Final Project Report

Author: Saurabh Gupta

Data Introduction

In this collaborative statistical analysis project, the focus is on predicting customer acceptance of coupons distributed through mobile recommendation devices within the marketing firm's context. As data analysts, the primary objective is to employ robust statistical analysis to discern patterns within the available dataset.

The dataset, gathered from over 10,000 respondents via Amazon Mechanical Turk, comprises demographic information, preferences, and likelihood of using coupons across various establishment types, such as bars, takeaway food venues, coffee shops, and restaurants categorized by price range. This dataset spans more than 20 diverse scenarios, encompassing different driving contexts, situational dynamics, and coupon attributes.

The demographic information collected from the sample includes gender, age, marital status, whether they have children, level of education, occupation, and income. From this information, the demographic profile of the sample was determined:

- Gender distribution is equal between male and female.
- Over 60% of the sample is between 21 and 40 years old.
- While 60% of the sample has a partner, 70% of them are married.
- More than 60% of the sample do not have children.
- Approximately 32% of the sample do not hold a degree, while around 60% hold a bachelor's degree or higher.
- 16% of the sample is unemployed. The three major occupations are computer & engineering, student, and sales & retail.
- Over half of the sample earns between USD 12,500 and 62,500 annually.

This demographic profile provides valuable insights into the sample population along with other important aspects about the dataset included below:

Driving Contexts:

- 44% of the sample need to drive less than 15 minutes to reach the establishment where the coupon is offered, while only 12% need to drive more than 25 minutes.

- 79% of the sample found that the establishment where the coupon is offered is in the opposite direction compared to their current destination.

Situational Dynamics:

- Half of the sample go to a non-urgent place when receiving the coupon offer.
- Around 60% of the sample were without a passenger when receiving the coupon offer.
- 43% of the sample go to a restaurant between 1-3 times per month with an average expense per person of less than $20, while 27% of the sample go to a restaurant with the same frequency but with an average expense between $20 and $50.
- 80% of the sample received the coupon offer during sunny weather with temperatures between 55 to 80°F.
- Approximately 40% of the sample receive the coupon offer in the morning, while another 40% receive it at night.

Coupon Attributes:

- The type of establishment with the highest share of coupons was the Coffee House with 32%, followed by restaurants with an expense of less than $20 per person at 22%, and Carryout & Takeaway with 19%.
- 56% of the sample received coupons with an expiration date of 1 day, compared to 44% with a 2-hour expiration date.

These insights will serve as a foundation for targeted goals, analysis, and interpretations, guiding informed decision-making and strategies in the context of marketing analytics.

Goals

This analysis had two distinct goals to utilize consumer characteristics and behaviors from a survey.

The first goal was to identify factors that would help the client target consumers with higher incomes to market their products to. In order to do this, it was necessary to explore the relationship between income and the variables collected in the survey and see if there are any significant relationships to explain income. This data was then used to create a prediction to see the effects of these different variables on income.
The identified characteristics that were deemed more likely to affect income included gender, level of education, and occupation. This was assessed based on the following assumptions: (1) there exists a difference in income between males and females, (2)

individuals who are employed with higher levels of education would likely have a higher income compared to those who did not graduate college, and (3) individuals employed in fields such as medicine or business would likely have a higher income compared to students or unemployed individuals.

The second goal was to determine what consumer variables have an impact on whether a consumer would accept a coupon. In this case, additional factors needed to be considered to determine which ones had an effect on the acceptance of a coupon. From the data, most coupons would expire either in 2 hours or 1 day. Additionally, the coupon may be given for restaurants, a coffee house, carry out restaurant, or a bar. The coupons would also be given out at times throughout the day from 7 am up to 10 pm. Based on these factors, different characteristics from the survey, including who was with the person (adult or child) being offered the coupon, were taken into account when performing the analysis. Additionally, the weather, temperature, coupon location, drive direction (opposite vs same), driving distance were considered as factors that could impact acceptance. These variables were then used to create a prediction to help identify the variables that would increase coupon acceptance.

Analysis

Manipulation of Relevant Demographic Variables

Prior to performing the analysis, the data was condensed by reducing the levels within each demographic variable. The process for this step is as described below.

Income: Based on the survey design and the notice privacy, income was collected as a range. To simplify the analysis, a midpoint was created for each income range, and labeled “incomemid”. For the low and high end incomes, the extreme values were chosen for analysis.

Income Range	Incomemid
<$12,500	$12,500
$12,500 to $24,999	$18,749.50
$25,000 to $37,499	$31,249.50
$37,500 to $49,999	$43,749.50
$50,000 to $62,499	$56,249.50

$62,500 to $74,999	$68,759.50
$75,000 to $87,499	$81,249.50
$87,500 to $99,999	$93,749.50
>$100,000	$100,000

Education: This variable was not manipulated further. There were 6 different levels of degrees from some high school all the way up to a Graduate Degree (Masters or Doctorate).

Gender: This variable was not manipulated further, and kept as male or female.

Occupation: The initial survey had 25 different responses and would have been difficult to analyze without regrouping. Following a recoding exercise, occupation was converted to “Employment” variable with only 4 levels: Student, Unemployed, Retired, and Employed (this combined all responses that were not the initial 3).

Age: Several ages existed for this variable, which were recoded to generate specific age ranges.

Lives Alone: The marital status (levels: Widowed, Unmarried partner, Single, Married Partner, or Divorced) and has children (levels: yes 1, no 0) variables were combined to yield responses on whether they lived alone or not.

Consumer Income Analysis

The data were initially explored via contingency tables to understand the distribution of survey results for selected variables and their performance in relation to the income of the target consumers. Multiple combinations of interactions were performed, some of which generated the data shown in Figure 1.

The FREQ Procedure

Frequency
Percent
Row Pct
Col Pct

Table 1 of income by agegroup
Controlling for gender=Female

income	agegroup					
	21 - 30	31 - 40	41 - 50	50plus	below21	Total
$100000 or More	131 2.01 17.31 5.39	330 5.07 43.59 18.50	165 2.53 21.80 17.10	66 1.01 8.72 5.92	65 1.00 8.59 29.68	757 11.63
$12500 - $24999	352 5.41 35.45 14.49	336 5.16 33.84 18.83	86 1.32 8.66 8.91	197 3.03 19.84 17.68	22 0.34 2.22 10.05	993 15.25
$25000 - $37499	541 8.31 42.20 22.27	302 4.64 23.56 16.93	241 3.70 18.80 24.97	176 2.70 13.73 15.80	22 0.34 1.72 10.05	1282 19.69
$37500 - $49999	394 6.05 38.82 16.22	201 3.09 19.80 11.27	111 1.70 10.94 11.50	287 4.41 28.28 25.76	22 0.34 2.17 10.05	1015 15.59
$50000 - $62499	370 5.68 51.25 15.23	220 3.38 30.47 12.33	0 0.00 0.00 0.00	132 2.03 18.28 11.85	0 0.00 0.00 0.00	722 11.09
$62500 - $74999	127 1.95 36.60 5.23	110 1.69 31.70 6.17	110 1.69 31.70 11.40	0 0.00 0.00 0.00	0 0.00 0.00 0.00	347 5.33
$75000 - $87499	84 1.29 18.18 3.46	131 2.01 28.35 7.34	186 2.86 40.26 19.27	61 0.94 13.20 5.48	0 0.00 0.00 0.00	462 7.10
$87500 - $99999	138 2.12 36.41 5.68	132 2.03 34.83 7.40	44 0.68 11.61 4.56	65 1.00 17.15 5.83	0 0.00 0.00 0.00	379 5.82
Less than $12500	292 4.48 52.71 12.02	22 0.34 3.97 1.23	22 0.34 3.97 2.28	130 2.00 23.47 11.67	88 1.35 15.88 40.18	554 8.51
Total	2429 37.31	1784 27.40	965 14.82	1114 17.11	219 3.36	6511 100.00

