Practical Application 1: Will the Customer Accept the Coupon?

Analysis produced by Delil Martinez as part of the Professional Certificate in Machine Learning and Artificial Intelligence

Link to GitHub repository: <https://github.com/delilx/Practical-Application-Mod-5>

Part I: Preliminaries

The csv file is read as indicated in the **prompt #1** of the assignment.

The set is said to come from UCI’s Machine Learning Data Repository (website\*). It consists of a dataframe with 12,684 rows and 26 columns. Worthy of note is the fact that the second column is labeled as “passanger”, which I quickly renamed to fix the typo:



Next: investigate the dataset for missing or problematic data (**prompt #2**):

A quick examination of the original UCI posting vs the information provided as part of the starter folder for this application suggests that the dataframe contains one additional variable denoted by ‘cars’, which has no added description or information. One may presume that it contains information on the model and make of the users’ vehicles. However, a basic inspection of the set to detect missing values reveals that the column corresponding to this variable contains 12,576 missing values, or 99.148534 % of the 12684 rows in the set.

Sixty-two of these 108 observations (57.4074%) accepted the coupon. The following table summarizes this classification, along with the responses for the variable ‘car’:

|  |  |
| --- | --- |
| Frequencies  A screenshot of a computer  Description automatically generated | Relative frequencies |

Deciding what to do about missing data (**prompt #3**):

While some diversity in the behavior of the different ‘car’ responses in terms of accepting the coupon is evident, with less than 1% of the information for this variable available, we cannot expect any solid insights from it. Hence, the variable ‘car’ will be dropped from the set used for my analysis.

Once the ‘car’ variable is removed, there are still some missing values, all of which correspond to missing responses to questions about how many times per month a user:

* goes to bars (variable ‘Bar’),
* goes to cafes/coffee houses (variable ‘CoffeeHouse’),
* buys takeaway food (variable ‘CarryAway’),
* eats at a restaurant with average expense less than $20 per person (variable ‘RestaurantLessThan20’), or
* eats at a restaurant with average expense between $20 and $50 per person (variable ‘Restaurant20to50’)

I see a couple of possibilities for handling these missing values:

1. Since the variables involved are binary (yes = 1/ no = 0), it does not make sense to consider mean or median imputation, as these summaries only make sense for quantitative variables. Instead, we can establish an assumption about the reason behind a missing value in one (or more) of these variables. For instance, assume that when a user skips entering a response it is because the answer is “never” (but user does not take the time to explore the options, and simply skips the question). Under this assumption, all missing values should be substituted by the response “never”.
2. A variation of this simplistic approach could be to fill the missing values with other possible responses based on the information from other columns. This would require additional analysis to establish a rule for determining the appropriate value to use for the imputation of each missing entry.
3. Dropping the observations with missing values. This is certainly the simplest approach, and it would only reduce the data frame by a total of 605 rows, leaving 12,079 observations, which is 95.23% of the original number of observations.

For this project, I will take the third approach: dropping all the rows that have at least one missing value. This produces a ‘cleandata’ set that contains 12,079 observations.

Overall coupon acceptance rate (prompt #4):

Once the dataset is cleared of all missing values, the overall coupon acceptance rate is found to be **56.9335%**.

A screenshot of a graph

Description automatically generated

Different types of coupons (**prompt # 5**):

While the most commonly coupon category is Coffee House, with a share of 31.592% of all coupons, the coupons that are accepted with the highest rate are those for “Carry out & Take away” (73.7719% acceptance rate), closely followed by coupons for inexpensive restaurants (those with average expense below $20 per person).

