**Practical Application 1: Will the Customer Accept the Coupon?**

**Data:**

The dataset used in this application (“coupons”) was based on 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).

**Exploratory Data Analysis:**

The goal of performing Exploratory Data Analysis in machine learning is to better understand the data before applying any model. This is a very important step as it allows us to identify patterns, develop hypotheses, take suitable steps to address missing values and understand the relationships between variables. These steps would help to properly prepare the data prior to modeling.

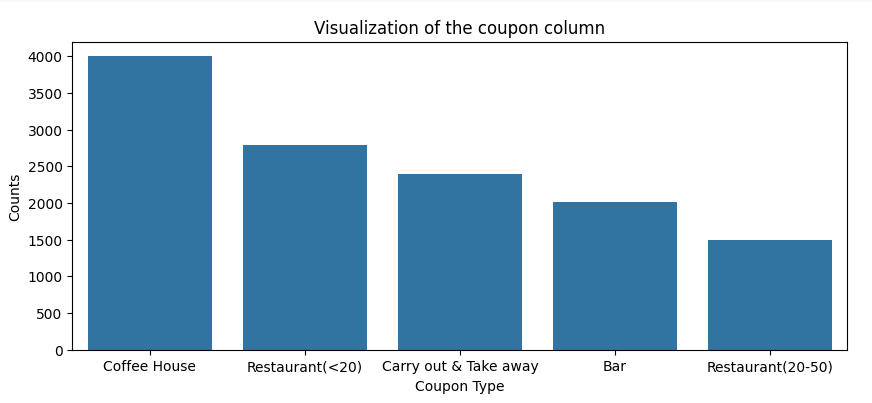
Data info analysis of dataset showed two type of data types, objects and Integers with some missing values. Object data types were transformed into string type prior to data analysis. The column labeled “car” indicated 12576 missing data points (99.15%). Since this is a significant number, it was decided to drop this column to achieve a more complete dataset for better modeling. Other columns that indicated missing values namely “Bar”, CoffeeHouse”, CarryAway”, “RestaurentLessThan20” and “Restaurent20To50” exhibited only about 0.84-1.71% missing values, it was decided to keep these columns for the data analysis.

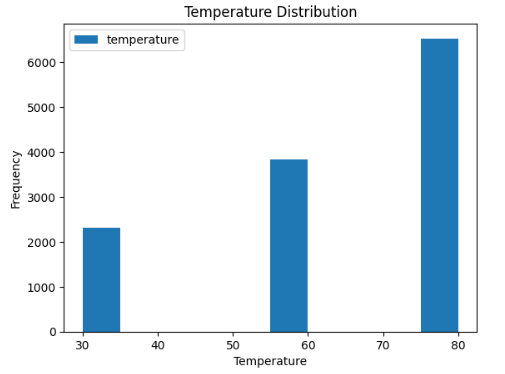
**Proportion of the Total Observations Chose to accept the Coupons**:

The data analysis indicated that the total observations chose to accept the coupon was 57.41%.

**Visualization of the Data on “coupon” Column:**

The following bar plots and histograms were prepared for the visualization of the data. The coupon types, CoffeeHouse and “Restaurent<20”combined together showed the highest count values. The other three coupon types combined appear to show a count value somewhat close to 50%. The highest contextual temperature attribute was 80°F.





**Exploration of “Bar” Related Coupons:**

This was performed by creating a new DataFrame that contains just the “Bar”coupons.

Results indicated that the accepted proportion of car coupon by passengers was 41.00%.

The comparison of the acceptance rate was made between those who went to a bar three times or fewer per month and to those who went more than three times per month.

The total acceptance of those who went to a bar 3 or fewer times per month was 666 while the total for more than three times per month was 153. These results indicated that the acceptance rate between those who went to a bar three or fewer times per month to those who went more than 3 times per month was 81.32%

**Comparison of 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:**

Results calculated showed that the acceptance rate of those who go to a bar more than once a month and are over the age of 25 was 72.80% while the acceptance rate of those who go to a bar more than once a month and are under the age of 25 was 27.80 % indicating a significant different between the two groups.

The same process above was used to 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. The results indicated that a 72.32% belongs to this group leaving 28.68% for the passengers who did not accept the coupons.

**Comparison of 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 ORgo to cheap restaurants more than 4 times a month and income is less than 50K.**

The Comparison of acceptance rates are shown in the Table below.

|  |  |
| --- | --- |
|  | Acceptance Rate |
| Drivers those who go to bars more than once A month, had passengers that were not a kid, and were not widowed | 71.32% |
| Drivers those who go to bars more than once a month and are under the age of 30 | 72.17% |
| Drivers those who go to cheap restaurants more than 4 times a month and income is less than 50K | 60.07% |

Since the last comparison is not based on the “Bar” column, one should use the original dataset, not the new one that contains just the bar coupons.

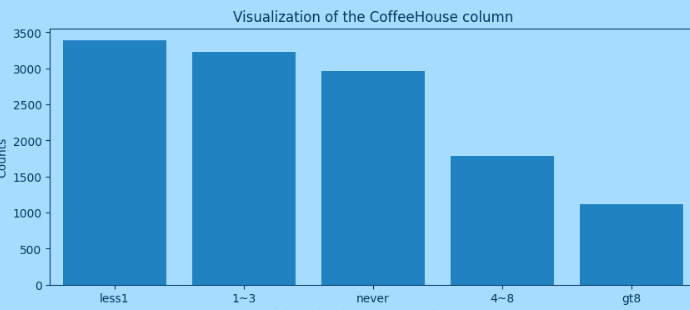
Based on these observations, one can hypothesize that drivers who accepted the car coupons were those who do not have kids as passengers and were under the age of 30 and were not widowed. A significant number of those drivers were earning lower income of less than 50K per year.

**Independent Investigation**

Using the bar coupon example as motivation, the CoffeeHouse coupon group was also evaluated to determine the characteristics of passengers who accept the coupons.

The results indicated that the acceptance rate of coffee coupons as 49.92% , a slightly higher percentage than bar coupons (41.00%).

The visualization graph shows that the most number of times one goes to a coffee house were less than one time or 1-3 times per month. Higher count value (around 3000 counts) were also shown for zero visits



Graph below shows the redemption location/destination of coffee coupons. Most of the drivers have redeemed their coupons at “No Urgent Place”.

The visualization graph shows that the most number of times one goes to a coffee house were less than one time or 1-3 times per month. Higher count value (around 3000 counts) were also shown for zero visits