Frequency
Percent
Row Pct
Col Pct

Table 2 of income by agegroup
Controlling for gender=Male

income	agegroup					
	21 - 30	31 - 40	41 - 50	50plus	below21	Total
$100000 or More	393 6.37 40.14 14.12	322 5.22 32.89 20.46	132 2.14 13.48 16.22	132 2.14 13.48 19.58	0 0.00 0.00 0.00	979 15.86
$12500 - $24999	490 7.94 58.47 17.61	110 1.78 13.13 6.99	149 2.41 17.78 18.30	67 1.09 8.00 9.94	22 0.36 2.63 6.71	838 13.58
$25000 - $37499	374 6.06 51.16 13.44	220 3.56 30.10 13.98	65 1.05 8.89 7.99	28 0.45 3.83 4.15	44 0.71 6.02 13.41	731 11.84
$37500 - $49999	351 5.69 44.43 12.61	219 3.55 27.72 13.91	88 1.43 11.14 10.81	110 1.78 13.92 16.32	22 0.36 2.78 6.71	790 12.80
$50000 - $62499	419 6.79 44.72 15.06	219 3.55 23.37 13.91	122 1.98 13.02 14.99	155 2.51 16.54 23.00	22 0.36 2.35 6.71	937 15.18
$62500 - $74999	175 2.83 35.07 6.29	110 1.78 22.04 6.99	120 1.94 24.05 14.74	50 0.81 10.02 7.42	44 0.71 8.82 13.41	499 8.08
$75000 - $87499	197 3.19 49.87 7.08	154 2.49 38.99 9.78	22 0.36 5.57 2.70	22 0.36 5.57 3.26	0 0.00 0.00 0.00	395 6.40
$87500 - $99999	137 2.22 26.55 4.92	154 2.49 29.84 9.78	72 1.17 13.95 8.85	66 1.07 12.79 9.79	87 1.41 16.86 26.52	516 8.36
Less than $12500	247 4.00 50.61 8.88	66 1.07 13.52 4.19	44 0.71 9.02 5.41	44 0.71 9.02 6.53	87 1.41 17.83 26.52	488 7.91
Total	2783 45.08	1574 25.50	814 13.19	674 10.92	328 5.31	6173 100.00

Figure 1. Contingency Tables for Income vs Age group as defined by Gender

When evaluating the income distribution across various age groups for either gender, it was found that the majority of the survey responses came from ages 21 - 50 while group representation for 50plus and below21 was lower amongst both genders. For the female below21 age group, no responses were received for the $50,000 - $99,999 income ranges, while the male below21 age group was missing responses for only $75,000 - $87,499 and above $100,000 income ranges. For both genders, 21-30 age range comprised a higher proportion of the survey within the respective groups. From the survey of the female group, the income range $25,000 - $37,499 ranked highest at 19.69%, while >$100,000 income ranked highest for the male group at 15.86%.

When examining income vs employment status as depicted in Figure 2, the female group had 67.33% employed individuals while the male group had 70.48% employed individuals. The second highest category for each group was unemployed for female and student for male candidates. Of the individuals surveyed that were employed, the highest income range for the female group was $25,000 - $37,499 at 21.42%, while the highest income range for the male group was >$100,000 at 17.95%.

Frequency
Percent
Row Pct
Col Pct

Table 1 of income by employment
Controlling for gender=Female

income	employment				
	Employe	Retired	Student	Unemplo	Total
$100000 or More	561 8.62 74.11 12.80	0 0.00 0.00 0.00	87 1.34 11.49 14.03	109 1.67 14.40 8.58	757 11.63
$12500 - $24999	774 11.89 77.95 17.66	66 1.01 6.65 27.85	88 1.35 8.86 14.19	65 1.00 6.55 5.12	993 15.25
$25000 - $37499	939 14.42 73.24 21.42	0 0.00 0.00 0.00	149 2.29 11.62 24.03	194 2.98 15.13 15.28	1282 19.69
$37500 - $49999	685 10.52 67.49 15.63	66 1.01 6.50 27.85	44 0.68 4.33 7.10	220 3.38 21.67 17.32	1015 15.59
$50000 - $62499	502 7.71 69.53 11.45	22 0.34 3.05 9.28	44 0.68 6.09 7.10	154 2.37 21.33 12.13	722 11.09
$62500 - $74999	237 3.64 68.30 5.41	0 0.00 0.00 0.00	22 0.34 6.34 3.55	88 1.35 25.36 6.93	347 5.33
$75000 - $87499	381 5.85 82.47 8.69	17 0.26 3.68 7.17	20 0.31 4.33 3.23	44 0.68 9.52 3.46	462 7.10
$87500 - $99999	197 3.03 51.98 4.49	22 0.34 5.80 9.28	28 0.43 7.39 4.52	132 2.03 34.83 10.39	379 5.82
Less than $12500	108 1.66 19.49 2.46	44 0.68 7.94 18.57	138 2.12 24.91 22.26	264 4.05 47.65 20.79	554 8.51
Total	4384 67.33	237 3.64	620 9.52	1270 19.51	6511 100.00

Frequency
Percent
Row Pct
Col Pct

Table 2 of income by employment
Controlling for gender=Male

income	employment				
	Employe	Retired	Student	Unemplo	Total
$100000 or More	781 12.65 79.78 17.95	22 0.36 2.25 8.53	110 1.78 11.24 11.41	66 1.07 6.74 11.00	979 15.86
$12500 - $24999	513 8.31 61.22 11.79	0 0.00 0.00 0.00	220 3.56 26.25 22.82	105 1.70 12.53 17.50	838 13.58
$25000 - $37499	577 9.35 78.93 13.26	0 0.00 0.00 0.00	110 1.78 15.05 11.41	44 0.71 6.02 7.33	731 11.84
$37500 - $49999	592 9.59 74.94 13.61	44 0.71 5.57 17.05	66 1.07 8.35 6.85	88 1.43 11.14 14.67	790 12.80
$50000 - $62499	679 11.00 72.47 15.61	126 2.04 13.45 48.84	110 1.78 11.74 11.41	22 0.36 2.35 3.67	937 15.18
$62500 - $74999	368 5.96 73.75 8.46	22 0.36 4.41 8.53	87 1.41 17.43 9.02	22 0.36 4.41 3.67	499 8.08
$75000 - $87499	373 6.04 94.43 8.57	0 0.00 0.00 0.00	0 0.00 0.00 0.00	22 0.36 5.57 3.67	395 6.40
$87500 - $99999	336 5.44 65.12 7.72	44 0.71 8.53 17.05	65 1.05 12.60 6.74	71 1.15 13.76 11.83	516 8.36
Less than $12500	132 2.14 27.05 3.03	0 0.00 0.00 0.00	196 3.18 40.16 20.33	160 2.59 32.79 26.67	488 7.91
Total	4351 70.48	258 4.18	964 15.62	600 9.72	6173 100.00

Figure 2. Contingency Tables for Income vs Employment as defined by Gender

Evaluating the comparison between income vs employment or age by gender, it was found that age is a much more complicated variable to evaluate the data by. This is primarily due to the fact that direct comparison between the male and female groups is complicated, and there could be several external factors such as education or home life that could be affecting why certain age groups are earning lower/higher. There is also missing survey data for an age group for both genders, which disproportionately affected the female group. While the same could be said for the retired groups for the employment evaluation, it equally affects both gender groups and therefore, can be ignored in favor of further analyzing employed, student, or unemployed groups.

There was additional interest in exploring whether the individual's living situation (alone or w/ partner/children) was a motivating factor in pursuing higher income opportunities. Based on data shown in Figure 3, $12,500 - $62,499 and >$100,000 income ranges constituted a high proportion of the survey with similar percentages, while $62,500 - $99,999 and <$12,500 income ranges fared much lower in comparison. All income groups except <$12,500 indicated a higher proportion of responses for living alone vs living with a partner/children. Based on the data, there could be a potential for individuals wanting to pursue higher incomes when living alone than when with a partner or children.

Frequency
Percent
Row Pct
Col Pct

Table of livesalone by income

livesalone	income									
	$100000 or More	$12500 - $24999	$25000 - $37499	$37500 - $49999	$50000 - $62499	$62500 - $74999	$75000 - $87499	$87500 - $99999	Less than $12500	Total
0	1256 9.90 14.89 72.35	1138 8.97 13.49 62.15	1331 10.49 15.78 66.12	1167 9.20 13.83 64.65	1219 9.61 14.45 73.48	636 5.01 7.54 75.18	639 5.04 7.57 74.56	650 5.12 7.71 72.63	400 3.15 4.74 38.39	8436 66.51
1	480 3.78 11.30 27.65	693 5.46 16.31 37.85	682 5.38 16.05 33.88	638 5.03 15.02 35.35	440 3.47 10.36 26.52	210 1.66 4.94 24.82	218 1.72 5.13 25.44	245 1.93 5.77 27.37	642 5.06 15.11 61.61	4248 33.49
Total	1736 13.69	1831 14.44	2013 15.87	1805 14.23	1659 13.08	846 6.67	857 6.76	895 7.06	1042 8.22	12684 100.00

Figure 3. Contingency Tables for Family Situation vs Income

To further analyze the data, a quantitative prediction model was developed for income as a target variable and the following variables as predictors: gender, education,and employment status. The reference point for the model was an unemployed female with some college education with no degree. The multiple linear regression equation was selected using a stepwise procedure with a significance level of 0.05.

