**Exploratory Analysis of Bar and Coffee House coupons acceptance rate**

**Will a Customer Accept the Coupon?**

**Context**

Imagine driving through town and a coupon is delivered to your cell phone for a restaraunt near where you are driving. Would you accept that coupon and take a short detour to the restaraunt? Would you accept the coupon but use it on a sunbsequent trip? Would you ignore the coupon entirely? What if the coupon was for a bar instead of a restaraunt? What about a coffee house? Would you accept a bar coupon with a minor passenger in the car? What about if it was just you and your partner in the car? Would weather impact the rate of acceptance? What about the time of day?

Obviously, proximity to the business is a factor on whether the coupon is delivered to the driver or not, but what are the factors that determine whether a driver accepts the coupon once it is delivered to them? How would you determine whether a driver is likely to accept a coupon?

**Overview**

The goal of this project is to use what you know about visualizations and probability distributions to distinguish between customers who accepted a driving coupon versus those that did not.

**Data**

This data comes to us from the UCI Machine Learning repository and was collected via a survey on Amazon Mechanical Turk. The survey describes different driving scenarios including the destination, current time, weather, passenger, etc., and then ask the person whether he will accept the coupon if he is the driver. Answers that the user will drive there ‘right away’ or ‘later before the coupon expires’ are labeled as ‘Y = 1’ and answers ‘no, I do not want the coupon’ are labeled as ‘Y = 0’. There are five different types of coupons -- less expensive restaurants (under $20), coffee houses, carry out & take away, bar, and more expensive restaurants ($20 - $50).

**Deliverables**

Your final product should be a brief report that highlights the differences between customers who did and did not accept the coupons. To explore the data you will utilize your knowledge of plotting, statistical summaries, and visualization using Python. You will publish your findings in a public facing github repository as your first portfolio piece.

**Data Description**

Keep in mind that these values mentioned below are average values.

The attributes of this data set include:

**1. User attributes**

- Gender: male, female

- Age: below 21, 21 to 25, 26 to 30, etc.

- Marital Status: single, married partner, unmarried partner, or widowed

- Number of children: 0, 1, or more than 1

- Education: high school, bachelors degree, associates degree, or graduate degree

- Occupation: architecture & engineering, business & financial, etc.

- Annual income: less than \\$12500, \\$12500 - \\$24999, \\$25000 - \\$37499, etc.

- Number of times that he/she goes to a bar: 0, less than 1, 1 to 3, 4 to 8 or greater than 8

- Number of times that he/she buys takeaway food: 0, less than 1, 1 to 3, 4 to 8 or greater

than 8

- Number of times that he/she goes to a coffee house: 0, less than 1, 1 to 3, 4 to 8 or

greater than 8

- Number of times that he/she eats at a restaurant with average expense less than \\$20 per

person: 0, less than 1, 1 to 3, 4 to 8 or greater than 8

- Number of times that he/she goes to a bar: 0, less than 1, 1 to 3, 4 to 8 or greater than 8

**2. Contextual attributes**

- Driving destination: home, work, or no urgent destination

- Location of user, coupon and destination: we provide a map to show the geographical

location of the user, destination, and the venue, and we mark the distance between each

two places with time of driving. The user can see whether the venue is in the same

direction as the destination.

- Weather: sunny, rainy, or snowy

- Temperature: 30F, 55F, or 80F

- Time: 10AM, 2PM, or 6PM

- Passenger: alone, partner, kid(s), or friend(s)

**3. Coupon attributes**

- time before it expires: 2 hours or one day

**Statistical Analysis**

Univariate analysis of each independent variable was carried out using descriptive statistics. Counts and percentages were estimated for categorical variables while mean and standard deviation for continuous variables.

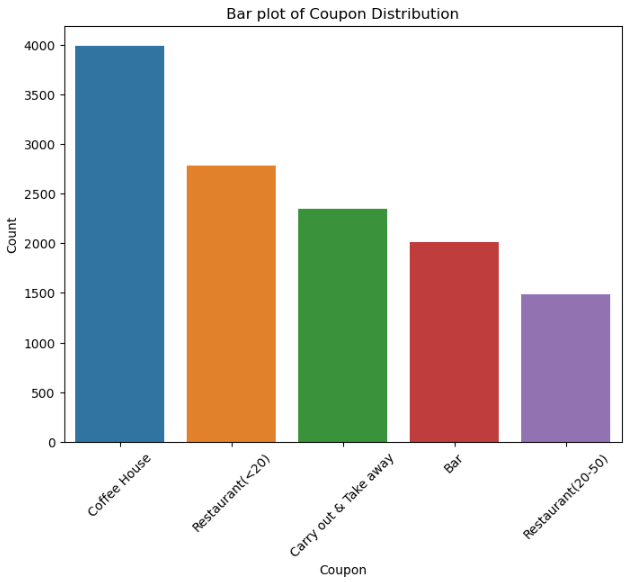
Also, distributions of independent variables were visualized. Selection of higher-level complex interaction patterns for further exploring was carried out using decision trees (Classification and regression tress) with entropy criterion.

Prior to analyzing the data, missingness was explored as well as exclusion of duplicated records.

**Results/findings**

Missing values in the variable car were highly significant hence the variable was excluded from the analysis while missing values in variables; Bar, RestaurantLessThan20, CarryAway, Restaurant20To50, and CoffeeHouse were few. For this analysis, no imputation was carried out hence the missing values (NaN) were replaced with blank.

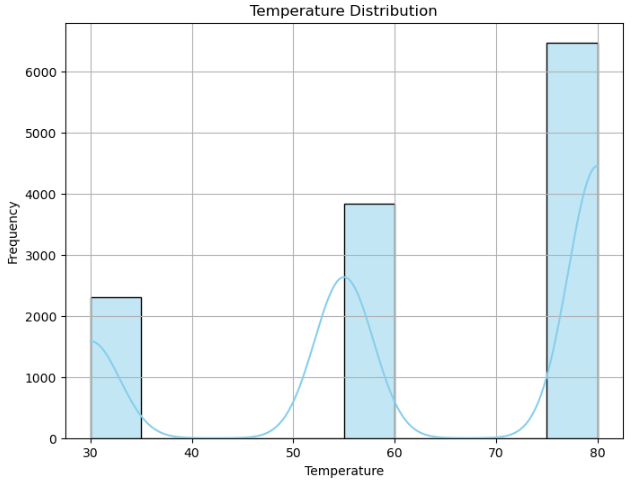
**Distribution of the coupons**



Coupons column is skewed to the right.

From the analysis, it was observed that 56.8% of the respondents accepted the coupon. Acceptance rate for bar coupons was 41% while for coffee house coupons was 49.9%.

**Distribution of temperature**



Temperature is multimodal. It was excluded from the analysis.

**Findings for Bar Coupons**

**Aim 1: Compare the acceptance rate between those who went to a bar 3 or fewer times a month to those who went more.**

***H0: Drivers who go to a bar more than 3 times a month are less likely to accept bar coupons compared to drivers who go to a bar 3 or fewer times a month.***

**Conclusion:** From the analysis it was observed that the acceptance rate among those who went to a bar 3 or fewer times a month was 52.8% compared to 76.9% of those who went more than 3 times.

**Aim 2: Compare the acceptance rate between drivers who go to a bar more than once a month and are over the age of 25 to the all others. Is there a difference?**

***H0: Drivers aged 25 and older who go to a bar more than once a month are less likely to accept bar coupons compared to all others.***

**Conclusion:** The acceptance rate among drivers who went to a bar more than once a month and were over the age of 25 was 69.5% while it was 33.5% among all others.

**Aim 3: Compare the acceptance rate between drivers who go to bars more than once a month and had passengers that were not a kid and had occupations other than farming, fishing, or forestry.**

***H0: Drivers who go to bars more than once a month with passengers other than a kid and are in occupations other than farming, fishing, or forestry are less likely to accept bar coupons compared to all others.***

**Conclusion:** The acceptance rate among drivers who went to bars more than once a month and had passengers that were not a kid and were in occupations other than farming, fishing, or forestry was 71.3% while it was 29.5% among all others.

**Aim 4: Compare the acceptance rates between those drivers 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.**

***H01: Drivers who go to bars more than once a month with passengers other than a kid and are not widowed are less likely to accept bar coupons compared to all others.***

***H02: Drivers who go to bars more than once a month and are under the age of 30 are less likely to accept bar coupons compared to all others.***

***H03: Drivers who go to cheap restaurants more than 4 times a month and earn an income less than 50K are less likely to accept bar coupons compared to all others.***

**Conclusion:** The acceptance rate among those drivers who go to bars more than once a month, had passengers that were not a kid, and were not widowed was 71.3% while for those drivers who go to bars more than once a month and are under the age of 30 was 72.2% and for those drivers who go to cheap restaurants more than 4 times a month and income is less than 50K was 45.4%. The acceptance rate among those drivers who belonged to at least one of the 3 categories above was 58.9%.

**Independent Investigation: Characteristics of passengers/drivers who accepted the Coffee House coupons.**

From the analysis, acceptance rate for coffee house coupons was 49.9%.

Using the decision tree classifier with entropy criterion, the predictor spaces below were selected for further analysis.

1. (driving destination ==No Urgent Place) X (time at or before 6PM) X (expiration == 2h)
2. (driving destination ==No Urgent Place) X (income outside the range "$75000 - $87499") X (occupation not Student)
3. (driving destination ==No Urgent Place) X (occupation == Student)

**Conclusion:** The acceptance rate of Coffee House coupons was highest among students who had no urgent driving destination/place (73.3%) followed by non-students earning outside the range $75000 - $87499 and had no urgent place to be (57.4%). For coffee house coupons expiring within 2 hours, drivers who had no urgent destination to go at or before 6pm accepted them at a rate of 53.3%.

The acceptance rate among drivers who belonged to at least one of the 3 categories above was 58.6%.

**Sensitivity analysis: Incorporating number of times been at a coffee house and gender.**

None of the predictor space in the CART model/decision tree selected gender. The interaction pattern below was further explored.

1. (Ever been at the coffee house) X (expiration > 2h) X (driving destination!=No Urgent destination)

**Conclusion:** The acceptance rate of coffee house coupons expiring within a day among drivers heading to work or home and have ever been to a coffee house was 62.8%.