amc_movie_posts, and tripadvisor_simple) and the click-through rate calculation sheet used for this take-home exam can be found using: "Modules" -> "Quizzes &Exams" tab.
4. While submitting the work, please be sure to include: 1) the answer sheets, 2) important SAS output tables and all SAS original codes, and 3) the completed CTR excel spreadsheet. Regarding SAS tables, please note that only table supporting your justification/conclusion are needed. Answers, CTR spreadsheets, and SAS codes can be submitted as separate documents.
5. The submission deadline of the take-home exam is May 6th, by the end of the day.Late submission will *not* be accepted or graded.
6. Please be sure to upload the completed work via Canvas, using the submission portal. Email submission will *not* be accepted or graded.
7. If there is any question you'd prefer further clarification, please feel free to reach out.

I. Customer Profiling & Preferred Segment Identification (15 points)

***Scenario (dataset: waze)*:**Waze managers want to understand customers' diverse preference on some features included in the mobile GPS application (e.g., routes, map layout, etc.). They also want to know customer's perception towards a variety of subscription plans. To that end, managers reached out to 1,400 respondents and collected their feedback, preferences, as well as each respondent's demographic and behavioral information. The coding information is provided in the following table:

Route Alternatives (routes)	Tags	Map Layout (modes)	Nearby	Subscription Plan (price)
1-3 routes (coded as 1)	Road Construction Ahead (coded as 1)	2D (coded as 1)	Fastfood restaurant (coded as 1)	$0.99 per month (coded as 1)

3+ routes (coded as 2) 0	Police Reported Ahead (coded as 2)	3D (coded as 2) 0	Gas station (coded as 2)	$1.00-$1.99 per month (coded as 2)
	Car Accident (coded as 3)		ATM (coded as 3)	$2.00 - $2.99 per month (coded as 3) 0
	Vehicle Stop on Shoulder Ahead (4) (coded as 4)			

To get customer insights, managers decide to run a cluster analysis – therefore, they reach out to a group of data analysts for assistance. When running the cluster analysis, analysts suggested that variables that are behavioral related should be prioritized and considered. As a result, in one of the segmentation exercises, they used the following predictors and create ***3*** different segments.

- whether or not the respondent ever used Waze before: (variable name: "waze_before")
- number of years using mobile GPS (variable name: "mobile_gps")
- number of map applications ever used (variable name: "number_of_apps").

1) Please conduct the cluster analysis using "***proc fastclus***" , report the CCC, and provide some brief comments regarding the quality of the segmentation exercise.

Answer.
The quality segmentation is bad, because our CCC is negative. The negative CCC value is critical as it suggests that the clustering model may not be appropriate for the data. CCC = -6.76

Results Viewer - SAS Output

Statistics for Variables				
Variable	Total STD	Within STD	R-Square	RSQ/(1-RSQ)
waze_before	0.50017	0.50040	0.000457	0.000457
mobile_gps	1.40598	0.53990	0.852746	5.790965
number_of_apps	0.80560	0.75847	0.114813	0.129704
OVER-ALL	0.97911	0.61024	0.612085	1.577885

Pseudo F Statistic = 1133.71

Approximate Expected Over-All R-Squared = 0.65132

Cubic Clustering Criterion = -6.756

Editor - Untitled1 *

```
proc fastclus data = tmpl.waze out = tmpl.waze_segment
maxclusters = 3 maxiter = 100;
var waze_before mobile_gps number_of_apps;
run;
```

Results Viewer - SAS Output

Statistics for Variables				
Variable	Total STD	Within STD	R-Square	RSQ/(1-RSQ)
waze_before	0.50017	0.50040	0.000457	0.000457
mobile_gps	1.40598	0.53990	0.852746	5.790965
number_of_apps	0.80560	0.75847	0.114813	0.129704
OVER-ALL	0.97911	0.61024	0.612085	1.577885

Pseudo F Statistic = 1133.71

Approximate Expected Over-All R-Squared = 0.65132

Cubic Clustering Criterion = -6.756

Editor - Untitled1 *

```
proc fastclus data = tmpl.waze out = tmpl.waze_segment
maxclusters = 3 maxiter = 100;
var waze_before mobile_gps number_of_apps;
run;
```


2) Assuming that managers decide to use the empirical results generated from this segmentation, please report 1) size of each cluster (number of customers included in each cluster) and 2) behavioral characteristics for each cluster and fill up the information in the tables below.

(*Note: for *categorical* variable, please report the number in the percentage format. For *continuous* variable, please report the mean. For number reporting, please round each number to its nearest tenth – the 1st digit after the decimal point).

Cluster	Size	Experience with Waze	Years of using GPS on mobile platform	Number of Mobile GPS used
1	517	50.5%	1.4	2.2

2	467	48.2%	4.6	2.2
3	456	50.4%	3.0	1.6

```
proc means data = tmpl.waze_segment;
class cluster;
var mobile_gps number_of_apps;
run;

proc freq data = tmpl.waze_segment;
table cluster*waze_before;
run;
```

The MEANS Procedure

Cluster	N Obs	Variable	Label	N	Mean	Std Dev	Minimum	Maximum
1	517	mobile_gps	Years of using GPS on a mobile platform	517	1.4294004	0.4954700	1.0000000	2.0000000
		number_of_apps	Number of Map applications used	517	2.1605416	0.7492568	1.0000000	3.0000000
2	467	mobile_gps	Years of using GPS on a mobile platform	467	4.5717345	0.4953581	4.0000000	5.0000000
		number_of_apps	Number of Map applications used	467	2.2291221	0.7360781	1.0000000	3.0000000
3	456	mobile_gps	Years of using GPS on a mobile platform	456	3.0175439	0.6252202	2.0000000	4.0000000
		number_of_apps	Number of Map applications used	456	1.6096491	0.7907675	1.0000000	3.0000000

The FREQ Procedure

Frequency
Percent
Row Pct
Col Pct

Table of CLUSTER by waze_before			
	waze_before(Experience using waze before)		
CLUSTER(Cluster)	0	1	Total
1	256 17.78 49.52 35.36	261 18.13 50.48 36.45	517 35.90
2	242 16.81 51.82 33.43	225 15.63 48.18 31.42	467 32.43
3	226 15.69 49.56 31.22	230 15.97 50.44 32.12	456 31.67
Total	724 50.28	716 49.72	1440 100.00

In addition to the behavioral characteristics, managers also want to know if there exists any variation in terms of customer's product feature propensity and sensitivity towards subscription fee. In the following table, within each attribute, please specify the feature option(s) that is ***most*** preferred by customers within each cluster (e.g., the reporting format can be articulated as: 1-3 routes, 2D, $2.00~2.99/month).

Cluster	Route Alternatives	Map Layout	Subscription Fee
1	3+ routes	3D	$0.99/month
2	1-3 routes	3D	$0.99/month
3	1-3 routes	2D	$1.00-$1.99/month

```
proc sql;
create table tmpl.customer_segment1 as select *from tmpl.waze_segment
where cluster = 1;
run;
```

```
proc logistic data = tmpl.customer_segment1 descending;
class routes modes Price/param = ref;
model Purchase =  routes modes Price;
run;
```

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	1.8375	4	0.7656
Score	1.8360	4	0.7659
Wald	1.8328	4	0.7665

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
routes	1	1.6708	0.1962
modes	1	0.0043	0.9478
Price	2	0.2048	0.9027

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	0.1068	0.2071	0.2658	0.6061
routes	1	1	-0.2286	0.1769	1.6708	0.1962
modes	1	1	-0.0116	0.1765	0.0043	0.9478
Price	1	1	0.0680	0.2170	0.0983	0.7539
Price	2	1	-0.0252	0.2188	0.0132	0.9084

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
routes 1 vs 2	0.796	0.563	1.125
modes 1 vs 2	0.989	0.699	1.397

```
proc sql;
create table tmpl.customer_segment2 as select *from tmpl.waze_segment
where cluster = 2;
run;
```

```
proc logistic data = tmpl.customer_segment2 descending;
class routes modes Price/param = ref;
model Purchase =  routes modes Price;
run;
```

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
routes	1	0.0046	0.9462
modes	1	4.7290	0.0297
Price	2	0.5713	0.7515

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	0.2815	0.2165	1.6911	0.1935
routes	1	1	0.0126	0.1868	0.0046	0.9462
modes	1	1	-0.4077	0.1875	4.7290	0.0297
Price	1	1	0.1517	0.2294	0.4375	0.5083
Price	2	1	0.00608	0.2310	0.0007	0.9790

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
routes 1 vs 2	1.013	0.702	1.461
modes 1 vs 2	0.665	0.461	0.961
Price 1 vs 3	1.164	0.742	1.824
Price 2 vs 3	1.006	0.640	1.582

```
proc sql;
create table tmpl.customer_segment3 as select *from tmpl.waze_segment
where cluster = 3;
run;
```

```
proc logistic data = tmpl.customer_segment3 descending;
class routes modes Price/param = ref;
model Purchase =  routes modes Price;
run;
```

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
routes	1	0.0126	0.9106
modes	1	2.0390	0.1533
Price	2	2.9379	0.2302

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-0.1317	0.2068	0.4055	0.5243
routes	1	1	0.0212	0.1889	0.0126	0.9106
modes	1	1	0.2699	0.1890	2.0390	0.1533
Price	1	1	-0.3184	0.2267	1.9722	0.1602
Price	2	1	0.0426	0.2340	0.0331	0.8556

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
routes 1 vs 2	1.021	0.705	1.479
modes 1 vs 2	1.310	0.904	1.897
Price 1 vs 3	0.727	0.466	1.134
Price 2 vs 3	1.044	0.660	1.651

Answer.

Cluster Analysis:

Cluster 1: “Beginner” Customers The first cluster consisted of 517 customers:

- **50.5% of customers in this cluster have prior experience using Waze.**
- **The first cluster’s average number of years of using GPS on a mobile platform is 1.4 years, which suggests that this cluster is mainly composed of new users.**
- **The first cluster also uses an average of 2.2 different mobile GPS applications.**

Cluster 2: “Expert” Customers The second cluster included 467 customers:

- **48.2% of the customers have prior experience using Waze.**
- **The second cluster’s average number of years of using GPS on a mobile platform is significantly higher at 4.6 years, suggesting that this cluster is composed of expert users.**
- **The second cluster also uses an average of 2.2 different mobile GPS applications, suggesting they prefer experiencing additional platforms despite their extensive experience.**

Cluster 3: “Average” Customers The third cluster involved 456 customers:

- **50.4% of the customers have prior experience using Waze, similar to the first cluster.**
- **The third cluster’s average number of years of using GPS on a mobile platform is 3.0 years, positioning them between the expert and beginner users.**
- **The third cluster uses fewer mobile GPS applications, averaging 1.6 apps, suggesting they may be satisfied with their current app.**

The three clusters present different levels of familiarity with GPS usage on a mobile platform and a variety of GPS applications used. These differences indicate distinct behaviors that could lead to differing needs and preferences regarding GPS app functionality.

3) Combining the empirical results from the two tables, which cluster(s) could potentially be Waze's preferred customer segment(s)? Please simply justify the reason.

Answer.
As per my analysis Cluster 2 could be identified as a potentially preferred customer segment for Waze for several reasons:

1. **High GPS Experience: Their significant experience with GPS apps (4.6 years) implies they understand and value the functionality of such applications, which can make them more appreciative of advanced features that Waze might offer.**
2. **Multiple App Usage: Their usage of multiple GPS applications suggests they are not yet fully satisfied with any single app or are interested in multiple feature sets. This presents an opportunity for Waze to capture more of their usage by introducing differentiated features or more robust functionalities that address gaps left by other apps.**
3. **Marketing Target: This segment's familiarity with technology and navigation tools makes them more likely to be receptive to new updates and features, making them ideal targets for marketing new functionalities and premium subscriptions.**

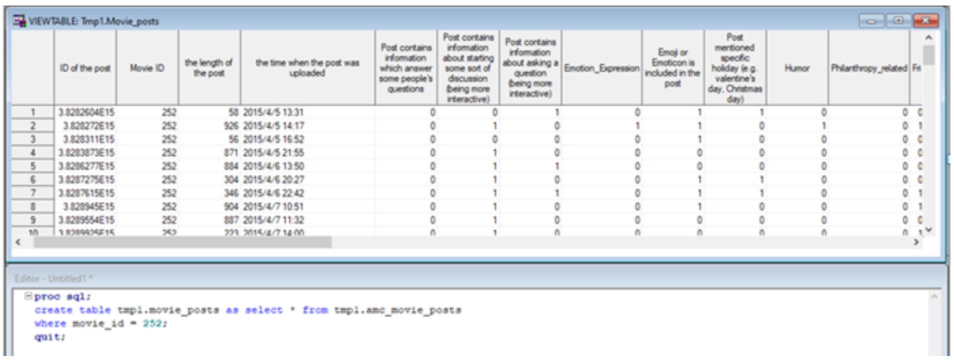
II. Social Media Analytics – Customer Digital Engagement (15 Points)

Scenario (dataset: amc_movie_posts):Movie producers of Fast & Furious wanted to understand how online posting (e.g., tone of voice) may affect customer's overall digital engagement (e.g., likes, comments, and shares). To this end, producers wanted to use one of their successful movies, which is "Fast & Furious 7" and its online posts promoted by the AMC movie theater via a social media platform, as an example to explore this relationship.

Although the movie was released many years ago (in late 2014), producers believed that the insights could still be valuable and can provide some important guidance regarding the use of the narratives, promoting shows in similar genres via movie exhibitors' social media accounts.

A team of data analysts were hired to conduct these analyses. The dataset, “amc_movie_posts”, is collected, with Fast & Furious 7’s online posts and consumer different level of engagement included (e.g., likes, comments, and shares). The identifier of this movie is 252 (movie_id = 252).

**Note/Hint:*to answer this question, we need to do some data preparation (e.g., generating a subset specifically showing online posts for the selected movie, using *proc sql*).



1) To get better understanding about this movie’s online posts, analysts conducted some descriptive analyses, looking at the number of posts, average performance of likes, comments, and shares.Please report the numbers in the table below.

**Note:* please feel free to choose any number reporting format when reporting the number in this question (e.g., integer, round the number to its nearest tenth or hundredth are all acceptable).

Movie Name	Number of Posts	Average Likes	Average Comments	Average Shares
Fast & Furious 7	80	71.89	195.78	185.59

The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
Number_of_Likes	80	71.8875000	168.0177804	1.0000000	1124.00
Number_of_Shares	80	185.5875000	796.3096003	0	6856.00
Number_of_Comments	80	195.7750000	1211.07	4.0000000	10821.00

Editor - Untitled1 *

```
proc sql;
create table tmpl.movie_posts as select * from tmpl.amc_movie_posts
where movie_id = 252;
quit;

proc means data = tmpl.movie_posts;
var number_of_likes number_of_shares number_of_comments;
run;
```

They also conducted two frequency analyses to check the number of posts whereby either pictures or videos were included, simply to make sure that pictures or videos embedded in the posts were not playing a dominant role (e.g., more than 50% of posts containing either videos or pictures), influencing consumer online engagement.

Movie Name	Number of Posts where *pictures* were included	Number of Posts where *videos* were included
Fast & Furious 7	17(21.25%)	1

```
proc freq data = tmpl.movie_posts;
table video_dummy;
run;

proc freq data = tmpl.movie_posts;
table Number_of_Pictures;
run;
```

The FREQ Procedure

whether or not the post includes a video				
video_dummy	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	79	98.75	79	98.75
1	1	1.25	80	100.00

The FREQ Procedure

Number_of_Pictures	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	63	78.75	63	78.75
1	9	11.25	72	90.00
2	1	1.25	73	91.25
3	1	1.25	74	92.50
6	2	2.50	76	95.00
9	4	5.00	80	100.00

2) Knowing that there are not too many pictures and videos included in the online posts, they now begin to analyze how informative and emotional appeals might affect consumer's three levels of engagement.Please conduct three regression models,specify the models, and simply describe the relationships between advertising appeals and online engagement.

Answer.

Linear Regression Model

Likes = 113.34 – 1.76 * total_informational – 12.37 * total_emotional

Comments = 94.38 -9.97 * total_informational + 37.98 * total_emotional

Shares = 147.05 +4.78 * total_informational + 9.79 * total_emotional

1. **Regression Model for Likes:**
   - **Model Equation: Likes = 113.34 – 1.76 * total_informational – 12.37 * total_emotional.**
   - **Interpretation: The intercept, 113.34, represents the expected number of likes when there is no informational or emotional content. Both informational and emotional contents have negative coefficients, suggesting that an increase in these contents tends to decrease the number of likes. Specifically, each unit increase in informational content reduces likes by about 1.76, and each unit increase in emotional content reduces likes by about 12.37.**
2. **Regression Model for Comments:**
   - **Model Equation: Comments = 94.38 - 9.97 * total_informational + 37.98 * total_emotional.**
   - **Interpretation: The intercept, 94.38, predicts the expected number of comments when both informational and emotional contents are absent. The informational content negatively influences comments, decreasing them by about 9.97 per unit increase. Conversely, emotional content positively**

impacts comments, increasing them by approximately 37.98 for each unit increase.

3. **Regression Model for Shares:**

    - **Model Equation: Shares = 147.05 + 4.78 * total_informational + 9.79 * total_emotional.**

    - **Interpretation: The intercept, 147.05, indicates the expected number of shares without informational or emotional content. Both coefficients are positive, showing that increases in informational and emotional contents tend to increase the number of shares. Specifically, each unit increase in informational content leads to about 4.78 more shares, and each unit increase in emotional content contributes to about 9.79 more shares.**

```
proc reg data = tmpl.movie_posts;
model number_of_likes = total_informational total_emotional;
run;
proc reg data = tmpl.movie_posts;
model number_of_comments = total_informational total_emotional;
run;
proc reg data = tmpl.movie_posts;
model number_of_shares = total_informational total_emotional;
run;
```

The REG Procedure
Model: MODEL1
Dependent Variable: Number_of_Likes

Number of Observations Read	80
Number of Observations Used	80

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	42697	21349	0.75	0.4751
Error	77	2187471	28409		
Corrected Total	79	2230168			

Root MSE	168.54883	R-Square	0.0191
Dependent Mean	71.88750	Adj R-Sq	-0.0063
Coeff Var	234.46194		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	113.34276	38.71306	2.93	0.0045
Total_Informational	1	-1.75935	5.30050	-0.33	0.7408
Total_Emotional	1	-12.36509	10.49381	-1.18	0.2423

The REG Procedure
Model: MODEL1
Dependent Variable: Number_of_Comments

Number of Observations Read	80
Number of Observations Used	80

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	470656	235328	0.16	0.8550
Error	77	115397200	1498665		
Corrected Total	79	115867856			

Root MSE	1224.19971	R-Square	0.0041
Dependent Mean	195.77500	Adj R-Sq	-0.0218
Coeff Var	625.30952		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	94.38457	281.17976	0.34	0.7380
Total_Informational	1	-9.97387	38.49843	-0.26	0.7963
Total_Emotional	1	37.98276	76.21835	0.50	0.6197

The REG Procedure
Model: MODEL1
Dependent Variable: Number_of_Shares

Number of Observations Read	80
Number of Observations Used	80

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	48101	24051	0.04	0.9637
Error	77	50046508	649955		
Corrected Total	79	50094609			

Root MSE	806.19765	R-Square	0.0010
Dependent Mean	185.58750	Adj R-Sq	-0.0250
Coeff Var	434.40299		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	147.05230	185.17115	0.79	0.4296
Total_Informational	1	4.78166	25.35317	0.19	0.8509
Total_Emotional	1	9.78832	50.19365	0.20	0.8459

3) Using the knowledge from the dataset and the empirical results derived from the regression model analyses, please answer the following questions:

- if movie producers were interested in seeing more conversation/discussion under the posts (elevating the comments), what type of appeals would you recommend them promote (informational or emotional)?

- If they were interested in making a post go "viral" (increasing number of shares), between informative and emotional appeal, which type(s) would you recommend?

Answer. Emotional Content: The regression model for comments shows a significant positive relationship between emotional content and the number of comments. Specifically, each unit increase in emotional content is associated with an increase of about 37.98 comments. This indicates that emotional appeals strongly resonate with the audience, prompting them to engage more through comments.

Informational Content: In contrast, informational content was found to have a negative effect on the number of comments. Each unit increase in informational content leads to a decrease in comments by approximately 9.97.

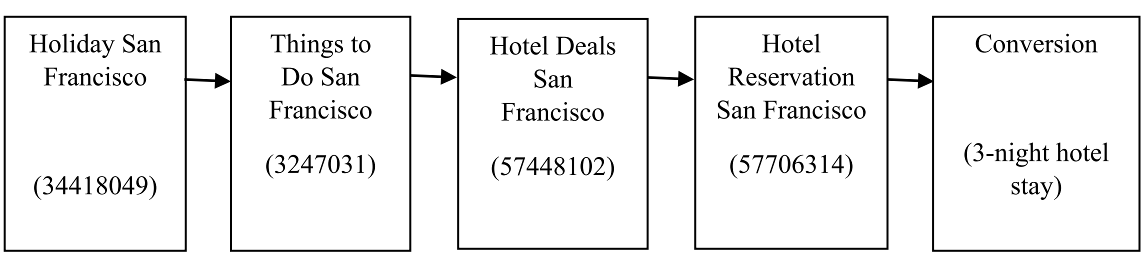
Therefore, if movie producers are interested in seeing more comments under the posts, they are supposed to increase emotional appeal. It has a greater effect on the number of comments than informational appeal.

4) Finally, based upon the empirical results from the regression models, can analysts conclude that this pattern applies the findings to all other similar movies? Please simply explain the reason.

Answer. An analyst can't conclude that the pattern observed in the regression models applies to all other similar movies. The lack of statistically significant relationships advertising appeal and online engagement shows the model poorly captures patterns in movies. The regression was also trained and built using specific Fast and furious 7 data based solely on the empirical results from these models.

III. Search Analytics – Attribution Modeling (20 Points)

Scenario (dataset: tripadvisor_simple): Managers at Tripadvisor.com want to promote its hotel chains in San Francisco California. To ensure certain level of visibility and click-through, they decide to use sponsored ads via a search engine platform. They know that customer's search paths on search engine often involve four critical touch points/ keywords (see details below). They want to know if an appropriate attribution model can be developed to help them better invest/assign credits to each touch point.



To that end, managers reached out to a group of data analysts, seeking insights and solutions. A dataset, called "tripadvisor" is collected from the search engine platform. Also, managers know that, on average, each hotel reservation during the holiday season includes 3 nights, the price for each night hotel stay, in a decent place, ranges from $80 to $220+. Analysts suggested that a median price can be considered, which is $150/night. Please follow the attribution modeling analytical steps, conduct the analyses, and answer the following questions.

1) Please conduct the binary logistic model (*proc logistic*) &specify the utility function.

Answer. U = -2.34 – 0.26 * rank + 1.04 *w_accurate + 0.95*w_phrase

Results Viewer - sashtml

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	2397.0905	3	<.0001
Score	2225.4906	3	<.0001
Wald	2186.1008	3	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-2.3388	0.0541	1868.7829	<.0001
rank	1	-0.2589	0.00566	2094.1697	<.0001

Editor - Untitled1 *

```
proc sql;
 alter table tmpl.tripadvisor_simple add w_accurate num;
 update tmpl.tripadvisor_simple set w_accurate = 0;
 update tmpl.tripadvisor_simple set w_accurate = 1 where wmatch = 63;
 quit;

proc sql;
 alter table tmpl.tripadvisor_simple add w_phrase num;
 update tmpl.tripadvisor_simple set w_phrase = 0;
 update tmpl.tripadvisor_simple set w_phrase = 1 where wmatch = 31;
 quit;

proc logistic data = tmpl.tripadvisor_simple descending;
 model click = rank w_accurate w_phrase;
 run;
```

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	2397.0905	3	<.0001
Score	2225.4906	3	<.0001
Wald	2186.1008	3	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-2.3388	0.0541	1868.7829	<.0001
rank	1	-0.2589	0.00566	2094.1697	<.0001
w_accurate	1	1.0369	0.0559	344.3970	<.0001
w_phrase	1	0.9516	0.0563	285.6893	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
rank	0.772	0.763	0.781
w_accurate	2.820	2.528	3.147
w_phrase	2.590	2.319	2.892

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	57.3	Somers' D	0.234
Percent Discordant	33.9	Gamma	0.256
Percent Tied	8.9	Tau-a	0.039
Pairs	2343983274	c	0.617

2) Given the empirical results from the utility function, with the use of the click-through rate estimation excel spreadsheet, what is the click-through rate for each keyword?

**Notes:*

a. Since some keywords' the ranking performanceis currently unknown, to answer this question, we need to conduct some analyses to get the average ranking performance for those keywords.

```
proc means data = tmpl.tripadvisor_simple;
 var rank w_accurate;
 run;
```

b. In the CTR spreadsheet, we need to choose the correct table to generate the correct CTR.

*c.***For parameter estimates reporting**, please round each number to its nearest hundredth (2nd digit after the decimal point) and have the information added to the CTR sheet. **For ranking performance reporting**, please round each number to the integer format. **For CTR reporting**, please round each percentage to its nearest tenth – the 1st digit after the decimal point (e.g., if you get a CTR = 15.13%, round the number to 15.1%).

wordid	Keyword	Match Type	Average Rank	CTR (%)
34418049	Holiday San Francisco	Broad	1	6.9%
3247031	Things to Do San Francisco	Phrase	4	8.1%
57448102	Hotel Deals San Francisco	Accurate	1	17.4%
57706314	Hotel Reservation San Francisco	Accurate	1	17.4%

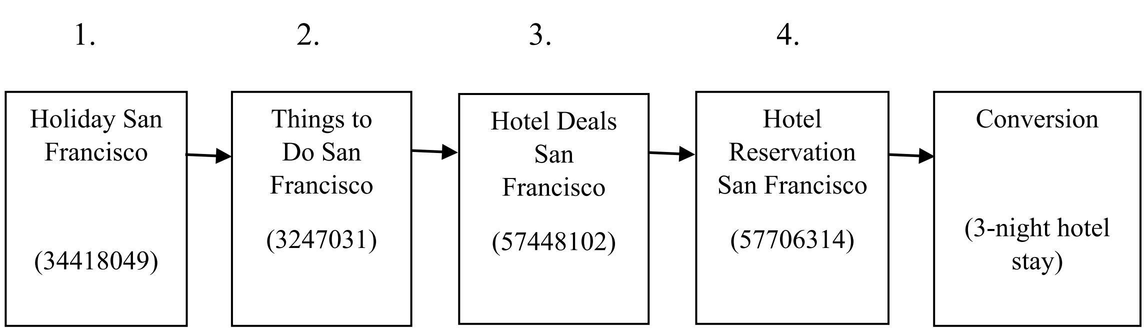
3) Please convert the CTRs into the proper percentage format:

Answer. Weight(%)=(CTR of Keyword/ Total CTR)×100

wordid	Keyword	Weight (%)
34418049	Holiday San Francisco	13.9%
3247031	Things to Do San Francisco	16.3%
57448102	Hotel Deals San Francisco	34.9%
57706314	Hotel Reservation San Francisco	34.9%

4) Using "click-through rate based" attribution model, how much money should managers assigned to each keyword? Please specify below.

*Note: please feel free to choose any number reporting format when reporting the credits to each keyword (e.g., integer, round the number to its nearest tenth or hundredth are all acceptable).



1. **$ 450 * 13.9% = $ 63**
2. **$ 450 * 16.3% = $ 73**
3. **$ 450 * 34.9% = $ 157**
4. **$ 450 * 34.9% = $ 157**

5) Finally, compared with the way how credits are assigned in this scenario, which of the following attribution approach shows similar credit allocation pattern? Please simply justify your reason.

A. First Click Attribution B. Last Click Attribution

C. Linear Attribution D. Time Decay Attribution

Answer. The option that comes closest among the ones provided is C. As opposed to the equal distribution, the example offers a proportional one – all in all, depending on CTR. However, it still seems closest to Linear, as it takes into account each touchpoint, and CTR is used to assess its direct impact, as opposed to an equal or time-decaying one. Thus, the correct option is C. Linear Attribution, an option with the spirit closest to the proportions of contributions to the overall outcome of each of the multiple interactions.