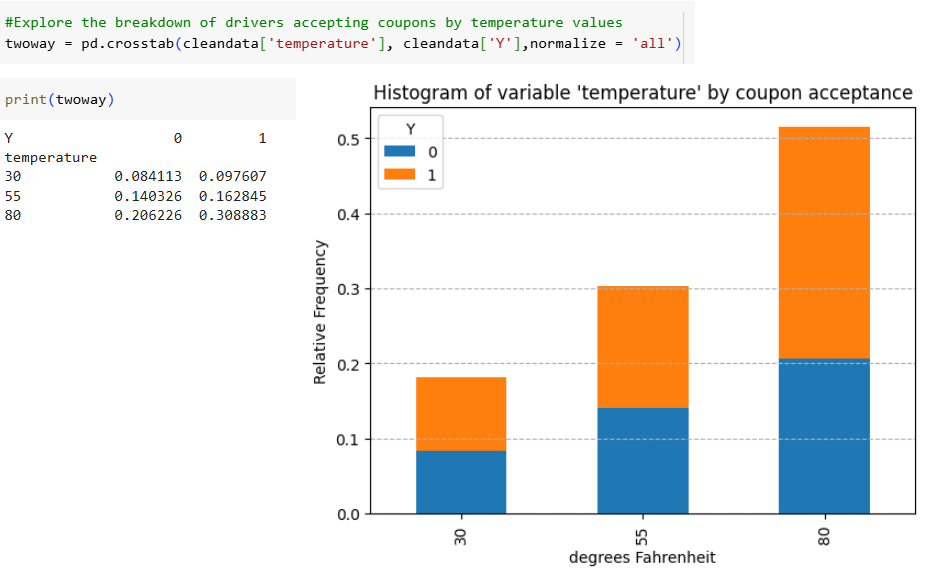
A graph of a bar graph

Description automatically generatedA graph of different types of coupons

Description automatically generated

Quantitative Variable: Temperature (**prompt #6**)

The values for this variable show the following distribution, also broken down by coupon acceptance:



Part II: Investigating the bar coupons

The reduced dataframe “barcoupons” contains 1,913 observations, corresponding to users who received a coupon for a bar. Overall, **41.1918**% of these coupons were accepted by users.

Now we can examine potential differences in the bar coupon success (acceptance) rate, depending on specific characteristics of the users.

An initial breakdown that makes a huge difference in the success rate is by the frequency with which the users who got the coupons go to a bar in a month. The results are:

|  |  |
| --- | --- |
| **Times per month the user goes to a bar** | **Coupon acceptance rate** |
| More than 3 | 76.1658% |
| 3 or fewer | 37.2674% |

Next, we look at a combination of variables: comparing users who go to a bar more than once a month and are over the age of 25, versus all the others. A quick note here about my understanding of the categories in the variable ‘Bar’, which represents the frequency with which users go to a bar in a month. The possible categories for this variable are: “never”, “less1”, “1~3”, “4~8”, and “gt8”. For this particular way of splitting the data, I am taking the categories {“1~3”, “4~8”, “gt8”} to correspond to “going more than once a month”, and {“never”, “less1”} to be the complement of that.

|  |  |
| --- | --- |
| **Characteristics of users** | **Coupon acceptance rate** |
| Go to a bar more than once per month AND over the age of 25 | 68.9826% |
| Everyone else | 33.7748% |

More complexity can be added by adding specific characteristics. Next we isolate users who go to bars more than once a month, had passengers that were not a kid, and had occupations other than “Farming, Fishing or Forestry”.

|  |  |
| --- | --- |
| **Characteristics of users** | **Coupon acceptance rate** |
| Go to a bar more than once per month AND have passengers that are not a kid AND have occupations other than farming, fishing, or forestry. | 70.9434% |
| Everyone else | 29.7903% |

Note: the restriction of the user’s occupation to be other than farming, fishing, or forestry does not seem to have any effect on the coupon acceptance rate:

|  |  |
| --- | --- |
| **Characteristics of users** | **Coupon acceptance rate** |
| Go to a bar more than once per month AND have passengers that are not a kid. | 70.9434% |
| Everyone else | 29.7903% |

Now a comparison is requested between three specific combinations of driver characteristics, namely, those who

* go to bars more than once a month, had passengers that were not a kid, and were not widowed OR
* go to bars more than once a month and are under the age of 30 OR
* go to cheap restaurants more than 4 times a month and income is less than 50K.

| **Characteristics of users** | **Coupon acceptance rate** |
| --- | --- |
| Go to bars more than once per month AND have passengers that are not a kid AND are not widowed. | 70.9434% |
| Go to bars more than once a month AND are under the age of 30. | 71.9512% |
| Go to restaurants with entrees that cost less than $20 per person more than 4 times a month AND have an income lower than $50K. | 45.6456% |