Null hypothesis: there is no significant relationship between income level and selected variables (education level, employment, gender)

Alternative hypothesis: There is a significant relationship between income level and at least one of the predictor variables.

The REG Procedure
Model: MODEL1
Dependent Variable: incomemid

Number of Observations Read	12684
Number of Observations Used	12684

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	1.223786E12	1.359762E11	172.84	<.0001
Error	12674	9.970811E12	786713814		
Corrected Total	12683	1.11946E13			

Root MSE	28048	R-Square	0.1093
Dependent Mean	52653	Adj R-Sq	0.1087
Coeff Var	53.27076		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	41765	772.70274	54.05	<.0001
education Associates degree	education Associates degree	1	3978.11817	943.05133	4.22	<.0001
education Bachelors degree	education Bachelors degree	1	12927	610.06360	21.19	<.0001
education Graduate degree (Maste	education Graduate degree (Masters or Doctorate)	1	16513	792.75874	20.83	<.0001
education High School Graduate	education High School Graduate	1	-9169.26988	1037.04427	-8.84	<.0001
education Some High School	education Some High School	1	-29336	3034.09932	-9.67	<.0001
gender Male	gender Male	1	6406.37349	504.71952	12.69	<.0001
employment Employe	employment Employe	1	3109.50166	735.38407	4.23	<.0001
employment Retired	employment Retired	1	92.66092	1435.25936	0.06	0.9485
employment Student	employment Student	1	-5674.34640	985.97674	-5.76	<.0001

Figure 4. Multiple Linear Regression (MLR) Analysis for Incomemid target variable

The stepwise procedure kept all of the variables that were inputted into the model. The resulting equation is:

$$\text{Income} = 41765 + 3978.12 * \text{Associates degree} + 12927 * \text{Bachelors degree} + 16513 * \text{Graduate degree} - 9169.27 * \text{High School Graduate} - 29336 * \text{Some High school} + 6406.37 * \text{Male} + 3109.50 * \text{Employed} + 92.66 * \text{Retired} - 5674.35 * \text{Student}$$

From the model that was constructed, individuals with any sort of degree make more money than those who do not. As expected, people with graduate degrees make the most amount of money, followed by those with bachelor's degrees and then associates degrees. Those with only some high school education make much less money than others, though this could also be due students being included in this survey. People who are employed make much more money than those who are not, which is to be

expected. Men in this model make 6406.37 more dollars than women who are of otherwise similar demographics. Analysis of the MLR shows that all tested variables had p-values below the significance level 0.05, except for “employment retired”. This leads us to rejecting the null hypothesis as there is significant association between the predictor variables and target variables. The only predictor variable that does not have a significant impact on is “employment retired”.

Coupon Acceptance Analysis

Aiming to determine what consumer variables have an impact on whether a consumer would accept a coupon, the first step is to define the variables:

- Response variable: Created a new variable Coupon_accepted. Convert Y (0 or 1) to Coupon_accepted binary outcome.

- Predictor variables used in this study:

Coupons:

For this variable, the survey indicated the type of coupon that could be offered to a consumer. The options were a Bar, Carry out & Take Away, Coffee House, and restaurant. The restaurant option was further separated into the cost of the meal, either < $20 or $20 – 50 minutes.

Expiration:

This variable indicated the time to coupon expiration. The options were either 2 hours or 1 day. It was hypothesized that this potentially could have a significant impact on whether the coupon was accepted.

Time:

The time of day the coupon was offered was also included in the data. Here, the options were throughout the day including 7AM, 10AM, 2PM, 6PM, 10PM. Along with expiration, it was hypothesized that when the coupon was offered to a consumer may have a significant effect on whether or not it was accepted.

Destination:

This variable described where the consumer was headed after they completed the survey/offered the coupon. The consumer may be traveling home, to work, or they had no urgent place to get to.

Direction:
direction_same / direction_opp - There are 2 separate direction variables that will be

described together. The consumer would potentially be offered a coupon to a location that could be in the same or opposite direction in which they were heading. These likely offer the same information, just in opposite ways where “0” = No and “1” = Yes.

Weather:

The weather was described at the time of coupon offering. This potentially could have an impact as well, as the person may or may not want to go somewhere based on the current weather. The options here were rainy, snowy, or sunny.

Temperature:
In addition to weather, the temperature at the time of offering was also noted. There were limited options here, with only an option of 30, 55, or 80.

toCoupon_GEQ15min:
Driving distance to the restaurant/bar for using the coupon is greater than 15 minutes.

Model 1

- Target variable: Coupon_accepted.
- Predictor variables: expiration direction_same, toCoupon_GEQ15min.
- NULL Hypothesis: coupon acceptance with expiration time and driving distance in same direction events has some relationship.
- Alternate Hypothesis: coupon acceptance with expiration time and driving distance in the same direction has some association to each other.
- Odds Ratio: The odds ratio quantifies the strength and direction of the association between the variables. An odds ratio less than 1 indicates a negative association, meaning that the odds of one outcome decrease as the other increases.

The following 2X2 matrix shows the statistical data from the contingency tables:

- Same direction and coupon acceptance as 0 for driving distance more than 15 minutes.
    - In this case (upper left), the odds ratio is 0.1994, indicating that the odds of taking more than 15 minutes to get to the coupon destination are approximately 0.1994 times lower for individuals whose direction is the same as expected compared to those whose direction is not the same as expected.

- Same direction and coupon acceptance as 1 for driving distance more than 15 minutes.
    - In this case (upper right), the odds ratio is 0.2198, indicating that the odds of taking more than 15 minutes to get to the coupon destination are approximately 0.2198 times lower for individuals whose direction is the same as expected compared to those whose direction is not the same as expected.

- Opposite direction and coupon acceptance as 0 for driving distance more than 15 minutes.
    - In this case (lower left), the odds ratio is 5.0150, indicating that the odds of taking more than 15 minutes to get to the coupon destination are approximately 5.0150 times higher for individuals whose direction is opposite to the expected direction compared to those whose direction is not opposite to the expected direction.

- Opposite direction and coupon acceptance as 1 for driving distance more than 15 minutes.
    - In this case (lower right), the odds ratio is 4.5497, indicating that the odds of taking more than 15 minutes to get to the coupon destination are approximately 4.5497 times higher for individuals whose direction is opposite to the expected direction compared to those whose direction is not opposite to the expected direction.

Frequency
Expected

Table 1 of toCoupon_GEQ15min by direction_same Controlling for Coupon_accepted=0			
toCoupon_GEQ15min	direction_same		
	0	1	Total
0	1355 1699.9	791 446.14	2146
1	2981 2636.1	347 691.86	3328
Total	4336	1138	5474

Statistics for Table 1 of toCoupon_GEQ15min by direction_same
Controlling for Coupon_accepted=0

Statistic	DF	Value	Prob
Chi-Square	1	553.5613	<.0001
Likelihood Ratio Chi-Square	1	545.5778	<.0001
Continuity Adj. Chi-Square	1	551.9573	<.0001
Mantel-Haenszel Chi-Square	1	553.4602	<.0001
Phi Coefficient		-0.3180	
Contingency Coefficient		0.3030	
Cramer's V		-0.3180	

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	0.1994	0.1731	0.2297
Relative Risk (Column 1)	0.7049	0.6811	0.7295
Relative Risk (Column 2)	3.5351	3.1544	3.9617

Frequency
Expected

Table 2 of toCoupon_GEQ15min by direction_same Controlling for Coupon_accepted=1			
toCoupon_GEQ15min	direction_same		
	0	1	Total
0	2228 2664.6	1188 751.43	3416
1	3396 2959.4	398 834.57	3794
Total	5624	1586	7210

Statistics for Table 2 of toCoupon_GEQ15min by direction_same
Controlling for Coupon_accepted=1

Statistic	DF	Value	Prob
Chi-Square	1	617.9586	<.0001
Likelihood Ratio Chi-Square	1	636.1673	<.0001
Continuity Adj. Chi-Square	1	616.5439	<.0001
Mantel-Haenszel Chi-Square	1	617.8729	<.0001
Phi Coefficient		-0.2928	
Contingency Coefficient		0.2810	
Cramer's V		-0.2928	

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	0.2198	0.1939	0.2492
Relative Risk (Column 1)	0.7287	0.7094	0.7485
Relative Risk (Column 2)	3.3152	2.9887	3.6774

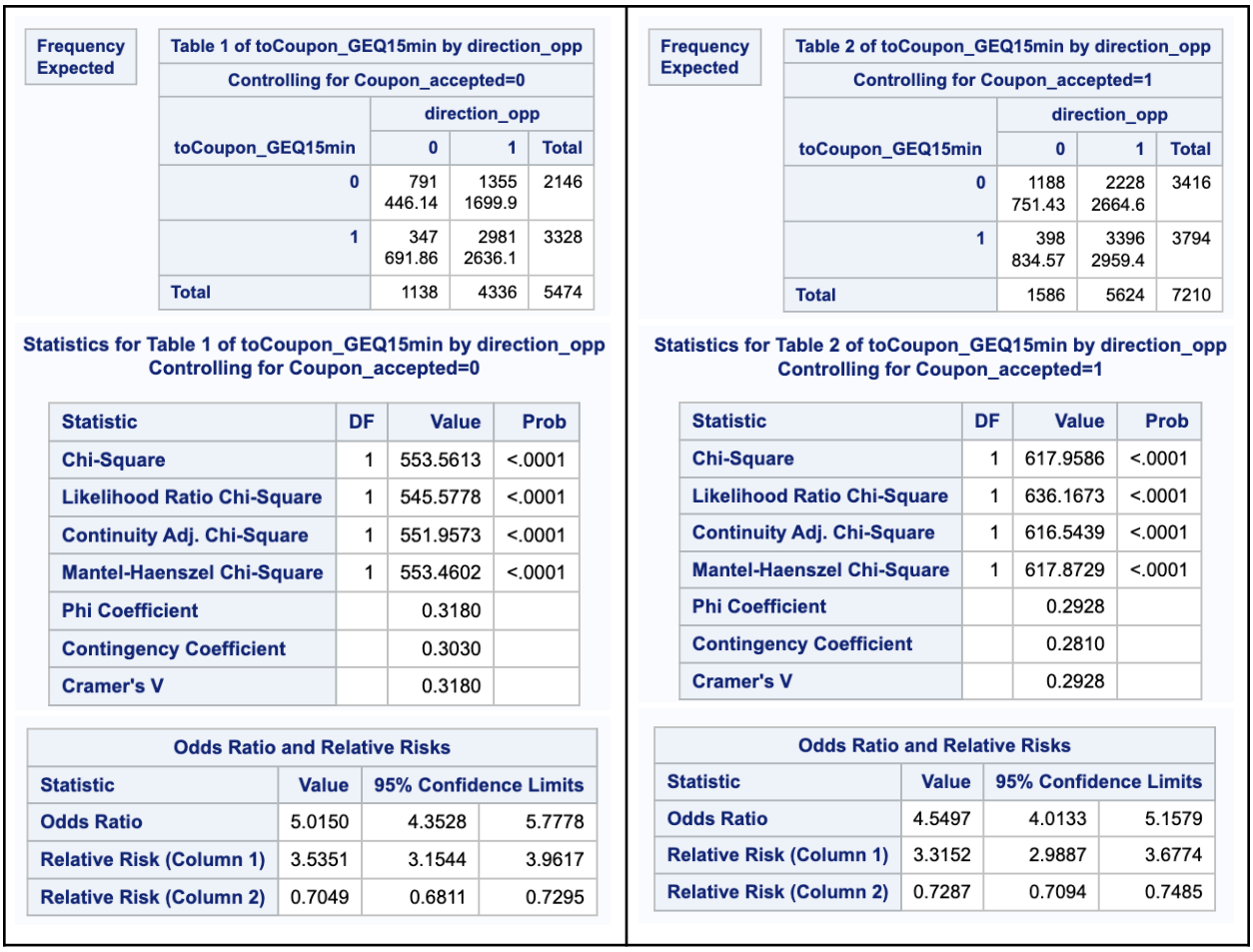


Table 1 of toCoupon_GEQ15min by direction_opp
Controlling for Coupon_accepted=0

Frequency
Expected

toCoupon_GEQ15min	direction_opp 0	direction_opp 1	Total
0	791 446.14	1355 1699.9	2146
1	347 691.86	2981 2636.1	3328
Total	1138	4336	5474

Statistics for Table 1 of toCoupon_GEQ15min by direction_opp
Controlling for Coupon_accepted=0

Statistic	DF	Value	Prob
Chi-Square	1	553.5613	<.0001
Likelihood Ratio Chi-Square	1	545.5778	<.0001
Continuity Adj. Chi-Square	1	551.9573	<.0001
Mantel-Haenszel Chi-Square	1	553.4602	<.0001
Phi Coefficient		0.3180	
Contingency Coefficient		0.3030	
Cramer's V		0.3180	

Odds Ratio and Relative Risks

Statistic	Value	95% Confidence Limits	
Odds Ratio	5.0150	4.3528	5.7778
Relative Risk (Column 1)	3.5351	3.1544	3.9617
Relative Risk (Column 2)	0.7049	0.6811	0.7295

Table 2 of toCoupon_GEQ15min by direction_opp
Controlling for Coupon_accepted=1

Frequency
Expected

toCoupon_GEQ15min	direction_opp 0	direction_opp 1	Total
0	1188 751.43	2228 2664.6	3416
1	398 834.57	3396 2959.4	3794
Total	1586	5624	7210

Statistics for Table 2 of toCoupon_GEQ15min by direction_opp
Controlling for Coupon_accepted=1

Statistic	DF	Value	Prob
Chi-Square	1	617.9586	<.0001
Likelihood Ratio Chi-Square	1	636.1673	<.0001
Continuity Adj. Chi-Square	1	616.5439	<.0001
Mantel-Haenszel Chi-Square	1	617.8729	<.0001
Phi Coefficient		0.2928	
Contingency Coefficient		0.2810	
Cramer's V		0.2928	

Odds Ratio and Relative Risks

Statistic	Value	95% Confidence Limits	
Odds Ratio	4.5497	4.0133	5.1579
Relative Risk (Column 1)	3.3152	2.9887	3.6774
Relative Risk (Column 2)	0.7287	0.7094	0.7485

Figure 5. Multiple Logistic Regression Model 1 - Output A

These statistics suggest that individuals whose direction is opposite to the expected direction have significantly higher odds of taking more than 15 minutes to get to the coupon destination and a significantly higher relative risk compared to those whose direction is not opposite to the expected direction, even when controlling for "Coupon_accepted=1" or "Coupon_accepted=0".

Logistic Model -
Expiration 2h and driving direction is same for driving distance more than 15 minutes
Number of Observations Read 12684
Number of Observations Used 12684

Class Level Information		
Class	Value	Design Variables
expiration	1d	1
	2h	0
direction_same	0	1
	1	0
toCoupon_GEQ15min	0	1
	1	0

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	289.3810	3	<.0001
Score	287.8465	3	<.0001
Wald	283.8486	3	<.0001

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-0.1681	0.0511	10.8314	0.0010
expiration	1d	1	0.5181	0.0365	201.9806	<.0001
direction_same	0	1	0.0237	0.0466	0.2591	0.6107
toCoupon_GEQ15min	0	1	0.3227	0.0385	70.1814	<.0001

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	49.6	Somers' D	0.171
Percent Discordant	32.5	Gamma	0.209
Percent Tied	17.9	Tau-a	0.084
Pairs	39467540	c	0.586

Odds Ratio Estimates and Profile-Likelihood Confidence Intervals				
Effect	Unit	Estimate	95% Confidence Limits	
expiration 1d vs 2h	1.0000	1.679	1.563	1.803
direction_same 0 vs 1	1.0000	1.024	0.935	1.122
toCoupon_GEQ15min 0 vs 1	1.0000	1.381	1.281	1.489

Figure 6. Multiple Logistic Regression Model 1 - Output B

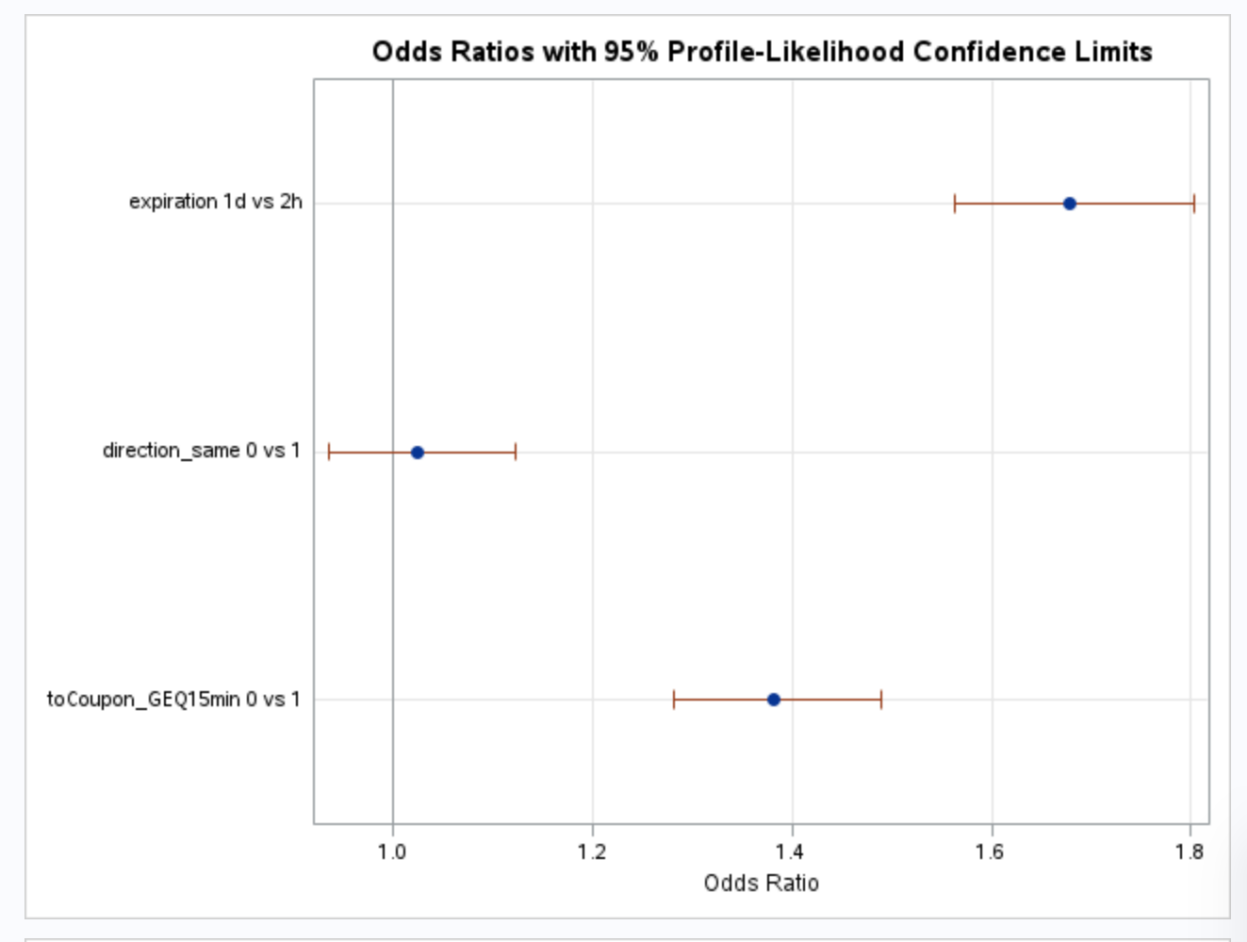


Figure 7. Multiple Logistic Regression Model 1 - Odds Ratio

Expiration (1d vs 2h): The odds of the event (e.g., taking more than 15 minutes to get to the coupon destination) are 1.679 times higher for individuals with an expiration time of 1 day compared to those with an expiration time of 2 hours.

Direction_same (0 vs 1): The odds of the event are 1.024 times higher for individuals whose direction is the same as expected compared to those whose direction is not the same as expected.

toCoupon_GEQ15min (0 vs 1): The odds of the event are 1.381 times higher for individuals who took more than 15 minutes to get to the coupon destination compared to those who did not take more than 15 minutes.

Model 2

- Target variable: Coupon_accepted.

- Predictor variables: coupon, weather * temperature, time * destination.

- Interaction terms: weather * temperature AND time * destination.

- NULL Hypothesis: coupon acceptance with weather * temperature and time * destination events has no association.

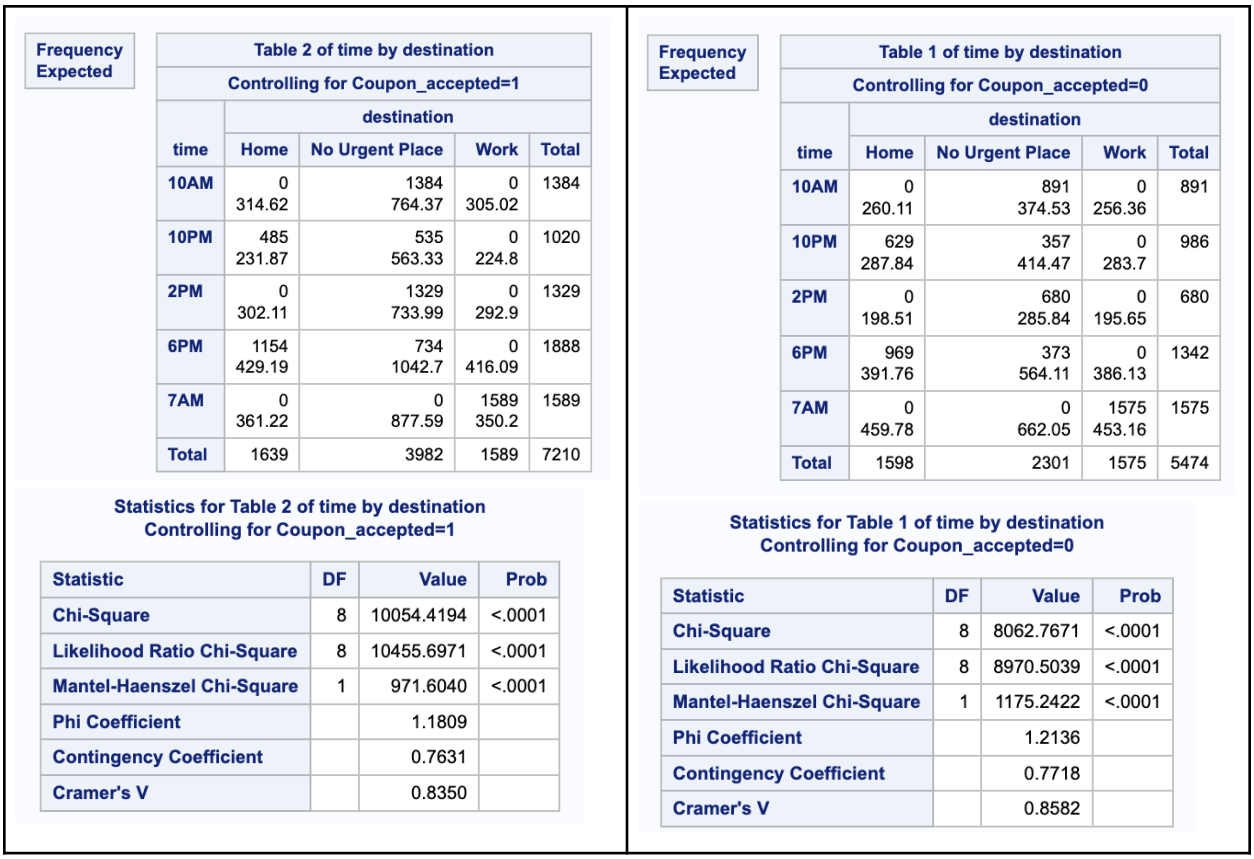
- Alternate Hypothesis: coupon acceptance with weather * temperature and time * destination events has at least one association.

Contingency tables: Coupon acceptance with Destination and Time.

- Association between "time" and "destination" for "Coupon_accepted=0".
    - The obtained chi-square statistic of 8062.7671 with 8 degrees of freedom is highly significant ($p < .0001$), indicating that there is a significant association between the two variables.

    - Similar to the chi-square test, the likelihood ratio chi-square test assesses the significance of the association between "time" and "destination". The obtained likelihood ratio 8970.5039 with 8 degrees of freedom is highly significant ($p < .0001$).

- Association between "time" and "destination" for "Coupon_accepted=1".
    - The obtained chi-square statistic of 10054.4194 with 8 degrees of freedom is highly significant ($p < .0001$), indicating a significant association between the two variables.
    - The obtained likelihood ratio 10455.6971 with 8 degrees of freedom is highly significant ($p < .0001$).



Frequency Expected

Table 2 of time by destination
Controlling for Coupon_accepted=1

time	destination			
	Home	No Urgent Place	Work	Total
10AM	0 314.62	1384 764.37	0 305.02	1384
10PM	485 231.87	535 563.33	0 224.8	1020
2PM	0 302.11	1329 733.99	0 292.9	1329
6PM	1154 429.19	734 1042.7	0 416.09	1888
7AM	0 361.22	0 877.59	1589 350.2	1589
Total	1639	3982	1589	7210

Statistics for Table 2 of time by destination
Controlling for Coupon_accepted=1

Statistic	DF	Value	Prob
Chi-Square	8	10054.4194	<.0001
Likelihood Ratio Chi-Square	8	10455.6971	<.0001
Mantel-Haenszel Chi-Square	1	971.6040	<.0001
Phi Coefficient		1.1809	
Contingency Coefficient		0.7631	
Cramer's V		0.8350	

Frequency Expected

Table 1 of time by destination
Controlling for Coupon_accepted=0

time	destination			
	Home	No Urgent Place	Work	Total
10AM	0 260.11	891 374.53	0 256.36	891
10PM	629 287.84	357 414.47	0 283.7	986
2PM	0 198.51	680 285.84	0 195.65	680
6PM	969 391.76	373 564.11	0 386.13	1342
7AM	0 459.78	0 662.05	1575 453.16	1575
Total	1598	2301	1575	5474

Statistics for Table 1 of time by destination
Controlling for Coupon_accepted=0

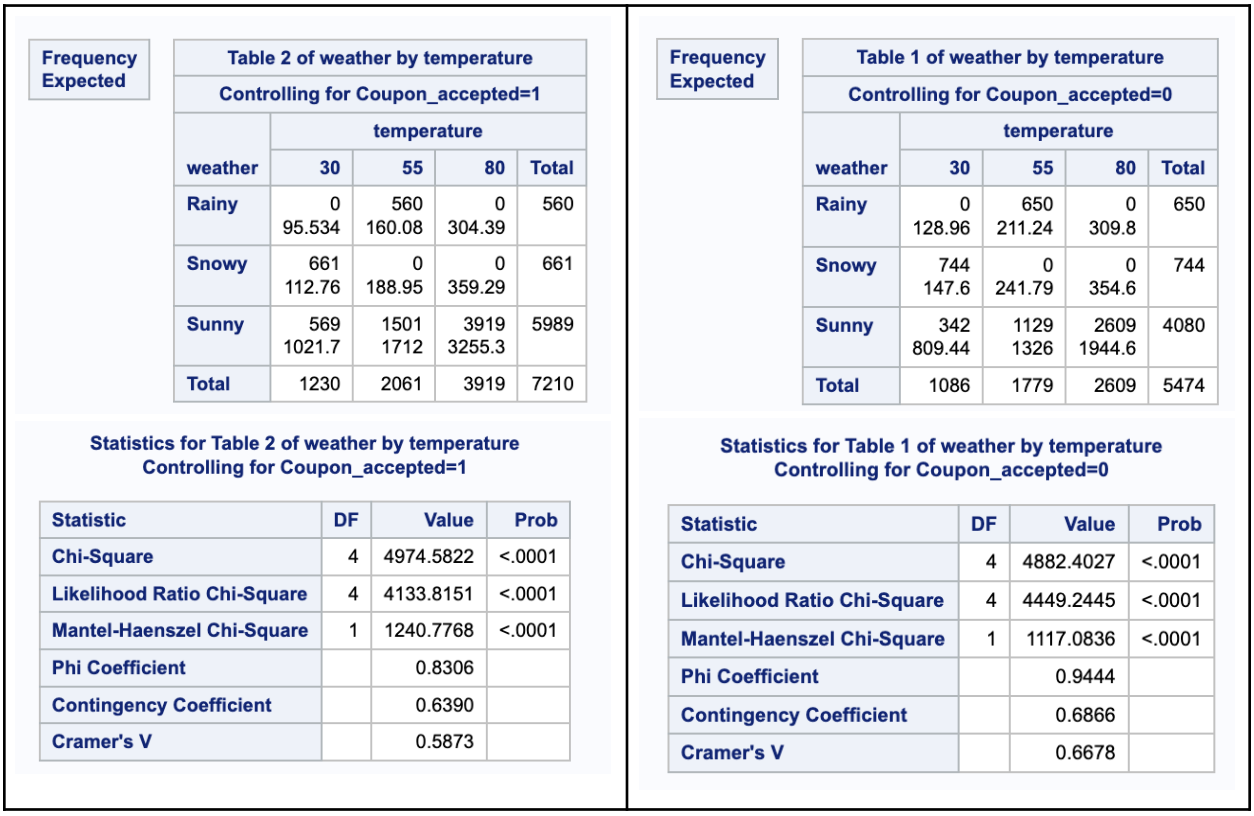
Statistic	DF	Value	Prob
Chi-Square	8	8062.7671	<.0001
Likelihood Ratio Chi-Square	8	8970.5039	<.0001
Mantel-Haenszel Chi-Square	1	1175.2422	<.0001
Phi Coefficient		1.2136	
Contingency Coefficient		0.7718	
Cramer's V		0.8582	

Figure 8. Contingency table Model 2 - Output A

Contingency table: Coupon acceptance with Weather and Temperature.

- Association between "weather" and "temperature" for "Coupon_accepted=0" -
  - The obtained chi-square statistic of 4882.4027 with 4 degrees of freedom is highly significant ($p < .0001$), indicating a significant association between the two variables.
  - The obtained likelihood ratio chi-square statistic of 4449.2445 with 4 degrees of freedom is highly significant ($p < .0001$).

- Association between "weather" and "temperature" for "Coupon_accepted=1" -
  - The obtained chi-square statistic of 4974.5822 with 4 degrees of freedom is highly significant ($p < .0001$), indicating a significant association between the two variables.
  - The obtained Likelihood ratio chi-square statistic of 4133.8151 with 4 degrees of freedom is highly significant ($p < .0001$).



Frequency
Expected

Table 2 of weather by temperature
Controlling for Coupon_accepted=1

weather	temperature 30	55	80	Total
Rainy	0 95.534	560 160.08	0 304.39	560
Snowy	661 112.76	0 188.95	0 359.29	661
Sunny	569 1021.7	1501 1712	3919 3255.3	5989
Total	1230	2061	3919	7210

Statistics for Table 2 of weather by temperature
Controlling for Coupon_accepted=1

Statistic	DF	Value	Prob
Chi-Square	4	4974.5822	<.0001
Likelihood Ratio Chi-Square	4	4133.8151	<.0001
Mantel-Haenszel Chi-Square	1	1240.7768	<.0001
Phi Coefficient		0.8306	
Contingency Coefficient		0.6390	
Cramer's V		0.5873	

Frequency
Expected

Table 1 of weather by temperature
Controlling for Coupon_accepted=0

weather	temperature 30	55	80	Total
Rainy	0 128.96	650 211.24	0 309.8	650
Snowy	744 147.6	0 241.79	0 354.6	744
Sunny	342 809.44	1129 1326	2609 1944.6	4080
Total	1086	1779	2609	5474

Statistics for Table 1 of weather by temperature
Controlling for Coupon_accepted=0

Statistic	DF	Value	Prob
Chi-Square	4	4882.4027	<.0001
Likelihood Ratio Chi-Square	4	4449.2445	<.0001
Mantel-Haenszel Chi-Square	1	1117.0836	<.0001
Phi Coefficient		0.9444	
Contingency Coefficient		0.6866	
Cramer's V		0.6678	

Figure 9. Contingency table Model 2 - Output B

- Coupon_accepted=0:
  - Chi-Square: 4882.4027
  - Likelihood Ratio Chi-Square: 4449.2445
  - Mantel-Haenszel Chi-Square: 1117.0836
- Coupon_accepted=1:
  - Chi-Square: 4974.5822
  - Likelihood Ratio Chi-Square: 4133.8151
  - Mantel-Haenszel Chi-Square: 1240.7768

Overall, both tables show a significant association between "weather" and "temperature", regardless of the value of "Coupon_accepted". This suggests that the relationship between "weather" and "temperature" remains consistent across different conditions of "Coupon_accepted".

Logistic Model -

Number of Observations Read	12684
Number of Observations Used	12684

Class Level Information					
Class	Value	Design Variables			
coupon	Bar	1	0	0	0
	Carry out & Take away	0	1	0	0
	Coffee House	0	0	1	0
	Restaurant(20-50)	0	0	0	0
	Restaurant(<20)	0	0	0	1
weather	Rainy	0	0		
	Snowy	1	0		
	Sunny	0	1		
temperature	30	1	0		
	55	0	0		
	80	0	1		
time	10AM	1	0	0	0
	10PM	0	1	0	0
	2PM	0	0	1	0
	6PM	0	0	0	0
	7AM	0	0	0	1
destination	Home	0	0		
	No Urgent Place	1	0		
	Work	0	1		

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	1131.7564	11	<.0001
Score	1095.7010	11	<.0001
Wald	1030.6205	11	<.0001

Figure 10. Multiple Logistic Regression Model 2 - Output A

These tests collectively provide evidence against the global null hypothesis (BETA=0), suggesting that the model with predictors significantly explains the variability in the outcome.

Analysis of Maximum Likelihood Estimates							
Parameter			DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept			1	-0.2754	0.0663	17.2332	<.0001
coupon	Bar		1	-0.1367	0.0715	3.6565	0.0559
coupon	Carry out & Take away		1	1.2888	0.0716	323.7433	<.0001
coupon	Coffee House		1	0.1410	0.0626	5.0734	0.0243
coupon	Restaurant(<20)		1	1.0704	0.0677	249.6465	<.0001
weather*temperature	Snowy	30	1	-0.4657	0.0664	49.1252	<.0001
weather*temperature	Snowy	80	0	0	.	.	.
weather*temperature	Sunny	30	1	-0.0601	0.0841	0.5110	0.4747
weather*temperature	Sunny	80	1	0.1060	0.0447	5.6341	0.0176
time*destination	10AM	No Urgent Place	1	0.2505	0.0557	20.2531	<.0001
time*destination	10AM	Work	0	0	.	.	.
time*destination	10PM	No Urgent Place	1	0.2548	0.0790	10.3870	0.0013
time*destination	10PM	Work	0	0	.	.	.
time*destination	2PM	No Urgent Place	1	0.4288	0.0591	52.6977	<.0001
time*destination	2PM	Work	0	0	.	.	.
time*destination	7AM	No Urgent Place	0	0	.	.	.
time*destination	7AM	Work	1	-0.2018	0.0497	16.4696	<.0001

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	65.7	Somers' D	0.338
Percent Discordant	31.9	Gamma	0.346
Percent Tied	2.4	Tau-a	0.166
Pairs	39467540	c	0.669

Odds Ratio Estimates and Profile-Likelihood Confidence Intervals				
Effect	Unit	Estimate	95% Confidence Limits	
coupon Bar vs Restaurant(20-50)	1.0000	0.872	0.758	1.003
coupon Carry out & Take away vs Restaurant(20-50)	1.0000	3.628	3.154	4.177
coupon Coffee House vs Restaurant(20-50)	1.0000	1.151	1.019	1.302
coupon Restaurant(<20) vs Restaurant(20-50)	1.0000	2.916	2.555	3.332

Figure 11. Multiple Logistic Regression Model 2 - Output B

Parameter Estimates:

- Intercept: This represents the baseline value when all other predictors are zero. The estimate is -0.2754 with a standard error of 0.0663.
- Coupon: Estimates for different levels of the "coupon" variable are provided. Each estimate represents the effect of that specific coupon type compared to the baseline (reference) category. For example:

    - For the "Carry out & Take away" coupon, the estimate is 1.2888 with a standard error of 0.0716.
    - For the "Bar" coupon, the estimate is -0.1367 with a standard error of 0.0715. .
- Weather*Temperature Interaction: Estimates for different combinations of "weather" and "temperature" are provided. This represents the interaction effect between these two variables. For example:
    - For the combination of "Snowy" weather and temperature "30", the estimate is -0.4657 with a standard error of 0.0664.
    - For the combination of "Sunny" weather and temperature "80", the estimate is 0.1060 with a standard error of 0.0447.
- Time*Destination Interaction: Estimates for different combinations of "time" and "destination" are provided. This represents the interaction effect between these two variables. For example:
    - For the combination of "10AM" time and "No Urgent Place" destination, the estimate is 0.2505 with a standard error of 0.0557.
    - For the combination of "7AM" time and "Work" destination, the estimate is -0.2018 with a standard error of 0.0497.

These estimates provide insights into the direction and strength of the relationships between predictors and the outcome variable. Additionally, the standard errors and significance tests help assess the reliability and significance of these estimates.

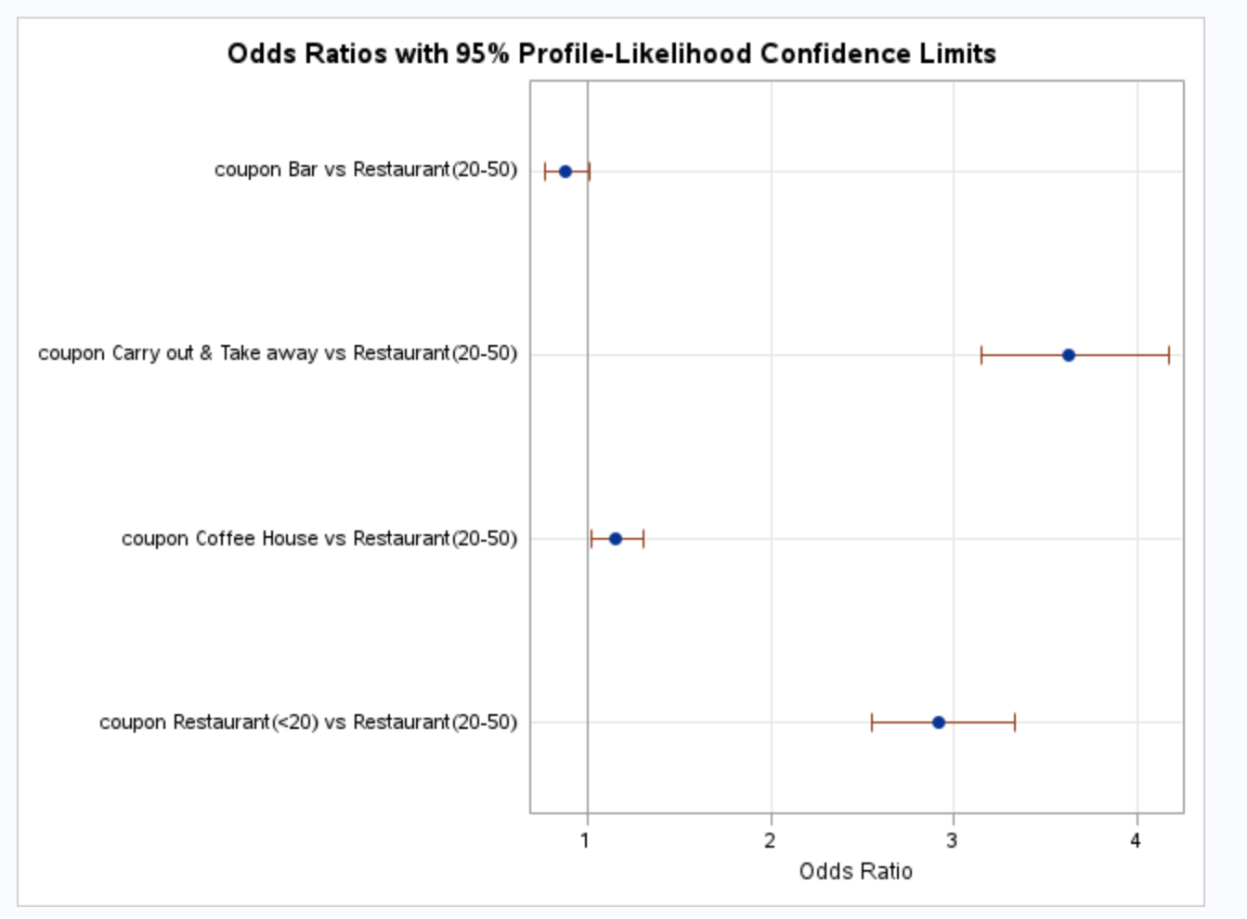


Figure 12. Multiple Logistic Regression Model 2 - Odds Ratio

These odds ratio estimates and profile-likelihood confidence intervals provide information about the relative odds of coupon types compared to the reference category "Restaurant(20-50)"

Odds Ratio Estimates:

- Coupon Bar vs Restaurant(20-50):
  - Estimate: 0.872
  - 95% Confidence Limits: (0.758, 1.003)
  - Interpretation: The odds of choosing the "Bar" coupon relative to the "Restaurant(20-50)" coupon are estimated to be 0.872 times the odds of choosing the "Restaurant(20-50)" coupon.
- Coupon Carry out & Take away vs Restaurant(20-50):
  - Estimate: 3.628
  - 95% Confidence Limits: (3.154, 4.177)

    - Interpretation: The odds of choosing the "Carry out & Take away" coupon relative to the "Restaurant(20-50)" coupon are estimated to be 3.628 times the odds of choosing the "Restaurant(20-50)" coupon..
- Coupon Coffee House vs Restaurant(20-50):
    - Estimate: 1.151
    - 95% Confidence Limits: (1.019, 1.302)
    - Interpretation: The odds of choosing the "Coffee House" coupon relative to the "Restaurant(20-50)" coupon are estimated to be 1.151 times the odds of choosing the "Restaurant(20-50)" coupon.
- Coupon Restaurant(<20) vs Restaurant(20-50):
    - Estimate: 2.916
    - 95% Confidence Limits: (2.555, 3.332)
    - Interpretation: The odds of choosing the "Restaurant(<20)" coupon relative to the "Restaurant(20-50)" coupon are estimated to be 2.916 times the odds of choosing the "Restaurant(20-50)" coupon..

Interpretation:

- The odds ratio estimates provide insights into how the odds of choosing different coupon types compare to the reference category "Restaurant(20-50)".
- Coupon types "Carry out & Take away", "Coffee House", and "Restaurant(<20)" all show significantly different odds compared to "Restaurant(20-50)", suggesting that they have a significant impact on coupon choice relative to the reference category.

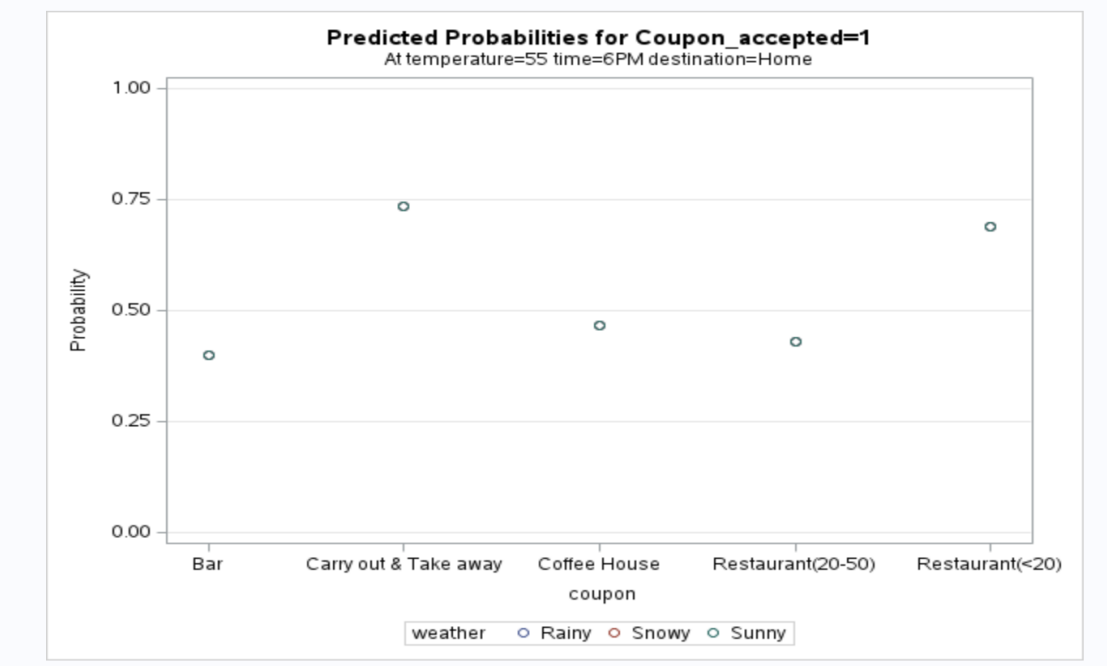


Figure 13. Multiple Logistic Regression Model 2 - Predicted Probabilities

Conclusion

The comprehensive statistical analysis conducted has yielded valuable insights, aligning with the defined goal of assisting the marketing firm's client in targeting consumers for their coupons, particularly those with higher incomes. Additionally, it has identified variables that significantly describe and predict the likelihood of a coupon being accepted, encompassing demographic, driving contexts, situational dynamics, and coupon attributes.

The tests collectively provide evidence against the global null hypothesis (BETA=0), indicating that the model with predictors significantly explains the variability in the outcome. This underscores the statistical significance of the predictors included in the model, affirming their association and efficacy in predicting the outcome variable.

As anticipated, the analysis revealed that, in general, men tend to earn higher incomes compared to women, and employed individuals earn more than retired individuals, students, and the unemployed. However, it was intriguing to find that the difference in income between employed men and women was only 3%, suggesting that gender disparities in income may be narrowing due to initiatives promoting equality in the workplace.

Contrary to expectations, age was found to be a challenging variable to analyze, with limited interpretive value in determining an individual's income. Instead, marital status and whether an individual lives alone emerged as significant factors influencing income levels. Individuals living alone were found to have lower income ranges compared to those living with a spouse/partner or children, underscoring the impact of various motivations and responsibilities on income pursuit.

Considering the first goal of this analysis was to identify factors that would help the client target consumers with higher incomes to market their products to, analysis findings show the majority of higher income earners will most likely be individuals with higher education degrees, male, and may potentially be living alone. While age in itself was a difficult variable to analyze, it should not be discounted completely because certain age demographics in specific situations have the potential to respond more positively to certain coupons.

Analyzing the impact of predictor variables on the likelihood of coupon acceptance revealed notable findings. In the first model, coupons with a 1-day expiration had higher odds compared to those with a 2-hour expiration, suggesting the influence of allowing consumers some thinking time to consider the offer. Coupons with the same directional requirement as the individual had higher odds compared to those with the opposite

direction requirement, indicating the influence of convenience. Additionally, establishments requiring more than 15 minutes to reach the coupon destination had higher odds compared to those with less than a 15-minute requirement.

The second model explored the impact of situational dynamics as well as coupon characteristics on the customer's decision to accept or reject a coupon. Analysis of the maximum likelihood estimates showed that when it came to situational dynamics, poor weather such as snow or low temperatures had a negative association with coupon acceptance rate. The only positive interaction in that scenario was when the temperature was higher and the weather was sunny was when the consumer was more likely to accept an offered coupon. When evaluating time vs destination, consumers were not interested in accepting any coupons when driving to work. However, when offered a coupon during a non-urgent drive, they were more likely to accept it especially during the afternoon, compared to early morning or late at night.

In analyzing the coupon type predictor variable, the interpretation of odds ratio estimates revealed significant differences among coupon types "Bar," "Carry out & Take away," "Coffee House," and "Restaurant(<20)" when compared to "Restaurant(20-50)." Nearly all options examined, where the dining locations' prices were below those of a restaurant priced between $20 and $50, demonstrated higher odds of acceptance. The only exception to this trend was the "Bar" category, which can be attributed to the inherently higher prices often associated with cocktails, resulting in bar tabs that are comparable to or even exceed the cost of dining at a restaurant priced between $20 and $50. This observation suggests that these coupon types exert a notable influence on coupon choice relative to the reference category.

As the second goal of this analysis was to determine what consumer variables have an impact on whether a consumer would accept a coupon, considering the analysis and findings, both situational dynamics and coupon types play different roles in determining coupon acceptance. For situational analysis, a non-urgent driving destination and time of day plays a major role in the consumer responding positively. When evaluating a type of coupon, consumer appears to respond more favorably when the coupon destination costs low as Carry out & Takeaway as well as Restaurants <$20 all showed higher odds compared to other coupon types.

Overall, recommendation to the marketing team is to consider the education and income demographics of their target market. Furthermore, situational dynamics (driving destination and time of day), type of coupon, and more importantly the cost associated with the location highly impact consumer behavior, and the marketing team should approach the target market accordingly.