**UCB Module 5**

**Practical Problem Write Up**

**Section 1 – Guided Analysis on Bar Coupon**

Initial Data Analysis and Observation

This analysis started with general inspection and investigation of the data within the given data set. As part of this analysis, I considered the completeness and quality of the various fields. It became obvious that there were some data fields that warranted some initial clean up and/or modification:

* Age – this field was not a numeric and it had some values that would not allow for conversion to numeric (i.e. 50plus, below21). For this field, it made sense to replace those two values with numeric values of 50 and 20. Given we didn’t know the specific value below 21 or above 50, setting them to a single value lost no data meaning and then allowed overall conversion to numeric.
* Bar, CoffeeHouse, CarryAway, RestaurantLessThan20, Restaurant20to50 – these fields were all assessed to have null values, but the volume of those values did not seem of a magnitude to warrant updating – and there was not a good way to “predict” what they should have been
* Car – this field had a relatively high value of nulls, but is not planned for use in analysis – if that changes, I will revisit.
* Bar – this field, as well as some others captures “number of visits” but does so in a text field which is not easily operated on as a numeric. Given the low number of possibilities, I made the decision to operate on these as text via “.isin” function.

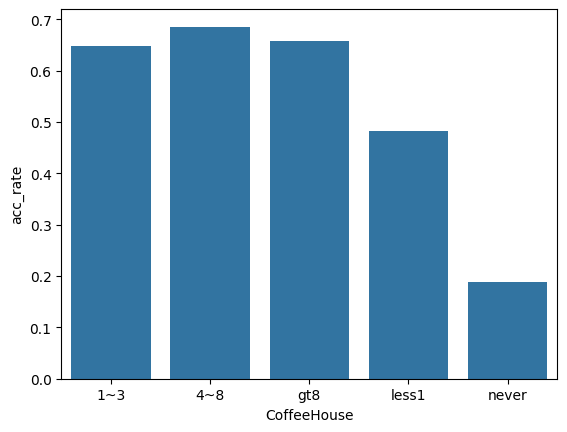
Hypothesis on Bar Coupons

Section 2 – Independent Investigation

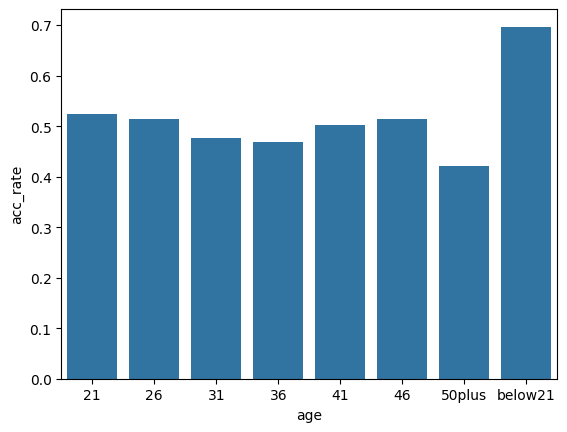
As part of my independent investigation, I chose to focus on the Coffee House coupons. I first looked at the overall acceptance rate of these coupons and found it to be 49.9%. This is slightly less than the overall rate of acceptance of all coupons which is 56.8%. In terms of structuring my analysis, I am going to assess the following:

* Visits to Coffee Houses – I would expect this to be a strong driver of the acceptance rate just based on utility of a coupon being offered.
* Age/income/education – these are dimensions that I don’t necessarily expect to have a strong correlation, but that I do think should be considered.
* From a utility/convenience standpoint, I will examine the “direction” aspect of the data to see if there is a convenience factor involved in acceptance

When considering the customers frequency of visits to coffee shops in relation to their acceptance rate, the acceptance rate increased to 67.5% for those visiting more than 4 times per month. This increase is notable, but not as large as might have been expected. See image below.



Further analysis by ages does show an elevated acceptance being driven by those 20 and under.



When considering education level, the data would indicate that this might not be as valuable a condition to consider given that 83%+ of the population has more than a high school or associate degree. Given this skew, the differentiation available is likely not as valuable. This distribution is shown below.

A screen shot of a computer

Description automatically generated

In looking at income as a factor in acceptance rates, the data is very inconsistent – with a range of rates that show low correlation to income levels. For example, the lowest acceptance rate (29.7%) occurs at an income range of $75,000-$87,499, yet that acceptance rate jumps to 55.7% at the next income range of $87,500-$99,999.

Lastly, I considered the proximity dimension looking at data related to the travel direction of the target customer. Working under the assumption that a value of “1” for “direction\_same” indicates that the customer is travelling in the direction of the target business. I found little to no correlation in this data – with a difference of only 4% between the acceptance rates – and rates that are very similar to the overall acceptance rate for all coffee coupons.

In summary, it would be my recommendation that this coupon program NOT be considered for a host of reasons. First, the majority of data gathered does NOT create ideal data for targeting customers. This is evidenced by the analysis above. Second, in a blind targeting, much of this data would be very challenging to have for a customer on a near real-time basis. The data that would likely be available (i.e. proximity to and direction from the business) are weak at best for prediction of acceptance.