**Conclusions:**

The data suggests that the following characteristics are associated with higher likelihood of a coupon for a bar being accepted by a driver:

* Driver goes to bars at least once a month
* Driver is under the age of 30

Characteristics that seem to negatively affect the likelihood of accepting a bar coupon:

* Driver has kids as passengers
* Driver does not have a high income

Finally, characteristics that do not seem to affect the likelihood of accepting a bar coupon(when combined with the characteristics in the first group):

* Whether driver is in farming, fishing or forestry
* Whether driver is widowed

Part III: Independent Investigation

For this part I chose to focus on coupons for coffee houses, as this category represents the highest share of coupons offered (3816 out of 12079, or 31.592%). The objective of the analysis is to shed light on the characteristics that lead to users accepting or rejecting the coupon they receive as they drive.

The overall acceptance rate for coffee house coupons is 49.6331%; this investigation will shed light on specific characteristics that differentiate this rate for different subsets of data.

First, an exploration of the coffee house coupon acceptance rate by the frequency with which drivers tend to go to coffee houses indicates that, nor surprisingly, an existing habit of going to cafes/coffee houses is associated with a higher acceptance rate:

A screenshot of a computer

Description automatically generated

A comparison using filters is as follows:

|  |  |
| --- | --- |
| **Characteristics of users** | **Coupon acceptance rate** |
| Go to a coffee house at least once a month | 65.8994% |
| Go to a coffee house less frequently than once a month | 34.0349% |

Other explorations that resulted in interesting differences in the rates of acceptance of coupons for coffee houses are:

|  |  |
| --- | --- |
| **Characteristics of users** | **Coupon acceptance rate** |
| Go to a coffee house at least once a month AND are driving a peer (partner or friends) | 76.7376% |
| Everyone else | 43.4908% |

Driving alone or with kids (or going to coffee houses less than once a month) results in a lower coupon acceptance rate.

It seemed reasonable to take into account the destination, with “No Urgent Place” being of particular interest, and distance from the coffee house offering the coupon:

|  |  |
| --- | --- |
| **Characteristics of users** | **Coupon acceptance rate** |
| Go to a coffee house at least once a month, are driving a peer (partner or friends), have a “non urgent” destination AND are less than 15 minutes from the coupon’s location | 80.9392% |
| Everyone else (at least one of the conditions above is not met) | 46.35% |

Questions of interest: Is age a factor here? What about the time of day?

|  |  |
| --- | --- |
| **Characteristics of users** | **Coupon acceptance rate** |
| Are 25 years old or younger, go to a coffee house at least once a month, are driving to a “non urgent” destination, in the middle of the day (10 am to 2 pm) AND are less than 15 minutes from the coupon’s location | 87.619% |
| Everyone else (at least one of the conditions above is not met) | 48.558% |

**Findings and Recommendations**

This analysis suggests that drivers who are most likely to accept a coupon for a coffee house:

* are young (25 or younger);
* go to coffee houses at least once a month
* are driving in the middle of the day (10 am to 2 pm)
* are not in a hurry (destination is “Not Urgent Place”)
* are close to the coupon’s location (less than a 15-minute drive), and
* (possibly) are driving with peer companions (a partner or friends)

Coffee houses interested in participating in this marketing campaign through coupons sent in real time to drivers will maximize their return by focusing on these characteristics.

**Other general findings:**

While it is clear that the combination of characteristics associated with higher coupon acceptance rates should be based on the type of coupon, here is a generic breakdown of general tendencies to accept coupons, according to the data included in the set:

* **By gender**: men tend to accept coupons slightly more frequently than women (59.235% vs. 54.751%)
* **By income**: all income categories have coupon acceptance rates roughly between 50% and 60%.
* **By marital status**: single people have a slightly higher coupon acceptance rate (61% vs other categories in the 50% range)
* People with no **children** have a slightly higher coupon acceptance rate over those with children (58.92% vs 54.05%)
* A pie chart with numbers and a triangle

  Description automatically generatedAnecdotally, the accepted coffee house coupons are associated with sunny **weather**: